# Comparaison linguistique et neuro-physiologique de conversations humain humain et humain robot

Charlie Hallart, Juliette Maes, Nicolas Spatola, Laurent Prévot, Thierry Chaminade

# Comparaison linguistique et neuro-physiologique de conversations humain-humain et humain-robot

**Charlie Hallart[$*] — Juliette Maes[$*] — Nicolas Spatola[**] — Laurent Prévot[***] — Thierry Chaminade[§*]**

[$] *Ordre alphabétique, contribution identique.*
[*] *Aix Marseille Univ, CNRS, INT, Marseille, France,*
[**] *Italian Institute of Technology, Genova, Italy,*
[***] *Aix Marseille Univ, CNRS, LPL, Marseille, France.*
[§] *Auteur pour correspondance.*

RÉSUMÉ. *Nous décrivons l'analyse d'un corpus de conversations humain-humain et humain-robot, pour lequel des données neurophysiologiques ont été acquises pendant les conversations. 21 participants ont été scannés en Imagerie par Résonance Magnétique fonctionnelle (IRMf) pendant qu'ils discutaient soit avec un humain, soit avec un robot prétendument autonome. En s'inspirant de ce qui est communément utilisé pour étudier les conversations orales, huit variables linguistiques adaptées aux spécificités du corpus ont permis de mettre en évidence les compétences linguistiques limitées du système de Magicien d'Oz utilisé pour contrôler le robot. Nous avons également adapté une variable d'alignement lexical qui nous permet d'étudier l'alignement conversationnel, plus important dans les interactions avec le robot qu'avec l'humain. Enfin, nous mettons à profit la disponibilité de données d'activité cérébrale pour étudier la réduction du contrôle cognitif associée à l'augmentation de l'alignement lexical du participant sur l'interlocuteur.*

ABSTRACT. *Here we describe the analysis of a unique corpus of conversations with a human or a robot, synchronised with neurophysiology data. 21 participants were scanned with functional Magnetic Resonance Imaging (fMRI) while talking either with a human or with a robot presented as autonomous. Inspired by what is commonly studied in natural oral conversations, eight linguistic variables adapted to the specifics of the corpus highlight the limited linguistic skills of the Wizard of Oz system used to control the robot. We also calculate a lexical alignment variable which allows us to study the phenomenon of conversational alignment, increased with the robot compared to the human. Finally, we take advantage of the availability of synchronized*

R1[e] soumission à *TAL 61-3*, le 20/06/2020

*brain activity data to suggest a reduction in cognitive control associated with an increase in the participant's lexical alignment with the interlocutor.*

MOTS-CLÉS : *Conversation, Humain, Robot, Neurosciences, Alignement lexical.*

KEYWORDS: *Conversation, Human, Robot, Neurosciences, Lexical alignment.*

## 1. Introduction

La conversation, activité complexe et ordonnée, est généralement décrite comme un échange de propos libres et spontanés entre plusieurs locuteurs autour d'un thème commun. Au cœur de l'interprétation des relations sociales humaines, la conversation permet aux individus de construire leur identité, de forger leur pensée et leurs idées, et d'établir leur rôle social. À défi déjà complexe, l'analyse conversationnelle voit grandir, avec le développement de l'intelligence artificielle et des interfaces Humain-Machine, l'importance d'étudier non seulement les conversations naturelles, mais également les conversations avec des agents artificiels. En effet, que les individus conversent avec l'assistant vocal de leur téléphone ou avec un robot humanoïde, les conversations entre les humains et les interfaces artificielles s'intègrent de plus en plus au quotidien. Il est pour cela primordial de comprendre les mécanismes qui sous-tendent ces nouvelles interactions.

Dans ce contexte, cet article décrit l'exploitation d'un corpus qui combine des données linguistiques, comportementales et neurophysiologiques synchronisées de 21 participants *(terme utilisé dans cet article pour parler du conversant installé dans le scanner IRM dont on mesure l'activité cérébrale)*. Le corpus rassemble les enregistrements de conversations naturelles qui ont eu lieu entre chacun des participants et un interlocuteur *(terme utilisé pour le conversant situé dans une pièce annexe)*, qui peut être soit un humain, soit un robot.

Nous partons de l'observation que l'interlocuteur artificiel possède des compétences linguistiques différentes de celles de l'humain. Le robot est contrôlé par un système de Magicien d'Oz doté de compétences linguistiques limitées (nombre de phrases restreint, intonation peu variable, latence dans le temps de réponse...). Ces limitations peuvent sembler fortes au regard de certaines démonstrations de robots humanoïdes pour le grand publique, mais il faut rappeler que ces démonstrations résultent de nombreuses heures de programmation et suivent des scénarios assez restreints. A l'inverse, il s'agit ici d'une réelle conversation où il faut que les réponses du robot soit adaptées aux interventions parfois imprévisibles du participant. Nous avons utilisé au maximum de ses capacités ce robot conversationnel. La voix utilisée, notamment, était une voix francophone d'*Amazon Polly*, assez naturelle mais peu expressive. Cette recherche s'inscrit dans un projet ayant pour but l'amélioration des compétences linguistique et sociale des robots pour en étudier les effets sur le comportement et l'activité cérébrale. Dans notre protocole, les participants sont amenés à croire que le robot est autonome, ce qui est renforcé par les capacités conversationnelles limitées et la latence de réponse du système de Magicien d'Oz utilisé pour le contrôle. En partant de la proposition que nous adoptons une posture différente selon que nous agissons avec un être doué d'états mentaux ou avec une machine (Dennett, 1987), les participants devraient se comporter différemment avec ces deux interlocuteurs. Pour ces raisons, nous posons l'hypothèse que la nature des interlocuteurs a un effet sur la production langagière des participants, et son évolution aux cours du temps.

Dans un second temps, nous nous intéressons au phénomène d'alignement conversationnel, un processus cognitif que la littérature linguistique décrit comme un principe fort et robuste des interactions humaines (Branigan *et al.*, 2000 ; Branigan *et al.*, 2007). Plus spécifiquement, l'alignement conversationnel se produit à un niveau verbal ou para-verbal lorsque les individus en interaction emploient un lexique commun, des structures syntaxiques identiques ou encore les mêmes patrons prosodiques (Pickering et Garrod, 2004). Notre corpus, constitué de conversations naturelles bidirectionnelles entre deux humains, mais également de conversations comparables entre un humain et un robot, nous permet d'étudier l'alignement selon deux directions. En effet, nous étudions l'alignement conversationnel du participant sur l'interlocuteur selon que ce dernier est humain ou robot, mais nous pouvons également étudier l'alignement conversationnel de l'interlocuteur sur le participant. En partant des travaux de Branigan *et al.* (2000) et Branigan *et al.* (2007) selon lesquels les locuteurs tendent à s'aligner davantage avec un robot dont les capacités langagières sont limitées, nous posons l'hypothèse que le participant s'aligne davantage avec l'interlocuteur robot plutôt qu'avec l'interlocuteur humain. En considérant ensuite que le comportement langagier du robot est limité par le paradigme du Magicien d'Oz, et que l'interlocuteur humain est ainsi le seul capable d'adapter finement son discours, nous posons l'hypothèse que l'interlocuteur humain s'aligne davantage sur le participant que l'interlocuteur robot.

Pour vérifier cette hypothèse, nous adaptons une variable d'alignement lexical qui nous permet d'étudier le lexique que les locuteurs utilisent afin de faire référence à des objets ou à des concepts communs. Les données de neuroimagerie associées à ces interactions nous permettent de tester que cette nouvelle variable d'alignement correspond à une réalité cognitive en recherchant ses corrélats cérébraux. Notre approche statistique est d'abord validée en l'utilisant pour vérifier qu'elle identifie correctement les corrélats cérébraux de la production et de la perception de language (Price, 2010).

Cet article commence par une présentation des données (Section 2) avant de présenter la définition et le calcul et l'analyse des variables linguistiques étudiées dans la section 3. Nous décrivons ensuite le travail effectué sur l'alignement entre les conversants (Section 4 avant de présenter les liens de notre travail avec la dimension neurophysiologique du jeu de donnée (Section 5). Nous terminons la discussion de l'ensemble des résultats dans la section 6 avant de conclure par un bref résumé.

