Pour un TAL responsable

Maxime Amblard

Pour un TAL responsable

Maxime Amblard*

* *LORIA, UMR 7503, Université de Lorraine, CNRS, Inria*
Campus scientifique F54506 Vandœuvre-lès-Nancy Cedex
maxime.amblard@univ-lorraine.fr

RÉSUMÉ. *L'intelligence artificielle (IA) a connu ces dernières années de grandes avancées qui ont résonné avec des préoccupations sociétales. Des instances ont été créées et ont commencé à structurer les problèmes posés par ces développements. Tant pour la société civile que pour de nombreux scientifiques, le champ de ces instances recouvre les problématiques du traitement automatique des langues (TAL). Dans cet article nous revenons sur certains aspects expliquant la relation entre IA et TAL, mais aussi sur les éléments qui les différencient. Nous revenons sur les problèmes d'éthique pour l'IA et également pour le TAL. Ces questions étant complexes, nous en donnons une lecture en contextualisant les problématiques. Enfin nous argumentons pour ne pas dresser l'éthique comme solution* de facto *à la réflexion, mais plutôt comme occasion de positionner les recherches dans des perspectives plus globales et nous revenons sur le problème de la relation entre utilisation de modèles numériques et faculté de les interpréter.*

ABSTRACT. *Artificial intelligence (AI) has evolved in recent years along with societal concerns. Various committees were introduced in order to brainstorm on the consequences of these developments. These authorities are also concerned by Natural Language Processing (NLP), not only as a subfield of AI but also as a specific field with which it interacts. In this article we review the links between AI and NLP but also where they differ. We focus on ethical clues for both of them. Finally we argue for not using ethics as a unique solution, but rather as the way to abstract over our researches. In the end, we go back on how to interpret machine learning methods in the context of NLP.*

MOTS-CLÉS : *éthique, intelligence artificielle, traitement automatique des langues, épistémologie.*

KEYWORDS: *ethics, artificial intelligence, Natural Language Processing, Epistemology.*

1. Introduction

Les recherches autour de l'informatique connaissent des accélérations phénoménales. Il s'opère ainsi un déplacement du champ scientifique qui questionne ceux qui le travaillent. En effet, l'informatique n'est plus seulement cette science qui s'intéresse au calcul, ou à ce qui est calculable, mais doit considérer ce en quoi elle modifie ce qui l'entoure. Exprimée ainsi, la phrase précédente peut s'appliquer à toutes les sciences, mais là où l'informatique devient spécifique, c'est que son évolution récente influence toutes les autres sciences. On parle par exemple de l'émergence des « humanités numériques ». Cette nouvelle thématique s'intéresse justement à la transformation des sciences humaines et sociales, des pratiques de la recherche et de l'interprétation des données en fonction de nouvelles informations que l'informatique permet d'obtenir. Il reste difficile de définir clairement ce que sont ces humanités numériques, ce qui explique la différence d'acceptation entre le monde anglo-saxon et le nôtre. Mais ces transformations ont également traversé les autres sciences. Sur ce point nous retiendrons que l'informatique opère une modification dans l'organisation d'autres champs disciplinaires. Sans préjuger de la conclusion à laquelle nous arriverons, il apparaît pour l'informatique que la problématique est aussi de comprendre comment les autres sciences la modifient.

Le spectre complet de l'informatique couvre de nombreuses questions, de problématiques très physiques à celles de la modélisation conceptuelle. Les capacités matérielles évoluant très rapidement, les machines deviennent de plus en plus capables. Cette aptitude n'est pas pour elle-même une réussite, et c'est l'objet dans lequel elle se déploie qui devient sujet de réflexion. Après avoir longtemps occupé le terrain des systèmes d'information, l'informatique se réalise ces dernières années dans le concept de l'intelligence artificielle (IA). L'apparition de l'IA comme objet est d'autant plus légitimée qu'elle a été définie par Alan Turing, considéré comme l'un des fondateurs de l'informatique. On regrettera que dans la période très récente, Turing ait occupé une place hégémonique, motivée par la célébration du centenaire de sa naissance et dans le même temps sa réhabilitation comme scientifique majeur du vingtième siècle. Cet engouement a laissé moins d'espace à d'autres branches de l'informatique tout autant fondatrices. Revenons sur l'IA qui revêt différentes formes dont l'une utilise la langue naturelle[1]. Un aspect est donc de positionner le TAL dans la perspective des grandes questions qui secouent l'IA. Un autre aspect que nous ne trancherons pas ici, mais d'une grande importance épistémologique, est d'arriver à positionner le TAL et l'IA l'un par rapport à l'autre.

Cet article propose de revenir sur les développements récents de l'IA et de positionner le TAL dans ce contexte. De fait, les problèmes d'éthique rencontrés par l'IA nous permettrons de positionner la question pour le TAL. On trouvera dans (Marquis *et al.*, 2014) une excellente mise en perspective historique et épistémologique de l'IA.

1. Langue utilisée par un humain dans un processus de communication. La terminologie s'oppose aux langages artificiels que l'on retrouve en mathématiques ou en informatique.

Turing pose dans (Turing, 1950) la définition de ce qu'est une IA. Il semble nécessaire de réintroduire (rapidement) sa définition pour pouvoir discuter les objets dans lesquels elle s'incarne aujourd'hui, ces aspects feront l'objet de la section 2. Il ressort que l'objet de cet article n'est pas de parler de l'IA pour elle-même mais de revenir sur le transfert de ces questions pour le TAL. Il apparaît très clairement qu'aujourd'hui, nous avons atteint un palier dans le développement des IA. De fait, cela génère un intérêt croissant du grand public et des médias autour de ces questions Mais derrière la fascination déclenchée par l'IA apparaît une méconnaissance des enjeux scientifiques. On arrive cependant à distinguer les notions d'IA forte et faible, et de discuter de la perception et de l'influence sociétale des deux, voir section 3.

À partir de ces constats d'existence de plusieurs formes d'IA, se posent deux questions. La première est de positionner les implications spécifiques du TAL dans cette relation à la machine ou à l'IA (voir section 4). Nous verrons en particulier sur un exemple la difficulté que nous avons à respecter les contraintes que nous nous donnons sur la question de la réidentification. La seconde est celle de la contextualisation de l'usage des IA, voir section 5. Ici se pose la question de qui est propriétaire des outils et pour quelles responsabilités. Un concept fondamental finit par émerger dans la relation que l'IA, et plus largement l'informatique, entretient avec les utilisateurs, celui de la loyauté. Par extension, se pose alors la question de la relation entretenue par l'informatique avec les autres sciences.

Finalement, la section 6 reviendra sur le problème d'inscrire un objet de recherche relativement à la question de l'éthique. Par ailleurs, un aspect n'a pas été posé dans la section précédente sur la contextualisation de l'objectif proposé par une IA. Il apparaît clairement qu'il y a un projet politique dès la définition de l'objet d'étude, de recherche ou industriel. La réussite économique, ou la sélection pour financement dans le cadre de la recherche, est donc la réalisation d'un projet de société. Avant de conclure nous reviendrons sur l'élément qui permet de donner une perspective éthique à notre champ, à propos de la possibilité d'interpréter des modèles uniquement construits sur des propriétés mathématiques sans analogie avec des aptitudes humaines.

2. Turing, Searle et l'IA

2.1. *Le test de Turing*

Il convient de revenir sur la définition du test de Turing introduit dans (Turing, 1950). Turing est connu pour avoir proposé une machine théorique qui a permis de construire une version physique capable de calculer effectivement et qui a donné lieu à la machine capable de décrypter les messages produits par Enigma pendant la Seconde Guerre mondiale. Au-delà du probable rôle prépondérant dans la gestion de la guerre, Turing fait surtout une proposition d'une force conceptuelle remarquable. Le principe de la machine de Turing introduit une différenciation entre le matériel et les instructions (ou le logiciel sous-jacent). Ces aspects sont longuement développés dans (Longo, 2009). Au-delà de ce résultat fondamental qui ouvre jusqu'à l'informa-

tique contemporaine, Turing se situe clairement dans une tradition issue de la logique. Il définit sa machine pour montrer que l'axiomatisation des mathématiques n'est pas possible, dans la suite de David Hilbert, et il le fait seulement quelques années après les résultats de Gödel (1931). Le pas fait par Turing est d'abstraire sur la faculté de calcul et de situer ces capacités dans un homme (*human computer*). Ici Turing mobilise une question qui sous-tend tout un pan de l'IA contemporaine autour de la notion d'incarnation. L'intelligence humaine existe parce qu'elle est située dans un corps, et une partie de sa faculté provient de la capacité du corps à interagir avec l'environnement.

On voit dans cette dernière remarque que ce qui meut Turing est bien de définir un objet technique capable de simuler une même faculté de calcul que l'homme. Pour Turing, il s'agit là de l'intelligence que la machine cherche à répliquer. Et s'il est possible de construire une machine capable de décoder les messages ennemis, cette machine doit bien être capable de simuler d'autres comportements aussi efficacement que des humains, voire davantage. Turing définira cela comme intelligence artificielle. Il est au centre du processus de calcul, le faisant passer vers l'intelligence. On se demande alors comment mesurer, calculer, qu'une machine est intelligente.

