

Enceintes connectées : Enjeux d'une nouvelle interface homme-machine

Virginia Haussauer

Mémoire de Master en Sciences Sociales

Session : Été 2019

Directeur : Boris Beaudé

Expert : Olivier Glassey

Table des matières

1. Introduction.....	5
1.1 Une technologie au centre des enjeux du numérique	6
1.2 Problématique et plan.....	7
2. Un objet en contexte.....	11
2.1 Accélération et pression temporelle	12
2.2 Un idéal d'efficacité	17
2.3 Économie et technologies des données.....	21
2.4 Un contexte à nuancer	25
3. La voix pour interface	29
3.1 La voix à distance	29
3.2 De l'interaction à la relation.....	31
3.3 La promesse d'une interaction « sans friction ».....	34
3.4 Le revers de la simplicité.....	40
4. Les enjeux de la personnalisation	43
4.1 Réduction des possibles	44
4.2 Vers une économie de l'intention ?	48
4.3 Le paradoxe de la vie privée	52
5. Des usages imaginés	58
5.1 Les nécessaires imaginaires	60
5.2 Vendre l'enceinte connectée.....	61
6. Conclusions	70
6.1 Limites et perspectives de recherche.....	72
6.2 Quels futurs pour les interfaces vocales ?	73
Références.....	75
Annexe	83

1. Introduction

Intégrés il y a une dizaine d'années aux systèmes d'exploitation des *smartphones*, les assistants vocaux se déploient aujourd'hui dans l'univers domestique au travers de dispositifs fixes nommés « enceintes connectées ». Ces haut-parleurs équipés d'un micro et d'un logiciel de reconnaissance vocale proposent de délester l'utilisateur de toute une série d'actions quotidiennes. Grâce à une simple commande vocale, l'enceinte connectée permet de consulter la météo ou l'état du trafic, d'envoyer un message, de lancer une requête sur un moteur de recherche, de commander un produit, ou encore de contrôler d'autres objets connectés. Conçus par Google, Amazon ou Apple, ces « agents conversationnels » ont déjà conquis des dizaines de millions de foyers américains et tentent aujourd'hui de séduire un public toujours plus large, auquel on promet un quotidien simplifié et « optimisé » grâce à l'interaction vocale.¹

Toutefois, l'adoption de ces objets intrigue tout autant qu'elle inquiète. Si les enceintes connectées semblent à même de faciliter de nombreuses tâches, elles suscitent également des craintes liées à l'intrusion d'un dispositif d'écoute dans la sphère privée. La presse spécialisée, tout comme les médias grand public, semblent partagés. « Enceintes connectées : faut-il les adopter ? », s'interroge le journaliste Nicolas Six dans *Le Monde*.² L'argument régulièrement avancé en faveur de leur adoption est le potentiel gain de temps réalisé grâce à la possibilité d'effectuer plusieurs tâches de façon simultanée. Toutefois, de nombreux doutes subsistent quant à la réelle efficacité du système et à la capacité des enceintes à garder confidentielles les conversations de leurs utilisateurs. Quelques cas médiatisés de failles sécuritaires ont répandu l'image d'un système peu fiable. Ainsi, l'enthousiasme pour des technologies

¹ Les fabricants ne communiquant pas leurs chiffres de vente, il est difficile de connaître le nombre exact d'enceintes connectées vendues. Selon un rapport de la société américaine *Consumer Intelligence Research Partners*, 66 millions d'enceintes connectées seraient installées aux États-Unis à fin 2018. Voir : <https://techcrunch.com/2019/02/05/report-smart-speaker-adoption-in-u-s-reaches-66m-units-with-amazon-leading/>, consulté le 20 mai 2019. Médiamétrie estime quant à elle à 1,7 millions le nombre d'utilisateurs en France. Voir : <https://www.mediametrie.fr/fr/plus-d17-million-dutilisateurs-denceintes-commande-vocale>, consulté le 20 mai 2019.

² Six, N. (2018). Enceintes connectées : faut-il les adopter ?. *Le Monde*, 13 juin 2018. Disponible sur : https://www.lemonde.fr/pixels/article/2018/06/13/enceintes-connectees-faut-il-les-adopter_5314162_4408996.html, consulté le 20 mai 2019.

vocales promises comme la nouvelle « révolution numérique » côtoie parfois la méfiance et le scepticisme à l'égard d'un dispositif au fonctionnement encore peu clair.

1.1 Une technologie au centre des enjeux du numérique

Bien que l'essentiel du débat médiatique se cristallise autour de quelques points – et se résume généralement à une pesée d'intérêt entre gain de temps potentiel et risque d'intrusion – l'ambivalence des discours autour des enceintes connectées témoigne plus généralement de la complexité des enjeux liés aux développements actuels des technologies de l'information et de la communication. L'enceinte connectée, objet au futur encore incertain, semble en effet au carrefour de diverses problématiques que soulève aujourd'hui le numérique.

Elle questionne tout d'abord la raison d'être de certaines innovations techniques. Pourquoi adopter une enceinte connectée ? A quel besoin répond-elle ? Si l'argument du gain de temps est couramment mobilisé, que dit-il de notre rapport à la technique et de la place qu'elle occupe dans notre société ? L'enceinte connectée interroge ensuite la façon dont nous pouvons et souhaitons interagir avec la technologie. Est-il possible de « dialoguer » avec une machine ? Quelles en sont les limites ? Désirons-nous converser avec des machines comme avec des êtres humains ?

Enfin, si l'on estime qu'elle présente certains bénéfices, nous pouvons encore questionner les coûts d'une telle technologie. Quel prix consentons-nous à payer pour faire usage de ces services ? Si l'achat d'une enceinte représente un investissement financier relativement peu conséquent, son coût se situe ailleurs.³ Au même titre que de nombreux services basés sur la gratuité, ces objets fonctionnent grâce à la collecte d'une importante quantité de données personnelles, garante d'une expérience personnalisée et d'une interaction « fluide », mais qui implique en retour une surveillance toujours plus étendue de nos pratiques. Les enceintes connectées interrogent donc nécessairement la part de notre sphère privée que nous sommes prêts à céder en l'échange de ces services.

Aussi, les dispositifs qui reposent sur un traitement algorithmique posent la question de l'autonomie de l'utilisateur. Si l'enceinte connectée propose de soulager l'utilisateur en effectuant pour lui un choix parmi un ensemble d'options, elle limite

³ Les enceintes connectées affichent des prix bien plus bas que les *smartphones*. Amazon Echo et Google Home sont toutes deux vendues 99 Euros.

inévitablement sa liberté d'action et réduit le champ des possibles. Au cœur de ces nombreux questionnements, l'enceinte connectée semble donc constituer un point de départ particulièrement intéressant pour explorer différents enjeux posés par les technologies numériques contemporaines.

1.2 Problématique et plan

En substituant la parole au geste, les concepteurs d'enceintes connectées promettent une interaction fluide et « sans friction » entre l'être humain et la machine. Il s'agirait d'une communication intuitive rendue possible par l'oralité. A la facilité d'interaction avec la machine, s'ajoute la promesse d'un service personnalisé, et donc optimisé. Au travers d'un système d'apprentissage continu basé sur la collecte et la mise en réseau de nombreuses données, l'enceinte connectée a pour but de connaître et d'anticiper les besoins de ses utilisateurs. Elle se présente ainsi comme un dispositif d'aide au « pilotage » du foyer, permettant de réduire la quantité d'effort et donc le temps nécessaire à la réalisation de certaines actions. Les termes « assistant vocal », « *virtual assistant* » ou « *voice assistant* » fréquemment utilisés pour désigner les logiciels intégrés à ces enceintes, suggèrent que leur fonction concerne principalement la délégation de certaines tâches à la technique. Celles-ci peuvent être d'ordre mécanique (allumer la lumière) ou cognitif (rappeler l'heure d'un rendez-vous). L'argument avancé est que le transfert de ces actions à la machine délivre l'individu à la fois d'une action physique et d'une charge cognitive, libérant ainsi du temps pour d'autres activités.

En proposant d'optimiser la gestion du temps, l'enceinte connectée semble s'inscrire dans une problématique plus large qui concerne la raréfaction des ressources temporelles. Depuis une vingtaine d'années, de nombreux auteurs ont décrit ce qu'ils perçoivent comme un changement majeur de nos sociétés modernes. Que l'on parle de « pression temporelle » (Szollos, 2009), « *temporal density* » (Wajcman, 2015), « culte de l'urgence » (Aubert, 2003) ou « famine temporelle » (Rosa, 2013), l'argument central est celui d'un sentiment de malaise provoqué par une surcharge informationnelle à laquelle l'individu devrait répondre par une accélération sans fin, consistant à effectuer davantage de tâches dans un temps toujours plus réduit. La technologie, souvent présentée comme une solution à l'accélération, semble pourtant participer de cette pression temporelle. L'envoi d'un e-mail plutôt qu'un courrier permet de transmettre rapidement un message, mais il implique également un retour

dans un temps plus bref. Ce paradoxe s'explique par un double effet de la technique : alors qu'elle permet d'augmenter la vitesse de succession de nos actions, elle crée également de nouvelles temporalités et espaces d'expérience. Hartmut Rosa (2013) parle ainsi d'un système de *feedback* dans lequel s'emboîtent accélération technique, accélération du changement social et accélération du rythme de vie. Le cycle de l'accélération est particulièrement manifeste dès lors que « l'accélération technique peut être vue comme une *réponse sociale* au problème de raréfaction du temps » (Rosa, 2013, p.41).

L'enceinte connectée est présentée comme une nouvelle solution à ce manque de temps qui toucherait désormais toutes les sphères de notre existence, jusqu'à l'intimité du foyer. En offrant une interaction fluide, personnalisée et « sans contact », cette technologie activable à tout instant permettrait d'accélérer, voire d'automatiser toute une série d'actions quotidiennes. Ainsi, l'ensemble du dispositif est présenté dans une optique de rentabilisation et de compression du temps. Cet idéal d'efficacité se traduit par des choix techniques aux implications nécessairement sociales et politiques. Car la technique n'est jamais neutre et s'accompagne toujours d'un certain « rapport au monde », comme le formulait Heidegger. S'il semble de toute évidence impossible d'aborder l'ensemble de ces choix, nous nous concentrerons ici sur les deux attributs centraux des enceintes connectées : l'interface vocale et la personnalisation des services. La voix constitue ici un aspect essentiel, car il s'agit du seul mode d'interaction proposé. Auparavant réservée à la communication humaine, la parole est désormais au centre de l'interaction homme-machine. La personnalisation n'est quant à elle pas nouvelle, mais se trouve au cœur du fonctionnement d'un dispositif pensé pour s'adapter aux besoins et préférences estimés de chaque utilisateur, afin de le suivre au plus près de ses pratiques. Ces deux caractéristiques constituent des points d'entrée particulièrement intéressants, car elles permettent d'aborder de nombreux enjeux que soulève cette technologie.

Par ailleurs, les choix qui ont conduit au développement de ces objets sont également façonnés par des logiques économiques, sociales et politiques propres à un environnement spécifique. La compréhension du contexte de création des technologies est donc nécessaire pour saisir la façon dont s'opèrent certains choix, qui participent à reconfigurer notre expérience du monde social. Il s'agit alors de mettre en lien le contexte de production des enceintes connectées, les choix techniques qui sont opérés et les enjeux sociaux qu'ils soulèvent.

La première partie de ce travail aura donc pour objectif de mieux comprendre les conditions d'émergence de ces dispositifs techniques au travers de trois dimensions : (1) un contexte global « d'accélération » et de raréfaction des ressources temporelles ; (2) une culture de l'efficacité et de la rationalisation du temps propre à l'environnement de conception de ces technologies (English-Lubeck, 2017; Wajcman, 2019) ; (3) des dynamiques techniques et économiques caractérisées par le développement du *machine learning* et d'une économie numérique basée sur la collecte massive de données.

Dans une seconde partie, c'est le principe de la voix comme interface qui sera interrogé. Nous verrons que le mode d'interaction proposé par les enceintes connectées s'inscrit dans une histoire plus longue de la synthèse vocale et de l'évolution des interfaces homme-machine. Toutefois, les systèmes d'assistance vocale proposés aujourd'hui se distinguent par différents aspects des technologies précédentes et posent de nouveaux questionnements. Ce chapitre visera à répondre aux interrogations suivantes : Quelles sont les promesses faites autour de ce mode d'interaction ? Dans quelle mesure les assistants vocaux participent-ils d'une nouvelle relation entre l'humain et la technique ? Quelles peuvent être les implications sociales du développement de technologies basées sur une interface vocale ?

La troisième partie de ce travail abordera les enjeux posés par la personnalisation toujours plus fine des services proposés par l'assistant vocal. Au cœur du fonctionnement de nombreux services numériques (sites d'achat en ligne, réseaux sociaux, plateformes de musique), l'individualisation des interactions est présentée comme une manière « d'optimiser » l'expérience de l'utilisateur en anticipant ses besoins ou ses préférences. Ce principe est rendu possible par la collecte et la mise en réseau de nombreuses données (comportement passé, localisation, interactions avec d'autres environnements, etc.) qui alimentent des algorithmes chargés d'effectuer un tri parmi l'ensemble des options possibles. Si cette personnalisation peut être perçue comme un filtre nécessaire dans un monde numérique où l'abondance d'informations dépasse nos capacités de traitement, elle limite également nos possibilités de réflexion et d'action. Par ailleurs, les assistants vocaux placent l'utilisateur dans une situation où le fonctionnement du dispositif dépend directement de la quantité de données personnelles qui lui sont fournies. Cette technologie questionne donc nécessairement notre rapport à la vie privée. Celui-ci répond à des logiques complexes que l'on tentera de mettre en lumière.

Contrairement aux études qui s'inscrivent dans le champ de la sociologie des usages, ce travail n'aura pas pour but d'observer la façon dont les individus s'approprient effectivement les technologies. La raison principale de ce choix tient au fait que l'introduction des enceintes connectées est relativement récente et leurs usages encore difficilement observables. Si les interfaces vocales sont présentées par les fabricants comme la prochaine « révolution numérique », leur adoption reste encore marginale en Suisse où les enceintes actuellement en vente ne proposent qu'une palette de fonctionnalités réduite.⁴ Une enquête ethnographique centrée sur les usages actuels et locaux des enceintes connectées (telle qu'elle pourrait être envisagée dans le cadre de ce travail) n'offrirait alors qu'une perspective très partielle. Toutefois, renoncer à une observation des usages ne signifie pas s'en détourner totalement. Car au travers des choix techniques, ce sont également des usages qui sont imaginés. C'est ce que Madeleine Akrich nomme le « script », soit un « scénario qui se veut prédétermination des mises en scène que les utilisateurs sont appelés à imaginer à partir du dispositif technique » (Akrich, 1987, p.3). L'objet, par ses spécificités fonctionnelles, engage certaines formes d'usages plutôt que d'autres. Ce principe avait déjà été formalisé par le psychologue James Gibson (1977) au travers de la notion d'*affordance*, désignant les potentialités d'actions qu'offre un dispositif.

L'analyse des caractéristiques des assistants vocaux nous amènera donc nécessairement à discuter les usages inscrits dans cette technologie. Par ailleurs, la notion « script », s'étend également aux discours qui accompagnent les objets techniques. Lorsque les produits sont présentés au public, les fabricants sélectionnent certains usages et mettent en scène l'utilisateur imaginé au travers de différents canaux (description du produit, publicités, modes d'emploi). La dernière partie de ce travail se concentrera donc sur les usages tels qu'ils sont actuellement pensés et communiqués par les différents fabricants. Car si les fonctionnalités des enceintes connectées sont multiples, nous verrons que la façon dont les usages sont imaginés peut varier selon les fabricants (Apple, Google, Amazon) dont les différentes cultures et modèles économiques traduisent également une approche différenciée de ces objets.

Ce travail, qui propose d'analyser un objet technique au travers de plusieurs entrées, mobilisera différents concepts théoriques, empruntant à la fois à la construction sociale des technologies (MacKenzie et Wajcman, 1985; Bijker, Hughes et Pinch, 1987), à la sociologie de la traduction (Akrich, Callon et Latour, 2006), à

⁴ <https://www.rts.ch/info/sciences-tech/9857484-les-enceintes-connectees-un-marche-juge-prometteur-en-suisse.html>, consulté le 6 mars 2019.

l'anthropologie des techniques et au champ des interactions homme-machine. Ces diverses approches permettront de mieux saisir ce qui se joue autour des enceintes connectées en articulant différentes échelles d'analyses, de la relation entre un individu et un dispositif technique à la prise en compte d'un contexte et d'enjeux socio-politiques plus larges. Il s'agira de tenter de saisir les liens entre un contexte de production particulier, une nouvelle forme d'interaction homme-machine, un principe de personnalisation qui repose sur une technique de *machine learning* et la collecte de nombreuses données personnelles, ainsi que la manière dont les usages et bénéfices de ces technologies sont pensés et communiqués.

2. Un objet en contexte

Dès les années 1980, le courant de recherche axé sur la construction sociale des technologies⁵ s'est attaché à montrer que la technique ne peut être dissociée des conditions sociales de sa conception (MacKenzie et Wajcman, 1985 ; Bijker et Law, 1992). Cette approche se distancie d'un déterminisme technique qui chercherait à observer les « impacts » de la technique sur la société. Le dispositif technique est davantage envisagé comme une « boîte noire » dans laquelle il est nécessaire d'entrer afin de saisir la façon dont le social est incorporé (*embedded*) à la fois au contenu des technologies et à leur processus d'élaboration (Williams et Edge, 1996). Toutefois, la notion de « construction sociale » recouvre, selon les auteurs, des phénomènes relativement différents (Sismondo, 1993). Certaines approches s'attachent à souligner, de façon générale, la nécessité de prendre en compte le contexte social, économique et politique de la conception des techniques (Nye, 1998). D'autres travaux, dans la lignée de la sociologie de la traduction initiée par John Law et Arie Rip, cherchent à comprendre la façon dont technique et société sont imbriquées en décrivant finement les relations entre les différents « agents » (humains et non-humains) impliqués dans de vastes réseaux socio-techniques (Callon, 1986 ; Akrich, Callon et Latour, 2006). De façon générale, les recherches qui s'inscrivent dans le champ de la construction sociale des technologies proposent le plus souvent une analyse centrée sur l'agentivité (*agency*) et n'accordent qu'une faible importance à l'environnement socio-politique dans lequel prennent place les technologies (Klein et Kleinman, 2002).

⁵ Le plus souvent désigné par son acronyme anglo-saxon SCOT (*social construction of technology*).

Ce chapitre, dédié au contexte de développement des assistants vocaux repose sur l'idée que la compréhension d'un objet technique passe également par l'analyse des dynamiques sociales, économiques et techniques propres à son environnement de conception. La première partie de chapitre tentera d'analyser la façon dont les assistants vocaux pourraient s'inscrire dans une tendance globale d'accélération et de raréfaction des ressources temporelles. Dans un second temps, il s'agira d'aborder plus spécifiquement le rapport au temps propre à l'environnement de conception de ces technologies. Nous terminerons par une analyse des logiques qui sous-tendent une économie numérique basée sur la collecte massive de données, ainsi que les progrès techniques réalisés dans le domaine du *machine learning*.

2.1 Accélération et pression temporelle

Depuis plus de deux décennies, de nombreux auteurs ont fait le constat d'un sentiment d'accélération globale du temps et de raréfaction des ressources temporelles. Les écrits de Paul Virilio (1995, 1996) ont alerté sur les conséquences sociales, politiques et culturelles d'une vitesse devenue le « moteur » du monde. Le « culte de l'urgence » (Aubert, 2003), après avoir investi le monde professionnel, toucherait aujourd'hui l'ensemble des sphères de la vie quotidienne. Hartmut Rosa (2010) identifie trois dynamiques de l'accélération : *l'accélération technique* qui concerne les domaines des transports, de la communication et de la production, *l'accélération du changement social*, c'est-à-dire les transformations qui touchent l'ensemble de la société au travers des institutions, du travail, des structures familiales, des valeurs, et *l'accélération du rythme de vie*, soit le sentiment de « famine temporelle » qui pousse les individus à réaliser toujours davantage d'actions en moins de temps. Cette dernière forme d'accélération, que l'on considère souvent comme caractéristique d'une société postmoderne, avait déjà été identifiée à la fin du XIX^e siècle par Georg Simmel dans son ouvrage *Les Grandes villes et la vie de l'esprit* (2013 [1903]). Il y décrit alors l'émergence d'une nouvelle temporalité propre à la métropole moderne. La ville, de par les nombreuses possibilités d'interactions qu'elle offre, soumettrait le citoyen à un phénomène d'« intensification de la vie nerveuse » qui le pousserait à accélérer toujours davantage le rythme de ses actions.

Aujourd'hui, cette tendance à l'accélération semble s'être encore intensifiée, alors même que nous disposons toujours davantage de moyens techniques pour gagner du

temps. Ces derniers semblent même participer de cette pression temporelle. Comment expliquer ce paradoxe ? Si la technique permet d'augmenter le nombre de tâches réalisées dans une unité de temps, elle crée en même temps de nouveaux espaces d'expériences. Comme le souligne Hartmut Rosa (2013, p.41), « l'accélération technique, qui est fréquemment liée à l'émergence de nouvelles technologies [...], apporte presque inévitablement tout une gamme de changements dans les pratiques sociales, les structures de communication et les formes de vie correspondantes ». Certains dispositifs conçus précisément dans le but de gagner du temps peuvent ainsi contribuer à créer de nouvelles pratiques, tout autant chronophages.

Dans son analyse du développement de l'industrie domestique aux États-Unis entre en 1860 et 1950, Ruth Schwartz Cowan (1983) montre comment des équipements conçus pour accélérer l'exécution de tâches ménagères n'ont pas permis d'alléger la charge de travail domestique. Cette contradiction s'explique par le fait que l'automatisation de certaines activités s'accompagne de l'émergence de nouveaux standards domestiques, nécessitant davantage d'effort. Ainsi, le gain de temps réalisé grâce au lave-linge, à la tondeuse à gazon ou au lave-vaisselle est compensé par l'effort investi pour répondre à de nouvelles attentes en matière de propreté, de soin et de confort. De la même manière, si la démocratisation de l'automobile aux États-Unis dès les années 1950 a permis à de nombreux individus d'effectuer plus rapidement certains trajets, elle a également contribué à la transformation de multiples pratiques sociales, de la consommation à l'habitat. Les logements se développent en périphérie des villes et les achats s'effectuent dans des espaces commerciaux éloignés des centres urbains. Le temps gagné grâce à l'augmentation de la vitesse de transport a ainsi été rapidement compensé par l'allongement des distances parcourues. Par ailleurs, l'attrait de l'automobile se traduit par une augmentation rapide du nombre de véhicules sur les routes. Malgré la construction d'imposantes infrastructures, le trafic continue d'augmenter, alimentant un cycle sans fin d'élargissement du réseau routier et de congestion. L'idée selon laquelle la voiture permettrait de gagner du temps peut donc être questionnée.

Le développement actuel des technologies de l'information et de la communication (TIC) s'inscrit dans ce processus d'accélération. L'instantanéité rendue possible par l'augmentation considérable de la vitesse de transmission de l'information a bouleversé notre rapport au temps (et à l'espace). Manuel Castells, dans son ouvrage *La société en réseaux* (1998), voit dans l'émergence des TIC la naissance d'un nouvel âge de

l'information caractérisé non plus par des logiques de « lieux » mais de « flux », dont les réseaux techniques forment les éléments centraux (Castells, 1998, p. 525) :

« Les réseaux constituent la nouvelle morphologie sociale de nos sociétés, et la diffusion de la logique de la mise en réseau détermine largement les processus de production, d'expérience, de pouvoir et de culture ».

