

**BULLETIN N° 149  
ACADÉMIE EUROPEENNE  
INTERDISCIPLINAIRE  
DES SCIENCES**



**Séance du mardi 9 novembre 2010:**

**Présentation par Franck VARENNE de l'Université de Rouen de :  
« La question des validations de modèles en sciences sociales.  
Le cas de l'iconicité croissante des simulations computationnelles. »**

**Prochaine séance mardi 14 décembre 2010- MSH, salle 015-18heures :  
Présentation par Daniel KAYSER de l'Université Paris XIII de :  
"Quelques réflexions sur le traitement automatique du langage naturel".**

# ACADEMIE EUROPEENNE INTERDISCIPLINAIRE DES SCIENCES

## FONDATION DE LA MAISON DES SCIENCES DE L'HOMME

**PRESIDENT** : Pr Victor MASTRANGELO  
**PRESIDENT Sortant**: Michel GONDRAN  
**SECRETARE GENERAL** : Irène HERPE-LITWIN  
**TRESORIER GENERAL** : Bruno BLONDEL  
**MEMBRES DU CA** Patrice CROSSA-RAYNAUD, Claude ELBAZ

**PRESIDENT FONDATEUR** : Dr. Lucien LEVY (†)  
**PRESIDENT D'HONNEUR** : Gilbert BELAUBRE  
**SECRETARE GENERAL D'HONNEUR** : Pr. P. LIACOPOULOS (†)

**CONSEILLERS SCIENTIFIQUES** :  
**SCIENCES DE LA MATIERE** : Pr. Gilles COHEN-TANNOUDJI  
**SCIENCES DE LA VIE ET BIOTECHNIQUES** : Pr François BEGON

**SECTION DE NICE** :  
**PRESIDENT** : Doyen René DARS

**SECTION DE NANCY** :  
**PRESIDENT** : Pr Pierre NABET

Novembre 2010

# N°149

## TABLE DES MATIERES

- P. 03 Compte-rendu de la séance du mardi 9 novembre 2010
- P. 07 Compte s-rendus de la section Nice Côte d'Azu
- P. 13 Annonces
- P.15 Documents

**Prochaine séance: mardi 14 décembre 2010 18h MSH, salle 015-18heures :**  
**Présentation par Daniel KAYSER de l'Université Paris XIII de :**  
**"Quelques réflexions sur le traitement automatique du langage naturel"**

**ACADEMIE EUROPEENNE INTERDISCIPLINAIRE DES  
SCIENCES**

Maison des Sciences de l'Homme, Paris.

**Séance du  
Mardi 9 novembre 2010**

**Maison des Sciences de l'Homme, salle 214, à 18 h.**

La séance est ouverte à 18 h. 00 sous la Présidence de Victor MASTRANGELO et en la présence de nos collègues Gilbert BELAUBRE, Alain CARDON, Gilles COHEN-TANNOUDJI, Françoise DUTHEIL, Claude ELBAZ, Michel GONDRAN, Walter GONZALEZ, Irène HERPELITWIN, Gérard LEVY, Pierre MARCHAIS, Pierre PESQUIES.

Etaient excusés François BEGON, Bruno BLONDEL, Brigitte DEBUIRE, Jean -Pierre FRANCOISE, Marie-Louise LABAT, Saadi LAHLOU, Jacques LEVY, Alain STAHL.

**I) Question diverses**

**A) Edition des actes sur le congrès « Emergences : de la fascination à la compréhension »**

- Notre Collègue Gilbert BELAUBRE nous informe de l'obtention d'un contrat définitif de publication auprès des éditions SPRINGER Methodos –Series de l'édition en langue anglaise des actes du congrès sur l'Emergence de décembre 2008. Le titre de l'ouvrage en anglais serait : « Emergences » .

- Notre Collègue a également cherché la collaboration de Paul HUMPHREYS et Marc BEDEAU pour faire un article de présentation de l'ouvrage.

Après la publication de l'ouvrage en anglais chez SPRINGER, il ne sera probablement pas difficile de trouver un éditeur français pour la version francophone.

**B) Subventions**

Une recherche sur l'obtention de subventions auprès de la commission européenne devrait être développée.

## II) Présentation par Franck VARENNE de l'Université de Rouen de :« La question des validations des modèles en sciences sociales. Le cas de l'iconicité croissante des simulations computationnelles »

Franck VARENNE , Maître de conférences en philosophie de la connaissance à l'Université de Rouen, est également Directeur du département de philosophie de l'Université de Rouen et Chercheur rattaché au Groupe d'Etude des Méthodes de l'Analyse Sociologique de la Sorbonne (GEMASS - UMR 8598 / CNRS / Paris 4 - Sorbonne) : [www.gemas.msh-paris.fr/](http://www.gemas.msh-paris.fr/) et Membre de la Société de Philosophie des Sciences : [www.sps.ens.fr/](http://www.sps.ens.fr/)

Il possède une formation à la fois scientifique et philosophique :

- Ingénieur de l'Ecole Nationale Supérieure d'Electricité (Option métrologie)
- Docteur en histoire et philosophie des sciences (2004)

Il est l'auteur de plusieurs ouvrages d'épistémologie sur les méthodes de modélisation et de simulation dans plusieurs domaines :

- *Les notions de métaphore et d'analogie dans les épistémologies des modèles et des simulations*, Paris, Pétra, 2006.  
<http://www.amazon.fr/exec/obidos/ASIN/2847430091/qid%3D1142437070/171-5081028-4722655>
- *Du modèle à la simulation informatique*, Paris, Vrin, 2007. [www.amazon.fr/Du-Mod%C3%A8le-%C3%A0-Simulation - Informatique/dp/2711619257/ref=sr\\_1\\_2](http://www.amazon.fr/Du-Mod%C3%A8le-%C3%A0-Simulation-Informatique/dp/2711619257/ref=sr_1_2) ou <http://www.vrin.fr/html/main.htm?action=loadbook&isbn=2711619257>
- *Qu'est-ce que l'informatique ?*, Paris, Vrin, 2009.  
[www.amazon.fr/Quest-ce-que-linformatique-Franck-Varenne/dp/2711621782/ref=sr\\_1\\_1](http://www.amazon.fr/Quest-ce-que-linformatique-Franck-Varenne/dp/2711621782/ref=sr_1_1)
- *Formaliser le vivant : lois, théories, modèles ?*, Paris, Hermann, 2010. [www.editions-hermann.fr/ficheproduit.php](http://www.editions-hermann.fr/ficheproduit.php) ou [www.amazon.fr/Formaliser-vivant-Franck-Varenne/dp/2705670890/ref=sr\\_1\\_9](http://www.amazon.fr/Formaliser-vivant-Franck-Varenne/dp/2705670890/ref=sr_1_9)

En prologue le Prof. Varenne indique une première caractérisation des fonctions des modèles « Un modèle est ce à quoi on se rapporte pour représenter quelque chose: d'où l'ambivalence entre le modélisant et le modélisé (Cf., également, Legay 1997). En sciences et techniques : le modèle est en général le modélisant.

Il fait un bref historique : synonyme de représentation réduite ou simplifiée (modèles physiques, maquettes...). Il cite notamment Maxwell (modèle des boules de billard en théorie cinétique des gaz) et Boltzmann (figuration non du réel mais d'un calcul : notion d'espace des phases). Au début du 20<sup>ème</sup> siècle : un mouvement de formalisation des modèles (travaux de Bouleau 1998) : les modèles mathématiques s'émancipent de la médiation de la Physique. Cela est dû au développement « interne » des mathématiques (logicisme puis théorie des modèles), à la diversification des axiomatiques utilisables, à l'essor indépendant des mathématiques dites descriptives (statistiques, analyse de données

multivariées...): le statut épistémique de ce type de mathématiques semble nouveau.

Il propose avec Minsky une caractérisation minimale du concept de **MODELE** plutôt qu'une définition : «*Pour un observateur B, un objet A\* est un modèle d'un objet A dans la mesure où B peut utiliser A\* pour répondre à des questions qui l'intéressent au sujet de A*».

Un modèle n'est pas nécessairement une représentation. Le modèle est un objet et sa fonction est la facilitation d'une médiation. Toujours dans le prologue, il définit cinq grandes classes.

1. Facilitation d'une observation, d'une expérience ou d'une expérimentation ;
2. Faciliter une représentation intelligible via une représentation mentale ou une conceptualisation ;
3. Faciliter une théorisation\* (dans l'élaboration, l'interprétation, l'illustration d'une théorie ; tester la cohérence interne d'une théorie ; faciliter l'application, l'hybridation des théories dans les systèmes hétérogènes : modèles polyphases...);
4. Faciliter la médiation entre discours autour d'un phénomène ; faciliter la formulation du questionnement (la communication, l'écoute, la délibération, la concentration, la co-construction des hypothèses de gestion de systèmes mixtes de type sociétés-nature)
5. Ne faciliter ni la formulation du questionnement ni la formulation de la réponse, mais la décision seule (ex. décision rapide dans un contexte effectivement complexe)

L'orateur dénombre ainsi au moins vingt fonctions différentes pour les modèles.

Toujours dans le cadre du prologue, il s'intéresse à la notion de « validation ». Il faut dit-il la distinguer de celle de « vérification », de « corroboration » (degré auquel une hypothèse théorique a résisté à des tests sévères et a ainsi fait ses preuves) et de « validation en logique formelle » (construction d'une proposition logique complexe de manière syntaxique à partir des axiomes et des règles de transformation de l'axiomatique considérée). Une validation et une corroboration s'opposent à une réfutation (montrer la fausseté, l'inconsistance...). Un modèle peut être inconsistant (auto-contradictoire) et néanmoins opératoire sur le terrain (donc « valide » en ce sens).

Il note un essor important des sciences sociales computationnelles depuis les années 1990.

Il évoque pour cela trois thèses interprétatives :

- Thèse faible : simple diffusion d'un nouveau type de formalisme inspiré de la Physique computationnelle et de l'intelligence artificielle distribuée ;
- Thèse moyenne : les simulations computationnelles proposent une nouvelle méthodologie aux sciences sociales ;
- Thèse forte : au-delà d'une nouvelle méthodologie, il s'agit de rien de moins que d'une manière inédite de fonder une nouvelle unité des sciences sociales.

Comment faut-il trancher ?, comment évaluer la portée épistémique de ces approches

Il nous propose pour ce faire une démarche d'épistémologie appliquée. Il définit tout d'abord le mot « computationnel ». Une computation est une opération pas à pas sur des symboles. Dans un sens plus général, « *une science computationnelle est une science qui utilise l'ordinateur comme machine à modéliser et à simuler au moyen de computation* ». Il précise toutefois que le mot « computationnel » est en fait polysémique.

Avant de parler de simulations en sciences sociales, il explicite les distinctions conceptuelles entre modèles et simulations. Les modèles en sciences sociales sont traditionnellement de trois types : modèle pour l'analyse de données, modèle prédictif, modèle explicatif. Le terme simulation dénote toujours au moins un processus de symbolisation à deux étapes : traitement pas à pas des symboles, la 2<sup>ème</sup> étape est l'observation des résultats de computation et, pour cela, de changement de niveau d'observation des symboles.

