HAL
open science

“ La madame de la voiture ne veut rien comprendre ” L’interaction vocale avec un ordinateur de bord d’automobile

François Perea

To cite this version:

François Perea. “ La madame de la voiture ne veut rien comprendre ” L’interaction vocale avec un ordinateur de bord d’automobile. Communication [Information Médias Théories] : revue québécoise des recherches et des pratiques en communication et information, 2020, 37 (1), 10.4000/communication.11371 . hal-03973285

HAL Id: hal-03973285

https://hal.science/hal-03973285v1

Submitted on 4 Feb 2023

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.

« La madame de la voiture ne veut rien comprendre »
L'interaction vocale avec un ordinateur de bord d'automobile.

François Perea. Université Paul-Valéry Montpellier III (France).

Les dispositifs et interfaces numériques constituent aujourd'hui une part importante de notre environnement. Cette écologie digitale est aussi vaste qu'hétérogène. Elle englobe des objets, réseaux, données… visibles ou invisibles, commandés par les usagers ou automatisés. Ainsi se réalise l'« intelligence ambiante » (Zelkha et al., 1998), reposant sur une *informatique diffuse* ou *ubiquitaire* (Weiser, 1991) c'est-à-dire décentrée des appareils et déployée dans les réseaux d'objets, sociaux et de données pour être accessible partout et en permanence. Cette intelligence est en outre capable d'*attentivité* et de *prédiction* en captant et traitant de multiples données pour une adaptation automatique et anticipatrice aux comportements des usagers, et favorise les *interactions naturelles*.

La domotique, ensemble des techniques électroniques, informatiques et des télécoms utilisé dans le contrôle et le fonctionnement des bâtiments, constitue l'un des domaines d'application privilégiés de ce déploiement. Les dispositifs sont désormais accessibles au grand public, soutenus notamment et fortement par les grandes firmes emblématiques du numérique, en prolongement des assistants personnels devenus ordinaires dans nos téléphones, nos tablettes ou nos ordinateurs (*Assistant Google*, *Alexa* chez Amazon, *M* chez Facebook, *Siri* chez Apple, *Cortana* chez Microsoft, pour ne citer que les GAFAM[1]) qui régulent notre quotidien et se déclinent sur les différents dispositifs ou dans des appareils spécifiques (*Google Home*, *Alexa* chez Amazon ou *Djingo* chez Orange par exemple). Ils apparaissent également dans des programmes spécialisés développés par les constructeurs (radiateurs ou stores « intelligents » par exemple).
Au-delà des murs de la maison, la domotique officie dans les autres espaces de vie. L'automobile, espace technique privé sur la voie publique, ne fait pas exception. En particulier, les dispositifs basés sur une interaction vocale humain-machine (désormais H-M) y sont développés (du GPS aux commandes vocales des véhicules récents) peut-être parce que la conduite occupe le regard et les gestes, rendant l'utilisation des écrans tactiles délicate si ce n'est dangereuse. Ces interactions sont diverses et variées et font se mêler interactions humaines, H-M, interactions entre machines. C'est à ce complexe interactionnel/interactif que s'attache le présent travail.

Après avoir présenté le cadre domotique de l'habitacle de l'automobile (1) et la perspective de la présente réflexion sur les dispositifs technolangagiers (2), ce travail explorera les différents aspects du formatage discursif et interactionnel à partir d'un dispositif concret (3) avant de conclure sur la « mécanisation » – pour reprendre à Sylvain Auroux – du langage telle que ce contexte d'observation permet de la saisir.

1. L'habitacle comme espace et cadre d'interaction

L'habitacle (du latin *habitaculum* : demeure) de l'automobile fonctionne en partie comme nos habitations. Pour Nicolas Dubois, la voiture est « un espace vécu comme un autre chez-soi » :

> La représentation de la voiture comme un potentiel support du sentiment d'être chez soi apparaît transversale dans la mesure où elle s'appuie sur différentes composantes réparties au sein de plusieurs valeurs principales. Ainsi, le sentiment d'être chez soi dans sa voiture emprunterait à sa valeur *utilitaire*, les aspects de sécurité, de modularité ou encore de centralité. De sa valeur *existentielle*, il en retiendrait l'expression de soi, la possession, l'image sociale. De sa valeur *ludique*, il en retire certainement la dimension d'ouverture sur le monde (Dubois, 2004 : 105).

Cet espace est hybride dans la perception et les représentations des conducteurs/propriétaires. Il permet de se déplacer dans l'espace public dans le respect du code de la route et des règles de vie collective, tout en demeurant un espace privé et personnel. L'habitacle est le lieu d'habitudes d'organisation (les rangements, les places) et de personnalisation (depuis le choix des options jusqu'aux éventuelles décorations en passant par les réglages) qui en font un prolongement de l'habitation. Comme cette

[1] L'acronyme *GAFAM* désigne les géants du numérique : Google, Amazon, Facebook, Apple et Microsoft.

dernière, l'automobile possède sa façade affichée (la carrosserie) et son intérieur où l'individu accueille les autres, à des places variées : tout le monde n'est pas autorisé à conduire comme tout le monde ne peut pas entrer dans la chambre à coucher par exemple. Elle abrite la vie de famille, amicale, et permet de faire un trait d'union entre les espaces fixes où s'ancre la vie des humains en une géographie propre (la maison, le lieu de travail, les écoles, les espaces de loisirs, les logements des amis, etc.).

Lieu personnel et lieu technologique, l'automobile est prédestinée à mêler les interactions automatiques ou pas, entre personnes et/ou entre objets, dans un espace réduit où elles se superposent parfois jusqu'à la confusion due aux possibles simultanéités et partages des écrans de médiation.
Le champ est vaste et nous n'aborderons pas ici les interactions entre les usagers à l'intérieur du véhicule, nous concentrant sur celles s'établissant entre le conducteur et le véhicule, qui peuvent inclure des échanges avec des « agents humains » (SMS ou appels téléphoniques via l'ordinateur de bord par exemple).