Cette question est difficile car elle ne peut que se réduire à la capacité de la machine à faire aussi bien, voire mieux qu'un humain pour une tâche donnée. Turing (1950) propose de tester la possibilité de se faire passer pour un humain. L'intelligence de la machine se résume donc à la tromperie d'un humain. Le principe de son test est qu'un humain pose une question sous forme écrite. La question est apportée dans deux salles distinctes. L'une accueille un homme qui répond à la question, à nouveau sous forme écrite, l'autre une machine, qui produit une réponse également écrite. Les deux résultats sont apportés à la première personne qui doit, au bout de quelques minutes d'échanges, déterminer qui est la machine et qui est l'humain. Si la machine arrive à se faire passer pour l'humain dans un nombre suffisant de cas, alors on considère que c'est une machine intelligente.

Ainsi Turing fonde l'intelligence artificielle et se prête au jeu de la prédiction historique en pariant sur l'existence de telles machines pour l'aube des années 2000. La prédiction s'avère fausse, tout comme bon nombre de prédictions scientifiques, mais là où Turing est resté prudent, c'est qu'il avait prédit que des machines avec 128 Mo de mémoire tromperaient des juges dans 30 % des cas. Depuis, de nombreuses propositions ont été faites pour faire évoluer cette notion de test d'intelligence.

2.2. *La controverse de Searle ou la place de la sémantique*

À la suite de Turing, de nombreux détracteurs sont apparus. Le plus célèbre est Searle qui avance que Turing se leurre en définissant le résultat du calcul comme de l'intelligence. Cette idée que l'intelligence réside dans le résultat prévaut aujourd'hui encore avec des avancées récentes de l'IA comme celles utilisées dans le jeu de go. Cela maintient une ambiguïté forte sur ce qui est attendu (tant comme résultat effectif

que comme simulation de la pensée). Cependant pour aller dans le sens de Turing, il n'y a *a priori* pas d'autre manière d'envisager le calcul qu'il voulait produire.

Searle avance un contre-argument connu sous le nom de la « chambre chinoise » dans (Searle, 1980). Les machines s'attachent principalement à la réalisation de surface de l'information impliquant qu'elles se concentrent sur les aspects syntaxiques des échanges, alors qu'un humain, supposé intelligent, mobilise des compétences sémantique. De fait, pour simuler l'aptitude à répondre à une question, il suffit de disposer de manuel expliquant comment gérer la langue pour produire la réponse.

Searle propose d'enfermer un humain dans une « chambre close » avec toute la documentation, dans sa langue, nécessaire pour répondre à toutes les questions qui lui sont posées en chinois. Si on décide d'apporter une question en chinois, cet humain va pouvoir appliquer les règles qu'il rencontre dans la documentation, sans comprendre le moindre symbole ou la moindre idée, mais il sera en mesure d'apporter une réponse. Si la réponse est correcte, donc si les manuels sont bien faits, la personne recevant la réponse peut croire (ou peut décider de croire) que ce qui a produit cette réponse est intelligent. Pourtant elle n'aura à aucun moment compris de quoi il s'agissait, et ne pourra ni reproduire, ni synthétiser, ni penser à partir de cette réponse. Pour Searle, il ne s'agit pas d'intelligence mais de la simulation de comportement qualifié d'intelligent. L'intelligence recherchée dans la machine n'est alors que son aptitude à simuler un comportement humain, et non à reproduire de l'intelligence (autant que l'intelligence puisse être un concept immanent).

Plusieurs contre-arguments à celui de Searle peuvent être avancés, en particulier sur les aspects linguistiques qui ne se résument pas à l'application de simples règles, ou encore que si la personne n'est pas intelligente, l'ensemble de la documentation qui conduit à produire la réponse l'est. Dans tous les cas on voit bien là la claire distinction entre la faculté de simuler une aptitude qualifiée d'intelligente pour les humains et celle véritable de produire de nouvelles connaissances.

Il faut malgré tout relativiser l'impact du test de Turing. Pour commencer, peu de chercheurs ont véritablement cherché à passer ce test, d'autre part, Turing ne l'a pas proposé comme méthodologie pour mesurer la capacité d'un programme à être intelligent, mais plutôt pour donner des perspectives explicites à ces questions, surtout, parce qu'en général il existe de bien meilleures méthodes pour évaluer la qualité d'un système informatique que de chercher à passer ce test, comme l'expliquent très tôt Russell et Norvig (2003) dans un livre qui fait référence sur la question de l'IA.

2.3. *La place du TAL dans le test de Turing*

Dans la description du test de Turing, la place du TAL revêt une importance capitale – qui plus est pour nous – et est très largement oubliée. Il est évident, mais très rarement mis en avant, que la personne au centre du test de Turing est en fait, non pas en train d'évaluer la capacité d'abstraction de son interlocuteur, ni sa capacité de synthèse, mais bien la transcription dans une langue d'un exercice de calcul.

L'avantage de passer par la langue est que cela permet d'utiliser un mode de représentation commun entre l'humain et la machine (ou plutôt entre tous les humains, puisque la personne doit déterminer qui est un humain et qui est une machine). Ainsi, il n'est pas question de comparer des modes de représentation de l'information, ce dont l'humain n'a généralement pas conscience. Il s'agit de maîtriser l'expression d'un résultat dans une langue. Il faut se rappeler que la langue est utilisée comme objet de pensée par les humains, ce qui ne sera pas le cas de la machine qui lui préférera nécessairement une modélisation numérique.

La qualité du raisonnement réalisé par la machine va donc être jugée par le prisme d'une autre compétence qui est celle de transformer ce qui a été trouvé en phrases. Mais alors, quelle est la vraie vocation du test de Turing ? Déterminer si une machine s'exprime comme un humain ou si elle est capable de réaliser des inférences (au moins élémentaires), de manière aussi efficace qu'un humain ? Pour Turing la question de la qualité du résultat se posait. Mais il faut aussi, ou d'abord, ou au moins, que la machine maîtrise le traitement automatique de la langue naturelle. Il semble donc étonnant que l'IA et le traitement de la langue n'aient pas constitué un seul et même champ de recherche. Pourtant historiquement, même si les deux branches se sont influencées, elles ont très longtemps seulement coexisté.

Ce point explique pourquoi une annonce récente a enflammé les médias et laissé de marbre les scientifiques. En effet, l'université de Reading en Angleterre a organisé un concours consistant à tester des agents conversationnels (chatbots). Les organisateurs ont annoncé que l'un des participants avait réussi à passer le test de Turing. Tout semblait vrai, mais pourtant, le test n'était pas passé. En réalité, le robot, prénommé Eugene Goostman était supposé être un adolescent ukrainien immigré aux États-Unis. Cette présentation induit deux faits : il est normal qu'un adolescent n'ait pas une connaissance encyclopédique et, le fait qu'il soit immigré présuppose qu'il ne maîtrise pas parfaitement la langue. Donc sous l'hypothèse de faire un robot qui ne s'y connaît pas en tout et qui s'exprime de manière approximative, peut-on faire croire à un humain que ce n'est pas un robot ? Certainement, mais l'impact est très limité.

Par ailleurs, une autre limite apparaît assez vite : comment considérer des robots ne maîtrisant pas la langue mais capables de réaliser des tâches complexes, c'est le cas des robots tueurs dont l'objectif est de localiser une personne dans un environnement physique non contraint pour la tuer. La tâche est très complexe puisqu'ils doivent se déplacer dans un univers réel, reconnaître sans erreur des individus (reconnaissance de formes, de visages...). Ces robots ne passent pourtant pas le test de Turing.

3. IA faible et forte

Au-delà de l'utilisation de la langue dans les systèmes d'IA, plusieurs résultats récents ont suscité l'enthousiasme, bien qu'appartenant à différents niveaux d'IA.

3.1. *Perception des enjeux de l'IA contemporaine et définition des IA faible et forte*

Les très récents résultats en IA qui ont été repris par la presse internationale sont ceux liés au jeu de go. On se rappelle que, dans les années 90, le système Deep Blue d'IBM, capable de battre le champion d'échecs A. Kasparov, avait suscité l'admiration. Cependant les chercheurs relativisaient l'enthousiasme général en rappelant, qu'au-delà de la performance technique et technologique il s'agissait principalement d'optimisation de calculs complexes. Même si le jeu d'échecs est très complexe, il ne définit qu'un problème dont la taille est très très grande. À cette époque, nombreux ont été à dire que définir une IA capable de battre les champions de go serait un problème autrement plus difficile. En effet, le jeu de go a la particularité de définir un problème dont l'espace d'états (les possibles) est de l'ordre d'un un suivi de cent-soixante-dix zéros. Bien que ce nombre soit énorme, cela pourrait sembler à nouveau être simplement un problème de calcul. Mais la machine n'est pas seulement limitée par le temps de réalisation des calculs mais aussi par l'espace dont elle a besoin pour mémoriser les données. Or, la taille théorique du problème correspond à l'ordre de grandeur supposé du nombre d'atomes dans l'univers. Il est impossible de mémoriser tous les états du jeu. La seule solution est de jouer avec une vision stratégique de la partie.

En mars 2016, AlphaGo, un système proposé par une filiale de la société Google, DeepMind, a battu le champion du monde de go. Le premier acte de ce match de l'homme contre la machine remonte au mois de janvier 2016 où AlphaGo a réussi à battre le champion d'Europe 5 parties à 0. Le rang modeste au niveau mondial de ce champion laissait croire que l'homme resterait plus fort que la machine, mais le match contre le champion du monde a été remporté par la machine (même si l'homme a gagné une manche). Il semble donc que nous disposions des outils théoriques pour construire une machine réalisant des opérations conceptuelles complexes de manière plus efficace que l'homme. En d'autres termes une vraie intelligence non humaine.