Ensuite il argumente sur la nécessité en sciences sociales d'utiliser directement la simulation numérique sans passer par une étape de modélisation. L'analyse de données bute souvent sur les phénomènes d'interaction entre variables. Elles ne sont pas seulement juxtaposées, elles se commandent les unes les autres. Il est alors préférable de construire directement un modèle « causal complexe » non linéaire, de le

simuler sur ordinateur et d'observer le résultat des simulations pour diverses conditions des paramètres. Ainsi la simulation sur ordinateur s'est généralisée à partir des années 1970 (sciences sociales, de la vie, de l'environnement...). Boudon, d'après F. Varenne avait prédit ce passage dès 1967. Il cite cet auteur au sujet du modèle causal rendu nécessaire par ce passage à la simulation comme pouvant expliquer et non seulement reproduire les corrélations existant entre variables.

Enfin il développe trois approches d'analyse par ordinateur désignées par « computationnelles » : la micro-simulation, les automates cellulaires et les approches par agents. La micro-simulation s'effectue simultanément à deux niveaux : le niveau de l'individu et le niveau agrégé (variables macro-économiques..). Les individus sont regroupés dans un échantillon représentatif de la population. A chaque pas de temps, on applique aux individus de cet échantillon un modèle probabiliste de transition. Une telle micro-simulation peut-être identifiée comme une simulation numérique sociale. *On numérise, on discrétise, mais on discrétise fictivement, cela de manière à pouvoir calculer pas à pas un comportement global de la société. L'individu y est à mi-chemin entre un symbole dénotant expressément une réalité sociale et un truc de calcul.*

La simulation par automates cellulaires est issue de la théorie des automates d'Ulam et Von Neuman. Un automate cellulaire est un réseau de cellules situé sur un maillage de calcul. Chaque cellule est caractérisée par un état interne (liste d'attributs) et une règle de transition de cet état. Celle-ci n'est fonction que de l'état des cellules voisines dans le maillage. Pour les sciences sociales, l'intérêt de cette formalisation est double. Elle permet d'imiter de manière plus réaliste le comportement de décision et d'action de chaque agent social. Par rapport aux micro-simulations, elle met en œuvre une *véritable interaction* entre cellules représentant les agents (effets de voisinage...). On peut ainsi simuler des phénomènes d'influence ou d'imitation de proche en proche. F. Varenne cite, comme exemple, le modèle de Schelling (Apparition des ghettos dans les villes américaines...).

Enfin, il nous expose les approches par agents ; ces derniers sont issus du domaine de l'Intelligence Artificielle Distribuée (IAD). Un agent est caractérisé par son autonomie, sa « capacité sociale » (interaction avec les autres agents), sa réactivité (réaction appropriée à des stimuli venant de son environnement) et sa proactivité (initiative propre dans les buts poursuivis). Il souligne que dans ce type de modélisation computationnelle, l'iconicité des dimensions psychologiques mais aussi physiques de l'agent sont considérablement accrues. Les agents peuvent être très hétérogène et évoluer au cours de la simulation. Il cite comme exemple : la simulation « Emergency of Society » (Cf. Doran & Palmer, 1994). Toutefois des critiques sont apparues. On doit tendre vers des agents plus intelligents : intelligence sociale, émotions...

En conclusion, il y a un apport épistémologique de la « science sociale générative ». On peut expliquer et prédire à la fois, mais sans toujours comprendre. Les simulations computationnelles proposent une nouvelle méthodologie aux sciences sociales (extension du domaine d'explication) et d'aller vers l'« individualisme méthodologique complexe », d'envisager des proximités méthodologiques avec celles de la Physique et de la Biologie. Cependant un couplage avec les techniques traditionnelles de modélisation statistique devra être envisagé pour « resserrer » le lien avec le terrain.

Après discussion générale sur ce très riche exposé, la séance prend fin.

---

\*Théorie: large ensemble d'énoncés –éventuellement formalisés et axiomatisés –formant système et donnant lieu à des inférences susceptibles de valoir pour tout un type de phénomènes donné ≠ modèle dans ce contexte = un ensemble de «contraintes particulières sur des interactions» (Livet 2007).

# Comptes-rendus de la section Nice-Côte d'Azur

## *Compte rendu de la séance du 16 septembre 2010*

*La simple observation, la description, le suivi des écosystèmes apportent plus à la compréhension du monde que les lourds procédés de mesures physico-chimiques qui accumulent des données souvent peu pertinentes qui, plus maniables et modélisables, semblent plus scientifiques.*

Bertrand Richer de Forges, 2010.

(140<sup>ème</sup> séance)

### **Présents :**

Jean Aubouin, Richard Beaud, Sonia Chakhoff, Patrice Crossa-Raynaud, François Cuzin, Guy Darcourt, René Dars, Jean-Pierre Delmont, Yves Ignazi, Jacques Lebraty, Maurice Papo.

### **Excusés :**

René Blanchet, Jean-Paul Goux, Michel Lazdunski, Jean-François Mattéi.

### **1- Approbation du compte rendu de la 139<sup>ème</sup> séance.**

**Le compte rendu est approuvé à l'unanimité des présents.**

### **2- Programme des activités.**

### **Conférences :**

La Mairie de Nice a réservé pour nous la salle du MAMAC, Promenade des Arts, de 16 heures à 18 heures, pour une série d'exposés centrés autour de « La diversité » :

#### **Le mardi 26 octobre**

*La diversité dans la domestication des espèces : P. Crossa-Raynaud*

#### **Le vendredi 12 novembre**

*La diversité en médecine par le Professeur Jean-Pierre Delmont*

#### **Le mercredi 15 décembre**

*La diversité et le génome par le Professeur François Cuzin*

Suivraient, en 2011, à des dates à préciser :

- *La diversité psychique chez l'Homme par le Prof. Guy Darcourt,*
- *La diversité en géologie par le Prof. René Dars,*
- *La diversité en économie par le Professeur Jacques Lebraty,*
- *La diversité dans l'art par Sosno,*
- *Philosophie et diversité : Fr Richard Beaud et/ou Jean-François Mattéi.*

### **Publications :**

**Les publications de l'Académie sur le site culturel de la Mairie de Nice sont en bonne voie.**

**Après une brève présentation du but de l'Académie et la liste des membres, on trouvera bientôt, sur ce site :**

- **le texte du colloque (2007) sur « Les climats de la Terre au cours des temps » qui demeure d'actualité et dont l'édition est épuisée,**
- **les conférences au MAMAC dans le cycle de « La diversité » au fur et à mesure qu'elles auront été prononcées.**

### **Publicité :**

**Au sujet de nos conférences au MAMAC, Madame Guillemot, qui va s'occuper de l'installation du site de l'AEIS à la Mairie, contactera Madame Caroline Magne-Ciotti, directrice de la communication. Elle nous conseille de contacter Nice Matin, les journaux gratuits, etc. On peut aussi envisager une annonce payante grâce aux crédits disponibles. Une affiche est en préparation.**

### **3- Débat : la diversité en économie par Jacques Lebraty.**

MANAGEMENT ET DIVERSITE

INTRODUCTION : Genèse d'une recherche

Notre équipe de recherche<sup>1</sup> travaillait sur le thème des organisations hybrides. Il s'agissait d'entreprises mêlant dans leur organisation et leur fonctionnement plusieurs types de logiques (centralisation et décentralisation, partenariats public-privé, économie capitaliste et coopérative, systèmes informatiques différents, cultures du personnel multiples etc.).

Au cours de cette recherche, nous avons rencontré le problème posé, selon la commission de Bruxelles, par l'existence du système coopératif bancaire, notamment français. Celui-ci, du fait des particularités de son statut entraverait le bon fonctionnement du système concurrentiel.

La controverse née à ce propos suscitait une interrogation fondamentale sur la nature même de la concurrence ; ce concept devait-il être compris comme le combat de clones ou comme une « compétition-coopérative » dans laquelle la diversité loin de constituer une entrave représentait la caractéristique nécessaire d'une véritable concurrence ?

Nous tournant vers l'abondante littérature sur le « management de la diversité » nous avons constaté qu'elle ne concernait que les ressources humaines : proportion des femmes dans les

<sup>1</sup> Dirigée par Les professeurs Jacques LEBRATY (Université) et Lyvie Guéret-Talon (Skéma)





***Compte rendu de la séance du 21 octobre 2010***  
(141<sup>ème</sup> séance)

*La simple observation, la description, le suivi des écosystèmes apportent plus à la compréhension du monde que les lourds procédés de mesures physico-chimiques qui accumulent des données souvent peu pertinentes qui, plus maniables et modélisables, semblent plus scientifiques.*  
Bertrand Richer de Forges, 2010.

**Présents :**

Jean Aubouin, Richard Beaud, Patrice Crossa-Raynaud, François Cuzin, Guy Darcourt, René Dars, Jean-Pierre Delmont, Yves Ignazi, Maurice Papo.

**Excusés :**

René Blanchet, Sonia Chakhoff, Jean-Paul Goux, Jacques Lebraty.

**4- Approbation du compte rendu de la 140<sup>ème</sup> séance.**

**Le compte rendu est approuvé à l'unanimité des présents.**

**5- Le mois écoulé.**

**Publication :**

Les publications de l'Académie, section Nice Côte d'Azur, paraissent désormais sur le site suivant :

<http://etudiants.nice.fr/culture-scientifique,140.html>

Est déjà disponible, outre la présentation des buts de l'Académie et la liste des membres, le dernier volume édité : « Les climats de la Terre au cours des temps ».

La première conférence du nouveau cycle : « La diversité » : *La diversité dans la domestication des espèces* par notre confrère Patrice Crossa-Raynaud sera bientôt disponible sur le site.

D'autres conférences suivront :

**Vendredi 12 novembre 2010 à 16 heures**

*La diversité en médecine*

par le Professeur Jean-Pierre Delmont

**Mercredi 15 décembre 2010 à 16 heures**

*Diversité et génome*  
par le Professeur François Cuzin

puis

le 26 janvier 2011

le 23 février 2011

le 23 mars 2011

le 25 mai 2011

le 22 juin 2011

**6- Débat : « Y a-t-il une finalité dans le monde » animé par Richard Beaud.**

Le sens de la vie, le sens de l'histoire, le sens du cosmos sont inclus dans le mot finalité.

Presque toutes les différentes écoles de pensée des scientifiques refusent plus ou moins fermement toute allusion à une finalité à travers les sciences de la vie ou chez les astrophysiciens.

Mais, souvent, sans s'en rendre compte, ces scientifiques font appel à des concepts philosophiques dès qu'ils essaient d'interpréter leurs observations.

Tout cela nous amène à la conclusion qu'il n'y a pas de science purement objective. Le scientifique le moins finaliste ne peut pas ne pas franchir une limite vers la réflexion philosophique. Ceci est particulièrement évident avec le créationnisme et la théorie du dessein intelligent.