L'activité principale de la conduite peut, par essence, être considérée comme une interaction du conducteur avec/via le véhicule dans/sur son environnement. Il y a là une double dimension actionnelle : agir sur le véhicule (depuis son intérieur) pour agir sur l'espace extérieur (et le faire « défiler »). L'expérience est donc celle de la fixité dans l'habitacle et de la mobilité de/dans l'environnement. Cette interaction complexe articule une activité cognitive importante et une activité corporelle complexe au service de la manipulation de l'automobile en fonction du contexte (autres véhicules et autres conducteurs, environnement routier et piétons, cyclistes…) et en vue d'un objectif (se rendre à un endroit précis la plupart du temps). Elle accapare ainsi l'attention du conducteur au détriment de toute autre activité sous peine d'une mise en danger. Les activités annexes (lire un sms ou passer un coup de fil sur son téléphone portable, déplier et lire une carte sur le siège passager…) sont donc proscrites et sanctionnées (amende, perte de points sur le permis) sauf à être intégrées et rendues fluides dans l'activité principale.
Pour ce faire, le dispositif généralement retenu est celui de l'interaction orale qui n'entre en contradiction avec aucune manipulation nécessaire à la conduite. Les derniers véhicules permettent de lancer les commandes vocales sans avoir à quitter des mains le volant qui accueille un bouton spécifique. Alors, le GPS parle (permettant de réduire le temps de regard sur la carte) et répond à la demande orale de trajet ; le SMS est lu et/ou dicté, le coup de fil passé en mode les « mains libres ». La voiture s'est ainsi dotée d'un centre névralgique : l'ordinateur de bord, connecté aujourd'hui au téléphone portable, est le dispositif d'intelligence domotique de l'habitacle destiné, notamment, à communiquer avec l'usager.
En réalité, le dispositif est double : il allie un ordinateur de bord chargé de communiquer avec le véhicule et ses fonctions intégrées (radio, GPS, etc.) à l'éventuel smartphone connecté qui ajoute des fonctions supplémentaires (appel, sms dictés… mais aussi les liens avec les applications compatibles : musique en ligne, détection de radar, etc.). Le conducteur se trouve donc, pour les cas qui nous intéressent ici, en lien avec deux dispositifs connectés entre eux (l'ordinateur de bord et le smartphone) qui gèrent eux-mêmes des applications variées voire d'autres appareils connectés et qui communiquent entre eux. Ce double lien permet l'ancrage au véhicule et au monde, prolongeant le *chez-soi* par une connectivité perpétuelle (Katz et Aakhus, 2002) aux êtres et aux choses en un entrelacement de dispositifs de médiation et d'échange.

2. Approche des IOHM et corpus

Ces dispositifs imposent un ensemble de règles de fonctionnement et de formalisation de l'oral comme de l'écrit et conduisent au développement de nouvelles pratiques langagières induisant de nouvelles conceptions de la langue.

2.1. Techniques et langage

En procédant à une objectivation de certains traits du langage dont il a fallu sélectionner certaines propriétés (sonores et sémantiques) pour les matérialiser dans des signes tracés et en permettant une conservation de l'écrit sous des formes diverses (volumen, codex…) l'écriture a conduit à développer

des techniques linguistiques que Sylvain Auroux (1994) nomme « grammatisation ». Marquée par des conceptions et pratiques nouvelles articulées au travail du temps (par la fixation de la parole jusqu'alors éphémère) et de l'espace (par l'inscription dans la bidimensionalité du plan d'écriture), la « raison graphique » (Goody, 1979) s'est déployée comme registre de pensée des sociétés concernées :

> L'écriture a une importance décisive, non seulement parce qu'elle conserve la parole dans le temps et dans l'espace, mais aussi parce qu'elle transforme le langage parlé : elle en extrait et abstrait les éléments constitutifs, ainsi la communication par l'œil engendre des possibilités cognitives nouvelles par rapport à celle qu'offre la communication par la voix (1979 : 221)

Les technologies langagières digitales opèrent de la même manière des transformations cognitives et sociales. Bruno Bachimont observe alors un passage à la « raison computationnelle » :

> Les anthropologues, en particulier Jack Goody, évoquent une *raison graphique* pour expliciter le fait que l'écriture induit un mode de pensée particulier et un rapport au monde spécifique. Nous parlerons, quant à nous, d'une *raison computationnelle* pour expliciter le fait que nous pensons différemment avec les outils numériques (2007 : 71)

Tout contenu numérique résulte d'une construction dynamique via un calcul, visant à rendre le phénomène saisi compatible avec le traitement informatique obligatoirement binaire. Cela engendre de nombreuses manipulations articulées aux représentations et aux traitements informationnels dans/par des programmes, des couches (logicielles) et des réseaux. Cette nécessité entraîne de nouvelles conceptions des phénomènes numérisés comme l'écriture a entrainé de nouveaux traitements de la langue. Les incidences cognitives des technologies numériques apparaissent ainsi, et Katherine Hayles d'écrire :

> à mesure que les médias numériques […] se généralisent, ils nous poussent à une communication accélérée, des flux d'informations plus intenses et variées, une plus grande intégration entre machines intelligentes et humains, et de plus grandes interactions entre langues humaines et code informatique. Ces modifications environnementales ont des conséquences neurologiques importantes (2016 : 52).

Les sciences du langage ont investi la question des liens entre leur objet privilégié et la technologie de longue date et de manière identifiante : l'écriture a ainsi marqué la nature et l'histoire des théories linguistiques. Cependant, l'informatisation a renouvelé l'approche : le tournant du XXème a été particulièrement marqué, pour Jacqueline Léon (2015), par les enjeux de la traduction automatique d'une part et de la linguistique de corpus d'autre part. L'évolution se poursuit dans le sens d'une « mécanisation du langage » qui, pour Sylvain Auroux, s'inscrit dans la continuité de la grammatisation apparue avec l'écrit et procède par objectivation, formation et externalisation des faits et comportements langagiers, spontanés ou réfléchis. Bruno Bonu (2014) souligne cependant que si :

> l'histoire des Sciences du Langage s'est focalisée, à juste titre, sur le rôle de l'écriture dans l'outillage de la langue et dans la technologisation progressive de la linguistique, via l'écriture, par la grammatisation et la mécanisation […] les relations entretenues par les chercheurs avec les technologies, notamment avec les « machines sonores » (Flichy, 1991) ont des conséquences qui n'ont pas été complètement explorées (2014 : 1).

