



**HAL**  
open science

# Intelligence Socio-Affective pour un Robot : primitives langagières pour une interaction évolutive d'un robot de l'habitat intelligent

Yuko Sasa

► **To cite this version:**

Yuko Sasa. Intelligence Socio-Affective pour un Robot : primitives langagières pour une interaction évolutive d'un robot de l'habitat intelligent. Intelligence artificielle [cs.AI]. Université Grenoble Alpes, 2018. Français. NNT : 2018GREAM041 . tel-01945238

**HAL Id: tel-01945238**

**<https://theses.hal.science/tel-01945238>**

Submitted on 5 Dec 2018

**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.

## THÈSE

Pour obtenir le grade de

### **DOCTEUR DE LA COMMUNAUTE UNIVERSITÉ GRENOBLE ALPES**

Spécialité : **Informatique**

Arrêté ministériel : 25 mai 2016

Présentée par

**Yuko SASA**

Thèse dirigée par **Véronique AUBERGÉ**

co-direction par **Gang FENG**

co-encadrement par **Yoshinori SAGISAKA**

préparée au sein du **Laboratoire d'Informatique de Grenoble (LIG)**  
dans l'**École Doctorale Mathématiques, Sciences et Technologies  
de l'Information, Informatique**

## **Intelligence Socio-Affective pour un Robot : primitives langagières pour une interaction évolutive d'un robot de l'habitat intelligent**

Thèse soutenue publiquement le **26 janvier 2018** devant le jury  
composé de :

**Mme Martine ADDA-DECKER**

Professeur, Université Paris III, LPP/CNRS, Président

**Mr. Mohamed CHETOUANI**

Professeur, Sorbonne Université, ISIR, Rapporteur

**Mr. Daniel LUZZATI**

Professeur émérite, Université du Maine, LIUM, Rapporteur

**Mme. Véronique AUBERGÉ**

Chargée de recherche, CNRS, Université Grenoble Alpes, LIG,  
Directeur de thèse

**Mr. Gang FENG**

Professeur, Université Grenoble Alpes, Gipsa-lab, Co-directeur de thèse

**Mr. Yoshinori SAGISAKA**

Professeur, Université Waseda, GITS, Co-encadrant de thèse

**Mr. Vincent RIALLE**

Maître de conférences-Praticien hospitalier Émérite, CHU Grenoble  
Alpes, Université Grenoble Alpes, Invité

**Mr. Florian NEBOUT**

Ingénieur R&D et Services Manager, Awabot, Villeurbanne, Invité





THÈSE  
pour obtenir le grade de  
DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

*Spécialité : Informatique*  
*Arrêté ministériel : 7 août 2006*

**Intelligence Socio-Affective pour un Robot :  
primitives langagières pour une interaction évolutive  
d'un robot de l'habitat intelligent**

*présentée par*

**Yuko SASA**

*dirigée par* **Véronique AUBERGÉ**  
*co-direction par* **Gang FENG**  
*co-encadrement par* **Yoshinori SAGISAKA**

*Thèse soutenue publiquement le 26 janvier 2018 devant le jury composé de :*

**Mme Martine ADDA-DECKER**

*Professeur, Université Paris III, LPP/CNRS, Président*

**Mr. Mohamed CHETOUANI**

*Professeur, Sorbonne Université, ISIR, Rapporteur*

**Mr. Daniel LUZZATI**

*Professeur émérite, Université du Maine, LIUM, Rapporteur*

**Mme. Véronique AUBERGÉ**

*Chargée de recherche, CNRS, Université Grenoble Alpes, LIG, Directeur de thèse*

**Mr. Gang FENG**

*Professeur, Université Grenoble Alpes, Gipsa-lab, Co-directeur de thèse*

**Mr. Yoshinori SAGISAKA**

*Professeur, Université Waseda, GITS, Co-encadrant de thèse*

**Mr. Vincent RIALLE**

*Maître de conférences-Praticien hospitalier Émérite, CHU Grenoble Alpes, Université Grenoble Alpes, Invité*

**Mr. Florian NEBOUT**

*Ingénieur R&D et Services Manager, Awabot, Villeurbanne, Invité*



*Dédicace,*

みんなを思い、いつも一人で頑張って生きている、祖母、岡本八重子様、  
今、何をしていますか？どこにいらっしゃるのでしょうか？  
何もできず日々を過ごす母、そして私たちも、  
心を通し、あなたと一緒にいます。  
あなたを近くで毎日支えている皆様に、感謝の気持ちを込めて。



## Résumé

Le Traitement automatique de la Parole (TAP) s'intéresse de plus en plus et progresse techniquement en matière d'étendue de vocabulaire, de gestion de complexité morphosyntaxique, de style et d'esthétique de la parole humaine. L'*Affective Computing* tend également à intégrer une dimension « émotionnelle » dans un objectif commun au TAP visant à désambiguïser le langage naturel et augmenter la naturalité de l'interaction personne-machine. Dans le cadre de la robotique sociale, cette interaction est modélisée dans des systèmes d'interaction, de dialogue, qui tendent à engendrer une dimension d'attachement dont les effets doivent être éthiquement et collectivement contrôlés. Or la dynamique du langage humain situé met à mal l'efficacité des systèmes automatiques. L'hypothèse de cette thèse propose dans la dynamique des interactions, il existerait une « glu socio-affective » qui ferait entrer en phases synchroniques deux individus dotés chacun d'un rôle social impliqué dans une situation/contexte d'interaction. Nous nous intéressons dans cette thèse à des dynamiques interactionnelles impliquant spécifiquement des processus altruistes, orthogonale à la dimension de dominance. Cette glu permettrait ainsi de véhiculer les événements langagiers entre les interlocuteurs, en modifiant constamment leur relation et leur rôle, qui eux même viennent à modifier cette glu, afin d'assurer la continuité de la communication. La seconde hypothèse propose que la glu socio-affective se construise à partir d'une « prosodie socio-affective pure » que l'on peut retrouver dans certaines formes de micro-expressions vocales. L'effet de ces événements langagiers serait alors graduel en fonction du degré de contrôle d'intentionnalité communicative qui s'observerait successivement par des primitives langagières : 1) des bruits de bouche (non phonétiques, non phonologiques), 2) des sons prélexicaux, 3) des interjections/onomatopées, 4) des imitations à contenu lexical contrôlé. Une méthodologie *living-lab* est ainsi développée au sein de la plateforme Domus, sur des boucles agiles et itératives co-construites avec les partenaires industriels et sociétaux. Un Magicien d'Oz – EmOz – est utilisé afin de contrôler les primitives vocales comme unique support langagier d'un robot majordome d'un habitat intelligent interagissant avec des personnes âgées en isolement relationnel. Un large corpus, EmOz Elderly Expressions –EEE– est ainsi recueilli. Cet isolement relationnel permet méthodologiquement d'appréhender les dimensions de la glu socio-affective, en introduisant une situation contrastive dégradée de la glu. Les effets des primitives permettraient alors d'observer les comportements de l'humain à travers des indices multimodaux. Les enjeux sociétaux abordés par la gérontechnologie montrent que l'isolement est un facteur de fragilisation où la qualité de la communication délite le maillage relationnel des personnes âgées alors que ces liens sont bénéfiques à sa santé et son bien-être. L'émergence de la robotique d'assistance en est une illustration. Le système automatisé qui découlera des données et des analyses de cette étude permettrait alors d'entraîner les personnes à solliciter pleinement leurs mécanismes de construction relationnelle, afin de redonner l'envie de communiquer avec leur entourage humain. Les analyses du corpus EEE recueilli montrent une évolution de la relation à travers différents indices interactionnels, temporellement organisés. Ces paramètres visent à être intégrés dans une perspective de système de dialogue incrémental – SASI. Les prémisses de ce système sont proposées dans un prototype de reconnaissance de la parole dont la robustesse ne dépendra pas de l'exactitude du contenu langagier reconnu, mais sur la reconnaissance du degré de glu, soit de l'état relationnel entre les locuteurs. Ainsi, les erreurs de reconnaissance tendraient à être compensées par l'intelligence socio-affective adaptative de ce système dont pourrait être doté le robot.

**Mots clés :** glu socio-affective, interaction homme-robot, habitat intelligent, gérontechnologie, prosodie, micro-expressions langagières, parole spontanée, TAP, systèmes de dialogue

## **Abstract**

*The Natural Language Processing (NLP) has technically improved regarding human speech vocabulary extension, morphosyntax scope, style and aesthetic. Affective Computing also tends to integrate an “emotional” dimension with a common goal shared with NLP which is to disambiguate the natural language and increase the human-machine interaction naturalness. Within social robotics, the interaction is modelled in dialogue systems trying to reach out an attachment dimension which effects need to an ethical and collective control. However, the situated natural language dynamics is undermining the automated system’s efficiency, which is trying to respond with useful and suitable feedbacks. This thesis hypothesis supposes the existence of a “socio-affective glue” in every interaction, set up in between two individuals, each with a social role depending on a communication context. This glue is so the consequence of dynamics generated by a process which mechanisms rely on an altruistic dimension, but independent of dominance dimension as seen in emotions studies. This glue would allow the exchange of the language events between interlocutors, by regularly modifying their relation and their role, which is changing themselves this glue, to ensure the communication continuity. The second hypothesis proposes the glue as built by “socio-affective pure prosody” forms that enable this relational construction. These cues are supposed to be carried by hearable and visible micro-expressions. The interaction events effect would also be gradual following the degree of the communication’s intentionality control. The graduation will be continuous through language primitives as 1) mouth noises (neither phonetics nor phonological sounds), 2) pre-lexicalised sounds, 3) interjections and onomatopoeias, 4) controlled command-based imitations with the same socio-affective prosody supposed to create and modify the glue. Within the Domus platform, we developed an almost living-lab methodology. It functions on agile and iterative loops co-constructed with industrial and societal partners. A wizard of oz approach – EmOz – is used to control the vocal primitives proposed as the only language tools of a Smart Home butler robot interacting with relationally isolated elderly. The relational isolation allows the dimensions the socio-affective glue in a contrastive situation where it is damaged. We could thus observe the primitives’ effects through multimodal language cues. One of the gerontechnology social motivation showed the isolation to be a phenomenon amplifying the frailty so can attest the emergence of assistive robotics. A vicious circle leads by the elderly communicational characteristics convey them to some difficulties to maintain their relational tissue while their bonds are beneficial for their health and well-being. If the proposed primitives could have a real effect on the glue, the automated system will be able to train the persons to regain some unfit mechanisms underlying their relational construction, and so possibly increase their desire to communicate with their human social surroundings. The results from the collected EEE corpus show the relation changes through various interactional cues, temporally organised. These denoted parameters tend to build an incremental dialogue system in perspectives – SASI. The first steps moving towards this system reside on a speech recognition prototype which robustness is not based on the accuracy of the recognised language content but on the possibility to identify the glue degree (i.e. the relational state) between the interlocutors. Thus, the recognition errors avoid the system to be rejected by the user, by tempting to be balanced by this system’s adaptive socio-affective intelligence.*

**Keywords:** *socio-affective glue, human-robot interaction, Smart Home, gerontechnology, prosody, micro-words, spontaneous and ecological speech, NLP, dialog system*



## Remerciements

Les contributeurs de cette thèse sont tellement nombreux que ma première peur serait certainement d'oublier quelqu'un. Je tiens donc avant toute chose, à remercier sans distinctions toutes personnes ou instances m'ayant accordé du temps depuis 2013 (et même avant) pour tous les travaux qui ont été réalisés dans le cadre de cette thèse.

Cette thèse est réalisée dans le cadre de l'École Doctorale MSTII (Mathématiques, Sciences et Technologies de l'Information, Informatique). Elle a été financée par le Labex Persyval-lab ANR- 11-LABX-0025-01, le projet Interabot BGLE n°2 ainsi que des contributions complémentaires, de la bourse de mobilité de l'ex-UJF, l'Université Grenoble Alpes, l'Université de Waseda, le LIG, l'action RHUM (Gipsa-lab), de l'Ambassade de France à travers le programme FadEx.

Les partenaires industriels et territoriaux : Awabot, (LIRIS et Voxler à travers le projet Interabot), Bien à la Maison, l'association AAPPUI, le foyer-logement Roger-Meffreys ainsi que toutes les institutions territoriales et collectivités locales, notamment les CCAS et associations grenobloises d'aides aux personnes âgées et personnes isolées, que je remercie pour leur participation et leur disponibilité.

Ma gratitude également au jury qui sera confronté à un manuscrit dense par rapport aux habitudes de la discipline.

\*\*\*\*\*

Il m'a toujours été dit que le meilleur est au début ou à la fin, alors qu'en fait il est omniprésent...

À Véronique Aubergé, toute ma gratitude que de simples mots sont loin de suffire pour l'exprimer. À celle qui m'a trouvé, à apprécier mes qualités et mes défauts scientifiques, et qui a su concrétiser des intuitions naïves ou de simples attraits réflexifs dans une vocation inaltérable en perpétuel remaniement d'idées partagées. J'espère que cette thèse sera digne de la confiance intellectuelle que tu m'as accordée et qu'elle aura su t'apporter sa modeste contribution pour l'avancée de ta pensée d'une science vécue et inspirée. La richesse de cette expérience professionnelle et personnelle est un présent qui ne peut se compter à mes yeux. Merci d'avoir persisté, de m'avoir convaincu de me lancer dans cette aventure, et d'avoir valorisé mes différences ainsi que mes cultures pour continuer à les cultiver. Entre toi qui ne taris pas de propositions et moi qui ne sais pas très bien dire non (mais en réalité, ne veut juste pas le dire), la quantité d'opportunités incroyables que tu m'as données depuis que nous nous connaissons ne se comptent plus. J'ai beau parfois me plaindre d'être submergée, je n'ai vraiment pas à m'appitoyer, tellement tes apports sont importants. J'espère ne pas trop avoir dénaturé ta pensée par mes propos encore naïfs et que les potentielles retombées des travaux engendrés te reviendront à juste mesure.

句坂芳典様、科学者が科学者でいられる為の「夢」、そして「目標」の大切さを教えてくださいまして本当にありがとうございます。先生が私にいつか質問されていましたね、「笹さんの夢は？」と。今では私の中で、人間の生命力の源は、周りとの触れ合いから生まれる糸。人は人と話す事により、その糸を変えていき、素地、形、長さ、弾力、色、その他と様々な特徴を持ち、変化をしていく。全てが切断してしまえば、人間は死んでしまう。また、人との関係性や文化が違えば糸の本質が変わるのではないかと、私の中に潜んでいる二つの文化が囁いています。そのような声を聞く応援も先生から頂き、将来も慎重に考え続けます。まだ未熟ですが、そんな私の科学者としての夢は、その見えない糸を観察できる、「めがね」を作る事だと思います。ロボットもその一部。私には人間を理解したり作ったりする能力はありませんが、一生をかけて編上げる糸の作り方、その中での変化の理解がしたいです。現代社会と共に最近見られる社会問題や SNS は、糸が痛められているからこそ、見られ現象だとも思っています。人間は本能的に生き延びる延長で、新たな IT、そして昔から存在するにもかかわらず、今注目を浴びている対話ロボットが発達しているのではないのかと疑問に思います。最後に、この論文は私が思う糸が存在していて、まだ相手と結ばれていない段階で、最初の綱目をどのように入れられるかをブユアーな韻律の形が土台で、徐々に言葉を扱いながら築き上げる器具とも言える研究の結果です。フランス語のみの論文で申し訳ないのですが、ご一読頂ければ幸せです。先生が紹介して下さった貴重な出会いも、近い将来すぐに役に立ちそうです。

## Remerciements

---

À Gang Feng, qui restera pour moi l'image de ce que la science peut faire de meilleur, à savoir, avoir le génie d'une simplicité extrêmement efficace, mais monstrueusement difficile à atteindre dans les complexités du monde et de l'humain.

J'ai été extrêmement honorée d'avoir été formée et suivie auprès de vous trois et j'espère pouvoir mettre à bon profit ces précieuses bases de la recherche scientifique. Mon seul regret serait certainement de n'avoir pas pu échanger davantage avec vous, sur des questions de fond, car nous n'arrivons à comprendre et à voir certaines choses que lorsqu'il ne nous reste plus de temps, puisque l'on erre dans la brume beaucoup trop longtemps. Mais au fond chaque chose en son temps.

Mais que serait une thèse sans les échanges issus de rencontres multiples et variées, ainsi que le soutien permanent des collègues et des proches ? Mes remerciements s'adressent :

\*\* tout d'abord aux personnes (que je ne connais parfois que par leurs écrits), mais dont j'ai pu hérité d'importantes connaissances pour démarrer mes réflexions sur ce travail, à savoir, Nicolas Audibert, Albert Rilliard, Takaaki Shochi, Amandine Fouard, Anne Vanpé, Fanny Loyau, Rosario Signorello, Lu Yan, Khoa Mac Dong, Juliette Khan ;

\*\* ensuite à toute l'équipe Getalp sans exception, permanents ou non permanents, tout d'abord en leur qualité de chercheur, car l'évolution dans leur milieu m'a appris énormément, et puis en leur qualité humaine, car chacun a un respect et une sympathie envers tous qui fait le point fort de la cohésion ainsi que de l'ambiance chaleureuse de cette équipe. Pour leurs conseils, suivis et commentaires avisés, merci à Laurent, Michel, François, Solange; l'aide de Benjamin dans la logistique finale de ma thèse ; pour leurs petites attentions et paroles réconfortantes Jean-Philippe (très efficace avec un joli coup de maître à la fin), Emmanuelle, Gilles, Didier, Christian. Mention spéciale pour le temps pris par Élodie, David et Moez pour leur participation active dans certaines tâches et discussions ;

\*\* aux collègues de nos belles plateformes, dont le soutien a permis la réalisation d'un nombre très important de tâches : Jérôme, Germain, Amr en particulier, mais aussi toute l'équipe du Fablab MSTIC, l'équipe d'Amiqua4Home avec Stan et Nicolas et leurs stagiaires ;

\*\* aux collègues de la faculté d'économie, de la Commission Recherche et Conseil Académique de l'UGA membres d'Unis dans la diversité, pour leur disponibilité et leur sympathie, leur disponibilité pour les remplacements dans les périodes pleines. L'expérience d'enseignement au sein de cette équipe pédagogique et l'immersion dans les méandres politiques et scientifiques institutionnels ont été une autre découverte intéressante du milieu académique ;

\*\* aux « Robodocs » et membres de Synergy que je ne nommerai pas tous, car ils sont très nombreux, mais se reconnaîtront. Ces groupes, mais surtout les personnes qui y contribuent m'ont permis d'avoir une expérience connexe de recherche, dans des ambiances toujours chaleureuses, tout en apportant des compétences supplémentaires très appréciables. Il en va de même pour Nathan, Zainab, Julien avec qui l'organisation des JJCR 2015 a été très instructive. Ces expériences passent aussi par les rencontres de l'école d'été du Labex SMART, de Social HRI et du FadEx, qui m'ont montré le niveau de rigueur et d'excellence attendu par la communauté scientifique. Cette attente fait peur et nous fait réfléchir à deux fois avant de se lancer dans la recherche, mais elle reste une stimulation unique et humaine irremplaçable ;

\*\* on ne peut dès lors oublier nos assistantes d'équipe, secrétaires et adminfi, sans qui il ne serait administrativement pas possible de partir faire ces échanges et avoir de telles expériences. Merci en particulier à Anne-Laure, Zilora, Amélie, Muriel, Élodie, Pascale, Corinne, Valérie, Ghislaine, pour votre gentillesse et votre patience face aux cas compliqués que je vous rapportais. Je profite de cette section pour également remercier Jean-Louis et François pour leurs nombreux soutiens techniques ;

\*\* aux différents interlocuteurs sociétaux, mais en particulier Sandrine Tie Bi, Gaëlle Fulgier-Garden, Hafida Dihili et Marie-Sophie Lasset, qui ont été extrêmement sensibles à notre problématique de recherche et dont les qualités humaines nous rassurent et nous motivent pour que nos travaux puissent avoir leur petite contribution à côté des leurs qui sont incomparables. Merci à Marie-Claire, Laurence, Laetitia, Julie, Adeline, Karine, Séverinne, Farida, Samia et Lise, les aidantes de nos sujets pour leur gentillesse, leur disponibilité et leur implication indispensable pour cette étude ;

\*\* aux « domusiens » qui ont partagé de près cette expérience de recherche et sans lesquels ce travail n'aurait pas été possible de réaliser et qui ont supporté mes ascenseurs émotionnels :

- à Brigitte, qui nous manque énormément depuis son départ, mais sans qui l'entame de ces travaux n'aurait pas pu se faire. Ses précieuses compétences ne peuvent se replacer et j'espère que nous arrivons à préserver les connaissances et son expertise qu'elle a laissées derrière elle ;
- à Nicolas, dont les soutiens technique et moral sont inégalables. J'espère très sincèrement qu'une personne avec des compétences d'adaptabilité d'ingénierie poussées avec une telle compréhension des expérimentations SHS ne sera pas perdue dans les méandres institutionnels... Les plateformes émergentes ont vraiment besoin de personnes comme toi ;
- à Sylvie, sans qui la logistique des expériences aurait été bien difficile et qui a toujours été une présence discrète réconfortante ;

## Remerciements

---

- à Tim et Jonathan pour leur monstrueuse efficacité et leur présence nonchalante qui m'a énormément motivé et sans qui la réalisation technique du magicien d'Oz n'aurait pas été possible ;
- à Gilles de m'avoir aidé à réfléchir et partager mes premiers pas dans cette expérience de thèse. Je regrette que l'on n'ait pas pu finir ensemble, mais nos premières discussions m'ont scientifiquement beaucoup aidée ;
- à Leandra ma grande sœur de cœur brésilienne, qui a été avant tout une collègue très attachante et efficace, qui m'a aidé à commencer mes premiers pas dans l'exploration phonétique et prosodique, à construire ma pensée autour des méthodologies, à organiser ma première conférence alors que je ne savais même pas encore ce que c'était que des *Proceedings* et a toujours su avoir les mots pour m'aider à me ménager ;
- à Adrien pour sa contribution sur l'interface, Leslie pour la découverte du monde de la gestualité à travers une expérience originale et sa connaissance d'Emox, à Jander pour la transition technique et sa bonne humeur, au regret de ne pas avoir été présente à ton arrivée sur la plateforme, à Sybille pour ses conseils avisés, à Pierre pour sa bonne humeur. À Maxence pour sa perspicacité, sa vivacité d'esprit et sa gentillesse, Noémie, Cécile, Marie-Claire pour leurs contributions discrètes, mais tout aussi importantes à Alain, pour la bouffée d'air réflexive revigorante qu'il m'a apportée à travers son beau projet ;
- à Ambre, pour son travail autour de la conversion de voix et ces questionnements pertinents qui à chaque fois m'ont fait prendre du recul sur mes propres réflexions. Mais aussi de toute l'aide transversale que tu m'as apportée sur la fin de ma thèse. Je sais que tu gardes en toi beaucoup d'incertitude malgré tes apparences sereines, mais j'espère que tu pourras extérioriser tes craintes et énervements sans détour à ceux qui sauront t'aider ;
- à Natacha, pour ta joie de vivre, tes bons gâteaux et ton enthousiasme. Les qualités et les défauts de chacun permettant de s'améliorer sans condition, les tiens m'ont permis de développer mes approches pédagogiques et mes compétences explicatives ;
- à Fred, qui me connaît depuis mes premières excursions au LIG, tu as toujours su être très pédagogue, un partenaire de discussion scientifique important, sans qui les données récoltées n'auraient pas pu être technologiquement exploitées ;
- à Liliya, ta force de travail, ta rigueur, la pertinence de tes réflexions ainsi que ton extrême gentillesse ont grandement contribué à l'avancée de cette thèse. Ton savoir-dire direct est une force à laquelle j'espère à nouveau pouvoir me confronter. Son dosage est certainement délicat à gérer, mais garde confiance sans excès, il saura être apprécié par les personnes qui savent le reconnaître ;
- à Romain, pour tes remarques toujours efficaces et pragmatiques. Tu n'as jamais déçu au moment où on avait besoin de tes nombreuses compétences et ton soutien moral sans faille m'a vraiment permis de passer le cap de la rédaction. Je ne pourrais que te souhaiter de pouvoir, à ton tour, avoir la possibilité d'officialiser tes recherches, avec la chance de pouvoir être soutenu au moins à hauteur de ce que tu donnes aux autres discrètement ;
- à Patrick, pour l'efficacité de tes actes et ta gentillesse dont tu as fait preuve à notre égard afin que cette thèse se termine dans les meilleures conditions possibles ;

\*\* à mes amis, toujours là pour avoir des mots de réconfort, avec une dédicace à celles et ceux qui ont vécu ou vivent des heureux événements de leur vie. Voici la raison de mes absences répétées, dont j'espère justifierai à vos yeux mes éloignements temporaires et surtout à toi Juline, auprès de qui mes millions d'excuses ne suffiront jamais, d'autant plus que ton soutien a toujours été infaillible. Mention spéciale pour Lucie, en espérant que la relecture de mon manuscrit ne *ponctuera* pas tes rêves trop longtemps...

\*\* à ma famille, qui a réussi à supporter mes angoisses et mes écarts d'humeurs, en espérant que ce travail expliquera bien des interrogations qu'ils ont eues à mon égard. Une pensée particulière pour ma mère qui a su être le pont entre mes attaches culturelles qui m'ont fait voir des choses à travers ses travaux et ma grand-mère Hisako, qui m'a soutenu sans relâche et suit de près mes travaux.

Et enfin à toutes les personnes âgées, consultées ou participantes à cette étude. Le temps que vous m'avez accordé n'a pas d'égal, et je vous remercie de la richesse de vos échanges. Comme de nombreuses fois où vous m'avez dit, « *nous ne voulons qu'aider la jeune génération* ». Cette aide est indéniable. Et très lucides, vous me disiez « *mais bon c'est pas pour demain, c'est pas pour nous, c'est pour ceux qui viennent* ». Je regrette que nous ne soyons pas suffisamment efficaces pour que cette aide, vous ne puissiez en bénéficier durant votre vécu. En revanche, j'espère que ces travaux pourront vous montrer le respect vis-à-vis de votre contribution ainsi que les perspectives auxquelles les générations futures pourraient être confrontées.

Une pensée collective emprunte de gratitude à celles et ceux qui ne sont plus...

## Remerciements

---

---

# Table des matières

Remerciements .....	9
Table des matières .....	13
Conventions typographiques .....	18
Abréviations .....	19
Introduction .....	14
Chapitre 1. Notions pour aborder l'interaction homme-robot .....	18
1.1. Des origines du langage à la dimension sociale de l'humain .....	18
1.2. Intelligence artificielle : bref historique .....	25
1.3. Des primitives de langage au robot : enjeux du Traitement Automatique des Langues 32	
1.3.1. Nouveaux défis de parole expressive en synthèse .....	32
1.3.2. Dynamique langagière en reconnaissance de la parole .....	34
1.3.3. Des théories à l'informatique des affects sociaux .....	36
1.3.3.1. De l'attachement aux marqueurs comportementaux .....	39
1.3.3.2. Du mimétisme aux comportements d'imitation .....	42
1.3.4. Signaux sociaux : continuité multimodale de la communication parlée .....	44
1.3.5. Micro-expressions vocales .....	46
1.3.5.1. Organisation, distribution et typologies .....	46
1.3.5.2. Micro-expressions vocales à la prosodie .....	52
1.3.6. De la voix à la personnalité : perspectives en robotique .....	56
1.4. Comprendre l'humain à travers le robot .....	58
1.4.1. La conscience de soi : corporéité et « embodiment » .....	58
1.4.2. Conscience de l'objet : lorsque le robot devient autre .....	61
1.4.3. De l'engagement à l'attachement : vers une « illusion empathique » .....	67
1.5. Dimensions sociales pour une robotique interactionnelle .....	73
1.5.1. Empathie : un paradigme complexe à multiples facettes .....	73
1.5.2. Spécifications de l'empathie dans l'approche robotique .....	78
1.5.3. De l'empathie à l'altruisme : création d'un lien social .....	81
1.5.4. Altruisme : une dimension propre de la dynamique d'interaction ? .....	83
1.5.5. Robotique comme instrument de modélisation .....	84
1.5.5.1. Hypothèse de la « glu socio-affective » .....	84
1.5.5.2. Hypothèse de « prosodie pure » graduellement gluante .....	85
1.5.5.3. La personnalisation vocale du robot .....	88
Chapitre 2. Enjeux et contextes sociétaux .....	92
2.1. Isolement relationnel : un choix méthodologique d'observation .....	92
2.1.1. De la question du rôle social au lien social .....	93
2.1.2. Isolement fragilisant : le cas des personnes âgées .....	97
2.2. Contexte socio-culturel et apport de la gérontechnologie .....	101
2.2.1. Problèmes de vieillissement de la population .....	101

2.2.2.	Technologies d'intelligence ambiante pour le maintien à domicile .....	103
2.3.	Robotique de services pour les personnes âgées .....	105
2.3.1.	Robotique d'assistance à la personne .....	105
2.3.2.	« Compagnons » : vers une robotique de services à caractère social .....	110
2.3.3.	De la robotique interactionnelle au choix du robot .....	111
Chapitre 3.	Méthodologie et protocoles expérimentaux .....	118
3.1.	Préambule autour des sciences de l'information et sciences du langage .....	118
3.2.	Vers des données écologiques et spontanées pour un apprentissage machine itératif .....	121
3.2.1.	Magicien d'Oz : continuité paradigmatique de la télé-présence à l'autonomie .....	122
3.2.2.	De l'induction à l'observation hypothético-déductive .....	124
3.2.3.	Observation de l'humain et living-labs .....	126
3.3.	Méthodologie expérimentale globale .....	129
3.3.1.	Co-construction tripartite : le corpus au cœur du système .....	129
3.3.1.1.	Hypothèses pluridisciplinaires pour le recueil de « <i>smart data</i> » .....	130
3.3.1.2.	Partenaire industriel : l'entreprise Awabot .....	132
3.3.1.3.	Éco-système humain : les aidants des personnes âgées .....	133
3.3.2.	Plateforme Domus-LIG .....	137
3.3.2.1.	Environnement : appartement intelligent .....	137
3.3.2.2.	Au-delà d'un environnement : une méthodologie itérative et intégrative de l'écologie des comportements .....	139
3.3.3.	Approches sous-jacentes aux boucles agiles .....	141
3.3.3.1.	Approche hypothético-inductive : scénarii expérimentaux prétextés .....	143
3.3.3.2.	Transversalité de l'approche hypothético-inductive : « levée de voile » et continuité expérimentale .....	145
3.3.3.3.	Approches inspirées de l'ethnométhodologie : annotations externes des valeurs informatives .....	146
3.3.3.4.	Approche empirico-déductive : étiquetage descriptif longitudinal des données .....	148
3.3.3.5.	Approche hypothético-déductive : la validation perceptive .....	150
3.3.3.6.	Boucles itératives : alternance et transversalité des approches scientifiques .....	151
3.4.	Protocoles expérimentaux .....	153
3.4.1.	Processus de recrutement .....	153
3.4.1.1.	Sélection des sujets .....	154
3.4.1.2.	Gestion de la tâche prétexte dans les entretiens .....	157
3.4.2.	Capture des données spontanées mais contrôlées .....	159
3.4.2.1.	Plateforme EmOz .....	159
3.4.2.1.1.	Accueil : déviation d'attention et indices d'isolement .....	163
3.4.2.1.2.	Entrée dans Domus : la question de l'amorçage .....	163
3.4.2.1.3.	Phase d'interaction personne-robot .....	164
3.4.2.1.4.	Retours des complices et débriefing en deux phases .....	166
3.4.3.	Auto-annotation : les contraintes d'adaptation pour la personne âgée .....	167
3.4.3.1.	Choix et conditions de l'auto-annotation .....	167
3.4.3.2.	Prétraitement des données .....	167
3.4.3.3.	Déroulement de la séance .....	168
Chapitre 4.	Corpus EEE : vers une modélisation de l'interaction homme-robot .....	172

---

4.1.	Caractéristiques générales du corpus .....	172
4.1.1.	Nature des données récoltées.....	172
4.1.2.	Profil des sujets retenus .....	173
4.1.3.	Panel des non participants entrés dans le protocole de sélection.....	174
4.2.	Étiquetage de données multi-sources .....	177
4.2.1.	Matériel interactionnel sélectionné.....	177
4.2.2.	Transcription des données .....	179
4.2.3.	Outils et organisation générale .....	180
4.2.4.	Étiquetage des productions vocales des sujets.....	183
4.2.4.1.	Commandes vocales – Cmd .....	183
4.2.4.2.	Nouvelles commandes – NwCmd.....	183
4.2.4.3.	Non-commandes – NonCmd.....	183
4.2.5.	Sous-étiquetage : focus sur les événements hors commandes.....	184
4.2.5.1.	Principes de segmentation des NonCmd.....	184
4.2.5.2.	Micro-expressions vocales de sujets – ME .....	184
4.2.5.2.1.	Micro-expressions « satellites » de commandes.....	185
4.2.5.2.2.	Micro-expressions « périphériques » de commandes.....	186
4.2.5.3.	Formes d’expressions complexes et dialogiques – FC .....	187
4.2.5.3.1.	Filtrage rapide de commentaires.....	187
4.2.5.3.2.	Énoncés dialogiques complexes .....	188
4.2.6.	Étiquetage multimodal des productions non vocales des sujets.....	189
4.2.6.1.	Nomenclature d’étiquetage manuelle : les limites .....	189
4.2.6.2.	Ciblage de la multimodalité par les auto-annotations.....	190
4.2.7.	Étiquetage des stimuli contrôlés par EmOz.....	191
4.2.7.1.	Des traces de l’interaction face-à-face aux traces du magicien .....	191
4.2.7.2.	Alignements des commandes domotiques .....	191
4.2.7.3.	Alignements des commandes robotiques scriptées.....	192
4.2.7.4.	Alignements manuels des mouvements robotiques à la manette.....	192
4.3.	Auto-annotations et marquage des « temps cognitifs ».....	194
4.3.1.	Intentions à travers les productions de sujets : FoT_ProdSuj.....	194
4.3.2.	Temps latents généraux : FoT_Gn.....	195
4.3.3.	Indices ponctuels communicationnels : FoT_IPC .....	195
4.3.4.	Impressions perceptives d’Emox : FoT_PE .....	196
4.3.5.	Alignements multiples d’auto-annotations chevauchées.....	196
Chapitre 5. Analyses des productions vocales : vers une prosodie du dialogue ? .....		199
5.1.	Aperçu quantitatif des données .....	199
5.1.1.	Proportion générale des productions vocales .....	199
5.1.2.	Proportion des actions domotiques et robotiques .....	200
5.1.3.	Proportion des étiquettes d’auto-annotation .....	201
5.2.	Des marqueurs de changements de glu .....	203
5.2.1.	Modification des formes morphosyntaxiques des commandes .....	203
5.2.1.1.	Les patterns de paraphrasage des commandes.....	203
5.2.1.2.	Cas de non-paraphrasage : indices équivalents.....	204
5.2.2.	Nature et organisation des « micro-expressions vocales ».....	206
5.2.2.1.	Proportion générale des événements hors commandes.....	206
5.2.2.2.	Observations sur les micro-expressions « satellites ».....	207
5.2.2.3.	Observations sur les micro-expressions « périphériques » .....	208

---

5.2.3.	Marqueurs remarquables de l'interaction socio-affective « gluante ».....	210
5.2.3.1.	Compliments et politesse .....	210
5.2.3.2.	Justification et marqueurs du <i>care</i> .....	210
5.2.3.3.	Phénomène de récupération d'erreurs.....	211
5.2.4.	Les observations acoustiques et prosodiques générales .....	212
5.2.4.1.	Augmentation générale du pitch .....	212
5.2.4.2.	La <i>breathiness</i> : dynamique subtile de la dimension « altruiste » ? .....	214
5.2.4.3.	Bref aperçu des profils prosodiques remarquables .....	216
5.3.	Evolutions temporelles des marqueurs interactionnels .....	218
5.3.1.	Chronogrammes des événements interactionnels.....	218
5.3.2.	Evolution de la perception d'Emox .....	220
5.3.2.1.	Description de « niveaux » de glu.....	220
5.3.2.2.	Visualisation temporelle des niveaux de glu mesurés.....	221
5.3.2.3.	Patterns d'évolution de la glu des sujets .....	222
5.3.3.	Evolution de « gluage » des commandes.....	225
5.3.3.1.	Visualisation temporelle des productions interactionnelles.....	225
5.3.3.2.	Patterns d'évolution des « processus de gluage » des sujets.....	226
5.3.4.	Evolution d'états latents du sujet.....	227
5.3.4.1.	Visualisation temporelle des durées de phases latentes .....	227
5.3.4.2.	Patterns d'évolution des états latents des sujets.....	228
5.3.5.	Détection de temps interactionnels forts ponctuels .....	229
5.3.5.1.	Distribution et durées des événements remarquables .....	229
5.3.5.2.	Comparaison inter-sujets des événements remarquables.....	230
Chapitre 6.	Système SASI : <i>Socio-Affective Speech Interaction</i> .....	232
6.1.	Modélisation et calculs de niveaux de glu socio-affective.....	232
6.1.1.	Algorithme Damereau-Levenshtein modifié .....	234
6.1.2.	Grammaire de coûts des variations de structures langagières .....	236
6.2.	Un prototype de reconnaissance d'états de glu-parole.....	238
6.2.1.	Reconnaissance de la parole et des caractéristiques des voix âgées.....	238
6.2.2.	Vers une solution de détection de « niveaux de glu » .....	240
6.2.3.	Proposition d'un algorithme de détection de glu.....	242
6.2.4.	Modèles multiples de langage en fonctionnement parallèle.....	244
6.3.	Positionnement du système dans le contexte actuel.....	246
6.3.1.	Quelles évolutions à partir des systèmes de dialogue actuels ?.....	246
6.3.2.	Vers une technologie hybride de dialogue évolutif et incrémental .....	248
Chapitre 7.	Discussions et Perspectives.....	251
7.1.	Pistes d'exploitation de données non traitées.....	251
7.1.1.	Entretiens : caractérisation de l'isolement et « entraînement » communicationnel .....	251
7.1.2.	Expériences annexes : améliorations de protocole .....	252
7.2.	<i>Acceptance</i> technologique et éthique .....	254
7.3.	Visualisation de données : comparer courbe « calculée » et courbe « mesurée » de glu .....	256
7.4.	Pistes de réflexion : de l'étiquetage à la synchronie .....	258
7.5.	Quelques réflexions sur le « bruit » du dialogue.....	259

## Table des matières

---

7.6. Pistes d'implémentation du prototype SASI .....	261
Conclusion générale .....	263
Bibliographie .....	267
Publications personnelles .....	305
Table des illustrations.....	307
Table des annexes.....	311
Annexes .....	313

## Conventions typographiques

*Terminologie étrangère alphabétisée* : cette thèse comporte de nombreuses terminologies étrangères dont les occurrences sont désignées en italique. Usage des guillemets anglais “*citations*” pour les citations non traduites en français.

« *Ceci est une citation* » : les citations longues d’auteurs sont réalisées entre guillemets dans une police Cambria, taille 10, italiques.

« Néologisme ou mot emphase » : certains termes sont des néologismes développés au sein de l’équipe. D’autres termes font une emphase ou se construisent sur des mots valises qui amènent à réflexion ou sont encore utilisés pour faciliter la compréhension de manière imagée. Toutes ces occurrences particulières de termes sont mentionnées entre cotes.

« *Transcription de sujet* » : dans les parties d’analyse de la thèse,

(Auteur, Année1/Année2) : le référencement suit la nomenclature APA. Parmi les citations, certains ouvrages ont fait l’objet de réédition bien que les notions ou les théories soient anciennes. Le manuscrit essaie donc de rendre compte de cette information en évoquant deux années, la première étant la plus ancienne dont la mention est faite dans l’ouvrage/article consulter dans l’édition de la seconde date mentionnée.

Les publications personnelles suivent un format distinct pour les distinguer de la bibliographie.

## Abréviations

### Abréviations françaises

ACA : Agents Conversationnels Animés  
AVC : Accident Vasculaire Cérébral  
AVQ : Activités de la Vie Quotidienne  
CERNA : Commission de réflexion sur l’Ethique de la Recherche en sciences et technologies du Numérique d’Allistène  
CTL : Centre des Technologies du Logiciel  
DHM : Dialogue Homme Machine  
EHPAD : Établissement d’Hébergement pour Personnes Âgées Dépendantes  
EPA: Établissement pour Personnes Âgées  
F0 : Fréquence fondamentale  
GdR : Groupement de Recherche  
GIR : Groupe Iso-Ressources  
IA : intelligence artificielle  
IHM : Interface Homme Machine  
LIG : Laboratoire d’Informatique de Grenoble  
NAC : Nouveaux Animaux de Compagnie  
OHLL : Origine de l’Homme, du Langage et des Langues  
OMS : Organisation Mondiale de la Santé  
RAP : Reconnaissance Automatique de la Parole  
SHS : Sciences Humaines et Sociales  
STC : Sciences des Technologies et de la Cognition  
TAL : Traitement Automatique de la Langue  
TAP : Traitement Automatique de la Parole

### Abréviations anglaises

*AAI: Adult Attachment Interview*  
*AI: Artificial Intelligence*  
*ASA: Acoustical Society of America*  
*ASR: Automatic Speech Recognition*  
*CDS: Care Dependency Scale*  
*DLA: Daily Life Activity*  
*DNN: Deep Neural Network (réseaux de neurones)*  
*EDD: Eye-Direction Detection*  
*FACS: Facial Action Coding System*  
*FoAK: Feeling of Another Knowing*  
*FoK: Feeling of Knowing*  
*GAFAMI: Google, Apple, Facebook, Amazon, Microsoft, IBM*  
*HCI: Human Computer Interaction*  
*HMM: Hidden Markov Model*  
*HRI: Human-Robot Interaction*  
*ID: Intention Detection*  
*IoT: Internet of Things (objets connectés)*  
*IST: Information Society Technologies*  
*ISTAG: IST Advisory Group*  
*JSTAR: Japanese Study of Aging and Retirement*  
*LEADS: Leeds Elderly Assessment Dependency Screening*  
*MC: Machine Consciousness*  
*MMSE: Mini Mental State Examination*  
*NLP: Natural Language Processing*

*OMLL: Origin of Men, Language and Languages*  
*SAM: Shared Attention Mechanism*  
*SPT: Subjective Perception of Time*  
*SSP: Social Signal Processing*  
*TD-PSOLA: Time Domain Pitch Synchronous Overlap and Add*  
*ToBI: Tones and Break Indices*  
*ToMM: Theory of Mind Mechanism*  
*WoZ: Wizard of Oz (magicien d’Oz)*

### Usages fréquents et spécifiques dans le manuscrit :

AA : auto-annotations  
BB : bruit de bouche  
Cmd : commandes  
FC : formes complexes  
*FoT: Feeling of Thinking*  
Gn : général  
IPC : indices ponctuels communicationnels  
PE : perception d’Emox  
ProdSuj : productions du sujet  
Interj. : interjections  
ME : micro-expressions  
NonCmd : non commandes  
NwCmd : nouvelles commandes  
PreLex : des sons pré-lexicalisés  
RAP : Reconnaissance Automatique de la Parole

*EEE: EmOz Elderly Expressions* (nom du corpus récolté)  
*EmOz: Emox Wizard of Oz* (nom de la plateforme magicien d’Oz développé)  
*HRI: Human-Robot Interaction* (pour désigner l’interaction Homme-Robot dans la terminologie française)  
*WoZ: Wizard of Oz* (magicien d’Oz)



## Introduction

Le langage parlé présente des variantes et une complexité qui caractérisent l'humain. Il s'agit d'un outil qui construit et régule les relations entre individus pour la persistance de leur(s) place(s) sociale(s). Le contenu du message est fondamental, mais « notre manière de le dire » peut changer entièrement l'effet de ce message sur un autre, et en particulier la relation que nous développons avec cet autre. Le rôle que l'humain prend dans chaque contexte socioculturel à chaque moment interactionnel est alors indispensable. Il va amorcer la nature de la relation que nous pouvons avoir avec l'autre. C'est ce rôle qui va pouvoir introduire cette communication qui nous modifie au fur et à mesure des interactions. En revanche, sans le canal qui nous relie, nous ne restons que des entités coupées du monde, ce qui est synonyme d'un état léthargique pour un humain qui est intrinsèquement social. La communication va donc renforcer et modifier le canal tout au long de nos interactions. Dans notre quotidien, nous constatons naïvement et subtilement ce phénomène, et parfois des *coachs* qui tendent à nous apprendre comment faire. Mais comment peut-on comprendre scientifiquement les mécanismes communicationnels résultants sur ce canal ?

L'observation humaine en est un point de départ, mais sans construction méthodologique, cette observation du comportement tel qu'il existe dans la nature ne peut donner lieu qu'à des spéculations empiriques où seule la statistique décide de la viabilité de la modélisation de cesdits comportements. Si plus encore, nous identifions des outils de langage spécifiques dont l'observation veut être ciblée, soit le contrôle inévitable que nous exerçons pour les produire change leur naturalité et biaise notre observation, soit il faut arriver à retrouver dans des situations écologiques des lieux où ces outils sont utilisés, mais nous retompons alors au cœur du premier problème. Il ne reste alors qu'à trouver des méthodologies permettant d'induire de la naturalité humaine tout en testant spécifiquement ces objets ciblés. Le robot social devient alors un instrument pertinent de modélisation, dépourvu de tout lien naturel avec l'humain, sur lequel le contrôle ne changera pas la naturalité du comportement puisqu'il n'en existe pas intrinsèquement. Il pourra ainsi être porteur des outils langagiers spécifiques que nous voulions tester pour en comprendre les effets sur l'humain. Mais l'humain reste complexe, et le fait de le confronter devant un objet avec un potentiel animisme induit des comportements spécifiques liés à cette nature interactionnelle. Ce comportement est liée à la nature même du robot, mais aussi à la perception que l'humain s'en fera à travers sont interaction avec lui. De plus, il sera influencé par l'écologie d'observation avec ses propres caractéristiques. Ainsi, nous tenterons de comprendre la construction du canal communicationnel à travers les objets langagiers les plus primitifs, supposés aussi les plus permanents.

Les deux premiers chapitres tenteront ainsi de retracer le paradigme scientifique de manière relativement transversale et pluridisciplinaire. Différentes perspectives permettront de poser l'hypothèse de l'existence d'une « glu socio-affective » qui serait la conséquence des mécanismes interactionnels représentée sur une dimension à part entière, telle qu'observée dans le *grooming*. Nous tentons donc de nous défaire de toute relation de dominance qui semble être un axe obligatoire dans les théories des émotions actuelles, en supposant que la relation peut se construire sur un tout autre axe que nous caractérisons pour le moment comme celui de « l'altruisme ». En particulier, le chapitre 1 s'attardera sur certains mécanismes de communication au regard des origines du langage qui est proche à la problématique d'attribution d'un langage propre au robot qui restera sa première entrée en communication avec l'humain. En effet, cet état de l'art tente de montrer les quelques

mécanismes interactionnels que nous connaissons de l'homme, de même que les processus susceptibles d'apparaître dans le seul fait d'utiliser un robot dans l'interaction.

Le chapitre 2, en introduisant les enjeux socioculturels, introduira les problèmes d'isolement, qui serait peut-être même l'une des raisons qui motiveraient le regain d'intérêt de la robotique à l'heure actuelle, alors qu'elle a toujours fait l'un des objets de fantasme de l'homme. En particulier, cet isolement est abordé à travers la motivation sociétale liée à la prise en charge de la fragilisation des personnes âgées, mais plus encore, ce public est un réel choix méthodologique d'observation puisqu'elle permet de situer de l'interaction dans un contexte contrastif où les effets des primitives choisies seront plus facilement visibles à travers cette situation « dégradée » d'un point de vue relationnel. Ceci ouvrirait même les réflexions sur des « processus de réparation » des mécanismes de communication mis en jeu, s'inscrivant parfaitement aux préoccupations de la gérontechnologie. À travers ces problèmes, ce chapitre présentera les objets langagiers supposés et définis comme des primitives possibles de l'interaction en face à face, permettant de construire et manipuler la « glu socio-affective ». Les objets que l'on qualifiera de « micro-expressions » présentent ainsi une prosodie (dynamique mélodique) socio-affective sous une forme relativement « pure » dans la mesure où elle ne porte pas les marques intonatives contraintes par les supports lexicaux du langage. Ces primitives choisies seront donc celles produites par un robot, avec une gradualité de présentation qui est supposée permettre la manipulation de la glu socio-affective. Les formes et les connaissances sur la nature des informations qu'elles véhiculent s'inscrivent ainsi dans les champs de l'*Affective Computing*, dont l'application s'étend à la robotique sociale, tout en sollicitant le domaine du Traitement Automatique des Langues. L'introduction des domaines de l'intelligence ambiante et de la robotique d'assistance permettra notamment un montage de scénario expérimental cohérent dans la méthodologie choisie.

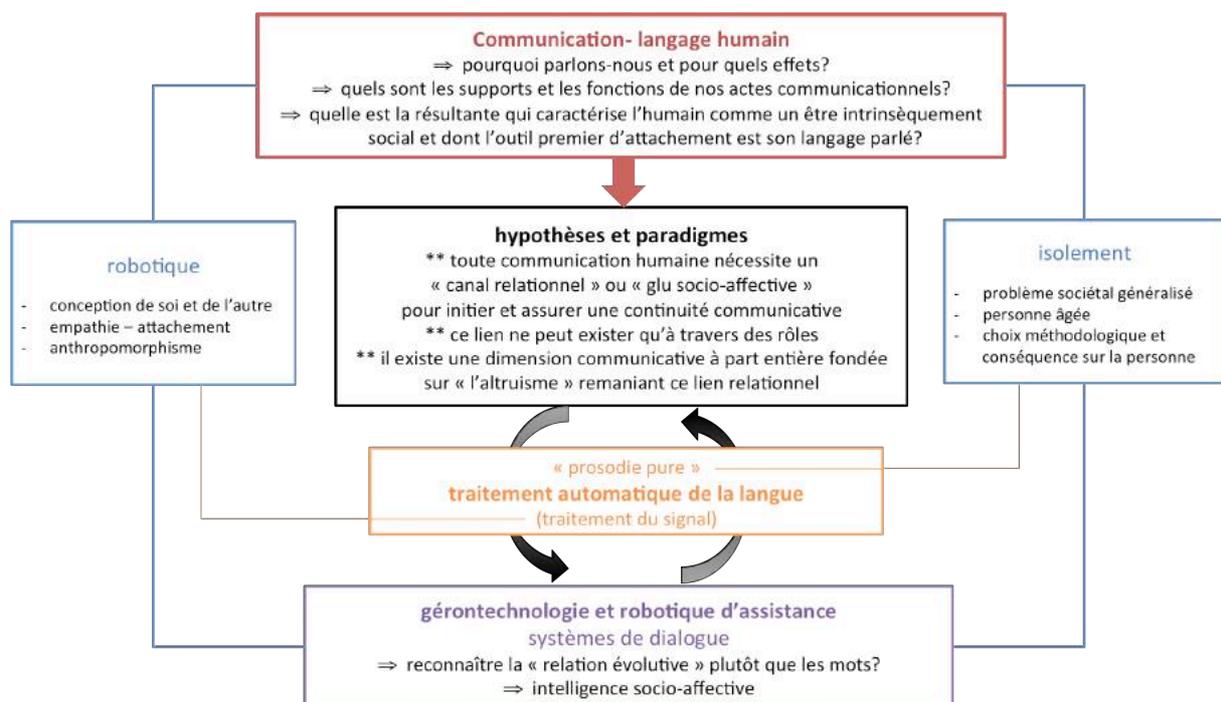


Figure 1 : Schéma simplifié des croisements de notions pluridisciplinaires de l'état de l'art

La Figure 1 ci-dessus tente de simplifier les ponts entre les notions à travers les champs disciplinaires abordés à travers le long état de l'art qui sera proposé. Cet aperçu, sera pourtant très incomplet, du fait de sa nature pluridisciplinaire de la robotique sociale et plus générale de la nature complexe de l'homme, le but étant ici de faciliter la compréhension de la

construction des hypothèses de base sous-tendant les méthodologies de recueil hypothético-déductif et d'annotation hypothético-déductive.

Cette méthodologie expérimentale est explicitée dans le chapitre 3. Elle dérive des approches de type *Living-Labs* et d'une lignée de méthodes d'observation de sciences humaines spécifiques. À travers des boucles agiles, elle propose une observation parallèle à la fabrication et l'évaluation de la technologie. Elle expose à la fois des notions de co-construction pluridisciplinaire et les enjeux de l'observation de l'humain, en favorisant la production de comportements interactionnels spontanés et écologiques dans un environnement contrôlé à travers des outils tels que le Magicien d'Oz. Plus spécifiquement, elle décrira l'outil EmOz et son usage, développés au sein de la plateforme expérimentale Domus-LIG. Les aspects méthodologiques sont d'ailleurs proposés comme l'un des moyens permettant d'appréhender les dimensions éthiques inhérentes à la robotique sociale.

Puis le chapitre 4 exposera la nature des données recueillies dans la construction du corpus EEE - *EmOz Elderly Expressions* - par les réflexions liées aux étiquetages multimodaux dans une approche descriptive. Les choix d'étiquetage sont ainsi au cœur de l'analyse d'un corpus de parole spontanée et nous en verrons les difficultés méthodologiques associées. L'enjeu de cette partie est ainsi de trouver des solutions de représentations acceptables en conservant l'évolution temporelle des outils interactionnels repérables dans la multimodalité, puisque c'est la dynamique de leur modification à travers le temps qui permettra d'appréhender les dimensions de la glu socio-affective. Ce travail est donc sous-jacent et indispensable pour la modélisation d'un système d'interaction homme-robot.

Ainsi, dans le chapitre 5, l'évolution de la « glu socio-affective » est abordée au regard des valeurs informatives portées par les objets étiquetés dans le cadre de méthodes dites d'auto-annotations. Le chapitre s'attardera donc aux observations relevées illustrant les dimensions impliquées dans la dynamique d'interaction, observée de manière transversale sur l'ensemble de la séquence expérimentale. Ce chapitre est à mettre en lien avec des notions telles que la synchronie interpersonnelle dans une vision holistique du dialogue-personne-robot, introduisant des réflexions autour de l'existence d'une « prosodie dialogique ».

Enfin, le chapitre 6 proposera le début d'une modélisation d'un système de dialogue incrémental (SASI - *Socio-Affective Speech Interaction*), dans une perspective dite *long-life*. Au cœur du système se trouvera l'outil de reconnaissance de la parole, dont le but ne sera pas d'obtenir le meilleur taux de reconnaissance, mais de reconnaître le niveau relationnel décrit par le degré de glu socio-affective. L'approche propose ainsi un prototype de reconnaissance automatique de la parole et un outil de pondération basée sur un outil de mesure de distance d'édition comme base de travail. Ainsi, le robot serait en mesure de générer des réactions en cohérence avec ce niveau relationnel détecté. Cela permettrait d'assurer une forme de continuité interactionnelle associée à une potentielle robustesse de fonctionnement à travers le temps. Théoriquement, il s'agit d'une étape évaluative de toute une boucle méthodologique, mais non d'une évaluation finale. En effet, l'approche s'appuie sur une récursivité continue et d'itération rapide, dont les perspectives sont très nombreuses, présentés également à travers des discussions sur des aspects non traités du corpus dans le chapitre 7.



# Chapitre 1. Notions pour aborder l'interaction homme-robot

## 1.1.Des origines du langage à la dimension sociale de l'humain

Dans les origines du langage il est souvent proposé une dichotomie entre communication et langage en questionnant le commencement du langage par rapport à la communication, car ses degrés de complexité et de variabilité ne semblent pour l'instant pas trouver d'équivalence dans les autres espèces. La réflexion est alors empreinte d'un aspect plus identitaire, à savoir comment définir l'Homme si le langage en est une propriété spécifique, et comprendre ce qui le différencie alors des autres êtres vivants.

« La cité est au nombre des réalités qui existent naturellement et (...), il est manifeste (...), que l'homme est par nature un animal politique, et que celui qui est hors cité, naturellement bien sûr et non par le hasard des circonstances, est soit un être dégradé soit un être surhumain, et il est comme celui qui est injurié en ces termes par Homère : sans lignage, sans loi, sans foyer.

Car un tel homme est du même coup naturellement passionné de guerre, étant comme un pion isolé au jeu de trictrac. C'est pourquoi il est évident que l'homme est un animal politique plus que n'importe quelle abeille et que n'importe quel animal grégaire.

Car, comme nous le disons, la nature ne fait rien en vain ; or seul parmi les animaux l'homme a un langage. Certes la voix est le signe du douloureux et de l'agréable, aussi la rencontre-t-on chez les animaux ; leur nature, en effet, est parvenue jusqu'au point d'éprouver la sensation du douloureux et de l'agréable et de se les signifier mutuellement. Mais le langage existe en vue de manifester l'avantageux et le nuisible, et par suite aussi le juste et l'injuste. Il n'y a en effet qu'une chose qui soit propre aux hommes par rapport aux autres animaux : le fait que seuls ils aient la perception du bien, du mal, du juste, de l'injuste et des autres notions de ce genre. »

La politique, Aristote. Livre I, Chap. 2.

Ces réflexions d'Aristote se prolongent dans les pensées des philosophes des Lumières tels que Rousseau par exemple. Notons la référence à :

- la « cité », à savoir le système global qui lie des individus entre eux en un ensemble cohérent, comme structure émergeant naturellement du fait de la nature intrinsèquement sociale de l'Homme. Ce processus de rattachement à un groupe d'appartenance en est une caractéristique propre, inaliénable dans sa définition.
- les systèmes grégaires existent également dans le monde du vivant, autre que celui de l'être humain, mais l'Homme est le seul à être doté de parole.
- la parole passe par la voix qui elle-même est porteuse d'informations, à commencer par des éléments très primitifs tels que les sensations physiologiques et les émotions, qui sont les prémisses de ce qui est échangé, et que l'on pourrait également associer à l'animal. Celle-ci vient naturellement à être modulée pour un échange mutuel, mais dont le contrôle vers une complexité linguistique dépend d'une « morale », soit des valeurs qui émergent par la perception associée à l'appartenance de l'Homme au système socioculturel dans lequel il arrive à s'identifier.

Au cours du XXe siècle, trois disciplines principales se sont emparées de ce problème, dans l'archéologie, la linguistique et la génétique des populations. Les connaissances issues de ces disciplines ne convergent que depuis une trentaine d'années, cette problématique ayant

longtemps été taboue, critiquée d'être spéculative plus que scientifique. Cette convergence a néanmoins permis de relancer ce questionnement malgré un certain scepticisme, à travers le projet OHLL (Origine de l'Homme, du Langage et des Langues), prolongé par le projet européen OMLL (*Origin of Men, Language and Languages*).

À partir de la référence (Hombert, Bergounioux, Bocquet-Appel, Coupé, et Collectif, 2005) et en adoptant une vue macroscopique à l'échelle des sociétés humaines, l'archéologie, la paléontologie et l'anthropologie nous donnent un aperçu global sur les origines du langage, en exposant deux points d'explosion migratoire qui correspondraient potentiellement à l'émergence du langage humain :

- homo erectus : sa migration rapide fait l'objet d'une répartition étendue de ses colonies, un phénomène qui ne s'observait pas chez ses prédécesseurs avec une telle ampleur. Cette forte vague de colonisation s'accompagne également d'avancées techniques et d'une augmentation de la boîte crânienne qui sont autant d'indices qui suggèrent l'apparition des premières formes langagières.
- homo sapiens archaïque : suivant le développement d'homo erectus, son apparition serait accompagnée d'une seconde explosion de colonisation.

La thèse d'émergence du langage, très largement acceptée par la communauté scientifique (Bickerton, 1992), mentionne le développement d'un protolangage sous l'égide de homo erectus. Il s'agirait d'un lexique diversifié, facilitant le repérage dans le monde/environnement et l'échange d'informations, avec dans un second temps l'apparition de syntaxe. Les spécialistes s'attèlent ainsi à déterminer, d'une part pourquoi et comment ce protolangage est apparu, d'autre part pourquoi il a été nécessaire qu'il se complexifie.

Les premières hypothèses suggéraient que ce protolangage était dû à la création d'outils de plus en plus élaborés, du fait de la difficulté à transmettre du savoir. Par analogie aux processus actuels de transmission de connaissance dans le domaine de l'artisanat, cette idée fut invalidée, car celle-ci est majoritairement basée sur le mimétisme ou la pratique manuelle plus qu'une explicitation langagière complexe. D'autres avaient émis l'idée d'un gain d'efficacité lors de la chasse, or une fois encore, l'analogie au monde animal avec l'observation de prédateurs variés écarta cette hypothèse. Les linguistes mettent alors en avant trois phénomènes du langage qui illustrent certaines particularités spécifiques chez l'Homme :

- la récursivité : capacité des langues humaines à s'emboîter les unes dans les autres (p. ex. évoquer les dire de quelqu'un) ;
- le temps et l'aspect : variabilité dans le temps discursif (p. ex. références du temps actuel, passé et futur dans un même récit) ;
- La modalité : changement de prise en charge de l'assertion qui peut être empreinte d'une dimension « morale ».

Ces spécificités ont ouvert la voie à de nouvelles réflexions sur les raisons de l'émergence et d'évolution de ce protolangage. (Dunbar, 2004) expliquait que ces phénomènes étaient liés à la volonté de renforcer les liens sociaux. Plus tard, (Victorri, 1999/2000) suppose que la fonction narrative (le fait de raconter l'histoire), fait partie de ces motivations. En effet, d'un point de vue anthropologique, toutes les sociétés ont vu émerger ce que l'on appelle des « mythes d'origine » qui racontent le passé et évoquent également les interdits sociaux. L'évolution de l'humain dans sa société suivrait ainsi les règles sociales établies qui sont des règles culturelles, à la différence des règles animales qui en suivraient des formes innées et seraient liées à la survie de l'espèce. Dans le même ordre d'idée, (Dessalles, 2000) suggère une émergence du langage par sa fonction dans l'organisation politique des sociétés humaines.

Par ailleurs, le langage est considéré comme le propre de l'Homme pour le biologiste, le psychologue, le paléontologue ou le philosophe. Il a fait l'objet d'un clivage important avec le

milieu animal, avec en particulier l'usage de la parole, mais c'est aussi le phénomène d'une évolution, tout comme la bipédie ou la main préhensile. Le caractère biologique du langage est donc supposé avoir eu un avantage adaptatif et supposé être en lien avec l'émergence de cette bipédie, modifiant la structure du conduit vocal par l'abaissement du larynx (Liebermann et Crelin in Hombert et coll., 2005) et libérant ainsi la cavité buccale permettant ainsi l'usage des mains. Ces articulateurs se placeraient suite à des phénomènes dits d'« exaptation » (possibilité de changer la fonction d'une structure initialement prévue pour autre chose) ou d'« adaptation » (développement d'une structure par le biais de la sélection naturelle), ces processus pouvant parallèlement s'opérer (Tecumseh et Fitch in Hombert et coll., 2005).

Les sons que l'Homme peut ainsi produire auraient permis de produire des éléments phonologiques de base pour ensuite gérer du lexique basique, tout d'abord pour nommer des éléments de son entourage, puis évoluer vers un langage de plus en plus élaboré syntaxiquement pour que l'humain puisse se référer au rapport qu'il entretient avec cet environnement. La conceptualisation à travers le langage peut ainsi s'exprimer. L'humain serait alors capable d'agir sur l'autre, partageant donc avec d'autres des perceptions, des intentions et des informations. De plus, une autre dimension associée au langage est celui du conditionnement et de la gouvernance des relations humaines vivant au sein d'une même communauté. Ce « dialogue social » amènerait ainsi à l'organisation des sociétés, voyant le langage évoluer en parallèle de l'évolution de la culture associée.

D'un point de vue phylogénétique, Chomsky ou Fodor ont supposé que le langage est doté de « modules » exclusivement dédiés à l'expression et la compréhension du langage, donnant des modèles tels que la grammaire générative, qui n'enlève rien à la qualité de la description organisationnelle syntaxique du langage, mais change le point d'entrée de l'accès à sa fonction. Ces courants sont notamment en opposition avec les réflexions behavioristes (p. ex. Skinner). Cette façon chomskyenne de voir le langage rejoint notamment des discussions en faveur d'un « gène du langage », FOXP2 (Enard et coll., 2002), considérant le langage comme une innovation soudaine, par une mutation génétique à l'origine d'un « saut évolutif ». Or l'existence d'un tel gène, de même que les réflexions autour de la phrénologie sont encore très largement discutées (Renneville, 2000), en opposition avec les études montrant des fonctionnements très semblables de nombreuses activités mentales entre l'Homme et d'autres animaux, notamment grégaires. Ces animaux seraient eux aussi dotés de capacités inférentielles, rejoignant notamment les partisans de la théorie de l'esprit. Philippe Vernier par exemple va critiquer les thèses basées sur le « gène du langage », prônant l'idée que plus que l'apparition de nouvelles structures cérébrales, c'est la concordance de structures et fonctions existantes également chez d'autres espèces qui a façonné le cerveau humain manipulant le langage.

Loin de réduire la singularité humaine, les confrontations des similitudes communicatives inter-espèces voient de nombreuses clarifications quant aux connaissances issues de l'embryologie. Elles font ainsi le pont entre l'activité cérébrale et le langage en fonctionnement par l'exploration de la mise en place organisationnelle du cerveau et du système nerveux. Sans réduire à ce seul paramètre l'émergence du langage chez l'humain, les observations telles que celles de Dunbar ont montré que la taille du cerveau (notamment du cortex frontal) est plus importante pour les êtres vivants qui s'organisent en groupes sociaux, composant ainsi un large pan des études neurophysiologiques. En revanche, la taille du cerveau devant avoir une limite, notamment du fait des contraintes de l'accouchement, celui du nouveau-né fait l'objet d'une immaturité relative. Le développement progresse alors durant les dix premières années de la vie, ce qui est appelé l'« altricialité secondaire ». Ce phénomène sous-tend ainsi l'une des particularités de la communication humaine, liée aux influences de l'environnement interactionnelle, faisant l'objet des travaux de l'acquisition du langage (Vernier, 2005).

