

JERRY FODOR 1986

La modularité de l'esprit : essai sur la psychologie des facultés

Paris : Les éditions de Minuit

et

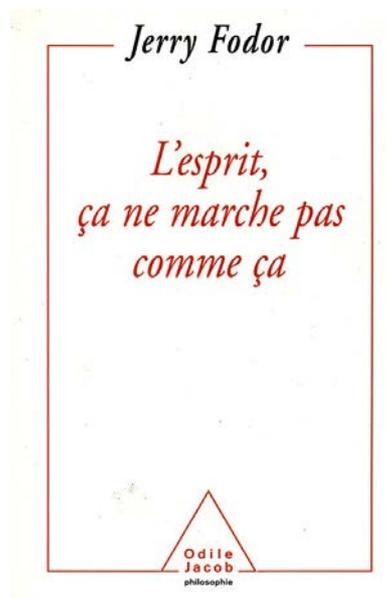
JERRY FODOR 2003

L'esprit, ça ne marche pas comme ça.

Paris : Odile Jacob (collection Philosophie)

2017, mars 2019

Les supports neurologiques possibles de l'intelligence artificielle



Une présentation pour le colloque de Martigny qui n'a pas été utilisée.



Les deux livres de Jerry Fodor *La modularité de l'esprit : essai sur la psychologie des facultés* (1983/1986 pour la tradition française) et *L'esprit, ça ne marche pas comme ça : portée et limites de la psychologie computationnelle* (2000/2003) permettent de poser la question, complexe, des origines de la démarche logiciste de Jean-Claude Gardin et de ses relations avec le positivisme logique, le cognitivisme, l'intelligence artificielle et empirisme classique. La situation est d'autant plus complexe que Gardin n'a jamais été très explicite sur les fondements épistémologiques de sa démarche qu'il considérerait avant tout comme une démarche pratique.

Le débat reste néanmoins d'actualité avec la mode de l'intelligence artificielle, une étiquette présentée à toutes les sauces par les médias qui, le plus souvent, ignorent de quoi il s'agit exactement. Ces questions ont touché le monde de l'archéologie. La Cité des Sciences a ainsi organisé fin septembre 2017 à Paris une série de conférences et de débats, dont une table ronde sur « *la nouvelle archéologie high tech* ». Parallèlement, le Forum économique rhodanien a consacré à Martigny le 15 septembre 2018 une rencontre dédiée à l'*intelligence artificielle* et a agendé dans ce cadre une séance consacrée à l'archéologie qui m'a donné l'occasion de réfléchir à ces questions.

Au moment où l'on affirme que les progrès de notre discipline dépendent avant tout de développements technologiques - drones, prospections à la télédétection laser de type Lidar, réalité virtuelle ou SIG - il serait peut-être temps de revenir aux fondamentaux de notre discipline. Les deux livres de Jerry Fodor soulèvent ainsi la question des sources épistémologiques du logicisme de Jean-Claude Gardin, d'où deux questions : quelles relations entre cette épistémologie pratique et la théorie computationnelle de l'esprit (TCE) de Jerry Fodor et quelle place pour l'intelligence artificielle ?

La théorie computationnelle de l'esprit (TCE) de Jerry A. Fodor

Pour aborder cette question il convient de se référer à deux livres successifs de Jerry A. Fodor qui ne présentent pas la même vision du fonctionnement du cerveau et qui soulèvent, de ce fait, certaines difficultés de présentation. Ce problème est d'autant plus difficile à résoudre que Fodor n'est pas explicite sur ce qui différencie son travail le plus récent de sa première thèse (tabl. 1, fig. 1).

Dans *La modularité de l'esprit : essai sur la psychologie des facultés* (1983/1986 pour la tradition française) Fodor oppose des structures modulaires horizontales en relation avec les sens à des structures verticales non modulaires correspondant aux mécanismes de la pensée.

Dans *L'esprit, ça ne marche pas comme ça : portée et limites de la psychologie computationnelle* (2000/2003 pour la traduction française) Fodor me paraît ajouter un étage intermédiaire qui assure la relation entre les sensations et la pensée globale, étage qui pourrait répondre plus strictement à la structure computationnelle proposée dans la perspective initiée par Alan Turing (fig. 1).

Classification	FODOR 1986 Niveau 1	FODOR 2003 (TCE) Niveau 2	FODOR 1986 / 2003 Niveau 3
fonctionnelle	Processus périphériques	Processus centraux Fixation de la croyance	Processus centraux Fixation de la croyance
Selon la « matière » traitée	Propre à un domaine	Propre à un domaine	Sans domaine propre
Selon le caractère computationnel	Cloisonné (local) Modulaire	Cloisonné (local) Modulaire	Quinéen et isotrope Général

Tableau 1. Position de Jerry Fodor quant à la structure cognitive du cerveau.

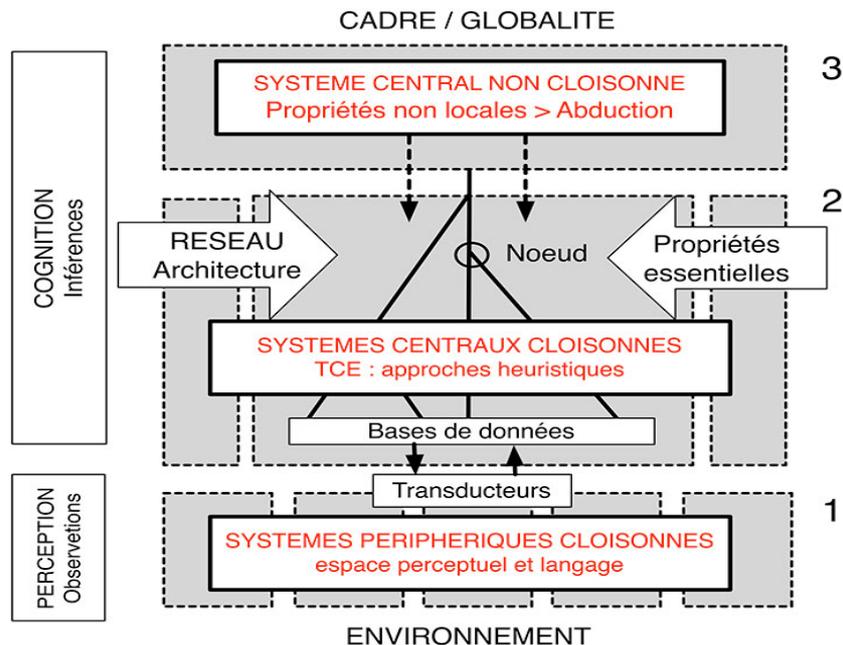


Fig. 1. Structure cognitive synthétique de l'activité globale du cerveau intégrant les données de Fodor de 1983 et 2000. Le niveau 2 des systèmes centraux cloisonnés correspond à la TCE *sensu stricto*. Schéma A. Gallay.

