

**École Nationale Supérieure des Télécommunications**  
Groupe des Ecoles de Télécommunications – ParisTech  
46 rue Barrault – 75634 Paris Cedex 13  
Tél. 01 45 81 77 77 – Fax 01 45 89 79 06 – [www.enst.fr](http://www.enst.fr)

**ENST 2005-S-001**

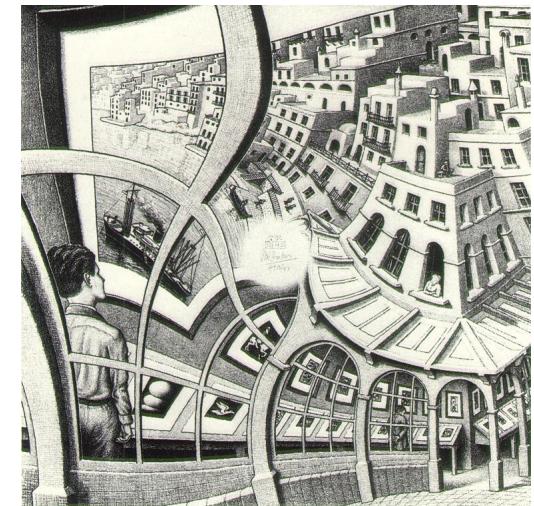
**Réflexivité et auto-référence dans les  
systèmes complexes**

**Journées de Rochebrune**



## **Réflexivité et auto-référence dans les systèmes complexes**

**12<sup>èmes</sup>  
journées  
de Rochebrune**



**Rencontres interdisciplinaires  
sur les systèmes complexes naturels et artificiels**

**du 24 au 28 janvier 2005**

Organisées avec le soutien  
de l'European Conference on Artificial Life (ECAL)  
de la Délégation Générale pour l'Armement (DGA)

**ENST 2005-S-001**

Ecole Nationale Supérieure des Télécommunications



## **12es Journées de Rochebrune**

**Rencontres interdisciplinaires sur les systèmes complexes naturels et artificiels**

**Du 24 Janvier au 28 Janvier 2005, Rochebrune, Megève, France**

# **Réflexivité et auto-référence dans les systèmes complexes**

Organisées avec le soutien  
de l'European Conference on Artificial Life (ECAL)  
de la Délégation Générale pour l'Armement (DGA)

**Mots clés** : réflexivité, auto-référence, mise en abyme, miroir, auto-observation, circularité, auto-similarité, récursivité, paradoxe

Textes réunis par **Serge Stinckwich**

**ENST 2005-S-001**

**École Nationale Supérieure des Télécommunications**



L'auto-référence caractérise un système ou un phénomène où il est fait référence à lui-même ou à une partie de lui-même. C'est un mécanisme que l'on retrouve dans de très nombreuses disciplines et l'on peut dire qu'il s'agit d'un phénomène transdisciplinaire par essence. Son étude et la compréhension des mécanismes nécessaire à sa mise en œuvre sont donc essentiels pour tous ceux qui s'intéressent à la notion de systèmes complexes. Son usage et son intérêt sont multiples et variés suivant le champ disciplinaire : jeu artistique de la mise en abyme dans la littérature et le cinéma, références à l'histoire de la peinture dans la peinture, outil de définition des objets conceptuels en informatique par la récursivité, mécanisme de simulation dans le droit, outil de compréhension et d'appréhension de la complexité enfin en général. Il est parfois impossible de ne pas l'appréhender dans certaines disciplines : est-il possible d'exprimer les concepts de la linguistique autrement que par l'objet même qu'elle étudie (i.e. la langue). Plus généralement dans les sciences humaines, l'observateur et l'observé sont en rapport de circularité. Son utilisation est parfois sujet à caution et discussions (par exemple : les paradoxes langagiers ou logiques) et peut également provoquer de grave crises épistémiques ou paradoxales en science (comme ce fut le cas en mathématique lors de la définition des ensembles). Son usage est donc à la fois simple dans sa mise en place et complexe dans la compréhension de ces conséquences.

## Différentes formes d'auto-références

- La récursivité est une première forme simple de l'auto-référence que l'on retrouve fréquemment : histoire à l'intérieur d'une histoire, tableau dans un tableau, ... Elle consiste à définir un concept à partir d'un concept plus simple, écartant par la même toute forme de paradoxe. La récursion est au cœur même de la matière : en mécanique quantique, les particules sont emboitées les unes dans les autres de façon récursive. Il est alors difficile de parler d'une particule sans parler d'une autre. C'est la problématique de renormalisation bien connu en physique.

- La circularité est une forme d'auto-référence, où la régression à l'infini n'est pas arrêtée. Elle peut être à l'origine de nombreux paradoxes comme celui du célèbre paradoxe du menteur, attribué au philosophe grec du IV<sup>e</sup> siècle avant JC, Eubulide.

- La réflexivité est un mécanisme de plus haut niveau puisque qu'un tel système a la capacité de s'auto-observer (i.e raisonner sur soi-même) et de s'auto-modifier (i.e. altérer sa structure (par exemple pour un programme) et altérer son comportement).

- L'auto-similarité désigne un phénomène ou une structure qui se répète à plusieurs échelles ou niveaux différents. C'est souvent une conséquence d'une définition ou d'un processus récursif.

## L'auto-référence dans différents champs disciplinaires

- Philosophie des sciences : L'approche réflexive est très générale en science où les méthodes et les modèles sont en permanence évalués. Ses enjeux sont de

différentes natures, avec des critères classiques : répétabilité, capacité à expliquer ou prédire des données, cohérence axiomatique et logique, réfutabilité, ... mais aussi des critères beaucoup plus opérationnels liés à l'utilisation des méthodes, modèles ou résultats, à leur efficacité en regard d'objectifs, à leur acceptabilité, voire à des critères d'ordre économique ou social, correspondant à une notion de retour sur investissement ou de perspective d'accroissement du bien-être. Quels sont les questions de déontologie, d'éthique qui peuvent alors se poser ?

- Philosophie : Certains concepts comme l'être, la réalité, l'identité ou encore l'existence ont un fort caractère auto-référentiel. Ces concepts se renvoient-ils à eux-mêmes par un jeu de miroir ontologique ? Les philosophes également logiciens ont mis en lumière les premiers paradoxes auto-référentiels. Quels sens donner à ces paradoxes ? Sont-ils résolus par mise en contexte ?

- Informatique : Écrire des méta-données pour rendre plus efficace la recherche d'informations ? Rendre les programmes plus simples et adaptables en leur fournissant des moyens de s'observer ? À quoi sert vraiment la réflexivité ? Y-a-t-il de véritables applications ?

- Sociologie : C'est parce, qu'à un moment de l'histoire (modernité), les sociétés se sont pensées comme production d'elles-mêmes (contrairement aux sociétés traditionnelles qui se pensaient comme ordonnées par l'au-delà) que la sociologie est apparue. Mais la sociologie classique, sans doute emportée par son désir de s'imposer comme une science, s'est un peu vite octroyé le monopole de cette pensée réflexive. Les acteurs sociaux, trop pris dans les déterminations sociales qui les définissent, sont incapables, nous dit-elle, de relever les "lois sociales". Seuls les sociologues, opérant une "coupure épistémologique" (prise de distance nécessaire) peuvent y parvenir. À un moment où nos sociétés se scrutent à un degré jamais atteint auparavant dans leur fonctionnement et leur changement, et où les individus sont sommés "d'être eux-mêmes" (avec la dose d'introspection que cela entraîne), peut-on encore dénier aux acteurs toute capacité de réflexivité et de compréhension sociologique ? Et si non, quelles conséquences méthodologiques faut-il en tirer ?

- Peinture, arts graphique : L'art est friant de ces boucles étranges issues du jeu de l'auto-référence et des miroirs : de l'art géométrique d'Escher avec ses mains qui se dessinent elles-mêmes à l'art publicitaire comme la célèbre vache qui rit qui porte des boucles d'oreilles ornées de sa propre figure. Pourquoi les artistes sont-ils fascinés par ces figures ? L'art ne parle-t-il que d'art ?

- Sciences juridiques : Les lois d'un système juridique forment un système complexe réflexif. L'auto-amendement (i.e la loi qui s'amende elle-même) par exemple a été étudié par le philosophe Peter Suber au moyen du jeu nommé Nomic. Dans ce jeu, changer une règle est un coup légal et les règles initiales régulent la façon de changer les règles. Tous peuvent changer, y compris le but du jeu. La réflexivité permet-elle de modéliser le jeu démocratique ?

- Logique : la logique est-elle une méta-théorie qui permet de décrire ou de formaliser toutes les autres ? peut-elle se substituer à d'autres modèles dits « informels » ou « semi-formels » ? peut-on faire le bilan de la formalisation en mathématiques et la formalisation en informatique peut-elle se penser de la même manière ? le langage et la logique sont-ils en situation de référence croisée ?

- Littérature, cinéma : La narration écrite ou filmée utilise des effets d'enchassement, de miroirs, c'est le film dans le film, la pièce dans la pièce au théâtre. Les auteurs en jouent allant parfois jusqu'à transgresser les règles permettant alors aux acteurs et aux personnages de traverser les miroirs d'une autre réalité ou encore d'interroger le lecteur ou le spectateur ce qui l'amène à douter de sa propre réalité. Alors pilule rouge ou bleue ?

- Géographie : Problématique de l'échelle et auto-similarité, jeu du miroir et de co-construction des Systèmes d'Informations géographiques. La terre (Gaïa) est-elle un système auto-organisé comme un sorte d'organisme vivant ?

- Sciences de la vie : Les systèmes vivants sont des systèmes complexes récursifs où les boucles de rétro-action et les mécanismes d'auto-observation, d'auto-modification sont importants. Est-il possible de lier leur fonctionnement à des phénomènes physiques ou chimique sans tomber dans le réductionnisme ? Que nous apportent la vision réflexive de ces systèmes ?

- Économie : L'économie connaît un grand nombre de références croisées, comme par exemple la "boucle prix-salaires" : les salaires sont indexés sur les prix, et ces derniers prennent en compte les coûts salariaux. Mais l'exemple le plus intéressant par les questions qu'il soulève concerne des acteurs en situation d'interaction "stratégique". Le processus auto-référentiel correspondant est bien connu depuis Keynes qui a utilisé la métaphore d'un "concours de beauté" pour mettre en évidence la nature auto-référencielle du comportement spéculatif. Comment modéliser l'auto-référence dans les décisions économiques ? quelles en sont les conséquences globales des représentations et des comportements individuels autoréférentiels ?

- Théorie des jeux : Au niveau formel, la théorie des jeux utilise largement des formes variées de réflexivité, tant au niveau de la rationalité des joueurs que de celui des conditions dans lesquelles les choix sont effectués (hypothèse de "connaissance commune" : chacun sait, que chacun sait que l'autre sait, et ainsi de suite jusqu'à l'infini...). Dans le cas du "concours de beauté", le concept de "profondeur stratégique" permet de contourner la récursion infinie par des arguments pragmatiques. Quel rôle joue la réflexivité et l'auto-référence dans la formalisation des choix individuels en interactions ?

- Sciences cognitives : La modélisation de la réfléxivité dans la prise de décision par des agents cognitif en interactions et de la coordination de ces derniers pose des questions générales relatives aux théories de l'action, de la connaissance, de

la socialisation, des règles, conventions et institutions... Ce sont autant de domaines qui peuvent être concernés par notre réflexion inter-disciplinaire sur réflexivité et auto-référence dans les systèmes complexes.

## **L'objectif des Journées de Rochebrune**

Depuis 1992, l'objectif des Journées De Rochebrune reste inchangé : il s'agit d'offrir un espace d'échanges et de débats interdisciplinaires à tous les chercheurs/chercheuses qui travaillent sur les systèmes complexes naturels ou artificiels. Rochebrune se veut un lieu d'échange de la recherche en train de se faire et de s'observer : les contributions ne sont donc pas nécessairement finalisées et l'on peut aussi parler de choses en cours, pour lesquelles les doutes et les questions pèsent aussi lourd que les réponses déjà apportées. Dans le cadre du thème de cette année, les participants des Journées de Rochebrune pourront s'interroger sur ce qui fait la nature de Rochebrune, ses succès, ses échecs et pourront dans un processus d'auto-modification tenter de le modifier pour les futures éditions. Il sera demandé aux participants un effort tout particulier de vulgarisation dans la présentation de leurs idées afin que le débat soit fructueux entre les champs disciplinaires.

# **Rochebrune 05 – du 23 au 28 janvier 2005 –**

## **Programme prévisionnel**

### **Lundi**

#### Introduction

- 16h00-16h45 : La réflexivité chez Maturana, Jean-Pierre Müller, CIRAD.

#### Informatique

- 16h45-17h30 : Modèles de composants logiciels réflexifs et dialogiques, Yasmine Charif, Nicolas Sabouret, CRIP5, LIP6.

- 17h30-18h00 : Pause

- 18h00-18h45 : Réflexivité et programmation du jeu de Go, Tristan Cazenave, Laboratoire d'Intelligence Artificielle, Université Paris 8.

- 18h45-19h30 : Comparaison entre les comportements réflexifs du langage humain et la réflexivité des langages informatiques, Anne Nicolle, GREYC, CNRS/Université de Caen.

- 19h30-20h30 : Repas

- 20h30-21h15 : Artrospection, Roger Cozien.

- 21h15-22h30 : Expérimentation réflexive 1, Denis Phan.

### **Mardi**

#### Adaptation et coordination

- 16h00-16h45 : Adaptation réflexive : Contrôleur de comportement et structure évolutive, Elpida S. Tzafestas, Institute of Communication and Computer Systems National Technical University of Athens.

- 16h45-17h30 : Coordination, cognition et réflexivité aux fondements des "croyances sociales", Denis Phan Université de Rennes 1 - CREM - UMR CNRS 6211.

- 17h30-18h00 : Pause

#### Société

- 18h00-18h45 : Définition juridique et Argument de la diagonale, Danièle Bourcier, CNRS-Centre de recherche en Sciences administratives, Paris, Sylviane R. Schwer, Laboratoire de Linguistique Informatique Cognition et Communication.

- 18h45-19h30 : Réflexivité et hypermodernité, Francis Jauréguiberry, SET – CNRS (UMR 5603), Université de Pau.
- 19h30-20h30 : Repas
- 20h30-21h15 : De la conception écologique à une écologie de la conception : un nouvel univers de pertinence et un cadre conceptuel général de la conception, Pierre de Coninck, École de Design Industriel, Université de Montréal.
- 21h15-22h : Mise en abîme et théorèmes de points fixes, Sylviane R. Schwer, Laboratoire LaLICC, UMR 8139 (CNRS - Université Paris-Sorbonne).

## **Mercredi**

### Linguistique

- 16h00-16h45 : La réflexivité dans l'évaluation en Traitement Automatique des Langues, Pierre Beust, GREYC, CNRS/Université de Caen.
- 16h45-17h30 : Réflexivité dans un processus d'interaction homme-machine. Exemple de résolution de charades assistée par ordinateur , Serge Mauger, Pierre-Sylvain Luquet, GREYC, CNRS/Université de Caen.
- 17h30-18h00 : Pause
- 18h00-18h45 : Réflexivité en linguistique, en discours et en grammaire, Nadine Lucas, GREYC, CNRS/Université de Caen.

### Art

- 18h45-19h30 : Le danseur de ballet en miroir, Francis Rousseaux, IRCAM-CNRS.
- 19h30-20h30 : Repas
- 20h30-21h15 : On the Design of a Musical Flow Machine, François Pachet, Sony CSL, Paris (republication avec autorisation de : Pachet, F. On the Design of Flow Machines. The Future of Learning, Tokoro & Steels Eds., IOS Press, 2004).
- 21h15-22h30 : Expérimentation réflexive 2 - Keiko Abe et François Pachet.

## **Jeudi**

### Espace

- 16h00-16h45 : Designing reflexive agents in a MAS model for urban mobility, Sylvie Occelli, IRES, Turin, Italie.

- 16h45-17h30 : Auto-similitude et auto-référence en urbanisme. Quelques exemples. Dominique Badariotti, Laboratoire SET, Université de Pau et des Pays de l'Adour.

- 17h30-18h00 : Pause

- 18h00-18h45 : Et si le contraire était vrai ? Remarques sur la tautologie comme figure du lien entre les pensées géographiques et aménagistes, Olivier Soubeyran, Université de Grenoble 1, UMR CNRS Territoires.

- 18h45-19h30 : Double visage du menteur, Jérôme Cardot, LaLICC, UMR 8139 - CNRS - Université Paris IV Sorbonne.

- 19h30-20h30 : Repas

- 20h30-22h : Débat Rochebrune 2006

## Vendredi

### Le soi et les autres ...

- 16h00-16h45 : Les paradoxes du regardeur, « ces biais par où les fictions s'introduisent », Myrielle Hammer, EVCAU.

- 16h45-17h30 : Le soi dans le regard des autres, un essai sur le système philosophique de Jean-Jacques Rousseau, Yûji Sakakura, Université de Rikkyo, Tokyo.

- 17h30-18h00 : Pause

- 18h00-18h45 : L'égalité à « soi » n'est pas l'identité : un commentaire de la Première Hypothèse du Parménide de Platon, Violaine Prince, LIRMM-CNRS, Université Montpellier 2.

- 18h45-19h30 : Self reference of serendipitists, Pek van Andel, Faculty of Medical Sciences, University of Groningen, Holland.

- 19h30-20h30 : Repas

### Physique

- 20h30-21h15 : Le mythe du bootstrap dans la physique moderne, Jerzy Karczmarczuk, GREYC-CNRS, Université de Caen.



## **Comité d'Organisation**

Serge Stinckwich  
Jean-Louis Dessalles  
Nils Ferrand  
Thierry Fuhs  
Francis Jaureguiberry  
Jerzy Karczmarczuk  
Denis Phan

## **Comité de Programme**

Keiko Abé  
Frédéric Amblard  
Evelyne Andreewsky  
Dominique Badariotti  
Pierre Beust  
Jacques Blanc-Talon  
Danièle Bourcier  
Noury Bouraqadi  
Guillaume Deffuant  
Hervé Glotin  
Francis Jauréguiberry  
Jerzy Karczmarczuk  
Anne Nicolle  
Jean-Pierre Müller  
Sylvia Occelli  
Christophe Parisse  
Denis Phan  
Vincent Régnard  
Francis Rousseaux  
Serge Stinckwich  
Martine Timsit



# Index des auteurs

- Badariotti D. 13  
Beust P. 23  
Bourcier D. 37  
Cardot J. 45  
Cazenave T. 53  
Charif Y. 65  
de Coninck P. 77  
Hammer M. 89  
Jauréguiberry F. 103  
Karczmarczuk J. 113  
Lucas N. 125  
Luquet P-S. 139  
Mauger S. 139  
Nicolle A. 151  
Occelli S. 163  
Pachet F. 175  
Phan D. 201  
Prince V. 221  
Rousseaux F. 233  
Sabouret N. 65  
Sakakura Y. 247  
Schwer S. 37; 257  
Tzafestas E. 269  
van Andel P. 283



## Auto-similarité et auto-référence en urbanisme.

### Quelques exemples...

**Dominique Badariotti**  
**Laboratoire SET, Université de Pau et des Pays de l'Adour**

A la fois corps social, entité fonctionnelle et construction architecturale, la ville est un objet complexe émergent, apparu en l'espace de quelques milliers d'années dans plusieurs foyers de civilisations au niveau du globe suite à la révolution agricole du néolithique. Ce type d'organisme s'est par la suite multiplié de par le monde, tout en se transformant et en se modifiant. La nécessité de s'adapter aux nouvelles technologies mises en œuvre par l'humanité et le besoin de mieux correspondre aux nouvelles attentes, aux nouveaux besoins et aux nouveaux modes de vie de ses habitants a représenté un puissant moteur de changement.

Certains auteurs distinguent l'*urbs*, la part physique de la ville, de la *polis*, la part sociétale, dans une vision analytique classique, qui a besoin de segmenter la touffe de l'organisme urbain pour l'étudier. Mais *urbs* et *polis* ne sont que deux représentations du même organisme et ont de ce fait subi un même destin : la ville physique s'est modifiée, mais la ville sociétale a changé aussi, au long des millénaires. Ainsi s'exprime la morphogénèse urbaine : un changement de fond se traduit aussi par un changement de forme, à plus ou moins brève échéance.

*Urbs* et *polis* ont donc évolué progressivement et se sont métamorphosés, au cours des siècles. Les données architecturales et urbaines d'origine ont été profondément transformées durant ce long exercice d'adaptation, qui a pu durer près de 10 000 ans pour les villes les plus anciennes : hormis quelques vestiges, que reste-t-il à Jéricho ou à Ur de la fondation urbaine néolithique ? Que reste-t-il à Bagdad de l'antique Babylone ? Et à Rome de l'ancienne capitale d'empire ? Pourtant, cette métamorphose n'a pas tout déconstruit et tout effacé : de nombreux aspects demeurent malgré les changements. Notamment, une chose remarquable demeure : c'est que la plupart de ces villes se sont transformées sur le même lieu, pratiquement sur le même site lorsque les aménagements ne l'ont pas trop défiguré. La continuité de la localisation tranche ainsi avec la variabilité de la forme locale de la ville.

Dans un monde en mouvement, et hormis quelques exceptions remarquables (Troie en Asie mineure, Angkor en pays Kmehr, Cartage en Tunisie, Ciudad perdida en Colombie,...), les villes originelles se sont en effet modifiées sur place, sur les lieux mêmes qui les ont vu naître, reconstruisant ainsi de nouvelles villes sur les anciennes sans briser le lien auto-référentiel qui les relie à leur origine. Il y a donc, du point de vue des localisations, une forme profonde de réflexivité dans le fait urbain.

Mais les aspects réflexifs ou auto-référents de la ville ne se limitent pas à la réutilisation du site des villes : d'autres aspects relevant tantôt de l'ensemble de villes tantôt du *détail* urbain ont une forte charge réflexive. Sans prétendre à l'exhaustivité, nous aimerions ici évoquer divers aspects auto-référentiels que tout un chacun peut observer dans l'armature urbaine de nos contrées ou dans les détails architecturaux de nos villes. Ces aspects peuvent être implicites, apparents, pratiquement évidents pour qui sait observer les systèmes de villes ou les paysages architecturaux et urbains de nos cités contemporaines : mais ils peuvent aussi être dissimulés, cachés, et plus difficile à mettre en évidence. Qu'ils soient apparents ou qu'il faille les expliciter, y compris par la mesure, ces aspects auto-référentiels ne sont pas neutres pour autant : ils pèsent sur la géographie des villes, ils contraignent l'urbanisme de la ville tout autant que son organisation.

### Aspects implicites, évidents, observables

La forme physique de la ville, son esthétique particulière qui forme son paysage, est un premier élément directement observable par tout un chacun et comportant des dimensions auto-référentielles quasi-évidentes. Or, que voit-t-on de la ville, sinon l'architecture de ses bâtiments et le dessin de son plan ? A ces deux aspects, on peut en rajouter un troisième, moins évident a priori, mais fortement présent dans l'espace de la ville tels que

nous le vivons : il s'agit de l'espace des noms de lieux qui composent les paysages réels et symboliques de nos cités.

### *L'architecture : récursivité et auto-similarité et réflexivité*

Parmi les liens implicites, évidents, nous trouvons plusieurs aspects qui ont trait à l'édification de la ville, à la construction physique et architecturale, qui s'est très souvent et très longtemps nourrie de références. Cet aspect référentiel se retrouve par exemple dans la conception architecturale de chaque époque, qui, partageant les mêmes références et les mêmes goûts, a produit des édifices de proportions semblables.

Ces références étaient parfois de nature esthétique, générant de véritables phénomènes de mode qui ont finit par modeler nos paysages urbains selon des références bien reconnaissables. Paysages cubistes des cités contemporaines, ambiances néo-gothiques des quartiers victoriens ou wihlemliens, rigueur classique, voire antiques, des quartiers des 16<sup>ème</sup>, 17<sup>ème</sup> et 18<sup>ème</sup> siècles : tous ces ensembles se retrouvent isolément dans nos villes et tissent comme un lien avec le passé (auto-référence).

Mais, l'architecture n'est pas qu'une affaire esthétique : ces références ont aussi puisé dans des gabarits similaires, profondément ancrés dans les connaissances et les pratiques des hommes de l'art. Toutes semblables et toutes différentes, les villes européennes par exemple ont été construites tout au long de l'histoire par des bâtisseurs qui se sont transmis par des voies traditionnelles un corpus de techniques de conception, de dessin et de réalisation qui ont créé des liens entre les divers ensembles urbains ainsi réalisés. Par exemple, l'utilisation du nombre d'or dans le dessin d'architecture en Occident a conduit à élaborer des bâtiments dont les proportions sont semblables. Cette similarité de proportions fait référence à des constructions mythiques : originellement aux dimensions du temple de Salomon, selon la légende du nombre d'or.

Les modes de transmission du savoir des architectes (mode essentiellement oral à l'origine, reposant sur l'apprentissage et le compagnonnage, et valorisant fortement l'historicité de la discipline) ont ainsi tissé par le passé un vaste ensemble auto-référentiel, que l'on retrouve dans le style des immeubles mais aussi dans leurs proportions. Le goût architectural des citadins s'en est trouvé fortement influencé. Par la suite, l'industrialisation de la construction du 20<sup>ème</sup> siècle a également généré des formes auto-référencées, parfois roboratives, en développant un peu partout les constructions standardisées : chorons, barres, tours, quartiers (plans de lotissement, de cités, de ZAC). Ceci pose la question du rôle du modèle dans l'architecture et dans la construction.

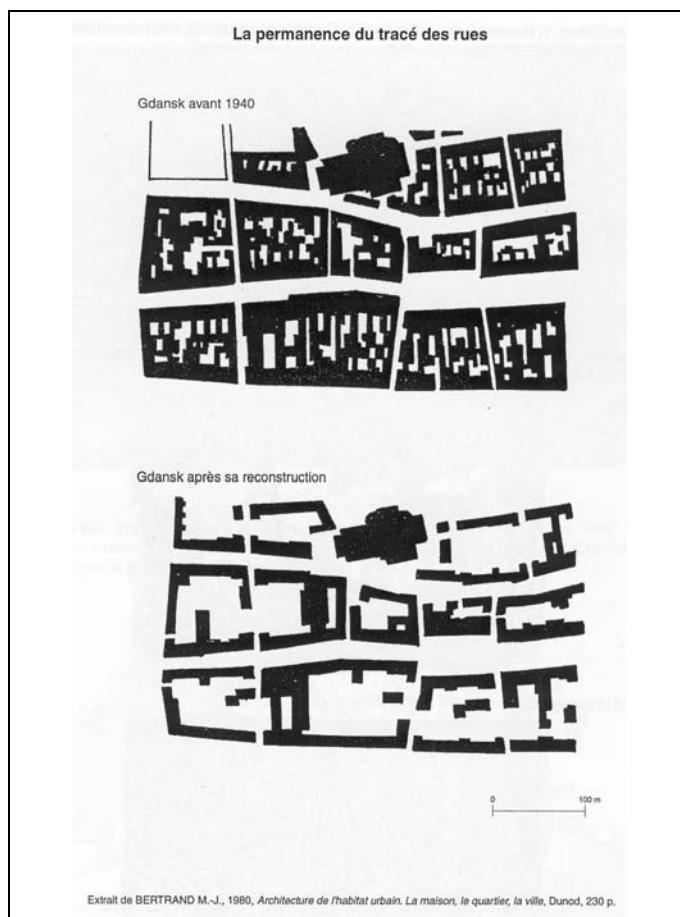
### *Le plan : récursivité, auto-similarité et réflexivité*

A un autre niveau, le plan d'ensemble de la ville est lui même fortement auto-référentiel. Pierre Lavedan (1926-1956), en son temps, avait émis la loi de la permanence du tracé des rues et des places dans les plans urbains à partir de l'étude minutieuse de dizaines de plans de villes européennes. Il avait en effet observé le maintien, à travers les siècles, des trames de rues antiques dans la plupart des villes d'origine grecque ou romaine. Les fouilles archéologiques on en effet pu montrer que les axes majeurs que forment le *cardo* et le *décumanus*, antiques voies principales reliant les portes de la ville du Nord au Sud et d'Est en Ouest, avaient été conservés, à l'instar de la place du *campo*, situé à la croisée de ces deux voies, et du tracé des rues secondaires dérivant de ces axes principaux.

Les axes plus récents ont subi des sorts comparables

Cette permanence (faut-il parler de résilience ?) que l'on peut à la limite comprendre tant que le tissu urbain n'est pas bouleversé par quelque destruction massive, est parfois même observable dans des situations de reconstruction complète. Si l'on étudie le cas de villes ayant subi de fortes destructions pendant les dernières guerres, comme celles de Dunkerque ou de Dantzig (fig. 1), on constate que les projets de reconstruction conservent quelque chose de la trame ancienne, alors même qu'aucune contrainte physique ne les y constraint. Et que dire de Varsovie où le projet avait été de reconstruire à « l'identique » une ville presque totalement anéantie ? Pourquoi dans ces cas avoir maintenu des lambeaux, sinon la totalité, de l'ancienne *urbs*, sinon parce que la *polis* le réclamait, le revendiquait.

**Figure 1 :** La permanence du tracé des rues à Dantzig



On le voit bien ici, la séparation entre l'*urbs* et la *polis*, la forme et son contenu n'est pertinente que pour l'analyse : mais elle ne correspond pas à une réalité.

#### *Les odonymes : récursivité urbs et polis*

Autre forme d'auto-référence observable dans le domaine urbain, concerne l'utilisation faite des noms de rue et de place (les odonymes). Ceux-ci affichent traditionnellement, dans certaines sociétés occidentales ou orientales, un lien particulier avec l'usage et le passé de ces lieux.

L'étude des noms de rue et de places montre en effet que dans les sociétés occidentales traditionnelles, les odonymes correspondaient la plupart du temps, avant la révolution française, à la fonction des lieux. Ainsi le marché aux herbes se tient-il place du marché-aux-herbes, les orfèvres travaillent-ils rue des orfèvres, une bastille se trouve-t-elle place de la Bastille, et l'église se localise-t-elle .... rue de l'église !

On affiche la nature réelle des lieux en faisant référence à ce qui s'y passe, créant ainsi un effet miroir : « le marché de la place du marché », « l'église de la rue de l'église », ... *Urbs* et *polis* convergent ici symboliquement : la nature physique de la ville et son usage social se trouvent mariés récursivement dans cette ville traditionnelle. Le marché en appelle à la place du marché et la place du marché fait référence au marché. Cette forme d'auto-référence n'est pas dénuée de sens pratique : un visiteur ne connaissant pas la ville pouvait se faire une assez bonne idée de la structure urbaine d'une ville inconnue simplement en s'y promenant et en observant les odonymes. Parallèlement une simple lecture du plan de la ville permettait de se donner une idée du fonctionnement de la ville, et de se renseigner sur la localisation de tel ou tel service.

Par la suite, à partir de la fin du 18<sup>ème</sup> siècle, la politisation et la banalisation de l'utilisation des odonymes (Milo, 1984) finira par distendre les liens récursifs entre noms de lieux et usages des lieux. On n'a jamais vu d'être suprême dans la rue du même nom, il n'y a pas de pinsons rue des Pinsons, De Gaulle n'a jamais mis les pieds place de général de Gaulle, l'avenue d'Alsace-Lorraine ne conduit pas en ces provinces, et la rue de l'Espérance est une sinistre rue de banlieue.

On le voit, des aspects autoréférentiels sont présents de façon presque évidente dans les organismes urbains .... de façon trop évidente serait-on tenté de dire ! En effet, dans la plupart des exemples cités plus haut, la limite entre la référence et l'auto-référence n'est pas totalement claire et nette. Lorsqu'un style architectural fait référence à un style similaire, est-on dans le domaine de l'auto-référence ou dans celui plus ordinaire de la référence ? La limite entre les deux termes est ici reste relativement floue et ne permet pas à l'auto-similarité de se manifester de façon si évidente que cela. Il nous faut donc trouver d'autres formes d'auto-référence, plus raisonnées, plus mesurées, pour compléter cette étude.

### Aspects moins évidents, cachés, à dévoiler, à expliciter

Les aspects auto-similaires ne sont pas absents du domaine de la ville : il s'y manifestent même à plusieurs niveaux et selon diverses facettes.

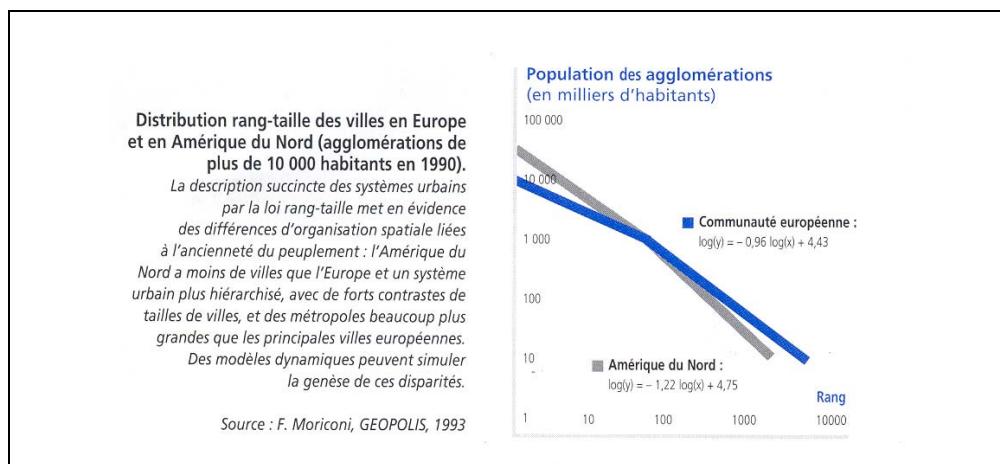
#### *L'armature urbaine : auto-similarité*

Dans le domaine de l'armature urbaine, l'autosimilarité apparaît très nettement dans la loi rang-taille, ou loi rang-dimension de Zipf (1949) appliquée à la distribution des villes. L'armature urbaine, que l'on appelle aussi la hiérarchie urbaine, désigne la structuration des villes à l'échelle d'un espace particulier, pouvant être une région, un pays, un continent, ...

Cette armature urbaine consiste en une répartition plus ou moins régulière des villes dans l'espace selon leur taille et leur fonction, répartition qui organise les espaces contrôlés par ces villes. Si l'on recense toutes les villes d'un espace particulier et qu'on les classe selon leur taille, on observe qu'en général ce classement suit une loi hiérarchique exponentielle où la population de chaque ville peut être déduite de son rang et de la connaissance de la population de la ville la plus importante selon la formule suivante :

$$P_n = P_1 / n^q \quad (\text{avec } P_n \text{ population de la ville de rang } n, P_1 \text{ population de la plus grande ville, } q \text{ paramètre caractérisant le degré de hiérarchisation})$$

**Figure 2 :** Loi rang-taille en Europe et en Amérique du Nord



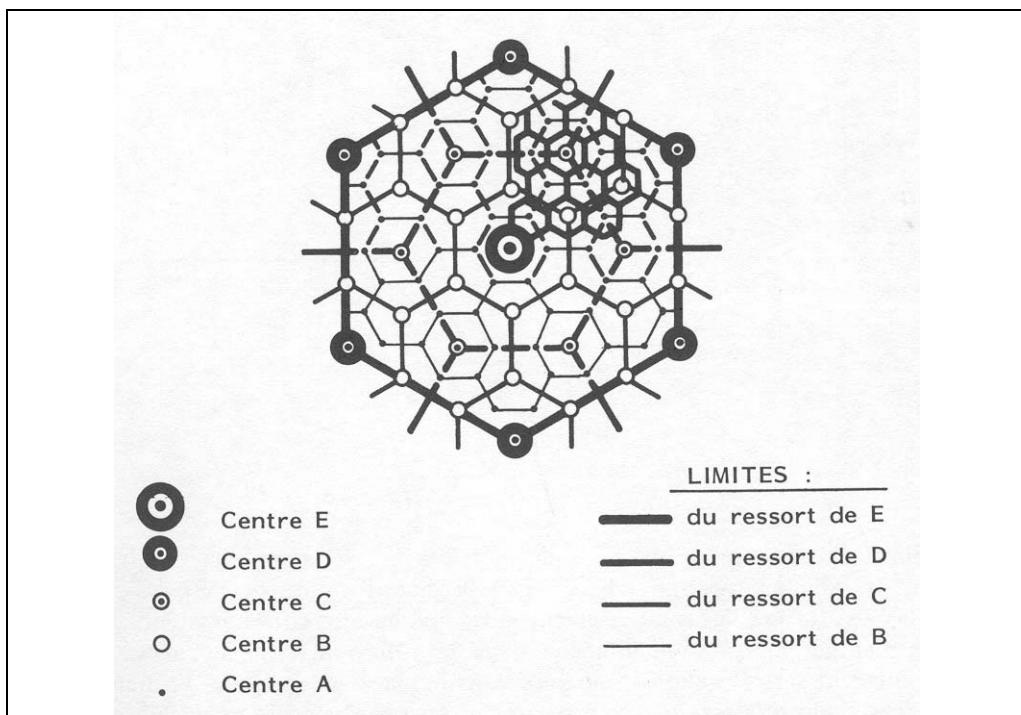
Cette loi est en fait une application de la loi de Pareto aux distributions urbaines : celles-ci s'organisent hiérarchiquement, conformément à la loi allométrique que l'on rencontre dans les systèmes biologiques et physiques. Cette organisation hiérarchique de villes est remarquable à deux titres : d'une part elle ne varie que très peu d'un espace à l'autre, ou d'un niveau d'analyse à l'autre ; deuxièmement, elle ne varie guère dans le temps, malgré le déclassement ou le reclassement de certaines villes. Il y donc là une véritable auto-similarité de l'armature urbaine à travers les échelles et en dépit des changements apportés par le temps.

#### *Les places centrales : auto-similarité à travers les échelles*

Un autre schéma d'organisation théorique de l'espace urbain montre également un caractère auto-similaire marqué : il s'agit de l'organisation hiérarchique des centres urbains, tel que le géographe Walter Christaller (1933) l'a présentée dans sa théorie des places centrales.

Selon la théorie de Christaller, la localisation des villes répond à des règles strictes qui tendent, dans un espace sans contrainte, à les disposer régulièrement. La géométrie de la répartition des villes est dans ce cas déterminée par leur degré de spécialisation fonctionnelle. Chaque ville assume en effet tout un ensemble de fonctions administratives et commerciales par rapport aux espaces environnants : elle est en même temps insérée dans une hiérarchie urbaine qui la relie aux villes de moindre niveau, placées sous son contrôle, et aux villes de niveau supérieur dont elle dépend pour certains équipements et certaines activités. La combinaison de tous ces éléments amène à proposer une structuration de l'espace sous la forme d'un treillage hexagonal qui se répète pour chaque niveau de ville (fig 3).

**Figure 3 :** Modèle de répartition des villes de Christaller



Walter Christaller propose ainsi une géométrie de la répartition des villes en 8 niveaux, qui s'emboîtent les uns dans les autres, et qui reposent tous intégralement sur une logique hexagonale. L'intégralité de l'espace est couvert par ce maillage, et tout point de l'espace est ainsi placé sous le contrôle d'une hiérarchie de villes déterminée. L'auto-similarité est ici induite par la répétition du treillage hexagonal à travers les échelles, conférant ainsi un aspect quasi-fractal à cette construction théorique. Toutefois, la fractalité de cette structure n'est ici qu'induite par la ressemblance entre les différentes échelles, elle ne repose sur rien d'autre et ne

correspond d'ailleurs pas à la pensée de Christaller, dans la mesure où son schéma est fondamentalement continu.

Néanmoins, ce schéma propose une vision auto-similaire de l'organisation spatiale des villes et de leurs zones de contrôle.

#### *Les marchés fonciers : auto-référence et réflexivité*

De puissant phénomènes auto-référentiels agissent aussi en ville et pèsent sur son fonctionnement et sur son organisation interne. Si on prend pour exemple l'étude des prix du sol urbain, on constate que ceux-ci sont en général fixés en référence aux prix pratiqués dans le voisinage pour des biens analogues. Toute nouvelle transaction se fait en fonction des prix observés jusque là, et va déterminer à son tour les prix observés pour les ventes ultérieures si la transaction se réalise.

Cet aspect auto-référentiel des marchés fonciers urbains a un impact morphologique très important sur la ville : en période de fluctuation des prix fonciers, il contribue à maintenir des spirales de valorisation ou de dévalorisation selon les quartiers. Dans le premier cas, l'impact transformationnel est élevé, car plus les terrains, immeubles ou logements urbains seront chers, moins ils seront rares (de nombreux propriétaires profitant de l'effet d'aubaine pour mettre en vente leur bien) d'où une modification profonde et accélérée des quartiers (modification morphologique, de nombreux immeubles étant détruits pour être reconstruits, mais modification sociale aussi, les anciens habitants étant souvent chassés au profit de nouveaux arrivants). Dans le second cas, on assistera à la dégradation inéluctable du bâti de certains quartiers, ce qui aura aussi des conséquences morphologiques et fonctionnelles importantes (taudification de ces zones et augmentation de la vacance de logements, ce qui débouchera bien souvent sur une politique de destruction ou de rénovation menée par les pouvoirs publics à partir d'un certain seuil de dégradation).

Au-delà même de l'auto-similarité, il semble bien que les marchés fonciers aient un aspect réflexif dans la mesure où toute nouvelle référence de prix foncier peut modifier les prix fonciers faisant référence à l'avenir. Le système des prix fonciers tend ainsi à s'ajuster en permanence, en altérant son comportement selon des phases de variations que de nombreux auteurs ont observé. A une tendance générale à la hausse depuis le 19<sup>ème</sup> siècle au moins (Granelle 1975) s'opposent ponctuellement – et presque régulièrement – des phases de dépréciation.

#### *Auto-référence et auto-similarité morphologique*

Plusieurs auteurs l'on remarqué (Batty 1991 ; François, Frankhauser, Pumain 1995 ; ...), la morphologie des villes présente des aspects fractals jusqu'à un certain niveau, ce qui implique une certaine auto-similarité morphologique à travers les échelles.

En effet, les agglomérations urbaines ont un aspect fractal dans la mesure où elles sont formées de quartiers urbains irréguliers et fragmentés, eux-mêmes issus d'îlots de même type, eux-mêmes constitués d'immeubles de même acabit, eux-mêmes composés de pièces irrégulières, elles mêmes fabriquées de briques de toutes tailles, (fig 4) ... Cette fractalité générale de la structure urbaine se traduit également par un lien hiérarchique entre les divers éléments urbains. On y trouve en effet peu de grandes parcelles et beaucoup de petites, peu de très grandes avenues et de nombreuses rues ou ruelles, peu de très grands immeubles et beaucoup de constructions modestes, ... Toute ville présente ainsi une structure auto-similaire et fortement hiérarchique à l'instar des figures fractales.

Cette fractalité des structures urbaines peut être localement ou globalement mesurée. Diverses techniques existent à cet effet (Frankhauser 1994). A titre d'exemple nous avons mesuré la fractalité de divers quartiers de l'agglomération transfrontalière de Sarrebruck – Forbach et nous avons fait varier la résolution de nos images pour voir si la mesure de la dimension fractale en était affectée. Les résultats sont éloquents : la variation de dimension fractale mesurée pour ces images est très faible (par exemple 0.48% pour le centre de Sarrebruck – tableau 2) et ce bien que les images en question n'apparaissent plus comme étant rigoureusement identiques pour l'œil humain.

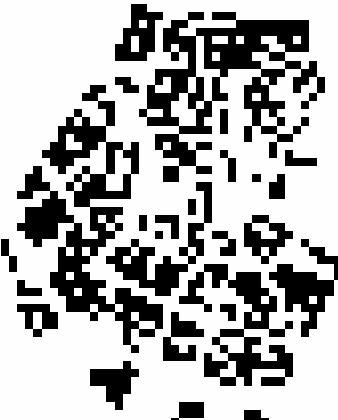
**Figure 4 :** « Fractaville » : Sarcelles, une image de la ville fractale



*Sarcelles, grand ensemble de Lochères en 1966.  
© J. Windenberger*

Dans cet exemple, la similarité est visuellement dégradée, car il s'agit d'un changement de résolution et non d'un changement d'échelle stricto-sensu. Malgré les différences entre les figures, la similarité est quand même bien présente du point de vue de la mesure, c'est-à-dire du point de vue de la répartition discrète des masses. D'où un calcul de la dimension fractale  $D$  très homogène, variant de 1.75 à 1.76, et ce quelle que soit l'image. La similarité à travers les échelles ainsi que l'aspect hiérarchique sont donc bien présent, même lorsqu'ils ne sont pas évidents, dans nos morphologies urbaines, leur donnant des caractéristiques fractales.

**Tableau 1 :** Centre ancien de Sarrebruck. Effet de la résolution sur la représentation de la morphologie urbaine et sur les dimensions fractales

<b>Résolution</b> <b>1 m</b>  Echelle 20 %  $D_{cor} = 1.749$		<b>Résolution</b> <b>2.5 m</b>  Echelle 50 %  $D_{cor} = 1.748$	
<i>Résolution</i> <b>5 m</b>  Echelle 100 %  $D_{cor} = 1.748$		<i>Résolution</i> <b>10 m</b>  Echelle 200 %  $D_{cor} = 1.753$	
<i>Résolution</i> <b>15 m</b>  Echelle 300 %  $D_{cor} = 1.760$		<i>Résolution</i> <b>20 m</b>  Echelle 400 %  $D_{cor} = 1.764$	

## Conclusion

Ainsi divers niveaux de récursivité se matérialisent dans la géographie et dans l'histoire des villes, positionnant les concepts d'auto-référence, de similarité, de réflexivité les uns par rapport aux autres et ouvrant à de nombreuses questions.

Par exemple l'autosimilarité est bien présente dans les structures urbaines : elle y est induite par l'auto-référence des constructions architecturales entre elles ou par l'autosimilarité des plans urbains. Mais quelle est la limite entre référence et auto-référence ? Lorsque l'on fait référence à un style architectural particulier, par exemple le gothique, ou à une composition urbaine typique (par exemple le hausmannien) est-on dans le domaine de la simple référence ou dans celui de l'auto-référence ?

Par ailleurs, comment se situent auto-référence et récurrence ? Un dessin d'immeuble récurrent est-il aussi auto-référent ? ou auto-similaire ? ou les deux à la fois ? Le fait que l'on améliore parfois ces immeubles d'une construction à l'autre suffit-il à établir la réflexivité ?

Enfin, ces ressemblances de paysages urbains, que l'on observe entre des villes finalement assez différentes, sont-elles voulues ou fortuites : faut-il y voir une construction raisonnée, un artefact déterminé par nos techniques de conception et de construction d'immeubles ou une émergence liée à la nature complexe de la ville ? Selon la réponse, quel est le poids réel de l'urbanisme dans la fabrication et l'évolution des villes ? A-t-il un simple statut adaptatif, se contentant de gérer marginalement les problèmes liés à la vie de l'organisme urbain, ou au contraire peut-il être plus directif ?

Au-delà des questions posées, cette étude permet également de souligner que *urbs* et *polis*, la forme et le fond, la morphologie et le fonctionnement des organismes urbains sont finalement étroitement liés. On retrouve là une idée, développée par D'Arcy Thomson, et reprise dans divers travaux de recherche. Cette idée est intéressante car elle vient en contre-point de ce qui a été réalisé en matière de recherche urbaine ces dernières années. Un certain désintérêt pour la géométrie associé à la spécialisation des approches a entraîné une relative désaffection pour les thèmes de recherche morphologiques :

## Références

- Batty Michael, 1991, Cities and fractals : simulating growth and form. In A.J. Crilly, R.A. Earnshaw, H. Jones, ed, *Fractals and Chaos*
- Christaller Walter, 1933, *Die zentralen Orte in Sueddeutschland*. Wiesbaden
- François Nicole, Frankhauser Pierre, Pumain Denise, 1995, Villes, densité et fractalité. *Les Annales de la Recherche Urbaine*, 67, 55-64
- Frankhauser Pierre (1994). La fractalité des structures urbaines. Paris: Anthropos - Economica.
- Granelle Jean-Jacques, 1975, *La valeur du sol urbain et la propriété foncière. Le marché des terrains à Paris*. Paris, Mouton.
- Lavedan Pierre, 1936, *La géographie des villes*, Paris, Gallimard
- Lavedan Pierre, 1926-1956, *Histoire de l'urbanisme*. Paris, Henri Laurens, 3 tomes
- Milo Daniel, 1984, Les noms de rues - in *Les lieux de mémoire*, sous la dir. de P. Nora, Paris, Gallimard, tome II vol. 3, p.283-320
- Pumain Denise, 1982, *La dynamique des villes*. Paris, Economica.
- Zipf Georges, 1949, *Human behavior and the principle of least effort*. New-York, Addison Wesley press.



# La réflexivité dans l'évaluation en Traitement Automatique des Langues

Pierre Beust

GREYC CNRS UMR 6072 – ISLanD<sup>1</sup> & Pôle ModesCoS<sup>2</sup> de la MRSN  
Université de Caen - Basse Normandie, 14032 Caen Cedex, France  
pierre.beust@info.unicaen.fr  
<http://www.info.unicaen.fr/~beust/>

Résumé : Cet article traite de l'évaluation dans le domaine du Traitement Automatique des Langues (TAL). Nous rappelons les grands principes des méthodes classiques d'évaluation en TAL et nous expliquons que ces métriques posent problème, surtout dans le cas de certains systèmes de TAL, dits individus-centrés. Ces problèmes tiennent essentiellement au fait que les métriques d'évaluation ne tiennent pas compte d'une propriété particulière des langues et des interactions homme-machine : la réflexivité. Nous proposons un cadre générique pour repenser la méthodologie d'évaluation des systèmes de TAL.

Mots clés : Evaluation, TAL, Système individu-centré, Sémantique des Langues

## 1. Introduction

L'objectif de cet article est de contribuer aux recherches sur de meilleures méthodes d'évaluation dans notre domaine de recherche à savoir l'informatique appliquée au langage naturel, domaine traditionnellement appelé le Traitement Automatique des Langues (TAL). Des méthodes pour rendre compte de la qualité du fonctionnement d'un système de TAL existent déjà mais elles ne font pas l'unanimité quant à leur possible adaptation à tout type d'application. Nous soutenons dans cet article que ces méthodes sont contestables notamment parce qu'elles ne tiennent pas compte d'une propriété essentielle des langues et des interactions homme-machine : la réflexivité.

Après avoir rappelé l'importance actuelle que tient la notion d'évaluation, nous montrerons comment elle est traditionnellement conduite en TAL. Nous dresserons une critique des méthodes classiques d'évaluation en TAL. Nous verrons alors que certains systèmes de TAL que nous qualifierons d'individus-centrés ne peuvent pas être évalués avec les mesures classiques. Enfin, nous proposerons quelques pistes pour une méthodologie d'évaluation dans le cas spécifique d'un traitement de la dimension sémantique des langues.

## 2. L'évaluation dans nos sociétés

Incontestablement, l'évaluation est une notion à la mode dans le fonctionnement actuel de nos structures sociales. Comme d'autres préoccupations actuelles telles que le principe de précaution par exemple, on retrouve l'évaluation impliquée dans un

---

<sup>1</sup> <http://users.info.unicaen.fr/~psluquet/Island/>

<sup>2</sup> <http://www.unicaen.fr/mrsh/modesco/>

nombre important de domaines de la vie publique : les sciences et technologies, les institutions et loi, les placements et transactions financiers, le commerce ...

Pour s'en convaincre, il suffit de soumettre une requête sur un moteur de recherche avec le mot clé « *évaluation* », ce qui en soi constitue une sorte d'évaluation, et d'examiner le nombre de réponses obtenues : approximativement 9 000 000 de réponses sur Google au 23/7/04 (1 460 000 sur Altavista à la même date). A titre de comparaison, on peut noter que le terme « *religion* » est crédité d'environ 8 000 000 de réponses et le terme « *politique* » d'environ 13 000 000 de réponses. Le nombre de réponses n'est évidemment pas un bon indicateur d'une présupposée importance d'un terme ou d'une notion car on pourrait tout au plus en conclure qu'il est plus ou moins utilisé dans la base de sites du moteur de recherche. Il est néanmoins un terme significativement important dans l'indexation du contenu de pages web. Il convient donc de rester prudent quant aux conclusions qu'on pourrait tirer de ces observations. Ceci prouve au passage, s'il en était besoin, qu'une évaluation peut être tout à fait contestable si son interprétation est erronée.

Il n'en reste pas moins que l'examen des premières réponses est instructif sur l'importance qu'on donne actuellement à la notion d'évaluation. Ainsi on trouve par exemple le site <http://www.evaluation.gouv.fr/> qui présente le portail de l'évaluation des politiques publiques. Ceci traduit bien que l'évaluation à valeur d'exigence sociale au point qu'il faille s'en justifier. De même dans le domaine législatif, on constate un recours de plus en plus important à l'évaluation dans les textes. On peut citer par exemple la loi organique de préparation de la loi de finances où il est prévu que les financements d'état fassent constamment l'objet d'évaluations conditionnant leur reconduction<sup>3</sup>. De même, le mouvement des chercheurs en mars 2004 et les textes en cours de discussion qui en découlent mettent particulièrement en avant l'évaluation des personnels de la recherche dans leur avancement de carrière ainsi que l'évaluation des structures de recherche qui les regroupent.

On peut également constater que d'une exigence sociale d'actualité on aurait souvent tendance à dériver un effet de mode. C'est par exemple le cas dans des méthodes publicitaires où l'évaluation est avancée comme un argument commercial. Ainsi on tente de vendre des services en les faisant évaluer et accréditer par des organismes de normalisation reconnus (c'est le cas de certaines normes ISO) où encore on cherche à vendre un yaourt plus qu'un autre parce qu'on argumente que des chercheurs en ont évalué (sans bien sûr dire comment) les retombées bénéfiques pour la santé. La mise en place d'une évaluation a ainsi pour but de donner l'impression d'une démarche sérieuse parce qu'on prend un recul sur ce qui est fait, parce qu'on met en place une démarche réflexive. Mais pour que cette démarche soit irréprochable il faudrait encore que la façon de mener cette évaluation soit claire. Pour qu'une démarche d'évaluation apporte une réelle valeur ajoutée, il convient de montrer en quoi elle est motivée.

---

<sup>3</sup> [http://www.minefi.gouv.fr/moderfie/avancee/outils-pdf/art109\\_juin.pdf](http://www.minefi.gouv.fr/moderfie/avancee/outils-pdf/art109_juin.pdf)

### 3. L'évaluation en TAL

Dans la plupart des domaines de l'informatique, l'évaluation des programmes a toujours été un souci premier car les conditions de leurs exécutions en découlent. Ainsi on a tout intérêt à savoir concevoir des programmes efficaces plutôt que des programmes qui en fonction des données qu'on leur soumet peuvent engendrer une explosion combinatoire, c'est-à-dire demander un temps et un espace mémoire tellement grands qu'il n'est même plus raisonnable d'en attendre la terminaison. Quantifier la qualité d'un programme du point de vue de sa complexité constitue un domaine de recherche au sein de l'informatique qu'on appelle *algorithmique* et dont le but est de caractériser les classes de problèmes ainsi que les solutions proposées en fonction principalement du temps « machine » et de l'espace mémoire utiles. L'algorithmique met donc l'accent sur les programmes mais les systèmes informatiques, qui plus est ceux développés en TAL, ne se réduisent pas à un jeu de programmes et d'autres aspects devraient être considérés pour une évaluation globale. Ainsi, la première chose à évaluer quand un programme produit un résultat est la qualité de ce résultat. D'autre part, quand des ressources sont nécessaires au fonctionnement d'un programme (bases de données, corpus de textes ...), il faut caractériser la taille mais aussi le temps « homme » qu'elles demandent pour leur élaboration. Enfin quand des modalités d'interaction homme-machine sont mises en place il faut savoir si elles sont efficaces en fonction du ou des utilisateurs et de la tâche visée.

La problématique de l'évaluation a été posée dès le début en TAL, principalement avec les travaux sur la traduction automatique et les rapports de Bar-Hillel (1960) et de l'ALPAC (1966) qui en mettaient en question la faisabilité. Depuis quelques années cette problématique est reposée comme un domaine de recherche en soi à travers différentes structures de recherche :

- des conférences internationales cherchant à mettre en compétition des systèmes de TAL concurrents (par exemple TREC<sup>4</sup> (Text Retrieval Conferences), MUC<sup>5</sup> (Message Understanding Conferences), LREC<sup>6</sup> (Language & Ressources Evaluation Conferences) ou SENSEVAL<sup>7</sup>)
- des organisations affichant cette thématique (par exemple l'ISKO<sup>8</sup> (International Society of Knowledge Organization) ou encore l'association ELRA<sup>9</sup> (European Language Resources Association) et son agence de distribution ELDA<sup>10</sup> (Agence pour l'évaluation et la distribution de ressources linguistique)).

---

<sup>4</sup> <http://trec.nist.gov/>

<sup>5</sup> [http://www.itl.nist.gov/iaui/894.02/related\\_projects/muc/info/muc\\_eval.html](http://www.itl.nist.gov/iaui/894.02/related_projects/muc/info/muc_eval.html)

<sup>6</sup> <http://www.lrec-conf.org/>

<sup>7</sup> <http://www.itri.brighton.ac.uk/events/senseval/>

<sup>8</sup> <http://is.gseis.ucla.edu/orgs/isko/>

<sup>9</sup> <http://www.elra.info/fr/>

<sup>10</sup> <http://www.elda.fr/>

- des actions communes à plusieurs organismes (par exemple l'action interministérielle TECHNOLANGUE<sup>11</sup> dont l'un des quatre volets est l'évaluation)

Dans le cadre particulier du TAL, les méthodes classiques d'évaluation (issues de la problématique dite de la Recherche d'Information) visent principalement la quantification de la qualité des résultats produits par un système. Deux taux sont proposés pour rendre compte de la réussite ou de l'échec d'un programme produisant un résultat à partir du traitement d'une données linguistique : le *Rappel* et la *Précision*. Le rappel est la proportion de bons résultats fournis par le programme par rapport aux bons résultats qu'il aurait dû idéalement fournir et la précision est la proportion de résultats corrects parmi l'ensemble des réponses fournies par le programme. Il en découle deux mesures duales du rappel et de la précision que sont le *Bruit* et le *Silence* : le bruit est la proportion de fausses réponses parmi l'ensemble des résultats d'un système et le silence est la proportion de bonnes réponses qui auraient dû être fournies par le système et qui ne faisaient pas partie de son résultat.

Considérons par exemple un système de recherche documentaire qui permettrait de retrouver des documents textuels à partir d'un corpus de textes thématiquement hétérogène et de quelques mots clés. Supposons que son utilisateur cherche des documents sur les langues vivantes en soumettant les mots clés *langue* et *vivante*. Si l'on sait que le corpus contient 50 documents traitant des langues vivantes et que le système donne 25 réponses, 20 correctes et 5 sur la préparation de langue de boeuf par exemple, alors on pourra quantifier le rappel à 40% (20/50) et la précision de 80% (20/25) et par conséquent le silence à 60% (100-40) et le bruit à 20% (100-80).

Les mesures de rappel et de précisions sont principalement mises en œuvre en TAL dans le cadre d'évaluations comparatives menées lors de campagnes où sont mis en rivalité différents systèmes. Ainsi on a très fortement recours au rappel et à la précision dans les conférences sur la compréhension automatique de message (MUC), la désambiguïsation de mots (SENSEVAL) ou l'extraction d'information (TREC) alors même que les objectifs opératoires spécifiques de ces conférences peuvent être très différents. Par exemple, il peut s'agir d'être capable de rechercher un passage d'un document où se situe la réponse à une question posée (tâche de Question/Réponse dans TREC) ou encore de remplir une base de données à partir d'une analyse sémantique de courts documents (par exemple lors de certaines conférences MUC, on a cherché à extraire des dates, des noms de lieux par exemple à partir de comptes-rendus d'attentats terroristes). Dans quelques cas particuliers, des variantes du rappel et de la précision sont mises en place. C'était le cas lors de l'action GRACE<sup>12</sup> (Adda & al. 1999) où il était question de comparer différents analyseurs morpho-syntaxiques du français contemporain avec les mesures de précision et de décision (en remplacement du rappel). La décision proposée dans GRACE mesure la proportion d'étiquetages stricts parmi l'ensemble de tous les

<sup>11</sup> <http://www.technolangue.net/>

<sup>12</sup> <http://www.limsi.fr/TLP/grace/www/presentation.html>

étiquetages (stricts et ambigus). Ainsi un taux de décision de 100% correspond à une analyse où toutes les formes sont étiquetées de façon non ambiguë, c'est-à-dire avec une seule catégorie.

## 4. Critique des évaluations classiques en TAL

Les définitions des taux de rappel et précision posent certains « problèmes ». Premièrement, contrairement à une évaluation d'ordre algorithmique (en terme de temps machine et d'espace mémoire) une évaluation en rappel et précision donne une place prépondérante à l'évaluateur humain car au final c'est lui qui décide si une réponse est bonne ou pas d'où une part d'individualité dans le résultat de l'évaluation. Deuxièmement, le calcul du rappel suppose qu'on sache déjà combien de documents correspondent à la requête donc une évaluation en rappel ne sera toujours qu'une approximation de la qualité d'un système calculée par rapport au résultat d'un agent humain sur un corpus de test et avec un jeu de requêtes test (forcément non exhaustif). Troisièmement, si l'on cherche à comparer différents systèmes en utilisant leurs taux de rappel et de précision, il faut savoir lequel des deux taux privilégier. Dans ce but une mesure synthétisant le rappel et la précision en fonction d'une variable de préférence a été définie par (Van Rijsbergen 1979) : c'est la F-mesure définie ainsi où  $R$  est le rappel,  $P$  la précision et  $\beta$  la variable :

$$F = \frac{(\beta^2 + 1)PR}{\beta^2P + R}$$

Comme  $R$  et  $P$ , la F-mesure doit être idéalement la plus proche de 1 possible. Lorsque  $\beta = 1$ , la précision et le rappel ont la même importance et on dit que la F-mesure est harmonique. Plus  $\beta$  est proche de 0 ; plus c'est la précision qui est privilégiée. Plus  $\beta$  est supérieur à 1 ; plus c'est le rappel qui est privilégié.

L'omniprésente utilisation des métriques classiques (rappel, précision, F-mesure et variantes) laisse croire que leur générnicité n'est pas contestable. Ce n'est pas à notre avis le cas. En premier lieu, il faut garder à l'esprit que ces métriques ne sont utilisables que lorsque qu'un corpus de test a pu être constitué et c'est loin d'être toujours le cas. Par exemple il n'est pas possible de calculer de façon incontestable le rappel d'un moteur de recherche sur Internet car personne ne peut établir en dehors de l'utilisation d'un moteur de recherche sur un échantillon de documents combien il existe précisément de documents qui répondent réellement à une requête. D'autre part, dans beaucoup de cas, quand un corpus de tests est réalisable, il ne faut pas sous-estimer le coût qu'il représente en temps et en investissement. A titre d'exemple, lors de l'action GRACE, une des étapes les plus importantes en temps a été de se mettre d'accord sur un jeu d'étiquettes morpho-syntactiques commun à tous les systèmes et au corpus de test alors qu'intuitivement trouver la catégorie morpho-syntaxique d'un mot dans une phrase semble ne pas poser de problèmes. On imagine dans ces conditions quelle serait la difficulté pour constituer des corpus de références

pour d'autres types d'évaluation beaucoup moins consensuelles (par exemple, il semblerait difficile de produire un corpus d'énoncés de dialogue avec étiquetage des actes de langage tellement la dimension interprétative subjective paraît ici première). Enfin il faut veiller à ce que le corpus de référence reste en accord avec les conditions réelles d'utilisation des systèmes pour lesquels il doit servir à l'évaluation. A ce sujet, (Lavenus & al. 2002) montrent bien la différence entre les questions/réponses du corpus de référence utilisés dans les conférences TREC par rapport à des vraies questions d'utilisateur en recherche documentaire. Par exemple, les questions du corpus de référence sont toutes des interrogatives canoniques courtes (par exemple « *What does a defibrillator do ?* ») alors que la majorité des demandes de « vrais » utilisateurs sont couramment des affirmatives complexes du style « je voudrais savoir ... ». Ainsi il convient de se rappeler de ce que montre (Spark Jones 2001) à savoir qu'une évaluation en laboratoire est fondée sur des présupposés concernant le contexte de la tâche, ce qui a un impact important sur le bien-fondé de l'évaluation et sur les résultats de performance. Pour considérer mieux la tâche, il faut remettre l'humain au centre de l'interaction avec les systèmes. C'est ce que proposent de faire (Lavenus & al. 2002) pour en déduire de nouvelles mesures. Nous sommes d'accord avec ces auteurs sur ce point de vue mais nous pensons qu'en fait de nouvelles mesures, c'est aussi et surtout de nouveaux protocoles d'évaluation qu'il faudrait chercher à mettre en place.

Dans les campagnes d'évaluation en TAL, les systèmes et modèles sont confrontés à un moment donné les uns aux autres sur les mêmes corpus de tests. C'est à notre avis une autre lacune de ces évaluations comparatives car il faudrait pouvoir évaluer les systèmes individuellement, ce qui éviterait déjà que le corpus de test soit en décalage avec les visées applicatives du système. D'autre part si l'on cherche à évaluer une application en soi, on va naturellement se poser des questions qui restent absentes des campagnes classiques d'évaluation, par exemple celle de la « durée de vie » du système<sup>13</sup>. Ce n'est pas parce qu'un programme informatique exécuté à deux moments distincts avec les mêmes ressources et sur les mêmes données donne en général les mêmes résultats qu'il n'est pas « périssable ». Evaluer à un moment donné ne garantit rien quant à une utilisation future quand on ne peut pas prédire les évolutions des données à traiter. C'est d'autant plus le cas dans les systèmes de TAL car les langues évoluent constamment et très rapidement qui plus est. Par exemple, les systèmes d'étiquetage morpho-syntaxique conçus il y a une dizaine d'années pouvaient assez raisonnablement faire l'hypothèse qu'un point était une marque de fin de phrase ou une marque de fin de mot dans le cas d'une abréviation. Cette hypothèse serait aujourd'hui considérablement fausse depuis que les adresses de sites Internet et les adresses e-mails dans lesquelles le point est fortement utilisé ont fait une apparition massive dans les écrits d'aujourd'hui, même les plus courants. Ceci est vrai en ce qui concerne la morpho-syntaxe et ça l'est encore plus en ce qui concerne

<sup>13</sup> Eric Bruillard dans [Siksou & al. 2003] montre que les logiciels ont leur « vie propre » qui est notamment caractérisée par les différences qui existent entre ce que les concepteurs du logiciel ont prévu qu'un « utilisateur idéal » fasse du logiciel et ce que les « vrais » utilisateurs dans des conditions réelles d'utilisation en font.

la terminologie et la sémantique tellement les mots changent en forme et en sens dans et par leur usage. Sur des corpus contemporains, des systèmes évalués sur des corpus plus anciens n'obtiendraient donc à coup sûr pas les mêmes taux de rappel et de précision. Une évaluation sérieuse d'un système de TAL devrait donc être menée en synchro-diachronie (et non simplement en synchronie comme c'est le cas dans les campagnes d'évaluation) car c'est ainsi que vivent les langues. Cette propriété des langues de se décrire sur les plans synchronique et diachronique à particulièrement été mise en évidence par Saussure. C'est bien parce que la réflexivité est première dans les langues vivantes que leur usage amène à les modifier elles-mêmes, passant ainsi d'un état synchronique à un autre état synchronique et ce sans cesse. En ne tenant pas compte de cette réflexivité au centre de la sémiotique des interactions langagières, les évaluations classiques en TAL ne font finalement pas la différence entre langues vivantes et langues mortes.

Si le cycle de vie de la donnée est à prendre en compte, celui du logiciel est aussi à considérer. Là encore les évaluations classiques sont restrictives car elles ne s'appliquent qu'à des systèmes considérés comme finalisés dont il convient de quantifier les résultats, mais il ne serait pas prévu d'évaluer la capacité d'apprentissage d'un système et la montée en qualité de ses résultats au fur et à mesure de cet apprentissage. Les campagnes d'évaluation ne s'appliquent donc qu'à des logiciels fournis « clés en main » et laissent de coté les systèmes dits « en amorce » et les logiciels conçus en laboratoires pour tester et corroborer des hypothèses sur du matériau attesté, dits « logiciels d'étude » (Nicolle 1996). C'est d'autant plus dommage que le TAL est un champs de recherche où aujourd'hui sont développés bon nombre de logiciels d'étude et de logiciels en amorce et pas uniquement dans des laboratoires de recherche (par exemple, le logiciel de reconnaissance vocale *ViaVoice*<sup>14</sup> d'IBM, bien qu'étant un produit commercial, est un logiciel en amorce étant donné qu'il nécessite une importante phase d'apprentissage de la prononciation de son utilisateur). Ceci pose la question du statut de l'évaluation par rapport aux spécifications du logiciel. Il n'est pas évident de chercher à évaluer, à un moment donné, les résultats d'un logiciel quand ses spécifications ne sont pas encore intégralement fixées, ce qui serait souvent le cas pour des logiciels d'étude où la réflexivité est au centre de la conception. Plutôt que de considérer l'évaluation comme étant forcément une tâche finale sans retour sur le logiciel (dont du coup l'intérêt peut être contestable et n'a finalement qu'une valeur de validation), il serait intéressant de chercher à l'intégrer à la conception même du logiciel en imaginant par exemple des méthodes d'évaluation à mi-parcours dans le cycle de vie du logiciel où l'utilisateur visé aurait déjà sa place<sup>15</sup>. En plus de chercher à savoir comment faire cette évaluation, on peut s'interroger sur sa nature : serait-ce déjà une réelle évaluation ou bien serait-ce encore une étape de l'étude de faisabilité ? Comme on le voit, un bon nombre de questions restent ouvertes et traduisent bien que l'évaluation

<sup>14</sup> <http://www-306.ibm.com/software/voice/viavoice/>

<sup>15</sup> Tenir compte de l'utilisateur très en amont dans le cycle de conception des logiciels amène à une meilleure prise en main des outils comme il a été montré dans le cadre du projet PIC (Nicolle & al. 2002).

est une problématique de recherche et pas simplement un enjeu technique qui ferait suite au travail de recherche.

## 5. Evaluer un système individu-centré

Dans le cadre de nos travaux de recherche sur l'instrumentalisation de la sémantique des langues au sein de l'équipe ISLaND<sup>16</sup> du laboratoire GREYC, nous nous intéressons plus particulièrement aux logiciels d'études que l'on appelle individu-centrés (également appelés anthropocentrés chez (Thlivitis 1998)). Un logiciel individu-centré est un système dans lequel les traitements réalisés dépendent intégralement des ressources construites en interaction avec son utilisateur. C'est le cas notamment des systèmes suivants :

- LUCIABuilder & LUCIASearch développés par Vincent Perlerin<sup>17</sup> qui ont pour objectif d'assister son utilisateur dans l'accès à des documents numériques textuels et dans l'exploration de leur contenu,
- ProxiDocs développé par Thibault Roy<sup>18</sup> pour la cartographie thématique de corpus.

Les systèmes individu-centrés s'opposent aux systèmes dits technocentrés où l'utilisateur n'a principalement qu'un rôle d'interprétation des résultats de la machine sans qu'il puisse déterminer la façon de les produire. Pour qu'un système soit effectivement individu-centré, il faut que ses traitements soient déterminés par l'expression d'un point de vue particulier, celui de son utilisateur, sur une tâche particulière. Il ne s'agit pas simplement que de permettre à l'utilisateur de personnaliser son application, ce qui reviendrait à prévoir d'avance une liste exhaustive de profils d'utilisateurs et de sélectionner l'un de ceux-là en fonction des choix faits. Dans les systèmes individus-centrés, les ressources utilisées sont produites de façon endogène dans la boucle d'interaction système-utilisateur à partir des corpus observés et analysés. Elles n'ont pas forcément l'objectif d'être adéquates à un plus grand nombre et d'être cumulatives car c'est leur spécificité en tant qu'expression de l'utilisateur qui prime.

Dans une approche individu-centrée, le problème de l'évaluation se trouve déplacé. Il ne s'agit pas tant d'évaluer une application que la faisabilité et l'efficacité, du point de vue d'une tâche, d'une interaction entre un utilisateur humain et un agent logiciel. Certaines propositions, en particulier dans le domaine du dialogue Homme-Machine, préconisent déjà des modalités d'évaluation adaptées à des conditions d'interaction entre l'utilisateur et le système (*taux de compétence* et *taux d'efficacité*<sup>19</sup> (Luzzati 1996)). Cependant, l'évaluation d'un système individu-centré ne peut être envisagée

---

<sup>16</sup> <http://users.info.unicaen.fr/~psluquet/Island/>

<sup>17</sup> <http://users.info.unicaen.fr/~perlerin/recherche/these.htm>

<sup>18</sup> <http://users.etu.info.unicaen.fr/~troy/>

<sup>19</sup> Le principe de ces taux d'évaluation est que l'utilisateur peut être amené à corriger les propositions de la machine dans l'interaction. Alors, moins il y a de corrections de l'utilisateur plus l'évaluation sera bonne. Il faut quand même remarquer que cela nécessite de faire vérifier toutes les réponses (aus si bien si l'évaluation est mauvaise ou optimale) ou bien de constituer un corpus d'interaction test pour chaque utilisateur.

sans une réflexion préalable. Le problème que posent les systèmes individus-centrés du point de vue de leur évaluation est principalement un problème de réflexivité. En effet, si c'est le point de vue particulier de l'utilisateur qui est primordial alors l'idée même de constituer un corpus de référence sur le résultat idéal qu'il conviendrait d'obtenir n'est plus tenable. En somme, il n'y aurait personne d'autre que l'utilisateur lui-même qui serait bien placé pour dire ce que devrait être le résultat de son utilisation du logiciel. Le problème n'est donc pas une question de comparaison à un quelconque résultat mais une question d'auto-évaluation, voire d'introspection ou d'un regard réflexif de l'utilisateur sur l'expression de son point de vue relativement à sa tâche et de ce qui en découle dans le système. Dans ces conditions les mesures de rappel, de précision et autres variantes ne sont pas applicables et il convient plutôt de se tourner vers les méthodes d'expérimentation pratiquées dans les sciences humaines : expérimentation avec plusieurs sujets, recueil des résultats et entretiens avec les sujets et enfin analyses.

## 6. Evaluation et instrumentation pour la sémantique des langues.

Comme on l'a vu précédemment avec l'exemple de l'action GRACE, l'évaluation d'un système de TAL est souvent complexe à mettre en œuvre même dans les cas où le ou les systèmes à évaluer visent un résultat sur lequel un consensus peut être trouvé quant à sa qualité (tâche d'étiquetage morpho-syntaxique ou encore de mise en relation syntaxique). Dans ces conditions, on imagine bien qu'évaluer la compétence interprétative d'une machine dans le cadre d'un traitement sémantique est de toute évidence un problème de taille.

La recherche du sens d'une production langagière ne peut pas par nature faire l'objet d'une évaluation en terme de rappel et de précision. Comme on l'a vu précédemment le rappel et la précision sont des métriques issues du champ de la recherche d'informations pour quantifier le succès d'une extraction de certains éléments recherchés parmi plusieurs. La problématique de la recherche du sens n'est pas à notre avis un problème d'extraction à partir du matériau linguistique, ni même de calculs sur une extraction d'informations à partir du matériau linguistique uniquement. Comme le montrent certaines théories linguistiques (la Sémantique Interprétative de (Rastier 1987) notamment) le sens n'est pas construit de manière compositionnelle à partir des constituants de la production langagière comme un mur est construit à partir de briques. Ce ne sont pas tant les caractéristiques propres des mots, des phrases ou des textes qui priment dans la signification des textes mais ce que les utilisateurs en attendent ou y projettent. Des critères externes aux textes peuvent être tout aussi importants. Le succès du moteur de recherche Google<sup>20</sup> nous en donne un exemple à propos du rapport entre le contenu d'une page et l'importance de cette page du point de vue des utilisateurs du moteur de recherche. Comparativement à d'autres moteurs de recherche, le classement par importance des

---

<sup>20</sup> <http://www.google.fr/>

pages Web répondant à une requête ne dépend pas avant tout de leur contenu. Pour AltaVista<sup>21</sup> par exemple, la pertinence d'une page dépend de critères liés à son contenu (présence répétée d'un mot clé de la requête dans le contenu, dans le titre, dans les métadonnées ...). Pour Google, c'est l'algorithme de *PageRank*<sup>22</sup> qui conditionne la pertinence d'une page avec, en plus des techniques classiques donnant une importance particulière à certaines zones (par exemple les titres), le principe suivant<sup>23</sup> : plus il existe de pages qui ont un lien vers la page P, plus P est pertinente (quelle que soit son contenu et quels que soient les mots clés de la requête). Les travaux menés dans le cadre du projet Princip<sup>24</sup> visant la détection automatique de sites Internet au contenu illicite (principalement des propos racistes ou antisémites) montrent également que le sens est de nature pluri-sémio-tique en tant qu'il provient de la présence conjointe de plusieurs facteurs dont certains sont extérieurs aux textes. Ainsi, du point de vue de la thématique des textes, des propos racistes ou anti-racistes sont parfois très proches à tel point qu'une détection automatique fiable uniquement basée sur la recherche de certains mots clés du texte n'est pas facile à obtenir. Si en plus on veut qu'elle soit fiable dans la durée, ça devient très difficile étant donné qu'on ne peut pas prévoir à l'avance l'usage de certains mots dans certains contextes. Il faut alors aller exploiter d'autres critères pour fiabiliser cette détection : la ponctuation qui traditionnellement n'est pas prise en compte (on remarque de façon significative que les sites racistes utilisent fortement le point d'exclamation et que les sites anti-racistes utilisent plutôt des points de suspension), le type de police de caractère utilisé (la police Arial semble significativement caractéristique des sites racistes), les couleurs de fond et de police de caractères utilisées (le rouge et le noir sont aussi significativement caractéristiques des sites racistes), les contenus des images entourant le texte (la thématique de l'animal dans les textes racistes est souvent corrélée avec des dessins montrant des animaux souvent connotés de façon péjorative, le rat par exemple).

Beaucoup de travaux en sémantique formelle (logique, DRT<sup>25</sup>, SDRT<sup>26</sup>...) ont depuis des années déployé beaucoup d'intelligence pour obtenir de façon compositionnelle un « calcul du sens » acceptable (à la manière des résultats des analyseurs morphosyntaxiques). Force est de constater qu'un tel résultat n'est toujours pas atteint à l'heure actuelle. Il ne s'agit pas à notre avis ici que d'un problème d'évaluation dont on n'aurait pas encore bien mis en place la méthodologie mais d'un problème beaucoup plus profond. Dès lors qu'on parle de « vrais » textes et pas simplement de petites phrases d'exemples artificiellement construites en dehors d'un contexte linguistique et pragmatique, il convient de se rendre compte que la dimension interprétative personnelle fait qu'il n'y a pas de consensus évident sur ce qu'est ou n'est pas le sens d'un texte. Il en résulte, à notre avis, que le sens ne peut

---

<sup>21</sup> <http://fr.altavista.com/>

<sup>22</sup> <http://www.google.com/technology/>

<sup>23</sup> Par rapport à d'autres moteurs de recherche, ce principe donne à Google un caractère résolument socio-centré.

<sup>24</sup> <http://www.princip.net/>

<sup>25</sup> Discourse Representation Theory

<sup>26</sup> Segmented Discourse Representation Theory

être modélisé à la façon d'un résultat calculatoire qui serait plus ou moins complété ou dégradé d'un interprétant à un autre. Le sens n'est pas de nature résultative ; c'est un processus sémiotique au centre de l'activité de l'interprétant qui est complexe notamment parce qu'il est réflexif. Le sens d'un texte provient de l'interprétation à un moment donné et dans une tâche donnée d'un sujet interprétant, ce qui est à notre avis un argument fort pour une instrumentation de la sémantique des langues individu-centrée. Dans ces conditions, évaluer un traitement sémantique, ce n'est pas quantifier le résultat d'une extraction ou d'un calcul à partir des propriétés d'un matériau linguistique mais c'est rendre compte de l'efficacité par rapport à une tâche menée par un utilisateur de l'interaction entre le système et cet utilisateur. Pour tester l'efficacité de cette interaction, il convient à notre avis de mener conjointement deux types d'analyses :

1. **Analyser techniquement le système** : est-il capable au cours de l'interaction de produire efficacement à l'attention de l'utilisateur des signes qui éventuellement feront sens pour lui ? Il s'agit ici de questionner la faisabilité algorithmique et ergonomique du système. Par exemple, les méthodes employées (calculs de fréquence, analyses statistiques, modèle vectoriel, ...) donnent-elles des résultats utilisables en un temps raisonnable ? Ces résultats sont-ils présentés à l'utilisateur de manière lisible avec des techniques de visualisation adéquates ?
2. **Analyser la satisfaction des utilisateurs**. Les signes produits par le système font-ils effectivement sens par rapport à différents utilisateurs en fonction de leur propre tâche ? Le rapport entre les productions du système et l'investissement en temps pour la constitution des ressources est-il favorable ? L'utilisation du système rend-elle le travail à plusieurs sur la tâche plus facile ?

Nous soumettons ces propositions à titre de cadre de réflexion générique pour l'évaluation des systèmes dédiés à l'instrumentation de la sémantique des langues naturelles. Elles ne constituent pas bien sûr en soi une méthodologie d'évaluation « clé en main » car, tant en ce qui concerne l'analyse technique que l'analyse de la satisfaction, il convient encore d'adapter les critères d'évaluation spécifiques (en quelque sorte les questions dont il faut chercher les réponses) en fonction du cahier des charges du système et de la tâche visée. Il conviendra également de déterminer de la même façon la fréquence de renouvellement des expérimentations d'évaluation.

## 7. Conclusions :

Au cours de cet article, nous avons voulu montrer en quoi la question de l'évaluation des systèmes de TAL est à l'origine d'un vaste champ de recherche. Il nous apparaît que c'est en tout cas une question qui dépasse de loin le classement de plusieurs systèmes en fonction de leur taux de réussite. Les différentes campagnes d'évaluation existantes en TAL ont beaucoup apporté pour mettre au devant de la scène le problème de l'évaluation. Elles ont aussi permis de produire des ressources importantes pour des travaux de TAL et de linguistique ultérieurs (corpus annotés, jeux d'étiquettes ...). Cependant, il convient maintenant à notre avis de repenser les

méthodes et les enjeux de l'évaluation pour ne plus uniquement l'envisager comme une mesure de performance de plusieurs systèmes mis en compétition.

En remettant notamment l'utilisateur au centre de la question, il nous semble qu'apparaissent des enjeux scientifiques et épistémologiques différents. La place prépondérante du sujet, de l'interaction, de la langue et plus largement de la culture et de la société interroge ici toutes les dimensions des sciences cognitives. L'évaluation considérée comme un processus sémiotique complexe et plus simplement comme un défi technologique propre à l'informatique ouvre un champ de recherche nouveau aux collaborations pluridisciplinaires.

La réflexivité est une propriété au centre de ce champ de recherche tant elle est omniprésente, que ce soit dans l'objet d'étude des systèmes à évaluer (la langue qui se construit et se modifie par elle-même), dans la définition des critères d'évaluation et la mise au point des protocoles (étant donné qu'une expérimentation d'évaluation ouvre en général sur une suivante) ou encore dans l'analyse des interactions homme-machine (étant donné que c'est l'utilisateur lui-même qui est probablement le mieux placé pour analyser son activité). Nul doute dans ces conditions qu'il faudra encore certainement consacrer beaucoup d'investissements en temps et en ressources humaines pour mieux prendre en compte cette réflexivité et en déduire d'autres protocoles d'évaluation ainsi que des méthodes de conception de logiciels nouvelles. C'est ce à quoi nous cherchons à contribuer en poursuivant nos travaux de recherche sur les formes d'instrumentation individus-centrées en TAL.

## Références :

- Adda G., Mariani J., Paroubek P., Rajman M., Lecomte J., 1999, Métrique et premiers resultants de l'évaluation GRACE des étiqueteurs morpho-syntactiques pour le français, actes de TALN'99, p. 15-24.
- Bar-Hillel Y., 1960, *The Present Status of Automatic Translation of Languages*, Advances in Computers, New York et Londres, F.C. Alt ed. Academic Press, vol. 1, 91-141.
- Beust P., 1998, *Contribution à un modèle interactionniste du sens*, Thèse en vue de l'obtention du grade de Docteur en Informatique de l'université de Caen.
- Bourrigault D., Aussenac-Gilles N., 2003, *Construction d'ontologies à partir de textes*, Actes de TALN'2003, vol. 2, p27-47.
- Kerbrat-Orecchioni C., 1988, *Sémantique*. In Encyclopedia Universalis, 693-699. Understanding and creating sentences. American psychologist. Vol. 18, 735-751.
- Lavenus K., Lapalme G., 2002, *Evaluation des systèmes de question réponse*, revue TAL, Vol. 43, n°3/2002, p. 181-208.
- Luzzati D., 1996, Le dialogue verbal homme-machine, études de cas, Editions Masson : Paris.
- Nicolle A., 1996, *L'expérimentation et l'intelligence artificielle*, Intellectica, vol. 96, no 1, pp. 9-19.
- Nicolle A., Saint-Dizier De Almeida V., Beust P., Jacquet D., Brassac C., 2002, *Étude des Processus d'Interaction en Conception Distribuée*, RIHM vol. 4, n°2.

Perlerin V., Beust P., 2003, *Pour une instrumentation informatique du sens*, in *Variation, construction et instrumentation du sens*, sous la direction de Siskou M., Hermès, ISBN 2746207524, pp. 197-229.

Rapport de l'ALPAC, 1966, Language and Machines. *Computers in Translation and Linguistics*. A report by the Automatic Language Processing Advisory Committee (ALPAC), National Academy of Sciences, National Research Council.

Rastier F., 1987, *Sémantique interprétative*, Paris, Presses Universitaires de France.

Rastier F., Cavazza M., Abeillé A., 1994, *Sémantique pour l'Analyse*. Masson : Paris.

Siksou M., Cadiot P., Nyckees V., Mauger S., Manes-Gallo M.-C., Léger A., Thlivitis Th., Rouillard J., Perlerin V., Beust P., Bruillard E., Nicolle A., André B., Kanellos I., Lassègue J., 2003, *Variation, construction et instrumentation du sens*, Hermès, ISBN : 2-7462-0752-4

Sparck Jones K., 2001, *Automatic language and information processing : rethinking evaluation*, in Natural Language Engineering, Cambridge, Cambridge University Press, n°7, p. 29-46.

Thlivitis T., 1998, *Sémantique Interprétative Intertextuelle*. Thèse en vue de l'obtention du grade de Docteur en Informatique de l'université de Rennes I.

Valette M., 2003, Détection et interprétation automatique de contenus illicites et préjudiciables sur Internet. Revue Texto. <http://www.revue-texto.net/>

Van Rijsbergen C.J., 1979, *Information Retrieval*, 2nd edition. University of Glasgow.



# Définition juridique et Argument de la diagonale

*Danièle Bourcier*

CNRS-Centre de recherche en Sciences administratives, Paris

[bourcier@msh-paris.fr](mailto:bourcier@msh-paris.fr)

*Sylviane R. Schwer*

Laboratoire de Linguistique Informatique Cognition et Communication

[Sylviane.Schwer@paris4.sorbonne.fr](mailto:Sylviane.Schwer@paris4.sorbonne.fr)

## Résumé

*L'argument de la diagonale a servi de base à de nombreuses démonstrations mathématiques. Il s'agit en fait d'une formalisation du vieux paradoxe du menteur qui a été l'un des arguments de poids dans la crise des fondements des mathématiques, à savoir tout décrire par la théorie des ensembles. De même, le législateur, quand il rédige une définition en extension ou en compréhension ne peut prétendre lister de façon complète la collection des éléments de la liste qu'il veut décrire. On commencera par décrire plusieurs exemples de réflexivité et d'auto-référence en droit. Puis la méthode de diagonalisation en mathématique sera décrite et confrontée aux travaux sur l'argumentation juridique.*

## Mots-clés

Droit – définition – modèle mathématique – réflexivité - diagonalisation -

## Présentation

La méthode de la diagonalisation a été développée par Cantor qui a démontré par exemple qu'une liste même infinie de nombres réels est nécessairement incomplète. Le même phénomène d'incomplétude se retrouve dans la définition juridique. Le législateur quand il rédige une définition en extension ou en compréhension ne peut prétendre lister de façon complète l'ensemble des éléments de la liste qu'il veut décrire. La conséquence est importante pour l'argumentation juridique car le juge, reprenant la définition du législateur pour qualifier des faits, peut toujours ajouter *sous certaines conditions* l'élément d'une classe qui n'est pas dans la liste.

Nous définirons d'abord quelques phénomènes de réflexivité en droit (1). Puis la méthode de diagonalisation en mathématique sera décrite et confrontée aux travaux sur l'argumentation juridique (2). Nous conclurons en montrant que cette méthode, utilisée dans les deux domaines même si elle ne repose pas sur les mêmes raisons mathématiques met en lumière que cette impossibilité d'énumérer n'est pas de type linguistique ou conceptuel mais logique.

## 1- Quelques exemples de réflexivité et d'auto-référence en droit.

L'auto-référence est une question traditionnelle en droit que nous observerons de différents points de vue, celui du législateur, celui du juge et celui du modélisateur (« le juge connexionniste »).

1- D'abord le droit s'auto-référence de façon explicite et continue. Toute nouvelle règle est encadrée par une méta-règle (M) qui légitime procéduralement la règle nouvelle. Ensuite cette nouvelle règle (N) cite cette méta-règle, indique qu'elle abroge ou modifie la règle ancienne (A). Autrement dit toute nouvelle règle doit s'insérer dans le cadre des références juridiques existantes et à jour. Parfois il y a des contradictions ou des erreurs. Des systèmes informatiques simulent désormais la projection de la mise à jour pour observer les impacts de N sur A (MAGICODE)<sup>1</sup>.

Certaines règles de niveau supérieur comme la directive européenne ont pour objectif d'être « transposées » dans un système juridique national. Il faut donc transformer A en insérant N, ou en abrogeant A et en lui substituant N (c'est plus simple).

Au niveau de l'écriture des textes, cette auto-référence se manifeste par des « références croisées » explicites, des liens transversaux ou verticaux (les cours et tribunaux sont organisés de façon pyramidale et les textes normatifs, de la Constitution à l'arrêté municipal aussi). Le droit s'auto-cite suivant des écritures relativement normalisées, ce qui signifie qu'on pourrait en partie opérer une mise à jour *dynamique*.

Il existe deux niveaux de références : des références internes à un sous-ensemble juridique (dans un code, dans une loi) et des références à un ensemble juridique plus large (un système juridique national, ou supranational).

La question d'un univers juridique unique et fermé n'est plus défendable. Y en a-t-il jamais existé un ? Ne peut-on pas considérer que le maire écrivant ses arrêtés éventuellement contraires à la norme supérieure agit dans son propre système municipal ? Auquel cas le système normatif national n'est pas unique et fermé. Mais on peut également considérer l'inverse. Le système 'maire' est intégré dans le système national (donc celui-ci reste unique) indépendamment de la légalité des arrêtés produits par le maire.

En étendant l'analogie, on pourrait supposer que le système normatif global reste unique, voire fermé (le phénomène juridique), mais qu'il évolue en fonction de l'évolution de ses composantes (les sous-systèmes nationaux). Donc ce n'est qu'une question de point de vue :

- soit micro : différents systèmes communiquent, se heurtent et disposent d'interfaces de communication ;
- soit macro : ces sous-systèmes ne sont que les composantes d'un système normatif unique qui 'phagocyte', intègre ou s'enrichit d'autres systèmes dès qu'ils sont mis en contact.

2- Mais là où la question se complique c'est lorsque le système de référence n'est pas bien délimité (ouvert) ou qu'il existe plusieurs systèmes de référence en concurrence ou à rapprocher. C'est ce qu'on appelle le pluralisme des sources juridiques ou polynormativité. Certains auteurs considèrent même que les règles juridiques sont en concurrence avec des règles techniques<sup>2</sup>. Actuellement il n'existe plus de système normatif -au sens strict- unique et fermé. Il existe plusieurs systèmes qui interfèrent (des « babylones juridiques » suivant le mot de J. Carbonnier). Au niveau européen par exemple, la cour de Justice peut avoir à décider en

<sup>1</sup> D. Bourcier, « Codification et légistique à l'heure du numérique », in *L'avenir de la codification en France et en Amérique latine, Actes*, Paris, Les Colloques du Sénat, 2004 pp. 176-181 : MAGICODE est une boîte à outils légistique pour transformer des textes juridiques existants en textes codifiés.

<sup>2</sup> L. Lessig, *Code and other laws of cyberspace*, New York, Basic books, 1999. Pour lui, quatre systèmes de contraintes déterminent les comportements sur internet : le marché, les normes sociales, l'architecture du réseau et, en dernier, le droit (p. 165).

fonction de plusieurs systèmes de normes. Il lui faut donc trouver une norme moyenne ou acceptable dans les divers systèmes nationaux.

Luhmann a défini comment divers systèmes sociaux (le droit, l'économie, le politique,...) pouvaient interférer entre eux *à la frontière*. Mais en général chaque système ne peut être défini comme tel que s'il est *relativement clos*. Il considère ainsi que tout système juridique peut être considéré comme « *normativement clos mais cognitivement ouvert* »<sup>3</sup>.

