

Dialogues vocaux entre clients et automates ou comment l’homme et la machine s’entendent dans la réalisation d’un service

Valérie Beaudouin ¹, Julia Velkovska ²

¹ Institut Télécom LTCI Télécom ParisTech – 46, rue Corvisart – 75013 Paris – France

² Orange Labs, Laboratoire SENSE –38-40 rue du Général Leclerc– 92794
Issy-les-Moulineaux cedex 9 – France

Résumé

Dans cet article nous exploitons en partie une base de données consignant les interactions vocales entre des clients et des automates vocaux dans le cadre de la relation de service. L’approche statistique nous permet d’identifier les facteurs, les traits qui favorisent l’accomplissement du service, tels qu’ils sont consignés dans la base. L’analyse ethnographique des interactions, basée sur l’analyse conversationnelle, et guidée par l’approche quantitative dans la sélection des échantillons, permet de mettre à jour d’une manière plus fine les causes des désalignements ou malentendus entre l’automate et l’humain et donnent des orientations pour développer de nouveaux indicateurs dans la base.

Abstract

This paper analyses human-machine speech interactions in the context of telephone customer service provision. The database contains audio recorded interactions between a customer and a machine, and a set of features, mainly technical, that describe the call. We operate a quantitative analysis of the database in order to identify the features that improve or decrease the success of this type of interaction. Based on this first analysis, a set of calls were selected, transcribed and submitted to in-depth analysis. The latter leads to a better understanding of the reasons of misalignment between the human and the machine and provides insight on what kind of features should be added in the database to increase its usefulness.

Keywords : natural dialogue, human-machine speech interaction, service relation, conversational analysis

1. Introduction

La relation entre les organisations et les clients, en particulier dans le secteur des services, est en pleine mutation. Les dispositifs médiatisés (téléphone, internet, systèmes d’information...) occupent une place grandissante dans la relation de service, ce qui conduit les firmes à proposer une offre de mise en relation multicanal. Dans des environnements où la compétition entre firmes se joue principalement dans la capacité d’innover en termes de services, l’accélération du processus de mise sur le marché de nouvelles offres accroît naturellement les occasions de contacts pour l’achat comme pour le service après-vente.

En multipliant la fréquence et les modalités de contacts, en augmentant les plages horaires de disponibilité, les entreprises sont confrontées à un accroissement de leurs coûts commerciaux.

Une des voies pour contrôler l'accroissement des coûts consiste à automatiser la relation de service en déléguant les tâches les plus répétitives et ayant le moins de valeur ajoutée des conseillers clients à des automates vocaux. Dans cette dernière configuration, le client interagit vocalement avec une machine.

Ce papier se concentre sur l'analyse de ces technologies, en analysant des expérimentations d'automates vocaux faites auprès de clients informés. Les données de ces interactions ont été consignées dans une base de données, conçue pour rassembler toutes les traces des expérimentations en vue de l'amélioration des dispositifs techniques. Cette base de données a fait l'objet de publication de rapports techniques internes.

La question qui guide notre travail est simple : connaissant la subtilité de l'art du dialogue, par quels mécanismes l'homme et la machine parviennent-ils à « s'entendre » pour que le service soit rendu ? Dit autrement, comment la machine parvient-elle à rendre le service demandé par le client ?

Nous commencerons par présenter le dispositif mis en place par la R&D d'Orange pour capitaliser les résultats d'expérimentations d'automates vocaux. Nous verrons en quoi le dispositif peut ou non nous permettre de répondre à nos questions.

Ensuite, sur des échantillons de la base, nous essaierons d'identifier et d'analyser les facteurs ou critères qui favorisent ou sont corrélés avec la réussite ou l'échec de l'interaction.

Enfin, en prenant appui sur cette exploration statistique, nous chercherons à explorer par une approche qualitative, relevant de l'analyse de la conversation, les mécanismes locaux qui favorisent ou défavorisent l'ajustement entre le client et la machine.

Nous montrons comment l'alliance entre une approche statistique et une exploration ethnographique nous paraît pertinente pour explorer cet objet singulier qu'est l'interaction vocale entre un humain et une machine.

2. Analyser la relation de service

2.1. Travaux sur la relation de service : d'une approche socio-économique à l'analyse des interactions

La relation de service constitue un domaine de recherche très actif dans les sciences sociales avec des approches à grains d'analyse variable. En socio-économie, des recherches ont tenté d'identifier les causes et les transformations de la relation de service (Bandt and Gadrey, 1994; Gadrey and Zarifian, 2002; Gadrey, 2003). D'autres travaux se sont attachés à analyser le fonctionnement socio-économique de la relation de services dans les centres d'appel (Cossalter, 2002; Flichy and Zarifian, 2002; Mounier, 2002).

