



De l'intelligence artificielle aux
modèles de définition de l'intelligence
le cas des variations dans
l'Anthologie Grecque

Marcello Vitali-Rosati

Publié le 15-04-2024

<http://sens-public.org/articles/1729>



Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA
4.0)

Résumé

Ce texte a l'ambition de proposer une sorte de révolution copernicienne dans les études sur ce qu'on appelle communément « intelligence artificielle ». La thèse qu'il s'agira d'illustrer est que au lieu d'essayer de créer des systèmes prétendument intelligents on doit se concentrer sur les multiples manières possibles pour donner des définitions précises et non ambiguës du concept d'intelligence – ou d'un autre concept. On passe ainsi de l'intelligence artificielle à des modèles de définition de l'intelligence. Pour illustrer cette approche et ouvrir la voie à une méthodologie scientifique de définition de modèles d'intelligence, cet article, après avoir posé les principaux enjeux théoriques d'une telle démarche, se concentrera sur un cas d'étude particulier, celui du projet de recherche « intelligence artificielle littéraire » qui vise à définir algorithmiquement le concept de « variation » tel qu'il se manifeste dans l'Anthologie Grecque.

Abstract

The aim of this text is to propose a kind of Copernican revolution in the study of what is commonly referred to as “artificial intelligence”. The thesis it aims to illustrate is that, instead of trying to create supposedly intelligent systems, we need to focus on the many possible ways of giving precise, unambiguous definitions of the concept of intelligence –or any other concept. In this way, we move from artificial intelligence to “intelligence definition models”. To illustrate this approach and pave the way for a scientific methodology for defining models of intelligence, this article, after setting out the main theoretical issues involved in such an approach, will focus on a particular case study, that of the “Intelligence artificielle littéraire” research project, which aims to algorithmically define the concept of “variation” as it manifests itself in the Greek Anthology.

Mot-clés : Intelligence artificielle, Définition, Machine, Intelligence humaine, ChatGPT, Anthologie Grecque, Modèles de langue, Sens

Keywords: Artificial intelligence, Definition, Machine, Human intelligence, ChatGPT, Greek Anthology, Language models, Sens

Table des matières

Introduction	4
« Intelligence artificielle » ?	5
Changer de paradigme	11
Le cas des variations dans l' <i>Anthologie Grecque</i>	19
Conclusion	24
Bibliographie	25

De l'intelligence artificielle aux modèles de définition de l'intelligence

Marcello Vitali-Rosati

Introduction

Ce texte a l'ambition de proposer une sorte de révolution copernicienne dans les études sur ce qu'on appelle communément « intelligence artificielle ». La thèse qu'il s'agira d'illustrer et, possiblement, de démontrer est que, au lieu de s'interroger sur l'intelligence artificielle, les sciences, et en particulier les sciences humaines, devraient se concentrer sur la production de « modèles de définition de l'intelligence¹ ». En d'autres termes, la question qui devrait être au centre des recherches n'est pas de savoir si une intelligence artificielle peut exister ou si des systèmes particuliers sont intelligents ou non, mais d'essayer de donner une, ou plusieurs « bonnes » définitions de ce qu'est l'intelligence – ou certaines formes particulières d'intelligence.

Il y va à la fois de la compréhension de ce qui se produit dans les dernières années – le développement (et surtout la mise à disposition pour le grand public) des grands modèles de langage (comme GPT3, Gemini, Bert, Alexa TM, LlamA), l'explosion du discours public et médiatique sur l'IA, l'ensemble des peurs et angoisses qui y sont reliées – et de la compréhension, plus importante et plus fondamentale, de qui nous sommes, de ce que sont les êtres humains.

Pour illustrer cette approche et ouvrir la voie à une méthodologie scientifique de définition de modèles d'intelligence, cet article, après avoir posé les principaux enjeux théoriques d'une telle démarche, se concentrera sur un cas d'étude

1. Pour une réflexion approfondie sur le concept de « modèle » et son application dans la recherche en Sciences Humaines, cf. Meunier (2017).

particulier, celui du projet de recherche « Intelligence Artificielle Littéraire » que je mène en collaboration avec Dominic Forest, Elsa Bouchard, Mathilde Verstraete et Yann Audin². Ce projet vise à définir algorithmiquement le concept de « variation » tel qu'il se manifeste dans l'Anthologie Grecque.

« Intelligence artificielle » ?

Pour commencer, il est nécessaire de faire quelques précisions sur l'expression « intelligence artificielle ». L'usage de cette expression et son omniprésence dans le discours médiatique – en particulier depuis novembre 2022, à la suite de la mise à disposition au grand public de chatGPT – est particulièrement significatif et il est important de le questionner pour bien comprendre la nature du problème. C'est une notion très vague, très floue, utilisée depuis longtemps à tort et à travers et qui met ensemble une série d'approches, méthodes, outils, environnements qui n'ont rien – ou très peu – à voir les uns avec les autres.

Russell et Norvig (2021), un des manuels de référence dans le domaine, présente l'IA comme l'effort de comprendre et de construire des entités intelligentes. Les auteurs montrent toute l'extension d'un tel champ :

Historically, researchers have pursued several different versions of AI. Some have defined intelligence in terms of fidelity to human performance, while others prefer an abstract, formal definition of intelligence called rationality—loosely speaking, doing the “right thing.” The subject matter itself also varies : some consider intelligence to be a property of internal thought processes and reasoning, while others focus on intelligent behavior, an external characterization. (2021, 31)

Comme dans le cas du mot « numérique³ », l'expression « intelligence artificielle » – la citation de Russel et Norvig le démontre bien – met ensemble des approches, des méthodologies, des outils et des démarches très différents tels que les modèles de langage, des systèmes experts pour résoudre certains

2. Le projet est financé par le programme « Développement Savoir » du CRSH depuis 2022.

3. J'ai montré, (cf. par exemple Vitali-Rosati 2021) que cette notion unifie dans un ensemble très pastiché une série hétérogène de phénomènes dont on peut légitimement se demander s'ils ont vraiment entre eux des traits en communs.

problèmes spécifiques, des applications se basant sur des données structurées et des logiciels de toutes sortes. S'il est vrai que, dans les dernières années, on a souvent limité la notion aux systèmes d'apprentissage profond, et, plus précisément, à des approches génératives, basées sur des réseaux de neurones, cette précision reste problématique et assimile des objets très hétérogènes. Quelle est la relation entre un réseau de neurones entraîné pour reconnaître des chiffres manuscrits – vieux problème informatique qui a été un des premiers succès des systèmes génératifs – et chatGPT ? Peut-on penser comprendre la spécificité d'un transformeur génératif pré-entraîné (ce qu'est GPT) en le mettant dans le même panier qu'une approche comme l'allocation de Dirichlet latente en traitement automatique des langues ?

Pourquoi alors parle-t-on d'« intelligence artificielle » ? D'où vient l'intérêt immense, presque obsessionnel, pour cette thématique – et pour cette expression ?

En suivant la suggestion de Russel et Norvig, nous pouvons nous concentrer sur la notion d'« intelligence » et tenter de comprendre le rapport entre intelligence humaine et intelligence artificielle qui semble être la question centrale qui détermine l'intérêt de cette notion.

Pour faire un peu d'ordre dans ces questionnements, je propose de partir d'une affirmation de Larry Tesler, ou plutôt du faux usage qu'en fait Douglas Hofstadter dans son célèbre *Gödel, Escher, Bach* en 1979 (1999).