## 2. Corpus

Les fondements théoriques qui sous-tendent le choix du paradigme expérimental et les procédures d'acquisition et de préparation du corpus, à la fois pour les données linguistiques et neurophysiologiques, ont été publiés ces dernières années (***, ***). Toutefois, et même si nous invitons le lecteur à utiliser ces références pour plus de détails, nous souhaitons rappeler les principaux points pour que cet article soit lisible de manière autonome.

| |
|---|
| Bonjour<br>Je m'appelle Furhat<br>Comment ça va ? |
| Oui<br>Non<br>Peut-être |
| C'est une poire jaune<br>La poire semble triste<br>Peut-être qu'elle est malade et elle devenue pourrie |
| Tu as une idée du message ?<br>C'est peut-être une campagne pour favoriser les fruits locaux<br>Ça pourrait être une pub pour des producteurs de fruits |

**Tableau 1.** *Exemple de phrases pré-enregistrées prononcées par le robot, groupées selon leur fonction dans la conversation : présentations, réponses génériques, descriptions d'une image (une poire dans la série des fruits pourris), et échanges sur le message de la campagne de pub*

### 2.1. *Cadre de l'expérience*

Afin d'étudier les interactions sociales naturelles, il est essentiel que les participants ne soient pas conscients du véritable objectif de l'expérience. À cette fin, l'expérience est présentée comme une expérience de neuromarketing, dans laquelle une entreprise veut savoir s'il suffit de discuter à propos des images d'une campagne de publicité à venir, soit avec une autre personne, soit avec une intelligence artificielle incarnée dans un robot conversationnel, pour deviner le message de la campagne (***, ***). L'interlocuteur humain est un expérimentateur, du même genre que le participant, mais présenté au participant comme un participant naïf comme lui. Le robot est une tête robotique conversationnelle rétroprojetée (Furhat robotics[1]). Il possède une apparence physique anthropomorphique incluant un visage, un genre, une voix et divers accessoires humains (perruque, casque audio, lunettes) pour ressembler à l'interlocuteur humain (voir Figure 1, gauche). Il est contrôlé par l'interlocuteur humain avec un système de Magicien d'Oz : il dispose d'un ensemble fini de réponses préenregistrées (voir Tableau 1) que l'expérimentateur choisit en appuyant sur des boutons virtuels sur une tablette tactile. Certaines réponses sont génériques et d'autres spécifiques d'une image ou du message de la campagne publicitaire. Ne pouvant pas répondre à des concepts ou des mot-clés qui n'ont pas été programmés, les échanges avec le robot sont donc nécessairement limités en termes de variété et de spontanéité. Il n'y a pas d'intonation sauf pour les phrases interrogatives et le système de Magicien d'Oz induit un délai qui perturbe la spontanéité des échanges. Ainsi, les participants discutent en fait avec la même personne dans les conversations avec l'humain et le robot mais la

1. `https://www.furhatrobotics.com` ; (Al Moubayed *et al.*, 2012)

| |
|---|
| Participant : du coup là c'est une poire |
| Participant : euh |
| Participant : un peu cabossée |
| Participant : et euh |
| Participant : du coup il semblerait qu'elle soit avinée |
| Participant : et |
| Interlocuteur : ouais |
| Interlocuteur : ouais ouais |
| Participant : et euh en fait euh |
| Participant : au départ je pensais qu'elle était triste mais au final non j'ai l'impression que c'est un petit sourire en coin |
| Interlocuteur : ah ouais |
| Participant : ouais je vois pas -fin bon tristesse -fin c'est neutre euh |
| Participant : c'est pas euh si triste |
| Interlocuteur : moi elle m'avait l'air euh je sais pas comment dire dépitée peut-être |
| Participant : comment |
| Interlocuteur : dépitée peut-être pour moi |
| Participant : ah oui peut-être ouais |
| Participant : oui un peu perplexe dépitée mais pas triste au final |
| Interlocuteur : ouais perdue aussi comme tu disais |
| Participant : oui peut-être |
| Participant : ouais |
| Interlocuteur : ouais |
| Interlocuteur : ça fait euh ça fait un gros une grosse différence avec euh les fruits de la première campagne |
| Participant : euh oui |
| Participant : totalement et aussi avec la framboise qui est plus euh |
| Participant : mh |
| Interlocuteur : gélatineuse |
| Participant : plus d'émotions -fin |
| Interlocuteur : ah |
| Participant : oui |
| Interlocuteur : moi je pensais à @ à l'aspect nourriture |
| Participant : à la texture |

**Tableau 2.** *Transcription complète des échanges d'un essai entre un participant et l'humain*

médiation par le robot appauvrit significativement la qualité de la conversation, alors que la croyance en son autonomie modifie la posture intentionnelle (Dennett, 1987) du participant.

| |
|---|
| Participant : alors là c'est une fraise qui est encore abîmée |
| Participant : et et qui avait l'air |
| Participant : perdue |
| Participant : et défoncée |
| Participant : pour moi |
| Participant : euh |
| Interlocuteur : comme les deux autres |
| Participant : euh non |
| Participant : pas trop non les autres ils avaient plus une expression euh de douleur ou euh |
| Participant : * |
| Participant : et euh |
| Participant : voilà |
| Interlocuteur : peut-être |
| Interlocuteur : cette fraise est déformée |
| Participant : oui |
| Participant : sur les côtés |
| Interlocuteur : la fraise est aussi pourrie |
| Participant : euh |
| Interlocuteur : qu'est ce que tu en dis |
| Participant : bah pourrie je sais pas mais déformée oui |
| Participant : abîmée |
| Interlocuteur : la fraise est un peu abîmée |
| Participant : ouh ben là c- |
| Participant : c'est comme |
| Participant : comme la poire en fait |
| Interlocuteur : comme la poire et la framboise |

**Tableau 3.** *Transcription complète d'un échange entre un participant et le robot. Il faut noter qu'il s'agit de l'essai suivant directement celui donné dans le Tableau 2*

### 2.2. *Acquisition des données*

Sur les 25 participants du corpus d'origine, seuls sont inclus dans les analyses présentées ici (15 femmes, m = 27.04 ans, e.t. = 8.19, [21-49]). Le premier participant est exclu puisque les données comportementales de sa quatrième session sont manquantes. Les participants 4 et 23 sont exclus pour manque de participation active à l'expérience. Le participant 19 est exclu pour cause de mauvaise audition de l'interlocuteur robot pendant l'expérience.

Les participants se présentent au centre IRM, où un expérimentateur leur introduit les deux interlocuteurs, l'expérimentateur humain et le robot. Après que l'histoire de neuromarketing ait été présentée, le participant est installé dans le scanner. Quatre sessions d'IRM sont enregistrées successivement. Dans chacune des sessions, le par-

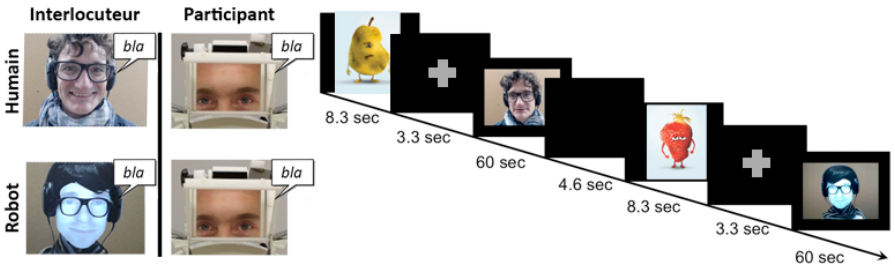


**Figure 1.** *Présentation des deux conditions expérimentales (Participant/Humain et Participant/Robot) à gauche, et organisation des essais à droite, avec les images des exemples donnés en Tables 2 (poire) et 3 (fraise)*

ticipant parle trois fois une minute avec l'humain, et trois fois une minute avec le robot, alternativement, en commençant chaque session avec l'humain. Chaque conversation d'une minute est appelée un essai. Au total, chaque participant réalise 24 essais, conduisant à 24 minutes de conversation enregistrées (douze fois une minute avec l'humain, douze fois une minute avec le robot). Avant chaque essai, le participant voit apparaître sur l'écran pendant 8.3 secondes une image qui illustre la campagne publicitaire dont il doit discuter (voir Figure 1). Dans deux sessions d'acquisition IRM, il s'agit de fruits et légumes déguisés en super-héros, et dans les deux autres sessions ce sont des fruits pourris. L'ordre de présentation des images est le même pour tous les participants, et chaque image est montrée deux fois par participant : dans une session d'une campagne publicitaire, avant de parler avec l'humain, et dans l'autre, avant de parler avec le robot. L'image est suivie d'une croix de fixation grise sur un fond noir (3.3 secondes), puis de la conversation d'une minute entre le sujet et l'un des interlocuteurs, et enfin d'un écran noir (4.6 secondes). Ceci se répète jusqu'à ce que le participant ait complété les six essais d'une session. Au cours de ces échanges, différentes données brutes sont acquises. Pour les analyses présentées dans cet articles, les données sont les enregistrements vocaux du participant (à l'aide d'un microphone à réduction active du bruit compatible IRM) et de l'interlocuteur, et l'activité cérébrale enregistrée par le scanner IRM.

### 2.3. *Préparation des données*

Les données brutes doivent être préparées pour les analyses automatiques décrites dans la suite de cette présentation. Nous nous concentrons sur les deux types de données exploitées ici, à savoir les données linguistiques et les données neurophysiologiques. Les enregistrements audio des participants sont filtrés pour réduire les bruits du scanner qui n'ont pas été éliminés par le filtre actif du microphone. Les données débruitées des participants et celles des interlocuteurs sont ensuite segmentées en unités inter-pausales (UIP), qui ont été définies comme des blocs de parole délimités par des

silences d'une durée minimale de 200 ms (Blache *et al.*, 2009). L'inspection visuelle du débruitage et de la segmentation a été effectuée à l'aide du logiciel de traitement de la parole oral Praat (Boersma, 2001). Les fichiers audio et les segmentations sont ensuite téléchargés dans SPPAS[2] pour transcription. Les transcriptions orthographiques des productions langagières des participants et des interlocuteurs (fichiers .TextGrid) sont réalisées manuellement sur des fichiers séparés pour les deux interlocuteurs, et ont été déposés dans l'entrepôt de données Ortolang[3]. Ce sont sur ces fichiers que nous réalisons nos analyses linguistiques. Les Tableaux 2 et 3 donnent des exemples de discussions reconstruites à partir de ces transcriptions.

