

Le robot, gardien de l'humanité. La Loi zéro dans les cycles des Robots et de Fondation d'Asimov.

Anthony VALLAT

Enseignant en sociologie et en philosophie
École Supérieure de la santé de Lausanne

Introduction

1.- L'importance d'Isaac Asimov. Je remercie les organisateurs de ce colloque sur les lois de la robotique d'Asimov et le Droit de m'avoir invité à présenter cette contribution. En effet, j'ai une histoire singulière avec Isaac Asimov. Encore lycéen, je me destinais aux études d'ingénieur en physique à l'EPFL, jusqu'à ce que je lise *Fondation*¹. Sous le choc d'une sorte de révélation, je me suis alors orienté vers des études de sociologie, espérant en apprendre davantage sur la psychohistoire. Hélas, à l'issue de ma licence, je n'ai pas été recruté par la Seconde Fondation. J'ai tenté de rembourser autrement ma dette à Asimov en publiant un essai en 2014 aux éditions ActuSF, *Utopie et Raison dans le Cycle Fondation d'Isaac Asimov*.

Il y a quelque chose de vraiment remarquable dans l'œuvre d'Asimov. C'est le point de départ de mes travaux sur Asimov. Si *Fondation* a été lu par des millions de lecteurs, c'est que cette œuvre a touché à quelque chose d'essentiel. Mon hypothèse est qu'à travers son œuvre, Asimov propose une pensée forte sur l'utopie, et que c'est une explication de son succès.

Cette communication va porter sur la question du robot gardien de l'humanité et la Loi zéro imaginée par Asimov. La voici : « Un robot ne doit causer aucun mal à l'humanité ou, faute d'intervenir, de permettre que l'humanité souffre d'un mal.² » Elle apparaît en 1985 dans

le roman *Les Robots et l'Empire*, pour compléter, ou dépasser, les trois Lois de la robotique. Pour rappel : « 1. Un robot ne peut porter atteinte à un être humain, ni, en restant passif, permettre qu'un être humain soit exposé au danger ; 2. Un robot doit obéir aux ordres qui lui sont donnés par un être humain, sauf si de tels ordres entrent en conflit avec la première loi ; 3.- Un robot doit protéger son existence tant que cette protection n'entre pas en conflit avec la première ou la deuxième loi. »

Je vais montrer comment la Loi zéro s'inscrit dans la dynamique de l'œuvre d'Asimov, et quels enjeux philosophiques sont interrogés (les robots, la raison, le langage, le dialogue, la prise de décision, l'humanité, l'utopie, ..).

I. Le Cycle de Fondation, une réflexion sur l'utopie

2.- Corpus sélectionné. Tout d'abord un mot sur la composition du Cycle de Fondation.

Notre lecture analytique sera basée sur le corpus suivant :

Fondation (*Foundation*, Gnome Press, 1951), volume qui inclut les nouvelles *Les Psychohistoriens* (1951), *Les Encyclopédistes* (1942), *Les Maires* (1942), *Les Marchands* (1944) et *Les Princes Marchands* (1944) ;

Fondation et Empire (*Foundation and Empire*, Gnome Press, 1952), volume qui inclut les nouvelles *Le Général* (1945) et *Le Mulet* (1945) ;

¹ *Fondation*, Gnome Press, 1951

² *Les Robots et l'Empire*, vol. 2, p.145, *Robots and Empire*, Doubleday, 1985

Seconde Fondation (*Second Foundation*, Gnome Press, 1953), volume qui inclut les nouvelles *La Quête du Mulet* (1948) et *La Quête de la Fondation* (1949) ;

le roman *Fondation Foudroyée* (*Foundation's Edge*, Doubleday, 1982) ;

le roman *Les Robots de l'Aube* (*Robots of Dawn*, Doubleday, 1983) ;

le roman *Les Robots et l'Empire* (*Robots and Empire*, Doubleday, 1985) ;

le roman *Terre et Fondation* (*Foundation and Earth*, Doubleday, 1986) ;

le roman *Prélude à Fondation* (*Prelude to Foundation*, Doubleday, 1988) ;

le roman *L'Aube de Fondation* (*Forward the Foundation*, Doubleday, 1992) ;

La présence de *Les Robots de l'Aube* et *Les Robots et l'Empire* peut surprendre, mais je montrerai en quoi leur apport au Cycle est décisif.

Pour toutes les citations, je me référerai aux deux volumes parus chez Denoël en 2006. Le premier, *Fondation*, regroupe les volumes *Fondation*, *Fondation et Empire* et *Seconde Fondation*. Le second, *Fondation Foudroyée*, regroupe *Fondation Foudroyée* et *Terre et Fondation*. Pour *Les Robots de l'Aube*, je me référerai aux deux volumes publiés en 1984 par J'ai Lu et, pour *Les Robots et l'Empire*, les deux volumes parus en 1986 chez le même éditeur.

3.- L'utopie comme projet littéraire d'Asimov.

Commençons par une citation remarquable.

« Le contrôle de soi et de la société a été abandonné au hasard, ou aux vagues tâtonnements de systèmes éthiques essentiellement basés sur l'intuition, l'inspiration et l'émotion. Il en résulte qu'aucune culture dont le coefficient de stabilité excède environ 55 % n'a jamais vu le jour, avec pour corollaire une

affreuse détresse humaine.³» (Isaac Asimov, *Seconde Fondation*).

Très tôt, Asimov a conscience des malheurs de l'humanité⁴. C'est ainsi qu'à travers le Cycle de Fondation, il va proposer une série d'utopies possibles, ou plutôt une série de paradigmes rendant possible l'utopie.