Le « *Cognitivement ouvert* » est la partie du droit la plus intéressante à étudier : en réalité il s'agit de voir comment dans un système clos, un changement est possible et comment les systèmes juridiques bien que clos, autoréférentiels, peuvent évoluer *dynamiquement*. En admettant que le système soit autoréférentiel et clos on pourrait supposer qu'il est stable. En fait il est attaqué à sa frontière (si on peut séparer le système normatif des systèmes sociaux l'influencant) et il doit adapter sa structure interne (concepts, catégories) afin de résister (de survivre) en tant que système.

Si l'on part du 'langage' du droit, on s'oblige à inventer une forme de traduction/translation permettant de gérer les interférences entre les différents systèmes de droit (références, renvois, visas). Si l'on part d'un méta-niveau (à identifier) permettant d'englober les différents langages, on crée de fait un langage permettant de faire communiquer les différents systèmes juridiques. A l'inverse, et c'est peut-être le point le plus intéressant : en construisant une "théorie de l'interface" entre systèmes juridiques, on est probablement sur la piste d'une méta-théorie du droit.

3- Pour effectuer ce travail d'*auto-référenciation dynamique*, le droit utilise plusieurs techniques : la définition, la classification. Il crée des objets « *artificiels* » que sont des concepts ou des catégories. Le juriste utilisera alors ces objets dans une série de raisonnements mathématiques (logiques modales ou même analogies), gérant des ensembles flous ou indéterminés. Si le raisonnement est « *faible* » (comme l'analogie), il a recours à l'argumentation qui n'est pas un raisonnement consistant mais fixe les limites conceptuelles interactives de l'acceptable, du raisonnable, du négociable.

D'ailleurs, certaines composantes définies par cette méta-théorie sont peut-être les *lieux communs*, *les briques fondamentales* des différents systèmes juridiques. Certains sont cités dans ce paragraphe : définition, catégorisation, objets conceptuels. Un point également intéressant est que ces éléments fondamentaux correspondent en langage informatique aux notions de variable, de fonction, d'objet, de classe et d'appartenance (qui est une fonction). Ce qui renvoie peut-être à un méta-niveau et qui montre que le droit est formalisable en langage informatique. Les travaux en Intelligence artificielle et droit l'ont montré d'ailleurs depuis une vingtaine d'années.

4- Plusieurs raisonnements portant sur des concepts ou des catégories peuvent être cités en exemple.

a) Un jour un juge fut contraint de raisonner à partir d'une catégorisation illégale dans un rapport de gendarmerie. Ce procès verbal avait rapporté qu'un certain monsieur X, qui avait été pris en train de voler des grenouilles dans un étang, avait commis un « *vol de grenouilles* ». Or cette qualification n'existe pas en droit pénal : le juge ne dispose que de la catégorie d'infraction pénale « *vol de poissons* ». Qu'à cela ne tienne, le juge fit un raisonnement qui lui permit de dire que les grenouilles étaient des poissons<sup>4</sup>. Comment réussit-il ce tour de magie? Il créa la catégorie, *fictive et fonctionnelle*, suivante : un poisson est 1) comestible et 2) susceptible d'appropriation. Or une grenouille a les mêmes caractéristiques. Donc une grenouille EST un poisson. Le tour est joué.

<sup>3</sup> D. Bourcier & P. Mackay (eds) *Lire le droit*, Paris, LGDJ, 1993, Introduction.

<sup>4</sup> D. Bourcier, « *Argumentation et définition en droit ou Les grenouilles sont-elles des poissons ?* » in revue *Langages*, n° 42, juin pp. 115-124

b) Un autre jour, le juge d'appel devait caractériser la marque d'un parfum comme une œuvre de l'esprit, qui à ce titre devenait protégeable par le droit d'auteur. Mais la catégorie des œuvres de l'esprit (loi de 1957) ne contient pas le parfum. En revanche elle contient une énumération d'« œuvres » comme la musique, les sermons, ou le livre. Musique et livre sont perceptibles par l'ouïe et la vue, et font donc appel à des « sens » (perceptions sensorielles) particuliers. Le juge a alors créé la catégorie intermédiaire : les œuvres protégées sont des œuvres perceptibles par les sens. Un parfum fait appel à l'odorat, donc il peut être protégeable comme œuvre de l'esprit.

En raisonnant par analogie et en ne s'intéressant qu'aux propriétés communes, on peut assimiler un poisson à une grenouille. Mais si l'on créé une catégorie et qu'on affecte l'exécution d'une norme à cette catégorie alors la norme affecte ses éléments par héritage. Les deux procédés sont différents. Et dans le deuxième, les grenouilles peuvent être affectées par la norme sans devenir des poissons.

5- La jurisprudence crée ainsi des catégories qui permettent au juge d'intégrer des éléments de fait ou de droit qui ne pourraient pas être classés dans les catégories existantes. Dans d'autres cas, les catégories et concepts perdurent et sont répertoriés comme une avancée « prétorienne » (c'est-à-dire créés de toute pièce par le juge et acquérant ainsi un statut juridique). La notion de « traditions » par exemple permet au juge en matière de police municipale d'accepter des exceptions au droit commun. La vente du muguet le premier mai sur le domaine public et sans taxes ou redevances liées à l'occupation privative du domaine public en est un exemple. On assiste donc à une émergence incrémentale de concepts issus des raisonnements des juges.

Ces concepts peuvent être utilisés pendant de nombreuses années puis finalement devenir inadaptés ou contradictoires. Il faut alors « changer le concept » et opérer « un revirement de jurisprudence ». Le terme est consacré et la doctrine ne manque pas de noter les décisions où le juge a effectué ce type de raisonnement.

Prenons l'exemple suivant. Pendant des années, la Cour de cassation n'a reconnu que le système de la *caution*, ce qui était très lourd car obligatoirement ce double contrat nécessitait pour le créancier de passer par le premier contractant (le défaillant) pour engager des poursuites vis-à-vis de la caution (la banque par exemple). En 1982, La Cour a créé « la garantie à première demande » qui permet au créancier de saisir directement la caution en cas de silence du défaillant. Cette catégorie a permis de trouver un chemin complètement nouveau pour éviter de « raisonner en rond ».

Le revirement de jurisprudence correspond à un changement brusque et complet. Mais s'agit-il de la façon de juger ou des décisions rendues ? Dans le premier cas, deux problèmes similaires sont analysés suivant deux raisonnements différents mais peuvent toujours avoir la même conclusion. Dans le deuxième cas, et pour qu'il y ait revirement, deux problèmes similaires doivent obligatoirement avoir deux conclusions différentes suivant que l'on se situe avant ou après le revirement.

Dans le premier cas, le revirement peut passer inaperçu, dans le second il paraît illogique.

6- La question se pose désormais d'utiliser les instruments de l'intelligence artificielle pour analyser si des raisonnements ne pourraient être élaborés de façon plus efficace par une machine. Il faut cependant se rendre compte que ce raisonnement doit s'appuyer sur des éléments de droit (arguments ou moyens juridiques) pour être justifiables. Le juge en effet est contraint par la motivation de ses décisions, ce qui est une *autre forme d'auto-référence*. Or les RNA ne sont pas explicites sur les catégories utilisées. C'est un handicap réel pour modéliser le raisonnement du juge.

Une expérience d'apprentissage de résolution de cas de jurisprudence<sup>5</sup> a été menée à partir d'un réseau de neurones artificiels (RNA). Il s'agissait de *décomposer* le mode de représentation du cas par le modèle (le cas était « entré » dans la boîte noire) et de le comparer avec la représentation explicite du juge spécialisé dans ce genre d'affaires. L'expérience a montré que la façon de résoudre le cas pouvait être presque similaire à celle du juge mais que dans certains cas, le réseau pouvait trouver un « chemin de résolution » plus performant. En effet les neurones cachés peuvent se spécialiser dans certains aspects de l'affaire et distinguer des « sortes » de catégories (des types comme on le verra plus loin ?) en se regroupant par couche de neurones. On a pu aussi remarquer qu'un nombre minimal de catégories peuvent être nécessaires pour « faire résoudre » un problème par un RNA (3 catégories en l'occurrence). Pourtant le juge avait besoin d'utiliser 4 catégories pour résoudre le même problème. Grâce à ces modèles et de façon inattendue, le juge peut donc acquérir une vue « critique » sur sa façon de raisonner : le RNA a montré dans notre expérience une capacité d'abstraction supérieure à celle d'un être humain.

L'expérience décrite (invention d'un nouveau chemin décisionnel par un RNA) correspond au premier cas. Un nouveau mode de résolution est inventé/construit mais aucune conclusion n'est modifiée. Les changements n'étant pas visibles, on peut supposer qu'il n'y a pas réellement de revirement. Mais d'un autre côté, et c'est justement le cas pour un RNA, le nouveau mode de résolution aura certainement un impact visible sur de nouveaux cas non prévus initialement et non rencontrés par les deux représentations. Dès lors le revirement est effectué mais ses symptômes ne sont pas encore visibles (revirement invisible ou caché). Pour le moment, il semble qu'on ne parle en droit que des revirements visibles.

Quelle que soit la méthode, lorsqu'on décompose le processus décisionnel du juge, on s'aperçoit qu'il existe deux façons d'opérer de tels changements : altérer la liste des éléments pris en compte au cours du raisonnement ou altérer l'influence de ces critères sur le raisonnement. Le problème est que ces altérations sont susceptibles de provoquer des différences de jugement, que l'on pourrait qualifier d'évolution ou d'enrichissement, avant de parler de revirement. Il y a un problème de 'degré à définir', de 'saut' dans une courbe représentant l'évolution d'un mode de raisonnement (il faudrait d'ailleurs trouver un moyen de représenter graphiquement l'évolution du mode de résolution d'un problème juridique en se fondant sur l'évolution des critères et de leur importance).

Il apparaît donc que les juges travaillent sur des collections qui ne sont pas *exactement* définies, ni sous forme explicite, ni sous forme implicite. Ce qui est complètement contraire aux mathématiciens qui, eux, oeuvrent dans des ensembles parfaitement définis. Donc il existe des collections qui ne sont pas représentables par des ensembles, ce qu'on peut montrer aussi par un raisonnement du type de la diagonalisation que l'on va exposer maintenant.

## 2- Présentation de la méthode de diagonalisation

Cette méthode, appelée aussi argument de la diagonale, est due à Georg Cantor (1845-1918) et se fonde sur la construction d'une véritable *diagonale* géométrique. Elle a servi de base à de très nombreuses démonstrations mathématiques. Cantor voulait prouver que toute énumération des nombres réels (toute liste - même infinie - indexée par des entiers) était nécessairement incomplète, c'est-à-dire que la notion de liste complète des nombres réels était en soi contradictoire.

Ce résultat prouve qu'il y a au moins deux types d'infinis : l'un décrit par les nombres entiers, qui permettent d'énumérer la liste complète des objets de ce type, l'autre décrit par les nombres réels - ou le nombre de points d'un segment de droite.

<sup>5</sup> F. Borges, R. Borges, D. Bourcier « Artificial Neural Networks and legal categorization », *JURIX 2003*, Amsterdam, Ios Press, 2003 pp. 11-20.

Décrivons la preuve de la non-énumérabilité de l'ensemble des nombres réels  $\mathbf{R}$  par Cantor. Il suffit de prouver la non énumérabilité des réels compris entre 0 et 1. Supposons qu'une liste infinie des réels entre 0 et 1 puisse être créée  $r=\{ r(1), r(2), r(3), \dots \}$ . Chaque réel  $r(n)$  compris entre 0 et 1 possède un développement décimal illimité  $r(n)=r(n,1)r(n,2) \dots r(n,3)r(n,4)\dots$ . On peut donc décrire  $\mathbf{R}$  sous la forme d'un tableau illimité dans ses deux dimensions :

$r(1) = 0,$	<b>r (1,1)</b>	$r(1,2)$	$r(1,3)$	$r(1,4)$	...
$r(2) = 0,$	$r(2,1)$	<b>r (2,2)</b>	$r(2,3)$	$r(2,4)$	...
$r(3) = 0,$	$r(3,1)$	$r(3,2)$	<b>r (3,3)</b>	$r(3,4)$	...
$r(4) = 0,$	$r(4,1)$	$r(4,2)$	$r(4,3)$	<b>r (4,4)</b>	...
...	...	...	...	...	...

Il suffit alors d'utiliser la liste des chiffres de la diagonale, c'est-à-dire  $r(1,1), r(2,2), r(3,3), r(4,4), \dots$  pour construire un nombre réel  $r=0,r(1)r(2)r(3)r(4)\dots$  compris entre 0 et 1 qui n'est pas dans la liste. Il y a de nombreuses façons de réaliser ce choix, par exemple, en décidant que  $r(n)=1$  si  $r(n, n) = 0$  et  $r(n)=0$  dans tous les autres cas.

Par exemple, si la liste commence ainsi

$r(1) = 0,$	<b>1</b>	0	1	1	...
$r(2) = 0,$	0	<b>1</b>	1	1	...
$r(3) = 0,$	0	0	<b>0</b>	1	...
$r(4) = 0,$	0	0	1	<b>0</b>	...
...	...	...	...	...	...

on a  $r=0,0011\dots$  qui est bien un réel compris entre 0 et 1 mais il n'est notoirement pas dans la liste. On peut certes l'insérer dans la liste, mais on trouvera alors, avec le même procédé de diagonalisation, un autre réel hors de la liste. Il est donc impossible d'établir une liste de tous les réels.

L'essence de cette méthode vient de l'usage du même entier dans deux contextes différents : l'un en index vertical et l'autre en index horizontal. Dans la preuve de Cantor, l'entier  $n$  décrit en index horizontal un réel et en index vertical le  $n^{\text{ème}}$  chiffre du développement décimal illimité de ce réel. Grâce à ce double emploi astucieux, on construit un objet qui n'appartient pas à la liste en question. L'astuce est d'utiliser la même chose (l'entier  $n$ ) pour coder deux objets de nature complètement différente, un réel et un chiffre de ce réel, l'un des objets étant une partie de l'autre.

La liste est supposée décrire une collection d'objets *in extenso*. La question est donc du statut de l'objet : ne fait-il pas partie de la liste parce qu'il n'a pas les qualités requises pour être dans la collection correspondante ou la collection est-elle trop riche pour être énumérée ?

Ce sont ces mêmes questions qui se posent au juge dans le cas de la grenouille.

Cantor utilisa la diagonalisation pour montrer par l'absurde que certains ensembles ne sont pas énumérables. La richesse du procédé de diagonalisation est qu'il peut être aussi utilisé pour prouver qu'un objet ne fait pas partie d'un certain ensemble énumérable, donc qu'il n'a pas

toutes les qualités requises pour être dans cet ensemble. C'est ainsi que l'on montre l'indécidabilité de la terminaison des programmes : il n'existe pas de programme *Prog*, qui détermine, pour chaque programme '*P*' et pour chaque donnée '*d*', si le programme '*P*' va traiter '*d*' en un temps fini. C'est par une méthode similaire que Gödel démontra son théorème d'incomplétude.

Dans la preuve de l'indécidabilité de la terminaison des programmes, on part du fait connu que l'ensemble de tous les programmes que l'on peut écrire dans un langage de programmation universel est énumérable, comme est aussi énumérable l'ensemble des données.

En effet, ces ensembles sont des chaînes de caractères finies. Remarquons qu'un programme peut être une donnée pour un autre programme.

On construit le tableau 'T' suivant : la cellule 'T(p,q)' contient 1 si le résultat de la fonction exécutée par le  $p^{\text{ème}}$  programme sur la  $q^{\text{ème}}$  donnée se calcule en un temps fini, sinon 0. Ce tableau permet donc de construire un programme qui n'est pas représenté dans ce tableau. Le programme *Prog* est le suivant : il donne pour la donnée '*d*' la valeur 0 si le  $d^{\text{ème}}$  programme donne une valeur pour la  $d^{\text{ème}}$  donnée, et 1 sinon. C'est-à-dire que  $\text{Prog}(d)=1$  si ' $T(d,d)=0$ ' et ' $T(d,d)=1$ ' si ' $T(d,d)=1$ '. Ce programme ne peut donc être placé dans la liste des programmes puisqu'il contredit systématiquement les valeurs de la diagonale. Ce programme n'existe donc pas.

Cette conclusion découle du fait que l'ensemble de tous les programmes possibles, bien que non énumérable en fait est un sous-ensemble d'un ensemble parfaitement connu, descriptible en une expression finie, à savoir l'ensemble de toutes les chaînes de caractères finies du langage dans lequel on a décidé de coder les programmes (Pascal, C, Cobol, ...)<sup>6</sup>.

### 3- Analyse et discussion de l'argument de la diagonale pour le droit

La diagonalisation peut être une représentation adéquate de la construction ou de la vérification de catégories en droit. Ainsi on peut voir comment un nouvel élément peut être en contradiction, non vis-à-vis de lui-même mais avec tous les autres. Il est alors possible d'élaborer une métro-auto-référence qui vérifie si cet élément peut être intégré dans le groupe. En regardant les exemples de la grenouille et du parfum, on voit que l'on a affaire à des problèmes de pertinence de définitions (sociales *versus* scientifiques).

Il s'agit alors de trouver la surclasse qui contient à la fois la grenouille et les poissons, en ne retenant que les attributs communs et en continuant d'appeler cette classe POISSONS. Mais cette surclasse est éphémère : elle sert à régler un cas et disparaît avec sa résolution. Que pourrait en faire le juge dans un cas ultérieur ? Comme le droit français n'est pas fondé sur le précédent et qu'un tribunal ne peut proposer de tirer d'un raisonnement une règle générale, le juge, dans une affaire ultérieure, ne pourrait utiliser la production de cette classe que comme argument et non comme moyen. Précisons que le moyen est un élément de raisonnement, l'argument un élément de conviction, suivant la distinction traditionnelle faite par la Cour de Cassation.

En mathématique, une liste est supposée décrire une collection d'objets *in extenso*. Pourtant on peut montrer par l'absurde que certains ensembles ne sont pas énumérables. La question est alors de savoir quel est le statut de cet objet qui n'appartient pas à la liste : ne fait-il pas partie de la liste parce qu'il n'a pas les qualités requises pour être dans la collection correspondante, ou la collection est-elle trop riche pour être énumérée ?

En droit, le législateur donne de nombreuses définitions mais le juge est souvent obligé d'étendre la liste des énumérations pour traiter un nouveau cas. On pose l'hypothèse que l'on peut utiliser le même procédé de la diagonalisation en droit qu'en mathématique non pour

<sup>6</sup> Jean-Michel Autebert, *Calculabilité et décidabilité, une introduction*, Masson, 1992.

prouver la non dénombrabilité des éléments d'une liste mais pour montrer que le droit est logiquement incomplet.

L'extension des définitions juridiques est difficilement calculable puisque ces définitions sont élaborées pour régler dans le futur des questions traitées aujourd'hui, sans réelle connaissance de l'évolution de la société. On ne sait donc pas réellement sur quels objets futurs elle va s'appliquer. Autrement dit, les définitions en droit ne sont pas closes *a priori* comme en mathématique.

On observera aussi que le droit utilise la langue naturelle et non un formalisme rigoureux comme les mathématiques. Pour cette première raison, le codage par la langue ne permet pas de clore une définition.

Mais allons plus loin dans la comparaison.

En mathématique la définition est sans intention : elle est explicite. Pour parler d'ensemble, il faut donner une propriété explicite qui regroupe tous les éléments dans la liste. En droit en revanche, il n'existe pas de définition complètement explicite notamment dans les univers « cognitivement ouverts » c'est-à-dire ceux qui contiennent un élément flou ou indéterminé (dont la définition implicite dépendra du contexte). Une grenouille peut être assimilée à un poisson dans un contexte A et pas dans un contexte B.

Ainsi une définition juridique a toujours un sens implicite même si explicitement l'objet en cause n'en fait pas partie. Autrement dit, il existe une définition explicite (vol de poissons) et une définition implicite non repérable par une machine, qu'un juge pourra utiliser dans un contexte donné (il n'y a pas de contexte dans une définition).

En Informatique, on sait qu'il existe des fonctions qui ne sont pas calculables. En droit, on ne peut tout expliciter : la méthode du juge va être de se mettre en accord avec la définition implicite, dans chaque contexte particulier.

Autrement dit, si tout était explicite et calculable, nous n'aurions besoin ni du juge ni de l'informaticien.

## Conclusion

La diagonalisation utilise la réflexivité sur un ensemble global ou une classe mais pas sur un objet propre. On a vu que ce procédé consiste essentiellement à prouver qu'un ensemble infini, dont les éléments se définissent par une suite infinie de termes, n'est pas énumérable, en construisant un élément qui possède toutes les propriétés communes aux éléments de l'ensemble mais qu'on ne peut identifier à aucun des éléments de l'ensemble.

En droit, les collections manipulées ne sont pas des ensembles mais relèvent d'une théorie du typage c'est-à-dire de collections qui sont dépendantes du contexte et reliées par une intention.

Il semble donc qu'en mathématique, l'autoréférence est quelque peu « destructrice » comme créant les paradoxes, alors qu'en droit, elle est « créatrice ».

---

### Note

Douglas Hofstadter, *Gödel, Escher, Bach Les Brins d'une Guirlande Eternelle*, Dunod, 1985. Voir aussi (suivant les indications d'Anne Nicolle) pour l'autoréférence en droit : "Ma Thémagie", ch 4 : « le Nomic, un jeu automodificateur fondé sur la réflexivité du droit », *Ma Thémagie*, Interéditions, Paris, 1988 (pour l'édition française), du même auteur.

# Double visage du Menteur

Jérôme Cardot

LaLICC, UMR 8139 - CNRS - Université Paris IV Sorbonne

<http://www.lalic.paris4.sorbonne.fr>

*Jerome.Cardot@paris4.sorbonne.fr*

---

## Résumé

Notre parcours à travers les figures de l'autoréférence rencontre le domaine du langage, pour explorer comment la sémantique tente de s'affranchir du monde des paradoxes ; celui des programmes informatiques, où l'autoréférence est à la base d'un principe fondamental d'organisation, la récursivité. Celle-ci étant liée à toute itération de transformations, nous envisageons encore des transformations géométriques permettant de construire des objets récursivement.

Ces trois domaines nous conduisent à mettre en évidence la nécessité d'un double regard sur les phénomènes réflexifs, permettant l'expression de paradoxes, ou l'interdisant.

**Mots clés :** Autoréférence, langue naturelle, point fixe, récursivité.

---

Quel est le premier, de la poule ou de l'œuf, de l'œuf ou de la poule ?

Si l'autoréférence est source classique de paradoxes tels que le paradoxe du menteur, elle peut être directe ou indirecte, médiatisée ou immédiate, directe quand Épiménide affirme « *Épiménide est menteur* », indirecte quand Platon dit « *Socrate dit la vérité* » alors que Socrate affirme « *Platon est menteur* ».

Immédiate ? Pas si sûr : elle cesse rapidement de l'être dans les théories typées et dans les processus temporels. Le menteur, lorsqu'il dit « *je mens* » est-il encore celui dont il dit qu'il ment ?

Si l'analyse de l'autoréférence appelle la vue d'un point fixe, celui-ci est visé par le biais d'une fonction monotone, sans quoi on peut tourner autour, attiré tantôt par l'œuf, tantôt par la poule.

Nous nous intéresserons donc à l'autoréférence dans les langages formels, où on l'empêche d'être directe pour sauver lesdits langages, et dans l'organisation des calculs informatiques, à travers la récursivité, et dans la construction de figures géométriques.

Et nous proposerons d'unifier la résolution des formes du paradoxe dans ces différents domaines, en y voyant l'introduction d'une différence entre le même et le même ; nous insisterons enfin sur la nécessité de conserver un double point de vue sur le paradoxe, à l'aide de deux représentations, l'une laissant le paradoxe visible, l'autre le résolvant.

## 1 Autoréférence et langage.

L'existence des paradoxes que l'on vient de rappeler n'est possible que parce que le langage contient, en puissance, l'autoréférence. Celle-ci se manifeste même d'au moins deux façons, à travers l'*acteur* du discours, le locuteur, ou à travers le *medium*, le langage lui-même. Et les tentatives de résolution des paradoxes pour les langages formels ne visent rien moins que limiter la puissance de ces langages – certes pour la bonne cause, pour leur donner de *bonnes* propriétés.

### 1.1 Autoréférence du locuteur.

Si la conscience de l'altérité est première, et rend nécessaire le langage pour communiquer avec autrui, la conscience de soi la suit de très près, dans un mouvement dialectique, et déjà réflexif, immédiat. Le langage permet donc de dire « *je* », il le permet même si bien qu'on peut entendre « *je* » même dans un discours où il n'est pas.

Ainsi le langage permet de parler de soi, d'entendre parler de soi : « *parlez-moi de moi, il n'y a que ça qui m'intéresse* » ou, symétriquement – et déjà en miroir –, « *ce qui n'intéresse que moi ne m'intéresse pas* ».

Et le langage est riche de ces pronoms, *moi, je, me*, de ces adverbes réflexifs, *ici, maintenant* qui placent – forcément – le locuteur au centre de tout discours, ce qui est après tout bien naturel puisque le langage doit bien être de quelque utilité directe au locuteur.

On rejoint ici les problèmes de la *deixis*, ou *indexicalité*.<sup>1</sup>

### 1.2 Autoréférence du langage.

Plus intéressante est la capacité du langage lui-même d'accomplir le même retour réflexif, de parler de soi, sans recours à un outil extérieur.<sup>2</sup> Et c'est avec le langage même qu'on exprime la grammaire et les analyses sur le langage, quels qu'en soient la complexité ou le niveau d'abstraction. « *La métalangue est dans la langue.* »

---

<sup>1</sup>(Perry 1999).

<sup>2</sup>C'est ce que (Authier-Revuz, Doury, and Reboul-Touré 2003) appelle « *le fait autonymique* ».

C'est là la source de tout paradoxe langagier, et de toute l'incomplétude des langages formalisés : Gödel a montré que l'arithmétique déjà est un langage suffisamment puissant pour être son propre métalangage, et pour former, en tant que son propre métalangage une formule autoréférentielle indécidable,<sup>3</sup> analogue au paradoxe du menteur, à la nuance près qu'il n'y est pas question de *vérité* mais de *dérivabilité*.<sup>4</sup>

### 1.3 Tentative de résolution.

Si les paradoxes autoréférentiels naissent de l'usage du langage comme métalangage, on cherchera à les éviter en interdisant cet usage ; plus précisément, Ladrière distingue entre paradoxes syntaxiques et paradoxes sémantiques : la théorie des types se présente comme un moyen d'interdire syntaxiquement l'autoréférence individuelle. Un objet ne peut se contenir lui-même, une proposition ne peut parler d'elle-même...

Les objets (de type 0) peuvent être sujets de prédicats de type 1, eux-mêmes sujets de prédicats de type 2... et la contrainte syntaxique interdit l'autoréférence.

Cependant l'autoréférence surgit à nouveau par le sémantique. Si l'interprétation sémantique nécessite, avec Tarski, une hiérarchie de langages qui différencient chaque niveau de langage de son métalangage, il n'en reste pas moins que la proposition de Gödel, dont l'interprétation métalangagière fait appel à l'autoréférence, est indécidable au sein même du langage initial (celui de l'arithmétique récursive).

### 1.4 Dialogue.

Retrouvons ici la langue naturelle et son objet principal, la situation de dialogue. Naturellement, un dialogue s'inscrit dans le temps. Par exemple :

- *A* dit à *B* ce qu'il souhaite lui dire ;
- *B* réagit en répondant (favorablement ou non) à la requête exprimée ;
- si *B* exauce le désir de *A*, cet exaucement transforme un *A* insatisfait en un *A* satisfait, et c'est *A* lui-même, par sa demande initiale, qui est à l'origine de cette transformation.

Les deux états de *A* peuvent très bien ne différer que par l'enrichissement apporté par le dialogue, par exemple lorsque *B* répond à une question de *A* en lui apportant une information que *A* ignorait.

Ainsi, une action de *A* est à l'origine d'une modification future de *A* (et heureusement, cette situation est possible et même fréquente !).

C'est ici le temps qui organise la différenciation entre *A* actuel, *A* futur, ou *A* passé, l'équivalent de la différenciation entre les différents langages dans la sémantique de Tarski.

---

<sup>3</sup>C'est-à-dire telle que ni la formule, ni sa négation n'est prouvable – on dit plutôt *dérivable* – dans le système.

<sup>4</sup>Cf. (Gödel 1931) et les commentaires de (Ladrière 1957), (Wagner 1998). La formule dont il est question affirme, à l'intérieur du langage de l'arithmétique, qu'elle n'est pas démontrable.

## 2 Autoréférence dans les processus.

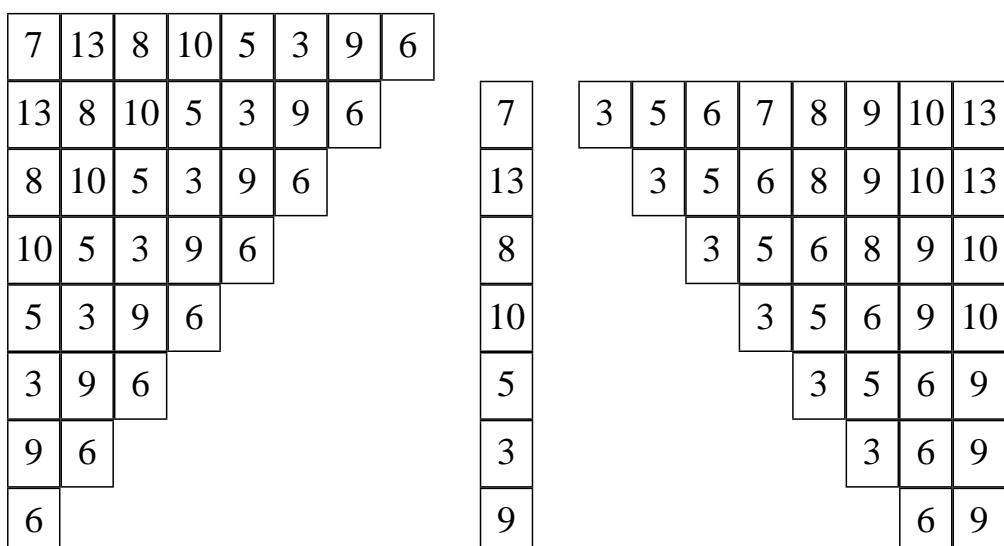
Aussi bien dans la vie courante qu'en ce qui concerne les processus calculatoires, le principe de ramener un problème donné à un problème analogue, plus simple, est une méthode courante. Par exemple, ranger un bureau (la pièce), c'est ranger les différents meubles, et pour chacun d'entre eux, c'est ranger les différents tiroirs et surfaces planes...

## 2.1 Processus récursifs.

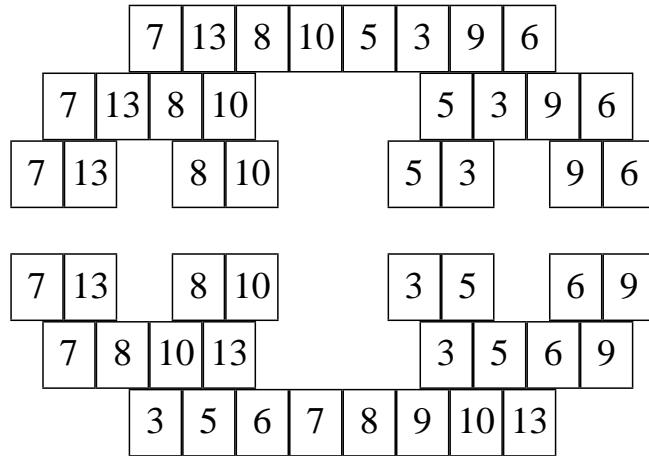
Dans les processus calculatoires, ce même principe s'organise avec des définitions récursives, qui ramènent progressivement le problème à résoudre à un problème élémentaire facile à traiter, moyennant une fonction d'itération, qui permet le passage d'un cas traitable à un cas un peu plus compliqué.

Les problèmes de tri constituent un exemple privilégié de problèmes récursifs, d'une part parce qu'on imagine aisément différentes organisations possibles de cette tâche, d'autre part parce que la résolution pratique de nombreux problèmes passe par une phase préalable de tri des données – qui peut être la phase la plus longue du calcul, d'où l'importance d'effectuer le tri aussi efficacement que possible. Comparons ainsi les méthodes suivantes :

1. trier une liste de  $n$  objets, c'est écarter l'un d'entre eux, trier les  $n - 1$  autres, puis mettre celui qui a été écarté à la bonne place ;



2. trier une liste de  $n$  objets, c'est la séparer en deux listes de même taille (ou peu s'en faut)  $\frac{n}{2}$ , trier séparément chacune des deux listes, puis les rassembler en une liste triée :



Pour ces deux méthodes, le cas simple auquel on cherche à se ramener est la liste formée d'un seul élément, déjà triée.

Si les deux méthodes sont récursives, la seconde est plus efficace que la première ; en effet, elle permet d'atteindre plus rapidement le cas simple visé : chaque itération divise par deux la taille des tableaux à trier, alors que dans la première méthode la taille du tableau n'est diminuée que d'une unité. En outre, les complexités des itérations de chaque méthode sont voisines (si bien qu'on ne regagne pas, dans la première méthode, par des itérations plus simples, ce que l'on perd par un plus grand nombre d'itérations).

Revenons alors sur la description de la deuxième méthode : « *trier une liste de  $n$  objets, c'est la séparer en deux listes de même taille  $\frac{n}{2}$ , trier séparément chacune des deux listes, puis les rassembler en une liste triée* ».

Assurément, et c'est là le principe même de la récursivité, il y a un lien très intime entre « *trier une liste de  $n$  objets* » et « *trier une liste de  $\frac{n}{2}$  objets* » ; la tâche à accomplir est la même... mais ce n'est pas tout-à-fait la même ! La preuve en est que la démarche présentée est effective, que le calcul termine, et il termine parce qu'à force de diminuer la taille du tableau à trier on finit par atteindre le cas le plus simple, qui se traite autrement.

En l'absence de ce « *cas le plus simple* » que l'on ne cherche plus à diviser, de ce *tableau atomique*, on devrait poursuivre la division, exactement comme on le fait dans le paradoxe de Zénon, sans pouvoir envisager de résolution par passage à la limite.

On a donc besoin, pour appréhender pleinement la démarche récursive, à la fois d'accepter son caractère réflexif, et de distinguer entre les itérations successives, comme entre les langages successifs de Tarski.

## 2.2 Transformation géométrique.

La géométrie est le domaine même de l'étude des transformations, et l'enchaînement de ces transformations ouvre lui-même la possibilité de la récursivité ; divers types de transformations présentent alors divers invariants, ce qui conduit à l'étude des *groupes* (par exemple les groupes de symétrie d'une figure, caractérisés précisément par le fait de laisser la figure globalement invariante).

Et si la transformation fait correspondre une partie de l'objet à l'objet entier, l'itération du processus permet d'engendrer les formes fractales. La figure 1 illustre ainsi les premières étapes de la construction de la courbe de von Koch.

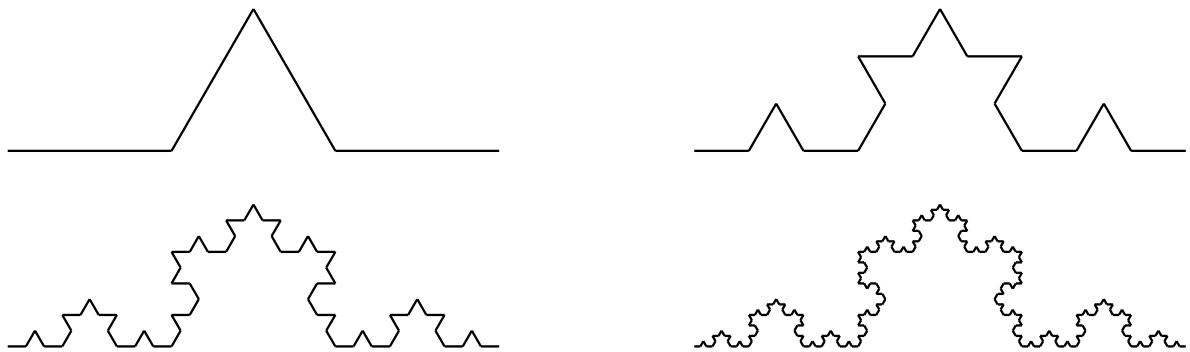


FIG. 1 – Courbe de von Koch

La courbe obtenue « *par passage à la limite* », ou encore le *point fixe* de la transformation, a une longueur infinie mais délimite une surface finie : des formes analogues pour des surfaces dans l'espace permettent d'atteindre de grandes surfaces d'échange dans de petits volumes, par exemple dans les poumons pour faciliter les échanges gazeux de la respiration.

Remarquons que lorsqu'on définit cet objet comme *point fixe* d'une transformation, résultat d'itérations successives de la transformation, l'itération se fait ici dans le sens d'une complexification, d'une complétion de l'objet (contrairement à la section précédente, où le calcul récursif simplifiait la tâche à chaque itération). De plus on ne peut pas ici partir de l'objet construit pour retrouver, en un nombre fini de transformations « *réciproques* » de celles de la construction, le segment initial à partir duquel il est construit, le « *cas simple* ».

En revanche on retrouve l'aspect paradoxal de l'objet à partir de considérations méréologiques : une partie de l'objet est semblable au tout, ou encore le tout contient un objet (et même plusieurs) qui lui est semblable, d'où l'on glisse vers « *l'objet se contient lui-même – strictement* ».

Et c'est bien la courbe-limite elle-même qui a cette propriété paradoxale, et non pas aucune des courbes intermédiaires obtenues lors des étapes successives de la construction.

Ici encore, on ne peut faire l'impasse sur aucun des aspects de l'objet construit par ce processus autoréférentiel :

- on ne saisit pas la courbe, comme point fixe de cette transformation, sans accepter de voir le tout dans la partie ;

- on ne réalise pas la construction sans en distinguer les étapes, correspondant à autant de niveaux de complexification de l’objet initial.

## Conclusion.

À travers les quelques domaines où s’illustrent l’autoréférence et la réflexivité que nous avons abordés, qu’il s’agisse de construction et représentation du sens à l’aide des langages formels ou naturels, de construction des processus calculatoires ou de construction d’objets géométriques, nous avons eu l’occasion de distinguer entre deux utilisation de la réflexivité, l’une simplifiant les objets qu’elle manipule, l’autre les complexifiant.

Par ailleurs, apparaît la nécessité de représenter les objets porteurs de réflexivité sous un double aspect : d’une part, distinguer dans la construction réflexive différents niveaux, différentes étapes, de façon à préserver les objets construits des risques véhiculés par les paradoxes ; d’autre part, accepter d’exprimer le paradoxe, à l’intérieur du système de représentation.

De la sorte, en se focalisant tantôt sur la version typée, tantôt sur la version non typée des objets, on pourra soit profiter de l’infine capacité du langage à se décrire, soit se prémunir d’« *excès* » de cette capacité, en établissant une distinction entre le *même* (l’objet) et le *même* (l’image de cet objet par un processus qui le laisse invariant).

Ainsi l’on pourra accepter le paradoxe du menteur, ou choisir d’en lever l’aspect paradoxal en distinguant entre le menteur qui parle et le menteur dont il parle, lui-même.

## Références

- Authier-Revuz, J., M. Doury, and S. Reboul-Touré (Eds.) (2003). *Parler des mots ; le fait autonymique en discours*. Paris : Presses de la Sorbonne Nouvelle.
- Gödel, K. (1931). Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme. *Monatshefte für Mathematik und Physik* 38, 173–198.
- Ladrière, J. (1957). *Les limitations internes des formalismes*. Paris : Jacques Gabay.
- Perry, J. (1999). *Problèmes d’indexicalité*. Stanford : CSLI.
- Wagner, P. (1998). *La machine en logique*. Paris : Presses Universitaires de France.



# Réflexivité et programmation du jeu de Go

Tristan Cazenave  
Laboratoire d'Intelligence Artificielle  
Université Paris 8, Saint Denis, France  
cazenave@ai.univ-paris8.fr

## Résumé

La réflexivité peut intervenir de multiples manières dans un programme de Go. Je présente ici trois aspects de la réflexivité dans la programmation du jeu de Go. Le premier aspect concerne l'auto-observation d'un programme de Go pour engendrer de nouvelles connaissances lui permettant de trouver plus rapidement des solutions aux problèmes qu'il doit résoudre. Le second aspect est l'utilisation d'un algorithme de recherche par ce même algorithme de recherche pour sélectionner les coups à envisager dans l'algorithme de recherche, ce qui lui permet d'être beaucoup plus efficace que l'algorithme simplement récursif habituellement utilisé. Le troisième aspect concerne l'utilisation d'un méta-niveau dans un programme de Go pour départager les recherches que l'on a intérêt à continuer de celles qui sont devenues inutiles.

## 1 Introduction

Une définition restrictive de la réflexivité, défendue par certains chercheurs en Intelligence Artificielle, est l'application d'un système à lui-même [Pitrat, 1990]. On peut ainsi utiliser un système qui découvre des connaissances pour découvrir des connaissances de découverte de connaissances [Lenat, 1983; Cazenave, 1999], ou utiliser un système qui contrôle l'utilisation des connaissances pour contrôler l'utilisation des connaissances de contrôle. Nous appellerons cette vision de la réflexivité, la réflexivité forte.

Une autre définition de la réflexivité couramment utilisée dans la communauté des chercheurs en langage de programmation [Tanter *et al.*, 2003] consiste pour un programme à avoir accès, lors de son exécution, à son propre état et à ses structures de haut niveau, et à pouvoir les modifier. On retrouve cette forme de réflexivité dans des langages de programmation comme Lisp, Java, Smalltalk, .NET, Ruby ou Python.

Dans cet article nous nous intéressons à la réflexivité comme la capacité que peut avoir un programme d'observer son propre comportement, et aux bénéfices

que peut apporter cette auto-observation pour rendre son fonctionnement plus efficace.

La réflexivité peut être utilisée dans un système complexe aussi bien pour permettre au système d'apprendre à trouver des solutions plus rapidement, que pour lui permettre de comprendre les raisons de ses succès ou de ses échecs et lui donner ainsi des indices sur les nouveaux essais à tenter, ou pour lui permettre de mieux contrôler son raisonnement en arrêtant avant qu'ils ne soient terminés les raisonnements qui sont inutiles ou qui ont une utilité trop faible.

Cette forme de réflexivité est liée au métaraisonnement dans les systèmes temps réel complexes composés de nombreux modules interdépendants. Le contrôle des ressources allouées à des algorithmes de recherche heuristique est par exemple un problème que l'on rencontre aussi bien dans les jeux [Russell and Wefald, 1989] que dans d'autres systèmes complexes [Pitrat, 1990; Russell and Wefald, 1991; Horvitz, 2001; Mouaddib *et al.*, 2002]. Ce type de problème était déjà abordé dans un des premiers démonstrateurs automatique de théorèmes [Pitrat, 1966] qui démontrait des meta-théorèmes de façon à améliorer son fonctionnement. C'est un thème récurrent et fondamental en Intelligence Artificielle [Ginsberg and Geddis, 1991; Russell, 1997; Minton, 1990; 1996].

L'article est structuré en trois sections : l'apprentissage par auto-observation, les recherches sélectionnant les coups à l'aide de recherches, les métaraisonnements sur l'utilité des recherches.

## 2 L'apprentissage par auto-observation

L'apprentissage par auto-observation [Cazenave, 1996b] consiste à mémoriser dans une trace les raisonnements d'un programme de résolution de problème pour ensuite analyser cette trace afin de créer de nouvelles connaissances qui permettront au programme de résoudre des problèmes plus rapidement. Ce type d'apprentissage a montré son utilité en programmation des jeux et d'une façon plus générale en résolution de problèmes.

### 2.1 Résolution de problème

Au départ un système d'apprentissage par auto-observation dispose d'une définition des règles du jeu, d'une définition des buts qu'il doit atteindre, et de métaconnaissances qui lui permettent de transformer les connaissances qu'il engendre.

La première étape de l'apprentissage est la résolution de problèmes. Lorsqu'il est confronté à de nouveaux problèmes, il essaie des coups et vérifie si un but est atteint après ces coups. Lorsqu'un coup atteint un but alors qu'aucune règle ne lui permet jusqu'alors de déduire que ce coup atteint le but, le système déclenche alors le processus d'apprentissage qui lui permettra de créer une règle

qui trouvera ce coup. Pour cela il a besoin de la trace des raisonnements qui lui ont permis de trouver que le coup atteignait le but. Le système mémorise donc toutes les règles qu'il a déclenché lors de la résolution de problème. Ce qui constitue la trace du raisonnement.

## 2.2 Analyse de la trace

Lorsque le système a détecté qu'il pouvait apprendre une nouvelle connaissance, il remonte dans la trace pour trouver les faits décrivant la position avant que le coup, qui sont à l'origine de la vérification du but après le coup. Une fois ces faits regroupés, il les assemble en une règle qui permettra de déduire plus rapidement le coup à jouer. Cette règle qui est au départ un ensemble de faits, est généralisée en transformant les variables instanciées en variable, puis ses conditions sont réordonnées de façon à être vérifiées efficacement [Cazenave, 1996a; 1998b].

## 2.3 Métaprogrammation

Les règles ainsi apprises sont transformées par un métaprogramme qui trouve tous les coups qui sont susceptibles d'empêcher d'atteindre un but à partir d'une règle déduisant un coup atteignant ce but [Cazenave, 1998a]. On appelle cet ensemble de coups des coups forcés (ce sont les seuls coups à envisager pour empêcher d'atteindre le but). Les règles sur les coups forcés sont utilisées pour engendrer des règles sur les situations où un but est atteint. Les règles sur les buts atteints sont elles mêmes utilisées pour apprendre les règles sur les coups qui atteignent un but, et ainsi de suite.

Pour éviter au système d'engendrer trop de règles, des métarègles de sélection des règles engendrées sont utilisées. On peut par exemple dans le cas du jeu de Go, ne retenir que les règles sur les coups forcés qui concluent sur un petit nombre de coups forcés.

Il est possible de remplacer le programme d'apprentissage par auto-observation par un programme qui engendre directement les règles par métaprogrammation. On peut aussi engendrer par métaprogrammation certains des métaprogrammes nécessaires à la génération des programmes [Cazenave, 1999], ce qui est une forme de réflexivité forte.

## 3 Les recherches sélectionnant les coups à l'aide de recherches

L'algorithme qui a en général le plus de succès dans les jeux de réflexion est l'Alpha-Béta. Toutefois, dans la plupart des jeux classiques, le nombre de coups légaux est relativement petit (de l'ordre d'une trentaine aux Echecs par exemple) ce qui permet à l'Alpha-Béta de faire des recherches relativement profondes (de l'ordre de 12 demi-coups actuellement) et donc d'avoir des programmes de bon niveau. Pour le jeu de Go, la situation est bien différente des autres jeux

classiques, puisque le nombre de coups possibles est de l'ordre de 250 et que les recherches sont en général plus profondes que dans les autres jeux.

On est donc amené au jeu de Go à utiliser des algorithmes sélectifs de recherche. Le problème est alors de sélectionner un petit nombre de coups sans toutefois oublier les coups importants. L'apprentissage par auto-observation permet d'apprendre des règles qui sélectionnent sûrement un petit nombre de coups. Une autre manière de sélectionner les coups est d'engendrer dynamiquement à chaque fois les explications sur la réussite d'un but, de façon à en tirer les coups forcés qui permettront d'empêcher l'adversaire d'atteindre ce but.

Les menaces généralisées [Cazenave, 2002] sont basées sur ce principe. A chaque noeuds Min de la recherche, le programme essaie des coups Max au lieu des coups Min. Puis il relance une recherche après chaque coup Max. Si un de ces coups Max atteint le but, il retient dans une structure de données les raisons qui font que le coup atteint le but. Puis il en déduit un petit nombre de coups Min qui sont les seuls susceptibles d'empêcher Max d'atteindre le but.

Le pseudo code pour la vérification des menaces est le suivant :

```

MenaceMin (menace, ordre, raisons) {
    Si butAtteint ()
        return 1;
    res = 0;
    mg = menaceGauche (menace, ordre);
    if (MenaceMax (mg, ordre, raisons) == 1) {
        res = 1;
        coups = trouveCoupsMin (ordre, raisons);
        md = menaceDroite (menace, mg);
        o = ordreMenace (md);
        for (c dans coups et tant que res == 1) {
            joue (c, MinColor);
            if (MenaceMax (md, o, raisons) == 0)
                res = 0;
            dejoueCoup ();
        }
    }
    return res;
}

MenaceMax (menace, ordre, raisons) {
    res = 0;
    coups = trouveCoupsMax (ordre);
    oteCoupOrdre (menace, ordre);
    for (c dans coups et tant que res == 0) {

```

```

joue (c, MaxColor);
initialise (r);
if (MenaceMin (menace, ordre-1, r) == 1) {
    res = 1;
    ajouteRaisonsCoup (c, r);
    raisons = r;
}
dejoueCoup ();
}
return res;
}

```

L'ordre d'une menace est le nombre maximum de coups de suite de la couleur Max qu'on a le droit de jouer en vérifiant la menace.

Lorsqu'on vérifie une menace, pour un noeud Min, il y a deux sous-arbres, le sous-arbre gauche et le droit. Par exemple dans la figure 1, dans le noeud qui est sous la racine de la menace (6,3,2,0), le sous-arbre gauche est une menace (2,1,0) alors que le sous-arbre droit est une menace (4,2,1,0).

L'algorithme de recherche mémorise dynamiquement les raisons pour lesquelles ce même algorithme de recherche avec un ordre inférieur a trouvé la solution à un problème. Ces raisons qui constituent une trace de l'exécution de l'algorithme de recherche sont utilisées pour trouver les coups de Min (et constituent un petit sous ensemble de l'ensemble des coups possibles).

Cet algorithme de recherche est non seulement beaucoup plus rapide qu'un algorithme Alpha-Béta naïf qui essaierait tous les coups possibles, mais il est aussi plus rapide et plus sûr qu'un algorithme Alpha-Béta sophistiqué qui n'envisage que certains coups en se basant sur des connaissances du jeu de Go. Ainsi pour la résolution du jeu d'Atarigo 6x6, l'algorithme de recherche utilisant la menace généralisée (6,3,2,0) est 270 fois plus rapide qu'un Alpha-Béta utilisant les mêmes optimisations de recherche [Cazenave, 2002] (tables de transposition, deux coups qui tuent, heuristique de l'historique).

La figure 1 donne des exemples d'arbres correspondant à des menaces généralisées. Dans ces arbres toutes les feuilles correspondent à des positions où le but est atteint pour le joueur Max. Chaque branche de l'arbre vers la gauche correspond à un coup gagnant pour Max. Chaque branche vers la droite correspond à l'ensemble de coups Min qui sont susceptibles d'empêcher Max d'exécuter la menace associée à la branche vers la gauche qui part du même noeud. On peut compter pour chaque arbre le nombre de coups de chaque ordre pour Max. Ainsi l'arbre le plus simple correspondant à un coup gagnant pour Max ne contient qu'un seul coup d'ordre un. On notera cet arbre (1,0). Le premier nombre dans cette notation correspond au nombre de coups d'ordre un dans l'arbre, le deuxième nombre au nombre de coups d'ordre deux, et ainsi

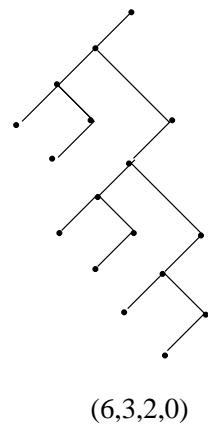
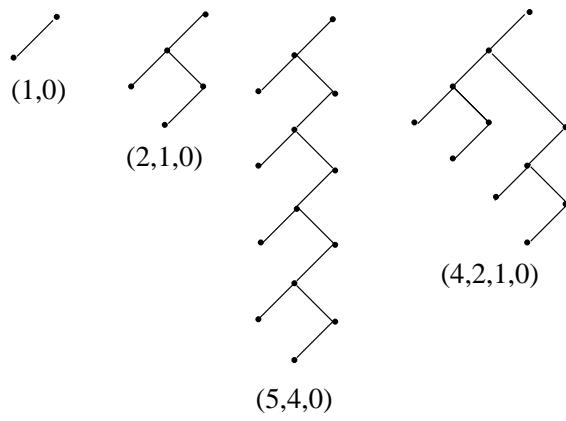


Figure 1: Arbres de menaces généralisées.

de suite jusqu'au premier zéro qui signifie qu'il n'y a pas de coups d'ordre supérieur à l'ordre correspondant au zéro.

Pour tous ces arbres, quand on est à un noeud où il y a une branche vers la droite et une branche vers la gauche, on a toujours le sous-arbre gauche inclus dans le sous-arbre droit. Cette contrainte sur la forme des arbres vient de ce qu'il est la plupart du temps plus difficile d'atteindre le but pour Max après un coup de Min qu'avant.

Pour les menaces généralisées complexes comme (6,3,2,0), plusieurs arbres correspondent à la même suite de nombres.

## 4 Les métaraisonnements sur l'utilité des recherches

Dans un programme de jeu de Go, l'efficacité du métaraisonnement dépend des résultats des recherches heuristiques, et la décision de continuer ou non les recherches heuristiques dépend du métaraisonnement et de raisonnement stratégiques sur des objets plus complexes que ceux traités par les recherches

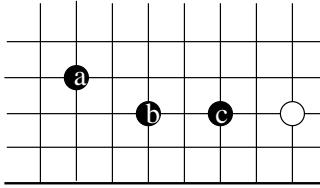


Figure 2: Trois calculs de connexion possibles.

heuristiques. Le méta niveau et le niveau de base sont interdépendants. Nous expliquons quelques unes de ces dépendances et la façon dont nous les traitons.

Un programme de Go effectue un grand nombre de recherches tactiques dont les résultats lui permettent de construire des objets complexes comme les groupes. Une approche basique consiste à faire toutes les recherches tactiques possibles, puis à construire les objets complexes une fois toutes les recherches effectuées. Une approche plus fine est d'entrelacer les recherches tactiques avec la construction des objets complexes. En effet, les objets complexes, même partiellement construits nous renseignent sur l'utilité des recherches tactiques restantes. En entrelaçant recherches tactiques et construction d'objets complexes, on peut ainsi décider d'abandonner des recherches couteuses et inutiles avant de les avoir terminées.

#### 4.1 *Construction des groupes*

Nous présentons ici un cas particulier qui est la construction des groupes à partir des résultats des recherches sur la connexion de deux chaînes.

La figure 2 donne un exemple où le programme trouve trois connexions à calculer. Les connexions à calculer sont les connexions de a avec b, de b avec c, et de a avec c.

Les groupes sont construits à partir des calculs de connexion. Si deux chaînes ne peuvent être déconnectées, comme c'est par exemple le cas pour les pierres b et c dans la figure 2, elles font partie du même groupe.

Le programme doit toutefois faire attention à la non transitivité des connexions [Cazenave and Helmstetter, 2004]. Dans le cas de la figure 2, a est connecté à b, b est connecté à c, et a est connecté à c. Il y a transitivité de la connexion, toutefois ce n'est pas toujours le cas.

Le programme qui détecte si les connexions complexes sont transitives est plus efficace que le programme qui calcule directement les connexions composées. Par exemple, toujours dans la figure 2, il est plus efficace de calculer que a est connecté à b et que b est connecté à c, plutôt que de calculer si a est connecté à c. On comprend la raison de cette efficacité accrue en raisonnant sur le coût de chaque recherche. Si le facteur de branchement est b pour une connexion (et donc 2b pour une connexion composée), et que la profondeur de recherche pour prouver une connexion est p (et donc 2p pour une connexion

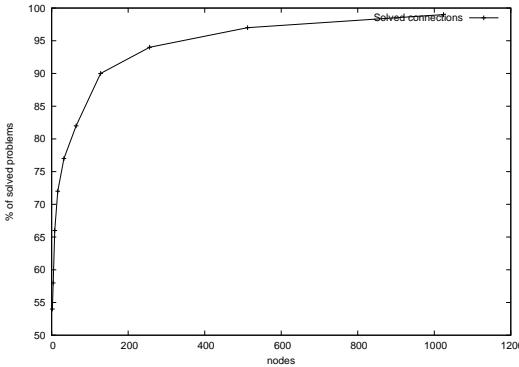


Figure 3: Profil pour les problèmes de connexion

composée). Chercher que a est connecté à b et que b est connecté à c coutera  $2b^p$  alors que le calcul de la connexion de a à c coutera  $(2b)^{2p}$ . Le calcul d'une connexion composée est beaucoup plus couteux que le calcul de chacune des connexions indépendamment. Les calculs de connexion sont indépendants lorsque les traces (les raisons des menaces généralisées) sont disjointes.

On construit les groupes en ajoutant au groupe les chaînes qui ne peuvent pas être déconnectées d'une chaîne appartenant déjà au groupe, tout en vérifiant qu'il y a bien transitivité de la connexion avec les autres chaînes du groupe.

#### 4.2 *Une implémentation simple*

Une implémentation simple d'un programme de Go consiste à effectuer tous les calculs tactiques les uns après les autres pour ensuite construire les groupes et utiliser les résultats des recherches tactiques dans des objets plus complexes.

#### 4.3 *Interactions entre la construction des groupes et les recherches*

Une implémentation plus fine consiste à entrelacer construction des groupes et recherches tactiques. Dans le cas des recherches de connexions, on peut fixer une suite de seuils pour les recherches. A chaque fois que toutes les recherches ont atteint un certain seuil, on déclenche le mécanisme qui permet d'arrêter les recherches inutiles. Par exemple dans le cas de la figure 2, dès que les recherches ont trouvé que a est connecté à b et que b est connecté à c, et que la connexion est transitive, on arrête la recherche sur la connexion entre a et c.

Les figures 3 et 4 donnent le pourcentage de problèmes résolus en fonction du nombre de noeuds cherchés pour les problèmes de connexion et de déconnexion. On voit bien que la progression est logarithmique. Il est donc probablement préférable de fixer une suite de seuils pour les recherches tactiques avec une progression exponentielle.

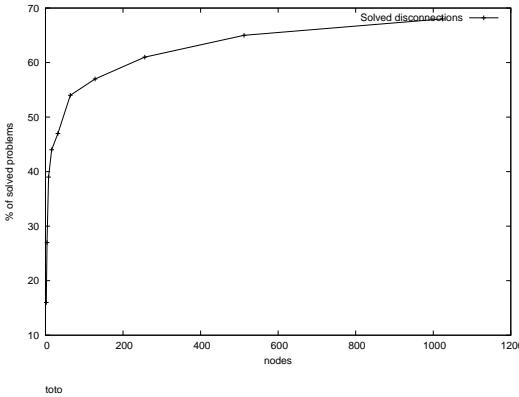


Figure 4: Profil pour les problèmes de déconnexion

#### 4.4 Autres interactions possibles

De même qu'on peut détecter que des calculs de connexions complexes sont inutiles parce que déductible de manière plus simple, on peut aussi détecter l'inutilité d'autres calculs. Par exemple si on a trouvé deux yeux indépendants dans un même groupe, il devient inutile de faire une recherche sur la vie de ce groupe.

#### 4.5 Ordonnancement heuristique des recherches

Une autre heuristique possible est d'ordonner les recherches par complexités présumées, en partant de la plus simple vers la plus complexe. Une bonne heuristique pour évaluer la complexité de la recherche sur une connexion est de compter le nombre de coups de suite qu'il faut jouer pour connecter les deux chaînes, c'est à dire l'ordre de la connexion au sens des menaces généralisées. Le programme peut ainsi continuer à éviter des recherches complexes inutiles puisqu'il disposera avant de les commencer des combinaisons de recherches plus simples. Par exemple, les problèmes de la figure 2 sont respectivement d'ordres 2, 1 et 3. Le programme commencera par les recherches d'ordre 1 et 2, et évitera la recherche d'ordre 3 car elle peut être déduite de la combinaison des deux recherches plus simples.

### 5 Conclusion

J'ai montré trois utilisations de la capacité qu'a un programme d'observer une partie de son propre fonctionnement en rapport avec la programmation du Go. Dans ces trois cas, l'utilisation de l'observation du programme par lui même lui permet de gagner du temps sur une implémentation n'utilisant pas cette auto-observation. Dans le cas de l'apprentissage par auto-observation, le programme crée de nouvelles connaissances qui lui permettent de résoudre les problèmes plus rapidement. Dans le cas des menaces généralisées, l'observation de la trace de ses calculs lui permet d'être beaucoup plus efficace que les algorithmes plus

classiques. Dans le cas des métaraisonnements sur l'utilité des recherches, le programme évite de perdre du temps dans des recherches inutiles en alternant les recherches avec la détection des recherches devenues inutiles.

## References

- [Cazenave and Helmstetter, 2004] T. Cazenave and B. Helmstetter. Search for transitive connections. *Information Sciences*, 2004.
- [Cazenave, 1996a] T. Cazenave. Automatic ordering of predicates by metarules. In *5th International Workshop on Metaprogramming and Metareasoning in Logic*, Bonn, 1996.
- [Cazenave, 1996b] T. Cazenave. Système d'Apprentissage Par Auto-Observation. Application au jeu de Go. Phd thesis, Université Paris 6, December 1996.
- [Cazenave, 1998a] T. Cazenave. Metaprogramming Forced Moves. In *ECAI 1998*, pages 645–649, Brighton, UK, 1998.
- [Cazenave, 1998b] T. Cazenave. Speedup Mechanisms For Large Learning Systems. In *IPMU 1998*, Paris, France, 1998.
- [Cazenave, 1999] T. Cazenave. Metaprogramming domain specific metaprograms. In *Reflection 99*, volume 1616 of *Lecture Notes in Computer Science*, pages 235–249. Springer, 1999.
- [Cazenave, 2002] T. Cazenave. A Generalized Threats Search Algorithm. In *Computers and Games 2002*, volume 2883 of *Lecture Notes in Computer Science*, pages 75–87, Edmonton, Canada, 2002. Springer.
- [Ginsberg and Geddis, 1991] M. L. Ginsberg and D. F. Geddis. Is there any need for domain-dependent control information? In *AAAI-91*, pages 452–457, 1991.
- [Horvitz, 2001] E. Horvitz. Principles and applications of continual computation. *Artificial Intelligence*, 126(1-2):159–196, 2001.
- [Lenat, 1983] D. B. Lenat. Eurisko: A program that learns new heuristics and domain concepts. *Artificial Intelligence*, 21:61–98, 1983.
- [Minton, 1990] S. Minton. Quantitative Results Concerning the Utility of Explanarion Based Learning. *Artificial Intelligence*, 42(2-3):363–392, 1990.
- [Minton, 1996] S. Minton. Is there any need for domain-dependent control information : A reply. In *AAAI-96*, 1996.
- [Mouaddib *et al.*, 2002] A.-I. Mouaddib, E. Grégoire, and J.-F. Dauchez. An intelligent system combining different resource-bounded reasoning techniques. *Applied Intelligence*, 17(2):127–140, 2002.

- [Pitrat, 1966] J. Pitrat. Réalisation de programmes de démonstration de théorèmes utilisant des méthodes heuristiques. Thesis, Université de Paris, 1966.
- [Pitrat, 1990] J. Pitrat. *Métaconnaissance - Futur de l'Intelligence Artificielle*. Hermès, Paris, 1990.
- [Russell and Wefald, 1989] S. J. Russell and E. Wefald. On optimal game tree search using rational meta-reasoning. In *IJCAI 89*, pages 334–340. Morgan Kaufmann, Detroit, Michigan, August 1989.
- [Russell and Wefald, 1991] S. J. Russell and E. Wefald. Principles of metareasoning. *Artificial Intelligence*, 49:361–395, 1991.
- [Russell, 1997] S. J. Russell. Rationality and intelligence. *Artificial Intelligence*, 94(1-2):57–77, 1997.
- [Tanter *et al.*, 2003] E. Tanter, J. Noyé, D. Caromel, and P. Cointe. Partial behavioral reflection: spatial and temporal selection of reification. In *OOPSLA 2003*, pages 27–46, 2003.



# Modèle de composants logiciels réflexifs et dialogiques

Yasmine Charif et Nicolas Sabouret

LIP6 8, rue du Capitaine Scott, 75015 Paris  
{Nicolas.Sabouret, Yasmine.Charif}@lip6.fr

## Résumé

Nous proposons à travers cet article un modèle de programmation de composants actifs dédiés à l'assistance des utilisateurs ordinaires de services web. Ce modèle permet aux composants de raisonner sur leur propre code en cours d'exécution. Nous montrons comment notre approche confère aux composants des propriétés de réflexivité, des capacités d'interactions dialogiques avec des utilisateurs humains qu'ils assistent et des propriétés d'interopération avec d'autres composants à l'aide de requêtes formelles. Enfin, nous concluons sur l'intérêt et la manière d'utiliser notre modèle dans les systèmes complexes.

**Mots clés :** Interactions entre composants, Raisonnement sur le fonctionnement, Composants réflexifs, Engagement ontologique, VDL.

## 1 Exposé de la problématique

Avec le développement de l'informatique pour le grand public, en particulier dans les services en ligne dans l'Internet, la conception de systèmes est confrontée à un double problème. D'une part, les systèmes deviennent de plus en plus complexes : ils sont issus de l'assemblage de composants hétérogènes qui n'ont généralement pas été prévus pour interopérer. Les travaux récents dans le domaine des services web sémantiques (SWS) [5, 11, 6] ont pour objectif de permettre aux experts de programmer une couche d'interopérabilité supplémentaire. À terme, les services devraient pouvoir produire automatiquement cette description sémantique de leurs fonctionnalités. Il est alors nécessaire que ces services soient capables d'accéder à leur propre code pour raisonner dessus.

D'autre part, l'*utilisateur ordinaire* n'est pas en mesure d'interagir directement avec les services. Il lui faut passer par des agents qui l'assistent dans la réalisation de la tâche [3]. Lorsqu'il se retrouve en situation d'échec, les études [12, 15] ont montré qu'il voulait pouvoir *poser des questions* au système à propos de *ce qu'il fait*. Une fois encore, il est nécessaire que les agents et les services puissent accéder à leur propre code, en cours d'exécution, afin de raisonner dessus. Mais les modèles de programmation agent classiques, comme JADE [1], sont fondés sur

la programmation déclarative dans laquelle les agents ne peuvent pas accéder à leur code à l'exécution et encore moins raisonner dessus.

Dans cet article, nous proposons un modèle de programmation d'agents assistants capables de :

- interagir avec un utilisateur humain en langage naturel pour la commande ou le contrôle ;
- raisonner sur son propre code **en cours d'exécution** pour répondre à des questions mais aussi pour **modifier** la définition de son comportement (réflexivité) ;
- interopérer avec des services web et d'autres agents assistants en s'appuyant sur un modèle dialogique de requêtes formelles.

Nous présentons tout d'abord le principe général de notre modèle de programmation ainsi que notre modèle formel d'interaction. Nous explicitons ensuite les capacités de réflexivité et d'interopérabilité des composants modélisés. Enfin, nous donnons un exemple de manipulation dynamique du code d'un composant en réponse à une interaction avec un autre agent.

## 2 Un modèle de programmation réflexif

Dans cette section, nous présentons notre modèle de programmation de composants sémantiques et actifs.

### 2.1 *Présentation de VDL*

Les travaux de recherches que nous menons consistent à définir des services web sémantiques et actifs (ASWS) [18], c'est-à-dire des agents assistants en ligne dans l'internet capables d'interagir à la fois avec les utilisateurs humains et avec des services web. Du point de vue de l'utilisateur, les ASWS sont des pages web actives<sup>1</sup>. Mais contrairement aux services classiques qui sont uniquement réactifs, les ASWS fonctionnent à la fois en mode réactif et proactif (ou autonome). Enfin, ils sont dits *sémantiques* car ils sont munis d'une représentation de leur propre fonctionnement sur laquelle ils peuvent raisonner.

Les ASWS sont définis dans un langage spécifique, appelé VDL (pour *View Design Language*) [20], vérifiant les propriétés suivantes :

1. Le modèle qui permet de décrire la partie opératoire des composants, *i.e.* les actions qu'ils effectuent, est aussi celui dans lequel s'effectuent ces opérations. Il est donc possible d'accéder en cours d'exécution à la description des actions des composants pour raisonner dessus en s'appuyant sur la sémantique opérationnelle du langage.

---

<sup>1</sup>Des démonstrations sont disponibles en ligne : <http://www-poleia.lip6.fr/~sabouret/demos>

2. Les descriptions de la structure et du fonctionnement des composants sont intégrées dans une même représentation qui s'appuie sur les modèles du web sémantique<sup>2</sup> [4]. Ceci permet aux composants de répondre aussi bien à des requêtes sur les données manipulées que sur les actions de manipulation.
3. Les composants peuvent interagir en langue naturelle avec l'utilisateur humain et en langage formel avec d'autres composants, en particulier pour répondre à un large éventail de questions portant sur leurs actions et leur exécution.

## 2.2 *Partie structurelle*

Le langage VDL est un langage de programmation de composants actifs (ou *agents*) basé sur la réécriture d'arbres XML. Le *corps* d'un composant, c'est-à-dire la description en VDL du monde opératoire qu'il manipule, est un arbre XML dont les noeuds sont des éléments appelés *concept*. Certains concepts modélisent les données manipulées par le composant alors que d'autres définissent les actions de manipulations. Nous parlerons alors de *mots clefs* du langage. Nous appelons *vue de l'agent* cet arbre de concepts.

Pour pouvoir interagir, les agents doivent non seulement partager un protocole d'interaction commun [8] (niveau syntaxique) mais aussi vocabulaire commun (niveau sémantique). Les ontologies sont des schémas de métadonnées qui permettent de définir ce vocabulaire, la signification de chaque terme, les concepts lui étant liés et les relations avec ces concepts.

Dans notre modèle, le programmeur VDL doit associer à chaque agent VDL une *ontologie de tâche*<sup>3</sup> regroupant les définitions des données manipulées. C'est-à-dire que tous les concepts de la vue, à l'exception des mots-clefs dont le sens est donné par la sémantique opérationnelle du langage, doivent être définis dans l'ontologie de tâche. Ces ontologies sont utilisées à deux niveaux. Premièrement, elles permettent à un module de traitement des langues de construire des requêtes formelles modélisant les questions et les commandes des utilisateurs ordinaires. Sans ontologie, cela serait impossible. Deuxièmement, elles répondent au besoin d'interopérabilité entre les agents.

## 2.3 *Partie fonctionnelle*

Il n'est pas possible de présenter ici en détail l'ensemble des mots clefs du langage VDL [20]. Nous voudrions cependant mettre l'accent sur les points importants pour la programmation de services web sémantiques et actifs (ASWS).

---

<sup>2</sup>Le web sémantique est une extension du web actuel dans lequel l'information est munie d'une sémantique.

<sup>3</sup>Une ontologie de tâche fournit un vocabulaire systématisé des termes utilisés pour résoudre des problèmes associés à des tâches qui peuvent appartenir ou non à un même domaine [14].

Une propriété importante du modèle VDL est que l'interprète du langage effectue à chaque cycle d'exécution une réécriture de la vue en un nouvel arbre XML, en fonction des mots-clefs du langage qui s'y trouvent, comme dans les logiques de réécriture [13]. Il est donc possible d'accéder à tout instant de l'exécution à la vue de l'agent, c'est-à-dire à la description complète du monde opératoire et des actions de manipulation.

Une autre propriété importante du modèle est qu'il intègre les deux modes d'exécution d'un composant :

- Dans le mode *réactif* (ou modal), les composants effectuent des opérations en réponse à des interactions de l'utilisateur (humain ou agent logiciel) comme cliquer sur le bouton "démarrer", demander un service, *etc.* C'est le mode de fonctionnement des services web classiques.
- Dans le mode *proactif* (ou non modal), les composants fonctionnent en dehors de toute interaction. Les actions sont régies uniquement par l'état interne de l'agent. Par exemple, un processus de cuisson, une fois lancé, ne nécessite pas d'interaction supplémentaire pour fonctionner.

Les actions en VDL sont donc une paire  $\langle P, M \rangle$  où :

- $P$  est l'ensemble des *préconditions* de l'action. Une précondition est une expression du langage VDL qui, lorsqu'elle est vérifiée, déclenche l'exécution de l'action. En mode réactif, certaines de ces préconditions sont des événements qui doivent être reçu par le composant.
- $M$  est l'ensemble d'*actions élémentaires* de modification d'un document XML. En VDL, il existe trois actions de modification élémentaires pour ajouter, remplacer ou supprimer un élément de la vue.

## 2.4 Exemple

Pour illustrer la définition d'actions en VDL, considérons un composant simple. Il s'agit d'un compteur dont le code VDL initial est le suivant :

```

<view>
  <variable type="integer" name="my-value">0</variable>
  <variable type="boolean" name="running">true</variable>
  <action><name>count</name>
    <guard><get><variable name="running"/></get></guard>
    <put> <path><variable name="my-value"/></path>
      <plus> <get><variable name="my-value"/></get>
        <value>1</value> </plus>
      </put>
    </action>
    <action><name>stop</name>
      <event><stop/></event>
      <put> <path><variable name="running"/></path>
        <value>false</value> </put>
    </action>
</view>

```

```

</action>
<action><name>start</name>
  <event><start/><go/></event>
  <put> <path><variable name="running"/></path>
    <value>true</value> </put>
</action>
</view>

```

Cet agent très simple manipule deux variables : *my-value* et *running*, dont les valeurs initiales sont respectivement *0* et *true*.

Il comporte également trois actions appelées *count*, *stop* et *start*. L'action  $count = < P_1, M_1 >$  a pour précondition  $P_1$  (de balise *guard*) qui est vérifiée si et seulement si *running* est à vrai et pour action élémentaire  $M_1$  (de balise *put*) qui augmente de 1 la valeur de *my-value*. Ainsi, tant que *running* est vrai, le compteur augmente sa valeur. Cette action définit donc un comportement **proactif**.

L'action  $stop = < P_2, M_2 >$  a pour précondition  $P_2$  (de balise *event*) qui est vérifiée si et seulement si le composant reçoit la commande (c'est-à-dire l'événement externe) *<stop/>*. Elle a pour action élémentaire  $M_2$  (de balise *put* aussi) qui met *running* à faux. Ainsi, dès que l'utilisateur envoie l'événement *stop* au composant, celui-ci s'arrête de compter. Cela définit donc un comportement **réactif**.

Enfin, l'action *start*, à la réception de l'une des commandes *<start/>* ou *<go/>*, met la variable *running* à vrai (et relance donc le processus *count*).

## 3 Un modèle de requêtes sur le fonctionnement

Dans cette section, nous présentons le modèle formel d'interaction qui permet à la fois de modéliser les questions des utilisateurs humains à propos du fonctionnement de leur agent assistant et de représenter les questions des agents entre eux sur leurs capacités actives et interactives.

### 3.1 *Principe général*

Le modèle de programmation VDL a pour objectif de définir des composants capables de se représenter leur fonctionnement, de raisonner dessus et d'interagir en langue naturelle avec l'utilisateur pour répondre à un large éventail de questions portant sur leurs actions et leur exécution. L'analyse du problème de l'interaction avec l'utilisateur en situation de demande d'aide [17] nous a conduit à proposer le schéma de principe présenté figure 1. Ce schéma illustre un utilisateur posant une question en langue naturelle au composant avec lequel il interagit. Un module de traitement de la langue naturelle (MLN) traduit cette question en une requête formelle. A partir de cette requête, un module de raisonnement sur le

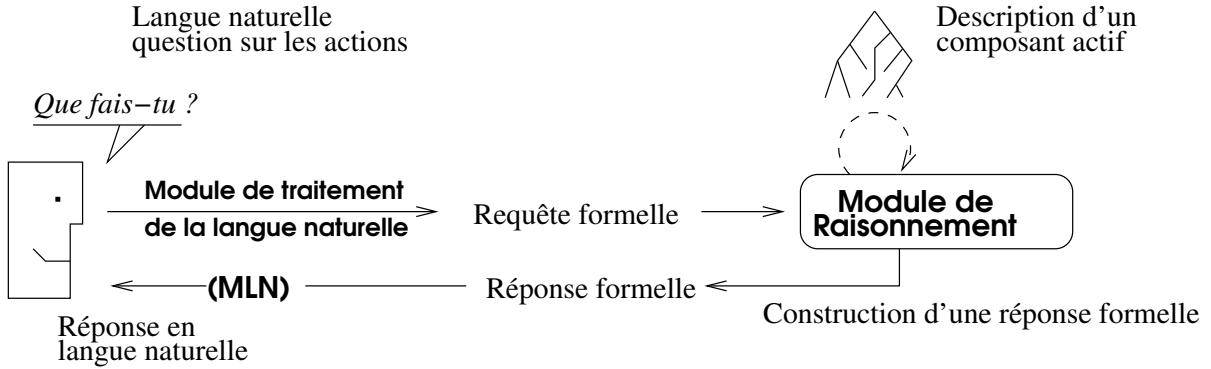


FIG. 1 – Modèle de requêtes sur le fonctionnement

fonctionnement (MRF) construit une réponse en s'appuyant sur l'introspection du code de l'agent à l'instant  $t$  rendue possible grâce au modèle VDL. La réponse est aussi une requête formelle. Le MLN est ensuite utilisé pour formuler la réponse en langue naturelle. Actuellement, ce module est encore en phase d'implémentation. L'interrogation du composant se fait à travers une interface de saisie de requêtes formelles.

### 3.2 Modèle de requêtes

Le modèle de requêtes que nous utilisons et que nous présentons brièvement ici fait suite à un travail de catégorisation [19] des questions qu'un utilisateur humain peut poser à un composant actif à propos de son fonctionnement. Il prend en compte à la fois :

- le performatif de l'acte de langage [21] de la question (questions « *pourquoi ?* », « *comment ?* », « *est-ce que ?* », etc) ;
- le type de fonctionnement considéré, *i.e.* l'activité du composant « *Est-ce que tu comptes ?* », son comportement « *Est-ce que tu peux aller plus vite ?* », son état « *Quelle est ta valeur ?* » ou ses capacités « *Qu'est-ce que je peux te demander de faire ?* » ;
- le traitement nécessaire, à partir de la description en VDL de l'état et des actions courantes de l'agent, pour construire la réponse ;
- et la possibilité de représenter des questions portant sur le passé “*Quand ais-je changé ta vitesse ?*” ou le futur de l'agent “*Quelle sera ta valeur dans 3 minutes ?*”.

Dans ce modèle, les questions et les réponses sont présentées sous la forme d'un sextuplet :

$$\rho = \langle \alpha, \tau, \sigma, \omega, \nu, \delta \rangle$$

où  $\alpha \in \{Assert, Ask, What, Why, How\}$  représente l'acte de langage,  $\tau \in \{is, do, can, order\}$  représente le type de fonctionnement considéré,  $\sigma$  est le sujet de la

requête, c'est-à-dire l'élément de la vue qui est questionné,  $\omega$  les arguments de la requête,  $\nu$  un marqueur pour les requêtes négatives et  $\delta$  la date de la requête (présent, passé ou futur).

Sans entrer dans les détails de l'algorithme de traitement [19], nous pouvons souligner que le performatif de l'acte de langage [21] (critère  $\alpha$ ) et le type de fonctionnement considéré (critère  $\tau$  de la requête) sont les deux éléments fondamentaux pour la construction d'une réponse adaptée. De nombreux exemples sont présentés dans [17].

## 4 Interactions dialogiques entre services actifs

Le modèle de programmation que nous avons présenté dans les sections précédentes permet aux agents d'accéder en cours d'exécution à la définition de leurs actions et de leur fonctionnement. Ils peuvent alors raisonner sur ces actions pour répondre à des requêtes formelles sur le fonctionnement. En s'appuyant sur ce modèle, nous voudrions permettre aux agents de construire dynamiquement les fonctionnalités d'interaction qui leur permettent d'interopérer.

### 4.1 Engagement ontologique

Pour permettre aux agents d'interopérer, il faut résoudre les contraintes d'intercompréhension au niveau syntaxique et sémantique. Il est donc nécessaire que deux concepts ayant le même nom aient également la même signification pour les agents qui les manipulent. Les recherches actuelles dans ce domaine proposent d'intégrer les ontologies initiales en une seule [7, 10, 16].

Cependant, contrairement aux approches « classiques » qui nécessitent que les agents disposent d'une ontologie de domaine commune incluant une taxinomie de toutes les **actions**, les agents VDL respectent un engagement ontologique minimal [9] qui ne contraint pas les programmeurs de composants à s'accorder sur les noms et définitions des actions manipulées. C'est-à-dire que les ontologies de tâche utilisées par les agents VDL contiennent uniquement la définition des concepts manipulés et non celle de leurs actions. Celles-ci sont en fait appréhendées par la sémantique opérationnelle des actions élémentaires et des préconditions qui les composent, plutôt que par le nom qu'un programmeur leur a attribué.

La seule contrainte est que les requêtes et les assertions doivent utiliser le vocabulaire défini dans l'ontologie commune. Les agents construisent alors leurs capacités d'interopération au moyen d'un *dialogue en langage formel*, c'est-à-dire d'un échange de requêtes formelles qui leur permet de s'informer sur les fonctionnalités d'un agent, de poser des questions sur son fonctionnement, *etc.*

## 4.2 *Un modèle d'interactions sur le fonctionnement*

Pour permettre aux agents d'interopérer, il est nécessaire d'intégrer dans VDL un langage de communication entre agents (ACL) [2] et de résoudre les contraintes d'inter-compréhension au niveau syntaxique et sémantique. Dans ce contexte, le modèle VDL prend en compte trois sortes d'**interactions élémentaires** pour l'envoi de messages :

1. Des *événements externes* pour déclencher l'action d'un autre agent ;
2. Des *actions* pour envoyer le code d'une action VDL à un autre agent ;
3. Des *requêtes* pour poser des questions (ou répondre) à d'autres composants concernant ses données, ses actions ou son comportement en utilisant le modèle de requêtes précédemment défini.

Un message à envoyer est défini dans l'élément **send-message** déclaré dans le corps d'une action. Les éléments *send-message* sont considérés par l'interprète VDL comme des actions élémentaires : ils ne sont déclenchés que lorsque l'action dans laquelle ils sont définis valide ses préconditions *event* et *guard*. Enfin, chaque envoi de message doit spécifier :

- l'adresse exacte du composant destinataire dans l'élément **dest**. C'est un identifiant unique formé du nom du composant, de l'identifiant de son utilisateur et de l'URL du serveur sur lequel il s'exécute ;
- et le contenu du message englobé dans un élément représentant son performatif (**event**, **action** ou **request**).

Les événements externes sont traités directement au niveau du code VDL lorsqu'ils sont reçus (ils activent les actions correspondantes). Au contraire, les actions et les requêtes sont traitées à un niveau méta dans l'interprète VDL.

## 4.3 *Traitemet des requêtes*

Lorsqu'un agent VDL reçoit une requête, il demande au module de raisonnement (*cf.* figure 1) de construire une réponse à partir des algorithmes développés dans [17]. Par exemple, supposons qu'un agent souhaite connaître les fonctionnalités d'un agent assistant servant d'interface avec un four comme l'agent *Mike*<sup>4</sup>. La question « *Comment fais-tu pour réchauffer ?* » peut être envoyée par l'agent à l'aide du code VDL suivant :

```
<send-message>
  <dest><name>Mike </name></dest>
  <request>
    <act>How</act>
    <type>do </type>
    <subject><view/> </subject>
```

---

<sup>4</sup><http://www-poleia.lip6.fr:8180/demos/jsp/agentframe.jsp?name=mike>

```

<object><heat/></object>
</request>
</send-message>

```

Une fois reçue par *Mike*, la requête est traitée au niveau de son module de raisonnement et une réponse formelle comprenant la définition de l'action « *réchauffer* » est renvoyée à l'agent initial :

```

<send-message>
  <dest><name>Four</name></dest>
  <action>
    <event><heat/></event>
    <put>
      <path><température/></path><value>180</value>
    </put>
  </action>
</send-message>

```

## 4.4 *Traitements des actions : un exemple de réflexivité*

Lorsqu'un agent reçoit une action, il doit l'intégrer à son propre code, dans la mesure où elle ne contient que des concepts figurant dans son ontologie. En effet, comme le montre l'exemple précédent, la définition d'une action ou d'une fonctionnalité peut être envoyée en réponse à une requête. L'agent demandeur va alors vouloir apprendre cette action, c'est-à-dire modifier son propre code pour y ajouter la description et les préconditions de la nouvelle action.

Cependant, nous voudrions que cette nouvelle action s'intègre correctement dans le code existant. En effet, tout ou partie de la définition de cette action peut déjà figurer dans le code de l'agent. C'est pourquoi, lorsqu'un agent VDL reçoit une action, il doit systématiquement la comparer avec ses propres actions. Il inspecte son propre code et le modifie s'il trouve des correspondances entre l'action reçue et l'une de ses actions ou rajoute simplement la définition de l'action reçue à son code si aucune correspondance n'est trouvée. Ainsi, le composant maintient l'optimalité de son code et s'adapte durant les interactions grâce aux propriétés de réflexivité que lui confère VDL.

Nous dirons qu'un élément  $t_1$  contient un autre élément  $t_2$  (noté  $t_1 \supseteq t_2$ ) si et seulement si  $t_2$  apparaît comme un fils direct ou indirect de  $t_1$ , ou si  $t_2$  est  $t_1$  lui-même. Soit l'action  $A_1 = < P_1, M_1 >$  (resp.  $A_2 = < P_2, M_2 >$ ). Soit un composant comportant l'action  $A_1$  et recevant lors d'une interaction l'action  $A_2$ . Lorsque ce composant compare ces deux actions, il peut trouver des correspondances entre : soit leurs préconditions  $P_1$  et  $P_2$ , leurs sous-actions  $M_1$  et  $M_2$  ou bien les deux. Il construit alors une nouvelle description VDL globale  $A'_1$  remplaçant sa propre action  $A_1$  (ou alors rajoute directement l'action reçue  $A_2$  si aucune correspondance n'est relevée), construite suivant cet algorithme :

- Si  $P_1 = P_2$  alors  $A'_1 = \langle P_1, M_1 \cup M_2 \rangle$  ce qui signifie que, lorsque la précondition  $P_1 = P_2$  est vérifiée, les modifications à effectuer seront  $M_1$  et  $M_2$  ;
- Si  $P_1 \supseteq P_2$  alors  $A'_1 = \langle P_2, M_2 \cup \{A_1\} \rangle$  ce qui signifie que l'action  $A_2 = \langle P_2, M_2 \rangle$  est rajoutée aux modifications  $M_2$  à effectuer lorsque  $P_2$  est vraie ;
- Si  $M_1 = M_2$  alors  $A'_1 = \langle P_1 \vee P_2, M_1 \rangle$  ce qui signifie que les modifications  $M_1 = M_2$  seront effectuées si la disjonction  $P_1$  ou  $P_2$  est vraie.

Cet algorithme simple donne un exemple des manipulations qui peuvent être effectuées grâce aux propriétés réflexives de VDL. Il montre qu'un composant VDL peut raisonner sur son code, l'optimiser, ou l'enrichir avec de nouvelles définitions lors de ses interactions. Il montre également que les actions échangées peuvent être utilisées pour apprendre de nouveaux comportements d'autres composants.

Notons que les sous-actions  $M_1$  peuvent représenter elles-mêmes du code pour l'envoi de message (*i.e.* contenir un *send-message*). Ainsi, l'agent peut apprendre de nouvelles interactions en cours d'exécution et donc construire dynamiquement les fonctionnalités d'interaction qui lui permettent d'interopérer avec ses pairs.

## 5 Conclusions et perspectives

Dans cet article, nous avons proposé un modèle pour la programmation de composants réflexifs et dialogiques. Nous avons tout d'abord présenté le langage VDL qui permet de sémantiser toutes les données et les actions des composants, de les rendre actifs (réactifs et proactifs) et capables de raisonner sur leur propre code. Nous avons ensuite présenté un modèle de requêtes qui permet de modéliser une large classe de questions qu'un utilisateur humain peut poser aux composants à propos de leur exécution. Nous avons présenté brièvement les fonctionnalités d'interaction des agents VDL puis nous avons montré comment le modèle de requêtes et les capacités de raisonnement des composants peuvent être utilisées pour permettre aux agents d'interopérer en utilisant un dialogue à l'aide de requêtes formelles. À l'aide d'un algorithme simple de maintien du code, nous avons montré que nos composants étaient capables d'apprendre, de regarder et de modifier leur propre code en réponse à des interactions.

Notre objectif est d'utiliser ce modèle pour l'interopérabilité entre les services dans le cadre des systèmes complexes en informatique en ligne sur l'internet. L'utilisation de VDL permet au système dans lequel les agents s'exécutent d'évoluer dynamiquement au fur et à mesure d'interactions entre eux et avec l'utilisateur. Pour ce faire, il est nécessaire de combiner les agents VDL à des services web sémantiques capables de se composer automatiquement. Les agents VDL peuvent alors être vus soit comme surcouche descriptive et active greffée à un compo-

tant (et lui conférant les propriétés dialogiques et réflexives de VDL), ou bien comme un service médiateur fournissant des traductions pour les composants en interaction.

Dans la réalité, c'est probablement la composition des deux visions qui résoudra cette combinaison. Dans une telle approche, l'utilisateur interagirait avec un composant médiateur VDL ayant connaissance des autres composants, eux-mêmes munis d'une surcouche VDL leur permettant de communiquer avec le médiateur.

## Références

- [1] JADE : Java Agent DEvelopement Framework. <http://jade.tilab.com>.
- [2] The FIPA ACL Message Structure Specifications. <http://www.fipa.org/specs/fipa00061/>, 2002.
- [3] J.F. Allen, D.K. Byron, M. Dzikovska, G. Fergusson, L. Galescu, and A. Stent. Towards Conversational Human-Computer Interaction. *AI Magazine*, 2001.
- [4] T. Berners-Lee. *Weaving the Web*. Harper San Francisco, 1999.
- [5] E. Christensen, F. Curbura, G. Meredith, and S. Weerawarana. Web services description language (wsdl). <http://www.w3.org/TR/wsdl>, 2001.
- [6] David Martin *et al.* OWL-S : Semantic Markup for Web Services. Technical report, DAML Organization, 2004.
- [7] A. Doan, J. Madhavan, P. Domingos, and A. Halevy. Learning to Map between Ontologies on the Semantic Web. In *Proc. 11th International WWW Conference*, 2002.
- [8] A. El Fallah, S. Haddad, and H. Mazouzi. Protocol Engineering for Multi-agent Interaction. In *Proc. MAAMAW'99*, pages 89–101, 1999.
- [9] N. Guarino. Semantic Matching : Formal Ontological Distinctions for Information Organisation, Extraction, and Integration. *A multidisciplinary approach to an emerging information technology. International summer school*, pages 139–170, 1997.
- [10] M. Klein. Combining and Relating Ontologies : an Analysis of Problems and Solutions. In *Proc. 17th International Joint Conference on Artificial Intelligence, Workshop : Ontologies and Information Sharing, Seattle, USA*, 2001.
- [11] H. Lauren, D. Roman, and U. Keller. Web services modeling ontology - standard (wsmo-standard). <http://wsmo.org/2004/d2/v0.2/>, March 2004.
- [12] P. Maes. Agents that reduce workload and information overload. *Communications of the ACM*, 37(7), 1994.

- [13] J. Meseguer. Rewriting Logic and Maude : Concepts and Applications. In *Proc. RTA 2000*, pages 1–26, 2000.
- [14] R. Mizoguchi, J. Vanwelkenhuysen, and M. Ikeda. Task ontology for reuse of problem solving knowledge. In *Proc. 2nd international conference on building and sharing of very large-scale knowledge bases*, 1995.
- [15] D.A. Norman. *The invisible computer : why good products can fail, the personnel computer is so complex, and information appliances are the solution.* MIT Press, Cambridge, 1998.
- [16] N. Noy and M. Musen. Algorithm and Tool for Automated Ontology Merging and Alignment. In *Proc. AAAI-00 Conference. Austin. USA*, 2000.
- [17] N. Sabouret. *Étude de modèles de représentation, de requêtes et de raisonnement sur le fonctionnement des composants actifs pour l'interaction homme-machine.* PhD thesis, Université Paris-Sud, 2002.
- [18] N. Sabouret. Implémenter des composants actifs dans le web sémantique. In *Proc. Journées Francophones de la Toile (JFT) 2003*, pages 49–58, 2003.
- [19] N. Sabouret and J.P. Sansonnet. Etude d'un modèle de requêtes sur le fonctionnement de composants actifs. In *Proc. Journées des Nouveaux Modèles de Raisonnement (JNMR) 2001*, pages 173–186, 2001.
- [20] N. Sabouret and J.P. Sansonnet. Un langage de description de composants actifs pour le web sémantique. *Revue RI<sup>3</sup>*, 3(1) :9–36, 2003.
- [21] J.R. Searle. *Speech Acts.* Cambridge University Press, 1969.

# De la conception écologique à une écologie de la conception : un nouvel univers de pertinence et un cadre conceptuel général de la conception

De Coninck, Pierre, Ph. D., Professeur agrégé,  
École de design industriel, Faculté de l'aménagement, Université de Montréal.  
CP 6128, succ. Centre-Ville, Montréal (Québec) H3C 3J7 Canada.  
Courriel : [pierre.de.coninck@umontreal.ca](mailto:pierre.de.coninck@umontreal.ca)

---

**Résumé :** Cet essai montre ce que les sciences de la complexité et la systémique apportent au processus de conception de projets. La présentation d'un «*univers de pertinence*» suggère que le sujet (dans le cas présent, le concepteur) est en constante inter-rétro-action avec l'objet, le projet et l'environnement. La tension qui s'exerce entre ces quatre pôles alimente un processus récursif qui se traduit par l'émergence en continu de nouveaux projets, mais aussi nouvelles connaissances et pratiques. Dans un second temps, un *cadre conceptuel général* permet de soutenir l'élaboration de projets transdisciplinaires ainsi que toute intervention afférente.