Le traitement des données de l'IRMf suit les procédures standard et a déjà été décrit en détails (***, ***). Nous utilisons une parcellisation cérébrale formée sur la base de données fonctionnelles et de connectivité cérébrale, de sorte que les régions d'intérêt représentent des volumes homogènes en termes de fonctions (Fan *et al.*, 2016). Dans chacune des 246 régions de l'atlas ainsi que dans un masque de l'hypothalamus développé par ailleurs (Wolfe *et al.*, 2015) l'estimation de l'activité est extraite utilisant la boîte à outils MarsBAR (Brett *et al.*, 2002) et l'ensemble de ces valeurs pour les 21 participants, 24 essais et 247 régions est utilisé pour les analyses statistiques.

## 3. Variables linguistiques

| Variables temporelles | Variables d'oralité | Variables de complexité |
|---|---|---|
| Temps de parole | Feedbacks | Complexité lexicale |
| Durée moyenne UIP | Pauses remplies | Complexité descriptive |
| | Marqueurs discursifs | Complexité syntaxique |

**Tableau 4.** *Liste des variables linguistiques étudiées*

### 3.1. *Description*

Nous calculons huit variables linguistiques, réparties en trois catégories et présentées dans le Tableau 4. L'objectif de ces métriques est double. Dans un premier temps, nous entendons décrire et caractériser les différences entre les productions des interlocuteurs humain et robot. Dans un second temps, nous souhaitons savoir si les participants adaptent leur production selon la production des interlocuteurs, ainsi que selon la nature de ces derniers.

Puisque les interactions enregistrées sont courtes, nous faisons le choix de travailler sur la minute entière pour chaque essai. Nous prenons en compte les temps

2. Version 1.9.9 :`www.sppas.org/` (Bigi, 2015).
3. anonyme

de pause à l'intérieur des minutes de conversation, et capturons ainsi la totalité des informations linguistiques pertinentes pour nos analyses.

3.1.1. *Variables temporelles*

Afin d'étudier d'abord la production linguistique des locuteurs dans son ensemble, nous commençons par extraire deux variables temporelles : le temps de parole des locuteurs et la durée moyenne des unités inter-pausales (UIPs) qu'ils produisent.

Le temps de parole, comme son nom l'indique, correspond à la durée de parole d'un locuteur sur l'essai, en secondes. Pour le calculer, nous sommons les durées de toutes les UIPs de l'individu au cours de l'interaction.

$$temps\,de\,parole = \sum_{i=1}^{N_{UIP}} duree\,UIP_i \quad [1]$$

Nous calculons ensuite la durée moyenne des interventions, en divisant le temps de parole du locuteur sur l'essai par le nombre $N_{UIP}$ d'UIPs produites.

$$duree\,moyenne\,UIP = \frac{1}{N_{UIP}} \sum_{i=1}^{N_{UIP}} duree\,UIP_i \quad [2]$$

3.1.2. *Variables d'oralité*

Les trois variables ensuite calculées sont des variables d'oralité. Nous réalisons en l'occurrence un travail sur les items lexicaux de feedback, sur les pauses remplies et sur les marqueurs discursifs présents dans notre corpus (voir Tableau 5 pour des exemples d'items lexicaux les plus fréquemment utilisés dans notre corpus).

Le feedback est un mécanisme linguistique qui permet aux locuteurs en interaction d'échanger des informations sur le processus de communication lui-même, c'est à dire sur la perception et la compréhension mutuelles (Allwood *et al.*, 1992 ; Bunt, 1994) et plus globalement sur la gestion du *Common Ground* (Clark *et al.*, 1983). Il est par exemple fréquent lors d'une interaction entre deux locuteurs que l'individu en position d'écoute produise de courts énoncés comme *"oui", "d'accord", "ok"* ou acquiesce d'un mouvement de tête, pour ratifier le discours de celui qui parle et pour signaler à ce dernier la compréhension de ses paroles. La variable que nous calculons réalise un ratio du nombre d'UIPs du locuteur commençant par un item lexical de feedback sur le nombre d'UIPs prononcés par le locuteur sur la minute.

$$ratio\,feedback = \frac{1}{N_{UIP}} \sum_{i=1}^{N_{UIP}} \chi_F(w_0^i) \quad [3]$$

où $w_0^i$ représente le premier mot de la $i^{eme}$ UIP et $\chi_Q$ est la fonction caractéristique de l'ensemble Q. Soit, pour l'ensemble F des marqueurs de feedback :
$\chi_F(w) = \begin{cases} 1\,si\,w\,est\,un\,marqueur\,de\,feedback\,(w \in F) \\ 0\,sinon \end{cases}$.

Les pauses remplies découlent des difficultés auxquelles se confrontent les locuteurs lorsqu'ils recherchent un mot, lorsqu'ils ont besoin de temps pour construire leur phrase ou encore quand ils ne savent pas quoi dire. En ce terme, ces éléments, que l'on peut qualifier de disfluences ou de feedbacks selon le contexte, perturbent le déroulement interactionnel et temporel du discours (Shriberg, 1994 ; Henry et Pallaud, 2003 ; Baiocchi, 2015). La variable que nous calculons réalise un ratio du nombre de marqueurs de pauses remplies (PR) produites par le locuteur par le nombre total $n$ de mots prononcés (W) par le locuteur sur la minute.

$$ratio\,pauses\,remplies = \frac{1}{n} \sum_{w \in W} \chi_{PR}(w) \quad [4]$$

La dernière analyse concerne les marqueurs discursifs. Aussi appelés connecteurs discursifs, ces unités linguistiques lient des propositions syntaxiques entre elles, permettant ainsi de marquer la relation entre les unités du discours (Schiffrin, 1987 ; Roze, 2009). Un locuteur qui souhaite produire un discours structuré va ainsi utiliser des connecteurs comme "mais", "parce que", "et". Pour calculer la variable, nous calculons le ratio du nombre de mots qui, parmi tous les mots prononcés par le locuteur sur la minute, sont des marqueurs discursifs (MD).

$$ratio\,marqueurs\,discursifs = \frac{1}{n} \sum_{w \in W} \chi_{MD}(w) \quad [5]$$

| | |
|---|---|
| Feedbacks | oui, ouais, non, ok, voilà, d'accord, mh, rire |
| Pauses remplies | euh, heu, mh, hum |
| Marqueurs discursifs | alors, mais, et, puis, enfin, parce que, parce qu', ensuite |

**Tableau 5.** *Liste non-exhaustive des marqueurs linguistiques de l'oral analysés dans notre corpus*

### 3.1.3. *Variables de complexité*

Pour finir, pour extraire les trois variables suivantes, nous utilisons l'outil d'enrichissement de données textuelles MarsaTag (Rauzy *et al.*, 2014). Entraîné sur des corpus français et oraux, l'outil nous permet de réaliser automatiquement la tokenisation, l'étiquetage morpho-syntaxique et la lemmatisation de l'ensemble des conversations. La préférence de MarsaTag sur Spacy, bibliothèque Python plus communément utilisée pour ces tâches, résulte de la comparaison que nous avons pu faire des performances de ces deux outils : en effet, Spacy n'étant pas entraîné sur de l'oral, l'étiquetage morpho-syntaxique, en particulier, est bien moins performant sur notre corpus. Le Tableau 6 présente les performances des deux outils, en se focalisant sur mots en "a" étiquetés "adjectif" par au moins un des deux outils. Sur l'intégralité du corpus, seuls 3 des 57 mots repérés par MarsaTag comme adjectifs sont mal étiquetés. En revanche, sur 53 tokens en "a" étiquetés adjectifs par Spacy, 15 d'entre eux n'en

sont pas. On remarque également que MarsaTag repère 4 adjectifs de plus que Spacy, et que parmi les 57 adjectifs repérés par MarsaTag, 17 n'ont pas été repérés par Spacy (contre 3 repérés par Spacy mais pas MarsaTag).

| Problème | MarsaTag | Spacy |
|---|---|---|
| **tokens en "a" faussement étiquetés en adjectif** | allo, aubergine, avengers | achète, agissais, ah, allais, allez, allo, amérique, annonces, apprécier, attends, au, auquel, aux, auxquels, avais |
| **adjectifs en "a" étiquetés par un seul des outils** | abandonnés, abîmés, acide, affaiblies, affaissée, aimée, allongé, ambigu, ambivalent, amusée, américaines, animé, arrondi, arrondis aseptisés, attachante, attaqué | abîmé, accroché, asiatique |

**Tableau 6.** *Problèmes rencontrés lors de l'étiquetage des tokens en "a" par les outils MarsaTag et Spacy. Les quatre listes de tokens sont exhaustives*

Cet étiquetage morpho-syntaxique nous permet de calculer nos variables de complexité. Estimant que le robot produit des énoncés linguistiques peu complexes lexicalement et syntaxiquement (voir Tableau 1), nous commençons par créer une variable de complexité lexicale qui correspond à la fraction du nombre de mots de contenu (noms + adjectifs + verbes, à l'exception des verbes auxiliaires, semi-auxiliaires et d'état) produits par le locuteur sur le nombre total $n$ de mots prononcés par le locuteur en une minute.

$$complexite\ lexicale = \frac{1}{n} \sum_{w \in W} \mathbb{1}_{adj}(w) + \mathbb{1}_{nom}(w) + \mathbb{1}_{vb\ act}(w) \qquad [6]$$

où $\mathbb{1}_{adj}(w) = \begin{cases} 1\ si\ w\ est\ un\ adjectif \\ 0\ sinon \end{cases}$ .

