

Le défi des données spontanées pour améliorer les interactions humain-robot : le projet Peppermint

Lucien Tisserand¹, Frédéric Armetta², Heike Baldauf-Quilliatre³,
Mathieu Lefort², Julien Perier-Camby², Brooke Stephenson^{1,2},

¹ICAR (UMR5191), CNRS, Labex ASLAN, Lyon, France

²LIRIS (UMR5205), UCBL, Labex ASLAN, Villeurbanne, France

³ICAR (UMR5191), CNRS, ENS Lyon, Univ Lyon 2, Labex ASLAN, Lyon, France

Correspondence: lucien.tisserand@ens-lyon.fr

1 Introduction

Notre présentation s'articule sous la forme d'un retour d'expérience d'une collaboration entre Analyse Conversationnelle (AC) et Intelligence Artificielle (IA), en particulier dans les domaines de la robotique sociale, de la multimodalité et de l'utilisation des *Large Language Models*. Cette collaboration a pour objectif d'améliorer la programmation et le design d'un robot commercial grand public – le modèle Pepper de Aldebaran¹ – à partir d'objectifs et de critères d'évaluation établis grâce à l'analyse systématique d'un corpus d'interactions humain-robot (HRI) spontanées.

Malgré les récentes avancées en *HRI studies*, les "échecs" restent à la fois fréquents et difficiles à caractériser (Förster et al., 2023). Doit-on se baser sur les échelles de satisfactions rapportées post-hoc par les utilisateurs, caractériser les réponses de la machine comme bonnes ou mauvaises quand bien même l'utilisateur est satisfait d'un échec mis en scène, ou se baser sur des indices mesurables dans l'interaction ? Plus fondamentalement, doit-on évaluer la machine selon les connaissances que nous avons du fonctionnement de l'interaction sociale dans une situation où l'humain sait qu'il ou elle s'adresse à une intelligence artificielle, ici incarnée coporellement par un robot humanoïde ? Et comment procéder ?

Au sein d'un projet interdisciplinaire de trois ans (2021-2024), financé par le LabEx ASLAN, nous avons essayé de caractériser (une partie de) des problèmes identifiés dans l'HRI et d'identifier des directions et solutions possibles afin de les éviter tout en s'alignant avec les orientations rendues manifestes par les utilisateurs tout-venant.

1. Voir la présentation commerciale de ses caractéristiques dans *IEEE Robotics & Automation Magazine* : <https://ieeexplore.ieee.org/document/8409927>

2 Objectifs et questions de recherche

Les HRI en situation de laboratoire font que l'utilisateur est invité à suivre strictement un scénario proposé par l'expérimentation. On pourrait alors croire que le suivi des règles d'un jeu informatique est la même chose que suivre les règles de l'interaction sociale. Or, dans une rencontre spontanée humain-robot, les usagers possèdent moins de connaissances d'arrière plan sur "l'objectif" du robot, son fonctionnement et le script à suivre. Ces personnes tout-venantes se réfugient donc dans des formats et normes d'action que celles-ci revendiquent comme adéquates afin de conduire une telle interaction. Ces pratiques réglées sont des ressources pour faire sens de leurs comportements en public, et non pas des règles au sens informatique du terme, c'est à dire, comme un jeu d'instructions (Button and Sharrock, 1995). En particulier, l'approche de l'AC, par l'analyse détaillée et multimodale de l'organisation de l'interaction permet de montrer le rapport qui est entretenu par les utilisateurs vis-à-vis des normes d'interaction avec une machine. Des pratiques normées spécifiques sont-elles mobilisées ? S'agit-il de transpositions des interactions sociales humaines, et le cas échéant, cela est-il problématisé ? Cela s'observe grâce au positionnement des silences, hésitations, *accounts*, rires, les réparations etc. dans l'interaction.

Nous avons donc proposé d'établir une recherche "dans le monde réel" (vs. en situation de laboratoire) afin de caractériser les échecs en fonction de ces rapports aux normes rendus visibles dans l'interaction. Dès lors notre démarche a permis de redéfinir les "échecs" ainsi que "ce qu'un système devrait faire" dans cette situation de façon empiriquement étayée en prenant en compte les attentes spécifiques liées à l'HRI et ce qu'il est possible de faire.

