

# Assistants personnels, et intelligents ?

Alexandre Roulet, David Schenkel, Victor Wojciechowski, Nathan Fourel

*Etudiants en ingénierie des médias, 1<sup>ère</sup> année, HEIG-VD*

Depuis les années 50, et l'avènement du NLP (Natural Language Processing), l'interaction entre l'homme et la machine grâce au médium vocal est rendue possible. Cependant, tout au long de l'existence de ces interfaces d'interaction vocale et ce depuis ses débuts, cette technologie a été lente, limitée et de facto peu utile. La capacité de la machine à nous comprendre est intrinsèquement liée à l'évolution des technologies et la multiplication des datas.

## I. INTRODUCTION

L'homme a toujours souhaité améliorer et simplifier son mode de vie. Il a souvent créé des machines pour l'aider dans son travail, et plus récemment des robots « intelligents ». Ceux-ci n'étant pas humains, il est par conséquent plus compliqué d'interagir avec et ce, quel que soit leur utilité, aussi bien professionnelle que privée.

## II. HISTORIQUE

Le langage est sans doute l'élément le plus important car il est à la base de toute communication. Le TALN (traitement automatique du langage naturel) est une discipline de l'informatique qui traite du langage. Nous pouvons y avoir recours lors de diverses traductions ou par exemple avec la reconnaissance vocale. L'idée est d'apporter de la cohérence autour du vocabulaire employé afin d'être le plus précis possible dans les interactions avec l'ordinateur. Un mot peut très bien avoir un certain sens dans un contexte donné et en avoir un autre dans un contexte différent. C'est un élément crucial lorsque l'on aborde le sujet des assistants vocaux. Il fait son apparition dans les années 50 alors même que le mathématicien et cryptologue Alan Turing crée un test mesurant le degré d'intelligence d'une machine à partir de la capacité à se faire passer pour un humain. Les résultats sont peu convaincants et même si de nombreux progrès apparaîtront au cours du 20<sup>e</sup> siècle, il reste très compliqué de faire comprendre toutes les subtilités d'une langue à des machines [1].

Cinq notions bâtissent les fondamentaux du TALN. Nous avons, dans un premier temps, la *morphologie* qui se concentre sur les mots. La *syntaxe*, qui permet l'analyse de l'agencement des mots dans une phrase. La *sémantique* qui offre la signification des mots et des expressions. La *pragmatique* qui aide à comprendre le contexte dans lequel se trouve un mot ou une phrase. Finalement, nous trouvons la *phonologie* analysant les sonorités vocales. C'est autour de ces points-clé que réside la discipline du TALN [2].

Différents outils de TALN existent comme par exemple, le « Natural Language Toolkit », produit en langage python. Nous trouvons également le kit « CSLU » qui traite le langage vocal.

Il reste bien évidemment beaucoup de progrès à faire avant que l'humain puisse aisément discuter en profondeur avec un ordinateur ou autre objet connecté. Non seulement parce qu'une machine n'a pas d'émotion à partager mais surtout parce qu'il reste compliqué pour elle de comprendre le sens de certaines expressions ou encore, de maîtriser l'ironie par exemple. En effet, la partie pragmatique doit être améliorée. Toutefois, il est déjà possible d'effectuer certaines requêtes auprès des machines, avec de bons résultats, comme celle qui consiste à transmettre un ordre vocal ou à transformer un texte écrit en paroles audibles. Ces avancées technologiques peuvent désormais s'appliquer également au marché des assistants vocaux.

## III. TECHNIQUE

Demandez quelque chose à un assistant vocal intelligent et vous le ferez travailler sur trois tableaux bien distincts qui, mis ensemble correctement devraient vous offrir l'expérience la plus agréable et la plus personnalisée possible.

Base de toute l'opération : vous allez parler. Banal.

Vos paroles devront être captées (analogie de l'oreille humaine et du tympan) c'est le rôle du *hardware*.

