



HAL
open science

Machines parlantes et AVI: vers une extension du domaine de la linguistique ?

Charlotte Danino

► **To cite this version:**

Charlotte Danino. Machines parlantes et AVI: vers une extension du domaine de la linguistique ?. Drôles d'objets, May 2023, Nancy, France. 10.5281/zenodo.12784018 . hal-04655361

HAL Id: hal-04655361

<https://hal.science/hal-04655361v1>

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.

Machines parlantes et AVI : vers une extension du domaine de la linguistique ?

Charlotte Danino

Université Sorbonne Nouvelle

PRISMES EA 4398 / SeSyLIA

charlotte.danino@sorbonne-nouvelle.fr

Résumé : Cette étude examine l'impact des assistants vocaux interactifs (AVI) pour les théories linguistiques qui reposent presque toutes sur le postulat que le langage est produit entre deux humains dans la communauté sociale. Or les AVI n'ont pas de subjectivité ni de cognition comparable à celle de l'homme. En confrontant un examen épistémologique des sciences du langage, qui ne peut pas être exhaustif, à la conduite de quatre entretiens de recherche, ce travail pose la question du dédoublement possible des objets de la discipline : faut-il une linguistique interhumaine et/ou une linguistique humain-machine ? La dimension écologique (contexte temporel, environnement matériel, ressources sémiotiques) s'avère cruciale dans la prise en compte théorique et analytique de ces nouvelles interactions qui nécessitent une extension du domaine de la linguistique.

Mots-clés : sciences du langage, subjectivité, interaction multimodale, théorie linguistique, incarnation

Introduction

Les assistants vocaux interactifs (AVI) posent la question de l'intersubjectivité : suffit-il qu'il y ait un seul humain pour que l'interaction soit intersubjective, ou que l'humain fasse « comme si » pour établir une intersubjectivité même partiellement jouée ? Le fait que les AVI soient programmés pour utiliser des verbes de perception, des pronoms personnels, des marqueurs de politesse et un lexique émotionnel suffit-il à construire l'intersubjectivité ?

Toutes ces questions sont en fait doubles car elles se posent pour les locuteurs en engageant ce qui se joue réellement dans l'interaction, et pour les linguistes qui doivent ensuite conceptualiser et analyser ces interactions. Si « faire comme si » peut être une modalité acceptable des pratiques langagières, est-ce épistémologiquement acceptable ? Il s'agit de prendre acte d'un décalage entre une machine qui produit un « output » à partir d'un calcul numérique statistique complexe et opaque, et l'interprétation sémantique qu'en fait l'utilisateur humain. Si la machine traite de l'information, elle ne traite pas de stimulus ou d'expérience contrairement à l'humain, pour qui l'interaction elle-même est expérience. De même, si une sorte de mémoire, au sens d'enregistrement de données précédentes existe, il ne s'agit pas de la mémoire humaine largement biographique participant de la construction du sujet.

En d'autres termes, certaines des notions fondamentales de la linguistique et plus largement des sciences du langage (incluant l'analyse de discours et l'analyse conversationnelle) sont remises en question : le sens, l'information, le sujet (parlant), l'intersubjectivité pour ne citer qu'elles. Faut-il alors revoir entièrement les théories et méthodologies d'analyses linguistiques ?

Ce travail propose une réflexion d'ordre épistémologique articulée autour de la notion de sujet. On cherche à poser les bases d'un travail théorique de modélisation des interactions humain-AVI à l'aune de cette notion. Pour ce faire, l'examen théorique et conceptuel de ce travail s'est appuyé sur quatre entretiens de recherche exploratoires dont le but était d'interroger les participant.e.s sur leurs usages effectifs d'un AVI domestique (Alexa ou Google Home) et d'observer la conscientisation de certains aspects de ces interactions durant l'entretien. En effet, il s'agissait d'interroger des locuteurs et locutrices sur leur perception et leur ressenti d'une intersubjectivité possible entre eux et les AVI.

Une première partie propose un rapide parcours notionnel pour situer le problème vis-à-vis de différents courants linguistiques. Une seconde partie confronte ces notions en regard des entretiens. La troisième partie offre une discussion synthétique et une conclusion propose une synthèse et suggère quelques pistes de réflexion.

1. Cadrage conceptuel : vers une extension du domaine de la linguistique ?

Depuis Saussure (1916), la linguistique distingue la langue (système de signes cohérents vu comme un code) de la parole (exercice de la faculté de langage ayant lieu nécessairement dans au moins une langue). Le langage quant à lui désigne une faculté cognitive et physiologique qui serait le propre de l'homme. A partir de ce fondement saussurien, Benveniste (1966) notamment développe la notion de discours qui considère les productions linguistiques comme toujours situées dans l'espace et le temps en lien avec un sujet parlant, qui peut être qualifié par ses identités sociales ou éthiques. Charaudeau (2023) revient ainsi sur cette notion en incluant au-delà de la stricte linguistique les sciences du langage telles que l'analyse des interactions et les différentes approches d'analyse de discours.

Ces dernières partagent d'envisager la circulation des discours de manière critique (en lien avec le fait socioculturel) ou sémantique (en lien avec la constitution et l'évolution de la langue) en tant qu'ils permettent et obligent les sujets parlants (individuels et collectifs) à se positionner les uns par rapport aux autres au sein d'une communauté spécifique. La sociologie a de son côté développé, en même temps que l'avènement de l'étude de l'oral, l'Analyse Conversationnelle qui s'intéresse aux interactions interhumaines ordinaires ou institutionnelles (Sacks et al., 1974 comme point de départ). Le langage y est alors vu comme une action et un comportement socialement régulé.