La vitesse et l'intensité de ces flux auraient pour conséquence de rendre le temps « intemporel », dès lors qu'il se caractérise par la simultanéité et l'instantanéité. Dans une perspective similaire, Johny Urry (2005) introduit la notion de « temps instantané » pour décrire la façon dont les TIC participeraient à faire émerger une nouvelle perception du temps basé sur la juxtaposition de moments si brefs qu'ils dépassent les capacités de la conscience humaine. La logique linéaire et mesurable du « temps-horloge »⁶ aurait disparu au profit d'interactions simultanées. Urry mobilise également la notion de « temps instantané » pour désigner une « métaphore traduisant l'importance de durées exceptionnellement courtes et fragmentées, même lorsqu'il ne s'agit pas littéralement d'instantanéité ou de simultanéité » (Urry, 2005, p.131). Il se réfère plus précisément à la temporalité produite par les médias, celle d'une « ubiquité instantanée » (Morley et Robins, 1995, p.131) qui juxtaposerait à un rythme effréné une multitude d'événements lointains, transportant l'individu d'une tragédie à l'autre selon des modalités qui lui échappent. Mais le « temps instantané » se manifesterait également par une temporalité quotidienne davantage fragmentée et des activités de moins en moins structurées collectivement. Urry perçoit ce changement temporel au travers de divers phénomènes, tels que : le caractère jetable de nombreux objets ; l'éphémérité des modes, des produits et des idées ; l'augmentation du nombre de contrats à durée déterminée et la flexibilisation des horaires ; des changements techniques et logistiques qui tendent à faire disparaître la distinction entre jour et nuit, jours ouvrables et week-end, foyer et lieu de travail, loisirs et activités professionnelles ; la transformation des modes d'alimentation marquée par la multiplication des

⁶ Urry (2005) identifie deux régimes temporels précédant le « temps instantané » : le « temps kairologique » et le « temps-horloge ». Le temps kairologique « consiste à puiser dans l'expérience du passé pour savoir quel moment précis de l'avenir convient pour tel ou tel événement » (p.118). Dès le XVIII^e siècle, le temps kairologique est progressivement remplacé par le « temps-horloge » et devient « désencastré de la vie sociale ». Le temps est dès lors quantifié et divisé en unités invariantes qui vont structurer de nombreuses pratiques sociales (travail, éducation, loisirs) et permettre une mesure synchronisée de la vie à l'échelle nationale, puis mondiale.

temporalités et des lieux dédiés au repas. Urry voit dans cette instantanéité une source de stress et d'anxiété, empêchant – pour les jeunes générations – la projection dans un avenir à long terme.

Selon Judy Wajcman, si les analyses proposées par Castells et Urry offrent un éclairage pertinent sur la façon dont de nombreuses sphères de la vie sociale sont marquées par une forme d'accélération, elles ne permettent toutefois pas de saisir comment cette accélération s'articule concrètement avec l'usage des TIC (Wajcman, 2015). Une attention plus fine portée à la technique permettrait de nuancer un récit global qui perçoit le temps et les logiques d'accélération de façon uniforme. De la même façon que les appareils ménagers ou l'automobile n'ont pas eu pour seul effet d'accélérer l'exécution des tâches domestiques ou la vitesse des déplacements, l'usage des technologies numériques n'implique pas uniquement une accélération des échanges et de la circulation de l'information.

L'exemple de l'e-mail est ainsi souvent évoqué pour décrire une technologie qui facilite la transmission d'un message, alors même qu'elle participe à augmenter sensiblement la charge de travail. Sa simplicité d'utilisation contribue à la prolifération de messages de toute sorte nécessitant un temps toujours plus long dédié au tri, puis au traitement de ce courrier. L'accroissement de la masse de courrier électronique s'accompagne également d'une réduction du délai de réponse attendu. Pourtant, rien dans la technologie *en soi* ne contraint le destinataire d'un e-mail à répondre rapidement. A la différence d'un appel téléphonique, la dimension asynchrone de l'e-mail laisse *a priori* la possibilité au destinataire de répondre lorsque cela lui convient. Le sentiment de devoir fournir une réponse dans un bref délai résulte moins des propriétés de la technique que de l'évolution des modes de travail et d'interactions sociales, à laquelle la technique participe. Dans une recherche menée auprès d'une grande société informatique aux États-Unis, Barley et al. (2011, p.903) montrent que l'e-mail sert avant tout de symbole permettant d'exprimer la surcharge de travail et la pression auxquelles doivent faire face les employés : « *As a symbol, e-mail became the interpretive scapegoat for the workers' perceptions that they were expected to do more than they could reasonably accomplish in a day* ». La nécessité pour les employés de répondre aux exigences qui leur sont imposées les contraint à traiter rapidement un nombre toujours plus important de tâches. Par ailleurs, contrairement aux réunions qui exigent une co-présence, les e-mails continuent à s'amonceler lorsque l'on est absent ou occupé à une autre activité, laissant ainsi la sensation de tâches jamais totalement achevées.

Les technologies mobiles renforcent encore cette pression, dès lors que les e-mails peuvent être consultés à toute heure et en tout lieu. Dans un monde où les individus sont confrontés à une compétition croissante et une augmentation des exigences de productivité, ils se voient contraints de suivre le rythme pour ne pas risquer de manquer une opportunité. L'injonction à la connexion et au renouvellement permanent n'a plus de limite de temps ni de lieu. Elle implique un mode de présence continu dont le sommeil semble constituer l'ultime obstacle (Crary, 2014). Cette course est d'autant plus ardue que, comme le souligne Hartmut Rosa (2013, p.43), « dans un monde en constant changement, il devient difficile de déterminer *quelles* options finiront par se révéler valables ». L'accélération du changement social entraîne ainsi un rythme de vie accéléré dont la réponse passe souvent par une solution technique destinée à augmenter l'efficacité de nos actions quotidiennes. Ces trois dimensions de l'accélération forment alors un système « fermé et autopropulsé ».

De nombreuses technologies continuent ainsi à être présentées par leurs concepteurs comme une réponse à la surcharge informationnelle qu'elles ont pourtant contribué à créer. Le développement des assistants vocaux semble également participer de ce cycle de l'accélération. Face à la raréfaction des ressources temporelles dont souffriraient de nombreux individus, l'assistant vocal apparaît comme une solution technique permettant de gérer de nombreuses tâches de façon simple et efficace. Avec Siri, « la journée va vous sembler plus légère » promet Apple.⁷ L'interaction vocale libère l'individu du contact avec la machine, lui permettant de réaliser plusieurs tâches simultanément. Par ailleurs, la personnalisation croissante des services proposés permet encore d'optimiser la gestion temporelle. Grâce au traitement algorithmique des nombreuses données dont il dispose, l'assistant vocal effectue en amont un tri parmi les options possibles pour ne proposer à l'utilisateur que les services qui semblent répondre à ses besoins. Le temps destiné à ce travail de sélection, désormais soustraité, peut être investi pour d'autres activités.

Cet idéal de vitesse, qui s'étend aujourd'hui à la gestion du foyer, correspond à une certaine vision du monde qui prône l'efficacité et la rentabilisation de chaque instant. L'idée selon laquelle l'optimisation de notre temps constitue un impératif auquel il s'agit d'apporter une réponse technique doit être comprise en regard du contexte spécifique de conception de ces technologies : la Silicon Valley. Les produits développés dans cette région sont marqués par une culture, des discours et des valeurs

⁷ <https://www.apple.com/fr/siri/>, consulté le 23 mars 2019.

qui lui sont propres. Le chapitre suivant s'attachera donc à comprendre en quoi les assistants vocaux ne sont pas uniquement l'objet d'une époque— celle d'une société contemporaine prise dans un cycle (paradoxal) d'accélération – mais également une solution technique envisagée par des acteurs portés par des convictions qui s'inscrivent dans un contexte sociologique spécifique.

2.2 Un idéal d'efficacité

Située au sud de la baie de San Francisco, la Silicon Valley est le berceau de nombreuses entreprises actives dans le domaine des nouvelles technologies. C'est le journaliste américain Don Hoefler qui, en 1971, baptisa « Silicon Valley » la région alors investie par de nombreuses entreprises dédiées au développement de puces électroniques à base de silicium (*silicon*). Le lieu devient progressivement un espace social, économique et politique où la technologie, envisagée comme moteur de changement, investit l'ensemble des sphères sociales. Dans une étude ethnographique consacrée à la Silicon Valley, Jan English-Lueck (2017) parle de « saturation technologique » pour décrire la dimension omniprésente (*pervasive*) de la technologie dans cette région. Plus que de simples outils, les dispositifs numériques participent d'une culture qui valorise une connexion permanente devenue symbole de statut social. Mais ce qui distingue la Silicon Valley d'autres lieux, c'est que la technologie y est à la fois intensivement « consommée » et conçue. Comme le souligne English-Lueck (2017, p.38) :

« In Silicon Valley, people have a unique relationship with these devices, applications, and services, for not only are they eager consumers, they are also the designers and producers. Their driving passion for technology makes them distinct from the people of other regions. Their technologies are their raison d'être and the economic prime mover, but the devices are also the chief production capital that enables the work itself ».

Si la technique est le moteur de la Silicon Valley, elle impose un rythme à ses concepteurs. Les cycles de production, toujours plus courts, nécessitent un haut degré de coordination et une communication intensive entre les acteurs de différents domaines (conception, ingénierie, développement, gestion de produit, design,

marketing, etc.). La vaste palette d'outils numériques à disposition leur permet d'échanger de diverses façons (e-mail, *voicemail*, appels, *instant messaging*, etc.) selon le contexte et le degré d'urgence estimé. Les sollicitations sont aussi nombreuses que diverses, il s'agit donc pour chacun de prioriser l'attention qui doit leur être portée. La gestion de ces différentes communications occupe une part importante de la journée et le temps « off » semble de plus en plus rare. Selon Judy Wajcman (2018), le travail de ces employés hautement qualifiés est caractérisé par un haut degré d'autonomie, des horaires flexibles, une connectivité permanente, des tâches effectuées hors du lieu de travail et la valorisation d'un emploi du temps surchargé mais efficacement géré.

Dans leur ouvrage *How Google Works*, Schmidt et Rosenberg (2014) – respectivement ancien CEO et Chef de produit de Google – proposent même un chapitre intitulé « *Overworked in a good way* » dans lequel ils affirment que les politiques visant à maintenir un équilibre entre vie professionnelle et vie privée peuvent être perçues comme insultantes envers des employés pleinement investis dans l'entreprise. Ils défendent une culture du travail basée sur un engagement intensif et une rationalisation de chaque instant. Tout ce qui pourrait constituer un obstacle à la productivité doit être écarté. Bien qu'anecdotiques, les propos de Mark Zuckerberg concernant ses choix vestimentaires illustrent cet état d'esprit. En 2014, lors d'une conférence publique, le fondateur de Facebook justifie ainsi sa décision d'adopter chaque jour une tenue identique :

*« I really want to clear my life to make it so that I have to make as few decisions as possible about anything except how to best serve this community. And there's actually a bunch of psychology theory that even making small decisions around what you wear, or what you eat for breakfast, or things like that, they kind of make you tired and consume your energy ».*⁸

Cette quête d'optimisation temporelle se traduit par une tentative d'automatisation de toute une série de tâches quotidiennes. L'idée sous-jacente est qu'une « vie bonne » est avant tout une vie efficacement gérée. Dans une récente étude menée auprès d'ingénieurs et designers actifs dans la Silicon Valley, Judy Wajcman (2018) note que

⁸ Friedman, V. (2018). Mark Zuckerberg's I'm Sorry Suit. *The New York Times*, 11 avril 2018. Disponible sur : <https://www.nytimes.com/2018/04/10/fashion/mark-zuckerberg-suit-congress.html>, consulté le 25 mars 2019.

la gestion du temps est une préoccupation centrale de ses enquêtés. La nécessité d'« optimiser son temps » (« *optimizing one's time* ») guide toute une série de décisions quotidiennes. L'affairement permanent est valorisé, tandis que les moments d'inaction sont considérés comme du temps « perdu ».

De nombreuses méthodes visant à proposer des solutions d'optimisation temporelle ont ainsi été développées. La plus connue d'entre elles, nommée « *Getting things done* » (GTD), propose un ensemble de stratégies pour cultiver un « art de la productivité sans stress » (Allen, 2001). Dans une perspective similaire, la méthode « *Inbox Zero* » créée en 2007 par Marlin Mann tente de rationaliser le traitement des e-mails. L'idée centrale est qu'il existe une façon optimale de gérer une boîte de réception électronique qui passe par différentes actions (archiver et supprimer des messages rapidement, classer les requêtes non prioritaires, etc.). Aujourd'hui, de multiples services numériques offrent des outils de gestion du temps. Dans son numéro de mars 2019, le magazine *Wired* proposait à ses lecteurs diverses solutions numériques pour gérer leur journée. « *Managing your workday shouldn't be a distraction unto itself. These apps make your phone handle the hard stuff* »⁹, promet l'article. Parmi les services conseillés, l'application mobile Google Assistant, jugée plus « productive » que le logiciel Siri : « *Siri is fine for queuing up song, but Google's Voice assistant gives your iPhone a bigger productivity boost.* »¹⁰ Les solutions de gestion du temps reposent toujours davantage sur des systèmes d'intelligence artificielle qui proposent une expérience personnalisée. Dans le même sens, en mai 2018, Google annonçait le lancement de *Smart Compose*, une fonctionnalité rattachée à Gmail qui suggère des phrases à l'utilisateur lorsqu'il rédige un e-mail.¹¹ Ce programme développé grâce au *machine learning* a pour objectif d'accélérer la cadence d'écriture en proposant plusieurs séquences de mots selon un contexte d'interaction, en partie déterminé grâce aux données dont dispose Gmail (contenu des e-mails, habitudes d'interactions avec le destinataire, etc.). Quelle que soit la méthode choisie, ces différentes approches perçoivent le temps comme une ressource individuelle qu'il est possible de maîtriser grâce à un bon usage de la technologie.

Une autre stratégie de gestion temporelle – qui concerne aussi bien les entreprises que les individus – est le recours à la sous-traitance. Le principe d'externalisation n'est

⁹ Horne, L. (2018). App pack desk mates. *Wired*, mars 2019, p. 32.

¹⁰ Ibid.

¹¹ Wu, Y. (2018). Smart Compose: Using Neural Networks to Help Write Emails. *Google AI Blog*, 16 mai 2018. Disponible sur : <https://ai.googleblog.com/2018/05/smart-compose-using-neural-networks-to.html>, consulté le 24 avril 2019.

pas propre à la Silicon Valley mais il est pratiqué de manière particulièrement importante dans cette région. Le fonctionnement de nombreuses entreprises actives dans le secteur de la « *gig economy* » repose précisément sur la sous-traitance. Les plateformes numériques proposent aujourd'hui toute une gamme de services (transport, livraison à domicile, ménage, blanchisserie, etc.) qui sont fournis par des travailleurs externes. Selon English-Lueck (2017), ce principe de sous-traitance est présent dans la Silicon Valley jusque dans les moindres activités quotidiennes. Le recours à des assistants personnels afin de se libérer des tâches peu valorisantes et chronophages serait une pratique courante parmi les classes les plus aisées. La sous-traitance est un moyen d'optimiser la gestion du temps en déléguant à d'autres ce que l'on ne souhaite pas faire soi-même.

L'idée que ces tâches puissent être davantage automatisées a contribué au développement de solutions techniques destinées à externaliser des fonctions non seulement physiques, mais également cognitives. Si le principe de délégation de certaines tâches à une machine ou à un logiciel n'est pas nouvelle, l'arrivée du *smartphone* marque cependant le début d'une forme d'assistance personnalisée basée sur l'enregistrement et la mise en réseau des traces numériques. Le développement des assistants vocaux, et plus récemment des enceintes connectées, prolonge cette logique de sous-traitance et de quête d'efficacité. Les fabricants font d'ailleurs largement appel à cet imaginaire lorsqu'ils présentent ces technologies. On peut ainsi lire sur les différents sites qui leur sont dédiés que Google Home est « à votre service »¹², Siri « en fait plus avant que vous le demandiez »¹³ et Cortana « s'occupe de tout, de manière proactive »¹⁴. En 2018, Google lançait un logiciel uniquement destiné à la prise de rendez-vous téléphonique, qui permet à l'utilisateur de confier cette tâche à un assistant vocal. Si le système est présenté par Sundar Pichai (PDG de Google) comme un nouvel outil destiné à augmenter la productivité des utilisateurs, il semble qu'il s'agisse avant tout d'un gain de confort (pour certains), la délégation de la prise de rendez-vous à l'assistant vocal nécessitant un investissement temporel plus important que d'effectuer l'appel soi-même. Le développement de ces services de sous-traitance algorithmique est présenté comme une manière de libérer chacun de tâches jugées subalternes et contraignantes.

¹² https://store.google.com/fr/product/google_home?hl=fr-FR, consulté le 26 mars 2019.

¹³ <https://www.apple.com/fr/siri/>, consulté le 26 mars 2019.

¹⁴ <https://www.microsoft.com/fr-fr/windows/cortana>, consulté le 26 mars 2019.

2.3 Économie et technologies des données

Bien que les assistants vocaux semblent nés d'une volonté d'automatiser toujours davantage de tâches pour gagner en efficacité, leur conception ne peut être détachée des contextes économique et technique dans lesquels elle s'insère. Le développement de ces objets répond à des logiques propres à une économie basée sur la collecte des traces numériques, rendue possible grâce aux avancées techniques qui l'ont accompagnée. S'il semble impossible de s'engager ici dans une description de l'ensemble des facteurs qui ont conduit à l'émergence des assistants vocaux, une transformation survenue au cours des années 2000, qui concerne à la fois la technique, les modèles économiques et les pratiques, constitue certainement une étape importante dans le développement des interfaces vocales. Cette mutation est celle du « Web 2.0 ». L'expression, formulée par Tim O'Reilly en 2004, exprime les changements profonds qui concernent à la fois le développement de nouveaux dispositifs techniques liés à l'évolution d'Internet et l'émergence de nouvelles pratiques. Jusqu'alors, le Web reposait essentiellement sur un principe de documentation. L'utilisateur avait accès à un vaste réseau d'informations, mais sa participation dans la production restait marginale et réservée à des initiés. Si les sites web ou les e-mails permettaient déjà d'échanger en ligne, le dispositif en lui-même n'offrait pas une connexion directe entre les individus. Dès le milieu des années 2000, l'apparition de services de mise en relation entre utilisateurs ou « plateformes », transforme progressivement le Web en un lieu d'échange, un espace participatif où les usagers ne sont plus de simples récepteurs passifs. La facilité d'utilisation et le coût dérisoire de ces services permet à un public toujours plus large de s'exprimer et d'échanger. Ces nouveaux réseaux sociaux numériques offrent des moyens d'interaction multiples, qu'il s'agisse d'un court message, d'un commentaire, de l'envoi d'une photo, voire d'un simple clic sur un bouton *like*.

Mais avec les plateformes, se développe également un modèle économique basé sur la captation et la revente des nombreuses traces laissées par les utilisateurs. En misant sur des services gratuits, efficaces et attractifs, un nombre restreint d'acteurs est parvenu, en seulement quelques années, à capter l'essentiel de ces échanges au travers de plateformes désormais difficilement contournables. Comme le formule Dominique Boullier (2016, p.83) :

« Ce moment que l'on présente comme l'âge des foules et de leurs contributions ouvertes est en fait devenu une carrière de données à ciel ouvert, données que l'on peut fouiller gratuitement (*data mining*) et qui génèrent d'autant plus de revenus que ces plateformes sont devenues des « points de passages » obligés (pour les utilisateurs et pour les marques) [...] ».

Cette masse de données collectées, qui constitue ce que l'on appelle communément le *big data*, concerne non seulement le contenu des réseaux sociaux, mais également tous les services capables d'enregistrer l'activité des utilisateurs. Il peut s'agir de traces visibles (*tweets*, publications, commentaires, *like*) ou invisibles (géolocalisation, données de navigation, requêtes sur des moteurs de recherche). Celles-ci ont d'autant plus de puissance que, déployées à large échelle, elles peuvent être associées et modélisées afin de constituer un profil individuel qui sera, soit directement exploité par la plateforme, soit revendu à des annonceurs ou partenaires. Google et Facebook ont su exploiter ce modèle jusqu'à devenir les principaux acteurs de ce marché. Par une capacité de traçage qui se déploie aujourd'hui au-delà de leurs propres espaces, ils étendent encore leur *hypercentralité* et disposent dès lors d'une perception inédite des pratiques individuelles (Beaude, 2012). Selon Shoshana Zuboff (2015), ce modèle représente un véritable changement de paradigme économique : l'ensemble des échanges informationnels répondraient désormais aux logiques d'un « capitalisme de surveillance » dont la richesse ne proviendrait plus d'une accumulation de biens ou d'argent, mais d'une maîtrise des données comportementales des individus.

Au cœur de ce modèle économique de profilage se trouvent des technologies d'apprentissage automatique (*machine learning*), dont le développement a été accéléré par l'accès à cette nouvelle abondance de données. Leur fonction consiste à fournir à une machine un nombre conséquent de données afin qu'elle puisse observer des régularités et corrélations qui vont permettre d'élaborer des modèles d'action. Il s'agit d'un processus composé d'étapes successives d'essais/erreurs, le plus souvent supervisé par des humains. Une fois le modèle établi, de nouvelles données sont fournies à la machine afin de réaliser la tâche attendue (traduction, identification d'un élément sur une photographie, classification, etc.). Contrairement aux approches statistiques classiques, le *machine learning* ne nécessite pas l'élaboration d'hypothèses en amont, celles-ci pouvant émerger des données elles-mêmes. Ce processus finit par rendre vaine toute tentative de théorisation, car « lorsque l'enjeu consiste à agir [...],

l'important est le résultat est non son explication » (Boullier, 2015, p.820). Le résultat apparaît d'autant plus solide que les données fournies à ces machines sont le plus souvent issues d'un traçage automatique des pratiques, et semblent *a priori* échapper à toute forme de subjectivité (Rouvroy et Berns, 2013). L'abondance de données disponibles participe également à l'impression de pouvoir obtenir une visualisation fidèle de la diversité des réalités sociales. Les modèles ainsi construits ont alors pour objectif d'inférer une intention, une préférence ou un comportement. Le champ d'application (potentiel) du *machine learning* est vaste et touche l'ensemble des sphères sociales. Il peut être mobilisé dans des processus de décision, classement, notation, recommandation ou catégorisation dans des domaines tels que l'analyse médicale, les assurances, la sécurité informatique, la justice, les transports, la traduction, le marketing ou encore les ressources humaines.

Le développement des assistants vocaux est intimement lié au *machine learning*, car ses caractéristiques centrales que sont l'interface vocale et la personnalisation des services font largement appel à des méthodes d'apprentissage automatique. L'interaction vocale entre l'humain et la machine est rendue possible par la mobilisation de systèmes de traitement automatique du langage naturel (TALN) qui ont pour objectif d'analyser, interpréter et reproduire la parole humaine. Le domaine du TALN, dont les premiers développements datent du milieu du XX^e siècle, a réalisé des progrès substantiels au cours de ces dernières années grâce à l'introduction de processus d'apprentissage profond (*deep learning*) basés sur des réseaux de neurones artificiels. Les algorithmes dédiés au traitement automatique du langage sont entraînés à l'aide de millions d'échantillons de texte ou parole préalablement répertoriés et « étiquetés » par des humains. En analysant une quantité conséquente d'exemples, la machine doit ensuite être capable de reconnaître de façon autonome le contexte d'interaction. La compréhension du contexte, qui permet d'offrir une réponse appropriée à la requête émise par l'utilisateur, est aujourd'hui l'un des enjeux-clés des assistants vocaux.¹⁵ Au-delà de l'interaction vocale, les progrès du *machine learning* participent également à améliorer l'ensemble du dispositif. Dans un récent article publié sur le blog d'Amazon, Rohit Prasad, vice-président et *head scientist* de l'entreprise, indique que les recherches menées actuellement autour du logiciel Alexa

¹⁵ Sarikaya, R. (2018). The Role of Context in Redefining Human-Computer Interaction. *Amazon Alexa Blog*, 7 décembre 2018. Disponible sur : <https://developer.amazon.com/fr/blogs/alexa/post/3ac41587-f262-4fec-be60-2df2f64b9af9/the-role-of-context-in-redefining-human-computer-interaction>, consulté le 31 mars 2019.

relèvent de cinq catégories : l'apprentissage de nouvelles compétences (*skills*) et la consolidation des compétences existantes ; la compréhension du contexte (à la fois global et propre à l'utilisateur) ; l'amélioration de connaissances des faits et événements ; l'optimisation de l'interaction vocale ; l'automatisation du processus d'apprentissage.¹⁶ Comme l'explique Rohit Prasad, l'ensemble de ces recherches font appel à des techniques de *machine learning* qui combinent puissance informatique et vastes bases de données.