Dès lors que nous percevons un continuum dans l'évolution du développement du système communicatif véhiculé par la parole, nous remarquons qu'il existe néanmoins une certaine rupture entre le langage parlé et les autres systèmes de communication, et qui essaie d'être comblée par le niveau cognitif. Les auteurs parlent ainsi de « langage articulé » pour se référer à celui qui emprunte le canal voco-auditif, par opposition à la « faculté du langage » qui repose en partie sur des capacités cognitives. Certaines thèses iraient même jusqu'à dire que le contenu et le contenant sont indépendants, notamment dans la perspective de l'émergence de langages gestuels. La présente thèse montrera d'ailleurs que cette rupture est critiquable, car le contenu possède des formes et des dynamiques semblables malgré la différence de modalité du contenant, notamment en considérant par exemple que la « prosodie » qui reflète une dynamique caractérisant des informations communicatives d'un acte langagier va au-delà du signal vocal et peut être considérée comme une dynamique multimodale.

Parmi les études s'intéressant aux origines non plus du langage, mais des langues, une série « d'explorations » que l'on nomme les « expériences interdites » ont, volontairement ou non, observé des individus, en particulier des nourrissons, en les privant de toute attention sociale. Ce sont des cas où les nouveau-nés sont entièrement isolés de contacts sociocommunicatifs et affectifs dès leur naissance. Les premières font référence aux récits de Psammétique rapportés par Hérodote dont le format est proche à l'une des expériences les plus connues, celle de l'empereur Frédéric II de Hohenstaufen. Cependant, il existe d'autres situations d'isolement forcés, encore plus récentes comme celles des « enfants de Ceauçescu ». Il s'agit de cas observés dans le cadre de conditions de vie extrêmes d'enfants délaissés au sein d'orphelinats en Roumanie par exemple, mais dont d'autres sont référencés dans les comptes-rendus de l'Académie des sciences américaine (Strivay, 2004). Ces situations ont d'ailleurs provoqué de nombreux débats sur l'éthique et la moralité. De plus, le rapport « Les enfants et les écrans »<sup>1</sup> de l'Académie des Sciences française met elle aussi en évidence de récents problèmes d'acquisition de langage chez l'enfant moderne éduqué à travers les écrans (ceux pour qui les contacts maternels sont quasi inexistantes), qui sont parfois diagnostiqués d'autistes alors que le phénomène semble progressive et longitudinal. Sans rentrer dans ces réflexions polémiques, la motivation de telles explorations était de découvrir quelle serait la langue naturellement émergente chez l'humain et quelle en est la composante communicative innée, qui le différencie notamment de l'animal.

Dans de telles conditions d'isolement, l'observation principale fut que les nourrissons ne parvenaient pas à survivre bien que leurs besoins vitaux comme l'alimentation ou le sommeil ont été préservés dans des conditions optimales. Des expériences plus récentes (Wismer Fries et Pollak, 2016/2017) tendent même à montrer que la privation d'affection engendre des modifications cérébrales chez le nourrisson. En effet, une corrélation est mise en évidence entre l'état d'isolement et la régulation de la production d'hormones liées aux comportements sociaux (notamment la vasopressine et l'ocytocine), dont la mesure des taux dans différentes conditions est à la base de l'étude mentionnée.

Les conséquences du manque de contact socioaffectif, qui dans les cas les plus graves amènent à la morbidité, sont également observables dans les conditions particulières de ce qu'on appelle l'hospitalisme, connu également sous le nom de *syndrome de Spitz* (M. S. Ainsworth et Bowlby, 1991). Ces termes désignent une « régression mentale des malades hospitalisés pour de longues périodes ». Elle s'observe sous forme de repli relationnel amenant notamment au décès des nourrissons qui sont séparés de leur mère au cours des dix-huit premiers mois de leur vie, dans le cadre d'une hospitalisation prolongée. L'évolution du tableau clinique lié à ce syndrome est rapide :

---

<sup>1</sup> Rapport 2013 « Les enfants et les écrans » par E. Postaire accessible sur : <http://www.academie-sciences.fr/pdf/rapport/avis0113.pdf> (consulté le 2 octobre 2017).

- des pleurs inexplicables et des signes de recherche d'attachement à tout adulte durant le premier mois de séparation ;
- des ralentissements voire même un arrêt de développement avec une perte de poids importante, des signes de « tristesse » et une recherche de contact avec une ardeur amoindrie durant le second mois ;
- un refus de contact avec des signes « d'anxiété, passant par une apathie (couché à plat ventre, insomnies, une fragilisation amenant à contracter facilement des maladies) et un retard psychomoteur qui se généralise dans le courant du troisième mois ;
- au-delà des trois mois, un regard figé, absent et l'absence de sourires, de pleurs remplacés par des gémissements, des mouvements de doigts anormaux, une perte de la coordination oculaire et productions de stéréotypies (production involontaire et continue des mots, gestes ou tics similaires).

Ces phénomènes ont eu des conséquences importantes sur l'organisation la prise en charge d'hospitalisation prolongée, mais ils montrent surtout qu'en l'absence de contact affectif maternel, amenant aux premiers liens d'attachement relationnel, le nourrisson dépérit en réduisant progressivement les signes socioaffectifs de communication.

À un stade moins extrême de rupture socioaffective, ne résultant pas systématiquement sur le décès des individus, mais relevant d'une scission socioculturelle importante, l'observation sur les « enfants sauvages » tels que Victor ou Genie, mettent en évidence des difficultés ou même des échecs de réhabilitation de ces enfants dans le tissu social humain, en fonction de leur condition de vie passée. Cela dépendrait notamment de la durée et de la gravité de leur isolement et/ou de leur éloignement des échanges socioculturels au sein du cercle humain. Ces enfants ont d'ailleurs souvent fait l'objet de fausse supposition de retards mentaux, à l'image des enfants sourds dont le rythme d'acquisition de l'usage des articulateurs et des processus linguistiques est différent de celui des entendants. Ce type d'observation a également été mis en évidence pour les enfants autistes (Guidetti, Turquois, Adrien, Barthélémy, et Bernard, 2004 ; Haag, 1996 ; Lheureux-Davidse, 2006). Ces signes d'attachement primaires apparaissent ainsi très tôt dans la vie de l'humain, tendant même à formaliser la détection des anomalies entre 3 et 30 mois via des outils tels que l'*Early Social Communication Scales* ou ECS (Seibert, Hogan, et Mundy, 1982), largement inspirés des théories développementales et des processus socialisation (Piaget et Cook, 1952 ; Piaget, 1954/2013 ; Werner et Kaplan, 1963/2014).

La complexification langagière de la communication animale vers le langage humain, dont les manifestations sont plus ou moins semblables et marquées selon les espèces, se caractériserait à travers certaines notions telles que :

- la réflexivité : elle fait référence à la métacommunication à savoir communiquer sur la communication. Cette dimension est supposée exister chez le primate sous forme de petites vocalisations ou *chucks* (Newman et Symmes, 1982) et elle est également abordée dans les travaux de (Bateson, 1951 in Ruesch, Bateson, Pinsky, et Combs, 2006), s'intéressant aux phénomènes de correction des chants de certains oiseaux adultes, par rapport aux cris émis par leur progéniture dès lors qu'ils seraient incongrus à la situation ;
- la capacité à élaborer un message à partir d'un autre message : il s'agirait d'une dimension moins observable chez l'animal, mais pointée chez des primates, le léopard, l'aigle, le python, des orques ou des baleines (Bain, 1989 ; Seyfarth et Cheney, 1997) ;
- l'ouverture : propriété définissant que la langue permet de construire de nouvelles expressions à travers sa structure grammaticale (le cas le plus complexe) et d'assigner de nouvelles significations aux éléments nouveaux ou anciens. Ainsi un comportement qui n'était forcément pas initialement dédié à être langagier, pourrait devenir communicatif. C'est notamment ce qui est observé dans le phénomène de *grooming*

(toilettage mutuel ou pansage) pour les chevaux, les primates ou les chauves-souris, jouant sur la relation de confiance et amicale (entre cavalier et cheval, ou entre les animaux) sans objectif de domination ; ou le *leaf-clipping* qui est un phénomène de « pinçage de feuille » visible chez les insectes et qui exprimerait une frustration sans représailles ;

- la double articulation et syntaxe<sup>2</sup> : elle ne semble pas totalement inexistante chez l'animal. (Hölldobler et Wilson, 1994) parlent notamment de « règles prescriptives » pour l'évoquer chez les animaux. De plus, elle semble aussi moins soutenue chez l'humain, puisque par exemple les langues des signes reconnues comme authentiques montrent que cette double structure du langage n'est pas exclusive ;
- enfin le déplacement et le symbolisme sont souvent considérés comme des caractéristiques de langage, l'illustration la plus révélatrice dans le cadre animal étant par exemple la danse des abeilles (1967 in von Frisch, 1993).

Bien que les différentes théories mettent en avant la singularité du langage humain, l'éthologie illustre ainsi une certaine relativité à cette exception langagière de l'Homme. Elle montre notamment que des comportements plus primitifs chez l'animal composent une information communicative importante. Ces comportements peuvent sous-tendre la complexité langagière qui s'élaborerait davantage pour exprimer d'autres dimensions dont notamment la fonction narrative, l'utilisation des temps ou des modalités. (Lestel, 1994) souligne ainsi que l'émergence du langage doit être recherchée « dans l'ensemble des comportements de l'animal, et non dans le sous-ensemble de ses communications ». Il évoque ainsi l'importance de la capacité de manipulation des règles, de partage et d'enseignement à la capacité de coopération via le langage, en mentionnant trois caractéristiques comportementales :

1. les procédures de détournement de l'intelligence de l'autre à son profit : initialement basées sur des phénomènes de mimétisme qui étaient à la base des ruses et des tromperies (Dawkins et Krebs, 1978 ; J. R. Krebs, 1984 ; Whiten et Byrne, 1988) manipulations menant à faire réagir un autre d'une certaine manière ;
2. la délégation individuelle de l'action et de statut : qui est en lien avec le rôle et la place sociale de l'individu, avec des comportements, notamment de type « indicateur » (Isak et Reyer, 1989 in Hombert et coll., 2005) ou « sentinelle » (Horrocks et Hunte, 1986 ; Connor et Heithaus, 1996 ; Mann et coll., 2001 in Hombert et coll., 2005) ;
3. les procédures de chaînes : qui nécessitent une acquisition rapide des « règles de la vie en société » en interagissant via un système langagier plus élaboré, mais qui peuvent s'appuyer sur des formes primitives initiant cette communication, par exemple le chœur des grenouilles mâles (De Waal, 1996 in Hombert et coll., 2005) ou les comportements plus connus de réapprovisionnement de nid par les fournis (Lestel, 1994).<sup>3</sup>

---

<sup>2</sup> Note pour les non-linguistes : cette notion (cf. A. Martinet) fait référence à un concept linguistique qui suppose que le langage humain est un système doublement articulé entre unités significatives et unités distinctives. La première articulation serait composée d'une unité de sens et d'une unité forme, ce que les linguistes comme Saussure appelle morphème (p.ex. l'une des formes du suffixe « in » indique une contradiction dans le mot « incompatible » et chaque forme composant ce mot porte un sens), la seconde articulation supposant que l'association des différentes unités se combinent entre elles pour en faire un énoncé significatif.

<sup>3</sup> Cette thèse s'appuie également sur des primitives langagières associées au robot, ce qui l'engage dans une interaction avec l'humain formant ainsi un système dans lequel il faut pouvoir l'intégrer rapidement dans l'écologie interactionnelle humaine.

Ces comportements soulignent deux approches de réciprocité pour ceux qui manipulent le langage : une réciprocité négative qui s'appuie sur le caractère dominant se basant sur des représailles dans le cas où des règles et interdits sociaux sont ignorés ; et une réciprocité positive appelée encore « altruisme réciproque », qui fait notamment l'objet de réflexion en biologie de l'évolution. Si nous considérons cette approche altruiste du langage, elle relève davantage de la possibilité de dialogue et s'apparente aux réflexions humaines de la maîtrise du langage. C'est donc sous ce deuxième angle que cette thèse tentera d'aborder le phénomène de l'interaction.

## 1.2. Intelligence artificielle : bref historique

« Si le cerveau humain était assez simple pour que nous puissions le comprendre, nous serions si simples d'esprit que nous ne pourrions pas le comprendre. »

(Pugh in Challoner et Haond, 2003)

Par la conscience, nous appréhendons l'autre à travers des formes de raisons/émotions impliquant une projection vers la notion d'intelligence plus ou moins grande ou proche de celle de l'humain. Elle est consciente ou inconsciente, et dans le second cas, plus ou moins volontairement. Pour le robot qui est un autre en devenir, il s'agit avant tout d'un artefact construit par l'homme. L'ingénierie modélise alors son intelligence par analogie ou par des approches logiques, sous un éclairage des sciences cognitives notamment. Dans le cadre de cette thèse, plutôt que de choisir l'une ou l'autre des approches qui seront présentées ci-après, cette partie essaiera de résumer brièvement les grands courants existants pour appréhender l'intelligence artificielle, car cette étude s'inscrit tout naturellement dans ce domaine, sans pour autant cheminer par les voies classiques de cette discipline.

La notion d'« intelligence artificielle » (IA) ou *artificial intelligence (AI)* est apparue en 1956. Elle est introduite par l'informaticien McCarthy qui la définissait comme suit :

« Tout aspect de l'apprentissage, ou toute autre caractéristique de l'intelligence peut en principe être décrit précisément que l'on peut concevoir une machine pour le simuler ».

Pour mettre au défi ces intelligences, les machines étaient donc mises en concurrence à travers des tests, par exemple de celui (Turing, 1950) qui confronte les traces de l'intelligence humaine à celle de la machine pour faire une discrimination de la nature de l'effecteur de la tâche, tout comme la chambre chinoise de (Searle, 1980). Plus récemment nous voyons également le *Coffee Test* (Wozniak in Adams et coll., 2012)<sup>4</sup>, le *Robot College Student Test* (Goertzel, 2012 ; Goertzel, Pennachin, et Geisweiller, 2014) ou encore le *Employment Test* (Nilsson, 2005).

Par ailleurs, il existerait une opposition entre la *weak intelligence* et la *strong intelligence*. La *weak intelligence* dite également *narrow intelligence* ou *applied AI* fait référence aux systèmes spécialisés, limités au sein d'un certain champ d'application restreint (p. ex. Siri dont la nouveauté). Par rapport aux anciens systèmes, sa spécificité est de s'appuyer sur une masse de données considérable – *Big Data* – augmentée par les possibilités techniques offertes par l'apprentissage machine, notamment de type *Deep Learning*. Elle est associée aux champs récents de la *machine consciousness* (MC) ou *synthetic consciousness* (Gamez, 2008 ; Reggia, 2013).

Par contraste, la *strong intelligence* est celle parfois appelée *full AI* ou également *Artificial General Intelligence (AGI)*, dont le dernier terme serait introduit par (Gubrud, 1997) puis décrit par (Goertzel, 2012 ; Kurzweil, 2013). À l'heure actuelle, les neurosciences et leurs approches présentent un engouement important, du fait des nouvelles possibilités calculatoires rendant techniquement possible la levée de certains verrous. Nous voyons notamment l'émergence de réflexions autour de l'hypothèse du point singularité (Kurzweil, 2005). Nous remarquons parmi ces études, le *Blue Brain Projet* développé sur la plateforme *Blue Gene* d'IMB<sup>5</sup>, le Google Brain<sup>6</sup> ou encore le NuPIC<sup>7</sup>. L'AGI introduit aussi quatre

---

<sup>4</sup> Proposition développée dans l'entretien « Three Minutes with Steve Wozniak », le co-fondateur d'Apple <http://www.peworld.com/article/134826/article.html> (consulté le 25/07/2017)

<sup>5</sup> Projet développé à l'EPFL en Lausanne (<http://bluebrain.epfl.ch>, consulté le 7/07/2016). Une présentation succincte est disponible dans une conférence TED donnée par Henry Markram donnée en 2009 : ([https://www.ted.com/talks/henry\\_markram\\_supercomputing\\_the\\_brain\\_s\\_secrets?language=fr](https://www.ted.com/talks/henry_markram_supercomputing_the_brain_s_secrets?language=fr), consulté le 07/07/2016)

<sup>6</sup> Équipe de projet (<https://research.google.com/teams/brain/>, consulté le 09/01/2017)

<sup>7</sup> Plateforme Numenta de Jeff Hawkins (<http://numenta.org>, consulté le 09/01/2017)

notions à travers lesquelles elle veut construire des réflexions éthiques. Ses réflexions sont orientées sous l'angle de la *superintelligence* (le risque de voir une machine surpasser celle de l'humain), la *consciousness*, et de la *self-awareness* (la question du soi et la conscience d'autrui – cf. section 1.4), la *sentience* (présente dans la dimension empathique – cf. section 1.5) et enfin la *sapience* (associée à la « sagesse »). Cette dernière n'est pas intrinsèquement abordable au système pour notre part, en revanche nous essaierons d'aborder ce point dans notre démarche méthodologique dont le partage de la décision collective nous permettra d'aborder la dimension éthique de ce problème complexe.

Historiquement, les constats de la complexité de l'intelligence, quelle que soit sa forme, amènent des auteurs tels que Minsky, Herbert Simon, Chomsky, Mc Carthy par exemple, à considérer que la simulation de la pensée équivaut à la pensée elle-même. Elle en deviendrait ainsi modélisable. C'est ce que la littérature aborde sous le terme de « computationnalisme de traitement symbolique ». Il se dessine alors, plus ou moins successivement, trois approches : descendante, experte et ascendante.

Les approches descendantes sont issues d'un courant qui s'appuie sur des symboles et dont les hypothèses sont cognitivistes. En effet, le symbole est une représentation qui dans ce cas reflète une intentionnalité, une conception qui serait désormais acceptée suite au rejet du behaviorisme selon (Varela, 1996) qui note que :

*l'« hypothèse cognitive prétend que la seule façon de rendre compte de l'intelligence et de l'intentionnalité est de postuler que la cognition consiste à agir sur la base de représentations qui ont une réalité physique sous forme de code symbolique dans un cerveau ou une machine. »*

Cela signifierait donc que ces courants centrent leur réflexion sur la causalité de l'intentionnalité sur des comportements qui en seraient issus. Les symboles représentent alors une réalité physique et sémantique, où l'ordinateur physique accède à la sémantique à travers la syntaxe du langage informatique qui reflète ou se projette parallèlement à cette sémantique. Les ordinateurs donnent ainsi un modèle mécanique de la pensée.

Cette approche descendante s'illustre par exemple avec le *Logic Theorist* de (Newell et Simon, 1956). Le principe est fondé sur l'usage de faits mathématiques comme point de départ afin de réaliser des déductions automatiques logiques par des démonstrations mathématiques. Si l'intelligence est considérée comme une suite de raisonnements logiques, cette approche est très certainement efficace, en revanche elle l'est beaucoup moins pour des tâches d'apprentissage dites « flexibles ». (Newell, Shaw, et Simon, 1959), développent ensuite des approches encore plus génériques qu'ils appellent le *General Problem Solver*. Ce système est basé sur des heuristiques en représentant des faits, des nombres et des concepts sous forme symbolique. Les cas les plus représentatifs de ce type de modélisation sont les systèmes de jeux d'échecs qui utilisent des arbres de décision basés sur ce principe.

Parallèlement, les travaux en sciences cognitives s'intéressent aux processus intellectuels dont une base reprend la théorie de la grammaire universelle (nouvelles réflexions in Chomsky, 2005). Parmi les systèmes qui en dérivent, nous avons par exemple le système ELIZA (Weizenbaum, 1966), qui est l'un des premiers *chatterbots* calés sur un dialogue de séance psychologique fonctionnant sous format de *feedbacks* compatibles, générant une illusion d'échanges. Ainsi, les travaux orientés sur des interactions personne-machine, s'appuyant sur des programmes descendants, tendent à constater que mieux sont les échanges si elles s'appuient sur des phrases courtes comme le montre le programme SHRDLU (Winograd, 1971). Ce système, qui recrée un monde virtuel, est l'un des travaux précurseurs du développement de la réalité virtuelle. Il fonctionne en faisant déplacer des objets géométriques sur une surface plane, où chaque déplacement correspond à des inférences logiques sur le monde virtuel. Il réalise ainsi une décomposition sémantique des données d'entrée. Basés sur ces approches cognitivistes, d'autres programmes de recherche

comme ICOT, au Japon se font remarquer. Ils utilisent notamment le langage PROLOG qui s'appuie sur la logique des prédicats dont l'usage est fait à travers un projet majeur en aérospatiale, afin de permettre un interfaçage de systèmes pour des interlocuteurs non spécialistes. Un projet homologue européen est par exemple le programme Esprit (dans Varela, 1996). Enfin, notons également que c'est cette approche descendante qui fut à la base des premiers systèmes du traitement automatique de la langue.

Les systèmes experts, de manière générale, sont basés sur le même principe que celui des systèmes descendants, mais ils sont associés à des bases de données qui sont analysées pour construire des réponses à une requête ou à un problème. Parmi les exemples illustratifs, le système DENDRAL, en physique-chimie a été développé pour déterminer des composants chimiques dans le domaine de la spectrométrie de masse et en résonance magnétique (Duffield et coll., 1969). Un autre exemple est le système MYCIN, qui est un outil d'aide aux diagnostics utilisé par les médecins, dont les solutions s'appuient sur des bases de données médicales (Shortliffe, 1976). À l'heure actuelle, ce type de système est surtout appliqué dans le milieu de la finance et dans le domaine médical, et plus récemment utilisé pour le développement de jeux tels qu'Akinator<sup>8</sup> ou dans des approches plus modernes (S. Russell et Norvig, 2009).

Les systèmes experts suivent donc la logique mathématique et s'appuient donc sur le « théorème d'incomplétude » de (Gödel, 1931/1971 in Raatikainen, 2013). Ce théorème décrit qu'il n'existe pas de théorie complète, totalement globale donc la formalisation du raisonnement logique dans son ensemble est considérée comme impossible. Par exemple, dans un dictionnaire, les mots ont des définitions. Ces dernières sont elles-mêmes composées de mots. Ainsi, si l'on considère qu'un système est fermé, c'est-à-dire que toute connaissance peut être prédéfinie, alors tout pourrait être abordé avec des règles logiques. Or l'humain, son langage (en perpétuelle modification) et le monde réel sont des systèmes ouverts donc non entièrement modélisables, car dans le cas contraire la description du monde tiendrait dans le dictionnaire.

À partir de ces réflexions, la notion d'apprentissage émerge dans l'avancement de l'intelligence artificielle, considérant que l'IA ne peut se construire si le système informatique lui-même n'expérimente pas le monde et construit sa base de connaissance qu'il rectifie et ajuste à partir de l'expérience faite. Cette réflexion est à la base des motivations de l'apprentissage machine ou *machine learning*.

L'approche ascendante a, quant à elle, émergé avec le courant de la cybernétique (cf. les *Macy Conferences – Cybernetics-Circular Causal and Feedback Mechanisms in Biological and Social Systems*). Son principe est de considérer que le cerveau est constitué d'éléments incarnant des principes logiques qui sont interconnectés par des portes logiques, en transformant ainsi le cerveau en une machine déductive. L'approche se construit donc par une analogie de la perception du monde comme le cerveau le ferait, à savoir des terminaisons nerveuses qui permettent d'acquérir des informations qui sont internalisées pour en former des représentations mentales à travers des neurones qui sont interconnectés. L'augmentation, le renforcement ou la diminution des connexions établies expliqueraient l'apprentissage qui en découle. Cette approche est donc connexionniste.

Parmi les initiateurs de ce courant, nous trouvons (McCulloch et Pitts, 1943) qui ont créé un circuit électronique à la base d'un fonctionnement biologique des neurones (dits neurones MP). Les neurones MP sont construits sur le même principe qu'un vrai neurone et ils fonctionnent sur une excitation positive ou négative de multiples connexions neuronales. La réponse est donc binaire (mais pondérée) et les limites sont comblées par le nombre

---

<sup>8</sup> Akinator.com, (<http://en.akinator.com/>, consulté le 18 novembre 2015), développé par Elokence (<http://elokence.com/fr>, consulté le 18 novembre 2015).

d'entrées, dont le calcul se fait par un intégrateur avec une unité de seuil. Cela donne ainsi une réponse si un certain niveau est atteint à partir de la somme pondérée. Ces neurones sont créés pour traiter des opérations logiques, dont le format sera délaissé, car des circuits plus simples pouvaient donner le même résultat.

Puis (Hebb, 1949/2005) neuropsychologue, suggère l'idée de nouvelles perceptions via le rôle des neurones dans l'apprentissage, basé notamment sur les changements observés dans les connexions synaptiques entre neurones du cerveau. Ces observations furent à la base du Perceptron de (Rosenblatt, 1958) (neurones artificiels qui savent apprendre). En s'appuyant initialement sur l'image, son principe réside dans le fait de soumettre plein de prototypes de formes afin de l'associer à une interprétation, un concept (p. ex. apprendre à reconnaître un A en présentant différentes manières de l'écrire pour que l'occurrence d'un A soit reconnue comme tel). Ainsi les poids d'entrée des neurones étaient automatiquement et graduellement modifiés en fonction des erreurs, puisque ces erreurs étaient remontées au système.

C'est sur ce principe que les réseaux de neurones ont été développés. Ces derniers ont beaucoup été explorés dans les années 60, en particulier dans le domaine militaire. Cependant en 1969 un article explicite une théorie cachée derrière le Perceptron qui freine durant une dizaine d'années les développements des réseaux de neurones. Il a notamment été évoqué que les ordinateurs classiques arrivaient à produire des résultats que ces réseaux n'arrivaient pas à générer. Dans ces années (Bryson, Ho, et Hestenes, 1969) proposent des solutions algorithmiques qui semblent être actuellement à la base de systèmes actuels.

(Minsky et Papert, 1969) avaient d'ailleurs émis des critiques sur les réseaux multicouches pour un gain d'efficacité, mais qui se sont avérés non valables, ce qui donna un regain d'intérêts de ces méthodes à partir des années 1980. En revanche, les problèmes ne semblaient pas pouvoir se résoudre fondamentalement avec l'effet de ces couches dites « cachées ». En revanche cette solution a ouvert toute une gamme des calculs possibles, notamment grâce à la remontée des erreurs à l'ensemble des couches intermédiaires plutôt que seulement à la couche d'entrée. Ce principe est ce qu'on appelle la « rétropropagation » qui donne un effet similaire à la mémoire à court terme. Le principe est très utilisé pour la reconnaissance de formes comme les visages ou pour les véhicules autonomes dans le cadre de la détection d'obstacles, ou encore dans la classification des galaxies en astronomie.

Il existe par ailleurs des réseaux de neurones autonomes dont les couches ne sont pas en série, mais fonctionnent en parallèle, dits également « distribués » (p. ex. le système 3DANN). Il s'agit d'une méthode qui est souvent utilisée sur les vaisseaux spatiaux et en robotique notamment. Ces méthodes rejoignent ainsi conceptuellement les méthodes de *Deep Learning* utilisées actuellement dans de nombreux domaines d'apprentissage dont le traitement automatique de la langue et la robotique (Deng, Yu, et coll., 2014).

Les limites et critiques de l'approche ascendante et descendante se rejoignent sur un manque de recul et sur le fait que les systèmes soient formalisés donc fermés. Sachant que le mode binaire n'est pas le reflet du cerveau, seule l'architecture de ces réseaux est conservée pour passer d'une logique mathématique binaire vers une « logique floue » qui fonctionne en gamme de valeurs. Cette logique est notamment celle qu'on dit « associée au bon sens ». L'exemple typique est le contrôle de l'ajustement des machines à laver en fonction de l'état de son contenu. Les réseaux ont alors un apprentissage réduit au profit du principe d'auto-organisation comme le permettent les réseaux bayésiens. Ces derniers consistent à retrouver des modèles à partir de situations où une grande partie de l'information est déficitaire. Ainsi ces réseaux émettent des hypothèses et des prévisions de décisions de « bon sens » en s'appuyant sur les informations et les relations en fonction de leurs types.

Sur l'ordinateur on simule le réseau de neurones de manière à ce que l'algorithme et réseau neuronal soient exécutés en parallèle et en utilisant les mêmes entrées, il retombe sur les mêmes sorties. Un réseau fondé sur un apprentissage non dirigé va lui définir lui-même son meilleur algorithme à savoir celui qui lui donne les meilleurs résultats. Cette perspective

est donc dans la même pensée que la théorie évolutive de Darwin la sélection naturelle et l'évolution des espèces par mutation génétique. En application aux réseaux de neurones, il s'agirait de développer plusieurs algorithmes pour un même problème et de croiser ceux qui fonctionnent le mieux pour en créer des « mutations ». On appelle ce type d'algorithmes les algorithmes génétiques (p. ex. issus des théories des automates basés sur des expressions régulières gérées par des automates à états finis chez (Conway, 1971 ; Goldberg, 1989) avec une utilisation accrue de ces techniques dans le champ de la robotique, en particulier celle de la robotique bio-inspirée. On voit alors apparaître des notions comme l'« émergence de comportement ». Ces approches bio-inspirées sont d'ailleurs celles qui motivent (Brooks et coll., 1991) pour faire la critique des autres types d'approches qu'ils considèrent comme réductrices et ne seraient pas adaptées pour véritablement modéliser la raison dans une perspective humaine.

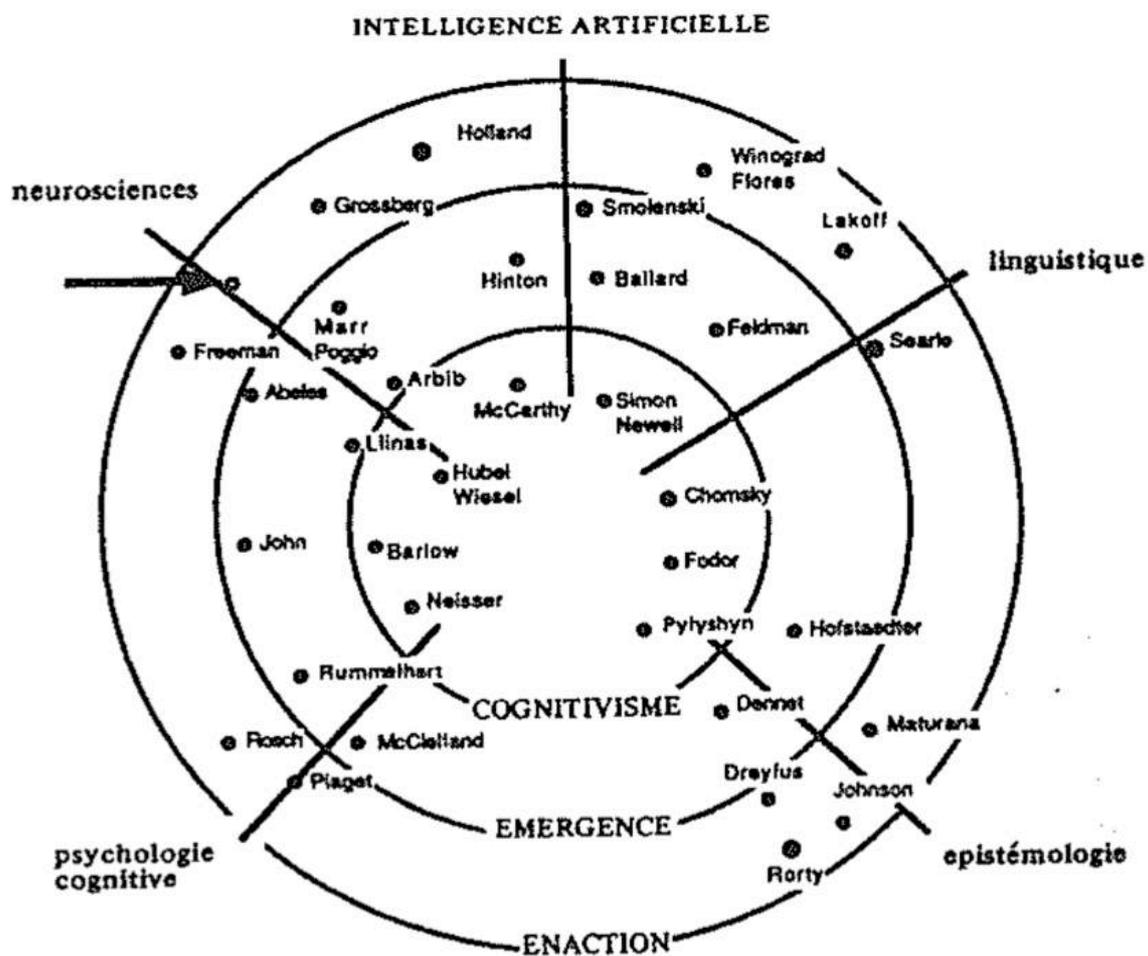


Figure 2 : Cartographie des sciences cognitives avec les champs de réflexion disciplinaire et les principaux auteurs (proposé par Varela 1989/1996 p.212)  
la flèche indique la position de l'auteur

Nous avons passé en revue très brièvement les principaux travaux issus de l'évolution de la cybernétique retraçant les changements notés au sein des sciences et des technologies de la cognition (STC), notamment dans leurs approches de l'intelligence artificielle. Cet aperçu a notamment explicité les deux courants principaux que sont le connexionnisme et le cognitivisme. (Posner, 1993) s'intéresse particulièrement à cette opposition, en proposant des réflexions qui vont bien au-delà de l'application robotique qui nous intéresse. L'aperçu qui est donné ici est loin d'être exhaustif, notamment car ce domaine peut à la fois s'appréhender par des perspectives de neurosciences, de psychologie cognitive, de linguistique ou encore par l'épistémologie. Varela propose une cartographie générale des approches. Elle est représentée Figure 2, afin d'illustrer les auteurs qui sont susceptibles d'intéresser les lecteurs en fonction de leurs spécialités.

Les approches précédemment évoquées ne font pas l'unanimité et elles font l'objet de critiques vis-à-vis des précédents courants (Brooks et coll., 1991). Elles apportent par exemple une perspective très différente dans l'approche de l'enaction (Varela, 1989, 1996) qui rentre dans la continuité des courants phénoménologiques de Husserl et Merleau-Ponty par exemple, en replaçant la question de la corporéité et de l'autonomie (cf. section 1.4.1).

L'enaction émerge également à la fin des années 90. L'auteur (Varela, 1989 in 1996) rappelle que le connexionnisme et le cognitivisme ne sont pas forcément illustratifs d'une scission nette puisque certains partisans de ces approches semblent noter une complémentarité des deux à l'instar de l'approche ascendante vis-à-vis de l'approche descendante, qui en font de même. En revanche, elles suscitent une insatisfaction, car l'évaluation de la cognition par ces approches ne s'appuie que sur l'adéquation ou non des représentations qui sont proposées par rapport au monde extérieur réel. Or Varela, parmi d'autres, suggère que la cognition émane des questions pertinentes qui surviennent au cours des événements de notre vie. Elles ne seraient donc pas prédéfinies, mais elles surviendraient au cours de nos décisions et nos adaptations vis-à-vis de notre situation et notre environnement en fonction de notre expérience passée.

Ces questions sont donc dites « enactées ». Un premier sens du terme est issu de l'anglais alliant une proximité entre les termes « acteurs » et « action ». Le second fait référence au sens « faire émerger » qui est un principe appliqué en phénoménologie notamment. C'est ainsi le courant de cette pensée qui est appelé enaction. L'enaction fait donc émerger ces questions en arrière-plan dont les critères de pertinence sont déduits par l'appréciation qui en est faite par notre « sens commun » qui est toujours contextualisé. Tout en s'intégrant dans la continuité des courants phénoménologiques, le caractère novateur de cette approche est de considérer que le monde n'est pas composé de briques d'éléments prédéfinis, mais qu'il donne suite à une interprétation construite sur la base d'un va-et-vient de savoirs, ou plutôt de boucles de perception et d'action.

Cette boucle s'opère dans un monde dans lequel nous sommes constamment immergés et duquel nous ne pouvons nous extirper. En effet, nous serions alors obligés de figer une situation dans une configuration particulière quasi statique pour pouvoir l'observer (*ceteris paribus* – toutes choses étant égales par ailleurs : il s'agit de la condition dans laquelle l'influence de la variation de la variable explicative sur la variable expliquée est examinée à l'exclusion de tout autre facteur). Elle est à distinguer du constructivisme ou des approches néo-kantiennes biologiques (p. ex. Lorenz et Jospin, 1967/2007). Si nous prenons par exemple la situation de la poule et l'œuf, l'enaction n'essaie pas de définir l'un ou l'autre des extrêmes dans un modèle pour partir de l'un ou de l'autre, mais suggère que les deux éléments sont corrélatifs et c'est cette codétermination qui est l'enjeu de cette approche. C'est donc l'expérience d'événements corrélés qui façonnent le monde, qui ne se construit donc qu'au fil de cette expérience qui fait évoluer les structures qui le composent, ajoutant une dimension de temporalité. En IA, cette approche n'a pas de réelle viabilité à l'heure actuelle, car il en manque des modélisations. Il n'y a pas notamment d'historiques de couplage de

facteurs pertinents et décisifs dans la transformation du système global. Par ailleurs, l'approche ne présente pas encore de stratégies d'évolutions de ces éléments corrélés.

En revanche, au-delà de l'IA, cette approche intéresse le domaine du langage quant à la manière de concevoir le dialogue qui n'est plus alors considéré comme un simple jeu d'échanges d'informations entre un destinataire et un expéditeur comme il est souvent décrit dans les théories de la communication (p. ex. Shannon et Weaver, Jakobson). En effet dans cette perspective il serait possible de considérer l'acte de langage comme l'action conjuguée et mutuelle entre les deux acteurs, qui les remodeleraient et modifieraient également le monde représenté communément auquel ces entités appartiennent.

### **1.3.Des primitives de langage au robot : enjeux du Traitement Automatique des Langues**

D'une part, les origines des langues tentent de montrer que la communication humaine est le fait de sa socialité dont le langage en fait une caractéristique complexe. D'autre part, l'intelligence artificielle se focalise sur la rationalité humaine par sa modélisation et sa représentation. Il existe donc un pont entre la représentativité de ce qu'est l'humain et de la manière dont cet humain est modélisé. Les systèmes de traitement automatiques de la parole, qui dérivent de l'IA, se situent quant à eux entre ces deux perspectives et se composent de la synthèse ainsi que de la reconnaissance de la parole. Ces deux domaines ont en effet pu mettre en évidence des mécanismes en essayant tout d'abord d'appréhender l'ambiguïté du langage naturel, par son expressivité (et non seulement son intelligibilité) ou par la gestion de la spontanéité de la parole humaine dont la dynamique engendre des difficultés de décodage puisqu'elle en change l'information véhiculée, qui n'est pas directement visible dans le message véhiculé. Or cette modulation subtile du langage est justement ce qui porte l'intelligence intrinsèque du comportement humain, qui n'est pas raisonnée, mais acquise dès lors qu'il commence à maîtriser le langage. Plus qu'état de l'art technique des méthodes développées dans ces domaines, les deux sections suivantes tenteront plutôt de réfléchir sur les enjeux actuelles de ces technologies, pour envisager une perspective robotique. De plus, différents aspects de la reconnaissance automatique et la synthèse expressive seront plus transversalement abordés à la suite du chapitre, de manière plus précise sur certaines notions.

#### ***1.3.1. Nouveaux défis de parole expressive en synthèse***

La synthèse de la parole est désormais très réaliste et intelligible, à l'image de la majorité des SVI (serveurs vocaux interactifs). Beaucoup des systèmes actuels sont basés sur le choix et la concaténation d'unités prosodiées naturellement dans des corpus (soit des rushes capturés en live, mais donc incomplets, soit des corpus contenant au minimum tous les diphtonges ou plus longs, dans toutes les variabilités prosodiques nécessaires au système de sélection pour la fabrication du style de voix). Ces derniers sont ainsi en mesure de copier les styles de voix<sup>9</sup> qui caractérisent des rôles sociaux (p. ex. la voix de sorcière, la voix du GPS, la voix de la SNCF... etc.) ou des personnes connus pour les rôles qu'ils incarnent (p. ex. la voix de J. Chirac quand il était président, J. Colombain l'animateur radio). Cette incarnation devient en soit une forme de robotique acoustique telle que nous le voyons dans le film *Her* (Jonze, 2014), où le système d'exploitation d'un ordinateur avec personnalisation vocale donne l'illusion d'un autre existant. Ces voix synthétisées véhiculent des expressivités spécifiques, qui sont en général actées par des personnes qui maîtrisent l'illusion expressive en assimilant empiriquement les éléments prosodiques qui maximisent leurs effets sur les personnes (cf. par exemple Dutoit, 1997 ; ou Erickson, 2005 qui fait une review des études de parole actée). L'expressivité prend également des raccourcis terminologiques sous des termes tels que la « parole émotionnelle » qui ne reflète pas véritablement la nature interactionnelle du langage humain (Audibert et coll., 2006 ; Morency, 2010 ; Sagisaka, Campbell, et Higuchi, 2012). Les techniques initialement développées à l'ATR (*Advanced Telecommunications Research Institute International*), la synthèse dite par corpus (Sagisaka, 1992 ; Sagisaka et coll., 2012) amène à utiliser des techniques telles que la sélection experte d'unités ou encore par calcul statistique de la sélection. Cela permet d'effectuer du « clonage de morceaux du corpus », aboutissant à des réalismes d'apparence importants, mais dont l'expressivité reste statique et limitée d'un point de vue communicationnel.

---

<sup>9</sup> Exemple entreprise Voxygen : <https://www.voxygen.fr/content/actualites>

Le fait d'avoir cette maîtrise est d'ailleurs bénéfique dans de nombreux cas par exemple la prévention d'une aphasie future diagnostiquée, où les patients peuvent conserver un corpus de leur voix pour qu'elle soit utilisée dans une synthèse dès lors qu'ils ne peuvent plus oraliser eux-mêmes (p. ex. Vocalix, Talk'activ). Malgré une qualité moindre de la synthèse sur le versant de l'intelligibilité ou de la naturalité du fait de la taille du corpus, celle-ci est beaucoup plus appréciée par les accompagnants du patient, car ils ont l'impression de reconnaître des éléments caractéristiques du patient, ce qui soutiendrait psychologiquement les aidants dont le soulagement semble également avoir des effets positifs sur le malade, enclenchant ainsi une boucle vertueuse.

En revanche, la synthèse est confrontée à des verrous quant à la maîtrise du changement dynamique qui intervient dans ses voix. En effet, bien que techniquement nous soyons capables de produire différents types de voix et d'expressivité, la voix synthétisée actuelle n'est pas en mesure de changer : 1) la transition dynamique qu'un humain adopte quand il change de rôle (p. ex. à égal contenu lexical, quel changement discrimine un animateur radio qui remercie ses auditeurs, de celui qui remercie sa boulangère après l'achat de son pain, de celui qui remercie des amis ou de la famille qui vient de lui faire un cadeau) ; 2) la dynamique interne qui s'effectue au sein d'un même rôle. Par exemple qu'est-ce qui change dans la parole de l'animateur radio quand il est confronté à différentes situations qui changent son état interne ? Ce dernier pourrait tantôt être joyeux, amusé ou tantôt énervé ce qui est un état statique relevant d'une caractéristique émotionnelle et affective à un instant  $t$ , mais qu'est-ce qui provoque le passage de l'un vers l'autre et comment cette transition s'effectue-t-elle ?

Par ailleurs connaître et véhiculer un certain état interne est une chose, mais que tentons-nous de communiquer à travers ses manifestations (qui font parfois même l'objet d'inhibition qui devient en soit une forme d'expressivité). Sachant que l'humain est toujours socialement impliqué dans ses rôles, quelles sont les subtilités qui permettent tantôt à l'animateur d'être joyeux et amusé avec une intention de communiquer sa joie pour mettre de bonne humeur à ses auditeurs et tantôt à l'animateur d'être joyeux et amusé en entendant une réflexion d'un de ces auditeurs pour lui communiquer le fait qu'il apprécie la boutade ? Dans le cas négatif, s'il est énervé il pourrait potentiellement exposer ouvertement cet énervement (p. ex. pour montrer son positionnement par rapport à un fait d'actualité ou montrer qu'il est dans le même état d'esprit que son auditeur qui est lui-même énervé pour montrer sa connivence), voire au contraire le cacher (p. ex. la situation sociale implique qu'il essaie de rester poli même s'il est énervé). Quelle est donc l'intention de communication qui est portée par l'expressivité des énoncés ? Les conséquences de la justesse et l'adéquation du comportement à cette situation sont certainement ce que reflète la fonction communicative puisque le choix de la forme langagière qu'il va adopter va influencer le ressenti et le comportement de l'autre (p. ex. calmer un auditeur ou dynamiser le débat face au fait d'actualité). Cette intentionnalité reflète ainsi le rôle que la personne adopte et les modifications vont remanier sa relation avec son interlocuteur. Ces dynamiques s'intéressent donc aux transitions portant sur le quand, pourquoi et comment nous changeons les formes « statiques » d'expressivité (dans le sens où une forme peut être inchangée pendant une certaine durée, car la situation nécessite qu'on reste dans ce registre d'expressivité). L'un des défis reste donc de savoir comment reproduire finement l'expressivité statique de manière intelligible, dont les enjeux futurs seraient d'essayer de comprendre comment ces dynamiques d'expressivités fonctionnent.

### 1.3.2. Dynamique langagière en reconnaissance de la parole

Concernant la reconnaissance de la parole expressive, la première difficulté a été de se confronter à l'étendue de la variation et de la variabilité des productions humaines. En effet, la taille du lexique, ainsi que le choix que les locuteurs font pour utiliser tel ou tel élément de vocabulaire posent tout d'abord des verrous technologiques, sauf si elle est contrainte dans un cadre restrictif de commandes (Forsberg, 2003 ; Gaikwad, Gawali, et Yannawar, 2010).

Cette reconnaissance est toujours confrontée à ce problème dans le paradigme de la gestion du langage humain générique. Cependant, la couverture de vocabulaire sera certainement bien plus efficace que l'humain puisque certainement aucun n'a la maîtrise du dictionnaire entier appartenant à une langue et même culture donnée. En revanche, les subtilités et la liberté morphosyntaxique et sémantique que l'humain rajoute un niveau de complexité, un aspect que le domaine de la traduction automatique est peut-être la mieux placée pour en cerner les nuances au regard des situations contrastives entre les différentes langues (Besacier, Barnard, Karpov, et Schultz, 2014). Puis à partir des années 90 à 2000, la capacité calculatoire des machines augmentant, le verrou technique associé à l'étendue du langage tend à disparaître, ce que nous voyons avec l'avènement des *Big Data* (Cambria, Rajagopal, Olsher, et Das, 2013) traité par des techniques de *Deep Learning* (Deng et coll., 2014).

De plus, les méthodologies de récolte de corpus servant à l'apprentissage des systèmes montrent que le focus sur certains domaines particuliers (soit gérer par un vocabulaire restreint) a de meilleurs taux de reconnaissance. Elles supposent une amélioration de la gestion des variabilités de formes, mais peut-être moins au niveau de leur variation selon le mode de récolte des données, spontanées, si elle est inductive, ou fermées, s'il prend en compte des bases de données existantes. Par exemple, si nous prenons une situation d'interactions politiques, le vocabulaire spécifique de ce domaine est bien plus restreint que le langage humain générique donc sa couverture sera meilleure. Grâce au verrou levé avec les *Big Data*, nous pourrions accéder à toute la variabilité possible en prenant en considération un maximum de parlants d'individus du domaine politique augmentant ainsi la variabilité du langage ; la variation de langage intrapersonnelle pourrait être comblée si nous multiplions le nombre d'heures de parole à apprendre de cette même personne, mais dans des situations différentes.

Cependant, la gestion de l'apprentissage utilise des méthodes statistiques et probabilistes sur de grandes données dont le traitement se fait à travers une « boîte noire », où actuellement nous ne pouvons intervenir que sur la qualité de la donnée d'entrée et les manipulations de pondérations sans véritablement pouvoir en comprendre les mécanismes intrinsèques (Sagisaka et coll., 2012). En effet, si nous suivons les lois statistiques et probabilistes classiques, les données viennent à être classifiées en fonction de leurs nombres d'occurrences ou de leur fréquence d'apparition selon des lois de distribution supposée. Par exemple, si une forme langagière est très fréquemment utilisée, nous pouvons avoir une certaine confiance à ce qu'elle appartienne à ce contexte d'interaction donc avoir une plus grande probabilité d'apparition impliquant un apprentissage de ce terme pour qu'il soit reconnu par ce système. Pour le cas inverse ou une forme est rare, elle aura beaucoup moins de poids donc sera peut-être définie comme un cas orphelin dont on déduirait la forme par le contexte par exemple. En revanche, elle est exposée à un plus grand risque d'être mal reconnue, car les variations de cette forme sont beaucoup moins connues que la première.

Or cette forme peut-être une nouvelle occurrence encore peu répandue, mais qui par la suite sera largement utilisée en ayant un fort potentiel communicationnel qui n'intervient qu'une seule fois. Elle peut aussi faire l'objet de micro-expressions qui sont des marqueurs importants pour la reconnaissance de la parole spontanée (Schuller et Batliner, 2013) et qui sont pourtant des événements beaucoup moins notables alors qu'elles sont nombreuses. Ainsi, soit du fait de la rareté, soit du fait de la valeur intrinsèque qu'ils véhiculent, la non-

reconnaissance de ces objets interactionnels spontanés pourrait être problématique notamment si la reconnaissance vocale devient une brique sur laquelle s'appuie un système de dialogue incrémental. Par analogie à une situation de débat, si une personne très discrète parle peu, mais dit des choses pertinentes en réagissant différemment, son intervention peut changer le comportement de ses interactants qui seront peut-être beaucoup plus attentifs à ce qu'elle a dit. Plus encore, un infime changement dans un flux interactionnel chargé peut également avoir un très grand impact sur la dynamique, et même représenter l'essentiel à communiquer alors qu'il semble à première vue noyé dans la masse informative. A contrario, les tours de parole enchaînés entre deux personnes éloquentes peuvent avoir un contenu de message très riche, mais présenter une dynamique interactionnelle plus monotone pouvant amener à une forme d'inattention au bout d'un certain temps. La quantité de données ne peut donc pas être corrélée, ni associée à une distribution similaire à la quantité d'informations qui en découle.

Techniquement parlant, il faudrait certainement pouvoir modéliser spécifiquement ces événements rares afin de ne pas les rendre aveugles au système. Encore faut-il savoir lesquels sont les plus pertinents et comment nous pourrions les reconnaître dans la masse de données, sachant que des indices existent également entre les tours de parole spontanée (Batliner, Kießling, Burger, et Nöth, 2011). En faisant abstraction de la couverture du langage, les systèmes peuvent également s'appuyer sur des formes morphosyntaxiques et sémantiques identiques, dans une situation d'interaction restreinte (soit à travers un apprentissage dit spécialisé).

Dans ce contexte restreint, il est possible de spécifier davantage certaines conditions particulières : âge, genre, modifications physiologiques, accents, hyperarticulation etc. Par exemple dans le cas de la personne âgée, les modifications physiologiques engendrent des modifications articulatoires et morphosyntaxiques (Aman et coll., 2012 ; Baba, Yoshizawa, Yamada, Lee, et Shikano, 2001). Dans ce cas le vocabulaire et la situation restent inchangés, et seule la caractéristique du sujet est différente. L'apprentissage spécifique à la parole âgée améliore alors les taux de reconnaissance du système. Cependant si ce même sujet est soumis à des conditions spécifiques qui changent son état interne (p. ex. situation de détresse induite), ce dernier va moduler des dimensions prosodiques qui vont faire chuter les taux d'efficacité de la reconnaissance vocale (Aman, Vacher, Rossato, et Portet, 2013). Les paramètres, caractérisant le sujet et les fluctuations internes de ce même sujet, changent tous deux des paramètres prosodiques, par exemple la qualité de voix, ce qui engendrera une difficulté d'analyse intrinsèque à l'écologie de la situation et au sujet. Cette interaction située va ainsi sans arrêt mettre au défi la robustesse de ces systèmes, d'où peut-être l'intérêt de reconnaître les dimensions de l'interaction, qui sont susceptibles de modifier l'orientation de la communication. Elles seraient potentiellement des points de manipulation et de changement relationnel plus fort. Nous voyons alors que la prosodie semble alors être un élément important pour les systèmes de traitement automatiques. À l'heure actuelle, l'intégration du modèle prosodique n'est encore que très balbutiante dans les systèmes eux-mêmes (Sagisaka et coll., 2012 ; Sagisaka et Greenberg, 2015) et poserait surtout le problème de savoir à quel niveau du système de la reconnaissance vocale l'intégrer, car ce dernier traite successivement un modèle acoustique, un dictionnaire de lexique et un modèle de langage. Devrait-elle être successivement appliquée à chacun des modèles, voire être une brique à part entière, mais dans ce cas dans quel niveau de couche l'intégrer?

La tendance serait de dire qu'elle est à la fin, mais cela suggère d'adopter la vision d'une grammaire générative. D'autre part, l'inclusion de la prosodie dans le système de reconnaissance vocale suggère qu'elle est intrinsèquement vocale. Or la parole, en interaction face à face, étant un phénomène multimodal, la prosodie et l'expressivité qu'elle véhicule serait beaucoup transversale et sous-jacente à toute production. Il s'agit là de la perspective que nous adopterons.

### 1.3.3. Des théories à l'informatique des affects sociaux

Suite aux réflexions posées sur les enjeux des systèmes de reconnaissance de la parole spontanée et expressive, nous voyons que l'expressivité est souvent exprimée à travers une parole dite émotionnelle et socio-affective dont la prosodie en est le porteur. Ainsi, avant de prendre toute décision de modélisation basée sur une théorie quelconque, la présente section explorera brièvement la valeur informative cachée derrière cette terminologie. Nous allons tenter d'introduire la manière dont ces notions se manifestent dans la parole notamment à travers la récente discipline de l'*Affective Computing* ou les sciences affectives informatisées.

L'*Affective Computing*, s'est développée pour intégrer la dimension émotionnelle et affective pour permettre d'appréhender une part plus importante de la naturalité du langage humain dans le cadre de l'informatique. Elle converge actuellement avec le traitement des signaux sociaux, en faisant émerger un autre courant qui est celui du *social signal processing*. Ces deux approches augmentent ainsi le cadre de réflexion d'une interaction face à face dans lesquelles s'inscrivent de multiples technologies (agents virtuels ou ACA, robots, avatars, immersion 3D, réalité virtuelle etc.). Ces approches nous permettront ainsi de voir quels sont les objets qui pourraient servir de primitives de communication à travers un robot. L'intérêt de cette section sera de pouvoir appréhender la manière dont les technologies qui dépendent de ces objets interactionnels modélisent les fonctions cognitives et affectives internes de l'humain, et quelles valeurs et fonctions véhiculent ces formes.

Ainsi une première distinction est celle entre les termes « émotions » et « affects / états affectifs » qui dans la littérature sont encore relativement confondus. Cependant, cette terminologie semble avoir suscité des discussions lors de la conférence LREC 2006 par exemple. Dès lors, les termes « états affectifs » semblaient plus appropriés pour désigner « les émotions, les sentiments, les attitudes, les humeurs et les attitudes interpersonnelles d'une personne » (Campbell et coll., 2006 dans Vanpé, 2011). Une première définition des affects sociaux (Aubergé, 2002) est reprise plus explicitement dans les travaux de (Shochi, 2008). Cette définition propose une distinction entre le contrôle automatique involontaire (émotions) et le contrôle automatique volontaire (affects sociaux ou attitudes) de ces phénomènes. Cette distinction permet la différence entre les phénomènes de la communication, référant aux premiers comportements, des phénomènes de langage dépendant des seconds. Dans les faits, l'état affectif est complexe à appréhender, car tous ces éléments sont mélangés à chaque instant d'une interaction et évoluent également dynamiquement. D'ailleurs ils sembleraient que les déclencheurs soient également multiples. Ce terme « affect » avait également été proposé par (Juslin et Scherer, 2005) dans lequel ils placent en plus le « stress » et les « traits affectifs de la personnalité ». De plus, ces éléments sont contrastés avec les « épisodes émotionnels » par rapport à une dimension de durée qui suggère que ceux-ci soient plus occasionnels et répondraient également à des événements spécifiques qui faisaient donc office de stimuli engendrant une réponse relativement immédiate (Lorenz et Jospin, 1965/2007 ; Scherer et Sangsue, 1996 in Kirouac, 2004).

Dans le cadre de cette thèse, nous emploierons d'ailleurs le terme de *Feeling of Thinking* (ou *FoT*) qui est un néologisme, mot-valise introduit par Aubergé (Loyau et Aubergé, 2006) qui dérive de la notion de *Feeling of Knowing* (ou *FoK*) de (Hart, 1965). Rattaché aux phénomènes de mémorisation, le *FoK* véhicule l'idée d'une information stockée, mais non accessible sur le moment alors qu'elle est reconnue, donnant lieu par exemple aux phénomènes du « mot sur le bout de la langue ». À partir de ce sentiment de « sentir qu'on sait » une généralisation plus large est illustrée par le *FoT*, introduit dans le cadre de l'analyse multimodale (voix, parole, langage, expressions faciales, gestualité) effectué par (Loyau, Aubergé, et Vanpé, 2006 ; Vanpé et Aubergé, 2006) sur le corpus E-Wiz du corpus de productions expressives spontanées (Aubergé, Audibert, et Rilliard, 2004b). Dans cette analyse, les sujets à l'origine de ces productions évoquaient via un « sentiment de penser », leurs états affectifs involontaires ou sociaux (émotions, attitudes, intentions) ainsi que leurs états mentaux (p. ex. l'anticipation ou la pensée qui accompagne un dire ou la concentration). Nous verrons d'ailleurs par la suite que cette notion n'est pas seulement un terme arrangeant, mais qu'elle est intimement liée à la notion de *backchannel* dont la rythmicité suppose l'existence d'une prosodie multimodale.

Ici nous ne décrivons pas les différents modèles d'émotions dont l'évolution historique est très intéressante, mais dont il existe des états de l'art bien plus complets basés sur les très nombreux travaux de psychologie (p. ex. *Handbook of Emotions* de Lewis, Haviland-Jones, et Barrett, 2010 ou encore le *Handbook of Affective Sciences* de Davidson, Sherer, et Goldsmith, 2009). Néanmoins, la conception générale des approches est rappelée rapidement, car les émotions et les affects sont porteurs d'informations dont l'humain ne peut se défaire dans l'interaction et que leurs fluctuations peuvent expliquer les transitions ou les changements de l'état relationnel de l'humain avec son interlocuteur, ce que par exemple (Bowlby et Ainsworth, 1991/2013) ont présenté dans leurs théories de l'attachement (cf. section 1.4.3).

Cela signifierait alors que les modélisations et les théories des émotions pourraient potentiellement donner des indices sur les dimensions possibles de la glu socio-affective. De manière générale, nous voyons qu'en psychologie où tous ces termes affectifs seront aussi associés au terme « émotions », il existe quatre grandes approches, représentées par les différentes zones du schéma Figure 3 ci-dessous, certaines basées sur une représentation plutôt catégorielle (émotions de base), d'autres, plutôt dimensionnelle (p. ex. Russell) et enfin celles dans une conception hybride positionnant des catégories le long d'axe dimensionnel (p. ex. le modèle *circumplex* de Plutchik).

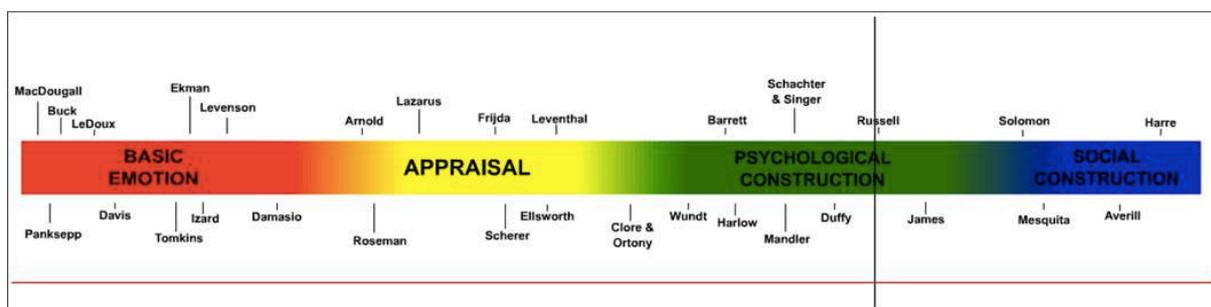


Figure 3 : Continuum des quatre grandes approches des théories des émotions avec les contributeurs principaux (proposition de Barrett-Feldman et Gross, 2011)

Les approches de *basic emotions* sont celles développées à l'instar du modèle darwinien dont dérive les *Big Six* (Cornelius, 2000), que sont les émotions de base (tristesse, joie, peur, colère, dégoût, surprise) définie par (Ekman et Friesen, 1969 ; Ekman, 1992) dont il propose une extension à seize dans les années 90. Leur nombre serait ainsi variable, mais limité et elles seraient universellement partagés et considérés comme des produits de l'évolution. Les auteurs seraient alors plutôt partisans d'une thèse innéiste des émotions.

À l'opposé, l'approche *social construction* considère que les émotions dépendent des contextes et des situations dans lesquelles elles apparaissent, car sinon elles seraient vides de sens, sans possibilité d'interprétation. Il s'agirait donc de formes issues de processus de construction sociale. Ces courants sont notamment initiés par des auteurs tels que (Averill, 1980) en s'appuyant sur les différences interculturelles, notamment par leurs manières lexicales de décrire les émotions. Elles montrent notamment des différences de finesse entre les cultures dotées de nombreux termes pour exprimer les émotions (quelques centaines) par rapport à d'autres qui ont un vocabulaire restreint (moins d'une dizaine).

Entre les deux, (Gross et Barrett, 2011) proposent qu'il existe des théories dépendant plutôt d'une construction psychologique continue. Ces dernières supposeraient que les émotions ne représentent pas des états mentaux spécifiques, uniques dans leur forme, leur fonction ou leur cause comme il serait possible de le concevoir pour la cognition ou la perception. Il n'y aurait donc pas de mécanisme dédié à produire uniquement des émotions. Ainsi tous les états mentaux émergeraient d'une construction continue dont les ingrédients ne relèvent pas seulement de la sphère émotionnelle. Cela n'empêche pas ces théories à définir les émotions dans des catégories, mais qui ne seraient pas des sommes de leurs parties, liées entre elles, intégrant alors des approches dimensionnelles. Par elles, nous avons l'exemple de (Russell, 2003) qui pose notamment la dimension de l'*arousal* (ou activation, niveau d'alerte) ainsi que la *valence* (positif-négatif rattaché à la notion d'attractivité/répulsion) qui sont celles qui sont très largement utilisées par la communauté de l'*Affective Computing*. À ces deux dimensions se rajoute la *potency-control*, en particulier très utilisée dans l'analyse vocale, désignant le niveau de contrôle exercé dans la régulation émotionnelle (Matsumoto, 2006), qui serait d'ailleurs dépendante de la culture. (Scherer, Schorr, et Johnstone, 2001) rapportent qu'à égal *arousal*, mais *valence* variable, il était difficile de discriminer les éléments socio-affectifs entre eux, par ces seules dimensions, ce que (Fontaine, Scherer, Roesch, et Ellsworth, 2007) tentent de combler par la dimension de *potency-control* qui est souvent dit être négligé ou controversé, de même que celle d'une quatrième dimension d'imprédictibilité. Cette *potency* est décrite comme faisant référence au ressenti de l'individu vis-à-vis d'un pouvoir ou d'un contrôle qu'il exercerait sur l'événement suscité (Osgood, May, et Miron, 1975 ; Wundt, 1907). Il s'agit de la dimension de dominance que nous considérons comme indépendante à celle de l'altruisme.

Sur l'ensemble des processus agissant sur les émotions, nous pourrions déduire des différentes théories existantes, que celles-ci peuvent principalement s'appuyer sur deux mécanismes : la simulation et l'inhibition (Gross, 1998). Concernant cet aspect de régulation, les auteurs tels que (James, 1884) puis (Damasio, 2010 et 1999), qui sont respectivement sur des frontières de deux approches, ont en commun une perspective physiologique de l'émotion. Ils considèrent tous deux que c'est l'expérience d'une modification physiologique qui est déclencheur de l'émotion. Ils vont ainsi à l'encontre des approches plus cognitivistes qui suggèrent que les manifestations physiologiques ne sont que la résultante de l'évaluation cognitive produisant une émotion qui déclenche ces manifestations. Cette distinction aurait une significativité pour la rythmicité que ces phénomènes pourraient engendrer dans l'interaction. Elle aurait également des conséquences sur la notion de corporéité associée à la robotique, notamment dans le cadre de la téléprésence où l'humain est amputé partiellement de ces aspects puisque son expérience passe à travers le corps du robot qui devient également dans une certaine mesure son propre corps. Cependant, nous ne nous positionnerons sur

aucune de ces théories à cette étape de cette thèse, car notre démarche se veut hypothético-déductive sur ces aspects.

Enfin, l'*appraisal* relevant d'approche cognitive est actuellement l'un des courants dominants dans les théories des émotions. Il considère que les émotions sont déterminées à partir d'une évaluation cognitive d'une situation par l'humain, p. ex. nous sommes en colère, car nous sommes offensés ou nous sommes tristes, car nous venons de perdre un proche. (Arnold, 1960) mettait en lien cet *appraisal* à un mouvement « vers » ou « contre » son interactant. Il suggérait qu'une évaluation positive engendrait des émotions positives avec des comportements de rapprochement, alors qu'inversement une évaluation négative suggérait un éloignement avec des productions d'émotions négatives. Cette évaluation dite « subjective » dépendrait du vécu et des connaissances socioculturelles de la personne, changeant l'interprétation qu'auraient un/des individu(s) face à une même situation, en produisant chacun différentes émotions (Lazarus, 1991). (Frijda, Kuipers, et Ter Schure, 1989) dans la même lignée, suggère par ailleurs que cette émotion reflète la motivation à l'origine des actions ou des expressions d'une personne. Cet humain, par l'analyse de l'environnement, manifesterait des traces d'anticipation et de préparation à l'action. Actuellement, ce sont les travaux de (Scherer, 1984 à 2001) qui s'inscrivent dans cette même lignée, qui sont pris comme référence. Son approche multidimensionnelle considère l'émotion comme un sous-système du processus d'interaction dont les fonctions seraient adaptatives. Ainsi à chaque modification de l'environnement externe ou interne de l'humain, un « composant d'émotions » permettrait de s'adapter à des changements de différents niveaux, ces adaptations pouvant se faire plus ou moins simultanément. Ce *component process model* serait alors constitué de cinq composants : cognitive, d'expression motrice, physiologique, motivationnelle et de sentiment subjectif qui sont évaluées par étapes temporellement successives dans ce qu'il appelle les SEC, *Stimulus Evaluation Checks*. Sans rentrer dans les détails de son modèle, ce dernier semble attirer l'attention de la communauté parole, car son approche, outre les mécanismes cognitifs, aborde la dimension expressive des émotions à travers ses manifestations vocales. Cette expressivité aurait alors des facteurs ayant différents effets, dont ceux qu'il appelle *push* (issus de l'externalisation de manifestations biologiques) et *pull* (manifestations issues des normes socioculturelles), qui apparaîtraient particulièrement au niveau de la prosodie vocale.