3 Méthodologie

Nous avons collecté un corpus vidéo d'interactions spontanées entre le robot Pepper et des usagers d'une bibliothèque universitaire dans le cadre d'un scénario crédible et représentatif de l'usage commercial répandu pour ce modèle : rendre des services basiques d'information. Dans cette perspective d'authenticité, le robot a été programmé selon les recommandations constructeur à partir d'une machine à états finis : l'automate détecte des mots-clefs (à des moments prévus pour les questions, salutations, remerciements, compliments des utilisateurs) et fournit des tours de parole scriptés en guise de réponse. Deux caméras grand angle et trois microphones ont permis de capter l'ensemble de la situation depuis le moment où les personnes s'approchent du robot. En effet, le robot fut placé dans un environnement passant à côté de la réception. Les usagers étaient libres d'interagir et repartir à tout moment, sans aucune instruction. Ainsi, plus de 700 interactions avec le robot ont été enregistrées, puis transcrites et alignées temporellement selon les conventions ICOR² de transcription de l'oral. Nous avons ensuite réalisé les analyses séquentielles approfondies d'une sélection de phénomènes récurrents. Ces analyses nous ont permis de déterminer un schéma d'annotation du corpus adossé aux transcriptions temporellement alignées, afin de récupérer systématiquement des occurrences de types de problèmes liés à la gestion de la séquentialité en tant que déploiement temporel continu (action, temporalité, réparation, valeur évaluative, enchâssements, adressage). Ces choix analytiques ont été réalisés en lien avec la possibilité de traiter informatiquement ces situations problématiques dans le cadre du projet et selon l'état de l'art en IA et HRI.

4 Revue de la littérature

Plusieurs études basées sur l'AC ont identifié un problème central dans le cas des interactions multipartites (constituant le cas le plus courant dans les espaces publics ouverts) : le sens, la temporalité et les mécanismes basés sur la séquentialité des tours, lorsque ces derniers sont adressés à la machine, dépendent en réalité de l'interaction qui a d'abord lieu entre les utilisateurs et utilisatrices (Gehle et al., 2015; Porcheron et al., 2017; Arend et al., 2017; Reeves and Porcheron, 2023; Tisse-

2. <http://icar.cnrs.fr/corinte/conventions-de-transcription/>

rand and Baldauf-Quilliatre, 2024). Or, ces interactions reflètent la complexité des actions sociales dans le monde réel dont l'interprétation échappe spécifiquement à la formalisation (Button and Sharrock, 1995). Alors qu'elles influencent les inputs adressés à la machine, on ne peut les prendre en compte telles quelles d'un point de vue informatique. Cependant, les études détaillées des interactions avec les robots et autres agents conversationnels ont mis en avant deux aspects qui montrent que les personnes s'orientent vers une simplification de l'interaction avec la machine. Premièrement, elles mobilisent des pratiques spécifiques normées concernant l'adressage à une machine, telles que le formatage par mots-clefs (emphase prosodique ou isolation syntaxique) des tours de parole (Pelikan and Broth, 2016; Avgustis et al., 2021; Velkowska et al., 2020). Deuxièmement, il a été montré que les utilisateurs se réfèrent prioritairement aux "normes" interactionnelles que sont les *paires adjacentes* (Schegloff, 1968), et transposent une partie de ces "normes" comme le type d'action attendue (Pelikan and Broth, 2016; Reeves et al., 2018; Fischer et al., 2019; Licoppe and Rollet, 2020). L'AC pourrait alors fournir des spécifications pour un système qui prendrait en charge ce système d'attentes, ce qui, dans notre cas, a débouché sur des pratiques d'annotations du corpus en fonction de ces "normes" d'interaction. Dans les études HRI, si le séquençage entre action est considéré comme important pour la compréhension de la progression d'un dialogue, les méthodes pour modéliser explicitement la séquentialité commencent tout juste à être explorées (Duran, 2023; Kunneman and Hindriks, 2022), mais à notre connaissance, aucune étude n'a entrepris de modéliser la séquentialité en tant qu'actions déployées dans une temporalité continue.

5 Résultats issus de la recherche

Notre communication discutera les défis de la collaborations en se focalisant sur 1) les particularités de l'interaction humain-robot, 2) ce que l'annotation séquentielle du corpus permet de modéliser, 3) les défis pour l'apprentissage automatique de la reconnaissance multi-modale des interactions en environnement ouvert, 4) vers l'exploitation des régularité identifiées par un LLM pour une interaction humain-robot plus naturelle.

Remerciements

Les auteurs remercient le LABEX ASLAN (ANR-10-LABX-0081) de l'Université de Lyon pour son soutien financier dans le cadre du programme français "Investissements d'Avenir" géré par l'Agence Nationale de la Recherche (ANR).