Puisque les dispositifs de traitement de l'oral sont de plus en présents, et notamment les interfaces d'interaction orale humain/machine, il est essentiel de se pencher sur le cas de ces machines aux incidences sur le langage des sujets parlants humains (il faut désormais le préciser).

2.2 Une approche linguistique des « objets parlants »

Notre approche est inscrite dans le champ de l'analyse de discours en interaction, en particulier dans l'approche Textes, Discours, Interactions (TDI) impulsée par Marie-Anne Paveau (2012a) et définie comme une linguistique symétrique (en référence à B. Latour, 1997) considérant les rapports entre humains et non-humains dans un continuum, et le social dans une dimension intersubjectives et interobjectives assemblées. Dès lors, « Les objets inanimés, qui ont une âme comme chacun sait, ont aussi une bouche : ils parlent, souvent, et beaucoup. Mais on ne les écoute sans doute pas assez » écrit Marie-Anne Paveau (2012b : 54).

Il s'agit alors de considérer les *objets* comme des locuteurs, peut-être et parfois même des *sujets* parlants. Cela ne signifie pas qu'ils sont équivalents ou dotés d'une subjectivité semble à celles des humains, mais que nous comptons dans la démarche d'analyse avec ce fait inédit dans l'histoire de l'Humanité : pour la première fois les hommes et les femmes peuvent parler à « autre chose » qu'un être humain et s'attendre à ce que ce destinataire lui réponde comme le ferait un humain. Le cas est différent des classiques incantations aux objets (la supplique à la voiture de démarrer) ou aux dieux (dans la prière) par exemple, car les locuteurs attendent de ces derniers cas des signes libérateurs ou révélateurs et non un échange ordinaire, *naturel* pour reprendre le terme technique consacré.

Ces objets parlants non humains sont donc bien distincts et particuliers bien qu'ils soient dotés, à des degrés divers, d'une forme de traitement du langage qui prend l'apparence d'une parole. Il s'agit là d'un simulacre (Perea, 2016) mais l'effet de subjectivité de ces objets parlants est déployé au-delà du don (de la programmation) de la parole comme ils sont parfois, dans une visée de marketing et/ou ergonomique, dotés d'un nom (Perea et Richard, 2018), d'un genre (Perea, 2018), etc.

Ces éléments participent à la continuité entre nature et artefact, les robots humanoïdes en sont le parangon) dont les incidences sont importantes sur le plan des comportements humains inscrits dans des environnements et des dispositifs d'interactions fortement affordants (Gibson, 1977). Il s'agit alors, dans ce contexte singulier, d'interroger « ce que disent les objets » (Paveau, 2012b) mais également « les efforts que déploient les humains pour interagir avec l'artefact » (Hutchby, 2001), c'est-à-dire notamment les adaptations langagières des locuteurs humains pour s'adapter aux locuteurs non humains.

2.3 Le corpus

Le présent travail s'attache au dispositif de gestion des interactions orales présents dans les véhicules. En l'occurrence, il s'agit du cas de *Connect Nav* tel que son principe et son fonctionnement sont observés dans un véhicule Citroën (C3 Aircross) de 2018 (il équipe les nouveaux véhicules de la marque et de Peugeot).
Ce dispositif est en lien avec l'écrit. Il complète les informations présentes sur l'écran tactile, comme c'est le cas avec les indications vocales et d'affichage d'une carte et des symboles liés au trajet lors de l'utilisation du GPS par exemple. Par ailleurs, l'écran affiche des conseils et des exemples lors de l'utilisation des commandes vocales. Ces exemples sont également présents dans le manuel imprimé des véhicules, comme dans cet extrait du manuel *C3 Aircross Citroën* (p. 219) :

CITROËN Connect Nav

Premiers pas

Exemple de "commande vocale" pour la navigation :
"Guider vers adresse, 11 rue de Rennes, Paris".
Exemple de "commande vocale" pour la radio et les medias :
"Jouer artiste Alizée".
Exemple de "commande vocale" pour le téléphone :
"Appeler Jean-Pierre".

Ce dispositif est activable sur l'écran tactile, véritable interface centralisatrice de l'habitacle, ou via un bouton situé sur le volant, ce qui évite au regard d'avoir à quitter la route.

3. Approche des Commandes vocales

Connect Nav permet à l'usager d'interagir avec le véhicule ou - via l'association d'un smartphone - avec des personnes.

Le manuel présente les thématiques des interactions possibles en cinq rubriques avec des exemples de commande :

AIDE
Aide | Aide commandes vocales | Aide navigation | Aide Radio | Aide média | Aide téléphone | Régler l'interaction vocale en mode <…> | Sélectionner le profil <...> | Oui | Non

NAVIGATION
Guider vers la maison | Guider vers le bureau | Guider vers destination préférée <...> | Guider vers contact <...> | Guider vers adresse <...> | Afficher POI <...> | Distance restante | Temps restant | Heure d'arrivée | Arrêter le guidage

RADIO MEDIA
Ecouter radio <...> | Qu'est-ce qu'on écoute ? | Jouer morceau <...> | Jouer artiste <…> | Jouer album <...>

TELEPHONE
Appeler <...> | Composer <…> | Afficher les appels | Appeler la boîte vocale

MESSAGES DE TEXTE
Envoyer un message rapide à <...> | Lire le dernier message

La présentation de ces commandes permet de se rendre compte du poids des exigences du traitement informatique dans ces interactions. Il ne s'agit pas (encore ?) d'une intelligence artificielle capable de converser mais bien d'une interface destinée à traiter des instructions vocales avec une capacité d'adaptation réduite. Alors, c'est bien à une forme particulière de traitement et de formalisation du langage avec laquelle nous avons affaire, qui conduit le locuteur à adopter/adapter son comportement pour se conformer au comportement prévu par le concepteur, dans une perspective loin du « mythe de la transparence des interfaces » (Scolaris, 2004, p. 84).
Ce mouvement d'ajustement langagier de la part du locuteur humain prend notamment deux aspects : le formatage interactionnel et le respect d'une matrice technique du discours.