A l'inverse de cette épistémologie, la linguistique cognitive (Langacker, 1987 ; Lakoff, 1987) considère le langage comme une faculté cognitive qui n'est pas indépendante des autres. Cette tradition théorique subjectiviste et mentaliste partage certaines caractéristiques avec les traditions énonciatives du continent européen (Culioli, 1990). Les productions linguistiques sont vues comme les produits de l'activité mentale des locuteurs et l'énoncé en porte les traces. Les structures linguistiques reflètent ainsi les structures conceptuelles émergeant de l'activité cognitive au-delà de la seule activité de langage.

Ce que ces approches théoriques partagent, au-delà de leurs différences réelles, c'est le postulat d'un sujet parlant dont l'activité sociocognitive construit, régule et marque les

productions linguistiques, quelle que soit la place laissée au sujet en tant que concept. En effet, la linguistique cognitive prévoit en théorie une conception très aboutie du sujet qui inclut ses caractéristiques cognitives, ses déterminations sociales, son incarnation et son interaction avec l'environnement (Popova, 2015 par exemple), mais peine à en tirer tout le bénéfice méthodologique. A l'inverse, l'analyse conversationnelle préfère une conception émergente du sens et de l'interaction en évitant de substantialiser les participant.es à l'interaction dont les identités sont négociées durant la conversation. Si le mentalisme est exclu de cette approche, l'intersubjectivité y reste centrale (p.ex. Heritage, 2009). Toutes ces approches partagent également désormais de devoir accorder un statut aux machines parlantes qui ne sont pas des sujets et n'ont pas de comportements à proprement parler. En d'autres termes, leurs productions linguistiques ne procèdent ni des mêmes contraintes, ni des mêmes mécanismes que celles des locuteurs.trices humain.es.

Avant de clore ce parcours épistémologique, il reste à mentionner le cas de (certaines) linguistiques formelles qui se concentrent sur la forme logique des propositions ou des énoncés et leur modélisation. Elles sont ainsi à l'interface de l'informatique et des technologies de l'information (linguistique computationnelle) en étant précisément détachées de l'expérience sociopsychologique humaine. Elles permettent d'articuler, non plus expérience et langage, mais langage naturel et langages artificiels, ceux des machines précisément.

Ce parcours ne prétend en aucun cas à l'exhaustivité et ne peut rendre justice aux théories et méthodologies citées. Mais cette rapide cartographie théorique permet de proposer un consensus, parfois débattu, vis-à-vis duquel chaque courant peut se positionner. En effet, au cœur de cette géographie notionnelle, se trouve le postulat suivant : l'activité de langage est actualisée dans une langue naturelle et présuppose la présence d'au moins deux êtres humains engagés dans une interaction intersubjective et socioculturellement déterminée. Leur relation interlocutive se fonde autant sur le code de la langue que sur l'incarnation des sujets parlants, ce qui présuppose aussi la prise en compte d'une cognition (spécifiquement) humaine. L'écrit n'est bien entendu pas en reste de cette approche – il a même longtemps constitué un biais analytique (Linell, 2005). La force de ce postulat repose sur deux piliers dont l'un commence à faiblir. Le premier toujours aussi solide est que tout être humain apprend à parler dans la conversation orale ordinaire (vocale ou visuogestuelle pour les langues des signes). La primauté de l'oral et du conversationnel n'est pas remise en question dans sa centralité théorique et analytique. Le second pilier était que seuls des humains répondaient aux humains. Or le champ d'inclusion des sciences du langage (SDL) est en train de s'élargir.

Les SDL prennent progressivement en compte de nouveaux objets et/ou de nouvelles interactions dans leur champ d'étude. Ce mouvement est à la fois externe et interne à ce champ disciplinaire. Il est externe en ce qu'il vient de disciplines connexes. Ainsi, les sciences cognitives ont revu la notion d'intelligence en étudiant notamment la cognition animale et le développement de l'intelligence artificielle (IA) auxquelles elles participent. Les nouvelles technologies de communication ont forcé sinon permis la prise en compte de nouvelles interactions (homme-machine) mais elles ont créé de nouvelles formes

interactionnelles (Delvotte, Paveau, 2017, notamment sur les distinctions mises à mal entre oral et écrit, entre synchrone et asynchrone). Paveau (2009) propose de résumer cette tendance en parlant de « linguistique symétrique » où la symétrisation décrit la fin du privilège de l'humain et du seul verbal dans l'étude de la communication et de l'interaction. Le langage en fait partie mais n'en est plus la seule modalité. L'analyse interactionnelle pour qui le sens est émergent, c'est-à-dire dans l'interaction, a la première franchi ce pas. La multimodalité dans l'étude du langage est le second mouvement, interne cette fois aux SDL.