À titre d'exemple, décrivons ici le tout premier paradigme, présent dans le premier volume de Fondation. Tout d'abord, il faut qu'un chercheur particulièrement brillant achève l'élaboration d'une sociologie définitive, la fameuse psychohistoire⁵. Cette science a pour vocation de décrire et de prédire le plus précisément possible l'évolution d'une société, en calculant les probabilités qu'ont les différents avènements de se réaliser. Ensuite cette science pourrait ainsi prédire comment orienter l'histoire de l'humanité pour l'amener à l'utopie. Et finalement, il faudrait des hommes raisonnables et de bonne volonté pour utiliser cette science pour amener l'humanité à l'utopie. Ces trois éléments sont le point de départ de *Fondation*, un premier paradigme d'une utopie possible.

Bien entendu, ce n'est pas si simple, et Asimov va mettre en scène une série de coups de théâtre pour montrer que ce paradigme n'est pas suffisant. Dans les nouvelles et romans suivants, il va enrichir le premier paradigme, et proposer de nouveaux paradigmes décrivant les conditions de possibilités de l'utopie.

Je ne vais pas ici détailler l'ensemble des paradigmes qui se succèdent dans le Cycle de Fondation. Cela excéderait l'ambition de cette communication⁶. Pour cet exposé sur la Loi zéro, je vais débiter par l'examen de *Fondation Foudroyée*⁷.

³ I. Asimov, *Fondation*, coll. Lunes d'encre, Denoel, 2006, p. 773.

⁴ Notons que les premières nouvelles qui composent le Cycle de Fondation ont été rédigées pendant la Deuxième Guerre Mondiale.

⁵ Asimov reconnaît d'ailleurs dans l'article Psychohistoire (in recueil *Mais le Docteur est d'Or*, Presses Pocket, 1996, Gold, 1995), qu'il aurait été plus

approprié de nommer cette science la « psychosociologie », mais qu'à l'époque de ses vingt-deux ans, il était fasciné par les livres d'histoire.

⁶ On trouvera dans mon essai *Utopie et Raison dans le Cycle Fondation d'Isaac Asimov*, éditions ActuSF, 2014 une description analytique et critique des paradigmes liés à l'utopie dans le Cycle de Fondation.

⁷ *Foundation's Edge*, Doubleday, 1982.

4.- Fondation foudroyée. *Fondation Foudroyée* est le quatrième volume du Cycle, roman écrit après une pause de 29 ans depuis la parution de la trilogie initiale, et paru en 1982.

À la fin de *Fondation Foudroyée*, une utopie possible est proposée, l'utopie Gaia. Gaia est une entité planétaire, une planète où les humains sont dans une forme de communion. Ils communiquent entre eux par télépathie, et savent ce que les autres veulent ou pensent. S'installe ainsi une forme d'utopie, que l'on pourrait qualifier sommairement de démocratie directe absolue.

À la fin du roman, Gaia, par le truchement d'une entité nommée Joie, raconte aux personnages principaux qu'elle est structurée et contrainte par les trois Lois de la robotique. Nous nous retrouvons donc non pas avec un robot soumis aux trois Lois, mais toute une société planétaire soumise aux trois Lois.

Gaia explique ensuite qu'elle est devant un choix décisif. Soit rester une utopie locale, c'est-à-dire une utopie localisée sur une seule planète. Soit s'étendre dans toute la Galaxie et englober ainsi tous les humains de la Galaxie dans une communication télépathique les uns avec les autres, projet qu'elle nomme Galaxia. Gaia doit opter pour l'une ou l'autre alternative.

Gaia va demander à un personnage humain, Golan Trevize⁸, qui a été sélectionné au début du roman par Gaia, de choisir à sa place.

En effet, Gaia explique qu'elle ne peut pas procéder elle-même à cette décision, parce qu'elle est bloquée par les trois Lois. Voilà où on en est à la fin du roman. Et coup de théâtre final, on découvre que Gaia est contrôlée par des robots !

C'est inattendu ! Jusqu'alors il n'y avait jamais eu de robots dans le Cycle de Fondation, aucun robot depuis les premiers textes de 1941. Le cycle des robots et celui de Fondation étaient jusqu'alors totalement disjoints.

Nous apprenons donc que c'est à cause des trois Lois que les robots ne peuvent pas décider de l'avenir de l'humanité. Mais dans cette fin de roman, il n'y a pas de justifications ou d'explication du pourquoi les trois Lois empêchent la prise de décision par Gaia.

Asimov n'est visiblement pas satisfait de cette fin un peu abrupte. Il se propose de développer les différentes questions en suspens. Pour cela, il va rédiger deux nouveaux romans, *Les Robots de l'Aube*⁹ et *Les Robots et l'Empire*¹⁰.

Ces deux romans ont été publiés entre *Fondation Foudroyée* et *Terre et Fondation*¹¹. Pour l'anecdote, Asimov précise dans sa biographie¹² qu'il a préféré écrire *Les Robots de l'Aube* et *Les Robots et l'Empire* que *Fondation Foudroyée* ou *Terre et Fondation*.

5.- Les Robots de l'Aube. Posons le cadre des *Robots de l'Aube*¹³.

L'action se situe plusieurs milliers d'années avant la Chute de l'Empire, quelques années après *Les Cavernes d'acier*¹⁴ et *Face aux feux du soleil*¹⁵. Il s'agit de la troisième enquête de l'inspecteur terrien Elijah Baley, accompagné pour la troisième fois par le robot Daneel. Le cadre est le début de l'expansion spatiale humaine. On trouve d'un côté une Terre surpeuplée, des milliards d'habitants, et de l'autre cinquante planètes colonisées et terraformées. Y habitent les Spatiaux, peu nombreux, qui vivent jusqu'à quatre cents ans. Ils sont assistés par une multitude de robots soumis aux trois Lois. Les Terriens n'ont ni la possibilité, ni le droit d'émigrer sur ces

⁸ Golan Trévize est présenté comme un individu qui pratique une très fine épistémologie, basée sur un exercice rigoureux de la Raison.

⁹ *Robots of Dawn*, Doubleday, 1983.

¹⁰ *Robots and Empire*, Doubleday, 1985.