L'idée essentielle est qu'à l'intérieur du vivant comme à l'intérieur du cosmos lui-même, il y a une force qui pousse vers l'avant. Il y a, dans l'univers, une tension dont le résultat ultime est l'apparition de la conscience et, plus encore, de la pensée (conscience de soi).

D'une certaine manière, on peut dire que le cosmos et les existants sont, du fait du dynamisme interne qui les anime, cause de leur devenir, cause de leur être.

Cette fonction de causalité a été particulièrement étudiée par Aristote (Livre II de la physique), par St Thomas d'Aquin et par Kant. L'homme cherche à mettre de l'ordre dans le monde.

Certains affirment que le monde est en marche vers toujours plus de conscience en relation avec le Tout-autre dont il postule l'existence. D'autres refusent ce postulat et veulent rester dans un déterminisme matérialiste. Ils ne sont en rien inférieurs. Mais refuser un sens entre les éléments observables, c'est déjà affirmer un sens. Tout est toujours tout philosophique.

C'est parce que l'Homme est essentiellement philosophe qu'il est possible de dialoguer et d'échanger.

\*\*\*\*\*

Prochaine conférence

**Vendredi 12 novembre 2010 au MAMAC à 16 heures**

*La diversité en médecine*

par notre confrère le Professeur Jean-Pierre Delmont

\*\*\*\*\*

**Prochaine réunion**  
**le jeudi 18 novembre 2010 à 17 heures**  
**au siège : Palais Marie Christine - 20 rue de France**  
**06000 NICE**

## *Annances*

### **I) En souvenir de la disparition de notre Collègue de Nice , Jacques WOLGENSINGER :**

*C'est toujours avec émotion que nous évoquons notre ami Jacques Wolgensinger... Son intelligence, sa culture et son humanisme mettaient en valeur son culte pour une indéfectible amitié. Tous ceux qui ont eu la chance de le connaître ne sont pas près d'en oublier les témoignages.*

*Maurice PAPO*

### **II) HISTOIRE DE LA PALEONTOLOGIE – REPRESENTATIONS DE LA PREHISTOIRE**

Séminaire de Claudine COHEN - EHESS 2010-2011

**Paléanthropologie et évolution dans la première moitié du 20<sup>e</sup> siècle :**

**Piltdown, l'impact d'un faux**

**Le séminaire a lieu tous les mardis, du 23 novembre au 15 février, de 15 h à 17 h, à l'EHESS, 105 bd Raspail salle 11**

De nombreux travaux récents ont porté sur l'histoire de la paléanthropologie et de la préhistoire du 19<sup>e</sup> siècle, mais la période de la première moitié du 20<sup>e</sup> siècle demeure beaucoup moins bien connue.

Pour explorer cette période complexe, l'affaire Piltdown offre un angle de vue particulièrement éclairant. Reconnue comme « la plus grande supercherie du XX<sup>e</sup> siècle », la fraude de Piltdown fut un épisode majeur de l'histoire de la paléanthropologie. Ce cas exceptionnel d'un faux qui eut un impact décisif sur les orientations et les débats d'une science plus de 40 ans durant, en fait un objet de choix pour l'épistémologue qui s'intéresse aux modalités du raisonnement, de la vérification et de la preuve, pour l'historien des sciences qui s'intéresse aux mutations des courants de la pensée scientifique, et pour le sociologue qui s'intéresse aux formes de la persuasion et à la réception publique des savoirs.

Le séminaire étudiera les différents épisodes de cette affaire, et interrogera son inscription dans le contexte intellectuel, politique, social anglais, français puis international, de la paléontologie et de la préhistoire. Il interrogera tout particulièrement la catégorie de « présapiens européen » aujourd'hui obsolète, et sa signification dans la paléanthropologie du XX<sup>e</sup> siècle.

**Mardi 23 novembre 2010 :**

Etat de la paléanthropologie et de la préhistoire au tournant du 20<sup>e</sup> siècle

**Mardi 30 novembre 2010 :**

La découverte de Piltdown . Acteurs, sites, objets, épisodes.

**Mardi 7 décembre 2010 :** *Présentation d'Eoanthropus dawsoni* à la communauté scientifique : (Londres, 18 décembre 1912)

**Mardi 14 décembre 2010:**

Controverses autour de la découverte de Piltdown en France et en Angleterre : (1912-1925)

**Mardi 4 janvier 2011**

La découverte de l'enfant de Taung (1925): Australopithèques et « chaînons manquants »

**Mardi 11 janvier 2011:**

Henry Fairfied Osborn et la paléanthropologie

**Mardi 18 janvier 2011**

Consolidation de la thèse des « présapiens » européens : Les découvertes de Swancombe (1936) et de Fontéchevade (1947)

**Mardi 25 janvier 2011 :**

Presapiens et racismes en France (1940-1945) (Montandon, Vallois)

**Mardi 1<sup>er</sup> février 2011:**

Carbone 14 : la révolution des datations

**Mardi 8 février 2011 :**

L'identification du faussaire

**Mardi 15 février 2011**

Conclusions : La signification de Piltdown pour la paléanthropologie et la préhistoire du 20<sup>e</sup> siècle

**III) SITE Programme Rencontres de la Connaissance Jean Luc BEAUMONT Section Nice :**

<http://delaconnaissance.unice.fr>

**IV) MANIFESTATION SCIENTIFIQUE DE L'ASSOCIATION TECHNION à L'UNESCO sur les nanotechnologies** proposée par

notre Collègue Claude ELBAZ [S'INSCRIRE A LA JOURNEE](#)

le mardi 14 décembre 2010 à la Maison de l'Unesco, 125 avenue de Suffren, Paris 7<sup>ième</sup>. Informations : <http://www.lesassos.com/sn/event/220/>

## Documents

p. 16 : En vue de sa conférence sur « Quelques réflexions sur le traitement automatique du langage naturel », Daniel KAYSER nous propose un texte intitulé: « Traitement automatique du langage naturel »

p.19 : Comme introduction au traitement du langage naturel, nous vous proposons les deux premiers chapitres du cours « Introduction au TALN et à l'ingénierie linguistique » d'Isabelle Tellier, professeur d'informatique à l'université d'Orléans que l'on peut trouver sur le Web.

p. 31 : Présentation plus détaillée du dernier ouvrage historique et épistémologique de Franck Varenne : ***Formaliser le vivant : lois, théories, modèles ?***, Paris, Hermann, 2010

## Le Traitement Automatique du Langage Nature<sup>2</sup>

Par Daniel KAYSER

Le langage est un sujet sur lequel des réflexions sont attestées depuis plus de 2000 ans, et l'apparition des ordinateurs a fait de ce dernier demi-siècle la période qui a peut-être produit les avancées les plus novatrices et les plus prometteuses, même si aucune percée décisive ne s'est encore produite.

L'enthousiasme un peu brouillon des premiers temps de l'informatique a fait entrevoir, par quelques réalisations spectaculaires, les potentialités de l'ordinateur notamment en matière de traduction automatique dès la fin des années 1940, et, dans les années 1960, de résolution de problèmes scolaires posés en langage naturel, de réponse à des questions simples sur de petites banques de données, d'aide à la recherche documentaire dans des bases textuelles, etc.

Ces réalisations atteignirent vite un palier à partir duquel on vit — on voit encore — deux attitudes se dessiner chez les chercheurs : les uns s'efforcèrent, principalement par des méthodes empiriques, à hausser peu à peu ce palier ; les autres, voulant résoudre au fond les difficultés rencontrées, se concentrèrent sur les théories qui leur permettraient de les surmonter.

Ces deux démarches sont plus complémentaires que contradictoires, et les principaux succès actuels sont le résultat de leur hybridation. L'une et l'autre ont apporté des progrès importants : la première a développé une *ingénierie de la langue* dotée d'outils logiciels qui, à l'intérieur de limites encore astreignantes, rendent d'indéniables services ; la seconde fait évoluer considérablement les idées que nous avons sur notre propre langage. Les informaticiens en quête de théorie se tournèrent naturellement vers la linguistique formelle. Or celle-ci, influencée par la démarche cartésienne, découpe cette étude en étapes (morpho-syntaxe, sémantique, pragmatique), alors que les besoins en traitement automatique du langage nécessitent d'intégrer ces étapes en un tout cohérent.

La *syntaxe*, qui a constitué le centre des travaux des linguistes dans la période 1960-90, repose sur un fait indubitable : on peut généralement juger si une phrase est correctement construite indépendamment de sa signification, et il existe des régularités remarquables dans ce jugement. Se borner à recenser ces régularités à la manière des grammaires traditionnelles n'étant pas satisfaisant, la linguistique s'est donnée pour tâche de découvrir des formalismes puissants pour énoncer des règles générales applicables à toute une langue, ou même à une famille de langues. Mais il faut garder à l'esprit que ces règles ne reposent que sur une intuition — forte dans certains cas, beaucoup plus faible dans d'autres — de l'acceptabilité des constructions de phrases.

Les travaux de *sémantique* pure ont donné des théories d'une grande élégance, dont on met trop rarement en question l'adéquation à leur objet : faire apparaître la "forme logique" d'une phrase est-il nécessaire pour trouver sa signification ? La réponse des sémanticiens est que l'étape suivante, la pragmatique, prenant en compte les connaissances sur le monde et la situation d'énonciation, rend compte de l'écart entre cette forme logique et ce qui est réellement exprimé.

Ainsi définie, la *pragmatique* est censée résoudre tous les problèmes que syntaxe et sémantique formelles ont rejetés hors de leur objet d'étude. C'est cependant de sa faisabilité que dépend la valeur, aussi bien scientifique que pratique, des principales théories linguistiques : s'il s'avérait que la

---

<sup>2</sup> Ce texte reprend des éléments de Daniel KAYSER : *Traitement du langage naturel : un regard sur son évolution et ses perspectives*. T.S.I. vol.19 nos 1-2-3 pp.291-301, janvier-mars 2000, Hermès



pragmatique ne puisse y parvenir, ou que les étapes de syntaxe et sémantique "pures" ne lui apportent qu'une aide très relative, tout l'édifice serait remis en cause.

L'approche ci-dessus, que l'on qualifiera d'*analytique*, est la plus représentative de ce que les informaticiens ont tenté de faire. Mais d'autres voies ont toujours existé. L'ordinateur est par exemple un excellent outil de traitement *statistique*. La théorie de l'information, contemporaine de l'informatique, a brièvement suscité l'espoir, fondé sur les succès de la cryptographie, de constituer une alternative pour déterminer la signification d'une phrase. Les travaux en statistique textuelle, s'ils ont depuis longtemps abandonné cette prétention, se révèlent particulièrement intéressants dans les domaines de la terminologie, de l'indexation, de la critique littéraire p. ex. pour présumer la paternité d'une oeuvre, et comme auxiliaire des approches traditionnelles. Après un premier apogée il y a une trentaine d'années (on parlait alors de "documentation automatique"), ils connaissent une nouvelle jeunesse avec le rapide développement d'Internet et l'amélioration des moyens de calcul.

Une voie qui, reprenant certains acquis de l'approche statistique, est souvent perçue comme alternative du courant analytique, est le *connexionnisme* ; ses succès se situent dans des domaines spécifiques (traitement du signal de parole, levée d'ambiguïtés), mais elle ne propose pas actuellement un cadre suffisamment complet pour rivaliser avec l'approche standard.