Au final, le développement des modèles s'appuie sur des éléments expressifs variés permettant de mettre en évidence plus ou moins facilement certains facteurs selon ce qui est observable. L'*Affective Computing* chercherait donc à savoir quelle est la nature des objets expressifs et dans quels mécanismes ils interviennent. Bien que ces processus soient nombreux, la section suivante fera un bref focus sur le phénomène d'attachement.

### 1.3.3.1. De l'attachement aux marqueurs comportementaux

Les émotions, au sens de la psychologie, sont étudiées à travers de multiples phénomènes tels que les défenses psychologiques (Freud, 1926/1959/2014), le stress ou la gestion de la menace (Lazarus, 1966), l'autorégulation (Mischel, Shoda, et Rodriguez, 1989) ou encore l'attachement. Ce dernier point va nous intéresser plus particulièrement dans le cadre de la construction relationnelle. Ainsi, les origines de la théorie de l'attachement remontent à la fin des années 60. Dans les faits, les observations comme le phénomène d'hospitalisme relevé par (Bowlby, Robertson, et Rosenbluth, 1952), les études de bébés primates isolés<sup>10</sup> de (Harlow, 1958) ou encore le phénomène « d'empreintes » observé chez les oies par (Lorenz, 1935/1985) sont également des études qui anticipent le phénomène d'attachement adaptatif par lequel passe l'enfant pour avoir une chance de survie. Il manifeste ainsi un besoin de forte proximité face une entité qui prend soin de lui, en situation de stress

---

<sup>10</sup> Vidéo illustrative de l'expérience de Harlow : [https://www.youtube.com/watch?v=\\_O60TYAIgC4](https://www.youtube.com/watch?v=_O60TYAIgC4)

ou de peur. La théorie de l'attachement découle de ces observations et elle est formalisée à travers le travail de (Ainsworth et Bowlby, 1991). Cette théorie est notamment explicitée dans la relation d'une mère avec son enfant dans un contexte de séparation précoce durant la Seconde Guerre mondiale, en se focalisant sur la perte, le deuil et la séparation.

Juste avant les années 70, Bowlby met en évidence une dimension de *care* ou de « prendre soin » dans son « modèle interne opérant ». Il suppose ainsi que chez l'enfant, il existe une dimension interactive qui naîtrait des attentes et des anticipations de qualités d'interactions qu'il a envers son « soignant » (soit une personne qui prend soin de lui au quotidien). Le système d'attachement est donc orienté pour maintenir l'accessibilité et la réactivité du *caregiver*. Ce phénomène dépendrait d'une certaine continuité mémorisée, le soin à long terme ne prodiguant pas une dépendance, mais une assurance suffisante chez l'enfant, lui permettant de s'ouvrir au monde. Dans ce système de comportements innés, l'attachement serait activé pour engendrer un rapprochement dans le cadre d'un contexte d'insécurité, qui quand il est comblé s'arrête au profit de mécanismes tels que l'exploration.

Dans la continuité de ces réflexions, (Ainsworth et Wittig, 1969) propose un prolongement expérimental de cette théorie en mettant en avant le concept de *Strange Situation*. Cette situation serait ainsi considérée comme un contexte d'évaluation de la relation là où l'attachement permettrait de comprendre le développement relationnel de l'enfant. Elle met ainsi en évidence trois variables : la proximité, l'évitement et la résistance. Ces variables émergeraient des comportements de l'enfant lorsqu'il retrouve le parent ou le *caregiver*, après avoir été séparé de lui dans des contextes variés. Ainsworth suppose par l'observation de ces variables qu'il existerait une « base de sécurité » qui dépendrait d'une entité adulte accessible qui a la capacité de compréhension des besoins et des inquiétudes de l'enfant. Cette base proviendrait ainsi d'un lien prévisible, stable et sécurisant. De plus, en plus des trois schèmes d'attachement, Ainsworth avait mis en évidence une variable « inclassable » que son élève Mary Main avait définie par la suite dans un schème « désorganisé » en se basant sur les comportements contradictoires ou de figement en explorant la relation à travers des contextes de maltraitance et de violence. C'est également elle qui développa la méthode de l'*AAI*, (*Adulte Attachment Interview*) pour étendre le travail de Bowlby qui avait noté l'existence d'attachements multiples ou encore celui de Ainsworth qui posa la question de la continuité des comportements d'attachement à travers le développement évolutif à différents âges dans le processus de socialisation et d'expression des émotions.

Les interprétations des manifestations des « émotions » ou plus généralement des affects semblent converger sur le fait qu'ils émanent de mécanismes sociaux. Ainsi, dans le cadre des interactions, l'informatique et la linguiste tend vers une pratique de *social signal processing*, à l'instar du réseau européen d'excellence HUMAINE<sup>11</sup>, la mobilisation au *workshop WASSS 2013*<sup>12</sup>, satellite de la conférence *Interspeech 2013*, ou encore la session spéciale *ASSS (Affective Social Speech Signals)* de la conférence *Speech Prosody 7* en 2014<sup>13</sup>. À noter qu'un autre courant dénommé *social signals* se développe également dans le cadre du *data mining* et de la web sémantique, recherchant ce type de traces non pas dans les interactions face à face, mais dans les interactions à travers les réseaux sociaux dans le cadre des *SEO*<sup>14</sup> (*Search Engine Optimization*). Pour notre part, nous resterons dans le cadre de l'interaction face à face.

Le traitement automatique des signaux sociaux ou *SSP* se développe à travers la notion d'intelligence sociale (Thorndike, 1920 ; Argyle, 1967/1994 ; Albrecht, 2006). Il concerne aussi bien des travaux de reconnaissance que de production de signaux et de comportements

---

<sup>11</sup> SSPnet du réseau HUMAINE : <http://sspnet.eu>

<sup>12</sup> WASSS 2013 : <http://wasss-2013.imag.fr>

<sup>13</sup> SP7 2014 : <http://fastnet.netsoc.ie/sp7/>

<sup>14</sup> *Social signals* dans les SEO : <https://moz.com/blog/your-guide-to-social-signals-for-seo>

sociaux. (Vinciarelli, Pantic, et Boulard, 2009) considèrent qu'il existe une différence de temporalité entre « signaux » et « comportements », les premiers étant très courts de l'ordre de la milliseconde (p. ex. les effets de *mirroring*), alors que les seconds sont associés à des comportements plus longs, allant de la seconde (p. ex. signes d'acquiescement, d'accord), à la minute (p. ex. formes de politesses), jusqu'à l'heure ou aux jours (p. ex. empathie). Ces signaux incluraient ainsi des éléments paralinguistiques (Schuller et Batliner, 2013) également dits non verbaux, mais nous préférons le terme de « non lexicaux », qui sont des objets observables au niveau des expressions faciales, des postures, de la gestualité ou des événements vocaux tels que des rires, autant d'indices qui sont utiles pour ce que (Pentland, 2005) appelle le *socially aware computing*.

En interaction face à face, c'est le langage articulé à travers la parole qui est le média principal de la communication et qui est une activité multimodale dont la pragmatique a essayé d'explorer les influences. Plus généralement, dans les années 1970, l'analyse conversationnelle a commencé à s'intéresser à la gestualité et aux mouvements du corps (Charles, 1981 ; Heath, 1986), fondé sur l'observation d'activités sociales spontanément produites en situation, avec la prise en compte contextuel dans son ensemble. Cette approche multimodale s'étend également aux mimiques faciales et aux regards (Rossano, 2012), en prenant en compte le corps entier dans ses postures (Goodwin, 2003 ; Schmitt, 2005 in Mondada, 2012) ou dans l'espace interactionnel en définissant les positions qu'il occupe (Hausendorf, Mondada et Schmitt in *Body – Language – Communication*, pp. 580-589, 2013 ; Mondada, 2009) intégrant une dimension dynamique dans l'observation de la gestualité, des mouvements du corps et des déplacements (Schlegoff, 1984 ; Mondada, 2005, 2007 ; Relieu, 1999 ; Haddington, Mondada, et Nevile, 2013). Un véhicule extrêmement riche se trouve également au niveau des expressions faciales (Ekman et Friesen, 1975), notamment à travers le modèle du *FACS (Facial Action Coding System)*. Ces aspects comportementaux et kinesthésiques ont également fait l'objet d'analyse par des courants comme *le mind-reading* (Burlingame, 1891 ; Rauscher, 2002), le mentalisme (Cassidy, 1995) ou la synergologie (Turchet, 2000) dans l'optique de lier les comportements physiques à des interprétations de faits cognitifs. À ces éléments, s'ajoute la voix, dans sa phonation et son esthétique (Kreiman et Sidtis, 2011), mais également dans son contrôle intonatif ou prosodique.

Dans cette approche SSP, selon (Vinciarelli et coll., 2009), une certaine intention communicationnelle serait ainsi rythmée par divers objets de natures différentes :

- affectives, attitudinales et cognitives,
- emblématiques (p. ex. gestualités ou expressions culturellement codées,
- manipulateurs (p. ex. actions sur un objet ou soi-même tels que pincement de lèvres ou grattements),
- illustratives (p. ex. des éléments accompagnant des faits langagiers tels que le pointage ou des haussements de sourcils,
- régulatrices (également appelés des « médiateurs conversationnels » véhiculés par exemple des échanges de regards, des hochements de tête ou des sourires).

Des auteurs tels que (Ambady et Rosenthal, 1992 ; Knapp, Hall, et Horgan, 2013 ; Mehrabian et Ferris, 1967) affirmaient que la part d'information constituant la perception sociale de l'autre et de la situation d'interaction était beaucoup plus importante dans les événements dits non verbaux par rapport aux verbaux. Pour (Mehrabian et Ferris, 1967), les événements verbaux ne représentaient d'ailleurs qu'environ 7% de cette perception sociale. Tout comme pour les émotions, nous ne choisirons pas ici d'adopter une taxonomie spécifique. En revanche, la continuité communicative permettant de réguler la relation sociale à travers des traces d'interactions multimodales est très riche. Elle serait également liée à une temporalité donnée, ainsi qu'à une rythmicité intrinsèque. Dans la section suivante, nous allons nous intéresser à quelques phénomènes où ces signaux sociaux apparaissent

préférentiellement pour augmenter l'empathie entre deux interactants, bien que ces mécanismes ne soient pas exclusifs au processus de socialisation.

Voici une proposition récapitulant les codes et les fonctions que (Vinciarelli et coll., 2009) proposent dans leur état de l'art des signaux sociaux :

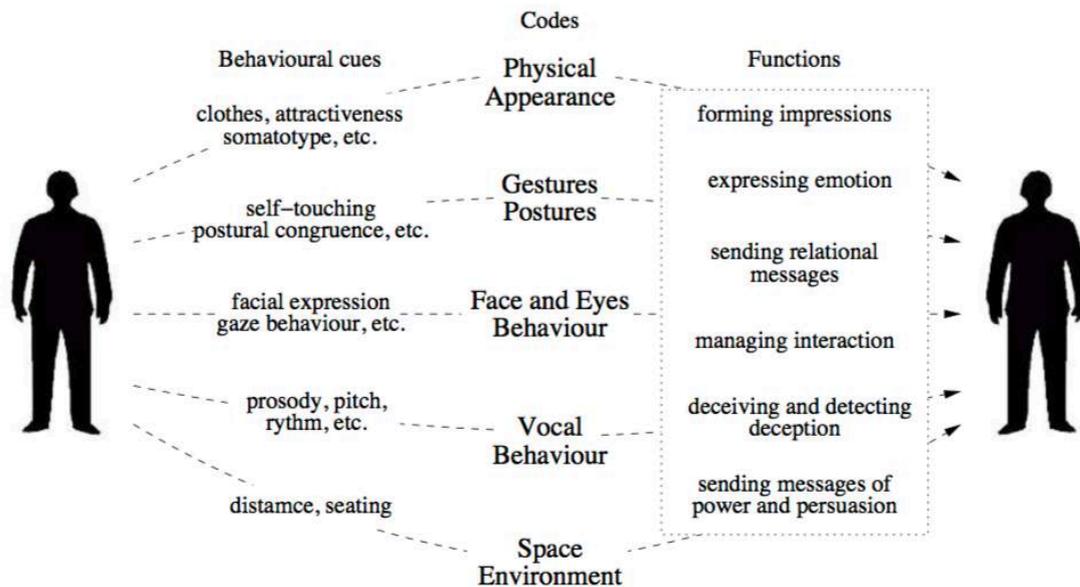


Figure 4 : Nature des comportements non verbaux, codes et fonctions influençant la perception sociale de l'humain vis-à-vis d'un autre (d'après Vinciarelli et coll., 2008)

### 1.3.3.2. Du mimétisme aux comportements d'imitation

Le mimétisme et l'imitation sont deux mécanismes distincts qui sont souvent proposés comme des processus de base dans la construction sociale, caractérisés par de nombreuses manifestations de signaux sociaux. De plus, ces notions sont fortement exploitées dans le cadre de la robotique sociale.

(Baudonnière, 1997) définit ainsi le mimétisme comme un phénomène qui recouvre toute situation où il existe une ressemblance entre une chose et un animal ou entre deux animaux soit deux espèces différentes. Il aurait un pouvoir adaptatif important (si utilisé au bon moment et au bon endroit). Il s'agirait ainsi d'un processus dynamique continu et évolutif qui serait à la base de mutation hasardeuse. En revanche, il serait trop souvent assimilé à certains comportements imitatifs.

En effet, selon Baudonnière, le comportement imitatif serait différent selon 3 raisons par rapport au mimétisme :

1. il implique une intentionnalité, même non consciente
2. il est sélectif : on n'imité pas n'importe quoi, qui et quand
3. l'imitation serait propre à l'homme alors que le mimétisme est universel aux espèces

Les deux phénomènes seraient ainsi associés à la dimension de socialisation, de cohésion sociale et de transmission de la culture qui pour les auteurs sont les fondements de l'humanisation. L'apprentissage par observation serait alors un entre-deux, une étape intermédiaire qui se situerait entre mimétisme et imitation.

Le mimétisme qu'il soit intra ou interspèce semble être passif ou actif, passant soit par une apparence, un trait biologique ou un comportement. Celui qui est observé chez l'animal par rapport à son environnement illustrerait un cas particulier de l'ensemble des mécanismes d'adaptation intervenant dans la co-sélection. Il s'agirait donc d'une stratégie de survie parmi d'autres, dont le point de vue diffère selon qu'on se place du côté de prédateur ou de la proie.

Parmi les mécanismes, le camouflage s'appuie par exemple sur une similitude visuelle. Du côté de la proie, c'est un effet de hasard qui jouera sur la mutation génétique dans le cadre de sa sélection naturelle en utilisant un leurre (qui peut aussi être utilisé par le prédateur pour attirer sa proie). En revanche, ce mimétisme peut aussi faire l'objet de rencontre fortuite entre des caractéristiques similaires retrouvées entre un organisme et son environnement et on parlera alors de mécanisme de coadaptation.

Entre des animaux, le mimétisme prendrait deux formes :

- un mimétisme intelligent qui est issu d'une intention consciente pour bénéficier de l'exemple d'un autre, qui s'observe surtout dans le cas d'un comportement atypique nouveau. Ce dernier a souvent été confondu avec le comportement d'imitation.
- et mimétisme instinctif qui est une tendance à copier des actes simples, une notion qui tente à être remplacé par la notion de « facilitation sociale ».

D'autre part, le cas où un comportement n'existe pas initialement, acquis graduellement par une première apparition qui se distribue ensuite de manière radiale, suggérerait une influence sociale du mimétisme. De même, les déclenchements de la ponte des pigeons, les constructions de nids ou les permanences des routes migratoires sont autant d'éléments qui illustreraient cette dimension sociale en éthologie. Ces comportements peuvent être individuels ou collectifs (p. ex. tendance à manger davantage en groupe qu'en étant isolé, avec plus ou moins d'effet de compétition). De plus, le mimétisme semblerait même se propager au-delà du vivant, car si l'on introduit par exemple des poules mécaniques avec un comportement propre dans une basse-cour, les vraies poules auraient tendance à sélectionner les mêmes grains que les poules mécaniques. Ainsi les comportements innovants seraient imités plus souvent, d'autant plus s'ils émanent d'un individu dominant même si ce processus d'apprentissage par observation ne relève d'aucune intentionnalité pédagogique.

Chez l'humain, les effets de mimétisme semblent d'autant plus apparaître lorsque la charge émotionnelle est importante, en particulier au sein des foules. Par ailleurs, (Chartrand et Bargh, 1999) ont également mis en évidence ce qu'ils appellent le *chameleon effect*, défini aussi sous les termes *unintentional mirroring*. Ce phénomène s'observerait dans le cadre d'une interaction fluide et réussie, caractérisée par une forte empathie où des comportements (vocales ou physiques) mimétiques viennent à apparaître et sont partagés inconsciemment par les interactants. Les auteurs ont ainsi proposé trois séries successives d'expériences pour illustrer ce phénomène :

- Premier cas (mimétisme en production) : dans une dyade positionnée en face à face, l'un des interactants doit faire preuve de maniérisme (sourire et toucher de visage pour l'un ou balancement de pieds pour l'autre). Ces comportements auraient engendré une augmentation des comportements mimétiques de l'autre (20%), mais encore plus dans le cas des balancements de jambes (50%).
- Deuxième cas (mimétisme en perception) : les mêmes sujets passent une autre entrevue avec un expérimentateur qui pour la moitié reste dans une posture neutre sans maniérisme. Dans le second groupe, il copie les maniérismes des sujets dès qu'il en voit une (p. ex. croisement de jambe et passage de main dans les cheveux). Puis il est demandé aux sujets de noter la fluidité de l'interaction avec l'expérimentateur et placer sur une échelle leur appréciation de l'expérimentateur. Les personnes ayant été imitées disent avoir beaucoup plus apprécié l'entretien et l'expérimentateur.
- Troisième cas : un questionnaire est proposé à certains de ces sujets pour mesurer leur empathie envers des personnes. Les sujets ayant été caractérisés comme les plus ouverts aux autres par leurs réponses au questionnaire étaient ceux qui montraient également une plus forte augmentation de mimétisme constaté dans le premier cas d'expérience.

Concernant l'imitation, pour (Baudonnière, 1997), ce comportement réside dans le fait de reproduire ce qu'un autre être a déjà fait avant et dont la reproduction peut être immédiate, décalée ou différée. Il nécessite donc que ce comportement imité soit décodé et interprété pour en proposer une reproduction semblable. L'imitation serait alors souvent associée à différentes notions telles que les mécanismes d'imprégnation ou empreinte, la contagion émotionnelle ou encore l'imitation néonatale (Meltzoff et Prinz, 2002).

En effet, dans le cadre de la psychologie du développement (Wallon, 1934/2015) a tout d'abord posé la question de la capacité d'imitation de certains gestes (p. ex. la protrusion de la langue) chez le nouveau-né, ce qui signifierait le bébé aurait très tôt une capacité de représentation. (Meltzoff et Moore, 2005) ont pensé que ce comportement était en fait une propriété sélective de certains comportements. Cependant (Jacobson, 1979) a montré que la protrusion de langue observée par Wallon relevait plutôt d'un cas de réflexes influencés par une certaine rythmicité et qui ferait partie de comportements néonataux qui disparaissent à l'âge de deux mois comme la marche automatique, l'agrippement etc.

En revanche (Legerstee, 1990) a montré que l'imitation vocale chez le bébé de trois mois (qui en soit serait déjà un phénomène surprenant) existe sur certaines voyelles ne nécessitant pas un trop grand contrôle moteur. En situation où l'enfant était exposé à un contexte d'association entre des paires de formes visuelles simulées d'un son et une production orale, son imitation apparaissait plus fréquemment dans le cas d'un bon couplage entre stimuli visuels et auditifs. Sur des travaux similaires d'imitations vocales, parfois même assez complexes (p. ex. triphones), le bébé d'un mois imiterait également. Il présenterait même une certaine sensibilité à la durée de l'imitation (le bébé allongerait ou raccourcirait ses vocalisations de la même manière que le modèle). De plus, l'imitation serait associée à de vrais sourires, des comportements qui contrairement aux réflexes archaïques (qui disparaissent au bout de deux mois) persisteraient en montrant ainsi une grande prédisposition de l'humain au langage. L'enfance et l'adolescence seraient ainsi des périodes riches de formes d'imitations, en s'ancrant dans le processus de socialisation ou d'apprentissage (explicite ou implicite). Elle présenterait une certaine conformité au groupe en faisant correspondre non pas l'image qu'on a de soi, mais l'image qu'a l'autre d'un « personnage ». Par exemple, l'acteur ne serait pas forcément dit « bon » du fait de la compréhension personnelle qu'il a d'un personnage, mais tendrait plus vers ce qualificatif si son jeu « colle » à la représentation collective de ce personnage.

Parmi les comportements spécifiques faisant l'objet d'imitation, le pointage propose une décentration supposée et une compréhension de la causalité sociale. Il est ainsi considéré comme l'un des meilleurs indicateurs d'un problème développemental. Enfin, les auteurs parlent également d'imitations vocales lorsque les comportements sont associés aux vocalisations d'individu de la même espèce alors que les mimétismes vocaux seraient réservés à des sons non fonctionnels qui ne font pas seulement partie des étapes précoces de la vie.

Finalement, une dernière distinction entre mimétisme et imitation serait que l'imitation n'a pas besoin de la représentation du point de vue de l'autre contrairement à l'imitation. Dans les expériences de psychologie du développement, l'imitation serait ainsi un phénomène qui n'apparaîtrait que sur des comportements qui sont potentiellement possibles d'effectuer dénotant une dimension de capacité, dans le sens où il n'existerait pas d'imitation de comportements trop complexes à réaliser. Elle serait également réalisée lorsque les comportements sont non habituels et originaux d'un point de vue exploratoire, ce qui l'associe à la notion de sélectivité.

### ***1.3.4. Signaux sociaux : continuité multimodale de la communication parlée***

L'approche SSP (*Social Signals Processing*) montre que les événements vocaux influençant la dimension sociale en interaction face à face sont très nombreux. Elle montre

par exemple à travers des phénomènes tels que l'ironie, que le contenu sémantique du message verbal n'est pas absolu et que la fonction langagière peut résider sur la modulation intonative qui lui est associée. L'état de l'art de (Vinciarelli, Pantic, Bourlard, et Pentland, 2008) évoque ainsi cinq composants majeurs associés aux comportements non lexicaux : la qualité de voix qui peut être englobée dans la prosodie, les vocalisations linguistiques et non linguistiques, les silences et les tours de parole. Chacun de ces éléments contribuerait ainsi à différents aspects de la perception sociale du message. Nous reviendrons en détail sur la dimension prosodique et les expressions de vocalisations qui sont au cœur de cette thèse.

Le premier composant de l'interaction, le silence, est souvent interprété comme de la non-parole. Or certains auteurs lui attribuent un rôle majeur (Zellner, 1994). À l'image du premier axiome de l'école de Palo Alto, « on ne peut pas ne pas communiquer » (Watzlawick, Beavin, et Jackson, 1967/2007), du fait de la permanence des signaux sociaux et des inférences que l'humain en fait. (Enninger, 1987) définit par exemple qu'il existe des silences deux grands types de silences, ceux spécifiques à la situation et ceux qui dépendent de la culture. Les différences interculturelles ont d'ailleurs fait l'objet de nombreuses études (Lemak, 2012), notamment interculturelles (p. ex. Hasegawa et Gudykunst, 1998 ; Nakane, 2006) illustrant les fonctions langagières de ces silences.

(Richmond, McCroskey, et Payne, 1991) définissent trois types de silence en parole :

1. le silence hésitant : qui exprime une difficulté d'expression soit du fait de la conceptualisation d'un message complexe, soit du fait d'une attitude de l'interactant qui n'est pas favorable à l'interaction ;
2. le silence psycholinguistique : qui relèverait du temps nécessaire pour décoder ou encoder un message, et serait plutôt visible en début d'interaction quand le locuteur « cherche ses mots » ;
3. le silence interactif : qui est illustratif d'informations inhérentes à l'interaction elle-même. Il peut par exemple être un signe de respect envers la personne, ou une manière de l'ignorer ou d'attirer son attention (p. ex. une salle de classe bruyante interrompue par le silence appelant au calme par un professeur). Il peut ainsi être activement accompagné d'autres indices d'attention mutuelle ou d'expressions faciales.

(Kivik, 1998) a quant à lui une approche plus linguistique en définissant le silence avec une fonction a) syntaxique permettant d'organiser les objets verbaux, b) sémantique puisqu'ils peuvent eux-mêmes être significatifs ou encore c) pragmatique, car ils peuvent organiser l'établissement des relations sociales dans l'interaction. (Jaworski, 1992) précise également que contrairement à toute production de parole qui ne dépend que d'un seul énonciateur, le silence ne peut exister qu'à travers la coopération de deux interactants. Cette remarque est valable si le silence est caractéristique de l'événement entre les tours de parole (Cutting et coll., 2005 ; Tryggvason, 2006) et non celui existant dans l'énonciation d'un locuteur qui semble même parfois en définir ses caractéristiques (Tannen et Saville-Troike, 1985). Cela souligne alors la question de la définition du silence, très peu consensuelle, et dont les approches sont multiples. Enfin, outre la signification ou la fonction qu'il véhicule, le silence peut être associé à d'autres types de marqueurs qui permettraient de prévoir son apparition, comme le suggère (Slifka, 2007) pour la phonation irrégulière par exemple.

Le second composant des signaux sociaux de parole est la notion de tours de parole ou *turn-taking* (Duncan, 1972 ; Psathas, 1995 ; Thórisson, 2002) qui permettrait une régulation et une coordination des transitions de locuteurs. D'après de nombreux auteurs, les enchaînements de ces tours de parole, et notamment leur fluidité, auraient autant d'importance si ce n'est plus que le contenu du message lexical transmis. (Cutler et Pearson, 1985) modélisent le tour de parole par la description prosodique des énoncés échangés par exemple.

La régulation concernerait les comportements qui tendent à maintenir, céder, refuser ou demander un tour. Ainsi d'autres signaux tels que le regard ou des événements vocaux nuancés (p. ex. une toux pour demander l'attention) permettraient de relever des points de fluctuation pertinents pour signaler les transitions de ces tours (Yule, 1996).

La coordination dépendrait davantage des temps entre ces tours de parole. Le temps de latence serait alors plus ou moins codifié, jouant sur une certaine durée de silence acceptable, variable en fonction des cultures (Cutting et coll., 2005 ; Tryggvason, 2006). La fluidité de l'interaction aurait ainsi des conséquences sur le niveau d'attention mutuelle ayant pour conséquence une modification de la synchronisation interpersonnelle entre les locuteurs (Condon et Ogston, 1966 ; J. Cosnier, 1996 ; Delaherche et coll., 2012). Enfin les chevauchements de la parole, souvent intégrés dans une approche de réciprocité dissymétrique, basée sur la dominance (Kerbrat-Orecchioni, 1998 ; Smith-Lovin et Brody, 1989 ; Zimmermann et West, 1996) seraient également des événements illustratifs des mécanismes de coordination. Il faut noter que dans le cadre d'une interaction classique (Shriberg, Stolcke, et Baron, 2001) estiment à environ 10% la quantité de paroles chevauchées (*overlap*) par rapport au temps de parole total.

### 1.3.5. *Micro-expressions vocales*

L'introduction des silences et des tours de parole permet d'anticiper l'organisation des signaux sociaux de parole, lexicaux et non lexicaux, ainsi que la prosodie qui leur est associée. La production continue de ces signaux sociaux s'intègre ainsi dans des fonctions dites de *feedbacks* (M. Schröder, Heylen, et Poggi, 2006) (signaux entre les tours de parole du locuteur) et de *backchannels* (Yngve, 1970 ; Heylen et coll., 2007) (signaux ayant lieu durant le tour de parole du locuteur) qui assurent la continuité communicationnelle. Les tours de parole et les silences forment ainsi un premier maillage autour des objets interactionnels multimodaux, avec derrière l'idée qu'une interaction n'est pas seulement un échange alterné de message sémantique, mais un flux informatif continu existe entre les interactants.

Afin de dénoter ces signaux d'interaction, cette thèse utilisera le terme de « micro-expressions » pour désigner l'ensemble de ces signaux qui sont majoritairement appelés des « événements paralinguistiques » (Schuller et Batliner, 2013) dans le cadre de leur traitement automatique, voire « non verbal » ou « para-verbal » dans la littérature. Nous n'adopterons néanmoins pas ces terminologies puisque notre paradigme scientifique considère ces objets et leur prosodie comme langagiers. Dans cette perspective, ils sous-tendent l'interaction et ils sont au cœur des fonctions communicatives et donc différents de simples objets périphériques au message lexical. Nous préférons ainsi adopter la distinction lexical/non-lexical comme (Ward, 2000) le propose à travers ses *non-lexical speech sounds*.

#### 1.3.5.1. **Organisation, distribution et typologies**

Ces objets non lexicalisés peuvent ainsi prendre plusieurs formes et dénominations, qui peuvent soit dépendre d'une description acoustico-phonétique, soit tenter d'adopter un point de vue plus fonctionnel, le problème étant qu'une forme, voire sa catégorie ne véhicule en général pas qu'une seule fonction, mais plusieurs (Scherer, 1994). (Li, 2005 ; Wilkins, 1992) ont tenté d'aborder ces sons par leur sémantique, en les considérant comme des objets du discours tels que l'on peut considérer des éléments lexicalisés. La contrainte d'une telle conception est que certains événements ne peuvent être pris en compte si la morphologie considérée n'a pas de sens propre en soi ou si son sens est variable en fonction du contexte. En effet, l'apparition de ces événements (par analogie au *stroke* gestuel<sup>15</sup> de McNeill, 1992

---

<sup>15</sup> La notion de *stroke* définit le marquage d'un « temps fort » de la dynamique gestuelle selon (McNeill, 1975). Si nous supposons que la dynamique interactionnelle vocale peut être analogique à la dynamique

par exemple) ou la fréquence/ la rythmicité de cet événement dans le discours peuvent également produire du sens d'un point de vue communicationnel, relevant alors de la notion de « convergence communicationnelle » (J. Cosnier, 1996). Ainsi, la synchronie/asynchronie d'événements dits « co-verbaux » modifient la dynamique interactionnelle changeant l'état relationnel des interlocuteurs, ce que nous pourrions également inférer à tous signaux sociaux.

Toujours dans cette perspective, à travers des points de vue sociologiques ou psychologiques, des exemples illustratifs existent également à travers l'usage des tics de langage (cf. p. ex. la communication orale *Les tics de langage, ces petits mots-moteurs*, 2011). Ceux-ci peuvent avoir des formes variables, allant des moins lexicaux tels que des « hum » ou « euh » à des formules lexicalisées idiomatiques telles que « il va s'en dire »<sup>16</sup>. Ils évoquent notamment deux de leurs aspects en particulier : les « mots d'époque » qui reflèteraient diverses stratégies communicatives (p. ex. évitement, connivence etc.) qui sont illustratives d'informations socio-affectives. Elles tendent à faire comprendre la direction de l'interaction sans formulation explicite de ces états, par ce que Yves Prigent appelle un « mot paresseux ». Le second aspect concerne les « mots-écrans » illustratifs de ce que Jean-Pierre Winter appelle le « tic d'appartenance », des formes qui masquent des états affectifs par l'utilisation volontaire de formules toutes faites, codifiées, voire même ritualisées, et qu'il distingue du « tic involontaire » qui serait une manifestation langagière sans conscientisation par la personne révélant des informations sur ce dernier.

Sans nous intéresser ici aux débats de l'évolution simplificatrice du langage qui fait l'objet des auteurs, les effets et les utilisations du tic du langage font échos aux actes de langage introduits par (Austin, 1975). Derrière cette idée, la communication est un acte en soi, qui amène à faire réagir l'autre. Elle converge sur la notion de pertinence, introduite par (Sperber et Wilson, 1991). Ces auteurs expliquent que dans l'interaction, l'humain en perception cherche les informations les plus pertinentes c'est-à-dire à trouver le meilleur rapport gain/effort qu'il doit mettre en œuvre pour aboutir à des effets cognitifs donnés. En production, la notion de *l'intelligent inferential activity* de (Grice, 1991) rejoint cette idée, en suggérant qu'un locuteur présente toujours des informations motivées par une intention qu'il essaie d'inférer à travers ses actes. Cette dimension informative et intentionnelle, (Ohala, 1996) s'en sert pour proposer sa taxonomie éthologique de signaux de communication qu'il distingue en « signe » (des formes produites par inadvertance) et en « signal » (des objets produits avec une intention affirmée). Les signes sont informatifs, et ils sont donc susceptibles d'être utilisés par son interlocuteur, qui en fonction de son état empathique choisira d'agir ou pas. Concernant le « signal », plus l'intention portée est marquée, plus il est sujet à variations interculturelles. Cette idée, (Wharton, 2003) la formalise à travers un continuum qu'il appelle le *showing-saying* présenté ci-dessous, Figure 5.

---

gestuelle qui lui est concomitante, nous pourrions supposer que l'apparition d'un événement non lexicalisé peut être considéré comme un temps fort marquant, dans la temporalité des événements vocaux.

<sup>16</sup> Cf. article de vulgarisation pour avoir des exemples illustrés : <http://www.psychologies.com/Moi/Se-connaître/Personnalite/Articles-et-Dossiers/Ce-que-nos-tics-de-langage-disent-de-nous>

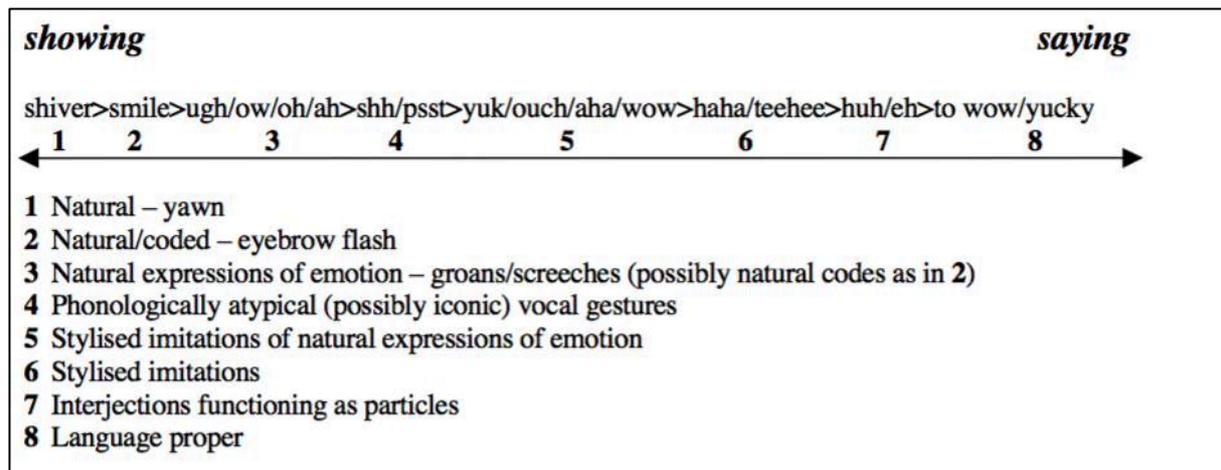


Figure 5 : continuum du *showing-saying* de Wharton (extrait de Wharton, 2003)

Ce continuum prend l'exemple d'une situation particulière qui est celle de l'expression de la douleur dont les formes non lexicalisées sont riches et variables. Wharton replace ainsi ces éléments expressifs en partant des éléments dits « naturels » qui illustrent des manifestations physiologiques qui sont visibles du point de vue du comportement corporel de l'individu (p. ex. les frissons, les expressions faciales). Ces objets tendent vers des productions de plus en plus « langagières » avec une norme socioculturelle qui codifie de plus en plus explicitement l'expression jusqu'à l'aboutissement d'une forme écrite partagée (p. ex. les interjections). Derrière cet axe, l'intention de communication évolue graduellement, en mettant en avant une dimension de contrôle de cette intention, le *showing* représentant ce que l'on voit malgré nous par des mécanismes automatiques conscients ou inconscients, mais involontaires, le *saying* illustrant ce que nous explicitons volontairement, à travers le langage qui et donc emprunt plus spécifiquement de la particularité socioculturelle. Elle rejoint la notion de *display rules* (Ekman et Friesen, 1971 ; Matsumoto, 2007 ; Kaiser, Wehrle, et Schenkel, 2009/2014). Elle suggère qu'il existerait des règles d'expressions implicites régulées par une norme socioculturelle orientant notre décision à plus ou moins montrer nos émotions et nos attitudes. Cette décision est prise soit en simulant les attitudes, soit en inhibant/neutralisant leurs effets, tout en changeant aussi l'intensité de ce qui est montré ou exprimé.

Il serait ainsi possible d'illustrer les signaux sociaux vocaux à travers cette approche graduelle du moins au plus linguistique. Sans être exhaustifs sur l'ensemble des terminologies utilisées, nous pouvons ainsi reprendre selon les classifications d'ors et déjà adoptés dans les travaux de (Vanpé, 2011; Signorello, Aubergé, Vanpé, Granjon, & Audibert, 2010; Sasa, 2013) :

Des « **bruits de bouches** » : il s'agit de son non phonétique plus ou moins vocalisé (p. ex. : rires, soupirs, des bruits respiratoires, des vocalisations, des bruits labiaux, des clics etc.). Les références en sont nombreuses et ils sont souvent définis comme une classe « fourre-tout ». (Sauter, Eisner, Ekman, et Scott, 2010) parlent par exemple de *nonverbal emotional vocalizations* désignées par le terme de *bursts* lorsqu'ils sont abordés d'un point de vue fonctionnel. (Scherer, 1994 ; Marc Schröder, 2003) parlent ainsi d'*affect bursts* dont les manifestations vocales ne seraient qu'une sous-partie intégrée dans un ensemble de modalités expressives (principalement voix et visage) qui font l'objet d'une synchronisation et reflèterait des informations émotionnelles ou socio-affectives. Les auteurs parlent également de « grognements » ou *grunts* (Ohala, 1996) repris sous la désignation de *conversational grunts* par (Ward, 2000) pour expliciter des événements relevés à fréquence relativement régulière dans des conversations américaines. Les auteurs évoquent également les

*vocal segregates* ou *vocal identifiers* (Crystal et Quirk, 1964 ; Smith, 1953 in Kendon, Harris, et Key, pp.285, 1975 ; Pittenger et Smith Jr, 1957 ; Trager, 1958) pour désigner des objets tels que « un-huh, shh, uh, ooh, mmmh, humm, eh, mah, lah » qu'on associe à des notions telles que la formalité, l'acceptation, l'assentiment, l'incertitude. Ils font également référence à des *vocal reflexes* (Crystal, 1969/1976) qui seraient de pures manifestations physiologiques sonores.

Des « **sons prélexicaux** » : bruits basés sur des sons phonétiques, voire même phonologiques, plus ou moins lexicalisés. (Goffman, 1981) évoquait par exemple les *response cries* comme des formes particulières d'interjections (p. ex. ouhla, aïe, wow, oops etc.). Il ne considérait pas ces productions comme des mots, car ils ne véhiculent pas véritablement un sens d'un point de vue linguistique. En revanche, ils correspondraient plutôt des formes ritualisées avec une acceptation socio-culturelle, qui permettraient de jouer un rôle communicatif. (Vasilescu, Adda-Decker, et Nemoto, 2008) évoquent les « hésitations vocaliques » qui sont des formes de voyelles apparaissant fonctionnellement dans un contexte d'hésitation, mais que les auteurs n'associent pas véritablement au système phonologique. Les *mind markers* (Poggi, 2001) désignent quant à eux tous les signaux, quelque soit leur modalité. Il s'agirait donc d'un objet multimodal, dont il existerait des manifestations vocales véhiculant l'état d'esprit des personnes, en particulier des informations sur l'incertitude et celles d'un ordre métacognitif. Les *fillers*, eux, ont souvent été étudiés dans le cadre interculturel sont des événements vocaliques relativement longs et stables qui seraient une manière d'annoncer un certain délai avant une reprise de parole (Candea, Vasilescu, et Adda-Decker, 2005 ; Dijkstra, Kraemer, et Swerts, 2006 ; Sadanobu, 2004). Les *filled pauses* (Maclay et Osgood, 1959 ; Batliner et coll., 2011) ou les *disfluency markers* (Brennan et Williams, 1995) en seraient des synonymes plus ou moins étendus.

Des **interjections** : leur définition est variable, mais elle a intéressé de nombreux auteurs. Ce sont des éléments qui ont une existence quasi lexicalisée, voire lexicalisée dans la langue. (Wundt, 1900 in Scherer, 1994) est ainsi l'un des premiers à en spécifier des fonctions. Il distingue une classe primaire uniquement dédiée à interrompre le flux de parole et une classe secondaire qui fait partie de la sphère linguistique. En repartant du concept de *response cries* de Goffman et du continuum de Wharton, l'idée de Wundt serait que des événements vocaux primaires (p. ex. « uh » dans une situation où le locuteur rate la cible d'une action) tend à devenir une interjection secondaire (p. ex. « oops » qui pourrait susciter une aide à la tâche ou) qui aurait une intentionnalité communicative plus forte à travers une formalisation linguistique plus marquée. Il peut donc apparaître seul dans un énoncé et être compris dans ce contexte. (Kainz, 1962 in Vanpé, 2011) reprendrait cette double distinction en rajoutant que les éléments primaires qui se manifesteraient sous formes relativement informelles (p. ex. cris, vocalisations, soupirs) sont avant tous des marques de communication émotionnelle donc se rapprocheraient de la conception de « bruits de bouche ». (Ameka, 1992) marque également cette transition à travers la notion de « geste vocal » en introduisant la dimension de la dynamique gestuelle associée à la production de ces interjections. Ameka définit également des critères pour discriminer ces interjections : morphologiques (non flexionnel), sémantique (véhicule d'informations sur état mental et affects), pragmatique (encodage d'attitudes et intentions communicatives). La définition linguistique de l'interjection, associant une sémantique affective, est quant à elle donnée par des auteurs tels que (Contini, 1989) par exemple.

Enfin nous pouvons également évoquer les **onomatopées** qui pourraient être considérées comme des formes spécifiques d'interjections. Plus ou moins conventionnelles, elles suivent directement les travaux du courant de symbolisme

sonore (Sasamoto et Jackson, 2016), particulièrement reprises dans une perspective grammaticale et sémantique (Akita, 2009 ; Kita, 1997 ; Toratani, 2005, 2015 ; Tsujimura, 2001) ou par une approche synesthésique (Ward, Huckstep, et Tsakanikos, 2006 ; Ward et Simner, 2003). Certains d'entre eux sont appelés des *sound-imitation words* (Ishihara, Komatani, Ogata, et Okuno, 2005) ce que l'on décrit comme des « phononymes » (la majorité des onomatopées français) désignant le *giongo* en japonais. Le *gitaigo* désigne des idéophones<sup>17</sup>, qui font référence à des états physiques ou émotionnels. Ces idéophones sont décrits par des « phénonymes » (des quasi-mots mimétiques qui représentent des phénomènes non verbaux) ou de « psychonymes », soit des mots mimétiques représentant des états psychiques) selon (Tomimoto et Nishioka, 2013). Dans le cadre de l'apprentissage de la langue japonaise, il est également fait mention des *giseigo* (sons émis par l'animal ou l'homme, p. ex. « cocorico », atchoum) ; les *giyougo* (associés à la caractérisation des mouvements et déplacements, p.ex. *byu-byu* japonais pour décrire un vent fort) et enfin des *gijougo* qui décrivent des impressions/sensations (p. ex. *moya-moya* associés à des moments d'inquiétude/rumination dont en français il serait possible d'approcher ces impressions par des expressions de type « avoir l'esprit embrumé » et « être dans les choux », *moya* ayant aussi comme acception « brume »). (Sperber et Wilson, 1987) rapportaient d'ailleurs sur l'usage de ces objets que dans des conditions appropriées, tout phénomène naturel ou artificiel peut être utilisé comme des représentations d'autres phénomènes qui leur ressemblent. Concernant l'impression sonore (Ramachandran et Hubbard, 2001) ont répété le protocole de (Köhler, 1929) et confirment la tendance universelle d'associer le son « kiki » à une forme au pointu, et le son « bouba » un élément plus en rondeur. Cela va également dans le sens des travaux de (Kagitani, Goto, Watanabe, et Sakamoto, 2014) qui montrent de manière systématique qu'il existe certaines formes phonétiques associées à des sensations de goût et de texture. Par exemple, l'épaisseur d'un liquide serait mise en corrélation avec des sons comme /n/, /d/ ou /m/ comme dans les onomatopées japonaises *neba-neba* ou *doro-doro*, alors que le /s/ est plutôt associé à une texture plus fine et plus légère comme l'indique le *sara-sara*. Ceci renforce l'exploration par l'association son/sémantique donnant lieu à des applications d'apprentissage comme dans le système JAMIOLAS<sup>18</sup> (Miyata, Ogata, Kondo, et Yano, 2008). Par opposition, la dimension inférentielle et pragmatique a très peu été étudiée. De fait, (Sasamoto et Jackson, 2016) se demandent si ces onomatopées ne sont pas des éléments de transition, à la fois des objets de *showing* et *saying* par le biais de la théorie de la pertinence (Carston, 2008 ; Sperber et Wilson, 1987). Selon les auteurs, ce pont serait fait par le degré de lexicalisation plus ou moins conventionnel et la dimension proprioceptive multimodale associée. Leibniz avait d'ailleurs proposé une théorie selon laquelle l'origine des langues était fondée sur la manipulation et l'évolution des onomatopées (Leibniz, 1973/2000). Ainsi l'onomatopée va au-delà d'une simple impression acoustique. Elle serait fortement liée à la perception corporelle à travers des inférences partagées. Elle serait d'ailleurs construite sur des rythmicités plus ou moins contrôlées (purement mimétique du son naturel ou plus ou moins lexicalisée par reduplication de syllabes symboliques, allongement, apocope etc.) (Tomimoto et Nishioka, 2013). Récemment, ces objets ont été utilisés dans des techniques de *coaching* divers comme dans le domaine du sport. D'une part, ils sont analysés dans le cadre de l'observation comportementale et le conditionnement psychologique (Fujino et Yamada, 2006), d'autre part dans le cadre

---

<sup>17</sup> Selon (Hagège, 2009) : « des idéophones, ou mots qui, comme le dit ce terme, offrent une peinture sonore d'une idée, pour symboliser un état, une impression sensorielle, une manière d'être ou de se mouvoir, une action qui n'est pas nécessairement elle-même reproductrice d'un bruit. »

<sup>18</sup> JAMIOLAS pour (J)apanese (M)imicry and (O)nomatopoeia Learning Assistant System) dont actuellement la version 3.0 est en développement.

leur production, pour faciliter l'accès à certaines formes de proprioception. Ils augmenteraient ainsi l'efficacité de la dynamique d'un mouvement en se basant sur l'impression de cette dynamique « prosodique » véhiculée par l'onomatopée, pour l'apprentissage (p. ex. le *shodo*, forme de calligraphie japonaise, assistée par machine dans (Hojo et coll., 2014) ou la description de « mouvements subtiles » (Fujino et Yamada, 2006). L'utilisation d'autres sons n'aurait *a priori* pas le même impact, en supposant qu'il existerait des structures acoustiques et prosodiques plus pertinentes que d'autres en fonction des situations. Ces onomatopées commencent également à être utilisées comme support de langage de robot dans le cadre d'interaction homme-robot en situation de collaboration (Jimenez, Kanoh, Yoshikawa, Furuhashi, et Nakamura, 2014).

Tous les événements vocaux présentés dans la section précédente sont très courts et peuvent sembler ponctuels, voire passer inaperçus. Certaines études longitudinales, dont le travail autour du corpus ESP par exemple, montrent que la proportion de productions non lexicales pouvait atteindre jusqu'à 49,5% d'un point de vue quantitatif (Campbell, 2005) ce qui rejoint également les propos de (Allwood, 1995) qui remarquait que quasiment la moitié des productions vocales d'un individu étaient constituées des signaux non lexicaux. Ils ont ainsi été étudiés principalement selon des fonctions spécifiquement présumées, en particulier émotionnelles (Campbell, 2004 ; Scherer, 1994 ; Schröder, 2003) dans le cadre de la phonostylistique (Fónagy, 1983 ; Léon, 1993) ; ou encore pragmatique (Ameka, 1992 ; Poggi, 2009 ; N. Ward, 2004 ; Wichmann, 2002). Ces sons ne donnent pas d'informations de contenus d'ordre sémantique, mais révèlent notamment la nature de l'interaction (Gravano et Hirschberg, 2009, 2011). Dans le cadre de l'interaction robotique, ces objets sont également largement exploités, mais de manière relativement empirique. (Breazeal, Kidd, Thomaz, Hoffman, et Berlin, 2005) leur donnent trois types de fonctions : interactionnelles, émotionnelles, et marqueurs de personnalité. Les signaux émergent de ces situations de communication s'appuient ainsi sur de nombreux indices multimodaux, étudiés dans le dialogue par une approche dite « étholinguistique » des conduites langagières (Colletta, 2007). Par ailleurs, la linguistique énonciative et discursive tend vers des analyses combinatoires de ces signaux de différentes natures, particulièrement à travers la prosodie. Elle montre notamment qu'il n'y a pas de relation biunivoque entre une forme et une fonction. En d'autres termes, les caractéristiques prosodiques dont on peut mesurer des valeurs absolues semblent faire référence à plusieurs fonctions à des niveaux de granularités différentes (Di Cristo, 1985).

### 1.3.5.2. Micro-expressions vocales à la prosodie

La prosodie sous-tendrait toute production vocale. Elle est souvent assimilée à l'intonation, qu'elle soit lexicalisée ou non, en suivant les contraintes du support linguistique qui peuvent être liées à la mécanique phonatoire et articulatoire d'une langue, caractérisant sa « mélodie » accentuelle (accents spécifiques, tons pour certaines langues). Au sein des productions verbales, la prosodie peut être associée à des fonctions structurant l'énoncé par des phénomènes d'emphase, de focus, de démarcation de frontière ou encore de « terminalité » qui désigne notamment la modalisation (p. ex. déclaration, interrogation etc.), caractérisant également la structure morphosyntaxique et morpho-phonologique intrinsèque des langues (Klewitz et Couper-Kuhlen, 1999).

La prosodie permettrait également de communiquer des intentions ou des attitudes, elles seraient chargées en informations émotionnelles, ou encore porteuses de caractéristiques d'âge, de genre ou d'identité (Cutler, Dahan, et Van Donselaar, 1997). De manière plus générale, des études longitudinales (Aubergé et coll., 2004b ; Audibert, 2008a ; Loyau et Aubergé, 2006 ; Vanpé, 2011 ; Signorello, Aubergé, Vanpé, Granjon, et Audibert, 2010 ; Biasi, Auberge, et Granjon, 2012 ; Sasa, 2013) ont montré à partir du corpus E-Wiz, que les fonctions des micro-expressions vocales et faciales, induites pour une production spontanée, sont riches : elles pointent sur des humeurs, des émotions, des intentions, des attitudes, des processus cognitifs et des états mentaux qui font l'objet du *FoT* (cf. section 1.3.3).

D'un point de vue acoustique, les informations véhiculées par la prosodie sont traditionnellement associées à trois paramètres principaux, explicités notamment dans (Lacheret-Dujour et Beaugendre, 1998). Les trois paramètres principaux sont ainsi :

- la fréquence fondamentale, notée  $F_0$ , déterminant la vibration périodique des plis vocaux qui constituent les cordes vocales ;
- la durée qui désigne le temps d'émission d'une production vocale, dont l'organisation temporelle peut être décrite par : le débit (nombre de syllabes par seconde), le tempo (accélération ou décélération du débit dans un groupe prosodique), les pauses, la durée d'un phonème soit l'unité minimale de son ;
- l'intensité qui évoque l'énergie, soit la force avec laquelle est émise la production qui résulte de l'amplitude de mouvement des cordes vocales, la force de l'air expulsé des poumons et de la variation de la fréquence de vibration des plis vocaux.

Se rajoute dernièrement un quatrième paramètre qui est celui de la qualité de voix (Campbell et Mokhtari, 2003 ; Sagisaka et coll., 2012) dont la modélisation reste un enjeu majeur dans le traitement automatique de la parole expressive et la reconnaissance de la parole spontanée.

La Figure 6 tirée de (Goudbeek et Scherer, 2010) permet d'avoir un aperçu général des différents descripteurs acoustiques qui sont mesurables pour déterminer la modification dynamique de cette prosodie. Elle peut également permettre de reproduire et modéliser l'esthétique prosodique d'une émotion, ce qui intéresse majoritairement la communauté de l'*Affective Computing*. En revanche, elle ne permet pas d'aborder fonctionnellement l'évolution de la dynamique prosodique associée à la variation de l'intentionnalité communicative.

Parameter	Description
Duration parameters	
Duration	Total duration of the utterances in s
Duration of the voiced parts	Duration of the voiced parts (excluding silences)
Duration of the unvoiced parts	Duration of the unvoiced parts (excluding silences)
Duration of the silent parts	Duration of the silence parts (threshold -25 dB)
Speech rate	1 / (duration) * no. of syllables
F0 parameters	
F0 mean	Mean F0 of the voiced parts of the entire utterance
F0 sd	F0 standard deviation
F0 maximum	F0 contour 95th percentile
F0 minimum	F0 contour 5th percentile
F0 range	Difference between F0 minimum and F0 maximum
F0 slope	The mean absolute difference between two adjacent points on the F0 contour, reflecting the mean local variability in F0
Intensity parameters	
Intensity mean	Mean intensity in dB
Intensity sd	Standard deviation of the intensity contour in dB
Intensity maximum	Intensity contour 95th percentile in dB
Intensity minimum	Intensity contour 5th percentile in dB
Intensity range	Difference between intensity minimum and maximum
Voice quality parameters (spectral balance)	
Energy below 500 Hz	The proportion of energy below 500 Hz
Energy below 1 kHz	The proportion of energy below 1 kHz
Hammarberg index	The difference between the energy maxima in the 0-to-2-kHz and 2-to-5-kHz range
Spectral slope	The slope of the regression line through the long term average spectrum
Spectral flatness	The quotient of the harmonic and geometric power spectrum means
Spectral skewness	The differences in spectral shape above and below the spectral mean
Voice quality parameters (variability)	
Harmonics-to-noise ratio	Degree of acoustic periodicity expressed in dB
Autocorrelation	The autocorrelation of the signal
Jitter	The mean absolute difference between consecutive periods, divided by the mean period
Shimmer	The mean absolute difference between the amplitudes of consecutive periods, divided by the mean amplitude

Figure 6 : Les paramètres acoustiques décrivant la prosodie (d'après Goudbeek et Scherer, 2010)

L'évolution de l'ensemble de ces paramètres permet ainsi de décrire une dynamique globale formant des courbes prosodiques appelées également des contours prosodiques. Pour la langue française (Delattre, 1966 ; puis Post, 2000 par exemple) par des approches différentes, proposaient dix formes intonatives de base décrivant les modalités d'énonciation. Cette modélisation d'une prosodie linguistique intonative est largement abordée et adaptée dans les différentes langues, dont (Rossi, 2000) par exemple, en retrace les différentes approches. Le modèle dominant reste celui de (Fujisaki, 1988) qui est une approche physiologique du processus de phonation et qui prend ainsi en compte l'évolution temporelle dont il rappelle que la rapidité des modulations de mouvement ascendant et descendant varie en fonction de l'unité d'observation prise en compte (l'énoncé, le lexique ou la syllabe). En effet, ces mouvements seraient plus lents dans le cadre d'un énoncé (la composante phrastique), plus rapide dans les niveaux de granularité inférieure (ce qui correspond aux composants accentuels). L'auteur considère alors qu'il est possible d'additionner ces deux types de dynamiques, dans le cadre d'une représentation logarithmique (en semitons) pour définir le contour de F0. Ce contour représenterait une fonction temporelle qui décroît globalement pour tendre vers le 0 à partir d'un point représentant la somme de l'ensemble des valeurs ascendantes du signal. C'est ce qu'il appelle la ligne de déclinaison qui sous-tend le contour prosodique pour représenter le mécanisme glottal de la production de parole,

temporellement situé. Cette approche s'inscrit dans ce que (Ladd, 1983 ; Ladd, Silverman, Tolkmitt, Bergmann, Scherer, 1985) appelle le *contour interaction theory* dont l'unité est celui de phrase d'énonciation. Ce modèle est en opposition avec les approches phonologiques, qui dans la littérature sont largement représentées par le modèle *ToBI (Tones and Break Indices)*. Ce modèle s'appuie sur les travaux longitudinaux de (Silverman et coll., 1992) jusqu'aux travaux de (Rosenberg, 2010) qui tendent vers l'automatisation de l'étiquetage prosodique. Cette convention de transcription prosodique développée dans le système anglais est largement adaptée à partir des années 90, pour différentes langues du monde. Il existe par exemple le J-ToBI puis X-JToBI (Maekawa, Kikuchi, Igarashi, et Venditti, 2002) pour le japonais, une version allemande de G-ToBI par (M. Grice, Reyelt, Benzmueller, Mayer, et Batliner, 1996), le ToDI (Gussenhoven, 2005) pour le néerlandais. Ce système s'appuie sur une première conjecture selon laquelle un contour prosodique peut avoir une représentation conceptuelle qui peut être séparée de la partie structurelle (donc qui se défait de la structure temporelle réelle et physiologique du mouvement prosodique). Ainsi dans ce modèle, le contour est découpé en segments de tons qui sont organisés de manière hiérarchique selon leur niveau de granularité de regroupement, marqué par des *break index* numériques représentant les frontières de ces segments. Ces derniers sont ensuite annotés dans un second temps, en termes de pitch relativement plus haut (H pour *high*) ou plus bas (L pour *low*), par rapport à la moyenne de F0 de l'énoncé (*pitch range*). Des règles sont ensuite définies pour décrire les mouvements accentuels de la production.

Alors que ces deux *mainstreams* d'approche prosodique se sont principalement appuyés aux trois paramètres prosodiques précédemment cités, la communauté prosodiste qui s'intéressait séparément depuis bien longtemps à la qualité de voix (que l'on assimile souvent au timbre de la voix), propose que ce paramètre soit considéré comme une quatrième dimension décrivant la prosodie (Campbell et Mokhtari, 2003). Il s'agit d'ailleurs d'un paramètre que (Scherer, 1986 ; Johnstone et Scherer, 1999 ; Roach, 2000 ; Gobl et Chasaide, 2000) associaient aux descripteurs des émotions comme elle apparaît également dans la Figure 6. La qualité de voix est également abordée avec des descriptions relativement précises dans le cadre du chant (p. ex. Garnier, Dubois, Poitevineau, Henrich, et Castellengo, 2004 ; Henrich, 2001), mais elle reste un problème complexe dont la définition est difficile (cf. état de l'art Audibert, 2008). Ainsi l'ASA (*Acoustical Society of America*), indique que le timbre est « ce qui différencie deux sons de même hauteur, de même durée et de même intensité ». Elle serait liée à la configuration et les propriétés physiologiques et physiques de la source sonore, soit au niveau des cordes vocales. Elle forme alors des impressions acoustiques définies par la phonostylistique (Fónagy, 1983 ; Léon, 1993) qui en présente des ontologies. (Ishi, Ishiguro, et Hagita, 2006) explorent également les paramètres de caractérisation de ces timbres, en se basant notamment sur des micro-expressions à travers des outils de reconnaissance automatique de la qualité de voix.

La prosodie a rapidement été investie par son versant oral. Or la dynamique d'un mouvement, qu'il soit oral ou visuel (gestes, expressions faciales, déplacement etc.), de même que la « tension » caractérisant la posture et du corps par exemple sont autant d'éléments qui peuvent faire l'objet d'une « prosodie audiovisuelle » comme l'introduisent (Swerts et Kraemer, 2005). Ces auteurs ont ainsi abordé la perception de FoK à degré informatif variable dans les trois conditions auditive, visuelle et audiovisuelle. Ils ont ainsi montré que la discrimination du FoAK (*Feeling of Another Knowing* qui correspond au FoK chez un autre que soi) était plus élevée en contexte de modalité couplée.

De récents travaux de synthèse expressive, tels que ceux de (Barbulescu, Hueber, Bailly, Ronfard, et coll., 2013) aborde cette question dans le cadre des ACA dans la continuité de ce qui avait été proposé par (Benoît et Le Goff, 1998 ; Beskow, 1997 ; Massaro et Egan, 1996 ; Pelachaud, Badler, et Steedman, 1996), par exemple. En termes de synthèse

expressive, les travaux orientés sur l'animation recherchaient surtout l'adéquation des mouvements des lèvres par exemple pour éviter les incongruences entre les stimuli auditifs et visuels, plus que sur les dimensions de dynamique commune à travers les différentes modalités.

La « prosodie visuelle » (Graf, Cosatto, Strom, et Huang, 2002) cible particulièrement les expressions faciales qui accompagnent la production orale. Malgré la tradition auditive dans le cadre des études en parole, en perception, la modalité visuelle est mentionnée comme ayant un rôle important dans l'intelligibilité du discours (Clark et Krych, 2004 ; Munhall, Jones, Callan, Kuratate, et Vatikiotis-Bateson, 2004). (Ferré, 2012) montre notamment dans l'analyse gestuelle co-verbalisée, qu'il existe des points de convergences rythmiques entre les temps marqués par une variation du débit de parole du locuteur et celle de ses gestes. En revanche, la prosodie audiovisuelle serait plus qu'une intégration de modalité. En effet, (Aubergé et Cathiard, 2003) ont montré par exemple que l'amusement s'entend au-delà de la conséquence acoustique liée au changement de configuration du conduit vocal, suggérant ainsi que la même valeur socio-affective aurait des contrôleurs différents. Le mouvement global de chaque geste (qu'il soit oral, gestuel etc.) associé à une fonction socio-affective devrait également avoir des caractéristiques communes à travers sa dynamique générale. (Girard-Rivier et coll., 2016) montrent dans le cadre d'une communication HRI uniquement gestuelle que des gestes associés à une même ontologie de mouvement présentent des variations dynamiques dont les caractéristiques sont pour certains similaires à ceux qui permettent de décrire la prosodie acoustique.

De manière paradoxale, la conception prosodique a surtout été abordée sur le versant oral, alors que la synthèse de micro-expressions a été très développée pour les expressions faciales, notamment dans le domaine de la robotique (cf. Fong et coll., 2003 pour un état de l'art sur ces aspects, ainsi que la section 1.3.6) et des agents conversationnels animés ou ACA (Pelachaud, 2005). En revanche, l'usage de micro-expressions vocales est encore balbutiant. Le statut informatif des énoncés vocaux non linguistiques commence tout juste à être exploré, avec une grande effervescence dans le cadre de la synthèse. Elle s'illustre par exemple par les débats dans le cadre de la table ronde “*roundtable on TTS*” (*Text-to-Speech*), de la conférence Interspeech 2012. De même, l'intérêt émerge également de la reconnaissance de la parole (Gatica-Perez, 2009), dans le courant plus large du *computational paralinguistics* (Schuller et Batliner, 2013) qui organisent notamment des challenges dans le cadre des conférences Interspeech : en 2009 “*emotion childrens' speech*”, en 2010 : “*age et gender, affect*” (*level of interest*), 2011 “*intoxication (+/- alcoholised), sleepiness*”, en 2012 « *personality (big five), likability, pathology*”, en 2013 : “*social signals, conflict, emotion, autism*” ; en 2014 : “*physical load, cognitive load*” et en 2015 : “*degree of nativeness, Parkinson's condition, eating condition*”.

### 1.3.6. De la voix à la personnalité : perspectives en robotique

Entre l'émergence des assistants personnels (*Siri* d'Apple, *Google Now*, *Amazon Echo*, *Cortana* etc.) et des robots, l'attribution d'une voix propre à ces objets technologiques ainsi que les caractéristiques de personnalisation deviennent des questions posées de manière récurrente. D'après des interviews de Ben Burtt, *designer* de la voix de R2D2, la voix de ce robot aurait été mixée par 50% de sons de synthétiseur et 50% de sa propre voix pour obtenir des inflexions plus « émotionnels » qui ont été comparativement étudiés pour le langage de Wall-E par (Jee, Jeong, Kim, et Kobayashi, 2010). Dès lors que les assistants et technologies caractérisés par des voix humaines commencent à avoir un plus fort ancrage culturel depuis leurs premiers développement dans les systèmes SVI (serveurs vocaux interactifs), le GPS ou la dictée vocale, le constat de (Nass et Brave, 2007) relevait que ce type de voix copiées de l'humain est associée à une perception culturelle de frustration et d'échec semble s'accroître du fait d'attentes surestimées, au profit des réflexions pour les voix non-anthropomorphiques. Elles sont alors principalement des voix féminines ou des voix d'enfants, et parfois masculines lorsque ces voix sont le support d'informations techniques.

Parmi les travaux explorant l'esthétique vocale en robotique, (Moore, 2017) suggère que la voix idéale d'un robot serait en cohérence avec sa masse corporelle. Il explore ainsi des vocalisations synthétisées sur des sons de nature mammifère avec un robot, Miro, dont le design est hybridé sur plusieurs formes animales. Il s'appuie sur un *mapping* d'affects à deux dimensions : l'*arousal* pour contrôler le débit d'air, l'amplitude et le tempo des vocalisations, et la valence pour régler la fréquence fondamentale et la qualité de voix, la variations de ces paramètres lui permettant d'explorer l'expressivité de cette voix biomimétique. Dans le travaux de (Wilson et Moore, 2017), la voix est explorée à travers différentes techniques de manipulations vocales dépendant principalement de techniques modifiant les dimensions temporelles et fréquentielles de la voix. Les auteurs rappellent l'existence d'outils commercialisés tels que Sonovox ou Auto-tuned permettant de travailler les productions vocales à partir de la source (c.-à-d. soit l'origine sonore de l'instrument vocal/sonore). Ils donnent aussi un aperçu de différentes techniques de modifications vocales et filtres applicables, plus ou moins complexes, directement appliqués sur des logiciels de conversion de voix tels que Voxal et dont certains standards permettent d'obtenir des voix de robot (c.-à-d. pitch légèrement augmenté avec addition d'harmonie et écho), alien (c.-à-d. diminution de pitch associé à un effet *chorus*) et cartoon (c.-à-d. une augmentation de pitch très importante avec un effet *chorus* et *tremolo*). À travers ces trois types de voix composant le corpus RAC, les voix sont étiquetées de manière à faire émerger les descripteurs de la *persona* incarnée par chacune de ces types de voix et l'étude montre que le design vocal va au-delà de l'intelligibilité ou la similitude à l'humain. Elle porterait un degré de justesse/convenance ou *appropriateness* qui dépend de l'apparence physique et les capacités cognitives attribuées aux entités étudiées.