Cela n'en fait toujours pas une IA capable de passer le test de Turing puisque le système en question est « hyper » performant sur la tâche spécifique du jeu de go, mais ne connaît rien à la langue naturelle. La question reste ouverte quant à son aptitude à apprendre des propriétés de la langue. Il se trouve que si les méthodes utilisées sont très prometteuses, le domaine du go est très spécifique. D'une part la machine a pu disposer d'une modélisation d'un grand nombre de parties, et donc de nombreuses données pour faire son apprentissage de base, mais surtout, le système a pu jouer contre lui-même pour auto-apprendre la compétence de jouer au go. Ceci explique pourquoi au bout d'un certain temps la compétence dépasse celle du meilleur humain pour la tâche. Pour la langue il faudrait que le système se fasse la conversation et que le résultat converge vers le même état que celui des humains.

Au-delà du grand succès autour du jeu de go, on a longtemps considéré qu'il existait d'autres tâches difficiles, comme par exemple la reconnaissance de formes ou le fait de se déplacer dans un espace réel. Pourtant ces deux aspects sont aujourd'hui utilisés dans des contextes où l'on exige que la compétence soit parfaitement maîtrisée : les voitures autonomes en sont un exemple. Il s'agit de voitures développées

également par Google tout comme AlphaGo, capables de se déplacer dans un environnement standard, avec des passagers humains.

Les voitures autonomes posent une nouvelle question intéressante : peut-on accepter de déléguer sa sécurité à une machine ? En effet, les humains qui entrent dans la voiture confient leur sécurité à la capacité de la machine à se déplacer correctement dans l'environnement. Il n'est pas vraiment nécessaire de savoir si cette intelligence est supérieure à celle de l'homme mais bien s'il est acceptable d'avoir un robot dans notre environnement normal à la place d'un homme. La question devient encore plus délicate si l'on ajoute à cette réflexion l'éventualité que ces systèmes soient à l'origine d'un accident portant préjudice, non pas aux passagers, mais aux personnes de l'environnement. La gravité du préjudice peut modifier notre degré d'acceptation, et ce, même si le système provoque moins d'accidents que ne le ferait un humain.

Un autre robot intéressant est Atlas, développé par Boston Dynamics, également acheté par Google (mais en cours de cession). Contrairement aux précédents exemples, il s'agit d'un système pensé pour avoir une allure générale humanoïde, sans toutefois ressembler exactement à un humain. Il y a évidemment une projection anthropomorphique sur une telle machine qui représente les fonctions d'un homme. De manière symétrique, il y a un intérêt à avoir un robot de forme humanoïde s'il doit interagir dans un univers non spécialisé, la plupart des interactions étant prévues pour être réalisées par des humains.

Que ce soit la voiture sans conducteur, le robot capable de se déplacer ou le joueur de go virtuel, aucune de ces IA ne rassemble des compétences variées et élaborées. AlphaGo est un succès incroyable parce que la communauté scientifique n'avait pas prédit son avènement avant une dizaine d'années. Mais AlphaGo peut seulement jouer à un jeu complexe. Ce jeu se définit par une séquence réduite de règles. Il peut simuler un comportement humain dans le choix de l'ordre dans lequel les règles sont appliquées, ça ne fait pas de lui une intelligence. D'ailleurs, de nombreux observateurs ont souligné qu'après ses victoires, il ne manifestait pas de réaction, ce sont ses concepteurs humains qui en avaient.

Pour l'ensemble des ces réalisations, il serait plus juste de les définir comme intelligence numérique plutôt qu'intelligence artificielle. Ces systèmes décident d'un comportement à adopter sur un versant réduit de l'intelligence. On parle aussi d'intelligence artificielle faible, en ce qu'elle réalise seulement une partie de l'intelligence *a contrario* de l'intelligence artificielle forte qui, elle, serait à l'égal de l'homme. L'IA forte est donc un système reproduisant les aptitudes des humains de manière globale (activités motrices, perceptives et cérébrales). Il est possible d'aborder le problème par des aspects très différents que ce soit par la physiologie, la psychologie, le raisonnement, la mémoire, les interactions, voire la procréation, jusqu'à réaliser tous ces éléments dans un unique système « fort ».

La projection de l'existence d'une IA forte mobilise des communautés contre elle, comme le mouvement des transhumanistes. Ils posent la question de ce qu'il adviendrait de l'humanité si une IA plus intelligente, capable de s'autoprogrammer, émer-

geait et était hostile aux humains. Serait-elle un risque pour l'humanité ? Ces idées lèvent des questions éthiques et morales majeures que nous ne pouvons pas ignorer.

Cette question qui semblait pouvoir être reléguée à plus tard reprend de l'importance avec le développement très rapide de l'utilisation des réseaux de neurones profonds. Cette technique assez ancienne bénéficie d'une visibilité accrue depuis les années 2010 et les travaux comme (LeCun *et al.*, 2010 ; LeCun *et al.*, 2015). Ce type d'approche a montré des résultats significatifs pour des thématiques diverses : reconnaissance de formes, d'images, de sons, de textes, etc. Ce type de système s'apparente à des formes d'IA forte, au sens de la définition précédente, bien que nous soyons encore loin d'un système capable de penser sur lui même.

3.2. *IA et organisation de la société*

Qu'elle soit faible ou forte, il persiste une peur envers les IA. Au-delà de la question de l'impact pour l'humanité de l'apparition d'une IA forte, nous pouvons déjà voir les conséquences de l'apparition des IA faibles sur l'organisation politique de la société.

Sur ces thématiques, l'exemple de la transformation des métiers que nous connaissons revient souvent. De manière inéluctable, la technologie transforme les savoir-faire, accompagnant l'apparition de nouveaux métiers et la disparition d'anciens. Il n'est pas question de les hiérarchiser, ni de tenter de décider si certains seraient plus utiles que d'autres. Il n'en reste pas moins vrai que les métiers évoluent.

Actuellement aux États-Unis, le métier de chauffeur (de taxi, de camion, de car...) semble particulièrement visé. Cela provient directement du développement de véhicules autonomes. Un exemple similaire a émergé dans les années 70 avec le travail de secrétariat. L'évolution de la technologie n'a pas supprimé les postes de secrétaire, mais a plutôt généralisé leur fonction, les démultipliant. On peut donc supposer qu'il y aura un transfert de compétences des chauffeurs vers d'autres métiers. Mais c'est une vision assez simplificatrice de la situation.

Des analyses mettent en avant le risque que 47 % des métiers aux États-Unis soient remplacés par des robots (Frey *et al.*, 2016). Le chiffre atteint une moyenne de 57 % pour les pays de l'OCDE, et il faut noter que les pays à forte croissance économique atteignent un risque très élevé : 72 % pour la Thaïlande ou 77 % pour la Chine. Il s'agit bien là d'une transformation en profondeur de l'organisation de la société par le travail. Mais dans le même mouvement, seulement deux postes sur cinq seraient transformés en un nouveau métier. Le transfert est donc loin d'être total et il s'agit alors de poser la question du partage du travail. Ainsi, contribuer à développer une IA moins faible participe à s'inscrire dans une évolution politique et sociale. S'il ne faut pas présupposer qu'il y ait un dessein politique derrière les recherches, on ne peut pas écarter qu'il doit, *a minima*, y avoir conscience de la dynamique sociétale dans laquelle cette recherche s'inscrit.

Il ne faudrait pas non plus supposer qu'il n'y a que des questions mineures ou théoriques derrières ces aspects. Il est clair que la question du développement d'IA plus fortes est un enjeu politique comme nous venons de l'introduire, mais également économique et industriel. De nombreuses estimations sur le volume financier du marché mondial de l'intelligence artificielle sont avancées pour 2020. Si aucune donnée ne semble consolidée, à chaque fois il s'agit de milliards de dollars. On peut raisonnablement penser que s'il y a une part de vérité dans ces valeurs, des entreprises vont certainement investir massivement sur ces questions.

Un autre aspect également intéressant est de constater que les acteurs économiques incontournables comme Microsoft Corporation, Google, IBM et Facebook sont exclusivement américains, et principalement de Californie. Il y a donc une concentration massive de moyens et de compétences dans une zone géographique réduite. De fait, le développement économique du champ scientifique est uniquement influencé par le modèle politique américain, ce qui induit des dynamiques et des choix spécifiques. On peut alors se demander quelle est la place des autres zones géographiques sur ces thèmes, et quelle est l'influence de leurs politiques.

Mais s'il existe une vision alarmiste de l'impact de l'IA sur la société, on peut déjà voir que le mouvement ne va pas dans cette seule direction. On observe par exemple que des robots sont remplacés par des hommes qui sont plus performants sur des tâches fines. Ce qualificatif de « fin » est en fait une manière de signifier que la robotisation nécessite parfois une interaction justifiée par des inférences complexes, ou, dit autrement, une forme d'intelligence particulière que possède l'homme. Par exemple, dans les entreprises automobiles de Mercedes-Benz où la mécanisation permet d'avoir des postes moins usants pour l'homme qui prend une nouvelle place dans la chaîne.

Au-delà des aspects économiques et politiques, l'aspect sociologique de l'acceptation des robots au sein de la société doit être traité. Cette question revêt plusieurs aspects car les robots et les IA sont déjà en son cœur sans que nous ayons le sentiment d'avoir vécu une transition sociétale majeure. Par exemple, les smartphones omniprésents sont constitués de briques de base d'IA, comme les logiciels de reconnaissance de la parole (capables de répondre à une question posée par l'utilisateur).