L'approche standard voit le monde comme constitué d'objets : l'archétype étant l'objet physique, auquel on ramène par "réification" les événements, les abstractions, etc. et on peut la qualifier de *référentielle* : la signification s'exprime comme l'attribution de propriétés à ces objets. Une autre approche, qui aurait pu davantage séduire les informaticiens, et qui semble plus proche de la vocation du langage, consiste à dire que la signification primordiale d'une phrase n'est pas référentielle mais *inférentielle*.

En énonçant une phrase, on ne cherche pas à dire une vérité sur un univers de discours préexistant, mais à amener son interlocuteur à tirer certaines conclusions. Accepter comme objectif primordial de rendre compte des inférences provoquées par la compréhension d'un texte, mène à représenter un *processus* dans lequel interviennent à la fois des inférences et des actions de construction de modèles, les univers sur lesquels ceux-ci se construisent étant constamment remis en cause.

Dès que l'on travaille sur des textes réels, on s'aperçoit en effet que le langage a un rôle bien plus riche que la simple référence à un univers préexistant. Supposer que les entités mentionnées dans un texte font déjà partie de l'univers de celui à qui il s'adresse, c'est nier à l'auteur la possibilité de faire surgir de nouveaux objets dans cet univers, ou de le découper différemment. Or faire envisager par un interlocuteur des modifications dans son univers est une motivation essentielle pour communiquer.

La nécessité pratique de traiter des corpus de textes réels pousse depuis quelques années à intégrer les résultats très hétéroclites hérités de l'approche traditionnelle. Elle fait également apparaître que certains phénomènes, tout aussi utiles à un traitement des langues que ceux qui ont été abondamment étudiés depuis des décennies, n'ont pas reçu l'attention qu'ils méritent, souvent parce qu'ils se situent à une charnière entre deux volets que la tradition sépare, et que de ce fait, les outils de chacun des volets ne suffisent pas à les résoudre.

Par exemple on sait qu'une analyse syntaxique *complète*, si tant est qu'on puisse définir exactement ce qu'on entend par là, est très coûteuse et parfois inutile pour les objectifs que l'on s'assigne ; on sait aussi qu'une analyse syntaxique au moins *partielle* est pratiquement indispensable quels que soient ces objectifs. Mais on ne sait ni caractériser le niveau d'analyse en fonction des objectifs, ni — au cas probable où ce niveau ne peut être déterminé à l'avance —, comment intégrer rationnellement un approfondissement de l'analyse syntaxique dans la chaîne des traitements à effectuer.

On peut faire une constatation semblable au niveau sémantique ; il est souvent approprié de réduire dans un premier temps les homonymies et les polysémies, de façon à voir si une analyse sémantique fruste atteint les objectifs visés ; si ce n'est pas le cas, on ne sait — à part quelques heuristiques grossières — ni quand ni comment affiner les représentations lexicales et grammaticales des composants du texte.

L'informatique a jusqu'ici traité les langues avec les méthodes qu'elle a excellemment développées pour ses applications traditionnelles : découper en modules, utiliser des structures arborescentes s'interprétant sur des modèles fixes. Or elle dispose d'outils lui permettant de s'affranchir d'hypothèses trop simplistes :

- systèmes de contraintes avec priorités qui, plutôt que de résoudre les niveaux un par un, recherchent un optimum global ;
- logiques non-monotones dans lesquelles des conclusions plausibles restent acquises jusqu'à ce que l'examen de nouveaux faits en fasse préférer d'autres ;
- langages à objets avec héritage multiple ;
- raisonnement à partir de cas où, au lieu de déduire une solution à partir des données d'un problème, on cherche à adapter des solutions connues pour être efficaces sur des problèmes voisins.

L'utilisation de ces outils — et de quelques autres, cette énumération n'ayant rien d'exhaustif — fera progressivement sortir le traitement des langues des ornières dans lesquelles un respect excessif des approches traditionnelles l'a trop longtemps maintenu. Mais on ne doit pas s'attendre à des progrès rapides : beaucoup de questions sont formulées en des termes qui doivent évoluer avant que des réponses puissent apparaître, et les reformulations sont des processus lents. Enfin, il ne faut pas oublier que le problème de la langue a toujours figuré parmi les préoccupations des hommes et que, s'il n'est pas mieux résolu, c'est certainement parce que c'est l'un des plus ardues qu'ils se soient jamais posés.

# Introduction au TALN et à l'ingénierie linguistique

Isabelle Tellier

professeur d'informatique

[Université d'Orléans](#)

[LIFO \(Laboratoire d'Informatique Fondamentale d'Orléans\)](#)

6 rue Léonard de Vinci, BP 6759

45 067 Orléans Cedex 2, France

responsable du [Master CCI](#)

Tel : 02 38 41 72 92

## Chapitre 1

## Introduction

Il ne leur manque que la parole ! Ce que l'on disait volontiers des animaux de compagnie il y a quelques années, on pourrait le formuler maintenant à l'intention des ordinateurs, nos compagnons familiers d'aujourd'hui. Le rêve de dialoguer "naturellement" avec des machines comme avec ses semblables est bien ancré dans l'imaginaire occidental. Tous les films de science-fiction se font l'écho de ce fantasme : impossible d'imaginer un robot un tant soit peu évolué qui n'aurait pas la capacité de parler. En 1950, Turing, le père fondateur de l'informatique, prédit que "dans 50 ans", les ordinateurs auront acquis cette capacité. L'échéance est passée sans que la prédiction ne se réalise. Pourtant, dans le domaine de la technologie numérique, beaucoup de rêves apparemment plus fous ont été largement dépassés. Aujourd'hui, les ordinateurs battent les grands maîtres d'échecs à plate couture, mais ils n'ont toujours pas les compétences langagières d'un enfant de 5 ans. Comment cela se fait-il ? En quoi la capacité de langage est-elle si difficile à "programmer" ? C'est à l'exploration de ces questions que nous invitons dans ce document.

Les travaux de recherche sur le sujet n'ont pourtant pas manqué, et ceci depuis les tout débuts de l'informatique. Une des premières conférences réunissant les pionniers de cette discipline, dans les années 50, portait sur la "traduction automatique". Le contexte politique était propice au développement de la thématique : l'époque était à la guerre froide et intercepter, décoder, traduire automatiquement les messages que s'échangeaient les "rouges" entre eux était pour les Américains un objectif stratégique. Mais l'histoire de ces tentatives est restée longtemps une succession d'échecs et de déconvenues. C'est l'histoire, en fait, d'une prise de conscience pour les informaticiens de ce que les linguistes savaient depuis longtemps : l'extrême complexité des langues naturelles.

Dans les années 70 et 80, les "systèmes experts" (programmes qui simulent, sous la forme de règles, les connaissances d'un expert, pour reproduire son mode de raisonnement) sont à l'honneur. Ils font les beaux jours de "l'intelligence artificielle" dont ils sont alors la principale vitrine. Le thème général du "traitement automatique du langage naturel" (TALN), lui, franchit plus rarement les frontières des laboratoires. Mais il fédère déjà les efforts de nombreux chercheurs. A l'image de ce qui se fait avec les "systèmes experts", on essaie alors de ramener l'usage du langage à des règles symboliques. Beaucoup de

concepts et de modèles évoqués dans les chapitres qui suivent datent de cette époque.

Les années 90 voient arriver des changements considérables : banalisation des ordinateurs personnels, avec des capacités de stockage et de traitement en progression exponentielle, apparition du Web. L'"ingénierie linguistique" naît à ce moment-là. Nous désignerons par ce terme la branche "utilitaire" des recherches en TALN, celles qui donnent lieu à des programmes applicables à des données réelles. Ces données ne manquent pas. Les textes représentent en effet une grande part de ce qui encombre depuis lors la mémoire des ordinateurs personnels, et constituent l'essentiel de ce qui est disponible sur le Web (l'émergence des sons et des images est plus récente). Tout ce qui peut aider à classer ou traiter les documents textuels, à extraire l'information qu'ils contiennent, devient un enjeu majeur. L'ingénierie linguistique se met ainsi au service de la "fouille de textes". Les approches symboliques laissent souvent la place à des méthodes statistiques.

Dans ce document, nous allons tout d'abord, dans le chapitre [2](#), parcourir un peu plus précisément l'histoire des liens entre la linguistique et l'informatique, en insistant notamment sur les "niveaux d'analyse" auxquels on peut soumettre le langage. Les chapitres [3](#) à [7](#) passent ensuite en revue chacun de ces niveaux. Pour chacun d'eux, nous adoptons une structure commune avec une partie "description linguistique", puis une sur sa "modélisation informatique". Une liste de ressources relevant de l'ingénierie linguistique (sites Web ou programmes gratuits) mettant oeuvre certains de ces modèles clôt chaque chapitre.

Le chapitre [8](#) est un peu à part. Il présente quelques-unes des tâches principales de la "fouille de textes" qui, sans être des tâches de traitement de la langue proprement dites, bénéficient de techniques issues de l'ingénierie linguistiques. Tout le parcours des recherches en TALN qui aura été évoqué auparavant prendra alors un autre sens. Peut-être n'aurons-nous pas de si tôt des ordinateurs de compagnie avec qui papoter, mais certains outils d'ingénierie linguistique, eux, sont d'ors et déjà intégrés dans des programmes informatiques que nous utilisons tous les jours...

# Chapitre 2 Traitement Automatique du Langage Naturel

## 1

## Introduction

Toutes les sociétés humaines découvertes de par le monde pratiquent au moins une langue. On en dénombre actuellement environ 5 000 différentes, dont beaucoup sont en voie de disparition faute de locuteurs. Même si l'acquisition du vocabulaire se poursuit tout au long de la vie, tout être humain normalement constitué et inséré depuis sa naissance dans un groupe social est capable, vers l'âge de 5 ans (donc bien avant qu'il ne maîtrise le "raisonnement"), de tenir une conversation courante dans sa langue maternelle. Aucun singe -et aucun ordinateur !- ne peut en faire autant. Parler est bien encore, à l'heure actuelle, "le propre de l'homme".

Pour désigner les langues humaines, on parle maintenant des "langues naturelles", parce que ce sont en quelque sorte des créations collectives spontanées auxquelles on ne peut pas attribuer de date de naissance précise. Les langues naturelles s'opposent ainsi principalement aux "langues artificielles" ou "formelles" que sont notamment les langages de programmation informatiques ou la logique mathématique. On peut très bien aussi classer parmi les "langues naturelles" les créations linguistiques intentionnelles comme l'Esperanto ou le Volapük, qui ne sont les langues maternelles de personne, ou encore certaines "langues des signes" inventées spécialement pour répondre à un besoin : leur point commun est qu'elles sont toutes destinées à être *utilisées par des humains pour communiquer*.

