

# L'intelligence artificielle peut-elle se passer du représentationalisme ?

**Francis Rousseaux**

LIP6, Université Paris 6  
francis.rousseau@poleia.lip6.fr

Les organismes spécialisés attendent des Systèmes d'Information et de Communication (SIC) une aide à la gestion des crises à caractère géopolitique, mais ces systèmes, propres à assister la décision programmée, s'avèrent inadaptés aux situations vécues comme critiques par ceux qui les mettent en œuvre.

Ceci est dû au caractère expérimental de leur conception, qui hérite trop brutalement des représentations mythiques des notions de crise et de décision.

Cette critique n'épargne pas l'intelligence artificielle (IA), aujourd'hui mobilisée dans des programmes de recherche sur les SIC visant à modéliser des simulateurs capables d'assister efficacement les organisations humaines.

Cela tient au fait que l'approche IA est confrontée à deux difficultés fondamentales.

D'une part, les agents artificiels engagés dans une coopération avec leurs interacteurs humains devraient s'inscrire jusque dans la sphère intentionnelle des actes d'interprétation, qui constituent l'activité irréductiblement humaine de dépassement des crises.

D'autre part, des signifiants artificiels devraient réaliser une sphère sémiotique effective qui simule efficacement l'entour d'un interprète humain impliqué en conscience dans un acte d'interprétation.

Un certain nombre de préjugés tenaces retardent la mobilisation concertée des chercheurs en IA face à ces difficultés identifiées.

Le premier de ces préjugés consiste à considérer qu'un acte d'interprétation est réductible à un jugement portant sur un corpus de faits plus ou moins avérés et conduisant à une certaine décision. En autorisant l'évacuation de la question de l'intentionnalité des motivations humaines, cette attitude laisse à penser qu'un homme est valablement approché par un agent rationnel.

Cette assomption a deux conséquences radicalement dirimantes.

La première conséquence concerne directement l'interprète, et aboutit à son exclusion implicite du système dans lequel il est réputé être immergé. Aussitôt l'être humain engagé dans une crise qu'il éprouvera comme telle, cette ambiguïté profonde se manifesterait explicitement, car le propre du vécu de crise est de concerner la personne.

La seconde conséquence concerne le concepteur de SIC, qui envisage la modélisation du système complexe multiagent sous l'angle d'une coopération rationnelle, en s'efforçant de doter les agents artificiels des connaissances appropriées à leur rôle présumé. Ces artefacts réactifs n'auront bien entendu aucune intention artificielle, mais s'ils sont ingénieusement réalisés, ils s'en trouveront en quelque sorte investis dialogiquement par leurs interacteurs humains. On dira alors qu'ils satisfont au test de Turing.

Reste que la coopération multiagent se réduira à un échange de données symboliques, bientôt débordé par l'illusion que son accroissement quantitatif est susceptible de se cristalliser en raffinement qualitatif. Cette naïveté conduit inmanquablement à un dévalement technologique et à une surenchère autoréférente, incompatible avec l'entreprise de construction du sens que constitue le dépassement d'une crise.

La juxtaposition de ces deux conséquences est pour le moins paradoxale, qui nous donne à imaginer un système confrontant des personnes dépossédées de leur humanité et des artefacts investis d'esprit.

Le deuxième préjugé consiste à défendre que l'entour d'une personne, où si l'on préfère son environnement perçu et vécu en conscience, est constitué des objets de la réalité, représentés comme tels.

Dans ce cadre épistémique, la simulation d'un entour est chose relativement aisée en théorie, puisqu'elle autorise l'élection empirique de sortes de clones sémiotiques, eux-mêmes objets naturels ou artificiels de la réalité, sur le critère qu'ils se trouvent être opportunément représentés de façon équivalente, voire plus performante, dans l'entour de la personne.

Cette attitude, caractérisée comme représentationnaliste, est lourde de conséquences, à vrai dire rédhitoires lorsqu'on vise une reconception du domaine de recherche SIC dans la perspective de l'aide au dépassement des crises.

