



Real Humans : des machines qui parlent comme des hommes, ou presque...

Maxime Amblard

► **To cite this version:**

Maxime Amblard. Real Humans : des machines qui parlent comme des hommes, ou presque.... Interstices, INRIA, 2013. hal-01281272

HAL Id: hal-01281272

<https://hal.inria.fr/hal-01281272>

Submitted on 1 Mar 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Real Humans : des machines qui parlent comme des hommes, ou presque...

Maxime Amblard

Cet article a été publié dans la revue *interstice* le 10 décembre 2013

Et si des robots ressemblant aux humains cohabitaient avec eux, comme dans la série *Real Humans* ? Il faudrait alors leur parler, qu'ils soient capables de comprendre ce qu'on leur dit et de s'exprimer à leur tour par la parole.



FIGURE 1 – *Real Humans* - 100% humains (Äkta Människor) / Photo Johan Paulin.

Série suédoise créée par Lars Lundström, diffusée sur SVT1 depuis 2012, en France sur ARTE depuis 2013.

Co-production : Matador Film, Sveriges Television, avec la participation d'ARTE France (durée des épisodes : 58 min.)

La série TV *Real Humans*, diffusée en France sur ARTE, met en scène des robots ayant une apparence humaine, appelés des hubots (contraction de human robots). Ces robots, aidant les humains dans leurs différentes tâches quotidiennes, ont une particularité : ils sont doués d'une intelligence très développée. Et l'un des enjeux narratifs de la série porte sur l'autonomie que leur confère cette faculté.

La situation imaginée n'est pas nouvelle, elle est analogue aux premiers fantasmes technologiques qui ont nourri l'imaginaire collectif depuis les an-

nées 1960. L'une des premières formes d'intelligence robotique mise en scène par le cinéma est HAL 9000, le robot du film 2001, l'Odyssée de l'espace de Stanley Kubrick, sorti en 1968 et inspiré d'une nouvelle écrite par Arthur Clarke en 1951. Dans ce film, Kubrick décrit notamment le voyage en vaisseau vers Jupiter d'une mission scientifique. Les astronautes coopèrent facilement avec HAL, le robot doté d'intelligence artificielle qui les accompagne. Jusqu'au jour où HAL se voit menacé de déconnexion. Là, il voit rouge et s'attaque directement aux spationautes. Cette vision proposée des usages de la technologie a suscité de nombreuses controverses.

Plusieurs films ou livres d'anticipation décrivent des conflits opposant les hommes aux machines. Il n'en reste pas moins vrai qu'au début des années 2000, le monde imaginé par Stanley Kubrick n'était pas advenu, ou pas encore ? Nous n'avions pas de mission habitée vers Jupiter, certainement pas la capacité de le faire et encore moins une intelligence artificielle capable de se rebeller contre l'homme. Sauf dans la fiction. Et, si 2001, l'Odyssée de l'espace a donné corps à un champ de réflexion, il en va certainement de même pour la série Real Humans.

Les scénaristes de la série Real Humans ont choisi de s'inscrire dans la lignée des fictions qui mettent en scène des robots à l'image de l'homme. Ils ont cependant introduit des éléments qui permettent immédiatement au public de distinguer les personnages humains des hubots : une rigidité de la démarche et du visage, et surtout, le regard. En effet, les hubots ne clignent pas des yeux ! Ce qui est technologiquement invraisemblable, puisque même les poupées pour enfants ont des yeux articulés... Alors, si on savait réaliser un système qui pense par lui-même et pour lui-même, on devrait être capable de concevoir des paupières qui se ferment. Mais c'est un souci constant des gens de l'image de faire en sorte que le spectateur ne se trouve pas exactement dans la même position que les personnages. Comme les hubots sont joués par des acteurs humains, un fort enjeu de la mise en scène est de gommer le caractère humain de manière homogène pour tous les personnages robots. Dans une autre fiction déjà ancienne, les envahisseurs, qui n'étaient pas des robots mais des extra-terrestres, eux, ne pouvaient pas plier le petit doigt. Un détail subtil, mais imparable.

Si cette série ne réinvente pas les enjeux, elle soulève de nombreuses interrogations, non seulement sur le développement technologique, mais aussi sur la condition humaine. Elle invite à se poser la question de la finalité de la technologie qui, dans le même temps, développe des systèmes suppléant à notre vieillissement ou nos handicaps, et des armes de plus en plus intrusives (conduisant à une surveillance permanente, comme dans la série Person of Interest) ou des drones.

De nombreuses questions ouvertes

La conception d'une intelligence artificielle d'aussi haut niveau embarquée dans une machinerie physiologiquement comparable à celle de l'Homme pose un grand nombre de questions de nature différente. Balayons-en rapidement quelques-unes.