Au niveau interactionnel, un champ s'est constitué, dans la suite des travaux en ethnométhodologie (Garfinkel, 2007 [1967]), sur les interactions au travail (Zimmerman, 1992) et de nombreux travaux ont porté sur l'analyse interactionnelle dans la relation de service (Baker et al., 2005), travaux regroupés sous le terme des *workplace studies* et influencés par les travaux de Goffman, en particulier par ses réflexions sur la relation de service dans *Asiles* (Goffman, 1968).

Que se passe-t-il quand la conversation est fortement médiée par la machine ? Nous renvoyons à Hutchby qui dresse une cartographie assez complète des différentes formes technokonversations (Hutchby, 2001), y compris la relation entre un homme et une machine. Licoppe et Relieu (2005) ont quant à eux analysé la manière dont la technologie (ordinateur et accès

aux applications du système d'information) s'immisce dans la relation téléphonique entre le consommateur et le conseiller, et souligné tout le travail d'ajustement que mène le conseiller entre son écran et la conversation. Des expériences d'interactions entre humains et machines ont été analysées par exemple dans les travaux de Suchman, dans l'interaction entre un humain et un photocopieur (Suchman, 2007), mais aussi dans les travaux de Cassell d'interaction entre un humain et un agent conversationnel animé (Bickmore and Cassell, 2005).

Le travail le plus proche du nôtre est celui de Woolfitt et MacDermind (1995), analyse d'interactions vocales entre un humain et un agent conversationnel. Cependant ce travail mobilise la technologie du magicien d'Oz, et nous éloigne de « naturelles ».

Le travail que nous présentons ici est la première analyse systématique d'interactions vocales entre un humain et une machine dans le cadre d'une relation de service authentique, autrement dit en situation naturelle. C'est également le seul travail à notre connaissance qui croise une approche quantitative avec une exploration ethnographique.

2.2. La base de traces vocales : consigner les interactions vocales

Le centre de R&D d'Orange à Lannion a construit sur une vingtaine d'années une expertise de haut niveau sur les techniques de la parole (reconnaissance et synthèse vocales), du dialogue et du dialogue intelligent (Sadek, 1999). Un des domaines d'application de ces technologies est la relation client. Il existe différents types de technologies derrière les automates vocaux, relevant d'une conception maximaliste (modéliser l'intellect) ou modérée (résolution de tâches concrètes) de l'intelligence artificielle (Rastier, 1991). Le noyau central du dispositif fonctionne soit à base de règles implémentées une à une, soit sur la base d'un système intelligent qui modélise l'activité avec davantage d'épaisseur.

Cela étant dit, ces systèmes de dialogue intelligent ne peuvent être conçus en laboratoire et déployés directement à grande échelle. La phase d'expérimentation avec des clients (la partie la moins maîtrisable) est centrale pour l'ajustement des règles, pour la prise en compte de cas non prévus... Il est d'usage que les expérimentations soient analysées de près par des spécialistes en ergonomie du dialogue, ce qui conduit à des ajustements quasiment en temps réel du dispositif.

Parallèlement à ce travail d'expérimentation et d'ajustement progressif des automates, la R&D a pris la sage décision de capitaliser sur ces expérimentations en les consignant dans une base de données. L'idée était de conserver la mémoire des expérimentations et d'autoriser des traitements secondaires sur les données. La Base des Traces Vocales regroupe ainsi une trentaine d'expérimentations de différents services. Pour chaque expérimentation, un ensemble de caractéristiques de l'interaction ¹ sont consignées :

- Des caractéristiques techniques comme la longueur de l'interaction, sa date de début, le nombre d'interventions client... ;
- Des traits codés manuellement comme le motif de l'appel et la qualification de l'appel : succès, échec, raccroché ;
- La bande sonore, à savoir l'enregistrement intégral de l'interaction.

Il nous paraît nécessaire d'insister sur ce dernier point. En effet, pour une partie des expérimentations, seule la voix du client a été enregistrée alors que celle de l'automate a été perdue. Comme nous le verrons dans la dernière partie, c'est dans l'ajustement, l'enchaînement

¹ L'interaction correspond à un appel et désigne l'ensemble des échanges entre l'homme et la machine, jusqu'au raccroché ou au transfert.

des tours de parole entre l'humain et l'automate, que se glissent les principaux problèmes interactionnels. Les expérimentations sans la « voix » de l'automate sont peu exploitables. Les expérimentations consistent en une transcription des interventions orales du client, à savoir ce qu'a reconnu la machine. Travailler sur ces retranscriptions serait très pertinent, si d'une part les deux « locuteurs » (humain et machine) étaient retranscrits, si d'autre part le système de transcription permettait de gérer les chevauchements. En particulier, certaines interventions du client ne sont pas entendues, si le client intervient là où il n'a pas sa place (cf. dernière partie) et ne sont donc pas retranscrites.