Niveau 1. Systèmes périphériques (Fodor 1986)

Les systèmes périphériques considérés comme « horizontaux » assurent la relation du cerveau à son environnement et regroupent les mécanismes de la perception et du langage. Ce sont des structures neuronales innées ayant des localisations précises dans le cerveau, connues des neurologues. La tâche de la perception est de représenter le monde de manière qu'il soit accessible à la pensée.

Les systèmes périphériques sont cloisonnés sur le plan de l'information.

Les systèmes périphériques ont pour fonction de fournir de l'information aux systèmes centraux. Une faculté horizontale est un système cognitif fonctionnellement identifiable qui opère sur des domaines de contenus différents.

La perception est insensible aux rétroactions en provenance de la pensée

Adopter une architecture computationnelle c'est affirmer qu'il existe une différence entre *perception* et *cognition*. La conséquence de cette architecture est que l'analyse

perceptuelle se trouve physiologiquement isolée de certains effets du système global. La perception de ce qui est imprévisible est effectuée par des mécanismes perceptuels opérant de bas en haut et globalement insensibles aux croyances et aux désirs de l'organisme.

Certains psychologues (et de nombreux théoriciens de l'intelligence artificielle) soutiennent que les processus périphériques opèrent de haut en bas, le codage perceptuel étant déterminé pour l'essentiel par les croyances et les attentes (conscientes ou inconscientes du sujet). Fodor avance au contraire que l'existence de certains effets de rétroaction sur les systèmes périphériques est incompatible avec leur modularité. Les calculs effectués par les systèmes périphériques ne peuvent librement mobiliser des anticipations perceptuelles ou des croyances de haut niveau. On a par exemple remarqué que beaucoup d'illusions perceptuelles persistent sachant pourtant que le percept est illusoire.

L'analyse perceptuelle de stimuli imprévisibles (dans le cas du langage et ailleurs) n'est possible que si :

1. les sorties des transducteurs sont insensibles aux croyances et aux attentes de l'organisme,
2. les systèmes périphériques sont capables de calculer une représentation en amont du stimulus utilisée par les transducteurs.

Les systèmes périphériques sont rapides parce qu'ils peuvent se permettre d'ignorer beaucoup de données.

Les systèmes périphériques sont propres à un domaine

Les systèmes périphériques regroupent six systèmes modulaires, un système pour chaque modalité sensorielle/perceptuelle traditionnelle (l'ouïe, la vue, le toucher, le goût, l'odorat), plus un système pour le langage. Ce regroupement, qui place le langage au même niveau que les sens, est inhabituel par rapport aux découpages traditionnels.

Les mécanismes computationnels des systèmes périphériques sont spécialisés parce qu'il existe des restrictions sur l'information qu'ils peuvent utiliser pour formuler des hypothèses. Pour la vision, de bons candidats seraient les mécanismes responsables de la perception de la couleur, de l'analyse de la forme et de l'analyse des relations dans l'espace tri-dimensionnel.

Chaque module correspond à un type particulier d'opérations. Les mécanismes perceptifs sont distincts de ceux du *langage ou de la mémoire*. La maîtrise du langage se décompose à son tour en sous-modules spécialisés : certains sont spécifiques à la grammaire, d'autres traitent des *informations sémantiques* contenant le sens des mots.

Pour Fodor, un module est donc :

1. spécifique à une opération précise,
2. son fonctionnement est autonome, rapide et donc inconscient,
3. il possède une localisation neuronale précise.

Les mécanismes cognitifs cloisonnés sont généralement dédiés au traitement de bases de données innées qui ne sont pas obligatoirement cloisonnées (par exemple à l'intégration d'informations innées avec des *entrées sensorielles*) et ce très tôt au cours de l'analyse perceptuelle). Les modules chomskien sont pas contre non cloisonnés. Un module est simplement un ensemble de connaissances propositionnelles innées. Les modules ainsi entendus n'engagent absolument à rien quant à l'architecture des processus mentaux.

Les opérations des systèmes périphériques sont obligatoires

La perception des choses ne peut être empêchée. Le monde extérieur s'impose à la perception indépendamment de toute volonté ou filtrage.

La sortie des systèmes périphériques est « superficielle »

Les philosophes des sciences se demandent depuis longtemps comment il convient de distinguer l'*observation* de l'*inférence* et les psychologues, la *perception* de la *cognition*. Il convient donc d'isoler un niveau de transducteurs à l'interface entre les processus perceptuels et les processus cognitifs, soit un niveau où la perception doit interagir quelque part avec les connaissances d'arrière-plan et les préférences de l'organisme.

Les systèmes centraux n'ont qu'un accès limité aux représentations calculées par les systèmes périphériques

Les processus cognitifs qui sous-tendent le comportement observable n'ont accès qu'au produit final des systèmes périphériques. Toutes les théories admettent que les systèmes périphériques calculent un certain nombre d'analyses intermédiaires des stimuli proximaux. L'important est que ces niveaux de représentation ne sont pas identiquement accessibles du point de vue du sujet. Le traitement perceptuel procède de bas vers le haut (chaque niveau de représentation est relié de façon plus abstraite que le niveau précédent à la sortie des transducteurs), mais l'accès aux représentations procède de haut vers le bas (plus on est loin des transducteurs, plus les représentations sont accessibles aux systèmes cognitifs centraux). L'inaccessibilité relative des niveaux inférieurs des processus périphériques résulte en partie des priorités qui régissent le transfert de représentations de la mémoire à court terme vers la mémoire à long terme. Un sujet ne perçoit pas de différence entre deux syllabes qui sont en fait très différentes d'un point de vue acoustique dans la mesure où les différences acoustiques phonologiques sont infra-phonétiques. Certains niveaux inférieurs de représentation semblent être totalement inaccessibles à la conscience.

Les systèmes périphériques sont associés à une architecture neuronale fixe

Il semble qu'une architecture neuronale caractéristique innée, « naturelle », soit associée à chaque système périphérique.

Les systèmes périphériques sont rapides

La vitesse des processus périphériques) contraste avec la relative lenteur des processus centraux typiques comme la résolution de problèmes. Ce n'est donc pas un hasard si ces processus psychologiques très rapides sont obligatoires.

Les systèmes périphériques présentent des défaillances caractéristiques

Les systèmes périphériques peuvent présenter des défaillances ce qui ne semble pas être le cas des systèmes centraux qui ne sont pas associés à une architecture neuronale spécifique.

Les systèmes périphériques présentent des caractères innés (cf Chomski)

On peut admettre en règle générale la vérité d'une lecture nativiste plus ou moins stricte de la cognition cloisonnée. Cette lecture ouvre la voie à des discussions sur les *processus évolutifs* à l'origine de ces structures mentales Si le nativisme de Chomski consiste avant tout en un nativisme relatif à la connaissance que manifestent les capacités cognitives, le rationalisme de la *Nouvelle synthèse* (Pinker 1997/2000 ; Plotkin 1997), consiste, lui, avant tout en un

nativisme relatif aux mécanismes computationnels qui exploitent une telle connaissance pour les besoins de la connaissance.

L'ontogenèse des systèmes périphériques suit un certain rythme, et une séquence d'étapes caractéristique.