En robotique sociale, des travaux ont également commencé à s'intéresser aux micro-expressions vocales. Elles sont en particulier utilisées pour des interactions avec des enfants (corpus NAOchildren de Delaborde et Devillers, 2010 ou travaux de Batliner et coll., 2004 sur Aibo). D'ailleurs, de manière très intuitive, à travers les œuvres culturelles occidentales, les robots tels que R2D2 ou Wall-E sont emprunts d'un fort potentiel socio-affectif, en produisant ce type d'objets. Par opposition, il existe aussi d'autres types de robots humanoïdes qui sont dotés de langage complexe, et qui malgré un modèle de raisonnement logique très développé, ne donnent pas forcément l'impression d'intelligence. Ils sont alors définis comme moins attachants ou moins « charismatiques », et parfois même hostiles, autant de termes qui sont utilisés dans les descriptions de la personnalité (cf. notion d'anthropomorphisme et *Uncanny Valley* section 1.4).

La personnalité peut ainsi être une dimension intégrée dans l'esthétique vocale du design robotique. En psychologie, elle est principalement décrite à travers une approche

lexicale de taxonomie (Allport et Odbert, 1936) aboutissant aux *Big Five* (McCrae et John, 1992) proposant 5 traits primaires que sont :

- l'extraversion (sociabilité, ouverture, confiance),
- le caractère agréable (amical, gentil, plaisant),
- le caractère consciencieux (aidant, implication),
- le neuroticisme (stabilité émotionnelle, l'ajustement)
- l'ouverture (intelligent, imaginatif, flexibilité).

D'autres alternatives telles que le *MBTI Myers-Briggs Type Indicator* (Myers, McCaulley, et Most, 1985) s'appuie également sur des questionnaires à échelles par approche lexicale. Elles restent cependant controversées d'un point de vue méthodologique. La question de la personnalité converge également sur la question de l'identité vocale, où la voix pose la question d'être une biométrie dans le cadre l'identification du locuteur (Ajili, Bonastre, Kahn, Rossato, et Bernard, 2016 ; Kahn, 2011). En effet, l'automatisation du séquençage de la parole par locuteur, appelée la diarisation (cf. état de l'art de Tranter et Reynolds, 2006 par exemple), montre que l'esthétique vocale en soit ne semble pas suffisamment discriminante pour identifier une personne dans une interaction multilocuteurs. Il existerait donc des comportements vocaux autres que cette esthétique, qui porte cette information très certainement associée à la dynamique prosodique d'un locuteur. De plus, il existerait peut-être plus de proximité du style de parole entre des personnes différentes exerçant un même rôle social, qu'une même personne changeant de rôle social au cours d'une même journée.

Par ailleurs, dans une approche symbolique de la voix, (Gussenhoven, 2002) montre qu'il existe des codes biologiques associés à la F0, inférant des informations universellement perçues, et qui portent les propriétés de celui qui produit ces sons. Il en définit ainsi trois codes : le *frequency code*, l'*effort code* et le *production code*. Le *frequency code*, dont de nombreuses illustrations sont relevées dans les travaux de (Ohala, 1994) par exemple, suppose notamment que la fréquence vocale infère des impressions sur la taille du larynx à partir duquel un son est produit. Il suppose par exemple qu'une F0 élevée donne l'impression d'un petit être avec toutes les inférences socio-affectives et intentionnelles liées. Une F0 faible suggère au contraire une impression acoustique produite par une grosse entité.

Ainsi, si des micro-expressions vocales venaient à être utilisées sur un robot, elles devraient être en mesure de conserver leurs enveloppes spectrales intactes tout en ayant la possibilité de modifier le *frequency code* pour avoir une cohérence avec le design corporel et le design vocal du robot. En effet, une infime modification comme celle du seuil de *glissando* (qui correspond à l'angle de la courbe spectrale – cf. par exemple Aubergé, Audibert, et Rilliard, 2004a), induit un changement perceptif important sur la fonction associée à la production. Le principe de copie prosodique est par exemple étudié dans le cadre du *morphing* par exemple (Mairesse, Walker, Mehl, et Moore, 2007), qui est également un outil souvent exploité dans le cadre de l'enseignement en langues étrangères (p. ex. *WinPitch Teaching*). Ainsi, associer une identité vocale, qui est un paramètre vocal de l'apparence du robot devrait pouvoir être personnalisé tout en conservant la cohérence avec l'impression globale qu'il fournira, sachant qu'il existe un symbolisme sonore associé à des paramètres acoustiques qui caractérisent également la prosodie.

## 1.4. Comprendre l'humain à travers le robot

La précédente section commence à introduire les signaux sociaux, et notamment les micro-expressions vocales, à travers l'objet robotique. Ceci montre que le robot peut donc être considéré comme un outil de mesure et de modélisation du langage, permettant de comprendre les mécanismes communicatifs de l'humain. Nous pouvons ainsi contrôler l'effet de certains objets interactionnels spécifiques en les faisant produire par le robot. En effet, il s'agirait d'une possibilité de reproductibilité de la situation interactionnelle, ce qui ne pourrait être directement possible avec l'humain puisqu'il en perdrait sa naturalité et sa spontanéité. Mais quelles sont les implications interactionnelles engendrées par cet outil robotique? La réflexion des sections suivantes s'intéressera donc à la notion de conscience, indispensable pour appréhender l'interaction, surtout lorsque la définition de cette conscience est mise en contraste avec un autre qui n'est pas humain, mais censé en représenter certains aspects.

À l'heure actuelle, les travaux sur la conscience se distinguent en trois types : la conscience de soi, la conscience de l'objet (ou plus général du monde extérieur) et la conscience morale chacune construisant en connivence l'ensemble de ce qui est aussi appelé la personnalité. Une déficience de l'une de ces consciences aurait donc des conséquences sur les autres (Vergnes, 2015). Dans le cadre d'une interaction homme-robot, la perception de cet humain par rapport à la conscience de son soi, va venir modeler la personnalité et le degré d'animisme qu'il va attribuer au robot, mais va également redéfinir sa propre conscience de soi à travers son interaction.

### 1.4.1. Conscience de soi : corporéité et « embodiment »

La conscience de soi est un premier point d'entrée pour concevoir la possibilité d'interagir avec autrui, celui avec qui la communication est potentiellement possible. Pour (Cooley, 1992) dont les théories sont à la base de nombreux travaux de psychologie sociale, le soi est considéré comme miroir, à travers la notion de *looking-glass self*, dont il définit : 1) l'image de notre représentation par rapport aux autres, 2) la conscience du jugement qu'il nous porte, 3) les résultants, positifs ou négatifs, qui en découlent. (Mead, 1934) modélisera à partir de cette idée le rôle qu'aurait autrui dans la constitution de soi, ce que la sociologie étudiera à travers des approches d'interactionnisme symbolique, comme dans les travaux de (Goffman, 2002) à travers la notion de *face*. Ce soi, peut ainsi être appréhendé à travers des normes esthétiques, de bienséance ou encore de performance pour faire l'objet d'une auto-estimation (Picard, 1983) ou encore pour s'autoévaluer en se comparant à autrui (Festinger, 1954). En effet Festinger, dans sa théorie de la comparaison sociale postule que l'humain a intrinsèquement un besoin d'auto-évaluation qu'il essaie de satisfaire en se basant sur des critères matériels, objectifs ou non sociaux dans un premier temps. Cependant en l'absence de critères de cette nature, l'humain s'appuierait sur une comparaison à autrui. (Heider, 1958) dénote d'ailleurs le caractère spontané et quasi inconscient de ce processus d'évaluation. Il étend également ce caractère aux mécanismes d'attribution de cause aux comportements observés chez autrui ou à soi-même. (Asch, 1946) a une approche gestaltiste et impressionniste sur la personnalité d'un autre perçu « au premier coup d'œil », en s'interrogeant comment une impression peut être globale et comment elle peut dépendre « d'une coloration donnée par certains traits », ce qui semblerait décisif pour l'appréhension de l'autrui. À travers un « effet halo », l'évaluation globale d'une personne pourrait selon lui être modulée par la présence d'un trait pertinent. (Bruner et Tagiuri, 1954 in Beauvois, 1982) ont exploré quant à eux la manière dont ces différents traits peuvent s'associer, par des inférences conceptuelles qui ne s'appuient pas sur la base d'une personne réelle. Ils portent leur attention sur « la manière d'être » d'un sujet, idées décrites dans les théories dites naïves ou implicites de la personnalité. Cette perception sociale du soi serait alors intrinsèquement

liée au corps véhiculant et modulé par les traits de cette conscience (Jodelet in Moscovici, 2015).

*« Dans le domaine de la perception sociale où dominent encore les modèles cognitivistes, nous avons affaire à des processus qui ne relèvent pas d'un mécanisme intra-individuel, mais d'une dynamique où le social et le culturel donnent sa matière et sa forme au psychologique. Ceci est rendu possible par l'introduction de la dimension corporelle dans l'approche psychosociale des relations interpersonnelles et de la formation des représentations de soi et d'autrui »*

(Extrait de conclusion Chapitre 2 par Denise Jodelet in Moscovici, 2015 –p.65)

Ainsi en psychologie sociale, les travaux de 1975 (in Argyle, 2013) ou (Frey, 1981) s'intéressent à la dimension du corps dans une perspective globale des processus communicatifs qui façonnent le soi. (Hall, 1991 in Garnier, 1991) s'intéresse également à ce corps qui intervient à travers les gestes et les mouvements dans la communication de manière kinésique (Hall, 1959), proxémique (1966, in Hall, 1990) ou chronémique (c.-à-d. le rythme corporel et la synchronie dans l'interaction – Hall, 1983). De plus, (1970, in Birdwhistell, 2010) rajoute à cela l'influence des facteurs culturels. Cette dimension de la corporéité sera donc abordée dans la suite, en lien avec les notions et les effets qu'elle implique dans le domaine de la robotique.

*« Je n'appréhende pas l'autre tout simplement comme mon double. Je ne l'appréhende ni pourvu de ma sphère originale ou d'une sphère pareille à la mienne, ni pourvu de phénomènes spatiaux qui m'appartiennent en tant que liés à l'ici. Mais, à considérer la chose de plus près, c'est son corps qui est constitué d'une manière originelle et est donné dans le mode d'un « hic absolu », centre fonctionnel de son action »*

(Husserl, Méditations Cartésiennes, 1929)

En philosophie, la phénoménologie du corps de Husserl marque un changement de pensées au cours du XXe siècle. S'opposant à la dualité corps/esprit proposée par Descartes ou Locke, elle est également défendue en critiquant ce dualisme fondé sur un corps de chair et un esprit l'animant, redéfinie dans la notion de « corps propre » de Merleau-Ponty ou « corps sujet » de Gabriel Marcel. Ces notions introduisent que toute expérience de la conscience de soi est totalement confondue avec ce corps dont on ne peut se séparer, car il en est intrinsèquement un prolongement : « Le corps est un Pour-Soi, et non un En-Soi dans le Pour-Soi » selon Sartre.

À partir d'une approche phénoménologique, Merleau-Ponty intègre à ce corps une notion de spatialité, mais surtout de temporalité qui lie ce corps à d'autres qui « coexiste », ce qu'il décrit à travers la notion d'intersubjectivité. Partant du postulat que le corps ne peut être considéré que s'il est situé. La temporalité dont parle Merleau-Ponty ne définit pas simplement le corps comme initiateur d'une rythmicité particulière ni comme la somme juxtaposée d'événements temporels et spatiaux expérimentés, le corps fait partie intégrante et il *est* temps et espace, lui permettant d'en appréhender physiquement les manifestations (Merleau-Ponty, 1976).

Ce point de vue est intéressant pour cette thèse puisqu'elle est sous-jacente à une vision scientifique plus large qui tente de comprendre comment l'humain peut être relié aux autres, la communication humaine semblant être un outil privilégié et spécifiant la nature de l'Homme afin d'y parvenir. En revanche, elle introduit aussi ambiguïté à laquelle tout roboticien se confronte, dont l'illustration la plus parlante est le *Géminoïd*. Ce robot entièrement humanomorphe d'Ishiguro, l'un des robots les plus élaborés techniquement et le plus esthétiquement proche de l'humain. Les expériences conversationnelles incarnées réalisées avec ce robot illustrent notamment la limite du dualisme cartésien en regard de la phénoménologie du corps (Grimaud et Paré, 2011).

En effet, à travers les âges, l'approche générale en robotique a été de considérer avant tout le robot comme un corps matériel (artistique, mécanique, mimétique) auquel il est possible de greffer des comportements ou insuffler un esprit ou une âme mimée par une intelligence artificielle dans le but d'animer et faire interagir cet objet. S'il est suffisamment cohérent et réaliste, cette structure pourra passer de l'état d'objet à celui d'un objet animé (automate) incarnant un autre (robot), directement programmé ou simulé par différentes techniques, à domaine et à motivation variable (Wood et Marty, 2005 ; Himbert, 2012 ; Tisseron, 2015)<sup>19</sup>. Cette perspective s'intègre ainsi dans un fort engouement de développement appelé la Robolution (Bonnell, 2010).

En revanche, les exemples tels que le film *Her* (Jonze, 2014), où seule la voix d'un système d'exploitation, amène un humain à le considérer comme un autre sans corps, voire même à de multiples corps technologiques. Ceci est une dimension rattachée à la notion d'ubiquité concernant les objets connectés ou *IoT (Internet of Things)*<sup>20</sup> introduits par (Zaslavsky et Jayaraman, 2015). Ce « robot vocal » en l'absence de corps pose ainsi les questions autour de la notion d'*embodiment* en robotique, définie comme suit par (Dautenhahn, Ogden, et Quick, 2002) : “*that which establishes a basis for structural coupling by creating the potential for mutual perturbation between system and environment*”. Cette notion établit un fort lien entre le corps et l'environnement, l'*embodiment* augmentant lorsque le corps et l'environnement se perturbent et se modifient mutuellement. Cette conception reste donc logique dans le cadre de la conception phénoménologique puisque l'environnement dans un sens large englobe les dimensions spatiale et temporelle, qui ne font qu'un avec le corps. De plus, d'après Dautenhahn et ses partisans, ce corps n'a pas besoin d'être physique, l'*embodiment* étant aussi une notion qui s'applique à des ACA (Agent Conversationnel Animé) par exemple. On dit alors que tout robot est *embodied*, mais que son degré d'*embodiment* change.

En robotique, les modèles qui décrivent cette notion veulent transposer des comportements sociaux dans une corporéité donnée. Cela implique des enjeux tels que l'émergence d'une « socialité intrinsèque » faisant appel aux processus empathiques des sujets du système interactionnel (Pfeifer, 2002). Toujours selon Pfeifer, les deux approches les plus communes de développement sont d'une part une approche bio-inspirée qui s'appuie sur la simulation et le mimétisme du vivant en se basant des connaissances éthologiques (soit sur la théorie de l'esprit ou la psychologie développementale) ; d'autre part une approche dite de « design fonctionnel » plus axée sur des l'ingénierie, dont la caractérisation sociale est bien plus superficielle et rentre dans les compétences de HCI (*Human-Computer Interaction design*). L'intérêt de ce type d'approche est qu'elle intègre une dimension itérative à la conception de design, représentant un point d'intérêt méthodologique, mais il ne sera pas conceptuellement suivi. Dans ce besoin d'associer le corps à une forme d'intelligence et de personnalité, de même que cette dimension intégrée du design, la notion d'*affordance* émerge (Severinson-Eklundh, Green, et Hüttenrauch, 2003). L'une des premières définitions explicites est initialement donnée par la psychologie à partir des travaux de (1966, in Gibson, 1977) qui

---

<sup>19</sup> L'histoire de la robotique depuis l'Antiquité à nos jours ne sera pas développée en guise d'état de l'art dans cette thèse, mais notons que cette histoire a son importance sur la dimension socio-culturelle et les connaissances que nous avons à tout à chacun vis-à-vis de cette technologie et de sa perception (p. ex. <https://humanoides.fr/hala-ou-quand-la-robotique-sempare-des-questions-socio-culturelles/>, consulté le 10/01/2017)

<sup>20</sup> Notions explicitées par exemple dans le Livre Blanc des Instituts Carnot ([http://www.instituts-carnot.eu/sites/default/files/images/AiCarnot-Livre Blanc-Objets Communicants et Internet des Objets-long.pdf](http://www.instituts-carnot.eu/sites/default/files/images/AiCarnot-Livre%20Blanc-Objets%20Communicants%20et%20Internet%20des%20Objets-long.pdf), consulté le 13 janvier 2015) ou au cœur des recherches du labex Persyval-lab (<https://persyval-lab.org>, consulté le 18 juin 2013) dont cette thèse a reçu le financement, orientés sur les « systèmes pervasifs » que l'on retrouve sous la terminologie anglaise de *ubiquitous computing (ubicomp)* qui fait aussi partie des enjeux des communautés travaillant sur les systèmes cyber-physiques (cf. 1<sup>ère</sup> édition *FaDEX on Cyber-Physical Systems, attended in 2016*, )

présente des ontologies d'*affordances*, et dont l'écologie et les concepts sont abordés par (Goldstein, 1981). Ces travaux de références sont très largement repris pour construire de nombreuses théories psychologiques autour de la perception, sur lesquelles nous ne reviendrons pas, mais dont des auteurs tels que (Chemero, 2003) font un état de l'art. Citons également (Norman, 2004) qui garde ce concept d'*affordance* rattaché à cette dimension design de l'objet. Ces dernières dérivent des concepts philosophiques de (Heidegger, 1962/1992) qui considère la perception comme un phénomène direct, telle qu'elle apparaît également dans l'approche physiologique de la théorie des émotions de (James, 1912/1976 in Titchener, 1914) qui est un autre point d'entrée pour la notion d'*affordance*. En robotique sociale, elle est rattachée au domaine de l'*Affective Computing* (R. W. Picard et Picard, 1997)<sup>21</sup> et son acceptation actuelle est réduite à « toutes les possibilités d'actions sur un objet » ou encore la « capacité d'un objet à suggérer sa propre utilisation ». Elle est souvent présentée comme synonyme du terme « potentialité ».

En parlant ainsi de comportements associés au robot, nous entrons dans une phase où ce dernier va commencer à produire des effets sur la perception de l'humain. De l'objet, il deviendra donc potentiellement un autre, ce qui initie alors des réflexions autour de l'animisme de ces objets qui sera la notion abordée dans la prochaine section .

### 1.4.2. Conscience de l'objet : lorsque le robot devient autre

« Les autres, ce sont plutôt ceux dont la plupart du temps on ne se distingue pas »

(Heidegger, Etre et Temps)

Cette citation d'Heidegger, disciple d'Husserl, illustre des réflexions autour de son concept d'altérité, soit le fait de concevoir un autre que soi. Il montre que la frontière n'est pas évidente, et qu'elle est même intégrée dans une continuité, tout comme le suggère le concept de la conscience de soi. Ce processus de distinction de soi et d'autrui est d'ailleurs exploré à travers les théories du développement humain, par différentes approches : cognitiviste (p. ex. Piaget), socioculturelle (p. ex. Vygotsky), comportementaliste (p. ex. Pavlov, Skinner, Watson), psychanalytique (p. ex. Freud, Erikson), humaniste (p. ex. Rogers, Maslow) ou écologique (p. ex. Bronfenbrenner). Les interactants de ces approches s'appuient sur des êtres vivants autres que l'humain, donnant un aperçu sur les mécanismes potentiellement exploitables dans l'interaction avec le robot, qui lui ne reste qu'une machine. L'enjeu de la compréhension des interactions de l'homme avec un robot reposera donc principalement sur une perception humaine située. Ainsi cette perception se basera sur des primitives de comportements (dans un sens très large en termes de modalités), modélisées et greffées sur cet objet qu'on appellera robot. En revanche, qu'elle est la différence entre un simple automate mouvant ou bruité, et un robot considéré comme un autrui pour l'humain ?

En anthropologie (Bird-David, 1999) mentionne l'œuvre de Tylor, *Primitive Culture* (1871 in Tylor, 1920) comme l'une des premières références développant le concept d'animisme à partir de l'introduction de l'alchimiste Stahl, au XVIIe siècle. Lui-même aurait hérité la notion de ses précurseurs classiques, la définition dérivée étant généralement associée à celle de (Harris, 1983 in Bird-David, 1999) : “*the belief that outside ordinary visible, tangible bodies there is normally invisible, normally tangible being : the soul (...)* each culture [has got] its own distinctive animistic beings and its own specific elaboration of the soul concept”. Cette définition coexiste avec d'autres, associées à une dimension religieuse (Hunter et Whitten, 1976) qui reste intéressante du point de vue culturel puisque sa

---

<sup>21</sup> Également voir les travaux de l'association HUMAINE (base de données disponibles) du réseau d'excellence SSPNET – European network of excellence in social signal processing (<http://sspnet.eu>, consulté le 13 septembre 2012).

pratique n'est pas obligatoirement associée à des croyances métaphysiques. Par exemple au Japon, la notion d'animisme pour la robotique adopte une approche éthique très particulière du fait de ses préceptes socioculturels (Kitano, 2006).

Il s'agit d'un concept qui semble avoir particulièrement intéressée la branche de l'ethnologie (Descola et Pálsson, 1996 ; Gudeman, 1986 ; Sahlins, 1976/2005). Les sciences sociales raccrochent d'ailleurs explicitement cette notion à celle de l'*embodiment* dans la définition qu'en donne (Gould et Cobb, 1965 in Bird-David, 1999) : "*the belief of the existence of separable soulentity, potentially distinct and apart from any concrete embodiment in a living individual or material organism*". Or un robot ne sera jamais vivant et tout être humain aura conscience qu'il ne s'agira que d'une machine. Pourtant, sa nature même ne pourra s'empêcher de subir un certain degré d' « illusion de vie » comme celle visant à être recréée par des techniques évolutives et empiriques, mais très judicieusement choisies entre mimétismes et accentuations de certains traits, par les professionnels de l'animation (Thomas, Johnston, et Thomas, 1995 ; Lasseter, 2001).

De manière générale, la littérature montre un certain regain d'intérêt autour des changements de perspectives de cette notion d'animisme. Récemment, dans la communauté anthropologue, le séminaire de l'EHESS-ENS 2014/2015 intitulé « La cognition animiste : une approche transdisciplinaire »<sup>22</sup> a été spécialement dédié à ce changement de dynamique afin de réactualiser les concepts développés autour de la cognition animiste. Parmi les approches contemporaines connues faisant suite aux conceptions classiques, sont évoqués les « schèmes ontologiques » de (Descola, 2006 in Journet, 2010) proposant quatre systèmes de sociétés humaines dont l'approche diffère entre naturaliste, analogique, animiste et totémique, la dernière étant par exemple largement développée par (Lévi-Strauss, 1962/2002), le naturaliste illustrant les sociétés occidentales, les sociétés d'« Archipel », dont par exemple, le Japon faisant office de civilisation analogique. Une autre approche mentionnée est également celle (Viveiros Castro, 2002 in Fausto, 2011) portant sur la théorie du perspectivisme et du multiculturalisme-monoculturalisme introduisant les effets de « vêtement » que l'animal porte, désignant le fait que ce dernier comporte une part d'humain que l'humain perçoit à travers et le ramène sur une dimension comparable à celle de l'humain. En revanche, il semblerait que de nombreuses théories sont présentées comme alternatives à l'animisme, et pour seulement les citer : Ingold, Bird-David, Rival, Harvey, Willerslev, Kohn, Karadimas, Santos Granero, Lenaerts. Sous l'angle de l'intelligence artificielle, il est également évoqué que nombre d'entre eux suivent une influence fodorienne (approche fonctionnaliste en philosophie de l'esprit, développant une théorie computationnelle de l'esprit par une représentation d'un langage de pensée innée comme Chomsky), ayant ainsi une approche symboliste (p. ex. Sperber) et bien plus encore connexionniste (p. ex. d'Andrade, Quinn, Shore, Bloch etc.). Cette communauté a d'ailleurs ouvert le débat sur les influences possibles modèles récents de l'esprit humain que sont la théorie de la cognition dynamique (chapitre de van Gelder, in Haugeland, 1997) et les théories du cerveau bayésien (dont les travaux évoqués ici sont ceux de Frith, Friston et Griffiths)<sup>23</sup>, en restant sensible à des questions telles que la modularité ou la spécialisation par domaine de la cognition humaine par une critique des approches constructivistes (p. ex. Boyer, Sperber, Atran, Bennardo) et modularistes (p. ex. Karmiloff-Smith, Sprinz, Barrett).

---

<sup>22</sup> Description du séminaire organisé par M. Fortier et J. Dokic en ligne (<http://enseignements-2014.ehess.fr/2014/ue/1001/>, consulté le 01/11/2016) et descriptif du syllabus (<https://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbntYXJ0aW5lbGllZm9ydGllcnxneDoyNjNkYmY5YTlkYWYyZDBl>, consulté le 01/11/2016)

<sup>23</sup> Une présentation abordée par le point de vue la psychologie cognitive expérimentale introduite par Stanislas Dehaene concernant cette approche du cerveau bayésien. « Le cerveau vu comme un système prédictif », délivré au Collège de France le 21 février 2012 (disponible sur : <http://www.college-de-france.fr/site/stanislas-dehaene/course-2012-02-21-09h30.htm>, consulté le 10/01/2017)

Outre les considérations anthropologiques, de nombreuses références, notamment issues de la psychologie évolutionnaire (p. ex. Guthrie, Boyer, Sperber, Barrett), tendent néanmoins à montrer que l'humain commence à inférer des concepts liés à cette dimension animiste à partir de très peu de choses, par exemple (Guthrie, 1993/1995 ; Vidal et Gaussier, 2014) qui évoquent une personnalisation des objets du quotidien par l'humain. En effet, dès lors qu'il est confronté à une forme d'attachement quelconque liée à la possessivité de cet objet (exemple de la voiture), il relèverait d'une dimension fétichiste, qui va même jusqu'à être explicitement exploitée sur l'Elfoïd d'Ishiguro (Kaerlein, 2015). Guthrie mentionne également le mouvement comme une dimension importante à cette inférence, ce que l'on retrouve d'ailleurs à travers la notion « d'animacité », le caractère de vie attribué à une chose à travers sa manière de bouger, mais dont les mécanismes d'attribution sont encore peu étudiés (Pantelis et Feldman, 2012). En revanche, elle permettrait d'appréhender la frontière perceptive de l'humain, en considérant qu'un être peut-être animé intrinsèquement par lui-même, ou en bloquant cette capacité comme pour les éléments qui bougent au gré du vent (Pratt, Radulescu, Guo, et Abrams, 2010).

Tout comme dans le cadre de l'animation, les pratiques comme celle des marionnettistes intéressent également la communauté robotique<sup>24</sup>. (Rusko et Hamar, 2006) évoquent deux aspects principaux dans cette pratique, celle du mouvement (*animation*) et celle de la voix (*déclamation*). Sur les aspects vocaux, ils évoquent deux approches :

*“the first one we have called « One voice – several speakers », is used to study voice imitation, and the second approach, that we call « One speaker – several voices », studies actor's voice when representing different characters. (...) it is very difficult to separate speech features reflecting personality from those that express emotions, affect and attitude.”*

En fonction de l'approche adoptée, les marionnettistes semblent alors accentuer et doser leur qualité de voix en particulier (l'un des quatre paramètres de la prosodie). Ils jouent ainsi sur ce qui est traditionnellement connu comme les cinq dimensions qui caractérisent la personnalité (Laver, 1991). De plus, attirant l'attention par des accessoires divers, ils créent une écologie d'interaction favorisant cette perception animiste à travers trois niveaux interconnectés :

*“personal characteristic of the traditional characters of the folk puppet theatre is a complex of three interconnected levels: semantic, optical and auditive. The semantic level characterizes the character as to its function in the play. The optical layer represents all components of visual representation of a puppet (face, costume, material and animation). The auditive layer includes speech, music and all the sounds generated by actor and his puppets.”*

Notons que le niveau dit « sémantique » relève de la fonction du personnage qui rejoint la notion « rôle » que nous attribuerons au robot dans notre approche.

Le marionnettiste (Rusko et Hamar, 2006) ou l'animateur (Lasseter, 2001), s'accordent ainsi sur le fait sont souvent les détails qui font la différence de perception, ce qui est également exploré par (Hoffman et Breazeal, 2007) dans le travail sur AUR, le robot répliquant de la lampe Pixar<sup>25</sup>.

Ces détails qui créent l'« illusion de vie », se retrouvent également dans la forme du mouvement, de la voix, de la parole, des mimiques, du timing, de la rapidité d'exécution d'une action (qui rend d'ailleurs difficile la communication de certains traits selon Lasseter) et de l'adéquation avec son environnement. Dans les deux cas, ces paramètres peuvent être réajustés pour amplifier ou attirer l'attention sur certains détails. La notion d'animacité peut

---

<sup>24</sup> Notons que lors de la dernière école d'été Social HRI (participation 24 au 28 Août 2015, <http://www.hrisummerschool.org>), la communauté de robotique sociale a invité Staci and Jonathan Hodge (*Furri-fingers, UK*) des marionnettistes dans le cadre du workshop “*puppetry and human-robot interaction*” pour explorer cet aspect.

<sup>25</sup> La présentation TED de Guy Hoffman “*robots with soul*” disponible sur ([http://www.ted.com/talks/guy\\_hoffman\\_robots\\_with\\_soul](http://www.ted.com/talks/guy_hoffman_robots_with_soul), consulté le 31/08/2014) donne des indices sur ce que peut être les indices de l'apparition d'animisme pour le robot

ainsi être étendue au-delà du mouvement, à travers les dimensions qui la caractérisent et qui seraient donc un enjeu important pour la robotique.

Enfin, ces traits activant la perception animiste chez l'humain se rapprocheraient peut-être de la notion de *patternicity* que (Shermer, 2008) introduit comme “*the tendency to find meaningful patterns in meaningless noise*”. Traditionnellement, ces effets (dont la paréidolie fait partie – p. ex. voir des formes dans les nuages, mais valables pour toute autre modalité perceptive) ont été considérés comme des erreurs de perception. Il serait soit de type I : croire que quelque chose est réel alors qu'il ne l'est pas, soit de type II : ne pas croire que quelque chose est réel alors qu'il l'est. D'après Shermer, ces traces que percevons dans divers stimuli qui nous entourent tendent à nous faire penser à des éléments interprétables. L'humain s'appuierait donc sur des éléments de mémoire de notre vécu pour produire des croyances et des illusions perceptives en trouvant des liens et une cohérence entre plusieurs éléments qui semblent ne pas en avoir s'ils sont pris séparément. Cet auteur, issu du courant sceptique, cite des propos du psychologue Whitson qui relie cette tendance illusionniste par le besoin humain de conscientiser un contrôle sur les éléments avec lesquels il interagit :

*“Feelings of control are essential for our well-being – we think clearer and make better decisions when we feel we are in control. Lacking control is highly aversive, so we instinctively seek out patterns to regain control – even if those patterns are illusory.”*

(Whitson in Shermer, 2009)

Si de tels mécanismes existent, il pourrait s'appliquer successivement sur deux aspects pour le robot : 1) l'animisme du robot ce que Shermer appelle plus globalement l'*agenticity*, qui découlerait de cette notion de *patternicity* en posant les bases cognitives des croyances humaines ; 2) l'anthropomorphisme du robot, qui par ses comportements et son apparence couplés ou non à des éléments de contexte, associera une dimension interprétable qu'il raccrochera à des choses humaines qui a une tendance à décrire le monde par des « attributions égocentriques » (Heider, 1958 ; Jones, Nisbett et coll., 1971) ou des « projections attributives » (Holmes, 1968 ; Murstein et Pryer, 1959). Ces notions réfèrent notamment à des terminologies de perception humaine pour décrire un autre animé (Viveiros Castro, 2002 in Fausto, 2011), amenant parfois à des biais de perception sociale (Ross, Greene, et House, 1977). Plus particulièrement en robotique, l'objet qui est animé est potentiellement perçu comme un « autre ». Cet effet a des conséquences particulières dans l'anthropomorphisme perçu, à travers la notion d'*uncanny valley* ou la « vallée de l'étrange ».

*« On croit que la physionomie n'est qu'un simple développement de traits déjà marqués par la nature ; pour moi, je penserais qu'outre ce développement, les traits du visage d'un homme viennent insensiblement à se former et à prendre de la physionomie par l'impression fréquente et habituelle de certaines affections de l'âme »*

(Rousseau – Emile, IV)

La citation de Rousseau nous rappelle que l'apparence peut aussi être le reflet de ce qui nous constitue et ce qui est perçu chez l'autre. Il s'agirait ainsi d'une extension de soi, mais définie à travers des caractéristiques qui nous en diffèrent. L'« apparence » peut ainsi se comprendre dans un sens large, pas uniquement visible, mais peut-être aussi dans la voix ou la nature de la motricité par exemple, soit toute dimension qui caractérise au premier abord des dimensions esthétiques d'un être animé.

En robotique, cette question d'apparence a d'abord été abordée au Japon, par les impressions que pouvaient donner des robots humanoïdes qui restent l'approche et la conception la plus courante dans cette société. Ainsi le roboticien Mori introduit la notion de *bukimi no tani*, ou *uncanny valley* ou vallée de l'étrange, en critiquant la mouvance de modélisation à vouloir toujours expliquer les phénomènes par une fonction linéaire. En effet,

la tendance était de dire que plus le robot ressemble à l'homme, plus l'humain aura une affinité importante avec lui. Or Mori suggère que cette relation apparence/affinité, à l'image de l'ascension d'une montagne, peut augmenter jusqu'à un certain point, mais qu'elle va chuter au point culminant qui présente ensuite un gouffre, à savoir que la trop grande proximité d'apparence entre l'homme et le robot génère une perception dérangement de l'objet<sup>26</sup>, car sa nature est différente.

Nous retrouvons ainsi classiquement deux représentations de cette vallée de l'étrange, une première axée uniquement sur l'apparence physique (schéma gauche – Mori, 1970) puis une version plus récente du même auteur qui intègre le mouvement de l'objet robotique dans (le schéma gauche – 2012).

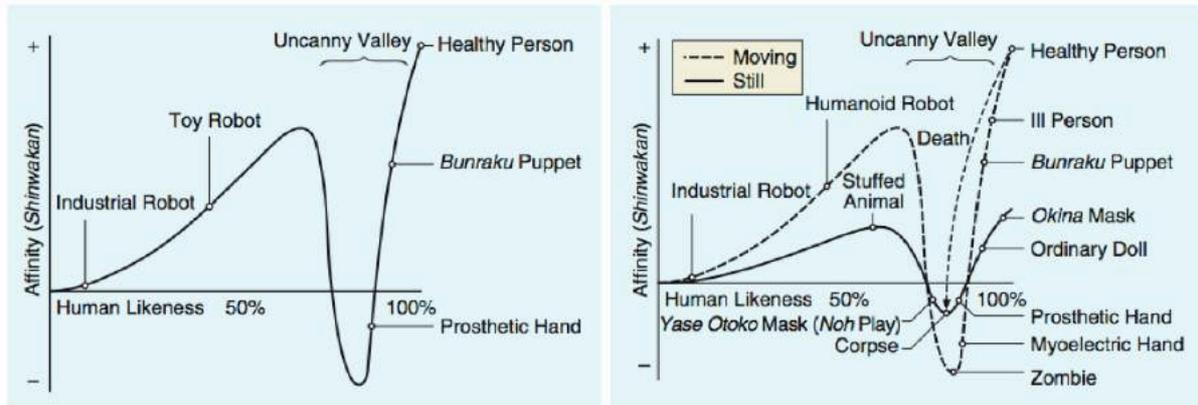


Figure : Représentation de la vallée de l'étrange par Masahiro Mori (à gauche version 1970 – à droite version 2012)

À l'instar de Mori et dans l'engouement de la robotique actuelle, cette vallée de l'étrange lance de nombreuses réflexions au début des années 2000 (Brenton, Gillies, Ballin, et Chatting, 2005 ; Gee, Browne, et Kawamura, 2005 ; Pollick, 2009), en particulier autour des robots humanoïdes (Chaminade, Hodgins, et Kawato, 2007 ; Kosloff et Greenberg, 2006 ; MacDorman et Ishiguro, 2006). Cette notion ouvre également à de nouvelles réflexions modélisation bayésienne (R. K. Moore, 2012) intégrant de nouveaux facteurs d'influences tels que le mouvement et la modalité visuelle (Giese et Poggio, 2003), axée sur les effets de mouvements d'expressions faciales (MacDorman et Ishiguro, 2006), sous l'angle de la psychologie qui suggère que cette notion dépend de ce qu'il appelle un « continuum de réalisme » (Ramey, 2005) ou encore dans une perspective neuro-anatomiquement au niveau du cortex pariétal (Saygin, Chaminade, Ishiguro, Driver, et Frith, 2012) dont les liens veulent notamment être établis par aux effets des neurones miroirs.

De manière générale cette littérature explore les facteurs, autres que ceux du mouvement, qui peuvent être associés à l'apparence. Cette apparence influencerait notre degré d'anthropomorphisme perçu, dont les incongruences perceptives tendent vers cette vallée. En effet, les auteurs recherchent lesquelles de ces dimensions influencent notre perception, et *via* quels couplages et avec quels degrés d'influence.

En revanche, si un robot est d'une apparence complètement éloignée avec des comportements qui lui sont propres, mais qui véhiculent un fort potentiel empathique avec des effets de *patternicity*, ne pourrait-il pas changer entièrement la forme et la dynamique de la courbe qui converge vers cette vallée de l'étrange ? L'un des problèmes de cette représentation est qu'il s'appuie sur des comportements mimétiques, ce qui reste une

<sup>26</sup> Voici des exemples de robots dont l'apparence suggère cette vallée de l'étrange : (<http://www.strangerdimensions.com/2013/11/25/10-creepy-examples-uncanny-valley/>, consulté le 10/06/2016)

perspective roboticienne très japonaise. De plus, il s'agit d'une orientation centrée sur le mouvement alors que d'autres facteurs d'influence existent.

En effet, en supposant qu'une dimension mécanique soit incompatible avec des formes humaines ou animales connues, mais très cohérentes avec un nouveau design complètement éloigné de ce que l'on connaît, quelles en seraient les conséquences sur la représentation de Mori ? Par exemple, si nous nous intéressons uniquement à un couple humanoïde/chien, la queue du chien est un élément non présent sur l'apparence humanoïde, mais le mouvement global de la queue du chien, qui frétille lorsque ce dernier a l'air content et veut une caresse, fait partie d'une dynamique de corps globale qui pourrait être imitée par l'humanoïde. Or dans un cas, alors que l'affinité du chien avec l'humain risque d'augmenter, le mouvement incongru de l'humanoïde associé à ce même mouvement va très certainement avoir un effet différent, voire même inverse, sur le degré d'affinité. Sinon, il faudrait pouvoir trouver les équivalents de l'effet fonctionnel porté par le remuement de queue du chien qui aurait une autre forme pour un humanoïde (p. ex. sautiller ou lever les bras, mais encore faudrait-il perceptivement le valider). Ces comportements auront alors pour effets de faire produire une action à l'humain, qui aura certainement des variantes différentes de réponse (socioculturellement distincts), mais dont la fonction communicative resterait la même dans les deux cas.

Il y aurait donc au moins deux définitions par rapport à ce que l'on peut considérer comme des patterns. Les premières seraient des patterns qui sont des instances mimées ou modélisables, à greffer sur un système, ce que par exemple la grande majorité des travaux de *machine learning* récoltent dans les données du *Big Data*. Ainsi une grande partie des variantes portant l'information fonctionnelle associée à la donnée serait récoltée, déduisant de manière calculatoire des prototypes, mais à travers une boîte noire statistique. La seconde définition du pattern serait par la dimension qui produit l'effet fonctionnel attendu sur la perception humaine. Dans cette perspective, l'apprentissage pourrait être orienté par les fonctions connues qui permettraient d'éviter certaines incongruences.

L'apparence même du robot a donc un effet difficile à contrôler en termes de perception humaine, où l'anthropomorphisme peut devenir un biais expérimental important. Ainsi, si l'apparence du robot suggère une ressemblance avec un élément qui a une existence socioculturelle pour l'humain, sa perception sera modifiée par le biais de ce filtre, qui sera d'autant plus important si le degré d'anthropomorphisme qui lui est associé est grand. Il aura par conséquent des effets communicatifs et manipulatoires importants.

La partie suivante s'intéressera donc à la dimension empathique pour mieux appréhender ce que Mori a défini sous les termes d'affinité qui se crée par l'illusion perceptive que l'humain redéfinit par un lien d'attachement découlant de cet anthropomorphisme (Chaminade et coll., 2007 ; Duffy, 2003) et qui fait notamment partie des enjeux de la robotique sociale actuelle.

### 1.4.3. De l'engagement à l'attachement : vers une « illusion empathique »

Dès lors que le robot tend à être envisagé comme un autre, il semble susciter des effets d'empathie chez l'humain. Dans le cadre de la robotique (Duffy, 2003) rapporte que cette empathie doit être exploitée à juste niveau d'illusion, donnant ainsi lieu à une forme d'« illusion empathique » (Aubergé, 2016). Les illustrations de tels phénomènes sont nombreuses dont nous pouvons en citer certaines :

1. des expériences qui demandent de brutaliser le robot Pleo provoquent des réactions de réticences voire de refus (Darling, 2012 ; Rosenthal-von der Pütten, Krämer, Hoffmann, Sobieraj, et Eimler, 2012),
2. faire ressentir une gêne lorsque l'on demande de toucher des « parties intimes » du robot Nao et peut avoir certains effets analogues que chez l'homme sur les mesures biophysiques<sup>27</sup>,
3. le déclenchement des boucles empathiques similaires à celles qui sont suscitées à la vue d'un humain lorsqu'il voit les doigts d'un robot androïde se faire trancher (Suzuki, Galli, Ikeda, Itakura, et Kitazaki, 2015),
4. des réactions acerbes d'internautes défendant les « pauvres robots » apparaissant dans les démonstrations technologiques de Boston Dynamics. La société avait pour but de montrer la robustesse des fonctions de déplacement de son robot Atlas, ou encore la capacité de maintien de sa position érigée et ses compétences préhensiles développées. Elle plaçait alors en position de difficultés son robot (le pousser, déplacer l'objet qu'il essaie de prendre, de leur donner des coups de pieds etc.)<sup>28</sup>,
5. de ressentir une souffrance et même de trouver la « situation inhumaine » (propos du colonel des *US Army*) à la vue des robots démineurs Packbots de Tilden, désintégrés morceau par morceau. La situation lui renvoie sa propre image de supérieur qui envoie un soldat pour une mission suicidaire (Tisseron, 2015).

Cette empathie est à la base des interactions qui vont s'établir avec le robot, dont les conséquences aboutiront à la notion « d'attachement ». C'est d'ailleurs autour de cette notion que la robotique s'appuie pour faire la différence entre objet animé et robot (Kidd et Breazeal, 2004).

À partir du moment où l'attachement se crée par des interactions, l'humain et le robot sont actifs. Des manifestations transparaissent donc par les décisions que prennent chacun de ces interactants l'un envers l'autre, faisant appel à une forme de conscience morale qui les lie entre eux, pouvant même aboutir à la notion de confiance ou *trust*, entre l'humain et le robot qui d'un point de vue expérimental est encore un point peu étudié (Billings, Schaefer, Chen, et Hancock, 2012). Cette confiance semble rattacher une certaine dimension de synchronisation entre l'humain et le robot (Billings et coll., 2012 ; Avril, Chetouani, et Sabouret, 2013 ; Tisseron, 2015) et en particulier sur l'aspect collaboratif du robot (Hoffman, 2013) à travers la notion de *fluency*. (Freedy, DeVisser, Weltman, et Coeyman, 2007) l'abordent sous l'angle de la modélisation. (Bauer, Wollherr, et Buss, 2008) ainsi que (Green, Billingham, Chen, et Chase, 2008) présente quant à eux un état de l'art de la robotique collaborative (également à travers la réalité virtuelle). En effet, la manière dont les robots réagissent selon la durée d'interactions (mais les effets semblent visibles dans de courtes

---

<sup>27</sup> Étude de Jamy Li, Université de Stanford, dont la méthodologie expérimentale reste critiquable. Article de vulgarisation dont et vidéo accessibles sur :

(<http://spectrum.ieee.org/automaton/robotics/humanoids/stanford-touching-nao-robot>, consulté le 15 mai 2016)

<sup>28</sup> Article de presses avec les accès aux vidéos qui ont suscités les réactions d'internautes : (<https://vieartificielle.com/empathie-robot-atlas>, consulté en février 2016)

interactions) en fait des entités dont l'humain commence à appréhender un certain schéma de comportements. Il semble alors anticiper les comportements du robot. La déviation par rapport à cette capacité d'anticipation serait un point de mesure à considérer pour quantifier la confiance de l'homme vis-à-vis du robot.

Cet attachement est par ailleurs fortement connecté à la notion d'*engagement* dont les principaux auteurs en robotique sont (Sidner, Lee, Kidd, Lesh, et Rich, 2005). Ce terme est largement utilisé dans la communauté, parfois même sous la forme d'un mot-clé qui désigne un paradigme d'étude qui semble encore être balbutiant<sup>29</sup>. Ainsi Sidner et ses collègues rapportent que dans la communauté robotique, le terme *engagement* a certainement été inspiré par Alan Bierman lors de la conférence User Modelling de 1999 puis qu'il aurait rapporté en 2002 l'idée suivante :

*"The point is that when people talk, they maintain conscientious psychological connection with each other and each will not let the other person go. When one is finished speaking, there is an acceptable pause and then the other must return something. We have this set of unspoken rules that we all know unconsciously but we all use in every interaction. If there is an unacceptable pause, an unacceptable gaze into space, an unacceptable gesture, the cooperating person will change strategy and try to re-establish contact. Machines do none of the above, and it will be a whole research area when people get around to working on it".*

L'*engagement* semble ainsi désigner des productions multimodales, en particulier le regard et la gestualité (mais pas exclusivement) qui se caractérisent comme des comportements qui amorcent, maintiennent et désengagent dans une interaction. Ils sont définis comme des « comportements d'engagement » par (Sidner, Kidd, Lee, et Lesh, 2004). Ce phénomène illustrerait une connexion établie au minimum dans une dyade, mais aussi dans de plus grands groupes d'humains avec un ou plusieurs robots (Matsuyama, Akiba, Fujie, et Kobayashi, 2015), mais désignerait surtout le processus qui permet l'établissement de cette connexion. Selon Sidner et ses collègues, ces interactions orientent alors l'observation sur deux axes : les capacités communicatives et la dimension collaborative de ces robots. Parmi les capacités d'engagement, qui selon les auteurs est avant tout linguistique, ils relèvent également les gestes, et les définissent par trois composantes : 1) les comportements d'engagement, 2) le management de la conversation (déterminer les intentions et buts dans le contexte interactionnel afin de les relier à l'état attentionnel des sujets), 3) des comportements collaboratifs, soit choisir que dire et que faire dans la conversation pour favoriser les objectifs partagés de l'humain et du robot, tout en interprétant les productions humaines.

Enfin, un autre domaine dans lequel cette notion d'*engagement* apparaît est celle de la psychosociologie qui prend pour point de départ la théorie de la dissonance cognitive de (Festinger, 1954). Cette conception s'inscrit dans le paradigme dit de soumission forcée (ou *forced compliance*) qui s'observe dans des cas de communication dite paradoxale où les études s'intéressent au changement d'attitude qui s'observe après un comportement contraire aux attitudes des sujets (dit comportement contrattitudinal, p. ex. tenir un propos contraire à ses convictions). Le cumul de comportements modifierait alors la dissonance globale, la rendant plus forte si ces comportements sont plus nombreux (Joule et Azdia, 2004). L'intérêt de ce concept est que là où la robotique essaie de trouver les éléments influençant l'*engagement*, la théorie de la dissonance cognitive suggère que certains points de rupture interactionnels engendrent des effets observables par des changements attitudeux qui modifie globalement la perception de l'état d'*engagement*. Cela suggère donc qu'il existerait des traces de ces points de rupture, en particulier ceux qui s'accompagnent d'un changement attitudeux qui est susceptible de décrire ce changement de niveau d'*engagement*.

---

<sup>29</sup> Cf. A l'école d'été *social HRI 2015* S. Thill ou encore T. Belpaeme invitait la communauté des jeunes chercheurs à enrichir ce paradigme que ce soit à travers l'angle des sciences cognitives ou en fonction de l'aspect collaboratif que le robots pouvaient avoir avec des publics spécifiques (enfants, personnes âgées mais pas seulement), avec pluridisciplinarité pour mieux définir ce que l'on appelle *engagement* en robotique sociale.

Cette notion d'engagement s'inscrit dans une conception beaucoup plus globale qui sous-tend une tendance qui concerne l'innovation technologique de manière générale, en regard de la norme publique et de ses représentations face à un *gold standard* (Wilkinson, Bultitude, et Dawson, 2011). La robotique qui rentre dans cette tendance peut associer à l'engagement une orientation spécifique selon trois configurations selon (Corrigan et coll., 2013) : 1) le *social engagement* (cas d'engagement entre deux interactions) ; 2) *task engagement* (l'engagement du sujet vis-à-vis de la tâche) ; 3) *social-task engagement* (formée de la triangulation humain-robot-tâche avec focus d'engagement bidirectionnel entre les trois). Les auteurs se penchent sur le troisième point dans l'étude proposée, qui décrit implicitement les deux premières situations. En revanche, cette conception dénote qu'il peut y avoir des changements de focus attentionnel des sujets et des robots qui influence l'engagement global tout en modifiant ponctuellement des dimensions par ces changements d'orientation.

Dans une approche non restreinte à la robotique, qui n'extrait pas spécifiquement les tâches exécutées, la notion de *user engagement* est utilisée. (Attfield, Kazai, Lalmas, et Piwowarski, 2011) en proposent huit caractéristiques :

1. *focused attention* : elle traite de l'orientation de l'attention, mais également de la perception du temps (*SPT : subjective perception of time*) en fonction du degré d'implication et d'attention,
2. *positive affect* : qui intègre une dimension de perception positive qui a des influences sur l'acuité et la motivation de la personne,
3. *aesthetics* : une composante impressionniste vis-à-vis de l'apparence perçue de son interactant,
4. *endurability* : rattachée à la mémoire d'une expérience antérieure qui compare l'état d'engagement actuel en l'orientant en fonction de ce vécu,
5. *novelty* : il s'agit d'un point intrinsèque à toute technologie qui peut possiblement maintenir alerte et curieux, sachant que cette dimension peut jouer sur la durée de familiarisation,
6. *richness and control* : elle relève du potentiel de l'objet en tant qu'outil technologique par le service qu'il apporte, mais il fait également la dimension de contrôle que l'utilisateur conserve pour utiliser ce service,
7. *reputation, trust and expectation* : il s'agit d'une résultante globale issue de l'expérience confrontant les comportements de la technologie à l'humain,
8. *user context* : qui rappelle que les interactions n'ont lieu que dans le cadre d'échanges situés.

Cet engagement a donc de multiples traces comportementales multimodales : regard, gestualité, expressions faciales, déplacement, position, proxémie, kinésie, voix, intonation, objets linguistiques, éléments dits paralinguistiques ou non lexicalisés (Malle et Pearce, 2001 ; Anzalone, Boucenna, Ivaldi, et Chetouani, 2015). De plus, (Malle et Pearce, 2001) considèrent que ces éléments sont observables ou non, intentionnels ou non (c.-à-d. ils croisent les propriétés pour les observer selon ces quatre catégories). Ils relèvent ainsi la difficulté d'accéder aux états mentaux des autres. En s'appuyant sur trois expérimentations, ils montrent que selon le profil des personnes (acteurs/observateurs), l'attention est portée sur des événements de natures différentes, les acteurs étant beaucoup plus attentifs à des éléments non ou moins observables. Inversement pour les autres, ils les relient au rôle de l'attention dans les phénomènes d'attribution et les comportements interpersonnels par une approche qui s'appuie sur des notions d'intimité et d'empathie. Cette étude souligne également un besoin de réflexion autour des aspects méthodologiques afin d'aborder cette notion d'attention. En effet, elle véhicule beaucoup d'implicites alors qu'elle reste incontournable pour les études autour de l'engagement en robotique. Cette remarque semble d'ailleurs valable pour d'autres disciplines telles que le traitement automatique des langues ou l'*Affective Computing*.

Lorsque la robotique cherche à mesurer et modéliser les traces de cet engagement à travers des métriques (Lehmann, Lalmas, Yom-Tov, et Dupret, 2012 ; Anzalone et coll., 2015), elle s'intéresse à un espace partagé d'interaction ou *shared social space* (Rauterberg, Dätwyler, et Sperisen, 1995) qui est une notion bien connue dans le cadre des interactions entre humains. L'un des modèles les plus connus est celui de (Hall et coll., 1968) qui définit des classes de distances : intime (0 à 18 pouces soit 45 cm), personnelle (18 pouces à 4 pieds soit à 122cm), sociale (4 pieds à 10 pieds soit 305cm) et publique (plus de 10 pieds), classés ici de la plus proche à la moins proche sur des mesures absolues de distances. En revanche, ces classes sont culturellement différentes comme relève (Patterson, 1968). Cette dimension de proxémique (Walters et coll., 2005) est également applicable dans des interactions avec un robot. Une exploration de différentes modélisations de cet espace d'engagement a été effectuée dans le cadre de la robotique sociale, à travers les comportements d'un robot réceptionniste (Michalowski, Sabanovic, et Simmons, 2006). Quels que soient les auteurs, l'engagement semble converger vers la notion d'attention qui est un point expliquant l'analyse multimodale appliquée en robotique.

Ainsi (Saito et coll., 2010) part de la définition selon laquelle l'humain possède une forme d'attention sociale permettant de détecter des focus attentionnels d'autres, orienter notre attention sur cette cible et en faire des inférences sur les objectifs partagés (Nummenmaa et Calder, 2009). L'une des théories qui a ainsi influencé le plus grand nombre de modélisations et certainement celle de (Baron-Cohen, 1995) focalisant sur le concept de l'attention conjointe ou *joint attention* dont les composantes seraient :

- la détection d'intention (ou *intention detection – ID*) : considérée comme équivalente à la notion d'animacité qui sous-tend la capacité d'identifier quelque chose comme un agent,
- la direction du regard (ou *eye-direction detection – EDD*) : qui dépend d'une représentation dyadique de la relation entre un agent et un objet à travers un schéma intentionnel dirigé par la direction du regard,
- un mécanisme d'attention partagé (ou *shared attention mechanism – SAM*) : qui est une intégration de la représentation dyadique précédente avec la perception de soi à cet instant, pour en donner une représentation tripartite de la situation,
- et des mécanismes relevant de la théorie de l'esprit (*theory of mind mechanism – ToMM*) : la théorie de l'esprit est décrite comme la capacité à attribuer des états mentaux à un autre, qui est également une conception très neuro-anatomique et relevant de compétences liées à la perception de comouvement (Premack, 1990). Celle-ci ne doit pas être confondue avec la notion d'empathie qui est définie comme la capacité de comprendre les états mentaux d'un autre en projetant que l'on est cet autre (cf. section 1.5 pour les détails de l'empathie dans l'interaction). C'est cette empathie et non par la théorie de l'esprit qui fait l'objet de notre approche.

Dans l'interaction sociale, l'approche de la coordination d'« attention » s'illustre par la convergence de l'éthologie, la sociologie et la psychologie du développement (Conein, 1998), dont l'auteur donne une perspective transversale de ce qu'il appelle les « sens sociaux ».

La sociologie a en effet étudié cette question à travers les travaux de (Goffman, 1961, 1963/2008) autour d'une conception d'interaction dite proto-sociale. Elle se caractérise par une coprésence physique (proxémie qui n'est pas obligatoirement associée à un partage attentionnel d'où l'usage du terme *unfocused interaction* par l'auteur). Les relations de proximité découleraient de l'interaction en passant par l'orientation des visages, du corps et des regards qui formeraient ainsi un contact attentionnel partagé. À partir de ces deux formes d'interaction, il définit deux types d'attention : une attention non coordonnée (dans un espace sans engagement) et une attention mutuelle. Cette approche converge les observations de la psychologie du développement. (Gómez, 1991) définit deux types de regards : un « regard déictique » où le regard de l'un est dirigé sur l'objet du regard de l'autre et un regard

bidirectionnel qu'il nomme le « regard mutuel ». Ces regards changeraient la relation avec l'objet lorsque ce dernier est animé, par une coordination mutuelle, qui n'existerait pas dans l'objet inanimé et qui s'appuierait sur un regard global. L'attitude peut alors être double, soit regarder sans être regardé avec un autre qui regarde aussi, soit regarder l'autre en retour. Dans ce second cas, le regard serait beaucoup plus focalisé sur le visage et les yeux comme dans une perspective gestaltiste. D'après (Murray et Trevarthen, 1985 in Trevarthen, 1998), les enfants de 2 mois réagissent négativement lorsque sur la présentation de films avec leur mère, les contacts visuels et les productions verbales ne sont pas coordonnés avec les leurs. Elle est également à rattacher aux expériences dites de *still face* ou l'étude des visages impassibles. Ce paradigme est proposé par (Tronick, Als, Adamson, Wise, et Brazelton, 1978). Une interruption des marques du partage d'attention de la mère qui adopte un masque d'inexpressivité à son enfant, avec lequel elle communiquait précédemment, va perturber le nourrisson et cela très tôt, dans les premiers mois de vie (cf. Mesman, van Ijzendoorn, et Bakermans-Kranenburg, 2009 pour un état des lieux des paradigmes associés).

Il existerait ainsi une socialité primaire qui s'appuie sur le mécanisme d'attention mutuelle adressant une intersubjectivité qui serait une forme primitive de l'interaction sociale, et cela durant les six premiers mois de vie. En revanche, une intersubjectivité secondaire apparaîtrait par la suite avec de nouvelles formes de coordinations. L'apparition des regards déictiques donnerait alors des formes dites d'attention conjointe, où l'autre avec qui l'on interagissait devient un point de passage vers une autre cible, qui n'est plus le regard de l'autre, mais celui de sa cible. En revanche, elle peut refléter un autre aspect qui illustrerait une forme d'anticipation de l'attention d'autrui. D'après (Butterworth, 1991), l'acquisition de ce type d'attention conjointe est progressive. À six mois l'enfant suit le regard de la mère qui regarde un objet, sans lui-même le regarder. Il suit également le regard de la mère lors de la survenue d'un événement pour progresser et construire sa référence sociale ou *social referencing* (Campos, Stenberg, et al, 1981). Les mécanismes attentionnels se construisent donc en différentes étapes :

- attention mutuelle (ou *mutual attention*) : ajuster son regard vis-à-vis d'un autre en situation face à face,
- attention conjointe (ou *joint attention* – en deux phases) : une première phase qui suit le regard de l'autre (phase 1) puis focalise sur la cible du regard de l'autre (phase 2)
- attention contrôlée (ou *controlled attention*) : contrôler la direction du regard de l'autre par son propre regard qui fixe une cible (capacité apparaissant vers les 18 mois).

L'intégration des capacités qui suivent l'acquisition de l'attention conjointe ouvre alors les situations communicatives à la base des mécanismes d'émergence de groupes sociaux. Nous n'aborderons pas ici ces aspects en détail, mais ils peuvent commencer à être abordés par des approches éthologiques et sociologiques par des auteurs tels que (Byrne et Whiten, 1988/1997 ; R. I. M. Dunbar, 1988 ; Heyes, 1998 ; Kummer, 1967 ; Wrangham, 1987). Dans ces conceptions, l'acquisition des compétences sociales amène à produire des attitudes de l'ordre de la tromperie (dissimulations, distractions, simulations, manipulations) qui dérivent d'une conception minimaliste (approche psychologiste) ou maximaliste et dominante (approche éthologique). Dans la première approche, les aptitudes sociales développées seraient des formes primitives de coordination interpersonnelle dont les aptitudes simples seraient innées (Jackendoff, 1992/1994). Dans la seconde approche, ses aptitudes sociales seraient à la base du développement de l'intelligence, qui d'ailleurs intègre cette perspective dans une hypothèse d'intelligence machiavélique dans sa description d'attitudes de tromperie (Byrne et Whiten, 1997 ; Humphrey, 1976). Ces compétences sociales relèvent alors de l'influence de l'intentionnalité de l'attention sociale, qui pour la majorité des psychologues ne serait pas un prérequis et qui relève ainsi des prémisses de la théorie de l'esprit. Au contraire, (Leslie, 1995) considère que l'attention conjointe nécessite des méta-représentations de

l'intention d'autrui à travers la notion d'*agency*. (Harcourt, 1988 ; Cheney et Seyfarth, 1992 ; C. Moore et Barresi, 1993) suggère que cette considération d'autrui dans un processus de coordination relève d'une prise de perspective, soit de passer d'une relation egocentrée à une relation exocentrée, par un changement successif dans le cadre l'attention conjointe. En effet, lorsque nous tournons la tête ou le regard pour suivre la cible de l'autre (perspective exocentrée), nous sommes ensuite amenés à poser notre regard sur cette cible, qui devient alors notre propre cible de regard (perspective egocentrée). Par opposition, il existe des différences d'intentionnalité dont la conception est différente dans la théorie de l'esprit. En effet, cette dernière introduit une nature intentionnelle avant même que l'on puisse avoir une représentation des états mentaux d'autrui. C'est notamment par ce biais que la notion de comouvement rejoint la théorie de l'esprit, pour laquelle (Premack, 1990) suggère que la perception du mouvement d'éléments intentionnels primitifs ne requiert pas de lecture mentale de l'autre. Il s'appuie ainsi sur une conception comme celle proposée par (Marr et Vision, 1982) sous forme de progression linéaire. En effet, celui-ci proposait d'abord un module d'interprétation d'objets intentionnels, puis une attribution de propriétés d'interactions à ces objets, qui enfin étaient expliquaient en termes d'états mentaux.

Pour rappel notre approche empathique de la communication ne suit pas cette perspective motivant la théorie de l'esprit et nous tendons donc à poursuivre les perspectives fonctionnelles dont la partie adaptative semble avoir été restreinte dans les anciens courants, ce qui aurait amené à une compréhension réduite de la connaissance sociale. Ce mécanisme semble d'ailleurs intervenir lorsque la technologie et sa maîtrise sont les points culminants de la réflexion, à l'instar de l'aptitude à vivre en groupe (Humphrey, 1976). Cette dimension adaptative introduit alors des théories telles que ceux de (Conein, 1998 ; Strum, Forster, et Hutchins, 1997) : « la prise en compte de la dynamique des systèmes d'interaction implique un changement de l'unité d'analyse considérant que ces systèmes ont des propriétés cognitives. Ce changement d'unité d'analyse conduirait à abandonner une caractérisation individualiste des compétences sociales ».

Finalement, l'éthologie montre que son approche a finalement été dominante, car elle permettait d'allier des aptitudes sociales à un développement cognitif supérieur, soit l'intelligence.

## 1.5. Dimensions sociales pour une robotique interactionnelle

### 1.5.1. Empathie : un paradigme complexe à multiples facettes

La section 1.1 illustre des aspects du langage humain qui révèlent son caractère intrinsèquement social. Cette perspective peut ainsi engager cet humain dans une interaction bidirectionnelle dont le versant positif a été évoqué sous les termes « d'altruisme réciproque ». Cet altruisme émanerait du processus d'empathie, que nous avons vu émerger dans la section 1.4 en s'intéressant aux phénomènes impliqués dans la perception de soi et de l'autre à travers des notions associées au domaine de la robotique sociale. La présente section va ainsi préciser cet aspect de réciprocité positive pour construire les hypothèses qui sous-tendent ce travail de recherche.

Dans sa première utilisation, l'empathie définissait le « ressentir de l'autre de l'intérieur », comme proposait par le psychologue allemand Robert Vischer 1873 sous le terme *Einführung*, traduit ensuite en anglais par Titchener pour donner le terme *empathy*. À l'origine, ce terme désignait la projection mentale de soi-même dans un « objet extérieur, une maison, un vieil arbre nouveau ou une colline balayée par les vents » auxquels l'humain s'associerait subjectivement.

En 1903, le philosophe Lipps a ensuite donné une extension à cette définition en y ajoutant un point de vue artistique. L'empathie est devenue alors un moyen de décrire « le sentiment (d'un artiste) qui se projette par son imagination non seulement dans un objet animé, mais aussi dans l'expérience vécue d'une autre personne ».

Le bouddhiste Ricard<sup>30</sup> rappelle quelques éléments de terminologie à ne pas confondre tels que la « contagion émotionnelle » qui appartient au domaine de la psychologie des foules qui désignerait la propagation d'une émotion d'un individu à un autre selon (Pacherie, 2004). C'est un phénomène qui est notamment observé en psychologie du développement également, motivé par le processus de différenciation entre moi et autrui.

Une autre opposition terminologique par rapport à l'empathie est la « sympathie ». Ricard fait appel à la notion d'impartialité pour insister sur cette opposition, car pour lui la sympathie est liée à un individu spécifique. (Pacherie, 2004) rejoint cette différence en suggérant que la sympathie met en jeu des fins altruistes pour créer un lien affectif avec l'autre, alors que l'empathie ne relève que d'une compréhension d'autrui (Wispé, 1986), mais n'a pas pour objectif l'établissement d'un lien affectif. L'empathie pourrait donc nourrir la sympathie qui n'en est pas une résultante nécessaire.

L'empathie se distinguerait également du phénomène de « simulation mentale » qui désigne l'exercice où nous nous mettrions à la place d'un autre en faisant « comme si » nous étions cet autre pour essayer de le comprendre selon différents procédés : imagination (à travers la perception visuelle de l'autre), doxastique (attacher des croyances qu'on impute à l'autre), conative (comprendre les actions, les intentions et les désirs d'autrui). Cette simulation mentale serait alors confondue avec l'empathie quand cette dernière porte sur la compréhension de la partie affective et émotionnelle d'un autre.