Toutefois, l'apparente automatisation ne doit pas occulter l'importante part de travail humain encore nécessaire au fonctionnement de ces systèmes. Les requêtes formulées par les utilisateurs d'assistants vocaux font l'objet d'un traitement manuel qui permet d'améliorer la pertinence des réponses fournies. En mai 2018, l'association La Quadrature du Net faisait part du témoignage de Julie, une jeune femme dont le travail consistait à retranscrire et vérifier de nombreuses conversations enregistrées par le logiciel Cortana (Microsoft).¹⁷ Il s'agissait principalement de courts extraits vocaux (entre 3 et 15 secondes) dont une première transcription était proposée par le système. Julie avait ensuite pour tâche de corriger le texte produit par la machine si celui-ci contenait des fautes de compréhension, de grammaire ou d'orthographe. Comme le rappelle Antonio Casilli dans son ouvrage *En Attendant les Robots* (2018), l'intelligence artificielle repose largement sur une forme de *digital labor* qui consiste à faire appel au travail humain pour réaliser des opérations que les machines ne peuvent effectuer d'elles-mêmes. Il s'agit pour l'essentiel de micro-tâches, comme l'annotation d'images, l'étiquetage de produits, la transcription de textes numérisés, le remplissage de questionnaires. Ces activités sont accessibles au travers de plateformes telles que *Amazon Mechanical Turk*, dont le but est de mettre en relation des clients avec des personnes disposées à réaliser des tâches très faiblement rémunérées – de l'ordre de quelques centimes. Des travailleurs peu qualifiés ou marginalisés réalisent ces tâches le plus souvent à domicile. Ce marché est particulièrement important dans certains pays d'Asie du Sud-Est, tels que la Thaïlande, le Bangladesh et le Vietnam où des « fermes à clics » ont vu le jour. Comme le souligne Dominique Cardon (2019, p. 344), « une sorte de colonialisme numérique s'est installé à l'échelle de la planète : le

¹⁶ Amazon. How Alexa keeps getting smarter. *The Amazon blog – Day One*, 10 octobre 2018. Disponible sur : <https://blog.aboutamazon.com/devices/how-alexa-keeps-getting-smarter>, consulté le 31 mars 2019.

¹⁷ Derrière les assistants vocaux, des humains vous entendent. *La Quadrature du Net*, 18 mai 2018. Disponible sur : https://www.laquadrature.net/2018/05/18/temoin_cortana, consulté le 31 mars 2019.

confort fonctionnel des services du web pour les usagers occidentaux est en partie soutenu par le travail souterrain délocalisé dans les pays du tiers-monde ».

Les grands acteurs du numérique font cependant appel à des travailleurs plus qualifiés, chargés d'évaluer la pertinence de résultats de recherche, de vérifier la transcription de commandes vocales ou le fonctionnement d'une application. Ces opérations peu valorisées ne sont pas effectuées en interne, mais confiées à des entreprises sous-traitantes principalement installées aux États-Unis. En 2017, le magazine *Ars Technica* enquêtait sur LeapForce, l'un des principaux prestataires de Google.¹⁸ Les employés de cette société ont pour mission d'exécuter une douzaine de tâches précises qui permettent d'optimiser le fonctionnement de différents services proposés par Google. Ils travaillent principalement à domicile en se connectant à un système nommé Raterhub et sélectionnent des opérations parmi les choix proposés. Il s'agit d'un emploi payé à la tâche et faiblement rémunéré, même si les candidats doivent préalablement réussir un exigeant test d'entrée qui nécessite la lecture attentive d'un manuel de 160 pages.¹⁹ Bien que central au fonctionnement de Google, ce travail reste pourtant largement occulté par le géant du numérique. La sociologue Sarah Roberts, qui a réalisé une vaste enquête ethnographique auprès des employés de ces sociétés, affirme que Google s'efforcerait de dissimuler le travail réalisé par ses sous-traitants afin de conserver l'image d'une intelligence artificielle fluide et automatisée, dont le fonctionnement serait détaché de toute activité humaine.²⁰

2.4 Un contexte à nuancer

Cet aperçu a permis de mieux saisir quelques éléments du contexte dans lequel s'inscrit le développement des assistants vocaux. La raréfaction des ressources temporelles, une culture de l'efficacité et de l'optimisation de chaque instant, un modèle économique basé sur la quantification et la marchandisation de l'ensemble des sphères sociales et les progrès réalisés par les techniques de *machine learning* sont autant de facteurs qui ont pu – à différents degrés – contribuer à l'émergence de ces

¹⁸ Newitz, A. (2017). The secret lives of Google raters. *Ars Technica*, 27 avril 2017. Disponible sur : <https://arstechnica.com/features/2017/04/the-secret-lives-of-google-raters>, consulté le 31 mars 2019.

¹⁹ Google Search Quality Rating Guidelines. Disponible sur : <https://static.googleusercontent.com/media/www.google.com/en//insidesearch/howsearchworks/assets/searchqualityevaluatorguidelines.pdf>, consulté le 31 mars 2019.

²⁰ Newitz, A., Ibid.

nouvelles interfaces vocales. Toutefois, certains de ces éléments, en apparence unidirectionnels, doivent être nuancés.

En premier lieu, il est nécessaire de souligner que la tendance à l'accélération ne concerne pas l'ensemble de la société de façon uniforme. On pourrait en effet reprocher aux théories de l'accélération leur perception homogène du temps, occultant la pluralité des temporalités. Comme le souligne Claude Dubar (2011, p.2) dans sa critique de l'ouvrage d'Hartmut Rosa (2010) :

« [...] le temps est multiple, il existe des temporalités distinctes et hétérogènes [...]. Tout ramener à un seul temps chronométrique, c'est s'interdire de repérer des contrecultures, des mouvements sociaux, des problématiques alternatives à l'hégémonie des structures technico-économiques du capitalisme mondialisé ».

Sans remettre en question la tendance générale à l'accélération, Michel Lussault (2014) porte également un regard critique sur la thèse d'Hartmut Rosa, dont il dénonce la vision monolithique du temps. Cette approche repose, selon lui, sur « des faits certes avérés pour extrapoler un métadiscours efficace mais relativement déconnecté de l'examen précis de ce que les êtres humains font et de la manière dont ils le font [...] » (Lussault, 2014, p. 66). Par ailleurs, si l'accélération semble être uniformément subie, elle constitue pourtant un privilège dans de nombreuses situations. Harvey Moloch (2016, p.120) souligne ainsi que la vitesse est synonyme de pouvoir dès lors qu'elle permet aux plus aisés de bénéficier d'un confort accru. A l'aéroport, comme en de nombreux lieux, les files d'attente sont réservées aux individus dépourvus d'un statut *premium*. En citant Barry Schwarz (1974, p.869), Moloch rappelle que la maîtrise du temps demeure une composante essentielle du pouvoir : « *Far from being a coincidental byproduct of power, then, control of time comes into view as one of its essential properties* ».

Ensuite, la focalisation sur la culture de l'efficacité ne doit pas faire oublier que l'histoire de la Silicon Valley est également traversée par une série de contradictions. Le développement d'Internet a toujours été associé à des valeurs d'émancipation et de liberté. A l'autorité étatique centralisée, les pionniers de la Silicon Valley préféraient l'idée d'un pouvoir partagé au travers d'un réseau d'individus. Pourtant, cette méfiance à l'égard des institutions ne s'accompagne pas d'une hostilité au marché (Turner, 2006). Les premiers partisans de la culture libertaire deviennent, au cours des années

1980 et 1990, les créateurs de nombreuses entreprises qui détiennent aujourd'hui un pouvoir considérable. Dans un récent article, le sociologue Olivier Alexandre (2018) démontre ce paradoxe au travers d'une enquête sur le festival Burning Man, autrefois haut lieu de la contre-culture et de l'esprit communautaire, désormais investi par les « *techies* » californiens. L'auteur souligne ainsi que de nombreuses critiques s'élèvent contre ce qui est perçu comme « l'hypocrisie d'un rassemblement de jeunes parvenus qui célèbrent l'espace d'une semaine des modes de vie alternatifs, l'échange non marchand et le retour à la terre tout en se livrant le reste de l'année aux formes les plus avancées d'exploitation dans le monde des start-up et du capitalisme de plates-formes » (p.90). Les solutions technologiques émanant de la Silicon Valley sont nourries par des idéaux libertaires, tout en s'inscrivant dans une politique économique néo-libérale, dont les conséquences sur le tissu social sont aujourd'hui indéniables.²¹

Par ailleurs, l'hyperconnectivité et l'affairement permanent valorisés par les entreprises technologiques sont aujourd'hui questionnés par de nouvelles injonctions à la déconnexion. Si la thématique n'est pas totalement nouvelle, elle occupe désormais un espace important dans le débat public. En France, le principe d'un droit à la déconnexion a donné lieu à un décret de loi qui impose aux entreprises d'instaurer des règles d'usage des outils numériques afin de garantir le temps de repos des employés. La déconnexion prônée actuellement ne relève pas d'un radical abandon des technologies, mais plutôt d'une mise à distance ponctuelle et partielle, d'une connexion maîtrisée (Jauréguiberry, 2014).

Cette forme de prise de contrôle est désormais également louée par des acteurs du numérique qui ont eux-mêmes participé à créer ce que certains dénoncent comme une forme « d'addiction » aux technologies (Turkle, 2015). En effet, au cours de ces dernières années, plusieurs anciens ingénieurs de la Silicon Valley se sont emparés de ce débat en soulignant les conséquences de dispositifs conçus pour garder l'utilisateur « captif » d'un environnement numérique. Selon le *Center for Humane Technology*, l'organisation créée à la suite de l'initiative *Time Well Spent* lancée par Tristan Harris, la solution réside dans un design davantage « tourné vers l'humain ».²² Si le concept reste flou, l'idée centrale est celle d'une conception moins insidieuse des services numériques. Dans la continuité de ce mouvement, plusieurs entreprises du numérique

²¹ Simmonds, C. (2017). The Silicon Valley paradox: one in four people are at risk of hunger. *The Guardian*, 11 décembre 2017. Disponible sur : <https://www.theguardian.com/us-news/2017/dec/12/the-silicon-valley-paradox-one-in-four-people-are-at-risk-of-hunger>, consulté le 2 avril 2019.

²² <https://humanetech.com/>, consulté le 24 avril 2019.

proposent désormais de nouveaux outils destinés à aider les utilisateurs à contrôler le temps passé devant les écrans. Apple propose une fonctionnalité qui présente des statistiques d'utilisation des applications et permet de définir des limites de temps quotidiennes. Google, Facebook et Instagram ont également développé des outils similaires.

Cette approche de la déconnexion, *a priori* critique à l'égard des modèles construits par les plateformes numériques, n'est pourtant pas totalement en contradiction avec les valeurs d'efficacité et de productivité prônées par ces entreprises. Tout d'abord, ces initiatives restent focalisées sur la problématique du temps et ne disent rien du contenu auquel l'utilisateur est exposé. Elles mènent à penser que la résolution de la question attentionnelle se situe simplement dans la détermination d'un temps d'utilisation « approprié ». Ensuite, le temps apparaît toujours comme une ressource individuelle qu'il serait possible de mieux « gérer » grâce à des outils technologiques. La responsabilité reviendrait donc à l'utilisateur qui dispose désormais de solutions pour contrôler ses activités en ligne. Toutefois, cette maîtrise n'est pas également répartie parmi la population. A mesure que les technologies se démocratisent et que leur coût se réduit considérablement, c'est le temps hors connexion qui devient un bien rare et convoité, que seules peuvent s'offrir les classes aisées. La capacité à s'extraire des sollicitations exercées par les technologies devient alors un nouveau signe de distinction. Comme le souligne la journaliste Nellie Bowles : « *As more screens appear in the lives of the poor, screens are disappearing from the lives of the rich. The richer you are, the more you spend to be offscreen* ». ²³

Dans cette quête d'une distanciation des écrans, les assistants vocaux présentent, pour les entreprises du numérique, l'avantage de contourner la problématique de la captation d'attention – liée avant tout à la dimension visuelle du contenu proposé par les plateformes – tout en gardant un lien étroit avec l'utilisateur. La course à l'attention ne disparaît pas pour autant, comme nous le verrons au chapitre 4, mais elle se réalise de façon plus discrète, sans solliciter directement l'individu. Si les assistants vocaux s'inscrivent dans un contexte d'optimisation temporelle en permettant le *multitasking* et la délégation de certaines tâches, ils apparaissent également comme une solution aux problématiques posées par une économie qui repose sur la maximisation du temps

²³ Bowles, N. (2019). Human Contact Is Now a Luxury Good. *The New York Times*, 23 mars 2019. Disponible sur : <https://www.nytimes.com/2019/03/23/sunday-review/human-contact-luxury-screens.html>, consulté le 24 avril 2019.

d'attention. En ce sens, ces technologies semblent répondre aux différentes injonctions (et contradictions) qui traversent le contexte socio-économique dans lequel elles ont été pensées.

3. La voix pour interface

Caractéristique centrale des enceintes connectées, l'interface vocale interroge notre rapport à la technique. Si le principe d'un objet parlant existait déjà auparavant (radio, GPS, automates), celui-ci se contentait de fournir une information, sans attendre de réponse en retour. L'assistant vocal introduit une nouvelle relation à la machine basée sur un échange. Il ne se limite plus à transmettre une indication ou guider l'utilisateur d'un service dans une succession d'étapes, mais propose de « converser » avec lui. Il s'agit par ailleurs d'un dialogue ajusté en temps réel et personnalisé. Les concepteurs promettent ainsi une interaction simple et fluide grâce à un mode de communication au plus proche du langage naturel. Ce chapitre propose d'explorer les enjeux liés à cette nouvelle forme d'interaction homme-machine en retraçant tout d'abord – de façon succincte – l'histoire de la voix de synthèse, puis en analysant les transformations qui accompagnent le développement de ces nouvelles interfaces. Ces changements se situent à la fois au niveau individuel et collectif, car le fonctionnement des interfaces vocales repose sur un processus de feedback entre les utilisateurs et le vaste réseau socio-technique auquel ils sont rattachés.

3.1 La voix à distance

L'invention du téléphone en 1876 est peut-être la première technologie à désincarner la voix – en la transportant sans le corps (Karpf, 2008). L'émetteur de la voix n'est plus directement observable. Comme le souligne Clotilde Chevet (2017, p.64), le téléphone « a rompu un axiome de communication fondamental, celui selon lequel l'œil peut repérer l'origine d'un bruit si l'oreille peut l'entendre ». Avant son arrivée, entendre des voix en l'absence de locuteur était un signe de mysticisme ou de folie. Au cours de ses premières années, le téléphone continue ainsi à véhiculer cet imaginaire mystique, l'idée qu'une voix puisse voyager au travers d'un fil relevant pour beaucoup d'une forme d'irrationalité. Les craintes et superstitions autour de ce medium sont donc nombreuses, mais rapidement apaisées par les voix des opératrices.

Celles-ci avaient pour tâche de mettre en relation un émetteur et son correspondant. En France, les « demoiselles du téléphone », comme les nommait Marcel Proust, se devaient de rassurer leurs interlocuteurs. Ainsi que le note Anne Karpf (2008, p.394) :

« le téléphone inquiétait parce qu'il donnait à des voix étrangères un accès direct à votre oreille ; utiliser des femmes comme intermédiaires entre les mondes public et privé paraissait une sorte d'antidote. Appelées pour domestiquer le téléphone, elles ont aidé à le faire passer dans l'esprit du grand public du statut d'intrus de la technologie à celui de moyen de maintenir le contact social par la parole [...] ».

On constatera qu'aujourd'hui également, les voix de synthèse – dont celles des assistants vocaux – sont essentiellement féminines.

A la suite du téléphone, d'autres « machines parlantes » voient le jour : le phonographe, le gramophone, puis la radio et la télévision. Mais c'est avec le développement des technologies numériques et de la synthèse vocale que la voix devient un véritable objet de communication, indépendant des contraintes « humaines » (Chevet, 2017). Ce système est particulièrement présent en situation de mobilité. Les gares, les aéroports et les métros sont des espaces où les voix guident les usagers en mouvement. Le plus souvent, une voix féminine est choisie, celle-ci étant perçue comme « plus agréable » (Mitchell et al., 2011). Au-delà de l'aspect pratique, la voix permet aussi de donner une identité et un « supplément d'âme » à des lieux de passage souvent froids, impersonnels et déjà sémiotiquement surchargés (Chevet, 2017). Les entreprises de transports se sont donc rapidement lancées dans le façonnage sonore de l'espace urbain. Dans les gares SNCF, les annonces au micro ont été remplacées dès 1980 par une voix enregistrée mais segmentée, qui permet de créer un système d'annonces séquentiel (Carron, 2016). Celle-ci est alors utilisée pour la plupart des annonces en gare et identifiée comme un élément-clé de l'identité du groupe. Dès 1994, les annonces vocales sont accompagnées d'un signal sonore qui permet d'attirer l'attention des voyageurs. Quelques années plus tard, celui-ci se transforme en « *jingle* » qui n'a plus pour unique objectif d'alerter les usagers, mais également de créer une identité de marque immédiatement reconnaissable. Ces quelques notes font partie d'un environnement acoustique développé au cours des années 1990 qui comprend le design de toute une série de sons. De la même manière, la RATP a développé divers outils sonores qui permettent à la fois aux voyageurs de

s'orienter et d'identifier la société. Les différents signaux sonores du passe Navigo, la carte des transports en commun parisiens, sont le résultat d'une collaboration entre un compositeur, une psychosociologue, ainsi qu'un technicien.²⁴ Chacun de ces sons doit non seulement alerter, mais également transmettre un message, et donc du sens. La synthèse vocale apparue par la suite permet de compléter les sonorités existantes et de répondre à certains défis auxquels se heurtait la voix humaine. Grâce à un système de « *text to speech* », permettant de générer une parole artificielle à partir d'un texte, il est possible de composer de nombreux messages dans un très court délai et de construire une identité vocale « sur-mesure ».

Si, pendant longtemps, les voix synthétiques étaient associées à l'espace public, l'arrivée du *smartphone* marque le début des interactions personnalisées et le passage de la voix dans le domaine privé. Dès 2011, les iPhones sont équipés du logiciel Siri, offrant à leurs utilisateurs la possibilité d'effectuer des commandes oralement, sans contact physique. Bien que l'existence des « objets parlants » ne soit pas nouvelle, le *smartphone* introduit une forme de dialogue encore impensable jusque-là. Cette nouvelle modalité d'interaction homme-machine est rendue possible par des systèmes de reconnaissance vocale, développés en parallèle des logiciels de *text to speech*. Sans interface graphique, les assistants vocaux promettent une interaction au plus proche de l'expression humaine.

3.2 De l'interaction à la relation

Les assistants vocaux sont présentés à l'utilisateur comme une nouvelle manière d'interagir avec la technique. Pourtant, il est nécessaire de rappeler que les objets ont toujours communiqué d'une certaine façon. Comme le souligne Dominique Boullier (2002, p.49), « ces objets de fait nous parlent depuis longtemps et nous les avons équipés en partie pour cela. Les machines portent avec elles des éléments de textes, sous forme de consignes, de signaux, d'affichages divers, à destination humaine ». Toutefois, ces éléments qui offraient un guidage à l'utilisateur semblent progressivement disparaître. Ils seraient rendus inutiles, dès lors que l'interaction ne relèverait plus d'une compétence technique, mais sociale. Les concepteurs demandent en effet aux utilisateurs d'interagir avec la machine comme ils le feraient avec un

²⁴ Delesalle, N. (2007). C'est quoi ce bruit ? Télérama, 4 mai 2007. Disponible sur : https://www.telerama.fr/divers/19019-est_quoi_ce_bruit.php, consulté le 3 avril 2019.

interlocuteur humain. Apple affirme ainsi que « Siri reste attentif à vos moindres besoins, comme un véritable assistant personnel ».²⁵ Les assistants vocaux introduisent un nouveau rapport à l'objet. Il s'agit de créer non plus seulement une interaction, mais une *relation* avec la machine. Ce principe concerne aujourd'hui tout un ensemble d'êtres artificiels (robots, agents conversationnels, etc.) auxquels on tente d'imputer des attributs humains. Selon l'anthropologue Denis Vidal (2012, p.72), il existe « un ensemble de tentatives convergentes pour donner à des machines un statut identifiable à celui d'un agent "autonome" quand ce n'est pas, plus radicalement, celui d'une "personne" ». Nous ne sommes plus dans une situation où une voix de synthèse lointaine guiderait l'utilisateur dans un espace urbain (gare, métro, aéroport) ou dans une procédure relativement simple et circonscrite (achat *via* un automate, assistance vocale téléphonique). La voix implique aujourd'hui un rapport personnalisé à la machine.

Cette relation n'est pas prédéterminée, mais se construit au fil de l'usage. Le mode d'emploi se révèle inutile, dès lors que les concepteurs promettent une « conversation » avec le dispositif. Celle-ci suppose alors le partage d'une même temporalité et la mobilisation de codes communicationnels propres à l'échange humain, tels que la formulation en « je ». Selon Clifford Nass (2005), les interfaces vocales mobilisant la première personne du singulier peuvent être beaucoup plus facilement perçues comme des êtres dotés d'une agentivité. De la même manière, le recours à des formules de politesse, telles que « je suis désolé » ou « je vous en prie » (Siri) font davantage référence à la conversation humaine qu'à un échange avec une machine. Toutefois, le simple usage de la parole constitue déjà une manière de faire entrer l'utilisateur dans un cadre interactionnel. Même si celui-ci est conscient de la dimension artificielle de la conversation, l'illusion d'un échange entre deux acteurs est nécessaire à l'usage du dispositif.

Dans certains travaux en anthropologie, le mode du « comme si » est considéré comme une modalité structurante des interactions entre humains et êtres artificiels (Borelle, 2018). Si celui-ci comporte tout d'abord une dimension ludique (Grimaud, 2012), il relève également de la nécessité pragmatique. L'« extension aux non-humains des modalités interactionnelles propres au dialogue entre humains » (Airenti, 2012, p.53) est une condition nécessaire pour initier l'interaction. Cependant, il ne s'agit pas de percevoir l'utilisateur comme la victime d'un « piège anthropomorphique » (Vidal, 2012) tendu par les concepteurs et destiné à provoquer une confusion entre humains

²⁵ <https://www.apple.com/fr/siri/>, consulté le 7 avril

et non-humains. Le paradigme CASA (*Computers Art Social Actors*), développé par une équipe de chercheurs constituée autour de Clifford Nass, a montré que les individus adoptent spontanément des comportements sociaux avec les technologies (attribution de catégorie de genre, politesse, réciprocité) tout en se défendant de toute perception anthropomorphique (Nass et Moon, 2000). Dans le même sens, Dominique Boullier (2002, p. 49) note que « nous sommes en fait déjà équipés des catégories et des formats anthropologiques pour accepter que des objets soient vraiment mus par des intentions programmées ». Cependant, « tout le modernisme a consisté à tracer une frontière hermétique entre cette rationalité ordinaire [...] et la rationalité scientifique supposée gouverner désormais toutes nos pratiques les plus quotidiennes (Latour, 1992) ».