La complexité descriptive, dont la formule est tirée de Ochs *et al.* (2018), correspond au ratio du nombre d'adjectifs et d'adverbes prononcés sur le nombre total de mots $n$ prononcés par le locuteur pendant l'essai.

$$complexite\ descriptive = \frac{1}{n} \sum_{w \in W} \mathbb{1}_{adv}(w) + \mathbb{1}_{adj}(w) \qquad [7]$$

Également inspirée de Ochs *et al.* (2018), nous calculons enfin une variable de complexité syntaxique en divisant le nombre de pronoms, de prépositions et de

conjonctions produits par le locuteur par le nombre de total mots $n$ qu'il prononce au cours de l'interaction.

$$complexite\ syntaxique = \frac{1}{n} \sum_{w \in W} \mathbb{1}_{pron}(w) + \mathbb{1}_{prep}(w) + \mathbb{1}_{conj}(w) \quad [8]$$

### 3.2. *Comparaison des productions des deux interlocuteurs*

Un modèle linéaire mixte a été employé pour l'analyse de variance sur l'ensemble des variables linguistiques pour les productions de l'interlocuteur.

L'intérêt de cette analyse ne porte pas sur l'interlocuteur a proprement parler, car nous savons que les productions sont différentes ; elle permet néanmoins de caractériser et quantifier ces différences. Nous nous sommes surtout intéressés au facteur temps, représenté par la variable *Essai*, avec l'hypothèse que l'interlocuteur humain s'adapte au participant au fur et à mesure que se construit une familiarité, alors que le robot ne peut pas s'adapter.

Ces analyses ont été produites dans R (R Core Team, 2013) avec le paquet lme4. L'avantage du modèle linéaire mixte, comparativement au modèle linéaire classique, est la prise en compte de la variabilité liée à différents facteurs non contrôlés (par exemple les participants). La sélection de la structure du modèle a été estimé par le maximum de probabilité restreinte (Restricted maximum likelihood) donnant une solution intégrant la nature de l'interlocuteur *(Interlocuteur)*, des douze essais successifs *(Essai)*, et l'interaction d'intérêt *Interlocuteur* $\times$ *Essai* en effets fixes. À cela, une variable aléatoire prenant en compte la variabilité des participants selon les essais *(Essai|Participant)* a été introduite dans le modèle. Cette variable aléatoire permet de prendre en compte la variabilité inter-essais et inter-participants et donc de fournir une meilleure estimation des effets fixes.

$$variable\ interlocuteur \sim Interlocuteur * Essai + (1 + Essai|Participant) \quad [9]$$

Les résultats des analyses statistiques (tests de Fischer) sont donnés dans le Tableau 7. Comme attendu, l'Interlocuteur a un effet très significatif sur toutes les variables : à cause du contrôle par le système de Magicien d'Oz, l'humain et le robot ont des comportements linguistiques différents. Il est probable que des systèmes de contrôle plus élaborés donneraient des effets différents. Nos résultats sont spécifiques à cette implémentation du système de Magicien d'Oz. Pour les variables temporelles, l'humain parle plus longtemps, ce qui est en partie causé par le délai introduit par le Magicien d'Oz. Il produit aussi des UIPs plus longues en moyenne que le robot.

Pour les variables d'oralité, l'humain produit des ratios plus importants de pauses remplies, de feedbacks et de marqueurs discursifs que le robot. Pour les variables de complexité, le robot obtient des moyennes plus importantes que l'humain pour la

complexité descriptive et lexicale, mais il est moins élevé pour la complexité syntaxique. Ceci s'explique aussi par des particularités du système de Magicien d'Oz, en l'occurrence ses phrases scriptées se rapprochent du langage écrit et évitent donc les disfluences. Les disfluences de l'interlocuteur humain augmentent le nombre de mots comptabilisés au dénominateur de ces variables et diminuent la proportion de signifiants par rapport au nombre total de mots. En corollaire, cette oralité de l'humain implique des structures de phrases plus complexes comme l'indique l'augmentation de la complexité syntaxique. En conclusion, les résultats confirment les hypothèses : le robot n'ayant à disposition que des phrases simples et pas de marqueurs de l'oralité, alors que l'humain parle librement, il a un contenu plus riche, proportionnellement, en terme de lexique mais moins élaboré en terme de syntaxe.

On notera un effet très significatif du temps (facteur *Essai*) et de l'interaction *Interlocuteur* $\times$ *Essai* pour les marqueurs discursifs. La Figure 2 indique qu'ils augmentent pour l'humain et diminuent pour le robot au cours du temps. Pour les trois variables de complexité linguistique, le facteur *Essai* et l'interaction *Interlocuteur* $\times$ *Essai* sont significatifs. La complexité descriptive augmente au cours du temps pour le robot et diminue pour l'humain ($p$ <0.001, Figure 2). Des dynamiques différentes apparaissent pour les autres variables de complexité ($p$ <0.05), avec une diminution au cours du temps pour le robot mais pas l'humain pour la complexité lexicale et une augmentation au court du temps pour l'humain mais pas le robot pour la complexité syntaxique. Aucun autre effet n'impliquant le facteur *Essai* n'est significatif pour les deux autres variables d'oralité, feedbacks et pauses remplies, et pour les variables temporelles. Ces résultats avec le facteur *Essai* impliquent que ces variables de complexité linguistique évoluent au cours du temps différemment selon l'interlocuteur. Une observation des expérimentateurs lors du recueil du corpus pourrait permettre de proposer une explication pour ces résultats. En effet, et sans que cela soit formalisé, en début d'expérience, les conversations concernent surtout la description des images, tandis qu'à la fin, elles portent plus sur le message de la campagne de publicité. En effet, au début les participants découvrent les images et décrivent ce qu'ils perçoivent avec l'interlocuteur, tandis qu'à la fin ils tentent d'accomplir l'objectif qui leur a été donné, à savoir trouver le message de la campagne publicitaire.

Pour l'humain, l'augmentation des marqueurs discursifs et de la complexité syntaxique et la diminution de la complexité descriptive peuvent s'expliquer par cette transition d'une phase descriptive vers une phase argumentative. Les effets opposés pour le robot (diminution des marqueurs discursifs et de la complexité lexicale, et augmentation de la complexité descriptive) s'expliquent par les phrases pré-enregistrées du système de Magicien d'Oz.

### 3.3. *Analyses des effets du comportement des interlocuteurs sur le comportement des participants*

Avec ces analyses en modèle linéaire mixte, dont les résultats (tests de Student) sont donnés Tableau 8, nous nous interrogeons sur les effets que les variables me-

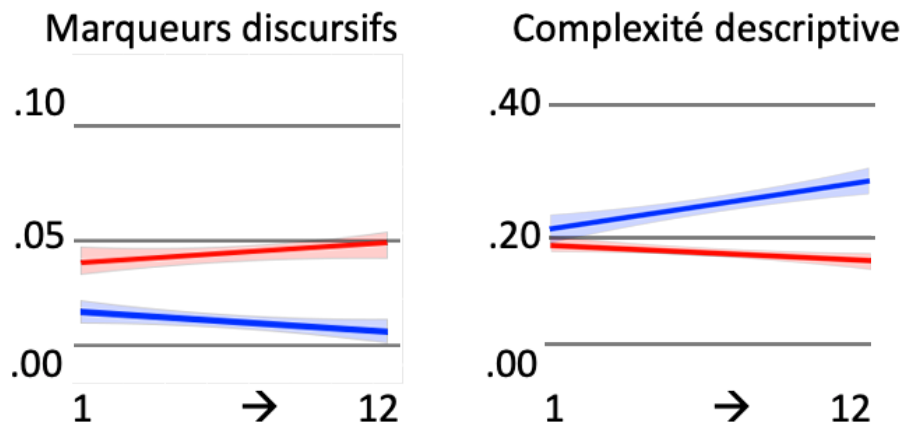


**Figure 2.** *Évolution des ratio au cours des Essais (interlocuteur humain en rouge et robot en bleu)*

| | **Interlocuteur** | | **Essai** | | **Interlocuteur × Essai** | |
|---|---|---|---|---|---|---|
| **Variable** | **F** | ***p*** | **F** | ***p*** | **F** | ***p*** |
| Temps de parole | 705.537 | **0.000** | 0.965 | 0.478 | 0.493 | 0.908 |
| Durée moyenne des UIPs | 217.297 | **0.000** | 1.552 | 0.110 | 1.354 | 0.192 |
| Feedbacks | 114.994 | **0.000** | 0.794 | 0.646 | 0.712 | 0.728 |
| Pauses remplies | 618.650 | **0.000** | 1.057 | 0.395 | 1.800 | 0.051 |
| Marqueurs discursifs | 211.748 | **0.000** | 3.431 | **0.000** | 6.533 | **0.000** |
| Complexité lexical | 124.758 | **0.000** | 2.164 | *0.015* | 1.900 | *0.037* |
| Complexité descriptive | 136.494 | **0.000** | 4.820 | **0.000** | 3.434 | **0.000** |
| Complexité syntactique | 148.811 | **0.000** | 2.344 | *0.008* | 2.332 | *0.008* |