3.1. Le formatage interactionnel

L'échange vocal entre le locuteur humain et le dispositif suppose que soient implémentés les possibles interactions et interactants. Nous retiendrons de ce travail de design programmatique deux aspects : les formats de la direction locutoire (qu(o)i parle à qu(o)i via qu(o)i), et la formalisation des possibles structures conversationnelles.

3.1.1. Les directions locutoires rendues possibles par l'interface rendent compte des possibles formats d'interaction possible. Les exemples de commande présentent de nombreux cas de figure, plus complexes qu'il ne paraît.

En règle générale, l'adresse correspond à la direction locutoire : la personne que je désigne comme mon interlocuteur est la personne à qui je destine mon message. Il y a bien sûr de nombreuses exceptions dans la communication humaine, souvent stratégique, comme lorsque l'on parle fort afin de se faire entendre par des tiers qui sont les réels destinataires du message.
La formule « H-M » et l'idée apriorique selon laquelle la formalisation nécessaire au traitement informatique suppose une simplification drastique, pourraient nous laisser croire qu'un Humain s'adresse à une Machine qui lui répond (ou pas). Les choses sont en réalité bien plus complexes.

On trouve d'abord le cas des *interactions H-M*, lorsque par exemple l'utilisateur énonce [Aide Radio] avec en retour une réponse de la part du dispositif.
Ce cas, qui pourrait paraître le plus simple, pose déjà la question de l'entrelacement des machines interlocutrices. En effet si les commandes d'aide (liées au fonctionnement du dispositif) et celles liées à l'ergonomie de base (telle *sélectionner le profil <…>*) renvoient au dispositif Connect Nav implémenté pour le constructeur, d'autres commandes invitent de nouvelles machines interlocutrices dans la communication. C'est le cas avec l'application de GPS qui, dans le cas d'espèce, est sous-traitée à l'entreprise néerlandaise Tomtom. Cela est également le cas avec les applications du smartphone qui serait connecté au dispositif central. Alors que la voix et le canal font que l'utilisateur ne s'adresse qu'à une seule machine (M), c'est en réalité un réseau complexe de dispositifs et programmes (m) implémentés dans le dispositif du véhicule ou à distance, qui est concerné par l'interaction :

H - [M(m-m-m…)]

Les *interactions humaines* viennent se greffer sur ce dispositif premier qui fait office de médiateur. C'est le cas pour les appels téléphoniques ou les SMS déclenchés par une commande vocale à Connect Nav (*Appeler <…>*, *Envoyer un message rapide à <…>*, etc.) connecté au téléphone et à son réseau assuré par l'opérateur.

H1 - [M(m-m-m…)] – H2

La sophistication du système est telle que ces formats sont emboitables. Ainsi, le lancement de l'écoute d'un message laissé sur le répondeur de l'utilisateur (commande : *Appeler la boîte vocale*), implique que Connect Nav (M) en lien avec une fonction du téléphone (m) mette l'utilisateur (H1) en contact avec une autre machine identifiée comme telle : la boite vocale (M2) dotée de « sa propre voix », qui pourra lui délivrer le message attendu de la part de H2.

H1 - [M(m)] <-> [(m)M2] – H2

La complexité de ces possibles est accrue par les possibilités de superpositions des interactions : l'utilisateur peut par exemple recevoir un SMS pendant qu'il est guidé par le GPS.
Les différents schémas présentés sont largement critiquables et imprécis. Ils ne visent qu'à rendre compte des processus complexes des adresses et des directions locutoires qui se cachent derrière l'apparente unicité de l'adresse à une machine dotée d'une voix unique, matériellement située dans l'écran de contrôle du tableau de bord.
Car cette multiplicité des machines interlocutrices n'apparaît pas dans un dispositif de surface à l'ergonomie simplifiée et aux compétences interactionnelles formellement réduites.

3.1.2. La structuration de l'interaction repose ainsi sur une réduction drastique car la gestion de la complexité du traitement automatique par les machines suppose que soient simplifiés les comportements humains reconnaissables et implémentables. L'un des aspects de cette réduction concerne en particulier la séquentialité (ouverture / clôture) et les formats conversationnels réduits la plupart du temps à un échange à deux termes (paire adjacente).

Le noyau des échanges utilisateur / Connect Nav est binaire et repose sur un type de paire adjacente particulier, dit *complémentaire*, de type question-réponse, à finalité pratique.
Ce modèle de commande / action se focalise sur l'action à réaliser.

Commande > Action

L'action peut être réalisée par le dispositif lui-même, « en interne », sur un modèle H-M, comme dans le cas de *Jouer morceau <…>*. Cette action peut également être « externe » en ce sens qu'elle suppose l'ouverture d'une nouvelle séquence interactionnelle hors du dispositif, comme avec *Appeler Aurore*.
Ce cœur du dispositif, faisant appel à des ressources internes ou supposant une connexion externe, est toujours inscrit dans une tension praxique visant :

- l'obtention d'une information (*Aide commandes vocales*, *Temps restant* ou encore *Qu'est-ce qu'on écoute ?*)
- la réalisation « automatique » d'une action (*Guider vers la maison*, *Jouer artiste <…>*, *Sélectionner le profil <…>*, etc.)
- la mise en relation avec une autre personne (*Appeler <…>*, *Envoyer un message rapide à <…>* ou *Composer <…>*).

