HAL
open science

Interaction homme-machine : l'oralité et l'écriture en question : le cas de Siri

Hugo Betelu

To cite this version:

Hugo Betelu. Interaction homme-machine : l'oralité et l'écriture en question : le cas de Siri. Sciences de l'information et de la communication. 2016. dumas-01613393

HAL Id: dumas-01613393

https://dumas.ccsd.cnrs.fr/dumas-01613393

Submitted on 9 Oct 2017

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

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

Copyright

CELSA
PARIS-SORBONNE

Master professionnel

Mention : Information et communication

Spécialité : Médias

Option : Médias et numérique

Interaction homme-machine : l'oralité et l'écriture en question

Le cas de Siri

Responsable de la mention information et communication
Professeure Karine Berthelot-Guiet

Nom, prénom : BETELU, Hugo

Promotion : 2015-2016

Soutenu le : 22/11/2016

Mention du mémoire : Très bien

École des hautes études en sciences de l'information et de la communication - Université Paris-Sorbonne
77, rue de Villiers 92200 Neuilly tél. : +33 (0)1 46 43 76 76 fax : +33 (0)1 47 45 66 04 www.celsa.fr

Remerciements

Je souhaite remercier l'ensemble des enseignants du CELSA pour leur dévouement au cours de l'année scolaire 2015/2016. Merci à eux pour leur temps, leur passion et leurs enseignements. Je garderai un excellent souvenir de la formation et des enseignements qui y sont dispensés.

Je souhaite particulièrement remercier Jerôme Denis pour son soutien dans l'élaboration de ce mémoire. Merci à lui pour sa réactivité et ses remarques qui m'ont permis d'avancer avec toujours plus de pertinence.

Merci à Clotilde Chevet, qui comme moi, est passionnée par ce sujet et avec qui j'ai beaucoup échangé. Merci pour son aide.

Merci également à l'ensemble de mes camarades, collègues et amis qui m'ont toujours soutenu dans mes projets professionnels, scolaires et personnels. Merci à eux pour leur temps, leur ingéniosité et générosité.

J'aimerais aussi dire un grand merci à l'Université française. Souvent critiquée et décriée, je suis fier d'avoir pu bénéficier de tels enseignements durant toute ma scolarité dans le supérieur. Elle m'a permis d'apprendre, de m'ouvrir aux autres et de cultiver un esprit critique. Me voilà enfin prêt pour la suite.

Enfin, je voudrais vous remercier tous une dernière fois.

Sommaire

Table des matières

Remerciements **1**

Sommaire **2**

INTRODUCTION **4**

PROBLÉMATIQUE **8**

CONTEXTE DISCIPLINAIRE **10**

HYPOTHÈSES **14**

MÉTHODOLOGIE **15**

I. Histoire, écriture et oralité des objets **17**

A. Aux origines de Siri **17**

1. Oralité et écriture : une interrelation historique 17
2. Outil et langage : indissociabilité 20
3. Technologies : de la reconnaissance vocale à l'assistant personnel 22

B. Le système Siri (technique, écriture et oralité) **25**

1. Processus technique de Siri 25
2. Quelle relation écriture-oralité en représentation ? 30

II. Interface de Siri : quel cadre pour l'échange homme-machine ? **45**

A. Les cadres instituants du logiciel Siri **45**

B. L'interface : lieu d'interaction par excellence ? **48**

1. Lieu de rencontres de deux paroles avec deux origines et deux natures 49
2. Lieu de dialogue ? L'écran comme lieu d'interaction 56
3. Lieu de conversation ? 60

III. Qui est qui, pour qui ? : perception et co-construction identitaire de Siri et de ses utilisateurs **70**

A. Écrits d'écran et oralité au service de la construction identitaire de Siri **70**

B. Une co-construction identitaire soutenue par des discours d'accompagnements **74**

C. Quel avenir, quelles évolutions et quelle place pour Siri ? **77**

Conclusion **81**

Bibliographie **87**

Résumé du mémoire **91**

Mots clés **91**

INTRODUCTION

Qu'il s'agisse de GPS, des dispositifs d'annonces en gare ou d'assistants personnels numériques (Siri, Google Now), les machines parlantes font désormais parties du quotidien. La mise en place de la technique de reconnaissance vocale s'inscrit pourtant dans une évolution longue de plus de 80 ans. Elle n'a donc rien d'une révolution, mais si son développement est considéré comme tel aujourd'hui, c'est parce que son déploiement a pris beaucoup d'ampleur ces dix dernières années et que la technique est accessible individuellement par les personnes équipées.

« Les machines nous parlent depuis 1930 »[1], avec depuis, de nombreuses évolutions technologiques mais aussi l'émergence de discours symboliques portés par l'industrie cinématographique (science fiction notamment) et de la musique qui a finalement contribué à l'évolution des machines parlantes.

Plusieurs enjeux sont liés au développement des machines parlantes. Le premier problème auquel se sont intéressés les scientifiques et les professionnels de l'industrie des télécoms est celui de la reconnaissance vocale. Le second est celui de la mémoire, grâce à laquelle la machine peut enregistrer les données vocales dictées. Le troisième, celui de la retransmission. Ainsi, dès l'Exposition Universelle de 1962, *« IBM présente un système (…) appelé Shoebox. Celui-ci fait office de calculatrice à reconnaissance vocale. Shoebox est capable de reconnaître les dix premiers chiffres, et six commandes parmi lesquelles l'addition, la soustraction et le total. Ainsi, il était possible d'énoncer un problème de maths dans le microphone et obtenir le résultat sur la machine reliée au dispositif. Chaque nombre reconnu est affiché, permettant ainsi de vérifier si celui-ci est exact. »*[2]

Dans le développement de ce type de technologies, 1988 est une date importante puisqu'elle correspond au moment où Apple lance une vidéo qui montre

---

[1] Mary Branscombe, *Reconnaissance vocale : la vie avant Siri* : http://www.tomsguide.fr/article/siri-reconnaissance-vocale-histoire,5-118.html (consulté le 21 mai 2016)

[2] ibid

« comment en 2011, nous all[i]ons pouvoir parler à un ordinateur et nous faire comprendre suffisamment bien pour tenir une conversation avec lui ».[3]La firme californienne passe alors un cap dans la mesure où ce « Knowledge Navigator » est présenté comme étant, non pas un *« simple outil de reconnaissance vocale »*[4], mais bien une intelligence artificielle ou un assistant personnel numérique.
Et alors qu'en 1984, Steve Jobs présentait son ordinateur parlant (le Macintosh), en 2011 c'est sur un téléphone que l'assistant numérique personnel Siri est dévoilé à la presse. Cette application informatique de commande vocale répond à de nombreuses requêtes.
Nous nous intéresserons donc à cet assistant numérique personnel au cours de cette étude.

Au delà des caractéristiques orales du système Siri, l'écrit reste très présent. *Oral* et *écrit* s'entremêlent, coexistent parfois et méritent alors une attention particulière dans la mesure où ils soulèvent des questionnements fondamentaux liés à l'histoire de l'écriture, de l'oralité, du langage et des techniques. En effet, *« l'étude diachronique de l'oralité, de l'écriture et des différentes étapes marquant l'évolution de l'une à l'autre offre un cadre de référence permettant de mieux comprendre non seulement la culture orale des origines et la culture écrite ultérieure, mais aussi la culture de l'imprimé (…) et la culture électronique – qui repose à la fois sur l'écrit et l'imprimé. »*[5]L'analyse du rapport oralité/écriture est donc nécessaire dans notre étude dans la mesure où Siri est un logiciel qui concentre tout un héritage culturel (langage, technique,…) lié à l'oralité et l'écriture.

C'est donc à travers une analyse sémiologique, complétée par une étude documentaire, que nous traiterons le sujet. L'étude sémiologique nous permettra de prendre le recul nécessaire à la compréhension de la relation entre écrit et

---

[3] ibid
[4] Ibid
[5] Walter J. Ong, Oralité et écriture, Les Belles Lettres, 2014, p. 22

oral dans le dispositif. En effet, nous analyserons l'ensemble des marques énonciatives orales et écrites qui permet la construction de la situation d'interaction entre l'homme et la machine.
D'autre part, la recherche documentaire, qui s'appuiera sur des captures d'écran du logiciel Siri et des articles de presse ou des discours du constructeur, permettront de compléter l'étude dans un objectif de déconstruction des symboles et imaginaires qui nourrissent les représentations liées à Siri. Mais, nous ne pourrons comprendre ces différents éléments sans effectuer une analyse technico-sémiotique du dispositif. Nous analyserons ainsi la manière avec laquelle les phases de l'interaction sont rendues possibles.

Le schéma[6] de Clotilde Chevet, doctorante au CELSA qui s'intéresse également au sujet, nous permet d'avoir une vue globale des différentes phases de l'interaction avec un assistant personnel numérique et donne une idée de l'interrelation qui existe entre oralité et écriture dans les instances de production, diffusion et réception.

[6] Clotilde Chevet, La part de l'imaginaire dans « l'interaction homme-machine » : le cas des assistants personnels numériques, Mémoire de recherche Master 2 au CELSA Sous la direction du Professeur Karine Berthelot-Guiet, p. 11

⇨ Conception et écriture de réponses en amont par l'entreprise ⇨ Choix d'une voix en fonction du destinataire visé	⇨ Echange oral entre l'utilisateur et « l'assistant personnel » ⇨ Incitation à l'interaction par le dispositif	⇨ Retranscription de l'échange ⇨ Sémiotique et représentation graphique du dialogue
EN AMONT	SIMULTANEMENT	

Schéma des différentes phases de l'interaction avec un assistant personnel numérique

Nous désirons poursuivre, dans la lignée de son travail, l'étude de l'assistant personnel numérique Siri avec une perspective dédiée aux interrelations entre oralité et écriture au sein du dispositif numérique.

PROBLÉMATIQUE

Siri est par définition un objet hybride dans la mesure où il délivre des réponses écrites et/ou orales à des demandes, exclusivement orales. Dès lors, les modalités de ce qu'on appellera la construction de l'interaction homme-machine sont complexes et nécessitent des analyses précises quant au rôle de l'écrit, de l'oral, et des deux à la fois, c'est à dire, en relation.

En effet, qu'en est-il de l'articulation entre les deux modalités de relation entre l'homme et la machine. Pourquoi le constructeur Apple vante-t-il les mérites d'un outil destiné à une pratique de l'oral alors même que l'écrit reste en grande partie le mode dominant au sein du dispositif et dans la relation avec l'utilisateur ?

Nous le comprenons aisément, Siri n'est pas un dispositif simple et renferme des contradictions, articulations et relations dont l'écriture et l'oralité sont les composantes à la réussite des échanges entre l'utilisateur et la machine.

Nous voudrions donc à travers cette étude de Siri répondre à la question suivante :

Comment oralité et écriture s'articulent dans la situation d'interaction homme-Siri ?

Dans l'histoire des objets parlants, Siri fait partie de ceux dont la caractéristique principale est de disposer d'un écran. En effet, le téléphone joue le rôle d'interface qui permet à l'usager de pouvoir engager des rapports d'échanges avec le logiciel Siri (nous parlerons alors, pour plus de simplicité dans la lecture, de « machine »). Ces rapports d'échanges se forment à travers des écrits mais également une oralité (cf. le schéma ci-dessus). C'est en cela que l'étude menée semble devoir prendre en compte la question de l'articulation entre écriture et oralité. Autrement dit, l'analyse ne devra pas porter d'une part sur l'écriture au sein du dispositif et d'autre part sur l'oralité de la machine/de l'usager, mais bien

de l'écriture et de l'oralité au sein du dispositif et dans le cadre d'une situation d'échange homme-machine.

Nous devons également apporter une précision importante quant à l'utilisation de l'expression *« la situation d'interaction homme-Siri »*. Nous devons tout de suite éclaircir ces propos car ils pourront au cours de notre étude poser des problèmes d'interprétationdans les analyses effectuées.

La morphologie du terme *interaction* nous indique que celle-ci constitue une action réciproque (« act ») entre deux inter-actants (« inter »). La forme interactive sous-entend donc qu'il existe une action réciproque. Mais c'est dans cette dernière formule (« action réciproque ») que les questions émergent : la réciprocité est-elle effective avec la machine ? L'interaction de la machine existe-t-elle ?

Cependant, les machines n'agissent pas avec intention ni sens lorsqu'elle s'exécute. La machine suit une succession d'état qui n'est que la résultante d'une somme d'actions réalisées en amont par l'homme afin que la machine puisse être en état de fonctionner.

Derrière le terme *interaction*, se cache en réalité une véritable métaphore qui vise à situer l'homme par rapport à la machine. Il se joue en fait un jeu de croyance extrêmement fort, dû notamment à l'impression que l'usager peut influencer le scénario alors que c'est le concepteur de la machine qui a défini en amont un nombre de scénarios possibles.

Dans le cas de l'assistant vocal Siri, c'est en effet ce qu'il se passe : les concepteurs ont anticipé en amont les questions que pourraient poser les usagers et ont généré des réponses toutes prêtes qui se donnent à lire à l'usager au moment opportun et qui sont le signe de ladite « interaction » (même si après des améliorations du logiciel, le machine learning a pris une place de plus en plus importante).

Clotilde Chevet explique clairement la situation d'énonciation pensée en amont par les concepteurs : *« Il est ainsi possible de poser parfois dix fois de suite la même*

question et d'obtenir dix réponses différentes. (…) La quantité de réponses proposées à la question « Qui es-tu ? » souligne la volonté de l'entreprise de proposer plus qu'un service, une relation. »[7].

Nous ne voulons pas nous perdre, dans cette étude, dans une discussion autour de l'interaction car elle présente un risque de conclusions normatives. Nous désirons plutôt questionner le modèle et les formes d'échanges que produisent écriture et oralité au sein du dispositif Siri. Autrement dit, nous chercherons à savoir d'une part comment la situation d'interaction est techniquement rendue possible, d'autre part, comment l'interface (graphique mais pas seulement) offre un cadre à ces échanges et enfin, nous tenterons de déterminer qu'est ce que « faire la machine » pour l'usager et qu'est ce que « faire l'humain » pour Siri à travers une analyse écriture-oralité.
Ainsi, dans le cadre de notre étude, nous envisageons le terme « interaction » comme étant un ensemble d'échanges entre la machine (Siri) et l'homme (l'usager du dispositif) qui est rendu possible par la concordance des moyens de communication (oraux et/ou écrits) entre inter-actants.

CONTEXTE DISCIPLINAIRE

Notre étude s'inscrit au carrefour de différents courants scientifiques.

Tout d'abord celui de la cybernétique, représentée notamment par les travaux de Norbert Wiener (et dans une certaine mesure John Von Neumann) qui a repensé la théorie de la communication en examinant *« l'objet ou le sujet sous l'angle de l'information »*[8].

---

[7]Clotilde Chevet, La part de l'imaginaire dans « l'interaction homme-machine » : le cas des assistants personnels numériques, Mémoire de recherche Master 2 au CELSA Sous la direction du Professeur Karine Berthelot-Guiet, p.49

[8]Daniel Bougnoux, Sciences de l'Information et de la Communication, Larousse, 1993

Norbert Wiener se place dans un champ vaste et complexe en voulant ajouter *« l'étude du langage et l'étude des messages en tant que moyens de contrôle sur les machines et la société »* à la *« théorie de la technique électrique pour la transmission des messages »*[9]. L'une des thèses de Wiener part du principe que *« dans le développement futur des messages et de ces dispositifs, les messages entre l'homme et les machines, entre les machines et l'homme, et entre la machine et la machine sont appelés à jouer un rôle sans cesse croissant »*[10]. Il introduit par ailleurs la notion de régulation, de contrôle et de rétroaction (feedback), qui nous intéresse particulièrement dans l'étude de l'assistant personnel numérique Siri. Enfin, son étude sur le mécanisme et l'histoire du langage nous donnera matière à réfléchir et analyser Siri tant ses travaux sont complémentaires de ceux de Jack Goody et Walter J.Ong.

Ensuite, les travaux de Jack Goody et WalterJ.Ong, justement, sur l'histoire de l'écriture et de l'oralité.

Leurs travaux, même s'ils se contredisent sur certains points, nous éclairent sur l'évolution des pratiques d'écriture, d'oralité et de lecture en même tant qu'ils nous font prendre conscience de la puissance du langage à transformer notre perception du monde et nos rapports aux autres.

Jack Goody a une thèse relativement claire : *« l'écriture, en tant que moyen d'archivage des informations et d'organisation du savoir en catégories, a permis le développement de la pensée logique, de l'abstraction, de l'objectivité et finalement de la science »*[11]. Il parle alors de « technologie de l'intellect » et place celle-ci comme un langage visible, signe d'une extériorisation qui permet à l'Homme de penser *(« l'écriture permet de détecter les contradictions en plaçant côte à côte différents comptes-rendus d'un même événement. Elle permet une analyse plus fine de l'usage de la langue et par là-même encourage la réflexivité »*[12]) et de

---

[9] Norbert Wiener, Cybernétique et société, Éditions du Seuil 2014, p. 47

[10] Ibid, p.48

[11] Emmanuel Isnard, « Jack Goody, Pouvoirs et savoirs de l'écrit », L'Homme, 2009, 274-277

[12] Jack Goody, Oralité et écriture, Communication et Langage n°154, Décembre 2007, p.8

requalifier son rapport au temps et à l'espace *(« Elle [l'écriture] nous permet de communiquer non seulement en face-à-face, mais également à des temporalités et dans des endroits différents. »*[13]).

Non seulement ses travaux nous paraissent pertinents dans une perspective historique, mais dans le cas de notre étude de l'assistant personnel numérique Siri, Jack Goody apporte également des enseignements précieux qui permettent une distanciation certaine vis à vis du développement des moyens de communication numériques, balayant ainsi les discours alarmistes qui voudraient que la société de demain soit 100% orale ou au contraire 100% écrite : *« Un nouveau moyen de communication ne remplace, nesupprime ni ne marginalise le précédent. Il s'y ajoute plus qu'il ne le supplante. On me demande parfois si l'arrivée des médias informatisés va entraîner la disparition des livres. Cela changera certainement la nature et le statut du livre, comme l'écriture l'a fait pour la parole, mais le livre ne disparaîtra pas ni a fortiori l'écriture. »*[14]

Walter J. Ong travaille les mêmes sujets à la différence qu'il place le langage comme étant un *« phénomène oral »*[15]. Selon lui, *« tous les textes écrits sont liés, directement ou indirectement, au monde du son (l'habitat naturel du langage) pour révéler leur sens. »*[16] Il devient parfois même caricatural lorsqu'il dit que *« l'écriture ne peut en aucun cas se passer de l'oralité »* ou encore quand il suggère une hiérarchie entre les deux langages : *« On peut qualifier l'écriture de système de modélisation secondaire dépendant d'un système primaire antérieur : le langage parlé. »*

Cependant il rejoint Jack Goody au sujet de la transformation de la pensée et du rapport au monde par l'écrit. Il s'intéresse également à la mémoire et aux rapports qu'entretiennent écriture et oralité dans les systèmes imprimés ou numériques. C'est en cela que ses travaux feront l'objet d'une attention particulière dans notre étude.