Le développement neurologique des systèmes périphériques peut être étudié sur le plan ontologique tant au niveau du développement paléontologique de l'humanité qu'au niveau de l'histoire de l'individu.

Niveau 2. Systèmes centraux : modularité locale des systèmes heuristiques (Fodor 2003)

Dans son livre de 1986 Fodor ouvre une petite fenêtre sur une étape qui serait intermédiaire entre les systèmes périphériques perceptuels et le système global de la pensée, question qui sera développée dans son livre de 2003 sous la dénomination de *théorie computationnelle de l'esprit* (TCE) :

« Il n'existe néanmoins en pratique des procédures heuristiques pour déterminer l'ensemble des effets que l'acceptation d'une hypothèse peut avoir sur le reste des croyances. Puisqu'il s'agit, par hypothèse, de procédures heuristiques, elles sont probablement locales – c'est-à-dire qu'elles ne sont pas sensibles à la totalité du système de croyances auxquelles elles s'appliquent. » (Fodor 1986, p. 149)

La théorie computationnelle de l'esprit (TCE) s'oppose au connexionnisme, une forme d'empirisme selon laquelle l'association demeure la relation causale de base entre les entités mentales particulières, alors que les pensées n'ont ni structures ni constituants. Contrairement à la théorie de la *Nouvelle synthèse* proposée par des auteurs comme Stephen Pinker (1997/2000) et Henri Plotkin (1997), Fodor n'admet pas que la totalité de la cognition soit massivement modulaire. Seuls les mécanismes répondant à une modularité locale composent, selon Fodor, la TCE (théorie computationnelle de l'esprit).

Les processus mentaux concernés sont conscients

La TCE concerne des états mentaux *intentionnels*, dont les croyances, les désirs et autres constituent de paradigmes et des états mentaux *conscients*, dont les sensations, les sentiments et autres représentent des paradigmes.

Le calcul selon Alan Turing est une bonne approximation des mécanismes impliqués

La grande découverte de Turing (1950) a été de montrer qu'il est possible de concevoir des machines pour évaluer n'importe quelle inférence formellement valide *indépendamment des significations*. Les machines sont en effet très peu douées pour découvrir ce que signifient les choses et ne le sont guère plus pour découvrir ce qui se passe dans le monde.

Le calcul se présente sous une forme syntaxique

L'idée centrale de la théorie computationnelle de la TCE est que les processus intentionnels sont *des opérations syntaxiques* s'effectuant sur des représentations mentales, soit sur des symboles. Il est possible de dissocier l'analyse des opérations effectuées des supports matériels (cerveau ou machine) qui le permettent. Sa structure locale peut encoder non seulement des *relations grammaticales* entre les phrases, mais également des *relations inférentielles*.

Les propriétés syntaxiques sont des propriétés « essentielles »

La syntaxe d'une représentation est l'une de ses propriétés essentielles (cf essence et non pas nécessité) et ne change donc pas lorsque la représentation se transporte d'un contexte à un autre. Le rôle cognitif d'une pensée est pour une part déterminée de manière plausible par des

propriétés essentielles (plus précisément syntaxiques) de la représentation mentale correspondante. Moins le rôle inférentiel d'une représentation mentale est déterminé par sa syntaxe interne, et plus elle dépend du cadre global, moins la représentation est « transportable » d'un système de croyance à un autre.

Le calcul prend une forme logique de type modus ponens (si p alors q)

Le raisonnement peut être réduit à des propositions abstraites de type « si p alors q » (équivalent à la notation de l'argumentation logiciste « si $\{P_i\}$ alors $\{P_i + 1\}$ »), soit à une opération en *modus ponens*. Étant donné une dérivation comprenant des formules de forme A et $A \rightarrow B$, le détachement de B s'effectue automatiquement par un processus architectural (en particulier, il ne requiert pas de prémisses ou de dérivations supplémentaires) si A et B sont des expressions primitives. La forme logique et les relations causales qui peuvent valoir entre ses constituants sont des paramètres *indépendants d'une attitude propositionnelle propre à un contexte*.

Les unités de traitement constituent des modules concernant des domaines d'information restreints

Le psychisme humain traite les informations sous forme de modules spécialisés propres à certaines questions choisies.

Spécificité et cloisonnement caractérise les modules activés

Chaque module est supposé avoir accès à son entrée du moment, et à sa base de données spécifique, et à rien d'autre. Il y a *cloisonnement architectural ou impénétrabilité du module*. Cloisonnement est le nom que l'on doit donner à cette restriction.

L'architecture des modules se présente sous forme d'un réseau

La TCE reconnaît deux types d'explications :

- Les *explications computationnelles sensu stricto* présentent des dérivations – des suites causales de représentations mentales – dont la dernière ligne est en général une spécification du comportement à expliquer.
- Les *explications architecturales* répondent à des questions portant sur la façon dont l'esprit, dans une telle dérivation, passe d'une ligne à l'autre.

Cette structure diffère des machines de Turing ou de von Neumann qui ne font pas la distinction entre programme et mémoire.

Les réseaux présentent une série de nœuds assurant la jonction des dérivations

Le réseau peut se définir à travers l'ensemble des nœuds qui le constitue. La position du nœud dans le réseau termine à quel type il appartient. Les types de nœuds et leurs positions dans le réseau définissent l'architecture du réseau.

Les dérivations s'appuient sur des bases de données

Les processus architecturaux de base, tout comme les processus computationnels de base, sont locaux et répondent uniquement à l'identité et à l'agencement des *représentations primitives*. Ces représentations primitives correspondent soit à *des bases de données innées*, soit à *des informations provenant de l'espace sensoriel*. La question est de savoir comment, dans leur domaine, les représentations se voient assigner les propriétés auxquelles elles sont effectivement sensibles.

Les calculs préservent la vérité des données

Les calculs définis par la syntaxe des représentations mentales préservent la vérité de façon fiable.

Les calculs aboutissent à des propriétés normatives

Les calculs et les architectures qui leur sont liées présentent des propriétés normatives.

Niveau 3. Systèmes centraux non cloisonnés de fixation des croyances (Fodor 1986 et 2003)

La notion de globalité est nécessaire à la compréhension du fonctionnement de l'esprit. Il doit exister des systèmes psychologiques non dénominationnels (c'est-à-dire n'ayant pas de domaine propre) dont l'une des fonctions est d'exploiter l'information fournie par les systèmes des deux niveaux inférieurs. Conformément à la tradition on les appellera des systèmes centraux.

Les systèmes centraux de fixation des croyances sont non-cloisonnés, leur fonctionnement est largement inconnu

Les systèmes centraux supérieurs sont pour l'essentiel non cloisonnés et c'est pour cela qu'il n'est pas plausible qu'ils soient modulaires. Plus des capacités cognitives sont, en un sens pré-théorique, supérieures et intelligentes, plus elles posséderont des propriétés globales. Plus un processus cognitif est global, c'est-à-dire isotrope, moins on sait de choses à son sujet. Sur les raisonnements très globaux, comme le raisonnement par analogie, on ne sait rien du tout.