**Tableau 7.** *Résultats des analyses statistiques (resultat du test de Fisher avec la formule [9]) sur les variables linguistiques pour l'interlocuteur, en gras les effets significatifs à p <0.001 et en italique à p <0.05*

surées pour l'interlocuteur ont sur les mêmes variables mesurées chez le participant, éventuellement en interaction avec la nature de l'interlocuteur. Les résultats doivent se comprendre comme il suit : un effet significatif de la variable de l'interlocuteur sur la variable du participant s'interprète comme l'influence automatique du comportement de l'interlocuteur sur le comportement du participant, alors qu'un effet significatif du facteur *Interlocuteur* s'interprète comme la conséquence de la différence de posture intentionnelle adoptée par le participant en fonction de la nature, humain ou robot, de l'interlocuteur. Une interaction significative signale que les effets "ascendants",

ceux de la variable de l'interlocuteur, sont différents selon la nature de l'interlocuteur, c'est-à-dire qu'ils dépendent d'un contrôle "descendant". Le facteur *Essai* a aussi été inclus dans le modèle (sans interaction avec les autres facteurs) mais ses résultats ne sont pas indiqués car pas discutés ultérieurement, ainsi que la variable aléatoire *(Essai|Participant)* décrites précédemment. La formule complète utilisée dans R pour ces analyses est :

$$variable\ participant \sim$$
$$variable\ interlocuteur * Interlocuteur + Essai + (1 + Essai|Participant) \quad [10]$$

| | **Var. Inter$^{eur}$** | | **Interlocuteur** | | **Inter. × Var. Inter$^{eur}$** | |
|---|---|---|---|---|---|---|
| **Var. Part$^{nt}$** | **t** | ***p*** | **t** | ***p*** | **t** | ***p*** |
| Temps de parole | -13.184 | **0.000** | -4.708 | **0.000** | -3.633 | **0.000** |
| Durée moyenne UIP | -1.558 | 0.119 | -2.556 | *0.011* | 0.770 | 0.441 |
| Feedbacks | -1.838 | 0.066 | -4.675 | **0.000** | -0.013 | 0.990 |
| Pauses remplies | 0.738 | 0.461 | 2.505 | *0.012* | -0.450 | 0.653 |
| Marqueurs discursifs | -0.491 | 0.623 | 0.656 | 0.512 | 0.357 | 0.721 |
| Complexité lexicale | 1.126 | 0.260 | 1.238 | 0.216 | -0.895 | 0.371 |
| Complexité descriptive | 2.357 | *0.018* | 0.598 | 0.550 | -0.594 | 0.552 |
| Complexité syntaxique | 2.285 | *0.022* | 1.881 | 0.060 | -1.581 | 0.114 |

**Tableau 8.** *Résultats des analyses statistiques (test de Student) sur les variables linguistiques du participant, en gras les effets significatifs à p <0.001 et en italique à p <0.05*

### 3.3.1. *Variables temporelles*

Nous observons un effet significatif de la variable de l'interlocuteur, du facteur *Interlocuteur* et de l'interaction entre les deux facteurs sur le temps de parole des participants. Cela signifie que le temps de parole des participants est influencé par le temps de parole des interlocuteurs et par leur nature. On observe en l'occurrence que plus l'interlocuteur parle, moins le participant parle (et inversement), ce qui s'explique de manière triviale par le fait que les locuteurs se partagent un temps limité d'une minute. Cette anticorrélation est plus faible pour le robot que pour l'humain, ce qui peut s'expliquer soit par des particularités du système de Magicien d'Oz (le robot présente des latences entre la sélection et la production de la réponse, cela conduit à des délais plus longs entre les prises de parole pour l'interlocuteur robot), soit par une différence de posture intentionnelle envers le robot qui réduit la propension à discuter. Il est possible que les deux effets interviennent. Concernant la durée moyenne des

UIPs, nous n'observons qu'un effet de l'Interlocuteur : les participants produisent des UIPs plus longues avec l'humain plutôt qu'avec le robot, ce qui va dans le sens de l'adoption d'une posture différente selon la nature de l'interlocuteur avec qui ils discutent, en l'occurrence en simplifiant leur discours avec l'agent artificiel.

#### 3.3.2. *Variables d'oralité*

Les seuls effets observables pour ces variables sont un effet de l'Interlocuteur pour les feedbacks et pour les pauses remplies, indiquant ici aussi que les sujets adaptent leur comportement selon la nature de leur interlocuteur. Les t-values montrent que les participants produisent plus de feedbacks avec l'humain qu'avec le robot, et plus de pauses remplies avec le robot qu'avec l'humain. Nous n'observons pas d'effet de la variable de l'interlocuteur sur la variable du participant, et pas non plus d'effet de l'interaction. La production de marqueurs linguistiques du participant n'est donc pas influencée par la production de feedbacks, de pauses remplies et de marqueurs discursifs de l'interlocuteur. D'un point de vue linguistique, ces résultats ne sont pas étonnants :

– un locuteur ne produit pas des feedbacks parce que son interlocuteur en produit, il produit des feedbacks pour signifier que l'énoncé de l'interlocuteur a été perçu et compris,

– il ne produit pas des pauses remplies parce que l'interlocuteur en produit, mais parce que le discours de l'interlocuteur est décousu et imprévisible (ou simplement parce qu'il cherche ses mots mais souhaite faire comprendre qu'il ne veut pas perdre le tour de parole),

– et un locuteur ne produit pas des marqueurs discursifs parce que l'interlocuteur en produit, il produit des marqueurs discursifs pour établir des relations entre les parties de son discours afin que sa parole ait du sens.

#### 3.3.3. *Variables de complexité*

Nous observons un effet de la variable des interlocuteurs pour la complexité syntaxique et la complexité descriptive, mais pas d'effet de l'Interlocuteur ou d'interaction. Cela indique que la quantité de connecteurs structurants (complexité syntaxique) et la quantité d'adjectifs et d'adverbes (complexité descriptive) des participants et des interlocuteurs sont corrélées indépendamment de la nature des interlocuteurs, en faveur d'une interprétation de type convergence ou alignement entre interlocuteurs. Pour la complexité lexicale, nous n'observons aucun effet de la variable de l'interlocuteur, du facteur *Interlocuteur*, ou de l'interaction. Les participants n'adaptent pas la quantité de mots de contenu qu'ils prononcent en fonction de la quantité de mots de contenu prononcés par les interlocuteurs, et produisent autant de mots de contenu avec l'interlocuteur naturel qu'avec l'interlocuteur artificiel. Ainsi, contrairement à nos hypothèses, il n'y aurait pas d'influence de l'interlocuteur sur le participant quant au ratio de mots de contenu prononcés. Cependant, et bien que ces résultats n'indiquent pas de corrélation entre la quantité de mots de contenu prononcés par les participants et par les interlocuteurs, nous décidons de nous focaliser sur lexique en présentant dans

la section suivante une variable d'alignement lexical. Cette variable nous permet d'explorer plus en détail le vocabulaire des locuteurs, en s'interrogeant spécifiquement sur les mots de contenus employés par un locuteur et repris par l'autre.

## 4. Variable d'alignement

### 4.1. *Description*

Puisque nous nous intéressons à l'alignement conversationnel présent dans notre corpus, et en particulier à l'alignement lexical, nous travaillons sur une variable d'alignement lexical à partir de laquelle nous étudions le lexique commun entre les locuteurs. Nous nous inspirons de la formule LILLA *(lexical indiscriminate local linguistic alignment)* proposée par Fusaroli *et al.* (2012) et reprise par Xu et Reitter (2015). De manière à calculer l'alignement lexical entre les locuteurs en interaction, cette formule repose sur le principe de l'effet d'amorçage, un principe décrit par Pickering et Garrod (2004) dans leur modèle d'alignement interactif selon lequel la production langagière des individus est directement influencée par les stimulations auditives auxquels ils sont exposés. Concrètement, cela signifie que sur le plan lexical, les participants devraient tendre à utiliser le vocabulaire des interlocuteurs après que ces derniers aient introduits de nouveaux mots dans la conversation. LILLA étant à l'origine une mesure normalisée du nombre de mots qui apparaissent dans un texte (*prime*) et qui sont repris dans une réponse (*cible*), la formule ne correspond pas exactement à notre situation. D'une part, nous souhaitons étudier l'alignement lexical des participants sur les interlocuteurs mais également l'alignement lexical des interlocuteurs sur les participants. Les locuteurs alternant les prises de parole au cours du dialogue, il se peut qu'un mot présent dans la conversation du *prime* ait été introduit précédemment par une *cible*. Dans ce cas-là, nous adaptons alors le vocabulaire utilisé comme référence pour le *prime* (P) afin ne pas prendre en compte les mots précédemment introduits par la *cible* (T). D'autre part, nous ne nous intéressons qu'aux mots de contenu (w) plutôt qu'à tout le vocabulaire, et nous utilisons ainsi les mêmes mots que ceux analysés dans la variable de complexité lexicale.