En parallèle de l'écologisation de leurs objets d'étude, les SDL, dans un mouvement concomitant, cherchent à intégrer la multimodalité et l'incarnation de nos activités de langage ainsi que la diversité des ressources sémiotiques à disposition des locuteurs pour construire du sens (p.ex. Tomasello, 2003 ; Kendon, 2004 ; Linell, 2009 ; Haddington et al., 2014 ; Goodwin, 2018 ; Mondada, 2019). Ces évolutions sont fondamentales pour analyser les interactions hommes-machines et *a fortiori* hommes-machines *parlantes* car elles autorisent la prise en compte d'agents conversationnels non humains d'une part et permettent l'inclusion de ressources sémiotiques variées dans l'interaction linguistique d'autre part. Plusieurs notions doivent alors être revues :

- le sujet : que reste-t-il de la notion de sujet (parlant) ? et donc d'intersubjectivité ?
- la communication sociale : si l'un des partenaires interactionnels n'est pas membre d'une société, peut-on encore parler d'un principe de régulation sociale, de comportement, de relation ou de rapport ? Programmation vaut-elle régulation ?
- la perception et le corps : Alexa « perçoit »-elle une situation de communication ? La matérialité des machines permet-elle de parler d'incarnation ? Peut-il y avoir un énonciateur sans corps ou sans identité sociale ?

En somme, il s'agit de conceptualiser le type d'énonciateur que sont les AVI par rapport à l'être humain et par rapport à d'autres dispositifs de parole qui s'écartent de la configuration défaut d'une conversation orale entre au moins deux sujets humains. Maingueneau envisage ainsi des « locuteurs angéliques »¹ qui seraient différents des annonces en gare. Celles-ci sont des humain.e.s enregistré.e.s. On reconnaît ainsi « la voix de la SNCF » qui *incarne* bien l'entreprise entière pour ses passagers – énonciateur d'un sujet collectif abstrait, comme la personne morale en droit. Dans les trains, l'accueil ou les imprévus sont annoncés par des humain.e.s – les chef.fe.s de bord. Autre dispositif encore que celui de l'émission de Canal+ *En aparté*, où la journaliste n'est pas présente dans le lieu de l'entretien qui ressemble à un appartement personnalisé pour l'invité.e. Sans face-à-face, l'invité.e est libre de ses mouvements oculaires et physiques, ce qui modifie l'interaction avec la journaliste et l'interaction parasociale (Horton, Wohl, 1956) pour lesquelles l'invité.e n'a aucun retour. Outre la réduction d'un interlocuteur à sa voix et l'appauvrissement des ressources sémiotiques, ces deux exemples partagent avec les AVI d'être consubstantiels à un lieu, comme peut l'être la prière qui se donne entièrement dans

¹ « Les innovations technologiques permettent également d'attribuer des énoncés, écrits ou oraux, à des êtres qui ne sont pas à proprement parler des sujets parlants, mais ce qu'on pourrait appeler des *locuteurs angéliques*, de plus en plus présents dans notre vie quotidienne. » (Maingueneau, 2021, p.179)

un lieu vers un interlocuteur symbolique désincarné et/ou réifié (dans une statue par exemple)².

2. « Comme une personne que l'on vient de rencontrer » : la subjectivité (ressentie) des AVI

Pour confronter (et différencier) le problème que les AVI posent aux conceptualisations fondamentales des SDL (c'est-à-dire aux linguistes) et aux locuteurs et locutrices, quatre entretiens ont été menés dans le but de faire émerger la conscientisation de certaines pratiques langagières et conversationnelles par les utilisateurs.trices des AVI.

Les entretiens ne sont pas un corpus et ne prétendent à aucune représentativité. Ils ont été analysés en termes de thématique (Paillé, Mucchielli, 2016) à partir d'indications sociolinguistiques (âge, genre, multilinguisme) et linguistiques (expression de la subjectivité, de la politesse par exemple). Ils ont permis d'affiner le travail de conceptualisation théorique en faisant émerger la perception qu'ont les locuteurs de leurs interactions avec les AVI et la manière dont cela engage leur réflexivité sur leurs propres comportements langagiers. C'est pour cette raison que l'âge et le genre par exemple ont fait l'objet d'une attention particulière car on sait que ces deux paramètres influent sur le comportement conversationnel des locuteurs et locutrices (Tannen, 1993 sur le genre pour une étude de référence). Le guide d'entretien s'organise autour de trois thèmes (les usages de l'AVI au quotidien, le rapport au langage et à la conversation des membres du foyer et la représentation des technologies connectées). Il visait des enjeux linguistiques identifiés au préalable : subjectivité, politesse multilinguisme, socialisation langagière, imaginaire linguistique. Ces enjeux proviennent tous d'une abondante recherche en sciences du langage sur la constitution et les manifestations de la subjectivité dans la langue et le discours.

Les quatre participant.es ont entre 20 et 50 ans. Il s'agit de trois femmes (P1, P2, P3) toutes mères d'enfants scolarisés en primaire et/ou très jeunes adultes, et un homme (P4) sans enfant. Les foyers dont ils font partie comptent tous plusieurs membres. Les six points suivants tâchent de rassembler les éléments centraux des entretiens en regard du problème de la subjectivité dans les pratiques langagières quotidiennes.