Les systèmes centraux de fixation des croyances sont isotropes et quinéens

La confirmation en science est isotrope et quinéenne.

1. Isotrope : toute vérité empirique (ou bien sûre déductive) préalablement établie peut être pertinente pour la confirmation d'une hypothèse scientifique.

Le raisonnement par analogie est la plus pure expression de l'isotropie. Par définition les systèmes cloisonnés ne raisonnent pas par analogie. Lorsque nous disons qu'un processus est intelligent, nous faisons justement allusion à ces propriétés globales.

1. Quinéenne : si la confirmation scientifique est quinéenne c'est parce que le degré de confirmation attribué à une hypothèse quelconque est sensible aux propriétés de l'ensemble du système de croyances.

L'un des dogmes de l'empirisme (Quine 1961/1953) était qu'il devait exister des relations sémantiques entre les énoncés théoriques et des énoncés d'observation. Plus une mesure de confirmation est isotrope, plus elle utilise des données d'origines variées pour prendre des décisions. De ce point de vue, la confirmation scientifique est un cas limite de l'isotropie ; c'est un modèle de la fixation non modulaire de la croyance. Les systèmes isotropes et quinéens ne sont pas cloisonnés, et, s'ils ne sont pas cloisonnés, ils ne sont pas modulaires. En fait, puisque tout ceci est une question de degré, il faudrait mieux dire que plus un système est quinéen et isotrope, moins il est modulaire. En bref, si l'opération des systèmes centraux est isotrope et quinéen, alors le caractère computationnel de ces systèmes devrait être différent de celui des facultés verticales propre au deuxième étage.

Cadre /globalité sont les deux caractéristiques de la pensée supérieure

Selon Fodor et contrairement à ce qu'avance le *Nouvelle synthèse*, on peut admettre qu'une bonne partie de la cognition n'est pas modulaire et ne possède pas les caractéristiques du niveau 2.

On peut admettre que chaque mécanisme computationnel modulaire présuppose des mécanismes computationnels moins modulaires que lui en sorte qu'il y a un sens dans lequel l'idée d'une architecture massivement (totalement) modulaire est contradictoire. Les fonctionnements de ces mécanismes nous restent inconnus.

Le *problème du cadre* est le nom que l'on donne à la question de savoir comment réconcilier une *notion locale de calcul mental* avec le *holisme manifeste de l'inférence rationnelle*. Le problème du cadre est pour une bonne part responsable de ce qui rend la cognition si difficile à comprendre. Il est peu vraisemblable que la *fixation de la croyance* soit réalisée par des calculs sur des informations limitées et locales. Une illustration de cette difficulté est fournie par ce que l'on appelle le *frame problem* en intelligence artificielle. Il s'agit du problème qui consiste à délimiter le cadre (*frame*) de concepts qu'il faudra éventuellement modifier au fur et à mesure qu'un certain type d'information devient disponible. En effet, la machine (robot) doit tenir compte des effets des calculs effectués pour déterminer sa conduite et ses attentes ultérieures, d'où la question suivante : quel est, en général, l'ajustement optimal entre mes croyances et mes expériences ? C'est précisément ce type de question qu'est censée répondre une *théorie de la confirmation* et la confirmation n'est pas une relation que l'on peut reconstruire en faisant appel à des propriétés locales des hypothèses ou aux données pertinentes pour leur évaluation. La M(TCE) admet en principe que les inférences abductives sont des calculs, autrement dit, que ces inférences sont intégralement dirigées syntaxiquement.

Au contraire la E(TCE) répond au principe E selon lequel seules *les propriétés essentielles d'une représentation mentale peuvent déterminer son rôle causal* dans la vie mentale. Il existe donc des processus mentaux – architecturaux ou computationnels ou les deux – qui répondent à des propriétés irréductiblement non locales des systèmes de croyances.

E(TCE) implique donc :

1. qu'il existe des *propriétés essentielles* qui ne varient pas au cours de la cognition,
2. que ces propriétés permettent *l'abduction* soit des raisonnements hypothético-déductif qui rendent la causalité du raisonnement possible (principe E).

La globalité est difficile sinon impossible à modéliser

S'il est difficile de modéliser l'impact de considérations globales dans la *résolution* d'un problème il est en général difficile aussi de modéliser l'impact de ces considérations globales quand il s'agit de *décider comment* résoudre un problème. La question de savoir *quelle quantité de contexte épistémique* la rationalité d'une inférence est sensible ne peut pas recevoir de réponse nette et bien délimitée.

L'abduction est le mode de réflexion propre aux systèmes centraux non cloisonnés

L'abduction ne peut s'expliquer par des processus computationnels, même si ceux-ci s'organisent en mettant en relation les modules cognitifs de bas niveau.

L'abduction ne peut être décrite selon les règles de la logique pure. De plus, elle outrepassé les lois de la modularité. En gros, l'abduction démontre la *démarche en partie « holiste »* de la pensée. Il s'agit d'un processus permettant d'expliquer un phénomène ou une observation à partir de certains faits, événement ou lois. Une abduction réalisable exige en pratique, qu'on ne consulte pas plus qu'un petit sous-ensemble des croyances d'arrière-plan pertinentes.

Logicisme et théorie computationnelle de l'esprit : les trois sources du logicisme de Jean-Claude Gardin

Le positivisme logique et l'empirisme classique constituent, avec le cognitivisme de Fodor les trois sources possibles du logicisme de Jean Claude Gardin (fig. 15).

Le positivisme logique de l'École de Vienne.

Pour ce courant les connaissances reflètent la réalité et nous en livrent une image conforme, et l'observation directe des faits est donc la plus parfaite des connaissances. La science ne comporterait que des processus inductifs.

Selon David Hume (1711-1776) c'est à partir de nos sens et des *sense data* et à travers leur généralisation que l'on peut construire une science (Hume 1982/1748). L'empirisme hérité de Hume a été profondément renouvelé par l'attention portée au rôle du langage dans la connaissance. Selon Ludwig Wittgenstein (1988/1921) le sens des propositions complexes est déterminé par le sens des propositions élémentaires (cf P0) dont elles sont composées.

Quelques années plus tard Hans Hahn, Otto Neurath et Rudolf Carnap fondent le *Cercle de Vienne* (manifeste de 1929) et se donnent pour tâche d'édifier une véritable conception scientifique du monde (Carnap 1932). Il s'agit de réduire les énoncés scientifiques et leurs concepts aux énoncés et aux concepts *du plus bas degré*. C'est par l'analyse du langage que l'on pourra établir s'il y a correspondance entre les énoncés théoriques et les énoncés *du plus bas degré*. Au terme d'un long débat, Carnap et Neurath sont pourtant parvenus à la conclusion que cela était un leurre. Ils n'abandonnèrent pas pour cela la conviction que les *sense data* sont à la source de toute connaissance. Selon Franck (2015) l'empirisme/positivisme logique a pourtant échoué à garantir la conformité de nos pensées aux observations.