La première conséquence se déploie au détriment d'une vision herméneutique de l'entour, et aboutit à la négation de l'acte d'apréntation. La structuration de l'entour est pensée comme un mécanisme nominal et décomposable par l'analyse, quand la sphère sémiotique est projetée sans autre forme de procès sur le monde physique.

Cette réduction radicale de la conscience à un processus de traitement de symboles signifiants en correspondance biunivoque avec des signifiés, en renonçant à la sémiotique triadique de Peirce, entretient la croyance illusoire en un contenu de sens immanent aux signifiants artificiels, en dehors de tout processus interprétatif.

Ces signifiants artificiels sont ainsi présumés porteurs de sens intrinsèque, exprimable aux formats habituels de l'acquisition des connaissances. Le pas est franchi qui autorise de combiner par calcul le sens des signifiants. Cela peut certes constituer une heuristique productive pour mobiliser l'investissement de sens interprétatif des hommes en présence effective de ces signifiants, comme l'a montré Bachimont. Reste qu'en situation de crise collective, quand la négociation constructive du sens est essentielle, la confusion des genres engendre typiquement la crise dans la crise.

La seconde conséquence du préjugé représentationnel réside dans le primat implicite de la perception sur l'action qu'il charrie invariablement, sans doute poursuivi par l'archétype de "la représentation comme image" que nous a légué Hume. Cette propension incorrigible se traduit par un privilège exorbitant accordé à la saillance du perçu sur la prégnance de l'action visée, pour prescrire un signifiant artificiel. Elle a également à voir avec le schéma linéaire de la décision, qui ne reconnaît l'action qu'au travers sa manifestation seconde dans la sphère physique, niant son aptitude intentionnelle à conformer l'entour dans l'immédiateté première.

Cette posture institue un malentendu chronique entre l'action et la perception, et empêche leur réconciliation dans la figure pragmatique du signe dynamiquement expérimenté. Inscrite dans la tradition, elle courbe et replie le signe vers son engramme hylétique et l'éloigne de son évocation noématique. C'est ainsi qu'en situation de crise transinstitutionnelle, les responsables risquent de s'enfermer dans de vaines tentatives de fusion, agrégation, ou corrélation de données factuelles, s'enfonçant ainsi dans leur dérégulation.

La juxtaposition de ces deux conséquences conduit là encore à une aporie retentissante, puisque des signifiants artificiels sont effectivement investis de sens interprétatif par des êtres humains, qui par ailleurs considèrent l'acte d'investissement de sens comme une simple représentation.

Qu'on songe enfin, pour clore l'exposé des difficultés auxquelles est confrontée l'approche IA des SIC, que les agents artificiels abstraits conçus pour une coopération multiagent "au niveau des connaissances" doivent in fine se réaliser "au niveau des symboles" dans des systèmes sémiotiques artificiels capables de supporter des interprétations constructives. On prendra ainsi la mesure de la complexité et de la diversité des enjeux conceptuels.

Cependant, il est possible de ressaisir ces difficultés de façon synthétique, dans une acception phénoménologique de la saisie, c'est-à-dire en visant l'origine essentielle de ces difficultés.

Qu'il s'agisse de la décision réduite à un processus linéaire, rationnel et déterministe comme Sfez le décrit, de la communication réduite à un échange de symboles entre agents rationnels, de la représentation comme réduisant le sens à un rapport biunivoque entre le signifiant et le signifié, ou plaquant le signifiant sur le signifié en privilégiant la perception à l'action, toutes ces thèses se corroborent en marche et font système. Toutes ont la même origine dans "la thèse du monde", celle-là même dont Husserl préconisait la suspension dans une césure époquale, considérant que seule sa "mise entre parenthèses" pouvait fonder la quête des essences se donnant par intuition originaire.

Suivons donc la méthode phénoménologique qui vise à saisir, par un retour aux données immédiates de la conscience, les structures transcendantales de celle-ci et les essences des êtres.

Si on baptise "forme consciente" un vécu éprouvé dans l'immédiateté, on conçoit la conscience comme le processus qui pulse le flux des formes conscientes. Pour rendre compte de la sensation immédiate de flux, on peut s'appuyer sur les propositions de Husserl en terme de rétention et protension dans la conscience intime du temps. Une forme consciente est toujours déjà-hors de la conscience comme immédiateté, et ressaisissable par elle (mais non en tant que telle) dans l'immédiateté, comme prégnante sur "l'immédiateté à venir".