Si l'on parle de robots plus élaborés ou humanoïdes, la question se pose de rendre plus fluides et acceptables les interactions avec eux. C'est d'ailleurs un sujet de recherche à part entière que de prendre en compte les émotions dans ces machines (à la fois identifier les émotions et également en afficher). Si on cherche à intégrer une dimension émotionnelle dans ces systèmes, c'est pour augmenter leur acceptation en simulant des comportements plus proches des aptitudes humaines. C'est par exemple le cas pour les machines qui maintiennent en activité physique et cérébrale des personnes âgées. S'agit-il toujours du même type d'interactions qu'avec le reste de la société ? Ne sommes-nous pas en train de participer à transformer les interactions sociales en les résumant à des interactions entre humains et machines ?

Dans le même temps, nous n'interrogeons pas – et peut-être même serions-nous effrayés par cette idée – les algorithmes de *trading* à haute fréquence qui prennent

des décisions en quelques millisecondes pouvant aller jusqu'à provoquer des famines dans certaines régions du monde. Ces machines sont intelligentes, capables de réaliser des tâches plus rapidement (efficacement ?) que les hommes. Il n'est pas question de sentiments. On voit bien dans ce dernier exemple que le projet politique sous-jacent à la forme d'IA que collectivement nous faisons émerger est déjà en place. Ce sont les algorithmes de *trading* qui sont intégrés à l'organisation de la société, non pas les algorithmes pour favoriser la diffusion de l'enseignement.

L'intégration dans la société des robots n'est pas si évidente que cela. Une expérience intéressante a été réalisée par un groupe de chercheurs canadiens avec comme prétexte d'étudier la possibilité que les robots pouvaient faire confiance ou non aux humains et donc croire en eux. La question n'en est évidemment pas une, les robots ne sont pas dotés d'affects ni de sentiments. Le groupe de recherche a construit un robot simplifié avec des bras et des jambes en mousse, des gants en plastique, des bottes en caoutchouc, un corps cylindrique et une tête dotée de quatre écrans à gros pixels. Ce robot, nommé HitchBot, avait pour objet de faire du stop et de parcourir des régions du monde. Il a pu réaliser des voyages au Canada, en Allemagne, aux Pays-Bas et aux États-Unis. Cette version IA.0 du voyageur est sympathique et a suscité un engouement important, relayé dans les principaux médias. Étant donné la taille et l'allure du robot, on pouvait l'assimiler à un enfant. L'aventure a duré un an et trois mois avant que le robot ne se fasse sauvagement attaquer à Philadelphie. On peut en conclure que les robots ne sont pas si facilement intégrables dans la société. Mais ce que ce robot ne savait pas faire c'était d'œuvrer pour maintenir son intégrité physique. De fait, il n'était pas adapté au monde[2].

4. Retour sur le TAL

Dans les sections précédentes, nous avons cherché à positionner certaines questions relatives à l'IA en général, sans nous arrêter particulièrement sur le cas du traitement automatique des langues. Une première approximation voudrait que le TAL soit une composante de l'IA et qu'il convient donc de préciser les arguments pour cette sous-partie. Mais la relation entre l'IA et le TAL est plus complexe qu'une simple inclusion.

4.1. *Où se positionne le TAL*

La position du TAL est assez semblable à celle de l'IA : de nombreuses réalisations sont déjà déployées dans la société (sur des thématiques qu'il n'est pas facile à décrire exhaustivement), et dans le même temps, la perception du TAL par les non-spécialistes les amène à croire que les outils peuvent davantage qu'ils ne font.

2. C'était d'ailleurs l'argument fatal à Hal dans *2001, l'Odyssée de l'espace* de S. Kubrick.

Les formes de TAL faibles, par analogie à la qualification pour l'IA, se retrouvent par exemple dans les téléphones mobiles sous forme de logiciels de reconnaissance de la parole couplés avec des moteurs d'inférences simplifiés. Il est ainsi possible de simuler une interaction en langue naturelle avec l'appareil sans que celui-ci n'ait véritablement interprété le sens de la question. S'il parvient à identifier suffisamment de contexte, il est en mesure de produire une réponse aux stimuli. L'appareil ne parvenant que très médiocrement à comprendre, ces outils ne sont pas de bons exemples en ce qu'ils réduisent la compréhension de la tâche[3]. Cela ancre dans l'imaginaire collectif l'impossibilité que nous avons à traiter la langue. Pour évaluer le caractère très limité du succès de ces outils, il suffit de regarder les innombrables vidéos d'appareils tentant d'entrer en communication l'un avec l'autre. C'est une preuve empirique que ces outils ne sont pas réflexifs, contrairement à AlphaGo qui peut jouer contre lui-même.

Il est pourtant admis que l'interaction avec une machine, et donc une IA, passe par la langue naturelle. Ainsi, la perception de l'amélioration qualitative des IA passe nécessairement par un TAL fort. L'existence de ce dernier est d'ailleurs un présupposé des images de l'IA dans la fiction. C'était le cas pour Hal, mais aussi plus récemment pour les hubots de *Real Humans*[4]. Au-delà d'ajouter des émotions dans le comportement des machines, utiliser le TAL faible actuel participe à définir la perception que nous avons de l'IA. De fait, le TAL apparaît comme une discipline de l'IA.

À ce point de l'argumentation, une question technique se pose. Et si l'hypothèse précédente était vraie ? Peut-être que la solution pour passer d'un TAL faible à fort réside justement dans le fait de considérer que ce n'est qu'un sous-problème de l'IA ? Dans ce cas, il est possible de s'inspirer des travaux en linguistique, mais probablement pas de chercher à les inclure pleinement, sauf à considérer que toute discipline est une sous-discipline de l'IA. L'IA actuelle est largement entraînée par le développement de méthodes numériques, qui sans être complètement décorrélées du type de données auxquelles elles s'intéressent, ne sont pas construites à partir d'elles. Il n'est pas aisé de se livrer au jeu des pronostics quant aux chemins qui permettront de faire évoluer le TAL, mais il semble que des résultats encourageants proviennent de l'utilisation des techniques mises en œuvre dans les IA actuelles. Dans un article récent Manning (2015) revient sur cette question. Il met en avant l'intérêt que les tenants des méthodes numériques du type *deep learning* portent actuellement sur le TAL. Ces derniers annoncent qu'ils le voient comme leur prochain défit et même qu'ils seraient surpris de ne pas parvenir à le résoudre dans les cinq prochaines années. En réponse à cet enthousiasme qui peut apparaître comme naïf, Manning (2015) argumente sur l'intérêt pour le TAL d'être considéré comme tel, et sur la nécessité de considérer les champs de recherche par leurs objets et non par leurs méthodes, afin de notamment contrer la tentation d'améliorer des indicateurs décorrélés de toute réalité.

3. Il est fréquent que la réponse soit en adéquation avec la question, ce qui permet de prêter au système une faculté d'interprétation. Mais cela ne reste possible que lorsque la question attend une réponse précise du type encyclopédique.

4. *Real Humans* est une série suédoise créée par Lars Lundström et diffusée en France entre 2013 et 2014. Elle mettait en scène des robots à l'apparence humaine au sein de la société.

Cet aspect pose une question cruciale sur l'interprétation des résultats. En reprenant l'exemple précédent des émotions, la reconnaissance peut être réalisée à partir de la voix. Les techniques utilisées proviennent très largement des méthodes numériques et les robots cherchent des indices à partir de la forme prosodique et non du contenu. Un acteur exprimant des émotions très contradictoires sur le même ton trompera facilement le robot. Cet exemple montre l'importance de ce type de travaux où la synergie TAL et IA est une nécessité, mais également la difficulté de ne pas avoir d'argument quant à la structure qui permettrait d'expliquer le résultat (et donc le comportement).

Une autre réalisation, qui a été largement médiatisée et présentée comme une IA contemporaine, est le système Watson d'IBM. Watson est un système de fouille de données et d'extraction d'informations couplé à des technologies de TAL. Pour (dé)montrer sa grande capacité, et marquer durablement les esprits, IBM l'a fait participer au jeu télévisé *Jeopardy*[5] où il a été confronté aux meilleurs joueurs humains. Watson n'avait pas d'interface particulière et était matérialisé (incarné ?) par un globe terrestre sur un grand écran. Il interagissait avec les autres protagonistes en langue naturelle. Watson a largement battu les autres joueurs et s'est imposé comme une forme d'intelligence numérique, un peu à la manière d'AlphaGo aujourd'hui[6].

Ce résultat est une réussite puisque le système a été capable d'interagir avec des humains en langue naturelle. Mais Watson n'a pas vraiment plaisanté avec les participants, ni avec le présentateur, ou fait référence à des réponses précédentes. Watson s'est concentré sur des énoncés très courts et a cherché à quoi les associer. Dès qu'une association émergeait de sa recherche, il devait l'exprimer sous forme de question non ambiguë. Dans ce contexte, toutes les tâches sont réduites à leur version simplifiée :

– l'analyse est réalisée sur des formes d'assertion de quelques mots ;

– la recherche se fait dans des bases de connaissances. En cas d'ambiguïté, un choix aléatoire est opéré et produit une question (dont l'assertion est une réponse) ;

– la génération est limitée aux structures syntaxiques les plus simples.

En revanche, le vocabulaire n'est pas réduit, et c'est là une réussite pour le système. Il faut pourtant bien considérer qu'il n'est pas en mesure de répondre à la question : « Est-ce que les poissons rouges battent les éléphantes au criquet ? ». La réponse à sa recherche (la question dans son cas) doit pouvoir être le résultat de l'application de règles. Pour cela, il faut donc que la modélisation du monde contienne explicitement que les poissons rouges ne jouent pas criquet, ou pas contre les éléphantes. Sans cela il n'est pas en mesure de répondre, contrairement à un enfant d'une dizaine d'années.