Le positivisme logique est le seul courant épistémologique auquel Jean Claude Gardin (1991/1987) se réfère explicitement en citant les noms de Carnap et Wittgenstein, mais c'est pour se concentrer sur des auteurs de ce courant qui ont été plus particulièrement intéressés par la question de la représentation des connaissances tels que Charles Morris (1938) et Charles Peirce pour la sémiologie des sciences et Stephen Toulmin (1958) pour la logique des champs.

L'empirisme classique

Pour ce courant l'observation ne nous livre qu'une connaissance tronquée du monde, il faut développer d'autres moyens pour le connaître véritablement. L'interprétation nous livre la forme des phénomènes. Il s'agit de ce que nous appelons aujourd'hui la structure. Multiplier les expériences nous permet d'observer les multiples façons différentes dont un phénomène se comporte, observations qui nous aide à découvrir progressivement ce sans quoi il n'émergerait pas.

L'induction est le nom donné par Francis Bacon à l'entièreté du processus expérimental orienté sur la recherche de la forme d'une propriété naturelle. La discussion savante est pourtant contrainte de recourir à des hypothèses. Elle va ensuite chercher à valider les hypothèses avancées, ou à les invalider au moyen d'inférences déductives. La démarche baconienne, qui est aux fondements des sciences modernes, réunit donc à la fois des processus empirico-inductifs et des processus hypothético-déductifs.

Charles Pierce a précisé cette question en distinguant trois processus :

L'induction montre que quelque chose *est effectivement*

La déduction prouve que quelque chose *doit être*

L'abduction ou démarche hypothético déductive sensu stricto, suggère que *quelque chose pourrait être*.

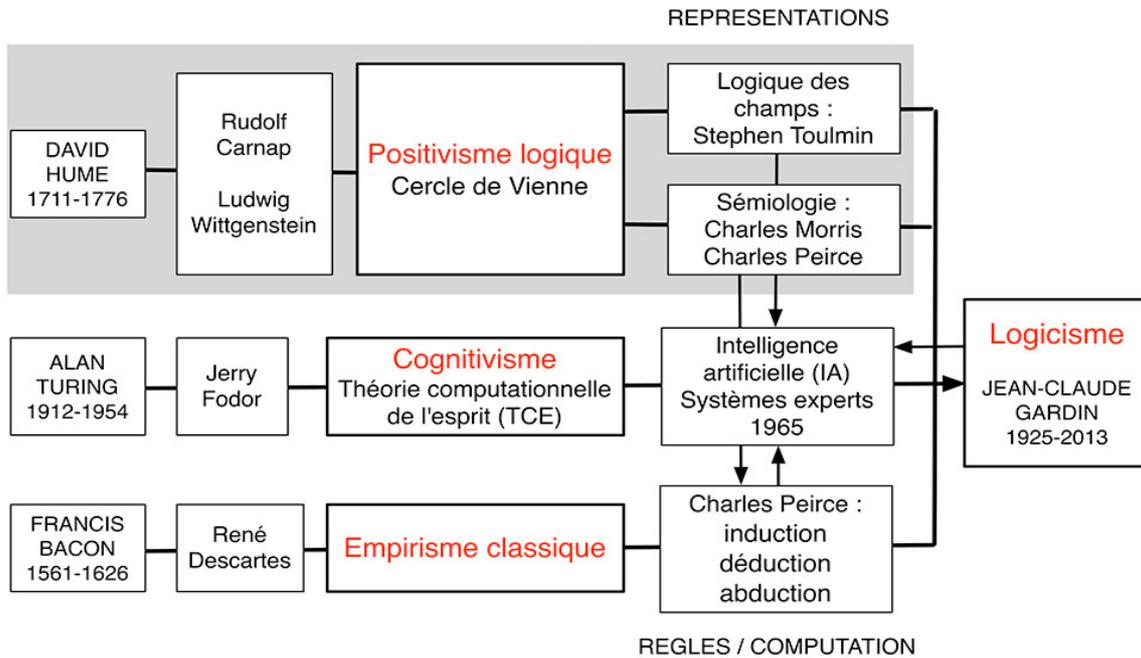


Fig. 2. Sources multiples du logicisme de Jean Claude Gardin qui n'est explicite que sur les systèmes sémiologiques étudiés par les tenants du positivisme logique, Charles Morris, Charles Peirce et Stephen Toumlin et sur les relations avec l'intelligence artificielle et les systèmes experts. Schéma A. Gallay.

Le logicisme et le cognitivisme face à la théorie computationnelle de l'esprit

Quelques rappels sur les composantes du logicisme nous permettront de mieux situer le statut de cette démarche considérée par Gardin comme une épistémologie pratique. Cette dernière concerne les fondements méthodologiques des constructions et non l'utilité empirique des produits (Gallay 2007) (fig. 3).

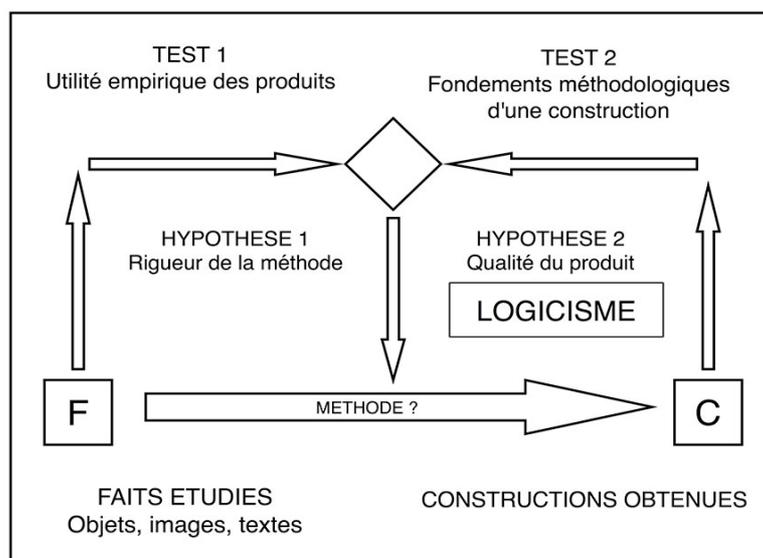


Fig. 3. Place du logicisme dans le cycle de la recherche. Complété de Gardin 1974, fig. 5.

Ambiguïté de l'analogie computationnelle

La mise en correspondance du fonctionnement « naturel » du cerveau et des procédures techniques destinées à asseoir un raisonnement mieux formé et plus efficace pose un certain nombre de problèmes, notamment sur le statut des modèles utilisés pour rendre compte du fonctionnement du cerveau. On notera que les analogies se développent dans deux directions :

- des procédures techniques vers un modèle de fonctionnement du cerveau dans la TCE de Fodor inspirée des machines de Turing,
- du fonctionnement du cerveau vers les procédures techniques de l'intelligence artificielle et les systèmes experts.