---

[13] Ibid, p.7

[14] Ibid, p.7/8

[15] Walter J. Ong, Oralité et écriture, Les Belles Lettres, 2014, p. 26

[16] ibid, p.28

Notre étude s'inscrit aussi dans le courant de recherche du GRIPIC, le groupe de recherche du CELSA Paris-Sorbonne. Les travaux d'Emmanuel Souchier viendront nourrir nos réflexions sur *« l'introduction de l'écran et des nouvelles technologies dans l'univers de l'écrit »*[17].

Nous analyserons notamment les transformations des *modes de faire* et des *modes de penser* qu'implique un dispositif numérique comme Siri, dans une perspective plus large de l'*écrit d'écran* et de sa pensée. Nous aurons également l'occasion d'introduire la notion d'*architexte* qui vaut également pour notre sujet d'étude (comment Siri se donne à voir ?). Enfin, dans une perspective plus globale, les travaux d'Emmanuel Souchier et d'Yves Jeanneret permettront de contextualiser, distancer et mettre en perspective nos analyses sur la place du texte, de la parole dans les écrits d'écrans, à travers l'*énonciation éditoriale*.

Enfin, il paraît logique d'introduire une littérature plus contemporaine et propre à l'intelligence artificielle moderne afin de pouvoir comprendre le contexte dans lequel s'inscrit notre étude. Nous nous concentrerons donc sur des textes divers qui ont pour thème commun les relations homme-machine. Il pourra s'agir de textes sur l'évolution du dialogue homme-machine (Jean Caelen), sur le traitement automatique de l'oral et de l'écrit (Daniel Luzzati) ou encore sur des textes plus larges comme ceux de Serge Tisseron.

Cette littérature nous permettra de contrebalancer les analyses portées sur Siri, en s'appuyant sur des exemples ou des analyses récentes. Car n'oublions pas que rien n'est statique, d'autant plus que les évolutions technologiques propres à ce secteur (intelligence artificielle, machines parlantes) sont rapides.

---

[17] Emmanuel Souchier, L'écrit d'écran, pratiques d'écriture & informatique, Communication et Langages n°107

HYPOTHÈSES

Pour répondre à ces questionnements, nos réflexions s'articuleront autour de différentes hypothèses :

Nous désirons poser une première hypothèse selon laquelle l'assistant personnel numérique Siri constitue un dispositif où écriture et oralité s'entremêlent en tout point, l'inscrivant ainsi dans l'histoire longue des évolutions des objets dans laquelle le langage est inscrit. Nous désirons montrer dans le cas de Siri les rapports qu'entretiennent écrit et oral.

Pour explorer cette hypothèse, nous voulons remonter dans l'histoire de l'écriture et de l'oralité pour questionner cette interdépendance qui fait de Siri un système technique avec une place à part entière dans cette historicité. Nous désirons par ailleurs montrer comment le processus technique joue de l'écriture et de l'oralité. Nous tenterons de fournir une représentation visuelle objective de celui-ci pour répondre à la question : Comment l'*interaction* est techniquement rendue possible ? Et comment se donne-t-elle à voir à l'utilisateur ?

La deuxième hypothèse est que l'assistant personnel numérique offre un cadre écrit et oral à « l'interaction homme-machine ».

Pour répondre à cette hypothèse, nous chercherons à montrer les différentes formes d'*interaction* que produisent écriture et oralité dans le dispositif. En effet, l'interface graphique, mais pas seulement, semble donner un cadre, engager la parole et les actions d'écriture de l'utilisateur. Nous chercherons à déterminer les marques énonciatives propres à Siri dans la mise en place d'une situation d'interaction.

Enfin, une dernière hypothèse semble intéressante à vérifier : d'un point de vue des usages, Siri semble « faire l'humain » lorsque qu'il adresse une réponse à l'utilisateur quand ce dernier semble « faire la machine » lorsqu'il pose une question à l'assistant personnel numérique. Nous soulevons ici une problématique qui touche à la question identitaire d'une part, mais également à une question de relation entre l'homme et la machine et de la considération de chacun des inter-actants dans la situation de communication.

Pour cette hypothèse, nous questionnerons les rapports entre écriture et oralité au sein du dispositif qui tendent à définir certains rôles. Nous chercherons également à comprendre comment est perçu/co-construit l'identité humaine de Siri et comment l'oralité et l'écriture participe ensemble à cette co-construction.

MÉTHODOLOGIE

Pour répondre à nos différents questionnements, nous avons mis en place un ensemble de méthodologies qui permettra de nourrir nos réflexions et d'effectuer les analyses en question.

Tout d'abord, nous utiliserons des captures d'écrans issues de plusieurs blogs ou Tumblr comme *Siri et moi*. Le choix semble justifié puisque ces espaces numériques rendent compte de la variété énonciative de Siri et soulèvent à la fois des paradoxes (réponses humoristiques…) mais permettent également de repérer quelques intentions d'un point de vue de la conception/production qui nous faciliteront l'analyse. D'autre part, ce blog participera également à définir les évolutions liées à Siri puisque depuis 2011, le design a évolué, les marques énonciatives ne sont plus forcément les mêmes et c'est en cela que nous

pourront analyser les intentions des concepteurs de placer Siri comme un véritable interlocuteur.

Notons néanmoins que dans le cas de notre étude, une certaine limite est directement posée par l'utilisation de captures d'écrans du logiciel puisque l'étude de l'oralité ne peut se faire à travers de simples captures d'écrans. Ainsi, nous pourrons faire référence à notre propre expérience avec le logiciel de reconnaissance vocale (dont les captures d'écrans permettront malgré tout de contextualiser l'analyse).

D'autre part, nous utiliserons des sites de presse et les discours d'entreprises (site web notamment) pour analyser comment le dispositif nous est donné à voir en tant qu'utilisateur. Nous tenterons de comprendre comment Apple tente de donner une identité à Siri auprès des utilisateurs à travers les discours médiatiques.

I. Histoire, écriture et oralité des objets

A. Aux origines de Siri

Siri est souvent présenté comme révolutionnaire, et c'est d'ailleurs de cette manière qu'il est donné à voir dès de son lancement en 2011, par le vice président d'iOS (de 2007 à 2012), Scott Forstall : *« Il [Siri] reconnaît le sens, pas seulement les mots »*[18]. Dès sa présentation à la presse le logiciel est placé au dessus de tous les autres systèmes de reconnaissance vocale puisqu'il est qualifié d' « assistant intelligent ». Mais Siri s'inscrit en réalité dans une histoire longue de plusieurs siècles, qui lorsqu'elle est prise en compte, permet de comprendre l'origine du dispositif ainsi que les relations qu'entretiennent écriture et oralité, dans le cas qui nous intéresse.

Il faut remonter dans l'histoire des télécommunications mais également dans la « grande histoire » de l'évolution des outils pour comprendre et justifier le fonctionnement de Siri.

1. Oralité et écriture : une interrelation historique

Depuis l'arrivée de ce qu'on appelle les médias informatisés, il est facile de penser que le monde entier puisse se diriger vers une société du texte, de l'écriture, du langage écrit : le code informatique est écrit (on peut partir du principe général qu'il ne se parle pas), les médias numériques – au sens

[18] Citation extraite d'une vidéo de présentation de Siri par Apple, le 4 octobre 2011 : https://www.youtube.com/watch?v=agzItTz35QQ (consulté le 29 mai 2016)

informationnel – produisent des articles à lire via des écrits d'écrans, les marques vendent des objets technologiques dont les usages sont fortement liés à l'écrit (Smartphones, ordinateurs, etc.). Mais il serait naïf de penser que nous allons vers une société du texte, alors même que le langage oral dispose d'une place encore importante, pour ne pas dire fondamentale, dans notre société.

Il n'y a qu'à voir les fonctionnalités des applications de messageries instantanées, qui incluent généralement toutes des fonctionnalités d'enregistrement de la parole. Rares sont les ordinateurs qui ne sont pas équipés de microphone. Le téléphone, même si ses usages ont muté, reste un outil qui confirme que l'oralité dispose encore d'une place de choix dans notre société.

Historiquement, Jack Goody l'énonce clairement, en retraçant l'évolution de l'écriture au sein d'un monde où l'oral était la référence langagière : *« L'introduction de l'écriture, ou d'un autre moyen de communication, ne supprime pas l'ancien. L'oral reste très important, de même le monde de l'imprimé l'est pour les médias informatisés. Leurs rôles ont changé, mais ils restent fondamentaux. »*[19]

Il semblerait que l'écriture soit alors considérée comme une technique, ou plutôt une *« méthode d'enregistrement de la parole »*[20] qui a permis de classer, d'archiver et organiser le savoir. Mais Walter J.Ong nous rappelle que l'écriture a ceci de particulier qu'elle est *« directement ou indirectement liée au monde du son (l'habitat naturel du langage). »*[21] Pour Jack Goody d'ailleurs, l'oralité n'est en effet jamais très loin de l'écriture : *« Parler est toujours une activité qu'il est important de cultiver, même dans les écoles dont l'objectif premier est d'apprendre à écrire. Parler reste une activité fondamentale. On peut le constater dans le cas de tribunaux qui, bien qu'ils soient, par beaucoup d'aspects, construits autour de la chose écrite (sur laquelle repose souvent la notion même de droit), exigent des témoignages oraux et la présence physique de l'accusé, du juge et du jury. »*[22]

---

[19] Jack Goody, Oralité et écriture, Communication et Langage n°154, Décembre 2007, p.10

[20] Ibid, p. 6

[21] Walter J. Ong, Oralité et écriture, Les Belles Lettres, 2014, p. 28

[22] Jack Goody, Oralité et écriture, Communication et Langage n°154, Décembre 2007, p.7

Historiquement, l'écriture et l'oralité ont toujours eu des liens forts et plusieurs arguments permettent de le vérifier.

Même si dans chacune de ses démonstrations, Walter J.Ong fait passer l'oralité avant l'écriture (ce qui semble être discutable dans la mesure où peu d'éléments laisse entendre une hiérarchie des langages aussi forte qu'il ne la pose), il apporte des éclaircissements intéressants qui servent nos propos. Il explique notamment, comme beaucoup l'ont certainement fait, le fonctionnement de l'art rhétorique qui dans son essence même pose la question de la relation entre l'écriture et l'oralité puisque *« l'art du discours, se référait essentiellement au discours oral, même si en tant qu'art (ou science) réfléchi et organisé, il était nécessairement un produit de l'écriture »*[23]. Précisons que Jack Goody propose, au sujet de l'écriture, le terme de « technologie de l'intellect » qui justifie la fonction de l'écriture. Elle permet en effet d'organiser, hiérarchiser les informations ou encore comparer des discours, ce qui a été fort utile dans le développement de l'art oratoire.

D'autre part, si l'on tisse le fil de la rhétorique jusqu'au bout, vient la lecture. Et avec celle-ci, la question de l'indissociabilité entre l'écriture et l'oralité, encore une fois. *« Lire un texte signifie le transformer en son, à voix haute ou dans notre imagination »*[24] affirme Walter J.Ong et nous pouvons difficilement le contredire sur ce point. Nous pouvons d'ailleurs faire l'exercice simple de penser à un mot et voir à quel point notre culture de l'écrit nous amène presque systématiquement à produire *« une vague image du mot épelé »*[25] dont il est question. Autrement dit, *« l'écriture nous fait assimiler les mots à des choses car nous voyons les mots écrits comme des traces visibles destinées à ceux qui les déchiffrent »*[26]. Dans le même sens, Jack Goody nous dit aussi que *« l'écriture est*

---

[23] Walter J. Ong, Oralité et écriture, Les Belles Lettres, 2014, p. 29

[24] Walter J. Ong, Oralité et écriture, Les Belles Lettres, 2014, p. 28

[25] Ibid, p.32

[26] Ibid, p.31

devenue un langage visible », [qui] peut être conservé de manière très différente. »[27]

Nous pourrions développer ces points de manière plus approfondis mais déjà nous constatons que l'écriture et l'oralité, dans une perspective historique sont en interdépendance constante. Le lien entre les deux types de langage est extrêmement fort. Et nous verrons rapidement que dans le cas de notre objet d'étude, que cette interrelation est d'une importance tout aussi capitale.

2. Outil et langage : indissociabilité

Mais les travaux de Jack Goody et Walter J.Ong ne sont pas suffisants à démontrer l'interrelation existante entre écriture et oralité d'un point de vue historique. De plus, nous aimerions nous diriger peu à peu vers des analyses qui permettent de comprendre notre objet d'étude, Siri. Pour cela, il convient de comprendre comment *« l'évolution des outils est marquée par un déplacement progressif de la mémoire »*[28] et qui, par là même, produit une reconfiguration des modes de faire, de penser et d'agir.

Emmanuël Souchier et Jack Goody s'accorde à dire que la fabrication d'outils par l'homme et *« l'intelligence sociale »* dont il dispose*« résultent de l'invention du langage »*[29]. Emmanuël Souchier ajoute : *« L'externalisation de l'outil va donc progressivement s'accompagner d'une externalisation des techniques de mémoire et des technologies intellectuelles, dont la première et la plus importante fut sans doute l'écriture. »*[30] Il y a finalement une extension de la mémoire de l'homme à

---

[27] Jack Goody, Oralité et écriture, Communication et Langage n°154, Décembre 2007, p.7

[28] Emmanuël Souchier, Mémoires – outils – langages Vers une « société du texte », Communication et Langage n°139, Avril 2004, p.41

[29] Jack Goody, Oralité et écriture, Communication et Langage n°154, Décembre 2007, p.7

[30] Emmanuël Souchier, Mémoires – outils – langages Vers une « société du texte », Communication et Langage n°139, Avril 2004, p.43

travers l'outil, comme s'il ne pouvait pas intégrer une *« sur-spécialisation qui finirait par l'handicaper dans son milieu naturel. »*[31] Dès lors *« l'outil externalise la mémoire de l'homme »*[32] mais celle-ci ne réside pas dans l'outil lui-même, c'est en revanche la mémoire qui permet de mettre en œuvre les potentialités de l'objet, précise Emmanuël Souchier.

D'autre part, comme nous l'avons précisé : sans langage, pas d'outils. Or *« le problème de la mémoire opératoire est dominé par celui du langage »*[33], on comprend donc à présent mieux pourquoi outil (qui nécessite la mise en œuvre de la mémoire) et langage sont en interrelation constante. Emmanuël Souchier l'explique de manière extrêmement lucide : *« L'apprentissage de l'outil ne peut se réaliser autrement que dans la sphère réflexive de ce que Leroi-Gourhan appelle « la mise en marche du symbolisme réfléchi ». On comprend dès lors la nécessité du recours aux sciences du langage (linguistique, sémiologie) lorsqu'il convient d'aborder des outils eux-mêmes constitués à partir des langages et qui leur sont pour la plupart dédiés, comme c'est le cas des médias informatisés. »*[34]

Dans une perspective plus liés aux dispositifs de communication, les outils et les langages ont toujours eu un lien fort.

Qu'il s'agisse des tablettes d'argiles mésopotamiennes, des télégraphes optiques ou des Smartphones, l'interrelation entre outil et langage a toujours existé, ce qui d'ailleurs fait même dire que le *« progrès technique est lié au progrès des symboles techniques du langage »*[35].

Ainsi, dans le cadre de notre étude, nous pourrons analyser plus aisément les relations qu'entretiennent outil et langages (parlé/écrit) d'autant plus que les médias informatisés sont sans aucun doute un domaine dans lequel la mise en

---

[31] Ibib, p. 42

[32] Ibid, p.44

[33] Emmanuël Souchier, Mémoires – outils – langages Vers une « société du texte », Communication et Langage n°139, Avril 2004, p.46

[34] Ibid, p.46

[35] Ibid, p.46

place de véritables langages est à prendre au sens propre, comme au figuré et qu'ils se situent *« à l'articulation de l'opératoire et du symbolique »*[36]

3. Technologies : de la reconnaissance vocale à l'assistant personnel

Avant Siri, Google Now ou Cortana, de nombreuses expérimentations ont été développées. Celles-ci ont donné lieu à la création d'outil de reconnaissance vocale puis, plus tard, les « assistants personnels » ont été créés.
A travers ces différentes évolutions, se trame en toile de fond, l'utopie d'un dialogue total, ou plutôt naturel, entre l'homme et la machine. Mais en observant de près les orientations de conception, nous nous apercevons que beaucoup d'obstacles restent à franchir et que l'interrelation entre écriture et oralité est toujours présente (comme une réponse aux difficultés rencontrées dans le cadre d'un dialogue homme-machine ?). Arrêtons-nous donc sur le développement de ces techniques vocales (et multimodales).

Dans le contexte du dialogue entre l'homme et la machine, deux domaines se distinguent nettement, selon Jean Caelen, Directeur des Recherches eu CNRS (Laboratoire LIG, Grenoble) : les services vocaux et la recherche d'information sur le web. Ces deux domaines servent une même cause : la transmission et l'échange d'information.
Mais si aujourd'hui encore les services vocaux comme Siri ne sont pas optimaux, c'est parce qu'il existe de réelles difficultés pour la machine d'intégrer le dialogue en langage naturel : *« La machine n'a pas la capacité d'anticipation ni la compréhension des intentions qui traversent le langage utilisés par les humains, l'utilisateur ne pourra pas parler par sous-entendus, ni introduire facilement de*

[36] Ibid, p.46

nouveaux mots, ni construire des sens dérivés, etc., choses que l'on fait spontanément en parlant à quelqu'un. »[37]

Nous pouvons déjà supposerici que Siri retranscrit les requêtes de l'utilisateur et ses propres réponses via l'interface dans un but de vérification du dialogue engagé. Le fait que la machine retranscrive à l'écrit (sur l'écran du Smartphone) notre requête orale permet de s'assurer que celle-ci a bien « compris » la requête comme nous l'entendons en tant qu'humain. L'écriture vient supporter l'oralité dans un but performatif (au sens de réussite de la construction de l'interaction homme-machine). Mais la machine retransmet également sa réponse orale à l'écrit. Ce choix de la part des concepteurs permet à l'utilisateur de s'assurer que la réponse donnée est cohérente avec sa requête (comme c'est le cas lorsque la machine donne à lire à l'utilisateur sa propre requête orale) et d'autre part, il permet de constituer un historique dans lequel l'utilisateur peut aller chercher (de manière éphémère), car rappelons que l'écriture est cette « technologie de l'intellect » qui permet d'archiver le savoir et donc les informations. Et Walter J.Ong d'ajouter : *« Les mots écrits sont des résidus. La tradition orale ne possède aucun résidu ou dépôt semblable. »*[38]

---

[37] Jean Caelen, Le dialogue homme-machine aujourd'hui, Dossier L'intelligence Artificelle, Centraliens n°580, Septembre 2007, p. 8

[38] Walter J. Ong, Oralité et écriture, Les Belles Lettres, 2014, p. 31

Sur cet exemple, la requête est donnée à la voix par l'utilisateur. La machine retranscrit le discours oral enregistré par l'utilisateur et donne une réponse orale et écrite.

Si les systèmes de traitement de l'oralité ne sont pas encore cent pour cent fiables, la recherche d'information sur le web, elle, a fait passer les outils de reconnaissance vocale en de véritables « assistants personnels intelligents », capables de répondre à toutes sortes de requêtes via les informations trouvées sur le web. Mais il se joue une interrelation entre écriture et oralité dont nous aurons l'occasion d'analyser les enjeux et composantes tout au long de cette étude.