Tout d'abord, où se situe l'intelligence ? La question n'est pas tant de savoir dans quelle partie de l'anatomie du robot l'intelligence doit être placée, mais plutôt de savoir si c'est un robot isolé qui est intelligent, grâce à une intelligence embarquée, ou si c'est une collection de robots qui résolvent une tâche en commun. Dans ce second cas, l'intelligence est dite collective. La résolution d'une tâche peut alors provenir de la somme des robots et non d'un seul, il s'agit d'une intelligence décentralisée.

Du point de vue technologique, comment reproduire les textures de la peau, les maintenir et les faire évoluer ? Comment produire des mouvements qui ressemblent à ceux des humains, ou du moins, qui s'adaptent à l'humain ? Il nous serait en effet inconcevable d'utiliser un robot d'aide qui ne sache réaliser que des mouvements violents, pouvant nous blesser... Cet aspect relève de la mécatronique.

Autre question technique, comment faire en sorte que ces robots aient une autonomie énergétique suffisante ? Cette faiblesse est d'ailleurs très bien retranscrite dans la série, où finalement pour survivre, les hubots restent à la recherche de courant électrique, comme les humains en situation de survie cherchent de la nourriture.

Comment adapter le robot à une situation non prévue, de manière analogue à l'homme ? Comment rendre le robot empathique ? Comment le doter d'états émotionnels, et les présenter à l'utilisateur ?

Du point de vue philosophique, est-il plus facile ou au contraire plus difficile pour l'être humain d'accepter d'être aidé par un robot qui lui ressemble ? Qu'est-ce qui différencie un robot humanoïde doté d'intelligence artificielle d'un être humain ? Faut-il considérer qu'un robot intelligent, humanoïde ou non, est aussi responsable qu'un humain ?

La problématique sur laquelle nous avons choisi de nous focaliser dans ce document est la suivante : comment faire communiquer ces robots au travers de la langue naturelle, c'est-à-dire à la fois pour la compréhension et la production de cette langue ? C'est en effet le processus d'interaction utilisé entre les humains et les hubots. Finalement, sans traitement de la langue, l'intelligence sophistiquée des hubots ne serait pas au service des humains.

Des robots communiquant par la parole

Ce qui est impressionnant - tel était déjà le cas dans 2001, l'Odyssée de l'espace - c'est que la forme d'intelligence artificielle présentée dans la

série Real Humans a une faculté de langage à l'égal des humains. Elle est capable tout autant de produire des énoncés dans une langue donnée que de les comprendre. Il est facile de constater que nous n'en sommes pas là aujourd'hui... Certes, l'apparition d'interfaces homme-machine en langue naturelle dans les téléphones portables pourrait le laisser croire. Mais ces systèmes sont efficaces uniquement parce qu'ils évoluent dans un univers restreint et ne proposent donc qu'un nombre très réduit d'interactions.

Mais alors, sommes-nous loin de ces échanges naturels avec les machines ? Quelles sont les étapes à franchir, qui nous permettraient de dialoguer avec une intelligence artificielle de manière fluide, comme nous pouvons le faire entre humains ?

Traitement automatique des langues

Pour une machine, la difficulté à comprendre la langue naturelle provient en partie du nombre important de phénomènes linguistiques en jeu dans une simple interaction. Et pour chaque sorte de phénomènes, plusieurs interprétations sont possibles. Par exemple, comment distinguer les deux phrases suivantes, en anglais, qui se prononcent d'une façon très similaire : « wreck a nice peach » ou « recognize speech » (« détruire une belle pêche » ou « reconnaître la parole ») ? Seul le contexte d'énonciation nous permet de discriminer la bonne version.

La grande différence entre la langue naturelle et les langages formels, tels qu'on les utilise largement en informatique, est qu'un énoncé en langue naturelle ne peut se comprendre qu'au travers de toutes ses dimensions, et chacune de ces dimensions apporte son lot d'ambiguïtés :

- phonologie : signal sonore
- morphologie : structuration interne des mots
- syntaxe : relation des mots au sein d'une phrase
- sémantique : interprétation des items de sens composés
- pragmatique : interaction avec le contexte
- et au-delà, en relation avec nos connaissances du monde

En tant qu'humains maîtrisant une langue, nous réalisons les opérations de désambiguïsation nécessaires sans même y réfléchir.

Le problème de l'ambiguïté

Sans entrer dans une liste à la Prévert des problèmes rencontrés, on peut comprendre qu'il n'est possible de déterminer la catégorie syntaxique d'un mot que dans le contexte de son utilisation, ici dans une phrase. Par exemple le mot « la » peut être :

- un déterminant, comme dans le syntagme « la pomme »,
- un pronom, dans la phrase « Jean la mange »
- un nom, eh oui, c'est possible : « le la était désaccordé ».