Dans cette perspective l'analogie qu'on peut proposer entre logicisme et TCE reste de convergence puisque Gardin ne se réfère pas explicitement à la TCE mais aux contraintes techniques entraînées par l'utilisation des ordinateurs pour édifier les bases de son logicisme et évacue toute référence à la structure naturelle du cerveau (Gardin 1989a, 1991-1, 1991-2). (fig. 4). Les filiations pourraient être les suivantes :

Fodor : machine (Turing) > cerveau (TCE)

Feigenbaum *et al.* : cerveau > intelligence artificielle et systèmes experts

Gardin : machine > logicisme > intelligence artificielle et systèmes experts

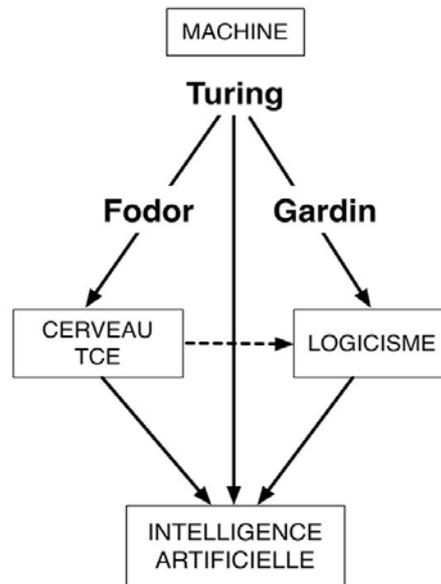


Fig. 4. Impact de la machine sur la connaissance du fonctionnement du cerveau et l'élaboration du logicisme.

L'importance des représentations

L'intérêt de Gardin pour la sémiologie se manifeste très tôt dans sa carrière. Il ne s'agit pas de la néo-sémiologie moderne, mais de l'étude des langages de la démarche scientifique dans la perspective du positivisme logique.

La perspective de Gardin s'écarte en effet résolument des débordements de la néo-sémiologie (Roland Barthes, Paul Ricoeur, Umberto Eco) qui considère toute manifestation humaine et même naturelle comme des signes selon la formule {sémiologie > monde empirique}. On ne

peut en effet concevoir l'émergence d'une nouvelle science unifiée qui prétendrait se surajouter ou même se substituer aux diverses disciplines existantes, tant dans le domaine des sciences humaines que dans celui des sciences de la Nature. Cette position rejoint parfaitement celle d'Alain Testart développée dans son livre *Pour les sciences sociales, essai d'épistémologie 1991*.

Nous pouvons distinguer trois types de discours :

- le langage naturel (LN), correspondant au langage discursif de tous les jours,
- les langages documentaires (LD), représentations métalinguistiques des objets (ou textes) en vue de leur identification dans une recherche d'informations,
- les langages scientifiques (LS), langages « contrôlés », subordonnés à des objectifs délimités et restreints, enrichis de termes techniques appartenant spécifiquement au domaine de recherche.

L'approche sémiologique concerne donc essentiellement les langages documentaires (LD), et les problèmes de formalisation de l'information dans la perspective d'une utilisation par les machines.

Structure de l'argumentation logiciste

Architecture en réseau de l'argumentation en modus ponens

Le cycle de la recherche peut se diviser en quatre étapes successives (Gardin 2001):

- les *compilations* (constructions compilatoires Cc), réunion des données de base spécifiques ou comparatives, pour servir la construction envisagée,
- les *ordinations* (constructions typologiques Ct) ou mise en ordre des données à travers des typologies, soit des mises en correspondance de caractéristiques intrinsèques (physiques, géométriques ou sémiologiques) et extrinsèques (de lieu, de temps, de fonction),
- les *interprétations* (constructions explicatives Ce) : résultats de la construction concernant la vie des hommes d'autrefois,
- les *pronostics* permettant l'administration de la preuve à travers un retour au niveau de faits empiriques nouveaux.

La démarche présente de nombreuses analogies avec les étapes 2 et 3 de la TCE de Fodor :

- L'architecture de l'argumentation se présente sous la forme d'un réseau avec des nœuds.
- la progression s'effectue en *modus ponens* avec des transitions de type si « p alors q ».
- les transitions si « p alors q » correspondent à des opérations de rétrodiction qui font appel à des référentiels externes plus ou moins maîtrisés qui annonce déjà l'étape 3 de la TCE (fig. 5 et 6).

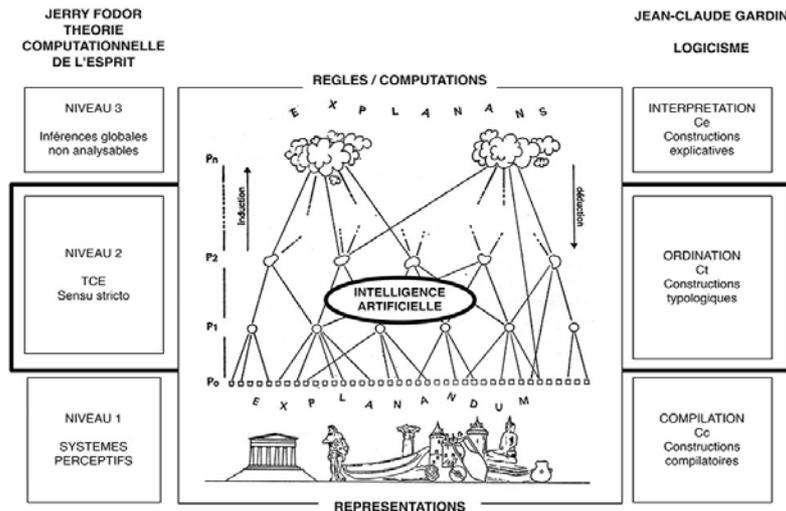


Fig. 5. Schématisation d'une construction logiciste et relations avec la TCE de Jerry Fodor. Les nuages de la partie supérieure de la figure illustre parfaitement les difficultés de compréhension des mécanismes cognitifs du niveau 3 de la TCE. Gardin 1979, fig.23 et 1991-1, fig. 3.

- les dernières étapes des P_n font intervenir un savoir extrinsèque mobilisé par abduction selon les mécanismes de l'étape 3 de la TCE (Gallay 2011).

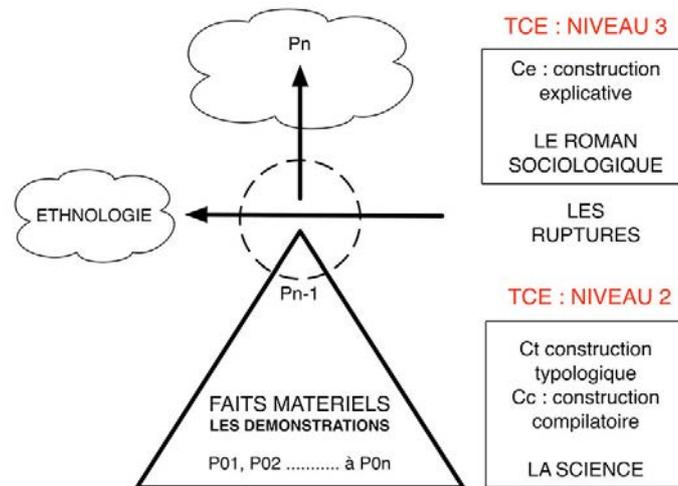


Fig. 6. L'interprétation des faits archéologiques dans la pratique ordinaire. Le triangle inférieur représente une construction archéologique fondée sur une série d'observations de base correspondant aux propositions P0. (TCE, niveau 2). A partir d'un certain seuil, qui correspond généralement à des explications « fonctionnelles » (politiques, sociales, religieuses) issues de l'ethnologie l'explication « décroche » par rapport à une explication fondée scientifiquement pour passer dans un domaine plus ou moins littéraire relevant souvent de la pensée vulgaire (TCE, niveau 3). Gallay 2011, Fig. 6.2.