Nous sommes typiquement en face d'un dispositif technique innovant, en phase d'évolution et d'amélioration.

2.3. Données analysées dans cet article

Nous nous sommes donc concentrées principalement sur deux services avec automate vocal. L'un, service A, propose un ensemble de services permettant de gérer sa ligne fixe (transférer des appels, payer sa facture, suivre sa consommation, consulter sa messagerie...) ; l'autre, service B, service d'assistance technique, permet d'identifier le client et le motif d'appel pour orienter vers les conseillers clients adaptés. Le service B est un aiguilleur d'appels.

Sur le service A, ont été analysées toutes les interactions enregistrées dans la base entre janvier 2006 et mars 2007, soit 16.777 appels, dont 10.281 appels ont été qualifiés manuellement (motif de l'appel et aboutissement). En revanche, nous n'avons pas pu analyser pour ce service la retranscription du motif d'appel tel qu'interprété par l'automate. L'exploitation quantitative porte sur des traits de description fermés (motif d'appel, aboutissement, durée, nombre d'interventions...). L'approche ethnographique porte sur des appels ou interactions sélectionnés à partir des résultats de l'analyse quantitative, qui ont été retranscrits manuellement à partir de l'enregistrement audio, en reprenant les principes de l'analyse conversationnelle.

Sur le service B, comme il n'y a pas eu de qualification manuelle des interactions, nous avons effectué un autre type de travail, qui a consisté à analyser les motifs d'appels énoncés par le client, tel que le système de reconnaissance vocale les a retranscrits, soit l'analyse de 4.716 interventions clients. Certaines interactions du service B ont également été transcrites à la main, main-oreille plutôt, et analysées en profondeur.

Si pour le service A on peut corréler le motif d'appel, le succès de l'appel à des indicateurs quantitatifs (longueur...), ce travail n'a pu être mené sur le service B (pas d'analyse manuelle du motif et de l'issue (réussite ou échec de l'interaction)).

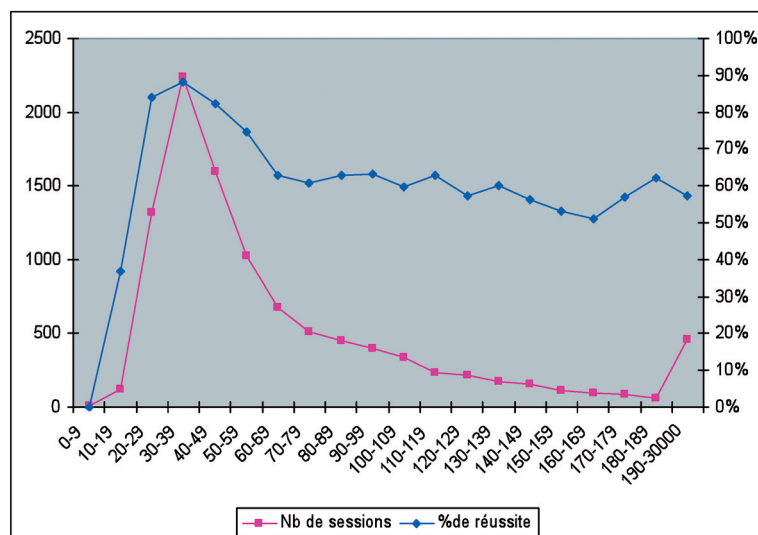
3. Facteurs d'échec ou de réussite : quelques traits quantitatifs

A partir des indicateurs disponibles dans la base, nous avons cherché à identifier les facteurs corrélés avec l'échec ou la réussite, dans la co-production du service.

3.1. Quels traits favorisent l'accomplissement du service A ?

Limitons-nous aux appels adressés au service A qui ont été qualifiés manuellement (identification du motif de l'appel et qualification en termes d'échec ou de réussite), soit 10.281 interactions. Soulignons que la notion de réussite ou d'échec relève d'une interprétation du codeur et correspond à une évaluation externe de l'adéquation de la réponse de la machine à la demande du client.

Fig. 1 présente la répartition des appels selon la durée. On note une concentration très forte des appels autour de la ½ minute et une longue traîne avec des appels pouvant durer très longtemps (courbe « nombre de sessions »). En croisant la durée des interactions avec le taux de réussite, il apparaît que plus l'interaction s'étend dans le temps, plus le taux d'échec augmente. Au-delà de 70 secondes, le taux d'échec est de 40%. De fait, il existe une courte fenêtre de tir, entre 20 et 50 secondes pendant laquelle le taux de réussite est très élevé (plus de 85%).



Clef de lecture : courbe « Nb de sessions » : 55 interactions durent moins de 10 secondes, et 660 plus de 190 secondes (1'40''). Quand la durée de l'appel se situe entre 20 et 50 secondes, le taux de réussite dépasse les 80%. En deçà, le taux d'échec devient très fort (60%) et au-delà le taux d'échec augmente avoisinant les 40%.