L'objet de la linguistique ou, comme on le dit plutôt désormais, des *sciences du langage*, c'est l'étude scientifique des langues naturelles et, à travers elles, du langage en tant que "faculté de langue" universellement distribuée dans l'espèce humaine. Derrière l'apparente diversité des langues humaines, les linguistes essaient de traquer des fonctionnements communs, des structures partagées, des *universaux*. Les linguistes ne sont pas nécessairement polyglottes ; ils cherchent plus à comprendre les *principes* qui régissent les langues qu'à multiplier les connaissances qu'ils ont de certaines d'entre elles (même si les deux ne sont bien sûr pas incompatibles).

Contrairement à une idée courante, la linguistique n'est pas prescriptive : elle ne dit pas comment "bien parler" ou "bien écrire". Les langues naturelles sont des systèmes vivants qui changent, interagissent, se transforment. Les linguistes se contentent de les observer *telles qu'elles se parlent et s'écrivent*, sans chercher à contrôler ou à limiter leurs évolutions naturelles.

Pour cette étude, l'informatique joue un rôle de plus en plus considérable, via le domaine du *traitement automatique du langage naturel* (TALN). L'informatique est une discipline scientifique récente, qu'il ne faut pourtant pas réduire à la simple utilisation d'ordinateurs et de programmes. Son nom la désigne comme la science du "traitement automatique de l'information". Elle est en fait l'héritière d'une longue tradition mathématique et logique de *modélisation du calcul*. Plus précisément, on peut dire que les fondements de l'informatique sont double :

- le codage des données à l'aide d'*éléments discrets* (les fameux 0/1)
- le codage effectif des traitements à l'aide d'*algorithmes*

C'est par ce biais qu'on va aborder le traitement automatique du langage. Introduire une démarche informatique dans un domaine revient en effet toujours à se poser les mêmes questions :

- quelles sont les données pertinentes de ce domaine, comment les coder ?
- quels sont les traitements pertinents de ce domaine, comment les coder ?

Maîtriser une langue requiert la manipulation de nombreuses données, et la mise en oeuvre de nombreux traitements. Les linguistes les ont progressivement mis à jour et caractérisés, les informaticiens ont progressivement contribué à les modéliser. Le TALN est né de leur interaction. Ce chapitre est destiné à faire un tour d'horizon (forcément simplificateur) de ce domaine.

## 2 Histoires croisées de la linguistique et de l'informatique

Les dates qui suivent ne constituent absolument pas une histoire exhaustive ni de la linguistique, ni de l'informatique. Ce sont plutôt des points de repère marquants qui concernent soit l'une de ces disciplines exclusivement (ce que l'on marquera par (l) pour "linguistique" et (i) pour "informatique" respectivement) soit les deux (ce que l'on marquera par (il) pour "informatique linguistique"). Souvent, les auteurs cités sont ceux qui ont introduit une distinction entre deux concepts, objets ou approches, féconde pour leur discipline. Ces distinctions sont en général destinées à favoriser une des deux branches de l'alternative, pour préciser la démarche suivie et les choix qu'elle entraîne. C'est à ce titre que ces auteurs sont évoqués.

**avant le XVIIIème siècle :**

les précurseurs de la linguistique sont les (très nombreux) auteurs de grammaires descriptives d'une langue donnée (l), les précurseurs de l'informatique sont les mathématiciens qui décrivent des méthodes générales de calcul (par exemple pour résoudre des équations) et les inventeurs de "machines à calculer" mécaniques comme Pascal et Leibniz (i).

**1660 :**

publication de la "Grammaire générale et raisonnée" (connue sous le titre "Grammaire de Port-Royal") d'Arnaud et Lancelot. Son ambition était de décrire les règles du langage en termes de principes rationnels universels (l).

**aux XVIII et XIXème siècles :**

c'est le règne de la linguistique comparative et historique. On compare les langues entre elles et on cherche à en déduire des lois d'évolution générales. Du rapprochement entre diverses langues (latines, grecques, perses, germaniques, celtes, slaves, etc), émerge l'hypothèse que toutes ont un "ancêtre commun" qui sera appelé plus tard "indo-européen" (l). Le XIXème siècle connaît aussi de grands progrès en mathématiques, et voit naître la logique "booléenne" (ou "propositionnelle") par Boole (1815-1864) puis la "logique des prédicats du 1er ordre" par Frege (1848-1925). Les débuts de l'automatisation du travail (métiers à tisser activés par des cartes perforées) inspirent aussi les premiers projets de calculateurs mécaniques. Les plus novateurs et visionnaires sont dûs à l'ingénieur et mathématicien anglais Babbage (1791-1871), qui a conçu les plans de machines ayant les mêmes capacités de calcul que les ordinateurs actuels. Elles n'ont malheureusement pas pu être construites de son vivant (i).

**1916 :**

(l) publication posthume (ce sont des notes de cours publiées par deux de ses étudiants) du "Cours de linguistique générale" du linguiste suisse Ferdinand de Saussure (1857-1913). Saussure introduit plusieurs distinctions et concepts importants :

- il caractérise le langage comme la construction sociale d'un *système de signes*. Un signe est l'association arbitraire entre un *signifiant* (défini comme l'"image acoustique" d'un mot) et un *signifié* (un concept, la représentation mentale d'une chose). Les signes font sens par les *rappports qu'ils entretiennent les uns avec les autres dans le système*.
- il considère que le *langage*, en tant que faculté générale de s'exprimer au moyen de signes, se distingue de la *parole*, qui serait plutôt l'utilisation concrète de signes linguistiques particuliers (et qu'il n'étudiera pas plus avant).

- il distingue les dimensions *diachronique* (évolution au cours du temps) et *synchronique* (rapports entre les signes à une époque donnée) du langage. Les études historiques et comparatistes se sont focalisées sur la première de ces dimensions, lui entend privilégier la seconde (nous aussi par la suite).
- il distingue deux axes d'analyse d'un discours, en tant que suite de signes : l'axe *syntagmatique* est celui de la succession linéaire des unités qui constituent le discours (un syntagme est une suite d'unités adjacentes) ; l'axe *associatif ou paradigmatic* provient des liens que les signes présents dans le discours entretiennent avec d'autres signes non présents dans le discours mais en rapport avec eux dans le système. Suivant l'axe syntagmatique "un petit chat" est un syntagme dont le sens provient de la combinaison des signifiés de "petit" et "chat", mais ces sens eux-mêmes sont associés suivant l'axe paradigmatic avec d'autres ("petit" par opposition à "grand", etc.).

#### années 30-40 :

(1) le "cercle de Prague" prolonge les analyses de Saussure et promeut une "linguistique structurale". Ses membres les plus connus sont Roman Jakobson (1896-1982) et Nicolas Troubetzkoy. On leur doit notamment l'invention de la *phonologie* : étude des sons élémentaires (les *phonèmes*) qui jouent le rôle d'unités distinctives dans une langue donnée. C'est aussi à Jakobson qu'on doit d'avoir identifié six *fonctions* permises par le langage dans un contexte de communication :

- la fonction *expressive* permet au locuteur d'exprimer ses sentiments ;
- la fonction *conative* permet d'agir sur le destinataire (donner un ordre...) ;
- la fonction *référentielle* permet d'informer sur le monde extérieur : il faut bien reconnaître que les modèles informatiques ont souvent tendance à limiter le langage à cette fonction ;
- la fonction *phatique* permet juste de s'assurer du bon fonctionnement de la "ligne" de communication ("allo"...)
- la fonction *poétique* met l'accent sur la forme du message plus que sur son contenu informationnel ;
- la fonction *métalinguistique* permet de parler du langage grâce au langage (comme le fait ce document !)

#### Alan Turing (1912-1954) :

mathématicien anglais. En 1936, il propose le dispositif plus tard appelé "machine de Turing" qui donne une caractérisation mathématique précise à la notion d'algorithme. Cette proposition peut être considérée comme la date de naissance de l'informatique (i). En 1950, il publie un nouvel article fondamental dans lequel il décrit un test pour juger de la capacité des machines à penser : ce test, appelé depuis "test de Turing" est fondé sur un jeu de dialogue. En quelque sorte, il énonce qu'une machine peut être dite intelligente si elle est indiscernable d'un humain dans une situation de dialogue courant. Il prédit qu'en l'an 2000, des machines réussiront ce test (ça ne s'est pas vraiment réalisé). (il)

#### 1945 :

Von Neumann (1903-1957), mathématicien et physicien américain d'origine hongroise, définit dans un rapport le plan de construction des ordinateurs, tels qu'ils sont encore conçus de nos jours (des prototypes plus rudimentaires ont été construits avant). (i)

#### 1952 :

première conférence sur la traduction automatique, organisée au MIT (Massachusetts Institute of Technology) par Yehoshua Bar-Hillel (1915-1975). C'est l'époque de la guerre froide, la compétition russes/américains bat son plein. L'informatique, elle, n'en est qu'à ses balbutiements et les premiers programmes de traduction doivent se contenter de dictionnaires bilingues et de quelques règles de restructuration élémentaires. La légende propage qu'en partant de "The spirit is willing but the flesh is weak" (l'esprit est fort mais la chair est faible), après un aller-retour russe-anglais, on obtint alors "The vodka is strong but the meat is rotten" (la vodka est forte mais la

viande est pourrie)... Le terme "Intelligence artificielle" (IA), lui, est inventé lors d'une autre conférence, en **1956** (il).

**André Martinet (1908-1999) :**

(l) linguiste français, connu notamment pour avoir caractérisé les langues naturelles par la propriété de la *double articulation*. Toutes les langues humaines sont doublement articulées parce qu'elles combinent des éléments (discrets) à deux niveaux différents :

- la "première articulation" est celle qui permet la combinaison "d'unités douées chacune d'une forme vocale et d'un sens" qu'il appelle des "monèmes" (le terme n'est plus vraiment employé : nous utiliserons plutôt celui de "morphème" mais, en première approximation, disons que ce sont les "mots"). On peut dire que ce niveau est celui de la *syntaxe*.
- la "deuxième articulation" décrit comment chaque "monème" est lui-même décomposable en une succession d'unités phoniques élémentaires dépourvues de sens, les *phonèmes*.

Il semble bien que toutes les langues humaines, y compris les langues des signes utilisées dans les populations ayant une déficience auditive et/ou articuloire (même si, dans ce cas, les unités employées ne sont pas de nature phonique), soient doublement articulées, alors que ce n'est le cas d'aucun autre système de transmission d'information, notamment de ceux en usage dans les autres espèces animales (nous détaillons cette affirmation dans la section suivante).

**1966 :**

Deux programmes célèbres datent de cette époque : "Eliza" de Weizenbaum (simulation d'un dialogue avec un psychologue, voir par ex : [Eliza](#)) et "Student" de Bobrow (résolution de problèmes mathématiques simples). Les deux étaient fondés sur la recherche de "mots-clés" dans les données qu'on leur fournissait, mots-clés qui servaient à remplir les "trous" de formulaires définis *a priori*, sans prise en compte du contexte d'énonciation. Par exemple, dans un dialogue avec Eliza, dès que l'utilisateur mentionnait un lien de parenté ("father"/ "mother"/"brother"...), le programme enchaînait en demandant "Tell me about your [father/mother/brother...]". Bar-Hillel critique cette approche en citant l'exemple suivant, où le sens des mots dépend de leur contexte : "The pen is in the box" (le crayon est dans la boîte)/ "The box is in the pen" (la boîte est dans le parc). Cette même année, paraît le rapport ALPAC (Automatic Language Processing Advisory Committee), commandé 2 ans avant par l'Académie des sciences américaines. Le rapport est très critique vis-à-vis des recherches menées dans le domaine du traitement de la langue à cette époque, et conclut qu'elles mènent à une impasse. Suite au rapport, les financements publics se tarissent en Amérique (il).