Ces choix sont primordiaux pour trouver les différentes relations syntaxiques qui existent. L'exemple suivant est très célèbre parmi les spécialistes du traitement automatique des langues : « La petite brise la glace » peut se comprendre avec comme sujet « la petite » ou bien « la petite brise », pour un petit vent froid. Dans le premier cas, le verbe est « brise », dans l'autre c'est « glace ». Le sens sera très différent en fonction de ce que l'on aura choisi au niveau syntaxique.

Un autre niveau d'ambiguïté vient de l'interprétation en fonction des connaissances du monde. Par exemple, « la pièce est dans le porte-monnaie » et « le porte-monnaie est dans la pièce » sont deux énoncés contenant les mêmes mots, avec les mêmes relations syntaxiques, et pourtant ici, le sens de « pièce » diffère considérablement. Dans le premier cas, il s'agit d'une pièce de monnaie et dans le second, d'un espace d'un bâtiment. Si cet exemple ne pose aucun problème d'interprétation à un humain, pour un système automatique, c'est d'une difficulté sans commune mesure, puisqu'il faut être capable d'abstraire de l'énoncé toutes les informations jusqu'à les mettre en relation avec nos connaissances du monde.

Actuellement, les recherches en phonologie, morphologie et syntaxe sont très avancées. Du point de vue linguistique, les phénomènes usuels des langues indo-européennes (comme le français ou l'anglais) sont nettement étudiés. Ce qui permet de définir des systèmes efficaces pour en rendre compte. Cependant, il existe encore de nombreuses langues, notamment en Afrique et en Amérique du Sud, pour lesquelles il n'existe quasiment pas d'étude (Koa-lib, Murucundu...). Du point de vue calculatoire, les outils abstraits pour en rendre compte évoluent. Il n'en reste pas moins que la question de la classe de langages formels qui recouvre la langue naturelle n'est toujours pas définie. On suppose qu'elle est faiblement sensible au contexte, sans pouvoir définir précisément en quoi elle diffère des langages sensibles au contexte.

Une fois résolus les problèmes de syntaxe, beaucoup de questions subsistent sur l'interprétation du contenu des phrases, ainsi que sur la relation des phrases entre elles dans la construction d'un texte cohérent. De nombreux phénomènes sont en jeu, tant dans la modélisation sémantique que pragmatique. L'une des difficultés actuelles est d'arriver à déterminer ce qui dans la modélisation est pertinent. Une fois cela déterminé, alors nous retombons sur le problème de gestion des ambiguïtés dans ce nouveau niveau de représentation. Ce qui pourrait se résoudre en partie, si nous avions une manière d'interroger une représentation de nos connaissances du monde... Cependant, de nombreuses équipes de recherche se concentrent actuellement sur la question de la modélisation de la pragmatique à partir des recherches en sémantique (que ce soit sur la sémantique lexicale, ou en utilisant des propriétés logiques...).

La réactivité des systèmes

On peut identifier deux approches pour répondre à ces problèmes d'ambiguïté : les approches symboliques (à base de règles) et les approches numériques (à base des probabilités). De manière générale, les approches symboliques ont explosé dans les années 1980, car elles permettaient de traiter quelques cas particuliers de besoins avec les ordinateurs domestiques. C'est à ce moment que l'on a vu apparaître les premiers correcteurs orthographiques. Leur grande force est qu'elles sont capables de rendre finement compte de propriétés linguistiques complexes. Mais leur grand défaut est qu'elles nécessitent une capacité et un temps de calcul très significatif. De l'autre côté, les approches numériques sont beaucoup plus performantes pour la recherche d'une solution, mais il devient très difficile, sinon impossible, d'introduire des propriétés linguistiques dans le traitement. Et ces systèmes se retrouvent quelque part entre langage formel et langue naturelle.

En revenant sur la question précédente de l'analyse syntaxique, les analyseurs syntaxiques (parsers) à base de règles sont actuellement capables de reconnaître une phrase standard d'article de journal en une quinzaine de secondes. Ce qui est une performance, mais qui peut paraître un temps extrêmement long pour un utilisateur de système. Les meilleurs systèmes peuvent descendre légèrement en dessous, lorsqu'ils sont très spécialisés pour une langue particulière, comme le français ou l'anglais. Reste malgré tout la question de déterminer parmi tous les résultats trouvés lequel est le meilleur. À nouveau, cela dépend très largement du nombre de mots présents.

De l'autre côté, les parsers basés sur des approches numériques obtiennent des performances remarquables, en dessous de la seconde. Il n'en reste pas moins que pour des phrases utilisant des structures linguistiques complexes comme dans les articles de journaux, la fiabilité du résultat proposé est très réduite. C'est en fonction du contexte d'utilisation que la granularité de la qualité de la réponse apportée va être acceptable ou pas.