¹¹ *Foundation and Earth*, Doubleday, 1986.

¹² *Moi*, Asimov, Denoël, 1996, I, Asimov, Doubleday, New York, 1994.

¹³ Pour toutes les citations, nous nous sommes référés aux éditions suivantes : *Les Robots de l'Aube*, paru en 1984 chez J'ai Lu en deux volumes (n° 1602 et n° 1603,) et *Les Robots et l'Empire*, paru en 1986 chez J'ai Lu en deux volumes (n° 1996 et n° 1997).

¹⁴ *The Caves of Steel*, Doubleday, New York, 1954.

¹⁵ *The Naked Sun*, Doubleday, New York, 1957.

cinquante mondes. Ils n'ont pas non plus les moyens de coloniser de nouvelles planètes. Précisons qu'il n'y a pas de robots positroniques sur Terre à ce moment-là.

Les Robots de l'Aube raconte comment Elijah Baley parvient, grâce à son enquête, à libérer les Terriens de l'isolement et ainsi à déclencher la seconde vague de colonisations extrasolaires. Il est aidé par les robots R. Daneel et R. Giskard. La révélation finale est la suivante : Baley découvre, par son seul raisonnement, que R. Giskard a développé des pouvoirs mentaux télépathiques. Il est capable de lire les émotions et pensées des humains et des robots, et est capable de les modifier. Ce pouvoir est arrivé suite à un accident technologique purement fortuit¹⁶.

6.- Les Robots et l'Empire. *Les Robots et l'Empire* se situe deux cents ans plus tard. Elijah Baley est mort, mais les robots Daneel et Giskard sont encore en activité. Il s'agit de l'un des rares récits d'Asimov où le lecteur accède au récit par le point de vue des robots.

Daneel et Giskard découvrent un complot des Spatiaux qui ont prévu d'isoler la Terre et de laisser les milliards de Terriens végéter dans leurs cités gigantesques en leur interdisant d'accéder à l'espace extrasolaire.

Ce complot est déjoué. Les Terriens obtiennent finalement la suprématie territoriale sur la Galaxie, et les Spatiaux se retrouvent isolés et voués au déclin. Avant de mourir, Giskard transmet ses pouvoirs à Daneel qui devra mettre en œuvre les projets utopiques qu'ils ont tous deux conçus pour l'humanité.

II. - Utopie et Lois de la robotique

7.- Les mondes des Spatiaux, des utopies grâce aux robots ? C'est par l'élaboration des mondes des Spatiaux qu'Asimov parvient à relier les thèmes de l'utopie et des trois Lois de

la robotique. Examinons comment cela se présente.

Les mondes Spatiaux, planètes terraformées pour être idylliques, sont des sortes d'utopie. Par exemple, voilà comment est présentée Aurora dans les *Les Robots de l'Aube* par le Dr Fastofle, le constructeur de Daneel et Giskard. :

« Le résultat est un monde qui produit tout ce dont nous avons besoin, en tenant compte de nos désirs... Voulez-vous que je vous dise vers quel idéal nous avons tendu ?

– Je vous en prie, dit Baley.

– Nous avons travaillé pour créer une planète qui, dans son ensemble, obéirait aux trois Lois de la robotique. Elle ne fait rien qui blesse les êtres humains, par action ou omission. Elle fait ce que nous voulons qu'elle fasse, du moment que nous ne lui demandons pas de faire du mal à des êtres humains. Et elle se protège, à des moments et dans des lieux où elle doit nous servir ou nous sauver même au prix d'un mal fait à elle-même. »¹⁷

Fastofle présente donc Aurora comme une planète obéissant aux trois Lois. Qu'est-ce que cela veut dire ?

Il y a un grand nombre de robots sur les planètes des Spatiaux. Les robots sont littéralement formatés par les trois Lois de la Robotique auxquelles ils ne peuvent échapper. Elles forment une sorte de devoir, un impératif. Quelle activité exerce un robot ? Il doit se préoccuper des êtres humains qui sont ontologiquement proches de lui. En vertu de la première Loi¹⁸, les robots sont programmés pour chercher à établir le bien-être des humains qui sont à proximité d'eux, et faire en sorte que rien ne leur arrive de mal. Et donc chaque humain entouré d'un grand nombre de robots est enveloppé dans une bulle de confort et de sécurité.

¹⁶ Le lecteur apprend cette révélation à travers une tranquille causerie entre Giskard et Baley (p. 246, volume 2). Comme souvent chez Asimov, nombre de scènes de révélations se passent dans un dialogue paisible, entre gens de bonne compagnie, confortablement assis et ravitaillés.

¹⁷ *Les Robots de l'Aube*, vol. 1, p. 136-137.

¹⁸ Pour rappel: un robot ne peut porter atteinte à un être humain, ni, restant passif, permettre qu'un être humain soit exposé au danger.

Cette bulle correspond à une définition minimale de l'utopie. Au minimum, on peut définir l'utopie comme une société dont les membres subissent le moins de mal possible. Chaque Spatial est donc au centre d'une micro-utopie locale. Toutes ces bulles additionnées transforment ces planètes en conglomérat de bulles d'utopies.

8.- De l'utopie locale à l'utopie globale. Les utopies générées par les robots ne sont que des utopies locales, elles ne s'étendent qu'aux humains en contact avec eux. Cela ne pose pas de problème dans les mondes spatiaux ; avec une forte densité de robots associée avec une faible densité d'humains, cette superposition des utopies individuelles englobe toute la planète, comme le dit plus haut le Dr Fastofle¹⁹. Mais ces utopies n'englobent qu'une fraction de l'humanité, celle susceptible d'être sous la surveillance bienveillante des robots.

Mais voilà qu'une rupture se produit dans cette architecture de l'effet utopique du confort social amené par les robots. C'est cette rupture qui va créer les conditions d'apparition de la Loi zéro.