Commençons par quelques remarques contextuelles : on peut trouver dans un livre de 1979 les mêmes discours et analyses qu'on retrouve aujourd'hui à propos de l'IA. En lisant Hofstadter on a l'impression qu'il n'y a pas eu beaucoup de changements depuis 40 ans, ou du moins pas autant que le discours médiatique et la rhétorique de la révolution voudraient nous le faire croire. Il n'y a rien d'absolument nouveau dans ce qui se passe dans les dernières années, rien de « révolutionnaire ». Des changements, certes, mais qu'on ne peut comprendre – comme d'habitude – que dans une continuité. Les discours qui se fondent sur la rhétorique de la révolution sont destinés à une myopie de fond et il faut apprendre à les regarder avec beaucoup de méfiance. En réalité, les discours n'ont pas changé, les préoccupations non plus et se concentrent sur la question de fond du rapport entre humain et machine tel que décrit par la citation de Russel et Norvig.

Ce qui est encore plus important de souligner c'est que, même du point de vue purement technique, les changements, certes indéniables, ne peuvent pas vraiment être considérés comme des révolutions. Il est vrai que certaines méthodes sont très récentes : notamment, dans le domaine du traitement des langues, les grands modèles de langage les plus récents – comme GPT3 ou Bert – sont fondés sur les transformeurs, une approche qui a vu le jour en 2015 ; l'idée de se baser sur l'« attention » – on reviendra sur cette notion – a été développée par un groupe de chercheurs travaillant chez Google dans le but de maximiser la performance de calcul par rapport aux réseaux de neurones récurrents ou convolutifs (Vaswani et al. 2017). Par ailleurs, la force de calcul d'aujourd'hui est incomparable avec celle d'il y a quelques années seulement. Les bases théoriques et mathématiques qui sont au fondement de ces méthodes sont beaucoup plus anciennes : des approches probabilistes utilisant des réseaux des neurones, existent au moins depuis les années 1960, et elles utilisent des idées mathématiques connues depuis des siècles (que l'on pense à la régression linéaire, utilisée déjà au début du XIX^e siècle par Gauss).

Revenons à Hofstadter et à la fameuse phrase de Tesler qu'il baptisa « le théorème de Tesler ». Tesler était un informaticien très actif dans des entreprises comme Xerox Parc (années 1970), puis Apple pour la plupart de sa carrière (1980-2000) et finalement Amazon et Yahoo. Hofstadter, dans son livre, cite une phrase que Tesler aurait eu l'habitude de dire et, affirme que pour cette raison, il appellera cet adagio le « théorème de Tesler ». La phrase est la suivante :

L'intelligence artificielle est tout ce qui n'a pas encore été fait.
(Hofstadter 1999, 601)

Cette idée a été souvent citée, réutilisée et est à la base de ce qu'on appelle l'*effet IA*. C'est une sorte de complexe du domaine, qui pousse toujours à penser que l'intelligence que l'on veut atteindre est plus éloignée que ce que l'on a réussi à atteindre. On développe un système qui fait des opérations super complexes, mais on se dit que l'IA est encore à conquérir. On réalise un algorithme qui sait jouer aux échecs, mais on se dit que l'intelligence artificielle n'a pas été atteinte, car le système n'a – par exemple – pas d'intuition. On crée un algorithme capable d'écrire des sonnets, mais on se dit que l'intelligence artificielle impliquerait qu'il ait aussi des sentiments.

Le théorème de Tesler est donc une volonté toujours renouvelée d'aller plus loin, une espèce de *Streben Tat* faustien dans le domaine de l'informatique (où le *Ewig-Weibliche* devient le *Ewig-Menschliche*), un complexe qui fait en sorte que la machine court toujours derrière l'humain sans jamais être capable de le rattraper. Dans cette configuration, l'intelligence humaine serait stable, un objectif très élevé que la machine devrait essayer d'atteindre ; l'intelligence humaine est un point de repère fixe, la machine court pour le poursuivre.

Or, en réalité, Tesler n'a jamais dit cette phrase. C'est lui-même qui l'affirme⁴ :

My formulation of what others have since called the "AI Effect". As commonly quoted : "Artificial Intelligence is whatever hasn't been done yet". What I actually said was : "Intelligence is whatever machines haven't done yet". Many people define humanity partly by our allegedly unique intelligence. Whatever a machine – or an animal – can do, must (those people say) be something other than intelligence. The theorem is cited in this Slashdot discussion and in these books :

Gödel, Escher, Bach : An Eternal Golden Braid by Douglas R. Hofstadter, Basic Books (1979, updated 1999), p. 601.

Foundations of Computer Technology by Alexander John Anderson, CRC Press (1994), p. 395.

Real-Time Systems Engineering and Applications edited by Michael Schiebe and Saskia Pferrer, Springer (1992), p. 398. (Tesler, s. d.)

Tesler a dit ce qui semble être complètement opposé à l'effet IA : selon sa formulation ce n'est pas la machine qui court derrière les êtres humains pour essayer de les rattraper, mais plutôt les êtres humains qui courent pour se déplacer par rapport à la machine.

L'enjeu n'est pas de développer des intelligences artificielles de plus en plus proches de l'intelligence humaine, mais plutôt de définir ce qu'est l'intelligence humaine, en nous comparant aux machines, et de la définir à chaque fois par la négative : si une machine sait faire cela, ce n'est pas ce qui est propre de l'intelligence humaine. Si la machine sait jouer aux échecs, l'intelligence

4. Cf. son site personnel à l'adresse http://www.nomodes.com/Larry_Tesler_Consulting/Adages_and_Coinages.html.

humaine doit être autre, peut-être de savoir parler dans une langue naturelle ; si la machine sait parler dans une langue naturelle, alors l'intelligence humaine sera plutôt dans l'intentionnalité ; si la machine peut avoir de l'intentionnalité, alors l'intelligence humaine résidera dans les sentiments... et ainsi de suite à l'infini.

Ce type de discours n'est pas nouveau et surtout ne concerne pas seulement notre rapport aux machines. Depuis toujours les êtres humains ont voulu se définir par rapport à ce qu'on a identifié comme « autre ». Et le point de cette démarche n'est pas tant de se distinguer, mais plutôt d'affirmer sa supériorité. Comme l'ont souligné plusieurs travaux dans le domaine des *posthuman studies* (par exemple Braidotti 2013 ; ou Hayles 1999), les êtres humains ont défini leur « essence » pour fonder leur supériorité par rapport aux animaux (l'homme est un animal rationnel, ou doté de langage, ou social...), les genres (justement : les « hommes », où on sous-entend que les femmes seraient moins humaines que les hommes), les peuples (avec la notion de race), puis les automates (que l'on pense à Descartes) et finalement, aujourd'hui, les ordinateurs. Le but du jeu est toujours le même : se placer vis-à-vis d'un autre **sujet** – un animal, un autre être humain, une machine... – et dire que nous sommes « mieux » ou « supérieurs », car nous ne nous « réduisons » pas à cela. Très souvent, dans ce jeu de supériorité, ce qui nous distinguerait de ce que nous avons en face n'est pas très bien défini, cela reste flou. Nous disons que nous avons un « plus ».