Ce modèle commande/action renvoie donc aux unités minimales praxiques. Cependant, la complexité de certaines actions ou du besoin de l'utilisateur peut excéder ce cadre premier. Ces échanges s'inscrivent alors dans des séquences plus larges qui intègrent des actions intermédiaires aux actions finales.
Apparaissent alors des interventions encadrantes d'ouverture (par le conducteur) et de clôture (par le dispositif) non régies par la *politesse* au sens de Penelope Brown et Stephen Levinson (1987) ou un

rituel de préservation des faces (Goffman, 1974) mais dans un souci de réalisation de l'action finale composée d'une série d'actions intermédiaires.

OUVERTURE *Le conducteur énonce* :	**ACTIONS INTERMEDIAIRES**	**CLOTURE** *Le dispositif énonce* :
« Guider vers Agde »	Série d'actions de guidage unilatéral	« Vous êtes arrivé. Votre destination est à droite »
« Ecrire SMS à Aurore »	Ecoute et écriture du sms, incluant une ou plusieurs validations et des échanges métalinguistiques (comme « recommencer, ajouter quelque chose ou envoyer ? »)	« SMS envoyé »

C'est donc bel et bien une structure conversationnelle qui est utilisée dans le cadre de ces interactions, en ce sens que le noyau primordial est constitué de paires adjacentes, articulées en échanges combinés en séquences. La comparaison ne peut cependant occulter la réduction drastique de ces échanges.

D'abord, les échanges sont dépourvus de tous les aspects rituéliques essentiels et primordiaux dans les interactions entre les êtres humains. Dans le cadre du dispositif Connect Nav, ils ne sont même pas simulés comme cela est le cas avec certains assistants vocaux, tels le dispositif Google Home qui se met en route lorsque l'utilisateur énonce « OK Google ».
Ensuite, l'échange ne peut fonctionner qu'au prix d'une distribution des rôles interactionnels figés : le conducteur est à l'initiative (actions d'ouverture) et le dispositif à la réponse finale (actions de clôture). Il aurait été possible d'imaginer des scénarios conversationnels différents. Par exemple, dans le cadre de l'échange qui se termine par « SMS envoyé », le conducteur pourrait énoncer « Merci » qui, outre le simulacre rituélique, aurait valeur de commande de validation et de finalisation du processus par l'utilisateur. Le « SMS envoyé » étant alors une pré-clôture par la machine qui attendrait la validation par le conducteur ou une relance pour une suite (envoyer un autre sms par exemple ou réaliser une autre action).
Enfin, les interactions sont centrées sur un contenu praxique et/ou informationnel exclusif répondant à l'exigence absolue d'une logique actionnelle. Les échanges sont ainsi limités pour éviter tout bruit communicatif, à la fois dans leur nature (purement pratique) mais également dans leur forme réduite. Il s'agit donc d'une suite d'instructions brèves et prévisibles que d'échanges conversationnels.

Ces réductions reposent sur un double traitement des données linguistiques qui compose la matrice technique du discours : la simplification de la complexité du langage naturel (« les mots ») en vue d'un meilleur rendement du fonctionnement technique des matériaux langagiers composites (« les technomots »).

3.2. La matrice technologique du discours

Les services de Connect Nav allient ce que l'on nomme *intelligence artificielle* et *internet des objets*, et reposent fondamentalement sur une gestion automatisée de données. Le traitement linguistique réalisé par ces interfaces est entièrement dévolu vers cette finalité.
Peu importe si les êtres humains manient et s'inscrivent dans le langage différemment depuis l'aube de l'Humanité ; ils devront se soumettre à cette matrice technologique s'ils veulent être compris par les machines. Mieux, lorsque les interactions humaines transitent aussi en partie croissante par ses machines (à l'instar de certains exemples supra), il faudra qu'elles satisfassent à l'exigence de la reconnaissance de cette matrice pour se réaliser.

3.2.1. Syntaxe

Cette matrice repose sur une syntaxe aussi normative, stricte que réduite. Les énoncés produits par la machine sont en nombre limité et pré-enregistrés : ils ne proposent pas de difficulté technique et linguistique. Le véritable enjeu se situe du côté des productions de l'usager qui doivent être reconnues et traitées de manière pertinente afin d'apporter la réponse adaptée, en particulier dans une situation de conduite potentiellement dangereuse.
La forme essentielle des énoncés est composée soit d'une forme nominale figée, soit d'un verbe d'action appliqué à un élément lexical imposé (*maison, bureau*) ou à compléter (représenté par <…>).

FORMES NOMINALES	
Aide	
Aide +	commandes vocales
	navigation
	radio
	média
	téléphone
Oui	
Non	
Distance restante	
Temps restant	
Heure d'arrivée	

FORMES VERBALES	
Régler	l'interaction vocale en mode <…>
Sélectionner	le profil <…>
Guider	vers la maison | vers le bureau
	vers destination préférée <…> | vers adresse <…> | vers contact <…>
Afficher	les contacts | les appels
	POI <…> [je précise : *point of interest*]
Ecouter	radio <…>
Jouer	morceau <…> | artiste <…> | album <…> |
Appeler	<…>
	la boite vocale
Envoyer	un message rapide à <…>
Lire	le dernier message
	Qu'est-ce qu'on écoute ?

Le catalogue des énoncés possibles est réduit à des formes en nombre limité, brèves et contraintes afin de faciliter la reconnaissance.
Deux types d'instructions se distinguent :
Les *formes figées* seront « récitées » à l'identique par le conducteur. C'est le cas de toutes les formes nominales et de certaines formes verbales pour lesquelles le contenu sémantique des formes lexicales complémentaires auront préalablement été renseignées par le conducteur (les adresses professionnelle et personnelle) ou seront fixes (*dernier message, les contacts*…).
Les *formes variables* comportent un élément variable. Il s'agit de formes verbales à l'infinitif (temps des instructions), à l'ordre figé et implacable destiné à un traitement sémantique par affinement. Ainsi « Guider » conduit à 5 possibilités dont deux sont figées (*vers la maison, vers le bureau*) et trois variables. Dans ce dernier cas, un complément lexical (*destination préférée, adresse, contact*) va orienter le système vers la base de données idoine.