References

- Béatrice Arend, Patrick Sunnen, and Patrice Caire. 2017. Investigating breakdowns in human robot interaction : A conversation analysis guided single case study of a human-nao communication in a museum environment. *International Journal of Mechanical, Aerospace, Industrial, Mechatronic and Manufacturing Engineering*, 11(5).
- Iuliia Avgustis, Aleksandr Shirokov, and Netta Iivari. 2021. "Please Connect Me to a Specialist" : Scrutinising 'Recipient Design' in Interaction with an Artificial Conversational Agent. In *INTERACT 2021*, pages 155–176. Springer Nature.
- Graham Button and Wes Sharrock. 1995. On simulacrum of conversation : Toward a clarification of the relevance of conversation analysis for human-computer interaction. In *The social and interactional dimensions of human-computer interfaces*, pages 107–125. Cambridge University Press, New York, NY, US.
- Nathan Duran. 2023. *Conversation analysis for computational modelling of task-oriented dialogue*. Ph.D. thesis, University of the West of England.
- Joel E. Fischer, Stuart Reeves, Martin Porcheron, and Rein Ove Sikveland. 2019. [Progressivity for voice interface design](#). In *Proceedings of the 1st International Conference on Conversational User Interfaces*, CUI '19, pages 1–8, New York, NY, USA. Association for Computing Machinery.
- Frank Förster, Marta Romeo, Patrick Holthaus, Luke J. Wood, Christian Dondrup, Joel E. Fischer, Farhana Ferdousi Liza, Sara Kaszuba, Julian Hough, Birthe Nessel, Daniel Hernández García, Dimosthenis Kontogiorgos, Jennifer Williams, Elif Ecem Özkan, Pepita Barnard, Gustavo Berumen, Dominic Price, Sue Cobb, Martina Wiltschko, Lucien Tisserand, Martin Porcheron, Manuel Giuliani, Gabriel Skantze, Patrick G. T. Healey, Ioannis Papaioannou, Dimitra Gkatzia, Saul Albert, Guanyu Huang, Vladislav Maraev, and Epaminondas Kapetanios. 2023. [Working with troubles and failures in conversation between humans and robots: workshop report](#). *Frontiers in Robotics and AI*, 10.
- Raphaëla Gehle, Karola Pitsch, Timo Dankert, and Sebastian Wrede. 2015. [Trouble-based group dynamics in real-world HRI : Reactions on unexpected next moves of a museum guide robot](#). In *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 407–412, Kobe, Japan. IEEE.
- Florian Kunneman and Koen V. Hindriks. 2022. [A Sequence-Based Dialog Management Framework for Co-Regulated Dialog](#). In *HAI2022 : Augmenting Human Intellect*, pages 143–156. IOS Press.
- Christian Licoppe and Nicolas Rollet. 2020. [" Je dois y aller " . Analyses de séquences de clôtures entre humains et robot](#). *Réseaux*, 2020/2(220-221) :151–193.
- Hannah R.M. Pelikan and Mathias Broth. 2016. [Why that nao? how humans adapt to a conventional humanoid robot in taking turns-at-talk](#). In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, page 4921–4932, New York, NY, USA. Association for Computing Machinery.
- Martin Porcheron, Joel E Fischer, and Sarah Sharples. 2017. [" do animals have accents?" talking with agents in multi-party conversation](#). In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*, pages 207–219.
- Stuart Reeves and Martin Porcheron. 2023. [Conversational ai: Respecifying participation as regulation](#). In *The SAGE Handbook of Digital Society*, pages 573–592. SAGE Publications Ltd.
- Stuart Reeves, Martin Porcheron, and Joel Fischer. 2018. ['This is not what we wanted': designing for conversation with voice interfaces](#). *ACM Interactions*, 26(1) :46–51.
- Emanuel A. Schegloff. 1968. Sequencing in conversational openings. *Directions in sociolinguistics*, 70(6) :1075–1095.
- Lucien Tisserand and Heike Baldauf-Quilliatre. 2024. [Rejecting a robot's offer: An analysis of preference](#). *Discourse & Communication*, page 17504813241271486.
- Julia Velkovska, Moustafa Zouinar, and Clair-Antoine Veyrier. 2020. [Les relations aux machines conversationnelles : Vivre avec les assistants vocaux à la maison](#). *Réseaux*, N° 220-221(2) :47–79.

A Appendice : résumé graphique du Projet Peppermint



FIGURE 1 – Schéma résumant la démarche du projet Peppermint.