Selon Ricard, l'empathie se déclencherait d'une part par la perception affective du ressenti de l'autre ; d'autre part par l'imagination cognitive de son vécu. Elle aurait deux composants : le premier, affectif, aurait un fonctionnement rapide qui apparaîtrait vers l'âge de un an. Le second, cognitif, sera un processus lent, délibératif et conscient pour comprendre aussi son point de vue en regard de la compréhension de celui de l'autre, qui survient vers l'âge de quatre ans et demi – dit aussi « compréhension empathique ». Ainsi l'empathie pourrait conduire à une motivation altruiste ou un sentiment de détresse lié à un évitement qui s'illustrerait par un repli sur soi afin de se détourner des souffrances de l'autre. Ces

---

<sup>30</sup> Billet de blog <http://www.matthieuricard.org/blog/posts/empathie-altruisme-et-compassion-1>

mécanismes dépendraient du versant affectif de l'empathie (Hoffman, 2008). Il peut également avoir un versant néfaste à l'empathie cognitive par l'usage d'une manipulation qui profite de la compréhension de l'autre, et qui n'aboutit pas forcément à de l'altruisme. De fait selon Hoffman, il existerait une empathie dite « mature » qui permettrait de réguler les effets de l'un et de l'autre pour maintenir un équilibre entre ces deux formes d'empathie. Cette conception est également celle qui est partagée par (Decety et Lamm, 2006).

Notons que (Tisseron, 2013)<sup>31</sup> souligne que la définition de la « compassion » selon Ricard est spécifique à la perspective bouddhiste et s'écarterait de la conception ternaire de l'empathie. L'empathie affective deviendrait alors majoritaire en ne laissant pas de place pour l'empathie cognitive, de même que dans la conception du modèle empathique défini par la neuroscience : pour (Bernhardt et Singer, 2012), les deux formes d'empathie ne peuvent coexister sous un même terme, car ils engendreraient l'activation de différentes zones du cerveau. Ainsi la compassion pour Ricard est une forme d'altruisme quand il y a confrontation à la souffrance d'autrui. Il caractérise ainsi l'empathie comme le fait de rentrer en résonance affective, soit une prise de conscience cognitive de la situation qui nous alerterait sur la nature et l'intensité des ressentis de l'autre, sans pour autant forcément ressentir la souffrance associée à la compassion. Nous voyons donc que l'empathie peut être adoptée selon différentes perspectives.

(Batson, 2009 in Decety et Ickes, pp.3-15) aborde la question de l'empathie en posant deux questions<sup>32</sup> qui pour lui sont celles qui orientent les divers travaux :

- Comment quelqu'un peut-il savoir ce que pense et ressent une autre personne ?
- Qu'est-ce qui pousse une personne à répondre avec sensibilité et attention à la souffrance d'un autre ?

La première attirerait l'attention de la philosophie, des sciences cognitives, de la neurophysiologie, la primatologie ou encore la psychologie du développement dont les travaux tendent vers la théorie de l'esprit. La seconde attirerait davantage les philosophes, les psychologues du développement et la sociologie pour converger vers la notion de *prosocial action* (p. ex. Eisenberg, 1987). Il mettrait ainsi en contraste d'une part les « théoriciens de la théorie » et les « théoriciens de la simulation ». Ainsi, ces communautés tendraient à définir l'empathie sous huit angles distincts :

1) "*Knowing a person's internal state, including his or her thoughts and feelings*"

Connue également sous le nom d'« empathie cognitive » ou *cognitive empathy* (Eslinger, 1998 ; Zahn-Waxler, Robinson, et Emde, 1992), elle reflète l'idée de justesse ou de précision de la dimension empathique, ce que (Ickes, 1993) appelle *empathic accuracy*. Certains caractérisent encore cette perspective comme un état interne de l'empathie (Preston et de Waal, 2002 ; Wispé, 1986). Cette approche dépendrait des facultés à trouver les indices qui reflètent cet état interne, leur verbalisation facilitant l'accès à cette information ou nous confortant sur les interprétations et les inférences déduites par le comportement de la personne à qui l'on fait face. Cette empathie est d'autant plus facilement accessible, lorsque la personne avec laquelle on interagit nous est proche.

2) "*Adopting the posture or matching the neural responses of an observed other*"

Associée à la notion « d'imitation » (Lipps, 1903 and Tichener, 1909 in Batson, 2009 ; Meltzoff et Moore, 1997) ou de *motor mimicry* (Dimberg, Thunberg, et Elmehed, 2000 ; Hoffman, 2000/2010), elle correspondrait à ce que (Gordon, 1995) appelle *facial empathy*.

---

<sup>31</sup> Billet de blog [http://www.huffingtonpost.fr/serge-tisseron/lerreur-de-matthieu-ricard\\_b\\_9343194.html](http://www.huffingtonpost.fr/serge-tisseron/lerreur-de-matthieu-ricard_b_9343194.html)

<sup>32</sup> Voici les versions originales de ces deux questions proposés par Batson : "*How can one know what another person is thinking and feeling ?* " et "*What leads one person to respond with sensitivity and care to the suffering of another*"

(Preston et de Waal, 2002) proposent une théorie unifiée de cette empathie en s'appuyant sur des représentations neurales du mimétisme et basée sur le modèle de boucle perception-action (PAM ou *perception action model*). D'après ce modèle, percevoir l'autre dans une situation donnée amènerait automatiquement à faire correspondre son état neural à celui de l'autre parce que la perception et l'action font partie d'un même circuit neural. C'est notamment ce qui a été montré par la présence des neurones miroirs (Rizzolatti, Fadiga, Fogassi, et Gallese, 2002). En effet, lorsque l'on perçoit un autre en train de faire une action, notre zone du cerveau dédiée à effectuer cette action s'active alors que nous sommes uniquement en train de la percevoir sans la réaliser. Cette dimension semble d'ailleurs constituer une grande partie des modèles d'interaction, qui lui confère parfois trop d'importance selon Batson. En effet, l'auteur rappelle que le mimétisme n'est pas le seul mécanisme qui sous-tend les interactions humaines et que les enjeux seraient de savoir à quel moment cette boucle est préférentiellement activée et à quel moment elle est mise en arrière-plan pour privilégier d'autres mécanismes tels que l'utilisation de la mémoire ou les connaissances générales (Singer et coll., 2004 ; Tomasello, 2009). De plus, il est souligné que la question de l'anthropomorphisme se pose, car en tant qu'être humain nous avons cette capacité d'inférence face à d'autres espèces ; et l'on pourrait aussi ajouter des objets tels que le robot.

### 3) "*Coming to feel as another person feels*"

Cette approche semble relever une ambiguïté quant à savoir le degré de similitude de l'émotion ressentie de l'un par rapport à celle de l'autre. En effet, les partisans de cette définition dont les philosophes (Darwall, 1998 ; Sober et Wilson, 1999), les neuroscientifiques (Eslinger, 1998 ; A. Damasio, 2003 ; Decety et Chaminade, 2003) ou les psychologues (Eisenberg et Strayer, 1987/1990 ; Preston et de Waal, 2002) disent que dans la perspective de cette forme d'empathie, il n'est pas nécessaire que celui qui perçoit l'autre ressente exactement la même chose, mais plutôt une chose (généralement une émotion) similaire ou proche, sans forcément spécifier ce qui joue sur cette proximité. (Hatfield, Cacioppo, et Rapson, 1993) proposent que ce ne sont pas seulement une correspondance d'émotions, mais une captation des émotions de l'autre qui sous-tend cette notion d'empathie, impliquant d'avoir un ressenti partagé des mécanismes physiologiques mis en jeu. Ils mettent alors en avant la dimension de l'*arousal* (cf. section 1.3.4) qui serait semblable dans le cadre d'attitude empathique. C'est cette empathie qui est appelée *affective empathy* (Zahn-Waxler et coll., 1992) ou *automatic emotional empathy* (Hodges et Wegner, 1997) parfois confondue à la notion de contagion émotionnelle dans la perspective de (Hatfield et coll., 1993) par exemple.

### 4) "*intuiting or projecting one self into another situation*"

C'est la perspective initiale qui avait fait l'objet de la définition d'empathie qui serait mentionnée dans (Lipps, 1903) et (Titchener, 1909) selon (Batson, 2009 in Decety et Ickes, pp.3-15). Associé à la notion d'esthétisme, c'est la forme que (Wispé, 1986) appelle *aesthetic empathy*. C'est une perspective de simulation en imaginant ce que l'on peut ressentir si nous étions dans une autre situation, un autre corps, y compris inanimé. Cette perspective est présentée comme une alternative aux approches basées sur la théorie de l'esprit (cf. section 1.4) qui sous-tend que toute personne doit être apte à faire des hypothèses (Barrett-Lennard, 1981) sur le ressenti d'un autre.

### 5) "*Imagining how another is thinking and feeling*"

Cette approche conçoit ce qu'on appelle *perspective taking* (Stotland, 1969 ; Decety et Jackson, 2004 ; J. (1927-) A. du texte Cosnier, 1994). Il s'agirait de prendre en considération ce qu'une personne dit et fait dans une situation donnée tout en s'appuyant sur la connaissance que l'on peut avoir sur son caractère, ses valeurs et ses désirs. Il s'agit donc d'une forme de décentration cognitive où l'on imagine ce que l'on percevrait à la place

d'autrui. Le focus n'est pas tant sur le fait de savoir ce que l'autre peut ressentir, mais plus sur la sensibilité que nous pouvons développer en état de connaissance de la situation de l'autre. Cette perspective est illustrative d'une approche thérapeutique (Barrett-Lennard, 1981) en adoptant une forme d'attention particulière qu'il appelle *empathic attentional set*, ce que (Wispé, 1986) décrit comme *psychological empathy* ou encore *projection* selon (Adolphs, 1999).

### 6) "*Imagining how one would think and feel in other's place*"

Dans cette approche, l'empathie est définie par une certaine aptitude à prendre imaginativement le rôle d'autrui afin d'en interpréter ses pensées, ses sentiments et ses actions. À l'instar de (Mead, 1934) dont la terminologie est également reprise par (J. (1927-) – texte Cosnier, 1994), les auteurs parleront de *role taking*. Cette approche fait l'objet de beaucoup de confusions terminologiques avec le point 5, mentionnant la notion de *perspective taking*. Or la perspective du point 6 semble être celle qui est la largement admise comme étant du *role taking*. Ce concept est finalement proche du mécanisme décrit en point 4, dans le sens il s'agit d'un cas de simulation. En revanche, la distinction se fait sur les orientations, l'une esthétique et l'autre portant sur la relation interpersonnelle.

### 7) "*Feeling distress at witnessing another person's suffering*"

En anglais le terme *distress* est associé à cette perspective qui traduit le désarroi, le bouleversement. (Hoffmann, 1982) parle de *empathic distress* et (Batson, 1991) de *personal distress* prenant plus la mesure de l'effet que peut avoir la souffrance d'un autre sur celui qui la perçoit chez l'autre. Cette souffrance émanerait de celui qui perçoit l'autre en situation de détresse et ne concerne pas celle de la personne en situation de détresse.

### 8) "*Feeling for another person who is suffering*"

Cette approche est orientée vers la personne en souffrance et non sur une vision égocentrée. Elle s'illustre par une congruence du niveau de la valence émotionnelle de chacun des interlocuteurs (savoir si le ressenti est positif ou négatif). En revanche, elle n'est pas liée à la nature émotionnelle (p. ex. se sentir désolée ou avoir pitié quand notre interlocuteur est lui en peine ou qu'il a peur. Les deux états ont une congruence négative). C'est ce qui dérive plus communément de la notion de sympathie (Darwall, 1998 ; Eisenberg et Strayer, 1990 ; Preston et de Waal, 2002 ; Sober et Wilson, 1999 ; Wispé, 1986).

Une perspective intéressante abordée dans l'approche 2 de Batson, et qui a fortement influencé les pensées autour de la cognition sociale, est l'apport des travaux des neurosciences autour de l'empathie et des « neurones miroirs ». Decety retrace notamment de manière critique les tendances et ses propres expériences, car il existe une véritable controverse quant aux rôles et fonctions de ces structures.

En effet, les capacités socio-affectives ont suscité de nombreux intérêts de philosophes, de psychologues ou encore d'économistes, et dans le domaine des neurosciences sociales, les auteurs tels que ceux présentés dans (Decety, 2010; Nadel et Decety, 2002) étudient les mécanismes neurobiologiques qui sous-tendent ces capacités. Dans ces problématiques, l'une des premières réflexions porte tout d'abord sur les origines évolutives de l'empathie. Ces origines amènent à la compréhension de la biologie des fonctions des processus : 1) autonomes, 2) de systèmes endocriniens (p.ex. phénomène d'attachement impliquant ocytocine ou vasopressine), 3) homéostatiques qui répondent aux besoins de communication et de sociabilité sélective. Et certaines caractéristiques de ces trois types de processus seraient partagées chez les mammifères. L'évolution neurobiologique des êtres vivants aurait notamment permis de développer des outils pour percevoir, comprendre, prédire et répondre à des états internes d'autrui. Dans ces réflexions phylogénétiques, MacLean (1985) proposait que ce serait avec l'apparition des mammifères (par divergence aux reptiles) que les

premières observations d'émergence de l'empathie auraient été observées. Ainsi trois comportements clés et parentaux (l'allaitement, la communication audio-vocale entre la mère et l'enfant, ainsi que le jeu) propres aux mammifères seraient apparus. Ces comportements seraient fortement dépendants de la division thalamo-cingulaire du système limbique, ce dernier étant impliqué dans l'évaluation et la régulation des émotions. Les structures du système limbique seraient reliées, par des connexions réciproques en fonction de rôles en lien avec la motivation, au système nerveux autonome intervenant dans les actions qui favorisent les réactions d'adaptation de l'organisme aux changements de l'environnement (interne et situationnel). Le développement du système nerveux autonome lui-même serait donc intimement lié à ces structures limbiques modulant les comportements d'engagement social. Cette division a notamment permis la formation du néocortex préfrontal impliqué dans les processus d'acculturation familiale, ces comportements et les organisations cérébrales n'ayant aucun équivalent chez les reptiles.

Ce sont donc des études portant sur le traitement de l'émotion, qui dans un premier temps, aurait mis en évidence la similitude des mécanismes neuronaux d'un observateur témoin de la perception d'une émotion vécue chez un autre. Dans les travaux d'imagerie cérébrale, d'abord chez le singe puis chez l'humain, certains neurones sensori-moteurs, dans les cortex prémoteur ventral, moteur primaire et pariétal postérieur étaient préférentiellement sollicités dans les processus de résonance motrice et ils furent appelés « neurones miroirs » (Rizzolatti, Fogassi, et Gallese, 2001). Mais selon Decety, l'enthousiasme de cette découverte aurait également donné une vision simpliste selon laquelle ces neurones miroirs seraient à la base neurobiologique de la cognition sociale et notamment des phénomènes d'empathie. En effet, de nombreuses études sur ces neurones tendraient à devenir tautologiques et systématiques, car la tâche permettant leur activation conclurait sur l'implication de ces structures dans ces mêmes tâches. Or, ne serait-ce que quantitativement, seuls 17% des neurones des aires mentionnés seraient véritablement « miroirs » et certaines études iraient même à donner des contre-exemples d'activation de ces neurones (p.ex. Murphy, Nimmo-Smith, et Lawrence, 2003; Phan, Wager, Taylor, et Liberzon, 2002). De plus, « chez l'homme, cette région est aussi associée à des processus computationnels divers qui n'ont rien à voir avec des propriétés sensorimotrices, comme le contrôle cognitif, l'attention sélective, la sélection de réponses et l'inhibition sélective. » (Decety, 2010). L'imbrication des différents mécanismes nuance cependant les critiques par les observations faites en neuro-imagerie fonctionnelle. En effet, les travaux de (Ruby et Decety, 2004) montraient que des changements hémodynamiques de régions somato-sensorielles sont remarquables au-delà de la simple activation, notamment si l'on précise la « perspective » dans laquelle se place un sujet (c.-à-d. la sienne ou celui d'un autre). Ainsi, les patterns d'activation psychophysiologiques seraient similaires entre les situations où les personnes pensent à une expérience émotionnelle personnelle (avec également une forte implication du cortex somato-sensoriel) et une expérience équivalente chez un autre. Finalement, nous pourrions retenir que dans les multiples travaux, un raccourci est souvent pris en confondant comme « activité miroir » toute activation corrélée entre une expérience émotionnelle personnelle et l'observation du même état chez un autre. Par ailleurs, les tâches proposées semblent confondre l'imitation volontaire d'une expression émotionnelle et l'expérience subjective de cette même émotion. Ainsi, la distinction effectuée par (Decety, 2010) est que l'observation de circuits neuronaux partagés ne constitue pas forcément une preuve que des fonctions similaires découlent de l'activation commune de certaines régions.

Pour notre part, cette critique générale de Decety vis-à-vis des études portant sur les neurones miroirs et l'empathie est importante. Elle a une incidence sur les hypothèses portant par exemple sur l'émergence du langage (p. ex. Gallese et Goldman, 1998), issues de l'observation de ce type de processus miroirs. Selon ces hypothèses, la résonance motrice à l'instar de ces neurones miroirs aurait fait l'objet d'une exaptation (cf. section 1.1 – c.-à-d. un détournement de sa fonction première) en faveur du développement du langage humain.

Par ailleurs, des paradigmes d'étude qui semblent alors plus pertinents pour la compréhension de ces mécanismes sous-jacents seraient par exemple les approches de perception de la détresse d'autrui (Decety, Echols, et Correll, 2010) ou encore celles qui étudient l'impact des lésions cérébrales sur les phénomènes d'empathie et de sympathie (Hornak et coll., 2003). La détresse présente une fonction protectrice par alerte d'un danger pour l'intégrité physique d'une personne. Son affichage permettrait d'attirer l'attention et d'inciter des comportements de réconfort et d'aide auprès d'un autre. Chez l'homme, la modulation de cette perception de détresse serait non consciente et elle serait amplifiée ou inhibée en fonction du contexte social par les critères qui déterminent nos appartenances à un groupe ethnique. Ainsi des expériences à multiples amorçages (p. ex. exprimer les causes de la contraction du HIV de patients – drogué ou non – chez qui les expressions faciales de douleurs sont évaluées par des sujets) montrent que l'intensité de la douleur perçue et les modifications hémodynamiques sont graduellement différentes selon la relation ou la représentation morale de la situation sociale (Cheng, Chen, Lin, Chou, et Decety, 2010; Decety et coll., 2010).

Nous ne nous étendons pas sur les cas d'observations lésionnelles, puisque le champ est très étendu et que nous n'avons pas d'expertise suffisante pour évaluer finement ces études. En revanche, plus que les événements corrélatifs, un focus pertinent serait de pouvoir inférer à la fois le poids fonctionnel et le rôle causal de la région affectée. Il s'agirait alors de regarder s'il existe bien des déficits hétérogènes des phénomènes sous-tendant l'empathie, plus d'une altération unique (Decety, 2010).

Finalement toutes les approches abordées dans cette section caractérisent surtout le point de vue que l'on adopte. En effet, il peut être plus ou moins égocentrique, jouant sur la limite perceptive du soi et de l'autre (cf. section 1.4.2), dans une approche soit cognitive, soit affective. Dans le cadre de la robotique, la dimension empathique semblerait être double entre *role taking* et *perspective taking*. En étant simpliste, la première intégrerait la réflexion orientant la modélisation du robot qui ne reste qu'une simulation, la seconde illustrerait les effets que cette technologie aurait sur l'humain qui rentrerait alors dans ce mécanisme s'il arrivait à percevoir le robot comme un autre. De plus, les perspectives mimétiques et l'imitation seront aussi des indicateurs qui permettront de jouer sur les deux versants de ces approches empathiques.

### ***1.5.2. Spécifications de l'empathie dans l'approche robotique***

En robotique, certaines notions dépendant de l'empathie sont préférentiellement mentionnées. (Tisseron, 2015) s'appuie par exemple sur la perspective de Decety et considère que l'empathie est une « réaction affective qui serait déclenchée par la réaction émotionnelle d'autrui » ou encore une « reconnaissance et (une) compréhension minimales des états mentaux de cette personne ». Ainsi reconnaître autrui s'apparenterait à percevoir un être semblable à soi tout évitant une confusion avec cet autre avec lequel s'établirait un partage affectif. Ce mécanisme d'empathie résiderait ainsi à une simulation mentale de la subjectivité d'autrui. Celle-ci aurait alors deux composantes :

- la « résonance motrice » : imiter mentalement les mouvements et expressions des personnes avec lesquelles on interagit. Cette représentation n'est pas équivalente, mais n'est pas sans rappeler le fonctionnement des neurones miroirs précédemment introduits. Cette composante illustrerait ainsi une capacité automatique, peu contrôlable et non intentionnelle ;
- la « flexibilité mentale » : avoir conscience que c'est autrui qui agit et éprouver une émotion ; nous y rajouterons les socio-affects. Il s'agit là d'une composante contrôlée et intentionnelle qui serait propre à l'espèce humaine ;

De plus, ces composantes semblent être influencées par deux autres facteurs :

- la « contagion émotionnelle » : il s'agirait du précurseur du processus d'empathie chez l'humain. Elle a notamment été mise en évidence avec l'exemple de bébés qui pleurent lorsqu'ils en entendent d'autres pleurer, mais à différentes intensités selon la source de pleurs (Martin et Clark, 1982 ; Simner, 1971 in Missonnier, 2014) ;
- l'« éveil empathique » : il s'agit d'une forme non verbale de communication permettant de savoir qu'une personne est en danger. Il représenterait des fonctions partagées par mammifères et deviendrait alors un avantage sélectif pour la survie. De plus, il permettrait une compréhension des émotions d'autrui en influençant la motivation du désir d'aider.

Ces éléments, (Tisseron, 2015) les met en situation sous quatre formes d'empathie afin d'explicitier la relation homme-robot en suivant la chronologie de développement de l'empathie chez l'humain au cours de sa vie :

1) l'empathie émotionnelle ou plutôt la contagion émotionnelle (observée vers un an et demi) : le mécanisme pour apprendre à distinguer les limites de soi et de l'autre sachant que la frontière peut être floue parfois, et cela même à l'âge adulte. Pour l'homme vis-à-vis du robot, elle consisterait à voir à partir de quel moment un humain pensera que la simulation des produits socio-affectifs du robot sera ressentie comme de vraies émotions qui sont propres au robot. Ce serait le passage de « il a l'air \*\* » à l'affirmation « il est \*\* ». Du point de vue du robot, cette empathie qui n'est pas ressentie dépend de la capacité de programmation à reconnaître les indices de l'interaction permettant de simuler une attitude empathique dans le but de pouvoir maintenir une continuité interactionnelle. En simplifiant le problème, l'analogie est donc faite avec l'autiste qui peut détecter des traits et des objets d'interactions multimodales appréhendables sur les expressions faciales, la voix, l'intonation, les gestes etc. tout en ne comprenant pas leur signification. Le robot les produit donc en suivant des modèles qui guident son comportement au vu de ce qui est reconnu dans l'interaction.

2) l'empathie cognitive (qui apparaît vers quatre ans et demi) : elle considère que la représentation du monde change. Par exemple, si nous avons un objet en face de nous d'un point de vue égocentrique, nous en voyons la face avant, mais du point de vue de l'interlocuteur il s'agit de la face arrière. Chez l'humain, cette capacité développée tend à attribuer des intentions et des traits de personnalité à partir du moment où le robot est perçu comme un autre. Elle illustrerait ainsi des formes de personnalisation du robot, qui relève d'une perception synthétique d'un ensemble de comportements. Elle ferait ainsi référence au « je comprends les raisons \*\* » de la perception du point 1). Au contraire, de par sa nature, le robot fonctionnerait de manière analytique puisqu'il en est encore à détecter indice par indice les éléments qui sont susceptibles de faire sens, tandis qu'une information perçue par l'humain n'est pas la somme des informations existantes concomitantes (p.ex. effet McGurk<sup>33</sup>). Il faudrait donc pouvoir développer des modèles globaux permettant d'illustrer la fonction communicative qui synthétise l'ensemble de ces comportements détectés chez l'humain pour répondre de manière pertinente.

3) le changement de perspective émotionnelle (*perspective taking* – observation privilégiée entre 8 et 12 ans) : suite à la perception qui découle des points 1 et 2, indiqués par les \*\* dans les paragraphes précédents, cette dimension empathique peut être résumé par l'idée « à ta place, je le serais aussi ». Cette capacité de se préoccuper de l'autre nous

---

<sup>33</sup> McGurk, H. (1981). Listening with eye and ear. Cette étude a montré que pour la perception visuelle d'un son (p.ex. : /ga/) et la perception auditive d'un son (e.g. : /ta/), la modalité audio-visuelle du son perçu ne correspond ni à l'un ni à l'autre ni de leur addition mais relève d'une synthèse qui donne lieu à une nouvelle information (e.g. : /da/). Voici un exemple qu'il est possible de tester : <https://www.youtube.com/watch?v=jtsfidRq2tw>

permettrait de faire preuve d'une empathie morale. Considérée comme un phénomène naturel, elle serait néanmoins à cultiver par l'éducation pour qu'elle puisse persister durablement. Du point de vue l'humain, le réalisme des comportements du robot, la durée du contact ou encore les signes illustratifs d'une empathie pertinente et située mettraient en conflit d'une part la conscience, qui sait que nous avons à faire qu'à un objet, avec d'autre part les tentations qui veulent poser cette réciprocité relationnelle. Du point de vue du robot, elle n'est que simulation, mais elle relève aussi d'une importante réflexion éthique puisque le choix de simulation ne repose actuellement que sur le choix de l'humain concepteur doté ses propres motivations ou idéaux. Or cette décision de comportement est une construction sociale, comme nous le voyons à travers la dimension éducative. Il serait donc peut-être plus judicieux d'envisager cette perspective de manière collective au moment même de cette conception.

4) l'empathie altruiste et morale : elle émane de la coexistence dans un tissu social dans lequel l'humain s'oriente naturellement sur un « bien-vivre ensemble ». Le postulat serait alors que personne ne veut préserver une situation conflictuelle si elle n'y est pas contrainte. Cette dimension de l'empathie dépendrait de trois facteurs : l'estime de soi, la faculté que nous avons d'aimer et d'être aimé, et enfin les droits citoyens. Elle serait tantôt décrite par le terme de « reconnaissance », ou encore rattachée à la notion « d'intersubjectivité ». Pour le robot, elle se reflète à travers les débats émergents autour de droits aux robots comme proposés par exemple par (Darling, 2012) au MIT puis plus récemment en France à travers : le projet de loi pour les droits aux robots proposé par le cabinet Lexing Bensoussan<sup>34</sup>, les réflexions éthiques du CERNA<sup>35</sup> (Commission de réflexion sur l'Éthique de la Recherche en sciences et technologies du Numérique d'Allistène) ou encore les 5 commandements proposés par Laurence<sup>36</sup> (Devillers, 2017) faisant partie de cette commission proposant une révision des trois lois d'Asimov<sup>37</sup> comme étant l'une des premières propositions du genre à travers l'œuvre littéraire *I, Robot*. Malgré la nature d'une interaction réciproque, cette dernière reste simulée et le statut des interactants reste donc très clairement unilatéral, car l'intelligence artificielle est encore bien loin de développer une capacité de réflexion morale. En effet, l'humain peut en avoir cette impression du fait de sa perception des comportements de cette technologie que lui-même associe à des valeurs empathiques altruistes et morales, mais qui ne sont pas le fait du robot qui ne reste qu'un objet. Bien qu'il soit nécessaire de développer les réflexions éthiques en parallèle des conceptions et développements, le chercheur en robotique n'a de possibilité que d'évaluer les effets que cette technologie peut produire sur l'humain, ou encore de relever les effets inattendus qui n'étaient pas prévisibles, et cela qu'elles soient bénéfiques ou néfastes. Cette évaluation passe notamment par la méthodologie afin de permettre un accès partagé à l'évaluation des risques, en concertation avec les partenaires industriels et sociétaux des projets. À terme, l'idée serait de permettre à la société civile, au sein de laquelle ces robots sont voués à évoluer, de discuter des décisions éthiques, en connaissance des effets et risques que ces technologies peuvent véhiculer.

Nous ne développerons pas davantage cet aspect éthique puisque qu'elle ne fait pas partie du domaine de compétence de cette thèse, mais notons que cette démarche réflexive s'inscrivant dans cette orientation d'une « science responsable » se développe par exemple à

---

<sup>34</sup> Articles vulgarisés contextualisant la proposition du rapport : <http://blogs.economie.gouv.fr/les-cafes-economiques-de-bercy/droit-des-robots-bensoussan/>

<sup>35</sup> Premier rapport éthique de la commission : [http://cerna-ethics-allistene.org/digitalAssets/38/38704\\_Avis\\_robotique\\_livret.pdf](http://cerna-ethics-allistene.org/digitalAssets/38/38704_Avis_robotique_livret.pdf)

<sup>36</sup> Interview exposant ces 5 commandements : [http://www.sciencesetavenir.fr/high-tech/robot/oubliez-les-lois-d-asimov-le-robot-assistant-de-vie-devra-repondre-a-5-commandements\\_104992](http://www.sciencesetavenir.fr/high-tech/robot/oubliez-les-lois-d-asimov-le-robot-assistant-de-vie-devra-repondre-a-5-commandements_104992)

<sup>37</sup> Les trois d'Asimov : 1) un robot ne peut porter atteinte à un être humain, ni, en restant passif, permettre qu'un être humain soit exposé au danger ; 2) un robot doit obéir aux ordres qui lui sont donnés par un être humain, sauf si de tels ordres entrent en conflit avec la première loi ; 3) un robot doit protéger son existence tant que cette protection n'entre pas en conflit avec la première ou la deuxième loi.

travers la Chaire Robo'Ethics inaugurée en 2016 par la fondation Grenoble INP dont le porteur industriel est Partnering Robotics<sup>38</sup>.

### *1.5.3. De l'empathie à l'altruisme : création d'un lien social*

En laissant de côté les débats moraux, le lien entre empathie et altruisme relève de multiples perspectives à travers le mécanisme interactionnel qu'il implique dans le processus de communication. Selon (de Waal, 2008), l'altruisme peut être une forme issue de l'évolution des mécanismes de l'empathie, partant du fait que cette notion souffre d'une distinction entre fonction et motivation. Dans une perspective évolutionniste d'un altruisme réciproque, les explications de tels processus pourraient se défaire de la motivation, car celles-ci émergeraient des effets du comportement (Trivers, 2004).

Selon d'autres perspectives, l'altruisme est un terme que nous retrouvons par exemple dans les écrits d'Auguste Comte. Il propose une dichotomie entre altruisme et égoïsme par leur différence de motivation et de perspective. De même, du point de vue de la psychologie sociale, une première approche est de considérer l'altruisme comme un concept qui relève d'un versant opposé à l'égoïsme, ce qui semble en réalité être une perspective influencée par une vision occidentale qui aurait centré la notion d'intérêt au cœur de son fonctionnement sociétal. La société ayant besoin d'une norme consensuelle pour se développer selon (Badhock, 1986 in Moscovici, 2015), la culture dans laquelle le monde occidental s'est construit considère comme norme l'égoïsme. Aucune définition qu'elle soit scientifique ou morale ne relève d'un consensus observable autour de ce terme. Néanmoins, la psychologie aurait tendance à associer l'altruisme à une dimension de sacrifice du soi donc une atteinte à l'*ego*, d'où la perspective de cette discipline à assimiler cet altruisme à la notion de sociabilité.

En effet, (Fultz et Cialdini, 1991 ; Rushton, Chrisjohn, et Fekken, 1981) montrent que le comportement des personnes peut être variable en fonction de la situation dans laquelle elles se trouvent et que le concept serait donc difficilement attribuable à une personnalité ou à une caractéristique individuelle par exemple, car une même personne peut agir d'une manière non altruiste dans certains cas, alors qu'elle l'aurait été précédemment. De plus, les situations, et même les relations entre les individus, montreraient des régularités de ces comportements, ce qui est en faveur de l'idée que l'altruisme est fortement rattaché à la sociabilité, soit au processus qui fait entrer et maintenir une personne dans le cercle social auquel il appartient.

Ainsi la psychologie sociale décrit trois approches pour décrire l'altruisme selon (Moscovici, 2000/2015) : un altruisme participatif, un altruisme fiduciaire et un altruisme normatif.

L'aspect dit participatif apparaîtrait dans des sociétés bien intégrées où il existerait une bonne division du travail. L'altruisme interviendrait alors sur un versant coopératif avec un engagement poussé de chaque individu dans la perspective de survie de l'ensemble social auquel il appartient. En transposant ce comportement aux relations existantes dans la société, celles-ci s'apparenteraient à celui de la famille ou de la patrie par exemple, mais à moindre degré pour des relations amicales. Il s'agirait d'un altruisme qui n'est pas directement adressé à autrui, mais qui reflète une relation implicite qui nous lie de telle manière à avoir un soi confondu à l'autre. Cela conduirait même à leur permettre d'interchanger leur rôle (p. ex. : l'enfant et le parent durant leur cycle de vie). L'altruisme reflèterait ainsi un lien particulier tout en le symbolisant ce lien.

---

<sup>38</sup> Robo'Ethics, la 7ème chaire d'excellence industrielle de la Fondation Grenoble INP : <http://www.grenoble-inp.fr/actualites/robo-ethics-la-7eme-chaire-d-excellence-industrielle-de-la-fondation-grenoble-inp-776507.kjsp> et le porteur industriel Partnering Robotics avec le robot Diya-One purificateur d'air : <https://partnering-robotics.com/home-fr.html>

Cet aspect montre donc qu'il est difficile d'établir un lien sans avoir au préalable de rôle qui définisse la nature potentielle de la relation qui nous relie. Dans le cas d'un robot, qui n'a pas de lien défini au préalable, il sera donc important de lui construire un statut spécifique, qui ne remplace pas celui d'un homme puisqu'il n'en est pas un. Il devra ainsi pouvoir revêtir des caractéristiques qui lui sont propres, mais vis-à-vis desquels l'humain pourra trouver suffisamment d'utilité pour estimer qu'il soit possible de rentrer en relation avec cet objet. Ainsi ce serait au fil de leurs interactions que le statut et ce rôle initial vont s'affirmer, voire se transformer du fait de la nature de ces échanges. Ces interactions seraient ainsi une manière de réduire le gouffre qu'il y a entre l'*alter* (autre) et l'*ego* (soi), permettant de « créer un monde d'intersubjectivité » et de nourrir cette intersubjectivité à travers une écologie d'interactions propres à cette situation. Dans cette perspective, l'autre reste une entité différente de soi. Ainsi un certain degré de confiance va influencer le comportement par l'établissement d'une tension dynamique de la relation (qui peut donc évoluer positivement ou négativement c'est-à-dire vers un gain de confiance ou de méfiance). Elle orienterait ainsi ce que nous nous autoriserions vis-à-vis de l'autre, sachant que cette attitude et cette relation changent en permanence au fur et à mesure de l'interaction. C'est ce lien établi à travers cette incertitude variable que (Moscovici, 2015) appelle l'altruisme fiduciaire. L'interaction nous permettrait alors de réduire la distance relationnelle entre les entités qui interagissent.

Par défaut, cette distance serait relativement réduite sur son versant participatif puisque le lien symbolique qui existe entre les interactants donne déjà un certain niveau de confiance qui semble beaucoup moins sujet à être remis en cause que dans le cadre d'une interaction avec un inconnu. Cette dimension serait davantage susceptible au changement dans le cadre d'une interaction homme-robot, puisque le robot est un objet dont la nature même est différente de l'humain, et que les représentations que nous en avons sont encore balbutiantes. Ce dernier point peut être un avantage d'un point de vue expérimental en robotique puisque la mesure de son impact pourrait être évitée si la représentation socioculturelle du robot est encore faible ou nulle. En revanche, la tendance actuelle de l'hypermédiatisation de la robotique tend à biaiser cette perspective. Le désavantage d'avoir une telle représentation renforcée serait alors l'a priori qui en découlera, puisqu'il : 1) sera plus difficilement remis à zéro, ce qui complexifiera d'appréhender le point de départ de la construction relationnelle et 2) aura probablement des effets sur les mécanismes observés donc nécessiterait un étalonnage, dont la méthode serait à construire.

À travers cette approche, l'*alter* (autre) et l'*ego* (soi) sont ainsi dans une relation évolutive, l'*ego* semblant être rattaché à des motivations qui semblent vouloir trouver une satisfaction prenant la forme « se sentir bien lorsqu'on fait du bien aux autres ». De ce type de mécanisme, il en découle des observations telles que présentées par (D. Krebs, 1975 ; D. L. Krebs, 1970 ; Piliavin et Piliavin, 1972). Certains exemples illustrent ainsi que les sentiments de culpabilité ou de tristesse peuvent accroître la volonté des gens à vouloir aider les autres, car elles permettraient d'améliorer son propre état d'âme (Cialdini, Baumann, et Kenrick, 1981 ; Weyant, 1978).

Sans pouvoir conclure quoi que ce soit sur les mécanismes induits, ces observations montrent que l'altruisme peut avoir une dimension affective influente sur la motivation qui pousse les personnes à entrer dans une telle relation, ce qui ne sera pas à négliger.

Enfin, l'aspect normatif de l'altruisme est décrit comme « codifié dans le monde social » (Moscovici, 2002/2015). Cette perspective est dite « impersonnelle » dans le sens où elle fait intervenir un « troisième acteur » qui est la norme d'une société/culture. Les valeurs socioculturelles impliquent alors des représentations spécifiques des relations égoïstes et altruistes propres aux sociétés, en ayant un « répertoire de situations » qui définit ces relations. Ce sera d'ailleurs ce qui se rapproche de la définition du « rôle » qui sera un point très important de l'hypothèse que nous poserons. Ainsi, cette norme définit qui peut être aidé,

par quel moyen et dans quelle situation. (Rushton et coll., 1981) soulignent que l'aspect normatif de l'altruisme a un lien avec une norme de responsabilité, qui selon le degré d'appropriation facilite l'entrée dans une relation orientée par cet altruisme. Ainsi, une forme de sélection, qui motive les personnes à s'engager ou non dans une telle relation, semble émerger. Cette décision humaine porterait ainsi sur l'évaluation de la condition qui introduit la forme d'aide, inhérente à l'approche altruiste (Amato, 1986 in Moscovisci, 2015). Elle représenterait alors ce qui est communément appelé la « bonne foi » des personnes. En effet, (R. F. Kidd et Berkowitz, 1976) montrent que si une personne est dans une situation qui découle d'une conséquence néfaste de ses propres actes, elle se rend « fautive » de cette situation. Sa demande d'aide prendra ainsi moins de valeurs aux yeux des gens qui seront à leur tour moins enclins à y répondre positivement.

Cette perspective est peut-être un autre point d'entrée pour aborder la dimension éthique et d'usage dans le cadre de la robotique interactionnelle. En effet, les technologies sont de manière générale développées pour augmenter l'Homme portent donc une dimension d'aide. Le robot qui interagit avec l'humain ne peut pas avoir, de par sa nature, une telle évaluation morale. En revanche, ses comportements vont eux susciter cette dimension morale chez l'humain interagissant. Or ces comportements ne sont pas effectués du propre chef de la technologie, mais dépendent de la décision des humains développeurs qui vont associer tels ou tels comportements sur leur robot. Cette évaluation morale n'est donc pas imputable à un seul individu puisqu'elle est construite par la forme de responsabilité qui découle des valeurs socioculturelles. Cette décision devrait donc être collective, en accord avec cet aspect normatif de l'altruisme qui pourrait certainement avoir des principes de valeurs plus universels si ceux-ci sont partagés interculturellement.

### ***1.5.4. Altruisme : une dimension propre de la dynamique d'interaction ?***

En réalité, les « formes d'altruisme » introduites par (Moscovici, 200/2015) ne suggèrent pas véritablement des ontologies, mais elles font plutôt référence à des dimensions différentes du même processus de création relationnelle basée sur l'altruisme. Ces dimensions semblent d'ailleurs se raccrocher à des phénomènes relevés par la communauté robotique. De plus, ces aspects auront certainement un degré d'influence variable en fonction de la situation et de l'état de maturité de la relation entre les interactants.

Nous pourrions alors concevoir cet altruisme sous forme d'une dimension propre, à part entière, sur un continuum dont les extrêmes représentent :

- d'une part la relation entre deux personnes (fiduciaire), rattachée à la notion de responsabilité,
- d'autre part la relation à l'échelle de la société (normatif) qui elle serait associée à la notion de solidarité.

Une expérimentation où des passants sont amenés à ramasser un billet de banque<sup>39</sup> illustre que ces trois dimensions de l'altruisme sont modifiées et engendrent une dynamique qui est tributaire de la situation et des effets de *priming* (Bargh, Chen, et Burrows, 1996) ou amorçage. Dans cette expérience, le sujet est un passant avec lequel différents effets d'amorçage sont testés. Le premier cas est de lui dire qu'il est très aimable avec insistance, puis de le confronter à une situation – une rue non passante – où il pourrait avoir un comportement altruiste. Sans les normes de la pression sociale, cette action serait par exemple de ramasser un billet qu'un autre passant (complice) aurait perdu pour le lui rendre. Avec

---

<sup>39</sup> Une vidéo vulgarisation illustre cette expérimentation 2001. Cette dernière est accessible à l'adresse : <https://www.processus-coaching.fr/2015/11/12/exp%C3%A9rience-du-billet-perdu-dynamique-de-l-altruisme/> (consulté le 12 mai 2013)

l'amorçage positif du départ, le sujet est amené à agir de manière altruiste, alors que sans amorçage, le sujet dévie plus facilement de cette tendance altruiste et afin de garder le billet pour soi.

Selon (Batson, 1991) avoir la maîtrise du langage nécessiterait d'appréhender un autrui, qui est également le reflet de ce que nous percevons de nous à travers l'autre. Ainsi la conscience du soi est liée à l'*ego*, qui dans la philosophie occidentale a été fondée sur des principes d'universalité : *"Everything we do, including everything we do to benefit others, is ultimately done for our own benefit"*. Il s'agit de l'image d'un égoïsme exclusif sous-tendant les propos d'Aristote, St Thomas d'Aquinas, Hobbes, Bentham, Nietzsche ou encore Freud, qui a intégré les dimensions relevant de l'altruisme au fondement de la morale. Celle-ci prend alors deux perspectives de développement : que l'égoïsme peut avoir un côté vertueux (p. ex. Nietzsche, 1888-1927 ; Rand, 1964) ou que les bénéfices des actes tournés vers autrui est plus avantageux que les actes réalisés pour soi (p. ex. Bentham, 1789-1876 ; Maslow, 1968 ou Rogers, 1961) : *"Other people, however dear, are simply complex objects in our world – sources of stimulation and reward, of facilitation and inhibition – as we each seek our own ends. But from perspective of altruism, we have the potential to be more interconnected, more closely tied to one another."*

À partir de ces observations, nous allons essayer de formaliser la façon dont nous pouvons envisager le lien social, la relation, qui sous-tendrait les interactions impliquant l'humain.

### ***1.5.5. Robotique comme instrument de modélisation***

#### **1.5.5.1. Hypothèse de la « glu socio-affective »**

Les sections précédentes ont suggéré que des liens dynamiques pourraient se construire dans une perspective altruiste, contrairement à la prise en compte d'une réciprocité négative, qui s'appuie sur une dimension de dominance qui est l'approche classique des théories des émotions. L'altruisme pourrait alors être considéré comme une dimension à part entière d'un processus de création de la relation, qui est nourrie et remaniée à travers les interactions sous-tendues. La résultante de la dynamique des processus communicatifs mis en œuvre, pour remanier le schème relationnel, est la notion que nous proposons ici sous le terme de « glu socio-affective » (Aubergé, 2013) qui lie les interactants entre eux, malléable et évolutive, qui s'instancient par la variabilité contrôlable dans et entre les modalités des signaux sociaux, et qui se modifient avec la transformation de la relation.

Dès lors, cette thèse tentera d'appréhender les dimensions sous-jacentes à ces processus. Elle partira donc de l'hypothèse que cette « glu socio-affective » existe et peut être construite entre deux entités et évoluer au fil des interactions.

Il faut ainsi pouvoir commencer à établir les conditions initiales permettant d'établir cette relation et donc cette glu. Nous avons pu voir à travers l'aspect participatif de l'altruisme qu'elle implique tout d'abord l'attribution de rôles aux interactants. L'altruisme étant inhérent à une forme de motivation tendant à fournir une aide, il faudrait que la situation interactionnelle puisse appréhender cette dimension avec préférentiellement une utilité sociale. Cela implique l'existence d'écologies dans lesquelles les personnes sont dans une situation de déficit, qui aurait d'autant plus de sens à observer si elle permet d'illustrer une situation suffisamment contrastive pour rendre compte des mécanismes qui sous-tendent la construction relationnelle.

Le chapitre 2 tentera ainsi de trouver des écologies d'observation utile en faisant notamment un choix méthodologique d'observer des situations d'isolement relationnel, un problème qui est de plus en plus remarqué dans notre société en changement. Cette situation serait peut-être même l'une des raisons pour laquelle nous voyons un nouvel engouement pour la robotique, alors que celle-ci a toujours fait partie des « rêves humains », et cela bien

avant l'Antiquité. De plus, l'interconnexion des individus prend de l'ampleur à travers l'évolution des nouvelles technologies de l'information et de la communication, mais elle semble également être accompagnée de l'émergence de besoins motivés par la socialité de l'humain. Quelles que soient les raisons, nous constatons l'existence de situations où la dimension relationnelle de l'humain est mise à mal. Ces situations sont ainsi des niches d'observation scientifique méthodologiquement intéressantes et où l'apport de la recherche pourrait avoir un bénéfice dans la perspective d'amélioration de certaines conditions de vie dégradées.

### 1.5.5.2. Hypothèse de « prosodie pure » graduellement gluante

Le second point d'hypothèse développé dans cette section concerne les outils permettant de manipuler la glu socio-affective. En effet, toute interaction étant située, il n'est pas véritablement possible de retrouver une situation dite « neutre » tant sur le versant socio-affectif que linguistique. Les réflexions autour de la seconde hypothèse se construiront donc principalement à partir des notions abordées dans la section 1.3. Cette dernière a tenté de montrer que le support lexical modifie la prosodie liée aux propriétés intonatives intrinsèques de chaque langue, il devient alors compliqué de l'appréhender avec ce support linguistique doublement articulé, notamment si l'intérêt est d'occulter la prosodie inhérente aux contraintes des structures phonologiques et morphosyntaxiques de la langue concernée au profit de la dynamique socio-affective qu'elle véhicule. En effet dans l'exemple de (Fujisaki, 1988/1997), la distinction entre la déclaration et l'interrogation (qui sont par exemple représentées par des courbes différentes dans les différents modèles intonatifs des langues) semblait montrer que ces informations étaient discrètes. Or cet auteur disait aussi que le degré de doute, généralement associé à une intonation interrogative, serait quant à lui continu. Les tendances actuelles de l'*Affective Computing*, s'appuient sur des modèles construits par comparaison croisée d'états émotionnels produits en différentes situations énonciatives. Elles permettent de discriminer plus ou moins l'état émotionnel ou socio-affectif des énoncés, mais contournent la question portant sur quand et comment ces états se modifient, et dans quel but communicatif ils sont produits. Ces approches suggèrent donc que la prosodie socio-affective s'ajoute à la prosodie linguistique intonative en les considérant comme des événements discrets. Cependant, du point de vue la psychologie et de l'interaction, nous avons pu voir que la dimension socio-affective semblait être un phénomène continu et qui ne s'ajouterait donc pas, mais sous-tendrait l'objet linguistique. Dans cet ordre d'idée, (Scherer, 1984) remarquait que l'appareil vocal humain est l'aboutissement d'une évolution phylogénétique pour communiquer notamment des signaux socio-affectifs suggérant un continuum de ces expressions tout au long de l'interaction, dans une perspective évolutionniste. Cette idée de continuité est également abordée d'une tout autre manière dans le cadre de la synthèse/génération de parole. En effet, (Greenberg, Shibuya, Tsuzaki, Kato, et Sagisaka, 2006 ; Sagisaka et Greenberg, 2015) montrent qu'à partir de plusieurs occurrences prosodiques du monosyllabe (/n/ en japonais), il est possible d'extraire des vecteurs caractérisant des dimensions de prosodie socio-affective. Ces éléments étaient ainsi transférables sur des énoncés plus longs pour étendre cette prosodie sur un objet linguistique plus grand, qui prenait alors la même valeur le monosyllabe initial.

D'un point de vue physiologique, psychologique ou cognitif et indépendamment de la modalité concernée, les paramètres prosodiques sont des éléments modulables à travers un contrôle plus ou moins volontaire (p. ex. un petit cri produit en situation de peur peut surgir de manière involontaire s'il nous surprend. Dans cette même situation, des vocalises peuvent volontairement être émises pour signifier cet état à l'interactant pour le faire réagir et ceci peut même faire l'objet d'un message encore plus explicite). Ce contrôle peut être aussi bien inconscient (dans le cadre d'un automatisme) que conscient (la production reflète ce qui est

cognitivement préparé). Ce contrôle permettrait ainsi de moduler les paramètres prosodiques permettant de générer les formes globales véhiculant les informations socio-affectives permettant de produire les fonctions communicatives. Cette approche fonctionnelle est celle que l'on retrouve dans le modèle prosodique de (Aubergé, 2002), où la prosodie est considérée comme un agent parmi de nombreux autres dans un système global multiagent communicationnel, le C-Clone (Aubergé, Audibert, et Rilliard, 2006). Dans ce modèle (cf. Figure 7), ces fonctions prosodiques sont présentes en deux temps : un temps émotionnel et un temps linguistique, dont le passage de l'un à l'autre semble abrupt (Signorello et coll., 2010). Durant le temps émotionnel, le contrôle de nos productions communicatives serait involontaire alors que dans le second temps, il générerait des fonctions attitudinales et linguistiques par un contrôle prosodique volontaire. Des objets les plus langagièrement primitifs, nous aurions un contrôle de plus en plus important jusqu'à aboutir aux objets doublement articulés. Ce contrôle va même encore plus loin, si nous venons à choisir le contenu lexical associé à des attitudes précises augmentant les effets des mots choisis, où la manipulation de l'autre par le langage devient extrêmement puissante. Éthiquement parlant, la manipulation d'objets langagiers portant en soi une sémantique, et qui seraient en plus associés à une prosodie socio-affective très marquée, peut au mieux produire des artefacts incompréhensifs sans effets, et au pire en produire d'autres facilement interprétables. Or nous n'aurions alors aucun contrôle sur les effets de ces artefacts en termes d'actes langagiers, dont les conséquences pourraient être extrêmement néfastes à l'humain en fonction des situations, tout comme ils pourraient être bénéfiques. Dans le cadre de cette thèse, nous nous limiterons donc aux objets non lexicalisés plus certains lexicalisés, mais sur un vocabulaire extrêmement contraint afin d'appréhender progressivement cette question.



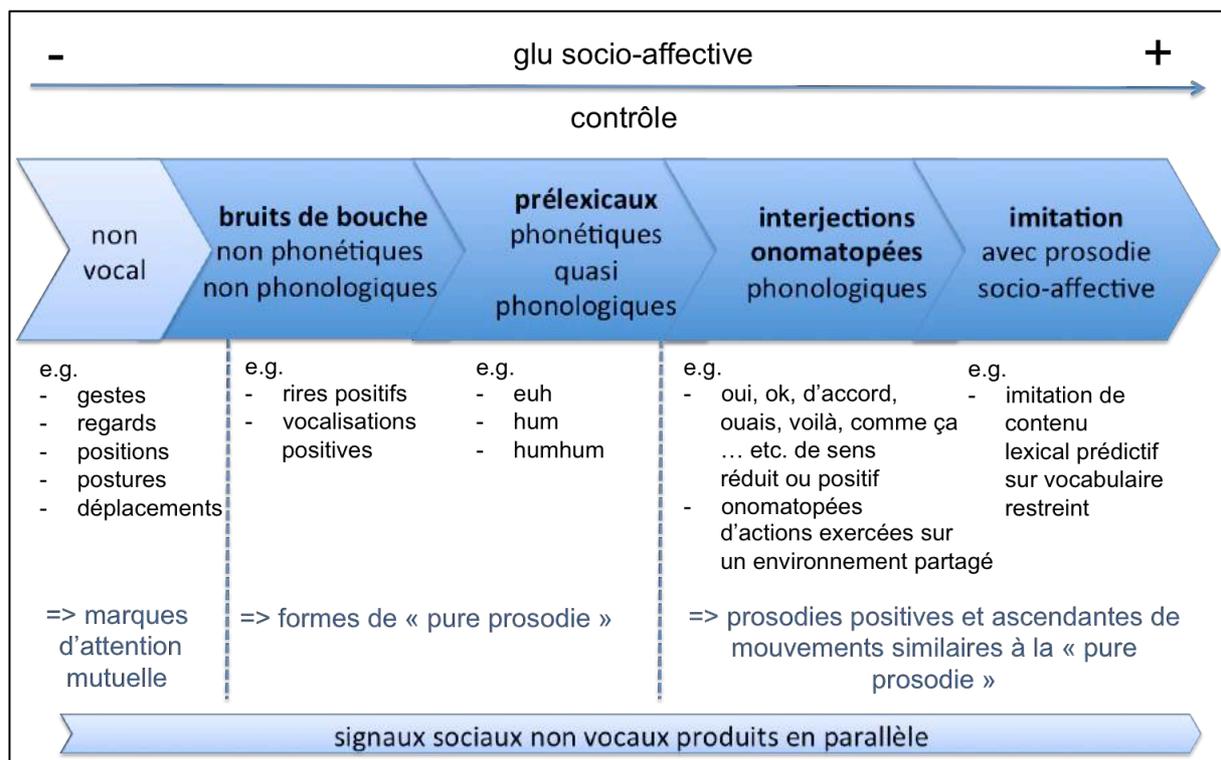
Figure 7 : Timing des fonctions de la prosodie (in Audibert 2008, d'après Aubergé, 2002)

Si nous faisons alors converger : 1) l'axe de contrôle de l'intention communicative des micro-expressions telles qu'illustrées dans le *showing-saying* ou les *display rules*, 2) le contrôle dans l'approche fonctionnelle de la prosodie d'Aubergé, ainsi que 3) la corrélation d'informations perçues de la prosodie de micro-expressions étendue sur des productions linguistiques plus complexes de (Greenberg et coll., 2006) ; nous pourrions supposer qu'il existerait une forme de « prosodie socio-affective pure » portée par les micro-expressions les plus éloignées de la structure linguistique. À partir de cette prosodie associée aux formes les plus primitives (et donc les plus permanentes) du langage, nous pourrions alors envisager que plus l'humain vient à contrôler sa production, plus cette production tendrait à devenir linguistique pour se doter au fur et à mesure d'un plus grand pouvoir manipulateur. Les éléments auraient alors un effet socio-affectif plus important au fur et à mesure, en renforçant progressivement l'état relationnel entre les interactants qui échangent par ce biais. Cette manière de rendre autonome l'agent prosodique va dans l'idée de (Demuth et Morgan, 1996), qui ont mis en évidence un effet d'amorçage (*bootstrapping*) d'une prosodie qui précède la syntaxe chez l'enfant.

Cet état général de la prosodie transparait de manière multimodale comme introduit par (Krahmer et Swerts, 2009), donc cette graduation devrait s'observer sur une prosodie multimodale qui véhiculerait des dynamiques cohérentes par rapport à la prosodie vocale. En

revanche, si nous commençons par axer nos observations sur la modalité vocale qui reste le principal véhicule de la communication face à face, nous pourrions suggérer l'hypothèse selon laquelle des primitives vocales de plus en plus contrôlées permettraient de construire et renforcer progressivement la glu socio-affective. Sachant que nous avons également supposé que cette glu pouvait se construire sur une dimension entièrement altruiste (positive ou négative), autre qu'une dimension de dominance. La prosodie socio-affective, portée par ces objets, pourrait donc être choisie parmi les plus positives et les plus universellement perçues d'un point de vue socioculturel, là où la *potency* serait nulle. La s.

La Figure 8 illustre ainsi l'hypothèse des primitives vocales de « prosodie socio-affectives pures » que nous souhaitons tester selon la gradualité : 1) de bruits de bouche (non phonétiques, non phonologiques) supposés initier le renforcement de la glu socio-affective, 2) des sons prélexicaux (phonétiques et quasi-phonologiques), 3) des interjections et onomatopées, 4) imitations (qui jouent sur le *chameleon effect* mais continuent comme tous les autres sons à porter la prosodie socio-affective censée manipuler la glu). Les prosodies en question sont tirées du corpus E-wiz, récoltées, étiquetées et perceptivement étudiées à travers des études longitudinales (Aubergé et coll., 2004b ; Audibert, 2008a ; Loyau et Aubergé, 2006 ; Vanpé, 2011 ; Signorello, Aubergé, Vanpé, Granjon, et Audibert, 2010 ; Biasi, Auberge, et Granjon, 2012 ; Sasa, 2013). Les réflexions théoriques amenant aux productions de ce corpus découlaient elles-mêmes d'acquis théoriques antérieurs.



**Figure 8 : Hypothèse de primitives vocales de *feedbacks/backchannels* à effet socio-affectif graduel dans le processus de construction de glu relationnelle en interaction face à face**

L'objectif de cette thèse est ainsi de voir si cette progression de primitives vocales socio-affectives, accompagnés parallèlement de signaux sociaux multimodaux, permettent de construire la glu socio-affective qui permettrait d'établir le premier « canal » de communication liant les deux interactants. Il faut alors que méthodologiquement leur production soit contrôlée (ce qui justifie l'utilisation d'un robot où l'état de glu est inexistant) et que ce mécanisme soit visible et utile (d'où interaction contextualisée nécessitant un rôle en lien avec l'isolement relationnel qui permet d'introduire une situation contrastive

d'observation). Les stimuli testés dans l'expérimentation sont consultables en Annexe 3). Il faut alors savoir si les micro-expressions vocales et surtout leurs informations socio-affectives peuvent être préservées lorsqu'un robot les produit et de savoir comment pouvoir le mettre en œuvre techniquement. Cela pose ainsi la question de la personnalité/personnalisation de la voix qui va faire transiter ces micro-expressions, ce que nous verrons dans la section suivante.

### 1.5.5.3. Personnalisation vocale du robot

Du point de vue physio-acoustique, il est possible de créer des sons de parole en mimant le mécanisme phonatoire humain avec des outils robotisés, ou dans le domaine de l'automatique, à partir d'un son pur modulé<sup>40</sup>, en mimant des mouvements et configurations d'articulateurs qui peuvent modifier la nature du son et certains paramètres prosodiques associés. En France, les travaux de l'IRCAM (cf. travaux de l'équipe Perception et Design Sonore) explorent de multiples façons de produire des sons notamment à partir de bruits d'objets ou de productions non lexicalisées basées par exemple sur des imitations (Susini et coll., 2006). Sachant que l'humain aurait la capacité perceptive d'associer un symbolisme sonore à une production (cf. section 1.3.5), l'esthétique vocale du robot donnerait des impressions perceptives inférant des caractéristiques de l'apparence dont la cohérence/adéquation du visuel du robot pourrait moduler la personnalité perçue de ce robot. Ainsi, d'autres perspectives de personnalisation vocale pourraient être de générer des courbes prosodiques sur une source sonore différente (c.-à-d. jouer une mélodie en changeant l'instrument qui produit le son à l'origine). L'esthétique acoustique de cette source serait alors à réfléchir en adéquation avec la taille ou la matière dans laquelle est fait le robot. En effet, par nature le robot n'a pas lieu d'être construit comme un objet ayant des cordes vocales comme l'humain, surtout s'il est non anthropomorphique. Ainsi la nature du son qu'il pourrait émettre dépend fortement de son « design corporel ». L'analogie serait de considérer un instrument de musique dont la matière de la source et les résonateurs sont différents du tissu biologique humain et produisent d'autres types d'esthétique de sons qui lui sont propres. Sur ce type de source sonore, un *morphing* prosodique pourrait être appliqué (p. ex. en apprentissage L2 comme Mareuil et coll., 2002 ; avec l'outil WinPitch LTL de Martin, 2005 ; sur le focus Aubergé et Rilliard, 2006 etc.). Cette technique permettrait de répondre partiellement au problème afin de trouver l'esthétique vocale cohérente, de la même manière qu'un marionnettiste rechercherait la « voix idéale » pour un nouveau personnage. Cette cohérence rejoint ce que (Moore, 2017) définit comme la mesure de *vocal appropriateness*. Le problème restant de savoir quel impact produirait une prosodie d'origine humaine sur une source sonore non humaine, de même qu'à vérifier si ces modifications permettent le maintien des valeurs informatives portée par la prosodie en question.

Certains de ces aspects ont notamment été abordés dans une étude exploratoire (Sasa, Aubergé, Franck, Guillaume, et Moujtahid, 2012) où un programme de traitement de signal hors ligne fut développé au sein de l'entreprise Awabot<sup>41</sup>. L'une des premières observations relevait que les micro-expressions vocales extraites de l'inventaire d'E-Wiz, ont montré que la personnalisation du robot était plus forte et affective qu'avec un robot muet. D'autre part, la

---

<sup>40</sup> Illustrations de robots synthétiseur de sons par mimétisme de production des mécanismes humains (Université Kagawa, Japon) : [https://www.youtube.com/watch?v=dD\\_NdnYrDzY&list=WL&index=4](https://www.youtube.com/watch?v=dD_NdnYrDzY&list=WL&index=4)

<sup>41</sup> Cf. travail collaboratif avec N. Moujtahid dont le principe été de modifier la fréquence fondamentale d'un enregistrement vocal sans en modifier l'enveloppe spectrale, une exploration préalable ayant montré que même si l'impression acoustique semble extrêmement proche, les outils de conversion existant ont tendance à faire des modifications de l'enveloppe spectrale pour améliorer le lissage, ce qui rend impossible les analyses des dimensions prosodiques qui seraient susceptibles de porter les fonctions modifiant l'orientation de l'interaction. Les essais consistaient donc à modifier la hauteur de voix de différente sur les mêmes enregistrements et en faire des mesures perceptives très surfaciques. Ces sons ont été testés sur le robot Emox de l'entreprise Awabot.

modification de la fréquence fondamentale, qui était d'ailleurs le seul paramètre modifié pour changer l'esthétique vocale du robot, avait montré une meilleure discrimination de la valeur socio-affective sur les hauteurs de voix que les juges disaient « mieux coller » à l'apparence du robot. Ce système de conversion vocale pouvait cependant être techniquement amélioré, car il changeait la nature du signal de départ. D'un point de vue technique, le fait que les objets acoustiques soient générés hors ligne, implique aussi l'existence d'un délai entre le traitement et la transmission du signal dans un contexte interactionnel. Cela changera grandement le rythme de l'interaction face à face. De même, les méthodes de *text-to-speech* qui peuvent s'appuyer sur des esthétiques sonores variables, nécessitent un pré-enregistrement de la voix et des énoncés à modifier. Le traitement nécessaire engendrera alors le même problème vis-à-vis du temps d'interaction. Ainsi, la génération automatisée de micro-expressions à prosodie contrôlée n'est techniquement pas encore possible de manière satisfaisante à l'heure actuelle.

Dans le cadre de cette thèse, si nous souhaitions reproduire en temps réel certaines formes spécifiques de prosodie socio-affective, la modification de la voix pourrait être une possibilité pour la personnalisation vocale du robot, à condition qu'elle puisse être correctement gérée et qu'elle conserve les informations fonctionnellement communicatives portées par le signal d'origine et notamment celles véhiculées par ses contours prosodiques. En effet, si le son à émettre existe et que nous ne souhaitons pas changer sa nature mais uniquement son esthétique vocale, l'une des solutions simple et directe serait de la rendre plus aiguë ou plus grave. L'une des méthodes de traitement de signal classique pour cela serait par exemple d'utiliser des algorithmes de type TD-PSOLA (*Time-Domain Pitch Synchronous Overlap and Add*) qui consistent à couper des périodes du signal pour les répéter à plusieurs reprises pour augmenter ou diminuer la fréquence de ces périodes, donnant ainsi un rendu de son plus grave ou plus aigu (Charpentier et Stella, 1986). Le problème est qu'une telle méthode dégrade le contour prosodique original puisque les périodes copiées sont choisies sur les portions les plus stables du signal, portant uniquement sur la nature du son (par exemple la voyelle) dans le but de discriminer le contenu du message lexical. De plus, il s'agit d'un « copier-coller » de bout de signal qui change donc la nature même du signal d'origine, notamment au niveau des transitions qui sont abruptes, avec pour conséquence des changements de pentes et l'apparition de plateaux sur un contour qui a l'origine une forme lisse qui suit la dynamique vocale de production. Par ailleurs, ils sont également très peu robustes pour reproduire et conserver la prosodie des bruits de bouches ou autres micro-expressions vocales (qui sont par exemple très peu voisés pour certains). Les outils de conversion de voix sont relativement nombreux, généralement payants, et nous en avons nous-mêmes utilisé un (cf. Voxal<sup>42</sup>) dans un cadre restreint d'interaction, uniquement pour des tâches connexes rendues nécessaires. Ces logiciels appliquent de nombreux filtres en modifiant la nature même du signal, créant ainsi des artefacts dont l'information prosodique risque d'être modifiée et donc de biaiser entièrement l'observation attendue.

L'idéal serait donc de développer une méthode qui conserve les contours prosodiques et l'information socio-affective permettant de construire la glu. La modification acoustique réfléchie s'appuie ainsi sur le même principe que celui qui permet d'augmenter ou de diminuer la vitesse du signal, tout en préservant sa nature. Une telle modification permet l'obtention d'un son plus grave ou plus aigu, par diminution ou augmentation du nombre de périodes du signal. Le signal obtenu est cependant plus long ou plus court par rapport au signal initial. Une correction temporelle doit donc être effectuée pour rapporter le signal modifié sur la même durée que le signal d'origine. Ainsi, tout un algorithme de traitement de

---

<sup>42</sup> Logiciel gratuit de conversion de voix temps réel : <http://www.nchsoftware.com/voicechanger/fr/index.html>. Le problème est la dégradation du signal acoustique qui est très peu utilisé dans le cadre de l'expérience, voir utilisé avec une exagération du mouvement prosodique testé en cas de besoin.

signal (développement G. Feng), est appliqué afin de conserver cette fréquence fondamentale nouvellement obtenue à l'appui d'une méthode appelée *time stretching*. Cet algorithme est travaillé sur des études successives de traitement de signal<sup>43</sup>. Ces approches évitent les manipulations de coupures-racolages utilisées par TD-PSOLA, en s'appuyant sur un module de double fenêtre glissante et chevauchante de traitement. Sans générer de coupure, elle transforme le signal au fur et à mesure de sa lecture, avec l'obtention d'un contour prosodique plus lisse, car elle ne dépend pas des zones stables du signal. La méthode permet donc de préserver l'information du signal d'origine avec son contour. L'avantage réside dans le fait que ce principe est applicable, quelle que soit la nature du son (c.-à-d. ici des micro-expressions vocales, qui sont courtes, pas toujours voisées, avec des caractéristiques acoustiques très instables). La modification de voix peut donc être appliquée, sur n'importe quel type de son qu'il prend en entrée contrairement à TD-PSOLA. La Figure 9 représente le principe général simplifié du *time stretching*. Un premier logiciel s'appuyant sur ce principe dans une version hors temps réel est également utilisé dans le cadre de cette thèse pour la création des stimuli du robot (cf. Annexe 1).