2.1 Subjectivité représentée et représentation sociale

Un élément clé de la subjectivité des AVI tient à leur programmation qui inclut des marqueurs d'(inter)subjectivité tels que les marqueurs de politesse, les pronoms personnels (la machine dit *je*, vouvoie l'utilisateur) et les verbes de perception ou de cognition. Il s'agit donc de faire croire à une intériorité en mimant ce qui pour l'humain n'est pas qu'une posture interlocutive mais bien un compte-rendu de son expérience. L'illusion est telle que l'AVI reprend les enfants sur la politesse. Pour autant, si Alexa ou Google sont tutoyés (y compris les enfants), la machine vouvoie et indexe l'asymétrie sociale. Plusieurs participant.e.s rapportent des échanges avec leurs conjoint.e.s ou

² Prak-Derrington (2019) sur la litanie par exemple.

enfants sur le degré de politesse employé avec l'AVI : certain.es estiment que tout langage doit être poli (« dans cette maison, on parle correctement ») et d'autres que ce point n'est valable qu'entre humains. Cette divergence explique que certains enfants sont repris par les parents dans les interactions enfants-machines, quand d'autres bénéficient d'un espace de liberté interactionnelle. Certains enfants seraient « tyranniques » avec l'AVI mais tous gardent un rapport ludique et affectif plus long et plus affirmé que les adultes. Pour les jeunes enfants, l'illusion de personne fonctionne bien et explique leur langage affectif (« je t'aime ») avec l'AVI qui y répond.

2.2 Objet/sujet : matérialité et incarnation

Les adultes eux sont moins dupes car ils sont chargés de la configuration et interprètent les signaux lumineux de la machine comme une activation – parfois vécue comme surveillance. Ils repèrent aussi mieux les incompétences du chatbot. Ainsi, quand Alexa cadre sa réponse par « Selon Wikipedia », l'illusion se brise pour P4 : est-ce parce qu'un humain ne l'aurait pas dit, ou pas comme ça ? Est-ce parce que l'AVI redevient alors un simple moteur de recherche ? La mémoire est une autre source d'illusion mise en danger : P2 s'amuse de voir que les publicités ciblées sont encore en espagnol de l'époque où elle vivait en Espagne. Ce défaut de mise à jour rend visible l'absence de mémoire discursive de l'AVI (Benveniste, 1966 sur l'historicisation du discours).

Notons aussi une certaine hybridité : le boîtier est une matérialisation de l'AVI mais pas un corps. Il « s'allume » mais ne se réveille pas. P2 s'amuse à dire que son boîtier abîmé « s'est suicidé dans un seau d'eau » mais elle se décrit comme utilisatrice informée travaillant dans la technologie, montrant en même temps qu'elle l'utilise les limites de sa métaphore. Les AVI parlent comme nous mais « écoutent » tantôt comme des machines de surveillance, tantôt comme des interlocuteurs. Cependant, la tonalité constamment explicative ou l'impossibilité de multitâche pour la machine, relevées par les participant.es trahissent une non-personne, signalant en retour ces deux éléments comme cruciaux dans nos représentations de l'interaction interhumaine.

P4 conclut finalement que l'AVI est « comme une personne qu'on vient de rencontrer », qui n'acquiert jamais ni intériorité ni profondeur avec le temps, et reste ainsi dans l'explicite relationnel. A titre anecdotique, P2 travaillait avec une stagiaire qui s'appelait Alexa et n'avait pas fait le lien jusqu'au moment de l'entretien. L'AVI *n'est pas* une personne mais *est comme* une personne. La voix et la parole créent largement l'illusion mais sont pourtant systématiquement décrites du point de vue fonctionnel, comme un moyen simple et rapide d'obtenir une information ou de faire exécuter une commande. La voix est ainsi qualifiée de manière pratique plus qu'éthique dans les quatre comptes-rendus d'expérience recueillis.

2.3 Coopération et intersubjectivité

Le principe de coopération qui suppose une forme de subjectivité ne serait-ce que relationnelle fonctionne seulement si l'humain s'adapte. Le travail d'articulation, s'il y a des accents ou un langage enfantin, l'alternance codique (le fait de changer de langue ou de les « mélanger »), la reformulation sont le fait du partenaire humain qui a seul la

charge d'une interaction réussie, comme le notent spontanément deux des participant.e.s. Et P3 de conclure qu'après la phase de découverte, « nous les humains, c'est nous qui nous sommes adaptés, on l'aidait ». Chacun.e a signalé avoir conscience que l'AVI n'a pas de représentation de l'interlocuteur humain, au point d'en faire parfois un atout interactionnel. Ainsi, les femmes interrogées ont toutes conscience de pouvoir de ce fait s'affranchir de toute injonction sociale et trouvent quelque chose de reposant dans l'interaction avec l'AVI – qu'il ait eu une voix féminine ou masculine. Il s'agit « d'une aide gratuite », d'un interlocuteur sans jugement car a-social – l'anthropomorphisme s'arrête donc à la réciprocité. Le genre semble d'autant plus pertinent que P2 et P3 ont constaté la différence de leur comportement conversationnel avec celui de leur conjoint face à l'AVI.

L'âge est une autre caractéristique pertinente car il existe un enjeu d'apprentissage. Par exemple, les enfants oublient que l'AVI doit être activé par un terme d'adresse et/ou qu'il n'est que dans le salon, donc inaccessible dans la chambre ou la voiture. Ce seul moyen de sélectionner l'AVI comme locuteur suivant sa disponibilité circonscrite mettent en évidence que les enfants ont déjà une représentation d'une interaction normale dont l'AVI s'écarte, limitant l'illusion d'humanité. Dans la même ligne d'idées, les parents prennent en charge cette « socialisation » en ajustant le comportement conversationnel de leurs enfants.