Pour les robots, la question se pose en permanence : comment, par des actions ou des inactions, prévenir au mieux les dangers qui menacent les hommes qui sont à leur contact ?

Les enjeux changent lorsque le rayon d'action des robots est modifié. C'est ce qui est raconté dans *Les Robots de l'Aube*. Suite à une involontaire modification technique, le robot Giskard se retrouve doté de pouvoirs mentaux de type télépathique : il peut lire les pensées et émotions des humains et les modifier. Son rayon d'action augmente. Par rapport à la première Loi, les humains dont il doit s'occuper sont beaucoup plus nombreux que s'il n'était pas télépathe. Il a désormais les moyens d'agir sur beaucoup d'êtres humains, et donc a beaucoup plus de gens sous sa responsabilité.

Donc à lui se pose le problème de l'extension de l'utopie locale qu'il peut générer.

Voilà un enjeu au cœur de ces deux romans, *Les Robots de l'Aube* et *les Robots et l'Empire* : la transformation d'une addition d'utopies locales en une utopie englobant tous les individus.

Voyons comment cela se déroule.

Dans *Les Robots de l'Aube*, une menace apparaît à laquelle Giskard est obligé de faire face. Il découvre par télépathie un complot visant à empêcher les huit milliards de Terriens d'entamer la colonisation de nouvelles planètes, complot dirigé par le Docteur Amadiro, politicien et ingénieur de haut vol.

Giskard découvre ce plan. Avec son potentiel d'action dû à ses pouvoirs, et l'impératif de protection des humains, Giskard est contraint de se prononcer. Doit-il agir ou laisser faire ? Le projet d'Amadiro apporte-t-il un mieux pour les humains en général, ou est-il néfaste ? S'il est certes positif pour les Spatiaux, il est cependant négatif pour les Terriens. Mais quelle partie de l'humanité compte le plus ?

9.- Robots et prise de décision. Giskard doit choisir entre l'inaction - laisser les Spatiaux développer des utopies locales à travers la Galaxie, avec des milliards de Terriens végétant dans une forme de dystopie -, ou l'action - chercher à établir une utopie globale pour toute l'humanité. Il décide de partager cette responsabilité avec R. Daneel.

Les romans *Les Robots de l'Aube* et *Les Robots et l'Empire* racontent leur quête des meilleures décisions à prendre pour l'humanité.

Ainsi apprend-on, à la fin de *Les Robots de l'Aube*, que Giskard a réussi à prendre les bonnes décisions, et à empêcher Amadiro de mettre son plan à exécution. Et grâce également aux manœuvres d'Elijah Baley, les Terriens vont réussir à lancer une vague de colonisations.

Les Robots et l'Empire se déroule quelque deux cents années après *Les Robots de l'Aube*. Les nouvelles colonies fondées par les Terriens menacent la suprématie des Spatiaux et

¹⁹ Dans *Terre et Fondation*, les protagonistes visitent quelques-uns des monde spatiaux des milliers

d'années plus tard, et le thème de l'utopie locale y sera à nouveau interrogé.

Amadiro (encore lui !) ourdit un vil dessein, celui de tuer les milliards d'habitants de la Terre en la rendant radioactive. Daneel et Giskard vont partiellement déjouer ce plan. Giskard décide tout de même de rendre la Terre radioactive, mais très progressivement, pour laisser cent cinquante ans aux Terriens pour partir dans l'espace. Sa motivation est que les humains se détachent de la Terre, et donc que l'expansion spatiale se fasse sans retour en arrière. Cette décision prise, Giskard meurt et transmet ses pouvoirs télépathiques à Daneel, désormais chargé d'amener les humains à l'Utopie.

10.- Mortelle prise de décision. La fin est très forte : Giskard meurt *parce qu'il a pris la décision de rendre la Terre radioactive en prenant consciemment le risque, certes très faible, de faire mourir des millions d'humains. Il meurt parce qu'il n'est pas sûr de sa décision.* Son cerveau positronique ne survit pas à une telle violation des Lois de la robotique.

Comment les deux robots en sont-ils arrivés à prendre ces décisions si difficiles ? Pour réussir cela, ils suivent tout un cheminement, une vraie quête intellectuelle. Le ressort central de ces deux récits se situe là.

Dans les quatre premiers volumes du *Cycle de Fondation*, certains êtres humains réussissent à prendre des décisions importantes pour la destinée de l'humanité, que ce soit Hari Seldon, Salvor Hardin, Preem Palver ou Golan Trevize, sans que leur cerveau se paralyse.

Par contre, lorsque Giskard prend la décision finale, soit laisser la Terre devenir radioactive pour que l'Utopie s'installe des milliers d'années plus tard, il en meurt !

Notre hypothèse est qu'Asimov révèle ainsi une différence fondamentale entre humains et robots.

11.- Le devoir de Giskard : un dilemme : respecter les trois Lois ou faire ce qu'il faut faire. Au départ, Giskard est confronté à une forme d'impasse : la meilleure chose à faire

entre parfois en contradiction avec les trois Lois. S'il choisit de faire ce qu'il faut faire, et donc de combattre le plan du Docteur Amadiro, il occasionnera fatalement du mal à des humains, en l'espèce Amadiro, et donc Giskard contreviendrait aux trois Lois.

On assiste à ses réflexions pour sortir de cette impasse.

La notion de « devoir » est un premier élément de la différence entre robots et humains. Cela se révèle dans une conversation²⁰ entre Giskard et Elijah Baley. Pour Giskard, le devoir est « ce qui doit être fait ». Elijah Baley lui répond que pour les robots, le devoir est d'obéir aux Lois de la robotique. Il ajoute que les êtres humains ont aussi des lois auxquelles ils doivent obéir. Giskard pose ensuite un constat et une question fondamentale : « Monsieur, je ne peux pas contrevioler aux Lois. Pouvez-vous désobéir aux vôtres ? »²¹

Réponse de Baley : « Je peux choisir de ne pas faire mon devoir, mais je ne le choisis pas, et c'est parfois la pulsion la plus forte. »

Les robots sont donc obligés d'obéir aux trois Lois, alors que les humains peuvent choisir de ne pas obéir à leurs propres lois.