En incluant ces réflexions, on obtient une formule comptabilisant, sur toutes les interventions d'un *prime*, le nombre des mots qu'il introduit dans la conversation et qui sont réutilisés par la *cible* dans la suite de l'échange. Afin d'obtenir une mesure entre 0 et 1, ce chiffre est normalisé par le produit du nombre de mots distincts prononcés respectivement par *cible* et *prime* (ne sont pas considérés les mots introduits par *cible* et répétés par *prime*). On peut donc formaliser le calcul de LILLA de manière mathématique :

$$LILLA(T,P) = \frac{\sum_{i=0}^{L} \sum_{w \in P_i \setminus \bigcup_{j>i} T_j} \delta(w, \bigcup_{j>i} T_j)}{\#(\bigcup_{i=0}^{L}(P_i \setminus \bigcup_{j<i} T_j)) * \#(\bigcup_{j=0}^{L} T_j)} \qquad [11]$$

avec $\delta(w, X) = \begin{cases} 1\ si\ w \in X \\ 0\ sinon \end{cases}$ .

$P_i$ (respectivement $T_j$) désigne l'ensemble des mots de la $i^{eme}$ intervention du *prime* (respectivement *cible*). Si $i$ n'est pas une intervention du prime, alors $P_i = \{\}$. $\bigcup T_j$ permet de regarder l'ensemble des mots prononcés par un participant, soit sur toute la conversation ($\bigcup_{j=0}^{L} T_j$), soit de manière plus ciblée, en ne regardant que les mots prononcés avant ($\bigcup_{j<i} T_j$) une intervention i donnée, soit prononcés après ($\bigcup_{j>i} T_j$). Le tour de parole 0 désignant celui d'un participant (qu'il soit ou non pris) ; $L$ dénote le dernier tour de parole. Pour permettre le calcul, les UIPs consécutives d'un même locuteur doivent être concaténées pour n'en faire qu'une (voir Tableau 9 pour un exemple). Après cette transformation, les tours pairs désignent ceux du participant, ceux impairs ceux de l'interlocuteur. L'alignement du participant sur l'interlocuteur est donc obtenu en ne considérant que les $j$ pair et $i$ impair ; l'alignement de l'interlocuteur sur le participant est obtenu pour $i$ pair et $j$ impair.

En accord avec les travaux de Brennan (1996) et Branigan *et al.* (2011) selon lesquels l'alignement est plus fort lorsque l'on s'adresse à un robot, notre hypothèse initiale est celle d'un alignement lexical plus fort du participant sur l'interlocuteur robot que sur l'interlocuteur humain. Ensuite, bien que la conversation avec le robot soit bidirectionnelle, elle n'est dans les faits qu'unidirectionnelle pour l'alignement lexical. En effet, le robot ayant un discours pré-enregistré et un lexique fini, il peut converser avec le participant mais il ne peut pas aligner son lexique si les mots de contenu employés par le participant ne font pas parti du vocabulaire du système de Magicien d'Oz. Ainsi, dans la direction d'alignement lexical de l'interlocuteur sur le participant, l'hypothèse est celle d'un alignement lexical plus fort de l'interlocuteur humain que de l'interlocuteur robot sur le participant.

### 4.2. *Analyse statistique*

Le Tableau 10 indique les résultats des analyses statistiques de LILLA selon l'équation 9. Sur la première ligne, le *prime* est l'interlocuteur et la *cible* le participant, c'est-à-dire que la variable mesure l'alignement lexical du participant sur l'interlocuteur. On observe un effet significatif du facteur *Interlocuteur*, et la valeur positive du test t indique qu'elle est plus grande lorsque l'interlocuteur est le robot. Comme attendu, le participant s'aligne donc plus sur le lexique du robot que sur le lexique de l'humain.

La seconde ligne indique les résultats lorsque le *prime* est le participant et la *cible* l'interlocuteur, c'est-à-dire l'alignement lexical de l'interlocuteur sur le participant. On observe aussi un effet significatif du facteur *Interlocuteur*, et la valeur positive du test t indique qu'elle est plus grande lorsque l'interlocuteur est le robot que l'humain. Contrairement à nos hypothèses, l'interlocuteur robot s'aligne plus sur le lexique du participant que l'interlocuteur humain.

| |
|---|
| Participant : euh c'est un **citron vert** |
| Interlocuteur : ouais et un **citron vert** avec aussi un un masque dég- euh dé- découpé euh sur le sur le zeste là en enlevant le le zeste autour des yeux |
| Participant : euh moi j'ai pas compris |
| Interlocuteur : j'ai l'impression qu'il y avait encore deux yeux |
| Participant : c'est le c'est genre **Tortue Ninja** ou quoi cette fois-ci |
| Interlocuteur : ouais et ben ouais c'est ce que j'allais dire c'est ce que j'allais dire c'était le bandeau là ils ont pas découpé les yeux ils ont découpé un truc autour des yeux |
| Participant : ouais |
| Interlocuteur : comme un bandeau |
| Participant : euh ça fait **Tortue Ninja** |
| Interlocuteur : donc après les **Tortues Ninja** donc l'aubergine Batman et le et le citron **Tortue Ninja** euh ça fait un truc euh *légumes* et super héros |
| Participant : c'est ça *légumes* et fruits en effet donc plus pour les **enfants** obligé |
| Interlocuteur : ah ouais et pour les **enfants** ouais ouais ça ça me paraît pour les **enfants** ça brille bien |
| Participant : ou pour les grands **enfants** ouais |
| Interlocuteur : ben ouais c'est vrai que ça c'est un peu intergénérationnel parce que les **Tortue Ninja** euh |

**Tableau 9.** *Transcription condensée d'un échange pour faciliter le calcul de LILLA : les interventions participant / interlocuteur sont alternées. Sont surlignés en gras les mots de contenu introduits par le participant et répétés, en italique ceux introduit par l'interlocuteur et répété. Au total, l'algorithme compte 17 mots de contenu introduits par l'interlocuteur parmi lesquels un seul token est repris par le participant ; dix mots de contenu sont introduits par le participant, cinq de ces tokens sont repris par l'interlocuteur. L'alignement LILLA du participant sur l'interlocuteur donne donc 0,00588 (assez faible), tandis que celui de l'interlocuteur sur le participant est 0,02066.*

| | **Interlocuteur** | | **Essai** | | **Interlocuteur × Essai** | |
|---|---|---|---|---|---|---|
| **Prime** | **t** | ***p*** | **t** | ***p*** | **t** | ***p*** |
| Interlocuteur | 3.661 | **0.000** | -1.846 | 0.065 | 1.400 | 0.162 |
| Participant | 2.835 | *0.005* | -2.097 | *0.036* | -1.679 | 0.093 |

**Tableau 10.** *Résultats de l'analyse statistique sur LILLA, en gras les effets significatifs à p <0.001, en italique les effets significatifs à p <0.05*

## 5. Variable neurophysiologique

L'objectif principal de cet article est l'application des outils d'analyse de traitement automatique à la description de notre corpus. Ce traitement automatique produit des

variables numériques qui caractérisent différents aspects de chaque conversation. Elles peuvent être utilisées pour identifier leurs corrélats cérébraux. En pratique, le cerveau a été parcellisé en 247 régions, dont nous avons extrait l'activité moyenne pour chaque conversation. La formule suivante est utilisée pour identifier les régions dont l'activité est modulée par l'interlocuteur, corrélée à la variable, ou significativement associée à l'interaction entre la variable et l'interlocuteur :

$$région_n \sim variable * Interlocuteur + Essai + (1 + Essai|Participant) \quad [12]$$

À noter que comme précédemment, la variable *Essai* est introduite pour capturer une éventuelle évolution temporelle du signal, mais ne sera pas décrite dans les résultats. L'important est d'identifier les corrélats cérébraux de variables comportementales et, éventuellement, comment ils sont affectées par la nature de l'interlocuteur (terme d'interaction entre variable comportementale et *Interlocuteur*).[4]

### 5.1. *Analyse du temps de parole*

L'analyse du temps de parole du participant et de l'interlocuteur correspond, respectivement, à la quantité de parole produite et à la quantité de parole perçue par le cerveau.L'objectif de cette première analyse est de valider l'approche utilisée avec des variables dont les corrélats cérébraux sont bien connus. On s'attend donc à une corrélation avec les systèmes de production langagière pour le premier (cortex moteur et gyrus frontal inférieur gauche) et de perception pour le second (lobe temporal).