D'un point de vue interactionniste, Siri se place comme une véritable *« interface multimodale »*[39], pour reprendre le terme employé par Jean Caelen. C'est à dire qu'il existe dans l'usage une multitude de modes d'interaction qui sollicite à la fois l'usage de la parole (requête de l'utilisateur et réponse de la machine), l'usage de l'écriture (lecture/observation, modification écrite de la requête orale) et les usages gestuels (tenue du téléphone, toucher de l'écran…).

[39] Jean Caelen, Le dialogue homme-machine aujourd'hui, Dossier L'intelligence Artificelle, Centraliens n°580, Septembre 2007, p. 8

B. Le système Siri (technique, écriture et oralité)

Dans cette sous-partie, nous voulons nous intéresser au processus technique du logiciel Siri. C'est à dire, que nous désirons comprendre comment lors de son utilisation, le logiciel fait intervenir écriture et oralité pour donner lieu à une situation d'interaction. Nous désirons dans cette approche qualifier cette interaction car Siri est un assistant numérique qui dispose d'une véritable spécificité par rapport à d'autres concurrents du type Google Now ou Amazon Echo, puisqu'il est équipé d'un écran qui donne à lire des informations concernant l'échange homme-machine.
Nous réaliserons ici une description du processus de Siri, puis une analyse des représentations proposées par le concepteur de ce processus de fonctionnement et enfin, nous nous arrêterons sur la reconnaissance vocale et le traitement du sens par la machine.

1. Processus technique de Siri

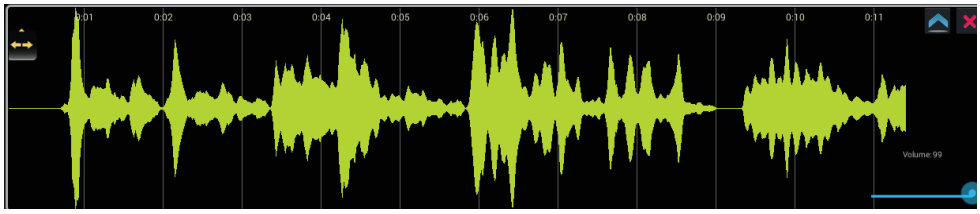
Plusieurs sources nous permettent de comprendre le fonctionnement technique de Siri et nous permettent de questionner l'utilisation des langages écrits et oraux pendant le fonctionnement du logiciel.

Du son à l'écriture numérique

Plusieurs sources concordent sur le fait que *« l'application ne fait qu'enregistrer la voix et ne réalise aucun traitement. Les données sont ensuite compressées au format Speex, un codec Open Source qui a été conçu pour compresser la voix et qui*

est déjà utilisé pour la VoIP (Voix sur IP est une technique qui permet de communiquer par la voix sur des réseaux compatible IP[40]) . *»*[41]

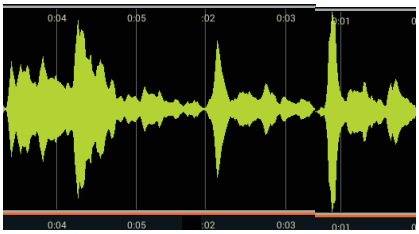
Dans cette première étape, la voix enregistrée est numérisée, donc écrite.



Ici, une représentation numérique courante de l'écriture numérique du son.

L'interrelation écriture-oralité se trouve dès le départ nécessaire à la mise en place de la situation d'interaction homme-machine. Siri est en effet conçu de sorte que sans oralité (de l'utilisateur), il n'y a pas de réponse qui peut être apportée de sa part. Or, à l'inverse, sans écriture (enregistrement du son au format numérique), aucun traitement de la parole et donc de réponse n'est possible non plus.

En même temps que le son est enregistré, l'écriture du son numérisé subit une séparation en différentes parties, pour isoler les mots retranscrits les uns des autres.



[40] Résultat Wikipédia pour la recherche « Voix sur IP » : https://fr.wikipedia.org/wiki/Voix_sur_IP (consulté le 3 janvier 2016)

[41] Des Français percent le secret de Siri, Simon Robic, 15 novembre 2011 : http://www.numerama.com/magazine/20578-des-francais-percent-le-secret-de-siri.html

Cette étape permet au logiciel de « chercher » les mots, les identifier (à travers une différenciation entre voyelles et consonnes) pour finalement pouvoir traiter la requête.La requête orale - divisée en plusieurs parties - se transforme ainsi en « texte numérique » (avec différents mots-clefs), interprétable, dans une phase ultérieure, par le logiciel à partir d'une base de données de mots-clefs (ex : « appeler », « envoyer », etc.).

Lors de cette première étape, l'utilisateur voit apparaître sa requête orale à l'écran, sous forme de texte numérisé. Nous analyserons dans un second temps les choix de représentation de l'énonciation par le constructeur et plus généralement ce qu'on pourrait appeler l'interaction dialogique homme-machine.

De l'écriture numérique au traitement de la requête utilisateur

A partir de cette étape, le logiciel tente de transformer en commande les mots qui ont été prononcés par l'utilisateur pour répondre à la requête utilisateur.

A ce niveau, il existe deux types de processus d'exécution de la recherche d'information : soit le logiciel est en face d'une requête appelée « local problem », soit il est confronté à un « global problem », soit les deux.

Dans le premier cas, la recherche d'information se fait à un niveau local (local data base), c'est à dire que les données sont disponibles directement depuis le téléphone (exemple : les numéros de téléphone des contacts du Smartphone, leurs noms, etc.).

Dans le second cas, les données nécessitent l'intervention des serveurs de l'entreprise. Une connexion au réseau est donc mise en place (Apple server data base) pour trouver l'information en question (exemple : la météo à Paris). Le fichier est *« envoyé aux serveurs d'Apple par le port TCP 443 via une connexion HTTPS* (on comprend donc ici pourquoi l'utilisation de Siri nécessite une connexion Internet) *et ce sont eux qui analysent le texte et traitent la requête »*[42]à

[42] Ibid

partir d'une base de données de mots-clefs et d'une comparaison de la requête vocale à un modèle statistique (rendu possible par un système de « machine learning », alimenté lui-même par la pratique des internautes) qui permet de traiter la requête utilisateur.

Une fois obtenue, l'information est renvoyée « en local », sur le Smartphone, où un mix des données (locales/serveurs) est effectué pour répondre à la requête utilisateur. Le logiciel exécute les commandes à partir des informations obtenues et traduit le résultat en texte numérique (qui s'affiche sur le Smartphone) et en synthèse vocale.

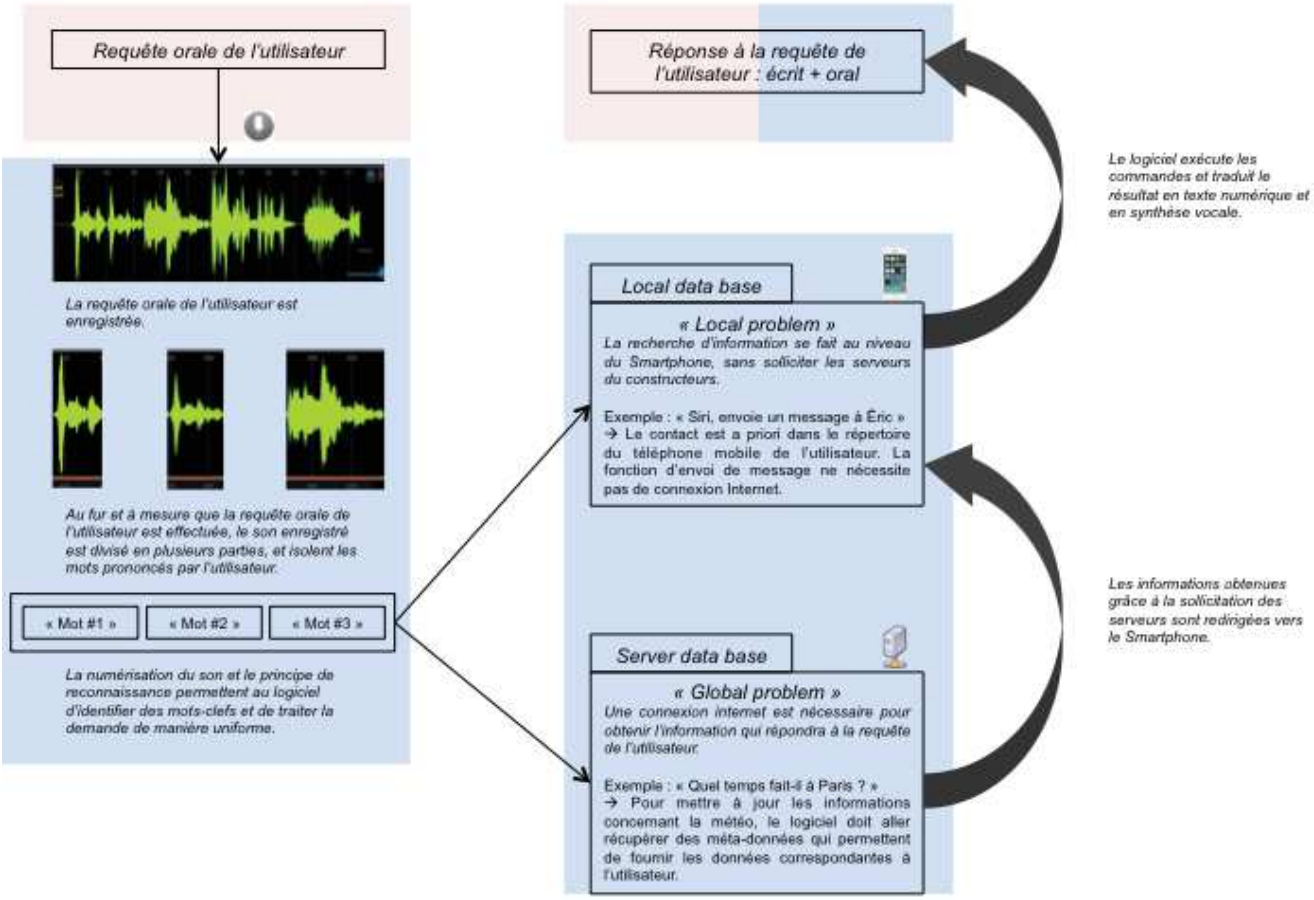
> **[L'origine des voix de Siri]**
>
> La synthèse vocale provient en réalité des voix humaines, Cyril Mazzotti pour la voix française. Les « speakers » ont enchaînées des *« semaines en studio à lire sur un prompteur des dizaines de milliers de mots [120 000 selon le décompte de Cyril]. (…) « Un exercice très particulier car il ne fallait laisser transparaître aucune émotion et s'en tenir à une lecture robotique des textes. Il s'agissait à la fois de phrases ouvertes et fermées, d'articles de presse et de mots sans queue ni tête »,* se souvient le comédien. (…) *Il s'agit de constituer une bibliothèque de tous les phonèmes. Puis de reconstruire artificiellement la totalité des termes courants,via un système de synthèse vocale de concaténation (du latin « concatenatio », enchaînement). »*[43]

Pour représenter cette description, nous avons réalisé un schéma qui permet de conserver en tête les différentes phases du processus technique et de se rendre

[43] Qui sont les voix de Siri ?, Isabelle Lesniak, Les Echos, 30 octobre 2015 http://www.lesechos.fr/week-end/business-story/innovation/021437753914-je-suis-siri-1170668.php?tR1B2Qx1HWahh8vg.99# (consulté le 3 janvier 2016)

compte de l'utilisation des modes oraux et écrits. Nous aurons l'occasion de s'appuyer sur ce schéma à plusieurs reprises dans nos analyses futures.

Les zones survolées en rouge signifient que l'oralité est le langage utilisé pour la phase du processus technique concernée. A l'inverse, les zones bleues indiquent que c'est l'écriture qui domine dans le processus en cours. L'oralité se trouve donc de bout en bout de la « chaîne ». Au départ, l'utilisateur fait sa requête à l'oral ; à la fin, le logiciel traduit la réponse sous forme orale. Mais entre ces deux temps, c'est tout un processus d'écriture qui se met en place : enregistrement du son, compression des fichiers créés à partir de l'oralité, transfert des données, analyse et renvoi.



Nous tirons deux conclusions de l'observation de ce processus technique.

D'une part, comme nous l'avions déjà mentionné au début de cette partie, l'oralité et l'écriture sont en totale interdépendance puisque sans l'une ou l'autre, le logiciel n'existe pas, ou du moins, il ne peut exécuter la tâche prévue par les concepteurs. Ce que l'on appelle couramment « interaction » est donc rendu possible seulement par l'articulation de l'oralité et de l'écriture ensemble dans un ordre qui permet à l'utilisateur d'obtenir la satisfaction désirée, à savoir une réponse à sa requête.

D'autre part, l'oralité et l'écriture permettent, grâce à leur complémentarité à travers le processus, de donner l'illusion à l'utilisateur d'un véritable échange conversationnel. Ce que nous voulons dire ici, c'est que tout ce processus technique qui joue à la fois de l'oralité et de l'écriture sert en quelque sorte la croyance selon laquelle l'homme peut parler à une machine. Toute la conception technique du logiciel est dirigée vers un seul et même but : créer une situation dialogique qui s'approche un maximum de celle qui pourrait avoir lieu entre deux hommes.

2. Quelle relation écriture-oralité en représentation ?

Nous désirons dans cette partie nous intéresser, non plus à la partie invisible du fonctionnement de Siri, mais bien à la partie visible par l'utilisateur. Nous reprendrons les différentes étapes du processus d'exécution type d'une requête (vue plus haut) afin d'observer les modes de représentations du logiciel « dans l'action » auprès de l'utilisateur et comprendre comment oralité et écriture vivent au sein du dispositif.

Une invitation au *dialogue oral spontané*[44]

Lorsque l'utilisateur désire utiliser Siri, il est confronté à une succession d'écrans qui invite au « dialogue ».

[44] Titre de l'article de Daniel Luzzati, chercheur de l'Université du Maine spécialiste de la question du dialogue homme-machine : « Le dialogue oral spontané : quels objets pour quels corpora ? »

Le premier écran auquel l'utilisateur est confronté présente un premier texte (« Que puis-je faire pour vous ? ») qui résume bien ce à quoi les constructeurs pensaient lorsqu'ils ont conçu le logiciel. Il correspond d'ailleurs parfaitement aux discours marketing avec lequel Apple a réussi à promouvoir Siri : *« Adressez-vous à Siri comme à une personne »*[45].

Cette première marque énonciative appartient au domaine de l'écrit. Elle est donnée à lire à l'utilisateur, mais elle suggère tout de suite le monde de l'oralité et du dialogue puisqu'en se plaçant comme un « assistant personnel », Siri se met au service de l'utilisateur. Ce dernier est alors invité à demander (de manière orale) ce qu'il souhaite, sans contrainte (en apparence, puisque l'on sait que la reconnaissance vocale nécessite de la part de l'utilisateur une énonciation claire).

[45] Clotilde Chevet, La part de l'imaginaire dans « l'interaction homme-machine » : le cas des assistants personnels numériques, Mémoire de recherche Master 2 au CELSA Sous la direction du Professeur Karine Berthelot-Guiet, citant la page « A propos de Siri » du site officiel d'Apple. Page consultée le 31 mars 2015. URL : https://support.apple.com/fr-fr/HT4992

Le deuxième écran arrive rapidement après le premier et invite directement l'utilisateur à parler, dans une perspective intimiste (« Allez-y, je vous écoute… »). C'est à ce moment précis que dans la représentation du logiciel à l'écran, le monde de l'écrit jouera un rôle secondaire, alors que l'analyse du processus technique non visible nous a prouvé le contraire (avec les différentes phases d'écriture qui se mettent en place lorsque l'utilisateur fait sa requête). Nous ne disons pas que l'écrit disparaît, bien au contraire, mais nous remarquons ici que dès le début de l'utilisation du logiciel, tout est fait pour faciliter certaines modalités d'actions de l'utilisateur, à savoir le mode oral.
La représentation du dispositif proposée oriente l'action de l'utilisateur vers certains modes. L'oralité est donc préférée à l'écriture et plusieurs éléments nous permettent de comprendre cette orientation de l'utilisateur par le logiciel.

D'une part, les ondes de couleurs apparaissent lorsque notre voix est détectée par le logiciel. Pour comprendre le rôle qu'elles jouent, reprenons par étape les différents écrans qui nous viennent aux yeux.

Sur le premier écran (« Que puis-je faire pour vous ? »), les ondes n'apparaissent pas tant que l'utilisateur ne parle pas. Mais dès qu'il commence à faire une requête, l'écran change : on lit « Allez-y, je vous écoute… » sur les deux écrans suivants et les ondes de couleurs apparaissent.

Nous pouvons interpréter ces ondes comme un *feedback* qui permet à l'utilisateur de s'assurer que la machine prend bien en compte le début de sa requête, qu'elle reçoit son message. L'utilisateur peut continuer à faire sa requête à l'oral, en sachant que Siri est engagé dans l'interaction. C'est ce que Jacques Cosnier appelle la « coordination interactionnelle », dans les études de l'interaction conversationnelle : *« Il existe des processus aux échanges verbaux qui permettent à chaque locuteur de gérer au mieux sa participation, c'est à dire d'accéder à la « félicité interactionnelle » : pouvoir expliquer sa pensée, la faire comprendre, (...). Pour ce, le parleur s'efforce d'être informé sur quatre points, que nous avons appelé les « quatre questions du parleur » : Est-ce qu'on m'entend ? Est-ce qu'on m'écoute ? Est-ce qu'on me comprend ? Qu'est-ce qu'on en pense ? »*[46] Par glissement, ces recherches sont interprétables dans le cas de l'interaction homme-machine. Les concepteurs de Siri avaient sans aucun doute compris la

[46]Jacques COSNIER, Synchronisation et copilotage de l'interaction conversationnelle. Protée, 33-39, 1992. Laboratoire d'Ethologie des Communications Université Lumière Lyon2 (FR).

valeur importante qu'accorde chacun des locuteurs dans la prise en compte de leur parole.

Les ondes de couleurs sont donc des marqueurs importants qui permettent à l'utilisateur de se sentir en conversation. Elles participent à la construction de la situation d'interaction. Chez les interactionnistes, c'est justement ces signes de « compréhension » et d'« échange » qui permettent de dire qu'il y a interaction. Nous pouvons donc dire que ces ondes participent à la mise en place d'une situation d'interaction entre l'homme et Siri.

D'autre part, et contrairement à d'autres applications mobiles, Siri ne propose pas d'accès au clavier pour écrire des requêtes. Cet élément nous fait dire une nouvelle fois que les concepteurs ont voulu mettre en avant le mode d'interaction orale, contre celui de l'écrit.