**Mots clés :** conception, design industriel, complexité, systémique, projet.

---

## Introduction

En ce début du XXI<sup>e</sup> siècle, le monde occidental est confronté à une situation de crise. Cette crise se retrouve tant sur le plan social et économique (crise de l'État-Providence, régression démocratique, mondialisation des marchés, répartition des richesses, etc.) que sur le plan cognitif (pluralisme, limites des rationalités classiques dans les situations complexes, etc.). Cette nouvelle perspective ébranle les valeurs qui ont guidé la société occidentale pendant plusieurs siècles et provoque une profonde remise en question de la place et du rapport que l'humain entretient globalement avec son environnement (développement durable, réchauffement de la planète, design durable, consommation responsable, simplicité volontaire). Ces répercussions se font sentir tant au niveau local que global.

Ainsi, depuis les deux dernières décennies, nous assistons à de véritables travaux de restructuration du savoir et des processus décisionnels (Klein, 1990; Le Moigne, 1990; Stengers et Schlanger, 1989; Genelot, 1992; Bonami et al., 1996; Bourgeault, 1999). La recherche d'un «métissage disciplinaire» (de Béchillon, 1994) a permis de passer d'une juxtaposition des savoirs (pluridisciplinarité) à l'interdisciplinarité, à partir des années 1950, puis à la transdisciplinarité, à partir des années 1970 (Casenave et Nicolescu, 1994; OCDE, 1972). Selon ce dernier courant de pensée, la connaissance ne se déploie plus à l'intérieur d'un plan déterminé et continu d'un possible unique et universel – l'*univers*, dans lequel il existe une solution idéale à tout problème (le *One best way* de Le Moigne, 1990) – mais plutôt à l'intérieur d'un «espace des possibles» (De Coninck, 1993), c'est-à-dire un espace dans

lequel plusieurs façons de concevoir et de résoudre une problématique coexistent – le *plurivers* de Morin (1990). Dès lors, la résolution d'un problème ne peut plus se baser sur le rationnel d'une seule discipline (rationalité limitée, March & Simon, 1969). Qui plus est, non seulement devra-t-elle concilier les autres rationnels scientifiques mais aussi considérer des dimensions comme, par exemple, les valeurs et les diverses visions du monde que défendent les différents acteurs (Kuhn, 1983). L'approche systémique et complexe a engendré un basculement dans notre façon d'agir et de penser. Aujourd'hui toute action doit articuler trois dimensions : un mode de raisonnement qui intègre le paradoxe, l'incertitude et la complexité ; une modélisation des problématiques en tant que systèmes à la fois ouverts et fermés; et une action prenant en compte les deux dimensions précédentes (Bonami, et al., 1993; Le Moigne, 1990).

Après avoir rappelé brièvement l'influence de l'approche de la science classique quant à la manière de concevoir un objet, nous terminerons la première partie de cette réflexion, par la présentation d'un *univers de pertinence* s'appuyant sur les sciences de la complexité et la systémique. Par la suite nous esquisserons un *cadre conceptuel général* visant à soutenir l'élaboration de projets transdisciplinaires ainsi que toute intervention afférente.

### ***L'univers de pertinence «objet/projet/sujet/environnement» : vers une (re)connaissance de la conception complexe***

D'une certaine façon, les designers subissent le contrecoup de la crise évoquée en introduction. En effet, les designers, comme tout professionnels de la conception, tels les ingénieurs et architectes, sont des initiateurs de changement *dans* et *de* la société. Ils doivent anticiper les nouveaux besoins et proposer de nouvelles réponses aux contraintes de l'usage et de la forme en considérant les technologies nouvelles et à venir. En outre, dans une perspective de développement durable et d'éco-conception, ils doivent également considérer les aspects contextuels (sociaux, environnementaux, politiques, économiques) présents et à venir du projet sur lequel ils se sont attelés. D'une certaine manière, ces concepteurs sont des *traducteurs* : ils sont à l'écoute de l'information qui provient des collègues et partenaires, de la société, de leurs environnements, pour ensuite la réifier, la cristalliser, en un produit ou en un système de produits à partir de leur vision du monde. Ainsi, si le design est souvent perçu comme une activité de résolution de problème, une activité créatrice, une activité systémique, une activité de coordination, le design est aussi une activité de *projection* dans un monde perçu complexe, à venir et incertain. Cette dernière activité est capitale dans le processus de design.

Lorsqu'un designer conçoit un produit, il effectue des choix sociaux et techniques. Ces choix seront déterminants non seulement pour tout le cycle de vie de ce produit, c'est-à-dire depuis l'extraction des matières premières jusqu'à son élimination, en passant par sa fabrication, sa diffusion et son usage, mais aussi pour les usagers (directs et indirects) et les environnements (sociosphère, biosphère, technosphère, noosphère). Ces choix déterminent, en quelque sorte, un «contexte de référence» à partir duquel toute la justification et toute la légitimation des actions subséquentes s'arrimeront. C'est pourquoi les prémisses sur lesquelles l'acte de conception se fonde suscitent un intérêt grandissant chez les chercheurs.

Or, entre la naissance de cette discipline (fin du XIX<sup>e</sup> siècle) et les années 1970, le design industriel a été fortement influencé par les notions et méthodes de la science classique. C'est ainsi que les théoriciens du design ont cherché à donner des lettres de noblesse à cette

discipline en lui associant des dimensions fonctionnalistes et rationnelles. Pour Ralph Caplan : «Le design est l'art d'aménager les matériaux et les événements dans une forme planifiée» (*By Design*). Alors que pour Victor Papanek (1974) : «Le design est l'effort conscient pour instaurer un ordre significatif». Pour sa part, H.A Simon a perçu une autre dimension du design alors qu'il affirmait que «tout acte intentionnel de création et de transformation de l'être humain et/ou de l'environnement est un acte de design» (Simon, 1969, p. 129).

Ces définitions illustrent bien toute l'ambiguïté que cette discipline entretient en étant autant une activité de *dessin* (plan, forme) qu'une de *dessein* (aménager, ordonner, planifier) (Quarante, 1984; Dorst, 1997). Le design est un type spécifique d'action dans le sens où il est à la fois *intentionnel* (Nelson et Stolterman, 2003) et *projectif* (Jonas, 2001).

La science classique se préoccupait essentiellement de l'observation et de la description des systèmes naturels ou abstraits. Son but était de découvrir ce «réel», que l'on suppose donné et ordonné et qui se cache à la raison, pour ultimement réduire l'ignorance et aller au-delà des apparences. Aussi, toute parcelle de réel se plaçant au travers du chemin du sujet-observateur, deviendra «objet» de recherche, comme le soulignent sans ambiguïté les racines étymologiques de ce concept. La relation que le sujet se doit de garder face à l'objet a été abordée par toutes les sciences car elle était l'enjeu de sa neutralité et de son objectivité; deux principes fondateurs du paradigme de la science classique analytique.

Mais récemment, la science de la complexité et l'approche systémique ont indiqué qu'il devenait difficile de maintenir cette position épistémique car «objet» et «sujet» sont consubstantiels (Morin, 1990, 2002, 2004; Le Moigne, 1990):

*«Le sujet, dans l'acte même de connaissance, produit l'objet. Inversement l'objet produit le sujet. (...) parce qu'il est au fondement de toute connaissance, c'est-à-dire, la possibilité pour un sujet de connaître quelque chose de la réalité. Sans objet, il n'y aurait ni connaissance ni sujet connaissant. (...) Objet et sujet sont désormais inséparables. La connaissance selon Morin, implique toujours la relation objet/sujet. Aussi, pour lui est-il extrêmement important, (...) de réarticuler cette relation sans cesse disjointe, désamorcée par la pensée simplifiante»* (Fortin, 2000, pp. 142-143.)

Sujet et objet sont donc mutuellement en constante inter-rétro-action. Mais cette relation bi-polaire ne peut suffire à saisir la totalité du champ d'intervention de la science ni du design. En effet, le sujet, en concevant un objet (produit, service, limites d'un système), lui attribue invariablement des finalités (Le Moigne, 1990). Cette notion de projet ou d'intentionnalité se retrouve à même la définition de «modélisation» (Le Moigne, 1990, p. 5); le projet fait donc partie intégrante du processus de connaissance. Aussi, au couple sujet/objet, il convient d'associer le projet.

Comme le souligne pertinemment Levy (1988), «le projet est la trame socio-environnementale à l'intérieur de laquelle l'acte de concevoir se produit». Le projet devient donc un processus de création, de construction et de réalisation d'une finalité particulière, qui en substance traduit soit des désirs, des besoins ou des valeurs, et qui est directement relié au sujet et à l'objet. C'est pourquoi selon cet auteur, il est d'une «extrême importance intellectuelle d'assurer que le processus de design soit un processus d'auto-design» (1988).

Par ailleurs, il conviendrait de ne pas oublier que le système tripolaire «*objet/sujet/projet*» est un système actif et évoluant qui prend son sens à l'intérieur d'environnements divers (bio-sphère, techno-sphère, socio-sphère, noo-sphère) et qui possède ses finalités propres, tout en étant consubstancial à ses environnements (Simondon, 1958; Roqueplo, 1983).

Si nous appliquons ici le «principe de reliance» de Morin (2004), il appert que le design d'un objet ou d'un système d'objets ne peut être dissocié ni du sujet, ni du projet, ni de l'environnement. L'objet fait partie intégrante du sujet, du projet et de l'environnement. S'il est possible de les distinguer, il devient par contre impossible de les dissocier et de les disjoindre, sous peine de simplification et de mutilation (Morin, 1982, 1990, 2002).

Ces quatre pôles sont de fait indissociables. Le quadripôle que nous venons de décrire s'appuie sur une épistémologie constructiviste et systémique (Morin, 1982; Le Moigne, 1990) qui agit comme cadre épistémique (Piaget, Garcia, 1983). Ce tétraèdre définit un «univers de pertinence» de la conception de projet (*Figure 1 : L'univers de pertinence*).

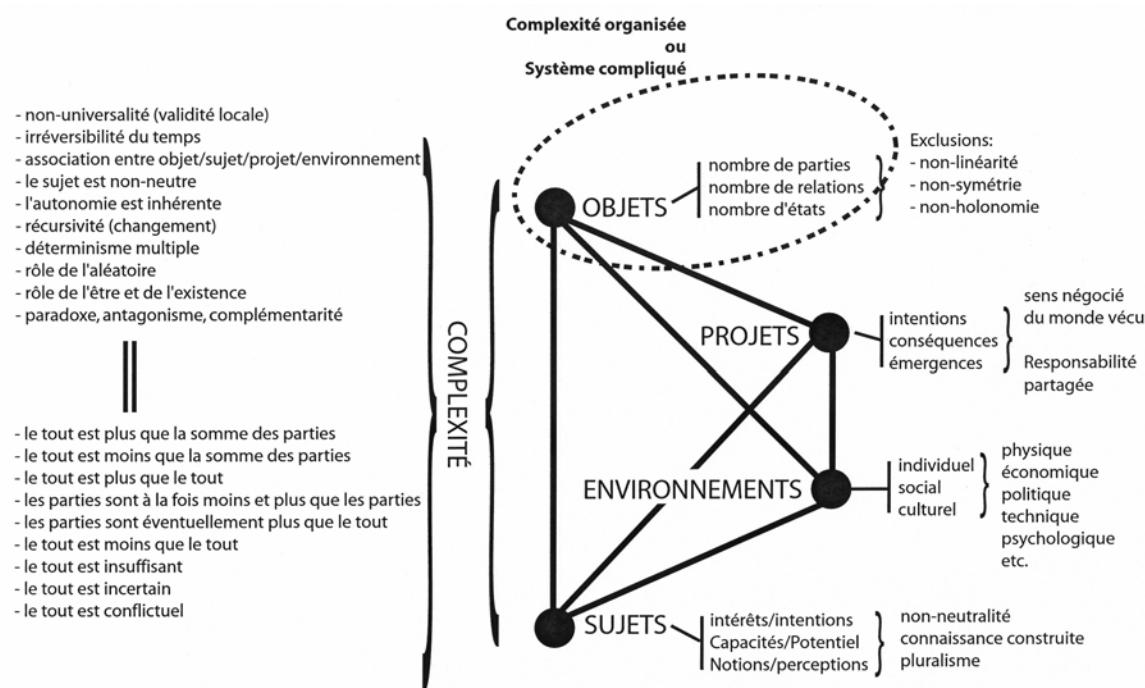


Figure 1 : L'Univers de pertinence (inspiré de Flood, R.L. (1987) et Morin, E. (1982))

Au sein de cet univers de pertinence, il apparaît que la définition la plus complète et pertinente du design est la suivante :

«Le design est la manipulation (par les êtres humains) de tout être; c'est-à-dire de tout phénomène (vivant et non-vivant), de tout objet, tout mécanisme, tout système et toute procédure.

*La finalité de cette manipulation de ces êtres est la création de concepts nouveaux et d'expériences nouvelles dont la totalité change le monde de manière récursive» (Levy, R)<sup>1</sup>.*

Cet univers de pertinence met en exergue que penser concevoir avec une approche «résolution de problème», c'est oublier que le design est moins une question de s'ajuster à un *statu quo* que de réaliser, de concevoir de nouveaux possibles et de découvrir nos réactions vis-à-vis ces derniers. Problème et solution sont intimement liés (De Coninck, 1997). Il apparaît dès lors qu'on ne peut plus concevoir, «designer», en ne pensant qu'à un problème perçu comme étant «prédéterminé», car ce dernier doit être dorénavant perçu comme «construit». De fait, «la compréhension des interrelations à l'intérieur du système sujet/objet/environnement/projet est la base sur laquelle il est possible d'assurer que l'acte de conception en sera un existentiellement créatif» (Levy, 1988). Si le scientifique crée des abstractions, des représentations, des modèles, des théories d'un monde dit «réel», et donc sur des objets déjà existants, le designer, quant à lui, crée des abstractions, des représentations, des modèles d'un monde «virtuel» en réponse à des besoins, une vision du monde (prospective). Il crée essentiellement des abstractions d'abstraction, en s'appuyant occasionnellement sur le savoir scientifique.

La formulation de la problématique, des objectifs et des besoins est devenu un enjeu majeur du processus de conception. Ces derniers sont chargés de présupposés, voire de préjugés (implicites), qui empêchent souvent de voir, de concevoir et d'explorer d'autres avenues possibles de résolution de problème. Il convient dès lors de considérer tout objet ou système technique comme la manifestation de différents cheminements décisionnels et de différentes dynamiques technologiques, économiques et sociopolitiques. Et il importe également de reconnaître la contribution des divers partenaires avec lesquels les designers industriels interviennent.

Ce virage ne nécessite pas seulement un changement d'ordre logistique mais aussi et surtout, un changement de mentalité (Crozier, 1995) qui impose un décloisonnement des connaissances et des champs d'activités. Il demande de définir tout problème ou toute solution au-delà d'une perspective disciplinaire, de considérer les multiples facettes que recèle chaque problématique et d'envisager que plusieurs solutions puissent apparaître acceptables à partir d'un cadre de référence dépassant le programme ou la discipline. Il convient donc de développer une vision systémique et complexe du processus de conception de projets.

Ce virage soulève d'autres enjeux majeurs. Ainsi, l'énonciation du «problème» que l'on perçoit dans la réalité est un problème en soi. Il devient nécessaire d'expliquer non seulement les critères qui ont été retenus et leur hiérarchisation mais aussi les acteurs susceptibles de participer aux diverses étapes du processus décisionnel dont, entre autres, celle de la conception, de l'évaluation et du choix. La structure décisionnelle dans laquelle le processus s'opère est donc également importante. Pour cela, il faut comprendre les différents niveaux de justification et de légitimation de la décision et évoquer : 1) les intervenants qui sont amenés à prendre une décision; 2) ceux qui subiront ou seront affectés par la décision; 3) la nécessité d'articuler, ou non, le premier niveau au second (De Coninck, et al., 1996).

---

<sup>1</sup> Cette définition n'avait jusqu'à présent fait l'objet d'aucune publication. Je remercie Monsieur Levy, professeur à l'École de design industriel de l'Université de Montréal, de 1972 à 1992, et également mon directeur de mémoire et de thèse, de m'avoir permis de la publier ici.

C'est donc au sein de cette conjoncture que le développement durable puise tout son potentiel d'éclairage et de compréhension de la complexité de la situation actuelle pour l'ensemble des réseaux de concepteurs, dont celui du design industriel. En effet, en visant l'équilibre entre croissance économique, protection de l'environnement et développement des personnes et des communautés, le développement durable peut être perçu comme un modèle de changement technologique et social; et c'est à ce titre qu'une multitude d'initiatives et d'actions tant au niveau local, régional que national s'y appuient. Toutefois, deux constats s'imposent : 1) il n'existe pas parmi les acteurs du développement durable de consensus quant à ce qui le caractérise : sa définition, les dimensions qui le composent (sociale, économique, culturelle, etc.) et les stratégies (moyens) qui y participent (responsabilisation, concertation, etc.); et 2) le développement durable confronte le champ de la conception des produits à la faiblesse de son cadre théorique et méthodologique face aux enjeux socio-techniques. En effet, les pratiques de conception traditionnelle, centrées sur la satisfaction des besoins du consommateur (lui conférant un rôle plutôt passif : information, sensibilisation, consommation), ne peuvent concevoir des pratiques favorisant l'émergence d'acteurs (et donc un rôle actif : co-création, co-formulation de projet de design) à la recherche de contrôle et de pouvoir sur ce qu'est ou pourrait être le projet ou encore les interrelations à un nouvel environnement.

### ***Un cadre conceptuel général de la conception de projets***

Depuis la publication du rapport Brundtland, les questionnements et la recherche concernant l'opérationnalisation du développement durable sont de plus en plus au centre des préoccupations des disciplines. Que ce soit en sociologie, en ingénierie, en santé publique, en géographie, en design, un virage important est amorcé, réclamant une vision globale et complexe, ainsi qu'une approche multidisciplinaire, voire transdisciplinaire.

Le concepteur, en général, et le designer industriel contemporain, en particulier, se situe résolument à la croisée de divers intérêts. Aussi, est-il important de faire appel à la contribution de multiples partenaires provenant de secteurs variés et de ne pas passer outre la considération des perspectives de la population à desservir. D'une certaine manière, les designers industriels sont appelés à orchestrer ces différentes visions du monde, ces attentes et ces projets dans le but de répondre aux besoins et aux aspirations des communautés actuelles et à venir. Face à la complexité des problèmes de conception de produits, il est nécessaire de développer des connaissances scientifiques et des pratiques transdisciplinaires (De Coninck, 1996) dans le champ de la conception des produits et services.

Comme nous l'avons mentionné auparavant, la pratique du design englobe toutes les activités de conception, de création de produits et de services qui ont pour but d'améliorer non seulement les usages et les comportements, mais aussi le cadre de vie et l'environnement et ce, dans un but d'augmenter le bien-être. Loin de se réduire à une activité de recherche formelle, stylistique, le design est aujourd'hui amené à dépasser l'approche «résolution de problèmes» (*solving problem*) pour se placer plus en amont, soit au niveau où l'on considère l'établissement des critères qui amèneront à la formulation du problème (*problem setting*) et, ultimement, proposer une réponse pertinente et globale à la réalité perçue complexe, sous forme de produits, de systèmes de produits ou encore de services. Le passage à l'analyse de la commande devient un enjeu important. Toutefois, ce changement requiert également un changement de perspective et l'adoption de nouveaux concepts et une nouvelle approche conceptuelle basée sur la vision systémique et une approche de la complexité.

Cet intérêt pour une analyse en amont du problème s'appuie sur différents aspects émergents touchant les quatre champs de la connaissance, que sont les champs épistémologique, téléologique, méthodologique et ontologique. Ainsi, au niveau du champ *épistémologique*, les sciences de la complexité et la systémique constituent le fondement d'un nouveau cadre paradigmique (voir Figure 1). L'émergence des sciences de la conception montre que l'approche traditionnelle fonctionnelle ne permet plus de saisir l'ensemble des enjeux qui caractérisent notre époque. Au niveau du champ *téléologique*, la vision du développement durable auprès de plusieurs sphères d'activités, qu'elles soient sociales, économiques (consommation responsable) ou encore de production (conception, fabrication, distribution, protection environnementale) agit comme un attracteur étrange vers lequel les actions semblent vouloir converger (Marchand, et al., 2004). Au niveau *ontologique*, la mort du produit est pensée dès la conception de ce dernier (analyse de cycle de vie) et les produits vont même jusqu'à se dématérialiser (Diani, 1991; Quarante, 1984, 2002). Enfin, au niveau *méthodologique*, nous avons vu que les pratiques et les processus de conception sont appelés à se modifier (Diani, 1991; Buchanan & Margolin, 1995; Laurel, 2003).

Ces changements au sein des quatre champs de connaissance sont importants car ils provoquent un questionnement de fonds non seulement de la pratique du design mais aussi de la légitimité même de ses interventions, ses stratégies et ses motivations. La vision et les pratiques traditionnelles du design ne sont plus adéquates car elles ne répondent plus aux enjeux du développement durable et celui du développement social.

Pour assumer ce rôle d'animateur et d'expert dans le champ de l'éco-conception, les designers doivent pouvoir se reposer sur un cadre conceptuel général qui leur permettra de réaliser des projets, de mettre en œuvre des procédures de participation, et ainsi répondre aux enjeux cruciaux du développement durable. Ce cadre suscitera une philosophie commune ainsi qu'une action concertée au sein de la profession, tout en permettant aux designers de se positionner par rapport à ses partenaires. Ce fondement de collaboration et d'échange entre les partenaires favorisera l'émergence de processus de co-création de projets et de co-formulation de solutions appropriées et de projets locaux; processus se fondant sur le dialogue et susceptibles de mener à une plus grande compréhension mutuelle, à l'émergence de nouvelles perspectives ainsi qu'à une responsabilisation accrue des acteurs.

Ce cadre conceptuel général (Figure 2) repose essentiellement sur quatre systèmes directeurs en interaction qui devraient être considérés lors de la réalisation de tout projet (Le Moigne, 1984; Levy et al., 1997) : 1) la *sphère d'intervention* réfère au processus de concrétisation des projets sur le terrain. Il s'agit du lieu où se transforment les acteurs, les connaissances (professionnelles, sociales, etc.), les pratiques et les intentions; 2) la *sphère d'intelligence* constitue un système collectif et interactif des connaissances soutenues et développées par les participants impliqués. Cette sphère contient essentiellement l'information et les connaissances auxquelles tous les participants devraient avoir accès afin de construire des compréhensions multiples des enjeux considérés; 3) la *sphère de finalisation* est le lieu où les idéologies, les missions, les orientations, les stratégies et les tactiques sont explicitées en vue de conceptualiser les projets. En plus d'énoncer les intentions, elle met aussi l'emphasis sur les enjeux, les contextes et les activités de conception mobilisés pour la formulation des projets; 4) la *sphère de pilotage* comprend les ressources et l'expertise pour aider les individus et les groupes à formuler et à expliciter les multiples facettes de leurs points de vue se rapportant aux problématiques considérées. Elle intègre les représentations

cognitives des partenaires, les orientations méthodologiques privilégiées, les capacités et les ressources disponibles et les finalités projetées.

Ce cadre conceptuel général, inspiré du modèle canonique de la décision de Le Moigne (1990, p. 131), se veut évolutif et convivial et s'articule autour d'un processus où les multiples intérêts des partenaires sont considérés (co-création de projets) et des terrains d'entente négociés afin d'arriver à des projets acceptables pour ces derniers et pour les consommateurs-usagers concernés (co-formulation de solutions acceptables). Ce cadre conceptuel général pourrait permettre de souligner quelques principes directeurs ainsi qu'une procédure pour identifier, structurer, concevoir et mettre en oeuvre des projets transdisciplinaires.

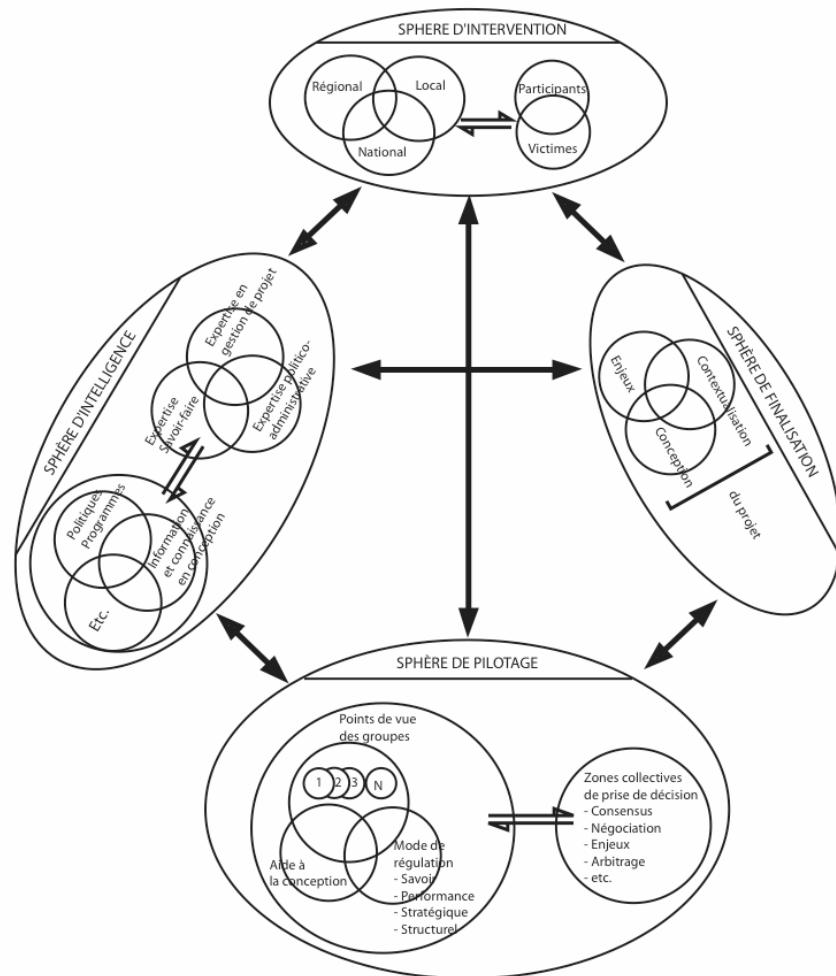


Figure 2 : Le cadre conceptuel général (d'après Levy, Denis, De Coninck, Gendron, 1997)

### **Conclusion : pour une écologie de la cognition, de la conception et de l'action**

Face à la complexité des problèmes de conception de produits, il est nécessaire de développer des connaissances scientifiques interdisciplinaires et de chercher à les intégrer à celles qui se rapportent à la culture des individus, des groupes et des communautés (et donc

de tendre vers une transdisciplinarité) afin de diminuer les impacts des produits sur les environnements (bio-physiques, usagers, communautés). Un cadre conceptuel implicite est à la base des justifications et des légitimations des actions entreprises par chacun des porteurs d'enjeux qui souhaitent tous contribuer au développement durable. Ce cadre offre une perspective qui leur est propre. Cette singularité de perspective empêche trop souvent l'élaboration d'une action globale convergente, malgré le partage d'un objectif commun déclaré. Plus souvent, ce cadre conceptuel constitue un obstacle au débat.

Bien que la concertation et l'approche intersectorielle apparaissent comme des stratégies pertinentes pour l'amélioration de la qualité des produits et qu'elles fassent l'objet d'un grand intérêt, les conditions de mise en œuvre sur le terrain d'une politique de concertation et d'une pratique intersectorielle requièrent de plus amples approfondissements.

Pour atteindre les objectifs du développement durable, il devient essentiel non seulement d'établir une nouvelle dynamique entre les décideurs, les promoteurs, les experts, les groupes d'intérêt et les citoyens mais aussi d'établir un lieu de confiance, où il sera possible de considérer et de débattre avant de décider. Ce lieu devra devenir un lieu de référence. Il faut donc convenir de structures et de processus qui permettront de mettre en scène des experts de différents secteurs d'activités, des représentants de groupes de pression, mais aussi des citoyens, sachant que tous ces intervenants se fonderont sur un référentiel paradigmique particulier, c'est-à-dire qu'ils considéreront, douteront et débattront à partir de leur propre système de valeur, d'idées et de conception du monde (De Coninck, 1996, 1997). La dynamique du cadre conceptuel général permet par le jeu de la confrontation des paradigmes, un certain décloisonnement des connaissances en vue d'une plus grande efficacité d'action en développement durable. En effet, la mise en perspective des quatre systèmes directeurs que sont les sphères d'intervention, d'intelligence, de finalisation et de pilotage permettra de modéliser et de théoriser les processus d'alignement des décisions émises par les modèles de gouverne démocratique ou latérale.

Il conviendrait par conséquent de se pencher plus particulièrement sur l'optimisation des mécanismes de concertation et d'approche intersectorielle en conception de produits et services pour le développement durable et social des communautés. Plus précisément, quelles sont les conditions de succès des processus de collaboration et d'échange entre partenaires (professionnels, décideurs, experts, groupes de pression, citoyens) pour la co-création et la co-formulation de projets ou de solutions. Les concepteurs sont dorénavant invités à établir de nouveaux liens de collaboration et de relations d'échange qui vont au-delà du schème traditionnel de pourvoyeur de services (Diani, 1991). Il ne s'agit plus de développer une rationalisation du choix d'une solution visant à convaincre la communauté et les consommateurs du bien-fondé du produit ou du service. Il convient plutôt de les intégrer dès l'étape de la définition de ce qui fait problème ou encore au moment de l'élaboration des solutions. Cette stratégie favorise un décloisonnement des connaissances (qu'elles soient sociales ou professionnelles) et, ultimement, une plus grande efficacité d'action en matière de développement durable.

Il s'agit également de déterminer comment les mécanismes de concertation viennent compléter les processus dominants d'aide à la décision basés, entre autres, sur des logiques réglementaires, économiques ou encore sur des connaissances objectives. En effet, les mécanismes démocratiques ne doivent pas être considérés comme un lieu de substitution au pouvoir formel, ni comme une panacée à la controverse, mais plutôt comme un lieu de référence ou, encore, comme un lieu de (re)formulation d'une problématique (à l'issue de

processus d'apprentissage et d'innovation). Il devient, par conséquent, un lieu duquel peuvent émerger des nouveaux choix potentiels de solutions répondant à un plus grand éventail de préoccupations sociales dans un contexte de transformation à l'échelle mondiale.

Si l'établissement d'un nouvel univers de pertinence permet l'élaboration d'une vision globale et systémique des produits et, ultimement, des produits écologiques et socialement acceptables, la présentation du cadre conceptuel général amorcera, quant à elle, une réflexion sur les bases d'une écologie de la conception complexe. Nous retrouvons ici un cadre propice à l'exercice d'une véritable «démocratie cognitive» que suggère Morin (2002, 2004).

L'enjeu du design est moins de devenir une science – une telle tentative serait vaine – que de voir comment les sciences peuvent venir en aide à cette discipline. À titre d'exemple, nous pourrions analyser comment les quatre champs de la connaissance enrichissent le processus de conception des produits et des services, dans une perspective de développement durable? ou encore, pour reprendre les termes de la définition du design proposée par Levy, comment le design peut créer et planter de nouveaux possibles et de nouvelles expériences susceptibles d'établir de nouveaux rapports au monde?

Si l'univers de pertinence a montré qu'il convenait de se placer en amont de la résolution de problème, le cadre conceptuel général a, quant à lui, souligné le caractère négocié de toute action. Selon le principe de l'écologie de l'action, «toute action commence à échapper à l'intention (l'idée) de ceux qui l'ont déclenchée dès lors qu'elle entre dans le jeu des inter-rétro-actions du milieu où elle intervient» (Morin, 1993, p. 152). Cette idée, émise pour la première fois par H. Arendt (1983), souligne que la patience s'avère de mise et que des conséquences, non prévisibles aujourd'hui, surgiront dans le milieu politique, social, environnemental et technique de demain. Le pari que nous pouvons prendre aujourd'hui est que l'établissement d'une écologie de la conception sera un atout majeur pour la mise en œuvre d'un développement durable et équitable à moyen terme.

## ***Bibliographie***

- Arendt, H. (1983) *Condition de l'homme moderne*. Calmann-Lévy: Paris.
- Blandin, B. (2002) *La construction du social par les objets*. PUF: Paris.
- Bonami, M., de Henin, B., Boqué, J.-M., et Legrand, J.-J. (1993) *Management des systèmes complexes*. De Boeck Université: Bruxelles.
- Bourgeault, G. (1999) *Éloge de l'incertitude*. Bellarmin: Montréal.
- Buchanan, R., Margolin V. (Eds) (1995) *Discovering Design*. The University of Chicago Press: Chicago.
- Callon, M. (1997) «Exploration des débordements et cadrage des interactions: la dynamique de l'expérimentation collective dans les forums hybrides», Séminaire du programme Risques collectives et situations en crise, *Actes de la huitième séance*, Grenoble, CNRS, 12 juin, pp.57-98.
- Cazenave, M., et Nicolescu, B. (1994) *L'Homme, la science et la nature*. L'Harmattan: Paris.
- De Béchillon, D. (1994) Une réflexion critique sur le concept de transdisciplinarité. *Trans-disciplines*, No 9. IRSAM/Université de Pau et des Pays de l'Adour: Pau, pp. 2-8.
- De Coninck, P. (2000) «Vers un cadre de planification participative : l'exemple de la consultation par consensus informé». Fiévet, C. *Invention et réinvention de la citoyenneté*. Actes du colloque international de Pau. Université de Pau et des Pays de l'Adour. Éditions Joëlle Sampy : Pau, pp. 471-480.

- De Coninck, P. (1997) «L'implication des citoyens ordinaires dans le processus d'aide à la décision en santé publique», *Ruptures*, vol. 4, no 2. Université de Montréal : Montréal, pp. 152-162.
- De Coninck, P. (1996) «De la disciplinarité à la transdisciplinarité : à la recherche d'une panacée ou d'une attitude ?» *Info-Stoper*, vol. 4, no1, Université de Sherbrooke : Sherbrooke, pp. 1-7.
- De Coninck, P. (1993) *Du mécanisme à la systémique : vers une modélisation complexe de la technique*. Thèse de doctorat. Université de Montréal : Montréal.
- Domenach, J.M. (1994) *La responsabilité. Essai sur le fondement du civisme*. Hatier: Paris.
- Dorst, K. (1997) *Describing Design : A comparison of paradigms*. Delf Univ. of Technology, Faculty of Industrial Design, Delf : Holland.
- Flood, R.L. (1987) «Complexity: a definition by construction of a conceptual framework». *System Practice*, 4(3), 177-186.
- Fortin, R. (2000) *Comprendre la complexité. Introduction à la Méthode d'Édgar Morin*. L'Harmattan/Les Presses de l'Université Laval.
- Genelot, D. (1992) *Manager dans la complexité*. INSEP Éditions : Paris.
- Klein, J.T. (1990) *Interdisciplinarity. History, Theory, and Practice*. Wayne State University Press : Detroit.
- Kuhn, T. (1983) *La structure des révolutions scientifiques*. Flammarion : Paris.
- Latour, B. (2003) *Un monde pluriel mais commun*. Éditions de l'aube : Paris.
- Latour, B. (1996) *Petite réflexion sur le culte moderne des dieux faïtiches*. Synthélabo/Les empêcheurs de penser en rond : Paris.
- Laurel, B. (Eds.) (2003) *Design Research. Methods and Perspectives*. MIT Press : Cambridge.
- Le Moigne, J.-L. (1990) *La modélisation des systèmes complexes*. Dunod : Paris.
- Levy, R. (1991) Critical Systems Thinking : Edgar Morin and the French School of Thought. *Systems Practice*, vol. 4, n°2, pp. 87-99.
- Levy, R. (1988) «Le projet : une projection de soi». *Informel*, 1, no 2, pp. 7-11.
- Levy, R., Denis, J.-L., De Coninck, P., Gendron, S. (1997) *Cadre conceptuel pour la formation de plans d'action à la Direction de la santé publique des Laurentides*. Juillet. 22p.
- Madge, P. (1997) «Ecological Design : A New Critique». *Design Issues*, vol. 13, no 2, pp. 44-54.
- Marchand, A., Walker, S., De Coninck, P. (2004) «Buying Time : Defining the Characteristics of Sustainable Consumption», *Creating a Culture of Sustainability, Highlands & Islands International Sustainable Development Conference and Exhibition : Conference Proceedings*, Novembre 3-5 (Accepted).
- Morin, E. (2004) *La méthode (6)*. Éthique. Seuil : Paris.
- Morin, E. (2002) *Pour une politique de civilisation*. Arléa : Paris.
- Morin, E. (1990) *Introduction à la pensée complexe*. ESF Éditeur : Paris.
- Morin, E. (1986) *La méthode (3)* La connaissance de la connaissance. Seuil : Paris.
- Morin, E. (1982) *Science avec conscience*. Fayard : Paris.
- Morin, E., Kern, A.-B, (1993) *Terre-Patrie*. Seuil : Paris.
- Morin, E., Le Moigne J.-L. (1999) *L'intelligence de la complexité*. L'Harmattan : Paris/Montréal.
- OCDE (1972) *L'interdisciplinarité. Problèmes d'enseignement et de recherches dans les universités*. OCDE : Paris.
- Papanek, V. (1974) *Design pour un monde réel*. Mercure de France : Paris.
- Piaget, J., Garcia, R. (1983) *Psychogenèse et histoire des sciences*. Flammarion : Paris.
- Quarante, D. (1984) *Éléments de design industriel*. Maloine S.A. Éditeur : Paris.

- Roqueplo, Ph. (1997) *Entre savoir et décision, l'expertise scientifique*. INRA : Paris.
- Roqueplo, Ph. (1983) *Penser la technique. Pour une démocratie concrète*. Seuil : Paris.
- Scolve, R. E. (2003) *Choix technologiques, choix de société*. Eds Charles Léopold Mayer : Paris. (Traduction française de *Democracy and Technology*, The Guilford Press : New York, 1995).
- Simon, H. A. (1969) *The Sciences of the Artificial*. MIT Press : Cambridge. (Traduction française : *Les sciences de l'artificiel*. Gallimard : Paris, 2004).
- Simondon, G. (1958) *Du mode d'existence des objets techniques*. Aubier : Paris.
- Stengers, I. et Schlangers, J. (1989) *Les concepts scientifiques. Invention et pouvoir*. Éditions La Découverte/Conseil de l'Europe/UNESCO : Paris.

# **Les paradoxes du regardeur, « ces biais par où les fictions s'introduisent »**

Myrielle Hammer,  
EVCAU, EAPVS  
14 rue Bonaparte  
75006 Paris  
myhamme@club-internet.fr

## Résumé :

Lorsque dans leurs toiles des artistes du Pop Art et de la Figuration Narrative s'interrogent sur leur pratique, qu'énoncent-il sous forme d'imagerie ? L'autoréférence, c'est celle qui évoque nos relations aux imageries de mass media. Elle met diversement en question notre regard : d'un point de vue théorique, d'un point de vue narratif. Elle ne se repère qu'à travers des parcours actifs de lecture. Entre le tableau, le regard, le peintre, les liens se transforment et deviennent paradoxes. Ces liens paradoxaux sont étudiés dans ces différentes œuvres dont la dernière est une photographie qui reprend à un autre niveau les opérations précédemment inventoriées.

## Mots-clefs :

Regardeur, Duchamp, Lichtenstein, Warhol, Rancillac, Huyghe, tableau, imagerie de mass media, Vache qui rit, photographie, décalage temporels.

## ***1- Des regardeurs et des tableaux***

« **Ce sont les REGARDEURS qui font les tableaux** »(1) proclame-t-on toujours avec Marcel Duchamp. Au fil des discours et des textes, cette formule, en vient à désigner autant les tableaux que les regardeurs, au risque de tout confondre. De son contexte d'origine jusqu'à nos jours, un glissement sémantique, en effet, s'est opéré, car cette « assertion » a été curieusement tronquée...

Duchamp, avait immédiatement précisé ce qu'il voulait dire, ajoutant :

**«On découvre aujourd'hui Le Greco ; le public peint ses tableaux trois cents ans après l'auteur en titre. »**

A brûle pourpoint, dans le sillage de cet auteur ou artiste, quelques œuvres du Pop Art redonnent force et sens à cette distinction entre le regardeur et

ces temps décalés. Et plutôt que de réfléchir sur l'art et l'auto-référence (2) en général, de me risquer sur le terrain des abstractions, voici à travers quelques exemples, une histoire de regardeur plus ou moins actif et présent du côté du tableau.

Mettre le tableau « Pop » sur la sellette à propos d'autoréférence, c'est déjà admettre un **dialogue corrosif** entre l'image et la peinture et faire passer ces relations par le crible de l'analyse, par des mots, des phrases, c'est redonner du corps, de la matière et du sens, à ce qui se voit, s'absorbe, se butine en guise de réflexion.

Lorsque les reproductions des œuvres sont absentes, dans le dans le corps de ce texte (3), le fil conducteur de cette étude correspond à un travail de description plus spécifique : il me faut reconstituer verbalement le tableau puis, avec ces premiers repères, tenter de le décortiquer en tout sens. Et pour préciser encore quel est le sens de cette recherche, il faut indiquer quelles sont les interférences diverses qui se nouent entre les images et la peinture. Certaines interférences sont à situer du côté d'une théorie: c'est le cas les deux premières œuvres, celles de Lichtenstein et de Warhol. Elles font retour sur elles-mêmes, montrent et démontrent. D'autres mènent vers des équivalents de discours à partir de comparaisons qui deviennent métaphores et dénonciation. Le jeu entre image et peinture chez Rancillac est provocation. L'image sert d' alibi. Mais cet alibi n'est pas neutre : il est valorisé comme autoréférence. On assiste avec Pierre Huyghe à une sortie du tableau, qui nous conduit vers une théâtralisation de notre regard réactivé sur le plan de la vie quotidienne. La remarque de Duchamp devient « prophétie ».

Mon travail de description n'est pas de rechercher un modèle à travers ces œuvres, mais de comprendre la nature de l'autoréférence à cette époque, celle des années 60 aux Etats-Unis et en France. Du Pop Art à la Figuration narrative. D'une aire culturelle à une autre. Afin de sortir des polémiques toujours sensibles aujourd'hui, (4), ce recours à une œuvre plus récente, celle de Pierre Huyghe, permet d'élargir mon point de vue. Il ne s'agit plus d'assortir deux tendances très voisines d'un certains nombre de caractéristiques convergentes et divergentes mais de savoir que l'on peut repérer des prolongements et des mises en cause de cet art des années 60

assimilant et réagissant, quant à lui, aux systèmes d'objets et d'images appartenant à une société de consommation qui par ailleurs a beaucoup changé.

Aussi la phrase de Marcel Duchamp servira de motivation à cet ensemble à « embrasser » et mémoriser, si possible, de façon la plus directe, après un récapitulatif concernant l'autoréférence en peinture.

## **2-Peinture et autoréférence**

Car ce qu'avance Duchamp remet en cause toute peinture et toute « peinture de peinture ». Le regardeur n'est plus seulement dans le tableau. Il est dans l'action et il se situe du côté d'un décalage temporel. Il est ailleurs. Il n'hante plus la peinture. Il la transforme. Il se transforme.

Un inventaire des figures qui se rapportent à ce thème récurrent de la mise en abîme dans l'art, a été fourni par Anne-Marie Lecoq et Pierre Georgel en 1983, dans « **Peinture dans la peinture** » (5). Cet ouvrage important prolonge l'article tout aussi important d'André Chastel, « **Le tableau dans le tableau** » (6). Ces publications sont utiles à rappeler brièvement. On peut se référer à d'autres travaux exemplaires.

Ceux de Foucault (7) à Lacan (8) et Jean-Louis Schefer (9) permettent de clarifier les termes qui font du tableau un espace stratégique. Le regard est texte, médiation, mouvement de lecture. Il est théâtralisé et théorisé. Des grilles potentielles de lecture sont alors prélevées chez Velasquez, Holbein et Paris Bordone. Ces typologies de « lecture » mènent vers ces « leçons » magistrales. Le visible devient texte moderne, autrement dit, tend aussi à rendre compte d'une particularité de notre modernité. Elle commence avec Baudelaire ; elle se reconnaît aussi grâce à Duchamp, ses ready-mades et sa « **Boîte en valise** » (1941). Cette « lecture » donc moderne des œuvres anciennes est, de surcroît, enracinée dans certaines œuvres des années 60. L'imagerie de mass media dans la peinture y joue un rôle qui peut être celui de « dénaturaliser » la peinture. Celle-ci, à son tour, cherche à être standard, triste, muette, vulgaire mais c'est compter sans une « pratique plus ou moins consciente » ou approfondissement sensible de cette conjonction entre « un art pictural » et un art de la communication pour le plus grand

nombre. D'autres forces, schèmes, ressources, modes de réflexion sont à analyser qu'on ne saurait soupçonner, imaginer, observer...*sans ce recours manifeste à l'autoréférence.*

### **3-Le miroir-tableau ou un « manque à regarder ?»**

**Magnifying Glass** de Roy Lichtenstein (1963), est une « icône » célèbre du Pop art. On peut facilement se rappeler cette oeuvre qui montre une loupe posée qui agrandit les point de la trame du fond du tableau. Les petits cercles colorés dans la loupe contrastent par leur taille avec de menus points composant la trame de cette même toile. Le manche de cette **Magnifying Glass** est découpé par le bord droit du tableau et l'objet, esquissé en grandes lignes rigoureuses qui suit la diagonale ramène le regard vers l'angle gauche du côté supérieur. Il ne s'agit pas d'un miroir grossissant mais bien d'une **loupe**. Ce qu'elle donne à voir, c'est ce **fond**. Par sa texture qui n'est autre que celle d'une surface composée de points d'imprimerie (the Benday dots), c'est le tableau qui se représente.. De tels points déjà, ont été déterminants dans cette peinture Pop, dès le tout début des années 60, avec les agrandissement des vignettes de bande dessinée. La loupe montre la toile qui montre ces points qui « nous » font rebondir vers les textures artificielles des autres tableaux. L'absence de figuration *dans* la loupe est contrebalancée par ces gros points qui la remplissent. Ces points font référence aux théories de Kandinsky, de Klee. Ils restituent aussi toute une histoire de l'art en accéléré. Et en même temps ils s'en distinguent radicalement : le point n'est pas cet élément actif qui devient ligne qui devient surface. Il est mécanisme de reproduction industrielle de l'image intercepté par une vulgaire loupe. Les points d'imprimerie engagent le regard du spectateur, son « point de vue » dans le tableau est capturé par l'automatisation. La structuration du regard héritée des courants picturaux abstraits est comme désintégrée et remplacée par un moyen de transcription mécanique banal. Le tableau cependant convoque et active simultanément une logique celle des peintres abstraits et celle des modes de production des imageries imprimées.

Lorsque Lichtenstein, compose quelques années plus tard, des miroirs-

tableaux avec ces mêmes points, il continue à s'adresser au spectateur de cette curieuse façon. Obtenu de façon mécanique, le *regard* s'identifie à la série de trame de points. Les points sont tous identiques. Nous devenons aussi égaux, tous semblables. Nous nous imaginons, en points réguliers, répartis en différents cadres et trames. Nous nous sommes dépersonnalisés. Nous ne nous voyons plus comme figure humaine mais comme quantité de points posés par une machine afin que nous parvenions à nous restructurer une image. Nous sommes pris dans un récit qui nous transpose en points puis en miroirs... Et ceux-ci offrent un « toucher » des yeux, parce que lorsque nous sommes devenus fonction de notre regard ; il nous reste encore la possibilité de sentir ces points comme une peau, non plus une peau artisanale mais une peau virtuellement et/ou industriellement représentée.

Le « sujet » paradoxal, c'est à la fois ce regard tactile et une représentation de loupe ou l'emblème du peintre. Le spectateur est invité à lire le sujet du tableau et en même temps à dissocier ses composantes, ainsi il en va de la figure comme du fond, puis d'une image que l'on perçoit par appropriation automatique ou balayage, un peu comme s'il s'agissait d'un **regard d'aveugle**. L'opération est assez complexe à décrire avec des mots alors qu'elle apparaît rudimentaire techniquement. Cette peinture figurative devient langage qui prend ses sources du côté de l'image imprimée. Il n'y a qu'une loupe comme figure et paradoxalement, ce que l'on découvre, c'est ce travail sur un regard, ainsi offert comme « objet » à observer.

Le fait de se référer aux moyens du peintre, en l'occurrence ces points d'imprimerie indique-t-il une volonté de se situer en s'opposant au traditionnel reflet illusionniste de l'art classique ? Dans un court article sur le Pop art, Roland Barthes, (« **Cette vielle chose l'art...** » 1980 in **L'obvie et l'obtus. Essais critiques III**, Seuil, 1982). (10) le remarque ; cette loupe qui agrandit, c'est celle des spéculations formelles des changements d'échelle du travail pop. Et il est vrai, comme il le souligne à juste propos, qu'une telle œuvre représente **tout** l'art pop.

Une décennie plus tard, Lichtenstein s'est dépeint sous les traits d'une figure composite, celle d'un buste dont la tête se termine par un miroir, c'est-à-dire par une toile ressemblant à un miroir. Ce tableau intitulé « **Self-portrait** » datant de 1978 réaffirme cette approche précise, toujours

**point par point**, quasi-objective d'un art de peindre qui neutralise la singularité du geste sensible traditionnellement attribué à l'artiste. De l'imprimerie en peinture, nous avons tout un vocabulaire et une grammaire qui se reconstruisent. Les points, hachures, grilles, plages colorées lisses déterminent les nouvelles conditions du lisible. Le tableau « **Self-portrait** » est un récit de mutations et métamorphoses toute en une subtilité régie par des grilles de points qui obéissent aux principes mathématiques de la théorie des ensembles. Lorsque nous subissons des « métamorphoses » de ce genre, on s'attend à voir quelque figure de conte de fée. L'ère de la machine, par les doigts « magiques » du peintre, produit des monstres équivalents, en apparence moins pittoresques et moins terrifiants.

*Transposer, raccourcir, agrandir, faire voir, imposer un regard tactile, interroger sa mémoire, reconnaître ses habitudes et automatismes visuels, associer, fragmenter, établir un saut qualificatif d'ordre sémantique, forcer les catégories subjectives de l'art par des techniques neutres, est-ce refouler tout l'imaginaire du côté d'une codification du visible en un lisible ? Le spectateur est-il perdu ? Son regard est-il reconfiguré ?*

#### **4- Do it yourself, Landscape, 1962**

Supposerait-on qu'Andy Warhol cherche à s'imposer, par répétition d'une même image de boîte de soupe Campbell ? Ou qu'il emplit la surface de son tableau par son multiple qui correspondrait aux soupes consommées et par ricochets aux spectateurs nombreux et affamés ? La « répétition » dépasse le cadre du potentiel interprétatif et finalement en revient au ready-made de ready-made...et à une valorisation du principe de la **répétition**.

Lorsque le peintre « expose » une toile avec son modèle de paysage déjà prêt à colorier, il invite à relire ce tableau comme « création » industrielle. C'est l'industriel qui dicte **les lois du visible**, et Warhol s'en fait l'observateur, le témoin, l'élève servile et besogneux. Mais par delà ces « actions », il apparaît comme celui qui tire les nombreuses ficelles.

Le tableau, acheté pour être rempli comme un formulaire, représente un paysage avec maison rouge, barrière de bois au premier plan à droite, route de village et campagne à l'arrière plan, le tout dominé au centre par un arbre. Ce paysage est à presque à moitié terminé, rompant les couleurs

rouge et bleu, marron et verte par d'importantes zones de blanc... Dans celles-ci, clairement délimitées par des contours en une quantité de régions sont toujours disposés selon des numéros qui permettent le report des couleurs . L'intérêt du tableau repose alors sur ces blancs laissés comme en réserve, -ils évoqueraient les aquarelles tardives de Cézanne. Et entre le charme des blancs fortifiant les couleurs et la banalité du tableau, cela ne peut que « jurer ». Le tableau de Warhol, (et ceux composant cette série), est donc un divertissement dérisoire. La peinture est déjà « ready-made ». Warhol en fournit ici la preuve. Alors que les sérigraphies de boîtes de soupe, de bouteilles de coca-cola parodient le « **all over** » de Jackson Pollock, c'est toute une vision « extralucide » de la consommation de la peinture qui s'ose.

Ces œuvres portent sur un regard, celui qui est objet quantifiable en plage de coloriage, en numérotation, en « passe-temps », en « mode paysage» en « talent »-ou « mètre – étalon », des surréalistes. On peut prolonger encore cette liste en affirmant que l'on achète sa propre vision du monde. Et Andy Warhol, sans en avoir l'air, nous renvoie cela, en pleine figure, par ce tableau. Le titre **Do it yourself, Landscape** récapitule l'achat, le travail, le regard, l'objet . Et on peut se demander si derrière cette invention, ne perce pas un sentiment de solitude, Warhol se cherche et cherche son regardeur, celui qui disparaît derrière presque deux siècles d'industrialisation. Et il lui faut « indiciblement » le retrouver.

La quantité, l' étendue de toile, les processus de fabrication, les mécanismes de reproduction sont les objets d'une démonstration du comment voir, peindre, observer pour toutes et tous. Le tableau réussit ce « libre-échange » entre un je et un **on** qui **s'amplifie**. Tout est déductible comme tout produit se monnaye. Le tableau montre comment ce processus devient souverain. Le regardeur est-il alors celui qui constate une absence *totale* de liberté ?

Le regard est préformé, prédigéré, programmé. Avoir su le re-produire en se prêtant au jeu, c'est commencer à sortir d'une vision machinale, et parvenir à **le faire voir**, c'est **ce** qui commence à retenir **l'attention et à inquiéter**. Warhol opère à froid dans cette société de consommation des années 60 qu'il a appris à mieux connaître. Et encore parcourir ces figures,

c'est se donner de nouveaux repères. En analyser quelques unes lorsqu'elles font retour sur elles-mêmes, ce serait alors gagner du temps.

*Calculer, cartographier, schématiser un à un les processus du travail pictural, par le biais d'une numérotation est-ce sortir du cadre des imitations en art ? Que devient le regardeur face à la codification et fabrication industrielle du parcours de son regard ? Quelque rêve ne s'inscrit-il pas lorsque le peintre y laisse une marge en plein cœur de sa toile ? Où encore une fois, ne sommes-nous pas dans un rêve qui nous interdit de dépasser de telles limites ?*

### **5-Le tableau comme pléonasme ?**

La situation face aux imageries de mass media en Europe, et notamment en France a été vécue différemment. Et pourtant, les artistes puisent aux mêmes sources. Même si on observe un décalage entre les dates, si la Figuration narrative s'impose avec quelques années de retard par rapport à l'Angleterre et aux Etats-Unis, les artistes abandonnent de façon plus progressive l'abstraction qu'ils pratiquaient en s'orientant vers la nécessité de s'engager politiquement, en se tournant du côté de l'URSS et surtout vers la Chine. La volonté de rivaliser avec le Pop Art est reprise par le critique G. Gassiot-Talabot (11) dans ses nombreux textes présentant les travaux de cet art centré sur l'imagerie.

Le tableau de Bernard Rancillac, **Sainte-Mère la vache** (1966) est-il l'équivalent de **Magnifying Glass** ? Le texte qui accompagne l'exposition « L'année 1966 » à laquelle appartient ce tableau tend à le démontrer car il s'intitule **L'image de l'image**, et il est signé par Pierre Bourdieu (12). Il démontre que les images peintes de Rancillac fonctionnent comme « pléonasme » ou type particulier de tautologie du genre : « prévoir à l'avance, monter en haut, au jour d'aujourd'hui ».

Des événements politiques d'une année sont transposés en images par Bernard Rancillac. Il en accentue le caractère d'affiche. Les couleurs vives et sourdes contrastent. L'affiche tableau est comme légendée ou cartographiée.. Dans **Sainte- Mère la Vache**, surgissent des contradictions déroutantes elle transparaissent de la façon la plus rigoureuse. C' est l'un

des exemples les plus frappants de cette série, avec cette ligne d'horizon qui divise la toile en un ciel exhibant en son centre l'image de la boîte de « *Vache-qui-rit* », célèbre pour sa mise en abîme. Sous ce ciel ainsi « illuminé » par ce fromage, nous avons le droit de voir une image de scène de famine dans un paysage désertique. Les composantes formelles de des deux types d'« imageries », publicitaire et politique sont essentielles. Sur le bleu-ciel clair du fond se détache très nettement le dessus du couvercle de la boîte ronde bleu foncé avec sa Vache qui rit. Celle-ci, avec son rire, surplombe un paysage jaune criard parsemé de quelques buissons verts qui diminuent selon une perspective rudimentaire vers la ligne d'horizon. Au premier plan de ce paysage, il est difficile de distinguer clairement les silhouettes de personnages, un enfant, un âne chargé de jarres, une femme en sari.

Reproduit fidèlement, le couvercle d'une de la boîte de *Vache-qui-rit* « se détache » d'un ciel bleu et donc prédomine. Les personnages, aux teintes assourdis, quant à eux, apparaissent mangés et broyés par la lumière d'un fond jaune criard.

Les images portent sur la misère humaine. La lecture synthétique du tableau devient ambivalente. Entre l'objet allégorique du tableau et sa lecture, il se produit un décalage et le peintre attend son regardeur au tournant. La Vache-qui-rit apparaît comme une transfiguration du monde occidental qui est révélé aux individus du bas, à moitié effacés et dans un état de dénuement. L'auto-satisfaction et auto-proclamation (la Vache-qui-rit dans la vache-qui-rit), ont pour corollaires la misère, et... ce *reste du monde*. La leçon (*the lecture*), ne s'arrête pas à ces quelques interprétations. Arrêtons-nous pour souligner l'aspect symbolique de l'image du ciel reprise par la mise en abîme de la boîte et celle signifiante des personnages informes, sans contour, s'estompant en plages ou taches de couleur. D'un côté, l'image qui est production d'elle-même, est *céleste*, de l'autre, elle se déduit des constituants du tableau, elle est « *viscérale* »... Et il nous faut *absorber* visuellement ces deux *états* de l'image en une seule, faire glisser les sens et les formes d'un cran, en remarquant que cette lecture correspond étrangement à celle des **Mythologies** de Roland Barthes.

La lecture est littérale, la lecture est symbolique. Mais elle est aussi plus : elle est enquête sur ces moyens nouveaux dont se dote le peintre, et son lecteur (ou spectateur).

Les images se confrontent. Le regardeur est pris à parti. Dans **Voir le voir**, John Berger (13) déjà remarquait ce genre de situation à propos d'une photographie montrant une famille en détresse, au Bangladesh reproduite juste au-dessus d'une réclame pour des sels de bains, avec comme énoncé : « Les choses arrivent après un bain ».

Cette « **juxtaposition fâcheuse** », le peintre l'utilise, révélant, de fait, comment il prend ses distances vis-à-vis de sa vision d'un Pop Art « narcissique », en quête de lui-même.

La portée allégorique du tableau, se lit par le biais du titre, « **Sainte-Mère la vache** », ouvrant vers toute une chaîne de significations religieuses. La vache, icône publicitaire, se désigne elle-même comme animal sacré en Inde. Elle nous conduit vers ces personnages des actualités catastrophiques. Ils miment, quant à eux, un épisode de la Sainte-Famille fuyant en Egypte. La toile de Rancillac nous immerge dans un système de références culturelles universelles. Par delà cet exercice, celui d'une démonstration d'une violence du voir, nous plongeons allégrement dans « un enfer de l'image », dans « **un monde, en proie à la puissance destructrice illimitée** », comme l'écrit Paul Virilio,(14) à propos de l'œuvre d'un autre artiste de ce même courant, Peter Klasen.

On pourrait multiplier les exemples en analysant d'autres œuvres, celles de Klasen, Adami, Monory, Télémaque, Cueco, etc.. Ils font aussi état de cette violence du monde social de l'image, comme dissolution du subjectif qui laisse le regardeur, plus ou moins perplexe et médusé. Les œuvres du Pop Art comme celles de la Figuration narrative prennent leur essor dans ce monde d'images qui se voient *pour* nous et non plus *en* nous.

*Récupérer des images célèbres, typiques, symboliques, politique, satisfaire au désir de lecture automatique, établir un saut qualificatif d'ordre sémantique, est-ce suggérer au regardeur qu'il ne vagabondera plus « gratuitement » dans l'œuvre, celle-ci étant alors l'équivalent condensé d'un texte ?*

**6-Quand la production de l'œuvre se démultiplie en moyens de production**  
... L'éloquence du banal, la stigmatisation de la peinture comme simple objet de commerce, le choc des images, les strates de sens et de formes

entremêlées forment toujours un contexte artistique qui au fil des ans s'est fortement enrichi. Et s'est limité au format du tableau qui déjà ne se suffisait plus à lui-même. Il fallait l'investir autrement pour sortir d'un certain champ d'indifférence, celui de la diffusion, des média, des images.

Lorsque l'on découvre une photographie, « **Chantier, Barbès-Rochechouart** », série des Posters, 1994, de Pierre Huyghe, on est « sidéré ». Car il ne s'agit plus d'une autoréférence comme précédemment, en peinture, mais d'une expérimentation savante et rationnelle réactivée de moyens quantifiés en regards, spectateurs, images d'ouvriers, ouvriers eux-mêmes manoeuvres. : une chaîne d'actes et de résultats qui nous entraînent du côté de Pierre Huyghe, nouveau type de maître d'œuvre.

Aussi commençons par une première lecture naïve de cette photographie qui représente un carrefour en travaux d'un boulevard parisien, vu en perspective. La singularité vient du panneau publicitaire qui se situe au premier plan à droite. Il oblige à observer deux zones, le « paysage urbain » proprement dit et ce panneau, qui joue le rôle classique du tableau dans le tableau. Dans le paysage urbain, nous distinguons les ouvriers du chantier. Ils sont au travail, avec leurs outils et autres brouettes. Mais il advient que le panneau publicitaire qui occupe une grande place dans cette oeuvre montre, elle- aussi dans une posture voisine, (il s'est passé du temps) ces mêmes ouvriers. Il pourrait s'agir d'une scène banale. Le fait qu'elle soit reproduite sous la forme d'une « affiche publicitaire » dans la photographie rend l'ensemble énigmatique.

La description peut paraître fastidieuse. Ce qui nous « dérange » ce n'est pas le procédé, issu du travail de Magritte, comme notamment, avec **La Représentation**. Non, c'est l'irruption d'une scène de la vie de tous les jours dans le cadre de la publicité à l'intérieur d'une « histoire ». Il s'agit d'un micro-récit, obtenu par décalage temporel. Le monde quotidien, la ville, devient le lieu de dédoublement du temps et de l'espace. Le fait de savoir comment Pierre Huyghe fabrique ce genre de situation n'ôte rien à la confusion qui s'empare de nous. Et c'est cette confusion que l'on apprécie.

Ces « jeux » entre image, photo, réalité quotidienne n'ont rien de gratuit.

Ces temps décalés, ces scènes légèrement différentes se laissent comparer. Les différences impliquent, de façon sensible, une perte des repères spatiaux, de relation de cause à effet, de cohérence narrative. La photographie prend une signification « insensée ». On s'interroge sur les places ou postures du... photographe. Les ouvriers, les acteurs qui les miment ceux-ci, tous sont des pièces dans la main d'un regardeur qui les joue comme sur un échiquier. Huyghe décrit, analyse ces rôles et ces regards pour qu'ils s'affolent. Non plus des regards d'aveugle mais des regards qui sont pris dans un enchaînement à la limite de l'absurde, une logique visuelle déconcertante pour des observateurs rassasiés d'images. L'œuvre doit faire diversion. On devine que pour un regardeur qui étouffe sous son monde d'images en croissance toujours exponentielle, il est nécessaire de tirer parti d'aberrations, de paradoxes, de hasards ou d'accidents. Tous est à réordonner, à recomposer, à transformer.

La description de la méthode de Pierre Huyghe est embarrassante. Elle implique un changement d'état. Des acteurs *vont* jouer les ouvriers. On *va les prendre* en photo. On *commence* alors à se placer sur les rangs du cinéma. Il y a tout un *travail* d'équipe. L'image qui devient film passe par ce *flou temporel* qui s'insinue entre divers temps de prise de vue et la photographie.

Les images vont et viennent dans des temps décalés. Le monde du quotidien n'est plus de tout repos et de toute éternité. Les temps saisis sous l'œil de la l'appareil photographique de l'artiste induisent des espaces labyrinthiques qui sont comparables aux enfers et aux prisons de Piranèse. Mais ce monde reste toujours nôtre avidement sureposé. Nos gestes, nos journées, nos désirs et nos projets collés bout à bout puis synthétisés rendent cette même saveur toujours plus anodine qui est celle stigmatisée dans « **Chantier, Barbès-Rochechouart** »,

*Passer du tableau à la photographie tout en devenant un réalisateur de film, faire du tournage l'équivalent d'une photo, résume toutes les opérations inventoriées dans les œuvres précédentes. Huyghe fait passer la peinture ailleurs, dans d'autres domaines, dans la cohérence organisationnelle de son travail. Serait-ce une ultime tentative de questionnement autoréférentiel en peinture ? Ou au contraire un nouveau départ ? Une œuvre encore en chantier ?*

## 7- Et de fil en aiguille, le regardeur...

Le monde des représentations nous appartient-il encore ? Notre expérience visuelle, esthétique ne se forge-t-elle plus qu'à travers des techniques et parfois des œuvres qui les mettent radicalement en question ?

L'autoréférence tient à ces mutations techniques. Les œuvres des années 60 changent de signification et d'usage. Elles rendent compte de mutations historiques de société. L'étrange pouvoir pédagogique, que l'on ignore encore, se lit dans ce genre d'œuvre. Le tableau ou force de travail artistique n'est pas simple plaisir esthétique, mais désir d'apprendre comment « on » apprend...

Les peintres des années 60 témoignent de cette étrangeté soudaine, celle de se sentir dépayrés et par l'image, et par contrecoup par la peinture, même. A ce propos, le message délivré par Duchamp ne reste-t-il pas d'actualité ? « **Ce sont les REGARDEURS qui font les tableaux . On découvre aujourd'hui Le Greco ; le public peint ses tableaux trois cents ans après l'auteur en titre** ».

Myrielle Hammer

### **notes**

-----  
in Sarane Alexandrian, Duchamp, Flammarion, 1976, p.84  
« Autoréférence », faut-il jouer encore sur le trait d'union ou assembler les deux parties de ce mot ? Cette hésitation orthographique indique un état de changement social de l'individu, du « moi, je » et toutes ses composantes... On observe comment en art, l'autoréférence devient le passage obligé menant vers toutes sortes de possibilités de renouvellement du sens des relations, comme celles que déclinées par Rosalind Krauss dans L'inconscient optique, éd. Au même titre, 1993

(3)

Aucune de ces œuvres à ma connaissance sont visibles sur le web mais il est possible de les retrouver dans les ouvrages suivants :

**-Magnifying Glass** de Roy Lichtenstein (1963), in Calvin Tomkins, L'art de Roy Lichtenstein, Altinea, 1994

**-Do it yourself, Landscape**, de Andy Warhol in Marco Livingstone, Le Pop art, Hazan, 1990

**-Sainte-Mère la vache** de Bernard Rancillac (1966) in Jean Louis Pradel, La Figuration narrative, Hazan, 2000

**-Chantier, Barbès-Rochechouart** », série des Posters, 1994, de Pierre Huyghe in Christophe Domino, A ciel ouvert, Scala, 1999

(4) Jean-Luc Chalumeau, La nouvelle figuration. Une histoire, de 1953 à nos jours, Cercle d'art, 2003

Anne-Marie Lecoq et Pierre Georgel, La Peinture dans la peinture, Adam Biro, 1983

(6) André Chastel, Le tableau dans le tableau in Fables, formes, figures, Flammarion, 1978

(7) Michel Foucault, Les mots et les choses, Gallimard, 1966

(8) Jacques Lacan, "Les quatre concepts fondamentaux de la psychanalyse" in Séminaire XI, Seuil 1973.

(9) Jean-Louis Schefer, Scénographie d'un tableau, Seuil, 1969

(10) Roland Barthes, « Cette vielle chose l'art... » in L'obvie et l'obtus. Essais critiques III, Seuil, 1982

(11) Gérald Gassiot-Talabot, La figuration narrative dans l'art contemporain, Galerie Creuze, 1965, n.p.

(12) Pierre Bourdieu, « L'image de l'image », in L'Année 66, Galerie Blumenthal-Mommaton, 1967

(13) John Berger, Voir le voir, BBC, Alain Moreau, 1972, p.163,

(14) Paul Virilio, Peter Klasen. Etudes d'impact, Expressions contemporaines, 1999

# RÉFLEXIVITÉ ET HYPERMODERNITÉ

Francis Jauréguiberry

Laboratoire SET (CNRS UMR 5603)  
Avenue du Doyen Poplawski 64 000 Pau  
francis.jaureguiberry@univ-pau.fr

## 1- Modernité et distance à soi

Au fondement même de la modernité réside le mouvement de dissociation identitaire de l'individu envers lui-même. Dédoulement et réflexivité sont au cœur de la modernité (Elias, 1975 ; Dumont, 1983 ; Touraine, 1992 ; Taylor, 1998). L'individu moderne se « regarde de l'extérieur » pour se penser dans son autonomie de la même façon que la société moderne se « met à distance » pour se concevoir comme œuvre à accomplir. Cette distance à soi est précisément la condition d'existence du sujet moderne. Celui-ci n'existe que dans un mouvement de décalage par rapport à ses déterminants impersonnels, que ceux-ci soient liés au corps (physique et pulsions) ou à la société (normes et rôles). C'est en « résistant » à ces déterminismes que l'individu devient sujet, l'enjeu étant de pouvoir vivre son autonomie et construire sa propre vie.

Les pères fondateurs de la sociologie ont tous mis l'accent sur le phénomène de la socialisation moderne comme condition d'existence de la société des individus. C'est parce que la société « formate » les individus à être autonomes qu'elle peut à son tour apparaître comme produit de leur mobilisation. Ce phénomène de socialisation comme processus de subjectivation est à la base même des distinctions désormais historiques entre communauté et société, sociétés traditionnelles et sociétés modernes. Contrairement à l'individu traditionnel qui soumet sa vie à un « destin » sous la forme d'un statut dictant ses comportements, de codes d'appartenances ritualisés, de morale communautaire et de croyances collectives, l'individu moderne se sent maître de sa vie et de ses choix dans un environnement fait de mouvement.

La façon dont a été interprété cette liberté individuelle a

toujours oscillé entre deux extrêmes. D'un côté, une vision « optimiste » qui pense que plus la socialisation moderne est forte et plus le processus de subjectivation l'est aussi. Autrement dit : c'est bien parce que l'individu est formaté à être autonome et critique qu'il devient capable de prendre de la distance vis-à-vis des structures qui, précisément, l'ont formaté. D'un autre côté, la vision « pessimiste » qui pose que la socialisation moderne ne fournit qu'une illusion de liberté. L'individu n'est que le support subjectif des attentes sociales qu'il a incorporées. Ce n'est pas lui qui désire et agit mais la société qui le « joue ». Mais, au-delà de leur divergence, ces deux positions reconnaissent la continuité entre l'individu et la société, entre l'acteur et le système. Le bon fonctionnement de la société dépend de la socialisation qui fait des individus des acteurs sociaux qui, à leur tour, produisent la société.

## **2- Hypermodernité et réflexivité**

C'est cette continuité qui, au tournant des années 70, a fortement été remise en question par la critique « post-moderne ». Le problème n'est plus celui de la plus ou moins grande acceptation par l'individu de sa place dans la société, sous des formes allant de la résignation à la révolte et du conformisme à la déviance, mais celui de l'unité de cette place. C'est, qu'entre temps, nos sociétés sont devenues — pour reprendre l'expression de Bauman — « liquides ». Les cadres sociaux dans lesquels nos sociétés modernes — et singulièrement la société industrielle — fixaient les individus sont désormais flous et éphémères, rendant impossibles leur acceptation ou rejet comme projet de vie. L'individu doit désormais se créer lui-même une trajectoire dans un environnement changeant.

Alors que, jusqu'à peu, les normes sociales disciplinaient l'individu et le portaient à accepter sa condition (produisant conformité ou, au contraire, révolte), elles l'assignent aujourd'hui à devenir « lui-même » sans autre référence que sa propre volonté. Alors que, jusqu'au seuil des années 70, un futur chargé d'espoir pouvait collectivement être entrevu à travers les idéologies d'émancipation, l'actuelle désaffection envers les utopies laisse place à un présent qu'il s'agit de conquérir personnellement à partir de ses propres ressources. On n'espère

plus collectivement dans le futur : il faut réussir personnellement dans le présent. L'individu est mis en face de ses propres réalisations. Démiurge de lui-même, il n'est plus l'objet d'un destin, mais seul maître de sa vie.

Ce n'est plus l'intériorité de l'individu qui se heurte à l'extériorité de la réalité moderne, mais l'intériorisation d'impératifs hypermodernes (entre autres ceux de « se réaliser » et de « réussir sa vie ») qui décolle l'individu de son extériorité (son image, son statut, sa définition sociale). N'étant plus seulement cantonné par le social à un rôle (disciplinaire mais qui était rassurant) et revenu des utopies d'émancipation collective (contraignantes mais qui offraient un espoir), l'individu hypermoderne est désormais mobilisé à se produire lui-même avec l'impression d'être toujours en décalage.

L'interprétation « post-moderne » consiste à célébrer ce mouvement comme une libération par la société des idéologies qui contraignaient les individus et produisaient la plupart du temps des catastrophes politiques, et comme une liberté pour les individus que ne sont plus redevables qu'envers eux-mêmes de leur conduite. En quelque sorte, nous aurions à faire à une société mouvante où chacun pourrait « zapper » d'une occupation à une autre dans le seul souci d'en vivre la meilleure intensité possible (Maffesoli).

Mais la situation est malheureusement bien moins simple... De nombreux auteurs, dont les plus connus sont Antony Giddens et Ulrich Beck ont au contraire montré d'une part qu'il n'y avait pas rupture avec la modernité mais dépassement de celle-ci en une « seconde modernité » ou « modernité réflexive » dans laquelle les trajectoires des individus et de la société elle-même apparaissent comme des choix et des prises de décision en situation de réflexivité, et, d'autre part, que l'extension de cette réflexivité produisait un certain nombre de problèmes dont l'un des moindres est celui de la monté des incertitudes, des doutes et, comme produit de cela, de la notion de risque.

En ce qui concerne le premier point, on peut rester sceptique sur la distinction sans doute un peu forcée effectuée par ces auteurs entre « première modernité » et « seconde modernité » ou « modernité avancée ». Certes, Giddens montre que les « croyances

modernes » (progrès, idéologies) ont fonctionné selon les mêmes principes que la tradition : la croyance. Beaucoup de modernes ont cru au progrès ou à l'avènement du socialisme comme les traditionnels croyaient au droit divin. De même Beck montre bien comment la famille, par exemple, a continué de fonctionner « à la transcendance » dans la mesure où les modèles qui la définissaient n'étaient pas questionnés mais continuaient à être vécus comme allant de soi. Mais, à trop vouloir forcer le trait, ils en oublient que la « première modernité » contenait déjà en elle-même tout ce qu'ils désignent comme étant le propre de la « seconde modernité ». La « première modernité » était aussi et déjà, on vient de le voir, faite de réflexivité. C'est même parce qu'elle a été, à un moment de l'histoire, capable de se regarder comme produit d'elle-même que la société est devenue moderne contre la tradition qui la pensait comme produit d'un au-delà méta-social. On assiste donc moins à un changement de nature de la modernité qu'à un approfondissement de celle-ci, et en particulier par l'extension de la démarche réflexive à des secteurs de la vie sociale jusqu'ici restés assez imperméables à l'auto-référence.

Pour ce qui est du second point, les « effets » de la « seconde modernité », amplement développés par Beck sous le concept de « risque », la critique pourrait suivre le même raisonnement : tout cela était déjà présent dès l'avènement de la modernité. Mais je voudrais mettre ici l'accent sur un autre point qui touche non plus la société prise dans son ensemble — si tant est que l'on puisse encore le faire, la notion même de « société » devenant problématique — mais les individus qui la composent. Et la notion de radicalisation de la modernité pensée comme approfondissement de la réflexivité des individus envers eux-mêmes me paraît bien adéquate au moment d'analyser nombre de leurs expériences. En cela, je donne raison, malgré les réserves qui viennent d'être faites, à Giddens et Bech contre les postmodernes. Il semble en effet que, lorsque la modernité se radicalise au point de devenir hyper, la conscience de soi engendre tout autant un sentiment d'insatisfaction producteur de vide ou d'angoisse qu'un sentiment de liberté producteur de jouissance. Dans notre société, où la norme est la réalisation de soi et l'impératif est celui de vivre avec un maximum d'intensité et de sensations, il arrive en effet que

l'individu ne se sente pas à la hauteur de l'injonction qui lui est faite de « réussir sa vie ». Rares sont en effet ceux qui possèdent les ressources à la hauteur de leurs espérances et les capacités nécessaires à la réalisation des performances escomptées.

### **3- Illustration : la manipulation de soi sur Internet**

La façon dont certains s'inventent des identités sur Internet illustre parfaitement cette faille. Que ce soit en empruntant un sexe, un âge ou un statut autre que le sien pour se dire dans une IRC (groupes de discussion en ligne), en habitant un personnage dans un jeu de rôle en ligne ou en s'identifiant à un avatar dans l'univers virtuel d'un MUD, il est en effet désormais possible d'être quelqu'un d'autre que soi tout en restant soi (de la même façon qu'il est possible d'être, avec son portable, ici et ailleurs à la fois). C'est cette identité empruntée, mais parfaitement habitée par l'internaute *et reconnue comme telle par ses correspondants*, que j'appelle le *sur-soi* des internautes. Il ne s'agit pas simplement d'une identité fantasmée, produit d'un rêve éveillé, mais bien d'une forme de réalité partagée. La tangibilité du *sur-soi* réside en effet dans le fait qu'il est vécu comme une réalité par autrui, quand bien même cet autrui-là serait-il lui-même la projection d'un autre fantasme. Ce qui intéresse l'internaute qui manipule son identité n'est pas tant d'habiter un personnage que d'expérimenter « l'effet que ça fait » d'être reconnu comme tel en communiquant.

Contrairement à ce qui se passe dans la vie *off line*, le milieu d'appartenance de l'internaute n'est pas là pour le dissuader, le limiter ou le censurer dans l'essai de nouveaux *soi*. L'internaute peut expérimenter autant de *soi* qu'il veut sans grand risque de sanction de la part de son entourage. Mieux : toutes les études que j'ai pu lire sur le sujet, comme mes propres observations, montrent que les *sur-soi* de l'internaute sont, la plupart du temps, et contrairement aux *soi* constitutifs de son personnage social, des rôles physiquement ou socialement intenables par lui dans la vie réelle. On peut donc faire l'hypothèse que c'est parce que ces rôles ne peuvent pas être réellement joués dans la société qu'ils le sont virtuellement sur Internet. Car, si l'on cherche à savoir quels sont, concrètement, les *sur-soi*, on s'aperçoit qu'il s'agit la plupart du temps de *soi* focalisant la réalisation de fantasmes, de pulsions et

des désirs contrariés. Il faut donc interpréter l'emprunt de ces *soi* virtuels à la fois comme une *re-création* (l'internaute devient démiurge d'une part de lui-même qui n'existe que dans son imaginaire) et comme une *récréation* (au sens scolaire du terme : comme l'élcolier pendant une pause, l'internaute échappe quelques instants aux règles de la classe — ici la société — et *joue à être ce qu'il veut*).

Deux issues extrêmes sont dès lors envisageables. D'un côté, l'enfermement dans le virtuel à cause du réel social. D'un autre côté et à l'inverse, le questionnement du réel à cause du virtuel. Dans le premier cas, c'est parce que le réel social désespère ou restreint qu'il est fui ou recomposé idéalement dans le virtuel. Dans le second, c'est parce que le virtuel autorise des expériences inédites et libère l'imagination que le réel en vient à être vécu différemment ou que de nouvelles exigences apparaissent.

La première issue, sans doute tragique, conduit à la dissolution de l'individu dans sa réalité virtuelle. Goûtant sans retenue les délices de la reconnaissance de ses fantasmes et attentes par le réseau, l'individu court le risque de s'enfermer dans une pratique compulsive d'Internet. Le *moi* de l'individu impose alors sous le poids d'un *sur-soi* qu'il ne peut plus maîtriser. L'individu ne part plus des *soi* qui le constituent socialement pour arriver au *moi* individuel, mais d'une « déception » vis-à-vis de son propre *moi* pour s'évader dans un *sur-soi* gratifiant.

Alain Ehrenberg a montré comment cette « tension entre l'aspiration à n'être que soi-même et la difficulté de l'être » pouvait conduire à la dépression. Selon cette perspective, la dépression est une « pathologie de l'insuffisance », une « pathologie de l'estime de soi » de l'individu qui ne parvient pas à vivre l'idéal qu'il s'est lui-même fixé. L'expérience de *sur-soi* est précisément, pour certains internautes, une façon d'échapper à cette conscience malheureuse de n'être que soi-même. Elle vise alors à combler le vide que connaît l'individu hypermoderne entre la conception surévaluée qu'il se fait de lui-même (*idéal du moi*) et la perception de sa réelle condition (*moi*). Plutôt que de se désespérer du fait de n'être « que lui-même », l'individu branché va, à peu de frais, se construire un *moi* beaucoup plus conforme à ses désirs par le biais d'un *sur-soi* valorisant ou gratifiant.

Dans son écoute constante envers lui-même, le *moi* du manipulateur de *soi* risque alors de ne plus être que quête d'émois. Ivre de lui-même, ce *moi*-là n'a plus besoin d'être mis à distance par l'individu. Ce qui est recherché par l'individu, c'est un accord parfait avec lui-même, c'est l'oubli de la distance critique qu'introduit le *je* et donc des tensions, limites, et souvent frustrations que cette distance révèle. Le *je* cesse alors d'être questionnement du *moi*.

Autant cette première issue à laquelle peut conduire l'adoption de *sur-soi* sur Internet peut paraître inquiétante, autant l'autre issue peut être considérée avec un certain optimisme. Elle renvoie à l'expérimentation critique et créative des limites du *moi* par le sujet. La réalisation virtuelle de ses fantasmes sur Internet charge en effet le *moi* d'une nouvelle dimension et par là même peut provoquer chez l'individu un nouveau questionnement sur ses limites identitaires : « *Jusqu'où suis-je moi ?* » Autrement dit : l'individu « s'essaie » à un ou des *sur-soi*, non pas pour s'y perdre, pour s'y oublier, mais au contraire pour mieux se situer, mieux se penser et mieux s'expérimenter dans sa capacité créatrice. Habiter un autre que *soi* demande en effet une grande capacité créatrice. En questionnant les limites du *moi*, l'emprunt de *sur-soi* interpelle donc le *je* de façon inédite. L'interprétation optimiste consiste à poser que le dialogue entre le *je* et le *moi* s'en trouve renforcé.

Bien sûr, les *sur-soi* sont sans doute, en grande partie, aux ordres de pulsions ou de normes, et donc l'objet de récupérations par leur mise en miroir sur Internet. Ils peuvent aussi répondre d'un désir d'intégration (statut) par le biais d'une présentation de *soi* conforme aux impératifs d'une organisation ou d'une communauté dont on attend des gratifications, en particulier sur certains MUD. Ils ont alors toute chance de conduire à du conformisme identitaire ou, dans le pire des cas, à du communautarisme. Mais il n'est pas interdit de penser qu'ils expriment aussi — et en quelque sorte « en creux » — l'aspiration de l'individu à être autre chose qu'un *moi* satisfait de ses statuts, intérêts ou rôles sociaux.

On peut en effet faire l'hypothèse que la manipulation de *sur-soi* traduit aussi la volonté d'échapper aux images de *soi* imposées par la société et révèle, sans doute maladroitement, le désir

d'exister « autrement. Non pas pour s'y perdre, pour fuir la réalité, mais pour mieux y tester son autonomie et son aspiration à la liberté. L'emprunt de *sur-soi* serait dans ce cas pour l'internaute une façon de se dédouaner des contraintes qui pèsent sur lui en tant qu'objet social et d'expérimenter, de façon détournée, son aspiration à être un sujet. Ce ne serait pas tant la transparence du *moi* qui serait alors recherchée que la capacité à s'étonner d'un décalage à soi-même.

On peut en effet considérer que les *sur-soi* sont tout autant des bases de résistance que des points de fuite, tout autant des actes créatifs que des abandons identitaires. La manipulation de soi par un individu sur Internet peut donc être lue aussi comme l'extériorisation transgressive de son exigence d'être un sujet créatif au-delà des déterminations, statuts et rôles qui le freinent socialement. Même si les changements sociaux dus à une telle pratique sont pour l'heure très faibles, une lecture optimiste de cette perspective conduit à penser que l'expérience de *sur-soi* peut fonctionner comme espace de recadrage de soi et des autres dans une attitude plus ouverte et créatrice. Le fait d'habiter ainsi une ou des altérités pourrait alors aider à mieux écouter, à mieux comprendre les autres dans leur complexité, ceux-ci cessant alors d'être appréhendés comme des objets pour être abordé comme des sujets avec les tensions et les faiblesses, mais aussi avec la force et l'inventivité que cela suppose.

#### **4- Les jeux incertains du je hypermoderne**

En offrant des espaces inédits d'expression et de manipulation de soi, Internet révèle et illustre bien certaines des tensions identitaires de l'individu hypermoderne. Ce n'est évidemment pas Internet qui transforme directement la gestion identitaire de l'individu. Il ne faudrait pas y voir un déterminisme « brutal » de la technique sur le social. Mais, en ouvrant de nouveaux champs du possible et en déstabilisant — même modestement — l'existant, une innovation technique peut fonctionner comme un *révélateur social*. Plus particulièrement, *la manipulation de soi sur Internet nous parle de la souffrance ou de la difficulté de l'individu hypermoderne à être un sujet capable de relever le défit de la gestion de son identité*. Ce défit est à la hauteur de l'enjeu : être soi, en dehors de ce qui,

jusqu'à récemment, avait informé son individualité sous la forme d'un destin, d'une appartenance ou d'une dépendance

Le résultat de cette nouvelle expérience est pour l'heure ambigu. D'un côté, le recours au virtuel peut perdre l'individu en ce qu'il risque de ne plus pouvoir se détacher de l'image gratifiante que ce même virtuel lui renvoie. Narcisse est alors condamné à se noyer dans le reflet de ses propres désirs qu'Internet lui renvoie. Mais, d'un autre côté, il peut aussi s'y trouver, c'est-à-dire devenir capable, par l'expérimentation d'une altérité radicale, d'accepter ses limites non comme censure, mais comme cadre possible de créativité, c'est-à-dire se penser comme sujet, à la fois dans ses déterminations et dans sa volonté. Seulement cette image-là ne peut se conforter que dans le dialogue avec autrui. Autrui recherché non plus comme faire valoir ou objet de satisfaction de l'ego, mais comme sujet, lui aussi à la recherche de lui-même.

## Références bibliographiques

- BAUMAN Z. (2000), *Liquid Modernity*, Cambridge, Polity Press.
- BECK U., GIDDENS A., LASH S. (1994), *Reflexive Modernization : Politics, Tradition and Aesthetics in the Modern Social Order*, Cambridge, Polity-Press.
- BECK U., (2001), *La société du risque. Sur la voie d'une autre modernité*, Paris, Aubier.
- DERY M. (1996), *Escape Velocity : Cyberculture at the End of Century*, Berkeley, Grove.
- DUMONT L. (1983), *Essais sur l'individualisme*, Paris, Seuil.
- EHRENBERG A. (1995), *L'individu incertain*, Paris, Calmann-Lévy.
- EHRENBERG A. (1998), *La fatigue d'être soi*, Paris, Odile Jacob.
- ELIAS N. (1975), *La dynamique de l'Occident*, Paris, Calmann-Lévy.
- GIDDENS A. (1992), *The Transformation of Intimacy*, Stanford, Stanford-University-Press.
- GIDDENS A. (1994), *Les conséquences de la modernité*, Paris, l'Harmattan.
- HABERMAS J. (1973), *La technique et la science comme idéologie*, Paris, Denoël.

- HILTZ S. R. et TUROFF M. (1993), *The Network Nation*, Cambridge, MIT Press.
- JAURÉGIBERRY F. (2003), *Les branchés du portable. Sociologie des usages*, Paris, PUF.
- LASCH C. (1980), *Le complexe de Narcisse*, Paris, Robert Laffont.
- MEAD G. H. (1963), *L'Esprit, le Soi et la Société*, Paris, PUF.
- RHEINGOLD H. (1993), *The Virtual Community : Homesteading on the Electronic Frontier*, Reading, Addison-Wesley.
- SENNETT R. 1979), *Les tyrannies de l'intimité*, Paris, Seuil.
- TAYLOR C. (1998), *Les sources du moi*, Paris, Seuil.
- TOURAINE A. (1992), *Critique de la modernité*, Paris, Fayard.
- TURKLE S. (1997), *Life on the Screem. Identity in the Age of Internet*, New York, Touchstone.
- VALLEUR M. (2003), *Sexe, passion et jeux vidéo*, Paris, Flammarion.
- WOODCOCK B. C. (2003), *An Analysis of MMOG Subscription Growth* (<http://pw1.netcom.com/~sirbruce/Subscriptions.html>).

# La récursivité et le mythe de *bootstrap* dans la physique moderne

Jerzy Karczmarczuk

Dept. d’Informatique, Université de Caen, France  
(<mailto:karczma@info.unicaen.fr>)

## Résumé

Nous discutons sur quelques exemples le concept d’« auto-construction » des structures ou entités *matérielles* dans le cadre de notre compréhension de physique moderne. Les hypothèses du *bootstrap* en physique ont montré plusieurs fois que l’approche réflexive à la description des états ou des structures peut résulter non pas seulement en une vision philosophique, mais fournir quelques outils calculatoires, et donc enrichir le « noyau dur » scientifique concerné. Cependant, il est toujours facile de donner – justement – trop d’importance conceptuelle à ces auto-références, cycles, etc. calculatoires, qui deviennent ainsi un peu mythiques (ou magiques). Notre objectif est de demystifier un peu ces mythes sur quelques exemples, mais en présentant leur beauté et leur rôle positif dans la construction des théories quantitatives.

## 1 Introduction

Les « hiérarchies étranges » comme les appelait Hofstadter dans *Gödel-Escher-Bach* [1], l’auto-référence, la cyclicité des structures – tout ceci exerce une attraction magique sur les physiciens. La vision « organique » de l’Univers composé d’entités qui s’auto-construisent, de l’Univers qui s’est auto-créé et qui constitue sa propre cause, remplace parfois la pensée religieuse dans le contexte de la Création (souvent sous-consciemment). Voici quelques exemples de spéculations, assez rationnelles et 100% scientifiques dans leur nature, mais contenant toujours des *a priori* impossible à vérifier :

- La cosmologie, les propriétés globales du monde résultent de sa structure locale, intrinsèque ; on sait que la densité de la matière détermine l’ouverture

ou la fermeture de l'espace ; que la dynamique locale *constitue* les propriétés du temps, etc. En même temps, les propriétés des micro-particules, leur masse, charge, etc., dépendent fonctionnellement des interactions avec l'Univers. C'est la géométrie globale de l'Univers qui « crée » la masse ; la topologie des dimensions internes (« occultes ») et la structure du *vide* quantique spécifient les charges électromagnétiques...

- Les particules interagissent en échangeant ... des particules ; il n'y a aucune différence entre la matière et les forces. Les particules excitées énergétiquement se désintègrent en d'autres particules, qui peuvent se recombiner ; il n'y a pas de différence entre le composé et l'élémentaire. Un électron est tel qu'il est, grâce à la présence d'un nuage de particules virtuelles autour. Parmi ces particules il y a des électrons, bien sûr. La profondeur d'une telle structure est infinie, comme dans une récurrence mathématique infinie.
- Un système à plusieurs corps : solide, plasma, matière nucléaire (p. ex. les pulsars neutroniques), se manifeste comme un *tout* macroscopique assez indépendant des propriétés des composantes à l'échelle micro. Dans un régime critique (proche d'une transition de phase) les détails microscopiques sont presque totalement écrasées par des forces agissant à longue distance. On peut trouver des identités qui connectent des modèles : système vu d'un mètre de distance, et vu d'un kilomètre. Mais, sachant que les distances microscopiques sont d'ordre de  $10^{-8}$  cm (atomes) ou  $10^{-13}$  cm (nucléons), les propriétés critiques de la matière « se construisent elles-mêmes » à partir des attributs indépendants de la structure microscopique, dépendants toutefois de la densité globale, de la dimension géométrique des structures, des connexions entre les éléments, etc. Cependant **toutes** les propriétés de ce corps doivent être le résultat des interactions élémentaires, on n'en doute pas, alors il y a un clivage conceptuel ou méthodologique dans la chaîne cause ↔ effet ...
- Dans la description classique de la Théorie Générale de Relativité, le monde se décompose en *matière* (properment dite) : les sources du champ gravitationnel, et ce *champ* avec ses équations non-linéaires. Or, les équations du champ sont si « tordues », qu'elles contraignent énormément les positions et autres propriétés des *singularités* du champ, souvent identifiées avec les sources. Alors, le comportement de la « matière » qui – apparemment – engendre le champ, en est la conséquence ! La gravitation détermine les attributs de ses propres sources ...