Le problème pour Giskard est d'harmoniser le « faire ce qu'il faut » avec son devoir, son obéissance inconditionnelle aux trois Lois. Pour ce faire, il ne dispose que de la faculté de raison.

12.- Solution : harmoniser le respect des trois Lois et faire ce qu'il faut faire. La première étape de Giskard est de définir « ce qu'il faut faire ».

Cette étape, autorisée par les trois Lois, est de réfléchir à ce qui peut être le meilleur pour l'humanité. Ce projet utopique est un raisonnement, une construction intellectuelle. Tant qu'il ne prend pas de décision au sujet de son application, Giskard est autorisé à simplement y réfléchir.

²⁰ Comme souvent chez Asimov, les représentations mentales des personnages se révèlent à travers les dialogues (*Les Robots de l'Aube*, vol. 1, p. 141).

²¹ Op. cit., vol. 1, p. 142.

Il dévoile son projet à Elijah Baley à la fin des *Robots de l'Aube*. Dans un premier temps, laisser les Terriens coloniser la Galaxie, sans robots pour les assister. Cela engendrera quantité de maux de toutes natures, mais « à la longue, à la fin, les êtres humains bénéficieront d'avoir travaillé par eux-mêmes²² ». Dans un deuxième temps, « peut-être un jour – un jour lointain dans l'avenir – les robots pourront de nouveau intervenir ».

Intervenir, oui, mais comment ? En utilisant la psychohistoire.

Pour Giskard, cette science sociale sera nécessaire un jour. Il la croit possible, car il pense qu'il existe des lois gouvernant le comportement humain, sur le modèle des trois Lois de la robotique, mais différente parce qu'« elles peuvent être si peu contraignantes qu'elles n'ont guère de sens, à moins que ces énormes populations ignorent le fonctionnement de ces lois²³ ». Dans le roman *Les Robots et l'Empire*, Giskard poursuit l'élaboration de cette psychohistoire en dialoguant avec Daneel.

Voilà qui justifie la lecture des *Robots de l'Aube*. Asimov y tisse des fils narratifs et thématiques qui rejoignent les révélations de *Fondation Foudroyée*, en montrant comment les robots ont amorcé le projet utopique, ici le Plan Seldon.

13.- Robots et prison des trois Lois. Mais Giskard est toujours devant son impasse : certes il a le droit de réfléchir à « ce qu'il faut faire », mais la mise en application de ce projet implique des décisions impossibles à prendre à cause des trois Lois. Cela sera l'enjeu du roman *Les Robots et l'Empire*.

Cette impasse se résoudra d'abord par l'emploi d'une faculté commune aux robots et aux humains : la Raison. Les deux robots emploieront leur raison, à travers de longs dialogues entre eux, qui se révèlent indispensables pour qu'ils puissent avancer dans leurs constructions intellectuelles. Cela montre l'importance qu'Asimov attribue au

langage dans le processus de clarification de la pensée.

Au début du roman, suite à une modification mentale qu'il a dû opérer sur Gladia (la Spatiale héroïne de *Face aux feux du soleil* et amante de Baley dans *Les Robots de l'Aube*), Giskard déclare que les trois Lois sont incomplètes. Il lui apparaît que la modification de l'esprit de Gladia « était la meilleure réponse possible dans la limite des trois Lois, mais pas la meilleure réponse possible dans l'absolu²⁴ ».

À la question de Daneel sur cette meilleure réponse possible, Giskard affirme qu'il ne peut l'exprimer tant qu'il est lié par les trois Lois. Et quand Daneel lui affirme qu'il n'y a rien au-dessus des trois Lois, Giskard met en avant une autre différence fondamentale entre humains et robots : « Si j'étais humain, je parviendrais à voir au delà des Lois. »

Selon Giskard, la solution serait de modifier les trois Lois. Mais « je ne peux les modifier – parce qu'elles se trouvent sur mon chemin²⁵ ».

Giskard est structuré par un paradigme, les trois Lois, qui surdéterminent son cerveau. Il doit donc utiliser sa raison pour redéfinir ce paradigme. Cela devient impératif lorsqu'il détecte dans l'esprit de Mandamus, un comparse d'Amadiro, l'existence d'un complot visant à tuer des milliards d'humains. Hélas pour connaître les détails du complot, il devrait endommager son esprit. Giskard ne peut pas le faire, prisonnier des trois Lois. À ce stade du récit, Giskard et Daneel sont au désespoir. Plus tard, ils débarquent sur Solaria, accompagnant Gladia. Ils seront attaqués par des robots. Grâce à ses pouvoirs télépathiques, Giskard triomphe sans peine.

Giskard et Daneel y font une découverte cruciale : la possibilité de redéfinir le paradigme des trois Lois. En effet, les robots Solariens qui les ont attaqués ont été programmés pour ne considérer comme humains que les natifs de Solaria.

²² Op. cit., vol. 2 p. 253.

²³ Op. cit., vol. 2 p. 253.

²⁴ *Les Robots et l'Empire*, vol. 1, p. 23.

²⁵ Op. cit., vol. 1, p. 24.

C'est Daneel qui avance le raisonnement : « *Si*²⁶ l'on pouvait cerner les instructions par des définitions et des instructions, *si* l'on pouvait détailler les instructions d'une manière suffisamment rigoureuse, *serait-il* aussi possible de tuer un être humain pour une raison sérieuse que de sauver la vie d'un autre être humain ? »²⁷

Giskard répond que c'est possible. Daneel s'appuie là-dessus pour poursuivre son raisonnement. Il en déduit alors que les trois Lois ne sont pas absolues. Elles pourraient être modifiées.

Giskard supplie Daneel ne pas aller plus loin.

