



HAL
open science

Interagir au quotidien avec un assistant vocal intelligent à la maison : effets sur la compétence interactionnelle des utilisateurs et potentiel éducatif dans des familles plurilingues

Maud Ciekanski

► To cite this version:

Maud Ciekanski. Interagir au quotidien avec un assistant vocal intelligent à la maison : effets sur la compétence interactionnelle des utilisateurs et potentiel éducatif dans des familles plurilingues. Drôles d'objets, May 2023, Nancy, France. 10.5281/zenodo.12784078 . hal-04655367

HAL Id: hal-04655367

<https://hal.science/hal-04655367v1>

Submitted on 22 Jul 2024

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.

Interagir au quotidien avec un assistant vocal intelligent à la maison : effets sur la compétence interactionnelle des utilisateurs et potentiel éducatif dans des familles plurilingues

Maud Ciekanski

Laboratoire ATILF, Université de Lorraine & CNRS.

Résumé

Les assistants vocaux intelligents (AVI) ont pris place dans nos foyers depuis une dizaine d'années. Basés sur des algorithmes d'apprentissage machine, ils ont pour fonction de faciliter certaines tâches quotidiennes grâce à une interaction vocale « naturelle » proche de la conversation humaine. Bien que de nombreuses études montrent que cette apparence de conversation naturelle est source de difficultés pour les utilisateurs, d'autres travaux soulignent les effets d'une utilisation soutenue des AVI sur le développement des compétences langagières, notamment chez les publics à besoins spécifiques comme les enfants ou les apprenants de langue, dans le cadre de leur formation ou en situation informelle. Par ailleurs, la compétence interactionnelle est une compétence complexe souvent difficile à acquérir par les allophones, même en situation d'immersion. Les AVI offrent-ils des possibilités complémentaires d'interaction langagière bénéfiques pour progresser en langue ?

Nous documentons l'usage des AVI domestiques dans des foyers plurilingues au sein d'activités quotidiennes individuelles ou collectives et les effets de ces interactions sur la compétence interactionnelle des utilisateurs, dont certains ont le français comme langue d'usage. Il en ressort qu'interagir avec les AVI est une opération complexe nécessitant de multiples ajustements dus à la double illusion d'un modèle coopératif et d'une langue-culture commune qui s'inscrivent dans des problématiques de « globalisation » (Robertson, 1992 ; Roudometof, 2021) des langues, de la culture et de leur circulation.

Mots-clés : assistants vocaux intelligents ; famille plurilingue ; apprentissage des langues en milieu naturel ; compétence interactionnelle.

Introduction

Les Assistants Vocaux Intelligents (désormais AVI) généralistes (ex. *Alexa*, *Google Home*, *Siri*, selon les marques) ont pris place progressivement dans les foyers depuis 2015¹. Basés sur des algorithmes d'apprentissage machine, ces objets ont pour fonction de faciliter certaines tâches quotidiennes grâce à une interaction vocale « naturelle » proche de la conversation humaine (Collins, 2018). L'intelligence conversationnelle du système repose sur sa capacité à comprendre les énoncés des utilisateurs à partir de mots-clés et à produire des réponses verbales pertinentes en recourant à l'Intelligence Artificielle ou à des pré-programmes sélectionnés. Les AVI traitent une grande variété de requêtes pour des activités de loisirs, des pratiques informationnelles ou des commandes domotiques. Sur sollicitation de l'utilisateur, ils produisent également des énoncés conversationnels

¹ Selon CSA-Hadopi (2019), 11 % (1,7 millions de personnes) de la population étaient équipés d'une enceinte connectée toute marque confondue, en France début 2019 (étude sur la population d'internautes de 15 ans et plus).

informels de type « small talk », comme le fait de raconter des anecdotes ou des blagues selon l'éphéméride par exemple, à partir de discours pré-fabriqués, et qui contribuent à l'illusion d'une conversation personnalisée, voire d'une certaine compétence sociale en créant une « communion phatique » (Malinowski, 1972) avec les utilisateurs. Ce sont à la fois des machines capables de remplacer divers instruments (calculatrice, station météo, radio, minuteur...) et des agents conversationnels.

De nombreux retours d'utilisateurs montrent pourtant que cette apparence de conversation naturelle est souvent source de difficultés (Gauducheau, Marcoccia, 2022), occasionnant des problèmes interactionnels qui nécessitent une « négociation » de sens et de forme, avec la mise en place de diverses remédiations par les utilisateurs (reformulation, répétition, changement de langues...) et par l'AVI (demande de confirmation, de répétition...) pour la bonne exécution des requêtes. Ainsi, le fait que les requêtes passent par des interactions orales impliquerait la maîtrise de nouvelles compétences langagières renvoyant à l'art de parler à un agent artificiel. En outre ces technologies ont d'abord été conçues pour des utilisateurs anglophones dans un environnement américano-centré. Si la traduction automatique permet à l'AVI de fonctionner dans d'autres langues que l'anglais, le fonds culturel reste par défaut majoritairement états-unien, ce qui suscite également de nombreux malentendus. Malgré ces contraintes et limites, certains travaux soulignent le potentiel des AVI pour le développement de la compétence langagière, voire interactionnelle, des locuteurs natifs d'autres langues, échangeant en langue-cible avec ces agents conversationnels sollicités à l'envi, dans un milieu familier. Ces nouvelles interactions du quotidien sembleraient propices à la mise en place d'apprentissages informels relevant « de l'ordre de la socialisation ou de l'acculturation » (Brougère, 2009, p. 24).

Dans quelle mesure ces pratiques langagières d'un nouveau type permettent-elles le développement de la compétence interactionnelle, c'est-à-dire les savoirs et savoir-faire qui régissent un engagement approprié dans les échanges sociaux, d'un point de vue linguistique (verbal et non-verbal), pragmatique et situé (André, 2021) ? Nous questionnons les effets de ces interactions répétées avec un AVI, dans des activités langagières du quotidien, sur le développement des compétences interactionnelles de locuteurs au sein de familles à plusieurs langues, où le français est maîtrisé de façon variable selon les membres. Nous documentons des pratiques peu étudiées au regard des AVI : les publics non-natifs, les interactions en français et des pratiques collectives informelles.