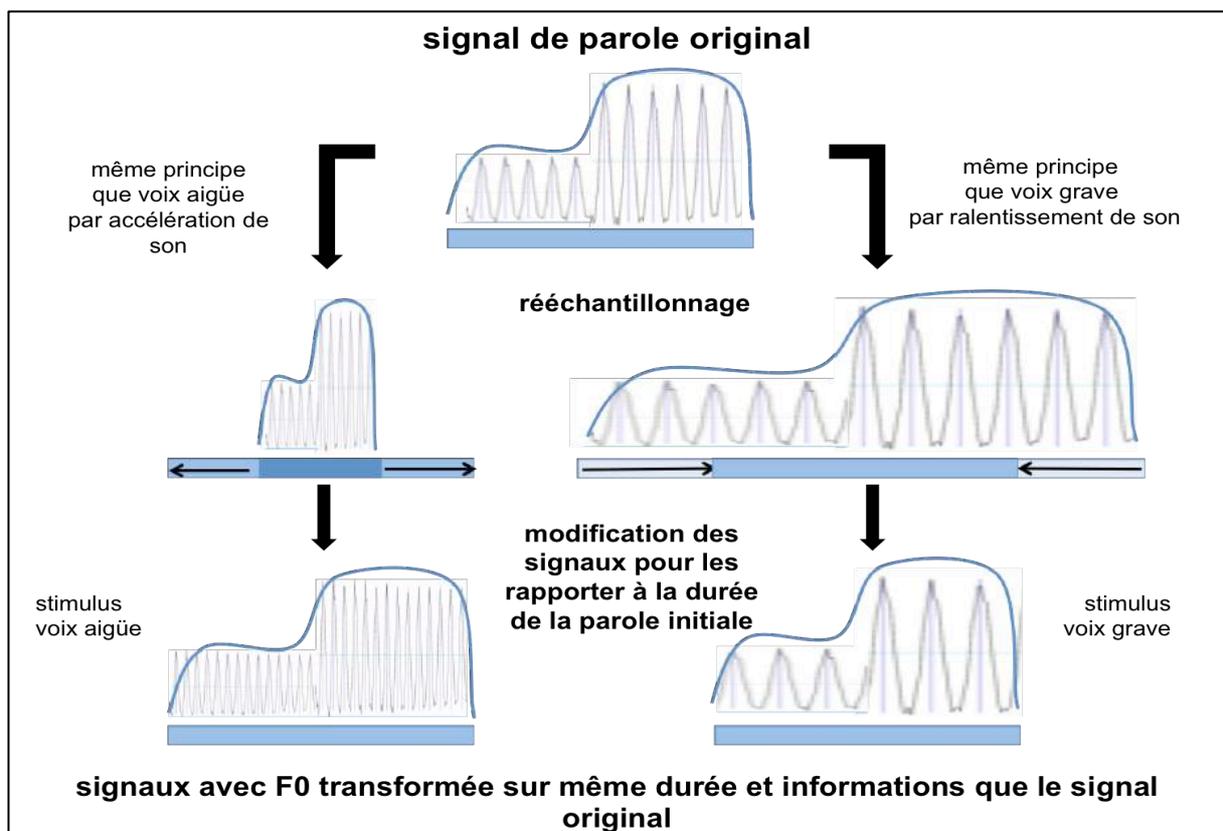


Figure 9 : Principe général simplifié du *time stretching*

La nouveauté dans le développement de ce programme de modification de voix reste avant tout la possibilité d'effectuer une transformation en temps réel. En revanche, le programme temps réel développé en parallèle de cette thèse n'a pas directement pu être implémenté sur l'interface de contrôle du robot pour les expérimentations de l'étude. Cependant, comme les sons portant les hypothèses composaient un corpus de stimuli qui n'avait pas à être transformé en temps réel, les primitives vocales testées ont été modifiées avant expérience par une modification de fréquence fondamentale conservant les contours

<sup>43</sup> Cf. travaux longitudinaux sur l'algorithme de G. Feng, adapté par (P. Prablanc, 2014) et (A. Davat 2015) sur des micro-expressions dans le cadre de leur stage de deuxième année filière SICOM au Gipsa-lab sous l'encadrement de F. Gang.

prosodiques. Le programme temps réel pourrait cependant être implémenté dans les outils expérimentaux, permettant alors d'avoir un moyen, de changer librement la hauteur de voix d'un locuteur maîtrisant sa production prosodique en direct. Il permettrait également de tester plus efficacement l'esthétique vocale d'un robot, indépendamment du contenu de sa production.

À l'heure actuelle, cette possibilité ouvre des perspectives pour l'incarnation de personnages, touchant à la notion d'*embodiment* en robotique et à la personnalisation de la voix d'un robot. Une voix dépend en effet de l'enveloppe physique qui la produit et les projections d'anthropomorphiques influencent la perception de l'objet et son animisme par la cohérence qu'il s'en dégage. Ainsi, depuis 2015, cet aspect est également exploré à travers le projet Aporia<sup>44</sup>. Ce projet qui pose des questions sur l'altérité est une convergence entre art et sciences. Il entre ainsi dans la continuité et dans les problématiques de méthodologies de co-construction collaborative semblable à celle qui a été développée en parallèle de cette thèse.

---

<sup>44</sup> Le projet Aporia dont le porteur est l'artiste plasticien Alain Quercia (cf. <https://vimeo.com/185304629>). Cet artiste se pose des questions autour de l'altérité à travers la pièce « Combat de Nègre et de Chiens » de Bernard-Marie Koltès, dont la représentation théâtrale fera intervenir ses œuvres sculptées. Elle sera augmentée par l'informatique à travers la collaboration du LIG, dont certains projets robotiques et d'interactions langagières rejoignent les mêmes problématiques que celle de l'artiste. Le programme de conversion de voix temps réel évoqué dans cette thèse est ainsi en perpétuel changement et permet également les possibilités expérimentales de la plateforme Domus-LIG notamment.

## Chapitre 2. Enjeux et contextes sociétaux

### 2.1. Isolement relationnel : un choix méthodologique d'observation

Le Chapitre 1 et notamment la section 1.5.5.1 portant sur l'hypothèse de glu socio-affective décrivaient la nécessité d'observer le processus de « manipulation » (au sens premier) relationnelle, par les primitives de langage qui sont proposées et évaluées dans ce travail, à travers une situation contrastive qui permettrait d'en comparer et comprendre les mécanismes. Concernant l'état dynamique de la relation, il s'agirait donc de pointer un lieu de l'espace socio-affectif où la glu socio-affective serait évolutive et observable.

Lorsque les systèmes sont éminemment complexes et encore peu modélisés, comme c'est le cas du système communicationnel, ce dernier doit pouvoir néanmoins être abordé de manière épistémologique. En ces termes, une méthode classique pour l'appréhender est de l'observer lorsqu'il est naturellement dégradé, puis de mesurer si et comment les hypothèses suivies permettent de le restaurer. Au minimum, c'est une manière de simuler un fonctionnement analogue (i.e. c'est le cas en biochimie avec les médicaments, en linguistique avec les pathologies du langage, etc.). En revanche, il n'est pas possible d'en déduire la réparation, ni même de savoir s'il suit ou non les lois du système naturel. Il s'agirait donc ici de trouver un cas d'observation où le mécanisme communicationnel, dont la glu est supposée être une trace observable, serait dégradé. Or la dégradation du mécanisme relationnel est en particulier visible dans les cas où des personnes déclarent, ou sont déclarées, souffrir d'« isolement relationnel ». Le chapitre 2, va donc essayer d'explorer ce problème pour en extraire un cas d'observation typique, où ce mécanisme communicationnel tendrait à être dégradé. Il s'agira également de contextualiser socialement ce problème ainsi que les solutions qui tentent d'être mises en œuvre pour le résoudre. Nous allons ainsi faire un focus sur le cas « pseudo-pathologique »<sup>45</sup> des personnes âgées en isolement relationnel qui sera le cas d'observation méthodologiquement choisi. Ce contexte est retenu car : (1) le problème de l'isolement des âgés, nous le verrons plus loin, est socialement et professionnellement identifié dans l'écosystème autour des âgés ; (2) cet écosystème est complexe, riche et « protecteur », ce qui nous apportera des garanties et des supports nécessaires dans la conduite expérimentale dont la base est avant tout de ne faire courir aucun risque expérimental à des sujets qui font dors et déjà face à une fragilisation plus ou moins prononcée. Nous explorerons alors le domaine de la robotique d'assistance et les environnements d'intelligence ambiante qui permettront de définir le second point théorique de l'hypothèse : le rôle social à travers lequel le robot devra interagir, qui permettra également de préciser les conditions du contexte expérimental d'interaction. Ainsi le robot sera ici un instrument expérimental d'observation écologique de la dynamique relationnelle. La conséquence de la modélisation de cette robotique interactionnelle socio-affective sera de fait une application mesurée et évaluée itérativement pendant l'observation/construction expérimentale de notre système.

---

<sup>45</sup> NB: ici le terme « pseudo-pathologique » n'est nullement utilisé dans une acceptation médicale ou psychologique. L'analogie serait plutôt celui d'un étranger parlant dans une langue seconde et qui serait partiellement démuné soit « pathologique » dans son utilisation de cette langue apprise car il ne la maîtrise pas entièrement, et surtout différemment d'un natif qui serait lui une situation témoin « non pathologique ».

### 2.1.1. De la question du rôle social au lien social

Parmi les phénomènes illustrant les changements sociétaux à caractères relationnels, les *hikikomoris* en sont devenus un point d'orgue, un phénomène de société à partir des années 2000. Si cette notion est relativement ancienne malgré une désignation officielle apparue en 1990 au Japon, elle est lancée à plus grande échelle avec la parution de l'ouvrage de (Saitō, 1998). Initialement, elle faisait uniquement partie de préoccupations gouvernementales. La définition de *hikikomori* telle que Saito (que la *MJSTS*<sup>46</sup> a adoptée en 2003) la proposait, désigne « un jeune qui s'est retiré chez lui et qui ne prend plus part à la société (études, travail, relations) depuis au moins six mois, sans qu'aucune pathologie mentale ne puisse être identifiée comme cause première ». Il concerne généralement des adolescents ou des jeunes adultes jusqu'à environ 35 ans voire plus, dont le retrait sociorelationnel peut s'étendre sur plusieurs mois jusqu'à plusieurs années (Kaneko, 2006), mais ce phénomène semble pouvoir extensible à tout âge. Initialement apparu au Japon, on estimait le nombre de *hikikomoris* de 260 000 à 696 000 en 2010 avec 460 000 personnes en devenir. Récemment, des cas auraient également été détectés en Oman, en Espagne, en Italie, en Corée du Sud (Kato, Shinfuku, Sartorius, et Kanba, 2011) ainsi qu'en France, qui a relevé 21 cas, mais dont 16 étaient atteints de pathologies psychiatriques concomitantes<sup>47</sup>.

Les causes suggérées à l'installation de ce phénomène seraient multiples et non consensuelles. Elles seraient : socioculturelles, économiques, cliniques (Tisseron, Chapitre 8 in Fansten, Figueiredo, et Pionnié-Dax, 2014) ou encore avec des traits tirant sur des spécificités comme à travers la notion d'*amae*<sup>48</sup> (une forme de trop grande tolérance de la part des parents, en particulier de la mère) comme le suggère (Galan, Chapitre 4 in Fansten, Figueiredo, et Pionnié-Dax, 2014) ; ou encore liées à des transformations rapides de la société japonaise alliant la pression sociale, la non-indulgence de l'échec voire à des formes de brimades d'une grande violence dans le milieu scolaire, avec des leaders qui imposent des niveaux de dominance telle que la victime ne peut pas trouver de soutien auprès de ses pairs, qui craignent d'avoir à leur tour des brimades. Ce phénomène est connu sous le nom d'*ijime*. (Suwa et Suzuki, 2001) relèvent que ce ne sont pas l'accumulation de faits culturels traditionnels existants qui a fait émerger un tel phénomène, mais bien des changements socioculturels qui en sont à l'origine.

Le retrait d'un jeune de son milieu le confronte à une perte d'une partie de son identité qu'il véhiculait à travers son rôle d'élève, mais les études actuelles montrent que les *hikikomoris* avaient en commun un fort investissement scolaire avant d'entamer un processus de désinvestissement (Guigné, Chapitre 2 in Fansten et coll., 2014).

Contrairement aux idées reçues, ce problème se différencie des phénomènes de *freeters*, *NEET* ou *otakus*<sup>49</sup> souvent associés à une surconsommation de jeux vidéos et divertissements informatiques qui les feraient converger vers des comportements asociaux (Borovoy, 2008 ; Inui, 2005). En effet, les nouvelles technologies ne seraient pas à l'origine de l'émergence des *hikikomoris*. Bien que jouissant de ces objets, la technologie est dans ce cas l'outil qui les

---

<sup>46</sup> Il s'agit du ministère japonais de la Santé.

<sup>47</sup> Article du monde traitant du sujet : [http://www.lemonde.fr/sciences/article/2012/06/07/des-cas-d-hikikomori-en-france\\_1714707\\_1650684.html](http://www.lemonde.fr/sciences/article/2012/06/07/des-cas-d-hikikomori-en-france_1714707_1650684.html) (consulté le 16/08/2013). Cf. également (Guedj-Bourdiau, 2011) pour les références de *hikikomoris* en France qui est l'une des rares à avoir publié sur le sujet.

<sup>48</sup> L'*amae* est un sentiment d'attachement volontaire, de dépendance indulgente ou encore le désir vital d'être aimé par son entourage qui selon (Doi Takeo, 1991) caractérise certaines relations sociales, notamment ?

<sup>49</sup> *NEET* : *Not in Employment, Education nor Training* vient désigner des jeunes qui se sont retirés du marché de l'emploi, terme anglo-saxon qui a été popularisé au Japon. *Freeter* vient quant à lui de la contraction de *free lance* et *arbeiter* pour pointer des personnes en emplois temporaires et précaires, arrivant dans ce statut par choix ou nécessité et vivant la plupart du temps au domicile familiale proche de l'image du « Tanguy » d'où souvent son assimilation erronée aux *hikikomoris*. Enfin l'*otaku* est un passionné addictif qui à l'heure actuelle est plutôt associé aux divertissements informatiques, *mangas*, *animes* mais qui en réalité peut concerner n'importe quelle passion, l'implication dans ce loisir pouvant parfois révéler des attitudes asociales.

maintiendrait dans le seul espace social réduit qui leur reste et qu'il cultive dans le virtuel, avec pourtant une incapacité à gérer cette même interaction en face à face (Tisseron, Chapitre 8 in Fansten et coll., 2014).

Nous allons donc nous attarder sur la notion de la dualité de la jeunesse que propose (Suzuki, Chapitre 3 in Fansten et coll., 2014) et inspiré de (Miura, 2001 in Fansten et coll., 2014) à travers la notion de jeunesse qu'il désigne par le terme *sei-shun* inspirant un mécanisme possible de socialisation, qui serait associée à trois traits : une conscience excessive de soi, un aspect « carnavalesque » qu'il relie à la capacité d'ouverture au monde extérieur, et enfin le lien avec le *Bildung*, dérivé de la culture allemande d'une culture de soi.

Les schémas ci-dessus illustrent le passage des différentes étapes par lesquelles cette jeunesse passerait pour construire son rôle social et comment elle peut être confrontée au phénomène qui le fait basculer dans une situation de *hikikomori*.

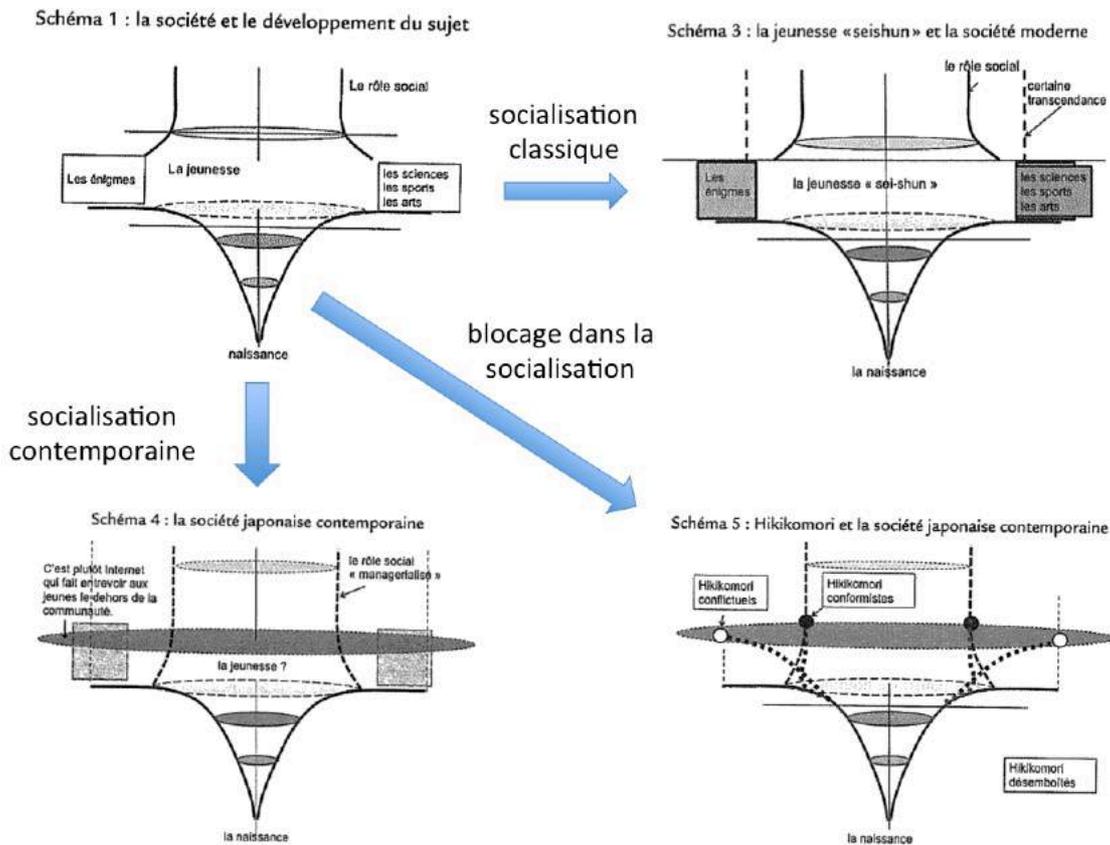


Figure 10: les changements de processus de socialisation inspirés de (Miura, 2001) (schémas tirés de Suzuki, Chapitre 3 in Fansten et coll., 2014)

Suzuki considère qu'à l'arrivée dans l'adolescence une personne fait face à une « béance de la sphère relationnelle » soit une possibilité quasi infinie de tisser son réseau de liens relationnels, le confrontant à des énigmes auxquelles il va essayer de trouver des solutions en essayant diverses activités qui vont lui permettre de construire les rôles sociaux qui vont le définir et le protéger de l'infinité de cette béance (schéma 1). Cette phase serait empreinte de deux facteurs contradictoires : 1) explorer pour voir ce qu'il existe au-delà de sa communauté, 2) un facteur qui l'intègre dans la société. Cette dichotomie est appelée la dualité de la jeunesse et elle ferait partie de l'expérience *sei-shun* de socialisation qui semble être une période relativement longue, visant des idéaux du rôle social sur une dimension fictive (schéma 3). En revanche, (Miura, 2001 in Fansten et coll., 2014) supposent que depuis les années 1970, le *sei-shun* a tendance à disparaître en remplaçant le rôle comme point de départ dans la conception du monde moderne, associés aux phénomènes de *compressed*

*modernity* (Chang, Kyun Sup, 2010) ou ce que Suzuki appelle aussi la modernité polyphonique. Les idéaux d'antan seraient désormais les points de départ dans la société néolibérale, alors que les idéaux actuels sont dits ici « managérialisés » et ne laisseraient plus le degré de liberté de l'exploration. Il tendrait plutôt à être favorisé à travers internet, pour que la jeunesse sorte de sa communauté (schéma 4). Pour Suzuki et ses collègues, le *hikikomori* serait alors quelqu'un qui reste bloqué dans le processus de socialisation, mais dans des lieux différents, suggérant alors trois types de profil *hikikomori* : ceux dont les trajectoires sont variables ou encore inconnues, ceux pris dans les mailles d'un conformisme ne leur permettant pas de connaître la dimension exploratoire, ceux qui se perdent dans cette tentative d'exploration (schéma 5).

D'autre part, le *hikikomori* ne serait pas une forme de revendication ou critique sociale, bien que parfois il semble être communautairement fédéré comme au Japon, du fait des caractéristiques socioculturelles de cette société (Teruyama et Horiguchi, Chapitre 5 in Fansten et coll., 2014). Selon (Pionné-Dax, Chapitre 11 in Fansten et coll., 2014) le phénomène ne relèverait ni d'un refus franc, ni d'un discours, mais d'une manière pour l'*hikikomori* d'esquiver les injonctions et les attentes sociales du cercle duquel il s'est retiré. Pionné-Dax décrit alors que « le retirant n'est pas seulement celui qui ne veut plus, il est celui qui ne veut plus qu'on veuille quoi que ce soit ». Ceci montre bien le problème n'est probablement pas individuel, mais relationnel avec un refus de confrontation avec les intentions communicationnelles d'un autre.

Enfin, certains psychologues, tel que (Fujita, 1971), relèvent que l'une des possibilités à l'origine du *hikikomori* est le *hitomishiri*. En japonais, *hitomishiri* désigne en premier lieu ce qui pourrait s'apparenter à une forme de « timidité », qui est notamment associée à l'enfant. Cependant, Fujita élargit sa portée à un phénomène plus global, incluant l'adulte, en le décrivant comme une spécificité relationnelle « viscérale » qui serait une réaction humaine en réponse à un changement « d'équilibre » lorsqu'un individu se trouve dans une situation où il doit exprimer ses émotions. La personne serait alors mise en danger par les conséquences du regard qu'un « autre, externe à son cercle » pourrait avoir sur elle : “*when one's personal relation with people who live in one's circle is broken by the presence of others who live outside that circle, the emotional situation becomes unstable, (...) one shows anxiety that one never expresses about those who are in the inside unity*”.

Un autre point relevé sur ce phénomène est que l'*hitomishiri* émerge très tôt, à partir du sixième mois de la vie. Des comportements sont encore plus visibles autour de 9-12 mois, et sont considérés de la même manière que les phénomènes de différenciations émotionnelles observées chez les nourrissons (cf. section 1.5). Puis s'en suit une période dite de *strange age* où jusqu'à l'âge de deux ans, ces comportements s'atténuent pour ensuite regagner de l'intensité, et en particulier lorsque les personnes sont confrontées à l'une des phases de « sociabilisation » ou de *sociality*.

Fujita rapportait à l'époque que cette caractéristique relationnelle était une spécificité de la culture japonaise. Cependant, l'émergence des *hikikomoris* dans d'autres cultures invalide cette perspective culturaliste pour laquelle (Furlong, 2008) s'oppose également. En revanche, nous pouvons peut-être considérer qu'elle manifeste une dimension du processus relationnel altruiste qui aurait induit les partisans d'une thèse proculturaliste de le *hikikomori*. En effet, la codification des normes sociales japonaises est intrinsèquement considérée comme étant altruiste, ce qui peut être associée à une manifestation fortement marquée dans cette culturelle, et cette explication n'irait pas à l'encontre des opposants à la perspective culturaliste puisque ce processus altruiste serait, dans notre perspective, une capacité humaine universelle sur laquelle une construction relationnelle peut s'appuyer pour produire de la glu socio-affective.

Avec l'apparition des *hikikomoris* certaines assimilations généralisées simplistes sont faites avec la génération *yutori* puis *satori* au Japon ou la génération Y en occident. Des

plaintes se font alors entendre vis-à-vis de cette jeune génération sous des propos tels que « les jeunes d'aujourd'hui nous manquent de respect ». Or il s'agit peut-être là de l'évolution des marqueurs illustrant le *kenson* (forme de code de respect ritualisé) qui est fortement en lien avec cette question de respect dans la culture japonaise. En effet, les jeunes générations revendiquent qu'ils préservent bien ce respect. Or le fait de ne pas restituer les marqueurs de cette communication ritualisée, tel que les anciennes générations la pratiquaient, poserait ainsi un premier problème intergénérationnel. De plus, l'incompréhension de la pertinence de tels us et l'évolution des valeurs socioculturelles différentes entre les groupes seraient également à la base de conflits d'opinions. En revanche, elle permet d'introduire la notion de *kenson* qui est une manière d'aborder le concept de « sens commun » des sociétés à travers des études portant sur les relations sociales dans une approche interculturelle, qui nous permettra de poursuivre nos réflexions autour de la glu socio-affective.

La relation sociale a donc été abordée par des études de cultures variées qui révèlent la notion de « sens commun » dont le *kenson* peut en être un dérivé au Japon (Hall et Noguchi, 1995). Le problème majeur selon (Hall, 1976) est de pouvoir déterminer les subtilités relationnelles propres à chaque communauté, ceci se reflétant dans des formes particulières de parler selon (Hymes, 1962) : “*every community has deeply patterned ways in which to relate to one another ; ways which are “only common sense”*”.

Dans la littérature anglophone, le terme japonais *kenson* est assimilé au terme *modesty* pour les Américains. Les meilleures illustrations du phénomène sont les formes de refus indirect, également désigné par le terme *enryo* qui est une attitude auto-restrictive et non contrainte permettant d'éviter les ruptures interpersonnelles. Il serait notamment à la base du refus japonais qui évite un non direct, qui pose souvent des problèmes de discrimination par les occidentaux et provoque du quiproquo interculturel. Cet *enryo* est en opposition avec la notion d'*amae* décrite comme une dépendance mutuelle également mentionnée comme possible paramètre à l'origine du *hikikomori*. Cette attitude émanerait ainsi pour une première part d'une mise en évidence de l'appartenance communautaire qui est proposée par (Hall et Noguchi, 1995) comme une déviation de la norme sociale dite de *compliance-gaining*. Cette norme, décrite en psychologie sociale, est une forme de manipulation visant à altérer intentionnellement le comportement d'un individu, là où la persuasion tente de modifier les attitudes et les croyances de la personne.

Toujours selon Hall et Noguchi, la notion de « sens commun » prendrait alors comme point de départ l'opposition entre la modestie et le *kenson*. Ainsi la modestie est considérée comme un acte individuel jouant sur la face de la personne (i.e. au sens de Goffman) et ayant des répercussions sur ses liaisons relationnelles ; alors que le *kenson* est quant à lui un acte d'abord social avec des répercussions sur l'individualité de cette personne. Il est même décrit comme une forme « d'alliance » ou de *social glue* qui permet de maintenir le *statu quo* d'une relation mais qui n'existe qu'au sein d'une production conjointe, ce qui place avant tout le processus dans un acte relationnel plutôt qu'individuel. Particulièrement visible en situation de « requête », ce processus permet de maintenir une harmonie et une équité dans la relation, même si cette dernière est initialement asymétrique en fonction du rôle qu'occupe chacun dans l'interaction. En revanche, cette relation ne s'appuierait pas sur la dominance, puisque les interactants tendraient à réduire les distances par des processus de compliments et de remerciements. Ces mécanismes sont même explicités, normalisés et ritualisés dans la culture japonaise mais sont aussi visibles à travers des pratiques comme le *salsipuede*<sup>50</sup> (Fitch, 1990),

---

<sup>50</sup> Le *salsipuede* n'est pas véritablement associé à une forme de ritualisation par les natifs, mais il semble prendre un schéma régulier de ce que l'auteur appelle une attitude de “*leave if you can*”. En effet, ce sont des échanges particuliers qui surviennent au départ d'une personne qui l'annonce à son hôte, qui réplique par une demande de rester, qui elle-même est contre-argumentée par une raison, sur laquelle l'hôte réplique en remettant en question la raison du départ. La particularité de ces échanges est qu'elle semble ne s'appuyer sur aucun rapport de force mais qu'elle relève d'une dimension d'entretien relationnel entre l'hôte et l'invité.

qui s'observe en Colombie par exemple. Le *kyoshuku* japonais en est également un exemple, qui illustre une forme de politesse extrême utilisée vis-à-vis d'un supérieur hiérarchique à qui nous voulons communiquer une contradiction tout en préservant sa face et son statut.

La dimension interculturelle de cet article suggère enfin que la relation interpersonnelle est appréhendée de manières différentes selon les cultures, dont les processus communicationnels focaliseraient différents aspects de la dynamique relationnelle. Ainsi toujours selon Hall et Noguchi:

- 1) pour les Américains, le langage passe avant tout par l'échange d'informations dont la connaissance et le partage transforment la relation entre individus, ce qui expliquerait peut-être la construction de modèles de communication de type shannonien<sup>51</sup>.
- 2) pour les Français, la conversation elle-même révèle ou exprime les liens qui unissent les interactants.
- 3) enfin pour le Japonais, l'acte communicatif est plus fortement motivé dans le sens de maintenir la relation plus que de la modifier ou de l'exprimer.

La nature de la glu socio-affective serait alors différente dans ces cultures et les enjeux de futures recherches pourraient ainsi porter sur les invariants qui caractériseraient la base universelle de la glu à travers diverses cultures (si tant est qu'elle existe), et quelles dimensions transforment socioculturellement sa nature pour l'intégrer dans les normes et les valeurs sociales du groupe d'appartenance. De la même manière, il pourrait s'agir de voir comment nous jouons de ces dimensions en tant qu'humains pour nous intégrer ou nous désolidariser d'un groupe dès lors que la culture change.

### ***2.1.2. Isolement fragilisant : le cas des personnes âgées***

Les relations sociales étudiées en interculturalité ont introduit la notion de « sens commun ». Cette notion serait rattachée à la nature des relations qui se construisent à travers des mécanismes socioculturels communicatifs pouvant donc être le produit de processus langagiers pour réguler les relations existantes. De plus, certaines de ces relations s'appuieraient sur des processus altruistes tels que le *kenson*, motivant l'hypothèse de l'existence d'une glu socio-affective. Le *hikikomori* quant à lui montre que le phénomène relationnel est mis à mal au cours d'un mouvement de construction identitaire sociale (le rôle social). Il est cependant difficile d'appréhender le rôle social dans ce cadre puisqu'il est encore en cours de construction.

Un autre cas qui permettrait d'observer ces problèmes de processus relationnel est celui de la personne âgée isolée. Au passage à la retraite, les personnes âgées sont susceptibles de rentrer dans un mouvement inverse au *hikikomori*, à savoir passer d'une situation de reconnaissance de son rôle social à perdre ce statut, ce qui attirent certaines d'entre elles dans un processus de désocialisation (Barthe, Clément, et Drulhe, 1990 ; Halimi-Falkowicz, Quaderi, et Joule, 2007 ; Pellissier, 2003). L'isolement initié par ce mouvement peut ainsi s'apparenter à différentes causes : la localisation géographique, les mauvaises expériences à l'extérieur de la maison, la réduction de contacts sociaux, les premières chutes, des maladies frappantes ou chroniques et/ou un enchevêtrement dans une routine journalière. Ces éléments objectifs caractérisant un isolement contextuel et situationnel ne posent pas systématiquement un problème. En revanche, s'ils sont associés à une perception d'isolement, ils peuvent induire un sentiment de solitude amenant à ressentir une forme de souffrance. Si cette solitude mêle ces faits concrets à des représentations personnelles, culturelles et émotionnelles, alors

---

<sup>51</sup> La théorie mathématique de la communication de (Shanon et Weaver, 1949), à l'instar de nombreux modèles de communication, part du principe qu'une interaction est un échange de message entre un émetteur et récepteur. Dans cette perspective, le message est le principal médium de l'information et chaque auteur propose des modèles fonctionnels alloués à ce support.

elle pourrait être la cause de graves fragilisations (Cornwell et Waite, 2009 ; Uchino, Cacioppo, et Kiecolt-Glaser, 1996).

En France, après l'âge de 75 ans, le risque d'impliquer une perte d'autonomie serait multiplié par un facteur sept. Le sentiment d'isolement lui aussi augmente progressivement, mais continuellement en parallèle de cette fragilisation (Pan Ké Shon, 2003). Dans la littérature gérontologique, le manque de relations sociales est dès lors décrit comme un risque majeur pour la santé, au même titre que les facteurs tels que la cigarette ou l'obésité (House, Landis, Umberson, et others, 1988). La perte d'immersion relationnelle se caractérise aussi bien quantitativement que qualitativement (Pinquart et Sorensen, 2001) et aurait alors un impact direct, aussi bien sur la santé physique que la santé mentale (Winningham et Pike, 2007 ; Tomaka, Thompson, et Palacios, 2006 ; Cornwell et Waite, 2009). Cet impact aurait à son tour des conséquences sur la qualité de vie de la personne en augmentant alors son risque de mortalité, voire de morbidité (Steptoe, Shankar, Demakakos, et Wardle, 2013), notamment si la perte d'autonomie impliquée engendre un niveau de dépendance trop élevée pour la personne âgée. Cette dernière est alors prise en charge dans des institutions, où l'isolement relationnel reste un problème global. En effet, les opportunités d'interactions sociales restent limitées par le manque de personnels et la lourdeur des tâches qui ne laissent que peu de temps à la dimension communicative (Neri et coll., 2012). Ce constat qui semble également valable au domicile de la personne. (Pinquart et Sorensen, 2001) rapportent que les institutions et les EHPAD<sup>52</sup> seraient même soumises à des risques d'isolement plus importants que les foyers-logements pour les personnes âgées.

Par ailleurs, l'ANESM<sup>53</sup> rapporte que l'amélioration de la qualité de vie des personnes dans les EHPAD nécessite un besoin de technologie permettant de faciliter la communication entre les différents personnels médicaux et médico-sociaux, mais également avec la famille et les pairs (ANESM, rapport 2011). Ces caractéristiques d'isolement ne sont pas spécifiques à la personne âgée, mais ces dernières sont plus facilement exposées à des situations fragilisantes et amenant à progressivement perdre les réflexes qui leur permettent de préserver leurs relations sociales. De fait, il existe un large spectre sociétal qui incitent l'amélioration de la communication dans le sens d'un maintien de liens sociaux (p. ex. programme CHIPS – Computers for Homebound and Isolated Persons, cf. Bradley et Poppen, 2003), mais paradoxalement (Turkle, 2012) soulève que les personnes n'ont parfois jamais ressenties autant de solitude alors qu'elles sont quantitativement très connectées à d'autres par les nouvelles formes de technologies de l'information et de la communication.

Au Japon, ce problème de vieillissement de la population est encore plus marqué. Le schéma socioculturel traditionnel de la prise en charge de la personne âgée par son cercle familial tend également à changer, induisant un effet notable d'une perception d'isolement (Kato, Shinfuku, Sartorius, et Kanba, 2017), amenant à la notion de *kodokushi*, une « mort d'isolement ». Par ailleurs, une augmentation d'une forme de petites délinquances de type vol à l'étalage s'observe de plus en plus chez la personne âgée. Lors des arrestations, ces personnes justifient leur acte par la possibilité de rentrer dans des « interactions directes avec des personnes » au poste de police. Il s'agirait d'un moyen de compenser leur solitude<sup>54</sup>. Ce type de phénomène prend alors le nom de *grey crime* et a également été remarqué aux Pays-Bas par exemple, un autre pays où l'assistance aux personnes par des moyens technologiques est très développée (van Alphen, Oei, et Herzog-Evans, 2010).

L'isolement chez les personnes âgées est donc un problème avéré et généralisé dans la société actuelle, mais il ne semble pas être un simple fait psychologique, il a des impacts physiologiques directs, modifiant les enjeux de la prise en charge de la personne.

---

<sup>52</sup> EHPAD : établissements d'hébergement pour personnes âgées dépendantes

<sup>53</sup> ANESM : l'agence nationale de l'évaluation et de la qualité des établissements et services sociaux et médico-sociaux

<sup>54</sup> Article de presse accessible sur : <http://www.courrierinternational.com/article/2009/04/23/plutot-la-prison-que-l-isolement-et-la-pauvrete>

Nous avons vu précédemment à travers les expériences interdites ou le problème de l'hospitalisme (cf. sections 1.1 et 1.3.3.1) que l'humain qui est un être intrinsèquement social. Il dépérit lorsqu'il est coupé de toute attache socio-affective dans les premiers mois de vie. Ceci semble également vrai à tout âge et contribuerait fortement à une fragilisation physique.

En effet, les relations sociales sont considérées comme une part ubiquite de la vie, véhiculant des fonctions sociales, psychologiques et comportementales (Uchino et coll., 1996). Ces auteurs montrent ainsi que la quantité, mais plus encore la qualité de ces relations ont une influence directe sur la morbidité (i.e pour un individu, il s'agit de l'état de maladie ou le déséquilibre psychique ou mental plus ou moins profond) et la mortalité voire plus exactement la létalité (i.e les conditions qui rendent nécessairement mortelles une plaie, maladie ou lésion). Les impacts les plus notables dans les études transversales sur long terme s'appuient surtout sur les paramètres physiologiques observés pour la réactivité cardiovasculaire (Allen, Blascovich, Tomaka, et Kelsey, 1991 ; Gerin, Pieper, Levy, et Pickering, 1992 ; Kamarck, Manuck, et Jennings, 1990). De plus, (Sorkin, Rook, et Lu, 2002) font également un focus sur la santé du cœur qui semble faire partie des marqueurs les plus importants. De même, l'activité immunologique semble évoluer dans le cadre de maladies à immunodéficience (House et coll., 1988). La dynamique des relations sociales aurait également un effet sur les structures hypothalamiques (Bovard, 1985 in House, 1988) et la sécrétion de cortisol (Sapolsky, Alberts, et Altmann, 1997 ; Stanton, Patterson, et Levine, 1985). La solitude serait aussi un élément altérant la régulation de la pression artérielle, ou encore le sommeil (Cacioppo et coll., 2002 ; Hawkey, Burleson, Berntson, et Cacioppo, 2003).

Jusque dans les années 1970, la question des conséquences des états de la relation sociale sur la santé physique et psychique des personnes n'avait eu que peu d'éclairage. Parmi les auteurs abordant le sujet, nous retrouvons l'étude sociologique de Durkheim sur le suicide corrélé à la qualité et la quantité des relations sociales (Durkheim, 1897/1951), mais cette dernière se limitait à ce seul risque. D'autres études dans le milieu médical auraient également tenté de mettre en évidence cette corrélation entre santé et relation sociale (House et coll., 1988). En revanche, celles-ci étaient méthodologiquement limitées du fait de leur nature rétrospective et ne pouvaient donc mesurer la portée de l'effet de l'isolement relationnelle puisque les constats ne rendaient compte que des observations post-diagnostiques des troubles de santé, sans informations sur la situation relationnelle antérieure de la personne. Cela laissait alors un flou autour de 1) l'idiosyncrasie de la relation sur l'état de santé, 2) l'effet causal entre l'état de santé et de l'état de la situation/perception d'isolement de la personne, sachant qu'il reste difficile de déterminer lequel est la cause et la conséquence de l'autre. Cette tendance a fait l'objet d'un tournant et d'un regain d'intérêt à partir du milieu des années 70, lié à trois changements (House et coll., 1988) : a) l'émergence de la construction théorique du *social support* ; b) une accumulation de prédicteurs empiriques de la mortalité qui semble caractériser quantitativement et qualitativement la relation sociale ; c) des démonstrations expérimentales de l'effet causal de la relation sociale sur des fonctions psychologiques, mais surtout physiologiques (directe et non conséquente à une dégradation psychologique préalable), aussi bien chez l'humain que chez l'animal.

Les études prospectives de suivi à long-terme, axées particulièrement sur la prédiction de la mortalité en fonction de l'état du réseau relationnel de la personne (Berkman, 1983 ; House et coll., 1988 ; Kaplan et coll., 1988 ; Orth-Gomer et Johnson, 1987 ; Schoenbach, Kaplan, Fredman, et Kleinbaum, 1986) montrent les effets bénéfiques du réseau, avec plus ou moins de disparités entre sexe et contexte urbain/rural. Les différences s'expliqueraient avant tout sur la méthodologie de récolte de données et de l'évaluation de la qualité du lien social qui reste encore un problème de nos jours.

Ces constats ont également été appliqués dans des protocoles expérimentaux impliquant des animaux (House et coll., 1988). Les effets physiologiques peuvent concerner des ulcères

(rats), hypertension (souris) et névroses (chèvre), ou encore s'observer sur une réduction de l'anxiété ou d'excitation physiologique (p. ex. sécrétion de certains acides gras libres). De plus, de nombreux exemples montrent des effets de modulation de l'activité et de la réactivité cardiovasculaire par la présence ou non de contact avec autrui. Le lien entre la relation sociale et mortalité, voire la morbidité, tend à être validé à travers les études longitudinales, mais elle doit s'affirmer davantage sur trois points : 1) les mécanismes et processus qui lient les relations sociales à la santé, 2) le degré d'exposition qualifiant la relation sociale, 3) les moyens de pallier ou au moins diminuer les effets délétères de l'isolement relationnel sur la santé.

Enfin, la théorie du *social support*, dont la notion émerge de la convergence des travaux de (Cassel, 1976 et Cobb, 1976), est apparue tout d'abord dans la littérature portant sur les maladies mentales. Elle est ensuite largement reprise dans les autres pratiques pour soulever les effets du stress et des facteurs psychosociaux en étiologie, notamment dans les pays industrialisés, où les enjeux de traitement de la pathologie simple commençaient à être reportés au profit des polyopathologies qui ne trouvaient plus suffisamment de réponses dans les seuls facteurs biologiques et génétiques. À partir de l'examen d'une trentaine d'études, Cassel et Cobb montrent que le maintien des relations peut avoir de multiples conséquences bénéfiques sur la santé, mais mettent surtout en évidence leur capacité à modérer et amortir les effets délétères des stress psychosociaux et des facteurs aggravants des troubles de santé. En effet, les mécanismes sollicitant les relations sociales amèneraient à favoriser certains comportements adaptatifs ou des réponses neuroendocriniennes palliant ces effets de stress et de dégradation physiologique.

Quelles qu'en soient les raisons matérielles ou les situations concrètes, l'isolement social et situationnel ne serait pas objectivement en soi la cause de la fragilisation physique d'une personne (Cacioppo et Cacioppo, 2014 ; Gupta et Korte, 1994 ; Shankar, McMunn, Banks, et Steptoe, 2011). En effet, il s'agirait plutôt de la perception d'isolement, d'un « sentiment d'isolement » donnant un point d'entrée au ressenti subjectif de solitude qui handicape la personne à maintenir ces relations, qui sont le propre de l'humain (Pellissier, 2003). La solitude en lien avec la santé psychologique des personnes serait associée à la description du bonheur et du bien-être subjectif (Myers et Diener, 1995).

Les précédentes études semblent converger vers l'idée que ce n'est pas le simple fait de se sentir seul, mais bien la sollicitation des processus d'interactions sociales qui auraient un impact sur les facteurs de la durée de vie des personnes, bien que les mécanismes en soient encore flous. Cela signifierait donc que ce sont ces processus qui seraient endommagés dans le cas d'un isolement relationnel. La personne âgée se trouve ainsi dans une phase de désocialisation ce qui pourrait être évité si ces mécanismes interactionnels pouvaient être entraînés et sollicités (peut-être même de manière artificielle), durant cette phase difficile. Ceci pourrait donc potentiellement permettre à la personne de conserver ses capacités relationnelles afin de préserver elle-même son réseau social et relationnel. Méthodologiquement parlant, il est également intéressant d'observer le processus de glu socio-affective auprès de ce public puisqu'il s'agit d'une situation qui est dégradée au niveau de la construction de cette dynamique. Contrairement au *hikikomori* qui présente une difficulté dans le processus de socialisation et qui donc n'a pas encore un rôle social stable, la personne âgée présente déjà quant à elle un ancrage fort dans ces rôles sociaux. Cependant, ses rôles sont fragilisés par les conditions de la situation d'isolement de la personne âgée. Cette fragilisation est néanmoins ce qui nous permettra de fixer un cadre méthodologique de l'analyse d'interactionnelles. En effet, notre hypothèse de glu socio-affective qui focalise sur le mécanisme interactionnel a ainsi un point d'appui pour observer la progression du processus relationnel qui évolue en fonction de ce rôle.

## 2.2. Contexte socio-culturel et apport de la gérontechnologie

Au regard des sections précédentes, l'étude privilégiera donc l'observation des personnes âgées en isolement relationnel. La partie suivante fera donc un rapide tour d'horizon du contexte sociétal impliquant ce public, notamment à travers la discipline de gérontechnologie qui tend à trouver des solutions technologiques pour soutenir le vieillissement de la population. En effet, notre étude devra s'intégrer dans ce contexte et montrer une cohérence vis-à-vis de cette discipline.

### 2.2.1. Problèmes de vieillissement de la population

D'après le dernier rapport de l'OMS de 2015<sup>55</sup>, la proportion des 60 ans et plus, dans la population mondiale, va presque doubler, passant de 12% à 22% entre 2015 et 2050. Elle dépasserait même le nombre d'enfants de moins de 5 ans en 2020. Par ailleurs, il est estimé que 80% des personnes âgées vivront dans des pays à revenu faible ou intermédiaire en 2050, où l'accessibilité aux soins est extrêmement réduite.

Ce vieillissement est donc très rapide, bien plus que par le passé, avec une importante mutation démographique qui s'accompagne d'un allongement de l'espérance de vie et une inversion de la pyramide des âges (cf. Figure 11 et Figure 12). Sur la base de ces changements, les projets actuels de « bien-vieillir », mettent en avant l'aide au maintien à domicile et à l'autonomie des personnes, en promouvant la dimension qu'ils appellent le « lien social »<sup>56</sup>.

Au Japon, 25% de la population est âgée de plus de 65 ans et selon son gouvernement, le pourcentage de la population de plus de 75 ans doublerait entre 2010 et 2060, en passant de 11% à 27%. D'après la base de données et les veilles de la JSTAR (*Japanese Study of Aging and Retirement*) l'inversion de la pyramide des âges, très forte et rapide, présenterait un ratio de dépendance estimé de 76% en 2050. Cette évolution motive l'augmentation accrue des projets de robotique de services et d'assistance orientées sur la réhabilitation, la prévention et la stimulation. À noter que même si le format de la prise en charge évolue, le *care* de la personne âgée au Japon est différente, beaucoup plus orientée sur une aide familiale explicitée dans (Sano in Ogawa et Retherford, 1993), car ce dernier est considéré comme un problème global et sociétal où il est de la responsabilité des proches directs et des structures communautaires de proximité ainsi que des quartiers de s'occuper de la personne âgée. De fait, l'assistance professionnelle ne rentre en jeu que bien plus tard dans la vie de la personne, une fois qu'elle montre des signes de très grandes dépendances contrairement à l'Europe ou en France où cette aide professionnelle accompagnante intervient très tôt dans le parcours de soin.

Malgré la configuration des services développés dans les pays occidentaux, la place de l'aidant familial<sup>57</sup> reste indispensable et dans son rôle unique (Ennuyer, 2007). Il existerait même une réticence des familles aidantes à solliciter le professionnel surtout dans le cadre des maladies neurodégénératives ou induisant des formes de démence (Coudin, 2004). Le fardeau ou *burden* (Bocquet et Andrieu, 1999) est estimé entre 7 à 17 milliards d'euros, si les actions

---

<sup>55</sup> Consultable en ligne : [http://apps.who.int/iris/bitstream/10665/206556/1/9789240694842\\_fre.pdf](http://apps.who.int/iris/bitstream/10665/206556/1/9789240694842_fre.pdf)

<sup>56</sup> De nombreux rapports existent en France sur ce sujet : Franco A. Vivre chez soi: autonomie, inclusion et projet de vie. Rapport présenté à Mme N. Berra, Secrétaire d'Etat en charge des Aînés, le 15 juin 2010 : [www.travail-solidarite.gouv.fr/IMG/pdf/Rapport\\_Vivre\\_chez\\_soi.pdf](http://www.travail-solidarite.gouv.fr/IMG/pdf/Rapport_Vivre_chez_soi.pdf) ; ; Novembre : 1er rapport Picard, Conseil Général des Technologies de l'Information, puis en 2008 : 2e rapport Picard CGTI : « Usage des TIC par les patients et les citoyens en situation de fragilité dans leurs lieux de vie » ; Rapport Rialle, 2007 - Technologies nouvelles susceptibles d'améliorer les pratiques gérontologiques et la vie quotidienne des malades âgés et de leur famille : <http://www.ladocumentationfrancaise.fr/var/storage/rapports-publics/074000390.pdf>

<sup>57</sup> Il est actif en France avec notamment l'association française des aidants : <http://www.aidants.fr> et de nombreuses structures locales propre à chaque région.

## Enjeux et contextes sociétaux

de ces aidants étaient rémunérées<sup>58</sup>. En effet, on compte officiellement en France 8,5 millions d'aidants officiels (mais qui sont certainement bien plus nombreux), dont un tiers meurent avant les aidés, ce qui pousse les gérontechnologues travaillant autour de la maladie d'Alzheimer par exemple, à trouver des moyens pour soulager le fardeau des aidants en essayant de dégager du temps des professionnels, mais aussi en développant des technologies pouvant apporter une aide directe aux aidants (cf. rapport Rialle, 2007 ; Rigaud et coll., 2011).

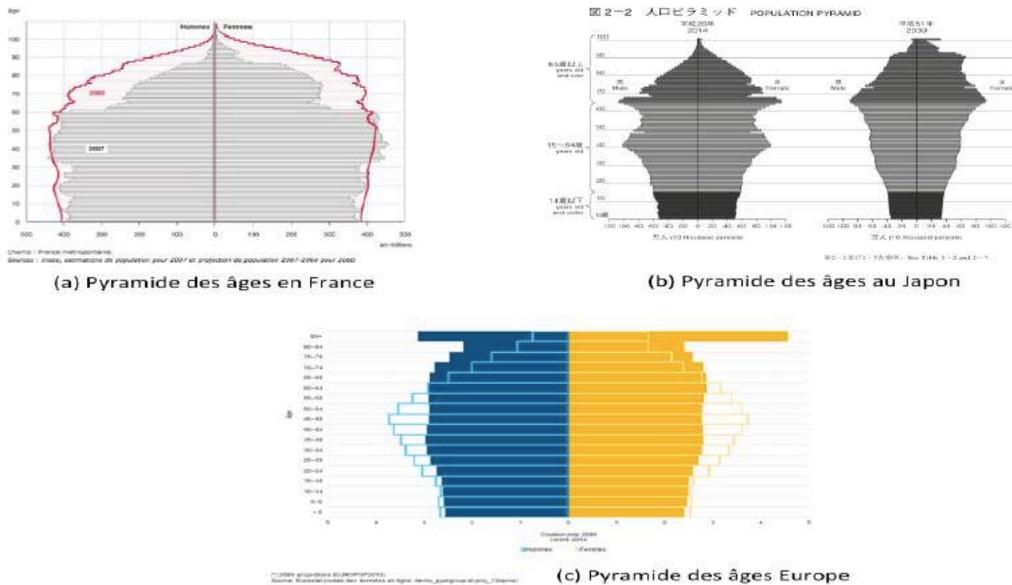


Figure 11: Pyramide des âges en France (a), au Japon (b) et en Europe (c) avec projections décennales à venir<sup>59</sup>

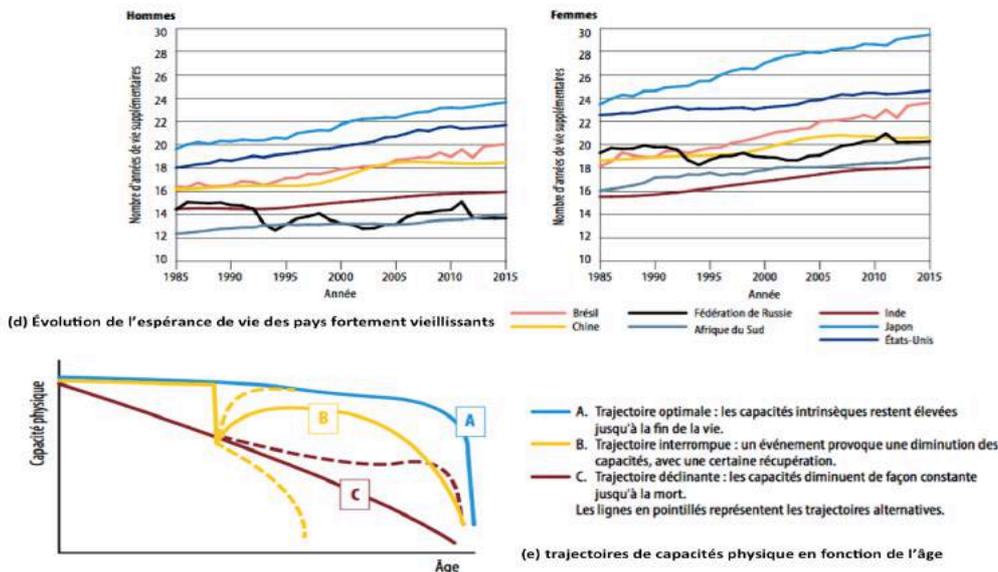


Figure 12: évolution de l'espérance de vie et fragilisation physique des personnes (source : Rapport OMS de 2015)

<sup>58</sup>Exemple parmi bien d'autres, voici un article de presse illustrant typiquement le problème évoqué : [http://www.lemonde.fr/economie/article/2013/10/27/les-aidants-invisibles-le-veritable-defi-de-l-independance\\_3503616\\_3234.html](http://www.lemonde.fr/economie/article/2013/10/27/les-aidants-invisibles-le-veritable-defi-de-l-independance_3503616_3234.html)

<sup>59</sup> Les sources consultées le 2 juin 2016 sont issues respectivement de : (a) : INSEE (<http://www.insee.fr/fr/ppp/bases-de-donnees/irweb/projpop0760/dd/pyramide/pyramide.htm>) ; ; (b) : Japan statistics (<http://www.stat.go.jp/english/data/nenkan/1431-02.htm>) ; ; (c) Eurostat ([http://ec.europa.eu/eurostat/statistics-explained/images/b/b0/Population\\_pyramids%2C\\_EU-28%2C\\_2014\\_and\\_2080\\_%28%29\\_%28%25\\_of\\_the\\_total\\_population%29\\_YB15-fr.png](http://ec.europa.eu/eurostat/statistics-explained/images/b/b0/Population_pyramids%2C_EU-28%2C_2014_and_2080_%28%29_%28%25_of_the_total_population%29_YB15-fr.png))

Dans le cadre des projets de programmes l'Horizon 2020 de la Communauté Européenne, qui présentent les perspectives et les défis de l'Europe, une veille globale est réalisée pour les pratiques d'assistance à la personne âgée<sup>60</sup>. Les recommandations issues de cette veille recommandent fortement une implication des aidants dans le cadre d'un vieillissement actif et préventif. La prévention doit alors porter sur le diagnostic de fragilisation et le déclin des fonctions aussi bien physiques que cognitives des personnes âgées. D'après ce rapport, la force de travail de la population actuelle diminuerait d'environ 42 millions dans les 50 prochaines années. Jusqu'à 2030 ce changement serait accompagné d'une intense vague de vieillissement de la population issu du *baby-boom* et composant la nouvelle vague appelée *le papy-boom* en France. Sur les quinze années à venir, l'Europe présenterait ainsi une augmentation du nombre de personnes âgées de plus de 65 ans, passant de 17,4% à 25, 6%, et cette proportion atteindrait même les 29,5% en 2060. Ainsi la population européenne âgée va également doubler, passant de 87,5 millions en 2010 à 152,6 millions en 2060.

Dans le cadre d'une évaluation normée, la plupart des pays ont essayé de développer leur propre échelle spécifique de dépendance dont par exemple le GIR en France (Belmin et coll., 2007). Il existe des mesures de degré de dépendance et de fragilisation plus générique également telles que le LEADS (*Leeds Elderly Assessment Dependency Screening de Slade, Fear, et Tennant, 2006*) ou encore le CDS (*Care Dependency Scale – Dijkstra, Buist, Dassen, van den Heuvel, et voor Gezondheidsvraagstukken, s. d., 2006*) proposé notamment par le groupe de travail européen sur la santé (EURECARE). Ces mesures s'appuient sur les actes de la vie quotidienne (AVQ) ou *Activity of Daily Life* (ADL) dont la réalisation ou la qualité d'exécution permet d'avoir une idée de la fragilité physique de la personne. En revanche, la dimension relationnelle et socio-affective est encore très peu accessible en dehors de questionnaires ponctuels ou d'une évaluation quantitative déduite de la fréquence de sorties, de visites ou de la pratique de moyens de télécommunication. Ainsi l'implication des aidants devient d'autant plus importante du fait de leurs connaissances empiriques du terrain. De fait, nous essaierons de développer des protocoles permettant d'intégrer au mieux ces acteurs en leur attribuant un véritable rôle dans la construction des méthodes ainsi que dans les expérimentations.

### ***2.2.2. Technologies d'intelligence ambiante pour le maintien à domicile***

Depuis le début des années 90, la gérontechnologie commence à gagner du terrain dans les pays occidentaux comme discipline à part entière (Bouma, Fozard, Bouwhuis, et Taipale, 2007). L'inversion de la pyramide des âges, aboutissant à de nouveaux patterns démographiques socioculturellement déterminés (Chesnais, 1990), amène à un déficit des solutions de prise en charge des personnes âgées dont la volonté majoritaire est de vouloir rester le plus longtemps possible à leur propre domicile<sup>61</sup>. Les « maisons expérimentales » appelées également *Smart Homes* ou « habitats intelligents » (Noury, Virone, Barralon, Rialle, et Demongeot, 2004) sont une des technologies proposées comme solution pour le maintien à domicile. Parmi les nombreux exemples sur lesquels nous ne pourrions être exhaustifs, nous avons par exemple le Georgia Tech's Aware Home (Kientz et coll., 2008) qui n'est pas seulement dédié aux personnes âgées, mais est pensé dans l'optique d'améliorer la qualité de vie des personnes et l'indépendance des personnes, dans des environnements réalistes. Ces technologies peuvent également être dédiées à apporter un complément de support physique et technique tel que dans le projet SmartBo (Elger et Furugren, 1998) ou encore de détecter des ADL (activités de la vie courante – *DLA* en anglais) comme proposé

---

<sup>60</sup> [https://ec.europa.eu/research/innovation-union/pdf/active-healthy-ageing/gp\\_a3.pdf](https://ec.europa.eu/research/innovation-union/pdf/active-healthy-ageing/gp_a3.pdf)

<sup>61</sup> Cf. rapports cité en note 56

dans le projet PROSAFE (Marie Chan, 2005). Ces systèmes informatiques embarqués, à cheval entre ubiquité et pervasivité, donnent lieu à un nouvel horizon de réflexions et de possibilités techniques, aboutissant à la notion « d'intelligence ambiante ». La Commission européenne en consultation avec le groupe de Conseil ISTAG<sup>62</sup>, soutient d'ailleurs cette problématique émergente en suggérant des scénarios et des méthodologies facilitant le développement de tels environnements de recherche. Des études en habitats intelligents sont également menées pour afin d'améliorer les systèmes de reconnaissance vocale qui leur sont associés, dont la commande vocale est un enjeu majeur dans le cadre du handicap (Vacher, Fleury, Portet, Serignat, et Noury, 2009 ; Portet, Vacher, et Rossato, 2012), de même que dans la détection de situation de détresse (Istrate, Castelli, Vacher, Besacier, et Serignat, 2006).

Néanmoins, l'acceptation et la pertinence de telles technologies restent un point sensible, notamment concernant les réflexions éthiques liées à la collecte de données et à la protection de la vie privée de ces utilisateurs qui fait tout l'objet de la cybersécurité. De plus, dans le processus de conception de ces environnements, des divergences importantes sont notables sur la perception des besoins pensés par les designers, les acteurs sociaux aidants et les personnes âgées elles-mêmes (Sponselee, Schouten, Bouwhuis, et Willems, 2007). L'acceptabilité et l'utilisabilité de ces technologies seraient d'autre part, non seulement liées à la motivation des personnes à rester à leur domicile, mais également à la qualité de leurs « liens sociaux ». Cette fragilisation, nous le rappelons ici, a des conséquences directes sur leur récupération physique, augmentant le risque d'exposition à une plus forte mortalité (Hawkley et coll., 2003 ; Luo, Hawkley, Waite, et Cacioppo, 2012 ; Uchino et coll., 1996).

L'enjeu est donc ici de s'appuyer sur ces « environnements intelligents » dont la motivation sociale s'accroît dans le domaine gérontologique, tout en essayant de trouver une cohérence méthodologique pour l'intégrer dans un scénario expérimental réaliste. Cet environnement intelligent sera d'ailleurs à associer à une autre perspective gérontechnologique qui est la robotique d'assistance.

---

<sup>62</sup> IST Advisory Group (ISTAG), 2010 European Commission report on Ambient Intelligence: <ftp://ftp.cordis.lu/pub/ist/docs/istagscenarios2010.pdf>

## 2.3. Robotique de services pour les personnes âgées

### 2.3.1. Robotique d'assistance à la personne

La robotique s'inscrit dans le domaine de l'intelligence artificielle dont les approches ont été explicitées sections 1.2 et 1.4 du Chapitre 1. Nous voyons notamment des approches qui réfèrent à la dimension du traitement de la « raison » humaine puis une évolution vers l'intégration d'une dimension « émotionnelle » qui en réalité est construite sur une position épistémologique bien en amont de celle dans laquelle nous nous inscrivons. Ce positionnement scientifique s'oppose pour rappel au dualisme corps/esprit qui reste valide dans la conception de nombreuses études actuelles, malgré la dimension intégrative des émotions par ces dernières, qui considèrent l'aspect émotionnel comme un complément d'aide au réalisme et à la modélisation de l'intelligence et de la raison. De notre point de vue, elle sous-tend et oriente le fonctionnement de la raison, ce qui va changer les perspectives d'approches et de méthodologie suivie. L'un des points importants pour appréhender l'interaction homme-robot tout en considérant notre hypothèse théorique de primitives relationnelles est donc l'association d'un rôle social au robot.

La robotique d'assistance est un vaste domaine émergent dont le public est bien plus large que celui des personnes âgées. En revanche, ici nous ferons un focus sur ce domaine, du fait de l'intérêt scientifique et la particularité de ce public, qui aura un fort impact sur la dimension communicative et relationnelle avec le robot que nous allons explorer.

La robotique d'assistance pour les personnes âgées est généralement catégorisée en deux domaines : les robots de réhabilitation (*rehabilitation robots*) ou les robots sociaux d'assistance (*assistive social robots*), cette deuxième catégorie étant elle-même subdivisée en « robots de services » (les « *nursebots* » comme Pearl, Care-o-Bot, iCat etc. ) et ceux dits « compagnons » qui ressemblent à des pets (Aibo, Paro, Karotz, Genibo, Zoomer etc.) et dont l'enjeu est l'interaction sociale (Bemelmans, Gelderblom, Jonker, et de Witte, 2012 ; Broekens, Heerink, et Rosendal, 2009). En gérontologie, la seconde catégorie des compagnons tend vers des robots *huggable*, que l'on peut prendre dans les bras, une action qui aurait des bénéfices physiologiques (p. ex. améliorations de pronostics vitaux, des test MMSE<sup>63</sup> de mémoire) et psychologiques (p. ex. relaxation, motivation), notamment sur des études à long terme d'environ 2 mois (Wada et Shibata, 2007). Dans les perspectives occidentales, une attention particulière est portée aux robots sociaux d'assistance autour des soins et du « *care* » (qui prend une dimension particulière dans le cadre de la sociologie du vieillissement). Ce *care* est encore majoritairement testé dans le cadre institutionnel (Shibata, Wada, Saito, et Tanie, 2005) à travers des activités de soins robotiquement assistées, avec également un effet social (p. ex. stimulation et une redynamisation de la communication entre le personnel et les patients). Dans ce cadre, la dimension interactionnelle est également explorée sur des Aibos, pour mesurer les niveaux hormonaux de la salive des personnes (Kanamori, Suzuki, et Tanaka, 2002) ou dans la comparaison des effets de ces robots avec une peluche en forme de chien chez des patients atteints de démence (Tamura et coll., 2004), ceci sur des études à court terme. Les résultats montrent des influences plus ou moins marquées de l'effet du robot, qui reste en revanche perçu positivement. Sur une étude avec le Paro, (Taggart, Turkle, et Kidd, 2005) observent des effets classiques d'*engagement*, de partage, d'affection ou de découverte, des points généralement soulevés dans les observations des études en robotique sociale, considérant ainsi la technologie comme une chose évocatrice, un artefact relationnel et non un simple objet. Entre la mise en route et le mode d'arrêt du robot, les dynamiques entre résidents semblent changer, l'interaction tendant vers un regain

---

<sup>63</sup> Le MMSE ou *mini mental state examination* est une mesure classiquement utilisée dans le cadre de la gérontologie pour mesurer les capacités de mémorisation de la personne

de dynamisme lorsque le robot est allumé. La présence de ce dernier semble aussi changer la nature des conversations, mettant en évidence un « effet observateur ». Par ailleurs, la question de l'acceptabilité de cette technologie n'est pas souvent exclusivement focalisée dans les études, mais elle ne semble pas non plus être un frein, soulignant en revanche des comportements variables en fonction de la génération et du genre notamment (Heerink, Kroese, Evers, et Wielinga, 2006). Cette technologie est d'ailleurs plus souvent difficilement appréhendée par les soignants que par les personnes âgées elles-mêmes (Weber, 2005). Une exploration méthodologique pour la mesure de cette dimension d'acceptabilité pourrait d'ailleurs être intéressante. Dans ce cadre, l'UTAUT – *unified theory of acceptance and use of technology* (Venkatesh, Morris, Davis, et Davis, 2003) est le modèle le plus communément utilisé par la communauté. Notons que malgré cette prédominance, il existe de rares exemples qui focalisent leur observation sur cette question comme l'étude de (De Ruyter, Saini, Markopoulos, et Van Breemen, 2005).