La conscience fait ainsi système clos et finalisé. On peut construire l'hypothèse que les formes conscientes émergent comme telles dans l'immédiateté, sous l'influence d'attracteurs de formes transcendants à la conscience, à partir d'un substrat de formes virtuellement conscientes qui est en retour transformé par l'acte d'émergence. On peut encore poser ce processus comme définitivement non décomposable par l'analyse, en s'inspirant des travaux de Bergson sur ce sujet.

Le flux des formes conscientes dans l'immédiateté se déploie paradoxalement au travers d'une médiatisation, et est ressenti comme conscience de quelque chose. Les formes émergentes sont vécues comme aprésentant quelque chose de transcendant avec une visée intentionnelle.

C'est ainsi qu'un fait de conscience a comme corrélat une forme dans la sphère sémiotique, témoin d'une activité et d'un souci éprouvé, vécu et expérimenté.

Ce corrélat est exprimable en avant-plan d'un arrière-plan de formes sémiotiques, le sens de cette forme étant immanent à la situation d'aprésentation et définitivement non immanent à la forme. Il est seulement exprimable par approximation dans son arrière-plan formel.

Le sens d'une forme sémiotique s'exprime par interprétation d'autres formes sémiotiques, renvoyant à un arrière plan complexe non infiniment précisable, conformément au cadre théorique proposé par Cardon. Non seulement le sens global d'une forme ne saurait être décomposable analytiquement en sens locaux, mais toute précision de sens local se fait au détriment de l'interprétation du caractère de localité, c'est-à-dire de la compréhension de la contribution du sens local au sens global.

Ultimement, une forme sémiotique est matériellement engrammée dans la sphère physique, les modes d'engramme étant sans doute éminemment variables. Mais en aucun cas la sphère physique n'est autre chose pour une forme sémiotique que son substrat matériel ultime.

On conviendra d'appeler arrière-monde la sphère physique, indiquant que cette sphère est définitivement hors du monde perçu et vécu en tant qu'il est interprété, tout en étant le substrat et l'engramme de ce monde.

Le monde du vécu est le monde de la conscience immédiate, qui engendre comme corrélat la sphère sémiotique, la sphère physique supportant et engrammant cette sphère. A la suite de Rastier, convenons d'appeler "entour" d'une conscience cette sphère sémiotique, augmentée de son "différentiel toujours immédiat" que constitue la sphère de l'aprésentation.

Dans ce nouveau cadre théorique, il devient possible de repenser à nouveaux frais une approche cohérente de l'IA en matière de SIC. C'est ainsi qu'Alain Cardon travaille actuellement à la modélisation d'une véritable intentionnalité artificielle.

Dans ce cadre, il est également possible de réinterpréter la "thèse du monde" et le mode de constitution de l'intersubjectivité qu'elle induit. En dernière analyse, un "objet du monde réel" est une forme sémiotique qu'on interprète comme étant supporté par la sphère physique, probablement à cause de son caractère spatio-temporel particulier. "Le monde réel" est lui-même une forme sémiotique souvent interprétée comme coïncidant avec la sphère physique, ce qui est une bizarrerie historique difficile à élucider.

Enfin, ce cadre permet de considérer le phénomène du représentationalisme depuis un lieu clairement différencié.

Le représentationalisme est avant tout une croyance, qui émerge et prend place dans la sphère épistémique de l'entour de celui qui y adhère. Comme telle, le phénomène de son émergence d'une part, et son être de croyance d'autre part, sont interprétables dans d'autres cadres épistémiques, surtout si ces cadres laissent place méthodologique à une démarche herméneutique, au sens de Ricœur.

Comme émergence dans la sphère des croyances, on pourrait penser que le représentationalisme ne modifie en rien le substrat à partir duquel il émerge. Autrement dit, cette croyance n'aurait aucun impact sur les phénomènes qu'elle interprète comme croyance.