1. Interagir avec un AVI : « qu'est-ce qui se passe ici ? »

Dans *les cadres de l'expérience* (1974, p. 21), Goffman souligne par la formule « qu'est-ce qui se passe ici ? » le malaise que l'on ressent lorsque nos interactions dérogent à l'ordre de l'interaction, sortent des normes qui régissent et régulent les actions humaines. Interagir avec un AVI, est-ce un semblant d'interaction, une fabrication, voire une illusion ou s'agit-il d'un nouveau cadre de l'expérience ?

1.1. Les enceintes connectées : un alter d'un nouveau type ?

Les drôles d'objets de cette étude sont les enceintes connectées à commande vocale dotées d'un appareil d'écoute permanent. Il s'agit de petits objets au design arrondi, de couleurs et textures variées, se fondant dans le décor de nos intérieurs. Ces objets multifonctionnels trouvent place dans les différentes pièces de l'habitat : cuisine, salle de bain, pièce à vivre... et il n'est pas rare d'en trouver plusieurs par foyer. Interactifs, ces objets réagissent à leur nom (ex. « Alexa » pour l'AVI d'Amazon), manifestent leur état d'activité de façon auditive (voix ambiante désincarnée) et visuelle (cercles lumineux), et sont dotés d'une certaine agentivité : ils font ce qu'on leur demande, font à notre place et parfois font ce qu'on ne leur demande pas et ne font pas toujours ce qu'on leur demande. Leurs manifestations, pas toujours prévisibles, contribuent à les doter d'une certaine altérité : ils revêtent à la fois des traits propres aux caractéristiques humaines (le langage naturel ou certaines conventions sociales) et se conforment à nos représentations protéiformes de l'altérité : il semble naturel de parler à une machine qui répond. En cela, ces objets mobilisent nos capacités à ajouter une dimension symbolique à leur fonctionnement.

Ces agents conversationnels à l'apparence d'objet empruntent des traits anthropomorphiques participant de leur personnification : ils ont un nom, un genre, et assurent des fonctions conversationnelles qui s'inscrivent dans une forme de relation sociale complexe, à la fois relation hiérarchique (il s'agit d'assistants) et relation de proximité (l'assistant connaît nos goûts et peut personnaliser ses réactions en fonction de l'interlocuteur...), créant l'illusion d'une relation interpersonnelle. Le design anthropomorphique de la voix joue un rôle essentiel pour la bonne acceptation de l'AVI par les utilisateurs. L'interface vocale décrit la personnalité de l'AVI : son ton, son champ lexical, son attitude, voire son apparence et son histoire imaginaires. Ces différents aspects jouent sur notre désir de les utiliser dans des expériences répétées et participent de l'illusion d'interaction que l'on peut ressentir, rendant ainsi poreuse la distinction entre les notions d'*interaction* réservée aux échanges entre humains et d'*interactivité* pour les échanges avec les machines. Le recours au langage ordinaire pour commander un objet transforme inévitablement nos relations aux objets. Le cerveau humain activerait les mêmes zones lorsque nous dialoguons avec un être humain ou avec un programme conversationnel (Nass, Brave, 2005). Les interfaces conversationnelles amèneraient donc les utilisateurs à se représenter des simulacres de sujet à partir d'un imaginaire partagé. Pour autant, si les utilisateurs mobilisent des pratiques langagières ordinaires, les réponses produites par les AVI n'ont pas un statut conversationnel dans la mesure où elles n'émergent pas de manière située, mais sont préconfigurées par les concepteurs (Porcheron *et al.*, 2018). Le fonctionnement de ces dispositifs est principalement basé sur le principe de la commande vocale à deux tours (requête de l'utilisateur et réponse du système) précédée par un mot d'activation comme formule d'adresse à l'objet. Néanmoins l'enjeu est qu'ils puissent générer aussi des énoncés susceptibles d'intéresser les locuteurs et de respecter *a minima* des règles humaines de conversation comme la séquentialité de l'interaction en tours de parole ou une certaine socialité (formules de politesse en ouverture et clôture des échanges). Il s'agirait donc d'une forme de

« phénomènes interactionnels hybrides » entre conversation naturelle et interactivité vocale avec une machine (Velkovska, Zouinar, 2018).

1.2. Parler à un AVI, est-ce parler à un mur ?

Suivant Goffman, « l'interaction est un processus temporel et séquentiellement ordonné qui prend place lorsqu'au moins deux individus se trouvent dans un espace perceptuel partiellement partagé, dans lequel ils recourent à des ressources sémiotiques, en vue de conduire collectivement une action conjointe, elle-même indexée à des pratiques sociales historiquement et culturellement constituées » (Filliettaz, 2022, p. 278). Interagir avec un AVI implique la mise en place de stratégies interactionnelles particulières chez les utilisateurs du fait de la double illusion d'une langue-culture commune et d'un modèle coopératif.

Les AVI commerciaux les plus fréquents sont conçus pour être utilisés par des locuteurs natifs anglophones. L'anglais est donc la langue-pivot des outils de traitement automatique des langues qu'ils utilisent. Grâce aux techniques de traduction neuronale, ils peuvent cependant s'adapter à certaines contraintes en termes d'accents, de politique de genre ou de codes culturels. Larssonneur (2021) montre que malgré le panel de langues disponibles, les choix opérés par les AVI répondent à une vision du monde américano-centrée, que ce soit en termes de corpus de référence ou de censure au sein des contenus collectés. Cela contribue à la fabrique d'un éthos discursif singulier (Amossy, 1999), à une standardisation de la langue, une standardisation des usages de la langue, au sein de chaque langue employée et par calque vis-à-vis de l'anglais américain. Peu d'études à notre connaissance questionnent les savoirs et les représentations du monde dont sont porteurs les AVI. Ils conditionnent pourtant le contexte de compréhension et d'interprétation des utilisateurs en façonnant le terrain commun, l'ensemble des informations partagées et mutuellement reconnues comme telles par les interlocuteurs nécessaire à toute interaction.