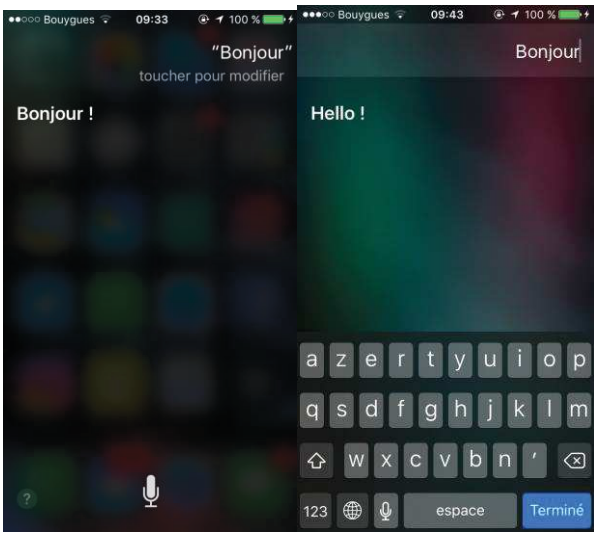
Voici différentes interfaces que l'on rencontre dans des applications mobiles dédiées aux échanges conversationnels :

Nous observons que l'alternative oral/écrit est toujours possible dans les différentes applications de messageries instantanées alors même qu'elles sont

destinées à des interactions homme-homme (quoi que depuis peu, nous pouvons dialoguer à l'écrit avec des machines, appelées *chatbots*). Le clavier est toujours présent et des petits pictogrammes de microphones permettent aux utilisateurs de s'exprimer à l'oral. Or Siri ne propose aucune solution qui permette d'écrire des requêtes. Aucun clavier n'est accessible pendant la requête. Après réponse du logiciel, le seul mode d'action possible est l'oralité, comme nous pouvons le voir ci-dessous, avec l'icône de la voix enregistrée : le microphone.

Pour autant, depuis peu, les concepteurs ont pris en compte les erreurs de retranscription du logiciel Siri et permettent désormais d'apporter des modifications écrites à la requête orale (« Toucher pour modifier »).

Mais nous noterons toutefois que l'apparition du mode d'action écrit n'est possible seulement après avoir fait une requête orale, ce qui place l'oralité comme principal mode d'action dans le dispositif Siri, une fois encore.



Quand l'écriture de la machine conditionne l'oralité humaine…

Nous voudrions nous arrêter pour cette première partie de l'analyse des représentations du logiciel sur un paradoxe du moins surprenant. Alors que l'oralité semble être le mode d'action utilisateur prioritaire et suggéré/imposé

par le dispositif, Siri ne recourt à l'oralité à aucun moment de la première phase de relation avec l'interface. Les phrases d'introduction « Que puis-je faire pour vous ? » et « Allez-y, je vous écoute… » ne sont jamais « dites » oralement par la machine. L'écriture prédomine. Deux explications peuvent être apportées pour comprendre ce choix de la part des concepteurs du logiciel.

D'une part, la place qu'occupe Siri dans la relation avec l'utilisateur/humain est symboliquement forte. Elle se traduit par la « figure du serviteur », pour reprendre les termes de Clotilde Chevet. Et d'ajouter : *« L'idée de soumission est essentielle dans le principe d' « assistant personnel ». Le slogan de l'application Siri n'est autre que « Vos désirs sont des ordres » et l'assistant lui-même dit être « l'humble serviteur » de l'utilisateur. »*[47]Ainsi, une relation hiérarchique, de pouvoir même, est établie avant même que l'échange n'ait pu commencer. La voix qui compte, c'est celle de l'utilisateur, qui est maître de la situation. Il peut en effet, quitter l'application sans rendre de comptes à Siri (chose moralement impensable en situation d'échange entre deux humains). De plus, la machine est au service de l'utilisateur, elle « attend » qu'il parle, qu'il lui donne une requête à traiter. Siri ne demande pas à l'utilisateur de parler, c'est l'utilisateur qui décide, lui seul, de parler ou non. L'oralité de l'utilisateur est en réalité une oralité totalement performative (précisément perlocutoire) dans la mesure où elle se traduira par des actes de la part de la machine (traitement et exécution de la requête).

L'utilisation d'une part des écrits d'écran par la machine et la place hiérarchiquement supérieure de l'oralité de l'utilisateur vis à vis de la machine renforce donc cette idée du serviteur à disposition de l'utilisateur.

D'autre part, si l'écriture est préférée à l'oralité dans cette phase d'utilisation de Siri, c'est certainement d'un point de vue pratique. En effet, si l'utilisateur ouvrait

---

[47] Clotilde Chevet, La part de l'imaginaire dans « l'interaction homme-machine » : le cas des assistants personnels numériques, Mémoire de recherche Master 2 au CELSA Sous la direction du Professeur Karine Berthelot-Guiet, p. 37

Siri et que la machine « parlait », alors même que l'objectif de l'utilisateur est de demander quelque chose à celle-ci, alors les risques de cacophonie, ou du moins de rencontres des différentes sources orales, seraient grands. Ils ne serviraient pas le bon fonctionnement de l'application et ils iraient à l'encontre de la promesse faite par les concepteurs (la figure du serviteur à disposition totale de l'utilisateur).

De plus, à quoi servirait de faire « parler » Siri lors de son ouverture et de lui faire dire qu'il nous « écoute » (« Allez-y, je vous écoute… ») alors même qu'il est destiné à cela. Nous pourrions presque dire que c'est dans sa nature (s'il n'était pas une machine) ; nous ne lui demandons finalement que cela.

Enfin, lorsque l'utilisateur s'apprête à ouvrir l'application Siri, il a déjà connaissance du « contrat de communication »[48] qui va s'établir durant l'interaction future avec la machine. Il sait que Siri est à son service et qu'il peut lui demander n'importe quel type de requête. Les rapports de forces, de pouvoir et de communication sont co-construits mais généralement connus des utilisateurs en amont de l'utilisation de l'application, ce qui permet une fluidité d'usage, sans chichis.

L'utilisation du langage, conditionné par le type de question

Nous allons désormais conclure cette sous-partie en analysant la suite du processus qui est donné à voir à l'utilisateur, après qu'il a ouvert l'application.

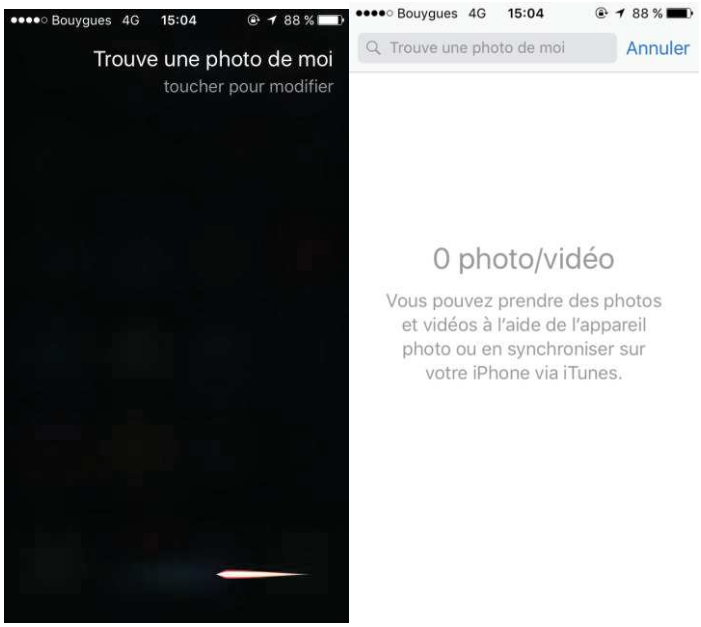
Nous allons prendre trois cas de figure concret d'utilisation qui pourront faire office de typologie. Ces situations correspondent, à peu de choses près, aux

---

[48] Terme que l'on doit à Patrick Charaudeau. Il contextualise l'approche du contrat de communication de la manière suivante : *« Tout acte de communication s'inscrit dans un cadre pré-structuré dont les particularités dépendent de la situation dans laquelle se déroule l'acte langagier. Ainsi, chaque situation se définissant par une certaine finalité (« parler pour atteindre quel but ? »), une certaine identité des partenaires de la communication (« qui s'adresse à qui ? »), un certain propos (« parler à propos de quoi ? »), un certain dispositif(« parler dans quelles circonstances matérielles ? »), il faut que ces particularités soient reconnues par les partenaires qui y sont impliqués, faute de quoi la communication échouerait »* in Ce que communiquer veut dire, Revue des Sciences humaines, n°51, Juin, 1995

différentes requêtes que nous avons étudiées dans la première partie (*local problem* et *global problem*).

Dans le premier cas (*local problem*), l'utilisateur demande à Siri de trouver des photos de lui-même. Après avoir ouvert l'application et lu les messages « Que puis-je faire pour vous ? » et « Allez-y je vous écoute… », il dit : « Trouve une photo de moi. »
Siri se met alors à traiter l'information et semble trouver une solution « en local ». Voici les deux écrans auxquels l'utilisateur fait face, successivement :



Durant la première étape de l'utilisation, la retranscription de la requête utilisateur s'effectue mot à mot (« Trouve », « une », « photo », « de », « moi »). Cette retranscription du discours utilisateur par mot correspond aux étapes de numérisation, division et reconnaissance vocale du logiciel, vue en début de partie (processus technique). L'information traitée, le logiciel s'exécute et

redirige l'utilisateur vers l'application iPhone « Photos » (deuxième écran) pour tenter de lui donner un résultat (ici, sans succès).

Plusieurs éléments sont à noter pour cet exemple. D'une part, Siri joue ici le rôle d'un simple exécutant puisqu'aucun son n'est émis de la part du logiciel. La réponse constitue en une réorientation directe de l'utilisateur vers une application censée donner la réponse à la requête utilisateur.

Il faut aussi noter l'absence de réponse « écrite » de la part de Siri. Le logiciel réoriente en fait directement vers le « lieu » de réponse (l'application « Photos »), sans commentaires écrits.

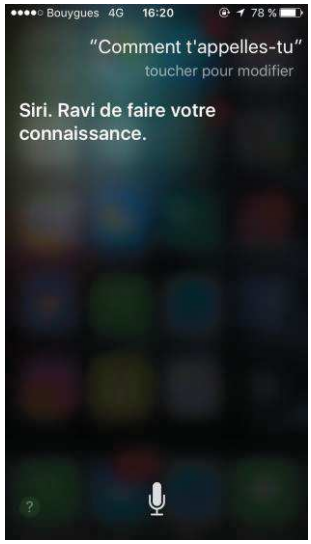
Avec une telle absence de langage de la part de Siri, ce que nous dit, a priori, cet exemple, c'est que l'objectif de Siri ne réside pas dans la construction d'une relation conversationnelle avec l'utilisateur mais il réside plutôt dans la « livraison » d'une réponse efficace, pertinente et rapide.

Dans le second cas (*global problem*), l'utilisateur demande à Siri de trouver des informations sur une célébrité mondialement connue. Il dit : « Qui est Obama ? ».

Lorsque le résultat est trouvé, les informations sous forme écrite préexistent à l'oralité, puis Siri « dit » : « Voici quelques informations : ».

Plusieurs éléments sont à noter. Tout d'abord, le fait que le résultat écrit soit accessible avant que Siri ne « parle » donne une idée de la position que les concepteurs ont voulu attribuer à la machine. Comme nous l'avons déjà vu, la machine joue le rôle de serviteur. Mais plus encore, Siri s'efface ici pour mettre en avant des résultats de recherche sous forme écrite qui proviennent d'une autre source que lui-même, ne donnant ainsi aucun espoir à une éventuelle construction de relation conversationnelle homme-machine. En effet, ce qui est optimisé ici, ce n'est pas la relation mais bien la pertinence des informations données. Aucune énonciation en relation avec l'information des résultats de recherche n'a lieu.
D'autre part, la phrase « Voici quelques informations : » est donnée à entendre et à lire à l'utilisateur comme le signe d'une réussite. Ce que veut dire Siri c'est « Tenez, j'ai trouvé des informations pertinentes, en un temps record et qui correspondent exactement à ce que vous recherchez ».

Il existe un troisième et dernier cas, assez courant. Il s'agit des situations où Siri donne une réponse sans l'intervention d'une application tierce, ni recherche d'informations sur internet. Le logiciel fait simplement appel à la base de données des mots/phrases enregistrées par les voix humaines de Siri.
Jusqu'à présent, nous avons étudié des cas dans lesquels Siri ne donnait pas de réponse écrite et orale en même temps (soit écrit, soit oral, soit aucun des deux). En effet, la redirection vers une application tierce et l'utilisation des résultats de recherche d'informations sur Internet constituaient jusqu'ici les seules formes de réponses de la part du logiciel.



Abordons donc le dernier cas de la typologie de réponse du logiciel. Dans l'image présentée ci-dessus, on découvre la question posée par l'utilisateur (« Comment t'appelles-tu »), ainsi que la réponse du logiciel (« Siri. Ravi de faire votre connaissance. »). Cette réponse est donnée à lire et à entendre, en même temps.

Au delà de la quantité de réponses que l'utilisateur peut obtenir en posant la même question, comme l'avait relevé Clotilde Chevet dans son étude sur les assistants personnels numériques[49], nous observons un élément particulier lié au type de question posée : seules les questions portant sur le logiciel lui-même, bénéficient de réponses orale et écrite, qui sont en réalités des réponses préparées à l'avance par les concepteurs. Les autres questions, nécessitant une recherche d'informations sur Internet ou faisant intervenir une application tierce ne bénéficient pas de réponses avec une oralité et écriture aussi complète.

On comprend ici à quel point le système Siri est verrouillé. Siri permet en effet de « répondre à tous nos désirs », mais seulement dans le cadre donné par les concepteurs. Cette maîtrise de l'expérience utilisateur, que l'on appelle méthode en « jardin clos », vaut également dans le champ du langage utilisé par le

[49]Clotilde Chevet, La part de l'imaginaire dans « l'interaction homme-machine » : le cas des assistants personnels numériques, Mémoire de recherche Master 2 au CELSA Sous la direction du Professeur Karine Berthelot-Guiet, p. 48 (« Définition des interlocuteurs »)

concepteur de l'assistant personnel. Le discours oral et écrit porté par la firme de Cupertino via l'interface est totalement sous contrôle.

Il est d'ailleurs impossible de demander à Siri de rechercher des informations le concernant via Internet :

Pour récapituler l'ensemble de ces cas, nous avons réalisé un tableau qui permet de garder en tête les différentes typologies d'interactions possibles et comment chacune d'entre elles fait intervenir le langage écrit et/ou oral.

Nature de la requête	Local	Global	Intern
Exemple de requête	« Trouve une photo de moi »	« Qui est Obama ? »	« Qui es-tu ? »
Traitement de la requête	Lancement d'une recherche d'une application tierce	Lancement d'une recherche internet	Lancement d'une recherche dans la base de données (Apple)

Réponse logiciel	Redirection de l'utilisateur vers l'application tierce	Phrase d'introduction + Affichage des réponses liées à la recherche Internet	Affichage de la réponse complète.
Oral/écrit	Aucune utilisation de l'oralité ou de l'écriture. L'affichage de l'application tierce fait office de réponse.	Utilisation limitée de l'oralité et de l'écriture : « Voici quelques informations »	Écriture et oralité sont utilisées par le logiciel pendant toute la durée de la réponse à la requête.

Qu'il s'agisse de l'analyse du processus technique ou de celle de l'interface utilisateur, l'oralité et l'écriture sont en constante interrelation dans le dispositif Siri. Mais plus encore, leur analyse nous a permis de tirer des conclusions intéressantes sur le fonctionnement de cette application.

L'interaction est techniquement rendue possible par un enchaînement de processus écrits et oraux qui permettent de donner lieu à un échange entre l'utilisateur et Siri. Techniquement, tout est fait pour que les échanges soient fluides et proches du genre d'interaction que deux hommes pourraient avoir ensemble lorsqu'ils discutent, par exemple.

Côté interface, beaucoup d'efforts sont consacrés à la création d'une interface qui représente des interactions fluides et proches de celles que pourraient avoir deux hommes, mais la réalité est toute autre. En effet, de nombreux éléments viennent prouver que l'interaction homme-machine (Siri) trouve sa force dans la capacité de chacun des utilisateurs à y croire.

Premièrement, les rapports hiérarchiques entre l'homme et la machine, imposés par les concepteurs via des signes langagiers oraux et écrits conditionnent l'échange entre les deux inter-actants ce qui ne favorise pas une situation dite

d'interaction. Siri, soumis, s'efface trop souvent pour laisser place à de véritables échanges conversationnels.

Deuxièmement, la course à l'efficacité et la performance du résultat, caractérisée par des exécutions de requêtes rapides et sans commentaires oraux ou écrits de la part de l'application ne permettent pas non plus de développer ce dialogue oral spontané que les utilisateurs pourraient attendre. Mais nuançons ce propos, car si Siri a été fabriqué comme cela, c'est certainement car il correspond à un besoin de servitude de la part des utilisateurs.

Enfin, le discours verrouillé de Siri (lorsque les questions concernent le logiciel lui-même) ne permet pas d'aboutir à un véritable échange conversationnel, comme nous l'avons vu.

Concluons en disant que dans l'utilisation de Siri, tout est fait pour que l'échange conversationnel ait lieu entre l'homme et la machine mais que par les contraintes qu'imposent les enjeux techniques, économiques et symboliques (oral/écrits) du dispositif, un doute persiste sur la possibilité de vivre une expérience dialogique qui pourrait se rapprocher de celle qu'on connaît entre deux hommes. Il va sans dire que de nombreux imaginaires circulent au sujet des éventuels types de relations qui pourraient exister avec les utilisateurs de Siri et qu'ils participent à des croyances selon lesquelles l'homme peut communiquer avec la machine comme il le fait avec l'homme. Le projet utopique du dialogue oral spontané est encore loin d'être réalisé.

Enfin, en ce qui concerne nos observations sur l'écriture et l'oralité dans le dispositif, nous pouvons résumer en disant que le langage participe à ce sentiment d'interaction conversationnelle avec la machine mais que d'un autre côté, la présence d'une interface, d'un écran et de réponses écrites ne favorise pas, ni ne naturalise l'échange conversationnel. Néanmoins la complémentarité de l'oralité et de l'écriture est nécessaire au bon fonctionnement du dispositif (exécution des requêtes, recherches et affichage d'informations…).

II. Interface de Siri : quel cadre pour l'échange homme-machine ?

Dans cette deuxième partie, nous désirons nous intéresser aux cadres de l'échange homme-machine en analysant l'interface plus en profondeur. Nous avons déjà analysé dans la première partie comment la situation d'interaction était rendue possible d'un point de vue technique mais également comment cette interaction se donnait à voir aux utilisateurs. Nous voulons ici rentrer en détail sur les cadres qu'imposent ou plutôt tentent d'imposer écriture et oralité en interrelation dans la situation d'interaction. Nous testerons donc notre deuxième grande hypothèse selon laquelle l'assistant personnel numérique offre un cadre écrit et oral à « l'interaction homme-machine » et essaierons de comprendre quelles formes d'interaction sont produites via l'utilisation du langage et comment elles sont données à voir à l'utilisateur.

A. Les cadres instituants du logiciel Siri

Comme tout média informatisé, Siri relève de la théorie générale des cadres. Comme le font savoir Emmanuël Souchier et Yves Jeanneret dans un de leur article sur l'écriture numérique, *« l'affichage d'un écrit à l'écran nécessite en effet la mise en abyme d'une série d'au moins quatre cadres successifs »*[50]. Leur travail réalisé pour l'exemple des écrits d'écran d'ordinateur, s'applique également pour notre objet d'étude Siri. Nous désirons donc faire une observation des différents

[50] Emmanuël Souchier & Yves Jeanneret, « Écriture numérique ou médias informatisés ? », Du signe à l'écriture, Pour la Science — Scientific American, Dossier 33, 2002

cadres qui se succèdent et en tirer des enseignements sur la manière dont ces cadres orientent les échanges entre Siri et l'utilisateur.