Comme le grand public considère que ces systèmes maîtrisent la langue, il les considère également comme du TAL (quasiment) fort. Nous ne sommes pas étonnés qu'il soit envisagé de les utiliser dans des contextes médicaux, au-delà des questions éthiques et morales.

5. Jeopardy est un jeu fondé sur la création de paire réponse-question à partir d'une assertion.
6. Il est intéressant de noter que les systèmes sont présentés comme intelligents lorsqu'ils remportent un succès à des jeux, comme si l'intelligence c'était de gagner aux jeux.

4.2. *De l'enjeu du TAL dans un projet politique*

Comme nous venons de le voir, le TAL n'est pas exempt de critiques, au contraire. Il se différencie de l'IA en ce qu'il ne cherche pas à construire une entité, physiquement à l'image de l'homme, mais en ce qu'il cherche bien à reproduire une capacité humaine. Or, il ne s'agit pas de n'importe quelle capacité, mais de celle qui permet à l'homme de communiquer et donc de le construire comme être social.

Les outils technologiques, leur perception, leur compréhension et leur insertion dans la société sont largement portés par la tromperie ou, exprimés de manière moins forte, par la méconnaissance des problèmes. Ainsi, les outils du TAL permettent la duperie des utilisateurs pour augmenter leur croyance en la réalité des IA.

Mais plus encore, comme il s'agit de se situer au cœur du système d'interactions sociales, les outils du TAL évolués seraient en mesure de comprendre, modéliser et inférer sur ce qu'un individu met en jeu dans ses interactions sur Internet. Le TAL devient ainsi une question de modélisation de l'individu dans l'entièreté de son processus de communication. Parmi les exemples, on a vu un appel à projet du ministère de l'Éducation nationale en 2008, où la thématique explicite était la détection de leaders d'opinion sur Internet et de lanceurs d'alerte[7]. Cette application semble très pertinente car elle nécessite de travailler sur du texte non contraint thématiquement et de parvenir à extraire du contenu sémantique des messages sur des tailles de messages différentes. Il s'agit bien de passer un cap technique et technologique par l'application de théories matures. D'autant que, si la problématique peut intéresser ce ministère, elle est également pertinente pour les industriels, tant pour suivre comment leur marque est décrite sur Internet que pour identifier les personnes en influant la perception.

Il se pose *de facto* la question de l'utilisation des résultats de la recherche. Si cette question prévaut pour tous les domaines, lorsque les mises en œuvre sont aussi directement présentées comme étant utilisées dans ce contexte, il apparaît nécessaire d'interroger les objets de ses propres recherches. Les moyens de tester la qualité des résultats de l'identification de leader d'opinion et d'influence numérique répondent, sinon directement à un projet politique, au moins à une vision normative des comportements sociaux (et donc de leur expression par la langue).

Cette idée de massification des applications fondées sur la langue comme compétence spécifique est aussi le vecteur d'un système d'organisation sociale. C'est évidemment le cas pour les outils actuellement développés pour l'apprentissage outillé. Malgré les nombreuses possibilités d'adaptations automatiques, ils conservent le présupposé d'une norme de comportement dans l'apprentissage. Le cas poussé à l'extrême est celui des enfants autistes qui répondent à des stimuli sous forme d'interactions avec des machines. Ainsi, ils apprennent à parler pour communiquer, ce qui

7. Il s'agissait d'un marché portant sur l'investissement de 220 000 euros lancé par X. Darcos et V. Pécresse en 2008 pour la « veille de l'opinion » dont l'objet était de surveiller l'opinion dans le domaine de l'éducation, de l'enseignement supérieur et de la recherche durant l'année 2009, année d'un mouvement social d'ampleur dans le domaine.

est un grand progrès dans leur sociabilisation, et la machine semble nécessaire, par exemple, pour répéter une même réponse un grand nombre de fois, (Wainer, 2012). Il n'en reste pas moins qu'il faut questionner le type de langue qu'ils apprennent, et de fait, le type d'interaction sociale. Cette forme semble très proche et fondée sur la langue naturelle, mais n'existe que dans un univers qui exclut les autres humains. Il s'agit alors d'une protolangue naturelle, résultat de la qualité des outils de TAL.

Un dernier exemple où le TAL comme objet de recherche entre dans une problématique politique est relatif à nouveau à la localisation du développement de ces outils. Le développement d'un TAL fort permettrait d'accéder à un très grand nombre de prises de parole et de positions simultanées sur l'ensemble du globe, et ces outils seraient massivement utilisés depuis les États-Unis. Il ne s'agit pas de chercher un dessein stratégique particulier mais de rappeler que l'accès à l'information est une partie du pouvoir, et que, de fait, la question de l'IA et celle du TAL, ne doivent pas être négligées par les instances politiques, au risque de se voir distancer par d'autres nations. Sous cette hypothèse, le TAL est bien un enjeu particulier d'un projet politique.

Mais il ne faut pas négliger qu'il est possible de réguler *a priori*, en explicitant la responsabilité du chercheur dans l'utilisation faite de ses résultats, et *a posteriori* par une régulation par les États ou une implication de la société civile dans l'évaluation des systèmes. Plusieurs pistes ont été avancées en ce sens tant dans (CERNA, 2014) qui est repris par le Conseil d'État dans son étude annuelle (Cassin *et al.*, 2015).

4.3. *Sur la question de la norme et celle de la réidentification*

Nous voyons émerger les enjeux d'un TAL fort dans notre impossibilité actuelle à résoudre certains problèmes. Dans le projet de recherche SLAM[8], nous avons décidé de travailler sur des données qui se sont avérées sensibles. Ce caractère n'ouvre pas un risque pour la société, au contraire de recherches en chimie ou en physique, mais pose bien une question morale car les données sont sensibles pour les personnes. De manière naturelle nous avons cherché à les protéger de ce qui pourrait permettre de leur associer un profil médical particulier.

Nous avons établi des profils singuliers par comparaison de comportements entre des cohortes d'individus. Ainsi, la manière dont les sujets s'expriment par la langue nous renseigne sur leur profil psychologique. On voit apparaître une vraie question d'interprétation de la déviance de comportement par rapport à la norme. Est-ce qu'embrasser une manière de s'exprimer légèrement différente de celle du plus grand nombre est symptomatique d'un dysfonctionnement ? Nous abordons la question selon plusieurs critères philosophiques, mais elle se pose pour tous les objets de recherche.

Ainsi, on peut s'interroger sur la nature de la compétence acquise par les outils du TAL développés à partir de corpus d'entraînement spécifique, comme celui de l'*Est-*

8. Schizophrénie et langage : analyse et modélisation, http ://discours.loria.fr

Républicain, et sur ce qu'ils sont capables de reconnaître de la langue. S'agit-il de la langue naturelle et non d'un usage tout à fait caractéristique du monde journalistique ?

Au-delà de la question de la norme, a également émergé la question de la possibilité d'identifier ou de réidentifier les personnes. Nous disposons d'outils très élaborés, capables de gommer un grand nombre d'informations en rapport avec les sujets. Mais dans notre cas, le corpus est principalement construit par des entretiens avec des patients qui racontent leur vie et leurs relations sociales. L'anonymisation consiste donc principalement à faire disparaître les entités nommées. Nous avons identifié une solution simple qui consistait à mélanger plusieurs entretiens pour limiter la reconstruction des biographies. Or, les phénomènes que nous étudions s'expriment justement dans la suite logique portée par la narration, ce qui invalide alors cette désidentification.

Même s'il n'est pas possible de découvrir explicitement les personnes, il est largement possible de retrouver de nombreux éléments biographiques. Nous avons exposé ces questions dans (Amblard *et al.*, 2014) lors de la journée ATALA sur l'éthique et de l'atelier ETeRNAL de TALN qui ont rassemblé sur la même thématique Eshkol *et al.* (2014), Mazancourt *et al.* (2015), Grouin *et al.* (2015). Nous avons présenté un extrait du corpus dans lequel on voit apparaître des éléments personnels significatifs. S'ils ne permettent pas de reconnaître l'individu (il n'en est évidemment pas question ici), ils permettent de limiter la recherche à un petit nombre de personnes. Sachant que les données personnelles sont de plus en plus facilement accessibles, il est alors pertinent de croiser ces données selon l'extrait. Il ne faut pas non plus négliger les effets de bord car s'il est possible de cacher en partie les identités, les choses se compliquent quand on sait qu'il est possible d'extraire des informations sensibles sur les familles des patients ou sur leur entourage.

Pendant longtemps, nous ne nous sommes pas tant inquiétés de laisser disponibles des informations relatives à nos usages et à nos pratiques. Mais la numérisation obligatoire qui accompagne nos vies fait que des documents contenant plusieurs éléments biographiques permettant de reconstruire nos histoires sont disponibles. La barrière de la langue qui permettait justement de disposer de l'information sans la rendre accessible aux algorithmes a donc des conséquences importantes. Si l'on reprend l'exemple précédent, le problème de la réidentification ne se pose pas pour ce qu'il contient, mais en fonction des données mobilisables pour la recherche d'indice biographique. Le niveau d'anonymat aurait été probablement acceptable dix ou quinze ans plus tôt.