Il s'agit là du fondement de ce traitement vocal qui vise à repérer l'action de recherche à effectuer dans une base de données en privilégiant une syntaxe simple. Les formes figées renvoient à une information univoque (une heure d'arrivée, un accès à une aide) alors que les formes variables visent une réduction progressive du champ des possibles base de données interrogeables :

VERBE INFINITIF (position initiale) > **COMPLÉMENT LEXICAL** (réduction médiane) > **ÉLÉMENT CHERCHÉ : <…>** (position finale).

Ces énoncés ressortent d'un principe technolangagier double.

3.2.2. Technomots

Observant les discours numériques, Marie-Anne Paveau observe une dimension *composite* de certains éléments linguistiques. Outre leur valeur en langue, ils possèdent une fonction technique supplémentaire :

> Certains éléments langagiers produits dans les univers numériques connectés ou non sont dotés d'une dimension technique intrinsèque, c'est-à-dire non séparable de l'unité considérée : ils sont technolangagiers. C'est le cas de presque tous les éléments cliquables, qui présentent à la fois les caractéristiques du signe classique, doté d'un signifiant, d'un signifié et d'un référent, et celles d'un élément dynamique et manipulable […] Cette propriété est bien sûr celle de l'hypertexte […] au sein duquel tous les énoncés sont des liens. […] L'écriture numérique présente donc, dans la matérialité de ses signifiants scripturaux, deux traits fondamentaux qui doivent faire l'objet d'une analyse : la dimension composite ou technolangagière des technomots, des liens hypertexte ou des URL, et la nature relationnelle de ces segments permise par la manipulabilité de l'écrit numérique tant en écriture (élaboration de liens hypertextes) qu'en lecture (manipulation hypertextuelle du texte lu). (Paveau 2018, p. 76)

Il est évident que les corpus oraux observés ici connaissent le même traitement numérique conduisant à développer une dimension composite à valeur technolangagière sur une nature linguistique première.

De ce point de vue, l'élément <…> est la désignation scripturale d'un élément oral en suspens, tel un champ à remplir ou d'un menu déroulant à choix sur une page web. Il entre dans la famille des éléments désignant un événement technolangagier applicable (« énoncer un choix ici ») comme le soulignement désigne le lien hypertexte sur le web ou le @ un destinataire sur twitter.

Seul souci, la matière sonore ne laisse pas la nature composite être circonscrite comme à l'écrit par un signe symbole (#, @...). Le dispositif Connect Nav afforde alors doublement l'usager :

- à l'écrit par un signe symbolisation l'emplacement précis du choix ;
- à l'oral par une contrainte syntaxique forte : l'élément choisi peut n'apparaitre que dans un environnement contraint d'un point de vue syntagmatique (en fin d'énoncé) et paradigmatique (les éléments qui précèdent limitent le champ des possibles, à un artiste, à un morceau ou à un album dans une base de données musicales par exemple).

Le dispositif technolangagier est donc double.
Tous les énoncés potentiellement reconnaissables sont des technoénoncés car ils valent pour impulsions instructionnelles. Ils sont d'ailleurs arbitraires et l'on aurait pu donner un signifiant verbal différent à l'instruction, poussant au paroxysme ce trait du langage, en reliant l'action, de mettre la radio à un énoncé tel que « choix n°21 ».
Certains de ces énoncés reposent un second principe technolangagier, celui du choix qui est noté à l'écrit par un signe (<…>) et à l'oral par un champ de contraintes syntagmatique et pragmatique fort.

La mécanisation du langage prend ici une forme à la fois particulière et emblématique. Alors que l'écriture avait conduit, pour Sylvain Auroux, à développer des outils ayant travaillé une conscience métalinguistique et des pratiques particulières (dictionnaires, grammaires), les dispositifs numériques induisent une vision du langage semblable à un assemblage d'étiquettes instructionnelles prises dans de nouvelles grammaires (ici, l'exemple d'une syntaxe opérative d'un point de vue technolangagier) et qui ont des incidences sur les pratiques des sujets parlants soucieux de se faire entendre.
Un dernier point de vue sur Connect Nav permet de fournir un exemple supplémentaire.

3.2.3. Calibrage de la parole

Le formatage numérique ne concerne pas que les patterns linguistiques composant la liste des possibles énoncés reconnaissables et traitables : il a également trait aux caractéristiques individuelles qui peuvent conduire à des écarts de « compréhension » de la part du système. Dès lors, l'usager ne doit pas seulement appliquer des modèles, il doit aussi adapter la nature propre de sa parole.
Les aspects phonétiques en fournissent un bon exemple. On sait la grande variété individuelle, géographique, sociale… des prononciations et des accents. Or, la variété constitue une difficulté dans la reconnaissance vocale, parfois même un empêchement. Les témoignages de ces ratés sont nombreux, à l'instar de celui de Sonia Paul, journaliste américaine :

> Ma mère est née aux Philippines, mon père en Inde. Les deux parlent anglais comme troisième langue. Ils habitent aux États-Unis depuis près de 50 ans et parlent l'anglais quotidiennement, couramment, mais avec des accents distincts et des expressions parfois différentes de celles d'un locuteur natif. Selon leur expérience, cela signifie que Siri, Alexa ou tout appareil utilisant la technologie de la parole aura du mal à reconnaître leurs commandes [traduit depuis l'anglais, 2017, en ligne].