14.- Emploi de la Raison et danger de mort. En effet, en raisonnant sur les trois Lois, Giskard et Daneel sont en danger de mort ! Leur cerveau positronique peut se paralyser. Pour eux, robots, utiliser la raison est une source de danger mortel : une construction mentale mal étayée, et ils meurent ! Voilà pourquoi ils ont besoin d'être deux à raisonner. Ce n'est pas le cas pour les êtres humains, qui peuvent se tromper sans risquer la mort.

« Encore un pas, ami Giskard. Le camarade Elijah l'aurait franchi.

– C'était un être humain. Il le pouvait. »²⁸

Les humains peuvent sans risque remettre en question les principes qui structurent le comportement. Encore une nouvelle différence entre humains et robots !

Et Daneel brise la prison des trois Lois en faisant un pas supplémentaire : « Ne serait-il pas possible que nous-même, *peut-être*, dans certaines conditions précises, puissions mod... »²⁹

Ici, Daneel s'arrête pour éviter la paralysie mortelle.

III. La Loi zéro : une transgression des trois Lois au service de l'humanité

15.- Loi zéro et Humanité. Voici maintenant un passage clé de *Les Robots et l'Empire*, qui va aboutir à la proclamation de la Loi zéro.

Plus tard dans le roman, Daneel sera contraint de poursuivre son raisonnement sous peine que lui et Giskard soient détruits, et que l'humanité sorte du long chemin vers l'utopie.

Giskard est confronté à Vassilia, fille du Dr Falstofle, roboticienne et comparse d'Amadiro. Celle-ci s'est enfermée dans une pièce avec nos deux robots. Nous assistons à un combat de langage et de raison. Celui qui argumentera le mieux vaincra. Si Daneel n'argumente pas suffisamment bien, lui et Giskard mourront et l'humanité n'accédera pas à l'utopie.

L'enjeu du combat sera de convaincre Giskard. En effet, Vassilia donne l'ordre à Giskard d'anéantir l'esprit de Daneel. Giskard est prisonnier de la deuxième Loi³⁰, qui l'oblige à obéir à un humain si la première Loi³¹ n'est pas en cause. C'est le cas ici : aucun être humain n'est menacé. Giskard doit obéir, prisonnier des trois Lois. Daneel doit donc modifier les trois Lois qui structurent Giskard. Il doit le faire par un raisonnement, de telle sorte que Giskard puisse s'automodifier en suivant ce raisonnement.

Vassilia lui donne l'ordre de se taire, s'appuyant également sur la deuxième Loi.

Daneel désobéit en invoquant la Loi zéro, une Loi qui supprime la première Loi, et donc aussi la deuxième.

Daneel décrit le cheminement de son raisonnement. Il commence par constater l'incomplétude des trois Lois. Selon lui, cela vient de ce que l'humanité est plus importante qu'un seul être humain. Il énonce alors la Loi

²⁶ Les italiques sont ceux du texte original.

²⁷ Op. cit., vol. 1, p. 214.

²⁸ Op. cit., vol. 1, p. 215.

²⁹ Op. cit., vol. 1, p. 215.

³⁰ Pour rappel, « un robot doit obéir aux ordres qui lui sont donnés par un être humain, sauf si de tels ordres entrent en conflit avec la première loi ».

³¹ Pour rappel, « un robot ne peut porter atteinte à un être humain, ni, en restant passif, permettre qu'un être humain soit exposé au danger ».

zéro : « Un robot ne doit causer aucun mal à l'humanité ou, faute d'intervenir, de permettre que l'humanité souffre d'un mal. »³² Pour étayer cela, Daneel s'appuie sur une métaphore qu'Elijah Baley lui a confiée avant de mourir : « Daneel, garde l'esprit fermement fixé sur la tapisserie et ne te laisse pas affecter parce qu'un fil s'estompe. Il y en a tant d'autres, tous précieux, chacun jouant son rôle... »³³

Daneel affirme donc que « la tapisserie de la vie est plus importante qu'un seul fil³⁴ ».

16.- L'humanité est-elle un concept abstrait ?

À cette question, Vassilia expose une objection majeure : l'humanité n'est qu'un concept, une abstraction ; elle n'a pas de valeur ontologique.

« Les trois Lois de la Robotique concernent les êtres humains en tant qu'individus et les robots en tant qu'individus-robots. Mais qu'est-ce que « l'humanité » sinon une abstraction ? Peux-tu toucher l'humanité ? Tu peux blesser ou éviter de blesser un être individuel et comprendre le préjudice ou l'absence de préjudice. Peux-tu voir le préjudice ou l'absence de préjudice causé à l'humanité ? Peux-tu le comprendre ? Peux-tu le montrer du doigt ? »³⁵

Elle demande ensuite à Giskard d'arbitrer le débat. Celui-ci confirme la validité de l'argument de Vassilia. Il se justifie en invoquant l'histoire de l'humanité. De grands crimes ont souvent été accomplis en se justifiant par ce concept d'humanité. Et c'est parce que l'humanité est une abstraction qu'on peut si aisément en appeler à elle pour tout justifier.

C'est donc une victoire dans ce débat pour Vassilia. Daneel tente un ultime argument pour convaincre Giskard, en lui demandant s'il est conscient des conséquences de son choix. Giskard confirme qu'il l'est, mais que « le danger dont tu parles ne constitue pas une certitude, mais découle d'une déduction. Et

l'on ne peut fonder nos actions au mépris des trois Lois pour autant³⁶ ».

Giskard est toujours prisonnier des trois Lois, et c'est le juste exercice de sa raison qui le retient prisonnier.

Vassilia fait appel à quatre robots et leur ordonne de démanteler Daneel. Et surprise, Giskard agit. Il paralyse les robots, endort Vassilia et efface une partie de sa mémoire.

Plus tard, en sécurité, Giskard et Daneel reviennent sur le débat.