Selon cette perspective, l'effort des concepteurs ne devrait donc pas porter sur la création d'une illusion parfaite, dont les usagers ne sont pas dupes. En outre, plus une machine revêt des caractéristiques humaines, plus élevées semblent être les attentes des utilisateurs de la voir également réagir de façon humaine (Vidal, 2012). La déception sera d'autant plus grande si elle ne peut tenir ses promesses. Denis Vidal (2012, p.70) parle alors d'un « pacte anthropomorphique » pour désigner « toutes les instances d'interaction entre des robots humanoïdes et leurs utilisateurs fondées sur l'apparence anthropomorphique des premiers, sans pour autant postuler que ceux qui interagissent avec eux ne devraient succomber à aucune forme d'illusion cognitive caractérisée ». Cette notion s'inscrit dans la continuité de travaux initiés dès la fin des années 1980 dans le champ de la sociologie des sciences et des techniques et qui s'opposaient au concept de « projection anthropomorphique » (Latour, 1988). Le « pacte anthropomorphique » relève donc d'un « instrument relationnel d'évaluation et d'ajustement aux objets qui nous entourent » (Grimaud et Vidal, 2012, p.19), plutôt que d'un procédé inconscient qui attribuerait à des entités non-humaines des facultés propres aux humains. L'assistant vocal propose précisément à l'utilisateur une forme d'accord qui leur permet de « converser », bien que ce dernier soit conscient de la nature artificielle et limitée de cet échange. Selon Nicolas Santolaria (2016, p.111), cette illusion est « nécessaire à la relation symbiotique qu'impose le dispositif ». C'est également ce rapport volontairement ambigu que décrit la journaliste Judith Shilevitz lorsqu'elle évoque son rapport à l'assistant vocal Alexa, dont la voix l'aide à trouver le

sommeil : « *Even though my adult self knew perfectly well that Sweet Dreams²⁶ was a glitch, a part of me wanted to believe that Alexa meant it* ». ²⁷

De façon générale, Denis Vidal (2012, p.71) estime que nous observons aujourd'hui deux conceptions de l'anthropomorphisme chez les concepteurs d'agents artificiels. Certains voient dans cette notion « l'indice d'une forme de carence cognitive de notre part, susceptible d'être exploitée par les roboticiens pour avoir prise sur nos agissements [...] ; d'autres pensent que l'anthropomorphisme peut aussi être mobilisé pour pallier les carences cognitives des robots comme celles des utilisateurs plutôt que chercher à les exploiter ». S'il semble délicat de se positionner de façon certaine quant à la manière dont l'anthropomorphisme est appréhendé par les principaux concepteurs d'assistants vocaux, l'apparition de logiciels tels que Google Duplex laisse penser qu'il s'agit principalement de tirer profit de l'illusion de réel créée par la machine.

3.3 La promesse d'une interaction « sans friction »

Le développement des interfaces vocales participe plus globalement d'une quête d'optimisation de l'expérience utilisateur, qui sous-tend la conception de tout service ou produit numérique. La simplicité d'utilisation constitue l'un des points centraux de cette expérience. Qu'il s'agisse de commander un repas, d'envoyer un message ou de louer un appartement pour un voyage, tout est mis en œuvre pour proposer à l'utilisateur une interaction rapide et fluide. Dans une économie numérique où les start-ups se livrent une concurrence acharnée, seuls les services et produits proposant le processus le plus simple peuvent espérer gagner le marché. Comme le formule l'entrepreneur Aaron Levie : « *If you're not the simplest solution, you're the target of one* ». ²⁸

Dans le champ des interactions homme-machine (IHM), on assiste ainsi à une « course à des expériences d'utilisation lisses, sans couture et sans friction, au nom d'une plus grande efficacité » (CNIL, 2017, p.9). L'interface doit permettre à l'utilisateur d'exécuter rapidement des commandes et répond donc à un principe d'efficacité qui s'inscrit dans un contexte plus large d'accélération et de raréfaction des ressources

²⁶ Sweet Dreams est l'une des *skills* du logiciel Alexa qui propose des petites comptines pour aider l'utilisateur à s'endormir.

²⁷ Shilevitz, J. (2018). Alexa, How will you change us?, *The Atlantic*, novembre 2018, p.96.

²⁸ Levie, A. (2012). The Simplicity Thesis. *Fast Company*, 5 février 2012. Disponible sur : <https://www.fastcompany.com/1835983/simplicity-thesis>, consulté le 15 avril 2019.

temporelles (voir chapitre 2.1). A l'aube des années 2000, Steve Krug, consultant en « usabilité », publie l'ouvrage *Don't make me think* dont le titre énonçait déjà l'objectif assigné à tout concepteur d'interface : le design doit être suffisamment intuitif pour permettre à l'utilisateur de se passer de réflexion. La simplicité est également l'un des principes fondamentaux d'Apple, qui a guidé l'ensemble des étapes de développement de ses produits, de la conception à la vente, comme le souligne Ken Segall (2012, p.4) :

« There's nothing subtle about Apple's love affair with Simplicity. It's everywhere you look. It's in the company's products, its ads, its internal organization, its stores, and its customer relationship. Inside Apple, Simplicity is a goal, a work style and a measuring stick ».

Aujourd'hui, la simplicité constitue, pour tous les concepteurs de services numériques, un principe directeur. Dans ce domaine, le fameux aphorisme « *less is more* » autrefois formulé par l'architecte Mies Van der Rohe demeure une référence.

Guidée par ce principe, l'interaction proposée par les assistants vocaux s'inscrit plus spécifiquement dans le champ des interfaces « naturelles » ou NUI (*Natural User Interface*). Si l'adjectif « naturel » décrit habituellement ce qui relève de la nature d'un être ou d'une chose, dans le domaine des IHM, il désigne avant tout une interface invisible, et qui le demeure tout au long de son utilisation. La dimension naturelle constitue davantage un objectif (tendre vers une interaction naturelle) qu'une réalité matérielle. Les NUI proposent une interaction corporelle directe, sans l'intermédiaire de périphériques de commande. L'idée est de réduire le temps d'apprentissage nécessaire à la maîtrise du dispositif technique en mobilisant des modes d'interaction déjà connus de l'utilisateur. Comme l'explique Florent Di Bartolo (2017, p. 208), « la transparence des interfaces numériques se construit sur des métaphores et des modèles de conception qui appartiennent aux objets auxquels nous sommes d'ores et déjà accoutumés ». Les systèmes basés sur des interfaces « naturelles » seraient ainsi dotés d'une *affordance*, c'est-à-dire une potentialité d'action inscrite dans le dispositif technique et perceptible par son usager (Gibson, 1977). A terme, il s'agirait pour ces systèmes d'interaction de compléter, voire de remplacer les actuelles GUI (*graphical user interface*), c'est-à-dire « un environnement visuel imagé avec lequel on interagit grâce à des fenêtres, des icônes, des menus et un dispositif de pointage » (Vial, 2017, p.61). Certaines entreprises voient même dans les NUI une façon de faire « disparaître » la technologie, comme en témoigne le slogan « *using voice to make*

technology disappear » de Snips, une start-up spécialisée dans le développement d'interfaces vocales axées sur le respect de la vie privée.

Bien que la voix soit certainement l'interface la plus naturelle, converser avec des objets ne va toutefois pas de soi. Dans l'une des premières recherches menées auprès d'utilisateurs d'enceintes connectées en France, Julia Velkovska et Moustafa Zouinar (2017) observent un fossé entre la simplicité promise par les concepteurs et la réalité des usages. Tout d'abord, la reconnaissance vocale reste encore perfectible. L'utilisateur doit parfois répéter sa requête à plusieurs reprises, ce qui peut le conduire à abandonner le dispositif. Si toutefois il persévère, il s'engage alors dans un processus de « construction de sens et de réaligement » (Velkovska et Beaudouin, 2014). Cela peut se traduire par différentes actions telles que reformuler, raccourcir ou clarifier l'énoncé, s'approcher de l'enceinte ou parler plus fort. Mais le travail d'adaptation va au-delà de la formulation et comprend toutes les activités dans lesquelles l'utilisateur s'engage pour faire fonctionner le système. Il s'agit de comprendre comment réagir en cas de mauvaise réponse, mais également de se familiariser avec le processus interactionnel imposé (activer la machine en prononçant le mot déclencheur, puis formuler sa requête à l'instant approprié, lorsque que l'assistant « écoute »). Par ailleurs, la fluidité d'interaction qui constitue le cœur du discours promotionnel autour des enceintes connectées crée des attentes qui peuvent parfois être déçues. Les chercheurs constatent que lorsque le système fonctionne de façon optimale, les utilisateurs tendent à lui attribuer des compétences qu'il ne possède pas. Un exemple est le recours à des termes comme « ici » ou « ceci » qui font référence à des éléments contextuels qui ne peuvent être compris par l'assistant vocal. Si les concepteurs savent que la compréhension du contexte demeure l'un des plus grands défis de ces systèmes, l'utilisateur, lui, ne perçoit pas nécessairement l'origine de cette difficulté d'interaction. Velkovska et Zouinar soulignent le paradoxe de cette situation : plus l'interface vocale semble bien fonctionner, plus les utilisateurs tendent à parler de façon « naturelle » et donc, augmentent le risque d'être mal compris par le système, sans en saisir la raison. Une amélioration possible du dispositif, selon les chercheurs, consisterait à aider l'utilisateur à identifier la source de l'incompréhension et à le guider dans la reformulation de sa requête.

Si le dialogue avec l'assistant vocal est encore loin de s'approcher d'une conversation humaine, c'est également parce qu'une interaction naturelle suppose la compréhension d'un ensemble de signes (intonation, gestes, expressions du visage) qui s'inscrivent dans un contexte et influencent la manière dont les réponses sont

élaborées. La dimension implicite du langage (insinuation, ironie, sous-entendus) ainsi que sa capacité à véhiculer des émotions relèvent de l'expérience humaine et peuvent difficilement être apprises par la machine. Comme le note l'orthophoniste Joana Revis, il existe dans l'échange oral un phénomène de « convergence interactionnelle » qui pousse chacun des interlocuteurs à adopter les habitudes discursives de l'autre afin de créer un espace commun de communication.²⁹ Un chuchotement entrainera une réponse dans une tonalité similaire. De la même manière, un passant interpellé par un touriste étranger adaptera son langage en optant pour des mots simples et un débit de parole plus lent. Si les assistants vocaux ne peuvent s'adapter aux utilisateurs, ce sont ces derniers qui vont modifier leur façon d'interagir afin de se placer à portée de la machine.

Afin de diminuer cette asymétrie, les recherches dans le domaine des interactions vocales se tournent également vers la prise en compte de paramètres qui relèvent non plus seulement du langage, mais d'éléments propres à la conversation humaine. En mai 2018, Google présentait Duplex, son nouveau logiciel d'assistance vocale destiné à la prise de rendez-vous par téléphone. Lors de sa démonstration publique, la voix de synthèse a surpris et inquiété par son degré de réalisme.³⁰ Le ton plat et monotone qui caractérise habituellement ces voix a été remplacé par une élocution rythmée, ponctuée de diverses hésitations et intonations propres au langage humain.³¹ Comme l'expliquent les ingénieurs en charge du projet, en se focalisant sur un type d'interaction précis, il est possible d'intégrer des éléments de langage plus subtils, conférant à la voix de synthèse une tonalité proche de la parole humaine :

« One of the key research insights was to constrain Duplex to closed domains, which are narrow enough to explore extensively. Duplex can only carry out

²⁹ Vallet, F. (2018). Joana Revis: « Notre voix porte en elle toutes les intentions qui sont les nôtres ». Laboratoire d'Innovation Numérique de la CNIL (LINC), 28 mars 2018. Disponible sur : <https://linc.cnil.fr/fr/joana-revis-notre-voix-porte-en-elle-toutes-les-intentions-qui-sont-les-notres>, consulté le 6 avril 2019.

³⁰ Goode, L. (2018). How Google's Eerie Robot Phone Calls Hint At Ai's Future. Wired, 5 mai 2018. Disponible sur : <https://www.wired.com/story/google-duplex-phone-calls-ai-future>, consulté le 6 avril 2019.

³¹ Google Duplex Demo (Google I/O 2018) : <https://www.youtube.com/watch?v=znNe4pMCsD4>, consulté le 6 avril 2019.

natural conversations after being deeply trained in such domains. It cannot carry out general conversations »³²

Dans les deux situations de prise de rendez-vous montrées en exemples (un restaurant et un salon de coiffure), la voix de synthèse est capable de réagir aux formules de son interlocuteur, d'adapter les temps de pause après certains mots, de jouer sur les intonations propres à certaines expressions, et même de ponctuer ses répliques de « hum » et « euh ». En imitant une interaction naturelle, Google Duplex a pour objectif de rendre l'expérience de la conversation avec un agent vocal plus « agréable ». ³³ Cependant, le degré d'élaboration de la voix de synthèse a suscité des inquiétudes et contraint Google à intégrer un message qui prévient l'interlocuteur qu'il interagit avec une intelligence artificielle. ³⁴

Le développement de l'interface vocale des enceintes connectées va également dans le sens d'une plus grande « convergence interactionnelle » entre l'humain et la machine. En octobre 2018, Amazon annonçait ainsi que son logiciel Alexa disposait désormais d'un « mode chuchotement » (*Whisper Mode*). ³⁵ Si l'utilisateur s'adresse à l'enceinte en parlant à voix basse, celle-ci lui répond de la même façon. Cette fonction doit ouvrir la voie à l'utilisation des assistants vocaux dans de nouveaux espaces. Toutefois, le caractère presque humain des chuchotements et autres expressions émises par les assistants vocaux peut également gêner l'utilisateur. Une certaine forme de « dysfonctionnement » semble donc également appréciée, comme le relève Nicolas Santolaria (2016, p. 87) au travers de ce témoignage d'une utilisatrice de Siri : « le fait que ça ne marche pas très bien ne me dérange pas. [...] Au contraire, s'il se mettait vraiment à tenir une conversation comme un humain, je trouverais ça inquiétant ». Ce sentiment de malaise ressenti face à un objet technique aux caractéristiques proches

³² Leviathan, Y. & Yossi, M. (2018). Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone. *Google AI Blog*, 8 mai 2018. Disponible sur : <https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html>, consulté le 6 avril 2019.

³³ *Ibid.*

³⁴ Statt, N. (2018). Google now says controversial AI voice calling system will identify itself to humans. *The Verge*, 10 mai 2018. Disponible sur : <https://www.theverge.com/2018/5/10/17342414/google-duplex-ai-assistant-voice-calling-identify-itself-update>, consulté le 6 avril 2019.

³⁵ Lee, D. (2018). You can now get Alexa to use its inside voice with Whisper Mode. *The Verge*, 18 octobre 2018. Disponible sur : <https://www.theverge.com/circuitbreaker/2018/10/18/17995092/amazon-alexa-whisper-mode-guard-context-aware>, consulté le 6 avril 2019.

d'un être vivant a été théorisé par le roboticien japonais Masahiro Mori en 1970. Nommé *Uncanny Valley* (« vallée dérangeante »), cette théorie aujourd'hui discutée, postule qu'un objet situé dans une certaine « zone » de réalisme – proche de l'humain, sans toutefois l'imiter parfaitement – peut susciter le rejet de l'utilisateur.

Ces différents exemples démontrent que la complexité de l'interaction vocale va au-delà de la difficulté technique. Dans des contextes d'interactions sensibles, comme lors d'utilisation d'applications relevant des données de santé, une mauvaise compréhension des commandes peut, par ailleurs, porter préjudice à l'utilisateur (Bickmore et al., 2018). A l'inverse, lorsque le dialogue semble trop réaliste ou que la parole de l'assistant vocal s'approche d'une élocution humaine, le risque est de provoquer un sentiment de rejet. En outre, un dispositif qui fonctionne (trop) bien soulève également davantage de questionnements liés à la captation des données personnelles, comme nous le verrons au chapitre 4. Enfin, l'argument d'une interaction fluide et « sans couture » tend à faire oublier les éventuelles contraintes que pose l'interaction vocale. Contrairement à la lecture, l'écoute impose un rythme et une durée souvent inconnue pour l'utilisateur, comme le souligne la designer Jess Thornhill : « *The absence of visual cues means that users enter into an interaction with no idea of how long it will take for them to complete the purpose of the experience [...]* ». ³⁶ L'utilisateur n'est pas libre d'accélérer ou ralentir la cadence des réponses énoncées par l'assistant vocal. Il n'est pas sûr non plus que la formulation d'une commande à voix haute soit toujours la façon la plus rapide d'exécuter une action. Dans tous les cas, la perspective d'une réelle conversation fluide avec la machine reste lointaine. Aujourd'hui, ces systèmes sont encore perfectibles, avec les avantages et les inconvénients que cela peut représenter, comme le formulait la journaliste Lauren Goode en décembre 2018 : « *virtual assistants still stumble, for better or worse* ». ³⁷

³⁶ Thornhill, J. (2018). Voice Design: a guide to user. Disponible sur : <https://uxdesign.cc/voice-design-a-guide-to-user-onboarding-cbc22543803>, consulté le 6 avril 2019.

³⁷ Goode, L. (2018). Your voice assistant may be getting smarter, but it's still awkward. *Wired*, 27 décembre 2018. Disponible sur : <https://www.wired.com/story/voice-assistants-ambient-computing/>, consulté le 6 avril 2019.

3.4 Le revers de la simplicité

Au-delà des caractéristiques fonctionnelles, la simplification des interfaces et de « l'expérience utilisateur » doit également être interrogée sous l'angle de ses implications sociales et politiques. En décembre 2018, le journaliste Kevin Roose posait la question suivante : « *Is Tech Too Easy to Use ?* ». ³⁸ Si la réduction des « frictions » peut faciliter l'utilisation de nombreux services et activités numériques, cette simplicité s'accompagne également d'effets parfois inattendus et peu souhaitables. L'auteur cite ainsi l'exemple de Facebook et Twitter, dont la diffusion facilitée de messages à un large public a eu pour conséquence – entre autres – de propager rapidement de fausses informations. Le système *autoplay* de YouTube est également régulièrement décrié pour sa propension à exposer l'utilisateur à un contenu toujours plus radical, le faisant rapidement basculer dans ce que la sociologue Zeynep Tufekci a nommé un « *rabbit hole of extremism* ». ³⁹ Par ailleurs, la simplicité ouvre également de nouvelles voies pour les opérations criminelles en ligne. Les mêmes processus et architectures qui ont permis l'accélération des échanges, des informations et du commerce, ont également facilité le hameçonnage et la propagation de logiciels malveillants. Comme le souligne Justin Kossly, ancien chef de produit chez Google : « *The Internet's lack of friction made it great, but now our devotion to minimizing friction is perhaps the Internet's weakest link for security* ». ⁴⁰

De façon générale, la notion de processus fluide est souvent mobilisée pour dissimuler des pratiques peu louables. Le « partage sans friction » (*frictionless sharing*) de Facebook a avant tout pour but d'inciter les utilisateurs à passer davantage de temps sur la plateforme, afin que l'entreprise puisse en extraire plus de données. Il en va de même pour YouTube, dont la fonction *autoplay* vise à augmenter le temps de visionnage. Amazon a également créé divers outils destinés à « faciliter » l'acte d'achat, et donc à réduire le temps de réflexion qui précède la transaction. Mais la simplicité ou la « commodité » (*convenience*) est également appréciée et choisie par l'utilisateur,

³⁸ Roose, K. (2018) *Is Tech Too Easy to Use?*. *The New York Times*, 12 décembre 2018. Disponible sur : <https://www.nytimes.com/2018/12/12/technology/tech-friction-frictionless.html>, consulté le 15 avril 2019.

³⁹ Tufekci, Z. (2018). *YouTube, the Great Radicalizer*. *The New York Times*, 10 mars 2018. Disponible sur : <https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html>, consulté le 15 avril 2019.

⁴⁰ Kossly, J. (2018). *The Internet Needs More Friction*. *Motherboard*, 16 novembre 2018. Disponible sur : https://motherboard.vice.com/en_us/article/3k9q33/the-internet-needs-more-friction, consulté le 15 avril 2019.

qui la préfère à un processus complexe. Dans une tribune du New York Times, le juriste Tim Wu dénonce ainsi la « tyrannie de la commodité », qui constitue, selon lui, à la fois l'un des plus importants moteurs des décisions humaines et une force structurante de l'économie, qui participe à alimenter des logiques monopolistiques. Plus une plateforme est facile à utiliser, plus elle gagne en puissance. La commodité apparaît à la fois comme une valeur, un idéal et un mode de vie qui façonne nos habitudes. Tim Wu rappelle également que nous sommes disposés à payer plus cher pour gagner en convenance. Si elle comporte des avantages indéniables, la simplicité tend cependant à faire oublier que la difficulté est constitutive de l'expérience humaine. La simplification, voire l'automatisation des processus limite assurément nos possibilités d'analyse critique comme le souligne également Kevin Roose, « *After all, "friction" is just another word for "effort", and it's what makes us capable of critical thought and self-reflection* ». ⁴¹

Dans le même sens, en février 2019, Dominique Boullier alertait sur les conséquences de ce qu'il désigne comme un « réchauffement médiatique » alimenté par les « machines à répliques » que sont devenues les plateformes. ⁴² La quête de fluidité va nécessairement de pair avec une augmentation de la vitesse et de la quantité d'informations transmises. Le bouton *retweet* est en cela un bon exemple de dispositif censé réduire les frictions (en simplifiant la publication de contenu) et qui contribue largement à intensifier la masse de messages échangés. Mais la vitesse et la simplification des mécanismes ne permettent plus de garder une distance critique. Selon l'auteur, « la conséquence immédiate de ce réchauffement en est, pour filer la métaphore, "la fonte généralisée de la calotte d'esprit critique" qui refroidissait tout l'espace public ». ⁴³

Sans interface graphique, les assistants vocaux occultent encore davantage les mécanismes qui sous-tendent leur fonctionnement. Plus ces technologies semblent bien fonctionner, c'est-à-dire « converser » de manière fluide, plus elles tendent à faire oublier la complexité du vaste réseau matériel et humain dont elles dépendent. Dans un récent article, les chercheurs Kate Crawford et Vladan Joler (2018) mettent en lumière l'impressionnant écosystème nécessaire à la conception, à la fabrication et à l'usage de l'enceinte connectée Amazon Echo. Celui-ci est découpé en 21 parties et

⁴¹ Roose, K. (2018). Ibid.

⁴² Boullier, D. (2019). Lutter contre le réchauffement médiatique. *Internet Actu*, 5 février 2019. Disponible sur : <http://www.internetactu.net/2019/02/05/lutter-contre-le-rechauffement-mediatique/>, consulté le 15 avril 2019.

⁴³ Ibid.

représenté sous la forme d'une grande infographie décrite comme « l'anatomie d'un système d'intelligence artificielle ». ⁴⁴ Sans prétendre à l'exhaustivité, cette visualisation révèle certaines des composantes de ce qui apparaît aujourd'hui comme une « boîte noire », soit un objet technique dont le fonctionnement reste, pour l'utilisateur, largement opaque. Extractions des ressources naturelles, exploitation du travail humain, infrastructure logistique, collecte de données personnelles et traitements algorithmiques sont quelques-unes des caractéristiques de ce vaste réseau. Si l'on pourrait relever que l'exploitation de ces ressources reste marginale en comparaison d'autres secteurs industriels, les auteurs soulignent avant tout le paradoxe entre la simplicité des gestes que la machine prétend remplacer et la quantité de ressources nécessaires à son usage. Qu'il s'agisse d'éteindre une lumière, d'allumer la radio ou de programmer un minuteur, chacune de ces actions ordinaires fait appel à un réseau de ressources vaste et complexe dès lors qu'elles transitent par une enceinte connectée, comme le relèvent les chercheurs : « *The scale of resources required is many magnitudes greater than the energy and labor it would take a human to operate a household appliance or flick a switch* » (Crawford et Joler, 2018, p.3).

En ce sens, le *cloud* dans lequel sont stockées les nombreuses données nécessaires au fonctionnement des assistants vocaux est une métaphore qui obscurcit largement la réalité opérationnelle de ces machines. Image par excellence d'un système « sans couture », voire transparent, le « nuage » implique en réalité une imposante infrastructure physique composée de lignes téléphoniques, de câbles sous-marins, de satellites et d'immenses *datacenters*. Selon l'auteur James Bridle (2018), ce manque de transparence ne serait pas uniquement le résultat d'une complexité technique mais procéderait d'un choix délibéré de certains acteurs. Ce qui « s'évapore » dans le *cloud*, c'est également ce qui relevait auparavant de notre propriété et de notre agentivité, soit notre capacité d'agir. Le *cloud* institue une relation de pouvoir à distance entre les détenteurs de ces infrastructures et les individus qui leur fournissent des fragments numériques de leur vie.

Le *cloud*, comme l'ensemble des systèmes destinés à optimiser l'expérience utilisateur en lui offrant un parcours « sans friction », entre ainsi en contradiction avec la nécessité de transparence demandée par de nombreux appels citoyens (Cellars et Masure, 2018). Même si les grands acteurs du numérique (Google, Facebook, Twitter, Apple, Amazon) fournissent aujourd'hui des « rapports de transparence », ces

⁴⁴ <https://anatomyof.ai/img/ai-anatomy-map.pdf>, consulté le 17 avril 2019.

derniers restent relativement opaques et n'offrent que peu d'informations permettant aux utilisateurs de saisir leurs modalités de fonctionnement. La transparence s'apparente alors davantage à la simple transmission d'une vaste collection de données brutes censées faire apparaître des « vérités » plutôt qu'à une réelle mise en visibilité des processus, comme le soulignent Cellars et Masure (2018, p.106) : « [...] ces initiatives versent dans une forme de *data washing* (« lavage de cerveau par les données ») : une mise à distance des informations controversées par un trop-plein de diffusion de données "brutes" ». Les données seules ne fournissent que peu d'information. C'est avant tout la manière dont celles-ci ont été élaborées et traitées qui importe et reste largement inaccessible.