En outre, la notion d'alignement est constitutive de notre compréhension des dynamiques à l'œuvre dans les interactions. Le déploiement dynamique de l'interaction a été étudié principalement sous l'angle de la collaboration ou de la convergence. La séquence minimale constituée par la « paire adjacente » est emblématique de l'interaction comme activité collaborative au cours de laquelle les participants adaptent leurs comportements l'un à l'autre en fonction de normes et d'attentes spécifiques. Comment comprendre alors les interactions avec un AVI ? S'agit-il de deux trajectoires interactionnelles (celle de l'utilisateur et celle de l'AVI) ou bien s'agit-il d'une seule trajectoire dans laquelle se confrontent parfois deux positionnements spécifiques ? Cela pose la question de l'orchestration de cette interaction conversationnelle plus ou moins alignée qui incombe à l'utilisateur.

Au regard de ces éléments, la situation de communication proposée par les AVI s'apparenterait à une communication exolingue « qui s'établit par des moyens autres qu'une langue maternelle commune aux participants » (Porquier, 1984, p.18) qui implique un travail accru d'intelligibilisation entre les partenaires. En outre, il s'agit de situations favorisant les malentendus qui ont pour origine le double codage d'une même

réalité, c'est-à-dire les interprétations/compréhensions contradictoires d'un même énoncé. En raison de l'asymétrie linguistique et culturelle entre les partenaires conversationnels, la communication exolingue exige autant d'efforts coopératifs linguistiques qu'extralinguistiques pour assurer la construction commune de l'interaction. Cette co-construction est-elle envisageable avec un AVI ?

1.3. Développer une compétence interactionnelle en interagissant avec un AVI

Les AVI offrirait cependant de nouvelles opportunités d'interaction en milieu naturel pour pratiquer la langue dans des activités du quotidien. Les AVI s'inscrivent dans l'essor plus général des outils numériques à interface vocale pensée pour faciliter leur usage (le mode oral est perçu comme naturel, rapide et économique car il étend le champ d'action). Cette interface vocale représente également un défi pour les utilisateurs allophones qui doivent s'adapter à la situation communicative et maintenir l'intelligibilité du message malgré les obstacles langagiers.

La recherche sur les AVI génériques dans le contexte de l'apprentissage d'une langue additionnelle est encore peu documentée. La majorité des études concernent l'anglais langue étrangère, situation linguistique dans laquelle le rôle de l'AVI est assimilable à celui d'un natif. Elles convergent sur le fait que les AVI (*Alexa* dans le cas de Dizon, Tang, 2020) sont perçus par les étudiants comme un outil générant un input pertinent en L2, donnant un feedback sur l'intelligibilité de la communication et en particulier sur celle de la prononciation, permettant de pratiquer la langue dans un environnement non-stressant (Moussalli, Cardoso, 2020). Les AVI sont vus comme un partenaire d'apprentissage et un tuteur infatigable disponible à tout moment et en tout lieu, en particulier dans les contextes et les milieux où l'éducation formelle et l'accès à des locuteurs natifs ne sont pas envisageables (Xiao *et al.*, 2023). L'étude expérimentale de Tai (2022) a comparé les effets de la conversation Agent conversationnel-Humain et Humain-Humain sur l'apprentissage de la L2. Le groupe avec l'AVI a obtenu des résultats équivalents à celui qui a conversé avec des natifs en termes de fluidité, contenu, vocabulaire, prononciation et grammaire. Toutefois, d'autres études montrent que les difficultés de l'AVI à reconnaître des énoncés moins standardisés en termes de prononciation, de structuration, de disflue, etc. génèrent plus d'échecs de communication. Ces échecs façonnent les perceptions qu'ont les apprenants de leur compétence langagière et affectent les résultats de l'apprentissage (Wu *et al.*, 2020). En outre, la compréhension mutuelle entre l'agent conversationnel et les locuteurs non natifs dépend fortement de facteurs contextuels, tels que le niveau de compétence de l'apprenant, son âge et le type de tâche dans laquelle l'agent conversationnel est intégré (Chen *et al.*, 2020).

Dans quelle mesure les interactions répétées avec un AVI aideraient-elles les utilisateurs qui recourent à une langue additionnelle pour interagir avec ces objets à développer une compétence d'interaction pertinente ? Nous entendons la compétence d'interaction en L2 comme le développement d'habiletés sociolinguistiques et l'émergence d'un répertoire de procédures interactionnelles permettant aux locuteurs de synchroniser leurs conduites et de les adapter aux besoins de l'interaction afin d'accomplir les actions visées. Interagir avec un AVI permet-il des transferts pertinents pour les interactions humaines ?

2. Méthodologie de la recherche

2.1. Cadre de l'étude

Notre étude s'intéresse à l'usage des AVI du point de vue de la sociologie des usages familiaux afin de comprendre les dynamiques d'usage individuelles et collectives, voire collaboratives, qui se dessinent avec et autour des AVI dans des tâches de la vie quotidienne qui mobilisent des interactions langagières. Nous nous sommes intéressée à des foyers « à plusieurs langues » (Zeiter, 2018) afin d'étudier les effets potentiels de ces interactions répétées avec un AVI sur la compétence interactionnelle des utilisateurs, notamment ceux qui ont le français comme langue d'usage. Notre parti pris était de privilégier les observations écologiques menées en milieu naturel, dans la sphère domestique comprise comme un espace d'apprentissages notamment langagiers. Le quotidien familial dans ses routines, ses rituels, ses activités diverses sans visées pédagogiques explicites contient un potentiel éducatif par un processus de co-production (Pain, 1990).

En outre, comme le souligne Simonian (2020), si la fonction d'un artefact est de fluidifier un mode d'organisation et de production de l'activité humaine, tout en créant une porosité entre connaissances et compétences nécessaires à sa réalisation, elle est aussi l'affaire de normes et de valeurs puisqu'elle porte en elle des processus de significations socioculturelles qui sont le produit d'une expérience individuelle et collective. Son étude implique une approche holistique afin de capter des faits empiriques qui permettent par cumulativité, confrontations, comparaisons de formaliser les pratiques d'usage et d'utilisation des AVI. Nous avons ainsi observé sur un mois les pratiques langagières avec et autour de l'AVI dans le microsystème familial. Nous souhaitions pouvoir identifier des dynamiques d'ajustement et de régulation, des évolutions dans les façons d'interagir et d'apprendre à interagir avec l'AVI.