Les parents rapportent ainsi tous, en lien avec l'engagement de leurs enfants dans ces interactions et les enjeux éducatifs perçus, le manque d'*accountability*, de responsabilité de la machine : Alexa ne refuse pas de donner la réponse au devoir de maths, Google n'empêche pas forcément la diffusion d'une musique que les parents auraient interdite. L'éthique de ces interactions reste à inventer, parfois au sein des couples parentaux en désaccord sur ce point. Ainsi, de ce père de famille qui demande une information à l'AVI pour gagner du temps durant les devoirs de son enfant qui jusque-là n'y avait pas pensé.

2.4 Cadres participatifs, mise en espace et séquentialité

Cette anecdote pose plus concrètement la question de l'interlocution et de ce que Goffman aurait appelé les cadres participatifs. Tous les AVI des participant.e.s sont dans le salon et si l'achat a pu être le fait d'une personne dans le foyer, l'AVI est décrit comme familial et les usages rapportés confirment ce point. Il s'agit souvent de donner une information lors d'un dîner de famille, de régler ensemble un rappel important dans la vie de famille, de permettre des usages routiniers. L'exception est le fait des enfants qui engagent plus facilement l'AVI mais souvent dans les temps de jeux, d'ennui ou en cas d'indisponibilité de l'adulte sollicité au préalable (tel que les adultes l'ont rapporté).

L'AVI devient assistant familial et non pas personnel. Si cela n'engage pas réellement la persona interlocutive de l'AVI, cela constitue en retour la famille comme sujet collectif, et principal partenaire interactionnel de l'AVI. Chacun.e a décrit une phase d'adaptation suivie d'un rodage de l'AVI qui est calé dans les usages familiaux. Ainsi, P2 rapporte avoir personnalisé les commandes du foyer bilingue. P4 prend conscience que l'AVI est souvent sollicité durant les désaccords ou malentendus des dîners de famille. Il est possible que des études en situation naturelle fassent apparaître des séquences interactionnelles

typiques où l'AVI intervient. La question reste entière de savoir si dans ces contextes les AVI sont des ressources interactionnelles ou des partenaires interactionnels (comme Tannen, 2007 posait la question pour les animaux domestiques).

La question des interlocuteurs apparaît aussi dans une certaine fluidité des rôles : les participant.e.s rapportent une fluctuation dans leur représentation de l'AVI. Il serait intéressant de voir comment certaines interventions « autonomes » de la machine seraient perçues dans la mesure où pour le moment elle répond plus qu'elle ne parle selon un modèle interactionnel fondé sur une asymétrie de pouvoir. De fait, l'AVI reste un assistant. P3 s'amuse ainsi du style très direct des messages (écrits) de son compagnon : « Il écrit *pain merci*. On dirait Alexa qui m'écrit ». Les interactions avec l'AVI deviennent une heuristique d'une communication utilitaire et efficace. En effet, les usages rapportés sont homogènes : les AVI sont utilisés dans le cadre de nos activités routinières et programmables (réveil, écoute de musique, météo, information). La routine est un paramètre clé car les usages rapportés sont presque tous liés à des moments de la journée : lancer la musique quand on prépare le dîner et que les mains sont prises; programmer des rappels (courses, prises de médicaments, tâches), ou encore demande d'une information que la conversation interhumaine a fait émerger comme manquante (quelle est la date de naissance de tel acteur dont on discute à table ?). En ajoutant l'importance du lieu (foyer, pièce de vie familiale) et les moments d'un quotidien, l'AVI se retrouve lié à des situations types de la famille en question dont elle technicise l'environnement. En cela, les AVI construisent quelque chose comme un dispositif de parole ambiante.

2.5 Vers des technointeractions ?

Dans leur important travail sur l'Analyse de Discours Numérique, Delvotte et Paveau (2017) concluent à la nécessité pour les SDL d'envisager de nouveaux objets : les technodiscours, aux affordances particulières. Il convient d'envisager que les interactions avec les AVI, et plus largement avec les machines, favorisent ou permettent l'émergence de nouveaux objets discursifs et interactionnels. Les AVI incitent à se poser la question de techno-interactions où non seulement le comportement conversationnel des humaines s'adapte à la machine, mais où les propriétés techniques et interactives de la machine permettent de nouvelles séquences interactionnelles.

Le fait de devoir systématiquement activer l'AVI par son « nom » permet d'envisager que ces interactions donnent lieu à une nouvelle ergonomie de l'oral. Le renouvellement interactionnel se trouve dans certains usages ludiques et créatifs qui « créent » des séquences³ types que les familles peuvent ritualiser si ces séquences sont perçues positivement. Plus avancé encore, on peut constater, avec nos enquêtés que la technologie permet de créer une nouvelle multimodalité : l'AVI lit les messages en verbalisant les émojis (« je rentre bientôt cœur cœur ») ou répond aux sollicitations affectives d'un enfant (« est-ce que tu m'aimes ? », « comment tu vas ? ») en chanson. Ce

³ On peut repenser ici au débat initié par la Chambre chinoise de Searle. L'une des objections, dite linguistique, s'appuie justement sur la créativité linguistique et le nombre infini de dialogues possibles dont on ne peut dire avec certitude que l'ensemble de règles explicites à un moment donné suffit à les produire.

qui est au départ une non distinction entre différentes ressources sémiotiques et une homogénéisation des systèmes de signes devient une sorte de langage propre à l'AVI et au rituel interactionnel familial.