Caractère local des règles d'inférences

Les considérations précédentes montrent que les opérations « si {P_i} alors {P_i + 1} », formule à laquelle on peut réduire les raisonnements, traduisent les pratiques discursives de nos constructions, mais ne sont en aucune manière l'expression de règles de raisonnements que l'on pourrait appliquer de façon aveugle et automatique dans des univers de discours quelconques

(Gardin, Lagrange *et al.*, 1981, rééd.1987). On retrouve ici le caractère local et cloisonné des transitions de l'étape 2 de la TCE (Gardin 1989b).

La rétrodiction, qui correspond à un transfert d'attribut, à l'origine purement local donc de niveau 2, ouvre néanmoins des perspectives se rapprochant du niveau 3 de la TCE dans la mesure où :

1. un transfert d'attribut originel peut s'appuyer sur d'autres transferts qui, à la limite, peuvent former des chaînes d'inférences sans limites (Gallay 2011, fig. 3.2) et
2. un transfert n'est possible qu'après élimination d'un référentiel Ox correspondant aux contextes au sein desquels le transfert n'est pas possible (Gallay 2011, fig. 3.3).

Argumentation logiciste et systèmes experts

Dès les années 80 Gardin constate la convergence formelle existant entre la manière dont on rationalise la recherche d'explications sous forme d'enchaînements de propositions répondant à la formule « si {Pi} alors {Pi + 1} » (ou « si {Pi} alors {Pj} ») et les systèmes experts qui voient le jour aux États-Unis dès la fin des années 70 (Gardin, 1987 1, 2 et 3; Gardin, Lagrange *et al.*, 1981). Plusieurs tentatives de formalisation du raisonnement sous forme de systèmes experts sont entreprises. Ces expériences ponctuelles s'écartent résolument des ambitions unificatrices des recherches en intelligence artificielle des années 60 pour s'intéresser aux raisonnements propres à chaque recherche, dans les domaines les plus divers (Gardin, 1989 -1 ; Gardin, Guillaume *et al.*, 1987 ; Francfort, Lagrange *et al.*, 1989).

Elles mettent en évidence les pratiques éminemment discursives des archéologues et le caractère "local" des démonstrations. Elles montrent par contre que le cumul des connaissances propres à plusieurs chercheurs permet d'enrichir considérablement l'approche des thèmes particuliers à chaque domaine de recherche (Lagrange, Renaud 1987).

Les procédures de réécriture du logicisme présentent de nombreuses analogies avec celles des systèmes experts et justifie de considérer ces derniers comme un outil permettant de prolonger et d'affiner la démarche (fig. 6). Nous avons en effet les concordances suivantes :

Analyse logiciste :	SI {Pi}		ALORS {Pj}
Système expert :	SI p	ET p > q	ALORS q
Base de faits	Règle		Base de faits enrichie

La conjonction des deux formes de savoir n'en caractérise pas moins la connaissance scientifique dans le sens de l'empirisme classique, qu'il s'agisse de centrales nucléaires ou de taille expérimentale du silex (Gardin, 1987-4, 2000) (fig. 7).

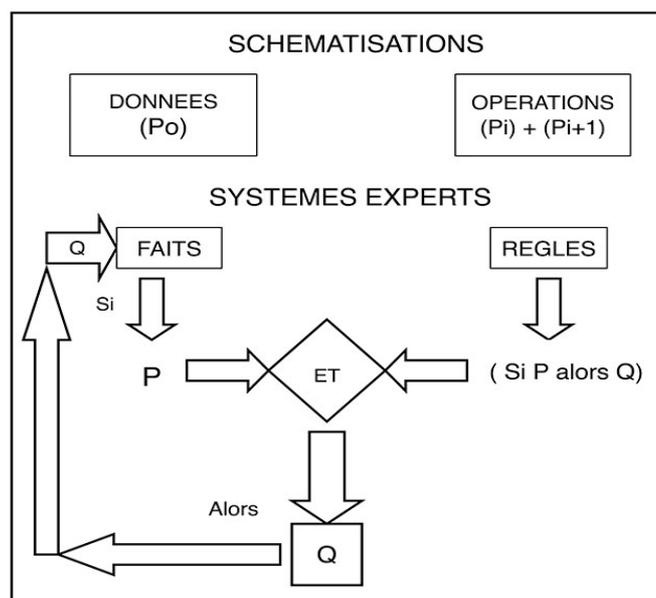


Fig. 7. Comparaison entre la démarche logiciste des schématisations et la structure d'un système expert. Les schématisations logicistes permettent de passer des données (P_0) à des interprétations (P_n) par des opérations de cumul de type « si $\{P_i\}$ alors $\{P_{i+1}\}$ ». Les systèmes experts pratiquent de même en appliquant aux faits, puis aux propositions dérivées de ces faits, des règles de nature identique « Si P alors Q ». Gardin, 1989-1, fig. 2).

Empirisme et déductivisme : importance de l'empirisme classique

Jean Claude Gardin a toujours insisté sur la totale complémentarité de l'induction et de la déduction ainsi que sur l'importance des processus abductifs. Cette situation est également propre aux systèmes experts avec les possibilités de chaînage avant ou arrière des inférences.

On se situe donc ici dans la perspective de *l'empirisme logique*.

Un bilan

L'analyse que l'on peut mener des sources épistémiques du logicisme de Jean-Claude Gardin montre que cette épistémologie pratique tire son origine de tous les grands courants de la réflexion épistémologique des XIX^e et XX^e s., positivisme logique, cognitivisme et empirisme logique (tabl. 2)

Gardin ne s'est jamais réellement préoccupé de réfléchir aux sources de ses théories, à part quelques allusions à la sémiologie des langages scientifiques qui s'est développée dans le sillage du positivisme logique (Gardin 1991) et des références plus appuyées et beaucoup plus nombreuses à l'intelligence artificielle et aux systèmes experts, deux domaines qui ne s'inscrivent pas directement dans la démarche de Fodor.

Dans cette perspective, il est étonnant, et pour moi inexplicable, que Gardin n'ait jamais proposé la moindre allusion aux travaux de Jerry Fodor et à la TCE. Peut-être considérait-il que les modèles neurologiques lui étaient inutiles et qu'il était même dangereux de s'y référer.

L'analyse proposée des thèses de Fodor montre pourtant qu'il est possible de comparer démarche neurologique et logicisme, la TCE présentant de nombreux points communs avec l'approche de Gardin.