Figure 1 : Répartition des appels selon leur durée et taux de réussite

Quand on regarde le nombre d'interventions client, dans près de la moitié (46%) des interactions enregistrées, le client n'intervient qu'une seule fois. Il intervient deux fois dans 19% des interactions.

En moyenne, les demandes clients adressées au service A aboutissent trois fois sur quatre. L'augmentation du nombre d'interventions du client dans la session s'accompagne d'une diminution du taux de réussite. Il semble y avoir un optimum d'efficacité du système quand le client n'intervient qu'une fois (Tab. 1).

	Nb d'interactions	% d'interactions	Taux de réussite
Une intervention	4.577	46%	82%
Deux interventions client	1.884	19%	75%
Trois interventions client	1.204	11%	72%
Quatre interventions client	737	7%	67%
Plus de 4 interventions	1.879	17%	57%
Total	10.281		74%

Tableau 1 : Répartition des interactions en fonction du nombre d'interventions client

En introduisant la variable « motif d'appel », on note une représentation très majoritaire d'appels pour un nombre très réduit de services. Les trois-quarts des appels relèvent d'un des quatre

motifs suivants : suivi consommation, transfert d'appel, paiement de facture et messagerie (Tab. 2).

	<i>nb sessions</i>	<i>%</i>	<i>nb interventions</i>	<i>médiane</i>	<i>durée moyenne</i>	<i>médiane</i>	<i>Taux de réussite</i>
Suivi consommation	3.437	33	2,2	1	00:00:50	00:00:36	89%
Transfert d'appel	1.706	17	4,4	3	00:01:20	00:01:04	83%
Paiement facture	1.547	15	2,1	1	00:00:58	00:00:45	92%
Messagerie	721	7	2,2	1	00:00:55	00:00:41	91%
Information	307	5	6,0	5	00:02:36	00:02:04	69%
Ne sais pas	1.442	14	3,0	2	00:01:11	00:00:59	7%
Hors périmètre	585	6	3,9	3	00:01:47	00:01:33	63%

Tableau 2 : caractéristiques des interactions selon le motif de l'appel

Pour les demandes très standardisées et régulièrement utilisées (paiement de facture, messagerie, suivi consommation) le taux de réussite tourne autour des 90%.

Quand il s'agit de demandes d'informations ou que la demande n'est pas dans le périmètre du service A, le taux de réussite est beaucoup plus bas. Enfin, pour les autres demandes, non ou mal identifiées (dans 14% des cas, le codeur ne sait pas identifier le motif d'appel), l'échec est quasi systématique.

Le cas du transfert d'appel est particulier puisque c'est le seul service intégralement traité dans le cadre du service A : cela explique le plus grand nombre d'interventions clients et la plus longue durée. En effet, pour les autres demandes, une fois que le service A reconnaît la demande, il bascule le client vers le service adapté.

Une régression logistique qui cherche à hiérarchiser les facteurs (motifs d'appels, nombre d'interventions clients et durée des appels) qui jouent sur le taux de réussite montre que les motifs sont plus structurants, et que toutes choses égales par ailleurs, quand on bloque l'effet motif d'appel, ce sont les appels entre 30 et 90 secondes qui ont le plus de chances d'aboutir.

3.2. Motifs d'appel au service B ou comment la machine oriente l'expression du client

Sur les données du service B nous n'avons pas pu déployer le même type d'analyse, car il n'y a pas eu de qualification manuelle des motifs d'appels, de l'échec ou de la réussite. En revanche la demande client, a été retranscrite par le système de reconnaissance vocale, avec les limites inhérentes au dispositif (erreur de reconnaissance...).

L'analyse porte sur 1675 interactions clients entre le 26/04/2006 et le 05/07/2006, qui correspondent à 4716 interventions clients, dont 1.638 motifs d'appels analysés. L'analyse des motifs d'appels retranscrits a été faite avec le logiciel Alceste (Reinert, 1990; 1993; Lebart and Salem, 1994), et donne une typologie en cinq classes des motifs d'appels : problème d'installation de la livebox / problème de connexion à internet (2 manières différentes de dire le problème : « je n'arrive pas à me connecter » ou « problème de connexion ») / problème avec la ligne fixe / autres problèmes.

L'analyse des ces motifs d'appels nous montre pour commencer que l'expression en langage naturel d'une demande ne va pas de soi. En effet, la moitié des réponses sont standardisées et suivent le *modèle prescrit par l'automate* (« je n'arrive pas à »). En effet, quand l'automate s'adresse au client, il lui dit : « par exemple vous pouvez dire « je n'arrive pas à installer ma livebox » ». Un quart des réponses sont sous la forme de *groupe nominaux* (« problème de

connexion à internet ») et empruntent le modèle de l'interrogation des moteurs de recherche. Enfin, seul un quart des réponses correspond à des énoncés « en langage naturel », *exprimés librement*, mais évidemment beaucoup plus complexes à traiter pour la machine.