Turing avait très bien saisi ce type d'attitude dans son célèbre article de 1950 (Turing 1950). Dans la section où il discute les objections qu'on pourrait soulever à l'idée que les machines peuvent penser, le mathématicien parle des « arguments à partir de différentes incapacités ». Voici le texte de Turing :

Arguments from Various Disabilities. These arguments take the form, "I grant you that you can make machines do all the things you have mentioned but you will never be able to make one to do X". Numerous features X are suggested in this connexion. I offer a selection : Be kind, resourceful, beautiful, friendly, have initiative, have a sense of humour, tell right from wrong, make mistakes, fall in love, enjoy strawberries and cream, make some one fall in love with it, learn from experience, use words properly, be the subject of its own thought, have as much diversity of behaviour as a man, do something really new. [...] No support is usually

offered for these statements. I believe they are mostly founded on the principle of scientific induction. A man has seen thousands of machines in his lifetime. From what he sees of them he draws a number of general conclusions. (1950)

La stratégie décrite par le mathématicien britannique est exactement celle dont parle Tesler : il s'agit de déplacer l'attention pour chercher ce que « la machine » ne saurait pas faire. Or le problème est que, évidemment, devant une machine, on trouvera toujours ce qu'elle ne sait pas faire, car il s'agit justement d'une machine. Une machine particulière qui implémente certains modèles particuliers qui lui permettent d'avoir des capacités particulières. Cette machine ne saura nécessairement pas tout faire, car elle n'a pas été modélisée. Pour donner un exemple, lorsqu'on analyse un LLM (que ce soit GPT, Gemini ou un autre), on constatera que l'algorithme n'est pas capable de discerner ce qui est vrai de ce qui ne l'est pas. C'est juste évident : la notion de vérité n'est pas implémentée dans l'algorithme. Ce n'est pourtant pas une « limite » des « machines », juste une caractéristique du modèle qui n'est tout simplement pas fait pour savoir ce qui est vrai, mais pour être capable de manipuler la langue naturelle. On peut avoir d'autres algorithmes – même beaucoup plus simples – qui connaissent la vérité et qui sont capables de fournir seulement des informations vérifiées et acceptées, par exemple, par une communauté scientifique. Demandez la liste des livres de Proust au catalogue d'une bibliothèque nationale et cette liste sera absolument correcte.

À chaque fois qu'on dit : « et ça, une machine ne pourra jamais le faire », nous mettons en place une double triche. En premier lieu, celle identifiée par Turing : nous induisons, sur la base des machines que nous connaissons – et souvent en nous limitant aux machines que nous avons devant les yeux – que « toutes les machines » ont les mêmes caractéristiques. Puisque ChatGPT ne peut pas aimer les fraises et la glace, alors les machines ne sauront jamais aimer les fraises et la glace. Mais le problème est qu'il peut exister d'autres modèles computationnels qui implémentent l'idée du plaisir de la glace. En second lieu, nous donnons souvent une définition très floue de ce « ça » que les machines ne sauraient pas faire.

Un exemple de ce type d'attitude est le discours de Searle dans son fameux article *Mind, brains, and programs* (1980), où le philosophe américain propose le *Gedanken Experiment* de la « chambre chinoise ». La thèse de fond de ce texte est justement que le sens est « plus » par rapport à la syntaxe. Cette

idée est ce que porte Searle à représenter dans une opposition deux pôles irréductibles, celui où se trouve le sens, la compréhension et l'humain et où se trouvent les machines, qui est fait de syntaxe et de règles. Searle affirme :

I understand stories in English ; to a lesser degree I can understand stories in French ; to a still lesser degree, stories in German ; and in Chinese, not at all. My car and my adding machine, on the other hand, understand nothing : they are not in that line of business. (Searle 1980, 419)

Or, le problème de l'argumentation de Searle est qu'elle se base sur une définition très claire et exhaustive de ce qu'est la syntaxe en laissant, au contraire, dans un flou stratégique ce qu'est le sens. Searle explique précisément ce qu'est la syntaxe en décrivant les règles et les tables de correspondance dont dispose la personne dans la chambre chinoise, mais en ce qui concerne la définition du « sens », il se limite à faire appel au sentiment que le lecteur peut avoir de « comprendre ». Le sens est simplement défini par la négative : tout ce que la syntaxe n'est pas⁵. On est clairement dans un cas d'effet IA : il y a quelque chose qui excède, mais quoi ? Souvent la réponse est très floue. Turing (1950), conscient de l'impossibilité de définir ce « plus » qui caractériserait l'humain, affirme que la réponse à la question « est-ce qu'une machine peut penser » ne peut être donnée qu'avec une approche pragmatique, justement parce que – selon lui – il est impossible de donner une définition formelle de ce qu'on entend par « pensée ». Il propose donc un test de comportement : si une machine agit comme un humain (on présuppose qu'il pense), alors elle pense.

Changer de paradigme

Dans le cadre de la réflexion sur l'intelligence artificielle, la question centrale semble être toujours la même : qui est mieux, l'humain ou la machine ? Si la réponse penche du côté de l'humain, on se dit que la machine doit se développer davantage. Si elle penche du côté de la machine, dans des crises d'angoisse, nous essayons de déplacer ce qu'est humain pour trouver une marge d'agentivité qui puisse nous garantir, encore une fois, notre supériorité.

5. Il existe plusieurs travaux qui tentent des définitions précises de « sens » et « sémantique » utilisables algorithmiquement. Cf. pour ne donner que deux exemples, les travaux de Pierre Lévy sur le langage artificiel IEMML (Lévy 2023), ou les travaux de Rapaport (2013).

Et si on changeait de question ? Et si on décidait d'arrêter cette course folle qui a pour but de démontrer notre supériorité ? Et si on arrêta de vouloir affirmer une spécificité humaine et qu'on essayait au contraire de comprendre les véritables raisons qui nous poussent à courir, en réfléchissant sur les différentes possibilités de compréhension et de définition de l'humain et de l'intelligence ?

Comme le dit Karen Barad dans son *Meeting the Universe Halfway* :

My posthumanist account calls into question the givenness of the differential categories of human and nonhuman, examining the practices through which these differential boundaries are stabilized and destabilized. (2007, 66)

Suivons donc la suggestion de Barad, arrêtons-nous, arrêtons d'essayer de montrer en quoi un prétendu « humain » serait supérieur à une prétendue « machine » et arrêtons d'essayer d'une part de rendre cette prétendue machine « intelligente » et de l'autre de montrer que cet « humain » reste toujours « plus intelligent ». Prenons cette course à contrepied pour nous poser la question : dans ce contexte, comment émerge la définition de ce qu'est humain ? Comment émerge la définition de « machine » ? Comment se stabilise la frontière entre les deux ? Comment peut-on questionner cette frontière et la penser autrement, la déstabiliser ? Comment émerge la définition d'« intelligence » ? Quelles en sont les implications politiques, éthiques, culturelles et sociales ?