Giskard s'explique. Il n'a pas été convaincu par le raisonnement de Daneel, mais le raisonnement a offert assez d'éléments convaincants pour réinterpréter la Première Loi, lui faisant bénéficier d'une marge de manœuvre.

Ils en déduisent que « penser suffisamment » à ce concept de Loi zéro peut affaiblir la Première Loi. Il y a donc une certaine force de la raison.

Pourtant, tous deux voient le manque de présence ontologique de l'Humanité comme un obstacle, qui leur interdit de s'affranchir définitivement de la Première Loi.

Cependant, ils persévèrent : « C'est là la difficulté, ami Daneel, avec ta Loi zéro. La Première Loi traite d'individus spécifiques et de certitudes. Ta Loi zéro traite de groupes abstraits et de probabilités. »³⁷

Daneel fait appel à l'expérience de Giskard : « Ne peut-on imaginer qu'il existe dans la Galaxie le bourdonnement de l'activité mentale de l'humanité tout entière ? Comment, dans ce cas, l'humanité serait-elle une abstraction ? C'est quelque chose de palpable. Rapproche cela de la Loi zéro et tu verras que l'extension des Lois de la Robotique est justifiée... justifiée par ta propre expérience. »³⁸

Ici, Daneel fait référence à ce qu'a ressenti Giskard pendant un discours de Gladia,

³² Op. cit., vol. 2, p.145.

³³ Op. cit., vol. 1, p. 276.

³⁴ Op. cit., vol. 2, p.145.

³⁵ Op. cit., vol. 2, p. 145.

³⁶ Op. cit., vol. 2, p. 146.

³⁷ Op. cit., p. 192.

³⁸ Op. cit., p. 194.

lorsqu'il a senti, grâce à ses pouvoirs mentaux, l'activité de la foule tout entière.

Giskard accepte le raisonnement. Encore une fois, un robot a eu besoin de la confirmation d'un autre robot pour accepter un bouleversement des principes. Daneel a désormais modifié son paradigme structurant en y incorporant la Loi zéro comme loi suprême. Cela se révèle dans le récit lorsqu'il choisit de protéger Giskard plutôt que Gladia lors d'un attentat.

Nous assistons au triomphe de la raison chez Daneel et Giskard, qui se sont auto-modifiés grâce à son usage rigoureux. Et cela grâce également au langage et au dialogue. En effet, le langage naturel a comme particularité fondamentale que les mots offrent une latitude dans l'interprétation de leurs définitions, contrairement au langage mathématique. Et c'est le dialogue qui leur permet de coconstruire les définitions qui permettent de transgresser les trois Lois. Je développe dans mon essai pourquoi le style d'Asimov, parfois décrié, est pourtant fidèle et cohérent à la philosophie du langage esquissée ci-dessus.

17.- Pour que tu t'y intéresses. Pour être complet sur cette thématique de la fécondité du dialogue, signalons la nouvelle *Pour que tu t'y intéresses* (*That thou art mindfull of him*), parue en 1974. Elle présente l'intérêt de faire un lien entre les premières nouvelles de robots et *Les Robots de l'Aube*. La problématique de la marge de manœuvre des robots au sujet de l'interprétation d'une des Lois de la Robotique y est déjà évoquée. Cette marge de manœuvre est ici conceptualisée comme la capacité de « jugement ». Le directeur de U.S. Robots, successeur de Susan Calvin, est face à un problème : un robot aura beaucoup de difficulté à obéir à la deuxième Loi s'il est confronté à un grand nombre d'êtres humains susceptibles de lui donner des ordres contradictoires. Il va ainsi demander à son plus récent robot, George 10, de réfléchir à une solution technique. Ce dernier demande à être accompagné dans sa recherche par un autre robot, George 9, pour prévenir la paralysie

mortelle par le dialogue rationnel. Ici, les deux George préfigurent en quelque sorte Daneel et Giskard.

Dans ce récit, y est également évoqué le thème de la redéfinition par les robots de ce qu'est un être humain, avec comme conséquence que les deux George se considèrent finalement comme supérieurs aux humains et se permettent d'envisager d'être les futurs maîtres de l'histoire humaine. Ainsi cette nouvelle présente déjà une réflexion d'Asimov sur des thématiques plus amplement approfondies par la suite.

18.- Application de la Loi zéro, une voie vers l'utopie globale. Une fois la Loi zéro acceptée, reste le problème de l'application de cette Loi. Désormais, les deux robots sont contraints d'installer l'humanité dans une utopie globale.

C'est ainsi que lors de la confrontation finale, Giskard devra procéder à des décisions critiques, hélas, sans psychohistoire à disposition.

Les deux robots se retrouvent sur Terre face à Amadiro. Celui-ci s'appête à rendre la Terre mortellement radioactive, provoquant des milliards de morts. Il est neutralisé par Giskard. Giskard prend ensuite la première décision qui mettra l'humanité sur la voie de l'utopie. Il décide que cette radioactivité mortelle mettra plutôt cent cinquante ans à s'installer. Cela laissera le temps aux Terriens de partir dans l'espace et d'abandonner la Terre. Mais cette décision entraîne la mort de Giskard. Il meurt parce qu'il y a des incertitudes à la base de sa décision. Il meurt parce que la raison l'exige. Sa décision n'est pas complètement justifiée par des arguments formellement inattaquables, car il n'a pas la psychohistoire à sa disposition.

Avant de mourir, il transmet ses pouvoirs à Daneel.

Et Daneel se retrouve seul en charge de la responsabilité de l'humanité ; une solitude et une responsabilité écrasantes. Je ne dévoilerai pas ici ce qui adviendra de Daneel dans les romans suivants, *Terre et Fondation*³⁹, *Prélude à*

³⁹ *Foundation and Earth*, Doubleday, 1986

*Fondation*⁴⁰ et *Aube de Fondation*⁴¹, afin de laisser au lecteur le plaisir de la découverte des dernières réflexions d'Asimov.