La simplification, voire l'automatisation des processus s'accompagne donc d'une forme de complexification et d'opacification qui occulte encore les logiques de pouvoir qui sous-tendent les services numériques. Les nombreux « objets connectés » prennent progressivement place dans notre quotidien sans que nous puissions accéder à leurs mécanismes. Pourtant, ces objets ne peuvent fonctionner sans une forme de surveillance qui doit aujourd'hui être rendue davantage visible, comme le relève Florent Di Bartolo (2017, p.218) : « tout dispositif technologique ne devrait pas avoir pour objectif d'échapper à notre regard, mais au contraire de pouvoir être observé [...] dans toute sa complexité ».

4. Les enjeux de la personnalisation

Avec l'interface vocale, la personnalisation des services est l'une des fonctionnalités centrales des assistants vocaux. Obtenir une information « sur-mesure » participe de la fluidité d'interaction mise en avant par les concepteurs. De façon générale, la personnalisation peut être définie comme un processus visant à « définir les besoins supposés ou réels de l'utilisateur en conduisant un ensemble de recherches sur celui-ci et sur son environnement, [...] ceci afin de définir les principes d'usages et d'interactions du service. » (CNIL, 2019, p.9). Il s'agit d'anticiper les préférences et les désirs de l'utilisateur pour le « guider » parmi les multiples options qui s'offrent à lui. Aujourd'hui, le filtrage proposé par de nombreuses plateformes paraît nécessaire pour s'orienter dans le flux d'informations quotidien. Il permet de choisir le contenu qui semble présenter un intérêt pour chacun. Comme l'explique

Dominique Cardon (2019, p.374), « l'évolution vers la personnalisation est à la fois une conséquence de l'abondance des informations disponibles et la réponse à une attente de nos sociétés où les choix et centres d'intérêt s'individualisent ».

Mais ce qui est présenté comme une amélioration de « l'expérience utilisateur » peut aussi drastiquement limiter ses choix. En effectuant un tri en amont, la machine prive l'utilisateur de sa liberté de comparer et d'évaluer chacune des options possibles. Par ailleurs, ce travail de sélection nécessite l'obtention de nombreuses données personnelles. La personnalisation est en effet d'autant plus efficace que le système dispose d'une quantité d'informations suffisantes pour déterminer le profil d'un individu et anticiper ses actions. Ce principe pose donc nécessairement des questionnements qui concernent la vie privée des utilisateurs et fait l'objet d'une négociation parfois complexe, comme nous le verrons dans ce chapitre.

4.1 Réduction des possibles

« Tout savoir. Sur tout. Partout »⁴⁵, voici l'une des nombreuses promesses faites par le logiciel Siri. Cela comprend des recherches relativement simples, comme celle d'un itinéraire, de la météo ou de l'état du trafic, mais également des questionnements plus larges sur l'actualité, la politique, l'histoire ou la science, par exemple. L'assistant vocal propose à l'utilisateur d'effectuer des requêtes de la même façon qu'il le ferait au travers d'un moteur de recherche doté d'une interface graphique. Mais à la différence d'une page web qui affiche des milliers de résultats (bien que filtrés et hiérarchisés), l'interface vocale n'offre qu'une réponse unique. Il serait effectivement difficile d'envisager la voix de synthèse lister toute une série de liens. L'expérience « fluide » promise par ces systèmes serait sensiblement diminuée. Cependant, ce mode de fonctionnement implique que l'utilisateur n'est plus libre de choisir sa source d'information, celle-ci est désormais sélectionnée en amont par la machine. Il importe alors de comprendre quels sont les enjeux d'un tel processus de sélection qui limite nécessairement l'horizon des possibles.

L'idée de ne proposer qu'une solution aux requêtes des utilisateurs n'est pas nouvelle et précède l'apparition des assistants vocaux. En 2007, l'ingénieur William Tunstall-Pedoe lance avec son équipe un site Web nommé « True Knowledge » qui

⁴⁵ <https://www.apple.com/fr/siri/>, consulté le 7 avril 2019.

offre des réponses uniques à tout un ensemble de questions. L'objectif de cette plateforme était alors de remettre en question le principe général des moteurs de recherche qui présente une liste de résultats basée sur des mots-clés, comme l'explique Tunstall-Pedoe (2010, p.80) :

« The motivation for the project was to tackle what might be regarded as the "holy grail" of Internet search, replacing larger and larger numbers of keyword-based lists of links with perfect, direct answers to naturally phrased queries on any subject ».

L'idée de cette plateforme repose donc sur le passage d'une quête par mots-clés à une recherche formulée de façon naturelle, en rédigeant des phrases proches d'une conversation orale. Par ailleurs, True Knowledge avait pour objectif d'accumuler une connaissance lui permettant de répondre aux questions sans devoir passer par d'autres sites. En mobilisant un système algorithmique complexe qui permet de classer et de relier ces informations, le site tentait de saisir le sens d'une question et d'offrir à l'utilisateur l'unique réponse qui semblait appropriée. Grâce à divers investissements, le site True Knowledge se transforme en 2012 en une application mobile nommée « Evi ». Disponible sur iPhone et Android, celle-ci vient directement concurrencer le logiciel Siri lancé peu auparavant.⁴⁶ Son succès rapide lui vaut d'être rachetée, seulement quelques mois plus tard par Amazon, qui s'appuya largement sur ce système pour développer son propre assistant vocal, Alexa.⁴⁷

L'apparition des assistants vocaux a donné sens aux recherches menées par Tunstall-Pedoe depuis les années 1990. Avec les interfaces vocales, le principe d'une réponse unique n'est plus une possibilité, mais une nécessité. D'autres acteurs du numérique travaillent également dans ce sens. En 2012, Google présentait son *Knowledge Graph*, un outil qui permet au moteur de recherche d'agencer et de structurer ses résultats en fonction de critères sémantiques. Ce système est supposé faire gagner du temps à l'utilisateur en lui fournissant directement l'information qu'il recherche, sans passer par la consultation d'autres sites. Comme annoncé sur le blog

⁴⁶ Metz, R. (2012). New Virtual Helper Challenges Siri. *MIT Technology Review*, 27 janvier 2012, Disponible sur : <https://www.technologyreview.com/s/426723/new-virtual-helper-challenges-siri>, consulté le 7 avril 2019.

⁴⁷ Bryce, E. (2017). How Amazon's Alexa was 'born' and where voice-controlled tech will take us next. *Wired UK*, 27 février 2017. Disponible sur : <https://www.wired.co.uk/article/amazon-alexa-ai-evi>, consulté le 7 avril 2019.

officiel de Google, il s'agit du premier pas vers une nouvelle forme de recherche orientée vers des résultats uniques :

« *This is a critical first step towards building the next generation of search [...]. We've always believed that the perfect search engine should understand exactly what you mean and give you back exactly what you want* ». ⁴⁸

Dans la continuité de ce principe, Google propose aujourd'hui toujours davantage de résultats sous forme de « *featured snippets* », soit de petits encadrés qui s'affichent en en haut de la page et qui contiennent un extrait de texte (et parfois une image) issu d'un autre site (Fig 1. et Fig 2.). Les résultats ainsi présentés devraient permettre à l'utilisateur d'obtenir une réponse suffisamment précise et complète sans quitter le moteur de recherche. Lorsqu'une recherche est effectuée au travers de l'interface vocale Google Assistant, ce sont très souvent les *featured snippets* qui sont « lus » à l'utilisateur.

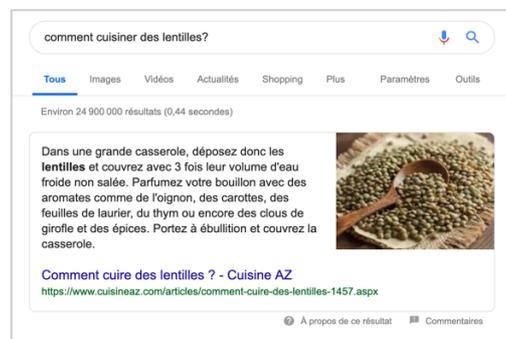
Figure 1.

Featured snippet extrait de Wikipedia



Figure 2.

Featured snippet extrait d'un site de cuisine



Toutefois, en offrant une réponse unique de façon automatique, le risque demeure important de transmettre une information erronée, sans que l'utilisateur puisse en connaître la source. En mars 2017, le site *The Outline* publiait de nombreux *featured*

⁴⁸ Introducing the Knowledge Graph: things, not strings. *Google Official Blog*, 12 mai 2012. Disponible sur : <https://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html>, consulté le 7 avril 2019.

snippets au contenu préoccupant.⁴⁹ Quelques semaines plus tard, Danny Sullivan, responsable des relations publiques pour le moteur de recherche, faisait également part de ces erreurs, qu'il attribue à un manque de vérification des sources :

*« Last year, we took deserved criticism for featured snippets that said things like “women are evil” or that former U.S. President Barack Obama was planning a coup. We failed in these cases because we didn’t weigh the authoritative nature of results strongly enough for such rare and fringe queries ».*⁵⁰

En choisissant parmi les informations disponibles celle qui semble correcte, ces entreprises endossent – qu'elles le veuillent ou non – le rôle de garde-fou, comme le souligne le magazine *Wired* : « *Whether or not these companies wish to play the role of Fact-Checker to the World, they're backing themselves into it* ». ⁵¹ Pour l'instant, Google se protège derrière une citation minimale des sources en introduisant certaines réponses vocales par « selon Wikipedia » ou « d'après le Guardian ». Mais d'autres systèmes, comme celui d'Amazon, répondent de façon beaucoup plus affirmative, en se passant d'introduction.⁵² Lorsqu'un utilisateur lance une requête, l'objectif d'Amazon est avant tout de le diriger le plus rapidement possible sur sa plateforme de vente en ligne. Toute entrave éventuelle à l'acte d'achat est donc supprimée.

En plus du risque important d'erreur, le principe d'une réponse unique donne l'illusion que la complexité du monde peut être réduite à des faits simples et objectifs. La contextualisation, la comparaison entre différentes sources, le développement d'argumentaires qui permettent d'appréhender une information parfois complexe ne sont plus à la portée directe de l'utilisateur. Il semble en effet peu probable de voir un usager de Amazon Echo ou Google Home se connecter à une interface graphique suite à la réponse donnée par l'assistant vocal afin d'en vérifier l'origine. Cet effort va

⁴⁹ Jeffries, A. (2017). Google's featured snippets are worse than fake news. *The Outline*, 5 mars 2017. Disponible sur : <https://theoutline.com/post/1192/google-s-featured-snippets-are-worse-than-fake-news?zd=2&zi=do3kqbxq>, consulté le 8 avril 2019.

⁵⁰ Sullivan, D. (2018). A reintroduction to Google's featured snippets. *Google Blog*, 28 janvier 2018. Disponible sur : <https://www.blog.google/products/search/reintroduction-googles-featured-snippets/>, consulté le 8 avril 2019.

⁵¹ Vlahos, J. (2019) Alexa, I Want Answers. *Wired*, Mars 2019, p.67.

⁵² Reese, B. (2017). Amazon's Alexa vs. Google's Assistant: Same Questions, Different Answers. *Gigaom*, 12 juin 2017. Disponible sur : <https://gigaom.com/2017/06/12/amazons-alexa-vs-googles-assistant-same-questions-different-answers/>, consulté le 8 avril 2019.

à l'encontre du principe de fluidité et de simplicité mis en avant par les concepteurs de ces technologies. L'utilisateur se retrouve alors enfermé dans une « bulle de filtres » (Pariser, 2011), un espace informationnel personnalisé, créé à son insu, et dont il ne peut que difficilement s'extraire.

4.2 Vers une économie de l'intention ?

Cette transformation progressive portée par le développement des interfaces vocales questionne également le modèle économique sur lequel reposent de nombreux services en ligne. Aujourd'hui, la majorité des plateformes sont financées grâce à la vente d'audience à des annonceurs, revers de la gratuité des services. La capacité de ces acteurs à mobiliser un large public a conduit les marques et entreprises à se tourner vers cet espace, principalement au détriment de la presse écrite. Google et Facebook restent les deux principaux vecteurs de ces annonces, bien qu'Amazon tente progressivement de s'imposer aussi sur ce marché.⁵³ Dans un espace saturé d'informations, toute organisation ou personne qui souhaite vendre des biens, des services ou communiquer une information tente d'être visible et d'attirer l'attention par différents moyens. Ce modèle n'est pas nouveau et a déjà été énoncé par Herbert Simon dans les années 1970, qui le nomme alors « économie de l'attention » (Simon, 1971). L'attention est un bien rare et convoité qui constitue – davantage que l'information – un puissant moteur économique (Goldhaber, 1997). Depuis le début des recherches en ligne, certains acteurs ont donc cherché à apparaître aux yeux de l'internaute en agissant sur la manière dont l'information est classée et ordonnée. Cette quête d'attention a notamment donné naissance au florissant marché du référencement, le SEO (*Search Engine Optimization*). Les professionnels de ce domaine proposent aux sites d'optimiser leur classement dans les résultats de Google, détournant ainsi le référencement « naturel » basé sur un principe d'autorité déterminant la « force sociale de la page » (Cardon, 2015) selon le nombre de liens qui s'y rapportent.

⁵³ Soper, S. (2018). Amazon Increases Ad Market Share at Expense of Google, Facebook. *Boomborg*, 19 septembre 2018. Disponible sur : <https://www.bloomberg.com/news/articles/2018-09-19/amazon-increases-ad-market-share-at-expense-of-google-facebook>, consulté le 9 avril 2019.

Lorsque la recherche ne s'effectuait qu'au travers d'un ordinateur, il s'agissait pour les sites de se trouver dans les dix premiers liens affichés. Avec l'apparition du mobile, l'objectif s'est réduit aux cinq premiers liens. Aujourd'hui, les requêtes vocales imposent d'apparaître en première position. Car dans la majorité des cas, c'est le résultat situé en « position zéro », soit juste en-dessous du contenu payant, qui est choisi par l'assistant vocal. Atteindre la position zéro nécessite une stratégie qui ne correspond plus aux méthodes connues de SEO. Le choix de bons mots-clés, par exemple, constitue moins un élément déterminant. L'attention devra être portée sur la manière dont les questions pourraient être formulées par les utilisateurs, dans un langage plus naturel.⁵⁴ Pour l'instant, il n'existe pas d'équivalent aux Google Ads – soit du contenu payant – pour les requêtes vocales. Mais ce n'est certainement qu'une étape intermédiaire. La commercialisation des résultats vocaux devrait se concrétiser et transformer fondamentalement le marché publicitaire en ligne, comme l'expliquait le publicitaire Jared Belsky en mars 2019 au magazine Wired : « *There's going to be a battle for shelf space, and each slot should theoretically be more expensive. [...] It's the same amount of interest funneling into a smaller landscape* ». ⁵⁵ Selon lui, cette course pour la première place sera particulièrement marquée pour les entreprises qui vendent des produits en ligne, notamment au travers d'Amazon. Avec le développement des recherches vocales, il s'agit moins d'attirer l'attention de l'utilisateur que de parvenir à lui proposer une seule et unique réponse pertinente.

En avril 2018, Sébastien Misoffe, directeur général de Google France, affirmait dans une tribune du journal Le Monde : « Nous croyons qu'Internet est une économie de l'intention et non de l'attention ». ⁵⁶ Ses propos font écho au concept d'« économie de l'intention », formulé par le journaliste Doc Searls dans son ouvrage *The Intention Economy* (2012). Celui-ci estime que la valeur ne réside plus dans l'attention en soi, mais dans l'*intention*, c'est-à-dire dans la capacité à répondre à la volonté exprimée de l'utilisateur. La mise en œuvre de ce principe repose sur un système qu'il nomme le *Vendor Relationship Management* (VRM), en opposition à la logique marketing du

⁵⁴ Poiré, A-S. (2019). Le référencement naturel à l'ère des assistants vocaux. *Infopresse*, 24 février 2019. Disponible sur : <https://www.infopresse.com/article/2019/2/24/le-seo-a-l-ere-des-assistants-vocaux>, consulté le 9 avril 2019.

⁵⁵ Vlahos, J. (2019). Ibid. p. 65.

⁵⁶ Misoffe, S. (2018). Big data : « Chez Google, nous croyons qu'Internet est une économie de l'intention et non de l'attention ». *Le Monde*, 19 octobre 2018. Disponible sur : https://www.lemonde.fr/idees/article/2018/10/19/big-data-chez-google-nous-croyons-qu-internet-est-une-economie-de-l-intention-et-non-de-l-attention_5372003_3232.html, consulté le 10 avril 2019.

CRM pour laquelle le prestataire détient l'ensemble des données relatives au client, avec pour but principal de le retenir « captif » (Kessous, Mounier et Mallard, 2005). Le VRM propose au contraire que le client détienne le contrôle de ses données personnelles et puisse les transmettre selon son bon vouloir à l'entreprise de manière à personnaliser un service ou obtenir une offre plus adaptée.

Si la déclaration de Sébastien Misoffe – dont le choix des mots n'est pas anodin – semble faire référence au concept proposé par Doc Searls, en particulier lorsqu'il précise que Google souhaite « proposer à l'utilisateur du contenu pertinent et non intrusif » ou encore que l'enjeu de l'entreprise consiste à « saisir pleinement les opportunités économiques offertes par le Web, sans transiger sur le respect dû à l'utilisateur en lui redonnant le contrôle sur ses usages numériques »⁵⁷, la manière dont le contrôle des données est envisagé par Google diffère largement du principe d'une « économie de l'intention » basée sur un système VRM. Il ne s'agit pas pour Google de mettre les données à disposition de l'utilisateur afin qu'il puisse choisir l'usage qui en sera fait. Son objectif consiste à le guider de façon toujours plus fine et discrète en anticipant encore davantage son comportement grâce à l'analyse de ses expériences passées. L'économie de l'intention est alors celle des volontés inférées, plutôt que des motivations ouvertement déclarées et partagées par les utilisateurs. Comme le soulignent Antoinette Rouvroy et Thomas Berns, en laissant s'opérer la collecte et l'usage de traces par défaut sans en connaître la finalité, « nous semblons nous mouvoir ici au plus loin de toute forme d'intentionnalité » (Rouvroy et Berns, 2013, p. 169).

Les assistants vocaux prennent appui précisément sur un discours qui présente, dans un univers *a priori* vidé de toute sollicitation marchande, un utilisateur invité à transmettre spontanément ses besoins, désirs et préférences afin que l'intelligence artificielle puisse lui fournir une réponse au plus proche de ses attentes. Il s'agirait d'un simple « dialogue » qui permettrait à la machine de proposer un service personnalisé. Cependant, les informations transmises oralement sont couplées à toute une série d'autres données qui vont permettre d'établir un profil et d'anticiper les propositions qui seront faites par l'assistant vocal. Les sollicitations publicitaires paraissent alors moins invasives, car elles sont directement intégrées aux actions quotidiennes sous forme de subtiles suggestions. Dans le film *Her*, Spike Jonze imagine même un monde où la publicité aurait disparu de l'espace public. Les marques existent encore, mais leur manière de communiquer s'est transformée. La relation commerciale procède d'une

⁵⁷ Ibid.

négociation directement gérée par l'intelligence artificielle qui connaît précisément le profil de son utilisateur, comme le relève David Carroll, professeur à la Parsons School of Design :

*« AI has obviated ads in favor of the deeply intimate personal assistant that knows everything about you and negotiates with other bots on your behalf behind-the-scenes for your products and services ».*⁵⁸

Un lien de confiance s'est créé entre l'utilisateur et son assistant vocal. Celui-ci « filtre » les informations pour ne proposer que le contenu qui apparaît pertinent. Cette surveillance quotidienne semble être acceptée dès lors que l'intelligence artificielle est perçue comme un guide bienveillant qui écarte les sollicitations non désirées. Bien que les assistants vocaux disponibles aujourd'hui ne permettent pas une telle prise de contrôle, leur développement semble toutefois aller dans le sens d'une anticipation toujours plus fine de nos actions. Selon Shoshana Zuboff (2019), ce processus – qui s'applique à l'ensemble des technologies prédictives – va aujourd'hui au-delà de l'anticipation, car il entend directement *agir* sur nos actions plutôt que de simplement les prévoir. C'est ce que l'auteure appelle « économie de l'action », soit des procédés destinés à intervenir dans des situations réelles. Comme elle le soulignait dans un récent article du Monde Diplomatique, « ces interventions visent à augmenter la certitude en influençant certaines attitudes : elles ajustent, adaptent, manipulent, enrôlent par effet de groupe, donnent un coup de pouce ».⁵⁹

Sous couvert d'offrir un service « sur-mesure » et efficace, les concepteurs de ces technologies procèdent à un profilage dont le processus demeure largement obscur, mais qui nécessite, dans tous les cas, la collecte de nombreuses données personnelles. L'économie de l'intention, telle qu'elle est envisagée par les principaux acteurs du numérique reste donc avant tout une économie de la captation des traces numériques destinée à prédire – voire modeler – les préférences, les décisions et les actions des individus dans un nombre toujours plus vaste de sphères sociales, transformant nécessairement leur expérience et perception du monde.

⁵⁸ Carroll, D. (2016). Spike Jonez's "Intention Economy". Disponible sur : <https://medium.com/@profcarroll/spike-jonezs-intention-economy-a0f5f4ea595d>, consulté le 11 avril 2019.

⁵⁹ Zuboff, S. (2019). Un capitalisme de surveillance. *Le Monde Diplomatique*, janvier 2019, p.11

4.3 *Le paradoxe de la vie privée*

La personnalisation des services implique la collecte d'une importante quantité de données personnelles. En ce sens, elle pose nécessairement des questionnements qui concernent le respect de la vie privée des utilisateurs. Mais comment qualifier ce qui relève ou non de la vie privée ? Donner une définition claire de cette notion reste difficile tant elle est mobilisée dans des contextes variés. Elle peut désigner à la fois « une revendication de la part des individus ou de la société civile, une situation de fait, une forme de contrôle, un droit garanti légalement ou encore une valeur morale en soi » (Rochelandet, 2010, p.7). La notion de vie privée varie selon les époques, les lieux et les personnes concernées. Définie à la fin du XIX^e siècle comme un droit à être laissé tranquille (Brandeis et Warren, 1890), le concept de vie privée a évolué avec le passage progressif à une société de l'information. Depuis les années 1960 et 1970, « le soi est exprimé, défini et influencé à travers et par l'information et les technologies de l'information » (Acquisti, 2011, p.110). La vie privée est toujours davantage associée à la protection des données personnelles, si bien que l'on parle parfois de *vie privée informationnelle* (Coll, 2014).

C'est précisément cette notion de la vie privée qui nous intéresse dans le contexte du développement des assistants vocaux. Pour proposer une expérience personnalisée, le dispositif doit pouvoir accéder à différentes informations (conversations, localisation, calendrier, e-mail, historique de navigation, etc.). Si l'on peut imaginer que le respect de la vie privée soit une préoccupation centrale à une époque où les dispositifs numériques tracent un nombre toujours plus important de nos activités quotidiennes, le comportement des individus face à cette intrusion dans leur sphère intime relève parfois de logiques peu rationnelles. C'est ce que traduit le célèbre « paradoxe de la vie privée », selon lequel nombre d'entre nous se diraient préoccupés par le respect de leur sphère intime, tout en rendant disponibles de nombreuses données personnelles (Lancelot Miltgen, 2011 ; Rallet et Rochelandet, 2011).