Quelles conclusions préliminaires pouvons-nous tirer de ces traitements à gros grain sur les interactions ?

- Il semble exister un *modèle dominant de l'interaction homme-machine* avec une seule intervention client et de durée courte. C'est à la fois le modèle le plus fréquent et celui qui conduit au taux de réussite le plus élevé.
- Plus la durée de la conversation s'allonge, plus les chances d'échec augmentent.
- Plus le motif est routinier et/ou clairement circonscrit – à savoir l'accès à un service spécifique - plus le taux de succès est élevé. Les taux de réussite les plus bas concernent les demandes d'information. La standardisation des motifs d'appels favorise la réussite de l'interaction.
- Enfin, l'automate joue un rôle fort de prescription sur la demande du client, en lui proposant des formulations de sa question. L'humain tend à parler comme la machine, à prendre prise sur le contenu proposé par l'automate. L'ajustement de vocabulaire qu'opère le client dans la formulation de la question suite aux « conseils » de la machine, ont un effet positif sur la concrétisation de la paire question-réponse.

Cette première approche quantitative sur des données d'interactions vocales ne donne pas une place aussi grande qu'on aurait pu le souhaiter au matériau textuel (retranscription de l'oral). Il permet cependant d'identifier des traits qui sont pertinents pour l'analyse des situations de dialogue.

4. Alignement ou désalignement : traits qualitatifs

Voyons à présent comment concrètement nous pouvons analyser les mécanismes d'ajustement et de désajustement entre le client et l'automate. Pourquoi l'humain et la machine cessent-ils parfois de coopérer ? Qu'est ce qui conduit à l'échec ? Pour ce faire, nous avons constitué des petits échantillons en prenant comme point d'entrée le motif d'appel, puisque c'est la variable la plus structurante de la réussite ou de l'échec de l'appel et en choisissant pour chaque motif d'appel une vingtaine d'interactions avec de 1 à 10 interventions clients, en ayant à chaque fois une interaction réussie et un échec. Ces interactions ont été pour la plupart retranscrites à la main et analysées, ce qui permet de voir concrètement dans le déroulement de l'interaction les points de difficulté et la manière dont ils sont ou non surpassés.

En termes de méthodologie, notons que l'analyse de conversation apporte un éclairage nouveau sur le déroulement des interactions entre l'être humain et la machine, complètement occulté par le traitement statistique. A l'issue de ces observations, la question se transforme et devient : peut-on mettre au point des indicateurs qui pourraient être mesurés automatiquement et seraient de bons critères pour identifier des interactions qui s'orientent vers l'échec, pour par exemple les faire basculer vers un conseiller humain ?

Il nous semble en effet que l'efficacité de ces dispositifs repose sur leur capacité à être hybridés avec des interactions entre humains, et en particulier à pouvoir basculer vers un conseiller humain, quand on entre dans un dialogue de sourd.

Nous concluons sur les conditions qui favorisent l'efficacité de la machine et qui impliquent une parfaite maîtrise de la machine par l'homme, ce qui contredit l'objectif initial d'un service dans lequel on peut exprimer librement sa demande.

4.1. Quelques difficultés interactionnelles sources de désalignements potentiels

Cette section identifie une série de problèmes interactionnels dans le déroulement des appels, qui peuvent être ou non résolus par des ajustements entre l'humain et la machine.

4.1.1. S'entendre

La première difficulté dans ces interactions entre un client et un automate tient au décodage du signal. Il est fréquent que la machine n'entende pas ou entende mal, ce qui arrive aussi au client. Dans l'exemple suivant, l'automate interprète la toux du client comme un nom de service.

Exemple 1. La machine interprète la toux comme un nom de service

A : (jingle) telco bonjour bienvenue sur le service A l'appel depuis une ligne fixe est gratuit (.) (jingle)
 (.) [avec les...
[(C tousse)]
 A : j'ai compris que vous demandiez un forfait si ce n'est pas cela dites annulez
 C : **annulez**

Les automates de reconnaissance vocale sont très sensibles au bruit et à l'environnement. Ils sont optimisés pour un locuteur seul dans un environnement calme. Dès qu'il y a des apartés, du bruit de fond (télévision ou radio), le système se trouve en difficulté. De même, il peut être perturbé par des prononciations spécifiques.

4.1.2. Trouver sa place

Dans ces interactions vocales par téléphone, c'est l'organisation qui définit quand le client peut ou non interrompre la machine. Elle définit les places autorisées et les places interdites. En particulier, elle interdit toute interruption dans le message d'accueil, qui est le lieu de véhicule de son identité, de sa marque. Ainsi dans l'exemple ci-dessous, la cliente tente-t-elle dès l'ouverture de donner son motif d'appel, sans être « entendue » la première fois, en étant « entendue » avec retard la seconde.