Le cas de chatGPT est assez significatif. Ce qui a le plus frappé l'imaginaire collectif à propos de cette plateforme est le fait qu'elle passe haut la main le test de Turing⁶. En d'autres termes, il est impossible – ou très difficile – quand on regarde le texte qu'elle produit, de savoir si ce texte a été produit par un être humain ou pas. Ce qui est très intéressant est que les chatbots, orientés par un fort effet IA, essaient d'atteindre ce résultat depuis longtemps. Ce que chatGPT fait correspond parfaitement à la définition du test de Turing, qui imagine poser des questions sous forme de texte à deux agents, dont l'un est une machine et l'autre un humain et de voir si on est capable de reconnaître qui est l'humain. La manipulation des langages naturels semblait la frontière

6. Je me réfère ici au test de Turing comme ce dispositif est normalement compris dans le monde de l'informatique. En réalité, une analyse attentive du texte du mathématicien britannique révèle une complexité et une ambiguïté importantes dans la définition de ce « test », cf. Sterrett (1999).

ultime pour les machines, car on sait que la complexité des langues naturelles empêche des modélisations sous la forme d'une série de règles. L'approche des systèmes experts, qui a dominé les années 1980, semblait démontrer que le monde du langage naturel est trop compliqué pour qu'une machine, qui ne sait que manipuler des règles et des unités atomiques, puisse arriver à le manipuler comme un humain. L'humain est donc supérieur, on pouvait penser, car il est capable de ne pas respecter les règles. Il est capable de faire des opérations plus complexes, qui ne peuvent pas être modélisées et, donc, qui ne peuvent pas être calculées – et finalement computées⁷. Cette doxa est un exemple parfait de la volonté de définir l'humain en opposition à autre chose pour démontrer sa supériorité : l'intelligence humaine n'est pas réductible à un ensemble de règles – autant complexes soient-elles – qu'il suffirait d'appliquer. C'est ce type d'échec qui justifie un discours comme celui de Searle : les machines ne peuvent que calculer, l'humain serait du côté du « sens » et le sens ne serait pas calculable, parce qu'il ne semble pas pouvoir être réduit à une série de règles.

Le développement des grands modèles de langue et l'augmentation importante de la puissance de calcul semblent changer un peu la donne. Certes, l'argument de Searle n'est pas touché par le fait qu'un agent comme GPT3 soit capable de passer le test de Turing, parce que Searle nie que le sens soit réductible à une question de comportement – sans pour autant définir ce qu'il est. Néanmoins, les grands modèles de langue démontrent qu'il est possible de modéliser de façon très satisfaisante les comportements linguistiques, qui ne semblaient pas modélisables, avec des approches comme celles des systèmes experts. L'idée est assez simple : au lieu de construire un algorithme qui implémente une série très longue de règles – des grammaires, des dictionnaires, des règles syntaxiques, etc. – on peut essayer de faire induire les règles à partir de l'observation d'un – très large – corpus. Dit autrement : au lieu d'« enseigner » la langue à partir d'une série de règles, on va essayer d'« exposer la machine » à la langue telle qu'elle est utilisée et les règles en ressortiront.

La métaphore de l'apprentissage est fondamentale, car ces approches naissent grâce à l'inspiration tirée de la linguistique et des neurosciences qui observent depuis des décennies les modes d'apprentissage de la langue (et pas seulement) de la part des êtres humains. Justement, on n'apprend pas à un enfant sa langue

7. Sur la question des modèles et du rapport entre calculable et computable, cf. Meunier (2014).

maternelle en lui expliquant quand il doit utiliser le subjonctif ou l'indicatif. Tout simplement on lui parle. Ensuite, l'enfant réplique et répète ce qu'il a entendu et il « induit » les règles. Évidemment, ce n'est qu'une hypothèse et, par ailleurs, certains linguistes – comme Chomsky, par exemple – ne sont pas du tout d'accord avec cette interprétation. Ce qui compte c'est qu'on a une idée, une piste, sur laquelle on peut fonder une modélisation. L'idée de l'induction des règles est justement à la base de certaines approches génératives⁸. L'idée est de donner à l'algorithme un nombre très élevé de textes et de lui faire calculer la probabilité que certains mots viennent avec d'autres mots (avant ou après). Sur la base de ce calcul, on transformera les mots en vecteurs, à savoir en une série de chiffres, qui représenteront leur « sens ».

Ce qui est fondamental à retenir ici c'est que cette approche donne de fait une définition formelle du concept de « sens » (ce que ne fait pas Searle) : le sens est la relation qu'un mot d'un dictionnaire a avec l'ensemble des autres mots du dictionnaire par rapport à un corpus d'usage (définition très structuraliste, héritière des travaux de Ferdinand de Saussure et de l'idée de langue comme système autonome). Le sens devient définissable formellement et, donc, calculable. Une machine peut le manipuler. La preuve : chatGPT est capable de parler comme un humain et il passe le test de Turing. Ce qui me semble avoir raison de Searle, dans cette situation, n'est pas tellement le fait que cette définition de sens est *la bonne*, mais que cette approche donne *une* définition claire et formelle de sens⁹ et que la non ambiguïté de telle définition est démontrée par une vérification : on peut faire passer le test de Turing à un algorithme qui l'implémente.

L'argument de Searle ne tient pas, tout simplement, parce qu'il ne se fonde pas sur une définition précise du concept de sens, mais sur l'idée, négative, selon laquelle le sens est tout ce que qui n'a pas été défini.

8. Il est intéressant de souligner ici l'ambiguïté du terme. Car « génératif » est aussi le mot utilisé par Chomsky pour décrire un modèle linguistique opposé : pour Chomsky « génératif » signifie déductif – il y a des règles linguistiques transcendantales qui sont applicables à toutes les langues. Ce qui est généré, pour Chomsky, ce sont les langues. Alors que dans le sens courant aujourd'hui dans le domaine de l'IA, « génératif » équivaut à « inductif » : ce qui est généré, à partir d'une induction, ce sont les règles.

9. Il y a donc d'autres définitions possibles ; les idées de « sens » et de « sémantique » proposées par Lévy (2023) ou Rapaport (2013) sont complètement différentes. C'est cette possibilité de multiplicité qui est importante à retenir.

Alors, faut-il se dire que l'intelligence artificielle a gagné, que la machine a rattrapé l'humain dans sa course ? Ou qu'elle n'a pas encore tout fait, qu'il faut aller plus loin – comme le veut le théorème de Tesler ? Ou encore que l'humain est ailleurs, que passer le test de Turing n'est pas ce qui le définit et, par exemple, qu'il « connaît le sens de ce qu'il dit » alors que chatGPT ne le sait pas ? Dans ce dernier cas, on est justement dans le jeu de l'humain qui essaie de se définir par opposition et qui affirme toujours qu'il est ailleurs, avec un ailleurs très flou, car comment peut-on définir ce que signifie « comprendre ce qu'on dit » ? Comment définir de manière formelle ce que sait ou pas chatGPT ?

Il y a une autre possibilité et c'est la piste que je propose ici. Nous pouvons nous concentrer sur la définition d'intelligence qui émerge dans le cas particulier de chatGPT. Nous pouvons nous arrêter sur l'idée de langue naturelle, étudier la définition de « sens » qui est implémentée dans l'approche particulière de cet algorithme. Par exemple, une caractéristique fondamentale de GPT3 par rapport à d'autres modèles de langue est qu'il se fonde, comme nous l'avons dit, sur le concept d'attention. Cette idée lui permet d'avoir une performance bien meilleure que d'autres modèles. L'idée, encore une fois est simple : si je dois deviner le rapport entre les mots qui composent une phrase, tous les mots ne comptent pas de la même manière. Cette idée permet de paralléliser le calcul sur le corpus – au lieu de le faire en séquence, mot à mot, on peut le faire en parallèle sur plusieurs phrases en attribuant des « poids » différents aux mots spécifiques.