On peut continuer ainsi longtemps, mais – soulignons le – *la physique n'est pas un domaine philosophique spéculatif*. Elle doit répondre aux questions quantitatives, donner des prévisions, et s'appuyer sur une solide base Popperienne de falsifiabilité. C'est la raison principale pour laquelle nous appelons ce phénomène de se tirer par ses propres cheveux, un *mythe*. Il *fonctionne* comme un mythe malgré sa rationalité et l'appui mathématique, il constitue une sorte de croyance ou d'une interprétation un peu métaphysique, ce qui est un peu étonnant, vu la logique typiquement Cartésienne des physiciens. Il n'y a rien de nuisible dans ce mythe, seulement le Léninisme orthodoxe voulait traiter Mach, et autres philosophes de la Nature qui croyaient en une sorte d'inséparabilité entre le local et le global, comme des criminels de la pensée... Le mythe de bootstrap est riche et inspirant, mais il doit être compris pour progresser. Nous allons essayer de désechanter ce mythe, en rappelant quelques principes communs à l'élaboration des théories basées sur des formalismes mathématiques.

Nous allons discuter aussi deux modèles de théories physiques, où les spéculations sur l'auto-référence fournissent une méthodologie de travail rationnel et fructueux du point de vue applicatif. Nous allons parler du *bootstrap statistique* dans la physique de hautes énergies, et du *groupe de renormalisation*. Le premier modèle a vécu son temps, il est considéré comme une étape vers une théorie plus basique, il est devenu un modèle phénoménologique qui est beaucoup moins fondamental que l'on n'y pensait dans les années '60.

Cependant l'autre, qui a engendré quelques prix Nobel, constitue toujours une des perles de la physique contemporaine, et possède plusieurs variantes, conceptuelles, calculatoires et applicatives. Mais il est très difficile à présenter aux non-spécialistes, donc nous allons le traiter *très* superficiellement.

## 2 Récursivité, et diagrammes de Feynman

Le cas le plus archétypique de la réfléxivité/auto-référence dans la physique ce sont des équations de la théorie quantique des champs, qui prennent parfois la forme des équations de point fixe :  $p = F(p)$ , où  $p$  est une entité théorique, p. ex. l'amplitude de propagation d'une particule, et  $F$  – une fonctionnelle compliquée qui réalise la théorie en vigueur. Déjà une telle forme suscite une spéculations naturelle : la physique, la matière, est-elle récursive, comme le forme ci-dessus? Notre opinion là-dessus est plutôt négative. La récursivité est une propriété de *description*. Si on cherche la solution de l'équation  $e^{-x} - x = 0$  on

peut procéder par la recherche itérative/récursive du point fixe, mais le nombre  $0.567432904\dots$  n'a rien de récursif *per se*. Nous sommes donc plutôt favorables à la thèse de Chomsky, selon qui c'est la *syntaxe du langage* qui contient des entités récursives, même s'il nous est difficile de prendre la récursivité comme quelque chose de psychologique et humaine [2].

Il est souvent possible d'appliquer la *théorie des perturbations* [3] et d'obtenir le développement du genre  $p = (\dots p \dots) + e(\dots p \ p \dots) + e^2(\dots p \ p \ p \dots) + \dots$ , où à droite on observe les occurrences « polynomiales » de  $p$  pondérées par les puissances de la charge  $e$ , considérée petite, et on commente : une particule, deux, trois... ce qui aboutit à l'assertion « mythique » : une particule est composée de plusieurs particules, qui, elles aussi... Ainsi on donne de la signification physique aux diagrammes de Feynman, comme celui dessus. Un électron engendre un photon virtuel qui se désintègre en paire : électron-positron, et finalement tout se recombine en présence d'un autre photon virtuel émis avant, et absorbé après la recombinaison.

Les physiciens sont placés devant un petit mystère de la Nature. D'une part on *sait* que la théorie des perturbations est incomplète (et, d'ailleurs, les séries infinies qu'elle engendre ne convergent pas...), d'autre part, les premiers termes qui contiennent un nombre petit de particules virtuelles donnent une raisonnable approximation ; on « voit » (indirectement) ces particules, et on a ainsi une tendance à les traiter sérieusement ; parfois on peut leur attribuer un effet observable particulier. Si on sépare des « bulles irréductibles » qui modifient l'amplitude de propagation d'une particule et se réduisent, et si on les itère comme sur la Fig. 2, on constate un phénomène intéressant.

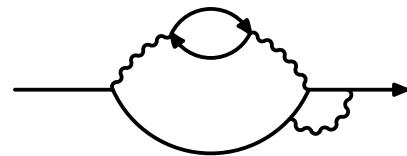


FIG. 1 – Propagation d'une particule « vêtue »



FIG. 2 – Renormalisation de la masse

D'abord, cette amplitude (« propagateur ») représentée par possède la forme mathématique  $d(q) = 1/(q^2 - m^2)$ , où  $q$  est la 4-impulsion de la particule, et  $m$  – sa masse. Ensuite on voit que si la « bulle » est appelée  $A(q)$ , sans la connaître nous pouvons écrire le propagateur modifié comme  $\bar{d} = d + dA d + \dots$

$dAdA + \dots = d + dA\bar{d}$ , une équation récursive qui possède la solution  $\bar{d} = d/(1 - dA) = 1/(q^2 - (m^2 + A))$ . Alors, dans le cas général, les particules virtuelles qui récursivement bâissent notre objet, modifient sa masse (si  $A(0) \neq 0$ ). C'est ainsi que les photons dont la masse est zéro, en traversant un milieu « intéressant» comme le plasma (matière ionisée, dont la température est de plusieurs milliers, voire millions degrés), deviennent plasmons, des particules avec masse. Dans ce contexte la «bulle» n'est pas la «ionisation du vide», la création des électrons/positrons virtuels, mais la ionisation réelle de la matière, le photon est constamment absorbé/réemis par des particules chargées, et il devient lourd, car il traîne avec soi ce nuage de polarisation.

Mais est-ce la récursivité de la définition du propagateur qui en est responsable? Elle n'est qu'une représentation «syntaxique» de la nonlinéarité des interactions.

### 3 «Fireballs» et le *bootstrap* statistique

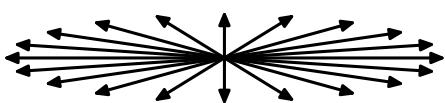


FIG. 3 – *Distribution des particules créées lors d'une collision de hautes énergies*

La collision de deux protons dont l'énergie est d'ordre de plusieurs GeV donne lieu à une “explosion”, produisant souvent des dizaines (parfois centaines) de mésons et autres particules. Le raisonnement statistique - thermodynamique vient à l'esprit immédiatement, on sait que la libération d'énergie dans un espace localisé, chauffe la matière qui commence à rayonner. Une collision de basse énergie produit surtout la radiation EM, les photons, l'explosion est une boule incandescente visible à l'oeil. Si l'énergie est plus haute, on voit les rayons gamma, mésons, anti-protons, tout... Déjà dans les années '60 on a constaté une chose bizarre : lors de ces collisions la distribution des impulsions n'est pas isotrope ; on voit une distribution qui qualitativement peut être dessinée comme sur la Fig. 3. On sait que pour des raisons cinématiques même une collision centrale doit propager l'énergie de la collision surtout le long de son axe, mais on attendait à ce que les «éclats» perpendiculaires à cet axe soient aussi importants. Or, même si les mésons émis transversalement sont nombreux, leur énergie est limité, leur distribution décroît exponentiellement. Les mésons simplement «ne veulent pas» avoir de hautes énergies, ne veulent pas être trop «chauds». Le «fireball» préfère être plus lourd que plus chaud.

La collision de deux protons dont l'énergie est d'ordre de plusieurs GeV donne lieu à une “explosion”, produisant souvent des dizaines (parfois centaines) de mésons et autres particules. Le raisonnement statistique - thermodynamique vient à l'esprit immédiatement, on sait que la libération d'énergie dans un espace localisé,

chauffe la matière qui commence à rayonner. Une collision de basse énergie produit surtout la radiation EM, les photons, l'explosion est une boule incandescente visible à l'oeil. Si l'énergie est plus haute, on voit les rayons gamma,

mésons, anti-protons, tout... Déjà dans les années '60 on a constaté une chose bizarre : lors de ces collisions la distribution des impulsions n'est pas isotrope ; on voit une distribution qui qualitativement peut être dessinée comme sur la Fig.

3. On sait que pour des raisons cinématiques même une collision centrale doit propager l'énergie de la collision surtout le long de son axe, mais on attendait à ce que les «éclats» perpendiculaires à cet axe soient aussi importants. Or, même si les mésons émis transversalement sont nombreux, leur énergie est limité, leur distribution décroît exponentiellement. Les mésons simplement «ne veulent pas» avoir de hautes énergies, ne veulent pas être trop «chauds». Le «fireball» préfère être plus lourd que plus chaud.

Dans la physique statistique plus classique c'est un phénomène totalement inconnu. Pour la matière classique la température d'un système thermique est proportionnelle à son énergie moyenne. Cependant, nous savons que pour la matière rayonnante, pour un bain thermique constitué de photons, la loi de Stefan-Bolzmann est respectée, et celle-la dit que  $T \propto \sqrt[4]{E}$ . La température « ne veut pas » croître trop rapidement, car l'énergie sert à fabriquer d'autres photons.

Or, les interactions fortes, nucléoniques, fabriquent plus que ça, on voit des mésons, baryons, et surtout des **résonances** : particules lourdes, instables, jamais observées, car de *très* courte vie ; on les observe indirectement, en constatant que deux ou trois particules visibles ensemble possèdent la masse qui correspond à ce resonance, et dont la distribution est très spécifique. Alors, la température peut (ou doit) monter encore plus lentement.

Resonances plus lourdes peuvent être constitués des plus légers, et ils sont difficiles à observer même indirectement, mais ils sont sans doute produits. Rolf Hagedorn (1919 – 2003)[4] s'est posé la question : *et si on pouvait dire, que l'essence des interactions fortes consiste à produire les résonances, qui produisent des résonances, qui sont constitués des résonances, qui...* – bref, que la matière hadronique soit **par excellence auto-référentielle**.

Une particule plus lourde a plus de « liberté » de se constituer, car on peut faire plus de choses si on a plus d'énergie. Ceci dit que les particules plus lourdes doivent être plus nombreuses, que le *spectre* des particules,  $\rho(m)$ , la densité (nombre) des objets en fonction de leur masse, doit croître rapidement avec  $m$ . Hagedorn a profité de ce spectre expérimental à basses masses (énergies), et il a extrapolé plus haut en utilisant sa philosophie auto-référentielle, nommée plus tard le **bootstrap statistique**. Il a constaté que ses formules pour  $\rho(m) \propto \delta(\text{petites masses connues}) + \sum_N \dots \int \prod_{i=1}^N \rho(m_i) \dots$  résultent en *spectre exponentiel*,  $\rho(m) \propto \exp(m/T_H)$ , où  $T_H$  était une constante expérimentale, autour de 150 MeV, ou  $10^{13}$  degrés Kelvin. (Les éléments manquantes dans l'équation contiennent l'intégration sur les attributs cinétiques, éléments géométriques comme le volume du fireball, et quelques facteurs combinatoires. La température de Hagedorn est une fonction de ces constantes. L'agrément avec l'expérience est remarquable).

Un tel spectre n'a jamais été observé ou proposé auparavant, puisque la matière plus classique (atomique) est composé d'entités fixes, le spectre descend vers zéro rapidement, il n'y a pas d'atomes trop lourds. À l'échelle nucléaire on a les nucléons, et les mésons, peut-être des hypérons pas trop lourds, et le spectre

s’arrête.

Mais la matière hadronique créé par ses interactions des nouveaux hadrons, toujours plus lourds, et ceci implique que l’énergie va être épuisée dans cette création, et non pas pour augmenter l’énergie cinétique. La température est limitée, la matière ne veut pas être plus chaude que 150 MeV !

Voici une conséquence de l’hypothèse d’auto-référence structurelle dans le monde hadronique. Si on postule – dans le cadre de la physique quantique, statistique – que la matière est composée des entités dont les fragments sont des animaux de même espèce que les objets entiers, on obtient après quelques manipulations formelles relativement faciles, mais trop spécifiques pour que l’on puisse les élaborer ici – le célèbre spectre exponentiel. Si on présente cette histoire de cette manière on a l’impression de dévoiler un mystère de la Nature, et de justifier la philosophie de réfléxivité dans le contexte discuté.

Et, réellement, le modèle de Hagedorn a fait beaucoup de bruit. Cependant actuellement notre philosophie est un peu différente, moins ésotérique.

- On sait – selon les « croyances scientifiques » actuelles – que la matière hadronique est composée de particules de plus bas niveau : quarks, gluons, etc. On ne peut plus considérer le bootstrap de Hagedorn comme quelque chose qui touche les fondements de la matière. Le modèle est devenue une description phénoménologique, superficielle, même si quantitative.
- Ainsi, le spectre exponentiel *doit* être produit différemment *aussi*, par exemple par la théorie combinatoire qui nous dit quelles sont les structures (plus ou moins) stables, produites à partir des quarks. On est loin de cet objectif, mais on progresse. Le modèle récursif est incomparablement plus simple !
- La température maximale de Hagedorn n’est plus considérée comme *absolue*. Au contraire, elle correspond « à peine » à la transition entre la matière hadronique « normale » (mésons, resonances), et le plasma de quarks et gluons libérés (comme on libère les électrons en chauffant les atomes, et, puis, on désintègre les noyaux dans une explosion nucléaire). On voit déjà des « jets » de haute énergie transversale, qui suggèrent des transitions de phase localisées. Mais selon les modèles fondamentaux actuels, nous ne verrons jamais des quarks libres, notre Univers se trouve globalement dans une autre phase...

Voici un joli et très inspirant mythe qui appartient au passé. En fait, il été handicapé dès le début, et Hagedorn en était conscient. L’auto-référence poussée aux limites, une cyclicité de description de hadrons, qui leur donnait un statut absolu, fondamental, était dès le début contraire à ce que nous avons appris pendant des

siècles : on trouve toujours quelque chose « en dessous »... Donc, maintenant on a une tendance de dire que Hagedorn a involontairement contribué à la théorie des quarks, ce qui était, dans les années '60, très loin de ses convictions !

## 4 Invariance d'échelle et groupe de renormalisation

Une expérience montrée souvent aux étudiants en physique consiste à refroidir un gaz comme le fréon dans un tube fermé. Il est transparent. Après le refroidissement il est de nouveau transparent (partiellement liquide). Mais autour d'une température dite critique il devient opaque. Il s'avère que dans des conditions critiques les corrélations entre les molécules sont d'ordre de grandeur comparables à la longueur d'onde de la lumière. Les « clusters » de cette taille se constituent et disparaissent, ce qui provoque une très forte dispersion/absorption de la lumière. On parle d'*opalescence critique*.

Vu de perspective microscopique les corrélations deviennent de longueur *beaucoup* plus grande que toute grandeur d'ordre atomique, on peut la considérer comme infinie, se comportant comme  $|T - T_c|^{-\nu}$  avec  $\nu$  autour de 1/3. Dans des conditions des courbes *réduites* densité/température, par exemple  $\rho/\rho_c$  en fonction de  $T/T_c$  s'avèrent similaires, voire identiques pour plusieurs substances. On voit le même comportement des substances magnétiques où la polarisation ou la susceptibilité magnétique deviennent également universelles. On parle d'*invariance d'échelle* de ces phénomènes.

Une spéculation particulièrement fructueuse commencée par Leo Kadanoff a donné le prix Nobel à Ken Wilson [5] et est connu sous le nom du groupe de renormalisation [6]. La similitude entre plusieurs systèmes critiques suggère que les détails, l'essence même de leur microstructure *cessent d'être importants*. Le système vu de la perspective macroscopique « tire ses propriétés par les cheveux » à partir de... bonne question : de quoi ?

On a découvert qu'il est possible de proposer *indépendamment des détails microscopiques du système* un modèle parfois assez qualitatif, *paramétré* par un certain nombre de variables qui peuvent dépendre de l'échelle sous laquelle le système est regardé. Par exemple, une substance magnétique est décrite par l'énergie d'interaction :  $H = J \sum_{i,k} S_i \cdot S_k$ , où  $S$  sont les moments magnétiques, la somme s'effectue sur les voisins, et  $J$  est le paramètre, le coefficient du couplage entre les spins.

Ensuite on dit : si on regarde ce système de plus loin, on doit observer *à peu-*

*près le même comportement*, seulement les paramètres varient.  $J$  devient une fonction de distance. Les spins peuvent devenir plus compliqués, par exemple Kadanoff a proposé de remplacer les spins individuels par des blocs de spins voisins, qui seraient traités ensemble comme un « macro-spin ». Ensuite on forme des blocs des blocs, etc., *ad infinitum*. Le modèle est reformulé – par exemple – pour fournir la probabilité d'un état énergétique (macroscopique), ce qui donne des valeurs des propriétés thermodynamiques.

On écrit des équations du genre :  $p(\lambda x) = F(\lambda, p(x))$  où  $p(x)$  est la quantité qui nous intéresse,  $\lambda$  un facteur d'échelle, et  $F$  – une fonctionnelle qui dit comment la propriété varie avec l'échelle. Mais : si on s'approche à la température critique on voit l'invariance d'échelle, le système doit se « stabiliser », par exemple on doit pouvoir dire quelque chose sur  $p(\infty)$  comme le point fixe de l'équation ci-dessus. Le fait que le système soit « auto-similaire », qu'il se ressemble à des échelles différentes, permet d'obtenir des réponses quantitatives, par exemple les coefficients comme  $\nu$  et les autres.

Nous avons commencé notre texte en soulignant le caractère anti-spéculatif d'une *bonne* physique. Peut-on proposer un modèle facile à comprendre par *tous* et qui donne des prévisions numériques? Essayons.

## 5 Percolation

La Fig. 4 montre un assemblage de briques noires (conducteurs) et blanches (isolateurs), placés aléatoirement dans un rectangle (pour l'instant regardez l'image à gauche seulement). La question est : est-ce que l'objet est conducteur (de haut vers le bas). La réponse est statistique et dépend de la probabilité  $p$  qu'une brique soit noire. La « théorie microscopique » demande l'analyse des chemins noirs entre les deux « électrodes », et elle est *très compliquée*.

D'habitude on utilise la simulation. Le système est très fragile. Il s'avère que pour  $p$  inférieur à 0.57 le système est *presque toujours* isolant, et pour  $p$  supérieur à 0.62 presque toujours conducteur ! La valeur critique de  $p$  est de 0.5927, quand la probabilité passe par ce seuil le système change brusquement ces propriétés, c'est une transition de phase.

La philosophie du groupe de renormalisation est incroyablement simple (et, il faut préciser toute de suite : elle donne des résultats approximatifs). Effectuons la construction de « blocs de Kadanoff » par la décimation, le remplacement de quatre briques voisines par une, « représentante ». Elle est noire si le bloc est



FIG. 4 – *Percolation ; système d'origine et sa décimation*

globalement conducteur, sinon blanche. Si on réduit le système à une seule brique (sa dimension d'origine doit être alors  $2^N$ ), on obtient la réponse voulue : si la macro-brique est noire, l'amas est conducteur. La Fig. 5 montre les configurations qui résultent en brique conductrice.



FIG. 5 – *Percolation ; blocs conducteurs de haut vers le bas*

Un simple raisonnement combinatoire montre que parmi 16 ( $= 2^4$ ) configurations possibles de 4 briques on a 1 configuration noire, dont la probabilité est de  $p^4$ , 4 avec une brique blanche, avec la probabilité  $p^3(1 - p)$ , et 2 configurations dont la probabilité est  $p^2(1 - p)^2$ .

Alors, après la décimation la probabilité des blocs noirs sera égale à

$$\tilde{p} = p^4 + 4p^3(1 - p) + 2p^2(1 - p)^2.$$

On constate très rapidement que si  $p = 0.67$ , alors  $\tilde{p} = 0.6963$ . Si on remplace à droite de l'équation  $p$  par  $\tilde{p}$ , et si on répète la procédure plusieurs fois, les itérations suivantes donnent  $0.7880, 0.8564, 0.9289, \dots$ . La convergence vers 1 est très rapide.  $p = 0.5$  donne  $\tilde{p} = 0.4375$ , etc., descente vers 0. Les points fixes stables sont triviaux. Vu de loin un tel système se comporte soit comme une brique noire (pour  $p$  grand), soit comme blanche. (Bien sûr, pour une configuration individuelle c'est un truisme, mais on parle de comportement statistique, ce qui est inattendu !)

Mais si  $p = 0.618034$ , on obtient  $\tilde{p} = p$ , c'est la valeur critique de  $p$ . Ce point fixe est instable, on ne peut pas l'obtenir par des itérations simplistes. Du point de vue expérimental le résultat n'est pas exact, il y a 4% d'erreur.

La leçon à retenir de cet exercice est la suivante. Sans connaître la « vraie » procédure permettant de trouver la densité de percolation critique, en postulant l'auto-similarité du système, nous sommes arrivés à trouver le point critique. Le modèle auto-référentiel est relativement primitif, donc le résultat ne peut être exact. Nous avons constaté que la spéculation et l'écriture d'une équation récurrentielle ne suffisent pas. L'itération de notre polynôme converge vers les points stables. Donc, pour pouvoir profiter des techniques de groupe de renormalisation ici et dans des cas plus complexes, le théoricien doit bien connaître des techniques numériques, algébriques, souvent des techniques de solution des équations différentielles rigides, etc.

## 6 Conclusions

Nous n'avons volontairement pas défini le mot « bootstrap », car sa définition reste extrêmement floue. Nous en avons donné des exemples, conceptuels, et un peu calculatoires, laissant le lecteur apprivoiser le terme selon sa sensibilité. En général, l'idée de se soulever dans l'air en se tirant par les lacés (ou par les cheveux, comme l'a fait baron von Münhausen) n'est plus traitée comme un paradoxe, et dans l'informatique possède une signification assez banale : le chargement d'un système d'exploitation par le système lui-même, à partir d'un noyau minuscule. Ou, si on veut, la compilation d'un compilateur d'un langage de programmation par ce même compilateur, écrit dans *ce* langage. On commence par une version minuscule, écrite dans un autre langage, et on itère, en enrichissant le compilateur. Dans la technologie on trouve d'autres exemples du bootstrap, par exemple l'alternateur dans une voiture produira du courant à condition d'avoir un

peu de courant pour alimenter les aimants ; avec une batterie complètement morte on ne démarrera pas, même en poussant la voiture. . .

De ce point de vue les variés boostraps dans les théories physiques constituent le même phénomène. À partir de quelques éléments rudimentaires comme quelques masses de particules connues, quelques idées sur le caractère des interactions, sur les ingrédients géométriques essentiels, on cherche une formulation réflexive, récurrentielle, contenant des autoréférences. Répétons : on *cherche une formulation...*

Dans plusieurs contextes ceci est possible et donne des résultats exceptionnels. Et, paradoxalement, ceci constitue un piège conceptuel. Quelques philosophes de la Nature ont essayé – sans mauvaise volonté – de mythifier le phénomène, d'affirmer que la physique **est** réflexive, que le bootstrap est une méthodologie rationnelle. Mais, jusqu'à nos jours ceci reste un mythe.

Dans ce texte nous avons volontairement effacé toutes les références à l'observateur, aux constructions mentales (vues de cette perspective), même si toute théorie physique n'est qu'une construction mentale. . . Mais le lecteur reconnaîtra qu'une ouverture vers la sémiotique, vers les spéculations concernant la création de la réalité par l'observateur/théoricien ouvre un puits infini. Nous avons voulu éviter cela à tout prix.

## Références

- [1] Douglas R. Hofstadter, *Gödel, Escher, Bach: An Eternal Golden Braid*, Basic Books, (1999).
- [2] M.D. Hauser, N. Chomsky, W.T. Fitch, *The faculty of language: What is it, who has it, and how did it evolve?* Science, **298**, pp. 1569–1579, (2002).
- [3] C. Itzykson, J.-B. Zuber, *Quantum Field Theory*, McGraw-Hill, (1980), ou M.E. Peskin, D.V. Schroeder, *An Introduction to Quantum Field Theory*, Addison-Wesley (1995), ou n'importe quel autre livre sur la théorie quantique des champs. . .
- [4] R. Hagedorn, Nuovo Cim. Suppl. **3** p. 147, (1965); Nuovo Cim. **56A**, p. 1027, (1968).
- [5] K.G. Wilson, Phys. Rev. **B4**, p. 3174; 3184, (1971). Voir aussi : *Problems in Physics with Many Scales of Length*, Rev. Mod. Phys. **55**, p. 583, (1983).
- [6] R.J. Creswick, H.A. Farach, C.P. Poole, *Introduction to Renormalization Group Methods in Physics*, Wiley, (1991).

# Réflexivité en linguistique, en discours et en grammaire

Nadine Lucas  
GREYC CNRS Université de Caen  
Nadine.Lucas@info.unicaen.fr

Résumé: La linguistique parle du langage, une forme de réflexivité en sciences, et les sciences de l'homme posent le problème du statut de l'observateur. Réciprocité, inversion des rôles, réflexion et réflexivité sont aussi des termes du métalangage en grammaire et en rhétorique. En étudiant les formes réflexives, on s'aperçoit que les distinctions habituelles entre sujet et objet, entre actif et passif se diluent. Les règles d'accord sont compliquées. En étudiant les verbes réfléchis sur quelques siècles, on constate des variations. En discours, les formes réflexives avec *se* et avec *lui-même* abondent dans les explications. Elles favorisent les transitions de topiques et mènent parfois à l'allégorie. Elles marquent les passages où le lecteur doit s'arrêter pour réfléchir.

Mots- clefs: réflexif, réfléchi, linguistique, français, analyse de discours, grammaire, se, lui- même

## 1. Introduction : La notion de réflexivité en linguistique

On a beaucoup glosé sur la réflexivité de la discipline : on parle du langage (avec le langage). Mais ceci vient d'une histoire, celle du discours scientifique en général, qui est discours sur le monde. Il semble que périodiquement le désir de l'homme de comprendre la nature doive être justifié. Galilée emploie ainsi une comparaison, souvent reprise, sur le livre du monde : "[*L'univers*] ne peut être lu avant que nous n'ayons appris son langage et que nous ne nous soyons familiarisés avec les caractères dans lesquels il est écrit. Il est écrit en langage mathématique, et les lettres sont les triangles, cercles et autres figures géométriques, sans le moyen desquelles il est humainement impossible de comprendre un seul mot." Opere Il Saggiatore p. 171.

Le postulat de la linguistique est que la langue est un objet de science, qui peut s'étudier objectivement comme les objets de la nature. Il ne peut s'installer, au XIXe siècle, que parce que les sciences de la nature ont gagné leurs lettres de noblesse. Actuellement la métaphore est inversée et on parle volontiers d'univers du discours, ou de l'univers du texte par référence à la physique.

En sciences humaines, la réflexivité est inhérente puisque des hommes étudient d'autres hommes pour définir l'Homme. La réflexion est toujours nécessaire, du fait qu'en anthropologie

comme en linguistique, l'observé et l'observateur peuvent changer de rôle. Par exemple, un linguiste français à la recherche des traces de la langue aïnoue au Japon peut rencontrer un linguiste japonais spécialiste du breton qui l'interrogera sur sa langue oubliée. La conscience de la précarité du rôle d'observateur se traduit à un autre niveau par des règles de déontologie, partagées par d'autres disciplines que la linguistique, ce qui implique une retenue (ne pas faire à autrui ce qu'on n'accepterait pas de subir). Il s'agit d'intégrer une possible réciprocité.

Pourtant, on observe aussi beaucoup d'incertitudes (faut-il nommer ses informateurs ?) voire de dérives : à un extrême, l'observateur s'approprie son objet d'étude comme extension de son monde et de ses valeurs : il faut étudier du "bon" français, celui de la classe dirigeante, et l'imposer comme norme de référence, tendance dénoncée par Bourdieu (1962) ou Yaguello (1988). À l'autre extrême, l'observateur ne se reconnaît plus comme appartenant à un groupe. Fuchs définit ainsi les difficultés de l'objectivité vs l'introspection et le solipsisme.

*On ne naît pas linguiste, on le devient, alors que l'on naît sujet parlant (en puissance). Et on devient linguiste en apprenant à se dissocier du sujet parlant que l'on est (situation légèrement schizophrénique, j'en conviens !). [...] Ce qui importe, c'est le regard que l'on porte sur les langues, regard distancié de l'observateur, qui transforme même la langue maternelle en une langue étrange, sinon étrangère. (Fuchs, 1995).*

## 2. La notion de réflexivité en grammaire

Pour en venir à des questions plus concrètes, les termes *réflexif*, *réflexivité* et *réfléchi* font partie du vocabulaire des grammairiens. Comme on le verra, ces termes sont associés à *réciproque* et à *irréfléchi* d'une part, à *pronominal* d'autre part.

### 2.1. Pronoms réflexifs et problèmes d'accord

Les formes réflexives couvrent des emplois pronominaux complexes. On parle de *réfléchi* ou de *réflexif* lorsque le pronom sujet et le pronom objet sont co-référents, mais aussi lorsque le sujet nominal est co-référent avec le pronom objet.

- 1) Je me suis levé. Je me suis aperçu qu'il pleuvait.
- 2) Pierre s'est levé.

On parle de *réciproque* lorsque l'action va de l'un à l'autre des partenaires (ou des uns aux autres) chacun d'entre eux étant

successivement (ou simultanément) sujet et objet (Ribeiro, 2000) comme dans "se battre".

3) Ils se sont battus.

Ces formes pronominales sont un casse-tête en orthographe parce qu'elles mettent en jeu des accords peu évidents. Elles sont très ardues pour les apprenants du français<sup>1</sup>. L'accord neutre avec l'objet fait partie de ces difficultés. C'est en fait une annulation de l'accord en genre et en nombre avec le sujet, qui est normal et attendu avec l'auxiliaire *être*.

4) Je suis épuisée. Je me suis fatiguée à courir les magasins. (accord);

5) Finalement, je me suis acheté une robe. (c.o.d. postposé donc pas d'accord)

6) Elle s'est lavée. (se = elle-même = c. o. d. donc réfléchi et c.o.d. antéposé donc accord);

7) Elle s'est lavé les mains. (s' = à elle-même = c. o. i. donc réfléchi et c.o.d. postposé donc pas d'accord)

On constate que la présence d'un complément d'objet direct suivant le verbe permet de désactiver l'accord avec le sujet. Quand on passe aux propositions subordonnées, on admet l'accord avec l'antécédent (et non avec le sujet de la proposition subordonnée).

8) La guitare qu'il s'est fabriquée (= qu'il a fabriquée pour lui-même \_\_ réfléchi);

Mais les choses se gâtent assez rapidement au pluriel, puisque ni la distinction entre réfléchi et réciproque, ni la règle sur le complément antéposé ne servent plus à distinguer les cas d'accord ou de désaccord.

9) Elles se sont battues (se = l'une l'autre = c. o. d. \_\_ réciproque);

10) Elles se sont plu (se = l'une à l'autre, les unes aux autres \_\_ réciproque);

11) Les devoirs qu'ils se sont corrigés (= qu'ils ont corrigés l'un à l'autre \_\_ réciproque);

12) Les événements se sont succédé (= ont succédé les uns aux autres \_\_ réciproque);

13) Ils se sont nui (= ils ont nui à eux mêmes; se = c. o. ind. \_\_ réfléchi).

On admet que si le pronom antéposé renvoie à un complément indirect, alors il n'y a pas d'accord. Généralement, les exemples de grammaire mettent en scène des personnes, et peu de

<sup>1</sup> Les exemples suivants et leurs commentaires sont en grande partie empruntés au GUIDE DE GRAMMAIRE FRANÇAISE: SYSTÈME VERBAL de Guilherme Ribeiro

[http://www.prof2000.pt/users/esjmlima/grafra/GUI\\_GRU-W.HTM](http://www.prof2000.pt/users/esjmlima/grafra/GUI_GRU-W.HTM)

commentaires sont faits sur les formes pronominales impersonnelles, liées à l'indéfini. Ce sont les "irréfléchis" dont certains, selon Ribeiro, sont "spéciaux", certains autres sont "à sens passif". Enfin, il existe des verbes essentiellement pronominaux, qui ne présentent pas de variation (transitif/ réflexif) comme *s'abstenir*. L'accord avec le sujet reste opérant dans ces cas.

- 14) Elle ne se serait jamais attendue à cela! (irréfléchi à sens spécial)
- 15) Cette pièce s'est jouée pendant six mois (s'est jouée = a été jouée  
— irréfléchi à sens passif)
- 16) Les voitures se sont bien vendues cet été. (irréfléchi à sens passif)
- 17) Ils se sont abstenus aux élections (essentiellement pronominal)
- 18) Elles se sont absentes sans autorisation (essentiellement pronominal)
- 19) On s'amuse bien. (essentiellement pronominal)
- 20) Il se fait tard.
- 21) Ça se passe bien.
- 22) Il se passe des choses.

Bref, les explications sont embrouillées. Il ressort en tout cas que l'opposition sujet- objet est oblitérée dans les constructions réflexives avec *se*. Et qu'elles sont fréquentes dans les tournures impersonnelles (Hériaud, 1980). Comme l'opposition actif- passif est également oblitérée, on parle de voix pronominale.

## 2.2. Verbes réfléchis

Les constructions verbales dites réflexives ou réfléchies, avec des verbes essentiellement pronominaux, (désignés en abrégé par verbes pronominaux), sont toujours un objet d'étude en grammaire. Elles présentent plusieurs degrés de difficulté. Nous verrons ici quelques interrogations soulevées par Larjavaara, (2002). Les verbes essentiellement pronominaux sont instables dans le temps, en diachronie, ainsi, au XVI<sup>e</sup> siècle et jusqu'au XVII<sup>e</sup> siècle les verbes *commencer*, *jouer*, *apparaître* ou *disparaître* sont tantôt intransitifs tantôt réflexifs ; on dit *s'apparaître* ou *se commencer* comme on dit *se terminer* (Haase 1925, cité dans Larjavaara, 2002).

- 23) Le Seigneur s'apparut à lui dans un buisson ardent (Mass., 2<sup>e</sup> sermon sur la Purification, I, 147)
- 24) Même débat, même jeu se commence (La Fontaine, Contes, 2, 6, 56).

L'inverse est également vrai, puisque des verbes comme *se moucher* ne se trouvent plus actuellement qu'à la forme réflexive,

alors qu'ils étaient soit transitifs soit intransitifs à l'origine (*Vous mouchez beaucoup ?* signifiant *Vous mouchez- vous beaucoup ?*). Enfin Clédat (1881, *ibid.*) cite une évolution possible de verbes uniquement transitifs à l'origine comme *renouveler quelquechose* qui donnent lieu ultérieurement à un emploi réflexif *se renouveler*.

Au XXe siècle, on voit toujours à l'œuvre ce phénomène et on note un usage de verbes transitifs qui deviennent intransitifs via un stade réflexif : *Les examens m'angoissent* puis *Ça m'angoisse* puis *Je m'angoisse* et enfin *J'angoisse* (Larjavaara, 2000). Larjavaara note que la forme réflexive, quand elle est concurrente avec une forme intransitive courante aux XVIe et XVIIe siècle, est contextuellement liée à l'imaginaire, voire au surnaturel. Le réel est décrit par une forme intransitive, le surréel est décrit par une forme réflexive, comme dans les exemples suivants tirés de la base Frantext.

26) Encor un nouveau crime apparoist à mes yeux ! (Scudéry, 1640)

27) Il ne se passe nuit que les morts ne s'apparoissent à elle, avec des formes estranges et un attirail épouvantable qu'elle leur donne (Guez de Balzac, 1631).

Pourtant cette distinction n'est pas décisive, puisque des apparitions (disparitions) miraculeuses ou imaginaires peuvent être accompagnées ou non du verbe réfléchi.

28) Cela dict, la Déesse se disparut et retourna en l'océan. (Flore, 1537)

29) ces fleurs devindrent ronces et disparurent soudainement comme fantosmes. (Brués, 1557)

Le critère d'agentivité (entendue comme le caractère animé et conscient du sujet) est proposé pour expliquer la forme réflexive. Mais ce critère n'est guère plus opérant, puisque Déesse ou diable sont agentifs et également dotés de volonté.

30) Le diable disparut. (Rabelais, 1552)

Curieusement donc, les constructions réfléchies échappent aux grammairiens comme aux philologues.

### 3. La notion de réflexivité en discours

Dans un sens très général, certains discours sont spontanés, d'autres sont réfléchis. Mais aussi, réflexifs, au sens où ils reviennent sur eux- mêmes et montrent que l'auteur réfléchit (≠ discours impératif). Dans beaucoup de textes et spécialement dans le discours scientifique, l'auteur prend du recul, compare, évalue, au moment de conclure. On trouve justement un grand nombre de verbes pronominaux ou de constructions réflexives dans ces

textes. Qui plus est, on les rencontre surtout dans les conclusions ou dans les passages de transition. Supposons donc que le terme "réflexif" vienne de l'observation du discours et se soit cristallisé sur une classe de formes verbales, ainsi que sur certains pronoms.

### *3.1. Les verbes réflexifs dans les explications*

Commençons par la forme *se*. Il est typique d'observer des verbes réflexifs dans une explication locale. Ils sont du type irréfléchi dans la plupart des cas.

#### 31) Vie et mort d'une bulle d'air.

Par temps calme, (force trois à six Beaufort), le vent arrache à la crête des vagues de petites gouttes d'eau qui trouent la surface de la mer en retombant (figure 1). Ce mécanisme de « perforation » conduit à la production d'une multitude de bulles d'air qui explosent en deux temps à la surface (figure 2). Premier temps : la paroi de la bulle **s'amincit** puis crève en projetant des gouttelettes d'environ un centième de millimètre (les gouttes de film). Deuxième temps : l'explosion initiale crée une onde de dépression qui **se propage** vers le bas, vient frapper la coupe concave que forme la paroi inférieure de la bulle et provoque le jaillissement dans l'air de plusieurs gouttelettes d'eau (les gouttes de jet) d'une dimension d'environ un dixième de millimètre.

[FOV 8] LABEYRIE, J. Le pétilllement marin, *Sciences et Avenir*, hors- série n° 58 (1994) pp. 48- 53.

32) Il existe d'autres sables mouvants — beaucoup plus rares et dans les déserts —, sans eau, mais avec de l'air. Là, les grains de sables ne sont pas en suspension. Ils **se touchent**, forment des voûtes et des trous. Des études récentes, menées par le physicien américain R. A. Bagnold, ont montré que des grains suffisamment fins (10 à 20 microns, autour de 1/100 à 1/150 de mm de diamètre) pouvaient former, en **se déposant** au hasard, un véritable gruyère de sable. Un peu comme les briques du jeu sur ordinateur *Tétris* quand on les laisse **s'empiler** toutes seules. Pleine d'air, à la moindre pression du pied, cette véritable poudre **s'effondre** sur elle-même comme un château de cartes. La poudre **se tasse** sous les pieds du marcheur qui **s'enfonce** dans le désert.

[FSV3] POUILLOUX, D. 1997. Les sables mouvants. *Science et vie junior*, (96, septembre), pp. 68- 69.

Le préambule de l'explication locale est marqué par une forme nominale (un titre de légende de figure dans l'exemple 31) ou une forme pronominale impersonnelle (*Il existe*, sans accord sémantique sujet réel- verbe, dans l'exemple 32). L'explication est ponctuée par une série de phrases à la voix pronominale. Elles ont pour effet de présenter les faits de manière non agentive, comme des constats. On remarque que les constructions réflexives du type irréfléchi font intervenir des pronoms objet antéposés (*se*). Mais les vrais réfléchis sont parfois soulignés par des pronoms renforcés (*sur*

*elle-même*). Ce pronom s'accorde avec le sujet de la proposition. Outre les formes négatives, comme dans 32, on rencontre aussi des formes interrogatives introduisant des explications de ce type.

### 33) Comment ces vagues déferlent- elles ?

Plus les fonds remontent, plus la vague **se** cambre. Au-delà d'une certaine limite de cambrure, la vague **se** brise, la crête **se** détache en un jet, peut **se** retourner sur elle-même et plonger ; la surface de l'eau devient un mélange d'eau et d'air.

[FOV 7] SCHMIED, L. La mécanique des vagues, *Sciences et Avenir*, hors-série n° 98 (1994) pp. 40-47.

Dans les deux passages 32 et 33, la forme réfléchie pronominale, *elle-même*, *eux-mêmes* marque une action réflexive, décrite sur le mode non agentif, comme un phénomène spontané.

## 3. 2. Propositions anaphoriques réflexives

Associés au verbe *être*, les pronominaux renforcés signalent certains passages conclusifs, notamment dans les exposés. Il est typique d'observer une forme de ce type au sortir d'une explication subordonnée. Dans l'exemple suivant, une explication étendue sur un phénomène A est donnée, suivie d'une explication plus courte sur un phénomène B.

### 34) L'énigme des crevettes folles

Sous l'eau calme des étangs de Languedoc-Roussillon, un vers plat parasite, le trématode *Microphallus papillorobustus*, rend les crevettes folles. Pas d'amour, mais tout simplement pour assurer sa reproduction, au cours d'un cycle que les chercheurs qualifient de complexe. Ce petit animal a en effet pour particularité d'orchestrer le "suicide" de son hôte, la crevette amphipode *Gammarus insensibilis*. Les larves s'enkystent dans le cerveau de l'hôte, entraînant chez lui des réactions suicidaires aux stimuli extérieurs. Au moindre bruit, alors qu'un gammarus normal plonge pour se cacher, le gammarus "fou" monte vers la surface... et la mort: il est dévoré par un oiseau d'eau, hôte définitif du parasite. Ce comportement a été découvert il y a dix ans par Simone Helluy au sein du Laboratoire de parasitologie comparée de l'université de Montpellier II. Mais *ce* sont les études récentes de l'équipe de François Renaud, directeur du laboratoire, qui ont montré la complexité des rapports entre ces parasites et les gammes, ainsi que les conséquences écologiques et évolutives sur la vie de leurs hôtes. Les études réalisées en milieu contrôlé dans le laboratoire ont ainsi prouvé que les gammes sains et les gammes fous ne se reproduisent pas ensemble. Une véritable ségrégation sexuelle est exercée par les mâles, qui choisissent préférentiellement des femelles saines quand eux-mêmes le sont, *ce* qui les rend plus compétitifs. "Les parasites, explique François Renaud, jouent un rôle primordial dans l'évolution des comportements liés à la biologie de la reproduction de leurs hôtes". Un autre trématode, *Maritrema subdolum*, exploite exactement le même spectre d'hôtes, mais

n'a pas d'effet direct sur la biologie ou le comportement de son hôte crustacé. Il s'enkyste dans la musculature des gammarides fous, profitant ainsi de la stratégie "mise en place" par l'autre parasite pour assurer sa reproduction. En d'autres termes, *Maritrema subdolum* fait du stop... du gammarefoustop! Ces stratégies et ces comportements montrent à quel point les parasites sont des facteurs essentiels dans l'évolution des espèces.

Philippe Dagneaux

[FSV6] DAGNEAUX, P. 1997. L'étrange des crevettes folles. *Le journal du CNRS*, (juillet/août 1997. n°91- 92.), p. 22.

L'explication A concerne la relation crevettes et parasite *Microphallus*. La proposition *eux- mêmes le sont* qui en marque la fin est une proposition anaphorique, de même que *ce qui les rend les plus compétitifs*, parce que ces propositions amènent à regarder en arrière. La proposition réflexive avec *eux- mêmes* contient aussi un pronom attribut antéposé *le*. Cette construction permet d'étendre l'accord au sujet ou à l'agent de la principale (*eux- mêmes* = *les mâles*) mais aussi de désactiver l'accord avec l'attribut (*le* = être sain). L'absence d'accord avec l'attribut est une façon de bloquer un développement. Le lecteur est donc conduit à retenir l'antécédent, et à faire une synthèse pour donner un contenu propositionnel à tous les pronoms sans référent immédiat : *le*, mais aussi *ce*.

Ce texte court nous permet d'introduire le principe général du discours.

« *Une histoire conduit une action d'un état initial à un état terminal à travers des transformations (...). Ainsi une histoire racontée ne comporte pas d'autre stabilité que de pouvoir intégrer des changements, (...) des péripéties, c'est- à- dire des renversements de nos attentes, qui néanmoins contribuent non seulement à l'avancée du récit, mais à son intelligibilité, l'épisode de la reconnaissance venant compenser celui de la péripétie* » (Ricœur, 1990).

Un des moyens de délimiter ces péripéties est l'usage des réflexifs, que l'on trouve dans différents types de discours, narratifs mais aussi expositifs ou argumentatifs. Si le texte est dense, l'objet dont on parle (topique) est immédiatement introduit ici, la relation hôte- parasite. Le topique est exprimé en tant que proposition (*un parasite rend les crevettes folles*). Le proto- rôle d'agent (Dowty, 1991) est porté par le parasite et le proto- rôle de patient est porté par la crevette. Dans l'explication subordonnée, ou sous- thème dérivé, la crevette, qui avait le proto- rôle de patient devient agent et sujet actif (*Le gammare fou monte à la surface*). La fin de l'explication subordonnée est signalée par un verbe réflexif de type réciproque (*se reproduire ensemble*). Mais il

peut y avoir plusieurs péripéties et plusieurs degrés d'enchâssement de discours. La fin de l'explication A est marquée par une phrase réflexive avec pronom renforcé (qui met en jeu deux propositions). Elle concerne un comportement particulier (celui des gammes sains mâles). Ce spécifique est associé (par personification ou allégorie) à un verbe à connotation 'humain' ici *choisir*. Pour remettre tout en ordre, au point de retournement du discours, il est alors nécessaire d'introduire un vrai personnage humain, sujet parlant, comme sujet syntaxique principal, et de rappeler la relation dont on parle, le topique principal, dans une complétive (ici, le discours rapporté). On peut alors entrer dans la seconde explication B concernant les crevettes et le parasite *Maritrema*.

### 3.3. Construction du sens par permutation des rôles

Dans l'exemple précédent, une figure de style est apparue, c'est l'allégorie, qui permet de rendre tangible une idée abstraite, en l'occurrence de personnaliser les gammes, qui sont censés choisir (être conscients d'être sains et de devoir éviter les femelles malades). On argue souvent que ce procédé est une licence, et qu'elle est justifiée dans un article de vulgarisation. Mais l'allégorie est un des procédés rhétoriques courant dans les textes denses, également à l'œuvre dans les explications savantes. Elle est souvent associée à un parenthésage des péripéties à l'aide des réflexifs. En témoigne un texte de P. G. de Gennes, prix Nobel, publié par l'Académie des Sciences qui les conjugue tous.

35)

#### « Explosion » à la fusion

Pierre-Gilles de GENNES

**Résumé** - Un polyéthylène cristallisé en solution semi-diluée est formé de chaînes peu enchevêtrées et très compactes. Lorsque le polycristal est fondu, les chaînes gonflent et s'enchevêtront en un temps très inférieur au temps de reptation dans l'état liquide. Nous tentons d'expliquer cette « explosion » par un mécanisme où la chaîne « invente un tube » en un temps de Rouse. Le rayon de giration en début de croissance  $R_g(t)$  est indépendant de la masse moléculaire, ce qui paraît observé.

[FPA1] GENNES, P.-G. de "Explosion" à la fusion, *Comptes rendus à l'Académie des Sciences de Paris série IIb*, **321**: (1995) pp. 363-365.

Cette expression étonnante, *la chaîne « invente un tube »*, est reprise dans une légende de figure *la chaîne construit des portions de tube AA' et BB'*. Dans le texte, l'explication fait intervenir des verbes réflexifs, une phrase définitoire réflexive avec le pronom renforcé. Nous présentons la partie II qui est au cœur de la

démonstration (La plupart des équations numérotées ne sont pas reproduites).

36) II. MÉCANISME D'INTERDIGITATION - Notre modèle est décrit sur la figure. Dans le polycristal initial, la molécule étudiée occupe une petite région  $AB$  sous forme de cristal à chaînes repliées. Pendant le temps  $t$ , l'extrémité  $A$  par exemple se libère et atteint le point  $A'$ . Entre  $A$  et  $A'$ , un certain « tube » s'est alors défini ; mais il n'est pas nécessaire d'explorer tous les tubes possibles à partir de  $A$  : seulement d'en engendrer 1. Si  $p$  est le nombre de monomères qui sont impliqués dans l'arc  $AA'$ , on peut comprendre la croissance de  $p$  par un argument simple, fondé sur la diffusion des défauts (« kinks ») depuis  $A$  vers  $A'$  (de Gennes, 1971). En  $A$ , la concentration linéaire de défauts  $c$  (nombre de défauts par monomère) est 0. En  $A'$  elle est d'ordre unité. Le flux de défauts le long de la séquence chimique est du type diffusif, avec un coefficient de diffusion  $W$ :

(1)

et le nombre de monomères « fondus » par unité de temps est proportionnel à  $J$ :

(2)

Donc :

$$(3) \quad p^2 \cong W t.$$

Ceci s'applique pour  $t < \tau$ , où  $\tau$  est un temps de relaxation de Rouse :

(4)

Il en résulte un rayon de giration  $R_g(t)$  de la forme :

(5)

où  $R \sim N^{1/2}$  est le rayon final de chaîne idéale.

[Dans l'équation (5) nous avons négligé la valeur  $R_g$  du rayon dans l'état initial.]

L'exposant 1/4 n'est pas celui proposé par Barham et Sadler (1991). Mais après examen de leurs courbes, les faits suivants émergent :

a) L'exposant 1/4 donne un début de montée raide, proche de ce qui est observé.

b) En début de croissance  $R_g$  est bien indépendant de la masse moléculaire, comme le prévoit la relation (5).

c) Le temps  $\tau$  définit la croissance ; il est lui-même beaucoup plus court que le temps de reptation :

(6)

où  $N$  est la distance entre enchevêtements ( $N_e \sim 10^2$ ). Pour  $N = 10^4 \tau_{\text{rep}} / \tau \sim N / N_e \sim 10^2$ .

### III Conjectures sur les propriétés mécaniques -

Le renforcement du pronom n'est pas indispensable dans le cadre de la phrase c) prise isolément *Le temps  $\tau$  définit la croissance ; il est lui-même beaucoup plus court que le temps de reptation*. Mais cette phrase réflexive liant deux propositions signale la fin de

l'explication concernant les enchevêtrements commencée dans la partie I. C'est le moment où la synthèse doit être faite par le lecteur. Ici, *le temps*, renforcé par *lui-même* est le facteur clé dans la construction de l'explication.

Dans ce texte, le scripteur se désigne explicitement et l'objet du discours (*proposer une loi cinétique*) est présenté dans l'introduction. Quoique le texte soit très court et très dense, il reste compréhensible à travers une hiérarchie des thèmes introduisant des péripéties enchaînées. Une forme réflexive simple permet d'introduire un thème, les enchevêtrements dans la partie I. Le recours à une forme réflexive passive permet d'introduire un sous-thème spécifique, le tube, dans la partie II. C'est le point d'enchaînement le plus profond du texte. Le tube est objet pour la chaîne, sujet allégorique, abandonné au profit d'un personnel indéfini (*on peut comprendre*). La partie II se termine par une discussion référant à la partie I et une phrase réflexive définitoire. L'explication couvrant les parties I et II étant close, la seconde mutation peut avoir lieu. Le retournement est observé dans la partie 3 où les enchevêtrements ne sont plus spontanés mais provoqués par un agent humain.

37) III CONJECTURES SUR LES PROPRIÉTÉS MÉCANIQUES - Supposons, pour simplifier, que la solution initiale était assez diluée (nombre d'enchevêtrements négligeable). Alors, après un recuit sur le temps  $t$ , on a réalisé sur chaque chaîne un nombre d'enchevêtrements  $2 p(t) / N_e$ .

Dans le résumé, cette explication est condensée sous une forme allégorique. On lit que *la chaîne « invente un tube »*, un effet de style souligné par les guillemets. Pour amener le lecteur à partager le modèle, la stratégie est de lui faire adopter fugitivement la place du polycristal, qui avait le proto-rôle de patient. *Entre A et A', un certain « tube » s'est alors défini ; mais il n'est pas nécessaire d'explorer tous les tubes possibles à partir de A : seulement d'en engendrer 1.* Il n'est pas nécessaire d'explorer pour qui ? ou pour quoi ? Peu importe. On imagine que la chaîne de monomères n'a pas le temps de s'organiser et finit enchevêtrée.

Dans les textes contemporains, en vulgarisation comme dans les articles de recherche, les principes expliquant les formes réfléchies sont invariants. Ces formes bloquent les développements et signalent les points de passage entre statuts alternés des topiques, agent et patient, définissant les péripéties. Les irréfléchis permettant l'accord au sujet servent en quelque sorte de parenthèse ouvrante, les réfléchis qui bloquent les accords servent de parenthèse fermante. Les phrases réflexives anaphoriques permettent la synthèse des péripéties avant la remise en

perspective du topique principal. Ces phrases prennent différentes formes, purement réfléchie (*Ce phénomène en lui-même... , Ce clivage est en soi indicatif... Ceci n'est pas synonyme en soi du rejet...* ), ou réciproque (*Ces deux opérations vont créer, l'une et l'autre ...* )

Dans les articles journalistiques aussi, les procédés rhétoriques sont à l'œuvre. Virtuellement n'importe quel verbe peut subir une transformation réflexive, et n'importe quel sujet antécédent peut être repris par un pronom renforcé, à la condition d'être bien amené. La fusion allégorique et la condensation amènent ici une forme *se surgir* qui, hors contexte, semble bizarre voire fautive. Mais en contexte, en fin d'article, elle est justifiée et passe comme une lettre à la poste.

38) Enfin se surgirait la question de l'opportunité politique d'une telle alliance avec Airbus, le groupe européen étant perçu aux Etats- Unis, et surtout au Congrès, comme l'ennemi juré de Boeing.

(L'AGFI, 1999- 05- 12 ETATS- UNIS: AIRBUS ET LOCKHEED MARTIN  
DISCUTENT D'ÉVENTUELLES COOPÉRATIONS)

Dans d'autres langues, les mêmes principes sont à l'œuvre, avec des moyens grammaticaux variés, mais qui ont pour effet de parenthéser le discours. Les procédés utilisés en clôture bloquent la propagation des accords avec le sujet ou le thème local (Jakobson, 1963, 1973 ; Katagiri, 1991 ; Huang, 2001).

## 4. Conclusion

Contrairement à une interprétation souvent proposée, la forme réflexive n'implique ou ne nécessite aucune forme d'agentivité intrinsèque du sujet : ni les vagues ni les polycristaux ne sont doués de volonté, ni du reste de vertu miraculeuse. Bien au contraire, les formes réflexives, qui ne sont ni actives ni passives, permettent les descriptions, les observations neutres, les constats. Elles apparaissent avec des pronoms indéfinis, aussi bien qu'avec des sujets animés ou inanimés.

En discours, la description et l'explication d'un phénomène complexe (physico- chimique, historique, biologique ou autre) est difficilement possible sans le recours aux formes réflexives qui signalent une transition vers l'actif, la première mutation de l'objet de discours en sujet d'action. Ce procédé entraîne une identification provisoire du lecteur à l'objet de discours devenu agentif. Ce passage peut être nécessaire pour rendre compréhensible ou imaginable au commun des mortels ce qui relève de l'invisible. L'auteur fait ainsi partager au lecteur le lieu de l'imaginaire, un monde où s'élabore une rationalité portée par le

discours. La seconde mutation, qui remet l'objet de discours dans la perspective initiale, s'accompagne aussi par des formes réfléchies.

Ainsi, à travers le métalangage, les rhéteurs ont signalé des formes qui accompagnent le raisonnement, la réflexion, l'effort de compréhension partagé par le scripteur et son lecteur. Au-delà, on peut noter que le vocabulaire des figures de style comme la parabole, l'hyperbole et l'ellipse sont de nos jours plutôt du domaine de la géométrie que de celui de la rhétorique. Qui se hasarde à penser que ce vocabulaire a d'abord été celui de l'étude du langage avant d'être celui des mathématiques rencontre le plus souvent une profonde incrédulité.

### Références bibliographiques

- BOURDIEU, P. (1962) *Ce que parler veut dire : économie des échanges linguistiques*, Fayard.
- DOWTY, D. (1991) Thematic proto-role and argument selection, *Language* 67, pp. 547- 619.
- FUCHS, C. (1995) L'art de faire parler le langage, *Lettre SHS*, 38, pp. 34- 36.
- HERIAU, M. (1980) *Le verbe impersonnel en français moderne*, Champion et Université de Lille.
- JAKOBSON, R. (1963) *Essais de linguistique générale*. Paris, Editions de Minuit.
- JAKOBSON, R. (1973) *Questions de poétique*. Paris, Seuil.
- KATAGIRI, Y. (1991) Perspectivity and the Japanese Reflexive "zibun", in Barwise, J. et al. (Eds) *Situation theory and its applications*, Stanford, Center for the Study of Language.
- HUANG, Y. (2001) Logophoric Marking in East Asian Languages, in GÜLDEMANN, T. & von ROCADOR, M. (Eds) *Reported discourse* Amsterdam, John Benjamins.
- LARJAVAARA, M. (2000) Présence ou absence de l'objet : limites du possible en français contemporain. *Humaniora* 312. Helsinki : Academia Scientiarum Fennica.
- LARJAVAARA, M. (2002) Le morphème qui n'y est pas: cas de quelques verbes dits pronominaux, *Romansk Forum*, 16:2, pp. 587- 596.
- RIBEIRO, G. (2000) Guide de grammaire française: système verbal.
- RICŒUR, P. *Soi-même comme un autre*. Paris : Seuil ; 1990.
- YAGUELLO, M. (1988) *Catalogue des idées reçues sur la langue*, Seuil.



# Réflexivité dans un processus d'interaction homme-machine

## Exemple de résolution de charades assistée par ordinateur

Mauger Serge  
GREYC - CNRS UMR 6072  
IUT de Caen – Campus III  
rue Anton Tchekhov  
14123 Ifs  
Tél.: +33 (0)2 31 52 55 15  
Mél: mauger@iutc3.unicaen.fr

Luquet Pierre-Sylvain  
GREYC - CNRS UMR 6072  
Université de Caen, Bd. Maréchal Juin  
14032 Caen Cedex, France  
Tél.: +33 (0)2 31 56 73 98  
Mél: psluquet@info.unicaen.fr

### Résumé

Dans le cadre des jeux que l'on peut faire avec le langage, le jeu de charades, pour rudimentaire qu'il soit, est intéressant en ce qu'il permet de renouveler la question des rapports entre phonétique et phonologie au sein d'un système linguistique. Les pages qui suivent présentent un générateur de charades. Nous y décrivons un premier niveau de résolution automatique de ce type d'éénigmes en exploitant les problèmes phonologiques posés par l'automate. La réflexivité intervient ici en tant que les entités lexicales prises en compte dans le calcul s'entre-sélectionnent et s'entre-valident sur la base d'une bijection phonologique. Dans un deuxième temps, l'analyseur est utilisé comme élément d'appoint au calcul sémantique nécessaire au décodage d'une charade dont chaque élément est donné selon une définition cryptée. Nous abordons en dernier lieu la question du dialogue homme machine dans la perspective de résoudre des séquences ludiques plus ou moins équivoques.

### Mots clefs

réflexivité, charades, phonétique, phonologie, calcul sémantique, énigme, jeux, isotopie.

## 1. Introduction

L'étude des jeux d'énigmes permet de renouveler l'étude du comportement interprétatif du sujet récepteur de messages et l'étude du langage lui-même. Parmi ces jeux, les charades obligent celui qui veut les résoudre à adopter une méthode de calcul phonologique et sémantique qui, pour partie, peut s'appuyer sur un dialogue Homme-Machine. En d'autres termes, l'automate sert de support à une intelligence artificielle qui laisse au sujet humain, le cas échéant, la responsabilité de la dernière étape du parcours : celle de l'interprétation et de la validation définitives des solutions.

Cette hypothèse permet de décrire le modèle d'une machine capable de poser une série de charades simples auxquelles sont corrélées des définitions elles-mêmes plus ou moins cryptées. Le rapport entre la concaténation formelle des séquences

phonétiques, correspondant à des lexies monosyllabiques et/ou polysyllabiques, fonctionne sur le principe d'une réflexivité qui contrôle interactivement les paradigmes organisés en base de données propres à sélectionner des réponses (éventuellement multiples).

## 2. Génération automatique de charades simples

### *Charades générées selon un modèle strictement phonologique*

Le jeu de charade part d'un principe conventionnel dans les jeux de langage, selon lequel il est possible d'exploiter de manière ludique les « plis » de la langue. Ce principe se fonde ici sur un type de connexion phonétique purement formelle : les mots, « contiennent », phonétiquement, d'autres mots<sup>1</sup>. Ce point rappelle, mutatis mutandis, l'hypothèse des anagrammes sans présupposer a priori d'enjeu interprétatif clairement attesté. Il constitue le fondement lexical du jeu.

La première étape, quand on construit une charade, suppose donc de choisir un mot polysyllabique, dont on ne retient que l'aspect phonétique et qui puisse être découpé en plusieurs « mots » plus petits, qui sont comme contenus dans le mot initial. Ainsi, « dinosaure » se découpe en « dîne » + « ose » + « or » ; « aimantation » se découpe en « aime » + « an » + « tasse » + « ion ») etc.

Conventionnellement le terme résultat est appelé « le tout » ou « entier ». Les mots contenus phonétiquement dans ce tout sont appelés « premier », « deuxième » et « troisième » (ou « dernier » dans le cas où le mot est découpé en trois mots), en fonction de leur ordre d'apparition, de gauche à droite, dans le tout. Le découpage doit respecter la règle selon laquelle chacun des éléments qui correspondent aux fragments du tout doit être lui-même un mot. Ce procédé peut donc ne pas être tout à fait conforme aux coupures de syllabes décrites en phonologie. Dans le cadre du jeu, et, partant, dans celui de notre étude, un même élément lexical se présente sous trois formes : une forme phonétique, que nous appelons « phonolexème », une forme orthographique, le « grapholexème », et, au moins, une forme développée, qui correspond à une définition, le « sémiolexème ». Dans d'autres circonstances, dans le rébus par exemple, qui est la forme iconique de la charade, la même entité lexicale pourrait devenir un « pictolexème » (un dessin). Nous dirons que ces formes sont des allomorphes de la même entité lexicale, qui, selon les circonstances, est convoquée sous tel ou tel de ses avatars.

I												II		
I 11	I 12	I 13	I 21	I 22	I 23	I 31	I 32	I 33	II 1	II 2	II 3			
phon	orth	déf	phon	orth	fait	phon	orth	déf	phon	orth	déf			
din	dîne	repas du soir	oz	ose	preuve d'audace	Or	or	métal précieux	dinozOr	dinosaure	animal préhistorique			

*Exemple de décomposition du terme « dinosaure »*

Il faut imaginer que, dans le cadre d'un fonctionnement automatique, les trois

1 [Starobinski, 1971]

allomorphes d'un paradigme, par exemple I 11, I 12, I 13, pour le paradigme I 1, se déplacent en même temps pour apparaître dans une fenêtre de lecture.

Le tableau suivant donne une représentation de ce qui a fait l'objet d'une implémentation expérimentale en machine, laquelle a permis de générer automatiquement un corpus de près de cinq cent charades<sup>2</sup>. Elle intègre des définitions allomorphes non représentées ici, pour des raisons de place.

I			II
I1	I2	I3	
<b>paradigme déroulant</b> <i>use (/yz/)</i> <i>à (/a/)</i> <i>âme (/am/)</i>	<b>paradigme déroulant</b> <i>ose (/oz/)</i> <i>ré(/re/)</i> <i>scion (/sjo~/)</i>	<b>paradigme déroulant</b> <i>à (/a/)</i> <i>âme (/am/)</i> <i>art (/ar/)</i>	<b>paradigme déroulant</b> <i>réunion (/reynjo~/)</i> <i>addition (/adisjo~/)</i> <i>amoral (/amoral/)</i>
<b>art (/ar/)</b>	<b>tee (/ti/)</b>	<b>chaud (/So/)</b>	<b>artichaut(/artiSo/)</b>
<i>chaud (/So/)</i> <i>dîne (/din/)</i> <i>dit (/di/)</i> <i>il (/il/)</i> <i>ion (/jo~/)</i> <i>lit (/li/)</i>	<i>une (/yn/)</i> <i>use (/yz/)</i> <i>à (/a/)</i> <i>âme (/am/)</i> <i>art (/ar/)</i> <i>chaud (/So/)</i>	<i>dîne (/din/)</i> <i>dit (/di/)</i> <i>il (/il/)</i> <i>ion (/jo~/)</i> <i>lit (/li/)</i> <i>mot (/mo/)</i>	<i>harmonie (/armoni/)</i> <i>dinosaure (/dinoz Or/)</i> <i>illusion (/ilyzjo~/)</i>

On obtient une charade quand la série des éléments qui apparaissent dans les fenêtres est phonétiquement identique dans I et II. Les éléments I et II du système s'entre valident réciproquement (exemple ci-dessus : « artichaut »).

Bien entendu, des mots orthographiquement dissemblables peuvent être homophones. Ainsi « vers » / « verre » / « ver » / « vair » / « vert », sont inscrits dans le système sous un seul phonolexème, renvoyant à plusieurs grapholexèmes et à autant de sémiolexèmes (définitions). Les sémiolexèmes (non représentées ci-dessus) enrichissent éventuellement le jeu en ajoutant une dimension sémantique à l'exercice mais ne jouent qu'un rôle secondaire à ce niveau. Cependant, dans la perspective d'un développement complet du système, ils seront fondamentaux en ce qu'ils constituent la manière conventionnelle de proposer une charade à quelqu'un. Par ailleurs, un même phonolexème et son avatar orthographique peuvent faire l'objet d'une série de définitions ordonnées par niveau de difficultés du point de vue de leur interprétation. Cet aspect du jeu correspond à des degrés de cryptage sémantique. Le calcul interprétatif de ces énigmes fait l'objet d'un programme additionnel dont nous avons présenté ailleurs le modèle<sup>3</sup>.

### *Exploitation du modèle implémenté et comparaison avec les charades déjà existantes*

L'algorithme qui a été exploité pour ce travail<sup>4</sup> génère des charades ayant comme caractéristique le strict respect de la phonologie du français: la réalisation du « tout »

2 [Luquet, 2001]

3 [Mauger 2003] Il s'agit là d'un calcul assisté par ordinateur de chaînes sémiques propres à chaque isotopie, dans le cadre d'une sémantique interprétative.

4 [Luquet 2001]

est le résultat de la concaténation de la forme phonologique des trois parties sans qu'aucun écart ne soit admis. A l'issue de cette expérience, un corpus de quelque cinq cents charades a été généré. Dans la démarche expérimentale qui caractérise nos travaux, il nous a paru nécessaire de comparer notre corpus avec une un second corpus constitué de charades "existantes", c'est-à-dire déjà en usage dans la communauté parlante. On peut trouver des charades sur le Web, quelques unes dans des magazines de jeux, et certains recueils vendus dans le commerce y sont exclusivement consacrés. Le livre *Mille et une charades*, d'Anatole Sliosberg<sup>5</sup>, a constitué un corpus suffisant pour les besoins de notre étude. Il est composé de charades écrites par quelques auteurs de renom tels que Voltaire, V. Hugo ou encore L. Etienne, mais la plupart d'entre elles proviennent simplement d'auteurs inconnus. La comparaison de ces deux corpus nous a permis de détecter des écarts phonétiques entre le jeu pratiqué dans le grand public et les résultats strictement phonologiques produits par la machine. Deux cents charades extraites du corpus ont été vérifiées. On observe dans les charades du public une dispersion phonétique telle que, par exemple, un [e] (fermé) et un [E] (ouvert) seront, jusqu'à un certain point, considérés comme semblables, voire interchangeables, du point de vue de leur valeur calculatoire. Même observation à propos d'un [o] fermé par rapport à un [O] ouvert. Ainsi, dans une charade où l'entier à découvrir est «botanique» on peut partir aussi bien d'une concaténation de type «beau+tanne+hic», que de type «botte+âne+hic». Pour l'entier «hypothèse» on découpe «hi + pote + aise» ou (entre autres) «hi + peau + thèse».

La plupart des écarts rencontrés concernent des rapprochement entre voyelles. A lui seul ce type d'écart représente plus de 80% de l'ensemble de ceux que nous avons recensés et qui se répartissent de la manière suivante :

Écart	Exemple	Nature (caractéristique <sup>6</sup> )	%	% cumulés
E -> e	Laid -> <b>léopard</b>	ouverture	47	47
o -> O	Faune -> <b>téléphone</b>	ouverture	12.5	59.5
2 -> 9	Jeûne -> <b>jeunesse</b>	ouverture	9	68.5
A -> a	Hâte -> amateur	ouverture	15	83.5

En arrière plan de ces exemples se trouve donc posée une question plus générale qui est celle de la laxité phonétique maximale autorisée dans le cadre du jeu. Ainsi la différence d'ouverture entre [e] et [E], habituellement considérée comme un écart d'un degré, permet l'assimilation (toujours dans le cadre du jeu) entre «et» et «haie», ou entre «marée» et «marais» alors que la différence d'ouverture entre [i] et [e], elle aussi en principe d'un degré, ne permet pas l'assimilation entre «marie» et «marée». On observe donc, dans la topique phonologique, une différence de qualité entre la coupure qui distingue d'une part [e] de [E] et celle qui sépare [e] de [i].

Les autres écarts que nous avons relevés dans le recueil de Sliosberg, ceux qui n'entrent pas dans les 80 % mentionnés ci-dessus, concernent essentiellement les

5 [Sliosberg 1989]

6 La caractéristique évoquée ici est l'un des sept attributs du phonème décrits par Saussure repris et exploités depuis par [Coursil 1992].

consonnes. La continuation de l'étude du corpus de référence aura, entre autres, pour tâche de vérifier s'il existe des rapprochements possibles entre consonnes identiques à une caractéristique près (par exemple celle de la nasalité) ; de confirmer les éléments de conclusions évoqués précédemment en ce qui concerne les voyelles ; de s'assurer du rôle des caractéristiques de voisement, d'arrondi, d'antériorité et de nasalité.

Une fois l'étude complète du corpus réalisée, nous pourrons passer à la validation des règles mises à jour par une nouvelle expérimentation. Celle-ci consistera à générer un nouveau corpus de charades, en définissant de nouveaux opérateurs sur les chaînes phonétiques et les phonèmes eux-mêmes. Les opérateurs implémentés dériveront de ceux que nous avons décrits ailleurs<sup>7</sup>, mais intégreront les approximations phonétiques que nous auront mises en lumière.

### 3. La charade comme énigme phonologique et sémique.

On ne peut limiter l'étude de la charade à cet unique point. D'abord parce qu'elle est construite comme nous l'avons dit sur plusieurs allomorphes du lexème, ensuite parce qu'il existe certains degrés, ou niveaux de force (par analogie aux mots croisés), dans la pratique de ce jeu. Une charade pourra par exemple présenter des traits humoristiques ou des définitions obscures, qui constituent autant de cryptage et posent des contraintes supplémentaires quant au choix des lexèmes formant la réponse. La charade suivante, attribuée à Victor Hugo en est un exemple :

*Je prends mon premier,  
au coin de mon second,  
en sortant de mon tout.*

#### *Utilisation de l'automate*

En dépit de l'obscurité relative de la formulation, on sait, ou on pose a priori, que cette charade se structure selon une forme « canonique » du point de vue des règles de concaténation phonétique. Elle correspond à la construction que nous avons décrite dans notre première partie et s'actualise sous le modèle initial :

I		II
I 1	I 2	
/x/	/y/	/z/ = /x + y/ (chaînage phonétique)

L'automate, en tant qu'il intègre idéalement tous les monosyllabes du lexique (dans les paradigme I 1 et I 2), tous les termes dissyllabiques (en II) et en tant qu'il ne travaille que sur le principe d'une règle d'addition phonétique, calculera nécessairement la solution. Celle-ci aura pour forme une chaîne phonétique de type : “ monosyllabe + monosyllabe = dissyllabe ”.

<sup>7</sup> [Luquet 2001]

Exemples de chaînes :

I		II
I 1	I 2	
/re/ (ré)	/po~s/ (ponce)	/repo~s/ (réponse)
/de/ (dé)	/mo~ / (mont)	/demo~ / (démon)
/mar/ (mare)	/Sa~/ (chant)	/marSa~/ (marchand)

Sur la base du seul calcul phonétique, la machine peut donc tourner jusqu'à épuisement des stocks de couples de monosyllabes retenus en I, corrélés à des dissyllabes en II. Elle produit nécessairement la chaîne cohérente, mais n'est pas en mesure de l'identifier comme telle.

Précédemment, (en première partie), la série des additions phoniques était indépendante des définitions qui se trouvaient attachées aux items de chaque paradigme. Pour le sujet, en cas de doute, la validation du résultat se faisait par la reconnaissance du terme polysyllabique et de sa définition. Il est clair que dans le cas présent, c'est à dire en fonction du libellé obscur de la charade, la validation de la réponse ne peut être faite en se limitant à cette seule concaténation phonique. Ainsi donc, si nous pouvons nous appuyer sur la certitude que la chaîne valide se trouve nécessairement proposée par l'automate tel qu'il est construit dans sa forme initiale, cette certitude se double tout aussi nécessairement d'une autre selon laquelle il faudrait spécifier à la machine un principe de sélection ou d'identification de la réponse correcte.

#### *Description des règles grammaticales à considérer dans le calcul.*

Du point de vue syntaxico-sémantique, la séquence, telle qu'elle se présente initialement a pour structure un système de règles de la langue qui s'analyse à deux niveaux.

Un premier niveau, intéresse une à une les chaînes que constituent les trois sous-séquences : *Je prends mon premier / au coin de mon second / en sortant de mon tout.* Un second niveau relève d'une cohérence globale de l'ensemble de la proposition : "*Je prends mon premier au coin de mon second en sortant de mon tout*"

#### *Description des règles de premier niveau.*

Le premier niveau d'analyse syntaxico-sémantique concerne les couples d'éléments formés à partir deux ensembles présentés ci-dessous. Le premier de ces deux ensembles regroupe un triptyque constitué de : "Je prends", "au coin de", "quand je sors de". On désigne ce triptyque par le symbole "A". Les éléments qui composent "A" seront désignés, pour leur part, par les symboles "a", "b", "c".

On peut représenter cette convention par le tableau suivant :

A		
a	b	c
je prends	au coin de	quand je sors de

Chacun de ces éléments est l'incipit d'une proposition dont la fin, se trouve

représentée par les termes “mon premier”, “mon second”, “mon tout”. Nous avons là un second triptyque qui sera désigné par le symbole “B” ; les éléments qui le composent seront désignés quant à eux par les symboles “1”, “2”, “3”. Soit le tableau suivant :

B		
1	2	3
mon premier	mon second	mon tout

Les éléments de chacun des deux groupes s’articulent grammaticalement et sémantiquement deux à deux, ainsi que cela apparaît dans la formulation même de la charade. On représentera cette articulation par le modèle :

a est en relation avec 1	que l'on note par la règle	<b>a R' 1</b>
b est en relation avec 2	que l'on note par la règle	<b>b R'' 2</b>
c est en relation avec 3	que l'on note par la règle	<b>c R''' 3</b>

Les notations “**a R' 1**”, “**b R'' 2**”, “**c R''' 3**”, sont symboliques et sténographiques et représentent les règles syntactico-sémantiques suivantes :

- **a R' 1** signifie que “**a**” est suivi d’un COD, représenté sous la forme codée “**1**”, et que ce COD appartient au paradigme sémantique des “objets” qui peuvent être saisis<sup>8</sup>.

- **b R'' 2** signifie que “**b**” sera suivi d’un locatif représenté sous la forme codée “**2**” et que la forme claire de ce “second” sera actualisée sous la forme d’un terme propre à désigner un lieu.

- **c R''' 3** signifie que “**c**” sera suivi d’un autre locatif représenté sous la forme codée “**3**” et que la reformulation claire de ce “tout” sera actualisée par un terme propre à désigner un lieu d’où l’on peut sortir.

Les règles, que nous avons appelées les “règles syntactico-sémantiques de premier niveau”, jouent dans le calcul de la charade, le rôle de définitions. En d’autres termes, elles servent à présélectionner pour le sujet, la catégorie grammaticale du (ou des) terme(s) qui pourrai(en)t être validé(s) comme réponse(s) au problème. On peut donc réfléchir à la manière d’améliorer l’automate de façon à ce qu’il ne sélectionne dans les paradigmes I 1, I 2 et II, que des termes répondant aux exigences syntactico-sémantiques conformément à “**a R' 1**”, “**b R'' 2**”, “**c R''' 3**”.

Le complément d’un automate analyseur / générateur de chaînes lexico-phonétiques serait alors, pour l’ensemble des items recherchés un module décrit par le tableau suivant :

I				II	
a R' 1 (je prends + C o d)		a R'' 2 (au coin de + locatif)		a R''' 3 (en sortant de + locatif)	
I 11	I 12	I 21	I 22	II 1	II 2
phon	orth	phon	orth	phon	orth
?	?	?	?	?	?

<sup>8</sup> Paradigme fort étendu, on s’en doute. On peut saisir une chose aussi bien qu’un être vivant ou une idée.

Sur cette base, et en tenant compte des règles d'addition phonétique, l'automate pourrait donc proposer des exemples du type:

I			II		
I 1 1	I 1 2		II 1	II 2	
phon	orth	phon	orth	phon	orth
/ li /	lit	/do/	dos	/lido/	lido

Cette séquence : *Je prends mon lit au coin de mon dos en sortant du lido*, est acceptable “dans une certaine mesure”. C'est-à-dire pour peu qu'on y ajoute un contexte / contexte de type surréaliste ou une explication acrobatique dans un hypothétique “monde possible”. Ceci constitue un exemple de résultat à vocation heuristique mais ne saurait être considéré comme satisfaisant dans le cadre du jeu. Ce qui veut dire que la communauté parlante exigera en réalité une solution qui réponde à deux critères.

- Un premier critère qu'on appellera “critère de congruence encyclopédique”, qui suppose que la solution doit se trouver en adéquation avec un modèle de description du monde ne nécessitant aucune explication particulière, parce qu'il sera conforme aux connaissances triviales. C'est là une forme implicite de boucle réflexive, propre à la langue qui, dans le même temps qu'elle exprime un message, informe de façon métalinguistique sur ce même message<sup>9</sup>.
- Un second critère, propre au jeu, veut que la réponse soit unique ou que l'on ne puisse, le cas échéant, trouver qu'un nombre très restreint de réponses.

Or, en l'état actuel de la configuration de l'automate, c'est à dire en admettant qu'on puisse lui ajouter un programme de sélection fondé sur les règles “**a R'1**”, “**b R''2**”, “**c R'''3**”, nous pouvons produire à nouveau un nombre lui aussi infini de phrases semblables à « lit » + « dos » = « lido ».

#### *Description de la chaîne solution*

Parmi toutes les concaténations qui permettent de construire les binômes de type « Lit » + « dos » = « lido », l'automate produit nécessairement la chaîne lexico-phonétique suivante :

I		II	
a R' 1		a R'' 2	
I 1 1	I 1 2		a R''' 3
phon	orth	phon	orth
/te/	thé	/atr/	âtre
			/teatr/
			théâtre

Cette chaîne correspond à la phrase : *Je prends mon thé au coin de l'âtre, en sortant du théâtre* qui a les mêmes caractéristiques phonétiques et syntactico-sémantiques que la précédente. Mais elle a la particularité, pour le solveur de charades d'être identifiée comme “la” réponse au calcul, ou, comme l'une des réponses absolument valides. En plus de respecter les règles de premier niveau, elle trouve sa véritable

<sup>9</sup> Voir entre autres les rappels que fait Anne Nicolle dans ce recueil: « Les langues humaines usuelles n'ont pas besoin d'un métalangage différent pour parler de leur propre fonctionnement ». Nous considérons pour notre part qu'un message est à la fois une information et de manière sous-jacente, un métalangage qui parle réflexivement de lui-même, en ce sens qu'on le considère comme grammaticalement bien formé.

cohérence en prenant en compte une règle de “second niveau” qui intègre les relations sémantiques de l’ensemble de la chaîne.

#### *Règle de second niveau et limites du calcul automatique*

La règle que nous appelons de second niveau est une règle sémantique particulière. Elle s’actualise dans le fait que les trois termes fournis automatiquement par le calcul phonétique se trouvent congruents non seulement par rapport aux couples coordonnés par les règles (a R’1), (b R’’2), (c R’’’3), mais aussi par rapport à l’ensemble de la chaîne constituée par la formulation de la charade. Une analyse sémantique simple, suffit à montrer que si nous estimons avoir trouvé la réponse -si nous estimons pouvoir valider la chaîne «thé» + «âtre» = «théâtre»- cela signifie qu’il est parfaitement cohérent de dire que l’on «*prend son thé au coin de son âtre (feu), quand on rentre chez soi et, pourquoi pas, quand on sort du théâtre*». Cette cohérence correspond à une homogénéité sémantique de la chaîne, que l’on appelle habituellement une isotopie.

L’isotopie est le principe organisateur qui fonde l’interprétation de la réponse. Cette interprétation s’appuie sur la prise en compte d’un continuum de valeurs sémiques entre tous les éléments de la phrase et non plus seulement entre les termes associés par couple. On peut formuler cette règle de cohérence isotopique (symbolisée RI) de la manière suivante : **RI ((a R’1) (b R’’2) (c R’’’3))** qui signifie que tous les couples de relation (a/1, b/2, c/3) sont non seulement soumis aux règles syntaxico-sémantiques de premier niveau (R’, R’’, R’’’), mais encore à une règle de cohérence isotopique. Cette cohérence est le résultat d’un processus qui permet d’intégrer tous les éléments de la chaîne-réponse dans un continuum interprétatif acceptable du point de vue des références encyclopédiques.

Les trois termes forment ainsi un “champ sémantique” qui les rassemble sous une même notion. Du point de vue d’une analyse componentielle on pourra assez facilement montrer quelle récurrence sémique existe entre “thé” et “âtre”. Admettons qu’il s’agisse d’un sème que l’on désignera par le concept de “réconfort” ou “plaisir”. On peut éventuellement admettre qu’il existe un rapport entre ces deux termes et le mot “théâtre”. Mais il est, cette fois, bien plus délicat de proposer d’en rendre compte par un sème commun.

Dans l’impossibilité même de nommer cet hypothétique sème commun à tous les éléments de la chaîne, il devient plus délicat encore de le considérer comme “ancré” à une (ou des) entité(s) lexicale(s) précise(s). Ducrot et Todorov<sup>10</sup> expliquent que “*dans la mesure où l’analyse sémique porte seulement sur des éléments du lexique [...] qu’elle représente comme des “paquets de sèmes”, elle ne suffit pas à assurer l’originalité de la combinatoire sémantique. [...] Pour que l’analyse sémique implique le caractère irréductible de la combinatoire sémantique, il faut qu’elle porte non seulement sur le contenu d’unités lexicales, mais, [...] sur le contenu de segment d’énoncés plus larges, voire sur des effets de sens (appelés sémèmes), c’est-à-dire sur des significations liées à un certain contexte ou à une certaine situation du discours. Les sèmes n’étant plus alors reliés à des mots ou à des morphèmes, les*

10 [Ducrot, Todorov, 1979]

*relations qui les unissent ne peuvent plus être parallèles aux relations syntaxiques.”* Or, dans la réponse que nous avons validée (théâtre), c'est bien l'expérience du continuum sémantique et de la cohérence de la proposition formulée à partir de “*thé + âtre*” qui nous sert de pierre de touche. Nous validons la reformulation de la charade sur la base d'une interprétation que nous jugeons satisfaisante. Cette interprétation sous-entend la prise en compte de la règle d'isotopie “*RI*”.

La règle de continuum (*RI*), qui permet de valider la chaîne, porte sur l'ensemble de l'énoncé, mais ne peut se décrire ni syntaxiquement, ni dans le cadre d'une analyse componentielle. On peut cependant en donner une représentation en considérant les trois termes « *thé* », « *âtre* » et « *théâtre* » comme catégorisés sémantiquement, par leur co-présence dans la proposition.

De fait, il est peu concevable d'implémenter *RI* dans le programme de la machine pour en faire une règle de sélection automatique de la réponse. L'isotopie constitue un continuum du point de vue de l'interprétation mais il est impossible de la situer dans les mots ou les lexèmes. Ceci doit tomber sous le sens si l'on admet qu'un continuum, par définition, n'est pas « *discret* ».

Sans entrer dans une démonstration de détail, il est clair que, à quelques nuances près, on aboutit à une situation semblable avec une charade comme :

*Mon premier  
coupe mon entier  
avec mon dernier .*

Le solveur d'énigme valide une séquence, parmi « *n* » possibilités concaténées par la machine. Mais il le fait d'un certain point de vue, c'est-à-dire, dans le cadre d'une « *lecture* » et d'une interprétation, en fonction de sa capacité à rassembler de façon cohérente ce que la machine produit par un calcul qui ne dépasse pas la simple corrélation de traces. Et ce, quand bien même les « *traces* » font l'objet d'une première sélection sur la base des règles grammaticales de premier niveau.

### *Rôles complémentaires de l'humain et de l'automate*

Ainsi que nous l'avons dit, le continuum sémique procède de valeurs construites dans le cadre de l'interprétation, autrement dit par le lecteur et, pour notre exemple, par le solveur d'énigmes .

A la suite d'un test, administré empiriquement auprès d'une trentaine de personnes, il s'est avéré qu'aucun des individus auxquels a été proposée la charade de V. Hugo n'a pu la résoudre, pas même après un assez long délai. En réalité, quelques uns seulement ont pu approcher la forme générale de la réponse, en proposant des approximations heuristiques semblables à « */li/ + /do/ = /lido/* ». En revanche, grâce à l'assistance de l'automate, on acceptera comme cohérentes des réponse comme: « *je prends mon vin au coin de la scène en sortant de Vincennes* » ou comme « *Je prends mon lait au coin de la table en sortant de l'étable* ».

Un automate correctement programmé est donc parfaitement, et même

nécessairement, capable de produire la chaîne « thé » + « âtre » = « théâtre » ou « chien » + « dent » = « chiendent ». Mais ces réponses n'ont aucune possibilité d'être validées plutôt que d'autres par le système automatique lui-même. Le partage des compétences entre l'homme et la machine est ici en relation de complémentarité. Il semble qu'on puisse faire à ce propos, deux remarques :

- L'ordinateur va plus vite mais ne comprend pas . Il gère des traces et se trouve dans la même situation que l'Anglais dans la chambre chinoise de Searle<sup>11</sup> : ce sont ceux qui sont à l'extérieur de la chambre (pour nous, ceux qui lisent sur l'écran de l'ordinateur) qui comprennent.
- L'humain valide (et donc comprend) le résultat mais est presque dans l'impossibilité de le calculer seul.

On voit ici l'automate prendre le rôle d'un adjvant sans doute indispensable à l'interprétation de ce message « obscur », dans la mesure où la réponse calculée automatiquement est proposée, parmi d'autres, à la vitesse de travail de l'ordinateur. C'est-à-dire incomparablement plus rapidement que par le « calcul » du sujet humain.

### *Le sujet intègre un analogue de l' automate*

Il faut bien distinguer, dans notre propos, ce qu'on appelle « l'individu » par rapport au « sujet ».

« L'individu » est une personne physique : Vous, moi, Pierre, ... Le « sujet » est le concept de l'analyseur de charade et plus largement de l'interprétant humain. L'incapacité de l'humain à calculer la chaîne-réponse est une incapacité réelle d'un individu en situation (c'est-à-dire quand on lui pose une charade de ce type). Du côté du « sujet »; cette incapacité n'existe pas. On veut dire par là que la « réponse », la « chaîne satisfaisante », existe comme élément de combinaison du système linguistique et que le sujet, en raison de sa compétence de locuteur, l'intègre virtuellement. Dire que le sujet est théoriquement capable de résoudre ce type de problème, c'est dire qu'il est, tout aussi théoriquement, capable à la fois d'un calcul de type phonétique et de l'évaluation réflexive du résultat sur la base d'une règle de type RI ((aR'1) (bR''2) (cR'''3)). Le sujet possède donc, en théorie, à la fois la compétence de l'automate et une compétence supplémentaire.

## **4. Conclusion : automate et compétence interprétative du sujet.**

La première de ces compétences est celle qui nous permet, théoriquement, d'une manière analogue à l'automate, de procéder à un calcul lexico-phonétique de toutes les chaînes constituant des charades. Cette compétence se trouve, dans la pratique, la plupart du temps grippée. C'est même sur ce grippage que l'on compte, quand on pose une charade complexe à quelqu'un, pour avoir le plaisir de lui donner la solution.

La seconde compétence en revanche, correspond à la capacité d'identifier la cohérence de la chaîne-réponse, c'est-à-dire de comprendre cette chaîne comme « texte », et de l'interpréter selon un principe de réflexivité isotopique. Cette

---

11 [Searle, 1980]

compétence n'est jamais longtemps prise en défaut. C'est elle qui est sollicitée quand on valide la séquence théâtre. Dans le pire des cas, il sera nécessaire d'expliquer à l'individu l'astuce qui lui a échappé pour qu'il réintègre sa dimension de sujet. Il ne comprendra qu'après coup, mais sauf pathologie particulière, il comprendra, alors même qu'il se trouve dans l'impossibilité du calcul phonétique initial. En d'autres termes : ce n'est pas l'automate qui est analogue du sujet mais, inversement tout sujet qui est par un certain côté un analogue de l'automate. En revanche aucun automate n'est un sujet.

Certes, l'individu en tant qu'automate « marche mal » et ne peut maîtriser l'ensemble des possibilités du système de sa propre langue. Mais, tout handicapé qu'il soit par sa lenteur de calcul, il est le seul à maîtriser l'interprétation des chaînes de signes.

## Références

- COURSIL, 1992, *Grammaire analytique du français contemporain*, thèse de doctorat, Université de Caen – Basse-Normandie
- DUCROT, TODOROV, 1979, Combinatoire sémantique, in *Dictionnaire encyclopédique des sciences du langage*, Seuil, Paris.
- ETIENNE, 1965, *L'art de la charade à tiroirs*, Pauvert, Paris.
- FUCHS ET AL., 1993, *Linguistique et traitement automatique des langues*, Hachette, Paris.
- JAKOBSON, 1973, *La charpente phonique du langage*, Seuil (réédition 1980), Paris.
- LUQUET, 2001, *Étude et modélisation du système phonologique de la langue française*, mémoire de DEA, Université de Caen – Basse-Normandie
- MAUGER, 2003, *Variation, construction et instrumentation du sens*, Réduction automatique des définitions cryptées, essai de sémantique combinatoire. Hermès Sciences. Paris.
- NICOLLE, BEUST, 1997, « ANADIA », un analogue machine de la mémoire, *cahier du GREYC*, Université de Caen – Basse-Normandie
- NICOLLE, 2005, Comparaison entre les comportements réflexifs du langage humain et la réflexivité du langage informatique (ici même).
- RASTIER, 1987, *Sémantique interprétative*, coll. Formes sémiotiques, PUF, Paris
- SEARLE, 1980, Minds, Brains and Programs, in *Behavioral and Brain Sciences*, #3.
- SEARLE, 1985, *Du cerveau au savoir*, Hermann, Paris.
- SEARLE, 1992, *Déconstruction, le langage dans tous ses états*, trad. J. P. Cometti, Combas, “L'Eclat”, Paris.
- SLIOSBERG, 1989, *Mille et une charades*, Nigel Gauvin, Paris.
- STAROBINSKI, 1971, *Les mots sous les mots*, Gallimard, Paris.

# Comparaison entre les comportements réflexifs du langage humain et la réflexivité des langages informatiques

Anne Nicolle

GREYC UMR 6072 CNRS (équipe ISLanD)  
Université de Caen

## Résumé :

Le langage humain est réflexif en ce sens qu'il n'a pas besoin d'un autre langage pour parler de son propre fonctionnement. Cette caractéristique du langage a été analysée en particulier par Hjelmslev et cette analyse est encore pertinente aujourd'hui. Elle est intrinsèquement liée à la multiplicité des interprétations possibles d'un énoncé. Tous les autres langages et toutes les notations mathématiques, logiques, musicales sont construites à partir du langage écrit. Chacun d'eux forme donc un langage dans le langage. Les interprétations sont univoques et ils ne sont généralement pas capables de parler de leur propre fonctionnement. Ils font appel pour cela au langage humain dans sa généralité. La formalisation du langage trouve ses limites dans le processus de création des notations, et nous en présenterons quelques conséquences. Certains langages informatiques ont des capacités réflexives : lisp et plus récemment Java ou XML. Ils perdent en même temps leur univocité d'interprétation. Cette nécessité de l'ambiguïté a été prise en compte dans le métamodèle UML. Les capacités réflexives des langages informatiques sont comparées avec les capacités réflexives du langage humain. La dynamité des langages informatiques et leur réflexivité sont des moyens de lever certaines difficultés de la formalisation du langage.

**Mots-clés :** réflexivité, TAL, Java, XML, UML

## 1 Introduction

La réflexivité étudiée dans cet article est une propriété de textes, textes en langue ou programmes informatiques. On fera abstraction de la réflexivité physique des miroirs, même si elle en est peut-être la source. On ne fera pas non plus référence à la réflexivité comme pratique humaine d'observation de soi et de prise de recul par rapport à ses propres travaux scientifiques. Hofschtader a montré des exemples de réflexivité en musique, comme les canons qui se lisent dans les deux sens ou comme la double interprétation musicale/langagière avec la notation alphabétique des notes. Il a montré des exemples dans les arts graphiques, comme chez Escher, qui mettent en évidence le point aveugle où la réflexivité prend sa source et trouve ses limites. Il a montré que la réflexivité des formules mathématiques était la source de l'indécidabilité et il a admirablement expliqué son fonctionnement : les formules indécidables exhibées par Gödel ont une double interprétation, à la même expression correspondent deux

significations, chacune d'elles a une valeur de vérité, donc l'expression en tant que telle porte deux interprétations contradictoires en terme de vérité. Si comme Peirce, on fait intervenir le troisième terme du signe, à savoir l'interprétant, il n'y a plus d'indécidabilité. Je ne reviendrai pas sur ces aspects, qui forment la toile de fond de cet article. Après avoir rappelé les caractéristiques des langues humaines qui permettent de s'en servir pour parler d'elles-mêmes, je présenterai quelques exemples de langages et de notations réflexives dans le domaine de l'informatique. Une discussion sur l'interprétant machine, comme introduisant le temps dans le texte, alors que les formules mathématiques en font abstraction, permettra de conclure.