Le premier cadre, est le cadre « matériel ». Dans le cas de Siri, il s'agit du Smartphone ou de la tablette de la marque Apple. Cette simple observation est plus intéressante qu'il n'y paraît puisqu'elle permet de replacer dans son contexte l'ensemble des analyses portées jusqu'ici, à savoir que le logiciel (ou l'application) Siri ne « vit » pas seul et lorsque le Smartphone est éteint, le lieu d'interaction n'existe plus. Ce Smartphone est le matériel-support qui accueille tout un ensemble d'écrits d'écran et qui les délimite matériellement (limite matérielle de l'écran et limite énergétique, car sans batterie, sans câble de recharge, pas d'interaction possible) mais aussi symboliquement (dans la mesure où le rapport qui se met en place entre le Smartphone et l'utilisateur est un rapport médiatisé, voire spectacularisé notamment par l'action de lecture, pour reprendre le terme d'Emmanuel Souchier[51]).

Le second cadre correspond à ce qu'on appelle le cadre « système ». Dans notre cas, il s'agit en fait du système iOS d'Apple, qui s'allume lorsque le Smartphone se met en marche et nous offre une possibilité d'action importante puisque c'est ce

[51] Emmanuel Souchier, L'écrit d'écran, pratiques d'écriture & informatique, Communication et Langages n°107, p.115

système qui nous permet ensuite d'accéder à de multiples applications via la mise en avant d'icônes. Peu de commentaires à faire sur ce deuxième cadre étant donnée que l'objet d'étude qui nous intéresse est Siri, pas le système iOS d'Apple. Néanmoins, notons tout de même que ce cadre système est un architexte qui permet *« l'existence de l'écrit à l'écran »* et qui permet *« l'exécution et la réalisation »*[52] de tâches par l'utilisateur.

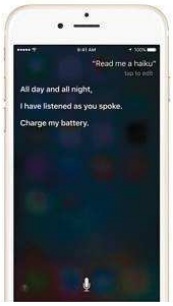
Le troisième cadre, correspond à celui du « logiciel », autrement dit Siri. L'utilisateur qui « entre » dans l'application en « demandant » son ouverture au Smartphone (via le maintien enfoncé du bouton central), bénéficie d'un ensemble de possibilité même si celles-ci sont délimitées par le fait que l'application-logiciel Siri est dédiée à un usage précis (requête vocale).

Quelques remarques concernant le cadre logiciel : lors de l'utilisation de Siri, impossible pour l'utilisateur de passer en mode horizontal (« paysage »), alors même que la fonction « verrouiller rotation » du système est désactivée. Ce choix des concepteurs correspond à la volonté de placer le logiciel comme étant un interlocuteur, au même titre qu'un humain. Cela se justifie par la culture d'usage du papier, qui en format « portrait » (vs. format « paysage ») refait surface ici. En effet, l'histoire de l'utilisation du papier en format « portrait » servait la plupart du temps à dessiner des visages (sauf dans le cas des portraits symbolistes qui sont dans des formats carrés ou paysage). Autrement dit, le cadre logiciel, est donné à voir à l'utilisateur comme étant un interlocuteur humain qui se tient devant lui (debout ?), à qui il peut s'adresser oralement.

---

[52] Emmanuël Souchier, Mémoires – outils – langages Vers une « société du texte », Communication et Langage n°139, Avril 2004, p.44

Autre remarque : la présence des ondes de couleurs dans ce cadre logiciel signifie que cet espace est paradoxalement celui de l'oralité pour l'utilisateur (qui peut dicter sa requête) alors même que l'écriture est le seul langage techniquement possible dans ce cadre, puisqu'il s'agit d'un écran.



Enfin, le dernier cadre, appelé « document », définit symboliquement l'espace de conversation, d'échange (écrits et oraux) et plus généralement d'interaction dans le logiciel Siri. Mais cet espace est surtout celui du texte à l'écran. C'est à partir de ce cadre que « l'écriture » et la lecture sont rendues possibles par le système, pour l'utilisateur. Ce cadre est « enveloppé » par tous les autres et concentrent l'ensemble des processus d'écritures et d'oralité vue en première partie de ce mémoire. C'est sur ce cadre en particulier que l'étude de la deuxième partie de ce mémoire s'arrêtera. Mais les autres cadres présentés ici nous serviront nécessairement dans notre analyse puisqu'ils participent eux aussi, d'une certaine manière à la mise en place de la situation d'interaction homme-machine.

B. L'interface : lieu d'interaction par excellence?

Siri est le théâtre d'une rencontre improbable entre un utilisateur lambda, qui provient de n'importe quel endroit et l'entreprise conceptrice et productrice du logiciel en question.

Le paradoxe de cette rencontre entre l'homme et la machine réside dans le fait que Siri est à la fois un inter-actant dans le dispositif de communication homme-machine mais également producteur (c'est le logiciel qui écrit et donne à lire les

échanges des différents protagonistes de la situation), support et média (c'est Siri qui permet à l'utilisateur de lire, écouter, voir. Via l'écran, c'est tout un ensemble de signes qui sont donnés à voir ; l'interface vocale vient compléter cette fonction support/média).

Nous chercherons donc dans cette partie à répondre à plusieurs questions qui concernent les formes d'interactions qui sont produites par l'interrelation écriture-oralité. Autrement dit, nous chercherons à comprendre comment le dispositif offre un cadre à l'interaction et comment elle engage, ou tente d'engager l'utilisateur à travers la parole, le dialogue écrit et les signes produits pendant l'utilisation de l'application.

1. Lieu de rencontres de deux paroles avec deux origines et deux natures

Parole de l'entreprise

La parole de l'entreprise ne provient que d'une seule source. Un nombre d'étape important a été nécessaire pour que se dégage une voix depuis cette application (comme nous l'avions expliqué dans la partie I. de ce mémoire). Mais plus encore, nous avons affaire à une application dont les sources de production sont hybrides : d'une part, le concepteur du logiciel Siri, d'autre part le producteur du support et enfin, la voix, qui provient d'une autre source. Mais ce qu'il faut noter dans cette remarque, c'est que l'utilisateur, reçoit cette parole comme si elle provenait d'une seule source, voire qu'elle provenait du dispositif lui-même. Le fait que l'ensemble de ces sources de productions soit réuni dans un tel support laisse entendre que le dispositif est comme tombé du ciel. Il participe à la création d'un imaginaire selon lequel la machine parlante est *une*. La voix de Siri pose en fait des questions fondamentales de rapport au temps, à l'espace et au corps. Marie-Pascale Huglo s'intéresse justement aux changement de perceptions et à la transformation de nos imaginaires face à la voix, suite à l'arrivée des technologies : *« S'intéresser à ces phénomènes, c'est prendre la*

mesure critique de leur impact sur le corps (détachement et extériorité de la voix, diversification des corps sonores, humains ou inhumains), sur la temporalité (la voix enregistrée est à la fois présence et archive, elle est « anachronique »), sur la spatialité (j'entends dans le creux de l'oreille la voix d'un absent), sur la communication (réseaux, effacement de la distance, « accessibilité » [terme bien contemporain] en tout temps et en tout lieu [public ou privé]), sur la mémoire (la voix de Malraux, c'est une diction, une intonation, une déclamation, de la « friture» aussi, qui datent). »[53]

Mais qu'en est-il de la parole de l'entreprise en situation ? Comment est-elle donnée à entendre et à voir à l'utilisateur ?

Le fait de poser la question de la sorte (comment une parole peut être donnée à voir ?) nous rappelle à quel point l'énonciation de la machine est hybride. En effet, en même temps qu'elle donne à lire la réponse à une requête sur l'écran, elle l'énonce oralement, via les enceintes du téléphone.

Ce qu'on peut d'ores et déjà dire, c'est que la voix de Siri, même si elle provient d'un enregistrement de voix humaine, elle ne s'apparente pas ou peu à une voix « naturelle ». La synthétisation de la voix est inéluctable dans la mesure où celle-ci *« oscille toujours entre intériorité et extériorité, entre émotion/sensation et sens, entre chant et parole, entre corps et langage »*[54] et que Siri, en tant que programme numérique ne dispose pas de « corps ». Une citation de Roland Barthes dans *Le grain de la voix* résume bien l'idée que l'on souhaite montrer ici : *« Ce qui se perd dans la transcription, c'est tout simplement le corps »*[55]. Roland Barthes parle évidemment ici de transcription écrite. Il veut dire qu'à partir du moment où la voix est (re)transcrite à l'écrit, elle perd son corps. Mais dans notre cas, nous avons vu que la voix était également écrite (enregistrement d'une voix humaine) mais surtout retranscrite lors de la réponse à une requête. Le dispositif

---

[53] Présentation, Marie-Pascale Huglo, Études françaises, vol. 39, n°1, 2003, p.7-11, URL : http://id.erudit.org/iderudit/006737ar, p. 9

[54] Ibid p.7

[55] Le grain de la voix Entretiens (1962-1980), Roland Barthes, Seuil

Siri, est finalement ce corps vide qui fait appel à une voix de source extérieure à lui-même pour simuler un discours dont le corps n'est autre que le cadre matériel du dispositif, à savoir l'objet Smartphone. Si l'utilisateur pense avoir affaire à un *unique* et *vrai* interlocuteur, il se trompe, mais d'autres éléments (paradoxalement de l'ordre de l'écrit) sont pourtant là pour prolonger cet imaginaire de la machine parlante.

Dans le système Siri, la parole de l'entreprise n'est pas seulement une parole orale. L'écran et la mise en écran de l'oralité des interlocuteurs sont donnés à lire via du texte qui s'affiche en même temps (en réalité, juste avant) que Siri donne la réponse oralement.

Ce que l'on peut dire de la retranscription écrite de la parole de Siri, c'est qu'elle s'apparente à une écriture humaine, c'est à dire qu'elle respecte les règles de ponctuation, l'orthographe et la syntaxe (voir l'exemple ci-dessus). Mais nul besoin de beaucoup de temps pour saisir que la parole de l'entreprise est avant tout une parole écrite en amont pas des hommes ! En effet, les réponses données par Siri, lorsque l'utilisateur pose des questions qui nécessitent une réponse au minimum élaborée, sont des textes qui ont été écrits par les concepteurs, en amont. En revanche, l'oralité associée à cette même réponse est loin de s'apparenter à une voix humaine, comme nous l'avons précisé plus haut. La voix

de Siri est une voix robotique, sans intonation. Il existe donc une dichotomie forte entre les discours oraux et écrits de Siri, qui lui donne une certaine image non-cohérente dans le sens où les réponses écrites sont finalement des réponses pré-écrites par des humains, alors que les réponses orales ne disposent pas de ce grain qui nous permettrait d'effacer cette frontière entre l'homme et la machine d'un point de vue du discours.

Néanmoins, c'est par ces tentatives écrites et orales qui reprennent dans une certaine mesure les codes langagiers de l'homme que Siri tente d'engager avec l'utilisateur une rencontre dialogique qui pourrait s'apparenter à celle de deux humains. Tout se passe en fait comme si la machine tentait d'humaniser son discours pour effacer cette barrière langagière entre elle et l'homme à qui elle « s'adresse ». Ces effets de simulation participent grandement à la croyance selon laquelle le dialogue homme-machine pourra être un jour naturel et spontané.

Parole de l'utilisateur

La parole de l'utilisateur est spontanée et naturelle. Elle est entièrement orale, contrairement à celle de Siri, qui est écrite en amont par les concepteurs.

Le questionnement de la place de la parole de l'utilisateur est intéressant puisque, dès que la requête est réalisée, la parole de l'utilisateur ne lui appartient plus. Elle est en quelque sorte confiée à Siri qui se charge de la traiter, comme bon lui semble, avec un objectif : répondre à la requête utilisateur de la manière la plus pertinente possible. Une analyse d'Emmanuel Souchier sur l'externalisation de la mémoire de l'homme par l'outil est intéressante à garder en tête pour expliquer ce phénomène de dépossession : *« L'extériorisation de l'outil [pour notre cas sur Siri, nous pourrions remplacer outil par langage] entraîne un véritable phénomène de dépossession individuel».*[56] Cette remarque vaut également dans le cadre de la retranscription écrite de la parole de l'utilisateur, qui est affichée sur l'écran, sans ponctuation (nous reviendrons sur

[56] Emmanuël Souchier, Mémoires – outils – langages Vers une « société du texte », Communication et Langage n°139, Avril 2004, p.45

ce détail plus tard) et qui participe du processus de dépossession de la parole utilisateur. Siri devient alors cette « boîte noire » dans laquelle l'utilisateur ne peut rien voir, savoir ni comprendre, comme l'explique Emmanuel Souchier à propos de l'écriture informatique (qui vaut également pour notre cas) : *« Le scripteur ne maîtrise désormais plus le parcours intégral de l'écrit. Les opérations, outils et processus placés entre la main et l'écran créent une rupture et interdisent la compréhension globale de l'acte d'écriture : espace du rêve, du fantasme ou de la domination, la boîte noire placée entre l'homme et son écrit est un espace du secret »*[57].

La voix de l'utilisateur se transforme, comme nous l'avions vu en partie I., en écriture pour permettre au logiciel le traitement de la requête. Autrement dit une des composantes de l'activité utilisateur pendant la rencontre des deux paroles est celle de lecteur, comme l'avait déjà remarqué Clotilde Chevet[58]. En effet, en même temps qu'il dicte une requête à Siri, l'utilisateur devient lecteur de son propre discours. Dans la deuxième phase d'échange, Siri donne à entendre et à lire une réponse à l'utilisateur. Cette activité de lecture de la part de l'utilisateur est très bien expliquée par Emmanuel Souchier qui parle de « scène double », à la fois spectacle et secret dans le cadre des écrits d'écran : *« La relation sociale à l'écrit est sans doute marquée par la dualité de l'écrire et du « donner à lire ». Il n'y a paradoxalement pas de pratique plus discrète et silencieuse que celle de l'écrit qui s'oppose aux arts d'exposition et à ceux du spectacle, mais il n'y a sans doute pas plus intimement exhibitionniste que celui qui se donne ainsi à lire. La réelle distinction de l'écrit repose sur un fragile équilibre entre l'élaboration*

---

[57] Emmanuel Souchier, L'écrit d'écran, pratiques d'écriture & informatique, Communication et Langages n°107, p.111

[58] Clotilde Chevet, La part de l'imaginaire dans « l'interaction homme-machine » : le cas des assistants personnels numériques, Mémoire de recherche Master 2 au CELSA Sous la direction du Professeur Karine Berthelot-Guiet, p. 11

et la lecture, exhibition qui suppose une absence du créateur au cours de la lecture. »[59]

Or ce qui caractérise cette action de lecture spectacularisée c'est justement la mise au second plan de l'utilisateur. En effet, celui-ci s'efface pour laisser place à une « scène performante »[60] dans laquelle il est spectateur non-omniscient (puisque le dispositif technique n'est pas visible). L'utilisateur est face à une double abstraction : celle du *« texte et de l'écriture informatique. Le lecteur ne perçoit en effet qu'une page-écran, l'avant et l'après n'ont pas pour lui d'existence sensible. Le texte qui suit et précède sont proprement illisibles. Absents de la scène visible, codés par et pour la machine, ils n'existent que virtuellement. L'écrit accède ainsi à un nouveau degré d'abstraction ».*[61]

Finalement l'utilisateur se trouve à l'origine de l'échange (dicte la requête à l'oral) et à la fin de celui-ci (lit et entend la réponse à sa requête), mais entre les deux, l'oralité de l'utilisateur est complètement prise en charge par la machine, avec une « *étape de traduction entre le canal et le récepteur, durant laquelle l'information peut être perdue et jamais regagnée »*[62].

[59] Emmanuel Souchier, L'écrit d'écran, pratiques d'écriture & informatique, Communication et Langages n°107, p.117

[60] Ibid

[61] Ibid, p.111

[62] Norbert Wiener, Cybernétique et société, Éditions du Seuil 2014, p. 108

D'autre part, comme nous l'avons dit pour la machine, l'homme, à l'inverse voit sa parole « machinisée » dans le sens où la retranscription écrite de sa parole se fait de manière à ce que la machine puisse interpréter le discours émis (comme nous l'expliquions dans la partie I. de ce mémoire). La retranscription se fait mot à mot, sans ponctuation. La qualité et l'authenticité de la retranscription orale de l'utilisateur est délaissée au profit de l'efficacité et la réalisation de l'échange (n'est-ce pas d'ailleurs utopique que de penser que la machine puisse interpréter l'oralité humaine avec exactitude, avec le bon sens ? Quand on connaît déjà les difficultés de l'utilisation du langage entre deux êtres humains… Cela l'est d'ailleurs certainement tout autant quand on décide de faire passer l'oralité à l'écrit, qui plus est un écrit numérique, traité informatiquement).

Dans cette partie nous avons étudié les paroles de chacun des inter-actants de la situation de dialogue entre l'utilisateur et Siri. Notre analyse nous conduit à une conclusion paradoxale de la rencontre de ces deux paroles. D'une part, les deux paroles, pour vivre ensemble se conforment aux normes langagières de l'autre inter-actant (la parole de la machine est un emprunt de voix humaine, alors que la parole de l'utilisateur est délaissée à la machine et traitée comme une « voix » (en réalité un langage écrit) interprétable par la machine). D'autre part, le paradoxe se trouve également dans la relation écriture-oralité, elle-même. Alors que la promesse même de Siri est de proposer un dialogue oral spontané à l'utilisateur, le recours à l'écrit d'écran se fait à différents niveaux : la retranscription écrite de la parole humaine, la réponse orale et écrite affichée à l'écran, etc. Tous ces écrits d'écran participent en fait, au delà de l'utilisation de l'oral et de l'écrit ensemble, à ce qu'on pourrait appeler la mise en scène de l'interaction homme-machine. L'écran est finalement le lieu de l'événement dans lequel l'interaction homme-machine est rendue lisible et visible. Il participe à l'élaboration de la croyance selon laquelle l'utilisateur et Siri dialogue.

2. Lieu de dialogue ? L'écran comme lieu d'interaction

Nous l'avons vu à l'instant, l'écrit joue finalement un rôle très important pour un service d'assistant personnel vocal comme Siri. L'écran est au cœur de la rencontre entre l'homme et la machine et les écrits qu'il donne à lire participe à la mise en scène de l'interaction. Nous pourrions presque dire que c'est l'écrit, par son arrivée à l'écran depuis un lieu inconnu et non visible, qui justifie ou qui incite l'utilisateur à se dire« ceci est une interaction entre l'homme et la machine ».
Mais qu'en est-il des marques énonciatives qui participent à la mise en place du cadre « document » ? Comment l'interface graphique donne un cadre à l'interaction homme-machine ? Quels signes sont à l'œuvre pour créer cette situation d'interaction ? Nous nous arrêterons dans cette partie sur l'ensemble des éléments qui favorisent la mise en place du cadre et qui tendent à créer une situation dialogique entre l'homme et la machine.