Selon le sociologue Sami Coll (2014), qui a mené une étude auprès des utilisateurs de cartes de fidélité, l'une des explications possibles de ce paradoxe réside dans le décalage entre la *vie privée objectivée*, c'est-à-dire une définition normative de la vie privée telle qu'elle est envisagée par ses défenseurs, et la *vie privée subjective*, soit une notion qui peut varier selon les perceptions de chacun. Les clients interviewés auraient bien une vision de la vie privée, mais celle-ci ne correspond pas toujours à celle du législateur. Par ailleurs, l'étude met en lumière une troisième dimension : la *vie privée*

vécue, c'est-à-dire la façon dont celle-ci s'articule au quotidien lors des interactions marchandes observées. La *vie privée vécue* peut encore différer des deux autres notions, comme le souligne Sami Coll (2014, p. 59) : « la *vie privée vécue* est sensible à des interactions très concrètes (demande en face-à-face de la date de naissance, demande de présentation d'une carte d'identité, etc.) qui ne font très souvent pas écho aux définitions subjectives et objectivées de la vie privée ». A l'inverse, les individus semblent parfois peu regardants face à des demandes qu'ils auraient eux-mêmes jugées sensibles dans leurs discours. Ces différences montrent plus généralement l'importance du contexte dans la perception de la notion de sphère intime. Acquisti et al. (2015, p.511) relèvent également l'influence de l'environnement dans lequel les questions de vie privée prennent place et proposent la notion de « *context-dependence* », qu'ils définissent ainsi :

« Applied to privacy, context-dependence means that individuals can, depending on the situation, exhibit anything ranging from extreme concern to apathy about privacy. Adopting the terminology of Westin, we are all privacy pragmatists, privacy fundamentalists or privacy unconcerned, depending on time and place. »

La notion de contexte semble importante pour comprendre les craintes que soulèvent les enceintes connectées. Au cours des trois dernières années, l'arrivée de ces nouvelles « machines parlantes » a fait l'objet de nombreux articles de presse, le plus souvent pour souligner les risques liés à leur intrusion dans notre sphère privée. Pourtant, des dispositifs d'interaction vocale existaient déjà auparavant, notamment sous forme de logiciels intégrés dans les *smartphones* (Siri, Google Assistant). Malgré un fonctionnement similaire, basé sur la collecte de données personnelles et l'enregistrement des paroles de l'utilisateur, ces interfaces vocales ne semblaient pas susciter autant d'inquiétude que les enceintes connectées. Si les raisons exactes de ces perceptions différenciées ne sont pas connues, il est tout de même possible d'envisager que le contexte et le mode d'interaction dans lequel se déploient ces différents dispositifs puissent jouer un rôle.

Tout d'abord, l'enceinte connectée ne disposant pas de support visuel, il est difficile de déterminer à quel moment celle-ci enregistre effectivement la voix de son utilisateur. Bien qu'ils doivent être activés par un « mot-déclencheur » (*wake word*), ces objets semblent être en permanence « sur écoute ». Quelques cas d'activations non

souhaitées, largement reportés par les médias, ont conforté l'image d'un système intrusif et peu sécurisé.^{60 61} Ensuite, contrairement au *smartphone*, l'enceinte connectée a pour *seule* fonction de prêter attention aux conversations de son utilisateur. Il ne s'agit plus d'une option parmi d'autres qu'il est possible d'ignorer ou de désactiver, l'enregistrement des voix est désormais au centre du dispositif. La méfiance apparaît plus prononcée à l'égard d'une technologie qui ne repose que sur un système d'écoute difficilement désactivable. Enfin, le fait que l'enceinte prenne place dans l'intimité du foyer apparaît également comme une source d'inquiétude.⁶² L'espace domestique est en effet considéré comme le lieu de l'intime par excellence (Zouinar et Fréjus, 2010). Cette problématique n'est pas nouvelle et avait déjà été relevée lors de l'apparition des *smart TV* équipées de systèmes de reconnaissance vocale.⁶³ Le principe d'un dispositif fixe installé dans la sphère privée paraît donc susciter davantage d'appréhension qu'un système mobile qui suit l'utilisateur dans chacun de ses déplacements, y compris à l'intérieur de son domicile.

Toutefois, ces craintes semblent avant tout constituer un frein à l'achat du dispositif plutôt qu'une résistance à son utilisation, une fois celui-ci déjà acquis. Dans une étude menée sur l'adoption et l'usage des enceintes connectées, Lau et al. (2018) montrent que les non-utilisateurs ne sont pas intéressés à acquérir cet objet, car ils n'en perçoivent pas l'utilité ou craignent une intrusion dans leur sphère privée. Les utilisateurs d'enceintes connectées, quant à eux, ne semblent exprimer que peu d'inquiétude à ce sujet. Leur décision d'adopter cette technologie est motivée, pour certains, par la promesse d'un confort accru, et pour d'autres, par la volonté de faire

⁶⁰ Six, N. (2018). Une enceinte connectée d'Amazon envoie une conversation privée par erreur. *Le Monde*, 25 mai 2018. Disponible sur : https://www.lemonde.fr/pixels/article/2018/05/25/une-enceinte-connectee-d-amazon-envoie-une-conversation-privee-par-erreur_5304453_4408996.html, consulté le 19 avril 2019.

⁶¹ Tual, M. (2018). Une enceinte connectée a-t-elle été « témoin » d'un meurtre aux États-Unis ? *Le Monde*, 19 novembre 2018. Disponible sur : https://www.lemonde.fr/pixels/article/2018/11/19/une-enceinte-connectee-d-amazon-a-t-elle-ete-temoin-d-un-double-meurtre-aux-etats-unis_5385647_4408996.html, consulté le 19 avril 2019.

⁶² Santolaria, N. (2019). « Alexa, passe-moi le sel ! » : comment les enceintes connectées s'incrument peu à peu dans nos foyers. *Le Monde*, 29 mars 2019. Disponible sur : https://www.lemonde.fr/m-perso/article/2019/03/29/avec-alexa-siri-ou-djingo-la-famille-sous-assistance-artificielle_5443154_4497916.html, consulté le 19 avril 2019.

⁶³ Gibbs, S. (2015). Samsung's voice-recording smart TVs breach privacy law, campaigners claim. *The Guardian*, 27 février 2015. Disponible sur : <https://www.theguardian.com/technology/2015/feb/27/samsung-voice-recording-smart-tv-breach-privacy-law-campaigners-claim>, consulté le 22 avril 2019.

partie des « *early adopters* ». Ces derniers ont ainsi acquis l'objet sans en connaître préalablement les fonctionnalités ou les risques. Les chercheurs notent également que les utilisateurs d'enceintes connectées ne possèdent qu'une compréhension limitée des enjeux relatifs à la collecte de données personnelles. Certains disent ne pas être concernés par ces questions, tandis que d'autres estiment que les constructeurs ne sont pas en mesure de collecter et d'analyser l'ensemble des conversations entendues par le système. De façon générale, les utilisateurs tendent à privilégier le confort à leur vie privée.

Divers arguments relevés dans cette étude font écho à certains concepts déjà explorés dans la littérature, notamment dans le domaine de l'économie comportementale. Les études empiriques menées par Spiekermann et al. (2002) ou Hann et al. (2002) ont par exemple montré que des individus qui se disent concernés par le respect de leur vie privée étaient prêts à divulguer des informations qu'ils jugeaient sensibles pour des raisons de commodité ou contre l'obtention d'une récompense relativement faible, mais fournie de façon immédiate. Par ailleurs, Alessandro Acquisti relève que les individus disposent d'une information souvent incomplète quant aux risques encourus. Il souligne ainsi que « les technologies de l'information peuvent rendre les atteintes à la vie privée omniprésentes et invisibles » (Acquisti, 2011, p.111). Les inconvénients ne sont perçus que de façon *ex-post*, à travers une expérience vécue. D'autres phénomènes, tels que l'illusion de contrôle ou le biais d'optimisme (Weinstein, 1989) – qui amènent les individus à penser qu'ils sont moins exposés à un danger que d'autres personnes – se traduisent également par une sous-estimation des risques liés à la collecte de données personnelles. De façon générale, les actions et les effets à court terme sont privilégiés à la protection des risques à plus long terme. Les premiers sont directs et immédiats (confort ou rapidité, par exemple), tandis que les seconds demeurent lointains et incertains.

Outre les différents biais cognitifs, un argument couramment mobilisé par les utilisateurs pour justifier leur manque d'intérêt pour la protection de leur vie privée est celui selon lequel « qui n'a rien à cacher n'a rien à craindre » (*nothing to hide, nothing to fear*). Cette justification est également régulièrement avancée par différents acteurs qui plaident pour une transparence des activités en ligne, pour des raisons économiques ou politiques. Eric Schmidt, ancien CEO de Google avait ainsi déclaré en 2009 : « *If you have something that you don't want anyone to know, maybe you shouldn't*

be doing it in the first place». ⁶⁴ Ces propos masquent l'évidente dimension subjective et contextuelle de ce qu'il convient de dissimuler. Des expressions politiques, sexuelles ou religieuses acceptées en un lieu peuvent faire ailleurs l'objet d'une sévère répression. A cela s'ajoute la difficulté déjà évoquée d'établir une définition stable et commune du concept même de vie privée. Néanmoins, l'argument demeure très présent dans les discours populaires sur la surveillance où il est souvent présenté sous la forme d'une pesée d'intérêts entre vie privée et sécurité (Solove, 2007). Selon ses défenseurs, la vie privée – souvent insignifiante – n'aurait que peu de valeur face à l'importance de garantir la sécurité des citoyens. Les utilisateurs d'enceintes connectées qui mobilisent cet argument estiment également que leurs conversations privées sont anodines et que cette intrusion n'est qu'un modeste prix à payer pour pouvoir profiter des nombreux avantages fournis par la technologie (Lau et al., 2018).

Malgré les diverses raisons qui peuvent conduire les utilisateurs à transmettre leurs données, le respect de la vie privée demeure une préoccupation centrale pour de nombreux individus et organisations. Certaines entreprises l'ont bien compris et la défense de la vie privée constitue parfois un argument marketing (Coll, 2014). Apple en a fait la démonstration lors du Consumer Electronics Show (CES) de Las Vegas en janvier 2019 en déployant un immense panneau publicitaire sur lequel était inscrit le message suivant : « *What happens in your iPhone stays in your iPhone* ». ⁶⁵ Depuis déjà plusieurs années, la marque se distingue de ses concurrents – dont le modèle économique repose sur la collecte et l'exploitation des données personnelles – en faisant du respect de la vie privée l'une de ses priorités. Dans le même sens, l'entreprise française Snips propose une alternative aux assistants vocaux développés par les grandes sociétés du numérique en misant sur un système *privacy by design* qui intègre la protection des données des utilisateurs dès la conception. ⁶⁶ Toutefois, pour espérer conquérir un large public, ces systèmes doivent proposer une expérience fluide et « confortable ». Comme le rappelle le chercheur David Clark, la commodité reste, plus que la protection des risques, un facteur déterminant :

⁶⁴ Esguerra, R. (2009). Google CEO Eric Schmidt Dismisses the Importance of Privacy. *Electronic Frontier Foundation*, 10 décembre 2009. Disponible sur : <https://www.eff.org/fr/deeplinks/2009/12/google-ceo-eric-schmidt-dismisses-privacy>, consulté le 22 avril 2019.

⁶⁵ Savov, G. (2019). Apple trolls CES with a giant dig at Android and Alexa privacy. *The Verge*, 5 janvier 2019. Disponible sur : <https://www.theverge.com/2019/1/5/18169781/apple-google-privacy-troll-billboard>, consulté le 15 avril 2019.

⁶⁶ <https://snips.ai/>, consulté le 22 avril 2019.

*« Unless we have a disaster that triggers a major shift in usage, the convenience and benefits of connectivity will continue to attract users [...] Evidence suggests that people value convenience today over possible future negative outcomes ».*⁶⁷

Des technologies axées sur la protection des données sont une partie de la solution au problème de la vie privée, car comme le formule Alessandro Acquisti (2011, p.126), « il est difficile de faire confiance aux individus pour prendre les décisions les plus conformes à leurs intérêts lorsqu'il s'agit de leur vie privée ». Selon l'auteur, ces solutions doivent cependant être combinées à d'autres mesures, telles que la sensibilisation et la mise en place de politiques régulatrices. Dans le même sens, la CNIL (2019) défend l'idée selon laquelle l'exercice des droits dont disposent les individus doit être facilitée, au même titre que le sont les expériences utilisateurs. L'information à ce sujet se doit d'être « simple, pratique et présente partout où elle fait sens dans des interfaces entre l'utilisateur et le service » (CNIL, 2019, p.41). Le règlement général sur la protection des données (RGPD) entré en vigueur en mai 2018 dans l'Union européenne vise précisément à donner aux utilisateurs les moyens d'un meilleur contrôle sur leurs données. Si ces mesures sont une première étape nécessaire pour une protection renforcée de la vie privée, elles ne paraissent pas contraindre, pour l'instant du moins, le développement d'assistants vocaux dont le fonctionnement repose sur la collecte d'informations personnelles. Comme le soulignait Sami Coll, quelques années avant l'introduction de ce règlement, « les politiques de protection des données (suivies par les entreprises ou par les gouvernements) semblent avoir pour objectif de rassurer les gens (et/ou le législateur) face à la propagation de la société de l'information, désormais au cœur du capitalisme moderne, sans trop porter d'entraves au marché » (Coll, 2014, p.62). L'équilibre démocratique est fragilisé par un système qui permet aux exploitants de données de recueillir toujours davantage d'informations, tout en gardant opaques leurs processus et fonctionnements. Le rapport de force qui s'instaure entre observateurs et observés

⁶⁷ Rainie, L. & Anderson, J. (2017). The Internet Of Things Connectivity Binge: What Are The Implications?. *Pew Research Center*, 6 juin 2017. Disponible sur : <https://www.pewinternet.org/2017/06/06/the-internet-of-things-connectivity-binge-what-are-the-implications/>, consulté le 22 avril 2019.

est d'autant plus asymétrique que la capacité de surveiller les autres sans faire soi-même l'objet d'une surveillance constitue l'une des plus importantes formes de pouvoir (Pasquale, 2015, p.3), désormais entre les mains de quelques-uns.

5. Des usages imaginés

Comme l'ont souligné les travaux menés dans le champ de la sociologie des usages, étudier la technique implique également de décrire « ce que les gens font effectivement avec des objets techniques » (Proulx, 2015, p.1). Née en France au début des années 1980, cette approche s'appuie notamment sur les réflexions de Michel de Certeau (1980) et l'étude des « manières de faire », soit la façon dont les individus adaptent et détournent des objets ordinaires pour leur assigner des usages auxquels ils n'étaient pas destinés. Ces travaux ont ainsi montré l'importance et la complexité de l'appropriation technique. Dans cette perspective, une recherche qui porte sur le dispositif que sont les enceintes connectées devrait nécessairement se pencher sur la façon dont les utilisateurs interagissent avec cet objet. Toutefois, comme évoqué en introduction, cette technologie a ceci de particulier que, malgré son importante visibilité médiatique, son usage reste limité. En Suisse, la technologie n'est pas encore pleinement disponible. Si les enceintes connectées sont déjà en vente, de nombreuses fonctionnalités ne sont pas compatibles avec ce marché, ce qui réduit drastiquement les possibilités d'usage. Une étude ethnographique dans un contexte local, telle qu'elle aurait pu être réalisée dans ce cadre, n'aurait donc offert qu'une vision très partielle de « l'objet dans l'action » (Akrich, 1993).

Cependant, renoncer à l'observation des usages ne signifie pas pour autant les délaisser. En effet, lors de la conception d'un objet, des usages et un contexte d'interaction sont pensés, comme le précise Madeleine Akrich (1993, p.4) :

« [...] les innovateurs eux-mêmes ont une définition "riche" de l'action qui ne peut être réduite à cette abstraction que constituerait le face-à-face de l'objet et de son utilisateur détaché de tout contexte : l'environnement dans lequel l'action est susceptible de prendre son sens se trouve spécifié, de même qu'un certain nombre

de compétences cognitives, psycho-motrices, mais aussi proprement sociales sont imputées aux acteurs ».

Par ailleurs, la conception d'un objet technique intègre également des discours qui portent sur les usages imaginés. Ces discours, au travers de leur dimension narrative, participent au processus d'appropriation de la technique. Pierre Musso note ainsi que « la fonctionnalité des objets techniques est indissolublement liée à une fictionnalité » (Musso, 2009, p.3). La mise en récit des techniques contribue à la création d'un imaginaire qui accompagne leur fonction utilitaire. Dans le même sens, Michel Faucheux souligne que « fabriquer un outil technique, c'est fabriquer du symbolique, raconter une histoire qui aide à donner forme et fonction à l'outil, qui situe l'acte et la production techniques dans le monde » (Faucheux, 2011, p.93). Mais les discours se construisent également lors de la diffusion de la technique. Ces formes de productions discursives sont parfois nommées « discours d'accompagnement » (Breton, 2002). Il ne s'agit pas de paroles d'experts, mais d'énoncés tenus dans l'espace public qui portent sur les usages pensés ou attendus. Ils peuvent être le fait d'acteurs très divers (journalistes, essayistes, politiciens, publicitaires, etc.) et relèvent, selon Philippe Breton (2002), de trois catégories : le discours prospectif, prescriptif et promotionnel. C'est cette dernière forme de discours qui nous intéresse ici, car elle offre une représentation de certains des usages imaginés par les concepteurs. Il s'agit d'une mise en récit composée de signes, d'images, d'histoires et de slogans qui fournissent un cadrage à l'objet technique (Musso, 2009).

Comme le soulignent Marianne Choueteau et Céline Nguyen (2011, p.47), les représentations des usages « agissent en premier lieu sur la pré-connaissance de l'objet [...]. L'appropriation n'est possible que si l'utilisateur a pris, même de manière minimale, connaissance de l'objet technique ». Le discours promotionnel est l'une des façons d'accéder aux usages que permet le dispositif technique. Il s'agit évidemment d'un discours particulier, qui a pour but d'informer, mais surtout de vendre. Lorsque le produit est relativement nouveau ou peu connu du public, l'annonceur ne peut toutefois faire l'économie d'une présentation – même sommaire – des caractéristiques de l'objet, de ses fonctionnalités et de sa manipulation. Mais contrairement au mode d'emploi, cette forme de communication vise moins à décrire un « programme d'actions » (Akrich et Boullier, 1991) qu'à proposer des imaginaires auxquels l'utilisateur puisse s'identifier.

5.1 Les nécessaires imaginaires

En France, la sociologie des techniques n'a porté qu'un intérêt limité à la question des imaginaires, contrairement à l'approche américaine pour laquelle les récits jouent un rôle-clé. Selon Antoine Picon et Patrice Flichy (2001, pp.9-10), cette dissymétrie tient à une tradition française qui, d'une part, ne propose que peu de ponts entre les différents champs disciplinaires et, d'autre part, met « systématiquement de côté tout ce qui ne tombe pas immédiatement sous le sens ou qui paraît moins solide que les dispositifs matériels et les faits socio-économiques ».

Dans le domaine des TIC, les imaginaires occupent pourtant un rôle central dans la conception et l'insertion sociale des objets techniques. Ils constituent une ressource et un cadre de référence pour l'innovation (Flichy, 2003) et imprègnent également les discours qui se construisent autour des techniques, influençant directement leur adoption et leurs usages. Certains de ces récits sont mobilisés de façon récurrente. Pierre Musso (2009, p.6) cite ainsi les « mythes dynamiques » sous-jacents à l'innovation technique décrits par Abraham Moles (1990), tels que le *mythe de l'ubiquité*, soit la capacité à être présent simultanément en plusieurs lieux (télécommunications, technologies mobiles) ; le *mythe de Babel* ou la recherche de langages ou de bibliothèques universels (moteurs de recherche) ; ou encore le *mythe de la recreation à l'identique* (dispositifs haute-fidélité ou haute définition).

Dans le même sens, le sociologue Victor Scardigli (1989) a établi une typologie composée de 7 couples de « miracles ou frayeurs » qui permettent de saisir l'imaginaire ambivalent autour de la technique » : 1) *liberté ou esclavage* : les technologies libèrent des contraintes professionnelles et domestiques en offrant autonomie et épanouissement personnel, ou elles asservissent l'individu et l'aliènent à la machine ; 2) *savoir ou inculture* : les TIC sont vecteurs « d'intelligence collective » et de connaissances nouvelles, ou elles abêtissent en se substituant à la réflexion humaine ; 3) *mémoire ou disparition* : la technologie permet d'enregistrer une vaste mémoire collective, ou ses défaillances présentent une menace pour le fonctionnement de notre société ; 4) *justice sociale ou inégalités* : les TIC permettent un accès facilité au progrès pour chacun, ou elles marginalisent les plus vulnérables ; 5) *lien social ou isolement* : la technologie accentue la fréquence des échanges et participe à la création de nouveaux liens sociaux, ou elle mène à la solitude et l'isolement ; 6) *prospérité ou précarité* : les développements technologiques créent des richesses et une « nouvelle économie », ou ils détruisent des emplois et aggravent la crise économique ; 7) *Mobilité ou*

immobilisme : les TIC accroissent la mobilité et la vitesse des déplacements, ou elles conduisent à l'immobilisme en créant de nouvelles formes de connexions entre différents lieux. Bien que cette catégorisation basée sur une dichotomie entre visions enthousiaste et pessimiste des technologies réduise certainement la complexité des imaginaires, elle permet tout de même de mettre en lumière les concepts sur lesquels s'appuient certains récits.

5.2 Vendre l'enceinte connectée

L'enceinte connectée est un objet relativement nouveau dont les fonctionnalités demeurent peu connues du public. Au cours de ces dernières années, les fabricants se sont donc engagés dans une communication visant à faire connaître cette technologie et à séduire de nouveaux utilisateurs. Bien que les différents concepteurs proposent des dispositifs relativement similaires, les stratégies et objectifs peuvent varier selon le modèle économique et la culture propre à chaque entreprise.

Amazon cherche avant tout à équiper les foyers d'un vaste réseau d'objets pilotés par la voix dont l'enceinte connectée constituerait le dispositif central. En septembre 2018, l'entreprise a annoncé le lancement d'une nouvelle série d'objets connectés.⁶⁸ La marque propose, entre autres, un micro-ondes, une horloge, une prise « *smart plug* », une caméra de surveillance ou encore un système de sécurité basé sur la détection de mouvements. Toujours dans la perspective de s'imposer sur le marché de la *smart home*, Amazon s'est récemment lancé dans le développement immobilier en établissant un partenariat avec Lennar, la plus grande société de construction de résidences privées aux États-Unis. Les habitations proposées sont équipées de différents systèmes connectés reliés au logiciel Alexa (enceintes, serrures, sonnettes, lumières, thermostats, etc.).⁶⁹ Au travers du développement des interfaces vocales, Amazon tente également d'établir un lien plus direct entre le consommateur et sa plateforme d'achat, même si cette stratégie semble pour l'instant peu convaincante.⁷⁰

⁶⁸ Welsch, C. (2018). The 14 biggest announcements from Amazon's surprise hardware event. *The Verge*, 20 septembre 2018. Disponible sur : <https://www.theverge.com/2018/9/20/17883242/amazon-alexa-event-2018-news-recap-echo-auto-dot-sub-link-auto-microwave>, consulté le 10 mai 2019.

⁶⁹ <https://www.lennar.com/wifi-certified>, consulté le 10 mai 2019.

⁷⁰ Velasco, C. (2018). It Seems Practically No One Is Using Alexa To Buy Things Off Amazon. *Tech Times*, 7 août 2018. Disponible sur:

Google, dont la principale activité demeure la vente de publicité ciblée, entend bénéficier des possibilités renouvelées de collecte de données qu'offrent les assistants vocaux. Si pour l'instant, les informations obtenues par Google Home ne permettent pas de transmettre du contenu publicitaire directement par ce biais, l'entreprise semble bien travailler en ce sens. En mars 2017, Google a tenté de diffuser une première publicité pour le film de Walt Disney « La Belle et la Bête » au travers de l'application My Day (un calendrier) de Google Home. Lorsqu'un utilisateur interrogeait son assistant quant au programme de sa journée, celui-ci indiquait au passage que le film était sorti en salles. Interviewé par le magazine *The Verge*, Google s'est défendu de diffuser de la publicité en expliquant qu'il s'agissait d'une fonctionnalité intégrée au calendrier qui permet de « suggérer à l'utilisateur certains contenus jugés opportuns ».71 Si la nuance semble subtile, ces propos s'inscrivent dans une stratégie plus globale qui vise à intégrer toujours plus finement du contenu publicitaire au sein d'autres activités. Bien que l'expérience – peu concluante – n'ait pas été renouvelée, elle témoigne toutefois de la volonté de Google de monétiser ces nouveaux espaces. Il s'agit désormais pour l'entreprise de trouver une façon pour que ces publicités s'insèrent de façon discrète, sans compromettre la promesse d'une interaction « fluide ».