19.- Différence entre humains et robots, et éthique kantienne. Les humains ordinaires ne sont pas comme les robots. Si les humains prennent une mauvaise décision, leur cerveau ne se paralyse pas, ils n'en meurent pas. Alors que les robots, s'ils prennent une mauvaise décision, qui trahit une Loi de la robotique, leur cerveau se fige, et ils meurent.

La philosophie de Kant offre des concepts qui permettent de décrire cela⁴². Kant partage l'esprit en deux espaces différents⁴³. Il y a l'espace phénoménal, soumis à la loi de la causalité et sensible aux données empiriques. Et l'espace nouménal qui transcende l'espace phénoménal, et est totalement indépendant des données empiriques. C'est là que s'exerce l'autonomie de la volonté, de manière inconditionnelle. La volonté peut y exercer la prise de décision en toute liberté, hors de toute la loi causale qui s'exerce dans le monde phénoménal, par exemple les humeurs, sensations ou sentiments. Dans cet espace, c'est la bonne volonté qui est responsable de la justesse de la décision éthique. L'esprit peut donc choisir entre deux principes ou deux valeurs en toute liberté. Kant considère que la nature fondamentale de l'homme est sa liberté de s'imposer des devoirs.

Dans les cycles des *Robots* et de *Fondation*, les humains jouissent de cet espace nouménal transcendantal, comme on l'a vu dans les dialogues entre Baley et Giskard et entre Giskard et Daneel. Ils peuvent désobéir aux lois ou à la morale. Et ils peuvent obstinément persister à vivre dans des sociétés où règne une forte détresse humaine.

⁴⁰ *Prelude to Foundation*, Doubleday, 1988

⁴¹ *Forward the Foundation*, Doubleday, 1992

⁴² Pour ce commentaire, je me réfère au vocabulaire et à la description proposés par Jeanne Hersch dans *L'Étonnement philosophique*, 1981, coll. folio/essais, Gallimard. Les pages qu'elle consacre à l'éthique kantienne résonnent profondément avec le dispositif spéculatif que sont les robots asimoviens.

⁴³ Notre exposé sur Kant peut paraître lapidaire et simplifié, eu égard à la complexité de la philosophie de

Alors que les robots sont prisonniers d'une réalité phénoménale qui leur impose une loi causale, un déterminisme. D'une part leur cerveau est structuré ontologiquement par les Lois de la robotique. D'autre part l'exigence de l'exercice juste de la raison est aussi inscrite ontologiquement dans leur cerveau, car tout manquement entraînerait leur mort. C'est cette combinaison qui conduit Daneel et Giskard à la proclamation de la Loi zéro, et au devoir impératif d'amener l'humanité à l'utopie⁴⁴.

Faudra-t-il nécessairement que surviennent les robots asimoviens pour que les humains aient la moindre chance de vivre dans une utopie globale ?

Conclusion

20.- Robots, Utopie, Raison et Science. À travers cet exposé, nous avons décrit comment Asimov met en scène la Loi zéro, et les enjeux philosophiques qui y sont liés, à travers les concepts de la raison, le langage, le dialogue, la prise de décision, l'humanité, l'utopie, et cet espace de liberté intérieure qui semble nous différencier des robots.

Dans les deux romans examinés, l'utopie est définie dans la Loi zéro : « Un robot ne doit causer aucun mal à l'humanité ou, faute d'intervenir, de permettre que l'humanité souffre d'un mal. »

L'utopie, c'est de ne souffrir d'aucun mal. Vision pour le moins sèche et désenchantée !

Quelle est la part de la raison et de la science dans ce paradigme de l'utopie proposé ici ?

Daneel et Giskard ne peuvent survivre que s'ils parviennent à concilier complètement un paradigme idéal, la Loi zéro, avec l'humanité réelle. Le danger mortel d'un mauvais exercice

Kant, mais une démonstration plus approfondie excéderait les ambitions de cette communication.

⁴⁴ Il serait intéressant de décrire aussi les contraintes des robots par les concepts d'éthique de la responsabilité et d'éthique de la conviction, tels que développés par Max Weber. Les robots semblent devoir assumer ces deux éthiques de façon absolue dans chacun de leurs actes.

de la raison les guette toujours. Une décision mal justifiée sur le devenir de l'humanité, et c'est la mort.

Respecter la Loi zéro sans mourir implique un instrument pour peser le coût humain des actions, une science dure qui guide les décisions ; une sociologie dure, la psychohistoire. Dans ce paradigme, la science est la seule voie de survie des robots. En effet, l'exigence de la juste conduite de la raison implique une attitude scientifique, une démarche qui doit faire correspondre au mieux les modèles et le réel. Sans elle, ils meurent.

Dans cette œuvre, nous ne sommes pas dans le *sense of wonder*. Ici, aucune mystique qui puisse faire rêver un lecteur en mal de spiritualité, nuls voyages sur des mondes extraordinaires foisonnants d'une vie extraterrestre luxuriante. Mais nous assistons à l'opiniâtre travail intellectuel de deux esprits aux prises avec des principes et des concepts, cherchant une assise ontologique à leurs constructions mentales. L'action rebondit principalement grâce aux dialogues, où les arguments sont potentiellement mortels.

Et, je crois qu'on reconnaît là une valeur centrale qui a toujours guidé Isaac Asimov : la quête du savoir et de la raison.

Par la voix d'Hari Seldon, peut-être Asimov a-t-il livré la clé d'une vie vouée à la recherche du meilleur pour l'homme.

« Pourquoi, se demanda-t-il, tant de gens passent-ils leur vie à éviter de trouver les réponses aux questions - et en premier lieu à éviter les questions ? N'y a-t-il pourtant rien de plus excitant que de chercher des réponses ?⁴⁵»

A. V.

⁴⁵ *Prélude à Fondation*, p. 260, *Prelude to Foundation*, Doubleday, 1988.