Nous avons déjà analysé les premiers écrans de Siri, ceux qui sont donnée à voir et à lire lorsque l'application est lancée (Que puis-je faire pour vous ? et Allez-y je vous écoute…). Ils sont déjà des invitations au dialogue pour l'utilisateur qui sait directement de quelle manière il va pouvoir utiliser le logiciel d'assistant vocal.

Que se passe-t-il une fois que l'utilisateur « dialogue » avec la machine ?
Voici le type d'écran auquel l'utilisateur fait face :

Plusieurs éléments participent à la mise en scène de l'interaction orale, pour ne pas dire du dialogue.

L'élément qui est en bas, centré, représente un microphone dont le signifié indique que l'utilisateur peut parler à tout moment. Siri est donc disponible pour dialoguer à chaque moment.

Le point d'interrogation en bas à gauche donne l'accès à toutes sortes de question qui peuvent être posées par Siri. Sur cet écran, on peut voir que les

différentes questions ont été choisies pour orienter l'utilisateur vers des usages précis de Siri. On pourra par exemple demander à Siri de nous mettre en « communication » avec l'extérieur (envoi de sms, appel…) ou lui demander des « informations » relatives à l'actualité par exemple ou dernière catégorie lui demander des « services » (comme par exemple trouver un restaurant italien). Ces catégories d'usages proposées par le dispositif orientent la manière avec laquelle l'utilisateur va poser ses questions. Le concepteur et l'application technique du logiciel dictent alors en partie la conduite, même si l'utilisateur est toujours libre de faire la requête qu'il veut. On peut aussi imaginer que le concepteur propose des suggestions de questions en fonction de la capacité de Siri à faire un tâche (il paraîtrait absurde que les concepteurs proposent une suggestion de question pour laquelle Siri ne peut pas répondre).

Arrêtons-nous maintenant sur les espaces d'énonciation. Il convient de faire un rapide historique des présentations d'interface de dialogues entre l'homme et la machine Siri pour comprendre ce qu'elles disent aujourd'hui des situations d'interaction qui se déroulent.

A gauche, une représentation de Siri en 2011, à droite la mise à jour de la représentation des espaces de dialogue, en 2013. On remarque rapidement l'évolution du design et la manière dont la représentation du dialogue entre l'utilisateur et Siri a évolué. Au départ, la représentation des échanges oraux était

enfermée dans des bulles (écrites), qui reprenaient l'imaginaire de la bande dessinée et du discours rapporté. La parole de l'utilisateur est en gras, entre guillemet. Clotilde Chevet expliquait de manière claire comment ces signes n'étaient pas des choix anodins de la part des concepteurs : *« Le fait de mettre la parole de l'utilisateur entre guillemets lui donne un statut très particulier, cela fait d'elle une parole seconde au sein du discours. »*[63] Dès lors, avec la mise à jour, c'est l'inverse. Les bulles ont disparu, le gras apparaît sur la réponse de la machine, non pas sur le discours de l'utilisateur cité par Siri. Cette transition peut s'expliquer de plusieurs manières possibles.

D'une part, la police grasse est passée du discours utilisateur à celui de la machine, répondant à un objectif d'efficacité et de lisibilité : ce qui importe dans l'utilisation de Siri c'est la réponse qu'il donne et le service qu'il apporte aux utilisateurs. L'utilisateur doit pouvoir lire la réponse de Siri de manière claire pour obtenir satisfaction.

D'autre part, si les bulles ont disparu, c'est que l'utilisateur est maintenant acculturé. Il ne faut pas oublier que lors du lancement de Siri, les utilisateurs de nouvelles technologies voyaient pour la première fois arriver un assistant vocal (l'un des plus poussé sur le marché) et que les concepteurs avaient très certainement adopté une démarche pédagogique dans leur première version grand public du logiciel. Ces bulles, aujourd'hui disparues, constituaient d'autres cadres instituants et permettaient aux utilisateurs de comprendre la nature dialogique de l'échange (ou du moins sa mise en scène) entre eux et la machine. Aujourd'hui, seuls les guillemets sont restés et permettent de signaler à l'utilisateur que la machine s'approprie son discours, qui ne lui appartient plus après la dictée orale, comme nous l'avons déjà évoqué.

Enfin, le design et le fond noir (gris dans la version 2011) participe également à la mise en scène du dialogue entre l'homme et la machine. L'écran noir,

---

[63] Clotilde Chevet, La part de l'imaginaire dans « l'interaction homme-machine » : le cas des assistants personnels numériques, Mémoire de recherche Master 2 au CELSA Sous la direction du Professeur Karine Berthelot-Guiet, p. 11

pourrions-nous dire, c'est ce théâtre dans lequel deux acteurs vont jouer, dialoguer et écrire une histoire, ensemble. Cette image est intéressante car elle participe au développement d'un questionnement important qui pourrait se résumer à ces questions : Qui est Siri pour moi ? Qui suis-je pour Siri ? Ces questions identitaires seront développées dans la troisième partie du mémoire mais déjà, elles émergent lorsque l'on réfléchit à la mise en scène de l'interaction via l'écran, dans laquelle des rôles sont déjà prédéfinis (nous l'avions expliqué en première partie) mais dont l'interprétation change en fonction de chacun des utilisateurs du logiciel.

Nous avons affaire ici à un écrit d'écran qui reprend l'ensemble des codes de représentations du dialogue, issu de la culture de l'imprimé (les bulles des bandes dessinées par exemple) et de l'écrit informatique (la succession des discours, leur architecture et disposition au sein de l'écran par exemple) et qui participe à la création de la situation d'interaction entre l'homme et la machine. Mais plus encore, l'écran est ici la représentation visuelle et écrite du dialogue oral. Siri se donne finalement à lire et à voir puisque qu'il est écrit.
Mais converse-t-on vraiment avec Siri ?

3. Lieu de conversation ?

La conversation en question

Valérie Patrin-Leclère nous prévient : *« Le substantif « conversation » agrémente aujourd'hui des pratiques si diverses qu'on peut se demander ce qu'il désigne vraiment. »*[64] La question se pose évidemment dans notre cas. Dans quelle mesure peut-on dire que l'interaction homme-machine (et précisément utilisateur-Siri) vaut pour conversation ? L'interaction vaut-elle pour conversation parce qu'elle convoque des énonciations discursives orales ou bien

[64] La communication revisitée par la conversation, Valérie Patrin-Leclère, Communication et Langages, n°169, Septembre 2011, p15

parce qu'elle est donnée à voir comme un ensemble d'échanges discursifs via des écrits d'écran, ou les deux ?

La question se pose dans la mesure où le terme renferme un ensemble de conceptions diverses, *« si vaste qu'il accueille des échanges ni verbaux ni même verbalisables »*[65]. Ainsi, la conversation peut être vue comme un *« dispositif caractérisé par l'intention de favoriser la communication entre les possibles participants. Le terme désigne donc tout à la fois un imaginaire et une pratique possible sans être nécessairement avérée »*[66]. En effet, comme nous l'avons vu dans nos analyses jusqu'ici, de nombreuses marques énonciatives orales et écrites participent à la mise en place d'une situation d'interaction. Cette dernière est effective si l'on s'en tient à cette définition: *« L'interaction (…) suppose simplement échange entre inter-actants animés, quelle que soit la nature et l'intensité des échanges »*[67], mais elle va finalement au delà.

D'une part, car *« c'est le dispositif qui est censé faire la conversation. La mise en condition suffit à qualifier la situation. Le dispositif serait performatif, la performativité ferait la performance »*[68]. Cela signifie que par la présence même du dispositif Siri et de toutes les marques énonciatives (bulles, dispositions des discours machine et humains, ponctuation…), la situation d'interaction vaut pour conversation.

D'autre part, la situation d'interaction tente de placer l'homme et la machine sur un même niveau, ce qui amène à faire entrer l'utilisateur en croyance, une croyance selon laquelle la communication est optimale et optimisée entre les différents interactants. Et Valérie Patrin-Leclère d'ajouter : *« Célébrer l'avènement de la conversation, c'est récuser le modèle reliant un énonciateur et un*

---

[65] Ibid, p.15/16

[66] Ibid, p.16

[67] Daniel Luzzati, Le dialogue oral spontané : Quels objets pour quels corpora ?, LIUM, Université du Maine, p.1

[68] La communication revisitée par la conversation, Valérie Patrin-Leclère, Communication et Langages, n°169, Septembre 2011, p18

destinataire pour mettre en scène des co-énonciateurs perpétuellement à égalité dans un échange idéalement symétrique. »[69]

L'analyse pragmatique de l'utilisation de Siri, ramène cependant les utilisateurs à une réalité décevante, tant la qualité conversationnelle n'est pas toujours au rendez-vous, comme le montre cet échange simple:

La première réponse est correcte, elle satisfait l'utilisateur, la remarque "C'est splendide", donnée par l'utilisateur pour commenter le premier résultat de Siri et la réponse de Siri met directement fin à la conversation car il n'y a pas de compréhension de la part de la machine.

Comment la "conversation" sert les intérêts économiques des concepteurs

Au delà du dispositif lui-même, la notion de « conversation » dans le cas de l'utilisation de Siri trouve une résonnance particulière dans les discours qui

[69] La communication revisitée par la conversation, Valérie Patrin-Leclère, Communication et Langages, n°169, Septembre 2011, p17

circulent dans les médias et les communications des concepteurs et participent de cet imaginaire conversationnel : *« Outre le fait que le « marketing conversationnel » réfère à une nouvelle donne technologique, sociologique et économique, sa dénomination est la trace d'un maquillage destiné à revaloriser symboliquement des pratiques à visée marchande »*[70]. L'action de promouvoir la conversation, dit Valérie Patrin-Leclère, *« ancre la fameuse demande : les conversations sont en effet faites en partie de questions et de réponses, qui prolongent et entérinent l'imaginaire de l'offre et la demande. »*[71]Et cette remarque prend une consistance particulière dans le cas de notre objet d'étude. La promesse faite par les concepteurs d'intelligence artificielle, douée de langage, c'est précisément de pouvoir répondre à n'importe quelle demande. La conversation que pourrait avoir un homme avec Siri se transforme donc symboliquement en une allégorie d'un marché qui se base sur l'offre et la demande. La « conversation » devient à la fois un pré-texte promotionnel pour les industriels du secteur, en même temps qu'elle est une pratique désormais de plus en plus intégrée dans les nouvelles technologies de l'information et de la communication.

Ainsi, nous constatons à quel point la notion de « conversation » dépasse le dispositif lui-même en devenant un élément de langage qui sert les intérêts économiques des concepteurs. Mais nous ne pouvons pas dire que c'est l'un ou l'autre des éléments observés ici qui fait que l'interaction vaut pour conversation. Nous admettrons que c'est l'ensemble de ces discours et marques énonciatives, dans le dispositif et en dehors de celui-ci, qui font que l'interaction vaut pour conversation.

Mais intéressons-nous maintenant de plus près à la nature conversationnelle de l'interaction utilisateur-Siri.

---

[70] Ibid

[71] Idid, p.18

La conversation comme modèle de rapport entre l'homme et la machine

Tout d'abord, il convient de prendre une certaine distance vis à vis de notre étude, notamment dans cette partie, dans la mesure où l'étude de l'interaction homme-machine, qui plus est faisant intervenir du langage oral, pose problème. En effet, Daniel Luzzati s'est posé la question de l'étude du dialogue oral spontané entre les hommes et les machines dans son texte *Le dialogue oral spontané, quels objets pour quels corpora ?* Et il se demande *« comment un objet aussi statique (le corpus), fût-il multiple sous forme de corpus, peut-il prétendre représenter un processus dont la caractéristique fondamentale est d'être dynamique, de véhiculer une expérience qui va bien au delà des signes et des référents manipulés ? »*[72] Il faut selon lui définir les mots utilisés pour tenter de déjouer cet *« univers d'ambiguïtés où l'interopérabilité des concepts et du vocabulaire jette un trouble permanent. »*[73]

Nous avons donc questionné la notion de « conversation » comme type d'interaction entre Siri et l'utilisateur, en prenant en compte tout un ensemble de signes internes et externes au dispositif, mais il semblerait intéressant de pousser la réflexion à un niveau supérieur en se demandant si l'utilisateur et Siri dialogue. Nous proposons cette notion, que l'on pourrait considérer comme étant un type de conversation, mais la différence se trouve dans l'approche : la conversation, comme nous l'avons vu, se co-construit via un ensemble de signes internes et externes au dispositif. Nous voulons questionner la notion de dialogue pour Siri en prenant comme référence la définition de Daniel Luzzati : *« Le dialogue présuppose que la nature de l'échange est sinon verbale, du moins langagière, c'est-à-dire avec un certain contenu. »*[74]

---

[72] Daniel Luzzati, Le dialogue oral spontané : Quels objets pour quels corpora ?, LIUM, Université du Maine, p.1

[73] Ibid

[74] Daniel Luzzati, Le dialogue oral spontané : Quels objets pour quels corpora ?, LIUM, Université du Maine, p.1

Mais ne nous méprenons pas. Parler de dialogue plutôt que de conversation ne résout en rien notre problème, ni ne répond à notre question sur la manière avec laquelle Siri interagit avec l'utilisateur.

En effet, au delà des questions sémantiques (dialogue, interaction, conversation), Daniel Luzzati introduit la notion de *« parole orale et spontanée »*[75] que nous avons pu déjà lui emprunter tant elle permet de mettre le doigt sur un élément fondamental, trop vite oublié lorsque l'utilisateur est dans l'action d'utilisation : Siri est une machine, l'utilisateur est un homme. Qu'en est-il, donc, de la préméditation et de la qualité naturelle ou non du langage utilisé par Siri ? Daniel Luzzati prévient : *« Dès que l'on bascule du côté de la parole, c'est-à-dire du côté de la parole orale et spontanée, on change de versant du langage. On quitte en effet l'univers de la phrase, celui de la langue préméditée, celui de l'erreur qui s'efface et se rature, pour basculer du côté des énoncés conçus et perçus dans le fil de leur énonciation, des énoncés non prémédités, dont l'émetteur est le premier auditeur, dans lesquels l'erreur se traduit par un supplément de message. »*[76]. Est-ce le cas pour Siri ? Nous devons admettre que Siri est un cas hybride : en effet, d'une part Siri ne quitte véritablement jamais l'univers de la phrase puisque chacun de ses actes de langage sont prémédités (les concepteurs définissent en amont les réponses possibles). D'autre part, tout est donné à lire et à entendre comme si les énoncés étaient non prémédités (de nombreuses phrases font justement appel aux codes de la parole orale spontanée, en reproduisant des erreurs et hésitations, propre à ce type de parole) :

---

[75] Ibid, p.2

[76] Ibid

Daniel Luzzati observe deux types de formes de dialogues langagiers dans les *« applications, notamment informatiques, qui tentent de pratiquer du dialogue »*[77] : Les illusions de dialogues, qui correspondent à des *« dialogues avec une machine qui simule la méthode psycanalytique rogérienne, qui consiste à renvoyer au patient ses propres paroles »*[78] et les dialogues fortement contraints, qui sont des systèmes *« capables de détecter plusieurs mots clés dans un même énoncé de parole continue, de suivre une structure de dialogue en fonction d'une planification préétablie »*[79]. Siri appartient à la deuxième typologie, si l'on se réfère à la partie I. de ce mémoire sur l'analyse du processus technique. Mais si les rapports utilisateur-Siri s'apparentent plus à un dialogue qu'à une illusion de dialogue, il n'en reste pas moins qu'il manque à ces systèmes de véritables capacités cognitives fondamentales que Daniel Luzzati énumère assez longuement mais dont la conclusion résume bien le tout : *« Toute entité dialoguante devrait être capable de s'adapter, d'acquérir, d'apprendre. Un dialogue est une expérience qui se produit, et non un programme qui s'exécute. Un dialogue conduit à modifier*

[77] Daniel Luzzati, Le dialogue oral spontané : Quels objets pour quels corpora ?, LIUM, Université du Maine, p.5
[78] Ibid
[79] Ibid

l'état dans lequel on se trouvait au départ, de telle façon qu'il ne puisse jamais se reproduire à l'identique. »[80]Autrement dit, Siri n'a pas encore atteint cet objectif.

En revanche, Daniel Luzzati touche du doigt ici (*« un dialogue conduit à modifier l'étant dans lequel on se trouvait au départ »*) un élément que nous avons encore trop peu évoqué, il s'agit de la performativité du langage. Jean Caelen qui étudie notamment la question du dialogue homme-machine nous rappelle que le « *langage est une forme d'action et, par extension, le dialogue est une séquence d'actions planifiées ayant pour objectif un but visé sous-tendu par une intention »*[81]

Autrement dit, *« le dialogue serait une interaction : il renvoie le langage à l'action et réciproquement, ils se co-construisent réciproquement »*[82]*.* Jean Caelen analyse finalement le dialogue comme *« une suite coordonnée d'actions (langagières et non-langagières) conduites conjointement pour satisfaire un but. »*[83]

Pour Siri, la *séquence d'actions planifiées* est réalisée (les concepteurs prévoient les réponses, etc.), cependant, on ne peut pas en dire autant pour l'*intention*, qui comme on l'a relevé, ne vaut pas (encore) pour les intelligences artificielles douées de langage. Enfin, la notion de *but* est très clairement présente dans le dispositif Siri dans la mesure où dès le départ, l'assistant vocal nous demande *« Que puis-je faire pour vous ? ».* Cette question est le point de départ du dialogue entre l'homme et la machine et signifie que dès l'émission des premières marques énonciatives, l'objectif sous-entendu est celui du *« dialogue réussi ».* Jean Caelen modélise le dialogue réussi grâce à un schéma qui donne à voir les différents échanges.

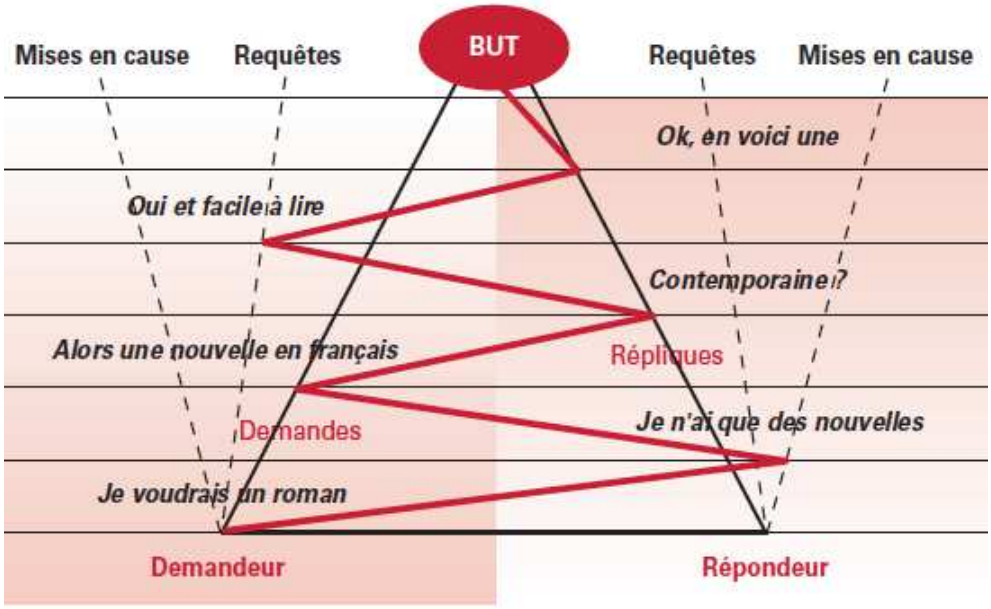
---

[80] Ibid

[81] Jean Caelen, Le dialogue homme-machine aujourd'hui, Dossier L'intelligence Artificelle, Centraliens n°580, Septembre 2007, p. 9

[82] Ibid

[83] Jean Caelen, Le dialogue homme-machine aujourd'hui, Dossier L'intelligence Artificelle, Centraliens n°580, Septembre 2007, p. 9



Dans le cas d'un dialogue réussi, le dialogue converge sur le but à atteindre. Ce schéma – qui représente un échange – montre les axes convergents (Demandes et Répliques) et les axes divergents (Mises en cause et Requêtes) du dialogue. Un dialogue réussi reste dans le cône de convergence.