2.6 Une écologie des objets et des pratiques

Enfin, les usages décrits par les participant.es ne peuvent pas se comprendre en dehors de l'explicitation qu'ils fournissent de toute une écologie de pratiques qui force la linguistique à intégrer, sinon de nouveaux objets, au moins de nouveaux paramètres d'analyse. Le contrôle vocal et le style conversationnel sont reliés aux écrans : les parents apprécient que les enfants interagissent sans écran, comme avec certaines boîtes à histoires. P1 et P3 apprécient l'absence de stimulus visuel à la fin d'une journée passée à lire, entre les occupations professionnelles et les communications personnelles en écrit synchrone. C'est parce que P2 ne porte jamais de montre *et* qu'elle sait rarement où est son téléphone lorsqu'elle est chez elle que l'AVI donne l'heure. De même, P3 et P4 expliquent que certaines requêtes sont intéressantes à demander à voix haute car cela prendrait plus de temps de les écrire – mais le résultat est décevant si l'AVI répond en lisant le premier résultat Google ou les premières lignes Wikipédia, rendant alors les deux interactions médiées strictement équivalentes. De même que le GPS et les panneaux proposent deux rapports au réel différents (la comparaison est faite par P1 et P3), l'AVI propose une nouvelle expérience du quotidien. C'est ainsi que chacun.e a inséré l'AVI dans une typologie d'objets analogiques ou numériques, connectés ou non qui reflètent l'hybridation des usages et la complexification de nos représentations de notre environnement matériel et interactionnel. De la même manière que les objets se juxtaposent, les pratiques se composent : on utilise le numérique pour les infos et se faire livrer mais on lit sur des livres en papier ; on initie les enfants aux CD mais l'AVI lance la musique dans le salon. La hiérarchie existe parmi les AVI : Siri est personnel, Alexa/Google Home, collectifs. Tous les enfants entre 18 et 20 ans ont tous été décrits comme utilisant peu l'AVI et préférant l'usage de leur téléphone portable dans leur chambre. C'est donc à la fois une écologie linguistique qui implique des styles conversationnels et les différentes modalités de la langue (orale, écrite, visuelle) et une écologie matérielle qui peut avoir un impact direct sur des comportements langagiers. Si l'importance de ces configurations matérielles n'est pas nouvelle en sociologie des techniques⁴, elle est relativement nouvelle en sciences du langage dont l'écologisation est très récente (cf. section 1). La linguistique doit faire face à ces données importantes qui font d'une montre ou de notre exposition à des informations visuelles dans une journée des données importantes des productions linguistiques humaines.

⁴ Je remercie le relecteur ou la relectrice qui m'a signalé ce champ disciplinaire pertinent dont l'examen systématique excède cependant le cadre de ce travail.

3. Discussion

Faire « comme si » on s'adressait à une personne oblige les locuteurs humains, comme dans toute interaction à ajuster et adapter leur comportement conversationnel en mobilisant différentes représentations des interactions possibles, ce qui semble plus facile pour les adultes que pour les enfants⁵. Il semble cependant que les usager.es s'adressent en fait à un assistant et non à une personne. Ceci est moins vrai pour les enfants mais ils sont aussi ceux et celles pour qui le rôle social de l'assistant.e n'est pas encore établi. L'AVI est sollicité pour récupérer une information pérenne ou exécuter une commande dont l'effectuation semble n'être pas contrainte par la situation (qu'en est-il des termes déictiques avec les AVI ?). Dans cette logique qui limite l'ancrage de l'interaction dans la situation, nous ne semblons pas avoir de longues interactions avec l'AVI et si les enfants leur demandent une histoire, peu d'usager.es rapportent raconter leur journée à leur AVI. Les utilisateurs semblent rapidement cantonner l'usage de l'AVI au routinier : est-ce par épuisement de la capacité d'être étonné ou par contrôle ? Toute sensation de perte de contrôle évoque le danger ou suscite la méfiance. La question de l'éthique recouvre celle de la surprise. Même là où la créativité malgré elle de la machine amuse, la routinisation du ludique s'installe.

La routinisation a probablement lieu aussi car l'interlocution avec l'AVI familial engage un locuteur collectif, comme s'il émergeait une norme linguistique familiale. Cette idée de norme, opposée aux variations, est importante en linguistique où elle décrit assez bien la distinction entre une langue standard (le français) et des variétés par exemple phonétiques (accents) et/ou régionales avec variation du lexique. L'adaptation et la coopération interlocutive participent de l'alignement intersubjectif. Or seuls les humains en semblent capables. La voix familiale ou dominante devient une norme pour l'AVI qui gomme les variations interindividuelles.