En Europe, les études rentrant dans le cadre du maintien à domicile augmentent de plus en plus. Cela s'illustre par des projets privilégiant l'aide des acteurs de services à la personne, aussi bien dans les institutions que dans les domiciles, et dont l'optique est de : 1) préserver la santé des aidants qui sont victimes de *burn-outs* fréquents, (avec pour rappel un plus fort taux de mortalité que les aidés eux-mêmes), 2) fournir du temps et des outils qui favorisent la concentration de leur attention sur le travail de la relation avec les personnes âgées. En effet, il s'agit d'une dimension qui vient à être négligée par état de fait, à cause des conditions de travail contraignantes où les activités d'aide portent préférentiellement autour des aspects pratiques, plutôt que relationnels. L'un des champs ainsi développés dans le cadre de cette assistance est notamment celui orienté autour du maintien à domicile afin de casser le cercle vicieux de leur fragilisation. Elle est alors souvent associée à des environnements intelligents ou dans un contexte dit d'intelligence ambiante (Baltus et coll., 2000 ; Beck, Edwards, Friedman, et Khan, 2003) dédiée à un soutien physique, cognitif ou social comme présenté dans le projet ACCOMPAGNY<sup>64</sup> qui s'appuie sur la plateforme Care-O-bot<sup>65</sup>. Le robot SRS<sup>66</sup> quant à lui est une application de contrôle téléopérée permettant d'agir sur l'environnement de vie de la personne âgée, le contrôle se faisant soit par la personne âgée elle-même ou soit par sa famille. L'interaction robotique peut également être orientée sur la perception de soi et les activités de stimulation cognitives comme dans le cadre du projet Mario<sup>67</sup> où le robot Kompai délivre une importante assistance médicale afin de fournir un soutien aux personnes affectées par des démences. L'apport est alors aussi bien pour l'aide aux activités de la vie quotidienne que pour assurer des interventions médicales ou aider à assurer l'activité médico-sociale des soignants. En revanche, cette situation a des tendances d'exclusion en isolant les personnes et les aides à domicile dans cet environnement, le robot d'assistance visant avant tout à apporter une aide technique pour faciliter l'activité de l'aidant comme dans le cadre du projet CompanionAble<sup>68</sup>. Enfin, notons également le projet GiraffPlus<sup>69</sup> qui travaille sur l'amélioration du réseau de soutien de la personne âgée, aussi bien composée des aidants familiaux que des professionnels dans le cadre d'une étude interculturelle avec un déploiement du robot au sein d'au moins 15 domiciles en Suède, en Italie et en Espagne. L'approche est associée à une activité de monitoring avec utilisation de capteurs biophysiques dont les mesures sont envoyées au corps médical avec alertes spécifiques en cas de détection d'anomalies, un suivi pour les détections de chute et une activité de téléprésence familiale et professionnelle pour la dimension sociale.

---

<sup>64</sup> <http://accompanyproject.eu>

<sup>65</sup> <http://www.care-o-bot.de>

<sup>66</sup> <http://srs-project.eu>

<sup>67</sup> <http://www.mario-project.eu/portal/>

<sup>68</sup> <http://www.companionable.net>

<sup>69</sup> <http://www.giraffplus.eu>



Figure 13: robots d'assistance développés dans le cadre de projets gérontechnologiques

Au Japon, où la robotique d'assistance a été abordée bien plus tôt du fait de l'évolution démographique, l'aspect de la réhabilitation est un domaine bien plus répandu. Ainsi des entreprises telles que Yaskawa Electric, Toyota ou Honda développent des prothèses ainsi que des outils d'assistance physique.

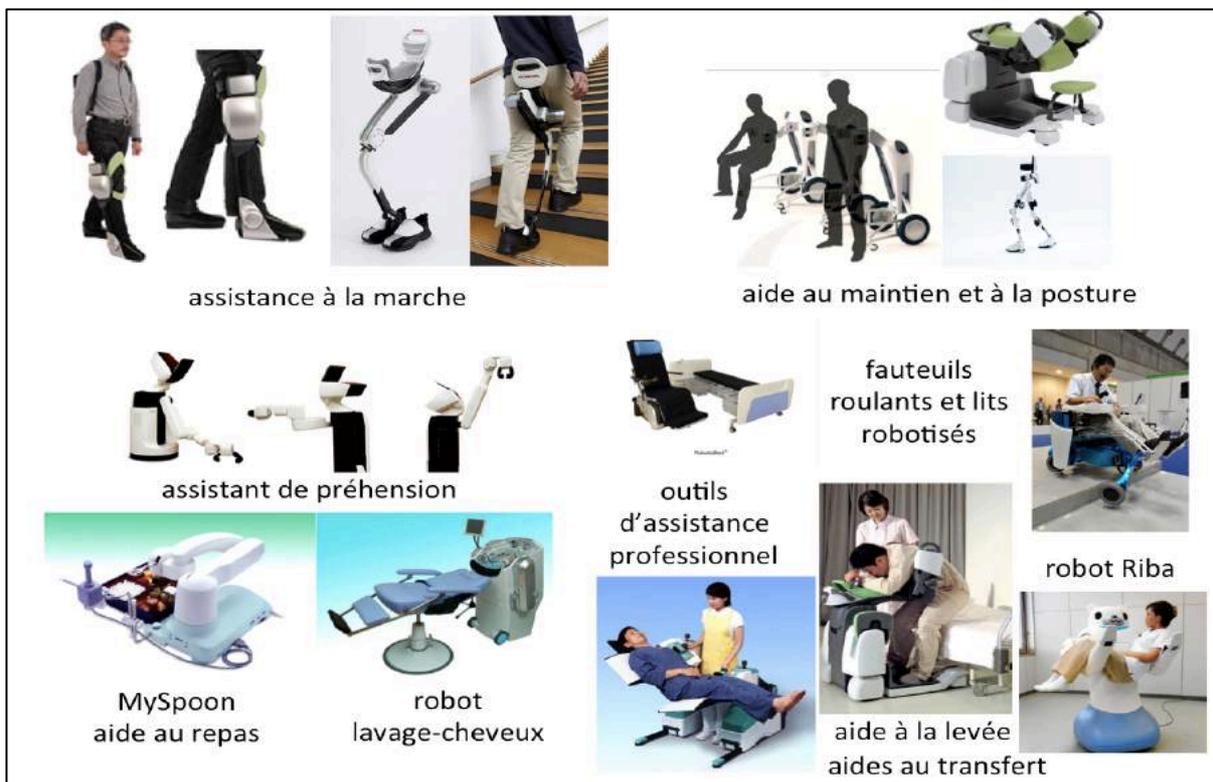


Figure 14: robotique de réhabilitation et d'assistance aux aidants et aux personnes âgées

Parmi ces exemples nous trouvons :

- Les assistances à la marche (p. ex. Independent walking assistance, 2011),
- Les assistances pour l'aide à la levée (p. ex. Care Assist Robot, 2011),
- Les supports du maintien et de la posture ainsi que les fauteuils roulants robotisés (p. ex. Wiglet or Mobiro),
- Les exosquelettes (p. ex. Cyberdyne),

- Les assistances au robot (p. ex. MySpoon, 2006 qui est un module complémentaire aux robots d'assistance),
- Les aides au soutien de la pratique professionnelle des aidants par exemple l'aide au transfert (p. ex. lifting RIBA de Riken, 2010),
- Les robots laveurs de cheveux,
- Les lits robotisés (p. ex. Panasonic) etc.

La dimension sociale est directement associée à l'acte de réhabilitation doublée d'une dimension préventive motivée par l'utilisation d'outils de stimulation cognitive avec un aspect d'entraînement (Tamura et coll., 2004 ; Kaneshige, Nihei, et Fujie, 2006 ; Inoue, Wada, et Uehara, 2011). Ces activités ont initialement commencé au sein de maisons de retraite avec le Paro (Wada et Shibata, 2007 ; Inoue et coll., 2011) dans l'optique d'avoir un effet bénéfique dans le cadre de l'accompagnement des démences. Plus récemment, nous voyons que la stimulation montre un intérêt pour l'exercice physique et le dialogue avec sa dimension communicative à travers des *coachs* robots ou des « compagnons » domestiques considérés comme des partenaires de vie (Iwamura, Shiomi, Kanda, Ishiguro, et Hagita, 2011). Les formes et design de ces robots sont extrêmement variables, allant du type « character d'anime » tel que Kabochan<sup>70</sup>, PaPeRo<sup>71</sup> ou Palro<sup>72</sup>, jusqu'aux formes humanoïdes, dont Pepper<sup>73</sup>, Wakamaru<sup>74</sup> ou Sota<sup>75</sup>.

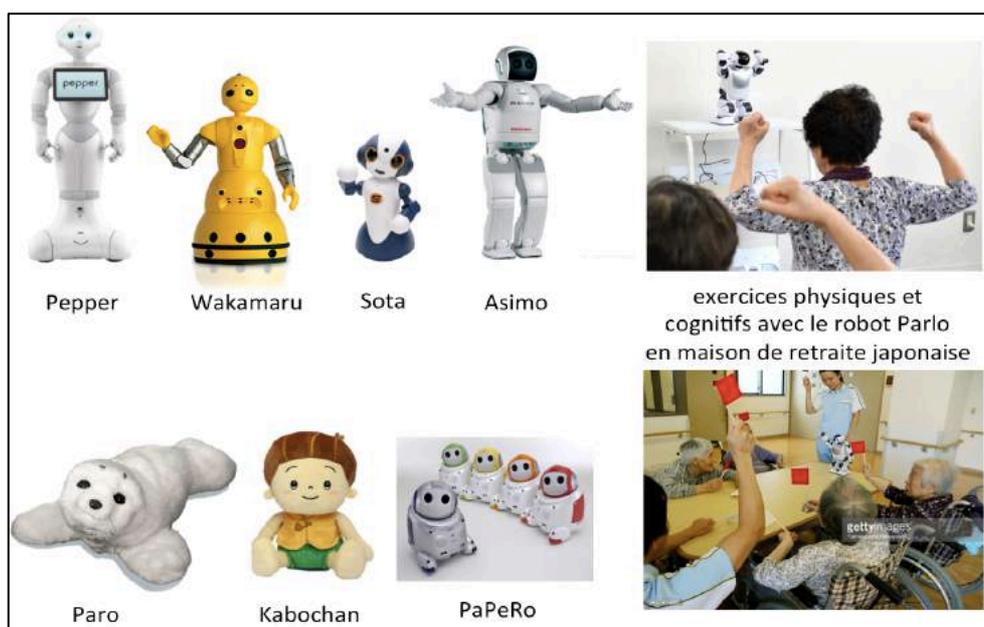


Figure 15: robots domestiques et professionnels gérontechnologiques d'assistance sociale (Japon)

De manière générale, la robotique sociale de réhabilitation et d'entraînement (Kiratli, 2006 ; Tapus, Maja, et Scassellatti, 2007), même si elle ne semble pas être actuellement la forme la plus répandue, montrerait des effets positifs à travers sa dimension de monitoring. Elle permettrait également de préserver la motivation des personnes, en les encourageant (Eriksson, Mataric, et Winstein, 2005 ; Kang, Freedman, Mataric, Cunningham, et Lopez, 2005 ; Kidd, Taggart, et Turkle, 2006) ou en faisant du coaching. Les études les plus

<sup>70</sup> [www.pipjapan.co.jp](http://www.pipjapan.co.jp)

<sup>71</sup> <http://jpn.nec.com/robot/en/functions/index.html>

<sup>72</sup> <https://palro.jp>

<sup>73</sup> <http://www.softbank.jp/en/robot/>

<sup>74</sup> <https://www.mitsubishi.com/mpac/e/monitor/back/0602/story.html>

<sup>75</sup> [http://www.roboticstrends.com/article/sota\\_home\\_robot\\_to\\_care\\_for\\_japans\\_elderly](http://www.roboticstrends.com/article/sota_home_robot_to_care_for_japans_elderly)

encourantes sont associées à la réhabilitation suite à un AVC – accident vasculaire cérébral – (Burgar, Lum, Shor, et Van der Loos, 2000 ; Krebs et coll., 2002). Ces interactions, impliquant l’usage de robots sociaux d’assistance, ont montré l’émergence de processus qui induisent des mécanismes d’association de personnalités au robot. Dans ces conditions, l’écologie d’entraînement semblait gagner en efficacité pour la réalisation des exercices proposés (Tapus et Mataric, 2006). Ainsi (Tapus et coll., 2007) rapportent que les principaux points d’intérêt ayant des effets sur la personne âgée, mais aussi sur la pratique clinique (et donc par extension gérontologique) sont :

- l’**embodiment** : qui semble avoir un fort impact dans ce domaine où cette dimension augmente par les situations où les personnes sont amenées à toucher le robot dans le cadre de leur interaction (cf. section 1.4.1)
- la **personnalité** : une dimension qui est encore très peu explorée, mais intéresse le domaine du fait de son fort lien avec les comportements observés. Elle motive notamment l’observation des traits indiquant les aspects illustratifs du style de vie de la personne, en essayant de déterminer celles qui persistent dans le temps et qui dénotent une caractéristique unique à cette personne.
- l’**empathie** : la notion renvoie à ce qui a été explicité en section 1.5.1 et section 1.5.2.
- l’**engagement** : la notion renvoie à ce qui a été explicité en section 1.4.3.
- l’**adaptation** : elle concerne aussi bien celle de la personne que du contexte, à court et long terme, mais prend une dimension très importante et surtout éthique dans sa méthodologie, car elle implique de faire face aux contraintes d’un public fragilisé et donc avec des effets amplifiés selon les cas, mais nécessitant une dimension d’apprentissage associé.
- le **transfert** : c’est la notion qui est rattachée à l’effet d’un comportement durable et à long terme avec une portée qui change le comportement dans une situation donnée. Les capacités et les comportements qui sont impliqués ne doivent pas uniquement être visibles dans la situation d’interaction testée avec le robot. Elles devraient avoir un impact sur la manière que les humains ont de se comporter entre eux (p. ex. changer la dynamique d’interactions entre patient et soignant/aidant, voir l’application des capacités sociales acquises par l’enfant autiste à travers ses futures interactions avec ses parents et des pairs etc.).

Par ailleurs, les nouvelles perspectives d’utilisation de la robotique intéressent beaucoup l’usage thérapeutique des personnes atteintes de démence (Inoue et coll., 2011 ; Kidd et coll., 2006 ; Marti, Bacigalupo, Giusti, Mennecozzi, et Shibata, 2006 ; Mitchell et O’Donnell, 2013), mais le cadrage médical nécessaire étant trop important et la maladie trop influente sur la dimension interactionnelle, cette thèse ne s’est pas placée dans ce paradigme.

En revanche, un apport bénéfique de la dimension sociale de la robotique d’assistance semble influencer la qualité des activités de soins (qui jusqu’à maintenant a surtout été à la base de constat empirique) sur l’état physiologique et psychologique des personnes (Kidd et coll., 2006 ; Pollack, 2005 ; Tapus et Mataric, 2006 ; Wada et Shibata, 2007). Ce constat tend à voir un rapide changement de perspective.

Cette évolution est notamment celle observable dans la terminologie associée à ce domaine. En effet, le terme de *l’assistive social robots* (Broekens et coll., 2009) réfère au robot de services avec une dimension de « socialité », comme ceux des *social robots* ou *sociable robots*. Or actuellement, la communauté parle de plus en plus de SAR (*social ou socially assistive robots*) qui existait de manière parallèle (Feil-Seifer et Mataric, 2005) et qui à première vue ne semble pas si différent, mais elle change en réalité la perspective d’approche. En effet, ce second terme tend à considérer la robotique de services orientée par un service à dimension sociale. L’ampleur de la dissémination de cette seconde acceptation tend donc à nous faire davantage réfléchir sur la dimension éthique (Feil-Seifer et Mataric,

2011) puisque le choix de la « socialité » tend à primer sur un service qui devient lui-même social, une tendance qui semble aussi être favorisée par le développement rapide de la robotique sociale et l'émergence de nombreux « robots compagnons » dans le cadre de la robolution. D'ailleurs, des auteurs tels que (Torrey, Fussell, et Kiesler, 2008) tendent à adopter cette seconde perspective en partant du concept de Smart Robot qui tend à devenir aidant, en analysant la perception humaine de l'aide apportée par ce robot social. Les auteurs cherchent ainsi à montrer qu'il existe un certain coût social à concevoir à travers trois paramètres : l'influence de la tâche centrale de l'aide apportée, les caractéristiques de l'aidé ainsi que celles de l'aidant. La fin de cette étude fait une emphase suggérant que le degré de cette « mécanique artificielle » doit être suffisant et remarquable, mais en contraste avec l'aidant humain pour en assurer une dimension complémentaire.

### ***2.3.2. « Compagnons » : vers une robotique de services à caractère social***

Suite à la section précédente, nous pouvons rappeler que notre approche utilise des primitives langagières à travers le robot qui apporterait donc une dimension sociale indéniable, car ses outils communicationnels sont supposés avoir un fort effet sur l'interaction et l'état relationnel qui en découlerait. En effet, notre objectif est avant tout d'observer l'effet de ces primitives. En revanche, ces primitives ne définissent pas le rôle, ce sera le service apporté qui le déterminera, de même qu'il précisera le contenu des primitives à utiliser en fonction de ce rôle. Ceci souligne donc l'importance de créer ce rôle social au robot, autour d'un service utile, tel que dans la perspective terminologique de (Broekens et coll., 2009). Ce service pourra alors être complété au fil des interactions par la modification de la dimension socio-affective du fait de la dynamique communicative outillée par les primitives testées. Ce ne sera donc que dans un second temps que l'interaction peut potentiellement aboutir à cette caractéristique de « compagnon » du robot, à socialité plus importante. Ainsi l'augmentation permettant d'avoir des effets bénéfiques sur le service qui peut être apporté à la personne, relèverait d'une motivation sociétale et non d'un objectif. Nous travaillerons donc sur un robot de services développant sa dimension sociale au fil de ses interactions avec la personne âgée.

L'un des éléments constitutifs importants de l'hypothèse de recherche (cf. section 1.5.5) réside dans ce rôle social qui est donc indispensable à tout établissement de glu socio-affective. Cette glu va ainsi elle-même remanier et transformer ce rôle au fil des interactions, un rôle évolutif qui lui-même aura une influence sur la transformation dynamique de la glu. Le problème est donc de trouver une fonction, d'utilité sociale, mais qui soit également propre au robot, par une activité qu'un humain ne peut intrinsèquement pas réaliser. De plus, d'un point de vue éthique, ce robot ne peut pas remplacer l'humain, de par sa nature et parce qu'il ne reste qu'une machine. En revanche, il n'aura de sens et de possibilité de perdurer que s'il arrive à bénéficier d'une existence sociale, qui va s'affirmer précisément par ce rôle, et que seules ses caractéristiques propres lui permettront d'accomplir, en augmentant l'existence socioculturelle de l'humain.

Cette perspective amène ainsi à réfléchir sur la typologie des robots dits « compagnons » (Turkle, 2006), développés par analogie d'apparence, renvoyant par exemple à l'image des « animaux de compagnie » (e.g Aibo, Pleo ou les jouets robotisés pour enfants). Par extension, ce terme « compagnon » définissant ces caractéristiques sociales s'est développé (p. ex. social robot, sociable robot ou cobot – pour désigner le robot collaboratif). Or cette dimension peut également être liée à l'apparition d'un réel besoin résultant d'un changement sociétal progressif, mais profond, tendant à combler des besoins insatisfaits de notre nature intrinsèquement sociale, qui s'accompagne également du développement des NTICs (Nouvelles Technologies de l'Information et de la Communication) ou encore à

l'instar de l'apparition des NACs (Nouveaux Animaux de Compagnie) (Bellangeon, 1984). Pour les animaux, ce statut de « compagnie » n'est officialisé qu'en 2000. Ce statut pousse même à l'extrême les modifications génétiques pour que les caractéristiques de ces animaux se rapprochent au plus d'un idéal de représentation socioculturelle maximisant cette fonction de compagnon. La transition des animaux de compagnie actuels est passée par une phase de domestication, associée à des rôles fonctionnels (bergers, de combat, de chasse etc.) et cela depuis 8500 ans avant notre ère (Vigne et Guilaine, 2004), mais son changement de statut « d'animal domestiqué » à « animal de compagnie » ne s'est réalisé que sur de longues années d'« apprivoisement » (Guichet, 2011). Par ailleurs, les valeurs attribuées aux traits et caractéristiques liés à cette capacité d'apprivoisement ne sont pas évidentes à déterminer, même pour un animal apprivoisé devenu compagnon (p. ex. le pitbull – Digard, 2004). Cette valeur est également socioculturellement dépendante (Millet, Digard, et Mezger, 1989), ce que nous voyons par exemple par l'acceptation de pouvoir moralement et culturellement « manger » un animal, qui change alors de statut d'être à objet (p. ex. le chien, le lapin, le cheval, la vache dont la perception comme « objet comestible » est variable en fonction des représentations et normes socioculturelles).

En revanche, ce terme de « compagnon » ne fait pas référence à un rôle pré-existant dans notre densité sociale (i.e. le métier ancien de « dame de compagnie » se réinvente actuellement mais, même si la relation y est pointée comme essentielle, il s'agit bien aussi concrètement de services à rendre à la personne). En effet, serait-il pertinent et acceptable de placer un tel robot à usage directement « compagnon » devant une personne en l'introduisant comme suit: « allez-y, maintenant parlez avec ce robot, ceci est votre compagnon » ? En revanche, il a été associé à cette terminologie de par sa dimension sociale qui ne découle que de son comportement qui se veut « amical » dans le cadre d'une société animiste, ou « collaborative » dans le domaine industriel, mais qui relève surtout du caractère empathique ou d'attachement qui le ferait dès lors devenir progressivement compagnon. En effet, l'humain lui-même ne développe ses affinités interpersonnelles qu'au fur et à mesure de ses interactions, un « ami » ne se présentant jamais comme un état de fait, notamment une fois que cet humain a assimilé la norme socioculturelle de l'environnement dans lequel il vit (à la différence du jeune enfant qui n'a pas cette inhibition de distance).

Dans cette thèse, le robot sera associé à un rôle de « majordome d'habitat intelligent » soit une télécommande d'actions domotiques contrôlables par commande vocale – ce qu'un humain ne sait pas faire. En effet, l'humain ne peut pas, à travers des canaux informatiques, actionner des protocoles de type KNX pour contrôler à distance des actionneurs domotiques. Il apporte ainsi par son rôle, un service domotique utile s'insérant dans le contexte des environnements d'intelligence ambiante, développés pour faciliter le maintien à domicile des personnes âgées. Ce rôle est ainsi rattaché à une motivation et un besoin sociétal cohérent avec les enjeux actuels. Dès lors, la « machine » de services majordomes peut être perçue comme un robot, c'est-à-dire un sujet qui rentrera inévitablement en relation avec la personne. Le travail de cette thèse est précisément de montrer que la nature de cette relation est manipulable. De plus, elle doit être comprise et contrôlée pour et par l'humain, ce qui en conséquence pourrait amener ce robot à entrer dans la gamme de ceux communément nommés compagnons, bien que cela n'en soit pas un but dans l'étude. En revanche, quel robot considérer parmi les multiples formes existantes ?

### ***2.3.3. De la robotique interactionnelle au choix du robot***

Cette dernière section d'état de l'art va résumer les diverses notions abordées à travers les précédentes parties des Chapitres 1 et 2. Elle fera cette présentation par un rapide survol des approches interactionnelles en robotique. Elle permettra notamment de justifier le choix

du robot utilisé dans cette étude, dont les implications sont motivées à travers les acquis théoriques précédemment présentés.

La robotique est un champ extrêmement étendu, et, quel que soit le champ d'application, cet objet est amené à interagir avec l'humain, directement ou indirectement. L'interaction homme-robot tel que défini dans le cadre de la robotique interactionnelle semble se diviser en deux paradigmes selon (Goodrich et Schultz, 2007) : la téléopération (cf. Vertut, 2013 – pour un état de l'art) et le contrôle supervisé (Jenkins, Nicolescu, et Mataric, 2004). Dans le premier cas c'est l'humain qui contrôle entièrement le robot par une manette, dans le second, l'humain supervise le comportement d'un système autonome et il intervient si nécessaire. En fonction de la coprésence des interactants, ces auteurs définissent deux types d'interaction possibles :

- *remote interaction* : associée à un décalage spatial et temporel de l'humain et du robot qui interagissent ensembles (p. ex. dans le cadre de l'aérospatial),
- *proximate interaction* : l'humain et le robot sont colocalisés et partagent le même environnement et temps interactionnel.

D'après Goodrich et Schultz, partant de ces paradigmes qui s'inscrivent dans le cadre d'une robotique autonome, une grande vague d'études a tout d'abord été notée dans les années 1980 décrite comme une robotique dite *behavior-based*. Cette dernière est basée sur les comportements suivant les courants d'émergence (cf. section 1.2). L'avantage par rapport aux méthodes antérieures résidait dans le fait que des réponses un peu plus élaborées pouvaient être envisagées au sein d'un environnement d'interaction changeant. Puis, une seconde vague, est venue des applications de systèmes à architectures hybrides en améliorant la réactivité de ces robots. Les systèmes évoluent alors vers des comportements anthropomorphiques alors qu'ils étaient jusque là focalisés sur les capacités de déplacement.

À l'échelle nationale, il existe un GdR (Groupement de Recherche)<sup>76</sup> autour de la robotique. Ce dernier est composé de huit groupes de travail dont le GT5 concerne les interactions-personnes / systèmes robotiques dans lequel nous retrouvons le champ de la robotique dite sociale.

Les développements de la robotique sociale, dont (Fong, Nourbakhsh, et Dautenhahn, 2003) proposent un état de l'art relativement complet, se focalisent sur un type d'interaction spécifique introduisant la notion de *socially interactive robots* qui sont considérés comme un sous-ensemble de cette robotique sociale. Les applications actuelles se présentent principalement sous forme de plateformes de recherche, de jouets, d'outils d'éducation ou d'aides thérapeutiques.

(Dautenhahn, 1999) propose une définition de la robotique sociale, qui semble rejoindre de manière générale les approches de la communauté :

*“Social robots are embodied agents that are part of a heterogeneous group: a society of robots or humans. They are able to recognize each other and engage in social interactions, they possess histories (perceive and interpret the world in terms of their own experience), and they explicitly communicate with and learn from each other.”*

Selon (Fong et coll., 2003), cette vision intégrative de cette technologie dans un « groupe social » introduit une notion qui est relativement spécifique dans sa terminologie

---

<sup>76</sup> Le site disponible du groupement de recherche est disponible sur : <http://www.gdr-robotique.org>. A noter que les JJCR (journées de jeunes chercheurs en robotique) qui rassemblent la communauté robotique des jeunes chercheurs sur l'ensemble des thématiques transverses étaient très peu représentées lors de l'édition 2014. Cela a sollicité mobiliser certains doctorants du LIG pour essayer d'introduire un plus grand nombre de jeunes chercheurs en robotique sociale lors de l'édition 2015 pour présenter notamment le groupe du travail de robotique sociale Robodoc-LIG avec son projet RACU (Sasa et coll., ICSR-2015), et d'ouvrir les appels à participation sur des disciplines en sciences humaines lors de l'édition 2016.

anglaise qu'on retrouve sous les termes *embeddedness* qui rejoint également le domaine des objets connectés (*IoT*). Ce terme fait référence à trois caractéristiques : 1) la situation dans un environnement social pour interagir avec des humains et des agents ; 2) le fait d'être structurellement intégré dans cet environnement social ; 3) avoir au moins une maîtrise partielle des schémas interactionnels humains, notamment les tours de parole. Ce dernier point reste l'une des grandes complexités de cette discipline puisque la connaissance de ces « schémas interactionnels » est encore très balbutiante et surtout extrêmement variable en fonction du contexte interactionnel, nécessitant très certainement une compréhension plus profonde du comportement communicatif humain.

Ce comportement communicatif de l'humain n'existe certainement pas dans des schémas génériques puisque l'humain est lui-même en perpétuel changement et adaptation de son environnement. Ainsi ce ne serait pas une amélioration unilatérale que le robot apporterait avec une augmentation de ces capacités interactionnelles à « coller » à celui de l'humain. Le système composé par l'humain et le robot devrait ainsi globalement aboutir à une continuité communicative, prenant en compte l'adaptation et les modifications que chacun va avoir l'un sur l'autre ainsi que sur leur environnement communicationnel, changeant de cette manière le système d'interaction dans son ensemble.

Selon (Breazeal, 2003), la partie sociale d'un robot dépendrait de quatre dimensions définies comme :

- *evocative* : qui s'appuie sur la tendance humaine à anthropomorphiser la technologie avec laquelle il interagit ;
- *interface* : à savoir avoir la capacité de produire des objets socio-affectifs et communicatifs proches de ceux produits par l'humain ou au moins interprétable par ce dernier de manière à assurer une continuité interactionnelle ;
- *receptive* : la capacité d'apprentissage du robot à partir de la capture qu'il peut faire à partir de son environnement ;
- *sociable* : qui fait référence conjointement à la notion d'*engagement*, spécifique à la robotique sociale, mais dans une approche de modélisation de cognition sociale.

Le fait que la robotique se soit confrontée à la notion d'*engagement* (cf. section 1.4.3), l'a assez rapidement sensibilisée à une approche de dialogue multimodale et située, en comparaison des approches classiques en linguistique (hors pragmatique) ou en traitement automatique du langage. En effet, ces disciplines avaient d'abord privilégié le contenu du message et sa forme influencée par le fort ancrage de la forme écrite du langage. De même, les interactions abordées par le point de vue des sciences de la communication apportent certes des dimensions fonctionnelles au langage, mais elle est surtout construite autour de la conception d'un message transmis entre l'émetteur et le récepteur, ce qui réduit fortement la formalisation du dialogue à un échange alterné de tours de parole, composés d'un contenu linguistique.

Ainsi la robotique dont les observations ont d'abord été empiriques, rappelle ainsi que toute interaction est située et que la dimension communicative peut être graduellement langagière.

*“Dialogue, regardless of form, is meaningful only if it is grounded, i.e., when the symbols used by each party describe common concepts. If the symbols differ, information exchange or learning must take place before communication can proceed. Although human-robot communication can occur in many different ways, we consider there to be three primary types of dialogue: low-level (pre-linguistic), non-verbal, and natural language.”* (Fong et coll., 2003)

Dans cette idée de langage *low-level* ou bas niveau, (Billard et Dautenhahn, 1998) décrivent par exemple que certains travaux ont une approche s'appuyant sur ce qu'ils appellent comme des « protolangages ». Dans leur étude, ils utilisent des données acquises de l'espace « visuel » recueilli par les capteurs du robot. Ils synthétisent ensuite un langage

primaire qui associe ces données acquises de manière spatio-temporelle pour en produire un langage d'apprentissage machine.

(Steels, 2001 ; Steels et Kaplan, 2000) partent quant à eux de l'hypothèse que la communication peut être le *bootstrap* du processus d'apprentissage social et constate que la signification initiale donnée au langage est fortement dépendante du contexte interactionnel. Enfin la question du langage naturel dépendant d'un nombre important de facteurs physiques et perceptuels est également abordée. Ces éléments seraient orientés par des dimensions socioculturelles dépendant de la situation, mais également de l'humain lui-même, ce qui rendrait difficile d'envisager à l'heure actuelle un parallélisme entre l'interaction homme-robot et celui d'un dialogue de langage spontané d'une interaction homme-homme (Severinson-Eklundh, Green, et Hüttenrauch, 2003).

Enfin, afin d'intégrer cette technologie dans la sphère sociale, notamment à travers le courant de la « robolution » (Bonnell, 2010 ; Himbert, 2012), l'anthropomorphisation socio-affective du robot social passerait par des apparences variées. Des exemples de ces designs sont représentés sur un continuum dans la Figure 16. L'esthétique et la forme associées aux robots pourraient être définies comme ci-dessous :

- Androïde : des formes tendant au maximum à se rapprocher à la morphologie humaines par mimétisme exact. Exemples : les robots Geminoïdes d'Ishiguro, l'imitation d'un enfant comme le robot Kaspar dans les études sur l'autisme ou la poupée My Little Baby robotisée.
- Humanoïde : qui hérite explicitement des traits socio-affectifs humains. (p. ex. Icube, ASIMO, Nao, Pepper, Romeo, Poppy, Robosoft, Twendy-One etc.).
- Animaloïde ou de type « pet » : c'est-à-dire héritant de l'animal intégré dans la sphère sociale dans un rôle éventuellement directement assimilable (Lestel, 2009) à un compagnon (p. ex. : Aibo, Genibo, Zoomer, Karotz etc.).
- Robot à « formes hybrides » : ce sont des formes d'artefacts qui sont reconnaissables par la présence de fortes composantes humaines fortes mais retransposer sur des formes cartoonistes ou *pets* (p. ex. : Reeti, Kismet, iCat etc.)
- Peluche ou personnage robotisé : ils représentent un transfert de la dimension de compagnon sur le personnage hérité culturellement du dessin animé ou de la peluche (p. ex. Pleo, Paro, Kabochan etc.).
- « Objet animé » : il s'agit d'un objet réel de la sphère sociale humaine qui n'est normalement pas animé, mais qui est retravaillé sur cette dimension pour devenir robot à dimension socio-affective (p. ex. AUR de G. Hoffman qui est une réplique de la lampe Pixar robotisée).
- Robot « mignon » : formes émergentes actuelles avec quelques traces d'anthropomorphisme ou de *pets* (dont les attributs sont dépendants de la culture) mais très réduites tendant à influencer la dimension de l'attachement par la forme du robot. (p. ex. : Buddy, Kuri, PaPeRo, Emox etc.)
- Robot « à forme propre » : ce sont des robots dont l'apparence atteint des formes propres au robot avec un anthropomorphisme d'apparence réduit au maximum, mais qui peut au contraire présenter un fort anthropomorphisme perçu de par son comportement socio-affectif (p. ex. : Jibo, Shimon, Diya-One etc.)
- Robot de téléprésence : la forme peut être plus ou moins explicitement anthropomorphique, mais généralement, il dispose d'une tablette projetant soit l'image d'un humain, d'un agent ou d'une entité dotée de composantes personnalisées (e.g : des yeux).

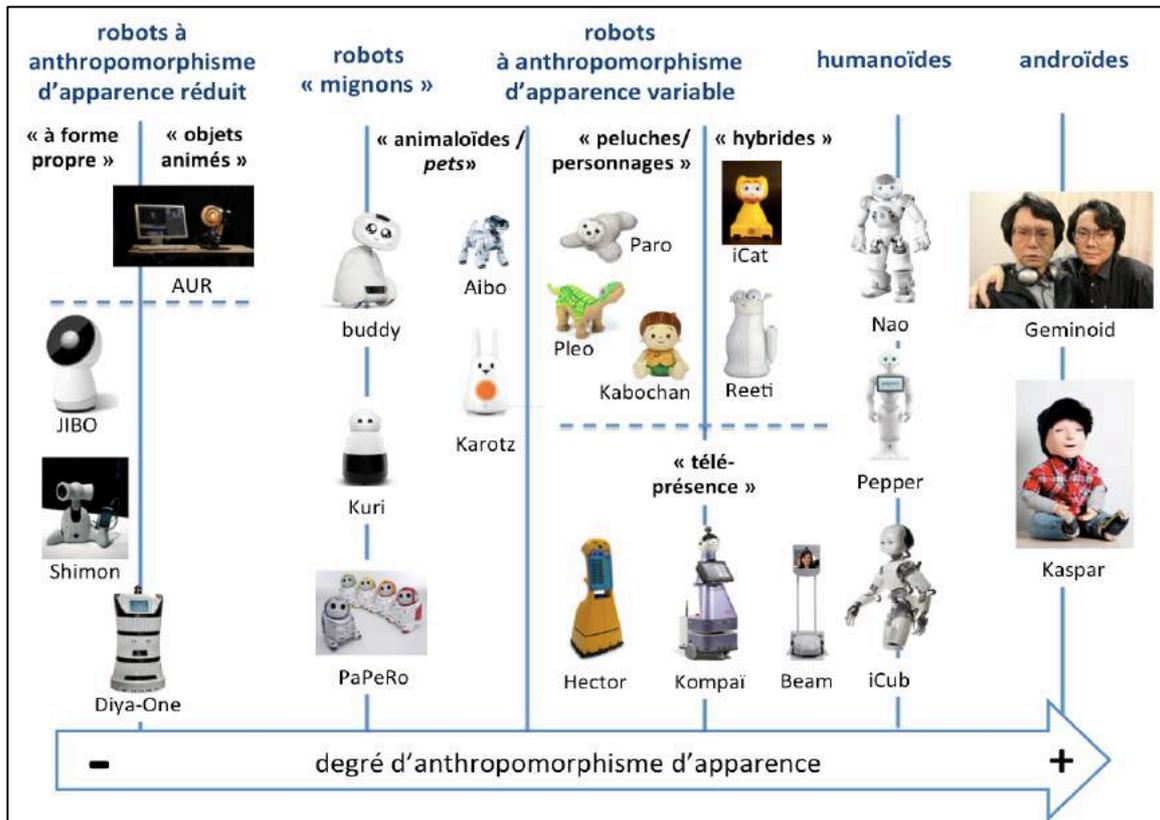


Figure 16: Exemples de robots sociaux sur un continuum d'anthropomorphisme d'apparence

Dès lors que les primitives vocales de notre hypothèse sont issues de productions humaines, l'anthropomorphisme est déjà conséquent, quel que soit le robot qui sera choisi. Le choix méthodologique a donc été de limiter cet anthropomorphisme pour le réduire à ce seul paramètre vocal. En effet, plus l'apparence physique présente des indices structurels évocateurs, plus il ouvre la possibilité d'une projection et d'une perception animiste et anthropomorphique, résultant sur des attentes de formes interactionnelles et d'intelligence dont les incongruités donneront lieu à des effets d'*Uncanny Valley* (cf. section 1.4.2). Ainsi, dans le cadre de l'étude, nous avons choisi de travailler avec le robot Emox, de l'entreprise robotique Awabot, dont le prototype utilisé est mentionné par « ancien design » dans la Figure 17 ci-dessous.



Figure 17: robot Emox (Awabot) (à gauche version sur laquelle les expériences de la thèse sont réalisées – à droite son nouveau design)

Emox a donc été choisi pour son faible anthropomorphisme d'apparence et sa non-ressemblance à un animal (ou en tout cas, qui reste moindre par rapport au Paro le phoque par exemple, ou au chien Aibo). D'ailleurs, pour réduire encore les effets de cette apparence, un nouveau design a été redéfini et développé (cf. Figure 17), au fil des échanges avec l'industriel Awabot, dans le cadre du projet Interabot<sup>77</sup>, pour lequel cette thèse a également apporté sa contribution.

De plus, a) les degrés de liberté visuels et de mouvement, b) la faible taille permettant un déplacement aisé dans un environnement du type appartement intelligent, c) la couleur relativement neutre et asexuée, d) la possibilité d'ouvrir le code Urbi du robot pour l'interfacer avec les programmes de la plateforme expérimentale, ce qui permettait de redéfinir des primitives interactionnelles du robot, le tout satisfaisait les besoins théoriques et techniques de nos contraintes expérimentales. En effet, l'étude outillée par ce robot nécessite un niveau de contrôle important de l'objet. Ainsi, de par cette petite taille et sa rondeur, une voix aigüe, très haute en fréquence fondamentale a été attribuée à ce robot. Les primitives portant la prosodie socio-affectives « gluante » (dont les stimuli sont directement issus du corpus E-wiz ou dont la prosodie a été copiée et amplifiée à partir des stimuli étiquetés de ce même corpus) ont été converties sur le principe présenté en section 1.5.5.3, pour obtenir un corpus de primitives à tester, d'esthétique vocale uniforme. Cette voix présente ainsi une fréquence fondamentale qui va au-delà de la limite de production possible pour le conduit vocal d'un humain (voire potentiellement possible pour un bébé ou un enfant), donnant un effet proche du cartoon, et réduisant davantage la dimension anthropomorphique du robot sur ce paramètre vocal. Cette hauteur de voix a d'ailleurs été perceptivement décrite comme « celle qui est le plus en adéquation avec le design du robot » et permettait un bon niveau de reconnaissance des valeurs socio-affectives associées aux stimuli de type micro-expressions vocales sur une étude antérieure à cette thèse (Sasa, Aubergé, Franck, Guillaume, et Moujtahid, 2012).

---

<sup>77</sup> Interabot est un projet Investissements d'Avenir BGLE n°2, porté par l'entreprise de robotique lyonnaise Awabot, en collaboration avec les laboratoires LIRIS, LIG et l'industriel Voxler.



## Chapitre 3. Méthodologie et protocoles expérimentaux

« On ne peut pas se mettre à la fenêtre pour se regarder passer dans la rue »

(Auguste Comte)

### 3.1. Préambule autour des sciences de l'information et sciences du langage

« L'intelligence est dans les données ». Cela relève d'une idée qui sous-tend le Traitement Automatique des Langues et les approches *Big Data*. Un exemple illustratif de cette pensée se trouve dans l'observation du changement de paradigme de la synthèse de parole. Cette dernière tout d'abord développée dans des perspectives très cartésiennes de méthodes par concaténation d'unités a évolué ensuite vers des méthodes de synthèse par corpus (Sagisaka, 1992). Ce corpus porte les manifestations de la connaissance humaine, capturées dans le signal de parole pour être exploitées de manière pertinente. Produire et coder ces données a ainsi toujours été un enjeu important dans les études d'interactions humaines et nombreux sont les états de l'art présentant les méthodes de recueil de corpus pour le traitement automatique de la parole depuis (Sagisaka, 2003 ; Campbell, 2004). L'expérience de la synthèse vocale démontre la complexité à extraire et à modéliser les processus de la communication parlée. Elle met également en évidence les limites des approches basées sur les données puisque les systèmes basés sur corpus sont d'autant plus performants dans la reproduction du naturel qu'une modélisation intermédiaire est explicitée. Or pour l'instant, ces systèmes s'éloignent du naturel lorsque les modèles sont statistiques (p. ex. à l'instar de la synthèse HMM par corpus, dont la qualité dépend de la présence de descripteurs pertinents mais dont le fonctionnement reste implicite).

Lorsque cette donnée est observée sous l'angle de l'*Affective Computing*, elle s'intéresse à la modélisation des émotions et des affects à travers les mécanismes de perception/production, à la fois naturelle pour l'humain, mais d'une extrême complexité pour la technologie (Cowie, 2009). La nature de telles informations donne ainsi la possibilité d'étudier des objets porteurs de ces informations recueillies : soit dans des interactions « spontanées », soient dans celles qui sont « actées/simulées ». Pour ce dernier type de données, qui ne fait pas partie de l'approche de cette thèse, le risque est de récolter des objets caricaturaux (Campbell, 2004) ou de faire face à l'effet de *display rules* (Matsumoto, 1990) qui est un biais personnel, mais surtout culturel tendant à inhiber ou exagérer certains types de comportements.

D'un point de vue méthodologique, l'une des difficultés, pour observer l'humain à travers des interactions naturelles et spontanées, réside dans le fait que l'observateur est de même nature que l'observé. En cela, même avec une attention consciente de vouloir s'extérioriser et objectiver ses observations, ce dernier ne pourra s'empêcher d'être constamment attiré par sa nature profonde, le poussant ainsi à traiter les informations qu'il perçoit de manière tout à fait naturelle à travers son filtre perceptif. L'étude de l'homme nécessite donc un niveau de contrôle de cette « intelligence naïve » humaine en adoptant un « regard éloigné » dont des principes pertinents peuvent être cités dans l'éthologie (Lévi-Strauss, 1983 ; Cosnier et Bourgain, 1993).

C'est d'ailleurs cette intelligence qui tente d'être exploitée dans les méthodes de *crowdsourcing* (Howe, 2006) ou prend encore part à l'émergence de phénomènes tels que

l'uberisation<sup>78</sup>. Cependant, ces dernières méthodes ne peuvent s'appuyer que sur des observations empiriques laissant très peu de place pour la méthodologie scientifique. En effet, l'évolution informatique facilitant l'accès aux données en masse, le codage de ces données devient un enjeu important pour le développement technologique. Ce codage peut ainsi adopter des approches différentes : « catégorielle » (Cowie, 2009), « dimensionnelle » (Schlosberg, 1954 ; Plutchik, 1980 ; Kaiser, Wehrle, et Schenkel, 2009) ou encore « hybride » (Schröder et Cowie, 2006).

Les méthodes de *Big Data* s'appuient sur l'expertise d'une quantité massive de données au départ, pour appliquer ensuite des règles mathématiques stochastiques issues de modèles supposés de comportements, permettant alors de redéfinir des distributions permettant d'en déduire des phénomènes. Contrairement à ces méthodes qui ont un plus grand risque de devenir tautologiques par l'application des méthodes de type crowdsourcing associant un étiquetage formaté sur grandes échelles (Howe, 2006), cette thèse n'a pas de modèle qui décrit au départ les dimensions constitutives du processus relationnel humain à travers la communication. Elle doit donc s'appuyer sur une méthodologie qui lui permette de capturer des données en maximisant les phénomènes qu'elle veut observer pour répondre à ses hypothèses. Nous faisons ainsi le choix de cibler ces phénomènes par le niveau de granularité le plus bas. La méthode prendra ainsi en compte les plus petits éléments informatifs mesurables dans l'interaction pour en observer les effets, soit les primitives vocales définies dans notre hypothèse de manipulation de glu socio-affective. Nous reconstruirons ensuite cette connaissance, par la compréhension des phénomènes dans une approche hypothético-déductive permettant alors d'en modéliser une partie. Ce ne sera qu'à partir de ce moment-là que les modélisations seront extensibles pour traiter des données à plus grande échelle, car nous aurons commencé à cerner les mécanismes sous-jacents. Le changement de niveau de granularité d'observation et de traitement de cette donnée est alors un enjeu, qui dépasse très largement les compétences de cette thèse, mais qui est important pour la communauté scientifique. En effet, les verrous d'une granularité donnée permettant d'anticiper un changement de l'échelle d'observation des données. Si les niveaux les plus bas tendent petit à petit vers le même ordre de grandeur que celui des données traitées par les *Big Data*, cette augmentation permettrait une meilleure compréhension de l'humain, mais surtout du changement d'échelle relationnel par l'étendue des effets de ses comportements. Pour notre cas, il s'agirait des comportements liés au maillage interpersonnel résultant de ses processus communicationnels.

De fait, le corpus de cette étude tentera de récolter préférentiellement les traces informatives permettant d'accéder à la compréhension du comportement humain dans ses processus communicationnels, en rendant plus visible le remaniement relationnel qui en découle. Le développement des nouvelles technologies devrait alors être construit au plus près d'une écologie réaliste tout en maximisant les chances d'observer les phénomènes dont la nature est définie par les hypothèses scientifiques. En d'autres termes, il faut induire des productions de l'humain, qui sont des données denses en information pertinente, qui n'est autre que celle qui pourrait répondre à nos hypothèses. Cette information est celle qu'il traite par son intelligence naïve dans une situation réaliste, spontanée et écologique, possiblement recréée à des fins expérimentales, afin de pouvoir contrôler les paramètres induits dans l'observation.

Or le seul moyen de cadrer scientifiquement et éthiquement l'observation de l'intelligence naïve de l'humain qui contrôle son comportement reste la méthodologie dont le fondement ne peut être que la méthode scientifique, celle appliquée de manière systématique dans les sciences dites dures (Kuhn et Meyer, 2008). Elle permettra ainsi de poser les jalons de fortes hypothèses portant sur les phénomènes qui sont les cibles de notre observation.

---

<sup>78</sup> Aperçu général du concept : (<http://www.creads.fr/blog/tendance-design-graphique/uberisation-travail-debat>, )

En effet, ces cibles sont difficilement atteignables par l'unique observation d'un contexte naturel existant sur le seul fait de l'empirisme, puisque l'on ne peut savoir si elles sont pertinemment existantes dans ce contexte, tout particulièrement sans le contrôle et la connaissance des conditions de sa mise en œuvre. De plus, pour ces données brutes « récoltées dans la nature » la caractérisation des paramètres d'influences du comportement observé ne peut forcément se faire qu'a posteriori en s'appuyant sur des grilles d'observations issues de connaissances théoriques préalablement acceptées par l'observateur, augmentant là encore le risque de la tautologie. Cela rend ainsi difficile l'évolution de ces théories, alors qu'elles ne sont fondamentalement que des pensées éphémères plus ou moins consensuelles à un temps donné. Cela ne favorise pas non plus l'observation de phénomènes non attendus et imprévisibles pouvant être à la base de nouvelles théories. En revanche, l'intérêt et l'avantage des courants actuels sont que l'apport des nouvelles technologie permet de valider ou d'invalider ces théories et hypothèses à travers des modélisations, dont on peut mesurer l'effet de l'utilisation que les usagers finaux en feront, et cela de manière rapide et itérative.

L'approche méthodologique choisie tentera alors de gérer : 1) la complexité de l'humain et de sa communication en s'appuyant sur des connaissances pluridisciplinaires et en tentant de faire converger ces apports dans une construction interdisciplinaire ; 2) une co-construction de cette méthodologie de recherche par une ouverture à une dimension collaborative à multiples acteurs par couches itératives ; 3) un contrôle de la subjectivité et de l'intelligence humaine par des cadrages protocolaires au sein de boucles agiles tout en essayant de préserver au maximum une écologie d'interaction contrôlée et proche de la réalité humaine.

### 3.2. Vers des données écologiques et spontanées pour un apprentissage machine itératif

En 1770, le *Mechanical Turk* développé par van Kempelen produit une illusion qui introduit une forme de dématérialisation de la tâche à travers une figurine mécanique (Standage, 2003), ancienne version de ce nous appelons la technique du magicien d'Oz ou Wizard of Oz – WoZ, inspiré du récit *The Wonderful Wizard of Oz* (Baum, 1987), ou des techniques de PNAMBIC – Pay No Attention to the Man Behind the Curtain (Bernstein in Newell, 1987). Ce concept est plus ou moins connu et adopté selon les disciplines. Émergeant des études sur l'intelligence artificielle, notamment dans le développement de CAL (calendar access language), contrainte à la difficulté de gérer la complexité du langage naturel humain, il s'appuyait initialement sur une méthodologie empirique et itérative de conception d'interface, posant ainsi un nouveau paradigme d'IHM (Kelley, 1984). Cette technique apparaît également dans le domaine du traitement automatique des langues comme les travaux du CNET dans le projet Esprit – PALABRE (Guyomard et Siroux, 1986). En robotique, il s'agit d'une pratique courante pour pallier les difficultés d'autonomisation de cette technologie puisqu'elle fait partie des premières manipulations permettant de contrôler sans limites autant de choses que nécessaire, telle que mouvements, navigation, parole, gestes, regards etc. du robot (Pettersson et Wik, 2014 ; Riek, 2012). Le contexte d'interaction est également un point contrôlable et à intégrer dans le WoZ, qui se déclinerait même sous plusieurs formes en fonction du point de vue scientifique adopté (Steinfeld, Jenkins, et Scassellati, 2009) :

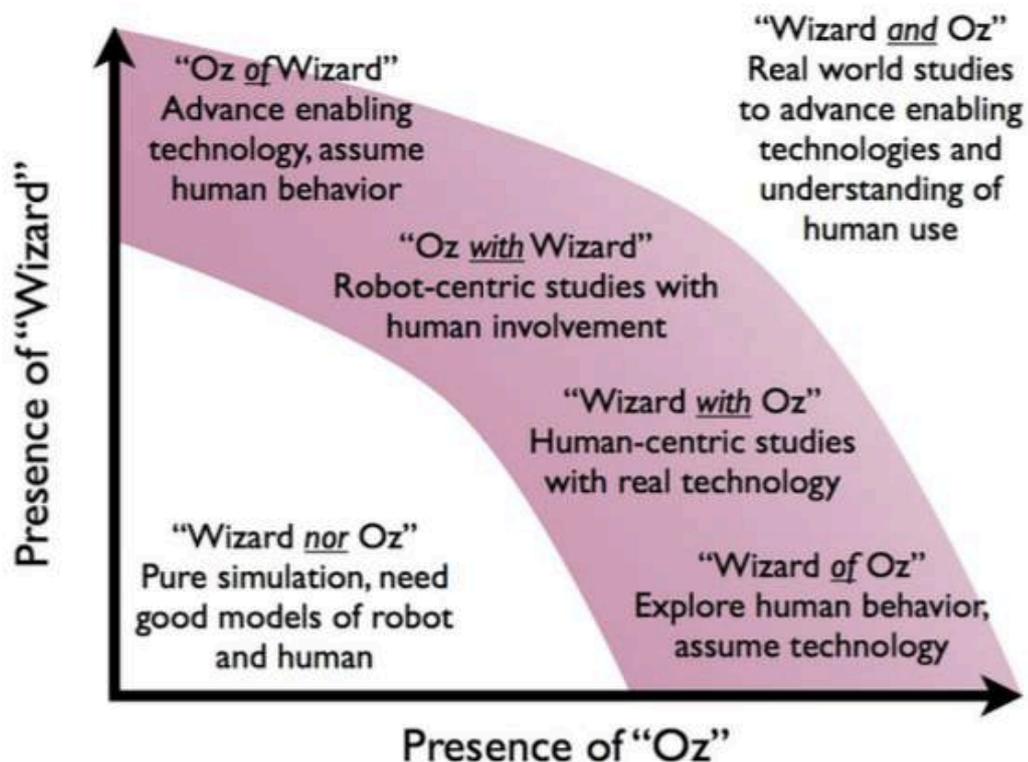


Figure 18 Conception de la combinaison Wizard/Oz selon (Steinfeld et coll., 2009)

D'ailleurs, dans le cadre des HRI s'intéressant à l'interaction sociale, cette technique placerait davantage le robot comme un proxy de l'humain que sous forme d'une entité, l'interaction étant non pas véritablement humain-robot mais plutôt humain-humain à travers le robot (Weiss, 2010). Cette technique a d'ailleurs fait l'objet de nombreuses réflexions

éthiques à l'étape de la « levée de l'illusion » au moment des débriefings d'expérimentation (Boden et coll., 2017), décrit encore sous les termes de *Turing Deceptions* (Miller, 2010). Dans le cadre de cette thèse, les considérations éthiques vont s'appuyer sur une dimension de co-construction intégrant cette technique au sein d'une méthodologie living-lab permettant d'encadrer collectivement les effets de la technologie et le protocole expérimental lui-même (cf. section 3.3). La limite relevée dans les WoZ en HRI est notamment le manque de critères permettant d'orienter la tâche, bien que les tâches plus difficiles puissent lever des verrous d'ingénierie, même à partir des observations empiriques réalisées (Breazeal, Kidd, Thomaz, Hoffman, et Berlin, 2005 ; Riek, 2012). En revanche, le Traitement Automatique de la Parole s'est rapidement intéressé à ces problèmes :

*“Although it is certainly better to rely on analyses of human-human interactions than to rely on intuitions alone, the fact remains that human-human interactions are not the same as human-computer interactions and it would be surprising if they followed precisely the same rules. The designer is caught in a vicious circle-it is necessary to know the characteristics of dialogues between people and automata in order to be able to build the system, but it is impossible to know what such dialogues would be like until such a system has been built. Turning to evaluation, we find similar problems. It is possible to evaluate perceptions of the system's usability but it is not possible to test its performance against a clear specification of what the system should be capable of, because this is unknown.”* (Fraser et Gilbert, 1991)

Pour ces auteurs, qui proposent notamment une taxonomie partielle des formats utilisables dans le cadre du TAL et du langage, le WoZ est un outil qui permet de : 1) simuler le futur système, en ayant conscience des limites de l'humain ; 2) spécifier le comportement du futur système ; 3) d'en produire une simulation convaincante, sachant que le système qui nous intéresse ici est celui de l'interaction englobante, modifiante et modifiée par ces sous-systèmes, dont les interlocuteurs. De même, cette technique n'est viable que sous le contrôle d'un scénario précis (Dahlbäck, Jönsson, et Ahrenberg, 1993 ; Fraser et Gilbert, 1991 ; Green, Huttenrauch, et Eklundh, 2004). Dans l'approche que nous suivons, il s'agit également d'une méthode qui permet d'induire des comportements spontanés et écologiques, observés chez l'humain, tout en gardant le contrôle sur les paramètres d'influence de la condition d'observation. Ces perspectives de Woz, semblables, mais à la fois différentes entre la robotique et le TAL dérivent ainsi un paradigme relativement spécifique dans cette étude, ce qui est l'objet de la section suivante.

### ***3.2.1. Magicien d'Oz : continuité paradigmatique de la télé-présence à l'autonomie***

Le robot est une technologie dont le niveau de contrôle par l'humain module le degré d'autonomie (cf. Figure 19). Du point de vue technique de la robotique, ce degré de contrôle semblera certainement totalement différent en fonction de la situation mentionnée et du point de vue de l'humain interagissant, il tend vers une ambiguïté transversale de la perception de soi et de l'autre. Pourtant notre étude relève bien d'une certaine forme de continuité paradigmatique dans une approche liant la robotique sociale et le TAL.

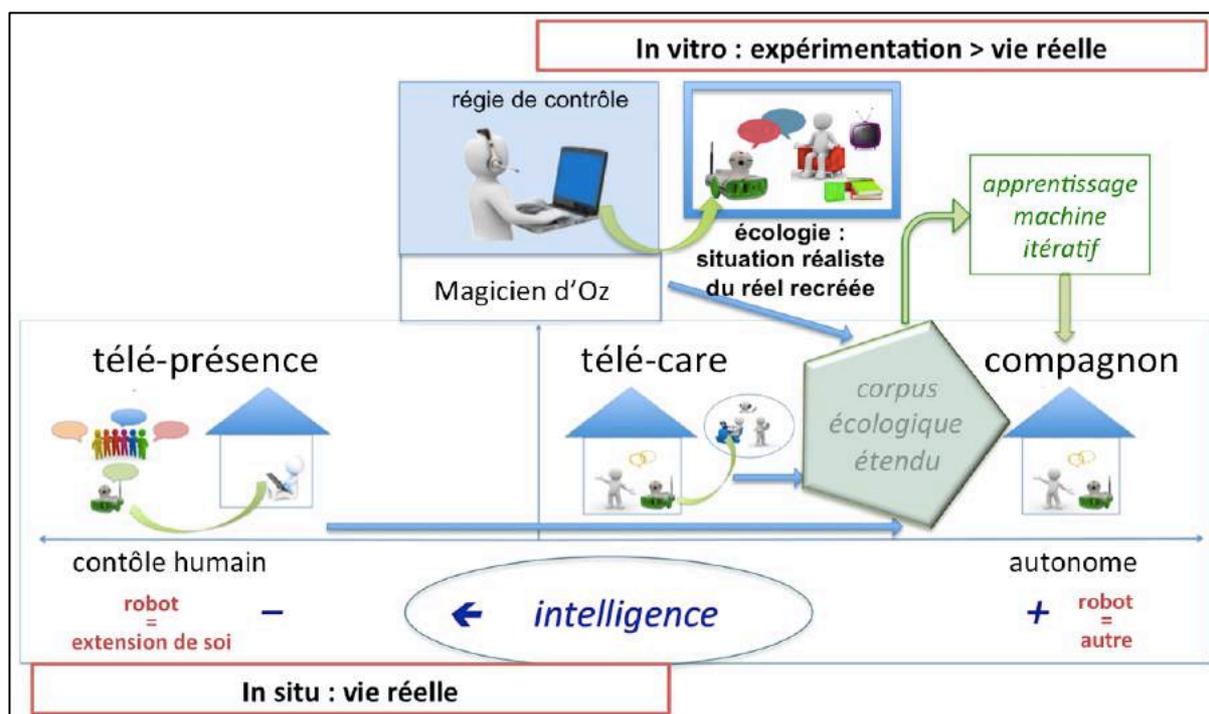


Figure 19 : Discontinuité perceptive de l'application robotique, continuité de la construction technologique (Aubergé, 2014)

En effet, sur un axe où ce contrôle humain laisserait la place à l'autonomie de la machine, la nature de l'humain-contrôleur peut changer à travers différents domaines d'application robotique. Cette technologie se situe ainsi à l'interface l'humain-contrôleur et les humains interactants, à travers :

- la téléprésence : entièrement contrôlé, le robot a un faible niveau d'autonomie et c'est l'utilisateur lui-même qui s'augmente en interagissant avec d'autres percevant l'humain-contrôleur comme leur interactant qui n'est pas dans leur espace physique commun.
- télé-care : des acteurs médicaux, socio-médicaux ou sociaux deviennent des contrôleurs du robot. L'interface met plus ou moins en avant l'humain-contrôleur, plaçant le robot dans une situation de « télésurveillance experte incarnée » lorsque les contrôleurs sont explicites, cas proches de la téléprésence. Mais ils peuvent aussi être dans une situation de « télésurveillance experte animée » lorsque le robot devient l'interface véhiculant exclusivement le suivi fourni par les contrôleurs professionnels, qui gardent alors le contrôle de toute réaction du robot, mais qui eux-mêmes ne sont plus les interlocuteurs directs.
- robot autonome : il s'agit de la situation où les réactions du robot sont entièrement gérées par le système lui-même, sans contrôle par l'humain et où l'humain interagissant avec le robot perçoit la technologie comme un « autre animé » qui est une entité propre.

La technique du magicien est ainsi à mi-chemin entre le contrôle et l'autonomie puisque d'un contrôleur identifié, on passe à un contrôleur caché, la frontière entre le soi et le robot en tant que contrôleur étant ténue. Il s'agit d'une condition d'interaction qui simule ainsi le comportement humain à travers le robot, mais qui devient également progressivement celui du robot lui-même. En revanche, comme le robot en Woz est présenté de manière à être envisagé comme un interactant potentiel à part entière, le corpus d'interactions récolté dans ces conditions induisant des productions spontanées vis-à-vis de cette écologie. La condition ainsi recréée de manière la plus réaliste possible des conditions réelles existantes. Ces

données composent alors la base pour un apprentissage machine, puisqu'elles reflètent la simulation des comportements que le robot aurait s'il était autonome, en fonction du rôle qu'il incarne dans ce contexte. Des itérations courtes de traitement/modélisation exploiteraient ainsi morceau par morceau ces données, augmentant cet apprentissage machine de manière itérative. Elle permettrait ainsi d'acquérir au fur et à mesure un degré d'autonomie et d'intelligence, à chaque fois qu'une dimension interactionnelle est modélisée. L'avantage de cette approche réside dans le fait que les contextes interactionnels restent inchangés, entre boucles itératives, ce qui permet d'une part l'accumulation de données pour une extension du corpus, et d'autre part une évaluation possible dans l'usage simulé de la technologie, qui se fait dans un contexte reproductible et surtout en parallèle du développement technologique.

Nous parlons ici de continuité, car pour l'humain qui interagit en face à face avec cette technologie, la perception du robot dans l'interaction est continue, avec un niveau d'intelligence perçue graduelle. Elle est même confondue dans le cadre du Woz, car cet humain est dans les mêmes conditions interactionnelles que face à un robot autonome. Alors que le robot de téléprésence existe déjà comme outil technologique viable dans la vie réelle, le passage à l'autonomisation est complexe puisqu'elle doit faire le transfert du contrôle humain sur la machine, tout en assurant la maîtrise des effets du comportement attribué au robot. La mesure de ces effets ne peut ainsi se réaliser que dans les traces des interactions naturelles et spontanées de cette situation de communication homme-robot, en plaçant ainsi le paradigme de recherche entre le robot autonome et le télé-care, afin de permettre le passage d'un contrôle en milieu expérimental au réel. Dans ce passage, la perception de l'humain interagissant reste exactement la même : qu'il y ait un contrôle ou non, le robot devrait être perçu de la même manière. En ayant donc en hypothèse scientifique des outils communicatifs avec un potentiel effet sur l'interaction, nous pouvons : d'une part, maîtriser la production et l'utilisation de ces outils dans une condition écologique, ce que permet l'usage de la technique du magicien d'Oz ; d'autre part, analyser et utiliser les données recueillies pour effectuer de l'apprentissage machine, et cela de manière itérative, car la mesure des effets des outils interactionnels ne peut se faire que progressivement du fait de la complexité de l'humain et de l'évolution de sa dynamique de communication.

### ***3.2.2. De l'induction à l'observation hypothético-déductive***

La motivation d'utiliser des techniques comme le magicien d'Oz a été explicitée du point de vue de la robotique et de l'informatique dans la section précédente. En revanche, elle relève aussi d'un choix et d'approches méthodologiques prenant en considération les disciplines de sciences humaines qui soulignent des contraintes incontournables, liées à la condition d'observation du comportement humain. Ainsi la prise en compte de ces contraintes aura directement un impact sur la qualité de la donnée récoltée et utilisée dans la modélisation technologique.

La communication humaine est un phénomène complexe, dont l'observation des fonctionnements globaux n'est rendue possible que dans des situations d'interactions restrictives et spécifiques. Ces mécanismes ne seraient donc observables que dans un contexte d'interaction situé, soit un environnement physique ancré dans un temps défini. Les outils d'observation doivent ainsi être choisis en fonction de la finesse requise pour décrire le comportement humain. Si nous observons ces phénomènes, directement dans la nature, tel qu'ils existent dans leurs formes brutes, tant est que nous puissions les reconnaître, nous n'avons aucune maîtrise du contexte et donc des paramètres susceptibles d'expliquer les processus mis en jeu dont résultent ces comportements. Idéalement, les phénomènes à étudier devraient donc être recueillis dans un contexte contrôlé tout en préservant l'écologie et le réalisme de la situation de communication (Labov, 1972, 1973 ; Wilson, 1987).

De plus, selon (Gadet, 2003), les mécanismes spontanés et naturels mis en jeu dans l'interaction ne sont susceptibles d'émerger que si le sujet se détache de toute méta-analyse sur ses propres productions. Le système attentionnel des personnes observées peut alors être réduit en le détournant des indices qui le rattachent au fait que :

1) il est dans la position d'un sujet d'observation (paradoxe de Labov), tout en sachant que cette situation génère elle-même sa propre écologie : *“To obtain the data most important for linguistic theory, we have to observe how people speak when they are not being observed.”* (Labov, 1973), *“Considerable care was taken to avoid the dangers of the ‘observer’s paradox’, i.e. the fact that the presence of the observer may destroy the phenomenon that s/he is observing.”* (Pooley, 1996). La présence de l'expérimentateur changeant l'écologie même du système : *“Social interaction can be observed through a two-way mirror ; tape recorders can be concealed or respondents can be asked their opinions about washing powder when the researcher is really sampling their vowel-sounds for future measurement. Social scientists may also disguise their purposes by pretending to be part of the group they are studying.”* (Cameron et coll., 1992). Il existerait alors des formats d'observations dites participantes incluant directement l'observateur dans l'interaction (Copans, 2008 ; Whyte, 1984).

2) il interagit en mettant en œuvre des mécanismes spécifiques, dont les processus vitaux et les plus résistants dans le temps sont rendus automatiques au cours de leur acquisition. Malheureusement, cet automatisme est biaisé, quand l'humain en reprend le contrôle en conscientisant cette action (p. ex. p. ex. la respiration devient irrégulière lorsque l'on se concentre sur sa régularité et plus encore, nous n'avons pas le contrôle des battements du cœur). Il en est ainsi de même pour la communication, qui est sujette à la création d'artefacts de par nature lorsqu'elle est conscientisée (Christensen, Sutton, et McIlwain, 2016).

