

Entre analyse conversationnelle et intelligence artificielle : défis méthodologiques pour modéliser une interaction avec Pepper

Lucien Tisserand (Sciences du Langage, Post-Doctorant CNRS, ICAR)
Antoine Bouquin (Sciences Informatiques, Ingénieur de recherche CNRS, LIRIS)
Sofiane Doulfaouar (Sciences du Langage, Assistant ingénieur CNRS, ICAR)
Justine Lascar (Sciences du Langage, Ingénieure d'études CNRS, ICAR)



Notre présentation porte sur les défis méthodologiques rencontrés et les solutions trouvées afin de mener un travail collaboratif de recherche entre Analyse Conversationnelle (AC) et Intelligence Artificielle (IA). Nous menons une étude exploratoire qui vise à restituer des analyses d'interactions de service humain-robot afin d'améliorer les interactions avec ce même robot (Pepper de Aldebaran). Sur le plan plus fondamental, notre recherche questionne empiriquement les attentes exprimées en temps réel par les humains envers ces objets qui viennent à la fois perturber et revendiquer une place dans un environnement connu (une bibliothèque universitaire). Car si les interactions humaines apparaissent servir de modèle et de repère tant pour les designers que pour les utilisateurs, des pratiques spécifiques d'adressage aux robots, mettant en scène la conversation (Button & Sharrock 1995), sont également analysées.

Notre démarche s'effectue pas-à-pas, avec des aller-retours entre les essais de programmation et mises en situation du robot et l'analyse des interactions spontanées issues de ces mises en situation. L'enjeu est de développer de nouveaux algorithmes et modèles qui prennent en compte la dynamique temporelle de prise des *tours de parole* (Levinson & Torreira 2015). Nous rendrons compte de cette expérience interdisciplinaire et des défis qu'elle a suscités, notamment 1) trouver des langages communs et expliciter nos différentes méthodologies de travail, et 2) gérer l'hybridation des approches qualitative et quantitative pour un objectif donné dès le recueil des données.

1) Des problèmes de traductions interdisciplinaires ont émergé dès l'écriture d'un premier scénario d'interaction programmable. Ce programme *ad-hoc* avait pour but de recueillir un premier jeu de données dont les analyses fournissent à la fois des résultats sur les spécificités (AC) de l'interaction humain-robot (IHR) mais aussi un corpus annoté systématiquement afin que celui-ci nourrisse un apprentissage machine pour la reconnaissance automatique (IA). Le scénario devait prévoir plusieurs possibilités de reconnaissance et d'action : Pepper devait être programmé pour orienter et informer les utilisateurs d'une bibliothèque universitaire à Lyon. Nous montrerons les effets concrets — dans la programmation et dans les interactions obtenues — des différences de représentation de la *séquentialité* question-réponse comme un enchaînement logique (IA) et comme un enchaînement temporel (AC). Entre le premier et second aller-retour sur le terrain, la solution a été que les linguistes apprennent le langage de modélisation des automates à états finis (cf. fig.1) afin d'en détourner les conventions.

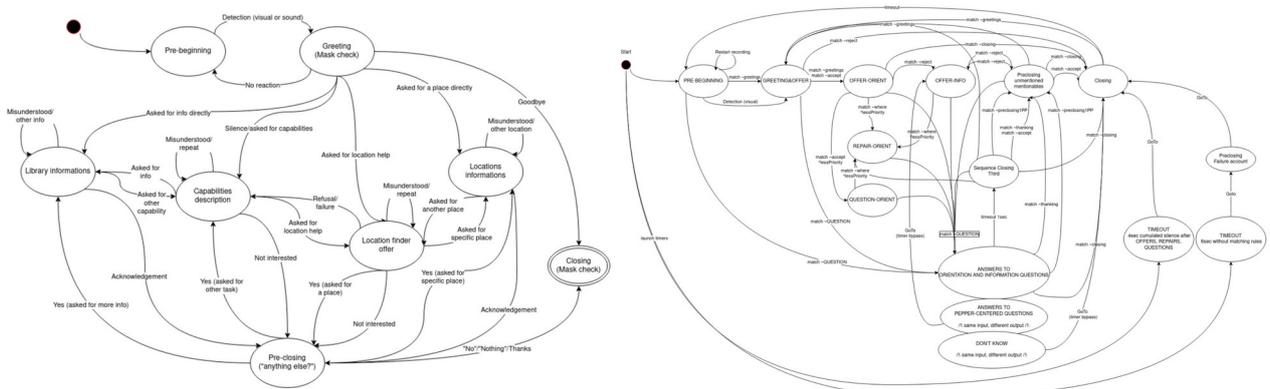


Fig.1 Diagrammes des deux versions successives de la machine à état

2) Notre méthodologie implique le recueil de données audiovisuels des interactions obtenues entre le robot et les humains dans un hall passant de la bibliothèque universitaire. Notre dispositif multi-caméras et multi-microphones nous permettra d'évoquer les enjeux liés à la complexité spatiale et acoustique de ces IHR. Les différents points de vue entre la caméra à la première personne du robot vs les caméras à grand angle, préservant l'écologie de l'interaction, permettent un montage épousant à la fois les objectifs analytiques de l'IA et de l'AC. Il en résulte alors la nécessité de synchroniser les données collectées pour les rendre accessibles en tant que données d'analyse et d'apprentissage.