Modèle projectif du dialogue – Jean Caelen[84]

Ce schéma vaut pour Siri dans la mesure où chaque type d'échange est représenté. Les « requêtes », qui sont le premier échange du demandeur, la « mise en cause » qui reflète l'échec d'une partie ou de l'entièreté du dialogue, et le but qui est atteint si le dialogue est réussi.

Mais revenons un instant sur la question de la performativité du langage et le rôle de l'écrit (que nous n'avons pas traité jusqu'ici) dans le cadre du dialogue utilisateur-Siri. Ce qu'on pourrait dire, c'est qu'au delà de l'oralité, la performativité du langage trouve une résonance importante à l'écran. L'utilisateur « dit », Siri « fait » et cela se « voit » à l'écran. La construction du langage par l'action se fait aussi à travers l'écran. Les architextes donnent vie aux écrits d'écran et créé « l'événement ». L'utilisateur parle, l'état des choses est modifié car une requête a été prononcée, et celle-ci est donnée à voir à l'écran.

[84] Ibid

Emmanuel Souchier explique de manière intéressante comment l'écriture numérique modifie la perception des images et des textes que nous avons : *« Le regard, paradoxalement enfermé par une fenêtre* [celle de l'écran]*, ouvre sur un univers dont on ne perçoit ni l'avenir ni le passé. C'est le règne du présent sans histoire ».* Autrement dit, en complément de l'oralité, ce sont les écrits d'écran qui sont donnés à voir par Siri, qui participent à cette forme de performativité du langage dans le dialogue utilisateur-Siri.

Dialogue-t-on avec Siri ? La réponse à cette question mérite une prudence certaine mais l'ensemble des marques énonciatives écrites et orales participent à la création d'un sentiment dialogique fort, même si celui-ci est souvent écourté par le manque de pertinence du système informatique.

D'un point de vue plus pragmatique, la question se pose tout autant dans la mesure où la durée des échanges entre l'utilisateur et Siri est assez courte. Cela veut dire que soit l'assistant vocal est assez réactif pour atteindre le but donné au dialogue, soit, comme on vient de le dire, que l'échec a lieu. Ce qui en revanche est certain, c'est que Siri ne joue pas le rôle de simple intermédiaire. Il réalise la médiation en même tant qu'il traite la requête (il est médiateur et interactant) mais il y a bien altération. Nous l'avons vu, de la conception du message à sa diffusion, les marques énonciatives sont nombreuses pour favoriser un univers dans lequel l'utilisateur peut sensiblement considérer qu'il est en dialogue avec la machine. L'assistant personnel numérique offre un véritable cadre à l'interaction homme-machine, qu'il s'agisse d'un cadre écrit ou oral.

III. Qui est qui, pour qui ? : perception et co-construction identitaire de Siri et de ses utilisateurs

A. Écrits d'écran et oralité au service de la construction identitaire de Siri

Nous l'avons vu à plusieurs reprises au cours de cette étude, de nombreux éléments participent à la construction d'une identité humaine de Siri. Nous nous appuierons sur les signes de l'interface qui permettent de soutenir cette hypothèse mais nous ferons appel à différentes sources telles que des témoignages d'utilisateurs via des forums ; les discours médiatiques et marketing viendront compléter notre étude sur l'identité de Siri.

La définition de l'identité sur laquelle nous nous appuyons est définie par Fred Dervin, qui a étudié notamment le GPS, les perceptions et la co-construction identitaire de celui-ci : *« Je soutiens que lorsque nous interagissons ou formulons des discours autour d'interactions vécues ou imaginées avec des objets , nous construisons également des expériences, la socialité et donc des identités »* [85]. Cette

[85] Le GPS a-t-il une identité ? Perceptions et co-constructions des GPS de véhicules et de leurs utilisateurs sur des forums de discussion, Fred Dervin, Synergies *Pays Riverains de la Baltique* n° 9 - 2012, p. 108.

approche correspond parfaitement à notre cas d'étude pour Siri qui inclut les notions de discours, interactions et objets (numériques).

Nous avions déjà repéré quelques éléments qui nous permettent de poser l'hypothèse selon laquelle la machine « s'humanise », ou plutôt que les constructeurs ont pensé la création de signes écrits et oraux pour tenter d'humaniser celle-ci.
Parmi eux, nous avions relevé la tenue à la verticale du Smartphone qui est la seule position avec laquelle l'utilisateur peut se servir de l'application, créant ainsi un sentiment de dialogue homme-homme.
D'autre part, la ponctuation des réponses de Siri respecte parfaitement les codes de l'écriture langagière de l'homme. De même que les hésitations, même si elles sont parfois ratées, participent à la création d'un imaginaire selon lequel Siri est un homme, avec ses failles, ses erreurs, etc.
Ces éléments à l'écran sont des signes qui donne à voir Siri comme un véritable interlocuteur. Dans ce sens, Siri tente de se faire passer pour un homme, en respectant les codes de l'écriture, mais cette imitation bien qu'artificielle, se renforce avec le complément de l'oralité. La voix de Siri est un élément du dispositif sur lequel il convient de s'arrêter, tant il participe à la création de l'identité de la machine.

« Le « grain », c'est le corps dans la voix qui chante , dans la main qui écrit , dans le membre qui exécute . »[86]Cette citation de Roland Barthes nous permet de comprendre l'importance de la voix dans la construction identitaire des individus, de comprendre comment la voix et le corps sont indissociables et par glissement, comment Siri, grâce à sa voix, aussi robotique soit-elle, se voit doté d'une identité humaine par les utilisateurs du logiciel.
En effet, le fait que Siri soit doté d'une voix lorsqu'il donne les réponses à l'utilisateur implique ce dernier dans une relation bien différente de celle qui

[86]Le grain de la voix Entretiens (1962-1980), Roland Barthes, Seuil

pourrait avoir lieu si Siri ne donnait que du texte à lire. La voix de Siri, c'est pour l'utilisateur une manière de pouvoir considérer le logiciel au même niveau qu'un interlocuteur humain. La voix de Siri, c'est aussi ce qui permet au cadre matériel de devenir un corps vivant, sensible et réceptif aux échanges avec l'utilisateur. La voix de Siri, c'est finalement ce qui permet à la fois d'humaniser la relation homme-machine, mais plus encore, de donner à Siri sa véritable identité.

Notons que la question identitaire est extrêmement complexe à en juger les possibilités qui sont données aux usagers des objets numériques parlants, comme le GPS ou les assistants personnels vocaux comme Siri.
En effet, les possibilités de personnalisation des voix des appareils numériques parlants sont nombreuses. Ainsi, on peut choisir une voix féminine, masculine, avec accent, sans accent…
Dans ses études d'analyses discursives autour de l'utilisation du GPS, Fred Dervin démontrent à quel point la question identitaire des objets numériques est prise très au sérieux par les utilisateurs : *« Ils masculinisent ou féminisent leur GPS quand ils en parlent. Tout d'abord, ils lui donnent un genre en faisant référence à lui par l'utilisation de pronoms personnels masculins ou féminins en anglais (she/he). Un autre phénomène largement répandu dans le corpus consiste à nommer le GPS. »*[87]
Pour ce qui est de Siri, nousaurions pu penser qu'une seule et même voix valait pour toutes les régions du monde, pour tous les utilisateurs. Or, il n'en est rien, les concepteurs du logiciel ont bien saisi l'enjeu identitaire et relationnel que renferme la voix d'un tel assistant personnel. Ainsi Clotilde Chevet a relevé que *« les concepteurs mettent au point des voix adaptées dans leur hauteur et timbre à la norme du pays . Ces différences d'un pays à un autre sont morphologiques mais également culturelles : ainsi au Japon il n'est pas rare que les femmes travaillent leur voix de façon à la rendre la plus aigüe possible , ceci étant un critère de*

[87]Le GPS a-t-il une identité ? Perceptions et co-constructions des GPS de véhicules et de leurs utilisateurs sur des forums de discussion, Fred Dervin, Synergies *Pays Riverains de la Baltique* n° 9 - 2012, p. 111.

séduction (...). L'élaboration de la voix de synthèse va donc prendre en compte des facteurs culturels pour s'adapter à son public. »[88]

Enfin, pour finir de démontrer à quel point la voix est indispensable à la création d'une identité humaine pour Siri, il suffit de remarquer à quel point les utilisateurs réclament de Siri un voix humaine, naturelle et fluide. De nombreux forums existent, sur lesquels beaucoup se plaignent de ne plus avoir la « belle voix » de Siri : *« Après restauration de mon iPhone, je constate que Siri est revenu à une voix compacte. »*[89]ou encore : « J'ai deux iPhone 4S, l'un a une voix humaine très agréable, l'autre pas. »[90]

Nous remarquons dans cette partie sur la question identitaire de Siri que la voix est un élément principal d'humanisation de la relation homme-machine. Plus celle-ci sera naturelle, plus la relation s'apparentera comme naturelle, pour l'utilisateur. Si Siri reprend les codes écrits et oraux de l'homme pour donner lieu à une situation d'interaction naturelle, les utilisateurs, eux, n'en attendent pas moins, alors même qu'ils savent qu'ils disposent d'une machine entre leurs mains. Finalement, les concepteurs n'ont fait que répondre à une demande, en même temps qu'ils perpétuent eux-mêmes cet imaginaire de dialogue oral spontané entre l'homme et la machine (le lancement de l'application se fait depuis peu via la parole, en disant "Dis Siri". Une manière de fluidifier encore plus la relation). Ainsi, nous assistons à une co-construction identitaire du logiciel Siri. L'utilisateur prend pour vrai les signes (écrits et oraux) qui participent à la naturalité du dialogue et d'autre part les concepteurs se chargent de mettre en place des techniques d'interactions qui s'apparentent le plus possible à celles qui existent entre deux hommes.

---

[88] Clotilde Chevet, La part de l'imaginaire dans « l'interaction homme-machine » : le cas des assistants personnels numériques, Mémoire de recherche Master 2 au CELSA Sous la direction du Professeur Karine Berthelot-Guiet, p. 59

[89] Siri n'a plus sa belle voix, Forum : http://www.edencast.fr/topic/siri-na-plus-sa-belle-voix/. Consultée le 31 août 2016.

[90] Ibid

B. Une co-construction identitaire soutenue par des discours d'accompagnements

Si le dispositif participe lui-même de la co-construction identitaire de Siri, il n'en reste pas moins que les discours marketing et médiatique autour de Siri favorisent la création d'une identité humanisée.

En reprenant les différentes versions des sites web d'Apple (via l'outil en ligne web.archive.org), nous pouvons analyser la différence et les évolutions de discours de la marque Apple pour la promotion et présentation de son assistant personnel. On identifiera plusieurs éléments qui participent à la co-construction de l'identité humanisée de Siri.

En 2012, Siri est promu via le site internet d'Apple comme étant un assistant personnel, un valet ou encore un génie qui répond à toutes les interrogations quotidiennes : « Your wish is its command »[91]. A cette époque, les concepteurs du programme étaient dans une démarche d'évangélisation et de pédagogie, les discours diffusés permettaient de faire comprendre le fonctionnement d'un tel assistant personnel aux utilisateurs. Ainsi, on pouvait également lire à propos de la manière dont il faut parler à Siri : *« It understands what you say. And knows what you mean. Talk to Siri as you would to a person. »*[92]En émettant des discours de la sorte, Apple tente d'effacer une bonne fois pour toute la frontière entre homme et machine. Désormais, on parle à une machine comme l'on parle à un homme. C'est du moins la promesse sous-jacente qui est faite ici.

---

[91] Titre de la page d'accueil du site internet http://www.apple.com/ios/siri en 2012 : https://web.archive.org/web/20120915020356/http://www.apple.com/ios/siri. Consulté le 31 août 2016

[92] Ibid : https://web.archive.org/web/20120915020356/http://www.apple.com/ios/siri

Aussi, le discours de promotion d'Apple passe en revue un certain nombre des attributs de l'homme. Après avoir précisé que l'utilisateur pouvait parler normalement à la machine (mettant donc en avant la capacité de la machine à dialoguer naturellement et spontanément), la marque vient mettre l'accent sur la capacité d'adaptation de la machine, grâce à une fonction cognitive et mémorielle forte : *« Siri not only understands what you say, it's smart enough to know what you mean. »*[93]. Ici, Apple place Siri comme un être doué de conscience, d'intelligence. Il est, selon le constructeur high-tech, capable de comprendre les sous-entendus, les blagues ou encore le sens de discours oraux complexes. Et d'ajouter : *« So when you ask "Any good burger joints around here?" Siri will reply "I found a number of burger restaurants near you." Then you can say "Hmm. How about tacos?" Siri remembers that you just asked about restaurants, so it will look for Mexican restaurants in the neighborhood. »*[94] Au même titre qu'un humain, Siri se rappelle de la dernière conversation que l'utilisateur a eu avec lui et peut ainsi reprendre les recherches en prenant en compte l'historique de conversation. En réalité cependant, cette fonction mémorielle ne fonctionne pas toujours de manière optimale.

Aujourd'hui encore, le discours marketing d'Apple à propos de Siri n'a pas changé. La promesse reste toujours la même, le ton étant seulement moins pédagogique puisque les utilisateurs de Smartphone de la marque à la pomme sont a priori maintenant acculturés à Siri. Apple place néanmoins Siri plus comme un ami et moins comme un valet à notre disposition, comme c'était le cas quelques années en arrière : *« Adressez-vous à Siri comme à un ami et il vous aidera à faire toutes sortes de choses. »*[95]

---

[93] Ibid

[94] Ibid

[95] Texte issu du site internet Apple consacré à Siri : http://www.apple.com/fr/ios/siri/. Consulté le 31 août 2016

Les discours médiatiques ne sont pas en reste et ont eux aussi contribué à développer un imaginaire selon lequel la relation conversationnelle homme-machine est équivalente à celle entre deux hommes.
Siri est d'ailleurs régulièrement considéré plus comme un proche de l'utilisateur que comme un assistant personnel numérique par les médias. Il est donné à voir comme étant un assistant personnel sur qui on peut toujours compter : *« Siri est un compagnon »*[96], *« Siri : l'ami électronique »*[97],*« un compagnon indispensable »*[98], etc.
Ces dénominations de Siri par les médias sont le signe d'une victoire pour Apple qui a réussi son évangélisation de Siri. Les discours marketing trouvent un écho importants dans les médias à tel point que les internautes, qui disposent de Siri et créent du contenu autour du logiciel n'hésitent pas à titrer leurs vidéos de la manière suivante : *«Siri mon ami et son humour», «Siri mon ami », « Mon ami Siri »*, etc.[99]

L'ensemble des discours d'accompagnement qu'ils soient marketing ou médiatiques favorisent cette idée selon laquelle Siri doit être considéré comme un interlocuteur au même titre qu'un humain. Il s'agit d'un jeu de va et vient entre les possibilités réelles du dispositif et les discours qui accompagnent celui-ci. Ce jeu vient finalement créer une croyance qu'on pourrait qualifier d'universelle et selon laquelle l'utilisateur dialogue avec Siri comme avec un humain, car le dispositif incite à agir de la sorte mais également car les discours d'accompagnement sont assez incitatifs (voire pédagogiques) pour que les utilisateurs acceptent d'utiliser le logiciel dans les cadres qui sont fixés par le dispositif.

---

[96] Article du Monde.fr : http://www.lemonde.fr/economie/article/2013/09/28/siri-eusement_3486436_3234.html Consulté le 31 août 2016
[97] Article de Protéger Son Image : http://www.protegersonimage.com/siri-lami-electronique-qui-garde-vos-donnees-pendant-2-ans/ Consulté le 31 août 2016
[98] Article de Mac4ever : https://www.mac4ever.com/actu/113093_ios-10-l-integration-des-apps-tierces-devrait-faire-de-siri-un-compagnon-indispensable COnsulté le 31 août 2016
[99] Recherche effectuée sur Youtube le 1 septembre 2016 : https://www.youtube.com/results?search_query=siri+mon+ami

C. Quel avenir, quelles évolutions et quelle place pour Siri ?

Alors qu'il avait été présenté comme révolutionnaire lors de son lancement, Siri a aujourd'hui perdu du terrain, ou plutôt, ses concurrents ont rattrapé leur retard : Amazon Echo, Cortana ou Google Now ont définitivement intégré la course à l'innovation dans le domaine des assistants vocaux personnels. La question n'est pas de savoir si l'un ou l'autre surpassera son voisin concurrent, mais plutôt de questionner la place que prendra l'assistant vocal personnel dans le futur.

Pour répondre à cette question, plusieurs éléments sont à prendre en compte. D'une part, dans un contexte de développement intense du machine learning, le critère de la base installée (c'est à dire, la capacité des constructeurs à séduire un public large et régulier) est extrêmement important. En effet, plus les utilisateurs seront nombreux, plus les logiciels en question pourront se nourrir des usages et donc affiner leurs réponses, leur pertinence. L'enjeu pour les constructeurs est donc de parfaire les systèmes d'apprentissage de la machine. Apprendre à apprendre est donc l'un des enjeux futurs pour les constructeurs. Sur ce terrain, il semblerait que Siri ne soit pas celui qui récolte le plus de point si l'on devait pondérer les différents acteurs du marché. En effet, malgré ses 100 millions d'utilisateurs quotidien, Siri apparaît quelque peu comme un *« fantasme technologique inachevé »*[100]. Ses réponses semblent parfois beaucoup moins précises que son concurrent Google Now, par exemple. Et pour cause : le moteur de recherche profite de l'ensemble des recherches utilisateurs pour affiner la pertinence de son assistant vocal, remportant ainsi un des meilleurs taux de pertinence sur ce marché.

[100] La loi des Siri, Gabriel Siméon : http://www.liberation.fr/futurs/2015/09/09/la-loi-des-siri_1379111, 9 septembre 2015, consulté le 21 novembre 2016

D'autre part, il est une tendance qu'on ne peut ignorer : les assistants personnels prennent une place de plus en plus importante et constituent désormais de véritables services à part entière pour les entreprises constructeurs mais également pour les marques qui utilisent ces techniques. L'exemple des chatbots comme outil de service de relation client est intéressant dans la mesure où l'automatisation de la relation homme-machine se déploie, pour ne pas dire se démocratise.