Cette partie, bien physique, représente l'objet auquel vous poserez une question, votre frigo, votre voiture, ou votre smartphone captera. Ensuite, celui-ci se chargera d'encoder votre voix (langage humain) en bits numériques (langage machine).

Cette partie n'est, en ce moment, pas la première priorité des fabricants car la plupart d'entre-eux préfère implanter leur solution d'assistant dans des objets créés par d'autre firme. En exemple nous citerons Amazon (alexa) qui s'implante dans le hardware (la voiture dans ce cas) de BMW et de Seat [3].

A la suite de cette première phase, c'est *le software* qui prendra la suite des opérations. Son but consiste à interpréter ce que vous avez dit. C'est la partie centrale de toute l'opération, qui se base sur la pièce maîtresse : *les datas* (les données). En reprenant l'analogie, les datas sont à la fois votre mémoire, votre capacité à parler ainsi que vos connaissances ; le software lui, c'est votre cerveau qui fait le lien entre toutes ces parties [2].

Les datas peuvent se séparer en trois sous-groupes : *les datas de service*, *les datas de communication* et *les datas de mémoire*. Celles de communication servent à améliorer l'interaction vocale entre la machine et l'humain. Quelle langue parlée, la tonalité, la retranscription de l'information. Celles de mémoire

servent à connaître l'utilisateur. Ces goûts, préférences, activités. Celles de services sont ici pour vous amener la connaissance que vous avez demandé (l'information). Il fait 20° à Genève, le restaurant ouvre à 19h par exemple.

#### IV. OLIGOPOLE VOCAL

Sur le marché, chaque produit, chaque offre est indépendant de l'autre, sans rechercher aucune interconnexion. Chaque assistant à ses forces, en fonction de l'écosystème et des bases de données qu'il a à disposition. Par exemple, l'assistant d'Amazon gère très bien les achats en lignes, puisqu'il se base sur les données d'Amazon, sa pléthore d'objets connectés et ses shops en ligne, tandis que l'assistant de Google est bien meilleur dans la gestion des fonctions générales (recherches, météo) et la gestion des services googles (agenda, mail).

La capacité de ces assistants relève des données qu'ils possèdent sur notre environnement et sur nous. Seulement, ces données ont une valeur et sont un avantage de taille sur le marché. Connaître ses clients et leurs habitudes permet de mieux cibler les publicités et les offres et donc de convertir plus d'action en achat. La vente des informations étant d'ailleurs un sujet récurrent depuis quelques années. Partager ces données avec la concurrence c'est d'abord proposer une expérience potentiellement plus complète mais également moins optimisée de son assistant vocal. Ensuite, c'est perdre un certain avantage par rapport à ces concurrents, car, rappelons-le, connaître le client et ses habitudes est un avantage énorme et pas uniquement dans le domaine des assistants vocaux. Ensuite il y a également un problème d'image.

Apple ayant eu une communication assez agressive concernant le fait qu'ils ne vendaient pas les données de ses utilisateurs ou Google ayant déjà une réputation assez moyenne concernant la gestion des informations privées, un retournement de communication semble difficile. Pour finir, comment se partageraient ces données ? Entre les « gros » du marché ? Accessible à tous ? Rendre ces données accessibles à tous les assistants vocaux revient à les mettre en ligne à disposition de tout le monde, ce qui paraît invraisemblable. Mais les partager entre les 5 mastodontes du marché (Google, Alexa, Bixby, Siri et Cortana) revient à asseoir l'aspect oligopolistique de ce marché, fragiliser les rapports de forces entre ces 5 entreprises et exclure certains aspects que des assistants vocaux professionnels ont (comme la gestion de projet) [5,6,7,8,9].

#### V. FUTUR

Avec toutes les informations désormais en notre possession nous pouvons formuler les questions suivantes ; Les fabricants nous laisseront-ils l'accès aux données qu'ils ont récolté lors de notre utilisation du service proposé ?