Ceci serait d'ailleurs entièrement vrai dans la vision représentationaliste tautologique. Si le représentationalisme était une méprise, il ne s'agirait que d'une mauvaise représentation, qui ne changerait rien de la réalité du monde représenté !

Ceci est moins vrai si on place le représentationalisme comme croyance interprétable comme telle à partir d'autres croyances. Si l'émergence du représentationalisme ne change rien en soi aux phénomènes qu'il conceptualise, il reste que la présence, dans la région épistémique d'un entour subjectif, d'une forme prégnante de type "représentationalisme", courbe en retour l'intentionnalité de cet entour. Cela se traduit en particulier par l'émergence de projets, de démarche, d'idées et de programmes de recherche emprunts de représentationalisme.

L'intelligence artificielle doit contribuer à l'effort de dépassement du représentationalisme. Cette discipline est bien placée pour mettre en évidence les graves apories découlant de cette croyance lorsqu'elle opère sur le monde. L'IA, dans son propre champ opératoire, est souvent en prise avec la "thèse du monde".

Le domaine applicatif des systèmes d'information et de communication invite à la mobilisation des chercheurs en ce sens.

## Bibliographie

- BACHIMONT Bruno BACHIMONT , "Herméneutique matérielle et Artefacture : des machines qui pensent aux machines qui donnent à penser", thèse de doctorat de l'Ecole Polytechnique, 1996
- BARBARAS Renaud BARBARAS , "La perception, essai sur le sensible", Optiques Philosophie, Edts Hatier, 1994
- BERGSON Henri BERGSON , "Matière et Mémoire", PUF, 1968
- CARDON Alain CARDON, "La complexité des systèmes d'expression du sens", rapport de recherche LAFORIA, 1996

- CARDON Alain CARDON, "Un modèle de couche communicationnelle d'un SIC prenant en compte les représentations mentales des acteurs", rapport interne LAFORIA, 1997
- DAVIDSON Donald DAVIDSON, "Actions et événements", PUF Epiméthée, 1993
- DEWEY John DEWEY, "Logique, la théorie de l'enquête", PUF, 1993
- HEIDEGGER Martin HEIDEGGER , "Chemins qui ne mènent nulle part", TEL Gallimard, 1962
- HEIDEGGER Martin HEIDEGGER, "Etre et temps", Vrin, 1990
- HUSSERL Edmund HUSSERL , "Idées directrices pour une phénoménologie", TEL Gallimard, 1950
- HUSSERL Edmund HUSSERL, "Méditations cartésiennes", Vrin, 1992
- LEIBNITZ Gottfried-Wilhelm LEIBNITZ , "La monadologie", Grasset, 1990
- MEUNIER Jean-Guy MEUNIER , "La théorie cognitive : son impact sur le traitement de l'information textuelle", dans "Penser l'esprit : des sciences de la cognition à une philosophie cognitive", PUG, 1996
- NEWELL Alan NEWELL, "The Knowledge Level", Journal of Artificial Intelligence 18, 1982
- PEIRCE Charles PEIRCE, "Textes fondamentaux de sémiotique", Méridiens Klincksiek, 1987
- PERROT Jean-François PERROT, "Des objets aux connaissances", Journée "Méthodes objets & Intelligence Artificielle, Frontières, Ponts et Synergies", Paris RIA, 1994
- RASTIER François RASTIER, "Représentation ou interprétation ? Une perspective herméneutique sur la médiation sémiotique", dans "Penser l'esprit : des sciences de la cognition à une philosophie cognitive", PUG, 1996
- RICCEUR Paul RICCEUR, "Le conflit des interprétations", Seuil, 1990
- ROUSSEAUX Francis ROUSSEAUX, "Contribution à une méthodologie d'acquisition des connaissances pour l'ingénierie des SIC : l'exemple de Cheops pour l'aide à la gestion de crises collectives à caractère géographique", Mémoire d'habilitation à diriger des recherches, rapport de recherche LAFORIA, 1995
- SFEZ Lucien SFEZ, "Critique de la décision", Presse de la fondation nationale des sciences politiques, 1992
- TURING Alan TURING, "Computing machinery and intelligence", Minds and Machines, Prentice Hall, 1964