Cette idée est une manière de définir l'intelligence humaine : le sens émerge non seulement grâce à une série de relations entre mots – où tous les mots seraient sur le même plan. L'émergence du sens comporte aussi des mécanismes d'attention et l'attention est définie de façon formelle – donc de façon non ambiguë. Encore une fois : le point n'est pas de dire si c'est « la » bonne définition, mais si c'est « une » bonne définition.

On peut reprendre la définition d'attention donnée dans l'article qui a lancé la technologie des transformeurs :

An attention function can be described as mapping a query and a set of key-value pairs to an output, where the query, keys, values, and output are all vectors. The output is computed as a weighted sum of the values, where the weight assigned to each

value is computed by a compatibility function of the query with the corresponding key. (Vaswani et al. 2017, 3)

Les auteurs de cet article ont essayé de donner *une* définition d'attention et ensuite de l'implémenter dans un modèle formel – la création de vecteurs qui permettent de représenter les différents poids de chaque objet.

Ce qui est intéressant de chatGPT n'est pas de savoir s'il est meilleur ou moins bon qu'un être humain. Ce n'est pas de savoir si la machine est intelligente. Ce n'est pas de savoir si l'être humain restera supérieur ou s'il se fera dépasser par la machine. C'est plutôt de comprendre comment on peut définir l'intelligence humaine et les implications de chaque définition.

Les modèles de langue montrent qu'il est possible de modéliser certaines habilités. **Dans le cas de chatGPT, ce qui est surprenant, c'est que les algorithmes sur lesquels il est basé modélisent des habilités qui semblaient être très difficiles à modéliser ; mais de l'autre côté, des habilités très faciles à modéliser ne sont pas prises en compte dans sa structure.** Pour le dire autrement : chatGPT sait très bien faire certaines opérations que les machines ne semblaient pas être capables de bien faire et, symétriquement, il en fait d'autres très mal que les machines savent très bien faire. Justement chatGPT est incapable d'appliquer des règles formelles. Il est incapable de faire une requête dans des données structurées pour sortir de l'information. Tout ce que savent très bien faire des systèmes experts très rudimentaires – comme une requête à une API, par exemple, pour récupérer la liste des livres publiés par Proust – chatGPT ne sait pas le faire. Le modèle probabiliste fait en sorte que chatGPT sait très bien faire des discours de comptoir, où il peut parler de n'importe quel sujet, sans aucune validation structurée de ce qu'il dit, toujours utilisant un langage clair et qui semble donc fiable. En effet, il met ensemble des mots qui, par rapport à un large corpus, ont de bonnes probabilités d'être ensemble et cela donne beaucoup de vraisemblance à ses propos. L'apprentissage supervisé qui a été rajouté par-dessus, en plus, lisse le discours pour donner lieu à des propos politiquement conformes et attendus.

ChatGPT modélise donc un certain type d'intelligence. Les algorithmes sur lesquels il est basé – le transformeur génératif pré-entraîné et ensuite le *reinforcement learning* – représentent l'implémentation formelle d'une définition particulière d'intelligence. Une étude approfondie – qui n'est pas l'objet de ce texte – pourrait nous permettre de spécifier de façon précise telle définition,

mais on pourrait la résumer plus ou moins comme suit : l'intelligence est la capacité d'associer à une série de mots quelconques une autre série de mots qui ait une haute probabilité de se trouver après ces premiers mots, cela en respectant un ordre et en pondérant chaque mot pour faire la différence entre des mots plus importants et d'autres moins. Cette association doit aussi tenir en compte l'alignement avec les attentes d'un public occidental et respecter les mœurs linguistiques de ce public¹⁰.

Ce qui est intéressant de chatGPT ce n'est donc pas tellement le fait qu'il soit « performant » – par rapport à quoi ? –, mais plutôt sa manière de définir l'humain. De fait il donne une définition particulière de ce qu'est l'intelligence en valorisant certains aspects et en en mettant d'autres entre parenthèses. Sans doute, l'intelligence qu'il propose peut être analysée et critiquée : elle en a très peu d'une intelligence « scientifique », n'a rien de rigoureux, ne donne aucune importance à la question de la vérité, de la vérifiabilité, de la logique d'argumentation. Elle ne donne aucune importance au savoir, ni à la connaissance, mais c'est sans doute une forme d'intelligence très valorisée par notre société contemporaine.

On peut – et on devrait – évidemment se poser la question : et si on voulait modéliser autre chose ? Si on voulait penser autre chose ? Si on voulait par exemple se concentrer sur la conscience ? En effet, chatGPT ne modélise pas la conscience. Mais on pourrait le faire. On pourrait se dire que ce qui nous intéresse est plutôt de comprendre les mécanismes de conscience – c'est ce que Turing mettait entre parenthèses dans son texte en disant qu'il était impossible de donner une définition formelle de conscience et ce sur quoi joue avec ruse le texte de Searle.

Or, contrairement à Turing, je suis convaincu qu'il est possible de donner *une* définition de conscience qui soit « bonne », à savoir non ambiguë et formalisable. C'est ce que fait par exemple Giulio Tononi dans un projet de recherche, où après avoir donné une définition formelle de « conscience » (son *Integrated Information Theory* (Tononi 2015)), il essaie de voir à quel point certains algorithmes sont conscients. Sa conclusion préliminaire (Tononi et Koch 2015 ; Tononi 2012) explique que des systèmes complexes basés sur des réseaux de neurones ne sont pas si conscients et que des technologies plus

10. Le *reinforcement learning* répond à la fameuse question de l'alignement. Les modèles de langue pré-entraînés ont souvent une haute « capacité », mais un très bas « alignement ». Cf. Ramponi (2022).

simples, qui ne passent pas du tout le test de Turing (parce que ce n'est pas leur but), le sont davantage. Cela porterait à penser que des compétences comme la manipulation de la langue naturelle ne sont pas nécessairement liées à la conscience. Ce qu'il faut souligner ici est que de telles recherches ont le mérite justement de produire des définitions précises de certains concepts. Encore une fois, on ne cherche pas *la* définition de « conscience » ou d'intelligence, mais on cherche à donner *une* définition possible et bien formalisée.

Cette approche change d'orientation : au lieu d'essayer de créer des systèmes prétendument intelligents – sans trop savoir ce qu'intelligence signifie – on se concentre sur les multiples manières possibles pour donner des définitions précises et non ambiguës du concept d'intelligence – ou d'un autre concept. On passe ainsi de l'intelligence artificielle à des *modèles de définition de l'intelligence*.

Je pense que c'est là le futur des sciences et en particulier des sciences humaines. Cela implique de repenser l'activité scientifique et d'arrêter de l'orienter vers l'applicatif – qui hélas est le seul aspect qui semble nous intéresser dans notre monde capitaliste – en se concentrant sur la recherche fondamentale.