Les résultats sont donnés sur la Figure 3 pour le temps de parole du participant et sur la Figure 4 pour le temps de parole de l'interlocuteur. La couleur correspond à la direction des corrélations, négatives en bleu et positives en rouge. On remarque donc la prédominance du lobe temporal (centré sur les gyri temporaux moyens et supérieurs dans les deux hémisphères), la principale région du cerveau humain pour la compréhension du langage, qui corrèle négativement avec le temps du parole du participant (en bleu dans 3) et positivement avec celui de l'interlocuteur (en rouge dans 4). Plusieurs commentaires peuvent être faits, d'une part que les parties ventrales des lobes temporaux, impliquées dans le traitement visuel, ne sont pas identifiées dans ces analyses, et d'autre part qu'on observe une latéralisation à gauche dans les parties les plus postérieures (jonction temporopariétale) qui correspond à l'aire de Wernicke qui joue un rôle majeur dans la compréhension du langage.

4. À noter que l'effet principal du terme *Interlocuteur* n'est pas directement intéressant en soi et difficile à interpréter pour la raison suivante : étant donné que quelle que soit la variable comportementale, l'activité de chaque région et l'étiquetage humain / robot de chaque essai restant identique, nous devrions obtenir toujours le même résultat pour ce facteur. Or, ce n'est pas le cas. En effet, une partie de la différence de réponse cérébrale entre interlocuteurs humain et robot s'explique par des différences de comportements entre ces deux interlocuteurs. Ainsi, si la variable décrit un comportement très différent entre humain et robot, elle capturera une partie des différences attendues pour le facteur *Interlocuteur*, rendant difficile l'interprétation des résultats de ce facteur.

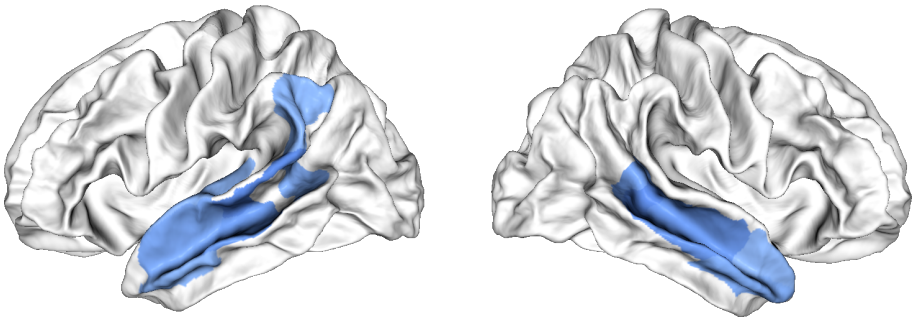

**Figure 3.** *Régions corticales corrélées avec le temps de parole total du participant*

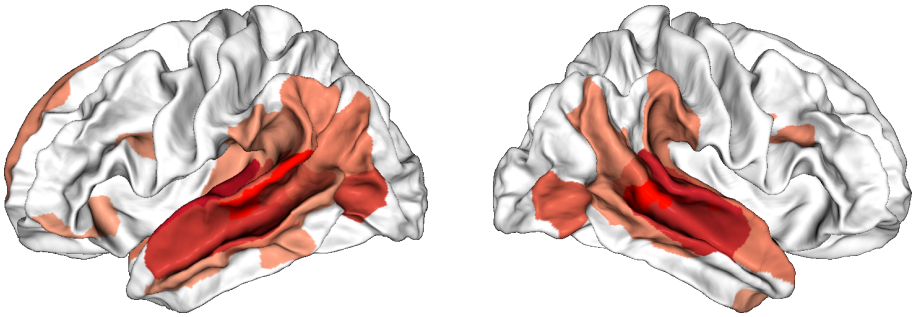

**Figure 4.** *Régions corticales corrélées avec le temps de parole total de l'interlocuteur*

Un autre résultat est la corrélation entre le temps de parole de l'interlocuteur et l'activité dans l'aire de Broca, au fond du sillon frontal inférieur dans les deux hémisphères. Ce résultat va dans le sens de l'implication de l'aire de Broca, connue pour son rôle dans la production verbale, dans la perception du langage. Par contre, il est plus surprenant de ne pas trouver de régions impliquées dans le production du langage qui corrèlent avec le temps de parole du participant, et bien que ce soit spéculatif, on propose une saturation du signal cortical associé à cette activité très demandeuse de ressources neuronales, si bien qu'il ne corrèle plus avec la quantité de langage produit.

### 5.2. *Analyse LILLA*

Les résultats précédents valident l'approche utilisé pour identifier les corrélats cérébraux des variables issues des analyses précédentes. Nous l'avons donc utilisée pour valider la variable développée dans le cadre de cette étude, LILLA. L'identification de corrélats cérébraux associés à cette variable suggère qu'elle représente un aspect per-

tinent du comportement langagier. Une région du gyrus parahippocampique gauche, mais surtout trois régions adjacentes au niveau du gyrus cingulaire central gauche sont identifiées comme ayant une activité significativement négativement corrélée à la variable LILLA. À noter que l'interaction *LILLA* $\times$ *Interlocuteur* n'identifie aucune région au seuil utilisé. Ces régions du système limbique ne sont pas connues pour leur implication dans le langage, mais la continuité anatomique implique donc que leur corrélation significative conjointe représente une implication fonctionnelle de cette région du cerveau. L'hippocampe est surtout impliqué dans les processus mnésiques, le gyrus cingulaire central est impliqué dans le contrôle des actions. L'augmentation de LILLA signifiant une utilisation plus importante des mots de l'interlocuteur, la corrélation négative avec l'activité dans ces régions suggère une utilisation réduite des ressources cognitives propres des participants (contrôle des actions et mémoire) lorsque les mots utilisés sont fournis par les interlocuteurs.

## 6. Discussion

Cet article présente l'étude, avec des outils de Traitement Automatique du Langage, d'un corpus unique combinant des conversations comparables avec un humain ou un robot, synchronisées avec l'enregistrement de l'activité cérébrale en IRM fonctionnelle. Trois séries de résultats ont été décrits, à savoir, (i) caractériser les différences, connues, entre les productions linguistiques des interlocuteurs humains et robots, ainsi que leur modification au cours du temps, (ii) mettre en évidence des relations entre les productions des participants et des interlocuteurs en utilisant justement le fait que les deux interlocuteurs ont des comportements, et donc des influences, différentes, et (iii) développer une variable d'alignement lexical pour ce corpus et valider sa pertinence en étudiant ses corrélats cérébraux.

Pour le premier point, il s'agit de vérifier et de quantifier des différences connues dans les productions de l'interlocuteur humain et robot, ainsi que leur éventuelle évolution au cours du temps (essais successifs). Il est important de noter que si ces différences trouvent leur origine dans des considérations techniques, elles ne sont pas pour autant des défauts de l'expérience à proprement parler. En montrant les capacités limitées du robot, elles servent à conforter la croyance du sujet dans son autonomie, alors qu'il est contrôlé par l'humain avec lequel ou laquelle ils ou elles interagissent vraiment. Les différences se retrouvent à différents niveaux : le robot parle moins et produit des UIPs plus courtes que l'humain. Le robot a moins de marqueurs d'oralité tels que les pauses remplies, les feedbacks et les marqueurs discursifs que l'humain, ce qui souligne le manque de spontanéité de son discours. Quant aux marqueurs de complexité, il est significatif que leur évolution au cours du temps amplifie les différences observées dans l'effet principal. Concernant le lexique (complexité lexicale et descriptive), les réponses scriptées du robot suppriment les disfluences. La même raison explique une complexité syntaxique plus importante pour l'humain, qui peut produire à volonté des structures grammaticales emboîtées complexes, que le robot n'a pas à disposition, et son augmentation au cours du temps pour l'humain peut être associée

au passage d'une conversation décrivant les images à une conversation argumentant sur le message de la campagne de publicité.

Le deuxième objectif de ces analyses est de caractériser les relations entre participants et interlocuteurs. En particulier, nous voulons utiliser le fait que les interlocuteurs ont des productions différentes pour distinguer les effets automatiques de convergences de ceux liés à la nature, artificiel ou humain, de l'interlocuteur. Nous nous plaçons dans le cadre théorique de la posture intentionnelle du philosophe Dennett (1987), qui postule que nous modifions notre comportement en fonction de la posture intentionnelle que nous adoptons selon la nature de l'interlocuteur. Les modèles linéaires évaluent quel facteur, soit la production, soit la nature de l'interlocuteur, explique la production du participant. Ils permettent aussi de mettre en évidence d'éventuelles interactions entre les deux. Après avoir vérifié la validité de l'approche avec une variable aux résultats connus, le temps de parole, les variables d'oralité et de complexité sont analysées à leur tour. Pour les variable d'oralité, c'est le facteur *Interlocuteur* qui a des effets significatifs (sur le ratio de feedbacks et de pauses remplies), indiquant que les participants adaptent leur comportement oral quand ils interagissent avec un robot, en accord avec le cadre théorique de la posture intentionnelle. Cette adaptation prend la forme d'une réduction des feedbacks et d'une augmentation des pauses remplies avec la machine. Mais la différence peut aussi s'expliquer par une augmentation des disfluences chez les participants dues aux limites des productions verbales du robot. Pour les complexités syntaxiques et descriptives, on a au contraire un effet de la production de l'interlocuteur et pas d'effet du facteur *Interlocuteur*. Ceci indique que pour ces variables, les alignements entre les interlocuteurs se font localement, au niveau de l'essai, et ne dépendent pas de la nature intentionnelle de l'interlocuteur.