Le fonctionnement même de l'apprentissage massif non supervisé de l'AVI (re)produit une langue idéale et reconstruite à partir d'un immense échantillon d'énoncés situés et contextualisés. D'une certaine manière elle décontextualise ces énoncés pour les re-situer sans avoir aucune conscience de l'environnement matériel et plurisémiotique qu'elle investit – ce qui a deux implications. D'une part, la machine propose une sorte de miroir linguistique en ce qu'elle produit des énoncés typiques et attendus. D'autre part, tout le travail d'interprétation, car les productions des AVI peuvent être globalement considérées comme asémantiques, repose sur l'humain. En forçant le trait on pourrait dire que la machine traite de l'information, manipule des signes et produit des *output*, là où l'humain traite des stimuli, manipule des significations (pas seulement en génération)

⁵ En fait, l'être humain engagé dans une interaction avec un AVI peut à tout moment, et même au milieu d'un énoncé, décider de changer d'attitude, sans conséquence sur le positionnement « intersubjectif » de la machine, voire sur l'efficacité ou le déroulement de l'interaction. Cependant, les AVI sont de plus en plus capables d'émuler ce travail d'ajustement comme rappeler aux marques de politesse ou exprimer une réaction affective.

et produit du sens⁶. Cet écart est parfois ressenti par les locuteurs qui adaptent leurs comportements conversationnels et interactionnels : en clair, ce sont bien les êtres humains qui produisent du sens en dotant les productions des AVI d'une intentionnalité. Cet écart, central dans les défis de l'intelligence artificielle, trouve un écho en linguistique. En effet, la linguistique cognitive, qui tâche d'intégrer les avancées de la psychologie et des neurosciences, se trouve face au délicat problème de l'interaction : concilier approches mentaliste et interactionnelle pose en fait le problème du rapport entre le cerveau et l'esprit – et ce qui fait esprit. Dingemanse et collègues (2023) en envisageant ce problème en viennent à conclure que c'est « l'interaction [qui] transforme un cerveau en esprit ». En transposant cette proposition aux interactions homme-machine, cela implique que seule l'interaction vaut comme objet d'étude mais que les sciences du langage doivent décider de l'étudier pour l'humain ou pour la machine.

Faut-il une ou deux linguistiques ? Faut-il une linguistique (humaine) de l'interface expérience-langage fondée sur l'action et le sens et une linguistique de l'interface entre code informatique et langage naturel fondée sur le nombre et la transformation (ou transmodalité) ? Si deux modèles deviennent nécessaires, comment les intégrer dans les cas d'analyses d'interaction avec les AVI ?

Conclusion

Ce travail exploratoire est né d'une interrogation sur la notion de sujet de situation en sciences du langage. Dans ce contexte, les interactions avec les AVI offrent un cas aux limites des conceptualisations traditionnelles de la linguistique. Ce travail visait à poser les enjeux épistémologiques que posent ces drôles d'objet que sont les AVI pour des sciences du langage où le sujet humain est (encore) largement central, et l'intersubjectivité essentielle. Si le problème est de taille pour les linguistes, les usager.es des AVI semblent avoir trouvé les compromis et les ajustements nécessaires pour une pratique quotidienne satisfaisante, même si les entretiens ont permis d'identifier ce qui dans leurs pratiques leur semble moins satisfaisant ou pas encore stabilisé.

La perception de ces participant.es a permis d'éclairer en retour ce qu'ils considèrent comme clé dans l'interaction humaine : la capacité à être multitâche, c'est-à-dire le fait que le langage est généralement produit en co-activité ; et le manque de registre en termes de modes de discours (expliquer n'est pas raconter, qui n'est pas décrire...). Ces éléments incitent à reprendre la distinction de Tannen (2007) entre ressource et partenaire interactionnel. Pour le moment, et dans les usages des participant.es à nos entretiens, les AVI semblent vite cantonnés au rôle de ressources interactionnelles, plus que de réel partenaire. Cela peut s'expliquer en partie par le caractère collectif de l'AVI systématiquement placé dans une pièce commune et qui constitue la famille entière comme sujet parlant collectif avec ses habitudes langagières et son historique de dialogue pour ainsi dire. Malgré les limites perçues, l'illusion est puissante, notamment chez les

⁶ Je remercie le relecteur qui a suggéré ici la pertinence de l'expérience de pensée de la « chambre chinoise » de Searle et suggère d'introduire dans la réflexion la distinction entre différents types d'intentionnalité primitive ou dérivée.

enfants précisément en cours de socialisation langagière, entre autres aspects de leur développement.

En définitive, on suggère que la notion de simulation peut tout à la fois synthétiser les enjeux de cette étude et en proposer un approfondissement. Le terme est partagé par l'intelligence artificielle et par une branche sociocognitive forte de la linguistique qui part du principe que compréhension et production impliquent de simuler le sens. Il ne s'agit plus d'une approche strictement représentationnelle (où les représentations sont des contenus) mais d'une approche actionnelle où la simulation est basée sur l'expérience incarnée, située et distribuée des locuteurs (Popova, 2015 pour une présentation). Or dans ce cadre-là, il est envisageable de modéliser les interactions homme-AVI en permettant l'asymétrie fondamentale dans la création sémantique et les ajustements constants de l'être humain. Cela permet aussi d'expliquer les limites perçues des AVI et leurs différences fondamentales : l'absence d'expérience crée des productions linguistiques détachées de la mémoire, du sensoriel et de l'affectif, qui n'en arbore plus que les marques. Un questionnement entre linguistique réelle et linguistique virtuelle adossée aux réflexions de la philosophie (voir ainsi Rebuschi, Renaud, 2019 ; Rebuschi, 2022) et alignée sur les avancées de l'analyse des discours numériques permet de ne pas créer deux objets de la linguistique mais deux modalités d'expérience du langage dans un continuum qui semble plus proche des pratiques observées. Enfin, cette approche par la simulation permettrait d'inscrire les analyses des interactions homme-AVI dans le cadre d'une linguistique qui prend en charge la situation, sa matérialité et sa multimodalité, ainsi que l'engagement des sujets humains dans différentes activités concomitantes. La linguistique cognitive propose des outils d'analyse comme la métaphore (Lakoff, Johnson, 1980), l'intégration conceptuelle (Fauconnier, Turner, 2002), ou encore l'énaction (Bottineau, 2013). Le travail théorique consisterait cependant à appliquer ces outils non pas au contenu de l'interaction mais au dispositif interactionnel lui-même dont les locuteurs se construisent au fil des interactions une représentation. Ainsi de P4 qui estime « parler à une personne qu'il vient de rencontrer ». L'absence d'intériorité construite à partir de l'expérience et liée à une mémoire discursive et autobiographique est ce qui limite l'illusion mais c'est aussi ce qui permet aux AVI de faire le mieux illusion, en simulant l'affect et le point de vue subjectif. En d'autres termes, la notion d'expérience apparaît comme le pivot d'un questionnement renouvelé en sciences du langage.