N'était-ce pas la critique que Socrate faisait contre les sophistes dans le *Phèdre* ? Au lieu d'essayer de jouer avec les mots pour convaincre quelqu'un à faire quelque chose, essayons d'abord de comprendre et de définir les mots, les concepts et les idées. À la différence de Platon, je pense qu'il n'y a pas une essence des concepts, il n'y a pas une seule définition, mais des définitions multiples. Cependant, il y a une différence fondamentale entre une bonne définition et une mauvaise : la bonne est claire, formelle et non ambiguë, la mauvaise est floue et idéologique. L'exemple par excellence est la mauvaise définition que Searle donne de « sens » en jouant justement sur un déplacement infini et régressif du sens par rapport à ce qu'on peut définir formellement. Le sens, pour Searle, est tout ce qu'il ne prend pas la peine de définir et en jouant sur cette ambiguïté, il peut utiliser le concept pour revendiquer une supériorité humaine. L'exemple de bonne définition est la notion de conscience pour Tononi : ce n'est sans doute pas la *seule* définition, mais c'est une définition non ambiguë, claire et bien formalisée.

Le cas des variations dans l'*Anthologie Grecque*

La piste des modèles de définition de l'intelligence peut être appliquée à n'importe quelle idée et n'importe quel concept : l'intelligence, le sens, la conscience, l'amour, l'intuition...

Dans le cadre d'un projet de recherche que je mène avec Dominic Forest, Elsa Bouchard, Mathilde Verstraete et Yann Audin, nous essayons de donner une définition formelle du concept de « variation » littéraire dans le cadre de l'épigrammatique grecque. Pour être plus précis, notre corpus est représenté par l'*Anthologie Palatine*, un recueil d'environ quatre-mille épigrammes composé au X^e siècle à Byzance, qui met ensemble des épigrammes de plus de cent auteurs ayant écrit entre le VI^e siècle av. J.C. et le X^e siècle de notre ère¹¹.

Une épigramme est une courte composition poétique destinée à être inscrite sur un sujet : un objet, un mur, une tombe, etc. L'épigramme devient ensuite un genre littéraire et les compositions ne sont souvent pas nées comme de véritables inscriptions, mais elles mettent en scène cette inscription (Cairns 2016). Cette émergence de l'épigrammatique comme un genre – datable du IV^e siècle av. J.C. (Tarán 1979, IX:1) et la fortune de ce genre à l'époque hellénistique implique la naissance d'une série de « pratiques érudites ». Comme l'affirme Tarán :

the *poetae docti* proudly took it upon themselves to study the work of their predecessors in order to incorporate it in their own, mainly in the form of glosses, so that the literature of the new age would both follow the long-established tradition and the same time entail innovation. (Tarán 1979, IX:2)

Cette démarche qui cherche l'équilibre entre tradition et innovation est ce qui donne la naissance aux « variations » : un auteur « reprend » une épigramme d'un de ses prédécesseurs et la réécrit pour démontrer son habilité poétique.

11. Dans un dossier thématique où l'*Anthologie Grecque* a une place fondamentale, je ne donnerai que des références sommaires, juste pour que l'article soit intelligible pour un non spécialiste. Pour plus d'informations sur l'*Anthologie Palatine*, cf. Cameron (1993). Pour une histoire du manuscrit, cf. Beta (2019). En ce qui concerne les éditions, nous ferons ici référence à celle des Belles Lettres.

Il y a plusieurs groupes de variations dans l'*Anthologie Palatine* – Tarán dit en avoir identifiés 120. Un exemple peut aider à mieux comprendre ce concept qui peut sembler trivial, mais qui est loin d'être de simple définition.

L'épigramme 13 du livre 6 est de Léonidas de Tarente, un poète qui a vécu entre le IV^e et le III^e siècles av. J.C, et il semble qu'elle était inscrite sur une fresque à Pompéi. La voici en grec et dans la traduction française de Waltz :

οἱ τρισσοὶ τοι ταῦτα τὰ δίκτυα θῆκαν ὄμαιμοι,
ἀγρότα Πάν, ἄλλης ἄλλος ἀπ' ἀγρεσίης :
ὦν ἀπὸ μὲν πτηνῶν Πίγρης τάδε, ταῦτα δὲ Δᾶμις
τετραπόδων, Κλείτωρ δ' ὁ τρίτος εἰναλίων.
ἀνθ' ὧν τῷ μὲν πέμπε δι' ἠέρος εὖστοχον ἄγρην,
τῷ δὲ διὰ δρυμῶν, τῷ δὲ δι' ἠϊόνων.

5

Les trois frères t'ont consacré, chasseur Pan, ces filets, pris par chacun à son genre de chasse : Pigrès, pour les oiseaux ; Damis, pour les quadrupèdes ; Cléitor, pour le peuple de la mer. Envoie-leur en échange une bonne chasse à l'un par les airs, au second par les bois, à l'autre par les grèves.

(*Anthologie grecque. t. III*, s. d., 31)

Cette épigramme est reprise maintes fois (6.11, 6.12, 6.14, 6.15, 6.16, 6.187, etc.). Par exemple, l'épigramme 6.11, de Sartrius, est d'époque romaine, la voici :

θηρευτῆς δολιχὸν τόδε δίκτυον ἀνθετο Δᾶμις
Πίγρης δ' ὀρνίθων λεπτόμιτον νεφέλην,
τριγλοφόρους δὲ χιτῶνας ὁ νυκτερέτης θέτο Κλείτωρ
τῷ Πανί, τρισσῶν ἐργάτιναι καμάτων.
ἴλαος εὐσεβέεσσιν ἀδελφειοῖς ἐπίνευσον
πτηνά, καὶ ἀγροτέρων κέρδεα καὶ νεπόδων.

5

Le chasseur Damis a consacré ce long panneau, Pigrès ce filet aux fines mailles pour attraper les oiseaux et le rameur de nuit Cleitor ce tramail à mettre les rougets : c'est à toi, Pan, que tous trois ont dédié ces instruments de leur travail ; sois propice à ces frères si pieux, accorde-leur leur provende de volatiles, de venaison et d'habitants des eaux

(*Anthologie grecque. t. III, s. d., 30-31*)

Le 6.12, de Julien L'Égyptien, d'époque byzantine (quelques 8 siècles après l'original) :

γνωτῶν τρισσατίων ἐκ τρισσατίης λίνα θήρης
δέχνυσο, Πάν Πίγρης σοὶ γὰρ ἀπὸ πτερύγων
ταῦτα φέρει, θηρῶν Δᾶμις, Κλείτωρ δὲ θαλάσσης.
καί σφι δὸς εὐαγρεῖν ἡέρα, γαῖαν, ὕδωρ.

De trois frères, reçois, Pan les filets de leurs trois genres de chasse : Pigrès t'apporte celui de la gent ailée ; Damis, celui des bêtes sauvages ; Cleitor, celui de la mer. En échange, accorde-leur de bonnes prises, dans l'air, sur terre, dans l'eau.

(*Anthologie grecque. t. III, s. d., 31*)

En lisant ces trois textes on se fait une idée de ce qu'est une variation : les trois épigrammes « parlent du même sujet », elles mettent en scène les mêmes personnages – avec les mêmes noms propres –, mais elles présentent cette situation en utilisant des mots différents. Cependant, cela se complique lorsqu'on lit l'épigramme 6.17 du Pseudo-Lucien :

αἰ τρισσαί τοι ταῦτα τὰ παίγνια θῆκαν ἑταῖραι,
Κύπρι μάκαιρ', ἄλλης ἄλλη ἀπ' ἐργασίης :
ᾧν ἀπὸ μὲν πυγῆς Εὐφρῶ τάδε, ταῦτα δὲ Κλειῶ
ὡς θέμις, ἢ τριτάτη δ' Ἀτθίς ἀπ' οὐρανίων.
ἀνθ' ᾧν τῆ μὲν πέμπε τὰ παιδικά, δεσπότι, κέρδη,
τῆ δὲ τὰ θηλείης, τῆ δὲ τὰ μηδετέρης..