L'absence de corrélation entre la quantité de mots signifiants prononcés par le participant et par l'interlocuteur nous a conduit au troisième objectif, développer une nouvelle variable basée sur des études antérieures d'alignement lexical. La variable LILLA indique qu'il y a une différence significative d'alignement entre le participant et l'interlocuteur, avec un alignement plus important du participant sur le robot que sur l'humain. Autrement dit, le participant emploie plus de mots signifiants introduits par le robot que par l'humain. En accord avec la littérature existante, cela confirme que les humains s'alignent davantage avec les interlocuteurs artificiels sur le plan lexical pour prendre en compte leurs capacités limitées : utiliser le même mot que le robot permet de s'assurer qu'il connaît ce mot et est donc capable de le comprendre, une contrainte pratiquement inversée avec l'humain, où des champs sémantiques permettent d'élargir le socle lexical commun.

La variable LILLA met également en avant une différence significative d'alignement entre l'interlocuteur et le participant, avec un alignement plus important du robot sur le participant. Alors que l'alignement lexical a été décrit par la littérature comme un phénomène robuste dans les interactions humaines, ces résultats vont à l'encontre de notre hypothèse. En effet, bien que l'interlocuteur humain a une parole libre tout au long de l'expérience, au contraire du robot qui est limité par les phrases

pré-enregistrées dans le système de Magicien d'Oz, ce premier a repris significativement moins de vocabulaire du participant que ne l'a fait l'interlocuteur artificiel. Nous pensons qu'il peut s'agir d'un artefact sous la forme d'un effet de report entre les essais avec l'interlocuteur robot. En effet, le résultat précédant suggère que le participant aligne son lexique sur celui du robot, il se peut que le lexique acquis du robot au cours d'un essai soit utilisé dans un essai suivant avec le robot. S'il s'agit effectivement d'un mot du robot mais qu'il est prononcé pour la première fois par le participant dans un nouvel essai et repris par le robot, on observe effectivement un augmentation de la mesure d'alignement du robot sur l'humain, mais qui ne représente pas un alignement au cours de l'essai, mais plutôt un alignement au cours de l'expérience dans la direction attendue, c'est-à-dire de l'humain vers le robot. Cette possibilité nécessite de développer une nouvelle approche pour prendre en compte les effets d'alignement lexical au cours des différents essais. Enfin, nous avons utilisé les données d'activité cérébrale pour valider la pertinence de la variable d'alignement lexical. Nous avons d'abord étudié les corrélats du temps de parole pour valider de l'approche utilisée. Nous avons identifié des régions du cortex limbique gauche négativement corrélée avec LILLA. Alors que ces régions sont impliquées dans des processus mnésiques et dans le contrôle de l'action, nous proposons que plus le participant s'appuie sur le vocabulaire introduit par l'interlocuteur, moins il utilise ses ressources propres pour choisir les mots qui seront utilisés dans la discussion : il s'agit d'une réduction du contrôle cognitif lorsqu'on s'aligne sur le lexique de l'interlocuteur, indépendamment de la nature de l'interlocuteur. Ces résultats suggèrent que la variable d'alignement lexical LILLA est pertinente pour caractériser un aspect des comportements langagiers.

## 7. Conclusion

Dans cet article nous décrivons l'analyse d'un corpus d'interaction de participants humains qui discutent avec un humain ou avec un robot. Cette analyse nous a permis de vérifier et de quantifier les différences de comportement verbal entre les interlocuteurs humain et robot. Nous avons aussi étudié les relations en terme d'alignement entre les productions du participant et celles des interlocuteurs (humain ou robot) afin de dissocier les phénomènes automatiques de type alignement d'autres phénomènes liés à la nature humaine ou robotique de l'interlocuteur. Enfin, l'analyse des données neurophysiologiques suggère que cette variable d'alignement décrit un vrai phénomène cognitif.

## 8. Bibliographie

***, « –Keeping Submission Anonymous– », ***.

Al Moubayed S., Beskow J., Skantze G., Granström B., « Furhat : A Back-Projected Human-Like Robot Head for Multiparty Human-Machine Interaction », *in* A. e. a. Esposito (ed.),

*Cognitive Behavioural Systems*, Lecture Notes in Computer Science, Springer Berlin Heidelberg, p. 114-130, 2012.

Allwood J., Nivre J., Ahlsén E., « On the semantics and pragmatics of linguistic feedback », *Journal of semantics*, vol. 9, no 1, p. 1-26, 1992.

Baiocchi L., Pauses remplies en interaction, mémoire de Master Théorie Linguistiques : terrain et expérimentation, Master thesis, Université d'Aix-Marseille, Aix-en-Provence, 2015.

Bigi B., « SPPAS - Multi-lingual approaches to the automatic annotation of speech », *The Phonetician*, vol. 111-112, p. 54-69, 2015.

Blache P., Bertrand R., Ferré G., *Creating and Exploiting Multimodal Annotated Corpora : The ToMA Project*, Springer-Verlag, Berlin, Heidelberg, p. 38–53, 2009.

Boersma P., « Praat, a system for doing phonetics by computer », *Glot. Int.*, vol. 5, no 9, p. 341-345, 2001.

Branigan H. P., Pickering M. J., Cleland A. A., « Syntactic co-ordination in dialogue », *Cognition*, vol. 75, no 2, p. B13-B25, May, 2000.

Branigan H. P., Pickering M. J., McLean J. F., Cleland A. A., « Syntactic alignment and participant role in dialogue », *Cognition*, vol. 104, no 2, p. 163-197, August, 2007.

Branigan H. P., Pickering M. J., Pearson J., McLean J. F., Brown A., « The role of beliefs in lexical alignment : Evidence from dialogs with humans and computers », *Cognition*, vol. 121, no 1, p. 41-57, October, 2011.

Brennan S. E., « Lexical entrainment in spontaneous dialog », *Proceedings of ISSD 96*, 1996.

Brett M., Anton J., Valabrgue R., Poline J.-B., « Region of interest analysis using an SPM toolbox. Presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2-6, 2002, Sendai, Japan », *Neuroimage*, vol. 13, p. 210-217, 01, 2002.

Bunt H., « Context and dialogue control », *Think Quarterly*, vol. 3, no 1, p. 19-31, 1994.

Clark H. H., Schreuder R., Buttrick S., « Common ground at the understanding of demonstrative reference », *Journal of verbal learning and verbal behavior*, vol. 22, no 2, p. 245-258, 1983.

Dennett D. C., *The intentional stance*, MIT Press, Cambridge, Mass, 1987.

Fan L., Li H., Zhuo J., Zhang Y., Wang J., Chen L., Yang Z., Chu C., Xie S., Laird A. R., Fox P. T., Eickhoff S. B., Yu C., Jiang T., « The Human Brainnetome Atlas : A New Brain Atlas Based on Connectional Architecture », *Cerebral Cortex*, vol. 26, no 8, p. 3508-3526, May, 2016.

Fusaroli R., Bahrami B., Olsen K., Roepstorff A., Rees G., Frith C., Tylén K., « Coming to Terms », *Psychological Science*, vol. 23, no 8, p. 931-939, July, 2012.

Henry S., Pallaud B., « Word fragments and repeats in spontaneous spoken French », *ISCA Tutorial and Research Workshop on Disfluency in Spontaneous Speech*, 2003.

Ochs M., Jain S., Blache P., « Toward an automatic prediction of the sense of presence in virtual reality environment », *Proceedings of the 6th International Conference on Human-Agent Interaction*, ACM, p. 161-166, 2018.

Pickering M. J., Garrod S., « The interactive-alignment model : Developments and refinements », *Behavioral and Brain Sciences*, April, 2004.

Price C. J., « The anatomy of language : a review of 100 fMRI studies published in 2009 », *Annals of the New York Academy of Sciences*, vol. 1191, no 1, p. 62-88, March, 2010.

R Core Team, *R : A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. 2013.

Rauzy S., Montcheuil G., Blache P., « MarsaTag, a tagger for French written texts and speech transcriptions », *Proceedings of the Second Asian Pacific Corpus linguistics Conference*, 2014.

Roze C., Base Lexicale des Connecteurs discursifs du français, mémoire de DEA, Master thesis, Université de Paris Diderot, Paris, 2009.

Schiffrin D., *Discourse markers*, Cambridge University Press, Cambridge, New York, 1987.

Shriberg E. E., Preliminaries to a theory of speech disfluencies, PhD thesis, Citeseer, 1994.

Wolfe F. H., Auzias G., Deruelle C., Chaminade T., « Focal atrophy of the hypothalamus associated with third ventricle enlargement in autism spectrum disorder », *NeuroReport*, vol. 26, n° 17, p. 1017-1022, December, 2015.

Xu Y., Reitter D., « An Evaluation and Comparison of Linguistic Alignment Measures », *Proceedings of CMCL 2015*, p. 58-67, 2015.