Enfin, Apple reste centré sur le développement de dispositifs techniques haut de gamme, et fait ainsi davantage valoir la qualité sonore de son HomePod que sa fonction d'assistant vocal.72 Le logiciel Siri, bien qu'intégré à cette enceinte, reste encore principalement lié aux fonctionnalités de l'iPhone. Toutefois, Apple investit également dans le secteur de la *smart home* au travers de son système HomeKit. Contrairement à Amazon, il ne s'agit pas d'élargir la gamme de produits, mais de proposer un système compatible avec une sélection d'objets issus d'autres marques. Le déploiement d'Apple dans ce domaine se fait de façon relativement discrète et maîtrisée, afin de ne pas compromettre la sécurité et la fiabilité des systèmes proposés.

Malgré des objectifs différents, certains imaginaires apparaissent de façon récurrente dans la communication autour des enceintes connectées. En ce sens, les discours publicitaires constituent un matériel intéressant, car il s'agit de mises en récit

<https://www.techtimes.com/articles/233017/20180807/it-seems-practically-no-one-is-using-alexa-to-buy-things-off-amazon.htm>, consulté le 2 mai 2019.

⁷¹Welch, C. Google Home is playing audio ads for Beauty and the Beast. *The Verge*, 16 mai 2017. Disponible sur : <https://www.theverge.com/circuitbreaker/2017/3/16/14948696/google-home-assistant-advertising-beauty-and-the-beast>, consulté le 23 mai 2019.

⁷²<https://www.apple.com/fr/homepod/>, consulté le 2 mai 2019.

qui comprennent un contexte d'interaction, des usages et des promesses qui font appel à différents univers symboliques. Lors de la diffusion d'une nouvelle technologie, l'absence d'usage préalable et la méconnaissance de ses caractéristiques laissent une marge de manœuvre importante au discours des fabricants. L'analyse de celui-ci semble alors importante, car il est à même d'influencer l'appropriation et les usages futurs du dispositif. Comme le soulignent Philippe Mallein et Yves Toussaint (1992, p.225), « d'une part il [le discours] reflète les attentes, voire les fantasmes, des concepteurs, d'autre part il reflète le cadre général dans lequel ceux-ci aimeraient voir prendre place la technique ».

Pour ce travail, neuf publicités issues de trois marques (Amazon, Google et Apple) et destinées à des marchés différents (France, États-Unis, Canada, Grande-Bretagne) ont été sélectionnées.⁷³ Le choix du corpus s'est porté sur des films introductifs qui présentent généralement le dispositif ou certaines de ses fonctionnalités pour la première fois. Plutôt que de s'engager dans une description fine de chaque séquence, il s'agit avant tout de comprendre, de manière plus transversale, la façon dont la relation entre technique et social est pensée au travers de la représentation des usages et des promesses. Cinq thématiques centrales ont été identifiées.

1. *Multitasking*

Sur les neuf films sélectionnés, quatre mettent particulièrement en avant la possibilité d'utiliser l'enceinte connectée tout en réalisant d'autres activités. Tout d'abord, les publicités pour Amazon Echo (Grande-Bretagne et France) montrent un utilisateur qui, tout en cuisinant, interagit avec l'enceinte connectée pour lancer une musique, activer un minuteur, obtenir des informations sur une recette ou encore ajouter des articles à une liste de courses. Ensuite, la campagne destinée au Canada met en scène un père de famille qui s'empresse de ranger le salon et de mettre les enfants au lit lorsque leur mère s'apprête à rentrer. Alors que tout le monde s'affaire, l'enceinte permet en même temps de baisser la lumière, d'éteindre la musique et de jouer une comptine aux enfants. Enfin, le spot réalisé pour Google Home (États-Unis) n'est pas uniquement focalisé sur le *multitasking* mais propose deux scènes dans lesquelles l'interaction par la voix permet la réalisation de plusieurs activités de manière

⁷³ Voir corpus détaillé en annexe.

simultanée : augmenter le volume de la musique alors que l'on prépare des boissons, lancer une requête concernant un ingrédient tout en cuisinant. Les actions peuvent non seulement s'enchaîner de manière fluide, mais également se superposer. La cadence des activités est alors accélérée par la machine. Si ces représentations font écho au contexte de pression temporelle analysé au chapitre 1, c'est ici davantage la dimension pratique d'une fonction « main libre » qui est soulignée, que le réel gain de temps permis par ce mode de fonctionnement.

2. Renforcement des liens sociaux

Comme de nombreuses technologies (téléphone, *smartphone*, réseaux sociaux), l'enceinte connectée est présentée comme un moyen de construire du lien entre les individus. Si cet argument est présent dans l'ensemble du corpus (à l'exception de la publicité pour le Home Pod), il constitue le propos central des deux vidéos pour l'enceinte Amazon Echo destinées au marché français. La première met en scène une jeune femme qui ne parvient pas à annoncer son homosexualité à ses parents. En découvrant l'enceinte connectée nouvellement installée dans le salon familial, elle décide alors de leur transmettre son message au travers d'une chanson aux paroles évocatrices.⁷⁴ La seconde publicité, dans un registre similaire, montre la façon dont une mère parvient, également grâce à la musique diffusée par Amazon Echo, à rétablir le dialogue avec son fils adolescent.⁷⁵ L'enceinte apparaît dans ces deux séquences comme le médiateur des relations entre parents et enfants. La technique ne permet alors plus uniquement une co-présence à distance, elle intervient lors de relations en face-à-face en créant de nouvelles formes et potentialités d'interactions sociales. L'enceinte se présente comme un médiateur de ces relations et la technique devient « organisateur de l'action » (Jouët, 1993).

Le spot pour Google Home (Grande-Bretagne) met également en avant le principe d'un renforcement des liens familiaux grâce à l'intermédiation de l'enceinte connectée. L'une des séquences présente un père qui, absorbé par sa tablette, ne prête pas attention à ses deux filles qui souhaiteraient jouer avec lui. Soudainement conscient de passer à côté d'un moment important, il dépose son écran, mais demeure hésitant. Une voix off lui suggère alors discrètement la solution : « *If only there was something that could keep us in the moment. Oh ! Hello Google Home* ». En apercevant l'enceinte, il

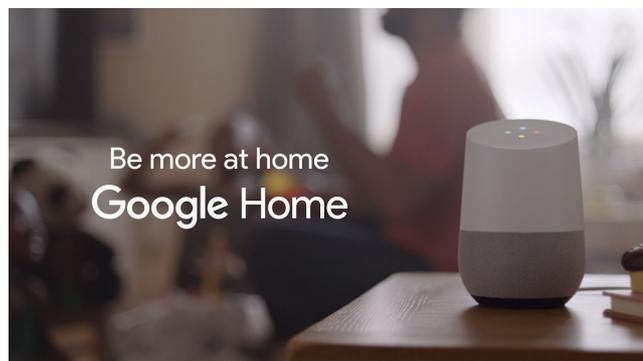
⁷⁴ Voir annexe, publicité n°3.

⁷⁵ Voir annexe, publicité n°4.

décide de lancer une musique et participe au jeu de ses enfants. Cette scène entre en résonnance avec la question de la captation attentionnelle évoquée au chapitre 4. L'enceinte apparaît ici comme la solution à une forme d'isolement et d'appauvrissement des relations humaines générés par les écrans. Si la technique est souvent présentée au travers de l'argumentaire de la sociabilité, cette séquence a de particulier qu'elle dénonce en même temps les effets de technologies encore largement présentes et promues par les mêmes acteurs. Dépourvue d'écran, l'enceinte connectée permettrait de profiter davantage des relations et interactions humaines qui prennent place au sein du foyer. C'est le message que transmet Google au travers de son slogan « *Be more at home* » (Fig. 3) décliné en une série de spots publicitaires.⁷⁶ L'enceinte répondrait au problème de connexion permanente en offrant une alternative, une forme de solution intermédiaire qui permet de se distancier physiquement des dispositifs numériques tout en gardant leurs services à proximité immédiate.

Figure 3.

Spot publicitaire pour Google Home (Grande-Bretagne)



Source : <https://www.youtube.com/watch?v=rgWQbiiLiLI>

⁷⁶ <https://www.72andsunny.com/work/google-homebemoreathome>, consulté le 3 mai 2019.

3. Vecteur d'information et de connaissance

Comme l'a souligné Victor Scardigli (1989), l'une des promesses souvent associées à la technique est celle d'un accès facilité au savoir. L'enceinte connectée, dont le fonctionnement dépend directement de sa capacité à rechercher et fournir une information pertinente n'échappe pas à ce constat. Si Amazon propose également une fonctionnalité de recherche, Google insiste particulièrement sur la pertinence des résultats que peut offrir son enceinte. La publicité montre ainsi la capacité de Google Home à traduire instantanément – de l'anglais vers l'espagnol ou le français – des phrases prononcées par l'utilisateur. Plusieurs scènes mettent également en avant l'aptitude de l'enceinte à répondre à une diversité de requêtes formulées. Il peut s'agir de questionnements généraux (Combien d'étoiles compte notre galaxie ?) ou de demandes directement en lien avec l'utilisateur (Quel est l'état du trafic entre notre maison et l'aéroport ?). Google a par ailleurs réalisé pour le marché britannique une série de courtes vidéos uniquement dédiées à la fonction de recherche.⁷⁷ Comme discuté au chapitre 4, le principe est celui d'offrir à l'utilisateur une réponse avant tout simple et rapide. Les questions choisies pour illustrer cet usage ne requièrent pas d'explications complexes. Il s'agit de requêtes qui portent sur des dates (Quel est le jour de l'anniversaire du prince William ?), des trajets (Quelle est la distance entre Leeds et Glastonbury ?) ou des définitions sans ambiguïté (Qu'est-ce qu'un *let* en tennis ?). De la même manière, les traductions proposées dans les différentes publicités dépassent rarement le cadre des salutations d'usage. Toutefois, en multipliant les exemples de requêtes variées, ces spots publicitaires transmettent l'image d'un dispositif dont la connaissance semble à la fois illimitée et indiscutable. L'accumulation des traces et leur mise en réseau suffiraient à produire une technologie omnisciente qui énonce des vérités sur le passé, le présent ou l'avenir de façon définitive. La contingence, le hasard et l'incertitude ont laissé place à un monde prévisible, et donc rassurant.

4. Pilotage et gestion du foyer

De par ses possibilités de connexion à divers objets domestiques, l'enceinte connectée est présentée comme un dispositif permettant d'optimiser la gestion et le

⁷⁷ <https://www.youtube.com/playlist?list=PLWLk3nKIQFCXrAYiOP36lbA9oLLP-aqrg>, consulté le 6 mai 2019.

« pilotage » du foyer. Elle s'inscrit alors dans la thématique de la maison « intelligente » (Harper, 2003) qui libère l'utilisateur de toute une série de petits gestes quotidiens, désormais remplacés par une simple commande traitée par un système centralisé. Les publicités pour Google Home montrent ainsi qu'il est possible, grâce à la voix, d'allumer et régler l'intensité de l'éclairage d'une pièce, de lancer un programme TV, d'activer un minuteur ou de contrôler la musique diffusée dans l'appartement. Amazon propose, en outre, de gérer les provisions du foyer en permettant à l'utilisateur d'ajouter des produits à une liste de courses. L'enceinte devient le centre de contrôle de l'ensemble de l'infrastructure nécessaire au fonctionnement du foyer.

En 1989, Victor Scardigli imaginait déjà une maison du futur pilotée par un « ordinateur domestique » qui gère tout : cuisson, réglages de l'énergie, sécurité, réapprovisionnement, rendez-vous... On voit à quel point cette prévision décrit avec justesse les services aujourd'hui proposés par l'enceinte connectée. Bien que les publicités analysées s'en tiennent généralement à la démonstration d'usages relativement basiques, qui n'incluent pas une multitude d'autres dispositifs, la mise sur le marché d'un nombre croissant d'objets connectés semble pourtant aller dans le sens d'un réseau domestique toujours plus étendu. Tournée vers l'efficacité et la rationalisation, la maison se rapproche alors de ce que Le Corbusier nommait « la machine à habiter », soit une habitation fonctionnelle qui programmerait nos conduites de façon efficace et neutraliserait l'imprévisible, tout en nous épargnant pénibilité et fatigue (Sauvage, 2007).

5. Intégration au quotidien

Dans une analyse du discours médiatique et publicitaire qui accompagne le développement de l'informatique dans les années 1980, Stahl (1995) a montré la prégnance de la référence à la magie pour présenter une technique dont le fonctionnement échappe au plus grand nombre. Les publicités pour les enceintes connectées semblent au contraire insister sur la dimension ordinaire et routinière d'une technologie qui s'insère discrètement dans les habitudes du foyer. Plutôt qu'une démonstration de fonctionnalités exceptionnelles, la communication met en lumière les capacités d'adaptation du dispositif à une variété d'expériences quotidiennes. Celles-ci ne sont pas radicalement transformées, elles sont simplement améliorées, fluidifiées ou « augmentées » par la technologie.

Alors qu'aucun quotidien ne se ressemble, l'enceinte connectée semble pouvoir proposer une solution qui convienne à une diversité de situations. Les publicités pour Google Home montrent – parfois successivement – un environnement urbain, une maison de banlieue, des personnes d'âges et d'origines variées (afro-américaines, asiatiques, hispaniques), différentes compositions familiales (couple homosexuel, famille monoparentale) et des expériences et situations diverses (réveil, anniversaires, repas, jeu, etc.). La technologie est avant tout un vecteur d'expression des individualités, comme le souligne le slogan « Home by you. Help by Google ».78 Google ne fait que mettre à disposition un service, qui est ensuite investi par chacun. Comme d'autres entreprises, elle propose « une offre standardisée à partir de laquelle peuvent s'exprimer les expériences singulières » (Beaude, 2019), afin de capter toujours plus finement les pratiques individuelles. La double signification du pronom « you » peut par ailleurs laisser entendre que l'enceinte accompagne à la fois des expériences individuelles et collectives. Installée dans un espace commun, l'enceinte est destinée à interagir avec différents utilisateurs.79 Elle distingue leur voix, connaît leurs préférences et s'adaptent à leurs besoins. La technologie semble si bien s'ajuster au quotidien, créant un enchaînement fluide et continu d'actions, que l'on en vient à oublier sa dimension à la fois prescriptive et intrusive.

Apple, une stratégie distincte

Si la publicité d'Apple semble largement absente de ces thématiques transversales, c'est que le spot destiné à promouvoir le HomePod ne ressemble en rien aux récits proposés par les autres fabricants. Plus proche du clip vidéo, voire du court-métrage, le film de 4 minutes réalisé par Spike Jonze ne fait que discrètement allusion à son objectif publicitaire. Mettant en scène la musicienne et danseuse FKA Twig, le film débute sans musique dans un espace urbain sombre et surchargé. De retour du travail, une jeune femme visiblement fatiguée par une longue journée demande à son HomePod de jouer une musique qu'elle aime. Alors qu'elle commence à danser, l'appartement sombre et vétuste se transforme en un espace surréel, graphique et coloré. Murs et meubles se déplacent, s'agrandissent, suivant les mouvements effectués

⁷⁸ Voir annexe, publicité n°6.

⁷⁹ Google Home propose aujourd'hui d'associer jusqu'à 6 voix différentes : <https://www.blog.google/products/assistant/tomato-tomahto-google-home-now-supports-multiple-users/>, consulté le 11 mai 2019.

par le personnage. S'en suit une longue performance de danse qui s'achève par un retour à l'état initial de l'appartement. La publicité se termine par un plan sur l'enceinte, suivi d'une simple mention « HomePod » et du logo de la marque. Au-delà des différences de forme, Apple se distingue de ses concurrents en axant sa communication sur une seule promesse : le HomePod propose une qualité sonore extraordinaire. L'ensemble du film est construit autour des effets produits par un dispositif offrant un son optimal. Le récit proposé est celui d'une société morose et oppressante à laquelle il serait possible d'échapper grâce à l'imaginaire véhiculé par la musique. L'enceinte n'apparaît qu'un court instant et se fait oublier au profit du son qui submerge et transforme l'espace. Apple montre ainsi qu'elle souhaite remettre la musique et le son au cœur de son activité. Toutefois, ce film ne peut à lui seul résumer l'ensemble de la stratégie de la marque autour des enceintes connectées. A l'instar de ses concurrents, Apple propose également d'autres types de communications plus classiques qui mettent davantage en lumière le dispositif et ses fonctionnalités.

Les trois principaux fabricants d'enceintes connectées font donc appel à des récits différents pour séduire de nouveaux utilisateurs. Amazon et Google proposent une expérience relativement similaire, qui souligne la dimension utilitaire de l'assistant vocal. Celui-ci permet de gérer les différents objets du foyer, d'obtenir des renseignements simples et pratiques, de réaliser plusieurs tâches de manière simultanée. Cet argument fait écho à ce que Victor Scardigli (1989) nomme « le pouvoir sur les contraintes », dont serait dotée la technique. L'une des promesses récurrentes est en effet celle d'une libération des obligations domestiques grâce au progrès technologique. Google insiste toutefois davantage sur la capacité de son enceinte à fournir une réponse pertinente à tout type de requête. L'entreprise met en avant un véritable guide « omniscient », qui offre traductions, recherche d'informations, d'itinéraires, de musiques, etc. La technique est donc présentée comme une manière d'accéder au savoir. Elle se substitue à une mémoire humaine parfois défaillante et vient pallier des capacités cognitives limitées. Elle permet une créativité et une imagination renouvelées. Mais au-delà d'un assistant personnel individuel, l'usage de la technologie s'accompagne également de la promesse de liens sociaux renoués. Les publicités d'Amazon et Google mettent en scène une enceinte connectée qui facilite le dialogue et l'interaction humaine. La technologie, installée au cœur du foyer, est également destinée à être partagée. Elle simplifie les relations, les rend plus « authentiques ». Seule Apple propose un récit relativement différent en s'éloignant des fonctions utilitaires ou sociales de l'assistant vocal. Il ne s'agit plus de

gérer des tâches quotidiennes, ni d'accéder à un savoir ou de renforcer des liens sociaux. L'enceinte n'apparaît pas comme une technologie destinée à faciliter ou adoucir le quotidien, mais à s'en extraire au travers d'une expérience « extraordinaire » vécue de façon individuelle.

En conclusion, nous soulignerons encore que si les outils semblent parfois manquer pour analyser les usages imaginés des enceintes connectées, c'est que technique et récit sont deux objets que la sociologie ne confronte que rarement (Nguyen, 2013). Pourtant, le récit est étroitement lié au monde technique, la communication n'étant que l'une des manifestations de cette relation. L'analyse de ces discours rappelle que la mise en lumière des récits permet de mieux saisir la façon dont l'utilisateur, le contexte d'interaction et les bénéfices des technologies sont pensés par les concepteurs et communiqués au public. Une future observation des usages réels confrontera nécessairement ces récits. Il convient donc de porter un regard attentif sur ces premières représentations.

6. Conclusions

En partant d'un objet technique encore peu étudié – l'enceinte connectée – ce travail avait pour but d'explorer différents enjeux que posent les technologies numériques aujourd'hui. L'attention s'est portée sur quatre points centraux : le contexte de conception, l'interface vocale, la personnalisation des services et les usages imaginés. Plutôt que de s'arrêter uniquement sur sa caractéristique *a priori* centrale – la voix – il s'agissait d'aborder l'objet dans son ensemble, afin d'en repérer les implications à divers niveaux. Ce choix conduit nécessairement à aborder des thématiques et des échelles d'analyse qui peuvent sembler relativement éloignées. Le défi est alors de lier de manière cohérente et pertinente des sujets aussi divers que l'accélération des rythmes de vie et la culture de l'efficacité, l'évolution des interfaces homme-machine, l'économie et les technologies basées sur la collecte de données, l'accès à l'information, la vie privée, ou encore les promesses des fabricants.

L'objectif était précisément de montrer que ces différentes thématiques ne sont pas cloisonnées et indépendantes. Elles se nourrissent mutuellement et leur articulation témoigne des logiques économiques, techniques et sociales qui soutiennent les dispositifs numériques actuels. Ainsi, si l'enceinte connectée est souvent présentée comme une réponse à l'accélération imposée par une société qui valorise la

productivité et l'efficacité dans toutes les sphères sociales (jusqu'à l'intimité du foyer), il est toutefois impossible de détacher son développement d'un contexte dans lequel la collecte de données personnelles est perçue comme nécessaire au développement économique et technique de nombreux services numériques. Ces derniers s'appuient sur la traçabilité toujours plus importante de nos pratiques pour proposer des interactions personnalisées, rendues indispensables dans un monde caractérisé par une abondance d'informations et une multiplication des possibilités d'action. Sans personnalisation, le travail de sélection et filtrage incomberait à l'utilisateur, rendant d'autant plus difficile la réalisation des objectifs de rentabilisation et de performance auxquels il est soumis.

Mais cette personnalisation implique également une nouvelle « gouvernance algorithmique » qui vise à modéliser, prévoir et guider nos comportements (Rouvroy et Berns, 2013). Cette forme de contrôle ne relève pas d'un accord, mais d'une adhésion tacite des individus. La prise de distance et la réflexivité sont d'autant plus délicates à mettre en œuvre que les processus sont automatisés et directement intégrés à un ensemble toujours plus vaste de pratiques. Questionner les résultats fournis – sous forme d'information, de « savoir » ou de suggestion – implique un certain coût cognitif et temporel. En proposant une action ajustée en « temps réel », la « gouvernance algorithmique » répond aux impératifs actuels de vitesse et de réactivité. Par ailleurs, le principe de « boucle de rétroaction » sur lequel reposent de nombreuses technologies numériques implique que leurs performances évoluent à mesure que l'on en fait usage. L'utilisateur est alors pris dans un système auquel il apparaît toujours plus difficile d'échapper.

De nombreux individus se disent inquiets face à cette nouvelle forme de surveillance généralisée de nos pratiques. Cependant, la rapidité et le confort que permettent ces dispositifs l'emportent le plus souvent sur les préoccupations liées au respect de la vie privée. L'argument selon lequel « qui n'a rien à cacher n'a rien à craindre » vient encore appuyer les nombreux biais cognitifs qui empêchent les individus d'agir selon leurs opinions déclarées. Dans le cas des enceintes connectées, la focalisation du débat (bien que légitime) sur les possibilités d'écoute et d'intrusion dans un espace privé, occulte d'autres risques, en apparence plus inoffensifs. Ils concernent un profilage toujours plus fin et discret des comportements. Cette traçabilité permanente opère un réagencement continu de notre environnement (Beaude, 2015) qui transforme considérablement nos relations, nos connaissances et nos possibilités d'action (Beaude, 2019).

6.1 Limites et perspectives de recherche

Au terme de ce travail, il semble nécessaire de souligner certaines de ses limites afin d'envisager de nouvelles pistes de recherche. Comme nous l'avons déjà évoqué, l'un des inconvénients liés au choix d'analyser un objet encore peu adopté, réside dans l'impossibilité d'en observer les usages. Ces derniers sont pourtant particulièrement importants, car ils permettent d'aller au-delà des discours et de saisir la façon dont les individus perçoivent la technique et interagissent avec elle. Comme le relève Madeleine Akrich (1993, p.3), il convient d'observer la transition entre « un espace de circulation économique, dans lequel une certaine description de l'objet sera prédominante, et un espace privé ou domestique dans lequel l'objet sera défini d'une manière différente de la précédente ». Une future recherche sur les enceintes connectées devra donc nécessairement intégrer les utilisateurs et leurs pratiques. L'analyse des usages est d'autant plus importante que ces derniers influencent directement le fonctionnement des enceintes. Contrairement à d'autres dispositifs techniques (téléphone, radio, télévision), le comportement de l'utilisateur est ici à même d'agir de façon immédiate sur l'ensemble du système, dont le « script » n'est que partiellement écrit.