5

Trois courtisanes t'offrent ces jouets, bienheureuse Cypris, chacune des revenus de son travail spécial : Euphro, de ce que lui rapportent ses fesses, Cleio de ce qu'elle gagne en suivant les

règles de la nature, Atthis enfin de ce qu'elle doit à son palais. En échange, souveraine, envoie à la première les profits que souhaite un garçon, à la seconde ceux que désire une femme, à l'autre ceux du troisième sexe.

(*Anthologie grecque. t. III*, s. d., p.33)

On pourrait se dire que cette épigramme est aussi une variation de 6.13. Il s'agit en effet de la même structure et il est clair que le Pseudo-Lucien est en train de jouer avec le texte original pour en proposer sa propre version, en le détournant et en le parodiant. Certes, il ne s'agit pas du même type de variation.

Laurens (2012) identifie pour cette raison trois types de variation : la variation rhétorique, stylistique et paradigmatique.

Revenons maintenant à notre question centrale. Il serait possible d'approcher l'*Anthologie Palatine* avec les « outils » de l'« intelligence artificielle ». Il serait donc possible d'essayer de produire un algorithme capable de trouver des variations et de l'utiliser à des fins heuristiques. Des approches de ce type existent (cf., par exemple Schubert 2020 ; Pöckelmann et al. 2020) et donnent des résultats assez intéressants. En suivant la méthodologie de Pöckelmann et Schubert, à partir de l'application d'un algorithme de type *Word mover Distance* (Kusner et al. 2015), nous pourrions faire chercher à la machine les variations et ensuite évaluer sa performance en regardant si les résultats trouvés sont plus ou moins bons – à savoir s'ils correspondent plus ou moins à ce qu'un.e helléniste pourrait penser. C'est une approche qui consiste justement à mettre en place une sorte de compétition entre « humain » et machine : si l'humain pense que l'épigramme x est une variation de l'épigramme y et que l'algorithme trouve cette relation, alors la machine est intelligente, sinon non. On pourrait aussi imaginer le cas où la machine trouve des variations « mieux » que l'humain : dans ce cas elle gagnerait la compétition d'intelligence.

Or, le problème d'une approche de ce type est qu'on retombe dans le cadre de la chambre chinoise de Searle : a-t-on bien défini ce qu'est une variation ? Ou cela relève de la subjectivité, floue et indéfinie, de l'humain qui lit au fur et à mesure les textes ? Comment « mesurer » la performance de la machine si les critères d'inclusion ou exclusion ne sont pas clairs ?

C'est là le fond de ma proposition : comment serait-ce possible de développer un système intelligent – dans ce cas précis, il s'agit de l'intelligence qui permet

de distinguer une variation d'une non-variation – sans avoir clairement défini ce qu'est l'intelligence – dans ce cas : ce qu'est précisément et formellement une variation ?

Je répète la thèse de fond de ce texte : au lieu d'essayer de créer des systèmes prétendument intelligents, on doit se concentrer sur les multiples manières possibles pour donner des définitions précises et non ambiguës du concept d'intelligence.

Dans le cadre de ce projet, il s'agit non pas de chercher à développer un algorithme capable de trouver des variations, mais plutôt d'essayer de donner *une* bonne définition de variation. Encore une fois, l'idée n'est pas de donner *la* bonne définition de variation, mais *une* bonne définition : à savoir une définition claire, formalisable et non ambiguë, une définition sur laquelle on puisse avoir des critères précis pour décider si deux épigrammes sont ou pas l'une la variation de l'autre.

Pour ce faire, nous sommes en train de réaliser un double travail : d'une part, nous essayons d'identifier toutes les épigrammes que nous considérons comme des variations, de les classer, de les baliser et de différencier plusieurs types spécifiques de variation ; d'autre part, nous essayons de préciser la définition de ce que nous entendons par variation et d'implémenter cette définition – ou plutôt cette série de définitions, car il y en a au moins une par type de variation – dans un modèle formel et, ensuite, dans des algorithmes. Si l'algorithme est capable de trouver celles que nous avons identifiées comme des variations, alors cela signifie que notre définition est « bonne » : à savoir claire, précise et non ambiguë. Nous aurons, en d'autres termes, été capables de fournir un modèle de définition de la variation. C'est justement un exemple de « modèle de définition de l'intelligence ».

Le processus révèle aussi un autre aspect fondamental : l'effort de modélisation demande une série d'allers-retours entre l'interprétation des textes, la théorie littéraire et l'implémentation des algorithmes. Ces allers-retours impliquent une codépendance du modèle formel et du modèle théorique, ainsi qu'une interaction constante entre syntaxe et sens. En résumé, la syntaxe modifie le sens et le sens modifie la syntaxe. La compréhension de ce qu'est une variation dépend bien sûr de la lecture des épigrammes et de leur interprétation, mais cette interprétation et cette lecture sont modifiées par la volonté de formaliser la définition du concept, ainsi que par les moyens spécifiques qui sont mobilisés dans cet effort de formalisation ; en même temps ce que

les algorithmes trouvent, au fur et à mesure de l'expérimentation, modifie la compréhension de ce qu'est une variation. Le sens ne se fait pas chez les êtres humains, comme s'ils étaient des entités séparables de la machine, mais dans les interactions entre approche fonctionnelle et approche herméneutique. Pour être plus précis, la variation émerge d'une série d'interactions (ou intra-actions, pour utiliser la terminologie de Barad) où les dispositifs techniques, les langages de programmation, les structures de la logique formelle, la langue, la tradition littéraire, les conditions particulières d'une expérimentation (comportant aussi l'ensemble des conditions matérielles, les machines utilisées, l'état physique et mental des chercheur·e·s impliqué·e·s, etc.) sont mélangés.

Dans le cadre de ces intra-actions émergent donc le concept de variation, ainsi que la possibilité de sa compréhension, mais aussi se stabilisent certaines frontières qui font apparaître des textes, des chercheur·e·s, des langages, des machines, des algorithmes, etc.

Le cas de la « variation » peut sembler plus simple et moins ambitieux que le cas de l'intelligence ou de la conscience, mais en réalité il s'agit de la même démarche, une démarche qui peut être appliquée à n'importe quel concept et à n'importe quelle « habilité ».

Conclusion

L'approche des « modèles de définition de l'intelligence » demande une réorientation du travail scientifique. Comme je l'ai dit, il faut d'abord et avant tout se concentrer sur la recherche fondamentale et mettre entre parenthèses l'application. Par exemple, pour les variations il ne s'agit pas de faire un outil capable d'en trouver pour rendre plus « simple » ou plus rapide le travail d'analyse littéraire – parce qu'on délèguerait des tâches à l'« intelligence artificielle » ; il s'agit au contraire d'utiliser les algorithmes pour tester des définitions littéraires formelles. On ne gagne pas de temps, on en perd : c'est là la caractéristique de la recherche fondamentale. Cette approche implique aussi un changement important par rapport aux compétences. La recherche en sciences humaines impliquera des connaissances mathématiques, informatiques et logiques très poussées et approfondies. On ne peut pas parler d'IA sans comprendre ce qu'est un transformeur et la différence entre un transformeur et un réseaux de neurones convolutif. Il s'agit – dans la continuité des *Critical code studies* (Marino 2020) – d'être capables d'interpréter les algorithmes et de comprendre leur sens, pour ce faire, il faut en saisir les

détails. Cela demande sans doute des efforts, mais on ne peut pas en faire l'économie si nous voulons continuer à dire des paroles qui ont une valeur et qui ne soient pas un discours de comptoir – à la chatGPT.