Il nous est souvent répondu que nous adoptons une posture trop protectionniste vis-à-vis de nos données. Il est vrai qu'il est peu probable que des corpus de TAL intéressent suffisamment pour qu'ils soient lus, voire qu'ils incitent à faire des recherches à partir des contenus. Le problème ne se situe pas sur ce plan. La démultiplication de la mémorisation des données et leur accessibilité rendent envisageable l'extraction automatique des éléments biographiques. Si cela n'est pas possible aujourd'hui, nous devons nous interroger sur ce que nous pouvons assurer comme anonymat et non-réidentification dans le temps. Si les données sont peu sensibles, il peut être acceptable de passer outre cette question, mais nous ne pouvons pas nous en contenter. Cet exemple montre la nécessité d'interroger nos pratiques à l'aune de nos compé-

tences actuelles, mais plus encore avec le regard de l'expert scientifique. Il ne faut pas nous cacher derrière de mauvaises pratiques car c'est en interrogeant nos activités scientifiques que nous trouverons les règles pour encadrer nos pratiques. Pour parvenir à un tel résultat effectif, il nous faut systématiquement interroger les perspectives de nos propositions de recherche.

5. Contextualisation

S'il nous faut nous interroger sur l'utilisation de nos recherches en TAL comme en IA, cela ne signifie pas pour autant que nous sommes responsables de toutes les utilisations faites de nos recherches. Mais nous ne pouvons pas nous en dédouaner totalement. Cette question de la responsabilité des usages est ancienne dans le rapport du scientifique avec son objet d'étude. Et si nous avons toujours pu nous retrancher derrière l'argument selon lequel c'était celui qui avait construit l'objet qui était responsable, cet argument est en train de nous échapper. Devenons-nous alors définitivement irresponsables ?

5.1. *De la responsabilité de ceux qui font*

Un exemple concret et récent de programme manipulant la langue (dans une vision réductrice du TAL) peut être trouvé dans une société qui met en vente des t-shirts *via* le site d'Amazon, en Angleterre. Les illustrations sont des slogans générées automatiquement à partir d'un détournement du célèbre *Keep calm and carry on* (Restez calme et continuez). L'idée est que la société n'imprime les t-shirts qu'à la demande. Ce qui importe est de disposer de demandes et pour cela, il est uniquement besoin d'avoir un message dans lequel un potentiel acheteur va suffisamment se reconnaître pour procéder à l'acte d'achat. Il suffit alors de tenter sa chance en produisant aléatoirement des slogans. L'infrastructure numérique est opérationnelle pour accepter des milliers de propositions sans intervention humaine. C'est ainsi que les algorithmes ont proposé d'arborer le message *Keep calm and rape a lot* (« reste calme et viole beaucoup »). Le message a ému le public et a été largement repris sur les réseaux sociaux. On peut raisonnablement penser que personne dans la chaîne de la vente ne soutient explicitement ce message, mais pourtant il faut bien répondre à la question de la responsabilité. Qui a produit un appel au viol ? S'il s'agit de celui qui a proposé l'algorithme, nous sommes bien responsables des outils et des théories que nous mettons à disposition du public.

Nous avons donc une responsabilité dans ce que nous proposons comme technologie à la société, mais également, et plus encore quand nous participons à modifier les relations à l'intérieur de la société. Ainsi, si mes recherches participent à rendre plus proche d'un comportement humain un robot, par exemple en lui donnant une faculté de manipuler la langue naturelle, est-ce que je deviens responsable de l'acceptabilité des robots dans la société ? On retrouve bien là une problématique de la science et de la technologie en général. Si je développe des outils de représentation de la séman-

tique de la langue, je ne cherche pas à faire mieux accepter des robots dans la société. La question n'est plus la même si mon objet d'étude est de travailler avec des psychologues pour rendre l'interaction verbale avec des robots plus acceptable. Dans ce second cas, la recherche participe à un projet politique différent. Ce type de question est envisagé par le biais du support aux handicaps, ou à l'accompagnement du maintien des personnes âgées à domicile. Autant dire des problématiques qui emportent l'adhésion générale.

La question est d'autant plus délicate que les approches du type *machine learning*, massivement utilisées, ne simplifient pas l'attribution de la responsabilité. Jusqu'à présent la technologie a principalement été fondée sur des approches qui prévoyaient le comportement de la machine. Ces nouvelles approches décalent le problème car le système apprend des données (des corpus) à qui ont ne peut pas transférer la responsabilité. La CERNA a justement lancé une réflexion sur l'éthique de l'apprentissage. De quoi, un système fondé sur un modèle d'apprentissage qui produirait l'exemple du t-shirt précédent, serait-il le symptôme ? Pour répondre à cette question il faudrait réinterroger la responsabilité de chacune des ressources mise en œuvre. Il y a deux attitudes, l'une est de considérer que l'apprentissage oblige encore davantage à une interrogation éthique de l'outil produit, l'autre est de rejeter systématiquement toute responsabilité sur ce type de question. Les outils finalement produits par les communautés scientifiques n'auront vraisemblablement pas le même type de comportement.

5.2. *IA et droit*

Au travers des exemples précédents, on voit l'importance de la question juridique. Le rapport du Conseil d'État précédemment mentionné est d'une grande richesse sur cette question (Cassin *et al.*, 2015). Sans revenir sur la nature de la relation entre le scientifique et sa production, il serait possible d'encadrer la notion de droit dans l'interaction avec des robots, par extension des réflexions pionnières sur la relation de l'homme avec l'animal. Les promoteurs d'une charte des droits des robots, (Bensoussan, 2015), animent les interrogations sur ces thèmes depuis plusieurs années. Les questions légales sur la nécessité de disposer d'un droit pour ou sur les robots sont complexes. L'article 6 de cette charte définit que c'est l'utilisateur du robot qui est présumé responsable des agissements du robot (avec des variations de responsabilité). Ainsi, après l'utilisateur, la responsabilité incombe au fabricant conceptuel, puis au fabricant des composants technologiques, et enfin au propriétaire.

L'expression des besoins quant à l'encadrement juridique de l'usage des robots diffère largement en fonction du contexte socioculturel. Par exemple, la Corée du Sud s'est très rapidement positionnée sur cette question car les robots y sont massivement acceptés (et même recherchés). Au contraire, les systèmes anglo-saxons ont une tradition juridique de la jurisprudence, préférant constater par l'expérience l'émergence de problèmes à trancher plutôt que des les prévoir théoriquement.

Les motivations pour construire des robots induisent des rôles que nous projetons pour eux, au-delà de l'exercice théorique de la considération d'un nouveau type d'interactions. Un cas prototypique, à l'extrême opposé de l'intégration du TAL, est celui des robots tueurs développés par les différentes armées comme nous l'avons mentionné. Ces robots sont capables de se déplacer dans un environnement physique réel et de reconnaître les visages et une fois la cible identifiée, de l'abattre. Il est conféré un droit de vie ou de mort aux robots. Il est alors nécessaire de définir qui endosse légalement la responsabilité de leurs actions par exemple en cas de bavure. Cette problématique s'inscrit à la suite de celle sur les implications éthiques et morales de l'utilisation de drones comme armes de guerre (Wareham, 2014) ou leur défense (Müller et Simpson, 2014). Les soldats pilotant ces engins sans être sur un terrain de guerre sont-ils responsables des conséquences de leurs actes ?

Si l'on peut considérer que le cas précédent est une vision excessive de la situation, le transfert à la société civile est malgré tout en train d'avoir lieu. Comment considérer le cas où un robot autonome dans une usine est responsable de l'action qui entraîne la mort d'un ouvrier, comme cela a été le cas en juillet 2015 dans une usine Volkswagen en Allemagne. Par ailleurs, l'argument qui consisterait à renvoyer cet aspect de l'éthique aux seuls roboticiens n'est pas convaincant.

Aborder cette question par les robots permet de partir d'une problématique fondée sur des interactions dans le monde physique, ce qui se rapproche aisément du fait des choses (notion ancienne en droit). Mais cela inclut aussi les implications des choix à l'origine des actes. Les outils du TAL se retrouvent face à ce même type de problématique. Ils opèrent des choix dans le monde numérique qui peuvent avoir des conséquences majeures dans le monde réel. C'est le cas des applications qui gèrent la santé des individus, et en particulier leur santé mentale. La projection de la faute commise dans ce contexte n'est pas la réalisation d'un acte physique contre un ouvrier mais celle résultant d'une erreur dans l'accompagnement de la prise de médicaments. Qui en sera tenu responsable ? Et si le patient a lui-même manipulé le comportement du système pour qu'il l'encourage à surdoser sa prise de médicaments ? Ces différents aspects interrogent de la même manière le TAL, même si cet aspect a moins été pris en considération pour le moment.

Un exemple de technique qui relève du TAL mérite d'être introduit dans cette partie sur le droit. Il s'agit plutôt de l'utilisation des techniques dans le domaine du droit que du droit de ces techniques, mais on voit bien le renversement qui s'opère. Les *block chains* sont des technologies qui permettent de déclencher automatiquement des contrats. Cette activité est possible car les textes sont fondés sur la langue. Ainsi, il n'est plus nécessaire de faire intervenir un spécialiste humain (un avocat) modifiant profondément son activité, (Veith *et al.*, 2016). Mais dans le même temps, l'avocat est un garant du contenu du contrat. Qui alors endosse cette responsabilité ?