Ces particularités des locuteurs, qui font partie de leur histoire, de leur identité, constituent ainsi un frein à l'usage d'une technologie même si des efforts sont déployés pour intégrer le plus de variables phonétiques possibles. En conséquence, les usagers désireux d'utiliser les dispositifs à reconnaissance vocale doivent calibrer leur accent pour se faire comprendre par la machine. Le sujet est tel que certains échangent trucs et astuces, à l'instar des membres d'un forum public pour les possesseurs de véhicules Citroën C5 et C6 (http://www.forumc6.com/t452-reconnaissance-Vocal.htm#p7207, consulté le 31 novembre 2018) :

PSEUDO	**MESSAGES** (extraits de l'échange de 2010, entre habitants du sud de la France)
Patrol.vite	La reconnaissance Vocale est bien pratique. Mais ma C5 ne reconnaît pas le chiffre 1, J'ait essayé avec différentes intonation, avec l'accent du nord , en anglet et c'est en español quelle comprend le mieux. Pensent que le problème venais de moi j'ai fait un test avec d'autre personnes, mais la non plus rien n'y fait. avez vous se genre de problème que puis-je faire ?
Fab.26	La je peux témoigner Corentin articule bien mais la madame de la voiture ne veut rien comprendre.
Philou	Pour que le 1 soit reconnu, il faut prendre la voix de Sarko... forcement !
Belial	Plus serieusement (quoique), j'ai un truc qui marche bien pour le 'un'. J'ai même fait un petit enregistrement video, mais je ne sais pas comment le poster.

> En gros, il faut faire un 'un' assez gutural, en essayant de faire un son assez grave. Ca marche bien, mais il faut eviter quand il y a du monde dans la voiture .
> Si vous avez un truc pour poster une video au format 3gp, je vous promet de la poster sur le forum.
> PS: Ce truc marche bien avec le RT3 de la C6, je pense que ca doit aussi aller pour la C5.

Les aspects phonétiques ne représentent qu'un aspect de ces pratiques langagières calibrées pour la machine. La réduction des phrases aux seuls termes informatifs, l'absence de formules rituéliques… constituent les éléments d'une nouvelle manière de parler dans ces contextes qui deviennent de plus en plus fréquents et étendus. Ces nouveaux usages se déploient ainsi lors d'interactions variées : routeurs téléphoniques invitant à énoncer des termes clés, dictée de sms au téléphone, chatbot en ligne…

Il est intéressant de souligner que, même lorsque les fabricants ne jouent pas sur la personnification des interfaces (à l'instar de Connect Nav et à la différence de Cortana par exemple), les usagers convoquent les habitus du langage interhumain. Ainsi, de Fab.26 dans les exemples ci-dessus : « La je peux témoigner Corentin articule bien mais la madame de la voiture ne veut rien comprendre ».

Dès lors, c'est une dimension souvent occultée des dialogues *naturels* homme-machine qui apparait. Si l'adjectif désigne d'ordinaire une caractéristique de l'interface qui utilise les ressources ordinaires de la communication humaine (ici, le langage oral), il masque une partie du travail de cette interaction. En même temps que la machine simule le langage naturel, l'humain doit mécaniser son langage.

4. En guise de conclusion

Marie-Anne Paveau souligne le poids des exigences techniques sur l'écriture numérique :

> Les contraintes que la technique fait peser sur l'écriture numérique sont de deux ordres : à un niveau macro, il s'agit des déterminismes des formats propres aux dispositifs d'écriture ; à un niveau micro, il s'agit de la nature des éléments langagiers en contexte numérique qui intègrent de manière intrinsèque une dimension technique (la dimension composite des éléments technolangagiers) (Paveau 2018 : 73).

Il est désormais évident que les mêmes contraintes s'exercent sur les dispositifs oraux, comme sur les dispositifs mixtes.

Dans le cadre des interactions entre homme et machine, ces contraintes sont en partie masquées par la mise en avant de la « naturalité » des échanges, censés dispenser les usagers humains de tout effort d'apprentissage d'un quelconque code informatique. C'est le dispositif, collant aux comportements humains, qui va prendre en charge ce travail de rapprochement : « Adressez-vous à Siri comme à une personne » trouve-t-on sur le site d'Apple (janvier 2017), alors que le téléphone Huawei Mate 10 Pro « doté d'une intelligence artificielle […] n'est pas artificiel, il est humain », selon la publicité télévisée de la firme (décembre 2017).
Les dispositifs tels que Connect Nav ne jouent pas sur cet aspect : les interfaces ne reposent sur un contrat communicationnel de personnification même si les utilisateurs, par affordance générale sur le langage qui est un fait humain, les personnifient quand même (« la voix de la Madame »). Au contraire, Connect Nav est caractérisé par sa sobriété et sa simplicité apparente. Il n'est pas présenté comme une révolution technologique ou comme un co-pilote virtuel.
Cette caractéristique met l'accent sur le travail d'adaptation permanent des sujets humains dans l'interaction, la demande d'un comportement langagier « mécanique » :

« HUMANISATION » DU COMPORTEMENT DE LA MACHINE → **DNHM** ← « MECANISATION » DU COMPORTEMENT HUMAIN

Cet aspect est bien moins célébré dans les discours publicitaires. Il est probablement bien moins pensé dans le cadre du développement des interfaces, qui privilégie le travail de l'outil aux praxis réelles et effectives des utilisateurs, et se contente trop souvent de les inviter à une utilisation améliorée ou de supposer des usages.

Ainsi, la technique impose un ensemble de règles et de formalisations nouvelles dont nous avons montré ici certains aspects liés aux formatages interactionnels (réduisant les apparentes directions locutoires)

et structurant l'interaction de manière prévisible (et donc programmable), à travers des routines interactionnelles et des réductions syntaxiques et lexicales strictes.
Ces structurations ont des conséquences diverses éloignées, et cela est normal, des conversations « naturelles ». Elles reposent sur :

- la simplification drastique des énoncés à des formes procédurales et codées, contraintes dans leurs dimensions syntagmatiques et paradigmatiques ;
- la normalisation (au sens strict du terme) des comportements langagiers, pour répondre aux canons calibrés du décodage phonétique, lexical et syntaxique automatique ;
- le traitement sémantique réduit à un traitement de texte, c'est-à-dire au sens littéral des énoncés inscrits dans des bases de (méta)données et soumises à statistique.