## 2 La réflexivité du langage humain

Les langues humaines usuelles, dites langues naturelles, permettent d'exprimer tout ce qui est exprimable dans les autres langages et n'ont pas besoin d'un métalangage différent pour parler de leur propre fonctionnement : les grammaires, les dictionnaires, les traités de rhétorique ou de stylistique en sont la preuve. Certains énoncés, dits autoréférents, peuvent parler d'eux mêmes : « cette phrase a cinq mots. ». Les langages restreints comme les formules mathématiques ou les langages logiques n'ont pas cette propriété : ils sont définis par un texte écrit dans une langue naturelle et tant pour les définir qu'en cas de problème d'interprétation, la discussion a lieu dans une langue naturelle. Vouloir fonder les mathématiques sur les mathématiques mène à l'indécidabilité.

Nous rappelons et commentons quatre traits de la structure fondamentale des langues, inspirés librement de [Hjelmslev, 47]. Ces traits explicitent les conditions de possibilité de la réflexivité des langues.

1- Les langues humaines transmettent des significations et il faut distinguer dans tout énoncé le point de vue de l'expression et le point de vue du sens<sup>1</sup> (ou signifiant/signifié pour Saussure, ou signe/objet pour Peirce [Morand, 2004]). Peirce a une autre analyse, qui met en évidence un troisième terme, l'interprétant pour qui le signe fait sens. Il insiste aussi sur le fait que le signifié est lui-même un signe, alors que pour Saussure c'est une image mentale. Comme pour parler du sens d'une expression, on utilise bien sûr une autre notation, ou une paraphrase, ou une traduction, le point de vue de Pierce sur l'objet du signe comme signe lui-même, en ce qu'il est partageable dans une communauté linguistique est précieux pour comprendre que le sens n'est jamais capturé par ses représentations.

Expression	8	8	8	huit	huit	1000
Sens			huit	card[     ]	7+1	eighth

---

<sup>1</sup> Dans la traduction française de Hjelmslev, le traducteur utilise le terme « contenu », qui est malheureux en français car le contenu est détachable du contenant (on peut boire un verre d'eau, il n'y a plus d'eau dans le verre après alors que le sens est toujours lié à l'expression. On ne peut pas l'en extraire).

L'expression renvoie au sens via une convention. Dans le dernier exemple, 1000 renvoie à huit en tant qu'il est codé en binaire. Le courant philosophique qui cherche à réduire l'énoncé à son expression, ou à définir des expressions logiques qui seraient le sens des expressions en langue, a pour origine le Cercle de Vienne et la question des fondations des mathématiques. Il est encore vivant en informatique dans la problématique de la formalisation. Cette question de la formalisation est souvent vue comme la réduction de l'énoncé à son expression, ou à une expression en logique, qui seraient les seules notions accessibles aux machines informatiques. Il n'en est rien et les trois termes du signe sont nécessaires pour comprendre ce qui se passe dans les relations des usagers de l'informatique avec les phénomènes que les machines produisent.

2- Les langues humaines s'analysent selon deux axes, un axe du procès, spatial ou temporel où la parole se déroule, un axe du système de la langue ou axe paradigmique qui définit les possibles à chaque position. Seul l'axe du procès est concret et observable, l'axe de la langue provient de l'analyse du phénomène et est construit pour en rendre compte. Il définit des catégories syntaxiques (article, nom, pronom, verbe...), et à l'intérieur de ces catégories des classes de termes substituables dans le contexte de la phrase ou du texte. Sur l'axe du procès, les énoncés sont soumis à l'ordre des positions et déplacer des éléments donne soit un énoncé incorrect, soit un sens différent : « le chat mange la souris », « le mange souris chat la », « la souris mange le chat »... De manière plus fine, on peut différencier les deux énoncés oraux ayant presque le même sens : « le chat, la souris, il la mange ! » et « la souris, le chat, il la mange ! ». Il y a même verbe, même acteur, même patient, mais la focalisation est différente et l'enchaînement conversationnel n'est pas possible de la même façon [Luzzati, 2000].

3- La relation signifiant/signifié n'est pas une relation de dénotation. Dans un inventaire, la relation entre la forme et le sens est une relation de dénotation. Par exemple, les nombres cardinaux forment un inventaire infini dont les expressions tant en langue qu'en notation arithmétique sont une dénotation. Dans les langues la relation entre l'expression et le contenu d'un signe n'est pas une pure relation de dénotation, chaque signe renvoie à un espace des significations différent suivant l'existence des autres signes qui prennent place autour.

« Combien as tu de bonbons ? »

« Beaucoup ! »

On peut, suivant les centres d'intérêt ou les métiers, savoir nommer les 7 couleurs de l'arc en ciel ou les 5000 couleurs de la palette d'un peintre. Les mêmes mots n'auront pas la même extension de signification dans les deux cas. De même, les modes et temps des verbes ne sont pas les mêmes d'une langue à l'autre. La création ou l'emprunt d'un nouveau mot, tout comme sa disparition, changent le sens des autres mots. La relation de dénotation ne permet pas de rendre compte des choix à faire pour la traduction. Le sens d'un mot est conditionné par les relations entre les mots commutables. Dans un énoncé, il dépend des autres mots de l'énoncé et des autres énoncés du texte, par formation d'isotopies. Il dépend même d'un intertexte [Kanellos, 99].

4- Les relations de réction entre les unités linguistiques d'un énoncé permettent ou interdisent leur combinaison. Elles construisent la structure de l'énoncé en termes

d'unités régissant d'autres unités (l'existence d'un adjectif implique l'existence d'un nom qu'il qualifie). C'est cette propriété des langues qui a conduit Chomsky à définir des grammaires génératives sur la base de règles de combinaison entre les unités pour former des arbres de syntaxe [Chomsky, 57]. Elle est considérée alors comme plus fondamentale que les autres, comme le cadre dans lequel les autres propriétés se placent. Les limites atteintes par le développement de cette théorie ont montré qu'il n'y avait pas de prééminence d'une des quatre propriétés sur les autres. Leur donner la même importance s'oppose à une vision des langues dirigée par la syntaxe. La conception d'analyseurs syntaxiques légers va ainsi dans le sens d'une meilleure prise en compte de la place relative de la syntaxe dans l'analyse des textes [Vergne & Giguet, 98].

Ces quatre propriétés construisent ensemble les compétences réflexives des langues. En restant au niveau de l'expression, on peut compter le nombre de mots d'une phrase, mais peut-on produire la phrase autoréférente qui en rend compte ? « cette super longue phrase a 7 mots. ». La création de cette phrase ne précède pas le compte des mots, elle en est concomitante. Construire des phrases autoréférentes est un jeu. La réflexivité vient de ce que les énoncés ou les textes, prononcés ou écrits, sont des objets du monde au même titre que les objets physiques. Ce sont des objets symboliques qui peuvent être enregistrés, manipulés, reformulés, décrits, commentés, déconstruits. Les descriptions, reformulations, commentaires peuvent à nouveau être pris comme des objets qui seront décrits, commentés, reformulés dans un processus sans fin.

### 3 La réflexivité des langages informatiques

L'enregistrement du programme en mémoire a transformé les machines à calculer en ordinateurs. Auparavant, il était écrit sur des bandes perforées. De ce fait, le programme et les données sont de même nature et un programme peut prendre comme données un autre programme. Ainsi, on peut afficher ou imprimer un programme, le compiler, le modifier. Ces opérations sont faites par des programmes qui peuvent se prendre comme objet : on imprime le programme qui imprime les programmes. La réflexivité est donc au cœur des processus informatiques comme elle est au cœur des langues naturelles.

La réflexivité, c'est la possibilité de s'observer soi-même, de rendre compte de son propre fonctionnement. Lisp est un langage réflexif puisque l'interprète Lisp est écrit en Lisp et s'interprète lui-même. Certaines fonctions (funcall, apply) peuvent être modifiées au cours de l'exécution d'un programme de manière à modifier l'exécution de l'interprète. Entre les années 60 et les années 80, la plupart des langages de programmation ne sont pas réflexifs. Beaucoup d'informaticiens s'interdisent d'utiliser cette possibilité pour des raisons d'indécidabilité ou de sécurité. Actuellement, la réflexivité est reconnue comme une notion importante et incontournable, nous en donnerons des exemples en UML, XML et Java. La réflexivité est apparue nécessaire et sans danger quand on a pu en donner des règles strictes d'interprétation avec le paradigme objet. En même temps, les langages informatiques ont perdu leur univocité d'interprétation des noms et des expressions avec le polymorphisme : une fonction ayant

le même nom peut être définie pour des types d'objets différents et on ne peut pas toujours savoir, à la lecture de l'expression d'appel de la fonction quelle est la fonction concrète qui va s'exécuter. On en trouve un exemple sur les arbres dans [Nicolle, 2000]. Pour illustrer les capacités réflexives des langages et notations informatiques, et pouvoir discuter leurs propriétés et leurs limites, nous considérerons un système de fiches décrivant des livres ou des avis sur ces livres, ou des avis sur les avis et ainsi de suite. Il est utile de pouvoir développer, sans notations particulières, autant de niveaux d'avis que nécessaire pour représenter une discussion sur des livres, une fiche d'avis peut porter aussi bien sur un livre que sur un avis donc le modèle est récursif. C'est un arbre qui comporte des noeuds, les fiches sur lequel des avis ont été donnés et des feuilles, les fiches d'avis sans fiches d'avis, qui arrêtent naturellement le système, comme dans une conversation, quand les points d'accord ou de désaccord ont été mis en évidence. On voit bien cet arbre dans la version XML des fiches ci-dessous.

### 3.1 Paquetages

La notion de paquetage (package) est commune à la plupart des langages de programmation modernes. Un paquetage regroupe des objets qu'il est logique de mettre ensemble relativement au projet. En UML, *Namespace* est une métaclassse pour définir les parties du modèle qui contiennent des éléments nommés et dans lesquelles un nom est unique (classes, fonctions, packages). Un élément d'un espace de noms est toujours visible dans cet espace et il assure l'unicité de chaque nom dans l'espace ainsi défini. C'est aussi un espace de visibilité par défaut, par opposition aux espaces *public* (visible de partout) et *private* (visible de la classe seulement). Pour accéder à un objet public d'un autre paquetage, on fait précéder le nom de l'objet du nom du paquetage, ou on importe le paquetage pour pouvoir utiliser toutes ses définitions publiques. Mais aussi, un paquetage est une manière interne de voir les dossiers du système hôte, ils portent le même nom sous les deux points de vue. Un paquetage peut comporter des sous paquetages, l'empilement des dossiers correspond à la récursivité des paquetages. Ce principe d'organisation augmente la lisibilité car l'organisation du logiciel est visible tant au niveau du système qu'au niveau interne aux programmes. La documentation est structurée de la même manière, en parallèle. En Java, la cohérence entre les points de vue interne et externe est assurés par les noms : la classe principale doit porter le même nom que le fichier dans laquelle elle est définie, au suffixe près, et le fichier compilé portera aussi le même nom.

### 3.2 Le métamodèle UML (*Unified Modeling Language*)<sup>2</sup>

UML est d'abord un métamodèle, qui unifie la sémantique de plusieurs méthodes de conception du logiciel. C'est ensuite une notation qui permet de rendre compte de ce modèle. Un métamodèle est un langage pour décrire des modèles. Il décrit les concepts qui seront utilisés pour produire des modèles. UML décrit le modèle objet de manière

---

<sup>2</sup> <http://www.uml.org/>

partagée par la plupart des grosses entreprises mondiales de construction de logiciel réunies dans l'OMG<sup>3</sup>. Il permet de donner d'une manière unique, définitive et partagée la syntaxe et la sémantique de tous les éléments de base de ce modèle (classes, attributs, opérations, composants...). Le métamodèle définit certains concepts et leurs relations, correspondant au noyau mis en commun, et il décrit la manière d'en définir d'autres pour rester cohérent. Il est donc évolutif et adaptable. Il définit les entités du paradigme objet sur 4 niveaux :

NIVEAU	DESCRIPTION	EXEMPLES
métamétamodèle	L'infrastructure de l'architecture de métamodélisation. Définit le langage pour la spécification des métamodèles	MétaClasse, MétaAttribut, MétaOpération
métamodèle	Une instance du métamétamodèle. Définit le langage pour la spécification des modèles	Classe, Attribut, Opération, Composant
modèle	Une instance du métamodèle. Définit un langage pour décrire un domaine d'informations	Fiche, FicheLivre, FicheAvis
Objets de l'usager	Une instance du modèle. Définit un domaine spécifique d'information.	f0, f1, f2, f3 (voir figure 4)

**Figure 1 : le paradigme objet**

Le métamodèle est défini avec la notation UML et les mécanismes d'organisation du modèle par 3 packages, le package de base (Foundation), et deux packages qui utilisent ce package, les comportements (BehavioralElements) et l'organisation des modèles (ModelManagement). Le package de base est lui-même composé de trois packages : les types de données (DataTypes), le noyau (Core) et les mécanismes d'extension, qui utilisent les types de données et qui s'utilisent mutuellement.

Les types de données définissent des objets qui n'ont pas d'identité. Ce sont des valeurs pures, comme les entiers et les chaînes de caractères, ou des types énumérés, comme les booléens qui ont deux valeurs : vrai et faux, ou des expressions, des noms, des adresses de machines (LocationReference), des projections (Mapping), des multiplicités (par exemple \*, 1..3). Les types de données sont un type particulier de classeur (Classifier) pour lesquels les opérations sont des pures fonctions, qui calculent des valeurs mais ne font pas d'effet de bord. Cette question des valeurs pures a été longtemps débattue dans la communauté objet par les tenants du « tout objet ».

Le noyau définit les constructions de base du métamodèle. Certaines sont abstraites donc ne sont pas instanciables, et elles définissent les constructions clés, souvent cachées, qui organisent le métamodèle (Element, ModelElement, GeneralisableElement, Classifier). Certaines sont concrètes, elles reflètent les constructions des modèles (classe, attribut, opération, association).

<sup>3</sup> <http://www.omg.org/>

Pour décrire un projet en UML, on utilise plusieurs diagrammes correspondant à plusieurs points de vue sur le projet : classes, activités... C'est le modèle sous jacent à ces diagrammes qui intègre les différentes perspectives et leur donne cohérence. Voici un exemple de diagramme de classes pour les fiches :

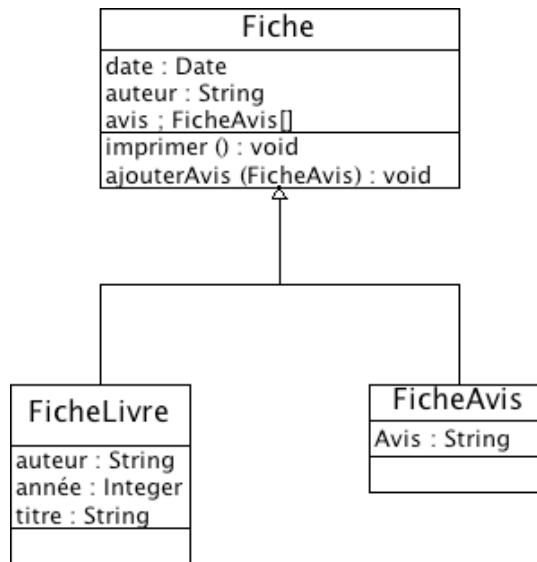


Figure 2 : diagramme de classes des fiches

UML est utilisé par les informaticiens pour concevoir les logiciels indépendamment d'un langage de programmation, et pour construire un modèle explicite, qui sera documenté de manière partagée. Mais ces caractéristiques en font un bon outil pour d'autres métiers : architecture, biologie par exemple

**La sémantique d'UML** est définie en utilisant 3 vues : une syntaxe abstraite, des règles et des contraintes sur les modèles valides. Elle est décrite en utilisant la notation UML elle-même. C'est une bonne méthode pour vérifier la consistance de cette notation, en l'utilisant pour un problème assez complexe. On peut avoir l'impression de tourner en rond, contrairement aux modèles formels en mathématiques qui sont définis à partir d'axiomes et de règles de dérivation de théorèmes, sans présupposer d'autres connaissances. Or, c'est un principe général pour la définition de langages par amorce (bootstrap). Les principales notions sont décrites dans le métamodèle par une définition métacirculaire.

*Element* est la métaclassse abstraite au sommet de la hiérarchie des métaclasses. Un élément est un constituant atomique d'un modèle. Elle a deux sous classes dans le noyau, *ModelElement*, qui fait partie des fondations, et *PresentationElement*, qui fait partie des éléments auxiliaires. *ModelElement* est une métaclassse abstraite pour les éléments nommés des modèles, par opposition aux éléments visibles qui présentent les éléments des modèles et qui eux ne sont pas nommés.

*Classifier* est une métaclassse abstraite qui décrit des types de données : valeurs pures, classes, interfaces, composants. Le diagramme des classifieurs dans la documentation UML 1.3 reproduit figure 5 montre comment UML est utilisé pour décrire cette notion.

### 3.3 La réflexivité en XML<sup>4</sup>

XML est un langage de description de données qui permet des échanges standardisés. C'est un langage de balises, comme HTML, ce n'est pas un langage d'affichage de données. Si on veut afficher des données XML, il faut définir une feuille de style qui interprétera les balises. En ce sens un fichier XML peut avoir plusieurs interprétations. Les balises XML sont définies librement par les usagers pour un usage particulier, par exemple, pour définir des fiches de livres et des fiches d'avis. La définition des balises et leurs propriétés syntaxiques est définie par un schéma XML. Lorsque un fichier XML fait référence à un schéma, le schéma fixe les interprétations valides. Les schémas XML sont définis en XML. On peut donc définir le schéma des schémas, mais peut-il se décrire lui-même ? Regardons des exemples de fiches de livres et de fiches d'avis.

```
<?xml version="1.0" encoding="Mac OS Roman"?>
<!--Fiche de livre-->
<fichelivre id= 'f0'
    <auteur> Pattie Maes </auteur>
    <titre> Computational reflection </titre>
    <annee> 1987 </annee>
    <ficheavis id= 'f1'>
        <auteur> Anne Nicolle </auteur>
        <date> 1992-06-23 </date>
        <avis> fondamental </avis>
        <ficheavis id= 'f2'>
            <auteur> Jean Legris </auteur>
            <date> 1992-07-01 </date>
            <avis> trop difficile à mettre en œuvre </avis>
            <ficheavis id= 'f4'>
                <auteur> Anne Nicolle </auteur>
                <date> 1992-08-22 </date>
                <avis> à justifier </avis>
            </ficheavis>
        </ficheavis>
    </ficheavis>
    <ficheavis id= 'f3'>
        <auteur> Anne Nicolle </auteur>
        <date> 1992-09-24 </date>
        <avis> à développer </avis>
    </ficheavis>
</fichelivre>
```

**Figure 3 : les fiches en XML**

Montrons maintenant un modèle de ce système de fiches sous forme de schéma XML.

```
<?xml version="1.0" encoding="utf-8"?>
<xss:schema xmlns:xss="http://www.w3.org/2001/XMLSchema">
```

---

<sup>4</sup> <http://www.xml.com/>

```

<xs:attribut name="identifiant" type="xs:string"/>
<xs:element name="ficheAvis">
  <xs:complexType>
    <xs:sequence>
      <xs:element name="auteur" type="xs:string"/>
      <xs:element name="date" type="xs:date"/>
      <xs:element name="avis" type="xs:string"/>
    </xs:sequence>
    <xs:attribute name="identifiant" ref="identifiant"/>
  </xs:complexType>
</xs:element>
<xs:element name="fiche">
  <xs:complexType>
    <xs:sequence>
      <xs:element name="auteur" type="xs:string"/>
      <xs:element name="titre" type="xs:string"/>
      <xs:element name="annee" type="xs:integer"/>
      <xs:element name="ficheAvis" ref="ficheAvis" min0ccurs="0"
max0ccurs="unbounded"/>
    </xs:sequence>
    <xs:attribute name="identifiant" ref="identifiant"/>
  </xs:complexType>
</xs:element>
</xs:schema>

```

**Figure 4 : le modèle des fiches en XML**

Remarquons que dans un modèle de fiche, chaque élément est défini avec son type, qui décrit les valeurs admissibles et le nombre des valeurs admissibles. Le système de description de fiches est récursif puisqu'un modèle peut apparaître dans les types de sa propre définition.

### **3.4 Le modèle objet de JAVA<sup>5</sup>**

Java est à la fois un langage de programmation par objets et une plate-forme de développement comportant un compilateur de classes, une machine virtuelle, des utilitaires de gestion de la documentation, de création d'applications autonomes et d'exécution des applets. Il a été créé par Sun en 1995. La machine virtuelle Java qui permet d'exécuter les applications et les applets est présente sur toutes les machines ayant un navigateur Web. Java a été conçu :

- pour assurer la sécurité des programmes circulant sur les réseaux et s'exécutant sur la machine hôte,
- pour assurer la portabilité des programmes sur toutes les plates-formes,
- pour faciliter la documentation et la réutilisation des classes.

---

<sup>5</sup> <http://java.sun.com/>

La syntaxe des instructions Java est très proche de celle de C, mais le modèle objet est très différent de celui de C++. C'est le modèle objet le plus abouti à ce jour, car il a profité de 20 ans de travail sur Smalltalk, CLOS et les autres extensions objet de différents langages.

**Les valeurs des variables** et les arguments des fonctions peuvent être soit des valeurs pures soit des instances de classes. Java distingue les types primitifs (char, int, double...) dont les instances sont des valeurs pures ('a', 5) et les classes dont les instances ont des variables et peuvent exécuter des méthodes. On appelle objets les instances des classes. Les classes sont elles-mêmes des instances de classes et peuvent donc être manipulées comme les autres objets. Les classes héritent de la classe abstraite Object par défaut.

Un objet est instance de sa classe, de toutes les classes dont sa classe hérite directement ou indirectement, de toutes les interfaces implémentées par sa classe et toutes les classes dont il hérite, et de toutes les interfaces dont ces interfaces héritent. Quand on déclare une variable ou un paramètre de fonction, on doit donner son type. Si le type est une classe, l'objet qui viendra dans cette place doit être une instance de cette classe ou de l'une de ces sous-classes. Mais une interface peut servir de type aussi bien qu'une classe. Pour créer un objet qui viendra dans cette place, on peut instancier n'importe quelle classe implémentant cette interface, ou une interface qui en hérite. La définition des types des places est donc très souple et en définissant le type d'une place par Object, on peut obtenir la même souplesse d'utilisation qu'avec les variables non typées de Lisp.

**La réflexivité** s'exprime de plusieurs manières, certaines sont usuelles et d'autres réservées aux spécialistes. On peut faire appel à un objet à l'intérieur de lui-même en utilisant **this**. Les classes et les interfaces sont des instances de la classe **Class** et **Class** est un type comme un autre. Il est possible d'examiner les classes, les interfaces et les objets à l'intérieur d'un programme. Par exemple, on peut connaître la classe d'un objet x par x.getClass() ou x.class. On peut interroger une classe pour savoir quels sont ses modificateurs (abstract, public...), ses attributs, quelles sont ses superclasses et les interfaces qu'elle implémente. On peut aussi obtenir des renseignements sur les méthodes et sur les constructeurs grâce au package **java.lang.reflect** qui propose des classes et des interfaces pour obtenir l'information réflexive sur les objets. On les utilise pour construire des compilateurs, des debuggers, des inspecteurs et des browsers.

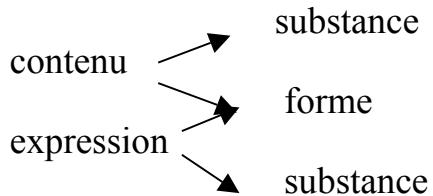
Dans les années 60, la sémantique des langages de programmation n'était pas bien établie et les compilateurs pouvaient avoir des interprétations différentes des mêmes constructions syntaxiques. Dans un premier temps, la sémantique a été définie formellement, par des règles qui s'accrochaient aux constructions syntaxiques. Mais il n'y avait pas de moyen interne de vérifier la sémantique. Avec le paradigme objet, la réflexivité des définitions sémantiques permet une certaine validation.

## 4 Conclusion

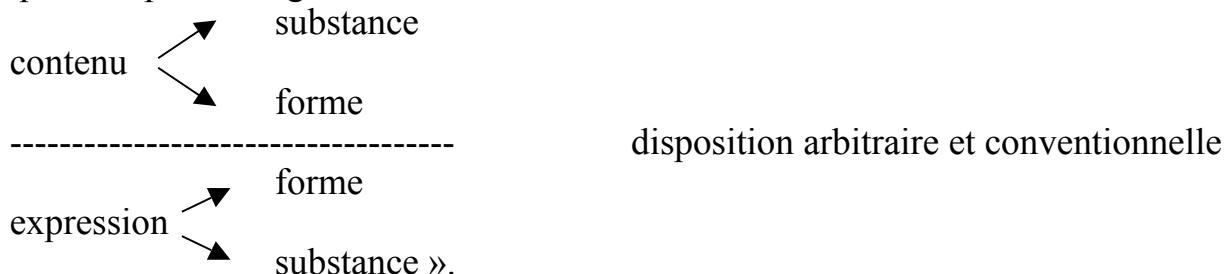
Dans le paradigme objet, on distingue bien le niveau des objets, le niveau des modèles et le niveau des métamodèles. Le modèle des modèles permet de construire un type pour les modèles, comme les modèles permettent de construire des types pour les objets. Ce

sont ces distinctions qui permettent d'implanter la réflexivité de manière productive. Par exemple, la notion de classe comme type concret est le résultat d'une activité d'abstraction sur les représentations des objets. Les métaclasses sont aussi le résultat d'une activité d'abstraction qui cette fois s'applique aux objets abstraits que sont les classes. Donc les classes et les métaclasses sont finalement de même nature. Un système d'objets décrit selon ce modèle est réflexif car les représentations d'objets et les modèles (classes, schémas) sont de même type, défini par le modèle des modèles ainsi que les attributs et les fonctions.

La réflexivité des notations UML ou XML est une réflexivité statique. En 2002, la définition du MOF, Meta-Object Facility, par l'OMG a unifié la description des métamodèles objet<sup>6</sup>. La définition de la réflexivité dans la spécification du MOF fait 31 pages, ce qui montre l'importance de cette notion dans les préoccupations des ingénieurs. La réflexivité des langages de programmation est une réflexivité dynamique, l'interprétation a lieu à l'exécution, dans le contexte du calcul en cours. On peut donc faire l'hypothèse que c'est ce niveau de modélisation qui permettra de rendre compte du fonctionnement des langages dans une machine. J'en ai posé les prémisses dans [Nicolle, 2002] en montrant comment le jeu entre forme et substance d'une part, expression et sens d'autre part, était la clé de cette modélisation. « Dans un langage formel, le contenu est intimement lié à l'expression à cause de l'identification nécessaire entre forme de l'expression et forme du contenu qui permet de calculer sur les expressions et de transférer les résultats sur les contenus.



En langue quotidienne, grâce à la disposition arbitraire entre expression et contenu, la question peut se régler naturellement.



Les modèles dits formels d'une part ne posent pas la question de la substance et d'autre part aplatissent le sens sur son expression. De plus, comme le sens est un processus, il ne peut être décrit que par un processus. L'interprétant machine, comme l'interprétant humain, introduit le temps dans le texte, pas le temps représenté par une variable, mais le temps « réel ». [Nicolle, 2004]

<sup>6</sup> <http://www.omg.org/cgi-bin/doc>

## 5 Références

- Noam Chomsky, 1957, *Syntactic Structures*. The Hague: Mouton & Co
- Louis Hjelmslev, 1947, *La structure fondamentale du langage*, traduction française 1971, dans *Prolégomènes à une théorie du langage*, Éditions de Minuit, Paris
- Douglas Hofstadter, 1985, *Gödel, Escher, Bach Les brins d'une guirlande éternelle*. Basic Books, Inc. 1979. Édition Française Interéditions, Paris
- Ioannis Kanellos, 1999, De la vie sociale du texte. L'intertexte comme facteur de la coopération interprétative. *Cahiers de Praxématique* n°33, p. 41,82
- Daniel Luzzati, 2000, Impératif, indicatif et subjonctif : de la morphologie à l'interaction, *Revue de Sémantique et Pragmatique*, N° 8, Pages 7-27
- Bernard Morand, 2004, *Logique de la conception. Figures de sémiotique générale d'après Charles S. Peirce*, Éditions L'Harmattan, Collection L'ouverture philosophique, Paris
- Anne Nicolle, 2001, *La question du symbolique en informatique*, in *La cognition entre individu et société*, ARCo'2001, pp. 345-358, Hermès, Paris
- Anne Nicolle, 2002, *Sciences de l'artificial, modélisation et rationalité*, *Revue d'intelligence artificielle*, vol. 16, n° 1-2, pp. 63-86, , Hermès, Paris
- Anne Nicolle, 2004, *Temps et processus*, Journées de Rochebrune « Le temps dans les systèmes complexes naturels et artificiels », Megève, France, ENST 2004 S 001, pp. 229, 239
- Jacques Vergne et Emmanuel Giguet. 1998. *Regards Théoriques sur le "Tagging"*. In proceedings of the fifth annual conference TALN 1998, Paris, France

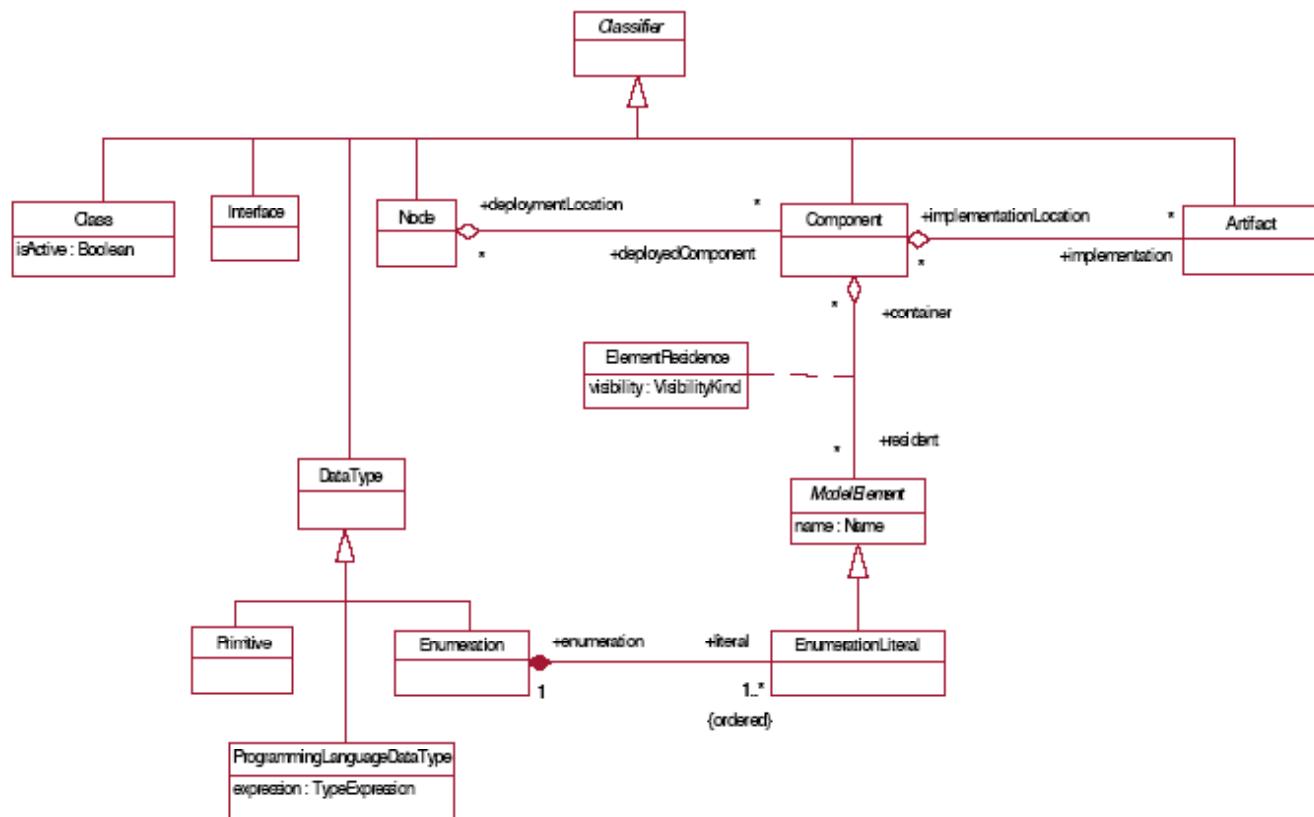


Figure 5 : diagramme de la sémantique des classificateurs

# Designing reflexive agents in a MAS model for urban mobility

Sylvie Occelli

IRES Istituto di Ricerche Economico e Sociali del Piemonte  
via Nizza 18,  
10125 Turin, Italy  
[Occelli@ires.piemonte.it](mailto:Occelli@ires.piemonte.it)

**ABSTRACT.** One crucial aspect for developing a MAS model for urban mobility is the definition of a strong notion of agency. In this respect, it is argued that the consideration of individuals' mental worlds plays a crucial role. First, a few insights into the notion of mental worlds are provided. A tentative definition is then suggested based on the coupling of a semantic and syntactic component. Finally, the introduction of mental worlds in current Activity Based Approaches is discussed. Some remarks about the next steps in the mobility research project conclude the paper.

Key words: MAS approach, urban mobility, mental worlds, urban agents.

## 1. Introduction

The definition of a strong notion of agency, see [29], is one major fundamental aspect for developing a MAS model for urban mobility. As pointed out in Occelli [23, 24] it requires to jointly considering an analytical-epistemological and a conceptual-ontological dimension. While increasing attention is being given to the former, the latter is to date still largely unexplored.

In particular, it is argued that one fundamental advantage of a MAS approach to urban mobility is the possibility to consider agents' mental worlds. This entails taking into account novel types of model variables, i.e. considering agents' psychological attitudes and learning abilities, while revising the categories commonly retained in describing urban behaviours, i.e. viewing urban activities as entities resulting from the consolidation of agent engagements.

These topics are under investigation in a MAS modelling study that has been lately launched at IRES as a part of a research project concerning the analysis and forecasting of regional mobility, see [24].

The present communication addresses some of the questions which are raised at the early stage of the model design, i.e. those dealing with the definition of some model entities and namely of agents' mental worlds. In the next section an exploration of the notion of mental worlds is addressed. A tentative definition is then suggested based on the coupling of a semantic and syntactic component. Finally, in the last section some suggestions about how to introduce mental worlds in the current Activity Based Approaches are briefly mentioned. A few remarks about the next steps for

implementing architecture of mental worlds for MAS based mobility analysis are suggested.

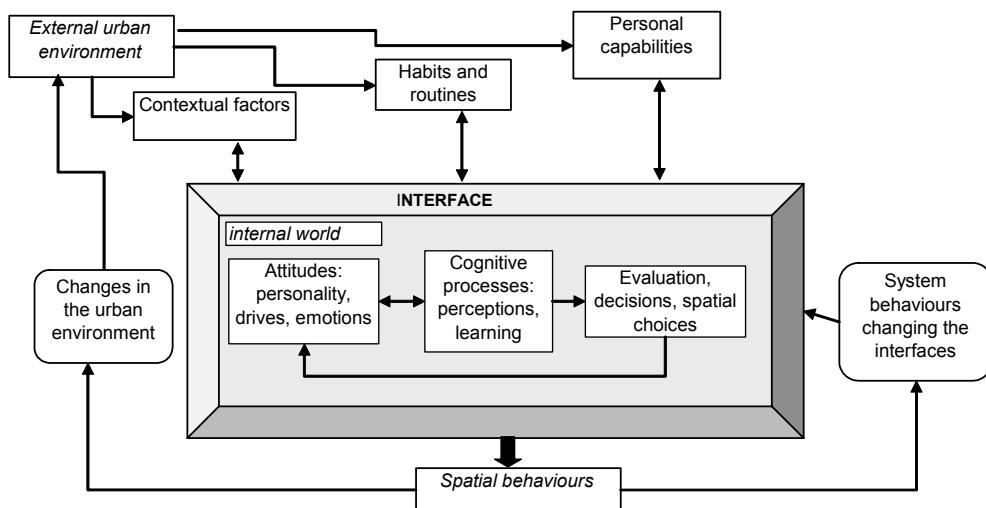
## 2. Defining mental worlds

A Mental World (MW), i.e. what is also generally referred to as a mental model, information processing mechanism, cognitive schemata, and conceptual blending, is a major challenging notion in many studies dealing with human cognition which nonetheless defeats any disciplinary bounded definitions.

An examination of the literature, see [8], [15], [17], and [27], shows that one general aspect shared by most of the notions is that MWs are cognitive entities involving both semantic and syntactic components, which are coupled together to account for an individual's understanding of his/her urban environment.

One rather straightforward application of this definition as far as spatial mobility is concerned is illustrated in the scheme of Fig. 1. It highlights the main groups of factors affecting spatial behaviour, i.e. contextual factors, habits and routines and personal capabilities. Also shown are two further elements:

- the working of the internal world, dealing with the process linking personal attitudes, cognitive processes and decision-making,
- and the role of the interfaces which filter / mediate the internal cognition process and the external environment.



**Fig. 1** An example of the application of the working of agents' mental worlds in a spatial environment

Associated with the latter are a few facts, which are increasingly important in today urban mobility, i.e.:

- the impact of the means of communication, i.e. the types of New Information and Communication Technologies, which affect the amount and quality of information

agents can extract from their environment, therefore influencing agents' spatial behaviours

- and the fact that travel decisions do not depend only on urban activities, agents' socio-economic characteristics and institutional constraints, but also on the interaction means themselves, i.e. transport modes and virtual (not physical) broadband connections, which are available in a certain area.

With a few exceptions represented by the behavioural studies, the relevance of MWs in guiding agents' spatial mobility or, more generally, individuals' behaviours in a spatial environment, was barely noticed. In particular, these studies paid attention to (see [14], p.2): a) the individual as opposed to the aggregates; b) the development of process-oriented explanations of behaviours; c) a multidisciplinary outlook; d) the notion that an individual's behaviour is based upon his/her perception and cognition of external environment; e) the concept of an individual being an actor as well as a reactor; f) the development of dynamic models of behaviour; g) the recognition of regularities in actual, as opposed to optimal behaviour and h) the search for behavioural postulates that are independent of spatial structure.

The generative role of Mental Worlds emphasized by Fig.1, however, has been only implicitly tackled by those studies.

If we hold the definition of Mental Worlds previously mentioned, then, these should be viewed as intrinsic features of human beings. Mental Worlds therefore should be considered as constitutive of the urban agency as they derive from two main fundamental facts:

- a. the fact that the drive for explanatory search for understanding (learning) is a constitutive trait in individuals, as human beings explain themselves and their circumstances while operating as observers, see [22]. It represents, therefore, a permanent spur for individuals to learn about themselves and their surrounding environment. Underlying it is a kind of meta-knowledge dimension capable of accounting for a wide set of situations resulting from both agents' internal and external cognition, i.e. as information can be extracted from the environment or derived from communication and interaction with other agents;
- b. the fact that any living system is capable to derive measures from the external environment and make sense from them, i.e. it is endowed with the so-called called operational closure [25]. This means that there exists a reflexivity dimension, which underpins agents' behaviors in an urban environment. It can be instanced in a continuum of situations, ranging from the assessment of agents' own decision-making (rules of action) to the observation of the outcome of agents' actions (percepts of the state of the world).

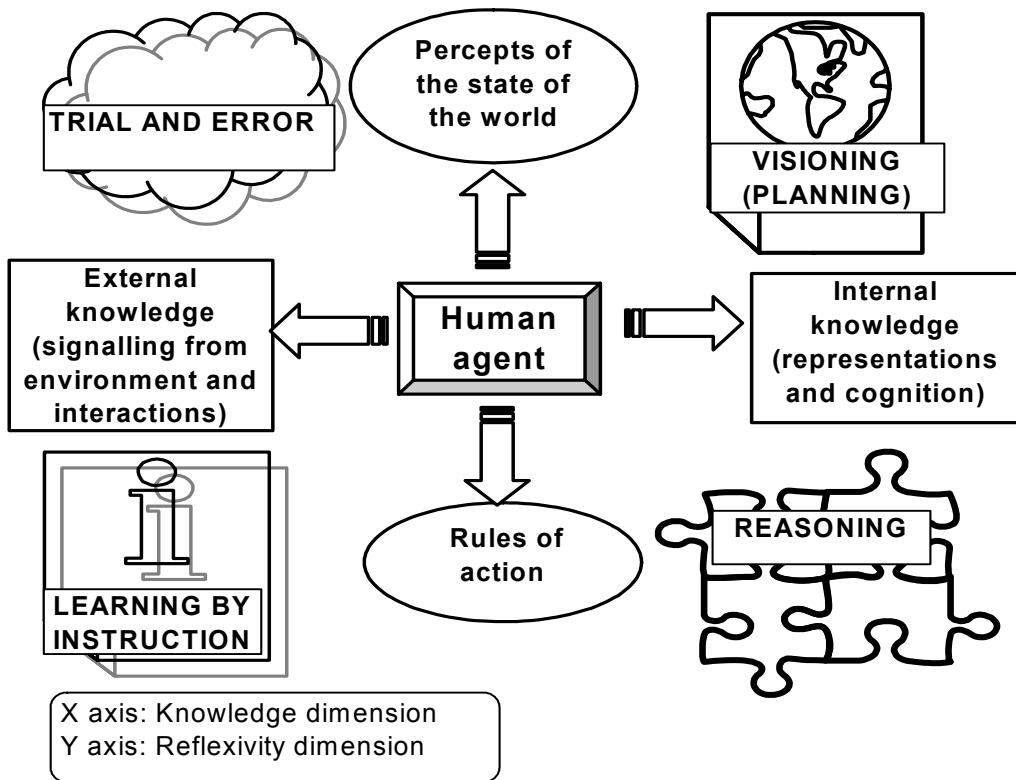
While the latter dimension points out a notion of reflexivity which is commonly referred to in devising an agent's profile, see [9, cap. 5], at least in a weak definition of agency, see [29], the former adumbrates a deeper notion. For an urban agent's, in fact, knowledge is not only associated with evolutionary processes, such as learning and adaptation, but also implies a reflexive activity about the type of knowledge endeavour involved in those processes. As argued in Chavalarias [7], the ability to

change that endeavour, i.e. to have some meta-rules guiding the knowledge process, turns out to be one major feature that distinguishes urban agents, and more generally human social agents from other types of agents.

The joint consideration of the knowledge and reflexivity dimensions allows us to identify some stereotyped cognitive stances. They can be related to the conventional architectures of the mind, consisting of a three-fold division between perception, central processing and action and a three layered levels of functions, where each level can have a different kinds of processing (as well as functional role) and higher levels can or can not dominate or control lower ones, see [28].

To fix the ideas, the identified types of cognitive stances have been labelled as graphically illustrated in Fig. 2 and can be briefly described as follows:

- Trial and error. This reflects the most basic kind of reactive behaviour based on stimulus action response. It typically occurs in emergence situations when unexpected events make it impossible to devise a strategy of action. This cognitive stance may also underpin those spatial behaviours which are more sensitive to established habits, i.e. those ways of acting, executed without reflections, which are developed through systematically repeating a behaviour having a positive consequence, see [12] and [16];
- Visioning. This can be related to what has been identified as the third of the mind's three I's, imagination, the others being, identity and integration, see [8]. Contrary to what is commonly held, in fact, imagination is not a character a few gifted individuals are endowed with, but a common one. It is at the basis of the anticipating attitude enabling urban agents to be pro-active, i.e. it may motivate them to take the initiative in a goal-directed behaviour and/or in thinking about agents' own goals. This cognitive stance might be also related to the higher layers of the mind's architecture, where more reflective (conscious) processes of decision-making take place;
- Reasoning. It stands for the purest cognitive attitude belonging to that realm of human experience, dealing with coherence, i.e. the process by which truth is ascribed to statements, which are consistent with other true statements. This is the focus of logics and forms the presuppositions on which most of conventional, and typically mathematical, approaches to urban mobility are developed;
- Learning by instruction. It underpins a whole set of processes in which episodic information are dealt with by instantiation of some generic memory structures previously created and retrieved, i.e. production rules, schemas, scripts.



**Fig. 2** Cognitive stances in agents' mental worlds

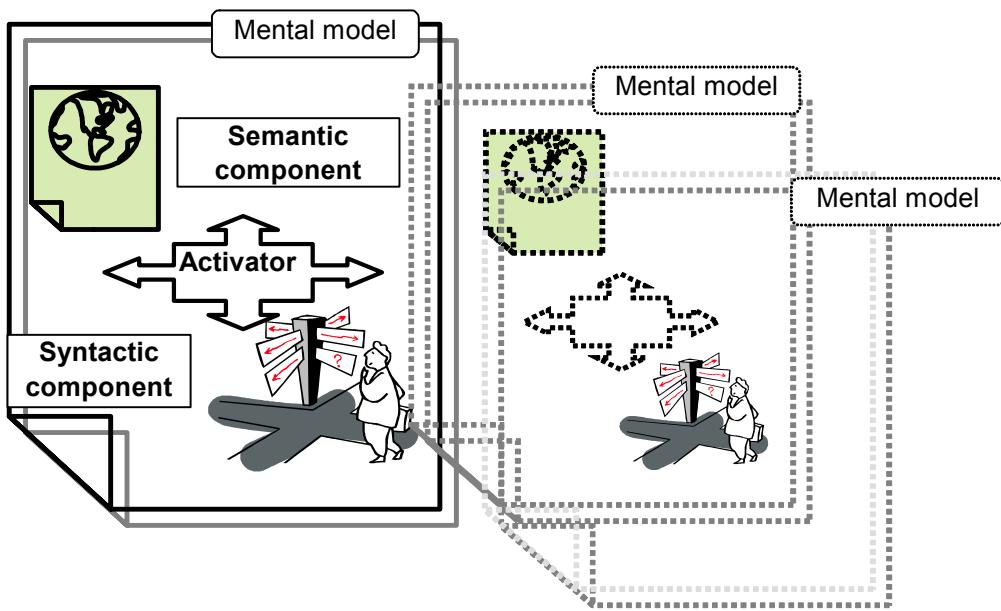
Of course, the above types of cognitive stances should be simply understood as conceptual pinpoints. In many practical situations a blend of cognitive stances is involved, although in certain cases one particular type may be prevailing, reinforced or simply activated, thus producing different behavioural outcomes.

They might however inspire more articulated descriptions as, for instance, in the case of metaphorical reasoning, where imagination and learning by instructions seem to jointly apply.

The arguments discussed so far mainly dealt with the syntactic components of MWs. It was emphasized however that semantic components exist which are coupled with the syntactic ones. It is the dynamic interlinking between syntactic and semantic components, which allows individuals construe their mental models and give meaning to a certain situation, see [7]. Because of their generative nature, not a unique, but a plurality of MWs may therefore co-exist. For the sake of illustration, one might suppose, as in Fig.3, that there exist two distinguishable components interlinked by a kind of activator, which makes it possible the generation of mental models and their integration.

As for meaning, it has been posited that it involves two main aspects, sense and reference, see [27]. The former brings to the fore the notion of coherence, and, as mentioned above, it is usually associated with the cognitive stance underlying reasoning. The latter implies a more general notion of consistency with some states of the world (and agent's own states as well) as perceived and acknowledged by the agents. It is the nature of these perceptions, which turns out mostly stirring for urban mobility phenomena.

As pointed out in the field of cognitive studies see [6], [8], the formation of these perceptions and the action response, which in term of observable behaviours these are likely to produce raise a number of questions.



**Fig. 3** A speculative representation of mental models and their components

A few aspects, which appear relevant also in activity and travel decision-making processes, are the following.

First, the definition of the very nature of needs should be given further insights. The needs of activity engagement and related travel cannot be defined without accounting for agents' underlying goals since only teleonomic goal-oriented systems have needs. In addition, the notion of need implies not only the idea of lacking or deprivation but also that of its awareness.

As a consequence, different levels of awareness in needs are likely to entail different goals. Tab. 1 illustrates a possible articulation as far as travels are concerned. A three level articulation is suggested, ranging from a simple to a more conscious perception of the need of travel.

**Table 1.** An articulation of the needs of travel

Level of awareness	Types of perception	Agents' profile	Likely behaviour
A) Perceiving the need of travel	Embodied perceptions, resulting from body sensory-motors	Agents need to travel as a result of a lack of spatial movements	They make travel
B) Feeling the need of travel	Perceptions are mediated by agents' internal and external beliefs	Agents are cognitively aware of their needs of travel. They have beliefs about the reasons to travel and a certain level of information about the system opportunities and constraints	They undertake a decision-making process for identifying their travel pattern in order to satisfy their perceived travel needs, given the existing constraints and available level of information
C) Sensing the need of travel	Perceptions are modified as a result of changes in agents' internal and/or external beliefs	Agents are aware that they need to engage in activities and that these may be actively managed in their collective system	They engage in (organize) their activities and travels

Feeling the need of travel supposes that a) travel needs depend on some perceptions (or at least partially perceived) travel outcome and b) these perceptions are mediated by internal models and external information. Sensing the need of travel recognizes the possibility that deeper effects are at work as the blend between the internal and external worlds depends on its turn from the cognitive stances of both individual and collective agents.

Second, owing to the intrinsic features of humans beings mentioned earlier, (at least for case B and C in Tab.1) there should exist certain reference values which agents take into account in evaluating their perceptions and related needs. These values can belong to both the agent's internal world, being driven by his/her beliefs and desires, and external world, being entrenched in the norms, rules, and social obligations constituting the individual's cultural world. Nonetheless, as Castelfranchi [6] points out, since mental models are rooted in both cultural and material embodiment, it cannot be excluded that certain travel needs are likely to be more cogent if instantiated as bodily perceptions.

Finally, and this holds in particular for the case C, the importance of the relationships between the internal and external worlds and of their co-evolution should not be underestimated. In this regard, there are at least two aspects worth being given further attention, i.e. the idea that an external cultural word is the consolidated outcome of shared individual mental worlds, see [8] and [20] and the Gibsonian notion of affordances, i.e. the potentialities held by objects for certain actions, see [18], which may play a role as an activator of the relationships between those semantic and syntactic components of Fig. 3.

### 3. Introducing the notion of Mental Worlds in Activity Based Approaches

Central to the Activity Based Approaches (ABA) is the notion of action space originally introduced by the Swedish human geographers in the sixties. The current applications aim to realistically account of mobility by providing a more detailed description of how individuals partake in activities (the so-called activity agendas), i.e. describing the access to urban activities (i.e. location, types of activity, opening times, and means of transport), daily scheduling, and set of family constraints individuals had to cope with in order to partake in the activities, see [5], [10] [19], and [21].

One major acknowledgment in the recent studies is that individuals do not simply partake in activities, depending on their activity space, bundle of constraints and consumer preferences. Rather, it is recognized that: a) individuals engage in activities and b) the decision-making process underlying this engagement, has its own causation, i.e. it depends not only on the changes in both the organization of activities but also on agents' overall states and modifications of agents' mental worlds.

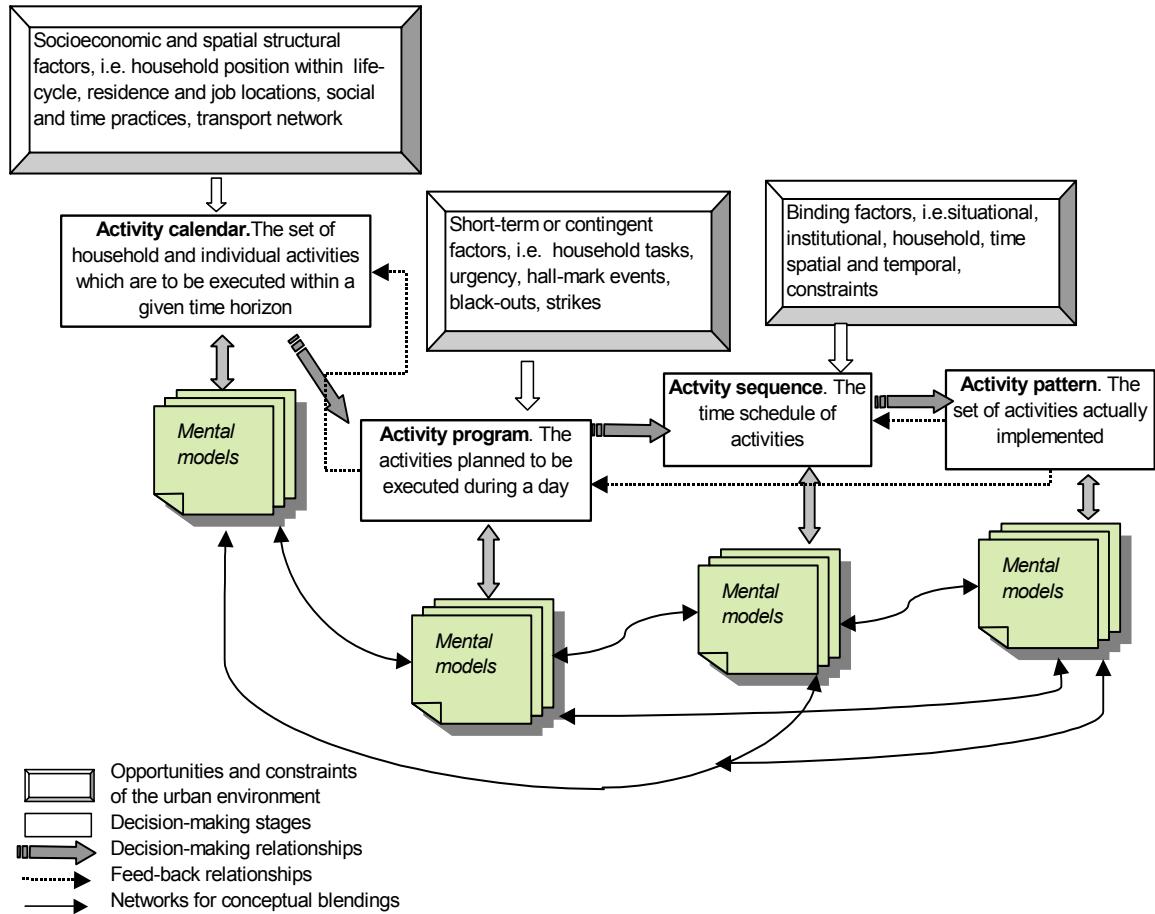
As a result, the current agent-oriented ABA models to mobility make an effort to explicitly include the decision-making process, see [1], [2], [3], [4], [5] and [26].

In particular, attention is focused on the fact that decision-making process unfolds over time as daily activity patterns depend on both short-term and long-term decisions. Activity agendas are interpreted as a hierarchy of successive levels reflecting the different temporal stages at which decisions concerning activity participation are taken. The scheme of Fig. 4 illustrates the identified stages.

An activity pattern is then understood as the outcome of the successive realization of a long-term activity calendar, a medium term activity programs and a short-term activity schedules.

It is worth noting that in several models being currently developed to describe that framework, the notion of agency typically refers to the different stages of the mobility decision-making process. A number of possibilities are foreseeable.

A first, rather straightforward, option is to assume that the framework provides a reliable insight into an agent's decision making process, i.e. this means to suppose that there is a correspondence between the observable articulation levels of the activity pattern and the cognitive stages of the decision making process. The latter would be hierarchically ordered and the mental models associated with its various stages correspond (map) to the levels of the activity pattern, see Fig. 4. The links between these levels, then, should have to be replaced by the networks of conceptual blending likely to be activated among the various mental worlds.



**Fig. 4A** descriptive framework of the formation of activity agendas (developed upon [2])

While appealing in term of the possibility of integration within existing agent-oriented ABA approaches, this interpretation is not without flaws. A fundamental question in fact is raised concerning whether some specific mental worlds, i.e. those resulting from a blend of the cognitive stances of Fig. 2, exist which may represent the different stages, in a meaningful way. The possibility that mental worlds could evolve, by means of individual learning, communication among agents or broader cultural changes would have to be addressed too. Even though admitting the conceptual plausibility of this option, its practical implementation in terms of data requirements and processing, however, is likely to be cumbersome, at least with the existing computing technology.

An alternative option is to focus on the decision making process as a whole and instead of looking for an explanation/interpretation of its relevant stages, explore the reasons behind that process and how these may evolve as a result of agents'

adaptation and learning. In this vein, Rind et al [26], for example, distinguish different types of learning:

- Learning about states of the world, i.e. improving an agent's ability to distinguish situations, and increase his/her familiarity with urban environment;
- Learning about the opportunity space, i.e. getting information that activities meet certain requirements;
- Learning about interpretations of historical trajectories, i.e. developing better knowledge about causal linkages in the environment;
- Learning about the decision rules, i.e. refining the existing rules as a result of an agent's experience or new insights.

To a closer inspection, in fact, learning and adaptation turn out to be the outcome of (the functioning of) mental models, thereby producing an updating of their underlying semantic and syntactic components, while yielding new meanings as well. This option actually elicits a need to extend the very function of the decision making process and recognize its twofold role of:

- prompting agents in their activity engagement, and namely, facilitating individuals' participation to activities, i.e. providing information about destinations and means of transport, making possible the access to activities;
- motivating individuals to actively construe their own activity engagement, as they can make sense of their perceptions of travel needs.

Also in this case, a temporal perspective is involved which, however, appears ontologically different from the one considered in the previous option.

Of course, while acknowledging this diversity, one may suppose to embed the latter within the former, and look for some relevant overlapping which would be both valuable for interpreting mobility needs and constructive for devising mobility policy measures.

## 4. Concluding remarks

This paper argued that the notion of mental models as cognitive entities involving both semantic and syntactic components is likely to be crucial for identifying a strong notion of urban agent in a MAS based mobility analysis. In particular, a claim was made that mental models play a role in the decision making process guiding the formation of individuals' activity patterns. In this direction, the possibility of their integration into existing descriptions of activity patterns was addressed and two options identified, i.e. linking the levels of the activity patterns with stages of the decision making process and motivating the role of the decision-making process. The opportunity to view these options not as opposing alternatives, but rather as different lines of enquire each of which is likely to yield valuable insights, was emphasized.

Whereas the former would be most helpful for supporting the management of mobility, i.e. devising more suited policy bundles for the increasingly diversified

mobility situations, insights into the latter would be essential for advancing the understanding of mobility demand:

- A theoretical one, related to the definition of the various profiles, underpinning the four types of cognitive stances in Fig. 2;
- An operational one, concerning the design of the agent architecture capable to account for the variety of cognitive profiles likely to be associated with the identified cognitive stances.

## References

1. Arentze T., Timmermans H (2000) *Learning Based Oriented Simulation System*. European Institute of Retailing and Service Studies, Eindhoven.
2. Arentze T., Timmermans H. (2002) Modeling the Formation of Activity Agendas Using Reactive Agents. *Environment and Planning b* 29, 719-728.
3. Arentze T., Timmermans H. (2003) Modeling Learning and Adaptation Processes in Activity Travel-Choice. *Transportation* 30, 37-62.
4. Bazzan A. L. C., Wahle J., Klugl F. (1999) Agents in Traffic Modelling-from Reactive to Social Behaviour. In: Burgard W., Christaller T., Cremers A.B. (eds.) *Advances in Artificial Intelligence (KI-99)*, LINAI 1701. Springer-Verlag, Berlin, 303-306.
5. Bonnel P. (1995) An Application of Activity-Based Travel Analysis to Simulation Change in Behaviour. *Transportation* 22, 73-93.
6. Castelfranchi C. (2000) Come studiare la mente (per quello che è). *Sistemi Intelligenti* 1, 39-66.
7. Chavalarias D. (2004) Métodynamiques en Cognition Sociale, Ecole Polytechnique. Filière Economie et Sciences Sociales, [www.Chavalarias.free.fr](http://www.Chavalarias.free.fr).
8. Fauconnier G., Turner M. (2002) *The Way we Think. Conceptual Blending and the Mind's Hidden Complexities*. Basic Books, New York.
9. Ferber J. (1999) *Multi-Agent Systems. An Introduction to Distributed Artificial Intelligence*, Addison-Wesley, Singapore.
10. Fuji S., Garling T. (2002) Application of Attitude Theory for Improved Predictive Accuracy of Stated Preference Methods in Travel Demand Analysis. *Transportation Research Part A* 37, 89-402.
11. Gabora L. (1997) The Origin and Evolution of Culture and Creativity. *Journal of Memetics* 1, [www.jom-emit.cfpm.org/all.html](http://www.jom-emit.cfpm.org/all.html).
12. Garling T., Axhausen K. W. (2003) Introduction: Habitual travel choice. *Transportation* 30, 1-11.
13. Garvill J., Marell A., Nordlund A. (2003) Effects of Increased Awareness on Choice of Travel Model. *Transportation* 30, 63-79.
14. W. C. Halperin W. C. (1988) Current Topics in Behavioural Modelling of Consumer Choice. In: Golledge R. G., and Timmermans H. eds. *Behavioural Modelling in Geography and Planning*. Croom Helm, London, 1-26.
15. Hutchins E. (1998) Mental Models as an Instrument for Bounded Rationality. [www.cogsci.ucsd.edu/hutchins](http://www.cogsci.ucsd.edu/hutchins).
16. Janger W. (2003) Breaking 'bad habits': a dynamical perspective on habit formation and change. In: Hendrickx L., Jager W., Steg L. (eds.) *Human Decision Making and Environmental Perception. Understanding and Assisting Human Decision Making in Real-life Settings*. Liber Amicorum for Charles Vlek. University of Groningen, Groningen.
17. Johnson-Laird P. N., Girotto V., Legrenzi P. (1998) Mental Models: a Gentle Guide for Outsiders. [www.si.umich.edu/ICOS/gentleintro.html](http://www.si.umich.edu/ICOS/gentleintro.html).

18. Knappett C. (2003) The Affordances of Things: A Post-Gibsonian Perspective on the Relationality of Mind and Matter. Cambridge Symposium, Rethinking Materiality: the Engagement of Mind with the Material World, 28-31 March, Cambridge.
19. Lee M.S., McNally M.G. (2003) On The Structure of Weekly Activity/Travel Pattern. *Transportation Research Part A* 37, 823-839.
20. Louça J. (2004) Organizational Culture in Multiagent Systems: Metaphorical Contributions to a Discussion. Conference of the North American Association for Computational Social and Organizational Science (NAACSOS-04), Pittsburgh PA.
21. McNally M. (2000) The Activity-Based Approach. Institute of Transportation Studies University of California, UCI-ITS-WP-00-4, Irvine California.
22. Maturana H. (2000) The Nature of the Laws of Nature. *Systems Research and Behavioural Science* 17, 459-468.
23. Occelli S. (2004) A MAS Perspective in Urban Modelling. Proceedings of SCI2004, 18-21 July, Orlando.
24. Occelli S. (2004) "Sensing" mobility: an outline of a MAS model for urban mobility, Proceedings of ESSA2004, 16-19 September, Valladolid.
25. Pattee H. (1986) Universal principles of measurement and language functions in evolving systems. In: Casti J., Karlqvist A. (eds.) *Complexity, Language, and Life: Mathematical Approaches*. Springer-Verlag, Berlin, 268-281.
26. Rindt C.R., Marca J.E., McNally M. G. (2002) Toward Dynamic, Longitudinal, Agent-Based Microsimulation Models of Human Activity in Urban Settings. Institute of Transportation Studies University of California, UCI-ITS-WP-02-5, Irvine California.
27. Schwamb K. B. (1990) Mental Models: a Survey. Department of Information and Computer Science, University of California, Irvine, Unpublished.
28. Sloman A. (2000) Introduction: Models of Models of Mind, The DAM Symposium: How to Design a Functioning Mind, 17-18 April.
29. Wooldridge M., Jennings N. R. (1995) Intelligent Agents. Theory and Practice. *Knowledge Engineering Review* 10 (2), 115-152.

# On the Design of a Musical Flow Machine

*François Pachet*

This paper addresses the issue of designing interactive systems that create flow experiences in users. I first describe an interactive musical system called the Continuator, which is able to learn the musical style of users in an agnostic, continuous fashion. I then describe experiments conducted with professional musicians and with 3 to 5-year old children and the Continuator. I show that these interactions are – almost – typical of the Flow phenomenon, as introduced by Csikszentmihalyi. I then focus on the abstraction of the design principles behind the Continuator and propose the notion of *Reflective Interactive System* as a class of applications which trigger Flow experiences. Based on the analysis of the various psychological experiments conducted so far, I identify the issue of flexibility in interaction protocols as a crucial step to enhance the efficiency of Reflective Systems as we envisage them today.

## 1. When are interactions interesting?

Virtually all things done by a computer program are interactive today. From web sites to word processors, from video games to entertainment robots, users are constantly engaged in various forms of dialogs with computer programs.

In most cases, these dialogs are designed to help users solve precise and well defined tasks. For instance, information retrieval systems are designed to allow users to find quickly information they look for. Many web sites are – at least in theory – often designed to minimize the number of clicks needed to find specific information. Video games propose interactive devices designed so that users can quickly issue

commands to move, shoot, or perform various actions in real time, without having to think about their input devices. In all these cases, the design principles put forward consist in satisfying precise criteria: utility, optimisation, conciseness or transparency. Of course there are some exceptions to the utilitarian view of interactivity: artistic installations in the domain of *digital art* (as found in shows such as Ars Electronica) consist often in providing the user with some sort of hopefully novel aesthetic or sensory experience, which does not necessarily correspond to a task or game to play with precise rules to follow. However, this lack of aim in purely artistic experiments makes these installations often hard to understand. Actually, the very idea of an aesthetic experience *as such*, decoupled from any relation to content is itself debatable.

Whatever the subject matter, all interactions at work in man-machine interfaces do not create equally *enjoyable* experiences for users. This very issue of what makes an interactive system appealing, enjoyable or attracting to users is systematically eluded. It is also the subject matter of this chapter: what makes an interaction interesting ?

This bold question is addressed under the form of a particular experiment in music interaction, called the Continuator project, whose goal is precisely to engage users in exciting, appealing interactions. We first introduce the project and the basic mechanisms of the system, and then describe some experiments involving children. Then we show the relations that these experiments may have with the theory of Flow, as introduced by Csikszentmihalyi (1990). In the last section we generalize from the experiments with the Continuator and propose the notion of *Reflective Interactive System* as a class of applications having the same properties than the Continuator. We propose that Interactive Reflective Systems are a possible way to build Flow machines. We conclude on proposing extensions and other examples of this class of system and issues remaining to be solved.

## 2. The Continuator

The Continuator project stems from major frustrations of the author regarding music interactive systems. The issue of interacting with a computer to make music has long been addressed by many researchers, with many different goals in mind. Today there seems to be two categories of music making systems.

On the first hand, purely interactive systems in which users may trigger various kinds of musical effects. These systems have been experimented for a long time by the pioneers of computer music. For instance, Jean-Claude Risset composed pieces with a Yamaha Disklavier

(a piano forte with a Midi input and output) in which the computer played various types of accompaniment based on his input (Risset & Van Duyne, 1996). More recently, the Korg company issued a synthesizer of a new kind, called the Karma, which proposes thousands of such musical effects (Kay, 2000). This notion of musical effect is particularly well adapted to describe the type of interaction at play: when a user presses a key, a chord, or some sort of predefined musical sequence, the system reacts by producing, in turn, a sequence based on the user input. The versatility of the machine makes it possible many effects ranging from arpeggiators to automatic accompaniments in many styles.

On the other hand, a lot of research has been devoted to understanding the notion of musical style, and building systems that produce music in particular styles. One of the first attempt in building an automatic composer is probably Mozart with his *Musikalischs Würfelspiel* in 1787 (Chuang, 1995). This composition took the form of a dice game. Each throw would determine a number. This number, associated with the number of the current measure would determine the next measure, out of a collections of pre-programmed measures composed by Mozart himself. From a table of 176 minuet measures, the device could generate, thanks to two 6-sided dices,  $1.3 \cdot 10^{29}$  different minuets. The device invented by Mozart was in fact a rudimentary implementation of so-called Markov models, invented much later, but using the same principle: transition tables containing probabilities about transitions between two events. Applying Markov models to analyze and generate music has become, since Mozart, a tradition in itself. Today, many systems have been proposed to analyze automatically corpuses of music material, under their score or Midi form, and to generate new material based on these analysis. The most spectacular of these attempts was probably done by Cope (Cope, 1996) with his EMI system. David Cope's system is able to generate convincing musical pieces in the style of virtually all classical and modern composers, ranging from Jean- Sebastian Bach to Scott Joplin. However, to achieve a reasonable level of accuracy and faithfulness to the style being mimicked, Cope has to enter by hand many information to the system, in particular concerning high-level musical constructs such as forms, beginning, endings, etc. which are not naturally captured by Markov models. However, whatever their level of accuracy music generation system are not interactive: the processes of analysis and generations are separated, with possibly many hand made parameter tweaking inside.

Unfortunately, purely interactive systems are never intelligent, and conversely, intelligent music generators are never interactive. The Continuator is an attempt to combine both worlds: interactivity and intelligent generation of music material in a single environment.

## The Continuator System

The Continuator system consists of one MIDI input (typically from a synthesizer) and one MIDI output (typically returning to the same synthesizer). Its operation in the standard mode involves no interface other than the MIDI instrument itself. The user plays musical sequences of any kind, either monophonic, polyphonic, in any playing style. When the phrase is terminated, the Continuator generates a musical phrase in response. This musical phrase has the characteristic of being stylistically similar to the phrases played by the user so far. Technically, it is a continuation of the last input phrase, hence the name of the system.

Although it is difficult to define the notion of *musical style* precisely, we have adopted, like our predecessors in music generation research, a notion of style consisting of the statistical distribution of notes, chords and musical elements in general as well as their ordering. The Continuator, like most of music style replication systems, is based on a Markov model of musical phrases, and the model of the style created by the system retains melodic patterns, harmonic progressions, dynamics and rhythmic patterns of the corpus used for learning. An important consequence of this approach is that the phrases generated by the Continuator are similar but different from the phrases played by the user. The Continuator may therefore be seen as an engine for producing variations of arbitrary musical material.

## Some Musical Examples

To illustrate the working of the Continuator, simple musical examples are given below (see Figure 1 and Figure 2). These examples are noted exactly as they are played, i.e. without rhythmic quantization. They show that the Continuator adapts quickly to arbitrary styles and is able to generate musical material that “sounds like” the user input on a relatively small scale. Issues related to capturing higher- level structure are not discussed here as they are not relevant for our purpose (refer to Pachet 2002a for more details). More sophisticated examples of music created by the Continuator can be found on the web site of the author. The most important aspect of the Continuator is the fact that the musical material generated always conforms stylistically to the input. Also, the Continuator keeps on learning from whatever input is given. As a consequence, the behavior of the system improves over time: if the user produces phrases which are stylistically consistent, but unique, the Continuator will learn more faithfully and will produce musical phrases that are increasingly accurate, with respect to the musical style of the user.



Figure 1. A simple melody (top staff) is continued by the Continuator in the same style (bottom staff).



Figure 2. A simple chord sequence (top staff) is continued by the Continuator in the same style (bottom staff)

## Implementation and Design

There has been considerable research done in the fields of artificial intelligence and information theory regarding the technical issue of learning a musical style automatically in an agnostic manner. Shannon introduced the concept of information based on the probability of occurrence of events in communications (messages) in his seminal 1948 paper (Shannon, 1948). This notion was soon after used to model musical styles, for instance by (Brooks et al., 1957). These early experiments showed that it was possible to create pieces of music that would sound like given styles by simply computing and exploiting probabilities of note transitions. More precisely, given a corpus of musical material (typically musical scores or MIDI files), the basic idea was to analyze this corpus to compute transition probabilities between successive notes. New music can then be produced by generating notes using these inferred probability distributions. One of the most spectacular applications of Markov chains for the generation of music is probably (Cope, 1996), although his musical results are not entirely produced automatically. A good survey of state-of-the-art Markov-based techniques for music can be found in (Triviño-Rodríguez et al. 2001), including in particular variable-length Markov models, which capture stylistic information more finely.

The Continuator system is yet another species in the world of musical Markov systems, although with novel features. In our context, we wanted to learn and imitate musical styles in a faithful and efficient manner and make the resulting mechanism useable as an actual musical instrument. This raised a number of technical issues, whose solutions were progressively integrated in the Continuator.

The architecture of the Continuator consists of two modules: an analysis module and a generator module. The analysis module takes as input MIDI sequences played in real time. The system contains three main parts:

- 1) A phrase end detector, which is able to detect that a musical phrase had “ended”. This detection is based on an adaptive temporal threshold mechanism. The threshold is inferred from the analysis of inter onsets intervals in the input sequence. As a result, if the input sequence is slow (or, rather, contains few notes per seconds) then the threshold is increased, otherwise it is decreased. This simple mechanism ensures that the continuation produced will be seamless, temporally.
- 2) A pattern analyzer. Once detected as completed, these input sequences are sent to a pattern analyzer, which builds up a Markovian model of the sequence. The complete algorithm is described in (Pachet, 2002b), and consists of a left to right parsing of the sequence to build a tree of all possible continuations for all possible prefixes of the sequence. To speed up learning, the system also learns all transpositions of the sequence.
- 3) A *global property* analyzer. Various global properties of the input sequence are also analyzed, such as: the density (i.e. number of notes per second), the tempo and the meter (location of strong / weak beats), the overall dynamics (i.e. loud or soft), etc. These properties are used to produce a continuation which is musically seamless with the input.

The generator is responsible for producing the continuation of the input sequence. The actual production of the musical material exploits the Markovian graph created by the analysis module (Pachet, 2002b). It essentially consists of producing the continuation on a note- by- note basis. Each note is generated using the Markovian probabilities inferred during the analysis stage. Technically it uses a variable- order Markov generation that optimizes the relevance of each single note continuation by looking for the longest possible subsequence in the graph. Special care has been taken to perform meaningful segmentations of the input phrases for the learning phase. Indeed, real- world input phrases are never composed of perfectly successive notes or chords. In order to “cut” input phrases into chunks, which are then fed to the learning system, a segmentation process is able to detect note or chord transitions and possibly cut across unfinished notes. The module also stores the possible “residual” discrepancy, and restores it at generation phase, so that the material retains the rhythmical “naturalness” of the original style.

This continuation sequence is, however, crude, in the sense that it does not necessarily have the global musical properties of the input sequence. Therefore, a mapping mechanism is applied to transform the brute continuation into a musical phrase that will be played just in time to produce seamlessness. Currently, the properties which are analyzed and mapped are tempo, metrical position, and dynamics. More details can be found in (Pachet, 2002a).

## Playing Modes

The basic playing mode of the Continuator is a particular kind of turn-taking between the user and the system determined by three principles:

- 1) Automatic detection of phrase endings. The Continuator detects phrase endings by using a (dynamic) temporal threshold (typically about 400 milliseconds). When a time lapse exceeds this threshold, the Continuator takes the lead, and produces a musical phrase.
- 2) The duration of the phrase generated by the Continuator is parameterized, but in most cases the duration is set to be the same as the duration of the last input phrase.
- 3) Priority given to user. If the user decides to play a phrase while the Continuator is still playing, then the system will stop and return to listening mode (and eventually apply again principle 1).

These principles, in the current implementation, are hard-coded in the system. Moreover, they are set without explicitly telling the users. Experiments with the system has shown that these rules are usually easily learned by the user in an implicit way – the behavior of the system is usually obvious, even for children.

## 3. Experiments with the Continuator

### Experiments with Professional musicians

Apart from the technical issues related to system design and the evaluation of the “stylistic similarity” of the music produced by the system, it soon appeared that there was an important dimension of the project that was more difficult to describe in a standard scientific paper, related to the *subjective impressions* of users playing with the system, or watching it in action.

Indeed, one impressive result of the experiments with professional musicians is that the very use of the system provokes intense

subjective impressions and reactions. Rarely but occasionally, these reactions are negative. For instance a composer from Ircam reacted aggressively against the system as soon as he understood the machine would create (compose) music, and would not even try it. His credo was that machines should only be used to do things humans could not do. Since the Continuator only generated music from human inputs, he therefore considered the system uninteresting and did not need further inspection. Another one considered *a priori* that the system was not interesting because it was not able to capture stylistic information when the user plays only one note: indeed, the Continuator analyses transitions between notes or chords, and a one-note sequence does not contain any transition. The remark was true, but the composer refused to forgive this limitation.

Apart from these two negative and ideological reactions, all the other reactions were very positive. The Continuator captures the attention of the audience beyond the traditional “demo effect” of many computer music presentations. In particular, a systematic *Aha* effect was noticed (Pachet 2002b) for professional users as well as beginners. *Aha* effects have been introduced in experimental psychology to characterize sudden moments of realization, understanding or inspiration. They are also being studied in neuroscience, where evidence of neuronal activity corresponding to *Aha* effects has been shown (Mogi, 2003).

Experimentation with the system invariably induces users to reflect on their own musical personality (Pachet, 2002c). Bernard Lubat, a Jazz pianist and drummer, at the forefront of progressive Jazz in Europe played with the system many times. He evoked with great precision (in particular during performances at Uzeste festival and during a concert at Ircam in October 2002, see Figure 3) how the system would speed up his own evolution in improvisation, allowing him virtually to “play ahead of his current thinking”. György Kurtág Jr., a composer and improviser, described the Continuator as a kind of “amplifying mirror”, and his regular use of the system throughout the year 2001 changed his way of improvising and composing music (this collaboration resulted in a composition performed at the Vienna Festwochen 2002 music festival, “The Hollow of the Deep Sea Wave »).



Figure 3. Bernard Lubat during his performance. At some point, he “launches” the Continuator with a characteristic gesture. Later, he makes gestures in a “pretend play” mode while listening to the music produced by the Continuator.

### Experiments with Children

When the first author started playing with his daughter and the Continuator (she was 3, see Figure 4), her positive reaction was quite surprising and significant. She started to become interested in playing the keyboard – she would laugh at the system’s “answers” and was able to focus her attention longer on musical playing.



Figure 4. A 3 year- old child playing with the Continuator at home.

The idea to push the experiment further with additional children was therefore quite natural. In most schools, music is still taught using outdated methods, in which children are confronted with formalisms before they have experienced the enjoyment of playing or listening to music. With a system like the Continuator, basic playing capabilities might be learned more easily, and earlier than with standard music education practice (piano lessons usually start at the age of 6 at the earliest in most conservatories). Most importantly, the Continuator – or in a general a class of systems able to learn and react - could develop a genuine *desire for music* in children, and consequently prepare them for traditional classical training in a more productive way. This need to bring more fun and interactivity in the classroom has long been advocated by various psychological studies (Webster, 2002; Delalande, 1984). More precisely, our vision of education as a pleasing experience falls within the boundaries of studies concerning *Flow* (Csikszentmihalyi, 1990), as discussed in the last section of this paper.

Preliminary experiments with children took place at a French kindergarten (Ecole Bossuet Notre-Dame in Paris). Later, more systematic experiments, involving in particular crossed experimentations, were conducted under the direction of Anna-Rita Addessi (University of Bologna) to investigate the issues brought to light in the first set of experiments further. A detailed analysis of these sessions from the perspective of experimental psychology is under way (Addessi & Pachet, 2004). Only the key observations are mentioned here.

The goal of these experiments was to test basic hypotheses about the effect of the Continuator on the playing abilities of 3 to 4 year-old children. More precisely, the following hypotheses were made:

1 – *Increase in attention span*: Children can play longer with the Continuator than just a keyboard alone.

2 – *Aha and surprise*: The Continuator produces noticeable Aha effects with children, as is the case with professional musicians (note that children rarely find things surprising, a flying pianoforte would probably not surprise them any more...).

3 – *Autonomy*: The Continuator motivates children to play music. For instance, a child, knowing that the Continuator is there, may express the intention of playing alone more often than with standard musical instruments.

4 – *Exploration and playing modes*: The Continuator pushes children to explore new playing modes. The results of the experiments with a normal piano suggested that children would usually stick to single playing modes, including playing with only one finger, playing clusters with two hands, playing ascending or descending diatonic scales (i.e. white keys), etc.

5 – The Continuator can develop various kinds of *attachment behaviors* in children, similar to what has been observed with the Tamagotchi or Aibo (Kaplan, 2001).

These experiments consisted of several sessions. Each child (3 to 5-year-old) was invited to play with a keyboard (a Korg Karma with piano sounds and no additional effect connected to a pair of amplified loudspeakers). The Continuator was set to a mode that played phrases in the same style as the child, and of approximately the same number of notes as the input phrase. The threshold for triggering the continuation was set to about 500 ms.

The child was left alone with experimenters in a familiar classroom. The protocol consisted in two phases: The child would first be told to play with the keyboard as he/she wanted, with no particular instruction. When the child stopped playing or express significant boredom, he/she would be told that the system would now try to play with him/her. At that point, the Continuator would be turned on. The session stopped when the child stopped playing. Several sessions involved also two or more children. Children were brought in one at a time by their teacher. The sessions were recorded with a video camera. A certain number of interesting points were observed:

- Aha effects were indeed produced and noticeable. Reactions of the children ranged from enhanced attention to surprise when the Continuator started to play initially.
- The Continuator did seem to augment the attention span of most of the children. On two occasions, I had the session had to be terminated because the duration exceeded 40 minutes. There were two exceptions: one child was very tired and played very little, with or without the Continuator. Another child seemed to enjoy playing the keyboard enormously with or without the Continuator.
- The children who engaged in long interactions (one of them played 30 minutes and had to be stopped) also appeared to develop the ability to listen with great attention.
- The interaction mode of the Continuator (stopping when the child plays, and playing when the child stops) induced “turn-taking” behavior from the children, without explicit directions. The emergence of turn taking in such a context is not a trivial phenomenon, as shown for instance by the works on complex dynamic systems by Ikegami (Iizuka and Ikegami, 2002).
- Some children exhibited a wide variety of playing modes. Apart from the classical playing modes (e.g. playing isolated notes, chords, arpeggios) the children invented new playing modes such as playing with the sleeves, kissing the keyboard, playing from behind, playing with the palm, etc. Some of them also theatrically accompanied the launching of the Continuator with particular gestures like raising hands at the end of a musical phrase (see Section 4.5.3).

## The Continuator as a Flow machine

The Continuator appeared in fine as a machine that promotes musical enjoyment in various forms. The system's ability to maintain children's attention for long periods of time - remarkable for this age group - and in general its ability to attract and hold the attention of users of all ages can be interpreted through the theory of Flow introduced by psychologist Mihaly Csikszentmihalyi (Csikszentmihalyi, 1990). Csikszentmihalyi's notion of Flow describes the so-called optimal experience as a situation in which people obtain an ideal balance between skills and challenges. Two emotional states of mind are particularly stressed in this theory: anxiety, obtained when skills are clearly below the level needed for the challenge, and boredom, when the challenges are too easy for the skill level. In the middle lies Flow. Other states can also be described in terms of balance between skills and challenges (see Figure 5). We can think of the Continuator as a Flow machine in the sense that it produces by definition a response corresponding to the skill level of the user. This approach also allows for the progressive scaffolding of complexity in the interaction, which is not the case for most pedagogical tools designed with a fixed pedagogical goal in mind.

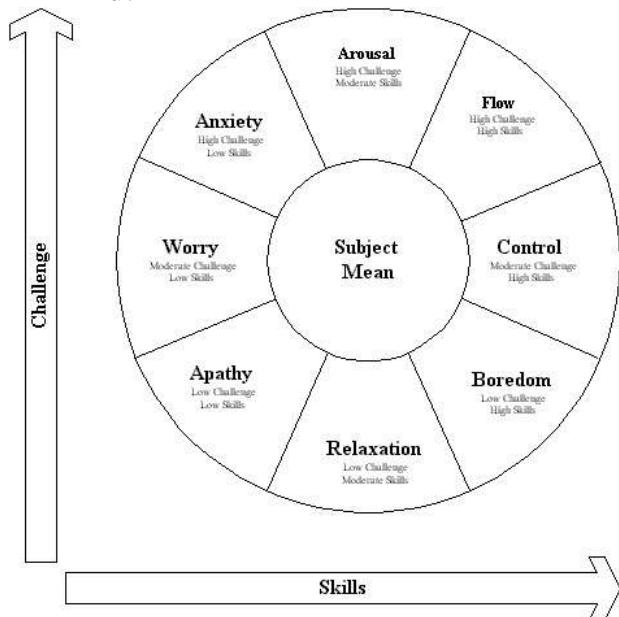


Figure 5. Csikszentmihalyi's Flow diagram describes various emotional states such as boredom or anxiety according to the relation between skills and challenges for a given activity (From Delle Fave, this volume)

More precisely, Csikszentmihalyi describes the state of Flow as consisting of several fundamental traits where the balance between challenges and skills is probably the most important. The other traits can be discussed in light of the experiments conducted:

*Focused attention.* The experiments show clearly that children are engaged in focused activity both when both playing and listening. The ability to listen and concentrate for several seconds and listen to music is remarkable at this age. As pointed out by Carla Rinaldi (Rinaldi, 2003),

listening is probably one of the most important abilities of children to discover the world around them. In her view, children are researchers constantly making up theories about the world and evaluating them. With the Continuator, we observed this phenomenon rather systematically, i.e. children engaging in deep, concentrated listening of the effect of their playing on the system. Figure 7 shows several examples of this phenomenon.

Another typical situation encountered in sessions involving two children was the phenomenon of joint attention. More precisely, one of the children would force the other to stop playing to listen in order to the system. This situation, which we call “aspetta” (the Italian word for “wait”), is illustrated in Figure 8. Of course this behavior was not observed with professional musicians, who so far experimented with the system alone.

*Ease of concentration.* This is particularly clear given the fact that no instruction is given to the children whatsoever. They play with the system in a self-motivated way, without any external constraints.

*Clear-cut feedback.* The Continuator produces clear feedback (in fact this is the only thing it does). The interaction in some sense is reduced to the analysis of the feedback produced by the machine.

*Control of the situation.* Children are in control of the situation most of the time. They understand quickly that they can interrupt the system whenever they want. The limitations in control are due to the difficulties that may arise when interpreting some of the system's outputs (see example in the next section).

*Intrinsic motivation.* The most striking result of the experiments is related to the intrinsic motivation of the children, who were not told anything about the rules of the system.

*Excitement.* Excitement is clearly shown most of the time in particular in the early phases of the sessions. We separate here excitement from surprise in the sense that the surprise effect is most often short in duration, whereas the excitement phase lasted much longer, sometimes for 20 minutes or more. Excitement was observed in most of the cases. Interestingly, the children were excited mostly by what the system was playing, rather than by what they were doing. Figure 6 shows some expressions of this excitement.

*Change in the perception of time and speed.* A systematic study concerning interaction times is under way, but it is clear already that at least for some of the children time did pass very quickly: some sessions had to be terminated by the experimenters when the time limit was reached.

There is, however, one Flow characteristic that does not apply directly to the Continuator experiments:

*Clear goals.* No goal was given explicitly to the children except to play until they were bored. Indeed, improvisation is generally not goal-oriented. Similarly, web sailing is usually non goal-oriented but many claim that it can nevertheless create Flow (King, 2003). More importantly, It can be argued that children did create spontaneously goals during their interaction. For instance, several sessions involved children trying to push the system to replicate a particular frantic musical style that they had played some minutes ago.



Figure 6. Various expressions of musical excitement. Excitement is mostly provoked by listening to the system, rather than by actually producing music.



Figure 7. Various gestures showing listening and concentration.



Figure 8. « Aspetta »: when one child forces the other to stop in order to listen to the machine.

### Phases to Reach Flow

In order to precisely define the role and importance of the Continuator in developing musical abilities, the phenomena described above deserve careful analysis on an individual basis (e.g. attention span, listening and concentrating, etc.). However, It is also worth studying the "life cycle" of these phenomena, or in other words, the progression of these Flow traits over time. We have observed that they always progress in the same order (see Figure 9).

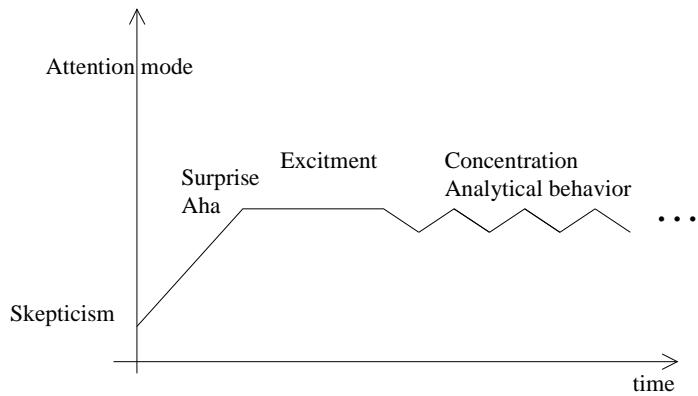


Figure 9. A tentative sketch of the “life cycle” of the interaction mode with the Continuator.

One of the most interesting aspect of this time line is the role of Surprise/Aha: this occurs only once, at the beginning of sessions, or shortly after, and it may be argued that these Aha constitute a sort of *initiatic* experience, acting as a phase transition: once this step is performed, the users change completely their attitude, both towards the system and towards the environment (the experimenters). They become at once involved and self interested.

### Observations Concerning Interaction Modes

During the analytical stage (i.e. after the initial periods of surprise and excitement), many other behaviors were also observed with clear indicators. For instance, several children developed spontaneously innovative playing modes. Besides rediscovering standard playing modes such as playing individual notes, arpeggios or chords, they would produce sometimes remarkable arpeggios and clusters, but also new modes such as playing with elbows, turned around with hands in back, kissing the keys with their lips, etc.).

An interesting phase in the interaction occurs when children, after having mastered the basics of the system, somehow abstract the concept of a “musical phrase”. This is indicated by a typical gesture demonstrating a pretend “launching” of their own musical phrase (as if it were a golf ball). This gesture determines the end of their musical phrase and also creates the expectation of the system’s response. It is remarkably similar to what professionals do themselves (see Figure 3, Bernard Lubat performing at Ircam and Figure 10). This gesture can also be interpreted as a desire to pretend- play, as described e.g. by (Wynants, 2004).



Figure 10. Children raising their hands and virtually « launching » the Continuator, after finishing a musical phrase.

Another phenomenon worth mentioning – and worth studying more in depth – is the children’s ability to engage in turn-taking behaviors in a spontaneous way. As mentioned before, the “rule of the game” (in this case a particular kind of master/slave turn-taking) was not explained to the children. However, they learned it, somehow implicitly, extremely quickly, after a few interactions.

This ability to learn the rule of turn making is nontrivial. In particular, a child who knows how it works must also have an understanding of beginnings and endings of musical phrases: of his, as well as the system’s. A few cases showed that this skill in ascribing an intentional ending to phrases generated by the system develops very early in the sessions. Figure 11 shows a typical scenario in which a child is confronted with an interesting situation regarding phrase endings. In this case, a particular continuation produced by the system started with the ending of the child’s phrase followed by additional notes. This situation, although rare, can occur because the Continuator has no particular notion of beginnings or endings. Figure 12 shows the situation graphically. The phrase played by the child is schematically represented by two consecutive chunks of a few seconds. The last chunk is repeated by the Continuator, and followed by another one.

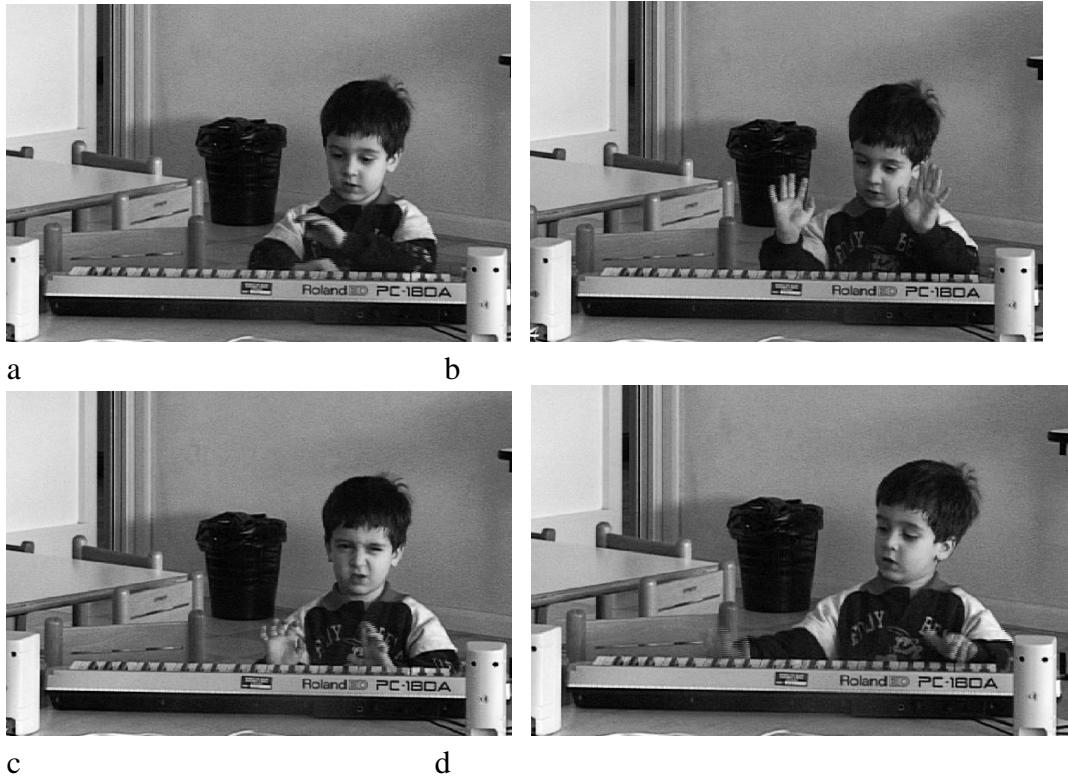


Figure 11. A child clearly showing its interpretation of the Continuator's phrase endings. (a) shows him just finishing a phrase and waiting for the answer. The answer generated by the Continuator turns out to contain the last part of his input phrase followed by some additional notes. At the end of the first part of the continuation (i.e. the repetition of the child's phrase ending) the child gets ready to play again, assuming that the phrase played by the Continuator is finished. It is not (c), and the child shows his misunderstanding with a facial expression. Eventually (d), the Continuator ends his phrase, and the child resumes playing.