Si oui, à quoi cela nous servirait-il de connaître ce que ces entreprises ont récolté sur nous ? Comme le dit Bernard Marr dans son livre « *Big Data In Practice* », (paraphrase et traduction de l'anglais) [10];

« Chaque seconde, l'humanité télécharge (upload) 1,7 mégabytes de données. Pour le mettre en perspective, il s'agit

1782579.2 Byte. Imaginez que chaque seconde, l'ensemble des êtres humains écrit 1'782'579,2 lettres de l'alphabet.

Chaque minute nous téléchargeons 300 heures de vidéos vers un serveur Youtube (upload), 3 millions de vidéos sur Facebook. En plus de cela d'ici 2020, l'humanité possédera 6 milliards de téléphones qui récoltent les données GPS de chacun. ».

#### VI. CONCLUSION

La conclusion est que cet ensemble de données n'a que très peu de valeur si on ne sait pas traiter celles-ci afin d'en retirer des informations utiles au développement de la société de consommation. Le consommateur lambda ne saura pas quoi faire de ses données, pourtant les GAFAs peuvent prédire les futurs modes de consommation, le comportement d'achat de chacun. Les états peuvent aussi se servir des données, prenons l'exemple de ce qu'il se passe actuellement avec la pandémie du Corona virus, la Confédération Helvétique a décidé de recourir au tracking des cartes Sim présentes dans un rayon défini. Google collecte aussi les données GPS de votre téléphone, vous pouvez vous en rendre compte en vous rendant sur Google MyActivity.

La récolte de données se fait à travers plusieurs outils comme mentionné ci-dessus, le smartphone en fait partie, mais il en existe encore bien d'autres peut-être même en utilisez-vous de manière journalière sans vous en rendre compte ? La carte cumulus de la Migros, la Supercard de la Coop.

Plus récemment nous avons constaté l'apparition de l'Apple Card, mais à quoi sert-elle ? Leur slogan est le suivant « *What you spend, when, and where. All beautifully clear.* »

L'Apple Card permet à son utilisateur de mieux se gérer financièrement, de savoir ce pourquoi il dépense ses sous et où il les dépenses. Afin de connaître ces éléments, Apple se doit de récolter une grande quantité de données afin de pouvoir les communiquer à son client.

En dernier lieu, nous pouvons nous poser la question de savoir quel est le type d'assistant intelligent que nous voulons posséder ? Un assistant vocal intelligent ou un assistant intelligent ?

#### RÉFÉRENCES

- [1] [https://fr.wikipedia.org/wiki/Traitement\\_automatique\\_du\\_langage\\_naturel](https://fr.wikipedia.org/wiki/Traitement_automatique_du_langage_naturel)
- [2] <https://www.ionos.fr/digitalguide/web-marketing/vendre-sur-internet/le-traitement-automatique-du-langage-naturel-taln/>
- [3] <https://www.lesnumeriques.com/voiture/amazon-place-alexa-chez-bmw-seat-n66967.html>
- [4] <https://www.cognifit.com/fr/parties-du-cerveau>
- [5] <https://www.silicon.fr/google-home-fin-vie-privee-159635.html>
- [6] <https://www.appvizer.fr/magazine/operations/ppm/assistant-vocal-professionnel>
- [7] <https://www.cardis.ch/fr/magazine-immobilier/2018/les-assistants-vocaux-bouleversent-la-domotique-dans-nos-maisons>
- [8] <https://communicationsetinternet.com/2018/09/27/interfaces-vocales-assistant-personnel-vocal-quelles-opportunités-pour-les-marques/>
- [9] <https://www.24heures.ch/news/news/assistants-vocaux-epreuve-marche-suisse/story/11782243>
- [10] Bernard Marr: *Big Data In Practice*, How 45 successful companies used big data analytics to deliver extraordinary results