**Noam Chomsky (né en 1928) :**

linguiste et activiste politique américain d'une productivité exceptionnelle, professeur de linguistique au MIT depuis 1961, qu'une enquête de 2005 désigne comme "le plus grand intellectuel vivant". A travers les différentes théories qu'il a contribué à élaborer, il a toujours argumenté en faveur de la *primauté de la syntaxe* sur tous les autres niveaux d'analyse du langage. Il a toujours adopté aussi une position innéiste et rationaliste : selon lui, les hommes disposent à la naissance d'un "organe du langage" de nature mentale. Dans cette perspective, apprendre une langue particulière revient à instancier une "grammaire universelle" innée. Il a élaboré de nouveaux concepts et promu plusieurs distinctions inédites :

- il distingue la *compétence* linguistique (connaissance des règles de fonctionnement d'une langue) de la *performance* (mise en oeuvre effective de ces règles en compréhension ou en production). On peut rapprocher cette distinction de celle entre *langage* et *parole* introduite par Saussure. C'est seulement la *compétence* qui sera l'objet de son attention, la *performance* relevant plutôt de la psychologie. La linguistique se doit en effet d'étudier le langage indépendamment de son substrat biologique (gènes, zones du cerveau, etc.) : on considère avoir affaire à un locuteur abstrait idéal.
- il considère que le but de toute théorie linguistique est l'explication des *jugements de grammaticalité* dont sont capables tous les locuteurs d'une langue (surtout quand c'est leur



langue maternelle). Pour lui, la *structure de surface* (syntaxe) d'un énoncé détermine sa *structure profonde* (les relations sémantiques qu'il exprime).

La chronologie des travaux de Chomsky a marqué l'histoire de la syntaxe des 50 dernières années :

- 1957 : publication de "Syntactic Structures" (structures syntaxiques), ouvrage fondateur. Dans les années 60, l'approche se raffine dans la théorie des "grammaires génératives et transformationnelles" qui deviendra ensuite dans les années 70 la "théorie standard" de la syntaxe.
- années 80 : l'approche "Principle and Parameters" (principes et paramètres) précise la nature de la "grammaire universelle" innée postulée par Chomsky : elle est constituée d'une liste de propriétés et de paramètres ne pouvant prendre qu'un nombre fini de valeurs. Ainsi, apprendre une langue particulière revient à acquérir son vocabulaire spécifique et à identifier la valeur des paramètres qu'elle instancie. Cette théorie sera reprise ensuite sous les termes "Government and Binding" (théorie du gouvernement et du liage, souvent abrégée en "GB").
- depuis 1995, Chomsky promeut un "programme minimaliste", qui est une reformulation de ses théories précédentes mais orientée par un principe d'économie.

Chomsky est aussi généralement bien connu des informaticiens via la *hiérarchie de Chomsky* qui caractérise des familles de langages de complexité croissante. Cette hiérarchie est issue de la mathématisation de la notion de "grammaire générative", qui remonte aux années 60. C'est le fondement de la *théorie des langages formels*, branche dans laquelle sont étudiées les propriétés des langages artificiels comme les langages de programmation informatiques. On évoquera cette hiérarchie au chapitre 5 dans ce document (il).

#### **1972 :**

l'informaticien Terry Winograd présente son programme intitulé SHRDLU (ce nom proviendrait de l'ordre décroissant de la fréquence des lettres en anglais : ETAOINSHRDLU...), permettant des interactions langagières avec un ordinateur sur un domaine restreint à un *monde de blocs*. Ce monde est constitué d'un nombre fini d'objets de forme géométrique simple (cubes, boules, cylindres, pyramides, etc.), disposés dans un environnement limité (l'équivalent d'une table). Les interactions se limitent à la possibilité de poser des questions sur l'état de ce monde simplifié ("Combien y a-t-il de cubes verts à droite de la boule rouge ?") et de donner des ordres permettant de le modifier ("Mettre le cylindre sur le cube bleu."). L'originalité de ce programme est qu'il ne se contente plus de mots-clés ou de "traitements de surface" rudimentaires : le monde étant parfaitement circonscrit, il pouvait être entièrement modélisé.

#### **années 70-80 :**

elles sont marquées en TALN par l'effervescence de la *sémantique formelle* pour représenter des connaissances et formaliser des raisonnements : nouvelles logiques (logique floue, logiques modales, etc.), "scripts" (Roger Schank), "frames" (Marvin Minsky), "réseaux sémantiques", "graphes conceptuels" (John Sowa) naissent à cette époque. La pragmatique, c'est-à-dire l'étude de l'utilisation du langage en contexte, dans des situations concrètes (déjeuner au restaurant ou réservation de billets de trains, par exemple), est aussi prise en compte dans ces modélisations. Les "systèmes experts", basés sur des modèles symboliques du même genre, constituent alors la vitrine de l'IA (il).

#### **depuis les années 90 :**

L'augmentation considérable de la capacité de stockage et de calcul des ordinateurs, ainsi que le développement exponentiel d'Internet, permettent l'émergence d'une *linguistique de corpus* fondée sur l'exploitation de textes au format numérique. Cette linguistique plus empirique, *fondée sur les données* plutôt que sur des modèles formels abstraits, fait un usage intensif de calculs numériques et statistiques. Cette évolution va de pair avec les progrès de "l'apprentissage automatique", branche de l'IA consacrée à l'écriture de programmes qui s'améliorent avec l'expérience, grâce à des exemples. L'idée est d'utiliser des corpus pour apprendre automatiquement à fixer les

paramètres de modèles (souvent statistiques) utilisables sur de nouvelles données. Cette démarche, prédominante dans la recherche actuelle, est abordée dans le chapitre 8 de ce document (il).

### 3 Les niveaux d'analyse du langage

La chronologie précédente a montré que ce n'est que progressivement que les sciences du langage ont précisé leur objet d'étude. Pour caractériser les *données* et les *traitements* pertinents de ce domaine, nous proposons le schéma de la figure 2.1.

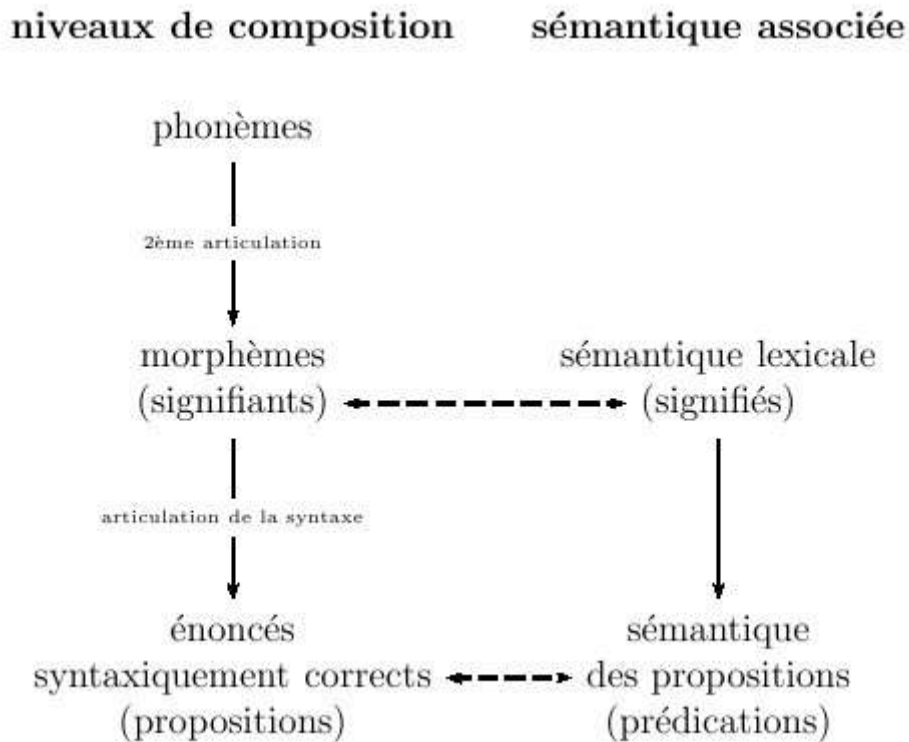


Figure 2.1 : hiérarchie des niveaux d'analyse des langues naturelles

Dans ce schéma, les principales unités d'analyse figurent de haut en bas, des plus simples aux plus complexes. Chacun des "noeuds" du schéma est constitué d'un ensemble de *données discrètes*. Les "flèches verticales descendantes" symbolisent des *règles de composition* opérant des traitements combinatoires sur les noeuds. Les "flèches bidirectionnelles horizontales en pointillés" traduisent, elles, des relations d'association qui ne sont pas bi-univoques (ou pas bijectives) entre données. La non univocité de ces flèches reflète les phénomènes d'*ambiguïté*, présents dans toutes les langues naturelles (sur lesquels nous reviendrons bien sûr par la suite).

Ce schéma permet surtout de mettre en évidence plusieurs des spécificités des langues naturelles évoquées précédemment :

- la dimension "verticale" du schéma est l'axe *syntagmatique*, tandis que sa dimension "horizontale" correspondrait plutôt à l'axe *paradigmatique*.

- la "double articulation" du langage se retrouve dans les deux niveaux de traitements combinatoires successifs qui occupent l'axe syntagmatique.
- ce schéma opère une claire distinction entre deux niveaux souvent confondus : celui de la *sémantique lexicale*, qui étudie le sens d'unités individuelles, et la *sémantique propositionnelle* qui étudie le sens d'énoncés complets, auxquels on peut attribuer une *valeur de vérité*.

Nous prétendons que ce qui caractérise les langues naturelles, c'est *l'ensemble des niveaux de description et relations présents dans ce schéma*. Précisons pourquoi.

Chacun des deux niveaux de l'axe syntagmatique traduit une combinatoire *ouverte* (au sens où la liste des éléments qu'elle produit est potentiellement infinie) d'*éléments discrets*. Ce dispositif original diffère fondamentalement de codages de type *analogique*. On cite souvent, pour illustrer la communication animale, le cas des abeilles, qui informent leurs congénères de l'emplacement d'une source de nourriture en modulant de façon *continue* l'amplitude et l'orientation de leur vol : leur danse fait un "huit", incliné suivant un angle qui indique la direction à suivre et dont la hauteur est proportionnelle à la distance à parcourir. C'est un exemple typique de codage *analogique* (mais tous les modes de communication animale ne sont pas analogiques). Comme le remarque le psycholinguiste canadien Stephen Pinker (né en 1954), les deux systèmes les plus sophistiqués de transmission d'information qui ont été sélectionnés par la nature, à savoir les langues naturelles et le code génétique, reposent tous les deux sur des unités *discrètes*. La nature a inventé le codage "numérique" bien avant les informaticiens. Sans doute est-ce dû à la fiabilité de la transmission ainsi permise...