Alors, le grand défi pour définir des outils pertinents et efficaces réside dans l'hybridation des méthodes, tirant parti à la fois de la rapidité de traitement des approches numériques et de la qualité de réponse des approches symboliques. Les signes d'une convergence des approches sont nombreux, qu'ils proviennent de prises de position d'éminents chercheurs de la thématique ou de groupes industriels majeurs.

La génération

On entend bien ici que les hubots sont capables de percevoir un signal, de le décomposer en morceaux de sons cohérents, d'inférer pour chacun un mot avec toute sa construction interne (genre, nombre, cas...) et que pour un ensemble de mots, ils sont capables de construire des phrases, d'en inférer des représentations du sens et d'interpréter le tout dans un ensemble de

connaissances du monde. Et de le faire en temps réel ! Nous venons de voir que cela n'est actuellement pas réaliste. Mais même si tel était le cas, le problème n'en serait pas réglé pour autant.

Si les ambiguïtés existent dans cet enchaînement de représentations vers l'abstraction, que l'on appelle sens de l'analyse, elles existent tout autant dans l'autre sens : d'une représentation abstraite vers sa concrétisation sonore. On appelle cet autre sens celui de la génération. Il est évident qu'on n'attend pas seulement des hubots qu'ils soient capables de comprendre ce qu'on leur raconte, mais qu'ils soient en plus capables de nous transmettre une information. Recommencent alors les difficultés et les ambiguïtés dans les passages d'un mode de représentation à un autre.

Dans une représentation simplifiée des choses, cette présentation suit le principe de la traduction automatique. Dans cet exercice, on part d'un énoncé, un système l'analyse jusqu'à en comprendre le contenu le plus abstraitement possible, puis le décline dans une autre langue. Chacune des langues — la langue de départ (source) et la langue d'arrivée (cible) — est alors vue comme régie par des spécificités. Le passage de l'une vers l'autre se fait alors par un langage pivot, le plus abstrait possible.

Tout utilisateur d'un moteur de traduction automatique a déjà pu constater que ces systèmes sont loin d'être parfaits. Rappelons qu'en 1964 le rapport ALPAC condamnait la traduction automatique sur l'exemple devenu fameux de double traduction anglais-russe-anglais : « the spirit is willing but the flesh is weak » (l'esprit est ardent mais la chair est faible) qui devient « the vodka is strong but the meat is rotten » (la vodka est forte mais la viande est pourrie). Malgré de grandes faiblesses, les systèmes actuels sont d'une qualité très supérieure.

Le problème de l'ambiguïté, abordé précédemment dans le sens de l'analyse, se retrouve complètement dans le sens de la génération. Les différentes possibilités peuvent être vues comme des paraphrases. La question de la génération a été évoquée dans un document sur la génération automatique de textes, où des analyses en langue naturelle à partir de données bancaires sont automatiquement produites. Ce qui apparaît clairement, c'est que la qualité du résultat proposé est très bonne, au sens où ce qui est produit est parfaitement acceptable pour un locuteur natif. Mais la contrepartie pour obtenir cette grande qualité est d'utiliser un domaine très restreint et une diversité linguistique très limitée. Ce qui signifie que les phrases suivent des structures très prototypiques où la coordination, les subordonnées relatives, ou encore les gérondifs sont absents. D'un point de vue qualitatif, seuls les contextes très techniques donnent de bons résultats, ce qui implique que dans un contexte plus littéraire, ces systèmes ne sont pas fonctionnels. Et encore moins dans un contexte réel.

Et demain ?

S'il y a un point sur lequel nous sommes certains de ne pas nous tromper, c'est que la mise en production de machines, ou de robots, capables de communiquer par la langue comme de réels humains n'est pas une chose envisageable à court et moyen terme. Un dernier indice pour confirmer l'intuition... si nous avions de telles machines, eh bien nous aurions la capacité d'analyser et de générer en temps réel, et nous aurions donc des systèmes de traduction de langue à langue de parfaite qualité. Or si certains types d'ambiguïtés peuvent être gérés presque en temps réel, il est actuellement impossible d'envisager que toutes le soient en même temps avec un résultat réaliste.

Et bien sûr, n'oublions pas toutes les autres questions soulevées. Pour faire de nos robots des hubots, il faudrait avant tout qu'ils soient capables de penser par eux-mêmes... Cet aspect a déjà été discuté sur Interstices à partir de l'idée reçue Un cerveau artificiel, c'est pour demain.

Gageons que la science-fiction a encore de beaux jours devant elle, même si la science lui grignote du terrain tous les jours !