Ensuite, la décision d'aborder un objet dont le fonctionnement repose sur des processus opaques ne permet pas, dans le cadre de ce travail, d'accéder à une analyse profonde des logiques de pouvoir inscrites dans la technique. Afin de contourner cet obstacle, une solution consisterait à se pencher plus attentivement sur la façon dont les algorithmes sont élaborés. C'est ce que proposent Cécile Méadel et Guillaume Sire (2017) au travers d'une « sociologie orientées programmes ». Selon les auteurs, il s'agit de dépasser les approches qui se contentent de « tourner autour du programme, se plaçant exclusivement du côté de ses utilisateurs ou ne s'intéressant qu'au travail de ses concepteurs, sans jamais l'observer directement » (Méadel & Sire, 2017, p.15). Toutefois, contrairement à un programme informatique qui repose sur un seul code identifiable, le fonctionnement de l'enceinte connectée fait appel à une multitude d'algorithmes, dont l'agencement demeure difficilement « décodable ». Cette démarche impliquerait alors de se concentrer sur certaines fonctions et usages précis, au risque de réduire l'analyse du dispositif à quelques-unes de ses caractéristiques.

6.2 *Quels futurs pour les interfaces vocales ?*

Si les grands acteurs du numérique prédisent avec assurance un futur piloté par la voix, il convient tout de même – en conclusion de ce travail – d’interroger l’avenir des enceintes connectées. Comme le relève Nicolas Nova (2011, p.19) dans une analyse des « flops » technologiques : « La confiance absolue dans la pertinence d’une innovation est aussi un aspect commun à la plupart des échecs technologiques ». De la même façon, l’emballement médiatique serait « un trait majeur des échecs récurrents ». Dès lors, comment juger de l’adoption des enceintes connectées ? Les chiffres de vente permettent-ils de qualifier la technologie de succès ? En regard du *smartphone*, ces résultats apparaissent modestes. Malgré une diffusion plus importante aux États-Unis, les enceintes connectées ne semblent pas avoir encore massivement séduit les foyers européens. Par ailleurs, ces chiffres ne disent rien de l’usage qu’en font les individus. S’agit-il d’une utilisation régulière ou occasionnelle ? Ses potentialités sont-elles pleinement exploitées ou son emploi se limite-t-il à quelques fonctionnalités basiques ? Ces questions ne peuvent être éclaircies qu’au travers d’une enquête qui porte précisément sur les usages situés de ces technologies.

Si les enceintes connectées ont dépassé l’étape de la conception et représentent aujourd’hui un réel secteur commercial dans lequel investissent les plus grandes entreprises du numérique, il demeure difficile de déterminer si leur usage se stabilisera à long terme, au même titre que le *smartphone*. Frédéric Fréry (2000) note que certaines technologies restent « éternellement émergentes » malgré des prévisions de croissance favorables, l’anticipation d’une forte demande et la mobilisation de nombreux acteurs. La domotique est précisément un domaine où les promesses sans cesse renouvelées n’ont jamais réellement abouti. La « maison intelligente » annoncée depuis les années 1980, ne s’est toujours pas concrétisée. Des innovations isolées, telles que le frigo connecté, n’ont connu aucun réel succès commercial. Toutefois, cette situation pourrait évoluer grâce à la participation de grandes entreprises dont la notoriété est déjà acquise. IKEA tente aujourd’hui d’investir le terrain de la maison connectée, en créant une gamme de produits compatibles avec les assistants vocaux d’Apple, Amazon et Google.⁸⁰ L’entreprise a également collaboré avec la marque

⁸⁰ Charara, S. (2019). IKEA has an ingenious plan to dominate the smart home market. *Wired UK*, 11 janvier 2019. Disponible sur : <https://www.wired.co.uk/article/why-ikeas-perfectly-positioned-to-own-the-smart-home>, consulté le 12 mai 2019.

Sonos afin de développer une enceinte intégrée à une lampe.⁸¹ En se concentrant sur des usages limités mais éprouvés, IKEA pourrait contribuer à démocratiser et légitimer la *smart home* auprès d'un large public.

Alors que la voix paraît investir un nombre toujours plus important d'objets, on peut également questionner l'avenir des interfaces visuelles. Les divers spots publicitaires pour les enceintes connectées présentent un monde où la voix seule permet d'effectuer de nombreuses tâches qui nécessitaient auparavant la médiation d'un écran. Pourtant, la disparition définitive des interfaces graphiques apparaît peu probable et les concepteurs semblent évoluer dans ce sens. Au cours de ces derniers mois, plusieurs marques ont annoncé le lancement de dispositifs qui combinent interfaces vocale et visuelle. Il s'agit généralement d'une enceinte connectée à laquelle s'ajoute un écran de la taille d'une tablette. Amazon (Echo Show, Echo Spot), Google (Home Hub) et Facebook (Portal) proposent chacun leur modèle de *smart display*. A la différence d'une tablette classique, l'écran sert avant tout de moniteur de contrôle et permet d'obtenir un support visuel lorsque nécessaire. La consultation d'un itinéraire ou des prévisions météo hebdomadaires, par exemple, reste difficilement transmissible par la voix.

Le développement des interfaces vocales semble donc aller dans le sens d'une multiplication de supports et d'une hybridation des modes d'interaction. L'enceinte connectée, telle qu'envisagée aujourd'hui, n'est peut-être qu'une étape vers une interaction vocale plus ponctuelle et rattachée à des pratiques pour lesquelles elle fait pleinement sens. Pour l'instant, bien que déjà commercialisé, le dispositif semble encore relever de la « promesse technologique » (Audétat, 2015). Seules de futures recherches pourront dire si certains usages auront su s'imposer, ou si les enceintes connectées s'inscriront plus largement dans ce que John Urry nomme un « *'failed' future* », soit l'anticipation d'un changement profond qui n'aura jamais su se concrétiser (Urry, 2016).

⁸¹ <https://www.sonos.com/fr-ch/symfonisk-by-sonos-and-ikea> consulté le 18 mai 2019.

Références

- Acquisti, A. (2011). Les comportements de vie privée face au commerce électronique : Une économie de la gratification immédiate. *Réseaux*, 167(3), 105-130.
- Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science*, 347(6221), 509-514.
- Airenti, G. (2012). Aux origines de l'anthropomorphisme : Intersubjectivité et théorie de l'esprit. *Gradhiva*, 15(1), 34-53.
- Akrich, M. (1987). Comment décrire les objets techniques?. *Techniques et culture*, (9), 49-64.
- Akrich, M. & Boullier, D. (1991). Le mode d'emploi : genèse, forme et usage. In D. Chevallier (Eds.). *Savoir-faire et pouvoir transmettre* (pp. 113-131). Paris : Éditions de la Maison des Sciences de l'Homme.
- Akrich, M. (1993). Les objets techniques et leurs utilisateurs, de la conception à l'action. *Raisons pratiques*, 4, 35-57.
- Akrich M., Callon M., Latour B. (2006). *Sociologie de la Traduction : textes fondateurs*. Paris : Presses de l'École des Mines.
- Allen, D. (2015 [2001]). *Getting things done : The art of stress-free productivity*. Londres : Penguin.
- Alexandre, O. (2018). Burning Man : l'esprit de la Silicon Valley : Un festival libertaire devenu libéral. *Revue du Crieur*, 11(3), 86-107.
- Aubert, N. (2003). *Le culte de l'urgence, la société malade du temps*. Paris : Flammarion.
- Audétat, M., Eds. (2015). *Sciences et technologies émergentes : pourquoi tant de promesses?*. Paris : Hermann.
- Barley, S. R., Meyerson, D. E., & Grodal, S. (2011). E-mail as a source and symbol of stress. *Organization Science*, 22(4), 887-906.
- Beaude, B. (2012). *Internet : Changer l'espace, changer la société*. Limoges : FYP Éditions.
- Beaude, B. (2015). Spatialités algorithmiques. In M. Severo & A. Romele (Eds.), *Traces numériques et territoires* (pp. 133-160). Paris : Presse des Mines.
- Beaude, B. (2019). Sur les traces numériques de l'individu. In Y. Calbérac (Eds.), O. Lazzarotti, J. Lévy & M. Lussault, *Carte d'identités. L'espace au singulier*. Paris : Éditions Hermann (à paraître).

- Bickmore, T. W., Trinh, H., Olafsson, S., O'Leary, T. K., Asadi, R., Rickles, N. M., & Cruz, R. (2018). Patient and consumer safety risks when using conversational assistants for medical information : An observational study of Siri, Alexa, and Google Assistant. *Journal of medical Internet research*, 20(9).
- Bijker, W. E., Hughes, T. P., & Pinch, T. (1987). *The social construction of technological systems : New directions in the sociology and history of technology*. Cambridge, MA. : MIT Press.
- Borelle, C. (2018). Sortir du débat ontologique : Éléments pour une sociologie pragmatique des interactions entre humains et êtres artificiels intelligents. *Réseaux*, 212(6), 207-232
- Boullier, D. (2002). Objets communicants, avez-vous donc une âme?. *Les Cahiers du numérique*, 3(4), 45-60.
- Boullier, D. (2015). Les sciences sociales face aux traces du big data. *Revue française de science politique*, 65(5), 805-828.
- Boullier, D. (2016). *Sociologie du numérique*. Paris : Armand Colin.
- Brandeis, L., & Warren, S. (1890). The right to privacy. *Harvard Law Review*, 4(5), 193-220.
- Breton, P. (2002). Les discours d'accompagnement. *Les dossiers de l'audiovisuel. Les nouvelles technologies : quels usages, quels usagers*, 103, 5-9.
- Bridle, J. (2018). *New dark age : Technology and the end of the future*. Londres : Verso Books.
- Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1), 1-12.
- Callon, M. (1986). Éléments pour une sociologie de la traduction. *L'Année sociologique (1940/1948)*, 36, 169-208.
- Cardon, D. (2015). *A quoi rêvent les algorithmes ? Nos vies à l'heure des big data*. Paris : Seuil.
- Cardon, D. (2019). *Culture numérique*. Paris : Presses de Sciences Po.
- Carron, M. (2016). *Méthodes et outils pour définir et véhiculer une identité sonore : application au design sonore identitaire de la marque SNCF*. Thèse de doctorat. Université Paris 6.
- Casilli, A. (2018). *En attendant les robots. Enquête sur le travail du clic*. Paris : Seuil.
- Castells, M. (1998). *L'ère de l'information. La société en réseaux I*. Paris : Fayard.
- Cellard, L. & Measure, A. (2018). Le design de la transparence : une rhétorique au cœur des interfaces numériques. *Multitudes*, 73(4), 100-111.

- Certeau, M. de (1990 [1980]). *L'invention du quotidien, tome I : Arts de faire*. Paris : Gallimard, 1990.
- Chevet, C. (2017). La voix de synthèse : de la communication de masse à l'interaction homme-machine. Dialogue avec le monde. *Communication & langages*, 193(3), 63-78.
- Chouteau, M., & Nguyen, C. (2011). *Mises en récit de la technique : regards croisés*. Paris : Archives contemporaines.
- CNIL. (2019). La forme des choix. Données personnelles, design et frictions désirables. *Cahiers IP Innovation et Prospective*, janvier 2019, n° 6, 47 p.
- Coll, S. (2014). La vie privée comme outil de gouvernance : Surveiller et fidéliser le lien marchand. *Les Cahiers du numérique*, 10(1), 45-68.
- Cowan, R.S. (1983). *More Work for Mother : The Ironies of Household Technology from the Open Hearth to the Microwave*. New York : Basic Books.
- Crary, J. (2014) 24/7. *Le capitalisme à l'assaut du sommeil*. Paris : La Découverte.
- Crawford, K. & Joler, V. (2018). Anatomy of an AI System : The Amazon Echo As An Anatomical Map of Human Labor, Data and Planetary Resources. *AI Now Institute and Share Lab*.
- Di Bartolo, F. (2015). Transparence et opacité des objets communicants. In I. Saleh, & V. Carayol (Eds.), *Internet des objets : Évolutions et Innovations* (pp. 199-220). Londres : ISTE Editions.
- Dubar, C. (2011). Une critique sociale du temps au cœur des préoccupations de Temporalités. *Temporalités. Revue de sciences sociales et humaines*, (13).
- English-Lueck, J. A. (2017). *Cultures@ siliconvalley*. 2nd ed. Stanford : Stanford University Press.
- Faucheux, M. (2011). Technique, récit et médiation. In M. Chouteau & C. Nguyen (Eds). *Mises en récit de la technique : regards croisés* (pp.93-104). Paris : Archives contemporaines.
- Flichy, P. & Picon, A. (2001). Présentation. *Réseaux*, 109(5), 9-14.
- Flichy, P. (2003 [1995]). *L'innovation technique : récents développements en sciences sociales, vers une nouvelle théorie de l'innovation*. Paris : La Découverte.
- Fréry, F. (2000). Un cas d'amnésie stratégique : l'éternelle émergence de la voiture électrique. Communication à la IXème Conférence Internationale de Management Stratégique, Montpellier, 24, 25 et 26 mai 2000.

- Gibson, J. J. (1977). The theory of affordances. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing : Toward an ecological psychology* (pp. 67-82). Hillsdale, NJ : Lawrence Erlbaum Associates, Inc.
- Goldhaber, M. H. (1997). The attention economy and the net. *First Monday*, 2(4).
- Grimaud, E. (2012). Androïde cherche humain pour contact électrique. Les cinétiques de l'attachement en robotique. *Gradhiva. Revue d'anthropologie et d'histoire des arts*, (15), 76-101.
- Grimaud, E., & Vidal, D. (2012). Aux frontières de l'humain. Pour une anthropologie comparée des créatures artificielles. *Gradhiva. Revue d'anthropologie et d'histoire des arts*, (15), 4-25.
- Hann, I. H., Hui, K. L., Lee, T., & Png, I. (2002). Online information privacy : Measuring the cost-benefit trade-off. *ICIS 2002 proceedings*, 1.
- Harper, R. (2003). *Inside the Smart Home*. London : Springer.
- Jauréguiberry, F. (2014). La déconnexion aux technologies de communication. *Réseaux*, 186(4), 15-49.
- Jouët, J. (1993). Pratiques de communication et figures de la médiation. *Réseaux*. 11(60), 99-120.
- Karpf, A. (2008). *La voix : Un univers invisible*. Paris : Autrement.
- Kessous, E., Mallard, A., & Mounier, C. (2005). « A la découverte du client ». L'engagement marchand dans différents formats de la relation commerciale. *Économies et sociétés*, 39(11-12), 2067-2086.
- Klein, H. K., & Kleinman, D. L. (2002). The social construction of technology : Structural considerations. *Science, Technology, & Human Values*, 27(1), 28-52.
- Lancelot Miltgen, C. (2011). Vie privée et marketing. *Réseaux*, (3), 131-166.
- Latour B., [Johnson J.] (1988). Mixing humans and non-humans together : The sociology of a door-closed. *The Sociology of Science and Technology*, 35(3), 298-310.
- Latour, B. (1992). *Nous n'avons jamais été modernes. Essai d'anthropologie symétrique*. Paris : La Découverte.
- Lau, J., Zimmerman, B., & Schaub, F. (2018). Alexa, are you listening? : Privacy perceptions, concerns and privacy-seeking behaviors with smart speakers. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW), 102.
- Lussault, M. (2014). L'espace à toutes vitesses. *Esprit*, (12), 65-75.

- Mackenzie, D., & Wajcman, J. (Eds). (1985). *The Social Shaping of Technology*. Milton Keynes, Philadelphia : Open University Press.
- Mallein, P., & Toussaint, Y. (1992). Diffusion, médiation, usages des TIC. *Culture technique*, 24, 219-226.
- Méadel, C. & Sire, G. (2017). Les sciences sociales orientées programmes : État des lieux et perspectives. *Réseaux*, 206(6), 9-34.
- Nguyen, C. (2013). Les liens entre technique et récit. In M. Chouteau & C. Nguyen (Eds). *Mises en récit de la technique : regards croisés* (pp.93-104). Paris : Archives contemporaines.
- Mitchell, W. J., Ho, C. C., Patel, H., & MacDorman, K. F. (2011). Does social desirability bias favor humans? Explicit–implicit evaluations of synthesized speech support a new HCI model of impression management. *Computers in Human Behavior*, 27(1), 402-412.
- Moles, A. (1990). La fonction des mythes dynamiques dans la construction de l’imaginaire social. *Cahiers de l’imaginaire*, 5, 9-33.
- Molotch, H. (2016). “Just Time” and the Relativity of Speed. In J. Wajcman & N. Dodd (Eds.), *The Sociology of Speed : Digital, Organizational, and Social Temporalities* (pp. 117-130). Oxford : Oxford University Press.
- Morley, D., & Robins, K. (1995). *Spaces of Identity*, Londres : Routledge.
- Musso, P. (2009). Usages et imaginaires des TIC : la fiction des frictions. In C. Licoppe (Eds.) *L’évolution des cultures numériques, de la mutation du lien social à l’organisation du travail* (pp. 201-210) Limoges : FYP Éditions.
- Nass, C. I., & Brave, S. (2005). *Wired for speech : How voice activates and advances the human-computer relationship*. Cambridge, MA : MIT Press.
- Nass, C., & Moon, Y. (2000). Machines and mindlessness : Social responses to computers. *Journal of social issues*, 56(1), 81-103.
- Nova, N. (2011). *Les flops technologiques : comprendre les échecs pour innover*. Limoges : FYP Éditions.
- Nye, D.E. (1998). *Consuming Power: A Social History of American Energies*. Cambridge, MA. : MIT Press.
- Pariser, E. (2011). *The filter bubble : What the Internet is hiding from you*. New York : Penguin.
- Pasquale, F. (2015). *The Black Box Society*. Cambridge, MA. : Harvard University Press.

- Proulx, S. (2015). La sociologie des usages, et après ?. *Revue française des sciences de l'information et de la communication*, (6), 1-14.
- Rallet, A. & Rochelandet, F. (2011). La régulation des données personnelles face au web relationnel : une voie sans issue ?. *Réseaux*, 167(3), 17-47.
- Rochelandet, F. (2010). *Économie des données personnelles et de la vie privée*. Paris : La Découverte.
- Rosa, H. (2010). *Accélération. Une critique sociale du temps*. Paris : La Découverte.
- Rosa, H. (2012). *Aliénation et accélération : vers une théorie critique de la modernité tardive*. Paris : La Découverte.
- Rouvroy, A., & Berns, T. (2013). Gouvernamentalité algorithmique et perspectives d'émancipation. *Réseaux*, (1), 163-196.
- Santolaria, N. (2016). « Dis Siri » *Enquête sur le génie à l'intérieur du smartphone*. Paris : Anamosa.
- Sauvage, A. (2007). Raisons d'habiter. Pour une modélisation anthropologique. In T. Paquot (Eds.), *Habiter, le propre de l'humain : Villes, territoire et philosophie* (pp. 69-88). Paris : La Découverte.
- Scardigli, V. (1989). Nouvelles technologies : l'imaginaire du progrès. In A. Gras & P. Breton (Eds.), *L'imaginaire des techniques de pointe. Au doigt et à l'œil* (pp. 97-114). Paris : L'Harmattan.
- Schmidt, E., & Rosenberg, J. (2014). *How Google Works*. New York : Grand Central Publishing.
- Schwartz, B. (1974). Waiting, exchange, and power : The distribution of time in social systems. *American Journal of Sociology*, 79(4), 841-870.
- Searls, D. (2012). *The Intention Economy : When Customers Take Charge*. Boston : Harvard Business Press.
- Segall, K. (2012). *Insanely Simple : The Obsession That Drives Apple's success*. New York : Penguin Group.
- Simmel, G. (2013 [1903]). *Les grandes villes et la vie de l'esprit*. Paris : Payot
- Simon, H. (1971). Designing Organizations for an Information-Rich World. In M. Greenberger (Eds.) *Computers, Communications and the Public Interest* (pp. 38-52). Baltimore : John Hopkins Press.
- Sismondo, S. (1993). Some social constructions. *Social studies of science*, 23(3), 515-553.

- Solove, D. J. (2007). I've got nothing to hide and other misunderstandings of privacy. *San Diego Law Review*, 44, 745.
- Spiekermann, S., Grossklags, J., & Berendt, B. (Octobre 2001). E-privacy in 2nd generation E-commerce : privacy preferences versus actual behavior. In *Proceedings of the 3rd ACM conference on Electronic Commerce* (pp. 38-47). ACM.
- Stahl, W. A. (1995). Venerating the black box : Magic in media discourse on technology. *Science, Technology, & Human Values*, 20(2), 234-258.
- Szollos, A. (2009). Toward a psychology of chronic time pressure : Conceptual and methodological review. *Time & Society*, 18(2-3), 332-350.
- Toussaint, Y., & Mallein, P. (1992). Diffusion, médiation, usages des TIC. *Culture Technique*, 24, 219-226.
- Tunstall-Pedoe, W. (2010). True knowledge: Open-domain question answering using structured knowledge and inference. *AI Magazine*, 31(3), 80-92.
- Turkle, S. (2011). *Alone together : Why we expect more from technology and less from each other*. New York : Basic Books.
- Turner, F. (2006). *From counterculture to cyberculture : Stewart Brand, the Whole Earth Network, and the rise of digital utopianism*. Chicago : The University of Chicago Press.
- Urry, J. (2005). *Sociologie des mobilités : Une nouvelle frontière pour la sociologie ?*. Paris : Armand Colin.
- Urry, J. (2016). *What is the Future?*. Cambridge : Polity Press.
- Velkovska, J. & Beaudouin, V. (2014). Parler aux machines, coproduire un service. Intelligence artificielle et nouvelles formes de contribution du client dans les services téléphoniques de l'après-vente. In A. Mallard & E. Kessous (Eds.), *Le travail commercial dans les télécommunications* (pp. 97-128). Paris : Presses des Mines.
- Velkovska, J. & Zouinar, M. (2018). *The illusion of natural conversation: interacting with smart assistants in home settings*. Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems. Montréal, 21-26 avril 2018.
- Vial, S. (2017). *Le design*. Paris : Presses Universitaires de France.
- Vidal, D. (2012). Vers un nouveau pacte anthropomorphique : Les enjeux anthropologiques de la nouvelle robotique. *Gradhiva*, 15(1), 54-75.
- Virilio, P. (1995). *La vitesse de libération*. Paris : Galilée.

- Virilio, P. (1996). *Un paysage d'événements*. Paris : Galilée.
- Wajcman, J. (2015). *Pressed for time: The acceleration of life in digital capitalism*. Chicago : The University of Chicago Press.
- Wajcman, J. (2018). How Silicon Valley sets time. *New Media & Society*. Disponible sur : <https://doi.org/10.1177/1461444818820073>
- Wajcman, J. (2019). The digital architecture of time management. *Science, Technology, & Human Values*, 44(2), 315-337.
- Weinstein, N. D. (1989). Optimistic biases about personal risks. *Science*, 246(4935), 1232-1234.
- Williams, R., & Edge, D. (1996). The social shaping of technology. *Research policy*, 25(6), 865-899.
- Zouinar, M. & Fréjus, M. (2010). Ergonomie des situations domestiques : nouveaux défis, nouvelles opportunités. In G. Valléry (Eds.), *Ergonomie, conception de produits et services médiatisés* (pp. 211-240). Paris : Presses Universitaires de France.
- Zuboff, S. (2015). Big other : surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, 30(1), 75-89.
- Zuboff, S. (2019). *The Age of Surveillance Capitalism : The Fight for a Human Future at the New Frontier of Power*. New York : Public Affairs.

Annexe

Détails du corpus de films publicitaires

	Produit	Pays	Année	Source
1	Amazon Echo	GB	2016	www.youtube.com/watch?v=sulDcHJzcB4&t=19s
2	Amazon Echo	FR	2018	www.youtube.com/watch?v=rdyYR0ljdYo
3	Amazon Echo	FR	2018	www.youtube.com/watch?v=Hnzy9Wh1TWk
4	Amazon Echo	FR	2018	images-na.ssl-images-amazon.com/images/I/D1-O9Ha+j-S.mp4.mp4
5	Amazon Echo	CAN	2018	www.youtube.com/watch?v=GY__yUY7XHA
6	Google Home	USA	2017	www.youtube.com/watch?v=rgWQbiiLiLI
7	Google Home	USA	2017	www.youtube.com/watch?v=RZNqSy-zFXo
8	Google Home	GB	2017	www.youtube.com/watch?v=EdMFmPEnhIU
9	Apple HomePod	USA	2018	www.youtube.com/watch?v=BLAnTQfFXOI