La maîtrise d'associations arbitraires "signifiant/signifié" au niveau lexical semble accessible à certaines espèces animales (principalement des singes) à qui on a pu enseigner l'usage d'un répertoire non négligeable de symboles : gestes empruntés à une langue des signes ou dessins abstraits arbitraires. Mais aucune espèce autre que l'homme n'a développé cette capacité dans la nature, sans enseignement explicite. Et (même si ce point est encore discuté) aucune non plus n'a semblé être capable de maîtriser complètement un niveau de combinaison syntaxico-sémantique plus complexe : la simple juxtaposition de symboles ne suffit pas à faire un langage doublement articulé.

Le niveau de l'association entre "unités signifiantes" et "sémantique lexicale" est un jalon important de notre schéma. Il constitue une étape fondamentale de l'acquisition de leur langue maternelle par les enfants, aux alentours de leur première année de vie. Certains auteurs argumentent aussi qu'il était peut-être la base d'un "protolangage" que nos lointains ancêtres auraient inventé avant d'avoir recours aux langues naturelles proprement dites. Evidemment, ce genre d'hypothèses est difficile à valider mais elle a le mérite de mettre l'accent sur la complexité considérable des langues humaines, dont on a du mal à imaginer comment elles ont pu émerger "d'un seul coup" dans une espèce particulière.

Pourtant, le critère de la "double articulation" est probablement insuffisant pour distinguer à lui tout seul les langues humaines de langages d'un autre genre apparus récemment : les langages de programmation informatiques (qui, il est vrai, étaient encore peu connus et diffusés en dehors des spécialisés à l'époque de Martinet). On peut en effet argumenter que ces langages sont, eux aussi, doublement articulés :

- la première articulation est celle des règles à suivre impérativement pour écrire un programme syntaxiquement correct ;
- la deuxième articulation caractérise la construction des unités lexicales de ce programme (soit mots clés du langage, soit identifiants de variables ou noms de fonctions) à partir des unités distinctives élémentaires que sont les caractères alphanumériques autorisés (pendant écrit des phonèmes).

Il est possible aussi d'associer une "sémantique" aux programmes informatiques, aux deux niveaux évoqués par le schéma. Sa nature n'a pourtant rien à voir avec celle véhiculée par les langues naturelles : son domaine se réduit à l'arithmétique, son "monde" est celui du calcul opérationnel, et elle exclut toute

ambiguïté. Les langages de programmation radicalisent en quelque sorte les propriétés formelles des langues naturelles. Pour donner aux ordinateurs l'incroyable pouvoir expressif de ces langues, à savoir leur capacité à référer à ce qui leur est extérieur, à dire des choses sur le monde, il reste à réduire cette description du monde à un calcul. C'est paradoxalement leur caractère flou et *ambigu* qui rend finalement les langues naturelles plus aptes à réaliser ce programme.

Notre schéma permet ainsi de bien situer les unes par rapport aux autres les distinctions évoquées précédemment, qui se focalisent chacune sur une portion de sa structure. Il est toutefois loin d'être exhaustif. Un schéma plus complet devrait aussi faire figurer, à titre de niveau intermédiaire, la combinatoire propre à la *morphologie*, à laquelle nous consacrerons bien sûr un chapitre. Ne sont pas évoqués ici non plus les niveaux d'analyse qui vont au-delà des propositions (analyse des discours, des textes ou des dialogues, pragmatique...). Pourtant, les hommes se racontent en permanence des anecdotes et des histoires, et c'est au bout du compte ce *comportement narratif* qui caractérise le mieux l'espèce humaine. Mais la figure [2.1](#), et tout particulièrement le "rectangle" qu'elle fait apparaître à sa base, circonscrit en quelque sorte le périmètre de ce document.

Le principal intérêt pour nous de ce schéma, malgré son caractère réducteur, est qu'il met bien en évidence les données (les noeuds du schéma) et les traitements (les flèches) qui vont pouvoir donner lieu à une modélisation informatique. Il reste à voir comment traduire en code informatique les informations qui y figurent. La prise en compte des données discrètes de la colonne "niveaux de composition" ne pose pas de grosses difficultés, puisqu'il s'agit de coder des données discrètes par d'autres données discrètes. La traduction sous forme d'algorithmes des modes de combinaisons de ces données est un problème nettement plus intéressant. Dans ce domaine, théories linguistiques et modèles informatiques doivent collaborer. Quant au codage de la dimension "sémantique" du schéma, la plus cruciale et la plus problématique, elle est l'objet de nombreuses théories dont nous allons aussi essayer de rendre compte. Mais avant cela, voyons comment concevoir l'architecture "classique" d'un système complet de "compréhension du langage".

## 4 La chaîne de traitements "standard"

Pour comprendre comment les humains utilisent une langue naturelle, il ne suffit pas d'avoir identifié les différents niveaux de connaissances impliqués dans la compréhension de cette langue : il faut aussi savoir comment ces connaissances sont exploitées. Ce champ d'étude ne relève plus à proprement parler de la linguistique, mais de la *psychologie*, voire de la *psycholinguistique*. La figure [2.2](#) propose une chaîne de traitements qui a une certaine plausibilité. C'est avec ce genre de schémas que les psychologues cognitivistes tentent de modéliser le fonctionnement de l'esprit humain.

---

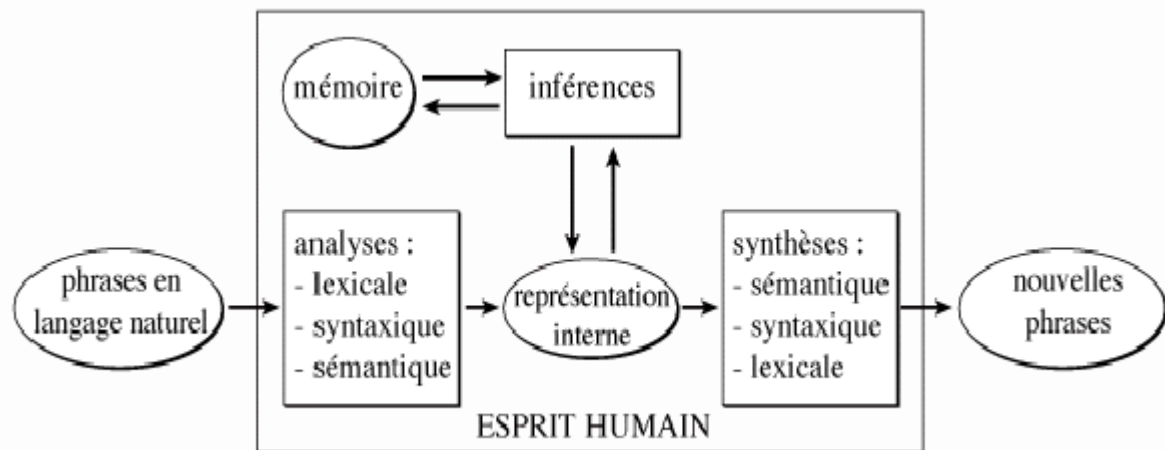


Figure 2.2 : chaîne de traitements classiques de compréhension du langage

Dans ce schéma, ici encore très simplificateur (il n'intègre pas, par exemple, la composante orale du langage), les données figurent dans des ovales, tandis que les traitements sont représentés dans des rectangles. Les deux principaux rectangles, intitulés respectivement "analyses" et "synthèses", correspondent aux deux tâches principales accomplies par les locuteurs d'une langue. Le fait de bien les séparer provient de l'observation de patients souffrant de lésions cérébrales qui affectent en particulier une de ces compétences et pas l'autre. Il n'est pas nécessaire pourtant de faire l'hypothèse que chacun des traitements évoqué dans ce schéma soit réalisé par une aire cérébrale spécifique. Il suffit de considérer qu'il met en évidence un enchaînement des *fonctions*, indépendamment de leur "implantation" dans un substrat biologique concret.

Suivant ce schéma, *comprendre un énoncé* revient donc à le transformer, via une "analyse", en une représentation interne, tandis que pour en *produire* ou en *générer* un, il faut traduire linguistiquement une telle représentation via une "synthèse". Chacune de ces tâches nécessite de prendre en compte l'ensemble des niveaux d'analyse identifiés précédemment, mais dans un ordre différent et en faisant chaque fois des hypothèses différentes sur ce qui est connu et ce qui doit être accompli.

Maintenant, pour construire un système artificiel complet avec lequel des humains pourraient interagir via une langue naturelle, une première approche possible consiste à tenter de reproduire une architecture de ce genre, en traduisant les "fonctions" en programmes. C'est en quelque sorte le projet du "traitement automatique du langage naturel" dans sa forme originelle. Il a donné lieu à de très nombreux travaux ces 50 dernières années, et nous allons prendre le temps dans les chapitres qui suivent de présenter ses principaux résultats. Nous verrons notamment que beaucoup des outils formels imaginés pour modéliser certains des traitements cités ici sont en fait exploitables à la fois en analyse et en synthèse (c'est le cas, par exemple, des automates et grammaires formelles). D'autres sont plus spécifiques.

Pourtant, le schéma de la figure 2.2 n'est instancié dans sa totalité dans aucun système informatique existant. C'est un cadre idéal théorique, très influencé par une conception *cognitivist* de l'esprit humain, où les "représentations internes" sont en général de nature symbolique. Pour une application particulière, on se limite la plupart du temps à implémenter une toute petite portion de cette architecture.

Mais il est aussi possible d'envisager une toute autre approche, qui ne se soucie pas de crédibilité psychologique, et se concentre sur l'efficacité pragmatique de ses programmes. Cette mutation est représentative de l'évolution qu'a connue l'"intelligence artificielle" ces dernières années. Après avoir longtemps tenté d'imiter le fonctionnement de l'esprit humain, les chercheurs du domaine essaient plutôt désormais d'exploiter au mieux les capacités de mémoire et de calculs de leurs machines. Des *modèles symboliques formels*, on est passé aux *modèles statistiques fondés sur l'analyse des données*. Nous aborderons cette mutation dans le dernier chapitre de ce document, centré sur l'ingénierie linguistique, telle qu'elle est exploitée en "fouille de textes".

Dans la suite de ce document, nous allons en quelque sorte passer en revue les niveaux d'analyse du schéma de la figure [2.1](#), en présentant chaque fois les théories linguistiques qui les décrivent et quelques-uns des modèles informatiques auxquels ils ont donné lieu. Les données et les traitements seront "parcourus" de haut en bas et de gauche à droite, en passant des unités les plus simples aux unités les plus complexes, sans cacher les difficultés propres à l'informatisation de chacun d'eux. Seul le dernier chapitre sera consacré aux applications qui utilisent certains des outils ainsi décrits.