2.2. Population d'étude

Les familles participant à l'étude ont été sollicitées sur la base de deux critères : l'utilisation en famille d'un AVI depuis au moins un an et la présence au sein de la famille d'au moins un membre allophone ayant le français comme langue d'usage. Notre échantillon d'étude comprend ainsi trois familles au profil homogène et dont le comportement langagier et interactionnel avec un AVI est en partie conditionné par deux facteurs : un positionnement social particulier (il s'agit de foyers dont les adultes relèvent d'une CSP+), et un rapport favorable aux technologies (les adultes qui ont acheté l'AVI travaillent en robotique, en informatique ou avec de nombreux outils connectés). Ces familles sont composées de trois à quatre personnes dont des enfants (de 10 à 11 ans) ou des adolescents (de 13 à 17 ans). Elles font un usage fréquent et régulier de l'AVI depuis trois à cinq ans. Cette pratique régulière dans le temps donne à lire des usages stabilisés générant possiblement des routines conversationnelles (Conein, 1998) qui se sont construites au fur et à mesure des interactions avec l'AVI au gré des utilisations que l'on en fait. En outre, les familles partagent des similitudes en termes de langue-culture

puisque le membre allophone du foyer est d'origine latino-américaine, résidant en France depuis 10 à 30 ans. Chaque membre allophone travaille dans un milieu multilingue français-anglais et a le français comme au moins l'une des langues d'usage dans le foyer (Tableau 1).

Familles composant le corpus d'étude	Langues du répertoire langagier des familles, par ordre de fréquence d'utilisation déclarée par les membres de la famille
FAM 1	français, anglais, espagnol
FAM 2	espagnol, français, basque, anglais
FAM 3	français, espagnol, anglais

Tableau 1. Répertoire langagier des familles étudiées.

Il s'agit en outre de personnes, enfants, adolescents et adultes, qui utilisent fréquemment les commandes vocales sur leurs appareils connectés (smartphone, tablette...) dans les langues de leur répertoire langagier. Si les familles partagent le même degré de connaissances techniques de l'artefact et ont des usages principaux similaires (musique, météo, information), elles diffèrent cependant dans les modes d'utilisation (plus ou moins individuel ou collectif, plus ou moins en français) et les modes de perception de l'artefact (un outil ou un alter). Selon cette perception plus ou moins symbolique de l'objet, les membres du foyer entretiennent des rapports plus ou moins socialisants avec l'AVI.

2.3. Protocole de recherche

Tracer le développement de la compétence d'interaction consiste à examiner des microcosmes actionnels à différents moments dans le temps afin d'identifier les moyens ou méthodes déployés par les locuteurs pour les gérer. Nous avons pour cela eu recours à une méthodologie mixte croisant :

- une approche éémique, captant le point de vue du sujet (première personne) avec une enquête initiale auprès de l'ensemble des participants sous forme d'entretien individuel semi-guidé d'une heure. Il s'agissait de faire connaissance avec les familles, de documenter les usages déclarés des AVI dans les foyers, la perception de l'AVI par chacun (femme, robot, machine, outil...), la ou les langues utilisées en fonction des tâches demandées, les difficultés de communication rencontrées et les stratégies interactionnelles mises en œuvre. Dans la mesure où la verbalisation de ses activités par chaque interviewé est nourrie par différents vécus, il était intéressant de réaliser ce recueil de déclaratif in situ dans l'espace de vie, parfois en demandant des simulations d'utilisation ;
- une approche éthique, apportant un éclairage plus extrinsèque par le recueil des interactions avec l'AVI sous la forme d'enregistrements oraux pendant un mois. Les enregistrements déclenchés par les participants ont tous eu lieu dans la pièce à vivre du logement où était connecté l'AVI. Le recueil d'interactions orales a été complété par l'historique de l'AVI qui donne à lire les interactions telles qu'elles ont été perçues par la machine. Il est ainsi possible de reconstruire l'origine de certains malentendus dans les interactions. Cet historique des échanges se présente sous la forme de transcriptions

écrites organisées de façon ante-chronologique, indiquant l'heure, le jour des échanges et l'identité de l'utilisateur si celle-ci a fait l'objet d'un réglage. Les participants avaient la possibilité de supprimer de l'historique les interactions qu'ils ne souhaitent pas divulguer avant sa transmission. Aucun ne l'a fait.

Croiser l'enregistrement des échanges avec l'historique permet de comprendre finement les stratégies phonologiques et prosodiques spécifiques qui peuvent être employées par les utilisateurs (chuchoter, parler moins vite, articuler davantage...), la façon dont les interlocuteurs traitent les situations d'incompréhension et comment ils gèrent l'artificialité des interactions pour parvenir à leur fin. À partir de l'analyse conversationnelle des enregistrements et des historiques (Kerbrat-Orecchioni, 2005), nous avons cherché à décrire les dynamiques interactionnelles mises en œuvre par les interactants (utilisateurs et AVI). Nous nous sommes intéressée en particulier à la gestion de l'inférence conversationnelle (Gumperz, 1989) particulièrement problématique avec les AVI. En effet, tout locuteur indique, de façon directe ou implicite, la manière dont un énoncé doit être interprété et montre par ses réponses, verbales et non verbales, comment il a interprété l'énoncé d'un autre locuteur.

Nous rendons compte dans cette micro-étude de l'analyse des interactions dans un seul des foyers étudiés (FAM3) car il constitue le recueil de données le plus complet. Il contient plus de 30 heures d'enregistrement audio, principalement au moment des repas, complétées par l'historique exhaustif des échanges transcrits par l'AVI. Contrairement aux deux autres corpus, FAM3 contient une plus grande quantité et variété d'interactions enregistrées avec l'AVI, ainsi qu'une plus grande quantité d'interactions en français, malgré les problèmes de communication rencontrés.