Je suis conscient que cette proposition va complètement à contre-courant des tendances actuelles. On privilégie toujours l'applicatif et c'est l'une des raisons du succès des approches génératives, mais ces approches pénalisent fortement des définitions d'intelligence qui valoriseraient les aspects les plus scientifiques de nos comportements. Si elles nous permettent des applications avec des résultats souvent époustouffants, elles ne nous font pas avancer du point de vue de la connaissance.

Bibliographie

- Anthologie grecque. t. III.* s. d. Budé. Paris.
- Barad, Karen. 2007. *Meeting the Universe Halfway : Quantum Physics and the Entanglement of Matter and Meaning*. Second Printing edition. Durham : Duke University Press Books.
- Beta, Simone. 2019. *Moi, un manuscrit. Autobiographie de l'Anthologie palatine*. Le Kremlin Bicetre : Editions Belles Lettres.
- Braidotti, Rosi. 2013. *The posthuman*. Cambridge, UK ; Malden, MA, USA : Polity Press.
- Cairns, Francis. 2016. *Hellenistic epigram : contexts of exploration*. Cambridge : Cambridge University Press.
- Cameron, Alan. 1993. *The Greek Anthology. From Meleager to Planudes*. Oxford : Oxford University Press.
- Hayles, Katherine. 1999. *How we became posthuman : virtual bodies in cybernetics, literature, and informatics*. Chicago : University of Chicago Press.
- Hofstadter, Douglas R. 1999. *Gödel, Escher, Bach : an eternal golden braid*. 20th anniversary ed. New York : Basic Books.
- Kusner, Matt, Yu Sun, Nicholas Kolkin, et Kilian Weinberger. 2015. « From Word Embeddings To Document Distances », *Proceedings of Machine Learning Research*, n° 37:957-66.
- Laurens, Pierre. 2012. *L'abeille dans l'ambre : Célébration de l'épigramme de l'époque alexandrine à la fin de la Renaissance*. 2e édition revue et augmentée. Paris : Les Belles Lettres.
- Lévy, Pierre. 2023. « Calculer la sémantique avec le langage IEML ». *Humanités numériques*, n° 8 (décembre). <https://doi.org/10.4000/revuehn.3836>.

- Marino, Mark C. 2020. *Critical code studies : initial methods*. Software studies. Cambridge, Massachusetts : The MIT Press.
- Meunier, Jean-Guy. 2014. « Humanités numériques ou computationnelles : Enjeux herméneutiques ». *Sens Public*, décembre. <http://www.sens-public.org/article1121.html>.
- Meunier, Jean-Guy. 2017. « Humanités numériques et modélisation scientifique ». *Questions de communication*, n° 31 (septembre):19-48. <https://doi.org/10.4000/questionsdecommunication.11040>.
- Pennington, Jeffrey, Richard Socher, et Christopher D Manning. s. d. « GloVe : Global Vectors for Word Representation », 12.
- Pöckelmann, Marcus, Janis Dähne, Jörg Ritter, et Paul Molitor. 2020. « Fast paraphrase extraction in Ancient Greek literature ». *it - Information Technology* 62 (2):75-89. <https://doi.org/10.1515/itit-2019-0042>.
- Ramponi, Marco. 2022. « How ChatGPT actually works ». *News, Tutorials, AI Research*. <https://www.assemblyai.com/blog/how-chatgpt-actually-works/>.
- Rapaport, William J. 2013. « Meinongian Semantics and Artificial Intelligence ». *Humana.Mente* 6:25-52. <https://api.semanticscholar.org/CorpusID:7632759>.
- Russell, Stuart J., et Peter Norvig. 2021. *Artificial intelligence : a modern approach*. Fourth edition. Pearson series in artificial intelligence. Hoboken : Pearson.
- Salton, G., A. Wong, et C. S. Yang. 1975. « A vector space model for automatic indexing ». *Communications of the ACM* 18 (11):613-20. <https://doi.org/10.1145/361219.361220>.
- Schubert, Charlotte. 2020. « Intertextuality and Digital Humanities », *Information Technology*, 62 (2):53-59. <https://doi.org/https://doi.org/10.1515/itit-2019-0036>.
- Searle, John R. 1980. « Minds, brains, and programs ». *Behavioral and Brain Sciences* 3 (3):417-24. <https://doi.org/10.1017/S0140525X00005756>.
- Sterrett, Susan G. 1999. « Turing's Two Tests for Intelligence ». *Minds and Machines* 10 (4):541-59. <http://philsci-archive.pitt.edu/8480/>.
- Tarán, Sonya lidia. 1979. *The Art of Variation in the Hellenistic Epigram*. Vol. IX. Columbia studies in the classical tradition. Leyde : E. J. Brill.
- Tesler, Larry. s. d. « Tesler's Theorem and other adages and coinages ». *Larry Tesler*. <https://www.nomodes.com/larry-tesler-consulting/adages-and-coinages>.

- Tononi, Giulio. 2012. *Phi : a voyage from the brain to the soul*. New York : Pantheon. <http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=740038>.
- Tononi, Giulio. 2015. « Integrated information theory ». *Scholarpedia* 10 (1):4164. <https://doi.org/10.4249/scholarpedia.4164>.
- Tononi, Giulio, et Christof Koch. 2015. « Consciousness : here, there and everywhere? » *Philosophical Transactions of the Royal Society B : Biological Sciences* 370 (1668):20140167. <https://doi.org/10.1098/rstb.2014.0167>.
- Turing, A. M. 1950. « Computing Machinery and Intelligence ». *Mind*, New Series, 59 (236):433-60. <http://www.jstor.org/stable/2251299>.
- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, et Illia Polosukhin. 2017. « Attention is All You Need ». In. <https://arxiv.org/pdf/1706.03762.pdf>.
- Vitali-Rosati, Marcello. 2021. « Le fait numérique comme « conjonctures médiatrices » ». *Communication & Langages*, n° 208-209:155-70. <https://papyrus.bib.umontreal.ca/xmlui/handle/1866/25755>.
- Vitali-Rosati, Marcello, Margot Mellet, Servanne Monjour, Antoine Fauchié, Timothée Guicherd, David Larlet, et Enrico Agostini-Marchese. 2021. « L'épopée numérique de l'Anthologie grecque : entre questions épistémologiques, modèles techniques et dynamiques collaboratives ». *Sens public*, n° SP1596 (juillet). <http://sens-public.org/articles/1603/>.
- Vitali-Rosati, Marcello, Servanne Monjour, Joana Casenave, Elsa Bouchard, et Margot Mellet. 2020. « Editorializing the Greek Anthology : The palatin manuscript as a collective imaginary ». *Digital Humanities Quarterly* 014 (1). <http://www.digitalhumanities.org/dhq/vol/14/1/000447/000447.html>.
- Wolfe, Cary. 2010. *What is posthumanism?* Posthumanities series, v. 8. Minneapolis : University of Minnesota Press.