## 5 Sites Web

Avant de commencer l'exploration des niveaux d'analyse du langage, faisons un détour par le "test de Turing". Ce test est un jeu au cours duquel un juge humain, en situation de dialogue médiatisé par une machine avec une entité distante, doit décider au bout d'un temps fixé à l'avance si son interlocuteur est un humain ou un programme. Son origine remonte à un article philosophique que Turing, l'inventeur de l'informatique, a fait publier en 1950. Selon lui, un programme qui "passerait ce test" -autrement dit qui ne serait pas identifié comme tel par le juge- devrait se voir reconnu les mêmes capacités que celles attribuées "naturellement" aux humains -en particulier celle de penser... Ce test fait donc des capacités linguistiques en situation de dialogue le critère principal de l'"intelligence".

Il existe à l'heure actuelle une compétition qui propose d'instancier le "test de Turing" dans un cadre contrôlé : elle s'appelle le "Loebner Price", du nom du généreux donateur qui offre un prix et une médaille au vainqueur. Les sites suivants offrent la possibilité d'une petite conversation à distance avec un programme ayant participé à cette compétition. A vous de voir si vous leur accorderiez le prix...

- [www-ai.jjs.si/eliza/eliza.html](http://www-ai.jjs.si/eliza/eliza.html)
- [www.ellaz.com/EllaASP8/Direct.aspx](http://www.ellaz.com/EllaASP8/Direct.aspx)
- [www.bearbot.co.uk/](http://www.bearbot.co.uk/)
- [www.pandorabots.com/pandora/talk?botid=f5d922d97e345aa1](http://www.pandorabots.com/pandora/talk?botid=f5d922d97e345aa1)

# *Formaliser le vivant : lois, théories, modèles ?*

Paris, Hermann, 2010

**Franck Varenne**

Université de Rouen & GEMASS (UMR 8598)

Peut-on formaliser le vivant ? Peut-on réduire une plante à une simple formule mathématique ? Goethe ne l'aurait pas admis. Pour beaucoup encore, cette question ne se pose même pas tant elle peut sembler provocante et contre-nature.

Dans une perspective à la fois historique et épistémologique, ce livre rend compte de travaux contemporains qui ont pourtant tous tenté de braver cet interdit. C'est en grande partie sur ce terrain, hautement problématique, que, dans les premières décennies du XXe siècle, on voit naître puis s'épanouir la pratique des modèles mathématiques appliquée aux sciences végétales. On voit en particulier que ces pratiques nouvelles de modélisation entrent en concurrence avec une tradition ancienne de théorisation mathématique des formes du vivant. C'est même devant les limites des essais théoriques récurrents que le tournant formel des modèles se confirme et permet des avancées incontestables.

À l'heure où toutes les sciences à objets complexes parlent beaucoup de modèles et moins de théories, est-ce le signe d'une victoire de la « modélisation » au détriment de la « théorie » ? Cette victoire est-elle définitive ? Cela a-t-il toujours un sens de les opposer ? Et qu'en est-il des « lois » ?

En proposant une analyse des travaux mais aussi des positions épistémologiques de certains scientifiques impliqués, en explicitant le sens de ce qui les rapproche, de ce qui les distingue ou les oppose, cet ouvrage montre que l'émergence, l'expansion puis la diversification des pratiques de modélisation formelle du vivant ont contribué, sur le terrain scientifique lui-même, à bousculer les rapports épistémologiques traditionnels entre théories, lois et modèles tels qu'ils nous ont été légués par la physique.

## SOMMAIRE DETAILLE

<b>Introduction générale.....</b>	<b>13</b>
<b>I. DES LOIS AUX MODELES.....</b>	<b>23</b>
Chapitre 1 – La mathématisation des formes du vivant, une curiosité.....	25
1. Morphologie et phyllotaxie géométrique.....	26
2. La géométrie spirale : une description précise mais sans explication ni application.....	28

Chapitre 2 – Une critique des mathématisations : les « hélices foliaires » de L. Plantefol.....	33
Chapitre 3 – Statistique et « loi mathématique hypothétique » chez R.A. Fisher (1921-1922).....	39
1. La signification de la mathématisation chez R. A. Fisher : condenser l'information.....	40
2. Infini hypothétique et modèle statistique : notion d'information et déracinement.....	45
3. Le rôle d'un infini hypothétique dans le modèle statistique.....	48
4. La discrimination des causes.....	53
5. Précision de l'indéterminisme : critique des formules mathématiques générales (1921).....	55
Chapitre 4 – La loi d'allométrie de J. Huxley et G. Teissier : de la mesure absolue à la mesure relative.....	63
1. Le problème des interprétations chimiques et .... métaboliques de la croissance.....	65
2. Sur les interprétations métaboliques : .... Bertalanffy (1932).....	71
3. Mathématiser l'élémentaire avant le complexe : .... J. Monod et G. Teissier (1935).....	74
4. Signification épistémologique du passage à .... l'allométrie : dialectique de la nature .... et déracinement.....	77
5. Prenant et Teissier : un physiologisme dialectique.	83
Bilan – Modèle, hasard et déracinement des formalismes.....	89
<b>II. RESISTANCE DES THEORIES AUX MODELES.....</b>	<b>93</b>
Chapitre 5 – La bio-mécanique de d'Arcy Thompson (1917-1942).....	97
Chapitre 6 – La bio-hydraulique de C. D. Murray (1926-1930).....	107
Chapitre 7 – La biophysique de N. Rashevsky (1931-1954).....	113
1. De la « biologie physique » de Lotka .... à la forme de la cellule.....	114



2. Forme et mécanisme de division de la cellule.....	116
3. Physicalisme unitaire et convergence .... avec le « positivisme logique » de Carnap.....	117
4. Morphogenèse des métazoaires et épistémologie .... des « principes formels ».....	120
5. Application à la forme des animaux .... puis des plantes.....	124

Chapitre 8 – La biophysique d'ingénieur de D. L. Cohn (1954).....	129
--	-----

Chapitre 9 – L'axiomatique du vivant de J.H Woodger (1937).....	135
--	-----

1. Philosophie et embryologie : l'épistémologie .... de Joseph Henry Woodger.....	135
2. La « méthode axiomatique » en biologie (1937).....	141
3. Un système axiomatique pour la biologie : .... le système (P, T, org, U,...).....	142
4. Impact sur l'embryologie : augmentation .... de la complexité sans recours au vitalisme.....	143

Bilan – Physicalisme et axiomatisme, deux stratégies de résistance aux modèles.....	151
--	-----

1. Statuts des divers formalismes .... avant l'ordinateur.....	153
2. Bilan général de la première époque, .... dite du déracinement des formalismes.....	158

### **III. NAISSANCE DES SIMULATIONS..... 163**

Chapitre 10 – Essor des modèles dans l'après-guerre..	167
---	-----

Chapitre 11 – La simulation comme calcul numérique : A. M. Turing (1952).....	171
--	-----

1. Le modèle chimico-mathématique.....	173
2. L'influence de l'embryologie chimique.....	175
3. La modélisation mathématique et le rôle .... du calculateur numérique selon Turing.....	178
4. Réception de l'article de Turing en embryologie...	181

Chapitre 12 – La simulation comme computation spatialisée : S. Ulam (1962).....	185
--	-----

1. Modélisation et genèse logique sans	
--	--

.... morphogenèse chez von Neumann.....	187
2. La réduction des mathématiques	
.... à une visualisation combinatoire chez Ulam.....	188
3. Quand le calculateur numérique est analogique,	
.... il simule.....	194
4. Des « systèmes de réaction binaire » au modèle	
.... de ramification végétale.....	198
 Chapitre 13 – La simulation comme génératrice de formes au hasard : M. Eden (1960).....	205
1. Lettres, mots, cellules et formes.....	205
2. Un stochasticisme biologique.....	211
 Bilan – Statut théorique des premières simulations.....	213

#### **IV. TOURNANT MATHÉMATISTE DES THÉORIES..... 219**

Chapitre 14 – La biotopologie du second Rashevsky (1954-1972).....	223
1. Fonction mathématique et fonction biologique :	
.... la « biotopologie » ensembliste.....	227
2. « Organisme primordial » et	
.... « propositions existentielles ».....	229
3. « Tranches » et « propriétés » du vivant :	
.... Woodger et le second Rashevsky.....	230
4. Une conséquence : l'oubli de la forme.....	233
 Chapitre 15 – La biologie relationnelle de Robert Rosen (1958).....	235
1. Re-spatialiser le formel : des « propriétés »	
aux « composants ».....	236
2. Application de la « théorie des catégories »	
à la représentation des systèmes biologiques.....	238
3. Érosion de la résistance aux modèles	
.... et reconnaissance des modèles mathématiques.....	243
 Chapitre 16 – Thermodynamique et topologie différentielle des formes.....	249
1. Arbres fluviaux et arbres botaniques (1945-1971)...	249
2. « Entropie généralisée » et phyllotaxie	

.... (1969-1973).....	253
3. Reconnaissance difficile de la dispersion	
.... des « modèles théoriques ».....	258
4. Une topologie de la morphogenèse en France :	
.... Thom et le modèle-paradigme (1968-1972).....	260
5. Structuration et institutionnalisation	
.... de la biologie théorique en France 1975-1981.....	266
 Bilan – Biophysique, biologie mathématique et phyllotaxie théorique devant l'ordinateur.....	 269

## **V. EXTENSION ET DIVERSIFICATION DES MODELES : LE CAS DE L'ECOLE FRANÇAISE DE MODELISATION..... 271**

Chapitre 17 – La rencontre entre biométrie et modèles de la biophysique.....	277
---	-----

1. Interdisciplinarité et indiscipline : le parcours	
.... de formation de J.-M. Legay (1947-1955).....	277
2. De Rashevsky à Legay : graphes	
.... et ramifications (1968).....	281
3. La biocybernétique et ce qu'en retient Legay :	
.... tout est lié (1967-1971).....	285
4. Le groupe « Méthodologie » de la DGRST :	
.... une rencontre avec l'écologie.....	286

Chapitre 18 – Le rôle spécifique de la philosophie française néo-marxiste.....	291
---	-----

1. L'accusation d'idéalisme.....	291
2. Informatique et Biosphère et la « Méthode	
.... des modèles ».....	296
3. Conséquence de cette épistémologie	
.... pour la simulation sur ordinateur.....	304
4. La ramification du gui : un modèle	
.... pour l'épistémologie des modèles.....	306

Bilan – Le modèle unique n'existe pas.....	313
--	-----

## **Conclusion – Multiplication des types et des fonctions des formalismes..... 317**

<b>Epilogue – Perspective pour la biologie théorique : des théories aux concepts ?.....</b>	<b>333</b>
---	------------

1. Une certaine continuité épistémologique.....	336
2. S'émanciper des horizons formels	
.... classiques.....	341
<b>Bibliographie.....</b>	<b>353</b>
<b>Index des matières.....</b>	<b>379</b>
<b>Index des auteurs.....</b>	<b>383</b>

**Un extrait de l'introduction générale est disponible sur le site :**

[Le choix des libraires](#)

**Cet ouvrage est disponible en librairie et sur tous les sites de vente en ligne, dont :**

[Site Hermann](#)   [Site Decitre](#)   [Site Amazon](#)   [Site FNAC](#)