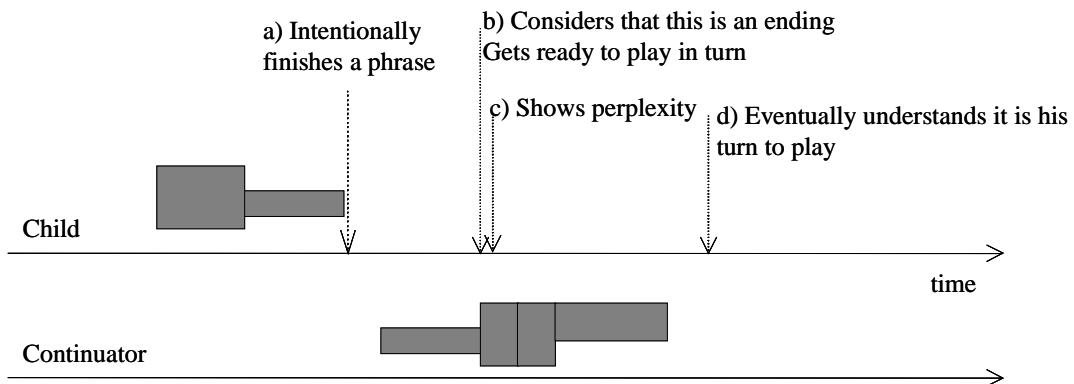


Figure 12. A short scenario demonstrating the complexity of the interpretation of phrase endings. The phrase structure is represented schematically using rectangle of various shapes. The important aspect of the scenario is that the last chunk of the input phrase is the first one of the continuation, thereby creating a false sense of ending for the child.

## 4. Designing Interactive Systems

The experiments conducted with the Continuator are interesting because they teach us something about the nature of human behavior, and in particular about what human beings find exciting. They also give us precious hints for the design of artificial systems that help to generate flow experiences. I now highlight two key ideas which appear crucial for the success of the Continuator: reflection and emergent interaction protocols.

### Reflection in Interactive Systems

A lot of research has been conducted into the properties that make interactive software usable. In the field of music, Sidney Fels (2002) analysed several interactive systems designed for creating novel aesthetic experiences and found that an ideal interactive system requires intimacy. This is caused by the sensation that the system is somehow an extension of the user's body, thus ensuring the efficiency of the user's actions with the instrument.

The Continuator has this property, but the presence of a learning component introduces another dimension of interaction. Because the system is able to learn and imitate the user's musical personality and style, the Continuator acts as a dynamic mirror, and I claim that most of the interesting behaviours observed in the experiments stem from this feature.

This reflective capability brings a number of crucial characteristics:

- *Similarity or Mirroring effect.* What the system produces "looks" like what the user herself is able to produce.
- *Agnosticity.* The system's ability to reproduce the user's personality is learned automatically and agnostically i.e. without human intervention. In our case for instance, no pre-programmed musical information is given to the system whatsoever.
- *Incrementality.* Interactive systems are not designed only for short demos. Because the user is constantly interpreting the output of the system, and altering his playing in response, it is important to consider the longer term behaviour of the system. Incremental learning ensures that the system keeps evolving all the time. Each interaction with the system contributes to changing its future behavior. Incremental learning is a way to endow the system with an organic feel, typical of open, natural systems (as opposed to pre-programmed, closed-world systems). [A preliminary version of the Continuator used an algorithm based on Lempel-Ziv for learning. The Lempel-Ziv algorithm is not complete: several learning steps are needed to ensure that all the information contained in a new sequence is actually learned. As a

consequence, the system would be less faithful (would learn more slowly), but this very slowness increased the organic feeling of teaching something, which is less the case in the current version (which learns at once all the information contained in each new sequence): paradoxically, to ensure the incrementality, less complete algorithms may be more efficient than complete ones.]

- *Seamlessness*. The system produces material that is virtually indistinguishable from the user's input. Note that this characteristic does not apply in the case of "classic" hyper-instruments, where the sonic effects are entirely produced through the system, and therefore not resulting in any material directly produced by humans.

One important consequence of reflection is that the "user's center of attention" is not directed towards the *end-product* (the music), but towards the *subject* engaged in the interaction. Engaging in an interaction with such a system is therefore a means of discovering oneself, or at least exploring one's ability in the domain at hand (in our case, musical improvisation). This natural, deep interest in exploring oneself - particularly during the early childhood years of childhood - may explain why the Continuator is so appealing and we can design probably other successful interactive systems using the same principles. In some sense, these systems are an extension of the "second self" (Turkle, 1984). Not only does the machine seem to "think," but it thinks like the user. An interesting consequence of this is the reversal of roles: the student becomes the teacher who teaches the machine about himself.

### **Emergent Interaction Protocols**

The experiments have also shown various limitations of the current design of the Continuator. One limitation concerns the need for another very general principle, namely that interaction protocols need to be flexible and emergent. Interaction protocols specify the rules of the game. They determine how and when the system decides to play. Like in conversations, these rules can vary, and a simple question/answering sequence is by far not the only possible interaction protocol: lectures, smalltalk (in the common sense meaning), exams, group conversations, baby talk, etc. are different situations in which interaction rules differ greatly.

The issue of interaction protocols is closely related to the idea of music as a conversation, put forward for instance by Bill Walker in his *ImprovisationBuilder* system (Walker, 1997; Walker & Belet, 1999). The *ImprovisationBuilder* is able to take turns with the player, and also to detect, in the case of collaborative music playing, whose turn it is, using a simple analysis of the various musician's inputs. However, this system as well as all the music interaction systems so far, are based on fixed interaction protocols. The question here is whether it is possible to find

mechanisms so that the machine adapts protocols automatically and finds the best protocol suited to the context.

Currently, several interaction protocols were designed and experimentally tested with the Continuator. Here are some of them, by increasing order of complexity (and represented graphically in Figure 13).

- *Turn-taking*. This mode is represented graphically as a perfect succession of turns, with no gap. The Continuator detects phrase endings, then learns and produces a continuation. It stops as soon as the user starts to play a new phrase.
- *Turn taking with delay*. The same as above, except that the Continuator stops only when the user *finishes* a phrases. This produces an interesting overlapping effect in which the user and the Continuator can play at the same time.
- *Single note accompaniment*. The Continuator produces an appropriate chordal accompaniment each time a note is played, and with the same duration. It stops the chord when the key is released.
- *Phrase-based accompaniment*. The same as above except that the chord is produced only at the beginning of a phrase.
- *Collaborative*. In this mode, the Continuator plays an infinite stream of music (based on material previously learned). The user can play simultaneously, and what he/she plays is taken into account by the Continuator, e.g. harmonically. The user's actions are here like high-level control, instead of a question to be answered.

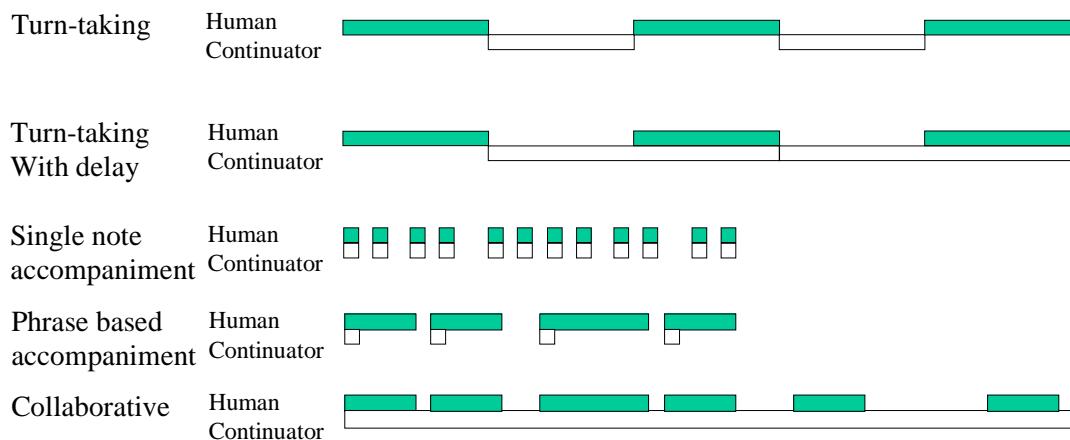


Figure 13. Various interaction protocols with the Continuator.

These various modes are in turn highly parameterised: The phrase length of the continuation in turn taking mode, the rhythm mode, the adaptation or not of the music produced to surface parameters such as dynamics, tempo, etc. In practice, it is easy to see that an infinite number of concrete interaction protocols could be defined, all tailored to a particular situation.

The main problem with these protocols is that they are fixed, therefore sometimes not appropriate. Several observations (with children and professionals) showed that this rigidity is a severe limitation:

- For instance, a child enters the room and remains seated looking at the keyboard. Nothing happens because the Continuator waits for a phrase before it starts playing itself. In this case, an initiative by the system so as to trigger an interaction would be more relevant than just waiting.
- During an interaction, the parameters of a chosen protocol may be unadapted. For instance, we have observed that short answers in early stages of the interaction are preferable, but longer ones are more adapted as the user gets involved in the playing. Conversely sometimes a continuation which is too long can confuse the user (as was the case with one of the children who did not know how to stop the system).
- During a session, different modes may be needed. For instance, while the turn-taking mode appears to be very natural at the beginning, several users (children and professionals) expressed the desire to play at the same time as the system. This requires indeed an explicit change of interaction mode (e.g. with a button), which breaks the fluidity of the interaction, and hence breaks the flow experience.
- Parameters other than the simple musical inputs should also be taken into account to influence the interaction protocol. The child's face in Figure 11 is a clear indication that the interaction was not as expected.

The need for the system to control its own interaction protocol is therefore crucial to enhance its capacity to adapt. More generally, a fully flexible system should be able, like humans, to infer on-the-fly the protocol induced by a particular situation, and follow it, as well as invent itself new protocols: a good music teacher would for instance switch from turn taking to, say, a teaching mode, and back to an accompaniment mode, etc.

To achieve such a control we need to make the system somehow *self-reflective*, i.e. able to manipulate explicitly the interaction protocol at a meta-level: select a protocol, create one, or infer one from the context. Such an endeavour is usually referred to as *reflection* in the world of language design, object-oriented in particular (Demers and Malenfant, 1995).

Building systems with such an architecture is not so simple. Successful attempts at designing self-reflective systems have usually been restricted to so-called structural reflection, i.e. the ability of a program to reason about its data structures. Conversely, systems able to switch between protocols (e.g. for mobile services) are able to do so only within a fixed list of predefined protocols.

To extend flexibility in protocols requires the representation of the intentionality in interactions. The speech act theory proposed by Searle (1969) allows to represent utterances as so-called illocutionary acts, such as stating, requesting, commanding, and so forth. Every speech act consists of an illocutionary Force  $F$  applied to a proposition  $P$ . (Covington, 1998). This theory has given birth to a number of languages such as KQML (Knowledge Query and Manipulation Language), which are used for specifying various interaction protocols between agents, notably for electronic commerce (Finin et al, 1994). However, speech act theory and KQML are primarily aimed at facilitating knowledge sharing and transfer, in contexts where the set of possible illocutionary forces is known *a priori*. A language such as KQML may be used to specify each interaction between the machine and the user. But KQML does not solve (or even address) the motivational issue nor the problem of interpreting the “messages”, or inferring optimal protocols.

## 5. Conclusion

We have introduced the bold question of what makes an interactive system appealing, interesting and attention grabbing, taking example from the experiments performed with the Continuator system, a musical interactive system. We have shown that the Continuator may be seen as a particular example of a Flow machine, i.e. a machine which generates flow-like experiences in users. The reflection principle, which causes a mirroring of user’s behavior after a learning process, appears to be of critical importance.

Current limitations concern primarily the lack of flexibility in interaction protocols. The Continuator proposes many different interaction modes, but is not able to detect automatically the best mode, nor to create new ones on-the-fly for adapting to unexpected situations. This is clearly an important research line for the future which can build further on a long history of research in computational reflection. Reflection for building Reflective Systems: although it seems a play on words, I believe this is not entirely coincidental.

## References

- Addessi, A.-R. & Pachet, F. (2004) "Musical style replication in 3/5 years old children: experiments and analysis", submitted to British Journal of Music Education.
- Brooks, Hopkins, Neumann & Wright. "An experiment in musical composition." *IRE Transactions on Electronic Computers*, 6(1), 1957.

Chuang, John. (1995). Mozart's Musikalisches Würfelspiel, web site: <http://sunsite.univie.ac.at/Mozart/dice/>

Cope, David. (1996). Experiments in Musical Intelligence. Madison, WI: A-R Editions, 1996.

Covington, M.A. (1998) "Speech acts, electronic commerce, and KQML," *Decision Support Systems* volume 22.

Csikszentmihalyi, Mihaly. (1990) Flow. The Psychology of Optimal Experience. Harper and Row, New York.

Cope, David. (1996). Experiments in Musical Intelligence Madison, WI: A-R Editions.

Delalande F. (1984), La musique est un jeu d'enfant, Buchet/Chastel, Paris.

Demers, François- Nicolas and Malenfant, Jacques (1995). Reflection in logic, *functional and object- oriented programming: a short comparative study*. In IJCAI '95 Workshop on Reflection and Metalevel Architectures and their Applications in AI, pages 29- -38.

Fels, S., (2000). Intimacy and Embodiment: Implications for Art and Technology, Proceedings of the ACM Conference on Multimedia, Marina del Sol, CA, pp. 13- 16, Oct.

Finin, Tim Fritzson, Richard McKay, Don McEntire, Robin (1994). KQML as an Agent Communication Language Proceedings of the 3rd International Conference on Information and Knowledge Management (CIKM'94).

Iizuka, Hiroyuki and Ikegami, Takashi. (2002) *Simulating Turn- Taking Behaviours with Coupled Dynamical Recognizers*, The 8th International Conference on the Simulation and Synthesis of Living Systems, University of New South Wales, Sydney, NSW, Australia 9th- 13th December.

Ingalls, Daniel (1981). Design Principles Behind Smalltalk, BYTE August 1981 Special Issue on Smalltalk.

Kaplan, Frédéric. (2001). Artificial Attachment: Will a robot ever pass Ainsworth's Strange Situation Test?. in Hashimoto, S., editor, Proceedings of Humanoids 2001: IEEE-RAS International Conference on Humanoid Robots, pages 125- 132.

Kay, Stephen. (2000). The Korg Karma music work station, Korg Inc.

King, Andrew (2003). Speed Up Your Site: Web Site Optimization. Chapter 2: Flow in Web Design.

Mogi, ken. (2003). Onceness, in *The Future of Learning*, Tokoro, M. and Steels, L. Eds, IOS Press.

Pachet, Francois. (2002a) The Continuator: Musical Interaction with Style. In ICMA, editor, Proceedings of ICMC, pages 211- 218, September 2002. ICMA. best paper award. Extended version in Journal of New Music Research, 31(4), 2003.

Pachet, Francois (2002b) Interacting with a musical learning system: the Continuator.

In C. Anagnostopoulou, M. Ferrand, A. Smaill, editor, *Music and Artificial Intelligence, Lecture Notes in Artificial Intelligence* (vol. 2445), pages 119- 132, Springer Verlag. September 2002.

Pachet, François. (2002c) Playing with virtual musicians: the Continuator in practice, IEEE Multimedia, April- June 2002.

Rinaldi, Carla (2003) The Joys of Preschool Learning, in the *Future of Learning: Issues and Prospects*, Tokoro, M. and Steels, L. Eds, IOS Press

Risset, Jean- Claude and Van Duyne, Scott. (1996). Real- time Performance Interaction with a Computer- Controlled Acoustic Piano, CMJ 20(1), pp. 62- 75.

Searle, J. R. (1969), *Speech Acts. An Essay in the Philosophy of Language*. Cambridge.

Shannon, C. E. (1948). "A mathematical theory of communication", *Bell System Technical Journal*, vol. 27, pp. 379- 423 and 623- 656, July and October, 1948.

Steels, L. (2004) Autotelic Agents: Robots that have fun. In: Hafner, V., F. Iida, R. Pfeifer and L. Steels (eds.) (2004) *Embodied AI*. Proceedings of the Dagstuhl Seminar. Lecture Notes in Computer Science. Springer Verlag, Berlin.

Triviño- Rodriguez, J. L. Morales- Bueno, R. (2001) Using Multiattribute Prediction Suffix Graphs to Predict and Generate Music, Computer Music Journal 25 (3) pp. 62- 79.

Turkle, S. (1984). The second self: Computers and the human spirit. New York: Simon and Schuster.

Walker, W. (1997) *A computer participant in musical improvisation*. In CHI 97 Electronic Publications.

Walker, W. Belet, B. (1999) Applying ImprovisationBuilder to interactive composition with Midi piano. In ICMC 99, Hong- Kong.

Webster, Peter R. (2002). Computer- Based Technology and Music Teaching and Learning, in R.Cowell- C. Richardson (EDS). The New Handbook of Research on Music Teaching and Learning, Oxford University Press, pp. 416- 439.

Wynants, M. (2004). Pretend Play as Learning: Case Studies from the Home. This volume

# Coordination, cognition, auto - référence et réflexivité aux fondements des « croyances sociales »

Denis Phan\*

\* Université de Rennes 1 – CREM – UMR CNRS 6211

Communication aux 12<sup>es</sup> Journées de Rochebrune, Rencontres interdisciplinaires sur les systèmes complexes naturels et artificiels : «Réflexivité et auto-référence dans les systèmes complexes», Megève, 24-28 Janvier 2005

## Résumé

L'objectif de ce papier est d'explorer les aspects cognitifs de la réflexivité des représentations dans la résolution de certains problèmes de coordination formalisés au moyen de la théorie des jeux, dans le but de définir ultérieurement une ontologie pour la modélisation multi-agents des « croyances sociales » dans les jeux de coordination. Dans une première section, on utilise plusieurs définitions formelles relatives à une dimension « sociale » des croyances, proposées par André Orléan (2002, 2004). On cherche ensuite à en évaluer la signification dans le cadre d'une confrontation entre l'approche « standard » de la théorie des jeux et l'approche « comportementaliste » (ou « expérimentale »). Deux situations particulières de coordination sont passées en revues. Dans les deux cas, la réflexivité des croyances est en question.

Dans les jeux de « pure coordination », les joueurs doivent se coordonner sans communication préalable (jeu non-cooperatif). Une caractéristique de ces jeux du point de vue de l'approche standard est la multiplicité des équilibres de Nash (de deux à une infinité, selon le jeu considéré). Du point de vue comportemental, le problème du « point focal » réside alors dans la nécessité de mobiliser des connaissances extérieures au cadre « standard » de la théorie afin de servir de référence « saillante » dans la sélection des équilibres. La notion de « croyances sociales » introduite par Orléan permet une interprétation féconde de ces résultats. Dans le « concours de beauté », les joueurs sont en concurrence pour évaluer un multiple ce que sera l'évaluation moyenne de l'ensemble des joueurs. Il n'y a, selon l'approche standard qu'un seul équilibre de Nash. Les résultats expérimentaux montrent que cet équilibre n'est jamais atteint dans le cadre d'un jeu unique. La réflexivité se manifeste d'abord à travers la question centrale de savoir ce que va produire en moyenne le collectif des joueurs. On est alors amené à poser la question « de second niveau » sur le rôle de la croyance des autres joueurs sur la réponse à cette question et ses incidences sur leurs choix. Ce bouclage récursif pourrait en théorie être infini, pourtant, l'économie expérimentale nous montre qu'il n'en est rien dans les raisonnements effectifs. Pour conceptualiser ces dimensions récursives du processus cognitif, on utilise les notions de « profondeur stratégique » et de « hiérarchie cognitive ». A un premier niveau, il s'agit d'évaluer le nombre d'étapes de raisonnement récursifs dont est capable un joueur. A un second niveau, il s'agit d'évaluer la capacité de raisonnement récursif qu'un joueur attribue aux autres joueurs, etc....Ici encore la notion de « croyances sociales » permet une réinterprétation intéressante du phénomène.

## Introduction

L'économie connaît un grand nombre de situations caractérisées par des références croisées, mais des problèmes particulièrement intéressants se posent lorsque l'on s'intéresse aux *croyances* des acteurs en situation d'interactions "stratégiques". Par

---

\* l'auteur remercie Frédéric Amblard et Jean Louis Dessalles pour leurs commentaires constructifs sur une version préliminaire de ce papier.

exemple, Keynes a utilisé la métaphore d'un "concours de beauté" pour mettre en évidence la nature auto-référentielle du comportement des spéculateurs :

« la technique du placement peut être comparée à ces concours organisés par les journaux où les participants ont à choisir les six plus jolis visages parmi une centaine de photographies, le prix étant attribué à celui dont les préférences s'approchent le plus de la sélection moyenne opérée par l'ensemble des concurrents. Chaque concurrent doit donc choisir non les visages qu'il juge lui-même les plus jolis, mais ceux qu'il estime les plus propres à obtenir les suffrages des autres concurrents, lesquels examinent tous le problème sous le même angle. Il ne s'agit pas pour chacun de choisir les visages qui, autant qu'il peut en juger, sont réellement les plus jolis, ni même ceux que l'opinion moyenne considérera réellement comme tels. Au troisième degré où nous sommes déjà rendus, on emploie ses facultés à *découvrir l'idée que l'opinion moyenne se fera à l'avance de son propre jugement*. Et il y a des personnes, croyons nous, qui vont jusqu'au quatrième ou au cinquième degré ou plus loin encore. » (Keynes, 1936, Chapitre 12 – V, souligné par nous)

Parmi les nombreuses questions soulevées par ce texte célèbre de Keynes, on notera l'idée des *niveaux* (les « degrés ») dans la « profondeur » du raisonnement, et celle de la focalisation sur un *objet intermédiaire*, « l'opinion moyenne » dont le statut ontologique (au sens philosophique) apparaît problématique par construction. C'est précisément cet « objet », système psycho-social complexe résultant directement des interactions entre les participants (et indirectement de la confrontation de leurs croyances dans leurs esprits) qui manifesterait des propriétés de « *réflexivité* ». En effet, les propriétés dynamiques de l'agrégat qui le caractérise fonctionnent « *comme si* » le système lui-même disposait *par lui-même* des capacités de s'auto-observer et de s'auto-modifier. Comme cet « objet » (qui est imparfaitement défini) n'a pas d'existence matérielle (et donc pas à proprement parler de « capacités »), la question de sa *nature* et de sa *manière d'être* doit être posée.

Si l'on se déplace au niveau formel, la théorie des jeux utilise largement des formes variées d'auto-référence et de réflexivité, tant au niveau de la rationalité des joueurs que de celui des conditions dans lesquelles les choix sont effectués (i.e. l'hypothèse de « *connaissance commune* » qui sera discutée plus loin). Mais la théorie standard ne recourt pas à des croyances « *sociales* » au sens où nous les entendrons ici. C'est précisément pour cela que nous avons choisi d'aborder la question de la possible réflexivité des croyances portant des « *objets sociaux* » à travers le filtre de la théorie des jeux (évidemment réducteur, il permet cependant de clarifier la réflexion).

Plus spécifiquement, l'objectif de ce papier est d'explorer les aspects cognitifs et réflexifs des représentations « *sociales* » dans la résolution de certains problèmes de coordination formalisés au moyen de la théorie des jeux. L'objectif ultime, qui n'est pas traité ici, étant de proposer une « *ontologie* » (au sens informatique discuté plus loin, distinct du sens traditionnel de la philosophie) pour la modélisation multi-agents des « *croyances sociales* » dans les jeux de coordination (Ferber, Phan, 2005)

La notion informatique d'*ontologie* ne faisant pas l'objet d'une définition canonique ni même d'un consensus, nous retiendrons d'abord la définition large d'un « *modèle d'organisation de la connaissance* ». Dans le cas qui nous intéresse la « *connaissance* »

repose elle-même sur une famille de modèles, méthodes, expérimentations et discours « savants ». On pourra alors parler de « *modèle d'organisation et de conception de la connaissance* ». Comme notre objectif premier reste la modélisation multi-agents des « croyances sociales » dans les jeux de coordination, il est utile de restreindre cette définition à une représentation rapidement opérationnelle dans ce dernier domaine, liée aux *modèles de conception* « *objet* » et « *agent* ». On retiendra donc l'idée *d'un modèle de représentation formelle d'un système de concepts articulés*, qui fasse sens à la fois pour le théoricien des jeux, le « *thématicien* » ou le théoricien en sciences économiques ou sociales, et l'informaticien. Ce système de concepts articulés devrait permettre à la fois d'organiser les connaissances existantes et d'en concevoir de nouvelles : notre *ontologie* serait donc ainsi elle-même dotée de propriétés réflexives ! Plus spécifiquement, il serait formalisé par des « *objets* » structurés par des relations de *composition* et *d'héritage*. Les « *objets* » ainsi définis seraient encore caractérisés par des types d'*attributs* et types de *relations* entre objets, l'ensemble de cette construction étant interprétable à tous les niveaux dans une *sémantique spécifique*. Le présent travail ne prétend pas construire une telle *ontologie*, mais cherche plus modestement à « mettre à plat » certains éléments du débat, du point de vue du théoricien en sciences économiques ou sociales, pour initier le dialogue avec les informaticiens.

Dans une première section, on utilise plusieurs définitions formelles relative à une dimension « sociale » des croyances, introduites par Orléan (2002, 2004). On cherche ensuite à proposer une interprétation dans le cadre d'une confrontation entre l'approche « standard » de la théorie des jeux et l'approche « comportementaliste » - ou « expérimentale » (Camerer, 2003). Deux situations particulières de coordination sont passées en revue : les jeux de « pure coordination » et le « concours de beauté ». Dans les deux cas, la réflexivité des croyances est en question. En effet, l'analyse réflexive du problème sous la forme récursive associée à la théorie standard, conduit à des résultats qui ne correspondent pas à ceux que l'on obtient par expérimentation. Une première notion de « *hiérarchie cognitive* » permet alors de conceptualiser une forme stylisée de procédure réflexive limitée qui permet aux agents de sortir de l'aporie liée à la récursivité généralisée. L'interprétation de ces procédures à l'aide d'une notion spécifique de « *croyance sociale* » proposée par André Orléan permet de mettre à jour des perspectives nouvelles. Dans les deux jeux considérés, la connaissance qu'ont les joueurs du « *background culturel* » des autres joueurs, peut faciliter l'attitude cognitive qui consiste à « se mettre à la place de l'autre », sans empathie (la « *décentration* ») pour éliminer des solutions qui apparaissent alors comme peu (ou pas) pertinentes. Cette attitude peut être considérée comme un *réducteur d'incertitude*, qui aide à la coordination. Mais elle demande aussi un investissement cognitif important. C'est pourquoi le recours à une entité abstraite représentative : « le groupe », à qui l'on attribue les caractéristiques que l'on cherche à inférer chez les autres, peut *réduire la complexité du processus cognitif*, et donc le *coût computationnel* que l'on pourrait y associer. Lorsque cette démarche est partagée, le « *groupe* » peut prendre une forme spécifique de croyance collective, qui se traduit par un objet abstrait doté d'une autonomie relative par rapport aux croyances individuelles qui en sont pourtant le vecteur.

## 1. Dans quelle mesure les croyances peuvent – elles être « sociales »

Dans des travaux récents Orléan (2002, 2004) propose une nouvelle définition des croyances « collectives » (*Collectives Belief*, in : Orléan, 2004), également appelées « croyances sociales » dans Orléan (2002). Cette initiative a entre autre deux intérêts pour notre démarche. Premièrement, elle *propose une présentation formelle (logique) de la dimension « sociale » des croyances*, qui a le mérite de clarifier les choses, en particulier dans la perspective de l’élaboration d’ontologies informatiques, au sens défini précédemment. Deuxièmement, elle soulève *la problématique de l’existence « d’objet sociaux »* (ici les croyances « sociales ») *dotés d’une autonomie relative par rapport aux connaissances individuelles qui en sont le support*. L’intérêt de l’existence de tels « objets » dans la théorie des jeux (et plus largement en économie) est loin d’être unanimement établie. En particulier, la théorie standard n’y a généralement pas recours. L’objectif explicite des articles cités d’Orléan est d’ailleurs de promouvoir leur intérêt dans le cadre d’une économie « cognitive ». Le *mode d’existence* de telles entités est par ailleurs discutée aussi bien chez les philosophes (parmi ceux qui cherchent des fondements ontologiques aux objets sociaux on citera : Gilbert 1990, 2000, 2003 ; Pettit, 1993, 2000, 2004 ; Searle, 1995 ; sans oublier Descombes, 1996, sur un thème connexe) que chez les sociologues. Boudon, 2003 en conteste l’autonomie, mais il existe une solide tradition selon laquelle les « faits sociaux » doivent être traités « comme des choses » (Durkeim, 1895 ; pour une ré-interprétation: Douglas, 1986 ; pour un fondement cognitif : Kaufmann, Clement, 2003 et sur l’ensemble de la question : Livet, Ogien, 2000). Dans la mesure où notre objectif ultime de modélisation multi-agents entraîne bien la *réification* d’un certain nombre de concepts, la recherche d’un positionnement dans cette discussion de l’ontologie (philosophique) et donc du mode d’existence de certains « faits sociaux » nous semble incontournable. La typologie d’Orléan nous offre précisément l’opportunité de clarifier ce débat dans le domaine de la dimension « collective » des croyances.

### 1.1 Croyances partagées et croyances communes.

Orléan nous propose quatre définitions. La « *croyance partagée* » permet de construire de manière récursive, la « *croyance commune* » (ré-interprétation dans l’ordre des croyances, de la notion de *common knowlegde*). Il introduit ensuite deux formes de « *croyances sociales* », elles-mêmes liées par une construction récursive. Dans ce but, il introduit la notation suivante :  $C_i Q$  signifie : « l’individu  $i$  croit que la proposition  $Q$  est vraie » que nous simplifierons en : « l’individu  $i$  croit  $Q$  ». On désigne ainsi les agents par leur indice  $i$ . La notation  $i \in G_e$  signifie qu’un individu  $i$  appartient à un groupe  $G_e$  ( $G_e$  est un *ensemble* d’agents). On notera  $N_e$  le nombre d’individus de ce groupe ( $N_e = \text{card}|G_e|$ )

Il y a « *croyance partagée* » dans un ensemble d’agents  $G_e$  lorsque l’on a :

$$(1) \quad C_i Q \text{ pour tout } i \in G_e$$

qui peut s’énoncer : « tous les individus de  $G_e$  croient  $Q$  ». La définition de la « *croyance partagée* » est homologue à celle de « *connaissance mutuelle* » en théorie des jeux (on l’a juste affaiblie en remplaçant « *connaissance* » par « *croyance* »)

Il y a « *croyance commune* » dans un ensemble d'agents  $G_e$  lorsque l'on a :

$$(2) \quad C_iQ ; C_jC_iQ ; C_kC_jC_iQ ; \dots \text{ ad infinitum} \text{ pour tout } i, j, k, \dots \in G_e$$

La définition de la *croyance commune* est ainsi homologue à celle de *connaissance commune* en théorie des jeux. Le philosophe David Lewis a défini cette notion de la manière suivante : « quelque chose » est connaissance commune si (1) tout le monde le sait, (2) tout le monde sait que : tout le monde le sait, (3) tout le monde sait que : tout le monde sait que : tout le monde le sait, et ainsi de suite *ad infinitum*. Cette définition logique a ensuite été traduite en théorie des jeux par Rober Aumann.

Il apparaît ainsi que la croyance commune peut être construite de manière récursive à partir de la connaissance mutuelle. En décomposant la première en niveaux hiérarchiques imbriqués, la « sous-croyance commune » de niveau (2) n'est pas autre chose qu'une « croyance partagée » sur la sous-croyance commune de niveau (1) qui est elle même simplement une croyance partagée. De même, la sous-croyance commune de niveau (3) est une croyance partagée sur la sous-croyance commune de niveau (2), et ainsi de suite... La caractéristique commune de ces deux définitions est qu'elles se fondent strictement sur des croyances individuelles. « Elles ne sont collectives qu'au sens où *tous* les d'individus [du groupe], d'une manière ou d'une autre, les ont fait leur » nous dit Orléan. De ce point de vue, elles sont conformes à la définition de l'individualisme méthodologique retenu par Boudon (2003).

## 1.2. Les croyances « sociales » de premier niveau : une tentative d'interprétation

Orléan nous propose d'introduire directement un « *objet collectif* » intitulé le « *groupe* », que l'on notera :  $G_s$  pour le distinguer de *l'ensemble d'agents*,  $G_e$ . Bien que cet « objet » ne soit pas un acteur comme les autres agents, au sens où il ne prend pas de décision par lui-même, les acteurs (les agents) lui prêtent des croyances qui sont formellement équivalentes à des croyances individuelles dans la formalisation introduite par Orléan :  $C_{G_s}Q$  signifie ainsi : « le groupe  $G_s$  croit  $Q$  ». C'est cette proposition qui est qualifiée par Orléan de « *croyance sociale* » : « dans ce cas, on attribue des croyances à une *entité abstraite*, à savoir le groupe lui-même » (souligné par nous).

La première justification qu'Orléan nous donne de l'introduction de cette entité abstraite est empirique et repose sur l'analyse du discours ordinaire. Il ne manque pas en effet de situation où les individus attribuent des caractéristiques anthropomorphiques à une entité abstraite représentant un collectif. C'est le cas par exemple avec cette phrase citée par Orléan : « *le marché croit que cette devise est sous évaluée* ». Bien entendu, on peut arguer qu'il s'agit là d'un « raccourci » et que personne ne croit que « le marché » soit doté de croyance. Quoiqu'il en soit, la simple utilisation de ce raccourci pourrait cependant introduire des *biais cognitifs* et il serait intéressant d'en explorer les conséquences. On sait par exemple que la dynamique d'un ensemble d'agents en interaction, en tant que *système complexe adaptatif* possède des propriétés spécifiques qui excluent généralement l'usage d'un agent représentatif (Anderson *et al.*, 1988 ; Weisbuch, 1989 ; Arthur *et al.*, 1997, Schuster 2001 ; Lesourne *et al.*, 2003, Phan, 2004). Nous ne discuterons pas ces questions ici, continuant à suivre le cheminement d'Orléan, avant d'en illustrer la portée à l'examen de quelques jeux significatifs, ce qui correspond également à la

démarche de cet auteur. Avec l'introduction des croyances attribuées au groupe, Orléan nous propose deux nouvelles catégories de croyances collectives : les « *croyances sociales* ».

Selon une première définition, l'individu  $i$  croit que « le groupe croit » que la proposition  $Q$  est vraie s'il croit qu'une « grande partie » du groupe croit que la proposition  $Q$  est vraie. Orléan ne propose pas de définition formelle, mais nous pouvons certainement traduire une des interprétations possibles de sa pensée en écrivant :

$$(3) \quad C_i C_{G_s} Q \stackrel{i}{\Leftrightarrow} C_i C_{G_p i} Q \equiv (C_i C_j Q \text{ pour tout } j \in G_{p i} / N_e \geq \text{card} |G_{p i}| > N_{p i})$$

C'est à dire : «  $i$  croit que *le groupe* croit  $Q$  » si, *de manière équivalente pour lui* ( $\stackrel{i}{\Leftrightarrow}$ ) il croit qu'un nombre de membres du groupe supérieur à un certain seuil  $N_{p i}$  le croit aussi, ce que l'on notera :  $C_i C_{G_p i} Q$  ( $i$  croit qu'une partie « suffisamment grande » du groupe,  $G_{p i}$  croit  $Q$ ). *La référence de base renvoie ici aux croyances individuelles* sur  $Q$  ( $C_i C_j Q$  pour tout  $j \in G_{p i}$ ) *et non au groupe en tant qu'objet social identifié*. Les conditions d'équivalence avec la *croyance sociale* de l'agent  $i$  :  $C_i C_{G_s} Q$  doivent encore être spécifiées. Notre formalisation va plus loin que l'article initial informel d'Orléan ne le suggère, puisque notre notation entraîne que les croyances de  $i$  intègrent explicitement un périmètre d'individus  $G_{p i}$  qui est une croyance spécifique à  $i$  et permet de renvoyer à des croyances sur les croyances individuelles des membres de ce groupe. Ce sous-groupe  $G_{p i}$  « existe » donc à ce niveau en tant qu'*entité ontologiquement subjective* propre à chaque agent, mais il est de fait distinct d'une référence commune au groupe dans son ensemble (noté  $G_s$ ) qui est supposée renvoyer pour tous les agents au même objet dont ils partagent la représentation. André Orléan fait remarquer que cette première notion de croyance collective « sociale » est proche de la croyance partagée. Si l'on ne considère que la partie droite de l'équivalence  $C_i C_{G_p i} Q$ , on pourrait même considérer que cette forme de croyance collective n'est qu'une forme affaiblie de cette croyance partagée. En effet, *l'étant* de cette entité  $G_{p i}$  se réduit à la liste des agents qui le composent. Mais la partie droite  $C_i C_{G_s} Q$  est d'*une autre nature* qu'il faudrait préciser, tout comme le mécanisme qui permet de rendre en pratique ces deux visions des choses équivalentes, pour les agents.

Ce n'est que qu'à un second niveau, par un « *raccourci de pensée* » qu'il peut y avoir ce que nous définissons comme une *subsumption* de la croyance sur en ensemble de croyances individuelles ( $C_i C_{G_p i} Q$ ) par une forme de croyance spécifique sur un « *groupe* » de référence commune ( $C_i C_{G_s} Q$ ). Cette opération substitue ainsi une entité abstraite à un ensemble d'agents, et la notation adoptée présente l'avantage de mettre en avant la logique « floue » sous-jacente à l'opération, qui traduit selon nous une première manifestation du « *raccourci de pensée* » qui a été évoqué précédemment. Plus spécifiquement, le flou se manifeste en particulier dans le passage des sous-groupe individualisés  $C_{G_p i}$  qui sont une construction subjective spécifique aux agents, et qui ont donc le statut d'une croyance personnelle sur des entités concrètes à la notion abstraite du « *groupe dans son ensemble* » :  $G_s$  qui est une « *entité collective* » en principe reconnue de tous, même si par définition, il n'est pas nécessaire que cette référence partagée repose sur une même évaluation des conditions d'équivalence pour les agents (i.e. les ensembles  $G_{p i}$  et les seuils  $N_{p i}$  peuvent différer selon les agents). Autrement dit, lorsque l'on parle de « *groupe* » (ou de « *marché* ») en tant qu'*entité collective*, tout le monde

fait *comme si* ce mot désignait bien sans ambiguïté la même chose pour tout le monde, bien que ces termes puissent renvoyer à des représentations différentes selon les agents.

Notons que selon Orléan (2002), la notion vague de « grande partie » devrait être « précisée par des travaux ultérieurs ». Dans Orléan (2004), cette difficulté est contournée en replaçant « grande partie » par « majorité ». Les exemples étudiés plus loin suggèrent que la notion (à préciser) de « grande partie » aurait l'avantage sur la notion de majorité de pouvoir capturer l'idée qu'il existe d'autres *seuils critiques* dans les groupes que la majorité, mais l'inconvénient de rendre ce critère très dépendant du contexte.

### 1.3. Subsumer un collectif d'agents par une entité abstraite : quelle signification ?

L'intérêt de l'introduction de la notion de « groupe » pour la formalisation multi-agents des croyances résiderait selon nous dans l'utilisation que l'on peut faire dans le temps de l'opération qui consiste à *subsumer*<sup>1</sup> la croyance individuelle  $C_i C_{G_i} Q$  sur le sous groupe  $G_i$ , dont « l'existence » est une construction ontologiquement subjective de l'agent  $i$  par la croyance  $C_i C_{G_s} Q$  sur le groupe  $G_s$  dont la construction est d'embrée « sociale » reste cependant à définir. En particulier, on remarquera que la condition d'équivalence individuelle (3) est une condition nécessaire de subsumption, mais elle n'est pas suffisante : encore faut-il préciser (modéliser) les conditions sous lesquelles cette subsumption peut se réaliser et se maintenir (ou être remise en cause). Autrement dit, à quelle(s) condition(s) les individus remplacent-ils des croyances sur d'autres individus par des croyances sur le « comportement » d'une entité collective représentative de ces derniers, le « groupe », dont la nature est tout aussi ontologiquement subjective, mais en principe partagée par tous ? On montrera dans la dernière section que cette formalisation du « raccourci de pensée » peut avoir des effets non neutres sur la dynamique des modèles (et corrélativement sur celle des comportements).

Le mouvement qui permet de considérer comme équivalent pour l'agent  $i$  le groupe vu comme un ensemble d'agents ( $G_e$ ) au groupe en tant qu'objet social ( $G_s$ ) est problématique. Selon la perspective de l'individualisme méthodologique orthodoxe, le groupe n'existe qu'en tant que représentation commode pour organiser la pensée. Comme la monnaie « voile » de la théorie standard, le groupe ( $G_e$ ) est une *abstraction neutre* au sens où il peut à tout moment être remplacé par la liste des agents qui le compose. Descombes (1996) oppose à cette vision logique de totalité abstraite celle de *totalité concrète*, irréductible à l'ensemble des éléments du collectif considéré, car les prédicts qui peuvent lui être attribués ne sont pas distribuables aux composantes. En effet, les propriétés relationnelles qui structurent le système ne sont ni de l'ordre de tout ni de l'ordre des parties, tout en étant constitutifs des deux (Descombes parle de holisme *structural*). La *dimension diachronique* permet alors de révéler la fonction particulière de la relation dans la forme d'existence de cet objet collectif : « *du point de vue historique et social, les collectivités ne sont justement pas des collections d'individus* ».

---

<sup>1</sup> du latin *sub*, sous, et *sumere*, rendre : - Lalande « penser un fait comme constituant l'application d'une loi » (notion de philosophie kantienne qui signifie aussi penser une entité comme comprise dans une entité plus large)

*puisque leur identité n'est pas fixée par une liste d'individu* » (Descombes, 1996, p.134). Commentant Wittgenstein, Descombes montre qu'une « bibliothèque » est plus qu'un ensemble de livres (totalité logique), car c'est un tout concrètement organisé, décomposable selon d'autres niveaux d'individuation que le livre lui-même, et dont le sens est lié dynamiquement à cette structure organisationnelle. L'existence de la bibliothèque et sa signification dépasse les éventuels mouvements qui portent sur les ouvrages qui la composent.

De la même manière, les marchés financiers d'Orléan sont également des formes durables, organisés pour coordonner les échanges. Ils ne peuvent être réduits à l'ensemble des agents qui y participent à un moment donné indépendamment du rapport qu'ils entretiennent avec cette structure (Orléan 2000). Cette *structure sociale complexe* est ainsi *porteuse d'un sens* qui se maintient dans le temps indépendamment des agents qui y participent et qui dépasse leur dimension individuelle. C'est peut être précisément pour cela que les agents lui attribuent spontanément dans le langage quotidien des qualités anthropomorphiques qui ramènent cette structure à une dimension plus interprétable.

Une conjecture intéressante sur une possible interprétation *cognitive* de la subsumption anthropomorphique (du « *raccourci de pensée* ») pourrait être dérivée de l'explication cognitiviste des croyances collectives proposée par Boudon (1995, 1997). Pour ce dernier, « les sujets tentent de se débattre dans une question complexe en émettant des conjectures raisonnables. C'est parce ces conjectures sont raisonnables qu'une grande majorité les adopte » (1995, p.28). Boudon appelle « raisons transsubjectives » les raisons perçues comme *suffisantes* (au sens de la rationalité limitée de Simon) par cette « grande majorité » pour être adoptées en lieu et place d'une représentation plus complète du phénomène complexe en question. On ne discutera pas ici des causes (éventuellement infra individuelles) pour lesquelles ce principe de raison suffisante est mis en œuvre par cette grande majorité des individus, et on se concentrera sur l'objet du processus cognitif lui – même : le *raccourci de pensée*. On sait que Boudon considère que la connaissance « *ordinaire* » et la connaissance « *scientifique* » renvoient à des mécanismes cognitifs de même nature et qu'il n'y a pas lieu d'établir « une ligne de démarcation franche entre la pensée scientifique et la pensée ordinaire » (op.cit. p. 48). Selon sa vision de l'individualisme méthodologique, Boudon considère qu'une « croyance collective ne s'explique que dans la mesure où elle fait sens pour un *sujet idéaltypique*, pour un individu *quelconque* » (op.cit. p.38). On est tenté d'appliquer ces principes de manière réflexive au-delà de l'usage qu'en fait habituellement leur auteur.

On pourrait ainsi dire que le *fait social* désigné par Orléan sous le nom de « croyance sociale », qui consiste à substituer une entité abstraite (le groupe) à un collectif d'agents lié par des relations complexes, est une opération cognitive de même nature que celle qui consiste à substituer à la diversité des individus une entité conceptuelle *idéaltypique* (au sens de Weber) pour *comprendre les raisons* qui motivent le comportement de ces individus. En d'autres termes, la subsumption cognitive d'un collectif d'individus par une entité sociale chez une « grande majorité » d'individus pourrait être fondée sur des « raisons transsubjectives ». Celles – ci se traduirraient par la croyance (en fait fausse) qu'il

est plus ais  de *comprendre* la dynamique (complexe) d'un groupe en la « personnalisant », c'est- -dire en r duisant l'ensemble des agents et des relations qui caract risent ce collectif   une entit  abstraite mais *porteuse de sens* pour les individus, gr ce   l'attribution de caract ristiques anthropomorphiques.

#### 1.4. Croyances sociales auto-r f rentielles de second niveau

La derni re d finition propos e par Orl an est d'embl e *auto-r f rentielle* : « l'individu  $i$  croit que le groupe croit   Q pour autant qu'il croit qu'une grande partie du groupe croit  g alement que le groupe croit   Q ». La nature auto-r f rentielle de cette d finition ressort de l'ind termination associ e   la r f rence   des *croyances du second degr * : puisque les membres du groupe ne se limitent plus « croire Q », mais « croient que le groupe croit Q »

$$(4) \quad C_i C_{Gs} C_{Gs} Q \stackrel{i}{\Leftrightarrow} C_i C_{Gpi} C_{Gs} Q \equiv (C_i C_j C_{Gs} Q \text{ pour tout } j \in G_{pi} / N_e \geq \text{card}|G_{pi}| > N_{pi})$$

Remarquons que cette notation induit des croyances de type (3) de second degr  sur la relation d' quivalence de l'agent  $i$  elle-m me, qui traduit une subsumption au premier degr . Ce qui signifie qu'il y a d j  eu remplacement au premier degr  des croyances de l'ensemble d'individus  $G_{pi}$  par une « croyance » attribu e   l'entit  sociale  $G_s$ . Le principal avantage de l'introduction directe de cette croyance attribu e   une entit  collective  $C_{Gs}$  peut  tre trouv  au niveau cognitif, puisqu'il permet d' conomiser des capacit s computationnelles, en particulier d' viter une r cursivit  d'une profondeur infinie, comme dans le cas de la croyance commune. Par contre, le prix   payer est celui de l'ind termination li e   l'auto-r f rence du second degr  : « le groupe croit que le groupe croit ». Mais cette ind termination est aussi une des sources de l'autonomie relative des croyances « sociales ». Autonomie car le pouvoir d'action des individus sur cet objet collectif (pourtant par nature ontologiquement subjectif) est d'autant plus faible que le groupe est grand, comme on peut le montrer avec des mod les formels d'interd pendance sociale ou l'ensemble form  par les agents en interaction peut  tre consid r  comme un syst me complexe adaptatif (Lesourne *et al.*, 2002, Phan *et al.* 2004). Mais cette autonomie n'est que relative, car cet objet collectif  mane bien des subjectivit s individuelles qui en sont le support. Il n'existerait pas sans elles et sa dynamique d pend en dernier ressort   la fois des interactions « externes » entre les agents du groupe consid r  et des raisonnements « internes » (en particulier la subsumption) de ces derniers. C'est ce que nous allons chercher   montrer   partir d'une  tude critique de certains jeux de population   la lumi re de l' conomie exp rimentale.

## 2. Jeux de coordination et « saillance   la Schelling »

Dans un jeu de « *pure coordination* », les joueurs ne sont pas en comp tition, mais cherchent simplement   se coordonner sur un « fait saillant » (ou « point focal ») *sans communication pr alable* (jeu *non-cooperatif*). Une caract ristique de ces jeux du point de vue de l'approche « standard » est la multiplicit  des  quilibres de Nash (de deux   une infinit , selon le jeu consid r ). Le tableau 1 repr sente le jeu de pure coordination le plus simple sous la forme dite « normale » (ou « strat gique »). Cette forme ne nous permet de conna tre ni l'ordre des interventions et les encha nements possibles, ni

l'information dont disposent les joueurs au cours de ces enchaînements. Mais elle nous permet d'identifier les équilibres non coopératifs.

## 2.1. Le cadre du jeu

Sur le tableau 1, les stratégies du joueur  $P_1$  figurent en ligne, et celles du joueur  $P_2$  en colonne. Ceux-ci doivent choisir entre deux stratégies dans l'ensemble :  $S = \{A, B\}$ . On désigne par  $S_1$  la stratégie du joueur  $P_1$  et par  $S_2$  celle du joueur  $P_2$ . Les gains associés à chaque couple de stratégies sont désignés par :  $\pi_i(S_i, S_j)$ . Ils sont représentés entre parenthèse dans les cases de la matrice de jeu, le gain du joueur  $P_1$  :  $\pi_1$  d'abord, puis celui du joueur  $P_2$  :  $\pi_2$  ensuite, soit :  $(\pi_1, \pi_2)$ . *L'équilibre de Nash est une situation où aucun joueur n'a intérêt à dévier unilatéralement.*  $(A, A)$  est un équilibre de Nash car si le joueur  $P_1$  dévie unilatéralement et joue B son gain passe de  $\pi_1(A, A) = 1$  à :  $\pi_1(B, A) = 0$ . Il en est de même pour le joueur  $P_2$ . Le lecteur vérifiera facilement que :  $(B, B)$  est également un équilibre de Nash.

Tableau 1 : jeu de pure coordination

	$P_2$ joue $S_2 = A$	$P_2$ joue $S_2 = B$
$P_1$ joue $S_1 = A$	(1, 1)	(0, 0)
$P_1$ joue $S_1 = B$	(0, 0)	(1, 1)

Formellement, dans le cas à 2 joueurs, pour le joueur  $P_i$ , la stratégie  $S_i^* \in S$  est un équilibre de Nash si la condition suivante est vérifiée :

$$(5) \quad \pi_i(S_i^*, S_j^*) \geq \pi_i(S_i, S_j^*) \quad \text{pour tout } S_i \in S \text{ et tout } i, j \in \{1, 2\}$$

Pour une stratégie donnée (quelconque) de l'autre joueur, on peut définir la *meilleure réponse* d'un joueur comme la stratégie qui maximise son gain, conditionnellement à cette stratégie donnée de l'autre joueur :

$$(6) \quad \forall i, j \in \{1, 2\}, \tilde{S}_i \mid S_j \in \arg \max_{S_i \in S} \{ \pi_i(S_i, S_j) \}$$

On peut alors également définir un équilibre de Nash comme un ensemble de choix tel que *la stratégie de chaque acteur est la meilleure réponse aux stratégies des autres acteurs* :

$$(7) \quad \forall i, j \in \{1, 2\}, S_i^* \in \arg \max_{S_i \in S} \{ \pi_i(S_i^*, S_j^*) \}$$

Par exemple, si mon adversaire joue A, ma meilleure réponse consiste à jouer A également. Si mon adversaire joue B, ma meilleure réponse est B. Dans ces deux cas c'est en effet la stratégie qui maximise mon gain. Ces trois définitions s'étendent facilement à n joueurs en remplaçant la stratégie de l'adversaire  $S_j$  par le vecteur  $S_{-i}$  des stratégies des adversaires du joueur  $P_i$ . Dans ce qui suit, on supposera que les joueurs jouent toujours leur meilleure réponse, conditionnellement à leurs croyances sur le comportement des autres joueurs.

Nous avons donc un jeu à deux équilibres de Nash. La question de savoir lequel des deux sera choisi est qualifiée dans la littérature de problème de « sélection » d'un

équilibre. Dans le cas à deux joueurs, *en l'absence d'une référence extérieure*, il est impossible de savoir lequel des deux équilibres sera sélectionné. Si, par contre, on introduit une référence extérieure, par exemple le fait que A vienne avant B dans l'alphabet, on peut imaginer, comme on le verra plus loin, que cet équilibre a plus de chance d'être sélectionné, en particulier si chaque joueur a des croyances (de premier degré ou au-delà) qui viendraient renforcer ce biais « culturel ».

## 2.2. Le « jeu de population »

Considérons maintenant le cas d'un « jeu de population » (Blume, 1997) où chaque joueur est apparié avec un autre joueur choisi aléatoirement (sans redondance) dans la population. Les joueurs doivent jouer en « *stratégie pure* », c'est à dire choisir A ou B. Que se passe-t-il si les joueurs ont des croyances *a priori* sur la distribution des stratégies dans la population ? Supposons que le joueur  $i$  croit qu'il y a dans la population une proportion  $\alpha_i$  de joueurs qui jouent A (de manière complémentaire :  $(1-\alpha_i)$  qui jouent B). L'espérance de gains de sa stratégie  $S_i$  dans un appariement aléatoire (*random pairwise*) sera donc égal à :

$$(8) \quad \pi_i(S_i, \alpha_i) = \alpha_i \cdot \pi_i(S_i, A) + (1-\alpha_i) \cdot \pi_i(S_i, B)$$

Le lecteur vérifiera que si le joueur  $i$  choisit sa meilleure réponse avec des gains définis par l'équation (8), alors il choisira A si  $\alpha_i > 0,5$  et B si  $\alpha_i < 0,5$ . On dit qu'il choisit ainsi sa meilleure réponse contre la « *stratégie mixte* »  $\alpha_i$ . Il importe de remarquer que chaque agent  $i$  à une croyance spécifique :  $\alpha_i$  et qu'il n'y a aucune raison que cette croyance soit la même pour tous les agents. Dans le cadre qui nous intéresse ici, les joueurs doivent choisir une stratégie pure dans un jeu de population avec appariement aléatoire et tout se passe comme s'ils faisaient face à un seul joueur qui joue A et (respectivement : B) avec une probabilité  $\alpha_i$  et (respectivement :  $1-\alpha_i$ ). Dans le cas d'un *jeu de population itéré*, les joueurs peuvent « apprendre » à partir de leur expérience. Si les agents révisent ainsi leurs croyances à partir des observations qu'il font à chaque nouvelle rencontre, ces croyances évolueront en fonction de l'histoire des rencontres et si l'environnement est stable, on pourra converger vers l'un des équilibres de Nash (A, A) ou (B, B) selon un processus qui dépendra à la fois de la distribution initiale des croyances et l'histoire des rencontres.

Si l'on attribue les croyances initiales à la « culture » et la révision des croyances à « l'histoire », la sélection d'un équilibre résultera donc de facteurs historico-culturels, dans une proportion qui dépendra de la part accordée à chacun de ses facteurs dans le comportement. C'est ainsi par exemple que les anglais se sont « coordonnés » sur la conduite à gauche, alors que les continentaux l'ont fait sur la conduite à droite.

Remarquons que les croyances décrites ci-dessus peuvent se formaliser à l'aide de notation utilisée en (3) pour la *croyance sociale*. Le lecteur vérifiera que si le joueur  $i$  croit qu'une « grande partie du groupe » croit que sa meilleure réponse est de jouer A ( $\text{card}|G_{pi}|/N_e \equiv \alpha_i > 0,5$  dans le cas où l'on adopte le seul majoritaire :  $N_{pi} = 0,5$ ), alors sa meilleure réponse sera de jouer A. En notant  $Q \equiv (\tilde{S} = A)$  et  $C_j(\tilde{S} = A) \equiv$  (le joueur  $j$  croit que sa meilleure réponse est de jouer A) on a :

$$(9) \quad \tilde{S}_i = A \text{ si } : C_i C_{Gpi}(\tilde{S} = A) \equiv (C_i C_j(\tilde{S} = A) \text{ pour tout } j \in G_{pi} / 1 \geq \text{card}|G_{pi}|/N_e > 0,5)$$

Mais aussi, en cas de *subsumption* :

$$(9 \text{ bis}) \quad \tilde{S}_i = A \text{ si } : C_i C_{Gs}(\tilde{S} = A) \stackrel{i}{\Leftrightarrow} C_i C_{Gpi}(\tilde{S} = A)$$

Mais d'où proviennent ces croyances « de premier niveau » ? En reprenant nos hypothèses, et en supposant que les croyances de  $i$  sur les croyances de  $j$  sont exactes, si le joueur  $j$  croit qu'il faut jouer  $A$  :  $C_j(\tilde{S} = A)$ , c'est parce qu'il croit lui-même, qu'il y a un groupe majoritaire  $G_{pj}$  qui joue ainsi ( $\alpha_j > 0,5$ ) donc :  $C_j C_{Gj}(\tilde{S} = A)$  et sous les conditions de la subsumption :  $C_j C_G(\tilde{S} = A)$ . Si nous réintégrions ce raisonnement dans les croyances de l'agent  $i$  nous *pourrions* re-formuler nos hypothèses sur les croyances initiales sous la forme de second degré des croyances sociales selon (4) :

$$(10) \quad \tilde{S} = A \text{ si :}$$

$$C_i C_{Gpi} C_{Gs}(\tilde{S} = A) \equiv (C_i C_j C_{Gs}(\tilde{S} = A) \text{ pour tout } j \in G_{pi} / 1 \geq \text{card}|G_{pi}|/N_e > 0,5)$$

Mais aussi, en cas de *subsumption* (croyances auto-référentielles) :

$$(10 \text{ bis}) \quad S_i = A \text{ si } : C_i C_{Gs} C_{Gs}(\tilde{S} = A) \stackrel{i}{\Leftrightarrow} C_i C_{Gpi} C_{Gs}(\tilde{S} = A)$$

Comme ces trois manières de formuler le problème des croyances diffèrent, peut-on dire qu'il y en a une qui soit plus pertinente qu'une autre ?

### 1.3. Distinguer les croyances partagées et les croyances sociales : le « point focal »

Pour savoir quelle est l'interprétation la plus adéquate à une situation de coordination, on peut considérer un jeu de coordination un peu plus sophistiqué, où il ne s'agit plus de choisir entre deux stratégies indifférenciées, mais où il faut choisir un nombre entier entre zéro et cent. Comme dans le jeu précédent, le joueur gagne un s'il trouve le même nombre que son adversaire, zéro sinon. Dans ce cas, il y a potentiellement 101 équilibres de Nash, puisque tous les choix sont des candidats possibles pour un équilibre, pourvu que l'autre joueur joue de manière semblable (la matrice stratégique de ce jeu serait une matrice « identité » avec des couples (1,1) sur la première diagonale et des couples (0,0) partout ailleurs). Pourtant, lorsque l'on expérimente ce jeu avec des joueurs « réels », ceux-ci arrivent assez bien à se coordonner avec des stratégies « saillantes », telles que choisir : 0, 1, 7, 50 ou 100. Pour bien comprendre le mécanisme cognitif qui permet de passer aux croyances sociales, il est commode, suivant Orléan, d'utiliser les résultats de Metha, Starmer et Sugden (1994). Ceux-ci proposent une variante du jeu précédent qui permet une meilleure focalisation (en évacuant les deux extrêmes : la borne inférieure 0 et la supérieure 100). Ils demandent ainsi aux sujets de l'expérience de *choisir un nombre entier supérieur à 1*. Ils réalisent cette expérience sur deux groupes de référence. Dans le premier groupe, ils demandent juste de choisir un nombre en dehors de toute situation de coordination. Dans le second groupe ils proposent le jeu de coordination (gain de 1 s'il y a coordination, 0 sinon). Leurs résultats sont reproduits dans le tableau 2.

Tableau 2 : le jeu de coordination de Metha, Starmer et Sugden (1994) p. 667.

Groupe test « P » (N = 90)		Groupe jeu coordination « G » (N = 90)	
Réponses	Proportion	Réponses	Proportion
7	11,4 %	1	40,0 %
2	10,2 %	7	14,4 %
10	5,7 %	10	13,3 %
1	4,5 %	2	11,1 %
$r = 28$	$c = 5,2 \%$	$r = 17$	$c = 20,6 \%$

$r$  = nombre de réponses différentes  $c$  = probabilité que deux individus donnent la même réponse

On peut interpréter les résultats du groupe test « P » comme une traduction des « préférences personnelles » des expérimentés, qui peuvent révéler des références « culturelles sous-jacentes. C'est par exemple le cas du nombre « magique » 7 qui arrive en tête. Lorsqu'il s'agit de se coordonner, 40% des joueurs choisissent le chiffre 1 qui sert ici de « point focal ». On remarque que le nombre de réponses différentes s'est réduit, et que la concentration des joueurs sur les deux premiers résultats les plus saillants (44,4% de la population) se traduit par une augmentation significative de la probabilité que deux joueurs tirés au hasard donnent la même réponse (de 5,2% à 20,6%). On remarquera enfin que l'inversion spectaculaire dans le classement du nombre 1 ne se fait pas au détriment du nombre « magique » 7, qui a été choisi par 3 % de sujets de plus dans la population (de 10,2% à 14,4%).

Hors de toute situation de coordination, les individus du groupe de contrôle « P » choisissent leur chiffre préféré. Dans la situation de coordination (le groupe « G »), si les individus avaient cherché à « projeter » sur les autres leurs propres croyances individuelles, où s'ils s'étaient demandé quelles pouvaient être les croyances individuelles des autres, cela aurait conduit, selon Orléan, au choix du chiffre 7, qui est intrinsèquement préféré par le plus grand nombre. Au lieu de cela, nous dit-il, dans cette situation : « *les individus réfléchissent ... en se plaçant à un niveau plus général d'abstraction, de façon à déterminer un principe capable de faire émerger, aux yeux de tous, un équilibre unique* » Ainsi, « *l'activité cognitive est tournée vers le groupe en tant qu'entité séparée* ». Ainsi donc, pour Orléan, la modélisation adéquate du comportement des joueurs correspond bien à la forme auto-référentielle (4) des croyances sociales. Une forme (3) des croyances sociales, basée sur un ensemble d'opinions individuelles aurait conduit selon lui à faire émerger le chiffre 7. Au lieu de cela « la force de la définition (4) tient au fait qu'elle fait totalement abstraction de la variabilité des préférences intrinsèques des individus [contrairement à (3)] pour s'attacher à définir une croyance propre au groupe  $G_s$  en tant que groupe ». Si cette démarche cognitive est problématique, c'est parce que la dynamique du groupe, en tant que système complexe adaptatif, peut s'écarte durablement de ce que serait celle d'un agent représentatif de la majorité. Le groupe dans son ensemble pourrait ainsi se trouver durablement dans un état contraire à celui souhaité par la majorité de ses membres (Galam, 1986 ; Orléan 1998a-c).

En résumé, dans un jeu de coordination, l'approche standard conduit à une multiplicité d'équilibres de Nash (de deux à une infinité, selon le jeu considéré). Du point de vue *comportemental*, les agents parviennent cependant assez bien à se coordonner, et les expérimentalistes font apparaître des « *saillances à la Schelling* » (ou « point focaux ») dans cet ensemble de solutions, ce qui facilite la sélection d'un équilibre dans des jeux itérés. Le problème du *point focal* est alors de mobiliser des *connaissances extérieures* au cadre « standard » du jeu afin de servir de référence dans la sélection des équilibres (des « stéréotypes »). La « culture » et/ou « l'histoire » des joueurs peuvent ainsi jouer un rôle. Mais à y regarder de plus près, ce sont les « croyances sociales » auto-référentielles qui en sont les supports cognitifs. Ce sont elles qui jouent un rôle décisif. En particulier, en l'absence d'un stéréotype « culturel » pré-existant, la forme auto-référentielle (4) des croyances sociales est certainement la plus appropriée pour faciliter l'émergence par un processus « historique » de tels « points focaux ». Nous allons maintenant essayer de proposer une explication cognitive à cette focalisation sur un objet collectif : « les croyances du groupe », plutôt que la construction d'une chaîne récurrente de références croisées sur l'ensemble des agents, selon l'approche de la « croyance commune ».

### 3. « concours de beauté », hiérarchie cognitive et profondeur stratégique.

Les premiers exemples habituellement cités de la formalisation d'un jeu non coopératif reprenant le raisonnement central du « concours de beauté » de Keynes remontent aux années quatre-vingt (Moulin, 1986, Simonsen, 1988). La première expérimentation publiée basée sur ce jeu est due à Nagel (1995). L'auteur cite ses sources d'inspiration (en particulier Roger Gesnerie) en introduction de l'article, qui se situe clairement dans une perspective *post anticipations rationnelles* (cf. Orléan 2002, sur le lien avec les anticipations rationnelles). Ce jeu se présente comme une « stylisation » du thème évoqué par Keynes. Il évacue en particulier la dimension « culturelle » dont on a eu l'occasion de souligner l'importance au sujet de la coordination sur des « point focaux », ce qui nous permet de nous concentrer sur les mécanismes d'auto-référence et de « *hiérarchie cognitive* » (Camerer 2003, pour une définition dans le champ des jeux expérimentaux, Phan, 2004a, pour une application aux SMA).

#### 3.1. Le cadre du jeu et les résultats de l'approche (standard) par la dominance itérée

Dans le « concours de beauté » de Rosemarie Nagel, un groupe de joueurs, qui ne peuvent pas communiquer entre eux doit choisir un nombre (réel) entre 0 et 100. Le vainqueur du jeu sera celui qui choisira la valeur la plus proche de la moyenne des choix, multipliée par un nombre  $p$  tel que  $p < 0 < 1$  (pour  $p = 1$ , on retrouve le problème de coordination précédent, où les joueurs doivent se coordonner sur un point focal). Si plusieurs joueurs ont choisi cette valeur, ils se partagent également le gain. Prenons  $p = 1/2$  et gain = 900€ pour se fixer les idées.

Suivant Camerer (2003) on peut classer les joueurs dans une *hiérarchie cognitive* selon la « profondeur stratégique » de leur raisonnement. Au « niveau zéro », le joueur joue n'importe quoi sans réfléchir, au « niveau un » le joueur est tenté de jouer 50, qui est la

moitié, mais s'il s'aperçoit que les autres font comme lui, il devrait jouer 50/2 et l'on se trouve alors au « niveau deux ». Si l'on considère maintenant un automate qui aurait été programmé pour appliquer itérativement le principe de la sélection des stratégies dominées, ce dernier va appliquer ce raisonnement de manière récurrente jusqu'à ce que la variation de la solution soit plus petite qu'un seuil fixé à priori (pour éviter une chute sans fin dans la récurrence infinie). Cette série converge vers l'unique équilibre de Nash, égal à zéro. Notons que zéro est aussi la seule solution de l'équation de récurrence linéaire d'ordre un :  $X(t) = 0.5 \cdot X(t-1)$ , qui nous sert à vérifier qu'aucun joueur n'a intérêt à dévier unilatéralement. Supposons que dix joueurs dotés d'une grande profondeur stratégique et appliquant mécaniquement la dominance itérée aient tous trouvé le bon résultat. Ils se partagent alors les 900€, soit 90 € chacun. Que se passe-t-il si un joueur « dévie unilatéralement » et joue par exemple 1 ? La moyenne pondérée sera maintenant de 0,05, qui est plus proche de zéro que 1 (le lecteur vérifiera que ce raisonnement est vrai pour toute valeur positive, même inférieure à un). Les neufs joueurs restant se partageront donc les 900€, ce qui fait 100€ chacun, et le déviant touche zéro !

Il n'y a donc, selon l'approche standard, qu'un seul équilibre de Nash égal à zéro. Les résultats expérimentaux montrent que cet équilibre n'est jamais atteint dans le cadre d'un jeu unique, mais que les joueurs peuvent s'en rapprocher dans une certaine mesure si le jeu est itéré (apprentissage). L'explication la plus simple est que les joueurs n'appliquent pas la méthode de la dominance itérée, ou ne « descendent » pas très profond. Une explication de second niveau suggère que des joueurs peuvent anticiper que d'autres joueurs ne descendront pas très loin. Au troisième niveau, le problème se ramène maintenant à former des croyances sur la profondeur stratégique moyenne des autres joueurs. Mais on pourrait alors dire à un quatrième niveau qu'il suffit de tout le monde croire que les autres croient que le niveau de profondeur stratégique ne soit pas très élevé pour que cette prévision se réalise (c'est le phénomène de prophétie auto-réalisatrice connu chez les sociologues au moins depuis Merton (1948), mais développé largement en économie dans le cadre de la théorie des anticipations rationnelles (cf. par exemple : Azariadis, 1981 ; Azariadis, Guesnerie 1982 cf. aussi pour la psychologie sociale: Snyder, 1984). Qu'en est-il réellement ?

### 3.2. Résultats expérimentaux, hiérarchie cognitive et « croyances sociales »

Les résultats des (économistes) expérimentalistes montrent qu'en fait, dans un premier temps, un peu plus de 50% des sujets ne descendent pas au dessous du niveau 1 de la hiérarchie cognitive définie précédemment. Les choix de la grande majorité de la population laisse apparaître une profondeur cognitive apparente de 1 ou 2. Si l'on retient notre exemple précédent de  $p = 0,5$ , la distribution des résultats obtenus par Nagel et Camerer est reproduite sur le tableau 3. Dans les expérimentations de Camerer, la valeur modale s'établit à 25. Chez Nagel, le mode correspond au choix de ceux qui se situent au niveau deux, soit la classe [8-10].

Tableau 3 - Hiérarchie cognitive et profondeur stratégique dans le concours de beauté

	Niveau 0	Niveau 1	Niveau 2	Niveau 3
Nagel (1995)	24%	30%	41%	6%
Camerer (2003)	16%	38%	47%	0%

Source : Camerer (2003) p.211 – La somme des deux lignes fait 101% à cause des arrondis utilisés par Camerer

Dans le problème de pure coordination, les joueurs doivent se coordonner entre eux dans des rencontres bilatérales. En l'absence de croyance sociale dominante de type (3) en un stéréotype reconnu, la *subsumption* d'un ensemble de croyances individuelles par une « croyance » de type (4) attribuée à une entité d'emblée « collective » - le groupe - peut améliorer la coordination en facilitant l'émergence d'un point focal. Ici la situation, bien que très proche (un paramètre  $p < 1$  au lieu de  $p = 1$ ) présente immédiatement une dimension collective dans la mesure où la coordination ne doit pas se faire avec un autre choix individuel, mais avec *un objet collectif inclus dans les règles du jeu*, sous multiple de la moyenne du choix de l'ensemble des joueurs du groupe. Par contre, même si une part significative des joueurs ont une hiérarchie cognitive limitée, les joueurs plus « raffinés » qui vont faire des raisonnements de niveau 3 et former des croyances sur la profondeur stratégique moyenne des autres joueurs, ne peuvent faire l'impasse sur la formation d'une croyance sur la part de joueurs dans la population qui descendra également jusqu'à ce niveau de profondeur stratégique, ce qui introduit de la réflexivité dans le processus de formation des croyances. Il faut donc être « un niveau au-dessous » du niveau moyen, mais se garder d'aller en dessous. Les résultats mentionnés par Camerer (2003) à propos des élèves de l'Université de Caltech et d'autres groupes particuliers confirment ce dernier point : la moyenne des résultats obtenus par ces groupes est « un demi niveau » au dessous de celui obtenu en général (comme dans le tableau 3, où la totalité de la population se trouve à un niveau hiérarchique inférieur à trois), ce qui montre qu'un part significative de ces groupes est certainement descendu au « niveau 3 ».

## Conclusion

L'introduction de *croyances* attribuées aux agents en situation d'interactions stratégiques peut conduire à des raisonnements récursifs très éloignés des pratiques de coordination effectives, comme le montre la confrontation des résultats de la théorie des jeux standard avec ceux de l'économie expérimentale. Nous avons montré que l'utilisation de « *raccourcis cognitifs* » tels que l'usage dans le raisonnement d'*entités sociales* pouvait aider à simplifier certains problèmes de coordination. Dans les cas sous revue, c'est précisément parce que la théorie des jeux standard n'utilise pas ce type d'entités que l'on peut aboutir à des résultats en apparence surprenants d'un point de vue pragmatique, comme la multiplicité d'équilibres, ou des raisonnements récursifs infinis. Un usage de ces *raccourcis cognitifs* consiste en effet à subsumer la représentation d'un ensemble d'entités par une « *entité sociale* » unique qui permet alors de limiter la profondeur stratégique des raisonnements, la complexité du problème et/ou la quantité d'information à traiter par les agents. Cette subsumption pourrait également donner l'illusion de *favoriser la compréhension* de l'agent en remplaçant d'un ensemble collectif au fonctionnement complexe par une entité unique dotée d'attributs

anthropomorphiques porteurs de sens. En conséquence, d'une part, la subsumption favorise la coordination par une *simplification des processus sélectifs*, mais elle n'est pas sans conséquences, d'autre part, sur l'issue de ces processus. En effet, les dynamiques auto-référentielles associées aux *entités sociales* ainsi introduites dans le processus jugement peuvent conduire à des dérives collectives ou encore stabiliser le système dans des situations sous-optimales, en particulier lorsque les dynamiques collectives produisent des effets complexes qui rendent la fiction commode de l'entité représentative inopérante ou même fallacieuse.

Dans les deux classes jeux considérés, la connaissance qu'ont les joueurs de « l'arrière plan culturel » des autres joueurs peut être considérée comme un réducteur d'incertitude, qui aide (le plus souvent) à la coordination, à travers le support de plusieurs niveaux de « croyances collectives ». Ces dernières facilitent l'élimination des solutions qui apparaissent alors comme peu (ou pas) pertinentes et la convergence éventuelle vers un point focal. La question de l'*émergence* de telles croyances reste cependant posée.

L'apparition d'un ensemble de comportements similaires au sein d'une population d'individus disposant d'un indicateur leur permettant de s'identifier mutuellement, peut être à la source de la formation de « croyances collectives » de ce type. L'étude de la nature de l'émergence d'un *comportement partagé* chez des ensembles d'agents dotés des signes distinctifs extérieurs et les conditions de formation de croyance associées à ce phénomène émergent sont traités dans des travaux complémentaires qui portent sur les cadres cognitifs de l'observation et de la détection des phénomènes émergents (Phan 2004b, Dessalles, Phan, 2005). Plus spécifiquement, il s'agit d'identifier les conditions selon lesquelles une approche en terme de « *croyances individuelles partagées* » pourrait conduire par « *émergence* » (et *subsumption*) à une approche en terme de « *croyances sociales* ». On montre alors que l'émergence de « *croyances sociales* » au sens défini précédemment peut être modélisée grâce à la conceptualisation / formalisation d'une autre forme de « *hiérarchie cognitive* », conçue pour formaliser à la fois des phénomènes «*d'émergence forte* » (au sens de Muller, 2002) et des processus d'observation et de *conceptualisation* de ces phénomènes par « *réduction de la complexité* » (Dessalles, 1992, Dessalles, Bonabeau, 1997).

## Bibliographie

- Anderson P.W., Arrow K.J., Pines D. eds. (1988) *The economy as an evolving complex system*, , Santa Fe Institute, Studies on the Sciences of Complexity, Addison-Wesley Pub.Co, Reading Ma.
- Amy B., Bonabeau E. Dessalles J. L. Ducret J.-J. Giacometti A., Grumbach A. (1992) *Emergence dans les modèles de cognition*, séminaire ENST 15-16 Avril S003, ENST Paris.
- Arthur W.B., Durlauf S.N., Lane D.A. eds. (1997) *The Economy as an Evolving Complex System II*, Santa Fe Institute, Studies on the Sciences of Complexity, Addison-Wesley Pub.Co, Reading Ma .
- Azariadis, C. (1981), "Self-Fulfilling Prophecies", *Journal of Economic Theory*, 25, p. 380-396
- Azariadis C., Guesnerie R. (1982) « Prophéties créatrices et Persistance des Théories », *Revue Economique*, 33, 5, p. 787-806
- Blume L.E. (1997) "Population Games", in : Arthur, Durlauf, Lane eds., op cit. , p. 425-460.
- Bonabeau E., Dessalles J-L. (1997) "Detection and Emergence", *Intellectica*, 25 p.89-94.

- Boudon R. (1995) *Le juste et le vrai, Etudes sur l'objectivité des valeurs et de la connaissance*, Fayard, Paris.
- Boudon R. (1997) « L'explication cognitiviste des croyances collectives », in Boudon, Bouvier, Chazel dir. *Cognition et Sciences Sociales*, Presses Universitaires de France, Paris, p. 19-54.
- Boudon R. (2003) *Raison, bonnes raisons*, Col. « Philosopher en sciences sociales », Presse Universitaires de France, Paris.
- Bourgine P., Nadal J.P. eds. (2004) *Cognitive Economics*, Springer Verlag
- Camerer C. (2003) *Behavioural Game Theory: Experiments on Strategic Interactions*, Princeton University Press.
- Descombes V. (1996) *Les institutions du sens*, Les éditions de minuit, Paris
- Dessalles J.L. (1992) « Aspects cognitifs de l'émergence » in Amy *et al.* op cit., p.47-59
- Dessalles J.L., Phan D. (2005) “Emergence in multi-agent systems: cognitive hierarchy, detection, and complexity reduction” (*in preparation*)
- Douglas M. (1986) *How Institutions Think*, Syracuse University Press. trad. Fr. *Comment pensent les institutions*, La Découverte, Paris, 1999
- Durkeim E. (1895) *Les règles de la méthode sociologique*, rééd. Presses Universitaires de France, Paris, 1973
- Ferber J., Phan D. (2005) “Cognitive Hierarchy in Multi-Agent Games: formal, conceptual and artefactual aspects” (*in preparation*)
- Galam S. (1986) “Majority rule, hierarchical structure and democratic totalitarism”, *Journal of Mathematical Psychology*, 30, p.426.
- Gilbert N. (1995) “Emergence in social simulation”, in Gilbert, Conte eds. *Artificial Societies, The computer simulation of social life*, UCL Press, London, p.144-189.
- Gilbert M. (1990) “Rationality, coordination and conventions”, *Synthese* 84, p.1-21.
- Gilbert M. (2000) « A propos de la socialité : le sujet pluriel comme paradigme » in Livet, Ogien, op cit. p.107-126.
- Gilbert M. (2003) *Marcher ensemble Essais sur les fondements des phénomènes collectifs*, Col. « Philosopher en sciences sociales », Presse Universitaires de France, Paris.
- Kaufmann L. Clement F. (2003) « Les formes élémentaires de la vie sociale » séminaire « *Formes symboliques* » : [http://formes-symboliques.org/article.php3?id\\_article=49](http://formes-symboliques.org/article.php3?id_article=49)
- Keynes, J.M. (1936) *The General Theory of Employment, Interest and Money*, London: Macmillan, trad. française : *Théorie générale de l'emploi, de l'intérêt et de la monnaie*, Payot, Paris, 1939.
- Lesourne J., Orléan A., Walliser B. dir. (2002) *Leçons de microéconomie évolutionniste*, Odile Jacob, Paris.
- Lewis D.K. (1969) ; *Convention, a philosophical Study*, Harvard University Press Cambridge Mass
- Livet P., Ogien R. eds. (2000) *L'enquête ontologique, du mode d'existence des objets sociaux*, Editions de l'EHESS, Paris.
- Moulin H. (1986) *Game theory for social science*, New York press, NY.
- Müller J.P. (2002) « Des systèmes autonomes aux systèmes multi-agents : interactions, émergence et systèmes complexes » *Rapport présenté pour l'obtention de l'HDR*, Université de Montpellier, novembre, 58 p.
- Orléan A. (1998a) “Informational Influences and the Ambivalence of Imitation”, in Lesourne, Orléan (eds.), *Advances in Self-Organization and Evolutionary Economics*, p. 39-56.
- Orléan A., (1998b) “The Ambivalent Role of Imitation in Decentralised Collective Learning”, in Lazaric, Lorenz (eds.), *Trust and Economic Learning*, Elgar Publishers, p. 124-140.
- Orléan A. (1998c) “The Evolution of Imitation”, in Cohendet, Llerena, Stahn, Umbhauer (eds.) *The Economics of Networks. Interaction and Behaviours*, Springer-Verlag, p. 325-339.
- Orléan A. (1999) *Le pouvoir de la finance*, Odile Jacob, Paris
- Orléan A. (2002) « Le tournant cognitif en économie », *Revue d'Economie Politique* N°5 p. 717-738.
- Orléan A. (2004) “What is a Collective Belief” in Bourgine, Nadal, op.cit., p.199-212.

- Mehta J., Starmer C., Sugden R. (1994a) "Focal Point in Pure Coordination Games: An Experimental Investigation", *Theory and Decision*, 36 p.163-185.
- Mehta J., Starmer C., Sugden R. (1994b) "The nature of Salience: An Experimental investigation of Pure Coordination Games" *American Economic Review*, 84, p. 658-673.
- Merton R. K. (1948) "The self-fulfilling prophecy", *Antioch Review*, 8, p. 193-210..
- Nagel R. (1995) "Unravelling in Guessing Games: An experimental Study", *American Economic Review*, 85, p. 1313-1326.
- Pettit P. (1993) *The Common Mind, An essay on Psychology, Society, and Politics* Oxford University Press
- Pettit P. (2000) « Défense et définition du holisme social » in Livet, Ogien, op.cit p.65-85.
- Pettit P. (2004) *Penser en société, Essais de métaphysique sociale et de méthodologie*, Col. « penser en sciences sociales », Presse Universitaires de France, Paris.
- Phan D. (2004) "From Agent-Based Computational Economics towards Cognitive Economics" in Bourgine, Nadal op.cit., p. 371-398.
- Phan D. (2004a) "Hierarchy of cognitive interactive agents and statistical mechanics : how Object Oriented Programming highlights the connection", in Coelho, Espinasse. Seidel .eds., *5th Workshop on Agent Based Simulation*, Lisbon, Portugal, May 2004 SCS Pub. House, Erlangen, San Diego, p. 69-76.
- Phan D. (2004b) "L'émergence dans les Systèmes Multi-Agents (SMA) : Modéliser et interpréter l'émergence avec des SMA, Support de discussion pour le séminaire « émergence » du GDR « économie cognitive », CREA, Paris, 4 juin et pour le séminaire « Complexité » de l' I.H.P.S.T, Paris, 2 décembre. <http://perso.univ-rennes1.fr/denis.phan/slides/emergenceSMA.pdf>
- Phan D., Gordon M.B, Nadal J.P. (2004) « Social Interactions in Economic Theory: an Insight from Statistical Mechanics » in Bourgine, Nadal op cit. p. 225-358.
- Schelling T.S. (1960) *The strategy of Conflict*, Harvard University Press, Cambridge, Ma.. trad. française : *La stratégie du conflit*, PUF 1986.
- Schuster H.G. (2001) *Complex Adaptive Systems, An Introduction*, Scator Verlag, Saarbrücken.
- Searle J.R. (1995) *The Construction of Social Reality*, Free Press ; trad. française : *La construction de la réalité sociale*, NRF-essais, Gallimard, Paris.
- Simonsen M.H. (1988) "Rational Expectations, Game Theory and Inflationary Inertia" in Anderson, Arrow , Pines eds. *The economy as an evolving complex system*, Santa Fe series Vol5 Addison-Wesley CA., p. 205-241
- Snyder, M. (1984) "When belief becomes reality", in Berkowitz ed., *Advances in experimental social psychology*, 18, Orlando, FL : Academic Press, p. 247-305..
- Weisbuch G. (1989) Dynamique des Systèmes Complexes : une introduction aux réseaux d'automates, InterEditions / CNRS



# L'égalité à « soi » n'est pas l'identité : un commentaire de la Première Hypothèse du Parménide de Platon

Violaine PRINCE  
Université Montpellier 2 et LIRMM-CNRS  
161 rue Ada  
34392 Montpellier cedex 5  
prince@lirmm.fr

## Résumé

Le philosophe Parménide d'Elée (Vème siècle avant JC) énonçait dans son *Poème* :  
« Il n'est rien, et jamais rien  
ne sera d'autre que l'étant »  
*De la vérité de l'être*. Fragment 14.<sup>1</sup>

En réécrivant les fragments du philosophe, Platon<sup>2</sup> « explique » dans son *Parménide* : *Il [ en parlant de l'un, ou de la qualité d'être ] n'est pas non plus le même qu'un autre et que lui-même, ni un autre que lui-même et qu'un autre. Autre que lui-même, il serait autre que l'un ; le même qu'un autre, il serait cet autre, et non plus l'un. Il ne sera pas non plus autre qu'un autre, tant qu'il sera l'un ; car il n'appartient pas à l'un d'être autre que quoi que ce soit, mais à l'autre seulement. Il ne sera pas non plus le même que lui-même, parce que la nature de l'un n'est point celle du même, et que par conséquent, l'un ne peut être le même que quoi que ce soit.*

*L'un n'est pas non plus semblable ni dissemblable à lui-même et à un autre ; car s'il a un caractère distinct de son unité, il deviendra plus qu'un, ce qui est impossible.*

*Il n'est pas égal ni inégal à lui-même ni à un autre : égal, il participerait du semblable et du même, ce qui ne peut être ; inégal, du dissemblable, ce qui ne peut pas être non plus.*

De ce fait, il est intéressant de revisiter les notions de :

- circularité (l'être sert à se définir)
- l'identité par opposition à l'égalité à soi. L'égalité des propriétés, la conservation des valeurs de ces propriétés est-elle une « preuve » d'identité ? Ou l'identité serait autre que cela ?
- la similitude : que signifie la similitude à soi ?
- le paradoxe : qu'est ce qui est ni semblable ni à lui-même ni aux autres, et qui « est » ?

Seront en particulier traités dans l'article, sous forme d'un commentaire de la réécriture platonicienne du texte de Parménide, les trois notions suivantes :

<sup>1</sup> traduction d'Yves Battistini (cette source sera indiquée par les lettres YB dans les citations des fragments du *Poème*. Jean Beauffret a aussi produit la traduction d'autres fragments qui ont été référencés en bibliographie.

<sup>2</sup> Traduction et adaptation d'Emile Chambry. La version proposée par Luc Brisson est aussi référencée.

- la relation d'*égalité* : Trois notions différentes d'égalité sont définies.
- la notion d'*intégrité* par rapport à l'égalité à « soi ». Un être change-t-il d'identité s'il change de valeur pour ses propriétés, s'il change de propriétés, s'il change de nom ?
- comment peut-on enfin définir *être* autrement que par *exister*, et exister autrement que par être, ou le problème de la *circularité*, qui rend l'existence bien plus importante qu'un simple quantificateur.

**Mots-clés** : Philosophie, Logique, Informatique, Etre, Identité, Egalité, Similitude.

## Introduction aux réfutations parménidiennes :Egalité, Similitude, Identité

Bien qu'il ne soit pas coutume de traiter un article de réflexion sous la forme d'un commentaire de texte, l'objectif de cet article est de proposer une visite « guidée » des hypothèses de Parménide, en particulier à partir du Parménide de Platon.

De l'auteur initial, Parménide d'Elée, ne subsistent que quelques fragments du Poème fondateur, fragments dont nous émaillerons notre texte pour fournir un appui au commentaire. Ce que nous possédons par ailleurs, c'est une réécriture du Poème, fournie par Platon, et donnée en introduction du dialogue Socrate-Parménide que Platon met en scène. Cette réécriture est particulièrement intéressante car elle est rigoureuse, assez complète, et vient combler les lacunes, que la texture fragmentaire du Poème qui nous est parvenue n'a pas manqué d'engendrer.

La thèse que nous défendons globalement, en s'appuyant aussi bien sur les fragments que sur la réécriture platonicienne, c'est que sur le plan épistémologique, Parménide a fourni des bases extrêmement intéressantes et modernes en ce qui concerne les mathématiques, la logique, ainsi que des développements qui sont soutenus par l'informatique. Les neuf hypothèses du Parménide traitent de nombreuses questions cognitives, en particulier la relation entre l'être et sa représentation (ou ses représentations), et fournissent des positions théoriques sur la topologie, sur des éléments pouvant apparaître comme des précurseurs à la théorie des ensembles, sur la logique, sur la mesure,etc.

Dans le cadre du thème particulier de ce colloque, centré sur la réflexivité, avec des notions afférentes d'identité, de similitude à soi, et d'égalité, nous nous limitons à un commentaire critique sur le traitement de ces notions, à partir des pistes fournies par la Première Hypothèse de Parménide. Nous avons, dans le résumé, fourni le texte platonicien ainsi que les fragments concernés à partir desquel nous procédon au commentaire.

L'être en philosophie (en particulier dans la problématique parménidienne) pose des questions intéressantes et afférentes à la réflexivité d'une relation et à la définition récursive. Seront en particulier traités dans l'article les trois notions suivantes :

- la relation d'*égalité* : problématique de l'égalité prédicative qui est un comparateur par opposition à l'égalité référentielle qui est un évaluateur, et enfin celle de l'égalité

dénominate, qui est en réalité une relation d'identification. La non coïncidence des trois égalités est courante, et a servi souvent dans des disciplines comme l'informatique à distinguer la comparaison de l'affectation.

- la notion d'*intégrité* par rapport à l'égalité à « soi ». Un être change-t-il d'identité s'il change de valeur pour ses propriétés, s'il change de propriétés, s'il change de nom ?

- comment peut-on enfin définir *être* autrement que par *exister*, et *exister* autrement que par *être*, ou le problème de la *circularité*.

La relation en langue « le même que » signifie souvent égalité des propriétés. Nous travaillerons donc en premier sur la relation d'égalité dans les disciplines symboliques. La relation de similitude à soi est réfutée dans le texte parménidien:

*L'un n'est pas non plus semblable ni dissemblable à lui-même et à un autre ; car s'il a un caractère distinct de son unité, il deviendra plus qu'un, ce qui est impossible.*

La non similitude signifie la divergence par rapport à soi, la similitude signifie l'existence de valeurs différentes mais proches, ce qui nous amènera à un problème d'égalité de valeur. La similitude, au sens parménidien, est différente de la notion mathématique, qui est réflexive (*aRa* est vrai).

*Il n'est pas égal ni inégal à lui-même ni à un autre : égal, il participerait du semblable et du même, ce qui ne peut être ; inégal, du dissemblable, ce qui ne peut pas être non plus.*

L'égalité ici est vue comme l'étape résultant de la similitude et de « le même que » (nous évitons sciemment d'employer le terme « identique »). Comme la similitude a été écartée, alors nous nous concentrerons sur le premier paragraphe qui traite « le même que ». Les premiers termes de ce paragraphe, remettent en cause la notion d'**identité** (c'est-à-dire action ou état permettant de reconnaître ou de rendre compte de l'intégrité d'une entité) et la confusion entretenue avec celle d'égalité à soi. L'auteur va se proposer de réfuter la notion suivante : **l'identité est définie par la constante égalité à soi** (le même que soi). En d'autres termes, on niera l'usage du troisième principe du discours scientifique aristotélicien qui est que :  $\forall a, a = a$  comme étant *une définition de l'identité de a*. L'égalité à soi devra être vue comme une propriété instantanée, et non pas comme un aspect toujours vrai dans le temps.

En outre, l'égalité momentanée à soi est une conséquence de l'identité, et non pas son fondement. Mais pour cela, il faut d'abord suivre le raisonnement décrit dans ce paragraphe, qui traite de manière générale du principe d'égalité.

En disant “ *il n'est pas non plus le même qu'un autre et que lui-même* ”, le texte réfute la collision entre coïncidence des propriétés de deux entités et coïncidence des identités de ces entités, qui est un principe que l'on pourrait croire corollaire du précédent. Ce principe s'énonce comme suit :

*si A et B coïncident entièrement dans leurs propriétés, ce que l'on indique généralement par  $A=B$ , alors A et B coïncident dans leur identité, et ce, jusqu'à ce qu'une propriété distinctive permette de les séparer.*

Ce dernier point est rajouté par nous. En effet, il est d'habitude omis. L'égalité est une relation en général non datée. En réalité, l'égalité est un principe qui admet des sens

différents, avec ce que cela suppose comme conséquences distinctes. Nous allons rapidement étudier ses diverses variations, en utilisant les trois dimensions de la triade aristotélicienne étendue, c'est-à-dire la représentation, les mesures et la trajectoire.

## La représentation symbolique : les trois égalités

La représentation symbolique d'une entité rend généralement compte de ses propriétés (attributs) et de leur valeur. Par exemple, pour un être vivant doté de vision, les objets matériels ont en général une couleur, qui est une propriété. Et un ballon particulier peut prendre la valeur « rouge » pour sa couleur, ou l'ensemble de valeurs « blanc, rouge, jaune » s'il est tricolore. On dira alors qu'il existe, pour la propriété couleur de l'objet « ballon » se trouvant sous les yeux de l'observateur, une égalité entre la désignation de sa propriété (couleur), et celle de la désignation de sa valeur (rouge), et par une convention extrêmement hasardeuse, on pourra écrire : couleur (de-ce-ballon-particulier) = « rouge ». (1)

On peut aussi l'écrire autrement sous forme :

couleur (de-ce-ballon-particulier, « rouge ») = « vrai ». (2)

Mais aussi sous la forme

couleur (de-ce-ballon-particulier, x) et = (x, « rouge ») (3)

et l'ensemble de la formule ainsi définie est évaluée à « vrai ».