Exemple 2. La cliente tente d'interrompre la machine trop tôt (service A)

A : Telco bonjour [bienvenue sur le (service A)]
 C : **[Payer ma facture]**
 A : l'appel depuis une ligne fixe est gratuit
 C : **Payer ma facture**
 A : (jingle) Prononcez le nom du service qui vous intéresse ou formulez librement votre demande sinon composez le code du service
 A : [si vous n'appellez pas de chez v... Veuillez patienter je vous passe le service telefact (jingle)
 [...]

Dans l'exemple suivant, le client n'arrive pas à occuper les places qui lui sont assignées par la machine. Il se tait pendant 6 secondes après un « c'est à vous », puis parle en même temps que la machine, sans se rendre compte qu'il est finalement entendu.

Exemple 3. Le client ne trouve pas sa place.

[...] A : D'accord avant de vous passer un téléconseiller merci d'exprimer la raison de votre appel par exemple vous pouvez dire : « je n'ai plus de tonalité sur ma ligne » c'est à vous
 C : **(6s)**
 A : Je ne vous [ai pas entendu pour vous diriger rapidement vers le
 C : **[euh il il il n'yap]**
 A : [téléconseiller approprié j'ai besoin de connaître la

C : [il n'y a plus de sonnerie
 A : [J'ai compris que vous appeliez pour « un problème avec votre ligne
 C : [il n'y a plus
 A : téléphonique » [si ce n'est pas cela dites « annulez »
 C : [problème
 C : (.)oh putain mais c'est pas possible de voir ça ((client raccroche))

En résumé, prendre sa place dans une conversation entre humains relève d'une compétence bien identifiée (cf. le turn-taking, Sacks et al., 1974, comme principe d'organisation de la conversation). Dans le cadre de ces interactions hybrides, homme-machine, les schémas usuels sont perturbés d'une part par un « droit » à l'interruption très développé et d'autre part par l'asymétrie fondamentale entre l'humain et la machine, puisque seule cette dernière a le droit de fixer les lieux où le client peut interrompre.

4.1.3. Partager les mêmes catégories

Le client et l'organisation ne partagent pas la même vision du monde. Le client voit les choses depuis sa perspective, un point de vue d'utilisateur de service centré sur lui ; l'organisation de son point de vue situé prend en compte le catalogue complet de ses offres, l'organisation de ses centres d'appels, la structure de ses applications...

Ces différences de point de vue sont évidemment visibles dans les interactions vocales et deux catégories majeures de difficultés peuvent être identifiées : les discordances de vocabulaire et des difficultés à s'accorder sur la logique de l'interaction.

Commençons par les écarts de vocabulaire. L'entreprise fabrique des services, leur donne des noms, mais le client ne s'en souvient pas toujours ou ne donne pas la même signification aux choses. Retrouver le nom du service que l'on souhaite n'est pas simple pour le client, et si dans le premier exemple ci-dessous, le problème est réglé par le client lui-même, qui en tâtonnant retrouve le nom, dans le second on assiste à un dialogue de sourds. Par *forfait*, la cliente entend l'état de sa consommation du forfait, alors que pour la machine, il s'agit d'une demande de souscription ou d'information à un service.

Exemple 4 : Le client retrouve seul le nom du service

A : que désirez-vous
 C : **bricogit**
 A : [je ne vous ai pas entendu (.) [si vous
 C : **[allofact** **[allofact**
 A : Veuillez patienter je vous passe votre service

Exemple 5 : Le client et l'automate ne s'alignent pas sur le nom du service

A : [...] Prononcez le nom du service qui vous intéresse
 [ou formulez librement votre demande
 C : **[forfait**
 A : j'ai compris que vous demandiez un forfait sinon dites annulez
 C : **for-fait**
 A : Dans la catégorie offres tarifaires, j'ai compris que vous demandiez un forfait...
 C : **Mais non !**

La deuxième gamme de difficultés tient au fait que la structure de l'organisation ou la structure de l'application peuvent être rendues visibles pour le client alors qu'elles ne sont pas pertinentes pour lui. Dans l'exemple ci-dessous, le client ne comprend pas pourquoi il faut dire d'abord « transfert d'appel », puis « transférer les appels ». L'arborescence du système lui est ainsi présentée, avec ses idiosyncrasies non adaptées au client.