Néanmoins, nous constatons une véritable propension à conserver l'écrit comme mode d'échange. La raison s'explique assez rationnellement : cette pratique de mise en place d'assistants numériques commerciaux via des applications de messagerie écrites correspondaux usages. L'écrit et la place de l'écrit n'a rarement été aussi important que depuis l'avènement des Smartphones et ordinateurs. Les possibilités de production et de duplication sont courtes et simples, les moyens de diffusion plus rapides et plus faciles. Les assistants personnels basés sur l'écrit sont donc plus faciles à mettre en place, en même temps qu'ils sont également plus discrets et moins intrusifs que les assistants vocaux. En effet, l'écrit se fait alors plus discret que l'oral, dans ce genre de situation. La notification écrite de l'assistant vocal sur votre smartphone qui vous rappelle que vous avez un train à prendre et que vous pouvez modifier votre place sera toujours plus discrète que la prise de parole de la machine, non désirée par l'utilisateur. Dans cette situation, l'écrit gagne en efficacité dans la mesure où l'utilisateur pourra consulter l'information au moment voulu, tandis que l'oralité viendra nécessairement perturber la situation. C'est en cela que Siri ne peut malheureusement pas surpasser les services qui utilisent des procédés écrits.

Mais alors que reste-t-il de la place de l'oral dans cette tendance qui se déploie année après année ? Siri pourrait-il être ce que ses constructeurs ont toujours

voulu qu'il soit ? Un assistant personnel numérique qui nous incite à penser que nous pouvons discuter, et solliciter ses services comme on sollicite un vendeur ou un concierge privé ? Un assistant numérique qui s'efface pour nous apporter des réponses seulement lorsque nous en avons besoin ?

Malgré les avancées et son déploiement depuis peu sur les ordinateurs MacBook Pro d'Apple, le logiciel Siri semble accusé le coup. Les constructeurs n'ont pas dit leur dernier mot mais il semblerait que la migration du logiciel vers d'autres machines soit perçu comme un aveu de faiblesse. Les constructeurs de Siri réussiront leur pari seulement s'ils prennent en compte les enseignements suivants :

Paradoxalement, en redéfinissant la position de l'écrit face à l'oral pour coller aux usages actuels. Nous aurions pu penser en effet que l'oral deviendrait le mode d'échange principal entre l'homme et la machine, réalisant ainsi le vieux fantasme de la machine parlante mais force est de constater qu'aujourd'hui soit l'acculturation n'est pas terminée (on constate d'ailleurs que peu d'utilisateurs dictent leur SMS par exemple), soit il existe des contraintes d'usage qui font que le mode oral ne permet pas d'aboutir à une satisfaction utilisateur optimale. Ajoutons d'ailleurs que l'écrit disposed'une force supplémentaire face à l'oral : un mot écrit reste toujours identique une fois rédigé tandis qu'un mot dit oralement pourra varier selon des accents, etc. Il en revient donc que les assistants personnels intelligents sont plus à même de satisfaire l'utilisateur via l'écrit que via l'oral. Siri devrait alors éventuellement reconsidérer les modalités d'échange avec ses utilisateurs en s'intégrant plus discrètement mais plus pertinemment avec un mode écrit, certainement plus adapté aux missions d'assistants personnels numériques.

Le nombre de capture d'écran disponible sur Internet qui montre les fautes d'interprétation et de captation du discours oral utilisateur sont suffisamment

nombreuses pour considérer qu'Apple a du chemin à parcourir avant de parfaire son outil d'assistant personnel. Mais le constructeur se retrouve face à une spirale infernale de laquelle il est difficile de sortir : si la pertinence de réponse du logiciel est faible, le nombre d'utilisateur diminue. Or, si le nombre d'utilisateur diminue, la pertinence de réponse ne s'améliore pas. L'enjeu résiderait donc dans le fait de modifier le logiciel pour obtenir des satisfactions sur la base utilisateurs déjà installée puis de réactiver une phase d'acquisition pour nourrir l'algorithme.

Enfin, il nous est impossible de ne pas parler de la future place de Siri, de son identité et de sa perception dans le temps. Etre un assistant personnel intelligent signifie être là au bon moment. Il se trouve que le logiciel Siri se rend disponible mais n'agit pas en mode pro-action. D'autre part, nous l'avons dit : l'oralité comme mode d'échange pour un tel service n'est peut-être pas le plus optimal dans la mesure où il y a un risque de parasitage. Il serait certainement plus intéressant d'avoir un assistant personnel plus discret, avec un mode d'échange plus écrit. Cela modifierait évidemment la perception et l'identité que l'utilisateur se fait de Siri mais elle correspondrait peut-être plus à l'idée que nous nous faisons de l'assistant personnel numérique. Il est vrai que nous sommes contradictoires : certains sont fascinés et contribuent à cette forme de fantasme selon lequel la machine pourrait un jour parler avec l'homme comme il le fait avec un autre. Mais de l'autre côté, nous souffrons déjà les sollicitations intempestives de la machine qui désire nous parler alors même que nous de désirons pas forcément l'entendre ou entendre la réponse par le mode oral. Siri est donc face à un problème de taille. Ne serait-il pas simplement réduit à être ce qu'il est déjà, un serviteur qui se doit de ne pas déranger mais être présent lorsque cela est nécessaire ?

Conclusion

Les concepteurs de Siri ont un objectif : assister la vie de toutes personnes équipées d'un Smartphone de la marque Apple pour les défaire de toutes les tâches quotidiennes triviales. Cet assistanat numérique, chacun des utilisateurs peut y avoir accès comme il le souhaite et les concepteurs l'ont pensé en conciliant les contraintes techniques du dispositif (téléphone mobile doté d'une interface, accès à Internet, microphones, etc.) et les usages potentiels que pourraient avoir les utilisateurs avec ce genre de service. La solution communicationnelle entre l'utilisateur et la machine la plus efficace qui ait été trouvée est celle de l'oralité. Mais l'oralité pour un assistant personnel « numérique » n'est pas une mince affaire, comme nous avons pu le constater dans notre étude sur Siri. Numérique, *« qui a rapport aux nombres »*[101]. Tout l'enjeu pour les concepteurs réside donc dans la cohabitation entre l'univers de l'écrit numérique (propre aux langages numériques) et l'univers de l'oralité naturelle et spontanée (propres aux hommes) pour aboutir à une situation dite d'interaction qui permettra de donner satisfaction aux deux inter-actants : l'utilisateur pour la réponse à la requête qu'il obtiendra, Siri (ou plutôt ses concepteurs) pour la réussite de ce challenge, qui aujourd'hui est devenu un sujet très tendance suite à l'arrivée des chatbots et autre assistants personnels oraux, notamment.

Lorsqu'on analyse les enjeux qui président à l'utilisation d'un programme numérique d'assistanat personnel, on ne se doute pas ou peu des combats menés par l'oralité et l'écriture qui composent ensemble le logiciel Siri. Notre étude a eu le mérite de mettre en avant un certain nombre de problématiques liées à la cohabitation entre l'oral et l'écrit qui peut se révéler tantôt fructueux (l'expérience de Siri, lorsque la réponse à la requête est juste, est très

[101] Étymologie de l'adjectif numérique,http://www.cnrtl.fr/etymologie/num%C3%A9rique, consulté le 31 août 2016

appréciable d'un point de vue utilisateur), tantôt infructueux (la frustration principale née d'une absence de compréhension, ou devrions-nous dire d'incompatibilité des langages écrits et oraux ; numériques et humains). Mais ces analyses nous ont certainement amené à poser beaucoup plus de questions qu'elles n'ont permis d'y répondre.

L'analyse du dispositif technique nous a permis d'apercevoir ce que cette boîte noire renfermait, ou du moins ce qu'elle laissait transparaître. Ce qu'il en ressort c'est une certaine représentation du monde d'aujourd'hui et de demain, qui ne fait que prospérer l'utopie selon laquelle hommes et machines arriveront un jour à communiquer, converser, dialoguer aussi bien (ou mieux) que deux hommes entre eux. Mais le paradoxe de nos analyses se situe dans les contraintes qui sont imposées par les enjeux techniques, économiques et symboliques du dispositif. Partagé entre efficacité, simulacre de l'expérience dialogique et cohabitation symboliques et pragmatiques de l'oralité et de l'écriture, Siri ne s'efface pas suffisamment, ni n'est assez performant pour qu'une expérience relationnelle et conversationnelle digne de ce nom ait lieu entre l'homme et la machine. Même si Siri présente l'avantage d'avoir un écran, ce qui permet d'afficher des résultats plus précis et longs, peut-être les concurrents tels que Amazon Echo (qui ne sont que des assistants personnels oraux) réussiront mieux le pari de l'expérience dialogique avec les utilisateurs (encore que supprimer l'écran ne résout finalement en rien les problèmes que posent l'écriture numérique et l'oralité ensemble dans un tel dispositif).

Nous ne disons pas que l'écriture et l'oralité ensemble constituent un désavantage ou un problème dans la réussite d'une telle entreprise, nous signalons en fait que le langage, qu'il soit écrit ou oral, pose problème. Il pose problème entres les hommes, mais il en pose finalement tout autant dans une relation homme-machine. Finalement, nous sommes aujourd'hui à une étape des évolutions où les industriels du secteur testent plusieurs méthodes pour tenter de réaliser l'utopie conversationnelle entre l'homme et la machine. Le futur nous dira donc (ou pas) ce qu'il en est.

Ce que nous retenons de la suite de notre étude se situe non plus dans le domaine de la compatibilité des langages des inter-actants, mais bien dans le langage lui-même. En effet, la parole de l'utilisateur étant déléguée au logiciel Siri, elle ne lui appartient plus et l'interprétation de la parole humaine par la machine ne pourra plus être modifiée. L'utilisateur devient spectateur et auditeur d'une conversation qu'il avait pourtant démarrée et se satisfera de lire ses propos rapportés, et d'une réponse orale (au son robotique) même si la requête a mal été interprétée par la machine. L'écran du dispositif devient alors le lieu de l'événement dans lequel l'interaction homme-machine est rendue lisible et visible. Ce qui est paradoxal dans cette partie de notre analyse (durant laquelle nous avions posé la question suivante : Quel cadre pour l'échange homme-machine (Siri) ?), c'est de voir à quel point les cadres qui rendent l'interaction homme-machine (Siri) possible sont les mêmes qui empêchent l'expérience dialogique d'être optimale ou du moins équivalente à celle qui existe entre deux humains. L'écrit d'écran et l'oralité (enregistrement de la requête et diffusion de la réponse audio) permettent de créer la relation entre l'homme et la machine en même temps qu'ils restreignent la situation de dialogue entre les deux interactants. Le dispositif Siri participe à la création d'une situation dialogique en donnant à voir des marques énonciatives nombreuses qui constituent des cadres à l'interaction homme-machine, sans que l'on puisse la qualifier de dialogue.

Nous voudrions néanmoins être clair dans notre conclusion : les mots ont un sens, mais ils ont également le sens que l'on veut leur donner. Ainsi, dire que l'homme et Siri dialoguent serait selon notre analyse de deuxième partie juste. Cependant, nous restons prudents dans l'utilisation de tels termes (interaction, conversation, dialogue) dans la mesure où nombre de chercheurs ont déjà tenté de définir ces notions et les débats concernant ces termes sont sans cesse renouvelés, notamment depuis que les échanges communicationnels entre

l'homme et la machine sont possibles. Nous avons préféré une formulation plus lourde (« situation dialogique »), mais qui a le mérite de mettre en avant la complexité de la situation.

Le langage définit notre identité. La parole distingue l'homme des autres espèces. Lorsqu'un logiciel comme Siri est doté d'une voix, se pose naturellement la question de son identité. Mais nous n'avons pas chercher à trouver la nature de son identité, ce qui nous ait apparu comme un projet vain ; nous avons plutôt cherché à comprendre, comment en situation d'interaction, l'homme considérait la machine parlante qu'est Siri. Force est de constater qu'un ensemble de signes co-construise une identité humaine à Siri. Qu'il s'agisse des écrits d'écran et de l'oralité, des discours d'accompagnements que la marque Apple a rudement mené durant le lancement du logiciel ou encore des discours médiatiques qui font passer Siri comme étant un « ami » de l'utilisateur, l'identité humaine que l'on peut accorder à Siri est le fruit d'une véritable co-construction. N'oublions d'ailleurs pas l'utilisateur lui-même, qui ne se cache pas de vouloir avoir affaire à un homme plutôt qu'à une machine.
La question identitaire de Siri est finalement une question de croyance (on entend croyance ici comme ce qui correspond à l'investissement des sujets dans une proposition et non pas comme étant l'objet du croire) . Chaque utilisateur sait bien qu'il s'agit d'une machine, mais ils se prêtent au jeu de la croyance. Comme l'explique si bien Frédéric Lambert dans son ouvrage « Je sais bien mais quand même » : *« Pour partager les règles communes, il faut négocier avec les autres nos croyances (…). Il nous faut jouer des langages »*[102]. Pour utiliser Siri, il faut donc négocier avec soi-même et les autres (Apple notamment) le fait que ce dernier est un interlocuteur (humain) alors même qu'au fond chacun sait qu'il s'agit d'une machine.

[102] Je sais bien mais quand même, Essai pour une sémiotique des images et de la croyance, Frédéric Lambert, Éditions Non Standard

« Chaque société produit les supports et les nouvelles formes de récits qui lui sont nécessaires pour offrir au plus grand nombre des raisons de croire ensemble »[103], alors laissons le soin aux utilisateurs de croire ou de ne pas croire aux récits qui font que l'on puisse considérer Siri comme ayant une identité humaine ou non. C'est en tout cas ce processus de croyance, en complément des discours d'accompagnement et des signes (écrits et oraux) proprement liés au dispositif que les utilisateurs s'autorisent à dialoguer avec Siri comme il pourrait le faire avec un homme.

Siri est une intelligence artificielle qui aujourd'hui s'est étendue aux ordinateurs de la marque Apple. En pleine évolution, les assistants personnels numériques se développent rapidement et l'arrivée des *chatbots* est un pas de plus vers une diversification des moyens de communication entre les hommes et la machine. Il se trouve qu'Apple a tout intérêt à redoubler d'effort pour faire évoluer son logiciel qui aujourd'hui reste relativement peu performant. L'avenir dira si Siri eu été une évolution ringarde ou avant-gardiste. Il aura, quoi qu'il arrive, eu le mérite d'être le premier assistant personnel numérique sur téléphone mobile, ce qui pour Apple n'est pas négligeable dans cette quête incessante de l'innovation technologique.

---

[103] Ibid

Bibliographie

Barthes Roland,Le grain de la voix Entretiens (1962-1980), Seuil

Bougnoux Daniel, Sciences de l'Information et de la Communication, Larousse, 1993

Ong Walter J. , Oralité et écriture, Les Belles Lettres, 2014,

Charaudeau Patrick, Ce que communiquer veut dire, Revue des Sciences humaines, n°51, Juin, 1995

Caelen Jean, Le dialogue homme-machine aujourd'hui, Dossier L'intelligence Artificielle, Centraliens n°580, Septembre 2007

Chevet Clotilde, La part de l'imaginaire dans « l'interaction homme-machine » : le cas des assistants personnels numériques, Mémoire de recherche Master 2 au CELSA Sous la direction du Professeur Karine Berthelot-Guiet,

Cosnier Jacques, Synchronisation et copilotage de l'interaction conversationnelle. Protée, 33-39, 1992. Laboratoire d'Ethologie des Communications Université Lumière Lyon2 (FR)

Dervin Fred, Le GPS a-t-il une identité ? Perceptions et co-constructions des GPS de véhicules et de leurs utilisateurs sur des forums de discussion, Synergies *Pays Riverains de la Baltique* n° 9 - 2012

Goody Jack, Oralité et écriture, Communication et Langage n°154, Décembre 2007

Huglo Marie-Pascale, Présentation,Études françaises, vol. 39, n°1, 2003, p.7-11, URL :http://id.erudit.org/iderudit/006737ar,

Isnard Emmanuel, « Jack Goody, Pouvoirs et savoirs de l'écrit », L'Homme, 2009, 274-277

Lambert Frédéric, Je sais bien mais quand même, Essai pour une sémiotique des images et de la croyance, Éditions Non Standard

Luzzati Daniel, Le dialogue oral spontané : Quels objets pour quels corpora ?, LIUM, Université du Maine

Patrin-Leclère Valérie,La communication revisitée par la conversation, Communication et Langages, n°169, Septembre 2011

Souchier Emmanuel, L'écrit d'écran, pratiques d'écriture & informatique, Communication et Langages n°107

Wiener Norbert, Cybernétique et société, Éditions du Seuil 2014

Webographie

Mary Branscombe, *Reconnaissance vocale : la vie avant Siri* :http://www.tomsguide.fr/article/siri-reconnaissance-vocale-histoire,5-118.html

Titre de la page d'accueil du site internethttp://www.apple.com/ios/siri en 2012 :https://web.archive.org/web/20120915020356/http://www.apple.com/ios/siri.

Article du Monde.fr
:http://www.lemonde.fr/economie/article/2013/09/28/siri-eusement_3486436_3234.html

Article de Mac4ever :https://www.mac4ever.com/actu/113093_ios-10-l-integration-des-apps-tierces-devrait-faire-de-siri-un-compagnon-indispensable

Texte issu du site internet Apple consacré à Siri
:http://www.apple.com/fr/ios/siri/.

Article de Protéger Son Image :http://www.protegersonimage.com/siri-lami-electronique-qui-garde-vos-donnees-pendant-2-ans/

Siri n'a plus sa belle voix, Forum :http://www.edencast.fr/topic/siri-na-plus-sa-belle-voix/.

Citation extraite d'une vidéo de présentation de Siri par Apple, le 4 octobre 2011
:https://www.youtube.com/watch?v=agzItTz35QQ

Des Français percent le secret de Siri, Simon Robic, 15 novembre 2011
:http://www.numerama.com/magazine/20578-des-francais-percent-le-secret-de-siri.html

Qui sont les voix de Siri ?, Isabelle Lesniak, Les Echos, 30 octobre 2015
http://www.lesechos.fr/week-end/business-story/innovation/021437753914-je-suis-siri-1170668.php?tR1B2Qx1HWahh8vg.99#

Résumé du mémoire

Ce mémoire s'intéresse aux relations entre oralité et écriture au sein de l'application d'Apple Siri. Assistant personnel intelligent programmé pour être utilisé à l'oral, sa particularité se trouve dans le fait qu'il dispose d'un écran qui produit donc des écrits. Cette observation ouvre l'analyse aux champs immenses que sont l'écriture et l'oralité dans l'ère des médias informatisés.

L'analyse menée dans ce mémoire se découpe en 3 parties :

- Une première s'attache à analyser le processus technique de Siri.
- Une deuxième tente de comprendre les cadres instituants qui font que l'oralité et l'écriture ensemble participe à la création de l'interaction homme-machine.
- La troisième s'intéresse à l'aspect identitaire (humaine) que les utilisateurs et les discours d'accompagnements concèdent à Siri.

De toutes ces analyses, l'étude de l'oralité et de l'écriture est un point commun qui permet de mettre en avant les problématiques de compatibilités des langages humains(écrits et oraux) mais également ceux qui désormais font partie de notre quotidien, les langages numériques (écriture informatique, écrits d'écran, architextes, oralité numérisée…).

Mots clés

Écriture, Oralité, Écrits d'écran, Interface, Interaction, Conversation, Dialogue, Siri, Voix, Numérique