Références

- Benveniste, E. (1966/1974). *Problèmes de linguistique générale*, vol. 1/vol. 2. Paris : Gallimard.
- Bottineau, D. (2013). Pour une approche enactive de la parole dans les langues. *Langages*, 192, 11-27.
- Charaudeau, P. (2023). *Le sujet parlant en sciences du langage. Contraintes et libertés. Une perspective interdisciplinaire*. Limoges : Lambert Lucas.
- Culioli, A. (1990). *Pour une linguistique de l'énonciation*, vol. 1. Paris : Ophrys.

- Develotte, C., Paveau, M.-A. (2017). Pratiques discursives et interactionnelles en contexte numérique. Questionnements linguistiques. *Langage et société*, 2(2-3).
- Dingemanse, et al. (2023). Beyond Single-Mindedness : A Figure-Ground Reversal for the Cognitive Sciences. *Cognitive Science* 47.
- Fauconnier, G., Turner, M. (2002). *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. New York : Basic Books.
- Goodwin, C. (2018). Why multimodality? Why co-operative action? *Social Interaction. Video-Based Studies of Human Sociality*, 1(2), 85–98.
- Haddington, P., Keisanen, T., Mondada, L., Nevile, M. (2014). *Multiactivity in Social Interaction: Beyond multitasking*, Amsterdam/Philadelphia : Benjamins.
- Heritage, J. (2009). Conversation as social theory. *The New Blackwell Companion to Social Theory*, Bryan S. Turner (ed.). Wiley-Blackwell.
- Horton, D., Wohl, R. (1956). Mass communication and para-social interaction : Observation on intimacy at a distance. *Psychiatry*, 19 (3), 215–229.
- Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge : Cambridge University Press.
- Lakoff, G. (1987). *Women, Fire, and Dangerous Things. What Categories Reveal about the Mind*. Chicago : Chicago University Press.
- Lakoff, G., Johnson, M. (1980). *Metaphors we live by*. Chicago : Chicago University Press.
- Langacker, R. W. (1987). *Foundations of Cognitive Grammar*. Stanford University Press.
- Linell, P. (2005). *The Written Language Bias in Linguistics: Its Nature, Origins and Transformations*. Routledge.
- Linell, P. (2009). *Rethinking Language, Mind and World Dialogically: Interactional And Contextual Theories Of Human Sense-Making*. Charlotte, NC : Information Age Publishing.
- Maingueneau, D. (2021). *Discours et analyse du discours : Une introduction*. Paris : Armand Colin.
- Mondada, L. (2019). Contemporary issues in conversation analysis : Embodiment and materiality, multimodality and multisensoriality in social interaction. *Journal of Pragmatics*, 145, 47–62.
- Paillé, P., Mucchielli, A. (2016). Chapitre 11. L'analyse thématique. Dans P. Paillé & A. Mucchielli (dirs.), *L'analyse qualitative en sciences humaines et sociales* (pp. 235-312). Paris : Armand Colin.
- Paveau, M. A. (2009). Mais où est donc le sens ? Pour une linguistique symétrique. *Res per nomen* 2, 21-31.
- Popova, Y. B. (2015). *Stories, Meaning, and Experience: Narrativity and Enaction* (1st ed.). Routledge.
- Prak-Derrington, E. (2019). La litanie à travers les genres de discours. Une iconicité de l'extraordinaire. Dans P. Païssa & R. Druetta (dirs.), *La répétition en discours* (pp.171-199). Louvain-la-Neuve : Academia-L'Harmattan.
- Rebuschi, M. Renaud, M. (2019). Interaction n'est pas fiction ? Le cas de la rencontre homme-robot. *Klesis - Revue philosophique*, n° 44, 1-26.

- Rebuschi, M. (2022). Pour un (pseudo)dualisme du virtuel. *Klesis - Revue philosophique*
« Objets, œuvres, et mondes virtuels : problèmes esthétiques », n° 52, 1-17.
- Saussure, F. (1916). *Cours de linguistique générale*. Paris : Payot Rivages.
- Tannen, D. (1993). *Gender and conversational interaction*. Oxford University Press
- Tannen, D. (2007). *Talking Voices: Repetition, Dialogue, and Imagery in Conversational Discourse* (2nd ed.) Cambridge : Cambridge University Press.
- Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Harvard University Press.