3. Résultats

3.1. La perception de l'AVI par FAM3

FAM3 rend compte des pratiques interactionnelles d'un collectif de quatre personnes (deux adultes, un adolescent, un enfant) interagissant depuis 5 ans avec l'AVI, soit en mode individuel (les activités les plus fréquentes listées lors de l'entretien sont l'écoute de musique, de la radio, questions/réponses au moment des devoirs notamment, socialisation avec l'AVI), en présence ou non d'autres membres de la famille, soit en mode collectif (les activités les plus fréquentes listées lors de l'entretien sont l'écoute de musique, les jeux, la météo, les questions/réponses notamment au moment des conversations dans le temps du repas). Les enregistrements audio ont eu lieu au moment du repas du soir et ne rendent donc pas compte de l'exhaustivité des utilisations de l'AVI. Contrairement à d'autres études portant sur les AVI en contexte familial (Porcheron *et al.*, 2018), il n'y a pas de prévalence de l'acheteur de l'AVI dans les utilisations qui en sont faites. Il ressort plutôt une distribution homogène et collégiale entre les différents membres de la famille, chacun ayant les mêmes droits d'activation et de désactivation de l'AVI, notamment lors des utilisations en présence des autres membres de la famille. Les utilisations les plus récurrentes concernent un nombre plutôt limité de fonctionnalités (musique, radio, alarme, météo, jeux, recherche d'information...) Certaines tendent à se

ritualiser (jeux ou recherches d'information au moment du repas), par exemple « on a fait la question du jour ? ». Toutefois, on note également des utilisations spécifiques selon les utilisateurs : seul l'enfant active la narration d'histoires, de blagues ou converse avec l'AVI (« est-ce que tu es mon amie ? », « bonjour »). Les activités langagières de socialisation sont ainsi plus fréquentes chez l'enfant qui est le seul interactant qui s'approche fréquemment de l'AVI pour lui parler. Si l'on constate une grande similarité dans les tâches réalisées avec l'AVI au sein de la famille, on remarque également des modes d'interaction différents selon les utilisateurs, qui utilisent plus ou moins de marqueurs de politesse par exemple. Tous les membres n'ont pas le même taux de réussite lors de leurs interactions avec l'AVI et ce indépendamment de leur compétence langagière : certains ont davantage de problème à être entendu (c'est-à-dire à activer l'AVI), d'autres à ce que l'AVI discrimine correctement l'énoncé prononcé, d'autres enfin à ne pas être interrompu par l'activation intempestive de l'AVI lors de leur tour de parole. Tous ces problèmes peuvent créer des impasses conversationnelles. Cela génère des trajectoires interactionnelles différentes au sein du foyer, donnant lieu à des dynamiques collaboratives entre les interlocuteurs pour assurer la réalisation des tâches langagières. Les entretiens et les enregistrements montrent aussi que l'AVI n'est pas seulement perçu comme un simple artefact mais peut donner lieu à des utilisations complexes. Il joue fréquemment un rôle de tiers, d'arbitre lors de conversations où s'affrontent plusieurs points de vue (même si les sources citées sont parfois jugées comme peu pertinentes et suscitent un regard critique, voire de la dérision chez les utilisateurs). Dans ce cas, on constate d'après Rabardel (1995) que l'AVI est intégré dans des schèmes d'action instrumentée où la réalisation de la tâche (convaincre son interlocuteur) est assistée par l'artefact qui revêt un statut d'autorité. De nombreuses interactions dans le corpus témoignent d'un rapport de confiance, voire de connivence avec l'AVI. Pour autant, les déclenchements de l'AVI non sollicités (auto-activation d'un signal visuel de fonctionnement) donnent lieu systématiquement à des verbalisations de défiance (ex. « tu nous espionnes ? »), témoignant d'un statut asymétrique par défaut dans les relations entre les utilisateurs et l'AVI dont les manifestations d'agentivité sont interprétées comme suspectes, voire menaçantes pour les utilisateurs.

Enfin, les entretiens individuels ont montré que le caractère féminin de la voix n'a pas d'incidence déclarée sur la représentation que les utilisateurs se font de l'AVI et sur leur utilisation ou mode d'interaction. L'envie d'utiliser l'AVI est motivée par le ton naturel et amical de l'interface vocale de l'AVI (plutôt qu'une voix robotique) et par l'expertise de l'AVI qui couvre une grande quantité d'informations aux niveaux international et local (le membre allophone apprécie que les références ne soient pas franco-centrées). L'enfant perçoit l'AVI comme un adulte différent de l'ordinaire avec lequel on éprouve moins d'insécurité langagière (« j'ai moins peur de parler avec elle » ; « elle ne me reprend pas si je fais des fautes »). Les utilisateurs de FAM 3 ont majoritairement une image positive de l'AVI qui influence leurs comportements interactionnels.

3.2. Description quantitative des interactions dans FAM3

D'après l'historique de l'AVI, le corpus FAM3 comptabilise 1200 tours de parole sur un mois d'utilisation. Les interactions recueillies comprennent majoritairement des paires adjacentes (ex. requête et exécution) mais aussi des échanges plus complexes, nécessitant plusieurs tours de parole pour la réalisation de la tâche. Il s'agit majoritairement d'énoncés courts qui génèrent soit une réaction par l'AVI (comportement interactif de l'artefact) ou une réponse (un énoncé) de l'AVI (comportement interactionnel de l'agent conversationnel). On dénombre ainsi 180 réactions non verbales (15 % du corpus FAM3) pour 1020 réponses verbales (85%) par l'AVI. Il en ressort que l'AVI est principalement sollicitée dans des activités langagières.