C'est la notion d'égalité logique. Elle a deux sens :

un sens tout à fait relationnel, qui associe des symboles du langage, comme dans le cas des formes (1) et (3), où d'ailleurs la forme (1) est impropre (plutôt un raccourci) et la forme (3) la façon canonique d'écrire les relations. C'est l'**égalité prédicative**.

Un sens **évaluatif**, comme dans le cas de la forme (2), qui évalue la validité (ou valeur de vérité) d'une formule, et ce sens est implicite dans les autres écritures, alors qu'il est explicité dans la forme (2). C'est l'**égalité référentielle**<sup>3</sup>.

Lorsque l'on cherche à savoir de quelle couleur est le ballon, alors on va se poser la question : quelle est la valeur d'une variable telle que cette variable soit une couleur (elle ne peut pas prendre par exemple la valeur « coléoptère »), et que ce soit la bonne couleur de ce ballon particulier, c'est-à-dire que l'égalité référentielle à « vrai » est assurée ?

On procèdera alors à ce qu'on appelle l'*unification*, c'est-à-dire la recherche de la **coïncidence** entre un nom de variable et une valeur, coïncidence qui sera formellement assimilée à une égalité. Dès lors, pour la logique, la représentation d'une entité x est tout ce que l'on sait de x, et l'on va supposer de facto que  $x = X$  et surtout que l'on peut réaliser une égalité entre un nom, X, et un ensemble ou des valeurs, c'est-à-dire  $\{a_1, a_2, \dots, a_n\}$ , ou une quelconque valeur v.

Remarquons qu'en logique, l'égalité prédicative est davantage une question, et n'a pas forcément valeur d'égalité référentielle, ou égalité dans le sens commun du terme. Le

---

<sup>3</sup> C'est ce qu'en logique on nomme la sémantique, ou l'interprétation. En logique classique, on reste vériconditionnel. Dans les logiques multi-valuées, ou floues, ce n'est plus le cas.

ballon de tout à l'heure peut être teint en vert, par conséquent, les formes (1) et (3) précédentes prendront implicitement la valeur « faux », et pour pouvoir continuer à les écrire, il va falloir introduire la négation explicite sur l'égalité prédicative. **Ainsi peut-on dire que la négation logique est apparue comme une nécessité lorsque l'égalité prédicative et l'égalité référentielle à « vrai » ne coïncidaient pas.**

Mais cela va au-delà de la seule égalité entre attributs et valeurs. Un objet logique A peut cesser d'être égal à B si A = B prend la valeur « faux » à la suite d'une quelconque modification de A. Ce n'est que si A et B sont atomiques, c'est-à-dire des objets de base, non interprétables (sans référence), des constantes, comme la valeur « rouge », que A=B signifie que A est identique à B, et cela, parce que l'univers mathématique des constantes, qu'il soit arithmétique ou logique, est univoque. La logique donc suppose l'égalité <attribut, valeur> comme étant toujours sujette à caution puisque falsifiable. En revanche, il lui est nécessaire de stabiliser au moins un point de la relation d'égalité, c'est celui où le membre gauche est le désignateur d'un ensemble. Dans ce cadre-là, l'égalité n'a plus le même sens, ce n'est plus une relation prédicative, ni une relation référentielle, c'est une **relation dénominative**. De manière globale, la logique utilisera des dénominations univoques, c'est-à-dire sans synonymie, ni polysémie.

En conclusion sur la logique, on peut dire que l'égalité désigne trois types de relations différentes : l'égalité prédicative, l'égalité référentielle, la relation dénominative.

En arithmétique, comme pour la relation dénominative de la logique, il y a aussi univocité. De même, l'arithmétique différencie les objets atomiques (indivisibles) des objets complexes, comme les expressions. Les objets atomiques ont un nom unique, car **leur nom** (en arithmétique) **est leur valeur**. La valeur 3 est unique dans son ensemble. Elle ne peut pas être représentée par 3 et par 3,1. En revanche, on a le droit de la représenter par (6/2), c'est-à-dire par une expression qui la calcule dans l'ensemble des rationnels. Si 3 est égal à 6/2, en arithmétique des rationnels, on acceptera de les confondre, c'est-à-dire de les rendre identiques par le signe =. Dès lors, l'égalité arithmétique est une collision de valeurs qui est toujours vraie dans l'ensemble considéré. Elle est plus stable que celle de la logique, mais surtout, elle est moins ambiguë. Ainsi, en arithmétique, l'identité se fait par égalité de valeurs prises, moyennant des conditions de stabilité sur l'environnement, le côté licite des opérations réalisées, et une convention pratique dans ce cadre-là. Si on oublie ces restrictions, et si on extrapole, alors on peut tomber dans un environnement dans lequel la convention n'est plus acceptable.

## **Les mesures : des coïncidences furtives et la différentiation de l'identité et de l'égalité ; le signe « := » par opposition au signe « = »**

En algèbre, l'égalité se fait par l'égalité des valeurs prises par les variables ou les objets algébriques. Il s'agit donc de coïncidences furtives ou possibles sur des **trajectoires** : on pourrait appeler cela des coïncidences de mesures, sans qu'il y ait

forcément égalisation des objets eux-mêmes. Comme en logique, l'algèbre suppose au moins deux relations d'égalité, celle qui associe un nom à un objet, qui est donc la dénomination, et celle qui associe un nom à une valeur, qui est donc assimilable à l'égalité prédicative. L'algèbre ne suppose pas explicitement l'existence de l'égalité référentielle<sup>4</sup>, sauf l'algèbre de Boole, qui est le croisement des deux domaines.

Cependant, il est tout à fait possible de transformer une algèbre en logique, car il existe des relations d'équivalence entre certaines algèbres et la logique, par conséquent, on ne s'étonnera pas de la coïncidence paradigmique entre les deux. En revanche, il est intéressant de savoir que l'algèbre admet la prise de valeur continue puisque les algèbres sur les réels sont évidentes. Cela signifie que la relation d'égalité se fait entre d'une part des objets symboliques de type « nom », et d'autre part, des objets de type « réels », ou « complexes », ce qui introduit un léger biais. En effet, l'univers des dénominations, même s'il est infini, est discret. Dès lors, **l'algèbre nous permet de voir que la dénomination est forcément une relation polysémique**, c'est-à-dire dans laquelle un nom admet plusieurs sens référents (dans ce cas, des intervalles continus). Non seulement la relation d'égalité est elle-même ambiguë, mais la relation dénominative, jusqu'ici univoque en logique (par convention), devient multivoque par nécessité. La polysémie est une propriété remarquable, et c'est elle que, dans un premier temps, l'informatique va tenter de réduire, car l'informatique est prise entre deux feux : celui du paradigme du continu qu'elle se doit d'approcher et le paradigme du discret qui lui est constitutif. Dès lors, il n'est pas étonnant de constater que, pour réduire l'ambiguité de la relation d'égalité, mais surtout, la polysémie de la relation dénominative, l'informatique propose une modification de la notion d'égalité par celle *d'affection*.

Quand un objet variable symbolique A prend une valeur  $v$ , on l'écrira  $A:=v$ , par opposition à  $A=v$ , qui reste une relation logique, et qui est soumise au test de vérité (égalisation référentielle). Ainsi, on peut dire, à un moment donné,  $A=B$  est vrai ou est faux, si  $A:=v$  et  $B:=v$  ont été réalisés ou non. Cela évite ainsi de supposer l'identité entre A et B alors qu'il n'y a que collision de valeur à un endroit de l'évolution des valeurs de ces variables. Cela, parce que l'informatique traite simultanément deux paradigmes mathématiques, celui de la logique, dont elle procède, et celui de l'arithmétique qu'elle permet de réifier. Là, identité et égalité ne se confondent pas, mais qu'est-ce qui permet alors de dire que A et B sont identiques? Tout d'abord, il faudrait qu'à tout moment, et quelles que soient les circonstances, A et B aient les mêmes valeurs. A ce moment-là, le programmeur s'apercevra ou pas de la redondance entre A et B. *La reconnaissance de la confusion entre deux objets symboliques se fait par la notion de redondance.* A priori, pour le programme ou le système, A et B restent distincts puisque, même prenant toujours les mêmes valeurs,

---

<sup>4</sup> En considérant des espaces de valeurs et non pas des espaces de symboles, l'égalité référentielle est alors confondue avec l'égalité prédicative. C'est que que l'on fait lorsque l'on écrit par exemple :  $ax + 3 = 5$ . L'égalité indiquée ici revient à dire : quelles sont les valeurs possibles que peut prendre x, et pour quelles valeurs de a, pour que x coincide avec  $(2/a)$ , c'est-à-dire pour que  $= (x, 2/a)$  soit vrai ?

ils ne sont pas au même endroit de la mémoire, dans la mesure où ils sont nommés différemment. Ainsi, l'informatique ancre l'idée selon laquelle **la reconnaissance de l'identité se fait par fusion des noms**. Nous avons donc une intéressante modification du principe d'identité :

*Pour un système informatique, une entité sera identique à elle-même tant qu'elle aura le même nom.*

Ce n'est donc pas une plaisanterie de savoir que, pour un système informatique, un individu change d'identité s'il change de nom. C'est un principe fondateur. En logique ou en algèbre, on peut désigner un même ensemble de propriétés par X ou par Y sans aucunement modifier son essence, puisqu'il est auto-définitoire, la relation de dénomination est univoque par commodité, mais pas par nature. En informatique, la relation de *dénomination* est essentielle (au sens propre) : elle touche (ou croit toucher) à *l'un*. C'est pourquoi il est absolument indispensable de désigner des entités informatiques représentant des individus humains par un nom différent de leur nom d'humain. Les êtres humains changent de nom sans perdre leur identité d'être, les enregistrements informatiques doivent toujours garder le même nom *informationnel* sous peine de disparaître en tant qu'entité abstraite. Nous plaidons donc pour une « désémantisation » des désignateurs d'enregistrements, parce que les entités vivantes ont un *un* qui n'est pas modifié par le changement de nom, alors que leurs correspondants informatiques, qui n'ont pas *d'un*, sont totalement tributaires du changement de nom dans leur identité.

Le poème de Parménide d'Elée met en garde contre les *noms* :

Alors ce ne sera qu'un nom, tout ce que les mortels ont nommé, le croyant vrai.

*De la vérité de l'être.* 14.(YB)

Puisque tout a été nommé lumière et nuit,  
puisque les noms qui relèvent de leur puissance furent appliqués ici ou là,...

*De l'apparaître.* 18.(YB)

Telles se montrent, dans notre accueil, les choses qui vinrent à l'existence et  
maintenant existent,  
et ensuite mourront après avoir grandi.

A chacune les hommes ont donné un nom qui la distingue.

*De l'apparaître.* 29. (YB)

Ainsi, c'est l'existence d'un *un* qui fait l'identité, et non pas une propriété, aussi intime soit-elle, telle qu'un nom. Mais la démonstration parménidienne, qui ne pouvait pas prédire les systèmes automatiques comme d'excellents contre-exemples de la confusion entre le soi et le nom (fortement ancrée dans la philosophie sémitique ancienne de laquelle les Eléates sont des « héritiers spirituels »), propose un détour par le principe d'altérité sur lequel nous reviendrons.

## **Les trajectoires : l'égalité par occupation du même espace et du même temps... Vérification de l'identité ?**

Alors que les systèmes symboliques abstraits sont chatouilleux sur les noms, le paradigme géométrique est chatouilleux sur la représentation en terme de

coordonnées espace-temps. Mais cela revient exactement au même parce que le nom symbolique est le désignateur d'un endroit de la mémoire d'un système à un moment donné, et d'un seul, et la collision nominale revient à une collision spatiale, qui est niée par le postulat d'unicité suivant : *deux objets distincts ne peuvent simultanément occuper les mêmes coordonnées dans l'espace.*

Si les trajectoires des valeurs en collision sont une conséquence de la collision nominale des variables dans les systèmes de type informatique, en géométrie, l'égalité se fait par coïncidence totale, point à point, entre deux figures géométriques. Quand ces objets géométriques contiennent une infinité de points, comme c'est le cas pour une droite ou un plan, il est difficile de vérifier que la coïncidence est toujours vraie. C'est pourquoi des théorèmes fondamentaux tels que le postulat d'Euclide ( dont une des formes est : « deux points dans un plan définissent une droite et une seule ») permettent de circonscrire la vérification de l'identité. Mais ce postulat est remis en cause par les géométries postérieures, en particulier les géométries supposant un environnement sphérique. Toutefois, elles ne réfutent jamais le principe d'unicité spatiale pour un objet non daté, ou l'unicité spatio-temporelle pour un objet pour lequel l'observateur affecte une dimension temporelle. Cela étant, la Première Hypothèse du Parménide de Platon semble tout à fait en phase avec le principe d'unicité, en disant :

*Autre que lui-même, il serait autre que l'un*

donc si l'*un* est, il ne peut-être un autre dans le sens de l'identité, en revanche, tout désignateur ou ensemble de propriétés ou de coordonnées peut être autre que l'*un* c'est-à-dire autre que la qualité d'être. Dès lors, **l'identité d'une entité est unique, alors que ses descriptions peuvent être nombreuses, et qu'elle peut occuper diverses places dans l'espace-temps, et posséder autant de noms que possible.**

En obtenant cette conclusion, on vérifie seulement une partie de la relation d'identité. Nous venons de décrire le fait suivant : **si l'*un* existe il a une identité unique.** La suite de la démonstration est une confirmation par l'absurde. En effet, le texte dit : *le même qu'un autre, il serait cet autre, et non plus l'un.*

Effectivement, A peut être égal à B, mais si l'*un* de A que l'on nommera  $\alpha$ , existe, et si l'*un* de B,  $\beta$ , existe, alors forcément  $\alpha$  est distinct de  $\beta$ , sinon A et B ne seraient qu'un même objet et A et B seraient simplement les différents noms d'une même entité.

## Identité et Intégrité du Soi

### **Réfutation de la circularité de la double négation**

La différence entre égalité et identité est renforcée par la réfutation de la circularité de la double négation. En effet, dans toutes les branches des mathématiques classiques (à l'exception de certaines logiques épistémiques particulières), nous avons la propriété suivante :  $\neg(\neg =) ==$  ce qui signifie que si quelque chose n'est pas distinct ( non non égal) d'autre chose, alors il lui est égal. Cela peut se vérifier aisément. En revanche, la relation d'identité, si elle admet la négation, c'est-à-dire que

l'on admettra que A est distinct de B si  $\alpha$  et  $\beta$  existent, n'accepte pas la circularité de la double négation. C'est cela qui est indiqué par :

*Il ne sera pas non plus autre qu'un autre, tant qu'il sera l'un ; car il n'appartient pas à l'un d'être autre que quoi que ce soit, mais à l'autre seulement.*

Cela revient à dire qu'il n'existe aucune relation possible entre  $\alpha$  et  $\beta$ . L'égalité n'est pas opérante sur les *uns*, ce qui signifie que *l'on ne pourra jamais écrire  $\alpha = \beta$ , parce que  $\alpha$  n'est pas de même nature que les propriétés égalisables*. L'identité n'est donc qu'une relation entre l'objet et lui-même, et l'**altérité dérive de l'existence d'un *un***.

### **Le principe d'altérité, ou de non égalité**

Le principe d'altérité est celui qui est défini par deux tentatives ayant échoué : celle qui consiste à égaliser deux *uns*,  $\alpha$  et  $\beta$ , et comme on l'a vu, elle n'est pas possible en ce sens qu'un *un* est obligatoirement distinct d'un autre *un*. La deuxième tentative est celle qui va égaliser l'*un* avec autre chose qu'un *un*, ou une qualité d'être, c'est-à-dire celle qui cherche à égaliser deux objets de nature incompatible, d'une part un *un*, d'autre part des propriétés, des mesures, des trajectoires, des noms. Celle-là aussi a été démontrée comme étant impossible. C'est donc par ce principe qui dit que :

**I' *un* est par essence unique, donc non égal à tout autre *un*. L' *un* est par essence sans relation avec ce qui n'est pas *un*** (c'est-à-dire, ce qui n'est pas une essence). **L'autre est défini comme ce qui ne peut être égal à l'*un*.**

La Première Hypothèse démontre qu'il est impossible à l' *un* d'être autre chose que lui-même. Elle montre que l'*un* n'est pas le même que lui-même, dans le sens de la relation d'égalité, et non pas dans le sens de la relation d'identité. Mais avant de développer le dernier volet de la démonstration, celui qui, après avoir fait le tour de la notion d'égalité dans les disciplines de nature symbolique, puis d'avoir étudié la divergence entre l'égalité et l'identité par la double négation et le principe d'altérité, nous voulons nous arrêter à une notion très importante, c'est-à-dire la notion de l'*autre*. Il semble y avoir deux *autres* possibles :

- ce qui est un *un*, mais distinct de l'*un*,
- et ce qui n'est pas *un* et qui peut être soit une **triade de matérialisation**<sup>5</sup> d'un *un*, soit un non-*un*.

Du premier des deux possibles, « ce qui est un *un* mais distinct de l'*un* », qui est l'essence d'une autre entité, on peut dire que l'on sait distinguer, que l'on ne peut égaliser, mais qu'en réalité ces *uns* sont de même nature, et cette nature est assez difficile à cerner. Elle n'a jusqu'à présent pas été qualifiable, quoique, dans le Poème, on ait les fragments suivants :

Même chose se donne à penser et à être...

Nécessairement doit être ce qui à dire et à penser se donne ; car il peut être.

Mais le non-étant ne le peut...

---

<sup>5</sup> Nous appelons **triade de matérialisation** d'un *un*, le triplet (représentation(s), mesures, trajectoire) qui sert à décrire l'entité matérialisant cet *un* et qui relève du discours scientifique ou du discours d'opinion.

*De la vérité de l'Etre. 5 et 6. (YB)*

Même chose que le penser [de l'être] et ce par quoi s'accomplit ce penser.

Jamais hors de l'étant en lequel elle s'exprime, tu ne trouveras la pensée.

*De la vérité de l'Etre. 12 et 13. (YB)*

Par conséquent, la pensée semble être la nature de l'être, mais la pensée est un mot qui doit être vu sous un angle large. Elle paraît englober aussi l'action (le fragment 12), et être produite par ce qu'elle constitue (fragment 13)<sup>6</sup>. En pratique, l'**être est l'existence, et cette existence n'est pas vérifiable au sens de la preuve, mais elle est reconnaissable par ce qui produit de la pensée et par ce qui en découle.**

Il était donc malheureusement facile pour Platon d'extrapoler et de supposer alors que le monde des Idées est le monde de l'*un*, et que c'est la forme parfaite de l'Etre, dont toutes les autres ne sont que des avatars affaiblis. Mais c'est méconnaître ou mal interpréter le début du fragment 6. Le modal « pouvoir » et le modal « devoir » sont à considérer dans le sens d'une recommandation. Un non-être peut, dans sa capacité, être dit. Mais on ne doit pas le dire, ou se le donner à penser; car alors :

... L'autre affirme : il n'est pas et il est nécessaire qu'il ne soit pas.

Route, sentier plutôt, te dis-je, où il n'y a que mensonge.

Car l'esprit ne saurait concevoir le non-être qui ne peut s'accomplir,  
ni les mots l'exprimer.

*De la vérité de l'Etre. 1 (YB).*

Telle est en tous cas l'injonction à laquelle Parménide d'Elée fait référence. Elle réclame l'attention des philosophes « sur ce qui est ». Ce qu'il faut retenir de ce discours est qu'un *un* peut être distinct d'un autre *un*, mais leur nature est unique, et c'est elle qui inspire ce que l'on en dit, c'est-à-dire la triade de matérialisation. En d'autres termes, la triade de matérialisation n'est pas ennemie de l'*un*, ce que l'on aurait pu penser dans la deuxième branche de l'alternative donnée ci-avant, au contraire, elle en procède. C'est d'ailleurs ce que tenteront de montrer les autres hypothèses (il y en a neuf en tout).

Par conséquent, il convient de différencier « ce qui n'est pas *un*, mais qui en procède », *id est*, la triade, de « ce qui n'est pas », le *non-un*. Quand la triade se réfère à un *un*, dont on sait maintenant que c'est l'existant, alors elle est recommandable, même si elle est partielle, même si elle est imparfaite. Mal connaître quelque chose, et mal le décrire n'enlève rien à sa qualité d'être. A partir du moment où il existe, « inengendré et impérissable », rien ne peut altérer sa qualité. On peut donc concilier encore Aristote et Parménide, la science descriptive et la recherche de la connaissance... à condition de ne pas substituer la triade à l'identité, la description à l'*un*, l'apparaître à l'*être*, ce que montre la Deuxième Hypothèse du Parménide de

---

<sup>6</sup> )On pourrait lui donner les dénominations d'esprit, d'âme et de soi, de *nefesh*, *rouh* et *neshama*, les trois dimensions de l'Etre dans le Zohar. Mais aussi celle d'animus, d'anima, de souffle, de vie, de présence organique, minérale, symbolique...

Platon. En revanche, plusieurs triades ont été créées pour ce qui n'est pas. On a malgré tout su trouver des mots pour exprimer le non-étant. Le prétexte invoqué est fallacieux : ce qui est, est pensable, et inversement. C'est l'« inversement » qui est faux. On peut tout à fait penser ce qui n'est pas. Ce n'est pas pour autant qu'il est créé, ce n'est pas pour autant qu'il existe, l'*un* est « inengendré » dit le Poème. Mais on le décrit. La pensée procède de l'*un*, cependant elle n'y est pas confinée, elle peut atteindre le non-étant. La pensée est libre...Et bien que Parménide parle de « l'esprit qui ne saurait concevoir le non-étant », en faisant probablement allusion au néant, étape ultime du non-étant, la route pour y arriver est jonchée des « mensonges » dont il dénonce la nuisance.

## En conclusion, Identité et Existence, deux mots pour un seul *un*

On finit donc par se demander si l'identité (non synonyme d'égalité) et l'existence ne sont pas deux notions équivalentes. On peut alors représenter l'*un* non pas par un symbole qui reste encore très près du nom, mais par le symbole  $\exists$  en post-fixe, qui est celui du quantificateur existentiel et qui signifie « il existe ». Si on parle de l'entité *x*, cela signifie par défaut  $\exists x$ , tel que *x* soit décrit par une triade  $\Delta x$ . L'identité de *x* est en lui-même.  **$\exists$  n'est pas seulement un quantificateur c'est la définition même de l'existence, et l'identité n'est pas autre chose que la relation, de ce qui existe, au monde et à soi.  $x\exists x$  signifie « *x* est *x* ».**

### La dernière étape ou la non nécessité de l'égalité à soi

*Il est important de se convaincre qu'il n'est tout simplement pas possible, pas pensable, d'immobiliser a priori « complètement » l'effet  $\eta_\Delta$  d'une opération  $\Delta R \rightarrow \eta_\Delta$ <sup>7</sup>...Etant donné... le postulat réaliste minimal<sup>8</sup>, une telle notion est illusoire, un monstrueux faux absolu, une brutale violation –d'embrée– de précisément le but poursuivi ici d'éradiquer toute fausse absolutisation.*

(M. Mugur-Schächter, 1995)

*Il ne sera pas non plus le même que lui-même, parce que la nature de l'*un* n'est point celle du même, et que par conséquent, l'*un* ne peut être le même que quoi que ce soit. L'*un* n'a pu être le même qu'un autre *un*, sa triade, ou ce qui n'est pas.*

---

<sup>7</sup> *Un opérateur épistémique  $\Delta$  défini sur  $R$  et qui produit comme objet d'examens ultérieurs une entité dénotée  $\eta_\Delta$  qui ne s'identifie pas à  $\Delta$ , ni ne l'inclut, mais qui par ailleurs est entièrement non restreinte, sera dénommé « découpeur ». L'endroit de  $R$  (la sorte de réel) sur lequel  $\Delta$  agit fait explicitement partie de la spécification de l'opérateur  $\Delta$ . Mais le contenu de cet endroit échappe à la définition de  $\Delta$ . En symboles :  $\Delta R \rightarrow \eta_\Delta \quad \eta_\Delta \subset \neg \Delta R$*

<sup>8</sup> *Postulat Réaliste Minimal : J'admet l'existence, indépendante de tout concepteur et de tout acte d'observation, mais EXCLUSIVEMENT l'existence par ailleurs strictement non qualifiée, d'une potentialité que des qualifications émergent si des actes d'observation se produisent.*

« x est x » ne signifie pas « x=x ». Si x est descriptible, sa description change dans le temps, et ne capture pas son essence. Si x possède des mesures, ces mesures changent, en valeur, en nature et en nombre, sans toucher aucunement au fait que x soit x. Si x évolue dans le temps, alors aucune évolution ne peut toucher ce qu'il est, mais cela ne veut en aucun cas dire que ce qu'il est est fixe.

Notons que « être le même » est une relation plus ambiguë que l'égalité. Le même en quoi ? Ce n'est pas en nature manifestement : la nature de l'*un* n'est pas celle du même. Nous étions effectivement déjà arrivés à cette conclusion. L'égalité n'est pas de la nature de l'*un*. Mais aussi quelque chose de plus intime encore, quelque chose qui fait que dire « x est x » est redondant, et qu'il suffit de dire “ x est ” et que ce n'est pas la peine d'écrire  $x\exists x$  mais seulement  $x\exists$ . On remplacera donc, le troisième principe du discours scientifique :

$$\forall x, x = x$$

qui n'est vrai que très partiellement sur x, par le **principe d'identité** qui est :

$$\forall x, \text{ si } x \text{ existe, alors } x\exists$$

qui signifie que **si une entité existe, alors elle possède une qualité d'être unique qui l'identifie et la rend distincte de toute autre tout en étant de même nature essentielle que les autres entités existantes.**

## Références Bibliographiques :

Yves Battistini, 1988 *Trois présocratiques*. deuxième édition revue et corrigée. TEL-Gallimard, Paris. (Première édition 1968).

Jean Beaufret, 1955 *Parménide, "Le Poème"*, Epiméthée-PUF. Paris.

Luc Brisson, 1994 *Platon, "Parménide"*. GF- Flammarion. Paris.

Emile Chambry, 1967, *Platon, "Théétète, Parménide"*. GF-Flammarion. Paris.

Miora Mugur-Schächter, 1995. Une méthode de conceptualisation relativisée : vers une épistémologie formelle apte à faire face aux complexités », *Revue Internationale de Systémique*, 9, pp. 269 à 303.

# Le danseur de ballet en miroir

Francis Rousseaux (Ircam-CNRS)

[francis.rousseaux@ircam.fr](mailto:francis.rousseaux@ircam.fr)

**Résumé.** Pourquoi, alors que les miroirs inversent la droite et la gauche, alors qu'ils nous projettent dans le monde inhabitable que dénonçait Borges, quand ils ne nous font pas perdre la tête comme au Nain de *l'Anniversaire de l'infante* mis en nouvelle par Oscar Wilde, pourquoi donc les miroirs semblent-ils si prisés par les danseurs et les chorégraphes de ballet ?

Cette investigation nous conduira sur la piste de la répétition, de la représentation et de la création de spectacle vivant, en passant par un travail philosophique sur la notion de *singularité*.

En répétition, les danseurs de ballet voudraient trouver dans le reflet spéculaire un lieu d'exposition qui ferait d'eux-mêmes le public de leur propre corps exposé, alors qu'ils cherchent dans le miroir, dans le même temps, quelque chose de leur corps propre, afin de le réfléchir dans le cadre du projet chorégraphique.

**Mots-clefs.** Miroir, répétition, représentation, disposition, singularité, création, exposition.

## 1° De speculo

“Inventa sunt specula ut homo ipse se nosset”. Ainsi s'ouvre la *Guerre des Gaules*, par une assertion de César à propos de l'invention des miroirs à des fins de (re)connaissance de soi.

Nous reviendrons sur les intuitions de César. Mais commençons modestement : à tout le moins, il n'est pas faux de dire que le regard réfléchit dans le miroir, et ouvre au regardant un accès à des vues ordinairement cachées.

### *Différents régimes d'usage du miroir*

Tâchons d'emblée de distinguer des grands régimes d'usage des miroirs :

- A l'exception d'une seule, les vues ordinairement cachées que délivre le miroir sont accessibles par d'autres moyens, comme un mouvement des yeux, de la tête ou du corps. Dans ces cas, l'usage du miroir dispense de ce mouvement (c'est le cas des rétroviseurs, qui permettent aux conducteurs des véhicules automobiles de

“jeter un œil” sur ce qui se passe derrière eux, sans avoir à se retourner). Le miroir est alors un *instrument* améliorant l'appropriation de l'environnement visuel du conducteur.

- Une seule vue cachée est absolument conditionnée à l'usage du miroir et inaccessible au regard sans cet artifice (ou des artifices voisins, comme le couple caméra/écran), et il s'agit de la vue réflexive (spéculaire) du regardant se regardant. Nous aurons l'occasion de revenir abondamment sur cette étrangeté, qui confère au miroir un usage singulier dont le mythe de Narcisse n'épuise pas les pouvoirs. C'est le miroir *vertige*.
- Cependant le cas du regard qui se porte sur une partie du corps propre ordinairement cachée ou difficile à voir, par le truchement d'un miroir, est en quelque sorte intermédiaire, puisque cette attitude conditionne l'appréSENTATION de cette partie du corps (au sens de Husserl), et ne peut pas se réduire à une simple appropriation visuelle. Le miroir se fait alors *prothèse*.

### ***Le miroir comme instrument de danse***

Lorsque les danseurs font usage d'un miroir pour contrôler les mouvements de leur corps propre ainsi que ceux, éventuellement, du corps de leurs partenaires ou du chorégraphe, il est probable qu'ils mobilisent concurremment ces différents régimes d'usage du miroir.

Il faut rappeler en effet que les danseurs et les chorégraphes occidentaux ont traditionnellement besoin de miroirs pour travailler, et rares sont les classes de ballet qui n'en sont pas équipées — même s'il existe de nombreuses pratiques et traditions de la danse qui ne mobilisent pas de miroir. Laban lui-même a écrit : “ Our body is the mirror through which we become aware of ever circling motions in the universe with the polygonal rhythms ” (page 26 de *The Language of Movement. A guidebook to choreutics*).

Mais dans l'imagerie classique du ballet, le reflet du miroir n'est jamais bien loin. Et nous adoptons ici le regard décalé de l'ethnologue qui regarde la classe de ballet sans le souci de classer les spectacles vivants, un peu comme Joann Kealiinohomoku écrivant "Une anthropologue regarde le ballet classique comme une forme de danse ethnique" (lu dans *Danse Nomade, Regards d'anthropologues et d'artistes*, Nouvelles de Danse, Eté 1998, pp. 47-67, traduction Agnès Benoit).

Ainsi, à l'heure de la mobilisation des nouvelles technologies dans la danse contemporaine (les créateurs de performances regardent volontiers du côté de la réalité virtuelle, cherchent à tirer parti de capteurs de geste et de retour d'effort, ont tendance à utiliser des images de synthèse et des techniques vidéo, et mettent quelquefois en œuvre des systèmes interactifs d'intelligence artificielle), une investigation du côté des parquets cirés entourés de glaces réfléchissantes et de barres fixes n'est peut-être pas aussi saugrenue qu'il y paraît.

## *L'attitude du danseur face au miroir*

Donnons quelques exemples des attitudes possibles des danseurs face aux miroirs, souvent combinées dans un même geste des yeux, de la tête et du corps :

- Le danseur apprend à synchroniser un déplacement complexe avec celui d'un partenaire (peut-être un rythme musical est-il perceptible en arrière-plan), le miroir est *instrument* de contrôle de l'environnement perceptuel de l'artiste. Car dans la danse, les corps se déplacent, ils sont mobiles.
- Le danseur apprend à aligner telle partie de son corps avec telle autre, le miroir est la *prothèse* qui permet de voir ensemble ces deux parties du corps propre, et de contrôler leur alignement en jouant de la musculature. Car dans la danse, les parties du corps font mouvement relatif, les corps de chair et d'os sont plastiques et élastiques.
- Si le spectateur appréhende globalement les déplacements et les mouvements des corps dansés, le danseur et le chorégraphe savent que "déplacements" et "mouvements" ne sont pas étrangers l'un à l'autre, et que la dynamique de la danse consiste à les relier. Tel mouvement du pied va conduire un déplacement du corps par contact avec le sol, tel déplacement va conduire par inertie à des mouvements d'amortissement engageant différentes parties du corps. Cependant, "déplacements" et "mouvements" ne sont pas vécus de la même manière par le danseur, qui vit *dans* sa chair la configuration interne de sa chair et *par* son corps le placement chorégraphique de son corps<sup>1</sup>. C'est pourquoi sans doute le miroir est aussi *vertige* pour le danseur, qui doit se trouver alors même qu'il se donne à voir.

L'ennui avec les miroirs, c'est qu'ils ne sont pas fidèles à nos attentes. Ils nous piègent comme ils piègent les alouettes. Pour chacun de leurs régimes d'usage en effet (*instrument*, *prothèse*, *vertige*), ils nous trahissent sournoisement à la moindre occasion. Et les danseurs en font certainement les frais, eux qui mêlent innocemment dans leurs gestes tant de confiance répétée dans les reflets.

## **2° Miroirs aux alouettes**

Les miroirs sont tout à la fois des *prothèses* invalidantes qui inversent insidieusement la gauche et la droite, des *instruments* passablement détraqués qui nous projettent dans un monde inhabitable, et des *vertiges* diaboliques qui abusent de notre confiance et menacent de nous faire perdre la tête.

Comment cela n'aurait-il pas de conséquences fâcheuses pour les danseurs ?

---

<sup>1</sup> Ce qui serait à travailler là, c'est la différence de *senti* entre un mouvement (des parties du corps) et un déplacement (du corps).

## ***Les miroirs sont des prothèses invalidantes***

Quand le danseur lève le bras droit, pourquoi son reflet dans le miroir lève-t-il invariablement le bras gauche ? Comment contrôler ses mouvements dans ces conditions de rupture de la latéralité ? Comment ne pas s'en trouver profondément désorienté ?

Mais au fait, pourquoi le même miroir n'inverse-t-il pas le haut et le bas ? Serait-ce une manifestation supplémentaire de ce pouvoir magique qu'on lui prête si volontiers dans les contes féeriques ? Une investigation s'impose.

De par notre mouvement propre, nous pouvons connaître les formes volumiques d'objets qu'il nous est impossible de déplacer. Pour cela, on opère par contournement de l'objet, ce qui équivaut au fond, pour le regard, à un déplacement de l'objet. Cette opération, on s'en souvient, peut toujours être vue comme une combinaison d'isométries positives (les isométries sont des transformations qui conservent les distances, donc qui ne déforment pas l'objet transformé, et elles sont dites positives lorsqu'elles conservent aussi les sens et les orientations internes de l'objet). Au sein de cette combinaison d'isométries (essentiellement des translations et des rotations) qui nous donne accès toujours partiel à l'objet auquel s'attache notre curiosité, la rotation est essentielle pour découvrir les parties cachées et aboutir à une connaissance finalement adéquate (Husserl parlerait ici de *donations par esquisses* pour indiquer que les perceptions visuelles de l'objet sont toujours parcellaires, et de *constitution par remplissement* pour signifier la synthèse que nous en faisons pour parvenir à une représentation unifiée).

Lorsque nous contournons un objet, par exemple un objet d'art dans la vitrine d'un musée qu'il est interdit de prendre en main, la rotation privilégiée se produit autour d'un axe vertical, les bipèdes que nous sommes évoluant à la surface de la terre sous attraction universelle. Je ne prétends pas ici que d'autres rotations n'ont pas cours (la position relative de nos deux yeux ne nous permet d'appréhender une profondeur de champ qu'à l'aide de repères verticaux, et il est parfois utile d'incliner la tête pour rechercher des profondeurs de champs). Mais je prétends que la rotation autour d'un axe vertical est incontestablement la rotation avec laquelle nous entretenons la plus grande affinité, dès lors que nous sommes en âge de marcher.

En face d'un miroir, il se passe alors la chose suivante : nous sommes enclins à reconnaître notre propre image *dans* le miroir, à une (petite) anomalie près; l'inversion de la latéralité. En effet, les isométries négatives comme la symétrie ne se donnent pas dans la nature en dehors de l'artifice "plan miroir", et elle restent comme une provocation au bon sens. D'où la fascination narcissique du miroir, qui peut sidérer (le fameux "stade du miroir" bien identifié par Freud). Mais il convient de se ressaisir : faisant sens à toute force avec les moyens connus, nous convoquons nos transformations fétiches et forçons notre compréhension de l'image appréhendée grâce à l'expérience de pensée d'une rotation d'axe vertical. Cette image que je vois, c'est un double de mon visage qui me fait face comme dans un face-à-face, et je

pourrais “occuper la place qu'il occupe” en faisant demi-tour sur mes talons et en me reculant d'un pas. C'est à ce prix que nous nous reconnaissions dans les miroirs.

Ce faisant, nous essayons de réduire par la pensée une symétrie plane (isométrie négative) à une rotation (isométrie positive), ce qui est géométriquement impossible, à moins de s'accommoder de l'illusion et de prétendre qu'il y a “inversion de la droite et de la gauche”. Pourquoi cette prétendue inversion est-elle le corrélat tout désigné de la (productive) confusion géométrique ?

Parce que l'image que nous découvrons “dans le miroir” vertical est quasi-symétrique par rapport à un plan perpendiculaire vertical médian, tout comme les visages que nous voyons dans la rue, et que dans ce cas particulier, *l'image géométrique* de la symétrie plane est similaire à l'image géométrique d'une rotation d'axe vertical de 180°. Mais en réalité, ce n'est pas l'image géométrique qui nous intéresse dans le reflet de notre visage (par ailleurs pas si symétrique qu'il ne paraît au premier regard), mais bien plutôt le fait qu'il s'agisse de *notre* visage, qui cligne des yeux et grimace quand je grimace, qui touche son oreille quand je touche mon oreille. C'est alors là qu'il faut bien invoquer une mystérieuse inversion de la latéralité.

Je crois même que c'est parce que les miroirs sont capables d'inverser la latéralité (et nous en savons plus désormais sur la supercherie à laquelle ils nous acculent) qu'on est enclin à leur prêter volontiers le pouvoir d'inverser également le temps vécu, cette autre référence intérieure. Dans *Le portrait de Dorian Gray* d'Oscar Wilde, c'est même l'image peinte, le fameux portrait de jeunesse relégué dans le grenier, qui vieillit en lieu et place du modèle, et qui ne doit jamais être vue de celui-ci, sous peine que le hiatus change de sens. Quant à Borges, qui aimait à lire la nuit les quartiers de lune pour s'orienter dans le temps du cycle lunaire, il regrettait que les miroirs le désorientent jusque dans cette activité d'interprétation.

Et on pourra remarquer encore (je dois cette idée à mon collègue Claude Secroun) que la verticale est une référence extérieure, déterminée par la gravitation universelle et ressentie au moyen de l'oreille interne, alors que la latéralité gauche-droite est une référence purement interne du corps humain, liée à un repère propre. Ainsi, l'opposition *je/univers* se cristallise-t-elle dans le regard narcissique au travers du miroir, lorsqu'il faut sacrifier l'ordre du *je* à l'ordre du monde.

### ***Les miroirs sont des instruments détraqués***

Ainsi l'image de son visage dans le miroir exige un déplacement *virtuel* pour être reconnue, et l'est du même coup au mépris de la géométrie, ce qui ne manque pas d'avoir des conséquences préjudiciables pour le danseur, qui doit sacrifier la latéralité de ses mouvements et se constituer prisonnier du monde des reflets.

Au moins pourrait-on espérer que l'usage des miroirs permette le contrôle efficace des déplacements *réels* des danseurs ... Mais il n'en est rien, comme nous allons voir.

Le problème vient cette fois du fait que la donation par esquisses ne fonctionne pas à travers un miroir. L'image spéculaire est en effet toujours définitivement réduite à une esquisse, en ce sens précis qu'elle ne peut jamais se donner comme esquisse *de quelque chose*. En effet, lorsque le regard change d'incidence pour explorer les dimensions cachées de l'espace spéculaire, c'est certes une nouvelle image qui se livre, mais qui ne se prête pas aux synthèses adéquates qu'on a l'habitude de pratiquer. Seuls les coiffeurs parviennent apparemment à pratiquer quelques synthèses actives, encore ne faut-il jamais trop les distraire. Quant aux jeunes filles qui se maquillent dans les trains, il faudrait davantage louer leurs efforts, car cette activité requiert sans doute beaucoup de talent.

Et multiplier les miroirs ne règle pas le problème : les jeux de miroirs restent miroirs, multiplient-ils les reflets à l'infini, et le Palais des glaces un endroit hasardeux. Borges prédit, lui, que même les sons finiront par être écrasés par les miroirs, ces objets hannis. Il déplore aussi la froideur de leur reflet, qui n'a pas de mémoire (nous reviendrons sur ce point). Le mieux serait d'ailleurs de relire son magnifique *Los espejos*, livré ici dans la langue de Cervantes<sup>2</sup> (c'est nous qui mettons en italique certains passages). Ici, c'est au sacrifice de l'immanence que conduit le miroir, à laquelle seul un homme préparé par une foi transcendante pourrait consentir.

Et là encore, je crois que l'écrasement des dimensions d'espace dans le reflet spéculaire laisse à espérer des pouvoirs magiques du miroir : en tout cas la Reine dans le *Blanche-Neige* de Grimm en est bien persuadée, qui demande à son miroir de faire converger tout le royaume dans la seule focale de sa question fatidique "Miroir, miroir magique, dis-moi qui est la plus belle ?". Et le miroir de répondre : "O reine, tu es très belle, mais, derrière les sept collines, dans le bois auprès des nains, Blanche-Neige est mille fois plus belle que toi."

Yo que sentí el horror de los espejos  
no sólo ante el cristal *impenetrable*  
donde *acaba y empieza, inhabitable*,  
*un imposible espacio de reflejos*  
  
sino ante el agua especular que *imita*  
el otro azul en su profundo cielo  
que a veces raya el ilusorio vuelo  
del ave *inversa* o que un temblor agita  
  
y ante la superficie *silenciosa*  
del ébano sutil cuya tersura  
repite como un sueño la blancura  
de un *vago mármol* o una *vaga rosa*,  
  
hoy, *al cabo de tantos y perplejos*  
*años de errar bajo la varia luna*,

me pregunto qué azar de la fortuna  
hizo que yo temiera los espejos.  
  
Espejos de metal, enmascarado  
espejo de caoba que en la *bruma*  
de su rojo crepúsculo disfuma  
ese rostro que *mira y es mirado*,  
  
*infinitos* los veo, elementales  
*ejecutores de un antiguo pacto*,  
*multiplicar el mundo como el acto*  
*generativo, insomnes y fatales*.  
  
*Prolongan este vano mundo incierto*  
*en su vertiginosa telaraña*;  
a veces en la tarde los empaña  
el hálito de un hombre que no ha muerto.

<sup>2</sup> Pour une traduction française de ce texte, consulter <http://membres.lycos.fr/jes/borges-biofr.htm>

Nos acecha el cristal. *Si entre las cuatro paredes de la alcoba hay un espejo, ya no estoy solo. Hay otro. Hay el reflejo que arma en el alba un sigiloso teatro.*

*Todo acontece y nada se recuerda en esos gabinetes cristalinos donde, como fantásticos rabinos, leemos los libros de derecha a izquierda.*

Claudio, rey de una tarde, rey soñado, *no sintió que era un sueño* hasta aquel día en que un actor mimó su felonía con *arte silencioso*, en un tablado.

Que haya sueños es raro, que haya espejos, que el usual y gastado repertorio de cada día incluya *el ilusorio orbe profundo que urden los reflejos.*

Dios (he dado en pensar) pone un empeño en toda *esa inasible arquitectura que edifica la luz con la terzura del cristal* y la sombra con el sueño.

Dios ha creado las noches que se arman de sueños y las formas del espejo *para que el hombre sienta que es reflejo y vanidad. Por eso nos alarman.*

### ***Les miroirs sont des vertiges menaçants***

Les miroirs sont des formes, qui comme toutes les formes “cachent ce qu'elles ne montrent pas” (Sylviane Agacinski). Cela signifie que le regard ne les traverse pas, et que les miroirs ne sont pas transparents, quoi qu'on dise lorsqu'on use de la formule “à travers le miroir” (Ingmar Bergman), ou qu'on se laisse aller aux invitations féeriques de Lewis Carroll ([3] pages 218-223).

Dans le cas du rétroviseur, les constructeurs de véhicules choisissent la taille, la position et la concavité de l'instrument pour optimiser ces contraintes, mais dans le cas du regard spéculaire, le phénomène peut prendre un tour sidérant, voire menaçant. Souvent pour le pire, quelquefois pour le meilleur : on se souvient du *Vilain petit canard* d'Andersen, “Et il pencha la tête sur la surface de l'eau, attendant la mort ... mais que vit-il dans l'eau claire ? Il vit sous lui sa propre image, mais qui n'était plus celle d'un oiseau gris tout gauche, laid et vilain. Il était lui-même un cygne. Peu importe qu'on soit né dans la cour des canards, si l'on est sorti d'un œuf de cygne ... ”.

Dans le registre du pire, Oscar Wilde a écrit un conte cruel, *L'Anniversaire de l'infante* (qui a inspiré *Le Nain*, un Opéra d'Alexandre Zemlinski), illustrant parfaitement *l'effet de dislocation* du miroir. Le Nain est un personnage qui n'a aucune capacité à la conceptualisation, n'ayant jamais été initié à la vie intellectuelle : il a été trouvé sur des rivages lointains et offert à l'infante d'Espagne, dont c'est l'anniversaire. Le Nain est laid et difforme, mais il l'ignore : il se déploie singulièrement et à son insu, et il ne sait faire que cela. Le Nain mourrait de se reconnaître, il se disloquerait de s'exposer à lui-même. Tout savoir de sa singularité serait immédiatement dé-singularisante, qui désagrégerait sa personne, *l'atomiserait* comme l'on dit parfois. L'infante s'en amuse, et apprécie le cadeau d'anniversaire. Mais elle a bientôt l'idée cruelle de lui présenter un miroir, au moment où le Nain, épris de sa jeune propriétaire, décide de mobiliser toute sa conviction pour séduire l'infante d'Espagne, incapable qu'il est d'une quelconque évaluation rationnelle de son entreprise. Il sera littéralement disloqué par la reconnaissance de son image dans le

miroir : impossible pour lui d'échouer *dans ses fins* (un concept qui lui est définitivement étranger) sans échouer *sa propre personne*. La situation s'imposait pour le Nain, en ce sens qu'elle se donnait sur le mode immédiat de la prescription impérieuse : l'infante devait l'aimer autant qu'il l'aimait. Nulle place ici pour la description de la situation, réduction inconcevable pour une singularité se déployant dans l'authenticité. Et sans description, pas de planification, pas de rationalisation, pas de réussite ou de succès, pas de regret ni de remords, pas davantage de prédition ou d'explication possible. En particulier, pas d'évaluation des chances d'aboutir à la séduction de l'infante, compte tenu d'un contexte humain rédhibitoire pour le Nain, socialement condamné à la déréliction. C'est alors que le miroir joue de son effet de dislocation, causant la chute irrémédiable du Nain, et sa mort.

Mais il arrive aussi que le miroir produise des *effets de collusion*, comme dans le cas des juges de Suzanne dans la scène biblique du *Livre de Daniel* (Daniel 13). Daniel est un enfant hébreu exilé à la cour de Nabuchodonosor, peu après la prise de Jérusalem par les Babyloniens en l'an 587 avant notre ère. Suzanne est la femme d'un notable estimé de Babylone, dont la maison abritait quelquefois les procès jugés par les deux vieillards qu'on avait cette année-là désignés comme guides du peuple. Suzanne, femme d'une grande beauté, aimait à se promener dans le jardin aux heures chaudes durant lesquelles tout le monde s'était retiré. Les deux vieillards se prirent à désirer Suzanne et, esclaves tourmentés de leur passion, se mirent à épier chaque jour ses promenades. Ils désirent Suzanne, mais ne le savent pas encore, et leur crainte de Dieu empêche toute velléité de ce désir de dévaler en convoitise opérante. Voici le passage du *Livre de Daniel* qui nous présente la scène de la collusion des vieillards : "Un jour, s'étant quittés sur ces mots "Rentrons chez nous, c'est l'heure du déjeuner", et chacun s'en étant allé de son côté, chacun aussi revint sur ses pas et ils se retrouvèrent face à face. Forcés alors de s'expliquer, ils s'avouèrent leur passion et convinrent de chercher le moment où ils pourraient surprendre Suzanne seule". C'est en effet après s'être reconnus guettant Suzanne, alors qu'ils feignaient tous deux de s'en retourner pour déjeuner, que la situation s'est imposée à eux sous la forme de la convoitise. En cette reconnaissance mutuelle d'une situation commune réside exactement l'effet de collusion du miroir. Honteux mais forcés de s'avouer mutuellement leur semblable dessein, ils convinrent de guetter ensemble l'occasion de surprendre Suzanne. C'est donc incidemment que les deux vieillards n'en feront plus qu'un, au point de parler *d'une seule voix* lors du procès de Suzanne qui aura bientôt lieu. Et il faudra tout le talent inspiré de Daniel pour les *confondre* (le juge *impartial tranche* les différends après avoir *confondu* les coupables, et l'ange inspirateur de Daniel lui fera apparaître en songe les vieillards *tranchés par le milieu*).

Ainsi, les miroirs provoquent la dé-singularisation. Et qui n'est pas armé pour se réfugier à son avantage dans le concept et la catégorie peut s'en trouver vertigineusement affecté, par effet de dislocation ou de collusion. Le vilain petit canard a cette chance qu'il se découvre appartenir à la catégorie des cygnes, mais le

Nain refuse d'accepter son appartenance au genre des prétendants indignes de l'infante, et les vieillards convoitant Suzanne n'ont pas la présence d'esprit de discerner l'instant où ils tombent sous la catégorie de ceux qui devraient derechef prier pour leur salut et en appeler vivement à Dieu, toutes affaires cessantes.

Dans certains cas, le miroir disloquera un éprouvé singulier incapable de traverser l'épreuve de l'évaluation rationnelle. Dans d'autres, il ouvrira à la révélation d'un désir prêt à se reconnaître et à opérer dans le champ descriptif de la planification. Le miroir, pour le meilleur ou pour le pire, révoque l'exclusive de l'éprouvé prescrivant, et convoque l'aventure descriptive.

### **3° Le miroir et la répétition**

Comment les responsables d'inversions étranges et de vertiges quelquefois fatals, comment les complices de l'institution d'un espace chroniquement inhabitable, en conflit avec l'espace chorégraphique, comment donc les miroirs se trouvent-ils tolérés et même désirés par les danseurs de ballet ?

#### ***Miroir et réflexion***

Si les miroirs sont tellement traîtres et dangereux, comment expliquer qu'ils soient si utiles aux danseurs de ballet désireux de répéter leurs mouvements, leurs déplacements et leurs chorégraphies ? En y regardant de près, on découvre que c'est seulement la *fonction de réflexion* qui est suspectée ici. Encore convient-il de rappeler que cette fonction a bien entendu sa productivité, et que la suspicion ne concerne que les modalités de son fonctionnement, même si ces modalités peuvent avoir des conséquences extrêmement profondes.

Précisons ce que nous entendons par "fonction de réflexion" du miroir, et voyons à quoi nous l'opposons.

La réflexion, c'est pour nous l'ouverture du vécu au symbole et au concept, qui passe toujours par la description. L'attitude réflexive est ainsi condition de possibilité de la prédiction et de l'explication, des projets et des plans, des succès et des échecs, mais aussi des satisfactions rationnelles, et de l'apprentissage du danseur.

En effet, le singulier d'une situation qui s'impose et qui s'éprouve immédiatement comme prescriptive jouxte toujours sa réduction descriptive, qui n'épuise jamais l'éprouvé mais lui ouvre un avenir et une médiation intersubjective. C'est cela la fonction de réflexion par le miroir, une dé-singularisation sur laquelle nous reviendrons, et qui est essentielle à la danse. Nous serions même tentés d'avancer que la *danse de ballet* est le nom qu'on donne aux spectacles vivants qui sont traditionnellement répétés face au miroir.

Danser consisterait alors à apprésenter son corps, se mouvant et se déplaçant dans un environnement chorégraphique, par le truchement d'un regard doté du pouvoir fantasmagorique que lui confère le reflet spéculaire.

Il faudrait comprendre comment la réflexion est indispensable au travail de répétition, ce qui n'est pas tâche facile. Nous avons dit que les miroirs cachent ce qu'ils ne montrent pas, mais il faut dire aussi qu'ils montrent ce qu'ils ne cachent pas. Le travail d'apprésentation et de contrôle de l'environnement chorégraphique par le biais du miroir permet de sélectionner et de concentrer l'attention, et par là d'engrammer le geste. Car les miroirs n'ont pas de mémoire, et seul le corps en a, qui élabore des gestes toujours plus lisibles en intentionnalité, de moins en moins marqués par la succession des présents instantanés qui le compose encore lorsqu'on commence à répéter<sup>3</sup>.

Mais si le reflet dans le miroir relève de la *réflexion*, il relève aussi, et à la fois, de *l'exposition*. Les deux attitudes induites sont en tension permanente et irréductible. Pour bien comprendre cela, il nous faut revenir à la notion de *singularité*, telle que nous la comprenons.

### ***Singularité et exposition***

Dans son exorde à la *Guerre des Gaules* Jules César, reconnaissant le rôle novateur du miroir comme condition de possibilité de l'introspection, pressentait peut-être que l'invention spéculaire touche à l'essence du singulier.

Remarquons en effet que le singulier ne manquerait pas de se dé-singulariser en se donnant comme tel : le singulier ne pourrait demeurer dans le savoir de sa propre singularité sans la trahir, et toute réflexion du singulier sur sa singularité l'annihilerait aussitôt. Le singulier présent à lui-même en signifierait ainsi la fin immédiate, et la donation d'une singularité sur le mode de la reconnaissance signifie sa dé-singularisation par catégorisation.

Posons donc qu'une singularité ne saurait se connaître immédiatement. Par conséquent, elle ne peut que procéder *en singularisation*, elle est procès de déploiement. Mais surtout, la singularisation est *ex-position* à, aux sens où l'on s'expose à quelque chose. On s'expose à un devenir ou à un risque : les deux sens n'en font qu'un, car quand le premier sens pointe sur l'altérité et la diversité, le second pointe sur le caractère aporétique d'un geste qui, parce qu'il n'est pas position mais ex-position (hors de), est par nature public et violent (viol de l'intimité, du latin *intimus* signifiant l'intérieur, le privé), et évoque l'adversité, la controverse et la polémique.

Paradoxalement, l'activité de déploiement d'une singularité est ainsi caractérisée par la distraction, la dissipation et le divertissement, au sens pascalien. Mais toujours dans l'ex-position.

---

<sup>3</sup> La vitesse d'un mouvement, par exemple, ne peut pas être acquise de façon linéaire, par simple augmentation ou ralentissement d'un geste déjà maîtrisé : la vitesse n'est pas un simple paramètre. En revanche, on se souvient des propositions d'un Stanislavski ([14]) insistant pour que l'intentionnalité habite chaque geste théâtral, en opposition à Diderot ([5]).

Nous sommes là devant une difficulté : le danseur de spectacle vivant ne s'expose véritablement au public que le jour de la création du spectacle et les jours de représentation. Comment donc peut-il travailler sa chorégraphie, si l'exposition lui fait défaut alors même qu'il doit s'y préparer pour la représentation ?

Pour tenter de répondre à cette question, il convient d'enquêter sur la *pratique de la répétition*, qui doit être aussi, paradoxalement, *préparation à l'exposition*.

### **Répétitions, création et représentations**

S'il est vrai qu'un spectacle peut être donné plusieurs fois de suite après sa création, cette reproduction là n'a rien à voir avec la répétition du spectacle.

Afin de pouvoir se produire devant les spectateurs, les danseurs s'investissent en effet dans un processus d'élaboration destiné à garantir la possibilité d'une présentation publique du spectacle qui soit à la fois efficace et reproductible.

A la différence du spectacle donné à voir au *public*, les répétitions constituent des étapes de travail préparatoires, *privées*. La *répétition* se distingue ici de la *représentation*, et “répéter le spectacle” se distingue de “se produire en public”.

Sans discuter ici la question subtile de la répétition et de son rapport à la différence (sur ce sujet passionnant, mieux vaut en effet se reporter au maître ouvrage de Gilles Deleuze [4]), on remarquera immédiatement un phénomène étrange : alors que la répétition est supposée ouvrir à la représentation par *déploiement progressif*, le passage du privé — qui caractérise la répétition, au public — qui caractérise la représentation, a nécessairement lieu *en rupture*.

Ainsi, et même en admettant que la répétition possède ce pouvoir étrange de condenser en différence capitalisable par progrès réfléchi (on l'a dit, l'approfondissement de ce point nécessiterait évidemment à lui seul un travail considérable), la question de la répétition, dans le domaine du spectacle vivant, reste hantée par la rupture privée-publique, qui ne peut en aucun cas être lissée par une quelconque solution de continuité.

Peut-on se voir comme on serait vu ? Peut-on *être son propre public* ? En travaillant devant un miroir pendant les répétitions, les danseurs cherchent certes à contrôler leurs mouvements, à synchroniser leurs déplacements et à éprouver le vertige de la perception spéculaire. Toutes choses qui relèvent de la réflexion, et qui supposent de traverser les multiples pièges tendus par le cristal réfléchissant.

Mais ces menaces du reflet ne sont pas de même nature que celles qu'ils préfigurent par ailleurs, terribles et imprévisibles dans leurs modalités, celles du public qu'il faudra affronter bientôt. Ainsi surtout les danseurs cherchent-ils à lisser le hiatus qui sépare toujours la répétition de la représentation, en étant spectateurs d'eux-mêmes par le truchement du miroir.

A ce titre, les miroirs annoncent déjà l'exposition au public. Ils *simulent* le public.

Plus précisément, disons que le propre regard du danseur simulerait celui du public, ce qui confèrerait à la danse un statut spécial dans le champ de l'intersubjectivité.

Reste que la danse contemporaine s'échappe précisément à nier cette hypothèse, et à exhiber d'autres modes d'apprésentation du corps propre, qui conduiraient à reléguer les miroirs loin des studios de répétition. Ceux des spectacles de danse ainsi conçus et répétés qui résistent au choc abrupt de la publication démontrent que quelque chose du regard du public était présent dès l'origine chez les danseurs; et cela aussi, il faudrait pouvoir le penser.

Fichte, dans son introduction au *Fondement du droit naturel selon les principes de la doctrine de la science*, cherche à répondre à la question "Comment une science philosophique réelle se distingue d'une simple philosophie de formules ?". Sa première assertion est alors "Le caractère de la rationalité consiste en ce que l'être agissant et l'être agi sont un et même; et par cette description la raison comme telle est circonscrite de façon exhaustive".

La danse comme spectacle vivant cherche à décliner la rationalité de Fichte dans le champ du mouvement, en chair et en os, des corps humains : corps propres et corps dansés. Et le miroir est un élément central de l'étrange organologie associée, en tant qu'il permet de constituer de la mémoire par engramme corporel, sans jamais lui-même rien retenir.

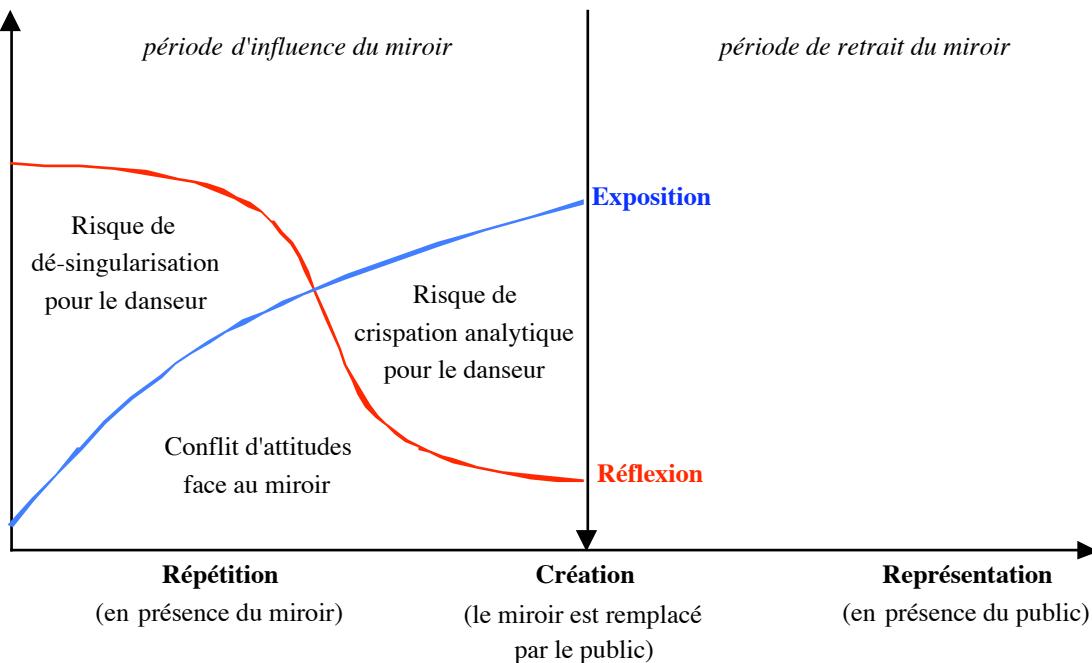
## 4° Dispositions

Pendant la période de répétition, les danseurs sollicitent les miroirs de deux manières contradictoires, à travers deux attitudes en tension irréductible.

D'une part, ils voudraient trouver dans le reflet un lieu d'exposition, qui ferait d'eux-mêmes le public de leur propre corps exposé (le miroir comme *ancre*). Cette attitude est nécessaire au déploiement en singularité du geste chorégraphique, et permet que quelque chose du public soit présent dès l'origine privée du travail. D'autre part, ils voudraient trouver dans le miroir quelque chose de leur corps propre, afin de le réfléchir dans le cadre d'un projet chorégraphique (le miroir comme *point de vue extérieur*).

Ces deux regards sont en fait juxtaposés et concurrents, la tension donnant lieu à des conflits d'attitude, et présentant des risques pour le danseur. Mais leur présence compétitive à travers le miroir est aussi une formidable tension productive, sans doute essentielle à la danse, au centre de la nécessaire disposition des danseurs.

Il semble d'ailleurs qu'on retrouve chacune de ces deux tendances, techniquement affirmées en oppositions à l'autre, dans des pratiques de danse qui ne sont pas des danses de ballet. Par exemple dans la mouvance "Fictive body" (voir Eugenio Barba et Nicola Savarese dans *L'énergie qui danse*, Lectourne Bouffonneries, 1995), le danseur cherche à être *lui-même* en même temps qu'un *autre*.



## Références bibliographiques

1. Antonin Artaud, *Le théâtre et son double*, Œuvres complètes, tome IV, Gallimard, Paris, 1964
2. Jorge Luis Borges, *Œuvres complètes*, La Pléiade Gallimard, 1993
3. Lewis Carroll, *Tout Alice*, Garnier-Flammarion, 1979
4. Gilles Deleuze, *Différence et répétition*, Presses Universitaires de France, 1968
5. Denis Diderot, *Paradoxe sur le comédien*, Larousse, 1934
6. Johann Gottlieb Fichte, *Fondement du droit naturel selon les principes de la doctrine de la science*, PUF, 1998
7. Eugen Fink, *Le jeu comme symbole du monde*, Les éditions de minuit, 1966
8. Witold Gombrowicz, *Ferdydurke*, Gallimard, 1966
9. Edmund Husserl, *Idées directrices pour une phénoménologie*, Tel Gallimard, 1950
10. Edmund Husserl, *La terre ne se meut pas*, Les Editions de minuit, 1989
11. Joann Kealiinohomoku, *An Anthropologist Looks at Ballet as a Form of Ethnic Dance*, Anthropology and Human Movement: The Study of Dances, vol. 1, pp. 15-36. Scarecrow Press, 1997
12. Agnès Minazzoli, *La première ombre*, Les Editions de Minuit, 1990
13. Friedrich Nietzsche, *Ainsi parlait Zarathoustra*, Gallimard - Folio / essais n° 8, deuxième partie : "L'enfant au miroir", pages 108-109, Paris, 1985
14. Constantin Stanislavski, *Ma vie dans l'art*, traduction, préface et notes de D. Yoccoz, L'Âge d'Homme, Paris, 1980
15. Andeï Tarkovski, *Zerkalo (Le miroir)*, film 106 min, couleur, avec Margarita Terekhova, Oleg Lankovski, Philip Lankovski et Ignat Daniltsev, URSS, 1974
16. Hermann Weyl, *Symétrie et mathématique moderne*, Champs Flammarion, 1964



# Le soi dans le regard des autres

## un essai sur le système philosophique de Jean-Jacques Rousseau

*Yûji SAKAKURA*

Université Rikkyo, Faculté des lettres

3-34-1, Nishi-Ikebukuro, Toshima, 171-8501, Tokyo, Japon

[sakakura@rikkyo.ne.jp](mailto:sakakura@rikkyo.ne.jp)

« La plus utile et la moins avancée de toutes les connaissances humaines, dit Jean-Jacques Rousseau (1712-78) dans la préface de son *Discours sur l'inégalité*, me paraît être celle de l'homme et j'ose dire que la seule inscription du temple de Delphes [« connais-toi toi-même »] contenait un précepte plus important et plus difficile que tous les gros livres des moralistes » (*OC*, t.III, p.122)<sup>1</sup>. En effet, la référence à soi-même reste toujours un des pôles les plus importants dans la tradition philosophique en occident. En un sens, celle-ci est qualifiée comme des interprétations variées de la formule socratique, « connais-toi toi-même ». Il s'agit dans cette communication d'une tentative d'analyse d'un aspect important de la philosophie moderne qui vise à se connaître soi-même, la quête de la nature humaine. Nous examinons, d'une part, comment le philosophe genevois comprend la nature humaine dans son système philosophique qui exige de distinguer « l'amour de soi » de « l'amour-propre », et d'autre part, quel sens y a-t-il dans cette distinction, en resituant le système philosophique de Rousseau dans le contexte philosophique et théologique des XVII<sup>e</sup> et XVIII<sup>e</sup> siècles.

### 1. Démonstrations de « la bonté originelle de l'homme »

Rousseau, qui a travaillé dans des domaines variés, avertit souvent ses lecteurs de l'unité de ses pensées. Il a écrit, par exemple dans la *Lettre à Ch. de Beaumont*, « le principe fondamental de toute morale, sur lequel j'ai raisonné dans tous mes écrits, [...] est que l'homme est un être naturellement bon, aimant la justice et l'ordre;

qu'il n'y a point de perversité originelle dans le cœur humain, et que les premiers mouvements de la nature sont toujours droits » (*OC*, t.IV, pp.935-936). Pour prouver « la bonté originelle de l'homme », il essaie de montrer que tous les maux attribués à la nature humaine viennent de l'extérieur de notre nature. Il aperçoit dans la nature humaine « deux principes antérieurs à la raison, dont l'un nous intéresse ardemment à notre bien-être et à la conservation de nous-mêmes [l'amour de soi], et l'autre nous inspire une répugnance naturelle à voir périr ou souffrir tout être sensible et principalement nos semblables [la pitié] » (*Discours sur l'inégalité*, *OC*, t.III, pp.125-126). Ici, il exige de distinguer « l'amour de soi » de « l'amour-propre » (*OC*, t.III, p.219).

Selon Rousseau, « l'amour de soi » fait prendre les soins nécessaires pour la conservation et la jouissance indifférente à autrui, c'est-à-dire, la jouissance qui n'est pas issue d'une comparaison avec les autres. Pour fixer la notion de cet amour, le philosophe suppose un homme qui n'a aucune relation avec les autres personnes, parce que, d'après lui, « le plus sûr moyen de s'élever au-dessus des préjugés et d'ordonner ses jugements sur les vrais rapports des choses, est de se mettre à la place d'un homme isolé, et de juger de tout comme cet homme en doit juger lui-même, eu égard à sa propre utilité » (*Émile*, *OC*, t.IV, p.455).

Au contraire, « l'amour-propre » qui fait s'aimer soi-même dans les relations avec les autres, et qui exige de se comparer aux autres, de s'y préférer, et de s'en distinguer, ne peut jamais être satisfait (cf. *OC*, t.I, pp.805-806, *OC*, t.IV, p.493). Cet amour est qualifié même comme « le principe de toute méchanceté des hommes » (*OC*, t.I, p.789). Un homme dirigé par cet amour cherche souvent un jugement ou un témoignage favorable de soi dans les yeux des autres, et perd sa liberté pour devenir l'esclave de l'opinion d'autrui. Il s'agit d'aimer le soi dans le regard des autres. Cela signifie que l'on cherche plus l'image de soi qui soit approuvée par les autres que le soi qui se sent heureux. C'est à cette « contradiction avec soi-même » (*OC*, t.IV, p.491) que Rousseau impute la misère et le malheur des hommes.

Cependant, malgré la théorie admise qui contraste le *mauvais* « amour-propre » avec le *bon* « amour de soi », et qui oublie le rôle que « l'amour-propre » joue dans la pensée rousseauiste, on trouve plus d'une fois que le mot « amour-propre » contient certaines valeurs positives dans les écrits rousseauistes, surtout, dans les deux derniers livres de l'*Émile* et la *Nouvelle Héloïse*<sup>2</sup>.

Il semble que l'on trouve de la confusion dans les écrits rousseauistes. En effet, d'un côté le philosophe insiste sur le contraste entre « l'amour-propre » et « l'amour de soi », en qualifiant celui-là comme le principe de tous les dérèglements, et de

l'autre, il dit que « l'amour-propre » est « indifférent par lui-même, bon ou mauvais par les accidents qui le modifient et qui dépendent des coutumes, des lois, des rangs, de la fortune, et de toute notre police humaine » (*La Nouvelle Héloïse*, IV-12, *OC*, t.II, p.491). On peut se demander ici pourquoi il a insisté sur la différence de ces deux amours ? Ce doute nous en inspire un autre : peut-on considérer « l'amour de soi » comme le fondement de la pensée rousseauiste, quoique le fassent bien des lecteurs ?

S'il me permet de simplifier, il y a deux interprétations influentes de la pensée rousseauiste, qui ne nous donnent pas toutes les deux de solution convenable aux énigmes mentionnées ci-dessus.

1<sup>o</sup> Bien des lecteurs qui insistent sur l'importance de l'État idéal proposé par l'auteur du *Contrat social*, trouvent dans les écrits rousseauistes une structure dialectique hégélienne, qui essaie d'établir une réconciliation de la nature et la culture qui a corrompu celle-là (cf. la figure 1). En effet, Rousseau essaie de montrer « dans l'art perfectionné la réparation des maux que l'art commencé fit à la nature », en s'efforçant « de tirer du mal même le remède qui doit le guérir » (*MS de Genève*, *OC*, t.III, p.288). Il faut avouer cependant que dans cet aspect, le système philosophique de Rousseau n'est pas bien structuré, à cause de la rupture profonde entre les trois catégories qui composent cette structure<sup>3</sup>. De plus, on aura une grande difficulté d'expliquer la philosophie politique de Rousseau par le principe de « l'amour de soi », parce que hormis le *Discours sur l'inégalité*, on ne trouve jamais le mot d' « amour de soi » dans ses écrits politiques.

2<sup>o</sup> Le message d'un prétendu « retour à la nature » est attribué souvent à notre philosophe. Bien que cette interprétation vulgarisée soit appréciée par les écrivains du romantisme et par les pédagogues qui ne s'intéressent qu'aux premières pages de l'*Émile*, elle trahit l'auteur qui dit que « la nature humaine ne rétrograde pas et jamais on ne remonte vers les temps d'innocence et d'égalité quand une fois on s'en est éloigné » (*Dialogues*, *OC*, t.I, p.935).

Il faudrait une autre lecture pour résoudre les énigmes dont il s'agit. Avant de la donner, il serait convenable de resituer le système philosophique de Rousseau dans le contexte philosophique et théologique aux XVII<sup>e</sup> et XVIII<sup>e</sup> siècles.

## 2. Moi est-il haïssable ?

Le sentiment qui cherche sans limite le bonheur physique, la richesse, la réputation est qualifié, dans le christianisme, comme une preuve du péché originel. Quelques théologiens de XVII<sup>e</sup> siècle, accentuent l'opposition augustinienne entre *amor Dei* et *amor sui*. On peut rappeler entre autres le *Traité de l'Amour de Dieu* (1616) de Saint

François de Roussy de Sales (1567-1622), le *Traité des passions* (1614) de Jean-Pierre Camus (1584-1652), et ainsi que les œuvres des jansénistes. En insistant sur « les misères de l'homme » et « la faiblesse de l'homme », ces écrits nient la possibilité du salut sans la grâce de Dieu. Les théologiens mentionnent « les misères », parce qu'il n'y a pas plus grande jouissance spirituelle que l'amour de Dieu. « La faiblesse », parce que l'homme ne sait même pas se connaître soi-même. « Nous sommes si accoutumés à nous déguiser aux autres qu'enfin nous nous déguisons à nous-mêmes » remarque La Rochefoucauld (1613-80) dans ses *Maximes* §119. Dans ce contexte, la seule vertu possible consiste dans la « Haine de nous-mêmes »<sup>4</sup>.

Il y a certains théologiens, d'autre part, qui ont voulu relativiser ces arguments simplifiés qui nient totalement l'amour-propre. Ils ont réhabilité une partie de l'amour-propre, partie compatible avec l'amour de Dieu, sous le nom de « l'amour-propre bien éclairé » ou « l'amour de soi-même » (cf. la figure 2) .

« On peut dire même que l'amour-propre entre si essentiellement dans la définition des vices et des vertus, que sans lui on ne saurait bien concevoir ni les uns ni les autres. [...] L'usage de notre langue est heureux [...], car elle nous fait distinguer entre l'amour-propre et l'amour de nous-mêmes. L'amour de nous-mêmes est cet amour, en tant qu'il est légitime et naturel. L'amour-propre est ce même amour, en tant qu'il est vicieux et corrompu »<sup>5</sup>.

Ces théologiens s'efforcent de favoriser « l'amour de soi » ou « l'amour intéressé » qui souhaite la félicité, parce qu'il s'agit d' « une manière de s'aimer soi-même, beaucoup plus noble, plus sensée que toutes les autres »<sup>5</sup>.

Malgré ces théologiens qui veulent protéger la Religion, les philosophes du siècle des Lumières se servent des vocabulaires élaborés dans les querelles théologiques (« l'amour-propre bien éclairé », « le plaisir », « l'intérêt »...) pour réhabiliter totalement la nature humaine (cf. la figure 2).

Le moi n'est plus haïssable. Ce sont les vices privés (vanité, luxe...), dit Bernard Mandeville (1670?-1733), qui fournissent le profit public et la prospérité de l'État, en sollicitant le raffinement des arts. Cet argument, qui a été bien critiqué, a exercé une grande influence sur les discours sur la morale au XVIII<sup>e</sup> siècle. Ce qui est important ici, c'est que l'ordre social est fourni indépendamment de la volonté et l'intention des individus qui ne se soucient que leur propre intérêt. Il faudra attendre cependant Adam Smith (1723-90), pour une explication théorique de l'invention de l'ordre social issu de l'intérêt personnel.

La laïcisation de la morale, basée sur la nature humaine, a altéré le sens de « la faiblesse de l'homme ». Il ne s'agit plus de la connaissance intérieure de soi-même.

Un homme est faible, parce qu'il manque de moyens de satisfaire ses désirs indéterminés. En considérant les désirs humains comme les fondements de la société, Helvétius signale : « C'est l'émulation qui produit les génies, et c'est le désir de s'illustrer qui crée les talents, c'est du moment où l'amour de la gloire se fait sentir à l'homme, et se développe en lui, qu'on peut dater les progrès de son esprit »<sup>6</sup>.

Dans l'*Encyclopédie*, L. de Jaucourt (1704-79) définit le mot « l'émulation » comme la « passion noble, généreuse, qui admirant le mérite, les belles choses, et les actions d'autrui, tâche de les imiter, ou même de les surpasser, en y travaillant avec courage par des principes honorables et vertueux », en le distinguant « d'une ambition désordonnée, de la jalousie, et de l'envie ». Ici, on ne voit pas l'incohérence entre l'éclosion des talents des individus et le progrès social.

Si l'on mesure le résultat de concurrences liées à l'émulation par la différence relative les uns avec les autres, il n'y en a pas de limite. Le désir de se distinguer augmente sans fin, comme le remarque Ch. L. de Montesquieu (1689-1755):

« Plus il y a d'hommes ensemble, plus ils sont vains et sentent naître en eux l'envie de se signaler par de petites choses. S'ils sont en si grand nombre que la plupart soient inconnus les uns aux autres, l'envie de se distinguer redouble, parce qu'il y a plus d'espérance de réussir. Le luxe donne cette espérance; chacun prend les marques de la condition qui précède la sienne. Mais à force de vouloir se distinguer, tout devient égal, et on ne se distingue plus: comme tout le monde veut se faire regarder, on ne remarque personne »<sup>7</sup>.

Ici Montesquieu a ajouté une annotation pour mentionner Mandeville. Les philosophes des Lumières insistent sur l'utilité de la vanité bien réglée dans la société civilisée, où le raffinement des arts, considéré comme le progrès, est fourni par le luxe, l'industrie, et le bon goût. De la même façon que la vanité engendre des désirs sans limite, le raffinement des arts pour servir à ces désirs n'a pas de limite non plus.

Si les philosophes des Lumières élargissent le champ de « l'amour de soi » pour le superposer à « l'amour-propre », Rousseau prend la direction contraire et s'efforce de rigoureusement limiter le champ de « l'amour de soi » (cf. la figure 2). Il nous reste à examiner le sens de cette limite.

### 3. Émulation et inquiétude

L'espoir du raffinement des arts est éclairé par la lumière de l'innovation liée aux sciences expérimentales, mais non sans certaines zones d'ombre. Le raffinement des arts ne fournit pas toujours de plaisir aux hommes, ni ne les rend forcément meilleurs, ainsi que le dit d'Alembert (1717-83):

« nous n'acquérons guère de connaissances nouvelles que pour nous désabuser de quelque illusion agréable, et nos lumières sont presque toujours aux dépens de nos plaisirs. La simplicité de nos aïeux était peut-être plus fortement remuée par les pièces monstrueuses de notre ancien théâtre, que nous ne le sommes aujourd'hui par la plus belle de nos pièces dramatiques ; les nations moins éclairées que la nôtre ne sont pas moins heureuses, parce qu'avec moins de désirs elles ont aussi moins de besoins, et que des plaisirs grossiers ou moins raffinés leur suffisent : cependant nous ne voudrions pas changer nos lumières pour l'ignorance de ces nations et pour celle de nos ancêtres. Si ces lumières peuvent diminuer nos plaisirs, elles flattent en même temps notre vanité ; on s'applaudit d'être devenu difficile, on croit avoir acquis par là un degré de mérite. L'amour-propre est le sentiment auquel nous tenons le plus, et que nous sommes le plus empressé à satisfaire »<sup>8</sup>.

On peut considérer que la première moitié de ce texte est écrite de la main de Rousseau. Cependant l'éditeur de l'*Encyclopédie* apprécie les arts qui flattent l'amour-propre ou la vanité. Les philosophes des Lumières font plus de cas du raffinement des arts qui élargit le décalage entre les nations « civilisées » et les peuples « sauvages », que des aspects matériels qui nous servent à subsister physiquement.

Partant de la nature humaine qui cherche la conservation et le bien-être, les philosophes des Lumières expliquent presque tous les comportements des hommes par le principe hédoniste. Le raffinement des arts peut être qualifié comme un outil pour le bonheur des hommes. Ce raffinement, cependant, élargit la possibilité des moyens pour satisfaire les besoins physiques et élémentaires des hommes d'une part, change la qualité des besoins même d'autre part, au travers du raffinement du goût. L'assaisonnement aromatique, par exemple, est pour un gastronome, qui ne se soucie pas de la famine. Il ne s'agit plus d'un moyen pour la conservation, mais, de celui pour attirer l'attention des autres .

Plus il y a les arts raffinés, plus il y a de différences dans la manière de jouir du résultat chez chacun à la fois qualitativement et quantitativement. L'inégalité éveille le désir de vivre comme les autres, qui n'est qu'une variation du désir de se distinguer. Pour les gens qui sentent l'insuffisance de leur propre puissance, il est absolument nécessaire, pour soulager cette inquiétude, de trouver l'approbation, l'applaudissement ou le respect dans le regard des autres. C'est sur les interactions entre les gens qui cherchent à donner une image favorable de soi aux autres, que les philosophes ont voulu installer une nouvelle morale dite « naturelle ». Dans ce système, être ridiculisé devient plus effrayant que mourir. C'est pourquoi les personnes vivantes dans la

société moderne cherchent à acquérir la culture et les arts raffinés qui attirent l'attention des autres, aussitôt que possible, autant que possible.

Dans la pensée moderne après Th. Hobbes (1588-1679), qui se base sur un individu dont le premier souci est de se conserver, toutes les institutions humaines sont dirigées vers ce but. Mais une fois que l'on a estimé que ce besoin fondamental est assuré, on fait moins de cas des jouissances physiques que du désir de se distinguer ou de l'envie de l'approbation d'autrui, qui soulage très souvent l'inquiétude ou sentiment de sa propre faiblesse. Dans ce contexte, l'important est de donner à ceux qui nous entourent une image favorable, ce qui nous permet de vivre en bonne harmonie.

Les philosophes des Lumières voient au cœur de la nature humaine le désir d'estime et l'inquiétude d'être mal vu, qui n'est que l'envers du premier. Les hommes sont influencés moins par leurs besoins physiques (provoqués par l'amour de soi selon Rousseau) que par les désirs appréciés dans une société civilisée (provoqués par l'amour-propre d'après Rousseau).

Selon Rousseau, « l'amour-propre », dont l'objet est le soi en relations sociales, devient bon ou mauvais selon la nature de celles-là. Pour bien maîtriser cet amour, notre philosophe arrange tout pour rendre ces relations convenables. C'est pourquoi il juge que « l'étude convenable à l'homme est celle de ses rapports » et qu'un être moral « doit s'étudier par ses rapports avec les hommes » et s'y employer « sa vie entière » (*OC*, t.IV, p.493). « L'amour de soi » d' « un homme isolé » nous indique la ligne de départ de cette étude. Le système philosophique de Rousseau n'est qu'une lutte acharnée pour transformer les relations entre les hommes<sup>9</sup> qui déterminent la forme de « l'amour-propre », et ainsi que la connaissance de soi-même, qui est dans le regard des autres.

## Conclusion

Il semble que les analyses jusqu'ici nous conduisent une nouvelle lecture des écrits rousseauistes. Tout son système est un diagnostic. L'important est la détermination de l'origine et la structure des maux dont les hommes souffrent. D'une part, il essaie de connaître « un état qui n'existe plus, qui n'a peut-être point existé, qui probablement n'existera jamais, et dont il est pourtant nécessaire d'avoir des notions justes pour bien juger de notre état présent» (*OC*, t.III, p.123). D'autre part, il insiste sur le fait qu' « il faut savoir ce qui doit être pour bien juger de ce qui est » (*OC*, t.IV, pp.386-387). Chez notre philosophe, la notion de nature et celle d'art perfectionné fonctionnent toutes les deux comme un miroir théorique qui reflète les hommes tels

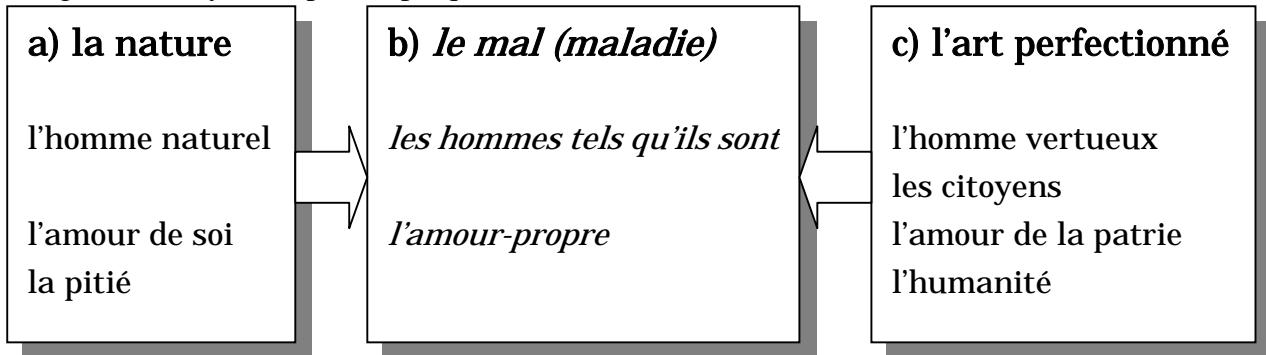
qu'ils sont (cf. figure 1). Ces notions n'indiquent qu'une fiction opératoire.

Si notre hypothèse est juste, il ne nous reste qu'à examiner l'efficacité du système philosophique de Rousseau dans les maux qui nous entourent.

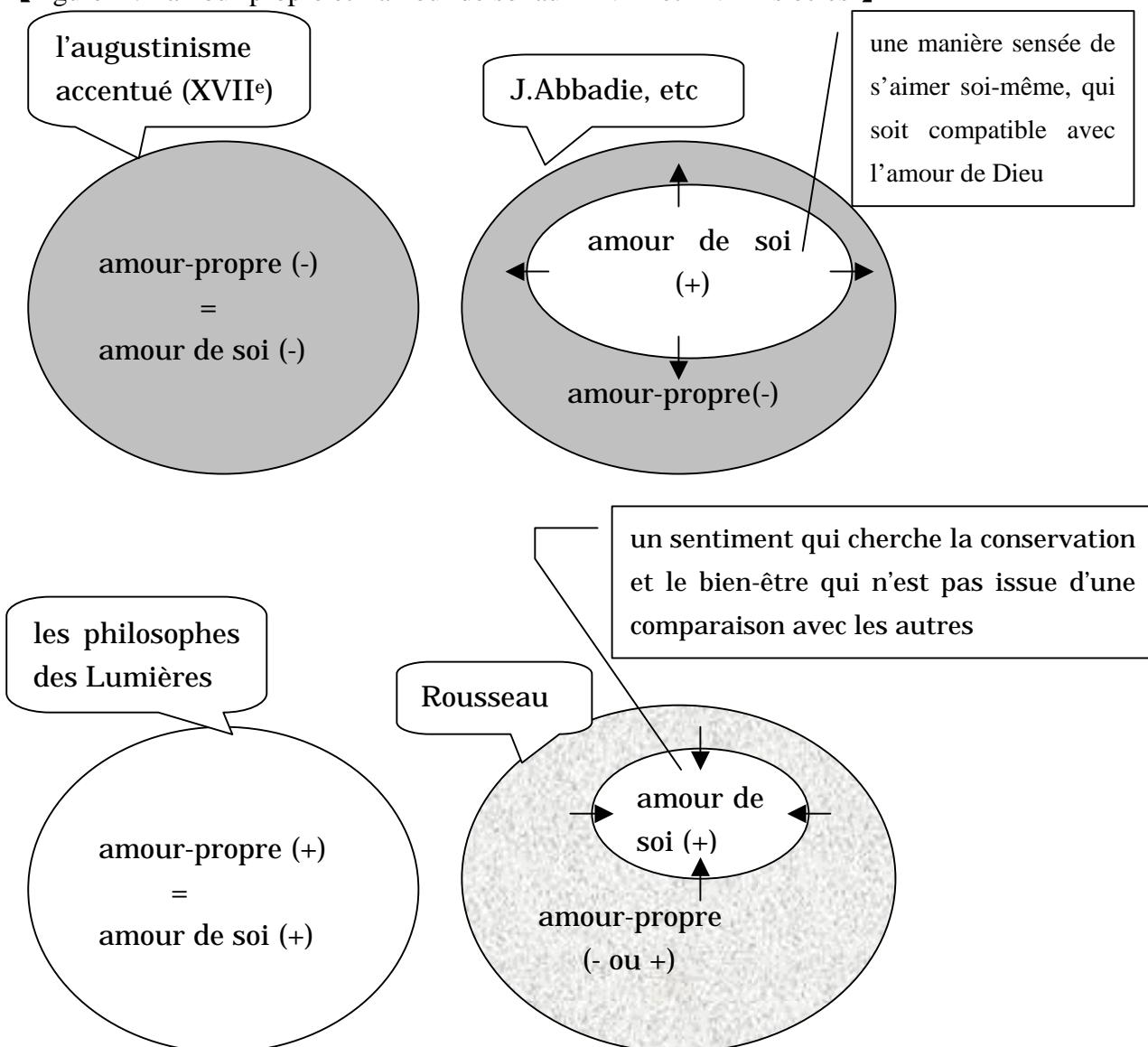
## Notes

1. Pour les textes de Rousseau, les indications de pages renvoient aux *Œuvres complètes* dans la collection de la Bibliothèque de la Pléiade, Gallimard, 5 vol., 1959-95. Le soulignement des mots dans les textes cités est de notre fait, [...] marque nos omissions, et [ ] nos ajouts. Nous avons modernisé l'orthographe.
2. cf. Yûji Sakakura, "Essai sur « l'amour-propre » chez Jean-Jacques Rousseau", in *Études J.-J. Rousseau*, Musée J.-J. Rousseau, Montmorency, N° 7, 1995.
3. Rousseau remet la cause du passage de l'état naturel à l'état social, par exemple, aux « accidents de la nature » (*OC*, t.V, p.402, cf. *OC*, t.III, pp.168-169, 533).
4. Jean-François Senault, *De l'usage des passions*, Fayard, 1987, p.201. cf. B. Pascal, *Les pensées*, §B468, 471, 476, 477, 485, 492, 545.
5. Jacques Abbadie, *L'art de se connaître soi-même*, Fayard, 2003, p.150. cf. Jean-Pierre Camus, *La défense du pur amour contre les attaques de l'amour-propre*, 1640, Paris, G.Alliot, 1640, pp.12-20. Nicolas de Malebranche, *Trois Lettres du P. Malebranche au R. P. Lamy*, in *Œuvres complètes*, Vrin, 21 vol, 1962-70, t.XIV, p.105. Marie Huber, *Lettres sur la religion essentielle à l'homme, distinguée de ce qui n'en est que l'accessoire*, Nouv. édit., Londres, 1756, 6 vol., t.II, pp.96-99, 102-106.
6. Helvétius, *De l'homme*, Londres, 2 vol., 1773, t.I, p.39.
7. Montesquieu, *De l'esprit des lois*, in *Œuvres complètes*, la Bibliothèque de la Pléiade, Gallimard, 2 vol., 1949-51 t.II, pp.333-334.
8. d'Alembert, "Réflexions sur l'usage et sur l'abus de la philosophie dans les matières de goût", in *Œuvres complètes*, Slatkine, 5 vol., 1967, t.IV, p.333.
9. Il s'agit, par exemple, dans la formule du pacte social proposée par Rousseau, comme le signale B. Baczko, « non pas d'une simple redéfinition de noms, mais d'une transformation des rapports entre les individus qui forment la communauté, ainsi qu'entre chacun des membres et la communauté elle-même » (*Solitude et communauté*, Mouton, 1974, p.312). Cependant, comme le remarque A. Philonenko, toutes les tentatives de remède proposées par Rousseau échouent dans ses écrits. *J.-J. Rousseau et la pensée du malheur*, Vrin, 3 vol., 1984.

【figure 1 : le système philosophique de Rousseau】



【figure 2 : l'amour-propre et l'amour de soi aux XVII<sup>e</sup> et XVIII<sup>e</sup> siècles】





# Mise en abyme et théorèmes de points fixes.

Sylviane R. Schwer

Université de Paris 13 en délégation CNRS au LaLICC-UMR8139

Université Paris-Sorbonne, ISHA

96 bld Raspail, Fr. 75006 Paris.

e-mail : Sylviane.Schwer@paris4.sorbonne.fr

## Résumé

Nous analysons dans cet article l'analogie formelle existant entre le procédé littéraire de mise en abyme et les calculs de points fixes en mathématiques et informatique fondamentale.

**mots clefs** : mise en abyme, points fixes, complétude, continuité.

L'autoréférence, véritable figure de style, fournit moultes paradoxes et constructions en abyme. En ce qui concerne les paradoxes, on attribue généralement à Epiménide, prêtre et poète grec de la fin du VII<sup>ème</sup> siècle avant notre ère, originaire de Cnossos, dans l'île de Crète, le premier paradoxe, connu sous le nom de paradoxe du menteur : *Il n'y a pas plus menteur qu'un Crétien, et vous pouvez me croire, car je suis Crétien moi-même.* Cet énoncé n'est pas réellement un paradoxe car il porte sur une collection d'individus décrite par l'un d'entre eux, inscrit dans le langage naturel, avec tout ce que cela comporte d'ambiguïtés liées au contexte (implicite) d'énonciation de l'énoncé. Paul, dans son épître à Tite (I, 12), dénature le texte écrivant "L'un de leurs propres prophètes [Epiménide] l'a dit : Crétien toujours menteurs". Cet énoncé n'est pas non plus auto-contradictoire. En effet, la boucle se brise facilement, et encore plus facilement selon l'interprétation non temporelle de l'adverbe "toujours". En revanche, la phrase d'Eubulide *la présente phrase est fausse* affirme elle-même sa propre négation et est en cela réellement paradoxale. Ce second énoncé est un jugement clos d'un individu (l'énoncé) sur lui-même. Ce sont ceux qui fournissent les vrais paradoxes.