5.3. *Autour de la loyauté des algorithmes et des systèmes*

S'il y a un élément sur lequel nous ne pouvons pas transiger, c'est celui de la loyauté des algorithmes. On retrouve dans (Cardon, 2015)[9] une motivation pour cette propriété. Actuellement il existe deux sortes de logiciels, les logiciels libres et les autres, que nous appelons fermés. On considère en général qu'il y a une différence de modèle économique derrière le choix de diffuser ou non le code d'un logiciel. Mais c'est méconnaître le logiciel libre parce que, d'une part il existe un modèle économique pour les logiciels libres, il est simplement décalé du logiciel à proprement parler vers celui de l'accompagnement de ses usages, et d'autre part, il y a une différence philosophique majeure entre les deux approches. Pour les logiciels libres le code de leur conception est accessible et permet de vérifier leur fonctionnement. Si individuellement nous ne sommes pas en mesure de faire cette vérification, la communauté peut expliquer et vérifier le comportement du logiciel, contrairement aux logiciels fermés.

Cette affirmation n'interroge pas les néophytes de l'informatique qui ne sont pas en mesure de faire les vérifications. Pour eux, le choix est souvent réduit à un logiciel convivial ou un logiciel demandant un apprentissage. Mais que se passe-t-il si le logiciel fait d'autres choses que ce pour quoi il est prévu ? Si votre logiciel transforme votre ordinateur en distributeur d'informations sur vos comportements alimentaires ou vestimentaires à votre insu, vous n'aurez comme témoignage que des publicités mieux ciblées. Si votre logiciel intercepte tous vos échanges électroniques, profils, messages courts ? Rien de grave sauf à considérer que la vie personnelle a une valeur.

Un autre élément semble pertinent dans la dynamique des relations machine et humains actuelle, autour de la personnification des machines qui justifie la proposition d'un droit des robots comme nous l'avons évoqué. Les machines, les robots et les algorithmes ne sont pas des personnes et ne se comportent pas comme telles.

Sans tomber dans une défiance envers les logiciels fermés, il conviendrait de parvenir à un niveau de confiance dans des logiciels présentant des garanties. Par exemple, en 2015, la Commission d'accès aux documents administratifs (CADA) a rendu plusieurs avis contraignant l'État à rendre public le code source de son calculateur d'impôts (ce qui est devenu effectif le 1^{er} avril 2016). On comprend la nécessité de pouvoir reprendre et vérifier la validité de cet algorithme. Par ailleurs, la loi n° 2016-1321 du 7 octobre 2016 pour une République numérique prévoit la communication des programmes pour en évaluer la loyauté. Il n'est pas possible de construire des machines complexes comme des robots sans pouvoir proposer une forme de loyauté quant à leur comportement. Pour y parvenir, il faut donc aussi que les outils du TAL se dotent d'une telle évaluation.

Les approches numériques posent un niveau supplémentaire de difficulté quant au problème de loyauté. Il n'est pas possible d'évaluer un programme fondé sur des méthodes d'apprentissage sans disposer également des jeux de données utilisés pour cet apprentissage. De fait, il faut rendre disponible ces jeux de données, qu'ils soient

9. Pour une recension de l'ouvrage de D. Cardon, nous renvoyons le lecteur à (Amblard, 2016).

constitués de données pour la science ou de données privées. La question de la loyauté des algorithmes implique aussi celle de l'évaluation de cette loyauté, et donc l'évaluation des systèmes qui les implémentent (programmes et données).

6. Institutions autours de l'éthique

Il ne faudrait néanmoins pas croire qu'embrasser une vision éthique et en profiter pour définir une norme est la solution. L'éthique ne doit pas nous faire tomber dans l'illusion de la conceptualisation des problèmes. La partie précédente a mis en avant plusieurs difficultés tant pour l'IA que pour le TAL : la chaîne de responsabilités, la considération de ces problèmes dans un cadre législatif et la loyauté des logiciels. Rassembler la communauté dans une instance en charge d'évaluer les propositions pourrait répondre au problème. C'est en tout cas la manière dont de nombreuses communautés se sont organisées face à la gestion de problèmes conceptuels et transverses difficiles à appréhender. On observe qu'il existe plusieurs comités d'éthique, généralement associés aux EPST (établissements publics à caractère scientifique et technique). Celui qui se rapproche le plus de ces problématiques est le COMETS (comité d'éthique du CNRS) qui a publié un rapport important sur l'éthique en sciences et technologies de l'information et de la communication, (Mariani *et al.*, 2009).

La Commission nationale de l'informatique et des libertés (CNIL) a pour mission la gestion des données personnelles. La loi n° 2016-1321 du 7 octobre 2016 pour une République numérique lui transfère des missions supplémentaires, ce qui est revendiqué par la CNIL dans son rapport d'activité (CNIL, 2015). Par ailleurs, INRIA a mis en place son propre comité COERLE (Comité opérationnel d'évaluation des risques légaux et éthiques) qui reprend les problématiques de la recherche en sciences du numérique, ouvert aux questions légales, la fonction restant assez proche de celle de conseil pour la direction. On trouvera dans le rapport (Dowek *et al.*, 2009) une analyse fine des enjeux qui nous intéressent. Ce rapport avec celui de la COMETS (Mariani *et al.*, 2009) ont préfiguré la création de la CERNA (Commission de réflexion sur l'éthique de la recherche en sciences et technologies du numérique d'Allistene), instance commune aux membres de l'alliance Allistene – l'alliance des sciences et technologies du numérique (CEA, CNRS, grandes écoles, INRIA, Télécom Paris Tech, universités). Son premier rapport (CERNA, 2014) présente de manière complète les problématiques de la robotique comme mentionné précédemment. Un second rapport sur l'apprentissage est en préparation.

Cette manière de procéder donne des résultats importants et est acceptée. Mais il ne faudrait pas laisser croire que le problème se résout par la mise en place d'une instance. Ces dernières sont nécessaires lorsque les communautés sont assez matures pour émettre des recommandations quant à la gestion des aspects éthiques de leurs objets. La communauté du TAL est en train de se diriger vers ces questions fondamentales. Il est probablement temps que l'ensemble des recherches, et donc des chercheurs qui portent ces recherches, se positionne sur leur impact sociétal désiré, potentiel ou probable. Un exemple dans cette direction est la Charte éthique et Big

Data, (Couillault *et al.*, 2014) qui ne vise pas à contraindre une perspective de recherche mais bien à donner une projection sociale et sociétale acceptable autour du TAL. On voit d'ailleurs que des rapprochements des problématiques ont lieu entre IA et TAL avec l'organisation de plusieurs journées d'étude en commun, par les associations scientifiques de ces grands domaines (ATALA et AFIA[10]).

Par ailleurs, les questions de la prise en compte des robots, de l'IA et du TAL dans une pensée englobante sont généralement présentées sous l'angle de la singularité de la relation entre l'homme et la machine. Cela laisserait entendre qu'il y aurait un combat sur l'occupation du territoire (que ce soit physique ou intellectuel) possible entre l'homme et la machine. Mais généralement, nous oublions dans cette réflexion que nous disposons déjà d'un conflit entre l'homme et l'homme, avec des volontés de destruction mutuelles et des enjeux stratégiques et politiques. Dans ce cadre, définir ce qui serait positif pour la communauté des hommes apparaît comme beaucoup plus délicat. Si l'on reprend l'exemple de la détection des opinions majoritaires ou des leaders d'opinion, la problématique ne revêt pas la même signification en temps d'état d'urgence. Mais dans le mouvement inverse, il est réducteur de présupposer qu'une relation s'installerait entre la communauté des machines et celle des humains.

7. Les approches numériques sont-elles interprétables ?

Nous devons encore aborder la relation qu'entretiennent TAL et IA avec d'autres sciences. Un grand nombre d'entre elles sont transformées tant dans leurs pratiques que dans leurs objets de recherche par les nouveaux usages de l'informatique. Une part de cette transformation s'appuie sur l'informatique comme outil, par exemple au travers des réseaux sociaux qui facilitent les relations dématérialisées entre chercheurs.

Au-delà des aspects techniques, l'informatique permet d'aborder la recherche plus globalement. En tant qu'outil, elle a déjà apporté des transformations majeures grâce à l'augmentation du volume de calculs[11]. Du côté du droit et des sciences humaines et sociales, cette influence se retrouve dans les humanités numériques, ou *digital humanities* comme nous l'avons précédemment introduit. Il est faux de croire qu'il n'y a pas non plus de mouvement de balancier vers l'informatique. Par exemple, elle est interrogée sur la fouille de données et l'extraction d'informations, qui relèvent des tâches qui nous intéressent car un grand nombre de disciplines du champ des lettres et des langues travaillent à partir de données exprimées en langue naturelle. De fait, avant de pouvoir réaliser de véritables recherches d'informations, il faut mettre en œuvre des méthodes de TAL. Mais il y a souvent un présupposé quant à l'existence de ces méthodes qui n'est pas toujours réalisé. En effet, s'il est raisonnable de produire des corpus segmentés en mots, annotés en marques morphosyntaxiques (*POS*) de manière automatique, il est moins réaliste de chercher tous les extraits correspondant à un

10. `http://www.atala.org` et `http://www.afia.asso.fr`

11. Il ne s'agit pas de modifier la notion de calculabilité, mais de réaliser des calculs dont la technicité nécessite un temps trop important.

concept particulier dans le corpus d'un auteur, par exemple les questions de logique modale chez Pascal, ou de témoignages chez Patrick Modiano. Dans ces cas, il nous faudrait disposer de représentations pragmatiques conceptuelles de plus haut niveau.