LOGICISME	CORRELATS
Sémiologie	Positivismes logiques
Argumentation en réseau	TCE étapes 2 et 3
Caractère local des inférences (cloisonnement)	TCE étape 2
Systèmes experts	TCE étape 2
Empirisme et déductivisme	Empirisme classique

Tableau 2. Origines épistémiques des composantes de l'approche logiciste.

Genève, le 1^{er} septembre 2017

Références

CARNAP R. 1932. Die physikalische Sprache als Universalsprache der Wissenschaft. *Erkenntnis* 2, p. 432-465.

FODOR J. 1986. *La modularité de l'esprit : essai sur la psychologie des facultés*. Les éditions de minuit. Trad. de *The modularity of mind*. Cambridge Mass. : Massachusetts Institut of Technology 1983.

FODOR J. 2003. *L'esprit, ça ne marche pas comme ça*. Paris : Odile Jacob (philosophie). Trad. de *The mind doesn't work that way*. Cambridge Mass : Massachusetts Institut of Technology 2000.

FRANCFORT H.-P, LAGRANGE M.-S., RENAUD M. 1989. *Palamède : application des systèmes experts à l'archéologie de civilisations urbaines protohistoriques*. Paris : Éditions du CNRS (LISH/UPR 315, document de travail 9).

GALLAY A. 2007. 25 ans de logicisme : quel bilan ? In : *Congrès du centenaire : un siècle de construction du discours scientifique en préhistoire* (26^{ème} congrès préhistorique de France, Avignon 21-25 septembre 2004), p. 23-36.

GARDIN J.-C. 1974. *Les analyses de discours*. Neuchâtel : Delachaux et Niestlé (Zethos).

GARDIN J.-C. 1979. *Une archéologie théorique*. Paris : Hachette.

GARDIN J.-C., LAGRANGE M.S., MARTIN J.-M., MOLINO J., NATALI-SMIT J. 1981, rééd. 1987. *La logique du plausible : essai d'épistémologie pratique en sciences humaines*. Paris : Maison des sciences de l'homme.

GALLAY A. 2011. *Pour une ethnoarchéologie théorique*. Paris : Errance.

GARDIN J.-C. 1974. *Les analyses de discours*. Neuchâtel : Delachaux et Niestlé (Zethos).

GARDIN J.-C. 1987-1. La logique naturelle ou autre dans les constructions de sciences humaines. *Revue européenne des sciences sociales (Genève)* 25, 77, p. 179-195.

GARDIN J.-C. 1987-2. Les rapports troubles de la sémiologie avec la linguistique. *Cahiers Ferdinand de Saussure*, 41, p. 65-74.

GARDIN J.-C. 1987-3. Questions d'épistémologie pratique dans les perspectives de l'intelligence artificielle. *Bulletin de la société française de philosophie*, 81, 3, p. 7-112.

- GARDIN J.-C. 1987-4. *Systèmes experts et publications savantes*. Londres : British library.
- GARDIN J.-C. 1991 (1987). Questions d'épistémologie pratique dans les perspectives de l'intelligence artificielle. In : *le calcul et la raison : essais sur la formalisation du discours savant*. Paris : Ecole des hautes études en sciences sociales (Recherches d'histoire et de sciences sociales, 46), p. 59-89. Extrait du *Bulletin de la Société française de philosophie*, 81,3, 1987, p. 69-112.
- GARDIN J.-C. 1989a. Artificial intelligence and the future of semiotics : an archaeological perspective. *Semiotica*, 77, 1-3, p. 5-26.
- GARDIN J.-C. 1989b. The rôle of « local knowledge » in archaeological interprétation. In : Shennan S. J. (éd). *Archaeological approaches to cultural identity*. Springer, One world archaeology 10, p. 110-122.
- GARDIN J.-C. 1991 (1987). Questions d'épistémologie pratique dans les perspectives de l'intelligence artificielle. In : *le calcul et la raison : essais sur la formalisation du discours savant*. Paris : École des hautes études en sciences sociales (Recherches d'histoire et de sciences sociales, 46), p. 59-89. Extrait du *Bulletin de la Société française de philosophie*, 81,3, 1987, p. 69-112.
- GARDIN J.-C. 1991a. *Le calcul et la raison : essais de formalisation du discours savant*. Paris : École des hautes études en sciences sociales (Recherches d'histoire et de sciences sociales 46).
- GARDIN J.-C. 1991b. The impact of computer-based techniques on research in archaeology. In: Katzen M. (éd). *Scholarship and technology in the humanities*. (Proceedings of a conference held at Elvetham Hall, 9th-12th may 1990). London : British library research, Bowker Saur, p. 95-110.
- GARDIN J.-C. 2000. Approches sémiotiques du raisonnement en archéologie : une contribution au problème du « pont ». In : Perron P., Sbrocchi L. G., Colilli, P. Danesi, M. (éds). *Semiotics as a bridge between the humanities and the sciences*. New York, Ottawa, Toronto : Legas, p. 27-48.
- GARDIN J.-C. 2001. Entre modèle et récit : les flottements de la troisième voie. In : Grenier J.-Y, Grignon C., Menger, P.-M. (éds). *Le modèle et le récit*. Paris : Maison des sciences de l'homme, p. 457-487.
- GARDIN J.-C, GUILLAUME O., HERMAN Q, HESNARD A., LAGRANGE M-S., RENAUD M. ZADORA-RIO E. 1987. *Systèmes experts et sciences humaines : le cas de l'archéologie*. Paris : Eyrolles.
- HUME, D. 1982 (1748). *Enquête sur l'entendement humain*. Trad. et édition établie par D. Deleule. Paris : Nathan.
- MORRIS, C. 1938. Foundations of the theory of signs. *International encyclopedia of unified science I, 2*. Chicago : University of Chicago press.
- LAGRANGE M.-S. RENAUD M. 1987. Cas n°6 : Superikon, essai de cumul de six expertises en iconographie In : Gardin, J.-C., Guillaume O. Herman P.O. et al. (éds). *Systèmes experts et sciences humaines : le cas de l'archéologie*. Paris : Eyrolles, p. 191-229.
- PINKER S. 2000. *Comment fonctionne l'esprit*. Paris : Odile Jacob 2000. Trad. de *How the mind works*. New York : Norton 1997.
- PLOTKIN H. 1997. *Evolution in Mind*. London : Alan Lane.
- QUINE W. 1953. Les deux dogmes de l'empirisme. In : Jacob P. (éd.). *De Vienne à Cambridge : l'héritage du positivisme logique*. Gallimard (Tel), p. 93-123. Trad. de Two dogma of empirism. In : Quine W., Willard van Oman (éds) 1961. *From a logical point of view*. Cambridge, Mass : Harward university press.

- TESTART A. 1991. *Pour les sciences sociales : essai d'épistémologie*. Paris : Christian Bourgois.
- TOULMIN N. 1958. *The use of argument*. Cambridge : the University press.
- TURING A. M. 1950. Computing machinery and intelligence. *Mind* 49, p. 433-460.
- WITTGENSTEIN L. 1988/1921. *Tractatus logico-philosophicus*. Paris : Gallimard.