Ce corpus audiovisuel d'environ 500 interactions représentant 8 heures de données a ensuite été transcrit, annoté et codé sous le logiciel ELAN® (Max Planck Institute, cf. fig.2). Ces opérations sur le corpus permettent :

- en AC, de découvrir et constituer en collection des phénomènes *séquentiels* centraux dans la gestion de ces IHR, par exemple la distinction des cadres participatifs entre humains et avec/au sujet de Pepper et la *réparation* (Kitzinger 2013) ;
- en IA, de permettre aux chercheurs d'apprendre au robot à reconnaître les caractéristiques multimodales de ces phénomènes et de proposer de nouveaux modèles d'IA qui tiennent compte de cette *séquentialité*.

Les enjeux et défis posés par cette interdisciplinarité se manifestent ainsi à toutes les étapes de notre projet, jusqu'à la phase actuelle où nous devons articuler la description qualitative des phénomènes et une approche quantitative pour l'apprentissage machine.

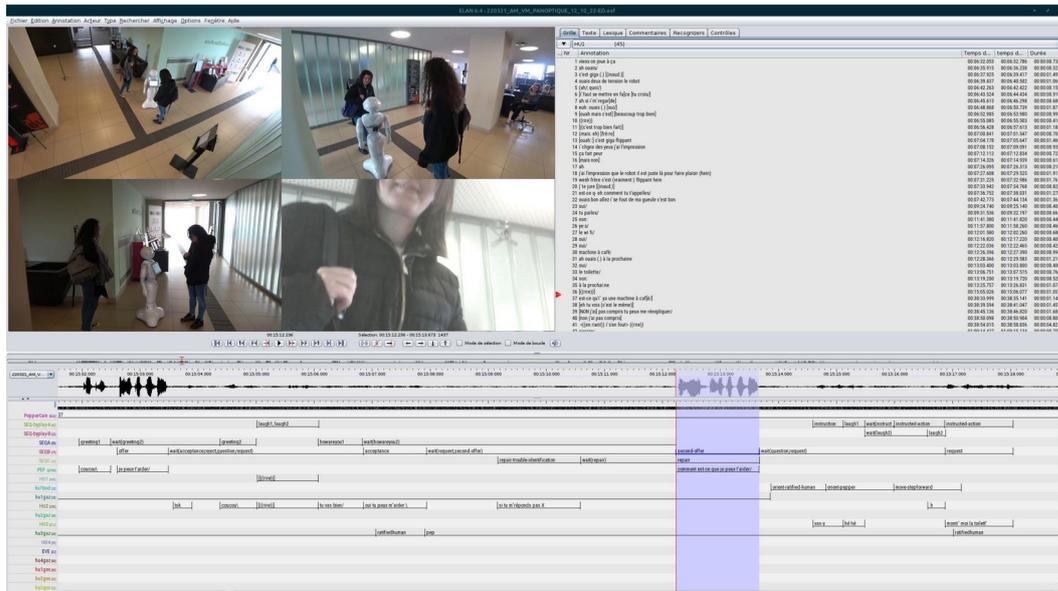


Fig.2 Capture d'écran du traitement du corpus sous ELAN®.

- Button, Graham, and Wes Sharrock. 1995. 'On Simulacrums of Conversation: Toward a Clarification of the Relevance of Conversation Analysis for Human-Computer Interaction'. Pp. 107–25 in *The social and interactional dimensions of human-computer interfaces*. New York, NY, US: Cambridge University Press.
- ELAN (Version 6.4) [Computer software]. (2022). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. <https://archive.mpi.nl/tla/elan>
- Kitzinger, Celia. 2013. 'Repair'. Pp. 229–56 in *The Handbook of Conversation Analysis*.
- Levinson, S., & Torreira, F. (2015). 'Timing in Turn-Taking and Its Implications for Processing Models of Language'. *Frontiers in Psychology* 6(JUN):1–17. doi: [10.3389/fpsyg.2015.00731](https://doi.org/10.3389/fpsyg.2015.00731).