En même temps que sont structurées des pratiques langagières, de nouvelles conceptions linguistiques (prêtes à alimenter de nouveaux dispositifs affordanciels) apparaissent et dessinent des visions du langage étriquées où les mots sont des étiquettes univoques, prises dans des codes qui suffisent en eux-mêmes à faire sens. Cette mécanisation du langage est séduisante pour certains et effrayante pour d'autres : elle permet de rationnaliser le comportement humain, de le rendre prévisible, implémentable, réductible à des données saisissables par des intelligences artificielles. Et Bruno Bachimont d'écrire :

> En considérant le fait humain comme un fait mobilisé dans une base de données, ces sciences adoptent de fait un positivisme naïf où le fait collecté est la seule positivité à interroger et d'où il faut partir pour appliquer les modèles de construction scientifique (2014, p. 77).

Ici, il ne s'agit pas, bien entendu, de sombrer dans un pessimisme digital et d'accuser la technique d'appauvrir le langage. Le fait humain est ici présent et prégnant, car ce sont bien des femmes et des hommes qui implémentent, codent et utilisent ces dispositifs. Il faut savoir se dégager d'une vision techniciste, promue par les discours publicitaires et les chantres de la révolution numérique réduite aux développements des processus informatiques, pour ne pas laisser le mode d'emploi devenir une norme langagière, même si l'interopérabilité et l'internet des objets poussent à de tels rapprochements nécessaires à soutenir la connexion de tout à tout dans les environnements digitalisés.
Aux usagers, inspirés par Michel de Certeau, de pratiquer le « braconnage » numérique et de se souvenir qu'interagir vocalement, c'est co-construire la réalité, formaliser la pensée à l'épreuve et au travail de l'autre, en adaptation mutuelle entre humains… avec ou sans prothèse numérique. Une affaire qui dure depuis que les êtres se parlent et qui n'a aucune de raison de cesser.

BIBLIOGRAPHIE

- AUROUX, Sylvain (1994), *La révolution technologique de la grammatisation*, Liège, Mardaga.
- BACHIMONT, Bruno (2014), « Le nominalisme et la culture : questions posées par les enjeux numériques », dans *Digital studies, organologie des savoirs et technologies de la connaissance*, sous la direction de Bernard Stiegler, éd. Fyp.
- BACHIMONT, Bruno (2007), *Ingénierie des connaissances et des contenus*. *Le numérique entre ontologies et documents*, Hermès / Lavoisier.
- BONU, Bruno (2014), « "L'autre" révolution technologique en sciences du langage : les cas du phonographe et du magnétophone à cassette » Dossiers d'HEL, SHESL, 2014, *Linguistiques d'intervention. Des usages socio-politiques des savoirs sur le langage et les langues*, pp.9. <http://dossierhel.hypotheses.org/>.
- BROWN, Penelope et LEVINSON, Stephen (1987), *Politeness*, Cambridge, CUP.
- DUBOIS, Nicolas (2004), « L'automobile : un espace vécu comme un autre chez-soi », thèse de doctorat, université de Nanterre - Paris X.
- FLICHY, Patrice (1991), « L'historien et le sociologue face à la technique. Le cas des machines sonores », *Réseaux* n°9, p. 47-58.
- GIBSON, James Jérôme (1977), *The Theory of Affordances. In Perceiving, Acting, and Knowing*, Eds. Robert Shaw and John Bransford.
- GOFFMAN, Erwing (1974), *Les rites d'interaction*, Paris, Ed. de Minuit.
- GOODY, Jack (trad. Jean Bazin et Alban Bensa) (1979 ; or. 1977) *La raison graphique. La domestication de la pensée sauvage*, Paris, Les Éditions de Minuit, coll. « Le sens commun », 272 p.
- HAYLES, N. Katherine (2015), *Parole, écriture, code*, Paris, éd. Les presses du réel.
- HUTCHBY, Ian (2001), *Conversation and Technology: from the Telephone to the Internet, Cambridge*, Polity Press.
- KATZ, James E. et AAKHUS, Mark (2002), *Perpetual Contact. Mobile Communication, Private Talk, Public Performance*, Cambridge et New York, Cambridge University Press.

- LATOUR, Bruno (1997), *Nous n'avons jamais été modernes. Essai d'anthropologie symétrique*, Paris, La Découverte.
- LEON, Jacqueline (2015), *Histoire de l'automatisation des sciences du langage*, Paris, éd. ENS.
- PAUL, Sonia (2017), « Voice is the next big platform, unless you have an accent », *Backchannel*, 20 mars, consulté le 31 octobre 2018 : https://www.wired.com/2017/03/voice-is-the-next-big-platform-unless-you-have-an-accent/#.5iveuiybx.
- PAVEAU, Marie-Anne (2012a), « Réalité et discursivité. D'autres dimensions pour la théorie du discours », *Semen*, n°24, 95-116.
- PAVEAU, Marie-Anne (2012b), « Ce que disent les objets. Sens, affordance, cognition », *Synergie*, Pays Riverains de la Baltique, n°9, 53-65.
- PAVEAU, Marie-Anne (2018), *L'analyse du discours numérique. Dictionnaire des formes et des pratiques*, éd. Harmann.
- PEREA, François (2016), « Nature et technologie langagière dans les dialogues oraux homme-machine », *Communication* [En ligne], vol. 34/1, mis en ligne le 26 août 2016, consulté le 24 août 2018. URL : http://journals.openedition.org/communication/6763.
- PEREA, François (2018), « Cortana est-elle une femme comme les autres ? Éléments de discours de fabrication du genre d'un artefact numérique », *Semen*, n°44, 61-80.
- PEREA, François et Richard Arnaud (2018), « De qu(o)i Siri est-il le nom ? », *Le discours et la langue*, t. 10.1, 221-232.
- SCOLARI, Carlos Alberto (2004), « Hacer clic: hacia una sociosemiótica de las interacciones digitales », *DeSignis*, n°5, *Corpus Digitalis. Semióticas del mundo digital*, p. 73-84.
- WEISE,R Mark (1991), « The Computer for the Twenty-First Century », *Scientific American*, vol. 265, n° 3, 3-11.
- ZELKHA, Eli, EPSTEIN, Brian, BIRRELL, Simon et DODSWORTH Clark (1998), "From Devices to "Ambient Intelligence", Digital Living Room Conference (published June 1998)