Les interactions avec l'AVI génèrent de nombreux problèmes de communication. On dénombre 26 bugs techniques sur l'ensemble des utilisations (non-activation de l'AVI suite à une sollicitation, répétitions en boucle, interruption impromptue de la communication, envoi d'un signal sonore manifestant la non-prise en compte de la requête...). Les échecs de communication (la non-réalisation de la requête) représentent 25 % des utilisations. Cela contribue à une histoire conversationnelle faite d'impasses et à la mise en place de comportements langagiers visant à anticiper ou contourner les problèmes. Les entretiens semi-dirigés montrent que les utilisateurs surestiment légèrement plus ce taux d'échec de communication à environ un tiers des échanges. On est assez loin des discours publicitaires vantant la capacité des AVI à interagir de façon « naturelle, fluide, simple et sans effort ». Au contraire, le corpus FAM3 met en exergue tout le travail réalisé par les utilisateurs pour parvenir à leur fin, ce qui génère un effort cognitif et langagier non anticipé. Ainsi, confrontés à un échec de communication, les interlocuteurs de FAM3 réessaient entre cinq à sept fois en moyenne avant d'abandonner leur requête. Ils correspondent plutôt à un profil persévérant qui se retrouve chez les différents membres de la famille. Cela montre un fort degré d'engagement des utilisateurs dans les interactions avec l'AVI et leur capacité à apprendre de leurs expériences d'interaction, par tâtonnement et par le développement de stratégies d'interaction au gré des problèmes rencontrés.

3.3. Dynamiques interactionnelles observées dans FAM3

Dans FAM3, l'analyse du comportement interactionnel de l'AVI montre des réactions de trois types :

- la demande est comprise et génère directement l'énoncé attendu ou passe par la préface de l'action (indiquer la source de la réponse...);
- la demande est incomprise et génère l'énonciation de l'incompréhension (signal sonore d'arrêt de fonctionnement ; demande de répétition ; demande de confirmation...);
- la demande peut être comprise de façon erronée ou approximative par l'AVI générant une action inattendue, inadaptée (il y a alors désalignement entre l'utilisateur et l'AVI).

Les incompréhensions ou les échecs dans la réalisation de la commande ou de la requête génèrent la verbalisation d'un travail de face de la part de l'AVI avec la formulation d'actes menaçants pour la face de l'AVI (formulation d'excuses : « désolé, je ne sais pas ») ou des énoncés informationnels commerciaux invitant l'utilisateur à recourir d'autres fonctions

et services associés non sollicités (ajout à une liste de course ; renvoi à des services payants...) qui sont alors ressentis comme des actes menaçants pour l'interlocuteur (les enregistrements montrent les réactions d'énervement des utilisateurs).

Les problèmes récurrents de compréhension de l'AVI relevés dans le corpus ont différentes origines : un problème de discrimination des phonèmes quel que soit le locuteur (natif ou allophone), la présence de chevauchement entre les locuteurs lors du tour de parole d'un utilisateur en situation de polylogue, une disfluece lors de l'énonciation interprétée par l'AVI comme la fin du tour de parole et qui génère une interruption intempestive par l'AVI ou encore une confusion entre les langues (mauvaise traduction de mots transparents recouvrant des réalités différentes, qui renvoient à des expériences culturelles différentes entre l'anglais et les langues employées par les utilisateurs) générant des malentendus sémantiques.

On remarque que les échecs de communication avec l'AVI, bien que représentant une minorité dans FAM3, interrogent la capacité même des utilisateurs à interagir, nécessitant de leur part la mise en place de stratégies conversationnelles que les locuteurs ne peuvent pas transférer de leur pratique interactionnelle habituelle. Pour Nass et Moon (2000), cela révèle notre tendance à déployer de manière spontanée, non réfléchie, des pratiques interactionnelles ordinaires lorsqu'on interagit avec des machines ayant des traits humains comme l'usage de la parole. Or parler à un AVI de façon « naturelle » – en hésitant, en utilisant des indexicaux, etc. – conduit fréquemment à l'échec du dialogue. Pourtant, la raison de l'échec de communication n'est jamais explicitée aux utilisateurs et reste obscure pour eux. Les utilisateurs se retrouvent ainsi dans une situation paradoxale : plus le système fonctionne correctement et mobilise des pratiques conversationnelles ordinaires comme l'utilisation d'indexicaux, plus les utilisateurs ont tendance à parler naturellement et plus cela risque de conduire à l'échec du dialogue sans pour autant pouvoir déterminer la cause de cet échec. Ainsi, on repère chez les utilisateurs, la mise en place d'un véritable travail tactique qui vise l'efficacité des échanges et qui repose sur les interprétations des expériences interactionnelles plus ou moins réussies avec l'AVI. Les entretiens semi-dirigés montrent que ces routines conversationnelles sont conscientisées et ont fait l'objet d'un apprentissage par tâtonnement. L'enfant explique que suite aux feedbacks négatifs de l'AVI, il doit simplifier ses énoncés, augmenter sa fluidité et « faire des phrases comme à l'écrit ». Les autres utilisateurs évoquent privilégier deux modes opératoires qu'ils estiment pertinents :

- une stratégie de contextualisation en plusieurs paires adjacentes : il s'agit d'amener l'AVI à comprendre le contexte par séquences de façon à l'aider à identifier la bonne inférence conversationnelle ;
- une stratégie qui s'adapte au script de fonctionnement de l'AVI à partir des occurrences les plus mobilisées, les formulations qui génèrent par expérience aux réactions optimales de l'AVI.

Ces deux modes opératoires permettent aux utilisateurs de communiquer avec l'AVI sans contexte partagé. On remarque également que les utilisateurs adaptent leur communication à ce qu'ils pensent être le fonctionnement de l'AVI : ils développent ainsi de véritables stratégies inter-linguistiques. Le corpus montre également la façon dont

l'AVI « collabore » avec l'utilisateur pour combler les lacunes ou pallier les échecs. Ces observations témoignent d'un effort de construction de la compréhension mutuelle et d'une coopération pour la réussite de la conversation à partir de très nombreux procédés d'ajustement – sollicitation, réparation, négociation, clarification, explication, simplification, facilitation, reformulation, auto-hétéro-correction, décontextualisation-recontextualisation... Ces phénomènes sont typiques des situations exolingues caractérisées par les manifestations de divergences de savoirs linguistiques et de savoirs socioculturels, sources de malentendus comme nous l'avons vu. Ces situations concernent l'ensemble des utilisateurs de FAM3, et non seulement le locuteur allophone ou l'enfant. L'AVI est lui-même porteur de cette altérité socio-langagière.