Exemple 6 : L'arborescence de l'application est non pertinente pour le client

A : [...] que désirez-vous
C : transfert d'appel
 A : transfert d'appel vous pouvez demander à transférer vos appels ou bien désactiver le transfert en cours enregistrer un nouveau numéro privilégié ou encore modifier ou supprimer votre numéro privilégié que voulez-vous faire
C : transfert d'appel
 A : transfert d'appel vous pouvez demander à transférer vos appels ou bien désactiver le transfert en cours enregistrer un nouveau numéro privilégié ou encore modifier ou supprimer
C : [raccroche]

4.1.4. S'ajuster dans le flou

La place nous manque pour entrer dans le détail et nous renvoyons pour plus de précisions à (Velkovska and Beaudouin, 2009) où sont explicitées les difficultés liées à l'ajustement flou entre l'automate et le client. Supposons que dans les systèmes de dialogue, le principe de coopération, qui, selon (Grice, 1975), régit la conversation, soit à l'œuvre. Les quatre maximes qui selon Grice gouvernent la conversation devraient être à l'œuvre : la maxime de quantité, qui pose qu'un tour de parole doit apporter autant d'information qu'il est requis, et pas plus, la maxime de qualité qui suppose que la contribution soit véridique, la maxime de relation qui pose qu'un tour doit être pertinent par rapport au précédent et enfin la maxime de modalité, qui implique la clarté des énoncés. Par rapport à ces principes de base, la violation des règles dans nos interactions porte principalement sur la maxime de quantité – l'automate est tantôt trop elliptique, tantôt trop bavard et sur la maxime de relation – la réponse de l'automate n'est souvent pas pertinente pour le client et inversement le client est parfois non pertinent pour la machine.

Le non respect des maximes conversationnelles crée des flottements dans les ajustements réciproques et donc dans la co-production du service.

Dans l'exemple ci-dessous, l'automate ne prend pas la peine d'expliquer le lien entre la demande du client et sa réponse : pour pouvoir accéder au service facturation, il faut en effet que la ligne soit abonnée au service. La machine est trop elliptique. Par ailleurs, sa manière de clôturer l'interaction en retournant à l'accueil est incompréhensible pour le client.

Exemple 7 : Une réponse de l'automate trop elliptique, n'explicitant pas le lien.

C : j'aurais voulu avoir (.) le service facturation
 A : [plusieurs services peuvent répondre&
C : [(les changements)
 A : &désolée mais la ligne XX XX XX XX XX n'est abonnée à aucun service modifiable depuis le service A\ je retourne à l'accueil, prononcez le nom du service

4.2. Maîtriser la machine : le travail du client

S'il est devenu commun de parler de co-construction dans la relation de service, voire du travail du consommateur dans la relation de service (Dujarier, 2008), l'écoute des interactions montre qu'il y a du côté des utilisateurs une grande variabilité dans la maîtrise du dispositif technique : plus le client prend en charge le travail de gestion de l'interaction, plus les chances que le service soit rendu augmente, et cela qu'il s'agisse du service A ou B.

L'utilisateur profane a du mal à trouver sa place, il intervient trop tôt ou trop tard, hésite. Il n'interrompt pas la machine, suit ses propositions même si elles ne sont pas pertinentes. Il

tâtonne pour retrouver le nom du service ou pour exprimer ce qu'il cherche. Il suit de manière disciplinée ce que lui demande la machine et a tendance à lui parler en « langage naturel ».

L'expert au contraire trouve sa place sans hésitation, n'interrompt pas pendant l'ouverture, interrompt immédiatement dès que la place est libre. Il connaît le nom de ce qu'il cherche, manifeste un savoir-faire dans l'enchaînement des actions à faire, comme s'il s'était construit une représentation de la séquence des actions, et enfin il n'intervient que par mots-clefs, voire en utilisant les touches.

Ces deux usages extrêmes signent des formats de coopération très contrastés. Du côté des usages profanes, les cas de désalignement entre la machine et le client sont très nombreux, la machine se met à bavarder excessivement et le taux d'échec devient très élevé. Inversement, pour les usages experts, l'alignement se fait correctement et le service est rendu en temps réduit. Le cheminement de l'usage profane à l'usage expert correspond à un travail d'apprentissage et de maîtrise de la machine. On est là face à une situation asymétrique d'une machine non apprenante face à un humain qui apprend et gagne en autonomie face à la machine.

5. Conclusion

L'analyse statistique de la base nous a permis d'identifier quelques facteurs qui favorisent la réussite ou l'échec des interactions vocales entre un humain et une machine dans le cadre d'une relation de service. Mais seule une analyse détaillée des interactions retranscrites permet de comprendre les difficultés que pose l'alignement entre un homme et une machine dans la construction d'un service. Nous avons en particulier noté que derrière des échecs peuvent se cacher des discordances de vision du monde, des problèmes d'identification de « place » dans l'interaction, des problèmes de pertinence relâchées dans les échanges. On voit surtout que la félicité de la relation nécessite une certaine connaissance du système par l'utilisateur.