Utiliser les outils du TAL dans ce type de contextes étend le champ des questions en mobilisant de nouveaux terrains de la discipline. Cela justifie de son utilité pratique et théorique. On observe qu'actuellement un travail de transfert technique et théorique vers les humanités est un vrai défi. Mais on constate également le mouvement inverse c'est-à-dire de mobiliser les sciences humaines et sociales pour justifier voire expliquer ce qui est identifié par les modèles mathématiques. C'est d'ailleurs un des éléments qui permettent de passer du TAL vers, par exemple, des interprétations cognitives.

Une vraie difficulté théorique et conceptuelle émerge. Les méthodes numériques à l'origine des IA de dernière génération, sont éloignées de modèles interprétatifs du comportement humain. Évidemment, les méthodes symboliques portent en elles-mêmes leur modèle d'interprétation (mais pas en relation avec le comportement humain). C'est à la fois ce qui permet de les développer massivement, mais également ce qui rend difficile de les adapter à des contextes généraux. La théorie mathématique à la base des méthodes d'apprentissage est de chercher des corrélations numériques dans de grandes masses de données. L'apparition de ces corrélations prend du sens (a une interprétation) à la surface de ses usages. Mais il reste plus compliqué de définir ce qui est appris quand l'apprentissage est réalisé par auto-apprentissage. Le réseau de neurones ainsi constitué montre sa capacité de jouer au go. Pour le moment nous ne savons pas transférer des SHS vers la modélisation. Si nous n'avons pas de modèle explicatif pour les modèles numériques, regarder ce qu'ils simulent ne suffit pas à comprendre ce qu'ils font.

8. Conclusion

Dans cet article nous avons souhaité revenir sur le positionnement actuel du TAL sur les questions qui relèvent de l'éthique. Pour cela nous avons cherché à le situer dans une perspective épistémologique par rapport à l'IA. En effet, l'IA est traversée par des questionnements sociétaux depuis de nombreuses années et la réflexion est déjà portée par plusieurs instances. Il semble que les questions pour le TAL aient été absorbées *de facto* par ces instances, bien qu'elles ne relèvent pas tout à fait des mêmes champs disciplinaires. Nous avons donc cherché à les situer plus spécifiquement.

Il ne faut pas négliger que ces questions sont posées dans un contexte particulier. Un mouvement international de défiance envers l'IA par des figures de la technologie comme Elon Musk. En face, les réactions sont souvent difficiles à justifier car motivées par des réactions naïves. On rassemble sous l'acceptation de transhumanistes ces détracteurs. Leur théorie est que l'apparition d'une IA de haut niveau aura des capacités supérieures à celles des humains et que sa première tâche sera de se retourner contre les humains. Il s'agirait donc de nous préparer à l'émergence d'un tel système

en élaborant des stratégies de lutte contre elle, en présupposant qu'une IA serait nécessairement contre les humains[12].

Face à cette vision radicale de la relation entre humains et machines, il ne faut pas non plus travailler à intégrer les robots et les IA massivement dans la vie des humains sans réfléchir davantage. Est-ce que je conserve le droit de refuser un robot à ma table ? Derrière cette question d'une grande simplicité se pose un problème éthique majeur. Devons-nous accorder à ce point une réalité sociale aux robots, parce qu'ils seraient capables de communiquer en langue naturelle avec moi ? Nous ne souhaitons pas défendre une thèse plutôt qu'une autre, mais bien pointer qu'un choix politique émerge ici, car la projection des choix que nous faisons entraîne vers des futurs différents.

L'un des arguments que nous avons repris est que tant pour le TAL que pour l'IA les outils et les méthodes participent à définir la relation des humains à la technologie. Les choix réalisés par les chercheurs participent, même inconsciemment, à un projet politique par rapport auquel ils doivent se situer. Nous ne cherchons pas à les définir ni à les qualifier, mais bien à revendiquer la conscience d'y participer. L'exemple de l'identification de leaders sur les réseaux sociaux arrivant entre les mains d'une dictature obligerait à se poser sérieusement la question de l'impact de nos recherches sans nous réfugier derrière notre statut de chercheur [13].

Nous ne prétendons pas avoir fait le tour des questions éthiques du TAL et de l'IA. L'intelligence, le raisonnement, l'interaction existent aussi dans le fait que les systèmes sont incarnés et pas seulement conceptuels. Par ailleurs nous n'avons pas abordé des questions aussi diverses que le problème du coût économique et écologique de faire tourner des systèmes intelligents qui eux non plus ne vivent pas dans les nuages (le *cloud*) mais sont réalisés dans des machines physiques. De nombreuses questions sont ouvertes et participeront à structurer les évolutions de la discipline dans l'attente d'une réflexion plus structurante pour la communauté.

9. Bibliographie

Amblard M., « Retour sur le livre de Dominique Cardon », *Interstices*, janvier, 2016.

Amblard M., Fort K., Musiol M., Rebuschi M., « L'impossibilité de l'anonymat dans le cadre de l'analyse du discours », *Journée ATALA éthique et TAL*, Paris, France, novembre, 2014.

Bensoussan A., *Charte des droits des robots*, Lexing, septembre, 2015.

Cardon D., *A quoi rêvent les algorithmes*, Seuil-La République des idées, 2015.

Cassin R., Sauvé J.-M., Stirn B., Toutée H., de Lamothe O. D., Pêcheur B., Martin P., Vigouroux C., de Saint Pulgent M., Seners F., Richard J., *Le numérique et les droits fondamentaux*, Les rapports du conseil d'état, janvier, 2015.

CERNA, Éthique de la recherche en robotique, Rapport de recherche, ALLISTENE, 2014.

CNIL, *Rapport d'activité CNIL*, La documentation française, 2015.

12. Ce qui n'exclut pas d'apprendre à se rebeller pour ne pas obéir à des ordres destructeurs.

13. Toute ressemblance avec des événements réels et corroborés serait parfaitement fortuite.

Couillault A., Fort K., Adda G., De Mazancourt H., « Evaluating Corpora Documentation with regards to the Ethics and Big Data Charter », *International Conference on Language Resources and Evaluation (LREC)*, Reykjavik, Islande, mai, 2014.

Dowek G., Guiraud D., Kirchner C., Métayer D. L., Oudeyer P.-Y., Rapport sur la création d'un comité d'éthique en Sciences et Technologies du Numérique, rapport technique, Inria, 2009.

Eshkol I., Baude O., Kanaan L., Maurel D., Dugua C., « Procédure d'anonymisation et traitement automatique : l'expérience d'ESLO », *Journée ATALA, Ethique et TAL*, Paris, 2014.

Frey C. B., Osborne M. A., Holmes C., Rahbari E., Garlick R., Friedlander G., McDonald G., Curmi E., Chua J., Chalif P., Wilkie M., *Technology at work v2.0 : The Future Is Not What It Used to Be*, CityGroup and University of Oxford, january, 2016.

Gödel K., « Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I », *Monatshefte für Mathematik und Physik*, vol. 38, n° 1, p. 173-198, 1931.

Grouin C., Griffon N., Névéol A., « Étude des risques de réidentification des patients à partir d'un corpus désidentifié de comptes-rendus cliniques en français », *ETeRNAL, TALN*, 2015.

LeCun Y., Bengio Y., Hinton G., « Deep learning », *Nature*, vol. 521, n° 7553, p. 436-444, 2015.

LeCun Y., Kavukcuoglu K., Farabet C. *et al.*, « Convolutional networks and applications in vision. », *ISCAS*, p. 253-256, 2010.

Longo G., « Parsing the Turing Test : Philosophical and Methodological Issues in the Quest for the Thinking Computer », *Laplace, Turing and the "Imitation Game" Impossible Geometry*, Springer Netherlands, Dordrecht, p. 377-411, 2009.

Manning C. D., « Computational Linguistics and Deep Learning », *Computational Linguistics*, vol. 41, n° 4, p. 701-707, 2015.

Mariani J., Besnier J.-M., Bordé J., Cornu J.-M., Farge M., Ganascia J.-G., Haton J.-P., Serverin E., *Pour une éthique de la recherche en STIC*, Rapport du Comets, novembre, 2009.

Marquis P., Papini O., Prade H., « Some Elements for a Prehistory of Artificial Intelligence in the Last Four Centuries », *ECAI 2014 - 21st European Conference on Artificial Intelligence, 18-22 August 2014, Prague, Czech Republic*, p. 609-614, 2014.

Mazancourt H. D., Couillault A., Adda G., Recourcé G., « Faire du TAL sur des données personnelles : un oxymore ? », *ETeRNAL, TALN*, juin, 2015.

Müller V. C., Simpson T. W., Killer robots : Regulate, don't ban, rapport technique, University of Oxford, Blavatnik School of Government Policy Memo, 2014.

Russell S. J., Norvig P., *Artificial Intelligence : A Modern Approach, second edition*, Pearson Education, 2003.

Searle J., « Minds, Brains and programs », *The Behavioral and Brain Sciences, Cambridge University Press*, 1980.

Turing A., « Computing machinery and intelligence », *Mind*, vol. 59, n° 236, p. 433-460, 1950.

Veith C., Bandlow M., Harnisch M., Wenzler H., Hartung M., Hartung D., *How Legal Technology Will Change the Business of Law*, The Boston Consulting Group, January, 2016.

Wainer J., Facilitating collaboration among children with autism through robot-assisted play, thèse de doctorat, University of Hertfordshire, United Kingdom, 2012.

Wareham M., « Pourquoi doit-on interdire les "robots tueurs" », *Revue internationale et stratégique*, vol. 96, p. 97-106, 2014.