4. Effets sur les apprentissages

4.1. Des apprentissages interactionnels informels collectifs

Le travail de l'utilisateur, indispensable au fonctionnement des AVI, repose sur des hybridations habiles des pratiques conversationnelles ordinaires et des savoir-faire relevant de *l'art de parler à une machine*. Nous avons vu que ce travail peut prendre différentes formes : réparation, reformulation, séquençement, résolution de problèmes. Qu'en est-il de l'apprentissage et les éventuels transferts de stratégies entre membres de la famille ? On relève des traces d'apprentissage liées à différentes dynamiques :

- les essais et erreurs amènent à une conscientisation des stratégies d'interprétation et de communication pour les plus âgés (3 membres sur 4), ceux qui disposent déjà d'une conscience langagière.
- l'étyage du collectif permet de modifier les stratégies de communication infructueuses (gommer les traces de l'oral au niveau de la syntaxe, effort de prononciation, contrôle du volume sonore de l'énoncé) et engager par exemple le locuteur allophone ou le plus jeune utilisateur à imiter ces nouvelles formes d'interaction.
- de nombreuses tâches langagières sont prises en charge collaborativement, lors des réparations notamment et l'interprétation des réactions de l'AVI par les individus (on observe la mobilisation d'un répertoire d'interprétation partagée selon les réactions les plus habituelles).

Toutefois les feedbacks négatifs de l'AVI ne déclenchent pas forcément des dynamiques d'apprentissage. Par exemple, les échecs dus à la prononciation du locuteur allophone sont vécus comme une stigmatisation par ce dernier et génèrent un faible sentiment d'auto-efficacité, voire un refus d'action (ex. « vas-y, dis le toi »). De même, l'enfant de FAM3 se retrouve souvent dans l'incapacité de déployer le savoir-faire interactionnel nécessaire pour faire face aux troubles qu'il rencontre, ce qui crée une situation d'exclusion et un sentiment de frustration. Ces situations démentent le caractère prétendument universel, accessible et intuitif de l'interaction vocale avec l'AVI.

4.2. Quels effets sur la compétence interactionnelle du locuteur allophone ?

Comme nous l'avons vu, l'utilisation de l'AVI offre une opportunité supplémentaire d'interaction en français pour le locuteur allophone de FAM3 qui apprécie la dimension

ludique des échanges malgré les difficultés récurrentes de communication. Ces difficultés sont aussi l'occasion de réfléchir au fonctionnement de la langue et d'ajuster son système linguistique en langue-cible, en particulier en termes de justesse lexicale. Le locuteur allophone apprécie le degré de tolérance linguistique de l'AVI qui lui permet de s'exprimer avec une certaine liberté et d'essayer la recevabilité de son système linguistique. Pour autant, les échanges avec l'AVI ne lui permettent pas de pallier ses difficultés, en particulier au niveau phonétique qui sont à l'origine de la plupart des problèmes de communication qu'il rencontre avec l'AVI. L'AVI est incapable de prendre en charge un étayage (Bruner, 1983) qui fonctionnerait comme un véritable outil de structuration des ressources langagières favorisant les apprentissages. Sans le cadre du « contrat didactique » (Py, 1992) entre les participants de l'interaction, le locuteur-apprenant ne peut réellement activer le processus d'acquisition sous-jacent à sa participation à l'interaction. Les signaux d'incompréhension que renvoient l'AVI ne constituent pas des feedbacks pertinents pour apprendre, dans la mesure où l'on ne comprend pas forcément l'origine du problème et que l'on n'a pas non plus forcément accès aux ressources pour y remédier.

Les données attestent surtout de la mise en place d'une compétence interactionnelle propre aux échanges avec l'AVI. Les entretiens montrent combien les participants de l'étude sont conscients, à des degrés divers, d'avoir appris à s'ajuster à l'AVI en termes de forme et de contenu et à gérer l'organisation des tours de parole (éviter tout chevauchement par exemple), à orienter l'attention de l'AVI, à contextualiser les topics, à privilégier certains lexiques mais aussi à ne pas recourir à l'ensemble des ressources multimodales de l'interaction (gestes, mimiques, postures corporelles...) qui permettent de se coordonner entre participants et de construire un discours partagé. La compétence interactionnelle étant située et contingente, il est difficile de conclure que s'entraîner à interagir avec un AVI peut avoir un effet sur le développement de la compétence interactionnelle dans d'autres situations d'échanges. D'ailleurs, le locuteur allophone déclare dans l'entretien qu'il ne perçoit pas ces stratégies interactionnelles comme pertinentes dans les autres situations d'interaction qu'il rencontre au quotidien. Il trouve au contraire qu'il est beaucoup plus facile pour lui d'interagir avec des humains qu'avec les interfaces vocales des objets connectés. Il apparaît que la difficulté à gérer les négociations conversationnelles avec l'AVI constitue l'un des freins principaux aux échanges avec un AVI. Or dans l'interaction, c'est bien la présence de l'autre et ses manifestations qui cadrent l'appropriation de la langue-cible. C'est dans le dialogue avec l'autre que s'élaborent les nouveaux moyens de dire, d'échanger et de construire un discours partagé.

Conclusion

Dans quelles conditions et pour qui les AVI offrent-ils une simulation de socialisation pertinente pour progresser en compétence interactionnelle en langue-cible ? Cette micro-étude rend compte du travail important des utilisateurs pour mener à bien leurs interactions avec un AVI lors de pratiques familiales et collectives face aux défaillances de

l'artefact. Si les problèmes de reconnaissance vocale constitue un grand volume de difficultés et occasionne chez les utilisateurs tout un répertoire de stratégies langagières (répétitions, reformulations, changements de langues, abandons...), le corpus montre que ces difficultés peuvent engendrer la mise en place de nouvelles formes d'interaction hybrides, alternant entre le modèle de la conversation naturelle et celui d'un certain art de parler à la machine (simplification des énoncés et leur structure, abandon des disfluences, mise en place de stratégie de contextualisation...). L'étude rappelle l'absence d'universalité de ces outils qui sont porteurs d'une culture, voire d'une certaine hygiène communicative (Cameron, 2012) générant une standardisation des échanges vers un prêt-à-parler, aux antipodes de la complexité et la créativité des échanges naturels. Notre étude montre également combien le feedback négatif (échec de communication) de l'AVI peut être incapacitant, et ce sont les dynamiques collectives familiales qui parviennent à pallier ces incompétences, en proposant des étayages (faire faire) ou des coopérations (faire avec) nécessaires à tout apprentissage de langue. FAM3 souligne qu'au-delà de l'agentivité suscitée par l'outil, c'est surtout l'environnement familial qui est capacitant, par l'entre-aide qui s'organise autour de l'AVI pour développer des formes d'interaction optimales et de conscientisation sur le fonctionnement de la langue. La place et le rôle des autres utilisateurs apparaissent ici comme déterminants.