Dans la poétique d'André Gide<sup>1</sup>, la mise en abyme est le procédé qui consiste à enchâsser dans un texte un fragment qui le représente, la mise en abyme est une autoreprésentation diminutive. On retrouve ce procédé de mise

---

<sup>1</sup>C'est André Gide qui a introduit ce terme dans une lettre du 15 novembre 1891 à Paul Valéry [ Gide & Valéry 1955]

en abyme, avec le phénomène de rétraction dans les théorèmes de points fixes en analyse mathématique et informatique. Un point fixe d'une fonction est un point que la fonction laisse fixe, c'est-à-dire qui est égal à son image via la fonction. Le théorème du point fixe est apparu, avant et après le travail de Brouwer, sous des formes étonnamment variées, quoique équivalentes, motivées par des questions très diverses d'analyse, de topologie, de combinatoire ou de théorie des jeux. Il existe en mathématiques plusieurs centaines de théorèmes du point fixe, et des livres entiers y sont consacrés [Mawhin 1985, Goebel & Kirk 1990]. Ces théorèmes sont fondamentaux en analyse, tant en théorie qu'en pratique. Citons par exemple, le théorème des fonctions implicites, les solutions de certaines équations différentielles, les théorèmes d'inversion locale. Ces théorèmes sont également fondamentaux en informatique puisque liés à la théorie des fonctions récursives.

Dans cette communication, nous nous proposons d'explorer le principe de construction en abyme et de le mettre en relation avec les théorèmes de points fixes en mathématique et informatique fondamentale. Dans la section suivante nous rappelons le concept de mise en abyme en littérature et témoignerons de sa présence en littérature, théâtre, cinéma et dans les arts plastiques. Dans la troisième section, nous étudierons quelques problématiques des théorèmes de points fixes avant de conclure. Nous montrerons sur deux exemples très simples, l'un en analyse, l'autre en informatique, que pour qu'on puisse affirmer l'existence d'un point fixe il faut (1) définir un espace topologie – c'est-à-dire une notion de voisinage – complet – c'est-à-dire *sans trou* – (2) se restreindre à des fonctions continues pour cette topologie – en informatique, il s'agit de topologies liées à une structure d'ordre – (3) demander à la fonction de rétracter son domaine de référence.

## 1 Mise en abyme

Cette dénomination, due à André Gide, provient d'un procédé héraldique qu'il a découvert en 1891. En effet [Littré],

l'abîme<sup>2</sup> est le cœur de l'écu. On dit qu'une figure est en abîme quand elle est avec d'autres figures au milieu de l'écu, mais sans toucher aucune de ces figures.

Bien que la figure en abyme ne soit jamais *la* figure du blason, le principe qui a captivé André Gide est l'image d'un blason, accueillant en son centre, une réplique miniaturisée d'elle-même. En recherchant le terme *abyme* dans le dictionnaire International des termes littéraires [1], on trouve la définition de la *mise en abyme* comme désignant la relation de similitude qu'entretient

---

<sup>2</sup>Abyme ou abîme : les deux orthographies se côtoient.

tout élément, tout fragment avec l'œuvre qui l'inclut, principe souvent décrit de façon imagée comme un effet de miroir. Cet emboîtement s'apparente à une auto-citation. Sémantiquement, le mot abyme invoque les notions de profondeur, d'infini, de vertige et de chute. Le terme est définitivement établi dans la critique littéraire en 1950 par Claude-Edmonde Magny [Magny 1971] dans le chapitre intitulé *La mise en abyme ou le chiffre de la transcendance*. L'expression sera conservée malgré les équivalents proposés par Pierre Lafille [Lafille 1954] : *métaphore spéculaire, miroir intérieur du récit, composition en abyme, construction en abyme*. Gérard Genette en 1972 propose également *structure en abyme*.

On y trouve alors l'analyse suivante dans laquelle on reconnaît plusieurs termes du langage logique et informatique.

Le concept, qui s'est imposé à grande échelle depuis sa prise en charge par le Nouveau Roman, appartient à la vaste problématique de la réflexivité (autoreprésentation, autoréférence) et est un des outils de base de la métafiction, cette écriture littéraire qui intérieurise un commentaire sur son écriture mais aussi sur sa lecture (ou sur sa représentation dans le cas du métathéâtre).

Dällenbach [Dällenbach 1977] propose la définition plus technique suivante de mise en abyme :

Est mise en abyme toute enclave entretenant une relation de similitude avec l'œuvre qui la contient. Est mise en abyme tout miroir interne (convexe) réfléchissant l'ensemble du récit par réduplication.

Il souligne également [Dällenbach 1977, p. 60] que la racine commune de toutes les mises en abyme est la notion de *réflexivité*, au sens philosophique du terme [Foulquier], c'est-à-dire du retour de l'esprit – ou du récit – sur ses états et sur ses actes. Un énoncé supportant la réflexivité fonctionne au moins sur deux niveaux : celui du récit où il continue de signifier comme tout autre énoncé, celui de la réflexion où l'énoncé intervient comme élément d'une métasignification permettant au récit de se prendre pour thème.

La mise en abyme peut donc se caractériser par les trois paramètres<sup>3</sup> : *< objet, type, portée >* ainsi définis par Dällenbach :

**objets** Trois éléments du texte peuvent être mis en abyme, suivant la discrimination de la linguistique jakobsonnienne : l'énoncé, l'énonciation et le code du récit. La mise en abyme de l'énoncé est le rappel du "résultat d'un acte de production". Elle peut être fictionnelle ("dimension référentielle d'histoire racontée") ou textuelle ("aspect littéral d'organisation signifiante"). Au contraire, la mise en abyme de l'énonciation

---

<sup>3</sup>les Définitions sont de Dällenbach, les dénominations me sont propres.

consiste dans "la mise en abyme du contexte ou des acteurs de la production et / ou de la réception".

**type** Trois valeurs qualitatives déterminent la profondeur de la construction :  
la *réplication simple* qui est un simple rapport de similitude  
la *réplication à l'infini* dans laquelle le fragment inséré inclut lui-même un fragment ayant cette relation de similitude  
la *réduplication spéciuse*, dans laquelle le fragment est censé inclure l'œuvre qui l'inclut.

**portée** Ce paramètre traduit la discordance entre l'ordre de l'histoire racontée et celui de sa narration : prospective si la construction "réfléchit avant terme l'histoire à venir"  
rétrospective, si elle "réfléchit après coup l'histoire accomplie"  
rétro-prospective, si elle "réfléchit l'histoire en découvrant les événements antérieurs et postérieurs à son point d'ancrage dans le récit".

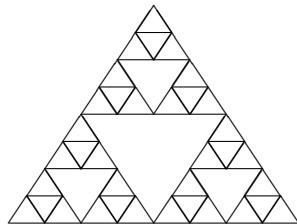
Un exemple très achevé de la mise en abyme se trouve dans "Manuscrit trouvé à Saragosse" de Jean Potocki (1761–1815) : une personne raconte une histoire où intervient la narration que lui a faite une autre personne, qui en cours de route relate à son tour un récit fait par une tierce personne, ...

En fait, cette figure si abondamment utilisée dans le nouveau roman, a déjà été décrite par Victor Hugo dès 1864 dans son *William Shakespeare* : "trente-quatre pièces sur trente-six offrent l'observation (...) une double action qui traverse le drame et le reflète en petit". Dans le théâtre, la mise en abyme se fait par des acteurs jouant des personnages qui sont eux-mêmes des acteurs. La mise en abyme la plus célèbre du répertoire dramatique théâtral est la pièce qu'Hamlet fait jouer devant le roi reproduisant les circonstances de l'assassinat du frère du roi par le roi lui-même dans le *Hamlet* de William Shakespeare. On la trouve également chez Molière, dans "L'impromptu de Versailles" [Etude]. Il s'agit dans cette pièce de théâtre de représenter le théâtre dans le théâtre, d'une part en jouant la pièce répétée mais également en exhibant l'avant répétition, c'est-à-dire la difficulté du metteur en scène pour faire jouer à ces acteurs la pièce à jouer.

Dans un cadre plus général, la mise en abîme est une représentation fractale dont les plus petits éléments reproduisent la structure de l'ensemble, avec une portée aussi longue que l'artiste le souhaite. Un exemple classique est le triangle de Sierpinski de la figure 1.

L'artiste se représentant lui-même en train de peindre dans son tableau est une pratique utilisée depuis longtemps en peinture, avec souvent un reflet pour créer une mise en abîme visuelle infinie. Du *portrait des époux Arnolfini* de Jan Van Eyck (1395 ?–1441) à l'œuvre graphique de Maurits Cornelis Escher (1898–1972) [Ernst 1992], en passant par Hans Memling (1430 ?–1494), Diego Velázquez (1599–1660), Salvador Dalí (1904–1989), les exemples abondent dans

FIG. 1 – triangle de Sierpinski : portée 3



lesquels les peintres ont utilisé les trois types de mise en abyme.

Les cinéastes ne sont pas en reste dans l'utilisation de ce procédé. La mise en abîme dans un film est une représentation de l'oeuvre dans sa propre narration. Un classique du genre est celui du film *La Nuit américaine* de François Truffaut, un film sur des acteurs qui tournent un film qui lui-même est leur propre vie. Cette technique est très prisée dans le film à caractère merveilleux. Généralement, le premier récit n'a rien de surnaturel alors que le second se traduira par un univers magique. Voilà donc comment débute le deuxième long métrage d'Elia Suleiman, *Intervention Divine*, longtemps intitulé *Chronique d'amour et de souffrance*. L'*Intervention Divine* est en fait celle de Elia Suleiman qui essaye d'écrire le scénario de *Intervention Divine* mais qui a du mal à y réussir à cause de son père mourant, de sa fiancée qu'il ne peut voir qu'à un barrage et des "événements" dans leur globalité... Cette construction en abîme n'est pas sans rappeler l'écriture de Milan Kundera dans un livre comme *la Lenteur*.

Enfin, les publicitaires ne sont pas en reste pour utiliser cette construction. Sur le modèle de la boîte de "vache-qui-rit", créée en 1922, sur laquelle on voit une vache rouge portant des boucles d'oreilles qui sont des boîtes de "vache-qui-rit". Concluons ce rapide panorama en remarquant qu'à l'ère où la télévision remplace le miroir, une étape initiale nouvelle dans la mise en abyme se dessine qui consiste à prendre comme objet de départ non plus un objet du thème exposé mais le spectateur lui-même, qui est l'actant principal que l'on fait entrer dans l'image à travers un personnage de la scène, qui peut être nommé ou non. Par exemple, dans la méthode interactive d'apprentissage de langues [Echolangues], un personnage, George, en premier plan, cadré de dos, regarde la télévision. L'élève regarde donc George qui regarde la télévision. Puis, par effet de zoom avant, l'élève prend la place de George regardant la télévision : il y a substitution de George par l'élève.

## 2 Théorèmes de points fixes

Puisque les mathématiciens focalisent leur intérêt sur la limite, nous ne considérons ici que des réplications à l'infini et des réplications spécieuses<sup>4</sup>. Dans cette partie, avant d'énoncer les théorèmes de points fixes, il nous faut rappeler quelques définitions et propriétés des domaines dans lesquels ils s'appliquent. En effet, le contexte est primordial, non pour obtenir un point fixe, mais pour d'une part s'assurer de leur existence et d'autre part avoir les moyens de les calculer, ou plus exactement de calculer l'un d'entre eux, possédant le caractère privilégié d'être soit unique soit le plus petit pour un ordre donné.

Etant donné un ensemble  $E$  et une fonction  $f$  dans  $E$ , ce que nous notons  $f : E \rightarrow E$ , on dit qu'un élément  $x$  de  $E$  est un point fixe de  $f$  s'il satisfait l'équation  $f(x) = x$ . Le graphe de  $f$  est une partie de l'espace  $E \times E$  à savoir  $\text{Graphe}(f) = \{(x, y) \in E \times E \mid y = f(x)\}$ .  $f$  admet donc un point fixe si son graphe à une intersection non vide avec la diagonale  $\Delta_E = \{(x, x) \mid x \in E\}$ .

La définition mathématique d'un point fixe est donc de type *réduplication spécieuse*, puisque c'est le même objet qui est nommé de part et d'autre du symbole de l'identité. Il convient ici d'ajouter un bémol, car en fait, cette identité est due au fait que le domaine de départ de la fonction égale le domaine final, et que l'on suppose donc implicitement que ces deux domaines sont confondus.

### 2.1 Théorèmes d'existence

Pour pouvoir affirmer l'existence d'un point fixe, il faut s'assurer que la fonction traverse effectivement la diagonale en un point de la diagonale, c'est-à-dire que la diagonale, donc  $E$  n'a pas de trou et que le graphe de la fonction ne saute pas par dessus la diagonale. Par exemple si  $E = \mathbb{Q}$ , l'ensemble des nombres rationnels et que le passage de  $f$  se fasse en  $\sqrt{2}$ ,  $f$  n'a pas de point fixe dans  $E$  alors qu'elle en a un dans  $\mathbb{R}$ , ensemble des nombres réels.

Pour dire que le domaine de la fonction est sans trou, on dit qu'il est *complet*. Pour dire que le graphe de la fonction est sans trou (pas de saut) on dit que la fonction est *continue*<sup>5</sup>.

**Théorème 1 (L'exemple de base)** *Si  $f$  est une fonction continue sur l'intervalle fermé  $[a, b]$ , elle possède au moins un point fixe dans  $[a, b]$ .*

En effet, le graphe de  $f$  se trouve dans le carré défini par les points  $a$  et

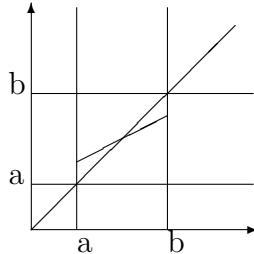
---

<sup>4</sup>Il serait intéressant aussi d'étudier le sens de l'objet et de la portée définie par Dällenbach, pour les théories de points fixes, mais nous ne le faisons pas ici

<sup>5</sup>Nous expliquerons ces notions topologiques dans le cadre des structures d'ordre

$b$ , soit  $a$  ou  $b$  sont points fixes soit *Graphe*( $f$ ) commence en  $a$  au dessus de la diagonale et se termine en  $b$  en dessous. Le graphe étant sans trou sur un domaine lui-même sans trou, il croise nécessairement la diagonale.

FIG. 2 – passage de la diagonale pour la fonction  $f(x) = \frac{1}{2}x + 2$



Ce théorème se généralise à tout espace métrique complet de dimension quelconque par le

**Théorème 2 (Brouwer-Kakutani)** *toute fonction continue de  $\mathcal{C}$  dans  $\mathcal{C}$  où  $\mathcal{C}$  est un convexe compact<sup>6</sup> de  $\mathbb{R}^n$  possède un point fixe.*

Ainsi, pour affirmer l'existence d'au moins un point fixe, il faut s'assurer d'au moins deux choses : (1) que le domaine sur lequel on travaille – ici l'intervalle  $[a, b]$  est *sans trou* (il faut aussi qu'il soit fermé) (2) que l'objet manipulé – la fonction – se comporte de façon non chaotique, i.e. qu'elle passe d'une valeur à une autre sans saut, i.e. que son graphe est sans trou.

Ces théorèmes, cruciaux en mathématiques, affirment l'existence de points fixes mais ne donnent aucun moyen de les trouver ni leur nombre. Pour obtenir ces informations, il faut réclamer plus à la fonction. En particulier d'être contractante, c'est à dire de réduire son domaine de départ. Cette notion est une qualité purement métrique, c'est-à-dire liée à la notion de distance<sup>7</sup>.

## 2.2 Un théorème d'unicité constructif

On va se placer dans le cas général d'un espace métrique complet, dont le prototype est  $\mathbb{R}^n$ . Dans ses conditions, la qualité pour une fonction de rétracter son domaine permet non seulement d'affirmer l'existence et l'unicité d'un point fixe, mais aussi de disposer d'un moyen de le calculer, avec l'approximation souhaitée.

<sup>6</sup>Un convexe compact est un ensemble qu'on peut déformer continûment sur  $\mathbb{B}^n$  est la boule de rayon 1 de  $\mathbb{R}^n$ ,  $\mathbb{B}^n = \{x = (x_1, \dots, x_n) \in \mathbb{R}^n \mid x_1^2 + \dots + x_n^2 \leq 1\}$

<sup>7</sup>Une distance  $\delta$  sur un espace  $E$  est une fonction qui à tout couple  $(x, y)$  de points de  $E$ , associe un réel positif ou nul  $\delta(x, y)$  satisfaisant les trois conditions suivantes : (i)  $\delta(x, y) = 0$  si et seulement si  $x = y$ , (ii)  $\delta(x, y) = \delta(y, x)$ , (iii)  $\delta(x, y) \leq \delta(x, z) + \delta(z, y)$ .

### 2.2.1 Fonctions contractantes et calcul en abîme

Une contraction d'un espace métrique  $(E, \delta)$  est une fonction  $f$  de  $E$  dans lui-même pour lequel il existe une constante réelle  $k$ , strictement inférieure à 1, telle que tous les points de  $E$  vérifient la relation  $\delta(f(x) - f(y)) \leq k\delta(x - y)$ . Une contraction est une fonction continue. On peut alors énoncer le théorème fondamental suivant

**Théorème 3** *Toute contraction  $f$  d'un espace métrique complet  $(E, \delta)$  possède un unique point fixe  $\omega$ . Ce point peut être calculé comme limite de toute suite de la forme  $x_{n+1} = f(x_n)$ , pour un choix quelconque de son premier élément  $x_0$  de  $E$ . L'approximation de ce point fixe  $\omega$  par  $x_n$  est au pire de  $\frac{k^n}{1-k}\delta(f(x_0), x_0)$  à savoir  $\delta(x_n, \omega) \leq \frac{k^n}{1-k}\delta(f(x_0), x_0)$*

Pour les fonctions pratiques, ce qui est essentiel, c'est la vitesse de l'algorithme utilisé pour calculer ce point fixe.

En fait, pour admettre un unique point fixe calculable, il n'est pas nécessaire que  $f$  soit une contraction, il suffit qu'une de ses itérées  $f^s = \underbrace{f \circ \cdots \circ f}_{s \text{ fois}}$  le soit.

Sur un intervalle fermé  $I$  de  $\mathbb{R}$ , si l'on impose en plus à la fonction  $f$  d'avoir un graphe sans "coin", c'est-à-dire d'être continûment dérivable sur  $I$ , alors on peut montrer que la suite  $(x_n)$  s'approche de la limite soit en restant toujours du côté de  $x_0$  par rapport à  $\omega$  si la dérivée est positive, soit en passant alternativement d'un côté à l'autre de la limite en l'encerclant de plus en plus près si la dérivée est négative. Dans ce dernier cas, nous avons le procédé de mise en abyme sur la suite des intervalles de bornes  $x_{2n}, x_{2n+1}$ . Dans tous les cas, d'une définition de type *réplication spécieuse*, le procédé constructif d'une des solutions, en restaurant le temps et l'espace, en fait une *réplication à l'infini*. Nous allons voir quelle structure spatio-temporelle, dans le cadre de travail de l'informaticien.

### 2.3 Ensemble ordonné complet avec plus petit élément (ECPO).

En informatique fondamentale, les objets sur lesquels on travaille sont essentiellement des fonctions programmables, les fonctions sont des fonctionnelles, c'est-à-dire des fonctions de fonctions. Le domaine de référence est l'ensemble ordonné complet avec plus petit élément, qui est un triplet  $\langle E, \leq, \perp \rangle$  tel que  $E$  est un ensemble muni de la relation d'ordre  $\leq$  pour laquelle  $\perp$  est son plus petit élément, i.e.

$$(\perp \in E) \text{ et } (\forall x \in E)(\perp \leq x)$$

et telle que toute suite croissante<sup>8</sup>  $(x_n)$  admet une borne supérieure  $\sup_n(x_n)$ . Par exemple, muni de l'ordre naturel sur les nombres, tout intervalle fermé de l'ensemble des nombres réels  $\mathbb{R}$  est un ensemble complet, ce qui n'est pas le cas dans l'ensemble des nombres rationnels  $\mathbb{Q}$ .

Une fonction est dite continue  $f$  de  $\langle E, \leq, \perp \rangle$  dans  $\langle E', \leq', \perp' \rangle$ , deux ECPO, si elle satisfait aux deux conditions suivantes :

- (1) elle est croissante i.e.  $(\forall x, y \in E)(x \leq y \Rightarrow f(x) \leq' f(y))$
- (2) elle commute avec la limite (on dit aussi qu'elle préserve la limite) i.e. pour toute suite croissante

$$(x_n), [f(\sup_n(x_n)) = \sup'_n(f(x_n))]$$

Dans ce contexte, le calcul du plus petit point fixe se fait simplement en calculant la limite de la suite  $u_n = f^n(\perp)$  comme l'asserte le théorème du point fixe dit de Kleene

**Théorème 4** *Si  $\langle E, \leq, \perp \rangle$  est un ensemble ordonné complet avec plus petit élément et  $f$  une fonction continue de  $\langle E, \leq, \perp \rangle$  dans  $\langle E, \leq, \perp \rangle$ , alors le plus petit point fixe de  $f$  est la borne supérieure de la suite  $u_n = f^n(\perp)$ .*

Ce théorème assure l'existence d'un point fixe au moins (le plus petit) et assure son calcul par approximations successives. Il est intéressant de montrer quelques ECPO pour calculer des fonctions connues. Nous allons présenter d'une part l'espace des fonctions récursives et d'autre part un espace permettant de représenter les nombres réels par approximations.

### 2.3.1 Fonctions définies récursivement

Nous prenons l'exemple de la fonction factorielle ainsi définie comme point fixe d'une fonctionnelle.

$$Fac = \lambda n. [ \text{ si } n = 0 \text{ alors } 1 \text{ sinon } n \times Fac(n - 1) ] = H(Fac)$$

La fonction factorielle est la plus petite solution de cette équation dans l'ECPO  $\langle E, \leq, \perp \rangle$  où

- $E$  est l'ensemble des fonctions partielles – non partout définies – de  $\mathbb{Z}$  dans  $\mathbb{Z}$ , ensemble des entiers relatifs,
- $\leq$  est l'ordre de prolongement ou de "mieux défini", on dit que  $f$  est mieux défini que  $g$  si quand  $g$  est défini sur la valeur  $x$ , alors  $f(x) = g(x)$
- $\perp$  est la fonction partout non définie.

En effet, en notant  $f(x) = \perp$  le fait que  $f$  est non définie en  $x$ , le calcul de  $Fac^9$  se fait ainsi :

---

<sup>8</sup>Une suite  $(x_n)$  de  $E$  est croissante si  $(\forall n \geq 1) : (x_{n-1} \leq x_n)$ .

<sup>9</sup>La fonction  $Fac'(n) = 0$  si  $n < 0$  sinon  $Fac'(n) = Fac(n)$  est un autre point fixe de  $H$

	...	-1	0	1	2	3	4	...
$\perp$	...	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	...
$H(\perp)$	...	$\perp$	1	$\perp$	$\perp$	$\perp$	$\perp$	...
$H^2(\perp)$	...	$\perp$	1	1	$\perp$	$\perp$	$\perp$	...
$H^3(\perp)$	...	$\perp$	1	1	2	$\perp$	$\perp$	...
$H^4(\perp)$	...	$\perp$	1	1	2	6	$\perp$	...
$H^5(\perp)$	...	$\perp$	1	1	2	6	24	...
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	...

Si l'on admet que la liberté d'une fonction, c'est la possibilité de choisir une valeur pour chacun des éléments de son domaine, tant qu'elle n'est pas définie sur l'un de ces éléments, elle a toute liberté de choisir, ce qui n'est plus le cas une fois qu'une valeur est calculée, car la relation d'ordre choisie oblige alors à s'en tenir à cette valeur. La fonction se calcule en contraignant une valeur de plus à chaque étape. La limite ainsi calculée se fait en perdant des degrés de libertés, en restreignant son champ des possibles, d'où l'image persistante de mise en abyme. En fait, le procédé itératif constitue le temps, nombré par le rang de l'itération. Le temps est alors, comme le dit Aristote, le nombre du mouvement suivant l'avant et l'après. A chaque occurrence d'unité temporelle est construit une nouvelle fonction partielle. L'itération se fait sur du semblable mais non de l'identique. Le temps permet de discerner chacune de ces occurrences, semblables mais non identiques. Nous sommes bien en présence d'un type de réplication à l'infinie.

### 2.3.2 Approximation des réels

Soit  $\mathbb{R}^+$  la droite réelle positive. Tout intervalle non vide  $I$  de  $\mathbb{R}^+$  est une approximation des nombres réels qui sont inclus dans cet intervalle. Soit  $\mathcal{I}(\mathbb{R}^+)$  l'ensemble des intervalles non vides de  $\mathbb{R}^+$ . On dit que l'intervalle  $J$  est une meilleure approximation que l'intervalle  $J'$  si  $J'$  contient  $J$ , et on note  $J' \leq J$ . On montre que  $\langle \mathcal{I}(\mathbb{R}^+), \leq, ]0, \infty[\rangle$  est un ECPO. Dans cet ECPO, on peut résoudre les équations  $x = f(x)$  pour les fonctions continues sur l'ECPO, comme  $f(x) = \frac{x+1}{x}$ . En effet si  $J' \leq J$  alors  $f(J') \leq f(J)$  et  $f(\cap_n J_n) = \cap_n f(J_n)$  pour toute suite croissante d'intervalles. Elle a une solution dans  $\mathbb{R}^+$  donnée par la suite d'intervalles calculée Table 1. Cette suite définit un réel unique  $x_0 = \frac{1+\sqrt{5}}{2}$ , c'est-à-dire le nombre d'or, limite d'une construction en abyme de type réplication à l'infini, dans un espace temps restauré.

On peut remarquer sur la Figure 3 que chaque étape de la construction se fait en s'appuyant sur une borne construite à l'étape précédente, c'est-à-

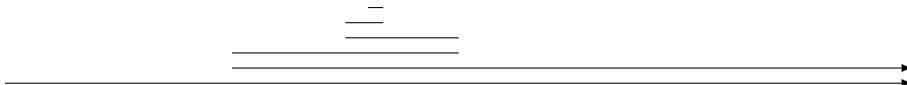
---

non calculable par ce procédé.

TAB. 1 –

	$f(\perp)$	$f^2(\perp)$	$f^3(\perp)$	$f^4(\perp)$	$f^5(\perp)$	$\dots$
$\perp = ]0, \infty[$	$]1, \infty[$	$]1, 2[$	$]\frac{3}{2}, 2[$	$]\frac{3}{2}, \frac{5}{3}[$	$]\frac{8}{5}, \frac{5}{3}[$	$\dots$

FIG. 3 – approximations de réels



dire que toutes les deux étapes, on construit un intervalle strictement emboîté dans celui construit à l'étape pénultième. C'est bien la suite des itérations de la fonction qui construit le temps, et à chacune des unités temporelles se construit le domaine résultant de l'itération, support du domaine initial de la prochaine itération. Il y a donc autant de domaines que d'étapes d'itération, donc autant de fonctions différentes.

### 3 Conclusion

Alors que dans les domaines artistiques, ces mises en abyme suggèrent simplement l'idée de chute infernale dans l'infini, soit en arrêtant le dessin dès que le mouvement est suffisamment bien lancé, soit en donnant simplement la définition récursive de base, le mathématicien va plutôt s'intéresser à la phase ultime, à savoir la limite : existe-t-elle ? Et si elle existe, est-elle dans l'espace de départ ? est-elle unique ? est-elle calculable ? La description du point fixe, sous forme d'équation  $x = f(x)$  est essentiellement statique. Tous ses efforts sont dédiés à la recherche de conditions sur le domaine, sur la fonction pour obtenir ce résultat. En revanche, l'artiste est essentiellement focalisé sur le dynamisme du procédé lui-même, qui permet, grâce à une "pichenette initiale", d'entraîner *en intention* le spectateur dans un mouvement infini, c'est-à-dire sans limite, d'où les images de vertige, de chute et donc d'abyme associées. Mais la suite des approximants de la limite calculée d'une fonction restitue ce mouvement, en recréant, par l'action même de ce calcul, le temps nécessaire à la dynamique de toute mise en abyme. C'est ainsi qu'aucun point fixe calculable n'est issu d'une réplication spacieuse, source de paradoxe, mais bien d'une réplication à l'infini.

**Remerciements :** Je remercie sincèrement Jean-Michel Autebert, Jérôme Cardot et les deux relecteurs anonymes pour les nombreuses discussions et ob-

servations autour de ce texte qui n'attend que d'être approfondi par un travail en commun.

## Références

- [1] Dictionnaire International des termes littéraires, article *abyme*.  
<http://www.ditl.info/art/definition.php?term=87>
- [Littré] Littré, dictionnaire de La Langue Française en 7 volumes.
- [Echolangues] Supplément de Accelerated English, Echolangues.
- [Etude] Etude de *l'Impromptu de Versailles*.  
<http://perso.club-internet.fr/yz2dkenn/impromptu-de-versailles.htm>
- [Dällenbach 1977] Lucien Dällenbach, *Le Récit spéculaire - essai sur la mise en abyme*,  
Seuil 1977
- [Ernst 1992] Bruno Ernst, *Les miroirs magiques de M.C. Escher*.  
Taschen 1992.
- [Foulquier] Paul Foulquier, Dictionnaire de la langue philosophique,  
Presse Universitaire de France, 1992 (6<sup>ème</sup>) édition.
- [Gide & Valery 1955] André Gide et Paul Valéry, *Correspondance 1890-1942*,  
Paris Gallimard, 1955.
- [Goebel & Kirk 1990] Kazimierz Goebel et William A. Kirk, *Topics in metric fixed point theory*.  
Cambridge University Press, 1990.
- [Lafille 1954] Pierre Lafille, *André Gide romancier*,  
Hachette (1954)
- [Magny 1971] Claude-Edmonde Magny, *Histoire du roman français depuis 1918*,  
le seuil 1971.
- [Mawhin 1985] Jean Mawhin, *Points Fixes, Points Critiques et Problèmes aux Limites*.  
Presses de l'Université de Montréal, 1985.

# Adaptation réflexive : Contrôleur de comportement et structure évolutive

Elpida S. Tzafestas

Institute of Communication and Computer Systems  
National Technical University of Athens  
Heron Polytechneiou 9  
Zographou 15773, Greece.  
[brensham@softlab.ece.ntua.gr](mailto:brensham@softlab.ece.ntua.gr)  
<http://www.softlab.ece.ntua.gr/~brensham>

## Résumé

Dans cet article, nous présentons le concept de l'adaptation réflexive comme déduit depuis trois modèles d'agents autonomes conçus spécifiquement pour des domaines différents de comportement : un agent explorateur adaptatif, une fourmis qui communique avec ses soeurs à l'aide de phéromones, un agent coopératif adaptatif. Ainsi nous comparons ces trois modèles en ce qui concerne leur mode d'adaptation en dehors de leur contexte détaillé d'opération. Nous constatons que dans les trois cas, il est question de variables "cognitives" constamment mises-à-jour et d'un mécanisme d'adaptation de deux niveaux, le premier agissant sur la variable cognitive et le deuxième sur le système d'adaptation de premier niveau. Le jeu entre les dynamiques différentes des deux niveaux d'adaptation est responsable de la performance finale de l'agent dans son environnement. Mais l'adaptation réflexive n'est pas simplement le contrôleur du comportement; elle peut servir également de structure évolutive. Plus précisément, nous pouvons imaginer une structure "paramétrée" qui sera adaptée de génération à génération de façon évolutive. Nous esquissons cette structure et nous discutons sa signification : d'une part, elle assure la convergence du comportement quant à ses performances, et d'autre part elle reste invariante à travers les problèmes posés et les applications.

## 1 Introduction : Adaptation

Dans cet article, nous passons en revue et nous comparons trois modèles d'agents autonomes qui résoudent trois problèmes typiques, l'agent explorateur (Tzafestas 1995, soumis 1), l'agent-fourmis (Tzafestas 1998, soumis 2) et l'agent tit-for-tat coopératif (Tzafestas 2000). Dans tous les cas, et malgré les différences apparentes entre eux, nous avons réussi à définir une mesure interne à l'agent et dépendante du problème qui représente l'état d'avancement de la résolution et un mécanisme d'autorégulation qui possède deux parties couplées. Le but de l'agent est de réguler la valeur de sa variable interne entre des limites, soit pour la ramener à une valeur de référence, telle que 0, soit pour l'empêcher de toucher les extrêmes. La régulation est positive, de manière que la variable de l'agent suive la tendance de son environnement qu'elle essaye de représenter, même si la valeur de cette variable ne correspond jamais à la réalité, mais elle assure une distance représentationnelle. À un deuxième niveau mété, un autre mécanisme de régulation "regarde" le premier niveau et régule ses taux d'adaptation de manière négative. La

puissance de l'agent réside dans l'existence de ce mécanisme de régulation de deuxième niveau qui est dépendant du problème et qui a conduit à une amélioration de la performance de l'agent par rapport au modèle de base d'un seul niveau. La conception générale de l'ensemble de ce système physiologique peut être résumée ainsi :

*Si (le monde diverge de la représentation de l'agent)  
alors (dans le futur) adapter afin de venir plus près au monde,  
sinon (dans le futur) adapter afin d'amplifier cette différence.*

L'adaptation devient ainsi réflexive dans le sens que le deuxième niveau d'adaptation regarde et agit sur le premier, autrement dit le système complet d'adaptation comme vu par l'extérieur se regarde et s'adapte de manière intrinsèque.

## 2 Étude de cas I : L'agent explorateur

### 2.1. *Le problème*

Un problème typique de robotique comportementale est celui de ***l'exploration*** : un ensemble d'agents (robots) débarque sur une planète avec la mission d'explorer sa surface pour des échantillons de minerai ayant certaines propriétés. Les robots arrivent dans un vaisseau spatial qui sert de base planétaire tout au long de la mission. La mission est accomplie quand toute la surface dans un certain rayon de la base est explorée, c'est-à-dire quand les agents ont "balayé" toute la surface en question et ont ramassé toutes les instances d'objets d'intérêt (voir par exemple Beckers et al. 1994). Les agents sont supposés de rentrer à la base lorsque leur mission est finie.

Ce problème d'exploration a été traditionnellement abordé du point de vue "fonctionnel" : "Comment un ou plusieurs agents balayent un espace délimité pour épuiser les sources d'intérêt ?". La réponse à cette question est un système de contrôle, une architecture, qui permet à l'agent de naviguer, percevoir, détecter du minerai etc., afin de balayer tout l'espace en question. Une solution comme celles rencontrées dans la bibliographie (par exemple Mataric 1992), avec un composant aléatoire et même sans apprentissage ou raisonnement spatial, assure statistiquement la couverture du champ d'intérêt et l'épuisement des sources de minerai.

Mais d'un point de vue plus "cognitif", cette fonctionnalité seule ne répond pas à la question essentielle : "***Comment les agents savent-ils qu'ils ont balayé tout l'espace, ou qu'ils ont accompli leur mission ?***". Pour répondre à cette question, il faut reformuler la description de la tâche de balayage, de manière à y inclure une expression, analytique ou autre, qui représente le critère de terminaison, c'est-à-dire l'épuisement des sources de minerai. Il suffit alors de définir une variable environnementale, la densité des sources de minerai (dénotée  $p_m$  par la suite), qui caractérise l'état du monde à un instant donné. Le but de l'agent explorateur-balayeur devient donc de ramener la valeur de cette variable à 0. Nous verrons qu'un agent ayant une représentation de cette variable constitue une solution simple à ce problème de description.

Troisièmement, nous cherchons à étudier l'opérationnalité du système, c'est-à-dire la relation entre l'architecture interne des agents et leurs performances, dans le but de trouver une architecture qui "optimise" ces performances-là. Le critère d'opérationnalité qui s'applique à la tâche de balayage est, bien évidemment, la durée de la mission : les agents sont plus performants s'ils se rendent compte de la terminaison de leur mission plus rapidement.

Dans les simulations effectuées, le monde sous exploration est défini comme un carré autour de la base centrale : la taille du monde est alors la longueur du carré (par défaut, les résultats reportés par le suite ont été pris dans un monde 25x25). Le système de contrôle de base de l'agent ainsi que les détails de simulation sont donnés dans (Tzafestas 1995, soumis 1), pour les deux cas d'un seul agents et d'agents multiples.

## 2.2. *La solution : Reformulation du problème comme un problème d'adaptation*

Nous abordons maintenant la deuxième question : “Comment les agents savent-ils qu’ils ont balayé tout l’espace pour rentrer définitivement à la base ?”. Ils ont besoin d’un moyen de détection du degré de compléTION de la tâche ou d’un critère de terminaison (balayage complété). Le seul paramètre de la tâche qui peut être utile pour le développement d’un critère de terminaison est la densité des sources dans le monde  $p_m(t)$ . Si l’agent connaissait d’avance sa valeur initiale  $p_m(0)$ , on pourrait définir comme critère de terminaison une formule du type  $\{p_m(0) * \text{sqr}(r) \text{ échantillons ont été ramassés}\}$  (où  $r$  est la longueur du côté du carré, ici 25). Cependant, ce critère n’est pas sûr, parce que, si un échantillon n’est pas détecté, l’agent ne terminera jamais (mais on pourrait laisser tomber un ou deux échantillons qu’on n’a pas trouvés).

La solution très simple à ce problème est d’estimer continûment la valeur de  $p_m(t)$  et, étant donné qu’elle tombe à 0 comme effet de bord de l’activité de l’agent, de prendre comme critère de terminaison  $p_m(t)=0$ . L’estimation de la valeur de  $p_m(t)$  nécessite alors une variable représentationnelle locale à l’agent ( $p_a(t)$ ) et peut se faire par l’intermédiaire d’une formule simple d’adaptation proportionnelle :

**Variable représentationnelle :**  $p_a(t)$

**Adaptation proportionnelle :**

*fenêtre d’observation w, taux r*

$$p_a(t) = p_a(t-w) + \text{diff} * r$$

$$\text{diff} = p_{\text{calc}} - p_a(t-w)$$

$p_{\text{calc}} = \text{nombre des échantillons ramassés} / \text{nombre des pas effectués}$   
(pendant la fenêtre de l’adaptation)

**Critère de terminaison :**

$$p_a(t) < e_p$$

où  $e_p$  un petit seuil (ici,  $e_p=0.001$ )

Le  $p_{\text{calc}}$  exprime l’estimation de l’agent lors de sa fenêtre d’observation et la loi proportionnelle assure que la mise-à-jour de l’estimation de l’agent ne se fait pas trop rapidement. Ce système de représentation et d’adaptation présente l’avantage de robustesse face aux perturbations/manipulations du type réinitialisation de la variable  $p_m(t)$  au cours du balayage. Dans la figure 1 est illustrée la co-évolution des deux variables  $p_m(t)$  et  $p_a(t)$  dans le temps. Comme on peut le voir sur la figure, **la variable représentationnelle permet à l’agent de résoudre son problème de terminaison dans tous les cas sans jamais prendre la valeur réelle qu’elle représente** (sauf un point de croisement). Les deux variables tombent progressivement à 0 sans jamais prendre la même valeur — on pourrait dire que celle de  $p_a(t)$  “suit” celle de  $p_m(t)$ . En effet, la montée rapide de  $p_a(t)$  au début du balayage est la conséquence de l’utilisation d’un capteur de détection d’échantillons à distance qui fait orienter l’agent vers les

sources de minerai en minimisant son comportement erratique de façon que la plupart des places visitées soient des places contenant des échantillons. La valeur de  $p_a(t)$  baisse ensuite puisque celle de  $p_m(t)$  baisse comme effet de bord de l'activité de l'agent qui trouve de moins en moins des échantillons.

### 2.3. Adaptation à deux niveaux

Nous avons ensuite voulu étudier la relation — s'il en existe une — entre les paramètres  $w$  et  $r$  du système d'adaptation et la valeur  $p_m(0)$ . Le système a été alors simulé pour des différentes valeurs de  $w$  et  $r$  et dans des différentes densités initiales de monde. La figure 1 donne les résultats de ces simulations pour trois ensembles des paramètres d'adaptation (adaptation rapide, moyenne ou lente).

L'adaptation rapide est plus opérationnelle que l'adaptation moyenne qui est à son tour plus opérationnelle que l'adaptation lente (toujours selon le critère de durée). Cependant, plus l'adaptation est rapide, plus elle démontre des fluctuations, et plus l'adaptation est lente, plus elle démontre des retards. Qui plus est, le même paramétrage donne des résultats différents dans les différentes densités de monde : la différence des résultats se voit dans la forme des courbes (pour plus de résultats voir Tzafestas 1995, soumis 1). Plus particulièrement, la réponse de l'agent aux différentes perturbations (la forme de la courbe d'évolution de  $p_a(t)$ ) diffère selon la condition de limite  $p_m(0)$  : pour le même paramétrage d'adaptation, l'agent finit plus ou moins vite sa mission selon la valeur de  $p_m(0)$ , c'est-à-dire l'intervalle entre le moment du ramassage du dernier échantillon et le retour définitif de l'agent à la base est d'une durée très variable. Il semble donc que, pour assurer l'opérationnalité de l'agent dans les différents mondes, il faut trouver un moyen de combiner les avantages de l'adaptation rapide en termes d'opérationnalité avec les avantages de l'adaptation lente en termes de régularité de courbe.

Plus précisément, il faut une adaptation rapide à la fin (pour terminer rapidement), mais lente lors du ramassage (pour éviter les fluctuations). Il s'agit alors de trouver un moyen de se stabiliser au bon paramétrage *en ligne*. Autrement dit, ***il faut un système de méta-adaptation***.

**Méta-adaptation :**

Si  $|diff| (= |p_{calc}-p_a(t-w)|) \leq f_p$ ,  
alors adaptation plus rapide

$r \Rightarrow r_{max}, w \Rightarrow w_{min}$

sinon adaptation plus lente

$r \Rightarrow r_{min}, w \Rightarrow w_{max}$

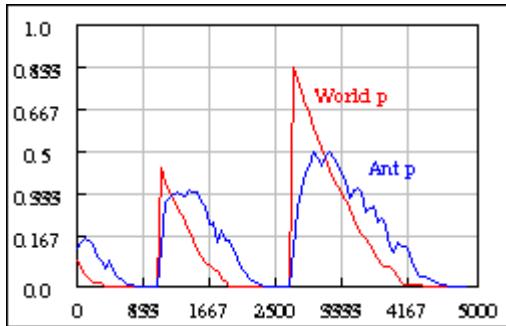
$r = r + r_r * (r_{max}-r), w = w + r_w * (w_{min}-w)$

La méta-adaptation doit affecter les paramètres  $w$  et  $r$  de façon que l'adaptation devienne plus rapide quand le  $p_{calc}$  est suffisamment près de  $p_a(t)$  et plus lente quand il est plus loin. Cette loi de méta-adaptation signifie que le monde paraît plus fiable quand il ne diffère pas trop de l'idée que l'agent en a, sinon il est pris moins au sérieux.

La figure 2 donne les résultats de l'application du système de méta-adaptation pour les trois densités du monde ; comme on peut facilement voir sur la figure, la réponse de l'agent (la forme de la courbe) est la même pour les trois densités exemplaires, autrement dit le résidu de

la durée de la mission après le ramassage du dernier échantillon est approximativement le même dans les trois cas.

Nous avons montré dans (Tzafestas 1995, soumis 1) que l'opérationnalité de l'agent qui possède ce système de méta-régulation ne dépend pas qualitativement des valeurs des paramètres  $w_{min}$ ,  $w_{max}$ ,  $r_{min}$ ,  $r_{max}$ ,  $r_w$  et  $r_r$ . De plus, la condition initiale  $p_a(0)$  n'a aucune importance non plus.



**Figure 1.** Performance de l'agent adaptatif de base pour trois densités initiales de monde, densité faible ( $p_m(0)=0.1$ ), moyenne ( $p_m(0)=0.1$ ) et élevée ( $p_m(0)=0.9$ ) ( $p_a(0)=0.15$ ). Moments de terminaison 1ère partie = 1019, 2ème partie = 2659, 3ème partie = 4897). ( $f_p=0.1$ ,  $w=30$ ,  $r=0.2$ , adaptation moyenne)



**Figure 2.** Performance de l'agent adaptatif avec du méta pour trois densités initiales de monde, densité faible ( $p_m(0)=0.1$ ), moyenne ( $p_m(0)=0.1$ ) et élevée ( $p_m(0)=0.9$ ) ( $p_a(0)=0.15$ ). Moments de terminaison 1ère partie = 489, 2ème partie = 1832, 3ème partie = 3497). ( $f_p=0.1$ ,  $w_{min}=15$ ,  $w_{max}=40$ ,  $r_{min}=0.15$ ,  $r_{max}=0.3$ ,  $r_r=r_w=0.2$ )

### 3 Étude de cas II : La fourmis et ses phéromones

#### 3.1. Le problème

La situation naturelle rencontrée dans les sociétés d'insectes et souvent modélisée est une variante du problème précédent où il existe peu de sources de taille importante distribuées dans l'environnement. La solution dans ce cas consiste à permettre au robot-fourmis de déposer de phéromones lorsqu'il se trouve chargé et de retour à sa base, c'est-à-dire à son nid. Un autre agent ou lui-même peut suivre ces traces pour arriver à la source d'intérêt rapidement. Une possibilité supplémentaire est de considérer que la trace de phéromones déposées par terre évapore lentement (Deneubourg et al. 1990, Solé et al. 2000).

La première motivation de notre approche est que le modèle ainsi défini ne doit pas être stable pour cause de dépendre d'une quantité infinie de phéromone si la nourriture est régulièrement renouvelée. Ce problème est étudié en détail dans (Tzafestas 1998, soumis 2).

Le modèle comportemental de base (Steels 1990, Drogoul et Ferber 1992) est le suivant :

```

Si (de retour au nid, c'est-à-dire chargé)
  Si (sur le nid) décharger
  Sinon {aller vers le nid, déposer 2 unités de phéromone}
Sinon  Si (sur une source) charger un morceau
  Sinon    Si (des phéromones ou un stimulus perçu)
            {suivre le stimulus, ramasser 1 unité de phéromone}
  Sinon aller au hasard

```

Le modèle de base dépose 2 unités de phéromone quand il retourne à sa base et ramasse 1 unité quand il suit la phéromone déjà sur le terrain ou un autre stimulus. Dans les simulations

effectuées, le monde sous exploration est défini comme un carré autour de la base centrale : la taille du monde est alors la longueur du carré (par défaut, les résultats reportés par le suite ont été pris dans un monde 20x20). La source de nourriture se trouve dans le coin de droite en bas, le nid se trouve dans le coin de gauche en haut et la population des robots-fourmis consiste de 5 robots. Les robots peuvent percevoir la nourriture ou la pheromone depuis une distance de 3 unités d'espace. Nous avons simulé ce système avec le comportement de base d'agent-fourmis et nous avons mesuré les quantités de pheromone dans le monde et dans chaque agent. Comme prévu, les quantités de pheromone des agents tombent sous 0, tandis que la quantité de pheromone dans le monde peut croître sans limite. Les valeurs de ces quantités dépendent des autres paramètres du problème (distance source-nid, nombre de robots, taille de source etc.) qui définissent le nombre nécessaire des voyages aller-retour source-nid qui sont nécessaires pour épuiser la source. Les mêmes résultats sont obtenus quand l'agent-fourmis ne ramasse pas de pheromone, mais elle évapore naturellement, ou si la pheromone possédée par chaque agent regénère naturellement aussi. Un deuxième problème émerge si l'on essaye de résoudre le premier (valeurs négatives) de façon non naturelle, en dotant depuis le départ les agents d'une quantité suffisante de pheromone, c'est-à-dire très grande. Ceci conduit à un système où les agents continuent à être attirés par la grande quantité de pheromone à une source de nourriture déjà épuisée depuis longtemps.

### 3.2. *La solution : Reformulation du problème comme un problème d'adaptation*

Les observations précédentes nous amènent à une nouvelle formulation du problème de marquage par pheromones : ***Nous cherchons un mécanisme de régulation de pheromone qui permet au chemins de pheromone d'être construits et renforcés rapidement autant que la source est présente et disparaître aussitôt après son épuisement.***

Un mécanisme simple qui permet la régulation de pheromone entre des bornes utilise deux valeurs  $pheromones_{min}$  et  $pheromones_{max}$  et la “loi d’effet” :

$$\begin{array}{ll} \text{Pour déposition} & pheromones(t+1) = pheromones(t) + r_1 * (pheromones_{min} - pheromones(t)) \\ \text{Pour ramassage} & pheromones(t+1) = pheromones(t) + r_2 * (pheromones_{max} - pheromones(t)) \end{array}$$

Ce simple mécanisme de régulation assure qu’aucun agent ne sortira des limites physiologiques. Cependant, la vraie valeur de la quantité de pheromones déposée ou ramassée à chaque cycle de la simulation dépend de l’état actuel de l’agent : un agent avec une grande quantité de pheromone déposera plus et ramassera moins qu’un agent avec peu de pheromones. Cet arrangement permet aux chemins de pheromone à être vite construits (puisque les agents déposent plus de pheromones au départ) et disparaître rapidement (puisque les agents vers la fin de la tâche possèdent statistiquement peu de pheromones, alors ils ont tendance à en ramasser plus). Dans les simulations effectuées nous avons fixé  $pheromones_{min}=10$  et  $pheromones_{max} = 100$ , pour tous les agents.

### 3.3. *Adaptation à deux niveaux*

Un taux élevé de déposition sera de bénéfice au début et au milieu de la tâche, quand tous les agents veulent créer et renforcer un chemin rapidement, tandis qu’un taux élevé de ramassage sera de bénéfice à la fin de la tâche, quand les agents veulent détruire aussitôt que possible le

chemin à la source épuisée. Puisque ces paramètres sont contradictoires, nous aimerions introduire un mécanisme d'adaptation qui "reconnaîtrait" si l'agent se trouve vers le début ou vers la fin de la tâche et mettrait à jour les taux d'adaptation de premier niveau comme il convient. Pour ce faire, il faut un critère de "reconnaissance" d'état d'avancement de la tâche. Le seul critère sous la main puisse être la quantité de pheromone dans l'environnement. Puisque cette quantité globale n'est pas mesurable par les agents, nous utilisons une mesure d'estimation, à voir la quantité de pheromone sur la position de l'agent. Cette mesure est utilisée comme suit :

**Pour déposition**

Si  $\text{pheromones}(t) \geq \text{estimation\_environnement}$   
 $r_1(t+1) = r_1(t) + r_{r1} * (r_{1\max} - r_1(t))$

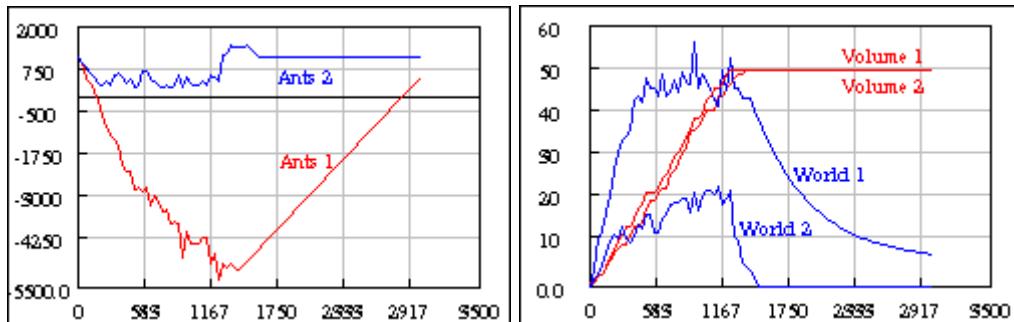
sinon  $r_1(t+1) = r_1(t) + r_{r1} * (r_{1\min} - r_1(t))$

**Pour ramassage**

Si  $\text{pheromones}(t) \geq \text{estimation\_environnement}$

$r_2(t+1) = r_2(t) + r_{r2} * (r_{2\min} - r_2(t))$

sinon  $r_2(t+1) = r_2(t) + r_{r2} * (r_{2\max} - r_2(t))$



**Figure 3.** Comparaison de modèle de base (1) et modèle adaptatif (2) (gauche) Quantité de pheromones des agents, (droite) Quantité de pheromones dans l'environnement et volume cumulatif de transport au nid.

Ainsi, le taux de déposition de pheromones augmente quand le robot possède plus de pheromone qu'il perçoit dans son environnement et baisse autrement. Inversement, le taux de ramassage de pheromones augmente quand le robot possède moins de pheromone qu'il perçoit dans son environnement et augmente autrement. La figure 3 donne un résultat typique d'application du modèle précédent. A notre surprise, la régulation des taux de déposition et de ramassage modifie le comportement qualitatif des agents, puisque la quantité de pheromone dans le monde prend rapidement une valeur très élevée, reste près de cette valeur durant la tâche, et retombe à 0 très rapidement quand la source est épuisée, en démontrant pendant toute cette période beaucoup moins de fluctuations que le modèle comportemental de base). Mais les résultats sont également affectés de façon quantitative, puisque la durée de la tâche (source épuisée et pheromones disparues) est bien inférieure à celle dans le modèle de base et celle dans le modèle avec adaptation à un seul niveau.

## 4 Étude de cas III : L'agent coopératif

### 4.1. Le problème

Une question majeure des recherches en vie artificielle et en biologie théorique est le comportement coopératif entre agents égoïstes. Le problème de coopération part de l'hypothèse que chaque agent a un intérêt personnel immédiat à défecter, tandis que le comportement

optimal joint serait de coopérer. Ce problème est traditionnellement modélisé comme un jeu de deux joueurs, le Dilemme Itéré des Prisonniers (Iterated Prisoner's Dilemma, IPD).

Ainsi, à chaque cycle d'interaction, les agents jouent au Dilemme du Prisonnier. Chaque agent peut coopérer (C) ou défacter (D). Les résultats obtenus pour chaque paire d'actions (agent,adversaire) sont donnés dans le tableau qui suit :

Agent	Adversaire	Résultat (score)
C	C	3 (= Récompense)
C	D	0 (= Stupide)
D	C	5 (= Tentation)
D	D	1 (= Punitio)

Les expériences habituelles avec des stratégies pour l'IPD sont soit des tournois (championnats) soit des expériences écologiques. Dans les championnats, chaque stratégie joue contre toutes les autres et les résultats sont regroupés par stratégie à la fin. Dans les expériences écologiques, une population de stratégies joue en tournois et chaque génération successive regroupe les meilleures finalistes de la génération précédente à des proportions relatives à leurs scores.

La première stratégie importante pour l'IPD a été conçue et étudiée par Axelrod (1984); il s'agit de la stratégie "Tit For Tat" (abrégé TFT).

*Commencer par coopérer,*

*Ensuite à chaque cycle d'interaction retourner la dernière action de l'adversaire.*

Cette stratégie a obtenu les meilleurs résultats dans les premiers championnats et a été trouvée relativement stable dans des contextes écologiques.

Une des meilleures stratégies rencontrées dans la littérature est GRADUAL (Beaufils et al. 1996) qui obtient les meilleurs résultats face à quasiment toutes sortes de stratégies conçues. Cette stratégie commence en coopérant et ensuite joue TFT, sauf que sa défaction n'est pas un seul D mais l'aveugle suite ( $nxD$ )CC, où  $n$  est le nombre des défactions de l'adversaire dans le passé (mesure cumulative). Ainsi, GRADUAL répond avec DCC à la première défaction de l'adversaire, DDCC à la deuxième, etc. La justification de ces performances est qu'elle punit de plus en plus l'adversaire, comme nécessaire, et ensuite l'apaise avec deux coopérations de suite.

Comme dans le cas de l'agent-fourmis, la première motivation pour notre modèle est la conviction que nous pouvons imaginer une stratégie comparable à GRADUAL quant à ses performances mais avec une mémoire qui ne soit pas irreversible et permanente. Au contraire, nous cherchons un modèle à base de TFT, mais plus adaptatif qui démontrerait gradualité comportementale et qui posséderait le potentiel de stabilité face à des mondes changeants (par exemple, lors de remplacements d'adversaire etc.).

Avant de procéder à la conception de notre modèle, nous avons étudié les résultats obtenus par GRADUAL. Les stratégies d'IPD rencontrées dans la littérature appartiennent à une des trois catégories suivantes :

- Des stratégies relativement complexes qui utilisent des informations de jeu; en général, il s'agit des stratégies coopératives jusqu'à ce que l'adversaire défekte, auquel cas elles adoptent une attitude de représailles (TFT, GRIM, GRADUAL, etc.).
- Des stratégies essentiellement coopératives mais qui commencent comme suspicieuses, par exemple en jouant quelque coups de D au départ, afin de tester leur adversaire. Cette famille de stratégies comprend entre autres la stratégie "suspicious tft" (STFT) et la stratégie "prober" de (Beaufils et al. 1996).
- Des stratégies qui sont clairement irrationnelles, parce qu'elles n'utilisent aucune information de jeu, mais elles suivent un rituel plus ou moins déterminé d'actions. Cette famille de stratégies comprend entre autres la stratégie aléatoire et toute stratégie aveugle périodique, telle que CCD, DDC etc.

Une stratégie maximisera ses résultats, si elle est capable de converger à un régime de coopération mutuelle face aux stratégies des deux premières catégories et à un régime de défection constante face aux stratégies de la troisième catégorie. Une déflection constante face aux stratégies périodiques est nécessaire pour atteindre un score maximal (voir Tzafestas (2000) pour une preuve). La stratégie GRADUAL remplit toutes les deux conditions, car elle répond avec deux C consécutifs après une longue série de déflections – ce qui donne la possibilité à STFT et les stratégies "prober" de reprendre coopération- et converge à ALLD face aux stratégies irrationnelles. Une solution du problème de la mémoire permanente doit démontrer les mêmes propriétés.

#### 4.2. *La solution : Reformulation du problème comme un problème d'adaptation*

La stratégie adaptative que nous cherchons doit être essentiellement tit-for-tat. En plus, elle doit montrer moins d'oscillations comportementales entre C et D. Pour ce faire, la stratégie doit avoir une estimation de la stratégie de l'adversaire, coopérante ou irrationnelle, et y réagir à la tit-for-tat. L'estimation doit être mise-à-jour continûment durant l'interaction avec l'adversaire. Cela peut être modélisé à l'aide d'une simple variable continue, "l'image du monde", qui prend une valeur de 0 (défection parfaite) à 1 (coopération parfaite). Les valeurs intermédiaires représentent des degrés de coopération ou de déflection. Le modèle tit-for-tat adaptatif peut alors être formulé comme un simple modèle linéaire :

##### ***Tit-for-tat adaptatif***

*Si (l'adversaire joua un C pendant le dernier cycle) alors*

*monde = monde + r\*(1-monde), r est le taux d'adaptation*

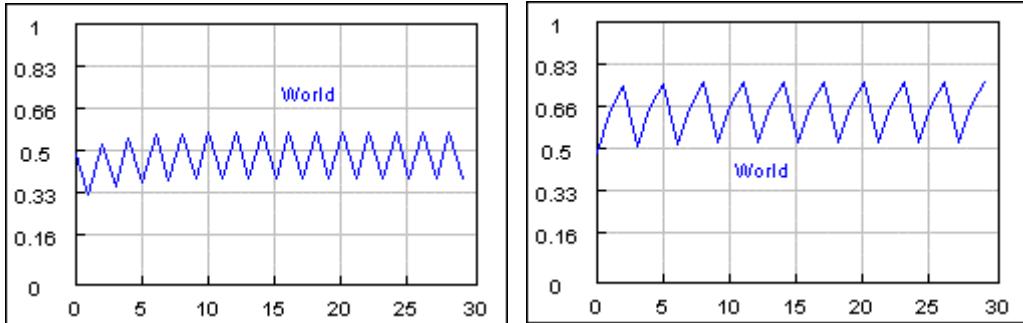
*sinon        monde = monde + r\*(0-monde)*

*Si (monde ≥ 0.5) jouer C, sinon jouer D*

Le modèle tit-for-tat habituel correspond au cas  $r=1$  (adaptation immédiate à la dernière action de l'adversaire). Bien évidemment, l'utilisation de valeurs  $r<1$  permettra l'émergence d'un comportement plus graduel et plus robuste aux perturbations. Les performances du modèle tit-for-tat adaptatif face aux trois types de stratégies présentés ci-dessus sont les suivantes :

- Face aux stratégies complexes qui utilisent des informations de jeu, le modèle coopère constamment et converge vite à une coopération parfaite.

- Face aux stratégies “suspicieuses”, le modèle joue exactement comme TFT, et la valeur de son “image de monde” oscille autour de la valeur critique de 0.5. (voir fig. 4 face à STFT).
- Face aux stratégies irrationnelles et périodiques, la valeur de son “image de monde” converge rapidement à des oscillations autour de la valeur critique de “nombre\_moyen\_de\_C/ nombre\_moyen\_de\_D” de l’adversaire.



**Figure 4.** Comportement de tit-for-tat adaptatif (à gauche) face à STFT ( $r=0.2$ , monde(0)=0.5) (à droite) face à CCD ( $(r=0.2$ , monde(0)=0.5).

#### 4.3. Adaptation à deux niveaux

La version précédente du modèle souffre de manipulabilité de sa variable “monde” par l’adversaire. Ceci se manifeste comme stabilisation de l’agent à un comportement oscillatoire (comme face à STFT) ou à un comportement parfaitement cooperatif face à des agents irrationnels (comme face à CCD). Pour palier à ce défaut, nous avons exploité notre observation que des taux différents de coopération et de défection (respectivement,  $r_c$  et  $r_d$ ) conduiraient à des résultats différents. Plus spécifiquement, nous avons observé que l’agent tit-for-tat adaptatif parvient à amener les agents tels que STFT ou le “prober” à coopérer avec lui si  $r_c > r_d$ , tandis qu’il parvient à exploiter au maximum les agents irrationnels en devenant défectif si  $r_c < r_d$ . Ainsi, nous avons besoin d’un critère qui permettra à l’agent tit-for-tat adaptatif de découvrir si l’adversaire utilise une stratégie de représailles ou s’il est tout simplement irrationnel, afin d’adopter le papamétrage convenable. La loi d’adaptation qui a été développée est la suivante :

*Pendant une fenêtre d’observation (window=w), enregistrer combien des fois (n) l’action de l’agent a coïncidé avec celle de l’adversaire. Dans des intervalles réguliers (tous les w cycles) adapter les taux comme suit :*

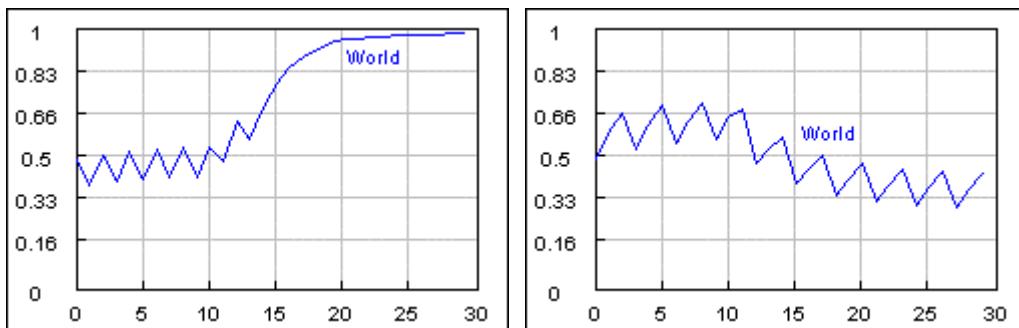
*Si ( $n > \text{seuil}$ ) alors  $r_c = r_{\min}$ ,  $r_d = r_{\max}$ , sinon  $r_c = r_{\max}$ ,  $r_d = r_{\min}$*

**Cette loi peut être interprétée comme :**

*Si (le mode est assez coopératif)\* alors  $r_c = r_{\min}$ ,  $r_d = r_{\max}$ , sinon  $r_c = r_{\max}$ ,  $r_d = r_{\min}$*

*(\*) à rappeler que “mon action = l’action de l’adversaire” est le critère dit “pavlovien” de coopération (Nowak et Sigmund 1993)*

A noter enfin que l’agent converge à un taux de coopération faible quand son monde est trouvé cooperatif, sinon à un taux de coopération important, c’est-à-dire, il utilise des rétroactions négatives au niveau de régulation de taux. Ce modèle avec de l’adaptation à deux niveaux, parvient à obtenir la meilleure performance contre les agents de tous les types, tout en possédant une mémoire en principe reversible.



**Figure 5.** Comportement de tit-for-tat adaptatif méta-régulé (à gauche) face à STFT (à droite) face à CCD. Dans les deux cas  $r_c(0)=0.2$ ,  $r_d(0)=0.2$ ,  $r_{\max}=0.3$ ,  $r_{\min}=0.1$ ,  $\text{monde}(0)=0.5$ ,  $w=10$ , seuil=2.

## 5 Adaptation, comportement, évolution

Dans tous les trois cas présentés auparavant, nous avons pu montrer que le comportement de l'agent dépend d'une variable critique qui dirige sa motivation pour agir. Cette variable est en réalité couplée avec l'environnement par l'intermédiaire même du comportement de l'agent. En essayant de réguler sa propre variable, l'agent essaye en effet de manipuler la variable correspondante globale. De plus, cette variable a une *valeur cognitive*, puisqu'elle représente l'idée que l'agent a pour l'état de son environnement. Sous cet angle de vue, l'agent essaye d'approximer son environnement, c'est-à-dire de s'adapter à lui. Les variables régulées sont dans ce sens critiques pour la "survie" ou l'opérationnalité de l'agent et correspondent ainsi aux *variables essentielles* de Ashby (1960).

L'opérationnalité du comportement est assurée par le mécanisme d'auto-régulation de deuxième niveau, qui agit sur les taux d'adaptation du premier niveau. Cette approche est donc compatible avec l'approche dynamique de la cognition (van Gelder et Port 1995) qui déclare la dynamique du comportement comme un facteur primordial des capacités cognitives.

En somme, nous avons montré que l'opérationnalité de l'agent dans un ensemble de problèmes est assurée grâce à un mécanisme de double régulation qui définit un système d'adaptation homéostatique de base, et un système de méta-adaptation qui agit sur les paramètres d'adaptation du niveau précédent. Notre perspective à long terme est de formaliser une théorie de régulation pour les agents autonomes. La notion d'adaptation qui agit sur elle-même, c'est-à-dire d'adaptation réflexive, semble jusqu'à maintenant un moyen puissant de modélisation qui est en ligne avec les recherches en régulation classique. Pour avancer sur cette voie, il faudra ultérieurement répondre à un nombre de questions :

- Comment peut-on identifier la variable cognitive critique pour chaque problème ? Et comment formalise-t-on la régulation (de premier niveau) pour chaque problème ? En effet, quels sont les problèmes qui peuvent être étudiés de cette façon ?
- Combien de taux de premier niveau faut-il ? Ou, combien de processus indépendants de régulation existe-t-il ? Par exemple, l'agent explorateur possède un taux d'adaptation, tandis que les deux autres agents en possèdent deux.
- Quel est le critère de méta-régulation ? Dans tous les trois cas étudiés, le critère est purement qualitatif et dépendant du problème. Peut-on trouver une loi générale conceptuelle qui regroupe tous les cas ?

- Quelle est la nature des dynamiques de méta-régulation ? Dans tous les trois cas, nous avons montré (Tzafestas 2000, soumis 1, 2) qu'une dynamique "bang-bang" (haut-bas) est suffisante, parce que ce qui compte est la relation entre taux et dynamiques et non pas les valeurs absolues des paramètres.
- Finalement, quel est le rôle et la valeur du "comportement dans le vide" (sans perturbation) ? Ce comportement est purement spécifique à l'agent en question et peut être différent d'agent à agent, à cause des paramétrages différents qui définissent ainsi la "personnalité" de chaque agent. Des expériences dans le vide (Tzafestas soumis 1, 2) montrent que l'étude dans le vide permet certaines prédictions de performance de l'agent.

Mais le mécanisme d'adaptation réflexive n'est pas simplement le contrôleur du comportement; il peut servir également de structure évolutive. Selon cette perspective, le "code génétique" d'un agent que nous faisons évoluer n'est pas un ensemble de paramètres numériques auquels l'évolution donnera des valeurs qui "optimisent" les performances de l'agent, mais l'ensemble de critères du mécanisme d'adaptation réflexive auquels l'évolution donnera une signification en les "branchant" directement aux capteurs de l'agent. Dans notre cas, le code génétique consiste à l'ensemble structurel suivant [perception de la variable du niveau 1, critère de divergence du niveau 2]. L'instantiation de cette structure pour les trois agents étudiés est comme suit :

	<i>Perception de variable du niveau 1</i>	<i>Critère de divergence du niveau 2</i>
<i>Agent explorateur</i>	$diff = p_{calc} - var$ $p_{calc} = \text{nombre des échantillons ramassés} / \text{nombre des pas effectués}$	$ diff  (=  p_{calc}-var ) \leq f_p$
<i>Agent-fourmis</i>	<i>Si déposition, diff = var<sub>min</sub> - var</i> <i>Si ramassage, diff = var<sub>max</sub> - var</i>	$(var - estimation\_environnement) \geq 0$
<i>Agent TFT adaptatif</i>	<i>Si (l'adversaire joua un C pendant le dernier cycle) diff = 1-monde</i> <i>Sinon diff = 0-monde = -monde</i>	$n \leq \text{seuil, où } n = \text{combien des fois l'action de l'agent a coïncidé avec celle de l'adversaire}$

En effet, la formulation de ce tableau est en liaison directe avec les questions posées auparavant, car les deux premières questions concernent la perception de la variable du niveau 1, tandis que la troisième question concerne le critère de divergence du niveau 2. L'analogie conceptuelle des trois mécanismes nous oblige à en dissectioner les différences (nombre de taux, directions de régulation, dynamique bang-bang ou graduelle). Dans le cas idéal nous aimeraions avoir un seul taux d'adaptation et une seule loi d'effet pour chaque niveau, comparaison à 0 et dynamique bang-bang dans le deuxième niveau. Il faut souligner ici que le mécanisme restera invariant à travers les générations, mais les structures utilisées changeront. Ceci permettra la "reprogrammation" évolutive des agents pour répondre aux besoins dynamiques d'adaptation à un environnement évolutif et notamment à un environnement dans lequel les agents co-évoluent continûment.

Le deuxième niveau d'adaptation formulé ainsi permettrait le passage rapide à une modélisation génétique, car on sait que les vrais gènes prennent des valeurs booléennes et régulent des vitesses de réactions chimiques à l'intérieur des organismes biologiques. Nous imaginons que cette perspective d'adaptation à deux niveaux permettrait la modélisation simplifiée d'un nombre de comportements biologiques complexes de tout niveau. Nous pensons plus particulièrement à des comportements tels que l'auto-organisation des réseaux de gènes (Raeymaekers 2002), le développement et la stabilisation des communications cellulaires

(Furusawa et Kaneko 1998), la répartition de tâches dans une société animale (Ratnieks et Anderson 1999), certains effets de spéciation (Dieckmann et Doebeli 1999) etc.

## Bibliographie

- Ashby, W.R. (1960). *Design for a brain - The origin of adaptive behaviour*. 2<sup>nd</sup> edition, Chapman & Hall.
- Axelrod, R. (1984). *The evolution of cooperation*. Basic Books.
- Beaufils, B., Delahaye, J.-P. et Mathieu, P. (1996). "Our meeting with gradual: A good strategy for the iterated prisoner's dilemma", *Proceedings Artificial Life V*, Nara, Japan.
- Beckers, R., Holland O.E. et Deneubourg, J.-L. (1994). "From local actions to global tasks: Stigmergy and collective robotics", *Proceedings Artificial Life IV* (R. Brooks and P. Maes, Eds.), MIT Press, 181-189.
- Deneubourg, J.-L., Aron, S., Goss, S. et Pasteels, J.M. (1990). "The self-organizing exploratory pattern of the Argentine Ant", *Journal of Insect Behavior* 3:159-168.
- Dieckmann, U., Doebeli, M. (1999). On the origin of species by sympatric speciation, *Nature*, 400(6742):311-2, 22 July.
- Drogoul, A. et Ferber, J. (1992). "From Tom Thumb to the Dockers : Some experiments with foraging robots", *Proceedings SAB 1992*, 451-459.
- Furusawa, C. et Kaneko, K. (1998). Emergence of rules in cell society : Differentiation, Hierarchy and Stability, *Bulletin of Mathematical Biology*, 60:659-687.
- Korzeniewski, B. (2001). Cybernetic formulation of the definition of life, *Journal of Theoretical Biology*, 209:275-286.
- Mataric, M. (1992). "Designing emergent behaviors : From local interactions to collective intelligence", *Proceedings SAB 1992*, 432-441.
- Nowak, M. et Sigmund, K. (1993). "A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game", *Nature* 364(1993):56-58.
- Raeymaekers, L. (2002). Dynamics of boolean networks controlled by biologically meaningful functions, *Journal of Theoretical Biology*, 218:331-341.
- Ratnieks, F.L.W., Anderson, C. (1999). Task partitioning in insect societies, *Insectes Sociaux*, 46:95-108.
- Solé, R.V., Bonabeau, E., Delgado, J., Fernández, P. et Marín, J. (2000). Pattern formation and optimization in army ant raids, *Artificial Life*, 6(3): 219-26.
- Steels, L. (1990). "Towards a theory of emergent functionality", *Proceedings SAB 1990*, 451-461.
- Tzafestas, E. (1995). *Vers une systémique des agents autonomes : Des cellules, des motivations et des perturbations*, Thèse de Doctorat de l'Université Pierre et Marie Curie, Paris, décembre 1995.
- Tzafestas, E. (1998). "Tom Thumb Robots Revisited : Self-Regulation as the Basis of Behavior", *Proceedings Artificial Life VI*, San Diego, CA.
- Tzafestas, E. (2000). "Toward Adaptive Cooperative Behavior", *Proceedings SAB 2000*, Paris, France.
- Tzafestas, E. (soumis 1). "Regulation Problems in Explorer Agents", submitted.
- Tzafestas, E. (soumis 2). "Pheromone regulation in insect foraging", submitted.
- van Gelder, T. et Port, R. (1995). "It's about time : An overview of the dynamical approach to cognition", in *Mind as Motion : Explorations in the dynamics of cognition*, by T. van Gelder and R. Port, MIT Press.



# Self reference of serendipitists

Pek van Andel,  
Medical Faculty,  
State University Groningen,  
The Netherlands  
m.v.van.andel@med.rug.nl

Il n'y a rien d'accidentel, et ce qui pour nous est accident n'est qu'un fait inconnu qui peut devenir, si on l'explique, l'occasion d'une découverte plus ou moins importante. (Cl. Bernard, 1865) [1]

This article is an exploratory attempt to show how the serendipitous aspect of scientific research is dealt with by authors in their publications and in their later comments, and also by others.

When something 'new' is found it often takes years before we know how new and true (in science), how new and useful (in technique) or how new and fascinating (in art) it really is. If we then look back on how the finding was done, the possibility for forming a legend is there. The serendipitous aspect of it can then easily become underestimated and even denied or overestimated and also invented by the finder and/or others.

Serendipity is by definition a finding that has not been anticipated upon by the finder himself. This non-anticipation is also the reason, why the finder is surprised by the unsought finding. Serendipity can be best described as a surprising observation followed by a correct abduction (= explication; abduction comes from the Latin *abducere* = lead away (*ab* = away + *ducere* = to lead). Or, to say it simply: the searcher observes a surprising fact or relation, then tries to think up an hypothesis to explain this surprising observation or relation (hypothesis from the Greek *hypothésis* = supposition). Of course this hypothesis has to be testable and tested independently from the surprising observation or relation. When the hypothesis is falsified, in a experiment for example, the observer will probably try to infer another hypothesis, by which the surprising phenomenon can be explained. Then he tests this new hypothesis, and so on, till he has found an explanation (or not) for the surprising fact or relation, that withstands one or more scientific tests. If he then publishes this positive finding, in most of the cases, he starts by first mentioning his last hypothesis, and does not write one word on how that hypothesis was born. He only mentions how this hypothesis was tested, and the results of it

and his conclusions. The reader may then get the idea, that the growth of knowledge goes directly from a good hypothesis to a correct thesis, from fiction to non-fiction, from fantasy to fact, from theory to practice, instead of (also) the other way around. That is a pity, because the reader is then also not informed about how the hypothesis was born and developed before being tested. So the article he read, does not give the whole truth, but only part of it, without referring to that at all. It has been written as a 'retrospective prophecy or falsification', in such a way that the reader can do the experiment again, if he wants, to see whether he gets the same results and conclusions as the article. The intended advantage of the conventional form of the scientific article is evident because anybody, who wants, can check the experiments, its reproducibility, methods, results and conclusions.

The unintended disadvantage of this main tradition of publishing is evident: an important part of the scientific research, namely the heuristic part, is not published and therefore lost, forever if not correctly mentioned elsewhere by the author or somebody else. Then nobody can learn anything from the used heuristic talents and methods of the author in question (*method* comes from the Greek *met[a] hodos* = the way along which).

But happily enough there are exceptions. There is a long tradition of a minority of researchers who do write about the heuristical aspects of their work, specially in chemistry. Chemistry is a big field, chemists think in processes, and so sometimes they also write about the process of their discovery, and the specific serendipitous elements and moments in the process of their search, research, development and application (if present).

Most examples of serendipity in science are reported by chemists. But also in other disciplines there is a tradition of writing about the serendipitous aspects of research and its findings. I collected examples of serendipity, as a flow experience and pet sin, just for fun: I wanted to learn from them. I was even warned that collecting accidental findings could not produce anything systematic, because serendipity is by definition something that happens 'at random', so how could I expect any system in that madness?

But an 'accidental' observation (leading to a new finding) means two different things: (1) an 'random' observation, and (2) an observation that falls to you without expecting it. When we deal with serendipity, we have to do with the second sort of observation. It has nothing to do with the physical or mathematical notion of 'randomness'. 'Accidental' in a case of serendipity is the psychological, subjective notion of the finder, who observes, discovers, invents, creates or imagines something he did not look for and therefore he calls it 'accidental' in the meaning of 'unexpected' and 'unsought'.

My collection of serendipities in science taught me that the serendipitous aspect of a finding can be (1) well documented, (2) overestimated, (3) fabricated or (4) underestimated, (5) hidden (or even denied) by the author and/or others. These five patterns have never been

made explicit and enumerated as such:

**1. Well documented serendipity.** The discovery of X-rays is a classic example of well documented, authentic serendipity. But Röntgen himself did not write one word about the unexpected, accidental, unsought (serendipitous) heuristic character of his finding in his famous first publication on X-rays *On a new kind of rays*:

A discharge from a large induction coil is passed through a Hittorf's vacuum tube, or through a well-exhausted Crookes' or Lenard's tube. The tube is surrounded by a fairly close-fitting shield of black paper; it is then possible to see, in a completely darkened room, that paper covered on one side with barium platinocyanide lights up with brilliant fluorescence when brought into the neighbourhood of the tube, whether the painted side or the other be turned towards the tube. The fluorescence is still visible at two metres distance. It is easy to show that the origin of the fluorescence lies within the vacuum tube. [...] A piece of sheet aluminium, 15 mm. thick, still allowed the X-rays (as I will call the rays, for the sake of brevity) to pass, but greatly reduced the fluorescence. [2]

But Röntgen's student Charles Nootnangle of Minneapolis, who had it from Röntgen himself, a few days later, reported:

By chance he happened to note that a little piece of paper lying on his work table was sparkling as though a single ray of bright sunshine had fallen upon it lying in the darkness. At first he thought it was merely the reflection from the elektric spark, but the reflection was too bright to allow that explanation. Finally he picked up the piece of paper and, examining it, found that the reflected light was given by a letter A which had been written on the paper with a platinocyanide [fluorescent] solution. [3]

Röntgen said finally in an interview some months after his discovery: 'I found by accident that the rays penetrated black paper.' [4] In short: Although Röntgen himself did not mention the serendipitous aspect of the discovery of what we now call X-rays in his scientific publication he reported in another context shortly afterwards.

**2. Overestimated serendipity.** Newton's apple is generally regarded as a typical example of an exaggerated or apocryphal story. Apocryphal means literally: a story of which it is not known whether it belongs in the bible or not. And in a figurative sense: its authenticity cannot be proven nor refuted. We can't prove the authenticity of the falling apple and whether this inspired Newton to discover the universality of gravity. But we do know is that there are several examples of researchers who at the end of their life tell 'the inside story', 'the how it really happened story', 'the story behind the story'. Then they have often arrived the 'hall of fame'. Then they dare to tell the eventually serendipitous aspects of their findings. Nobel Prize acceptance lectures, for example, are good sources for such stories. Next to a 'context of discovery' and a 'context of justification', I think there is also a 'context of fame', which allows the finder to tell the 'inside story', the 'how it really happened story' or the 'story behind the story'. Of course those stories are not necessarily true. Although even the apple story should not be true

it is any how a very didactic way to explain the discovery of the universality of gravity. ‘*Si non è vero, è ben trovato!*’ [5]

Anyhow, the confrontation of the student Isaac Newton with a falling apple is a classic scientific anecdote (*a n ek dotein* (Greek) = not e- dite), a not published story. The story has been called, with some wit, the second intervention of the apple into human history. But is the story true? Or stays it an apple of discord? Dr William Stukeley, who knew Newton well in his old age, is one of the four independent sources of this story:

On 15 April 1726 I paid a visit to Sir Isaac at his lodgings in Orbels Buildings in Kensington, I dined with him, and spent the whole day with him, alone [...].

After dinner, the weather being warm, we went into the garden and drank tea, under the shade of some apple trees, only he and myself. Amidst other discourse, he told me he was just in the same situation as when, formerly, the notion of gravitation came into his mind. It was occasioned by the fall of an apple, as he sat in a contemplative mood. Why should that apple always descend perpendicularly to the ground, thought he to himself? Why should it not go sideways or upwards, but constantly to the earth centre? Assuredly, the reason is, that the earth draws it. There must be a drawing power in matter: and the sum of the drawing power in the matter of the earth must be in the earth’s centre, not in any side of the earth. Therefore does this apple fall perpendicularly, or towards the centre. If matter thus draws matter, it must be in proportion of its quantity. Therefore the apple draws the earth, as well as the earth draws the apple. That there is a power, like that we here call gravity, which extends itself through the universe.

And thus by degrees he began to apply this property of gravitation to the motion of the earth and of the heavenly bodies, to consider their distances, their magnitudes and periodical revolutions; to find out that this property, conjointly with a progressive motion impressed on them at the beginning, perfectly solved their circular courses; kept the planets from falling upon one another, or dropping all together into one centre; and thus he unfolded the universe. This was the birth of those amazing discoveries, whereby he built philosophy on a solid formation, to the astonishment of all Europe. [6]

The apple in this case is the metaphor (example or model) for the moon. Did the falling apple led Newton to his revolutionary synthesis, his ‘discovery’ of the universality of gravity? Newton himself did not write about how the falling apple hit the fertile soil of his imagination. But we do know that he published his *Principia* (1687) twenty- one year after the presumed apple (1666), and also that revealed the story about the apple sixty years after it happened. The overestimation of the role of the falling apple might be linked with the fact that in our culture the apple is a symbol of knowledge. [7]

Anyhow, as the Frenchmen Paul Valéry wrote: ‘Il fallait être Newton pour apercevoir que la lune tombe, quand tout le monde voit bien qu’elle ne tombe pas.’ [8]

**3. Fabricated serendipity.** The most quoted, but alas often wrongly cited, remark on the role of chance in the discovery of electromagnetism was made in 1854 by Louis Pasteur in his opening speech as Dean of the new ‘Faculty of Sciences’ at Lille, France:

C’était dans cette memorable année 1822. Oersted, physicien Suédois [Danois], tenais en mains un fils de cuivre réuni par ses extrémitées aux deux pôles d’une pile de Volta. Sur sa

table se trouvait une aiguille aimantie placée sur son pivot, et tout à coup il vit, (par hasard diriez- vous peut- être , mais souvenez- vous que dans les sciences d'observation le hasard ne favorise que des esprits préparés) il vit tout à coup l'aiguille se mouvoir et prendre une position très différente de celle que lui assigne le magnetisme terrestre. Un fil traversé par un courant électrique fait devier de sa position une aiguille aimantie. Voila, messieurs, la naissance du télégraphie actuel. [9]

The first ‘accidental version’ of the history of Oersteds discovery was fabricated by Ludwig Wilhelm Gilbert in his *Annalen der Physik*, 1820. Gilbert was an ennemy of *Naturphilosophie* (a speculative approach to natural science) and his *Annalen* was the organ for a number of attacks on *Naturphilosophie*. And Oersted’s belief in the unity of powers of nature was characterisitc of the German Romantic school of *Naturphilosophie*. [10] Here the ‘made up’ version of Gilbert:

Was alles Forschen und Bemühen nicht hatte geben wollen, das brachte ein Zufall Hern Professor Oersted in Kopenhagen, während seiner Vorlesungen über Electricität und Magnetismus im verflossenen Winter. Er und die würdigen Naturforscher, in Gemeinschaf mit welchen er den Fund verfolgte, haben durch ihre Versuche die folgenreiche Entdecküng völlig bewährt, dass der geschlossene galvanisch- electrische Kreis starker Apparate einen mächtige, bisher ungeahnte Wirkung auf den Magnetnadel äussert. [11]

The earliest of account on how Oersted came upon his discovery of electromagnetism was published in Danish before April 1821 in which Oersted stated:

I was brought back to them [experiments on the effects of an electric current on a magnet] through my lectures on electricity, galvanism, and magnetism in the spring of 1820. The auditors were mostly men already considerably advanced in science; so these lectures and the preparatory reflections led me on to deeper investigations than those which are admissible in ordinary lectures. Thus my former conviction of the identity of electrical and magnetic forces developed with new clarity, and I resolved to test my opinion by experiment. The preparations for this were made on a day in which I had to give a lecture on the same evening. [...] I called attention to the variations of the magnetic needle during a thunderstorm, and at the same time I set forth the conjecture that an electric discharge could act on a magnetic needle placed outside the galvanic circuit. I then resolved to make the experiment. Since I expected the greatest effect from a discharge associated with incandescence, I inserted in the circuit a very fine platinum wire above the place where the needle was located. The effect was certainly unmistakable, but still it seemed to me so confused that I postponed further investigation to a time when I hoped to have more leisure. (All my auditors are witnesses that I mentioned the result of the experiment beforehand. The discovery was therefore not made by accident, as Professor Gilbert has wished to conclude from the expressions I used in my first announcement.) At the beginning of July these experiments were resumed and continued without interruption until I arrived at the results which have been published.

I presume that Gilbert didn’t like the *Naturphilosophie*, he depicted Oersteds classic discovery of electromagnetism as ‘accidental’, and unrelated with Oersteds belief in *Naturphilosophie*.

**4. Underestimated serendipity.** In general the role of serendipity in science, technology and art is underestimated. This is mainly and unintentionally

caused by the way we rationalize *a posteriori* about theoretical and experimental research and its results, when we publish. The not strictly rational, logical, chronological, or searched components (like chance, fortuitous, accident, erroneous, surprising, unsought, (n)ever thought or dreamt of, unknown, *etc.*), which have also led to these results are therefore underestimated and sometimes even banned from the theater and totally hidden behind the décor. The next step is that pure rationality becomes the norm, not only regarding the results, but also regarding everything that has led to these results. Scientists then report their results as following directly and logically from their initial hypothesis, omitting possibly crucial serendipitous events. They write their article in such a way that the reader, when he wants to, can redo the search, can *research* it, to see whether the results are reproducible. It is quite rational and efficient to publish that way. But reading and interpreting such articles as 'the real story' about a finding, can unwillingly brainwash the reading researcher in such a way and to such an extent that he neglects during his own research the flowers along the road that can form a nicer bouquet than the cultivated ones he is looking for. This can cause a loss of serendipity: the aim and/or the plan spoils the journey. A successful researcher has therefore an open eye for sought findings and a second eye open for unsought findings. As the German Nobel prize winning chemist Von Laue commented: '*Es gibt manchmal Verdienst ohne Glück, aber niemals Glück ohne Verdienst!*' [12] Peter Medawar advised in his essay *Hypothesis and Imagination* to listen at the keyhole if you want to know how researchers actually work:

What scientists *do* has never been the subject of a scientific, that is, an ethological inquiry. It is no use looking to scientific 'papers', for they not merely conceal but actively misrepresent the reasoning that goes into the work they describe (..) Nor is it much use listening to accounts of what scientists *say* they do, for their opinions vary widely enough to accommodate almost any methodological hypothesis we may care to device. Only unstudied evidence will do - and that means listening at a keyhole. [13] The avarage experimental scientific article is written, accepted and published in such a way that the experiment is in principle reproducible, with the unintended consequence that the serendipitous aspects of the investigation are underestimated, because this they are irrelevant in this respect.

**5. Hidden or denied serendipity.** Many cases of serendipity are not reported because the scientist didn't want to attribute his success to an accident. To create an image of logic many examples have been buried intentionally by investigators who refuse to concede that human frailty was a factor in their discoveries. And the form in which discoveries are reported in scientific journals conceals often the role which chance may have played. As a result, the literature on serendipity is meager and case histories are not readily accessible. The scientist ought to realize that: 'Le hasard ne sert que ceux qui savent le capter.' [14] It is thus not insulting to the integrity of a discoverer to admit that his finding was accidental, for serendipity does not depend entirely on chance. [15] Chance merely provides the most suitable

circumstances, but it must be recognized opportunely and turned to advantage by a creative intellectual process. [16]

**Unpublished serendipity.** Two famous chemotherapeutic drugs were discovered because they happened to be present as impurities in other substances which were being tested. Scientists associated with the work have told the stories of these two discoveries but have asked not to publish them as other members of the team may not wish the way in which they made the discovery to be made public. [17]

**Later published serendipity.** McLean, looking for blood clotting factors, discovered heparine as a anticoagulant (a factor preventing bloodclotting), but didn't 'tell it like it was' until 41 years later, as he said himself: 'I had in mind, of course, no thought of an anticoagulant, but the experimental fact was before me; and I retested again and again until I was satisfied that an extract of liver (..) possessed a strong anticoagulant action (..).' [18]

**Summary.** Serendipitous aspects of findings in the professional scientific literature are in general underreported or not mentioned at all. The main reason for this is that such articles are written (and accepted) for publication in such a way that the study can be precisely reproduced: to check the results. But happily a minority of authors has always been explicit about serendipitous facets of their study in their articles and/or afterwards in their comments. A whole spectrum of documented, overestimated, fabricated, underestimated, hidden or even denied serendipity, was found in articles, and also later comments them by the authors themselves and/or others. The best source for the serendipity of a finding is and stays the author himself, who refers to it in his scientific article or in a comment.

#### References:

- [1] Cl. Bernard, *Introduction à l'étude de la médecine expérimentale*, Paris, 1865), translated by H.C. Greene, New York, Dover, 1957, p. 166: 'Nothing is accidental, and what seems to us accident is only an unknown fact whose explanation may furnish the occasion for a more or less important discovery.'
- [2] *Nature*, Jan. 23, 1896, p. 274, an English translation of his German article: '*Über eine neue Art von Strahlen*'.
- [3] (Ch. Nootnangle, 'How Röntgen Discovered the X-Rays', *The Electrical Engineer*, New York, 22, 125, 5. August 1896, as quoted in *The Faber Book of Science*, ed. J. Carey, Faber and Faber, 1996, p. 181.
- [4] W.I.B. Beveridge, *The Art of Scientific Investigation*, Mercury Books, London, 1961, p. 126- 7.
- [5] Giordano Bruno, 1548- 1600.
- [6] W. Stukeley, *Memoirs of Sir Isaac Newton's Life*, 1752, cited by J. Carey in *The Faber Book of Science*, Faber & Faber, London, 1996, p. 30- 1.
- [7] R.S. Westfall, *Never at rest*, A biography of Isaac Newton, CUP, Cambridge, 1980, p. 15.

- [8] P. Valéry, *Analects, Collected Works*, 14, Ed. J Matthews, London, Routledge & Kegan Paul, 1970: 'One had to be a Newton to notice that the moon is falling, when everyone sees that she doesn't fall.'
- [9] From a copy of Pasteurs manuscript, send to me by the medical historian Prof. Dr. M. Grmek (†), Sorbonne, Paris.
- [10] R.C. Stauffer, *Isis*, 1953, 44: 307- 310, *Isis*, 1957, 48: 33- 50). Here the 'made up' version of Gilbert:
- [11] Vorbericht von Gilbert in his Ludwig Wilhelm Gilbert in his *Annalen der Physik*, 1820, 66: 291- 294.
- [12] (P. van Andel, 'Anatomy of the Unsought Finding. Serendipity: Origin, History, Domains, Traditions, Appearances, Patterns and Programmability', *British Journal for the Philosophy of Science*, 45, 1994, 631- 48, Oxford University Press.
- [13] Quoted by F.L. Holmes, The fine structure of scientific creativity, *History of Science*, 19, 1981, p. 62.
- [14] Ch. Nicolle, *Biologie de l'invention*, Alcan, Paris, 1932, p. 30: 'Chance serves only those who know how to win her.'
- [15] B.E. Schaar: 'Serendipity - A Comedy of Errors.' *The Chemist*, p. 411- 418, Sept. 1958.
- [16] R.E. Rossman, 'The History and Significance of Serendipity in Medical Discovery', *Transactions & Studies of the College of Physicians of Philadelphia*, 4th series, V. 33, 1965- 6, Waverly Press, Inc., Baltimore, p. 119.
- [17] W.I.B. Beveridge, *The Art of Scientific Investigation*, Mercury Books, London, 1961, p. 163.
- [18] J.H. Comroe, *Retrospectroscope, Insightys into Medical Discovery*. Menlo Park, CA: Von Gehr Press, p. 90- 2.

Finally, a superb and recent book about serendipity is recommended:

- [19] R.K. Merton & E. Barber, The Travels and Adventures of Serendipity, A Study in Sociological Semantics and the Sociology of Science, Princeton University Press, 2004, 313 p.