Nous avons été ici confrontées à une situation inédite. Dans cette base, on a à la fois le matériau brut (l'enregistrement audio de l'appel) et des traits techniques assez éloignés du matériau linguistique. Il nous semble que le dispositif technique d'enregistrement des interactions homme-machine devra tenter de combler ce fossé, en introduisant davantage de paramètres linguistiques et pragmatiques, comme des retranscriptions annotées, rendant compte du déroulement du dialogue.

Ce premier papier sur l'analyse d'interactions entre humains et automates suggère des pistes pour la constitution de futures bases de données d'interactions. Premièrement, il faudrait avoir un enregistrement complet des appels, avec les deux locuteurs. Deuxièmement, il est important d'avoir une retranscription des interactions, intégrant les questions de chevauchement. Pour cela, il faut prévoir de combiner la retranscription du discours du client avec des techniques de reconnaissance vocale et la synthèse vocale de la machine. Il est important de pouvoir mesurer l'écart entre ce que dit le client et ce qu'entend la machine. Enfin, pour améliorer les dispositifs, il pourrait être pertinent de demander au client en fin d'échange le motif de son appel et si sa demande a abouti. Ces éléments permettront de progresser dans les techniques de dialogue homme-machine.

Références

Baker C., Emmison M. and Firth A. (2005). *Calling for Help. Language and social interaction in telephone helplines*. Amsterdam-Philadelphia : John Benjamins.

- Bandt J. d. and Gadrey J. (editors) (1994). *Relations de services, marchés de services*. Paris : CNRS
- Bickmore T. and Cassell J. (2005). Social Dialogue with Embodied Conversational Agents. In Kuppevelt, J. v., Dybkjaer, L. and Bernsen, N., *Advances in Natural, Multimodal Dialogue Systems*, New York: Kluwer Academic, pp. 23-54.
- Cossalter C. (2002). Les plateaux téléphoniques dans les banques et les assurances. Continuité et rupture dans la dynamique d'organisation. *Réseaux*, 20, 114 : 23-50.
- Dujarier M.-A. (2008). *Le travail du consommateur. De McDo à eBay : comment nous coproduisons ce que nous achetons*. Paris: La découverte.
- Flichy P. and Zarifian P. (editors) (2002). *Les centres d'appels*. Réseaux-Paris : Hermès Sciences
- Gadrey J. (2003). *Socio-économie des services*. Paris : La Découverte.
- Gadrey J. and Zarifian P. (2002). *L'émergence d'un modèle du service : enjeux et réalités*. Paris : Liaisons.
- Garfinkel H. (2007 [1967]). *Recherches en ethnométhodologie*. Paris : PUF.
- Goffman E. (1968). *Asiles*. Paris : Editions de Minuit.
- Grice H.P. (1975). Logic and Conversation. In *Syntax and Semantics, Speech Acts*, London: Academic Press, III, pp. 41-58.
- Hutchby I. (2001). *Conversation and technology : from the telephone to the internet*. Cambridge : Polity Press.
- Lebart L. and Salem A. (1994). *Statistique textuelle*. Paris : Dunod.
- Licoppe C. and Relieu M. (2005). Entre système et conversation. Une approche située de la compétence des téléopérateurs dans les services d'assistance technique. In Kessous, E. and Metzger, J.-L., *Le travail avec les technologies de l'information*, Paris : Hermes-Lavoisier, pp. 117-199.
- Mounier C. (2002). Du statut d'agent de réservation à celui de conseiller client. Analyse d'une transition professionnelle dans deux centres d'appels. *Réseaux*, 20, 114 : 153-177.
- Rastier F. (1991). *Sémantique et recherches cognitives*. Paris : PUF.
- Reinert M. (1990). ALCESTE : Une méthodologie d'analyse des données textuelles et une application : Aurélia de Gérard de Nerval. *Bulletin de méthodologie sociologique*, 26 : 24-54.
- Reinert M. (1993). Les «mondes lexicaux» et leur logique. *Langage et société*, 66 : 5-39.
- Sacks H., Schegloff E.A. and Jefferson G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50 : 696-735.
- Sadek D. (1999). *Design considerations on dialogue systems: from theory to technology - the case of Artemis*. IDS-99, 173-187.
- Suchman L. (2007). *Human-Machine Reconfigurations. Plans and situated actions 2nd edition*. Cambridge : Cambridge University Press.
- Velkovska J. and Beaudouin V. (2009). Alignments and misalignments in human-machine speech interaction: losing some relevance tiens between turns. In *11th International Pragmatics Conference*, Melbourne, Australia.
- Wooffitt R. and MacDermond C. (1995). Wizards and Social Control. In Thomas, P., editor, *The Social and Interactional Dimensions of Human-Computer Interfaces*, Cambridge : Cambridge University Press, pp. 126-141.
- Zimmerman D.H. (1992). The interactional organization of calls for emergency assistance. In Drew, P. and Heritage, J., editors, *Talk at Work*, Cambridge : Cambridge University Press, pp. 418-468.