Références bibliographiques

- André, V. (2021). Des corpus d'interactions dans la formation linguistique des migrants. *Savoirs*, 56, 77-96.
- Amossy, R. (1999). *Images de soi dans le discours. La construction de l'éthos*. Delauchaux & Niestlé.
- Brougère, G. (2009). Vie quotidienne et apprentissages. Dans G. Brougère (dir.), *Apprendre de la vie quotidienne* (pp.21-31). PUF.
- Bruner, J. K. (1983). *Le développement de l'enfant : savoir-faire, savoir-dire*. PUF.
- Cameron, D. (2012). *Verbal Hygiene*. Routledge.
- Collins, H. (2018). *Artificial Intelligence: Against Humanity's Surrender to Computers*. Cambridge UK, Polity Press.
- Conein, B. (1998). « La notion de routine : problème de définition ». *Sociologie du travail*, 4, 479-489.
- Dizon, G. (2020). Evaluating intelligent personal assistants for L2 listening and speaking development. *Language Learning & Technology*, 24(1), 16-26.
<https://doi.org/10125/44705>
- Druga, S., Williams, R., Breazeal, C., Resnick, M. (2017). Hey Google is it OK if I eat you? Initial Explorations Child-Agent Interaction. *Proceedings of the 2017 Conference on Interaction Design and Children*, 595-600.
- Filliettaz, L. (2022). Enquêter et produire une action de formation par l'analyse collective des interactions. Dans B. Albero & J. Thievenaz (dir.). *Enquêter dans les métiers de l'humain. Traité de méthodologie de la recherche en sciences de l'éducation et de la formation*, Tome III, (pp. 270-282). Editions Raison et Passions.
- Goffman, E. (1974). *Les cadres de l'expérience*. Editions de Minuit.
- Gauducheau, N., Marcoccia, M. (2022). Dialoguer avec Alexa : analyse de récits d'usage. Dans M-F. Agnoletti & E. Salès-Wuillemin. *Communication réelle et virtuelle : nouvelles*

- perspectives en psychologie sociale de la communication*. Presses Universitaires de Dijon.
- González-Lloret, M. (2021). L2 pragmatics and CALL. *Language Learning & Technology*, 25(3), 90-105. <http://hdl.handle.net/10125/73451>
- Gumperz, J. (1989). *Engager la conversation. Introduction à la sociolinguistique interactionnelle*. Minit.
- Kerbrat-Orecchioni, C. (2005). *Le discours en interaction*. Armand Colin.
- Larsonneur, C. (2021). Alexa, Siri : la diversité linguistique au prisme des agents conversationnels. Dans J-C. Beacco (dir.), *Traduction automatique et usages sociaux des langues. Quelles conséquences pour la diversité linguistique* (pp.179-197). Observatoire européen du plurilinguisme. <https://doi.org/10.3917/oep.beacc.2021.01.0179>
- Malinowski B. (1972). Phatic Communion (première édition 1923). Dans J. Laver & S. Hutcheson (dirs), *Communication in Face to Face Interaction* (pp.146-153). Penguin Books.
- Nass C., Moon Y. (2000). Machines and mindlessness: Social responses to computers, *The Journal of Social Issues*, 56(1), 81-103. <https://doi/10.1111/0022-4537.00153>
- Nass, C., & Brave, S. (2005). *Wired for speech: How voice activates and advances the human-computer relationship*. Boston Review.
- Pain, A. (1990). *Éducation informelle, les effets formateurs dans le quotidien*. L'Harmattan.
- Pekarek-Doehler, S. (2006). Compétence et langage en action. *Bulletin VALS-ASLA*, 84, 9-45.
- Porcheron, M., Fisher, J., Reeves, S., Sharples, S. (2018). Voice Interfaces in Everyday Life. *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1-12. <https://doi.org/10.1145/3173574.3174214>
- Porquier, R. (1984). Communication exolingue et apprentissage des langues. Dans B. Py (dir.), *Acquisition d'une langue étrangère* (pp. 12-48). Armand Colin.
- Py, B. (1992). Acquisition d'une langue et altérité. *Cahiers de l'Institut de linguistique et des sciences du langage*, 2,113-126.
- Rabardel, P. (1995). *Les hommes et les technologies. Approches cognitives des instruments contemporains*. Armand Colin.
- Robertson R. (1992). *Globalization: Social Theory and Global Culture*. Sage.
- Roudometof, V. (2021). Qu'est-ce que la glocalisation ?. *Réseaux. Communication-Technologie-Société*, 226-227, 45-70. <https://doi.org/10.3917/res.226.0045>
- Simonian, S. (2020). Approche écologique des environnements instrumentés : comprendre le phénomène d'affordance socioculturelle. *Savoirs*, 52, 93-108.
- Velkovska J., Zouinar M. (2018). The illusion of natural conversation: interacting with smart assistants in home settings. *CHI'18 Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, New York, ACM.
- Velkovska, J. Zouinar, M., Veyrier C-A. (2020). Les relations aux machines « conversationnelles ». Vivre avec les assistants vocaux à la maison. *Réseaux. Communication-Technologie-Société*, 2(3), 47-79.
- Xiao, F., Zhao, P., Sha, H., Yang, D., Warschauer, M. (2023). Conversational agents in language learning. *Journal of China Computer-Assisted Language Learning*. <https://10.1515/jccall-2022-0032>.
- Zeiter, A-C. (2018). *Dans la langue de l'autre. Se construire en couple mixte plurilingue*. ENS Editions.