Le recueil doit ainsi faire l'objet d'une méthodologie permettant de minimiser ces biais en recréant expérimentalement des écologies de communication spontanée et naturelle. L'un des outils permettant de reproduire et d'induire le début d'une communication est le robot ; d'une part, car il s'agit d'une technologie suffisamment nouvelle avec laquelle l'humain n'a pas encore établi de code commun laissant donc une liberté d'évolution de la dynamique interactive ; d'autre part, car cet objet permet de maîtriser les outils de communication de manière reproductible afin de pouvoir en mesurer les effets sur l'humain.

Rappelons que l'approche méthodologique proposée dans cette thèse est en continuité directe avec une approche hypothético-inductive expérimentale pour produire des données qui ensuite sont traitées par une démarche d'observation de base empirique, mais hypothético-déductive contrôlée. Cette approche a précédemment été utilisée par (Aubergé, Rilliard, et Audibert, 2005 ; Loyau, Aubergé, et Vanpé, 2006 ; Vanpé, 2011) en s'appuyant successivement sur :

- des hypothèses fortes de départ pour faciliter la maximisation de capture des phénomènes potentiellement pertinents à observer, sans dénaturer leur naturalité, dans une écologie réaliste, mais contrôlée : la constitution du corpus à l'état brut.

- une méthodologie empirique basée sur les connaissances éthologiques, s'appuyant sur le principe de *sit and watch* (Garfinkel, 1967/2005 ; Pléty, 1993 ; Tinbergen, 1963) afin de recueillir un maximum de données analysables en minimisant les *a priori* sur ce qui va être trouvé une fois le corpus recueilli : le codage et l'étiquetage.

- ces codages sont confrontés à des annotations externes et à des validations perceptives afin d'éviter les travers de la tautologie liée à l'expérimentation (à savoir fermer et contraindre les choix de codage en s'appuyant *a priori* sur des grilles d'observation issues de modèles théoriques). Il rajoute ainsi un niveau de contrôle de la subjectivité de codage, qui ne peut être évitée de par la nature intrinsèque des objets observés. Les validations perceptives s'appuient sur une approche ultime (Tinbergen, 1963) permettant de déterminer les valeurs fonctionnelles des objets interactionnels (Signorello, Aubergé, Vanpé, Granjon, et Audibert, 2010) ou encore d'évaluer leur reproductibilité falsifiable intraculturelle (Biasi, Auberge, et Granjon, 2012) ou interculturelle (Sasa, 2013a).

- soumettre à expérimentation les données en modélisant leur fonctionnement et déduisant les invariants à partir de lois hypothético-déductibles : l'évaluation des théories par prototypage technologique.

Le corpus E-Wiz (Aubergé, Audibert, et Rilliard, 2006) avait ainsi permis d'effectuer les premières explorations et formalisations de cette méthodologie appliquée pour son recueil et son traitement. Ce corpus était dès lors contraint à un cadre d'interactions homme-machine spatialement statique, mais permettait de passer par des états évolutifs de locuteurs (par induction émotionnelle et socio-affective), contraints dans la dynamique (par le format de la tâche et l'environnement expérimental), *in vitro* (contrôle en chambre sourde) et authentique (corpus spontané avec une écologie expérimentale réaliste). Les travaux successifs ont donc permis d'établir les bases des contraintes des conditions de recueil (paradigme de magicien d'Oz perturbé par une scénarisation nécessitant un détournement d'attention), du codage interne et externe ainsi que des méthodes de validation puis d'évaluation (Loyau et coll., 2006 ; Vanpé, 2011). Cet ensemble forme ainsi une démarche scientifique centrée sur le corpus servant à la fois à comme base d'observation empirique contrôlée, de support d'hypothèses et de modélisation théorique et comme matériel pour une évaluation prototypée. Ce sera donc à partir de cet héritage méthodologique que sera construite la méthodologie mise en place pour les travaux de cette thèse.

### 3.2.3. Observation de l'humain et living-labs

Les données sont donc au cœur de nos recherches, aussi bien du point de vue de l'analyse, du matériel de conception/développement, ou de l'évaluation, d'une perspective de sciences humaines, de sciences du langage, d'informatique ou de sciences de l'ingénierie des technologies de l'information et de la communication. D'un point de vue épistémologique, notre approche hypothético-inductive à empirisme déductif cherche à privilégier une récolte fine de données et replace ainsi le contexte de recueil dans un environnement *in vitro* contrôlé, tout en simulant un réalisme de contexte écologique. De fait, cet environnement lui-même, mais surtout la pratique et les motivations scientifiques qui en découlent vont directement influencer la méthodologie globale adoptée. La prochaine section abordera donc l'évolution des contextes et conditions expérimentales précédemment décrites par l'introduction d'environnement et méthodes permettant de confronter l'observation humaine à des situations plus réalistes, toujours en laboratoire, mais dont la claustration est amoindrie par un aménagement plus proche d'une situation réelle. Elle pourra même s'étendre au-delà du laboratoire en exportant la méthodologie à l'écologie réelle de l'observé. L'évolution de ces approches est d'autant plus importante à comprendre, car l'environnement expérimental choisi (la plateforme Domus du LIG), a lui-même changé d'approche méthodologique au fil des années. En effet, bien que le contexte physique, voire même les acteurs restent inchangés, la méthodologie des laboratoires d'usage (ce que Domus était anciennement) et celle des living-labs (ce vers quoi Domus tend, mais différemment, avec une approche plus expérimentale) sont à distinguer.

Le développement des technologies s'accompagne d'un contexte économique qui se veut favorable pour répondre aux besoins sociétaux, ce qui amène au début des années 70, des réflexions autour de l'usage ou *usability* en anglais. Cette notion peut être abordée par deux types de ressources : des groupes d'usage ou des laboratoires d'usage, les groupes pouvant faire partie intégrante des laboratoires. Les laboratoires d'usage sont ainsi environnements physiques dont les paramètres sont plus facilement contrôlables. Ils sont dirigés à travers des règles organisationnelles ou de management, investissant des environnements et méthodes dont on trouve des états de l'art (Nielsen, 1994). Les approches sont nombreuses : test d'usage, enregistrement de session design ou de design participatif, ou encore des évaluations heuristiques d'interfaces d'utilisation, dont (Buchanan, 1992) souligne les limites. Aux États-

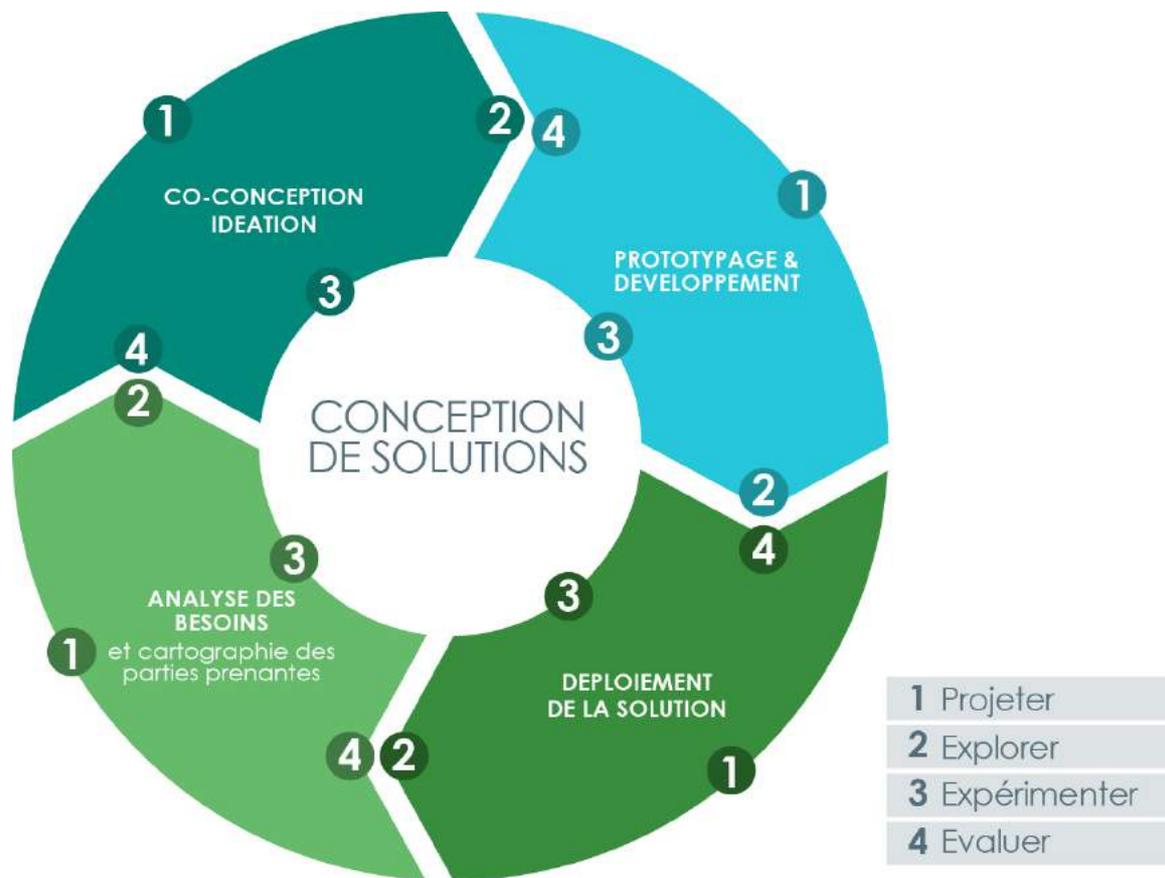
Unis, dans les années 90, ces environnements tendaient à se multiplier. Des plateformes fonctionnaient alors souvent de pair avec des entreprises, notamment dans le domaine des télécommunications, favorisant ainsi l'ouverture de la recherche au milieu industriel. Le format classique est ainsi la présence de l'environnement d'usage, une régie de contrôle permettant une observation hors champ d'usage et d'éventuels postes d'observation externes afin de limiter les interférences entre les sujets observés et les expérimentateurs (en général membre de l'équipe de développement du matériel testé). Dès lors, même si ces sessions sont enregistrées sous différents angles avec de multiples sources pour une analyse postérieure, l'une des possibilités mises en avant dans cette approche est l'observation en temps réel, le traitement automatisé des données récoltées étant un autre enjeu de ces environnements dont l'autre avantage est la richesse quantitative et qualitative des données recueillies grâce à l'aménagement technique associé. L'objectif de tels environnements est ainsi de pouvoir observer et/ou demander l'avis des utilisateurs quant à leur usage de certains outils voire même de l'environnement entier lui-même qui est outillé, qui d'ailleurs peut se présenter comme une copie conforme de situation de vie réelle. L'objectif premier de cette approche est ainsi dirigé par la motivation de développer et d'améliorer les outils finaux. Or les risques de ces approches malgré le format de l'expérimentation, sont les usagers eux-mêmes qui sont : 1) conscients d'être observés et 2) réceptifs aux buts de l'observation, ce qui augmente les biais qui nous ont orientés sur les choix inspirés d'ethnométhodologie déplaçant ainsi le paradoxe de l'observateur dans une nouvelle écologie d'observation. Cette démarche est peut-être donc adaptée si la technologie est une fin en soi, pour des objectifs d'amélioration ergonomique, mais elle n'est malheureusement pas suffisante pour la mise en place d'une méthodologie d'observation complexe.

Parmi ce type d'environnement, notons qu'il existe des plateformes orientées pour des utilisateurs spécifiques tels que LUSAGE qui propose un environnement dédié à l'étude des personnes âgées avec des troubles cognitifs (Pino et coll., 2009) par exemple. Notons d'ailleurs que certaines pratiques de cette plateforme tendent elles aussi vers le concept de Living-Lab, dont la frontière avec le laboratoire d'usage peut être parfois ténue.

Le concept de Living-Lab (LL) (Eriksson, Niitamo, et Kulkki, 2005 ; Schumacher et Feurstein, 2007) a commencé à se développer. Il porte des fonctions sur deux dimensions : une première horizontale en tant que lieu d'expérimentation physique (le cadre partagé avec le laboratoire d'usage), une seconde verticale qui révèle une pratique méthodologique (qui elle est totalement différente du laboratoire d'usage).

En France, à l'instar de Robert Picard et Antoine Vial, il existe actuellement une trentaine de LLs référencés, dont de nombreux sollicités dans le domaine de la gérontechnologie. Le forum LLSA (Living-lab en santé et autonomie) rassemble notamment les principaux membres de cette communauté en France (ACTIVAGEING, Autonom'Lab, CEN STIMCO, CEREMH, CMRRRF de Kerpape, i2ml, Connected Health Lab, Etape, CIC-IT evalab, Experiment'HAAL, Kyomed, LIO, MADoPA, PROMETEE, StreetLab, uCIL). Cette approche s'insère plus généralement dans ce qui est appelé les PSH (Personal Health Systems) quand celle-ci s'applique spécifiquement au domaine de la santé.

L'approche LL repose de manière générale sur cinq principes fondamentaux (Bergvall-Kareborn, Hoist, et Stahlbrost, 2009) : l'ouverture (nature des partenaires à impliquer dans la co-construction), l'égalité d'influence (entre les différents partenaires impliqués), le principe de réalité (celle liée à la contextualisation contrôlée et écologique), la création de valeurs (aussi bien économique que sur l'impact social – tangible pour tous les partenaires) et la pérennité (avec une vision à long terme) qui s'intègrent dans quatre phases cycliques basées sur : la co-confrontation, la co-exploration, la co-expérimentation et la co-évaluation.



## LE PROCESSUS LIVING LAB

Figure : schéma du processus *living-lab* proposé au sein de la communauté LLSA

La plateforme Domus du LIG, sur laquelle nous avons choisi de travailler, avait initialement été conçue comme plateforme d'usage (Jambon, Meillon, Perrot, Caelen, et Harsché, 2009), mais depuis 2012, elle s'est peu à peu transformée en un environnement de type LL par l'intégration d'une méthodologie de co-construction faisant converger : 1) la recherche pluridisciplinaire institutionnelle, 2) les industriels et 3) les partenaires sociaux, également connus sous les termes 3P Framework (pour public-private partnership ou encore PPP, 3P ou P3) (Barlow, Roehrich, et Wright, 2010 ; Cunningham et Cunningham, 2008). Le PPP est un business model qui a connu un essor important à partir des années 2000, particulièrement prisé pour les projets dans le domaine du bien-être et des services de santé. Les hypothèses et les modèles portant sur ces réflexions théoriques ne peuvent alors être validés ou invalidés qu'à travers leurs applications durant le processus de développement de cette technologie qui est l'objet des partenaires industriels. Leur utilisation dans des scénarios co-construits avec les partenaires sociétaux fait également émerger les problèmes éthiques et l'intégration cohérente dans leur contexte sociétal. Par ce biais, les aboutissants techniques, envies, contraintes et besoins spécifiques de chaque partenaire peuvent faire l'objet de discussions dès la phase de conception, faisant participer les futurs utilisateurs au sein d'un encadrement partagé durant la phase de développement, parallèlement associée à des évaluations, l'ensemble étant régi par des boucles agiles centrées sur l'utilisateur final. Nous allons donc décrire en détail les étapes de la méthodologie expérimentale adoptée, construite par la convergence des héritages des anciens travaux explicités section 3.2.2 et les orientations living-lab de la présente section qui augmente les anciennes approches par une dimension itérative co-construite.

### 3.3. Méthodologie expérimentale globale

#### 3.3.1. Co-construction tripartite : le corpus au cœur du système

À l'image de l'approche 3P mentionnée dans la section précédente, la méthodologie de la plateforme living-lab Domus-LIG intègre trois acteurs qui centrent leurs actions et leurs observations sur un utilisateur final de technologie (soit la personne âgée et le robot). Le schéma simplifié Figure 20 illustre ainsi ces 3 groupes acteurs intervenants dans cette démarche de recherche-développement.

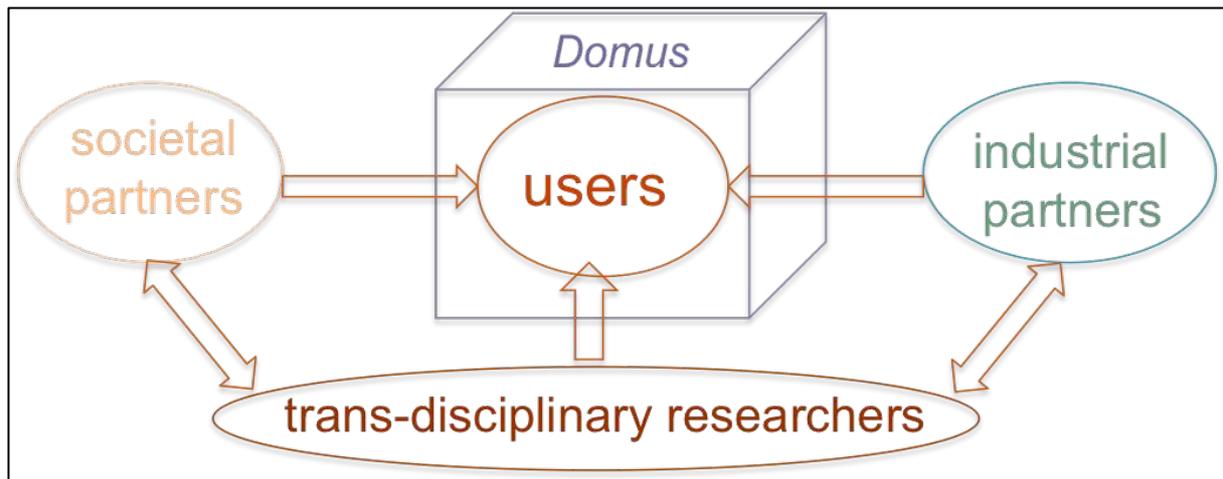


Figure 20 : Adaptation de l'approche 3P à la plateforme Living-Lab Domus-LIG

**Équipe de recherche transdisciplinaire** : plus la richesse disciplinaire sera importante, plus les hypothèses théoriques de départ pourront être étayées par l'éclairage de différentes perspectives et plus les solutions de traitement ainsi que d'analyses de données seront riches. En revanche, elle ne consiste pas en une simple répartition de travail disciplinaire autour du même objet d'étude. Elle implique une acceptation par les spécialistes de chacune des disciplines de passer par une phase de « désapprentissage » temporaire de leurs acquis afin de se confronter aux contraintes des disciplines avec lesquelles ils collaborent. Chacun pourra ainsi acquérir les bases de chacune des disciplines représentées, à proportion plus ou moins importante selon ses besoins. Cette démarche permettra donc de solliciter de nouvelles connaissances qui permettront au chercheur ainsi bousculé de revenir sur sa discipline mère, lui permettant l'acquisition d'une capacité de prise de recul différente qui peut donner lieu à l'élaboration de nouvelles méthodologies non plus pluridisciplinaires, mais véritablement transdisciplinaires.

**Partenaires industriels** : ils permettent de conserver le pont établi entre le milieu de la recherche et le milieu industriel, devenu indispensable dans la recherche s'inscrivant dans le domaine des STICs notamment. Cette orientation s'illustre également par la multiplication des projets investissements d'avenir. L'industriel apporte ainsi la base technologique de travail qui pose de multiples réflexions et interrogations étayant les questions de recherche. Cette collaboration permet également de rythmer les temps de recherche par les contraintes de ses marchés. Pour l'industriel, il s'agit de pouvoir cibler ses futurs utilisateurs ainsi que leurs besoins et surtout de mesurer les effets et les risques que sa technologie peut engendrer par l'expérimentation en laboratoire.

**Partenaires sociétaux** : leur présence permet de faire le pont avec le milieu industriel dès l'étape de conception et de pouvoir permettre les échanges sur les besoins et les contraintes. Mais leur implication ne s'arrête pas à une dimension de pré-enquête ou de conception participative. Ils sont sensibilisés à la motivation scientifique de l'étude pour en être un acteur « expérimentateur » jouant tantôt le rôle de complice de l'expérimentateur avec des fonctions spécifiques durant les expérimentations, tantôt le rôle d'un sujet éclairé (qui connaît le montage de la co-construction contrairement au sujet). Ils peuvent également réguler le suivi de l'expérience afin d'assurer la sécurité des sujets en détectant les éventuels problèmes grâce à leur connaissance privilégiée de l'écologie réelle des sujets dans leur quotidien. Cette participation peut également servir à lever des appréhensions vis-à-vis de la technologie ou de la recherche, ou au contraire à les confirmer si elles prennent une orientation dangereuse. En effet, ils sont les utilisateurs secondaires, mais également l'un des décisionnaires-intégrateurs de la technologie dans la société. Les sections suivantes vont donc dérouler concrètement chaque point développé dans cette conception tripartite dans l'approche pour la mise en place expérimentale de cette thèse.

### 3.3.1.1. Hypothèses pluridisciplinaires pour le recueil de « *smart data* »

Dans la perspective hypothético-inductive de notre démarche, l'initiation de la boucle méthodologique nécessite donc d'avoir à la base de fortes hypothèses construites à travers l'état de l'art pluridisciplinaire des chapitres 1 et 2 (section 1.5.5). Ainsi, l'hypothèse principale de glu socio-affective est fondée sur plusieurs points :

1. Il existerait une « glu socio-affective » inhérente qui sous-tend toute interaction impliquant l'humain. Elle serait construite sur une dimension à part entière, totalement altruiste et indépendante de la dimension de dominance. Cette glu est nécessaire à la construction d'un canal relationnel qui fait l'objet d'un remaniement permanent à travers les processus interactionnels et permet d'établir une continuité communicative (qu'elle soit courte ou longue), entre les interactants composant les sous-systèmes d'un système global d'interaction situé. C'est à travers ce canal que les expressions langagières seraient véhiculées.
2. Une interaction ne peut avoir lieu sans rôle social défini pour ses interactants. Ce rôle est défini dans une relation de réciprocité avec la glu, car :
3. un rôle est nécessaire pour instancier toute communication (nous ne pouvons surgir devant un autre et commencer à parler sans un rôle qui nous prédéfinit) et sans rôle il ne peut donc y avoir établissement de glu ;
4. ce rôle ne peut se définir sans glu : la glu est résultante des processus communicationnels qui se définissent à travers le rôle de chacun. Elle se réajuste et se modifie subtilement et perpétuellement au fil des interactions. Ainsi sans glu, il ne peut y avoir de rôle, car c'est à travers la glu que le rôle s'affirme ou évolue.
5. La prosodie socio-affective est porteuse de valeurs informatives et fonctionnelles, autre que celles contenues dans le support lexical du message, dont il existe des manifestations de « prosodie pure », sous la forme de micro-expressions vocales. Ces primitives engendrent ces mêmes valeurs socio-affectives capables de construire et modifier la glu socio-affective. Ces primitives vocales sont graduellement effectives par rapport à la forme qu'elles prennent, en partant d'un niveau de contrôle prosodique moindre à un niveau

plus élevé et donc plus linguistique. Le postulat sous-jacent étant que le langage complexe a un grand pouvoir de manipulation sur l'autre et que c'est le niveau de contrôle d'intention communicationnelle qui augmente le formalisme langagier d'une production. Ces différents niveaux sont présentés dans la Figure 21.

Ces hypothèses génèrent donc des contraintes épistémologiques qui font l'objet de choix méthodologiques d'un cas d'observation contrastif. En effet, afin d'observer les manifestations et les facteurs d'un processus dynamique et évolutif, il est nécessaire d'avoir une condition qui puisse rendre observable ce mécanisme de remaniement de glu. Une possibilité pour cela est de voir si les primitives supposées avoir un effet fonctionnel sur ce mécanisme ont des effets dans une condition où ce processus est endommagé.

Le cas choisi est donc celui de l'isolement relationnel des personnes âgées, une situation qui est socialement observée, fréquente et problématique, et qui existe dans un contexte réel. Ainsi si les primitives ont véritablement une valeur fonctionnelle sur la glu, elles devraient potentiellement avoir un effet réparateur sur le processus communicationnel altéré. Ces effets devraient donc être observables à travers des paramètres langagiers de la personne âgée observée puisque la glu est supposée sous-tendre toute interaction. De fait, une personne âgée souffrant d'un isolement relationnel soumise à des stimuli de primitives est supposée produire des traces interactionnelles comportant des points de rupture et des invariants observables influençant les directions de la dynamique communicationnelle. Les dimensions mises en évidence caractériseraient alors cette glu socio-affective dont on pourrait définir des variables potentiellement modélisables pour un système de traitement automatique de parole qui augmenterait l'autonomie du robot par une intelligence socio-affective.

Afin de tester ces primitives, il nous faut donc un instrument de mesure permettant de contrôler leur production, qui dans l'absolu serait idéalement produite par un humain. Or ne pouvant contrôler les productions d'une interaction sans en perdre la spontanéité, nous faisons ici le choix d'un robot pour lequel ce contrôle est possible et reproductible. De plus, cet outil est suffisamment nouveau pour ne pas être soumis à des processus d'interactions formatés ou formalisés, ce qui laisse la possibilité d'explorer la manière dont les sujets vont interagir spontanément avec cet objet. En partant ainsi d'un niveau quasiment vierge de glu de fait de la nature de cet instrument, nous réduisons au maximum l'anthropomorphisme du robot à sa seule dimension vocale. Cela permet de contourner les biais de la perception anthropomorphique issue de la projection empathique, qui serait d'autant plus importante et même dérangeante si les indices de patternicité étaient proches de celui de l'humain, à l'image des effets de l'*uncanny valley*.

Ainsi le format de l'étude implique :

- Un robot produisant des primitives porteuses de prosodie socio-affective présentées de manière graduelle selon leur niveau de force « gluante » (ces sons sont issus de productions humaines, dont l'esthétique vocale est modifiée et uniformisée – cohérente avec l'apparence physique du robot).
- Ce robot est doté d'un rôle de majordome d'un habitat intelligent domotique à commande vocale, qui ne remplace pas un rôle humain, mais qui est propre au robot. Ceci est important d'un point de vue éthique, mais aussi pour lui associer une identité propre qui pourrait définir son utilité et son existence dans le tissu social de l'humain afin qu'il perdure.
- Des sujets âgés en isolement relationnel vont donc interagir avec ce robot en lui fournissant des commandes vocales domotiques, le robot simulant la mise en route de l'actionneur domotique. Mais surtout, à chaque itération, ce robot va répondre graduellement par des feedbacks vocaux constitués des primitives

définies dans le cadre de l'hypothèse de glu socio-affective comme présentées en Figure 21 :

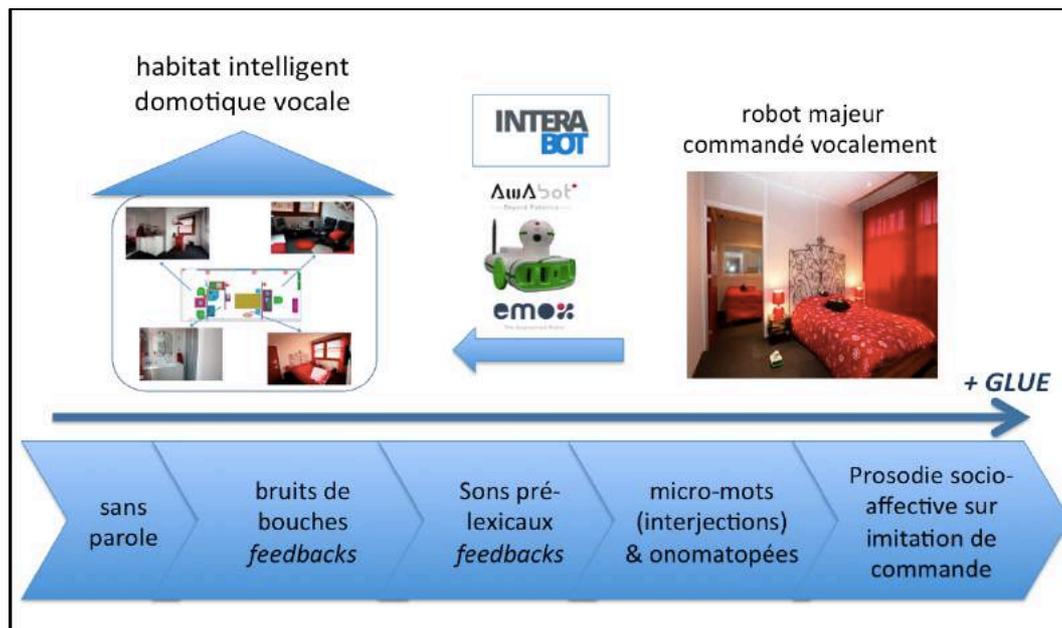


Figure 21 : Hypothèse de primitives vocales graduelle pour l'établissement d'une glu socio-affective

L'état de l'art tentait donc d'illustrer la dimension transdisciplinaire qui est à l'image de l'ensemble de l'équipe de recherche qui a contribué à cette thèse. Cette dernière ne sera pas présentée en détail, car elle est elle-même évolutive et le nombre de personnes impliquées est important. En revanche, en suivant la logique de la méthodologie adoptée, les sections suivantes introduiront les partenaires impliqués dans cette étude et la dimension réflexive qu'ils ont apportée à cette thèse par leur contribution à travers cette méthodologie.

### 3.3.1.2. Partenaire industriel : l'entreprise Awabot

La société Awabot, est une entreprise lyonnaise créée par Bruno Bonnell (Robopolis, Infogrames, Infonie) en 2010. Il s'agit du partenaire industriel avec lequel nous avons commencé une collaboration en 2012. Initialement orientée vers la conception, le développement et la production de robots à destination du grand public et du monde de l'éducation, elle présente actuellement une activité d'expertise-conseil et services robotiques ainsi qu'une activité de revente des robots de téléprésence Beam venus des États-Unis.

L'intention initiale d'Awabot était et reste certainement toujours de rendre accessibles au plus grand nombre les objets connectés et intelligents de même que de faciliter leurs interactions avec les humains. En 2012, l'entreprise a développé une première gamme de robots « compagnons » sous le nom d'Emox qui est l'instrument à travers lequel nous avons réalisé cette étude. Finalement son implication avec l'équipe de recherche est relativement longitudinale, puisque notre premier contact s'est effectué en amont de cette thèse. La collaboration avait notamment consisté en une étude exploratoire sur le langage et l'impression générale du robot (Sasa, Aubergé, Franck, Guillaume, et Moujtahid, 2012) puis axée sur une étude de faisabilité de protocole dans le cadre de l'utilisation de Domus (Sasa, 2013b).



Figure 22 : Entreprise robotique Awabot (B. Bonnell, Villeurbanne – France)

L'entreprise recherchait alors une identité vocale et communicative à donner à ce robot. Sa philosophie, tout comme l'approche de design de R2D2, suppose qu'un robot ne devient pas moins communicatif par un design qui s'écarte d'une apparence non anthropomorphique ou d'un langage naturel et complexe. Le but initial de cet industriel dans ces années-là était d'intégrer Emox dans les foyers comme robot compagnon en 2015, ce qui avait permis de poursuivre une collaboration dans le cadre d'un projet investissements d'avenir. Dans ce projet porté par Awabot, trois autres partenaires ont été sollicités, dont deux laboratoires de recherche, le LIG et le LIRIS ainsi qu'un second partenaire industriel, Voxler. Le LIG s'est engagé sur l'étude des aspects langagiers et communicationnels, dans le cadre du Traitement Automatique de Parole et du dialogue personne-machine, en apportant une expertise de méthodologie expérimentale centrée sur l'observation du comportement humain, au sein de l'équipe Getalp. Les travaux du laboratoire LIRIS apportent par ailleurs une expertise poussée en *machine learning* dans le cadre du traitement visuel. Il s'est en particulier attelé aux aspects de la gestualité pour lesquels le LIG a collaboré directement dans la définition d'ontologies en regard d'une récolte de gestes induits spontanés implicites comparés à des productions conscientes et explicites (Girard-Rivier et coll., 2016 ; Guillaume et coll., 2015). Voxler quant à lui est spécialisé dans les applications et jeux notamment de chants pour rajouter la dimension ludique et intégrative dans le quotidien des personnes. Ce suivi longitudinal a montré que la co-construction permettait de faire aussi bien l'évolution du design du robot, d'apporter des boucles de prototypages rapides qui permettaient d'évaluer en condition réelle des briques technologiques parallèlement développées et de voir les conséquences que l'industriel prenait par rapport à son positionnement sur le marché.

### 3.3.1.3. Éco-système humain : les aidants des personnes âgées

L'identification de potentiels partenaires sociétaux est une première étape chronophage de ce travail de recherche. En effet, l'étude tente d'aborder un public complexe : des personnes âgées, fragilisées, mais dont la fragilisation est fortement liée à leur état d'isolement. Or l'isolement ne concerne pas uniquement une condition contextuelle, elle doit faire l'objet d'un sentiment perçu avec les effets altérants potentiels sur la dynamique communicationnelle et relationnelle de la personne. Pour le gérontologue, le profil de sujet recherché relève déjà d'une complexité y compris pour des disciplines comme la sociologie, car il reste difficile de : 1) formaliser la détection du public âgé souffrant d'isolement (d'un point de vue relationnel, mais parfois aussi contextuel) et 2) trouver et accéder auprès de ces personnes, notamment pour pouvoir remédier à cette situation. En effet, les actes de prévention ou autres démarches que les acteurs sociétaux aimeraient pouvoir appliquer sont nombreux. En revanche, l'un de leur premier frein est de ne pas pouvoir repérer les personnes isolées dans un besoin non exprimé mais de fait. En effet, s'il s'agit d'aborder des personnes demandeuses, se manifestant auprès des services alloués, les moyens sont nombreux. Il en va de même pour celles qui sont entourées pour être aidées sur le strict nécessaire (c.-à-d. réaliser les actes de vie quotidienne les plus indispensables). Cependant, nous perdons facilement la trace des cas les plus problématiques, silencieux et difficilement repérables. Il ne s'agit dès lors pas de généraliser que toute personne vivant seule et qui ne réclame pas d'aide revêt un caractère d'isolement problématique, tout est là encore une question de perception et de conséquences de son état isolement.

En revanche, dans notre entourage, un nombre non négligeable de personnes vivent dans des situations critiques extrêmes, et cela sans que nous nous en rendions compte. Il ne s'agit peut-être que d'un premier constat empirique remarqué dans cette phase exploratoire,

notamment à travers les échanges avec les partenaires sociaux, mais la situation n'en est pas moins alarmante. En effet, il est classique de voir une personne âgée avoir une mauvaise expérience à l'extérieur de son domicile (car à partir d'un certain âge elles deviennent aussi des cibles plus importantes de formes de petites délinquances. Ceci peut avoir d'importance conséquence psychologique (et parfois physique), à l'image de leur première chute. S'ensuit alors un effet progressif d'isolement contextuel ou une réduction quantitative de liens sociaux tendant vers un isolement relationnel qui, à un certain stade, aura un effet délétère, car la souffrance perceptive de cet isolement augmente silencieusement et rapidement. Le plus vicieux est certainement que cette situation apparaît de manière tellement anodine que parfois, les personnes impliquées ne se rendent pas compte ou s'aveuglent devant leur situation. Elles peuvent parfois même faire l'objet d'un déni puisque dans leurs dires, aucun problème apparent ne semble les troubler. Or, généralement, cette souffrance ressentie dérange les personnes. Elles sont perturbées à l'idée de devoir demander de l'aide. Ce blocage peut survenir par peur de déranger, ou les amène à se plaindre fortement. Quelque soit le comportement, cela les enferme dans un contexte de vie où ils n'attendent plus que « leur heure vienne » (des paroles qui ont été entendues beaucoup trop de fois durant cette étude). Leurs tentatives à renouer échouent alors souvent, à cause de malaises communicationnels.

Ici, une première exploration donnera les lieux d'intervention des partenaires sociaux par lesquels les sujets de l'étude ont été recherchés. Les informations présentées ci-dessous tentent donc d'identifier ces potentiels partenaires sociétaux, mais visent également à clarifier et justifier certains choix méthodologiques pour le recrutement.

**EHPAD et structures médicalisées.** Il s'agit certainement des premiers types de structures qui intéressent les chercheurs visant le public des personnes âgées, mais elles sont à considérer avec parcimonie, car chaque structure est très spécifique et personnalisée en fonction de la philosophie de la direction de l'établissement. Elles sont tantôt très médicalisées avec différents types d'organisation (p. ex. tout un service interne, sollicitation de professionnels médicaux et/ou sociomédicaux externes, en plus ou non de ses propres services, degrés de liberté de visite/sorties variables), de vie communautaire (p. ex. le partage entre les résidents sont soit favorisés, soit imposés ou même non obligatoire, par plus ou moins forte sollicitation à participer ; intégration de la dimension intergénérationnelle, présence familiale etc.). Ces structures sont d'ailleurs très sollicitées en plus d'une gestion interne très difficile, mais certaines restent très ouvertes et intéressées à l'expérimentation. Si l'étude a obligatoirement une attache aux sciences humaines, il sera très certainement nécessaire que les caractéristiques de chaque structure soient étudiées au préalable pour que les partenariats établis soient en adéquation avec les hypothèses et la démarche méthodologique engendrée. Ces institutions se prêtent donc très bien aux études, mais son écologie intrinsèque aura un impact sur les comportements des personnes à observer.

**Les CCAS / clubs / maison des habitants.** Les problématiques et dynamiques scientifiques intéressent en général ces organisations, ce qui est très favorable pour le chercheur. Si l'on tente de trouver avant tout des personnes sur critère d'âge dont les caractéristiques personnelles et socioculturelles sont variables, il peut être intéressant de passer par ces structures. Dans le cadre de cette thèse, elles étaient néanmoins difficilement exploitables, car : 1) les bénéficiaires sont des personnes qui sont dans une démarche relationnelle active ; 2) le suivi par les aidants est trop global à l'ensemble de la communauté, et pas suffisamment individualisé (même si en général les responsables et animateurs connaissent très bien les personnes auprès desquelles elles interviennent). Ainsi les contraintes de cadrage ne pouvaient être appliquées pour la réalisation de la présente recherche.

**Les aidants familiaux :** il en existe de nombreux réseaux. Certains ont d'ailleurs parfois directement contacté l'équipe de recherche pour faire participer un proche. C'est un aspect rassurant pour la recherche puisqu'elle se positionne dans un contexte de demande sociétale, mais la relation de l'aidant familial avec son parent était souvent bien trop complexe à gérer. Un protocole ainsi qu'une méthodologie spécifique ont besoin d'être développés, pour impliquer au mieux ces aidants, car leur présence sera dans le futur indispensable pour construire la dimension éthique de co-construction comme abordée dans notre étude. En revanche, elle induit des effets tout autres que celui des aidants externes. Le plus gros frein observé est certainement à cause du changement de perspective du rôle social de l'enfant vis-à-vis de son parent, dont la tendance s'inverse, générant ainsi des biais quant à la nature relationnelle avec la personne âgée. De plus, l'aidant avait un fort impact sur la motivation de participation de la personne, ce qui est un point important de méthodologie. En effet, l'aidant avait tendance à obliger son parent à participer, et ce dernier répondait par des attitudes de rébellion (c.-à-d. refus de participer parce que c'est imposé), de dépendance (c.-à-d. où la personne n'osait plus prendre aucune décision en présence de son parent, avec parfois des propos contradictoires constatés en fonction de la présence ou non du parent dans la même pièce), ou encore de détachement complet (c.-à-d. « c'est toi qui as pris la décision, mais ça ne me concerne pas »). Mais d'autre part, l'aidant familial avait également tendance à supplanter la décision de participation que le parent avait prise, mais a posteriori du rendez-vous de recrutement (p. ex. crainte de responsabilité, de la nature de l'expérience, du besoin de contrôle de l'emploi du temps de son parent etc.)

**Le CPA :** Le Centre de Prévention des Alpes est un catalyseur de contacts variés et sa dimension préventive a un véritable intérêt dans les démarches co-constructives de la recherche. Il est dommage que dans le cadre de cette thèse, un partenariat n'ait pu aboutir. Le premier frein étant l'indisponibilité des responsables de cette structure.

**Les associations développées autour des personnes âgées isolées :** elles sont certainement très nombreuses. En revanche, celles avec lesquelles il a été possible d'échanger sont notamment : les Petits Frères des Pauvres (isolement plus ou moins important idéal, mais très peu de disponibilité), DIGI (qui s'occupe des collocations intergénérationnelles entre jeunes étudiants et personnes âgées) et EPPI (Écoute Partage Personnes Isolées – qui traite des cas d'isolement graves). Les intervenants directs de ces associations connaissent très bien les problèmes relevés dans les conditions et les motivations sociétales de notre étude. Ces associations ont pu nous confirmer l'existence de situations difficiles d'isolement. En revanche, pour ce public, l'engagement nécessaire pour la mise en place d'une étude expérimentale est bien plus complexe d'un point de vue pratique. Cela renforce d'autant plus le défi de telles approches, car l'état d'isolement des personnes âgées est celui qui se rapproche le plus de nos idéaux et besoins méthodologiques. Il a en revanche été impossible de l'envisager dans le cadre de cette thèse, du fait de la lourdeur protocolaire d'un déplacement sur un plateau expérimental. Un nombre alarmant et conséquent de ces personnes déclaraient n'être plus sorties de leur domicile depuis plus de deux ans. En effet, comme les services et les visites à domicile sont nombreux de plus en plus à se développer dans notre contexte social actuel, cette situation, pourtant extrêmement inquiétante, ne semble pas rare.

**Les commerçants, le voisinage et les contacts de proximité :** ils ne sont pas à négliger, car malgré l'individualisme déclaré des sociétés occidentales, le cercle restreint de notre écologie quotidienne conserve une attention particulière pour les personnes âgées. De fait, sans forcément passer par des procédures de porte-à-porte ou de veille informationnelle auprès des mairies, il est également possible d'envisager une approche à travers des zones de vies ou d'activités, ponctuées de changements, afin de retrouver les personnes isolées. En

effet, le passé relationnel plus ou moins riche d'une personne marque les esprits des acteurs de cet environnement de vie. Ainsi sans forcément très bien connaître la personne, les échanges à travers la boulangerie, le salon coiffure, le parc du coin ou l'épicerie de quartier permettent de cibler relativement efficacement les changements relationnels et les habitudes d'une personne, permettant ainsi de repérer des sujets potentiels. Le protocole de l'étude prévoyait la présence d'un partenaire social accompagnant, dans le cadre de la co-construction. Cependant, les personnes peuvent également être isolées sous d'autres formes : en couple, entre voisins, sans intervenants familiaux ni professionnels (notamment s'ils sont en bonne santé). Tous ces profils n'étaient donc pas négligeables et ont également été sollicités dans l'étude, car leurs caractéristiques relationnelles restaient en adéquation avec le profil du sujet idéal attendu.

**Les structures de services à la personne** : même remarque que pour les structures de type EHPAD, elles relèvent de la spécificité de chaque établissement (p. ex. le statut associatif ou la dimension lucrative de l'activité change le rapport du bénéficiaire et de l'aidant). En revanche, elles ont particulièrement été sensibles à la démarche de l'étude. La dimension collaborative est ainsi la plus importante avec ces acteurs, car leur expertise de terrain est véritablement mise à profit dans le cadre de la co-construction. En revanche, elle nécessite que la problématique scientifique soit bien intégrée et qu'elle motive ces acteurs. Par ailleurs, leur sollicitation était un premier critère efficace de filtrage pour cibler les profils des sujets, car la présence d'aidants externes est le point d'entrée d'une fragilisation potentiellement relationnel, car elle contourne les aidants familiaux. Si d'une part les personnes ne souffrent pas d'une dépendance physique ou cognitive importante et que d'autre part, elles n'ont pas de cercle familial aidant, elles ont potentiellement plus de chance d'être dans la phase transitoire d'isolement qui nous intéresse.

**Les foyers-logements et EPA** : Il s'agit d'une résidence où chaque habitant a son propre appartement entier et non seulement une chambre. Elle donne cependant la possibilité d'avoir sur place des repas ou des activités collectives proposées, permettant ainsi d'éviter les déplacements contraignants dans des lieux divers d'activités socioculturelles ou sportives. Ces activités ne sont cependant pas obligatoires et laissent ainsi une liberté, permettant aux résidents de conserver une totale indépendance, tout en leur laissant la possibilité d'accéder à des services facilitants et d'être plus ou moins suivi par son environnement proche. La spécificité de l'établissement est là encore très disparate, mais cette structure est présentée à part, car elle présente plusieurs intérêts dans le cadre de l'étude : 1) elle représente une structure transitoire, entre le domicile privé et l'établissement médicalisé. Il s'agit d'une solution qui est souvent choisie, lorsque les personnes commencent à ressentir une fragilisation (physique ou morale) qui s'accompagne alors d'une phase de glissement vers un isolement ; 2) elle est également, dans la plupart des cas, ouverte aux personnes externes au foyer, habitant encore dans leur domicile privé, afin qu'elles puissent participer aux activités collectives. En effet, ces personnes externes sont dans une dynamique d'isolement, souvent en attente de placement dans ces établissements. Elles sont d'ailleurs presque plus pertinentes dans le cadre de l'étude, par rapport aux résidents de longue date, qui sont déjà dans une dynamique de reconstruction relationnelle. Ainsi, le fait d'avoir des listes d'attente permettant d'accéder à des futurs résidents, ou encore le turn-over rapide de nouveaux résidents, favorisent le ciblage des profils d'isolement attendus. De plus, l'expertise de l'équipe intervenante du foyer permet de définir précisément les caractéristiques des personnes âgées. En effet, la taille plus restreinte du foyer-logement et la proximité des intervenants avec les résidents ont été plus propices à notre démarche de recrutement que dans les établissements de type EHPAD, qui sont d'ailleurs de plus grande taille et traitent de cas de fragilisations physiques et cognitives trop importantes pour les contraintes de l'étude.

Les partenaires sociaux avec lesquels nous avons donc préférentiellement travaillé sont : l'entreprise de services à la personne Bien à la Maison, l'association de services à la personne AAPPUI ainsi que le Foyer-Logement Roger Meffreys de la commune de Gières, établis non loin du lieu d'expérimentation. L'implication de ces différentes structures a été une exploration très instructive du point de vue épistémologique et est un élément fondamental, irréductible, de la conduite de notre expérimentation.



Figure 23 : Identification des partenaires sociétaux – à gauche Bien à la Maison – au centre l'association AAPPUI – à droite le foyer-logement Roger Meffreys de la commune de Gières

Finalement, au regard de cette synthèse, les trois partenaires sociétaux choisis étaient, de manière générale, ceux qui permettaient de cibler au mieux nos critères de recherche sur la dimension d'isolement relationnel. Leur participation a ainsi été active et totalement bénévole avec une compréhension et une motivation importante de leur part. Pour le chercheur, ces deux aspects ne sont pas à négliger dans le choix de ses partenaires, car le travail collaboratif est chronophage et lourd, aussi bien pour le partenaire que pour l'expérimentateur. Les protocoles représentent en effet un travail longitudinal en plusieurs étapes, par l'établissement d'une relation de confiance, un point également valable pour le partenaire industriel.

### 3.3.2. Plateforme Domus-LIG

Nous venons de voir l'orientation très globale de la méthodologie, en définissant les principaux acteurs de leur distribution tripartite dans une approche de co-construction. En revanche, nous n'avons pas encore abordé concrètement les démarches de cette méthodologie développée au sein de la plateforme Domus-LIG, ni l'environnement expérimental qui permettent de mettre en œuvre les protocoles expérimentaux. Ces aspects feront donc l'objet des sections suivantes.

#### 3.3.2.1. Environnement : appartement intelligent

Domus, en tant que lieu d'expérimentation, est une plateforme de 200m<sup>2</sup> qui comporte :

- 1) un bureau dont l'agencement peut être travaillé, qui n'est pas anéchoïque, mais reste acoustiquement isolé pour effectuer des enregistrements audio de qualité ;
- 2) un plateau qui est aménageable, dans diverses mises en scène (salle de réunion, scène théâtralisée, café/espace d'accueil, aménagement pour ateliers de conception ou de créativité, salle de communication etc.), comportant des caméras, des micros ainsi que des éclairages modulables et contrôlables ;
- 3) un appartement de 40m<sup>2</sup> qui dispose de trois pièces, chacune dotée de caméras et de micros (cuisine, chambre, salon/bureau – plus une salle d'eau non équipée de caméras), ainsi que de multiples capteurs et actionneurs. Ces éléments sont contrôlés par de nombreux protocoles dont KNX, UPnP, RFID, X2D, et ARTNET ;
- 4) une régie reliée à l'ensemble des espaces précédents permettant d'acquérir des données variées, et de contrôler à distance différents composants de la plateforme. La Figure 24 illustre la disposition de ces différents éléments et donne un aperçu visuel des pièces de l'appartement.

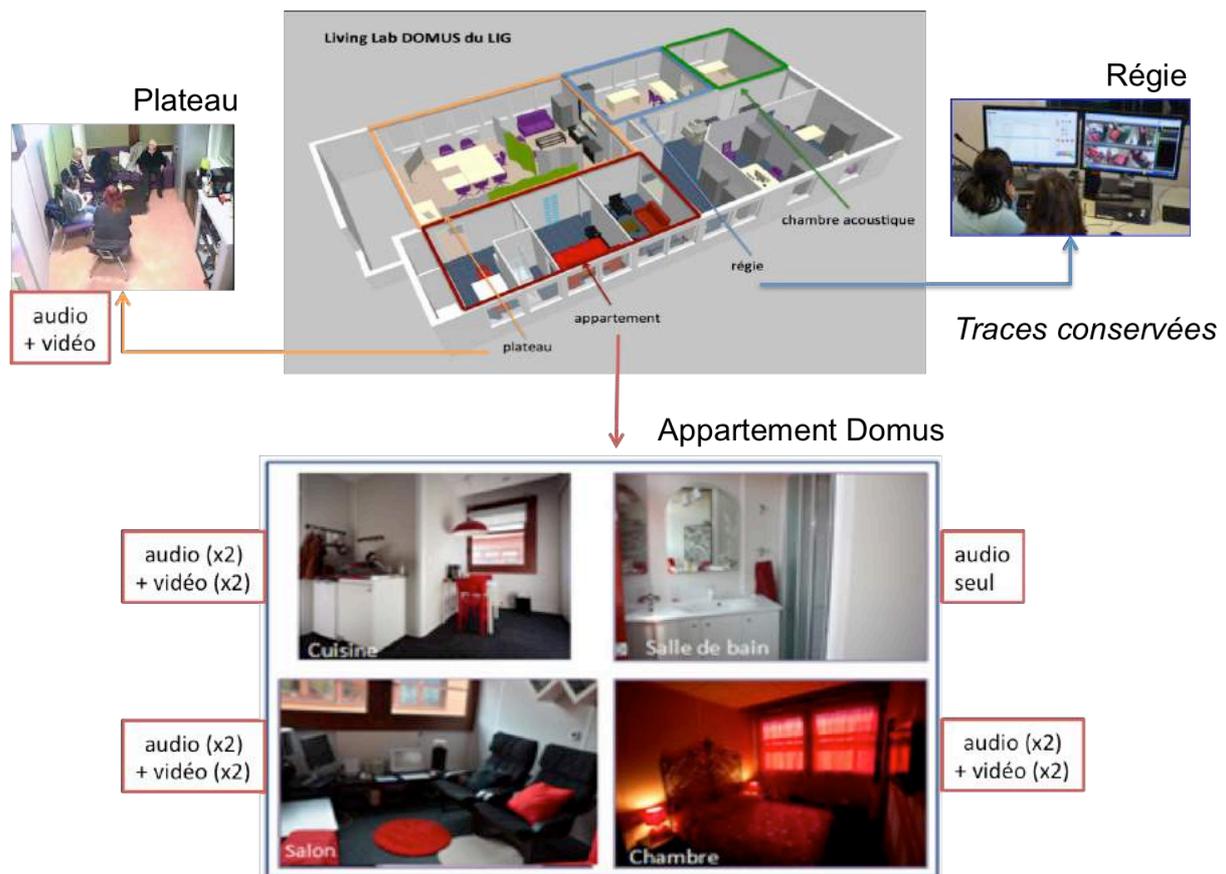


Figure 24 : Plateforme expérimentale Domus du LIG et son appartement intelligent

La phase expérimentale principale de notre étude est réalisée dans l'appartement intelligent Domus. Cet environnement sollicite un ensemble technique conséquent, permettant de récolter des données multisources, dont les procédés sont décrits par (Lecouteux, Vacher, et Portet, 2011). L'appartement est ainsi doté de six caméras AXIS, de type surveillance, qui sont positionnées au plafond des pièces. Elles enregistrent des fichiers vidéo au format.asf. Domus est également connecté également sept canaux audio sur récepteur à fréquence radio. L'acquisition est effectuée à partir de microphones omnidirectionnels sans fil (Sennheiser eW500) et distants pour la récolte des sons de l'environnement. À cela se rajoute un huitième canal audio, qui peut être relié à un micro indépendant, porté par le sujet (micro-cravate ou micro-casque<sup>79</sup>). Cette mise en place technique et les premières expériences contextualisées ont été réalisées à travers des projets externes à cette thèse, comme le projet ANR SWEET-HOME décrit par (Portet, Vacher, Golanski, Roux, et Meillon, 2013) avec des adaptations effectuées et supervisées par N. Bonnefond<sup>80</sup> qui a permis de faire fonctionner l'ensemble du système le couplant au middleware OpenHab. Cela rend ainsi la plateforme contrôlable à distance par smartphone ou tablette par exemple. Les canaux audio sont acquis à travers le système fonctionnant en temps réel, AuditHIS, développé par (Vacher et coll., 2011 ; Vacher, Fleury, Portet, Serignat, et Noury, 2009) et qui fonctionne sur un environnement GNU/Linux. Ce système présente cinq modules (acquisition/prétraitement ; détections d'événements sonores ; discrimination parole /son ; classifieur GMM/HMM ; extractions de mots-clés et

<sup>79</sup> Le modèle de micro-casque utilisé dans notre étude est un SENNHEISER HSP4 -ew-3 statique cardioïde, jack 3.5.

<sup>80</sup> L'exploitation du middleware openHAB sur la plateforme Domus-LIG et son utilisation en expérimentation sciences humaines a fait l'objet d'une présentation par N. Bonnefond s'intitulant *Ruling the experiment : how openHAB really empowers Living Lab platforms*, à la conférence EclipseCon 2015. Vidéo de présentation accessible sur : <https://www.youtube.com/watch?v=pKI9lBiNXgw>

génération de messages), dont nous n'avons exploité que le module d'acquisition, sans passer par les chaînes de reconnaissance et traitement automatique de la parole.

### 3.3.2.2. Au-delà d'un environnement : une méthodologie itérative et intégrative de l'écologie des comportements

Dans ce concept de Living-Lab, la plateforme Domus-LIG n'est pas uniquement un lieu d'expérimentation type laboratoire d'usages mais avant tout une méthodologie de capture et de modélisation de l'écologie des comportements humains. Elle s'appuie sur les connaissances des approches *Living-Labs* dans une perspective plus expérimentale. Elle permet alors d'augmenter les approches hypothético-inductives et empirico-déductives qui ont été décrites section 3.2.2 et section 3.2.3. Cette méthodologie introduit ainsi :

- une possibilité d'augmenter le réalisme et d'introduire une dynamique interactionnelle complexe par un environnement expérimental *in vitro* dont l'apparence reproduit le réalisme et les conditions d'un contexte *in vivo* (ici l'apparence d'un vrai appartement). L'association des technologies domotiques et robotiques hybridées sur cette plateforme permet ainsi de multiplier les possibilités de scénarios expérimentaux, en recréant de multiples situations écologiques. Le corpus est ainsi au centre de la démarche ;
- la définition d'un cadre contrôlé pour faciliter les réflexions éthiques inhérentes au processus scientifique de développement/déploiement technologique qui intègre des partenaires industriels et sociétaux avec une équipe de recherche pluridisciplinaire, par des méthodes de co-construction selon l'approche 3P définie section 3.3.1. Cette dimension demande d'ailleurs une maturation méthodologique afin de réussir à intégrer la société civile avec une dimension participative dans le cadre d'une démarche éthique qui évalue les risques des effets de la technologie modélisée ;
- le traitement par boucles itératives agiles et rapides, permettant de développer des briques technologiques en modélisant des sous-ensembles des processus décrits rapidement par le corpus, dont l'étiquetage et les valeurs associées évoluent continuellement. Cela introduit une approche par essai/échec permettant un développement parallèle à l'évaluation des modèles dans des contextes expérimentaux reproductibles. Cette co-fabrication parallèlement évaluée permet notamment l'exploitation de la richesse multiplateforme du LIG.

Les utilisateurs finaux, et donc l'humain, sont au cœur de ce système collaboratif. Ils vont produire la richesse du corpus recueilli qui sera le matériel centralisant tous les échanges de ce système méthodologique. L'utilisateur, soit le sujet de l'expérience, n'aura pas un rôle participatif dans le sens où l'objectif n'est pas de lui demander directement son avis par leur usage d'une technologie, comme dans les approches des laboratoires d'usage ou dans des méthodes de conception participative. Il est l'objet de l'observation, dont la naturalité et la spontanéité des comportements interactionnels sont induites et recueillies, formant des données exploitables pour un apprentissage machine, à modéliser dans un système de dialogue HRI par exemple. Tout l'enjeu de la collaboration des différents acteurs est donc de recréer une écologie interactionnelle la plus pertinente possible et en adéquation avec les hypothèses de recherche. Elle doit respecter la perspective finale de la technologie attendue par l'industriel tout intégrant les contraintes de la situation socioculturelle réelle par la confrontation des problèmes de terrain mis en évidence par les acteurs sociétaux. Cela implique que le sujet reste complètement aveugle et détaché des discussions pour éviter que ses comportements ne soient biaisés par leur conscientisation.

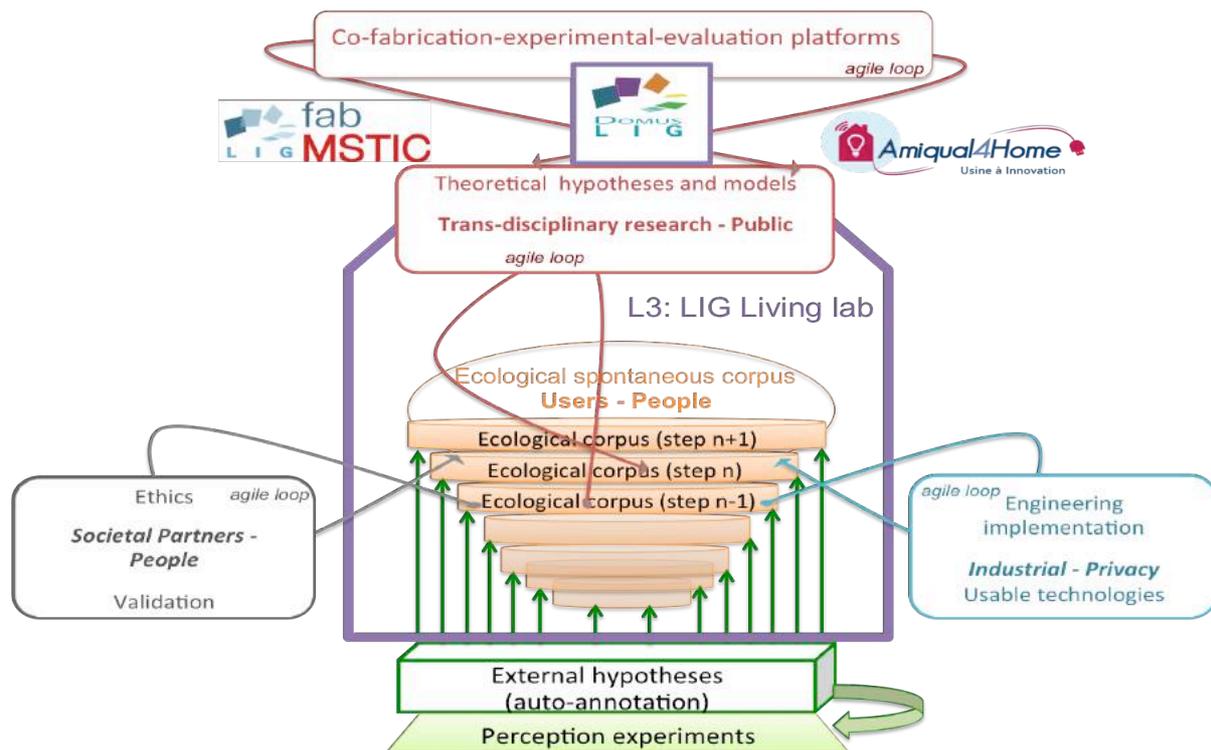


Figure 25 : Schéma de la méthodologie Living-Lab expérimental de Domus-LIG (Aubergé, Sasa, Maisonnasse, 2015)

La méthodologie utilisera donc de nombreux prétextes et éléments détournant l'attention du sujet. Elle utilise des scénarii d'expérimentation précis. Enfin, l'itération par les boucles agiles successives permet de rajouter une dimension évaluative qui n'est pas en bout de chaîne de la production technologique. Elle devient elle-même itérative et elle est réalisée en parallèle de la conception et du développement technologique, par réinjection de prototypes dans les mêmes conditions expérimentales. Ce prototype est établi à partir d'un stade précédent du corpus, dont les conditions de recueil et d'observation sont reproductibles. La plateforme Domus-LIG travaille ainsi en collaboration avec deux autres plateformes :

**Le Fablab MSTIC<sup>81</sup> :** le Fablab (laboratoire de fabrication) est un concept né au MIT en 2006 sous l'impulsion de Neil Gershenfeld (Mikhak et coll., 2002). Au LIG, il existe ainsi une plateforme, qui permet d'intégrer la dimension de prototypage sur des éléments moins coûteux, remaniables et rapides, qui s'actualisent au fur et à mesure des itérations des boucles agiles de la méthodologie. Pour les partenaires, c'est également un très grand avantage que les chercheurs accèdent à ce prototypage. Pour l'industriel, en fonction de la maturité de sa technologie, le prototype lui permettra de tester certaines briques et d'en faire des évaluations, pour aider sa prise de décision quant au développement réel dans sa chaîne de fabrication. En effet, s'il devait gérer entièrement ce développement qui serait plus contraint, ce prototypage serait moins rapide, potentiellement très coûteux et pour un simple test sans garantie de production derrière. L'industriel peut aussi se défaire de certaines contraintes techniques pouvant avoir des répercussions sur les expérimentations (p. ex. dans notre étude, le Fablab a notamment été sollicité pour recréer des batteries compatibles avec le robot, ce qui a été utile puisque l'industriel avait arrêté la production du modèle utilisé et ne pouvait donc plus en fournir). Pour le partenaire social, c'est une manière de co-concevoir et voir concrètement des outils sans avoir à attendre le produit. Il peut également mettre en évidence les points à corriger au regard d'un usage réel qu'il expérimente à travers cette méthodologie.

<sup>81</sup> Site du Fablab MSTIC, accès via : <http://fabmstic.liglab.fr> sous la responsabilité J. Maisonnasse

L'équipex **Amiqua4Home**<sup>82</sup> : cette structure comporte deux aspects. La première consiste en l'accompagnement au développement d'équipement expérimental pour les chercheurs (notamment à travers son atelier Fablab). La seconde est la possibilité d'effectuer des activités de *monitoring* comme le permet la plateforme Domus-LIG, mais dans un habitat plus complet construit sur structure d'un ancien appartement duplex. Loin d'être en concurrence, son expertise et son environnement sont complémentaires à Domus-LIG, qui lui est dédié à des études de court terme. Amiqua4Home lui permet d'effectuer des études à plus long terme, et pour un public non fragilisé, car il est construit sur deux étages avec des escaliers. Dans le cadre de cette thèse, il a notamment fait l'objet de collaboration pour créer des démonstrateurs pour des salons. Cette plateforme peut également co-concevoir des « mini-kits », des boîtes à outils pratiques qui peuvent permettre techniquement l'exportation du laboratoire au-delà de ses murs, pour intégrer l'environnement réel des sujets dans des formats d'expérimentation *real-life* mais contrôlés, permettant ainsi une extension du laboratoire. Cela permettrait donc d'envisager des perspectives d'évolution pour la méthodologie Domus-LIG.

### 3.3.3. *Approches sous-jacentes aux boucles agiles*

La schématisation de la méthodologie (cf. Figure 25) montrait que cette dernière s'appuie sur un corpus évolutif et extensible à différents temps, résultant de boucles de traitement agiles. Une boucle illustre donc une étape de modifications du corpus dépendant de phases intrinsèques à la démarche expérimentale. Cette approche suit ainsi de manière transversale ces phases définies dans les anciennes approches héritées (cf. section 3.2.2). Une boucle d'itération peut donc être plus ou moins rapide en fonction des contraintes de l'étape de traitement du corpus. Lorsque ce traitement est plus long, cette phase s'étend sur plusieurs boucles, pour permettre des prototypages et modélisations intermédiaires des briques technologiques qui en découlent.

Une autre façon de voir cette méthodologie globale est de décomposer ce qui est représenté par la ligne violette de la Figure 25, à travers les approches scientifiques, qui sont multiples et temporellement variables selon l'étape expérimentale concernée. La Figure 26 ci-dessous résume ainsi globalement ces étapes (cf. flèches cadrées numérotées) et les approches méthodologiques concernées à ces étapes (cf. les cadres fléchés tendent du blanc au rouge en fonction la force des hypothèses théoriques qui sous-tendent l'approche). Les codes couleur sont en adéquation avec celles de la Figure 25.

Ce schéma dont la version complète se trouve en Annexe 2, met en évidence cinq étapes expérimentales (cf. numéros de 1 à 5) qui illustrent les approches adoptées. Ainsi la méthodologie débute par une approche hypothético-déductive qui est sous-jacente à la construction ainsi qu'à la réalisation d'un scénario établi sur un prétexte (étape [1]). Cette approche est transversale en fonction du *timing* de « levée de voile » du prétexte (étape [2]). Elle permet le recueil de données spontanées et écologiques constituant un premier lot de comportements interactionnels induits par de fortes hypothèses (corpus T0). Cependant, ce corpus ne sera véritablement exploitable que s'il est associé à des valeurs informatives, qui elles aussi relèvent de cette approche transversale, mais qui sont également inspirées des pratiques d'ethnométhodologie dans une étape d'auto-annotation (étape [3]). Pour un sujet, chacune de ces étapes expérimentales relève d'une séance unique, non reproductible pour ce même sujet, car le passage par ces étapes modifie le sujet, qui ne peut alors plus revenir à des états antérieurs de conditions précédentes. Les protocoles prennent en considération cette impossibilité de non-retour en arrière, amenant à recueillir toutes les données possibles à chaque étape.

---

<sup>82</sup> Site de l'équipex Amiqua4Home, accès via : <http://amiqua4home.inria.fr/fr/>

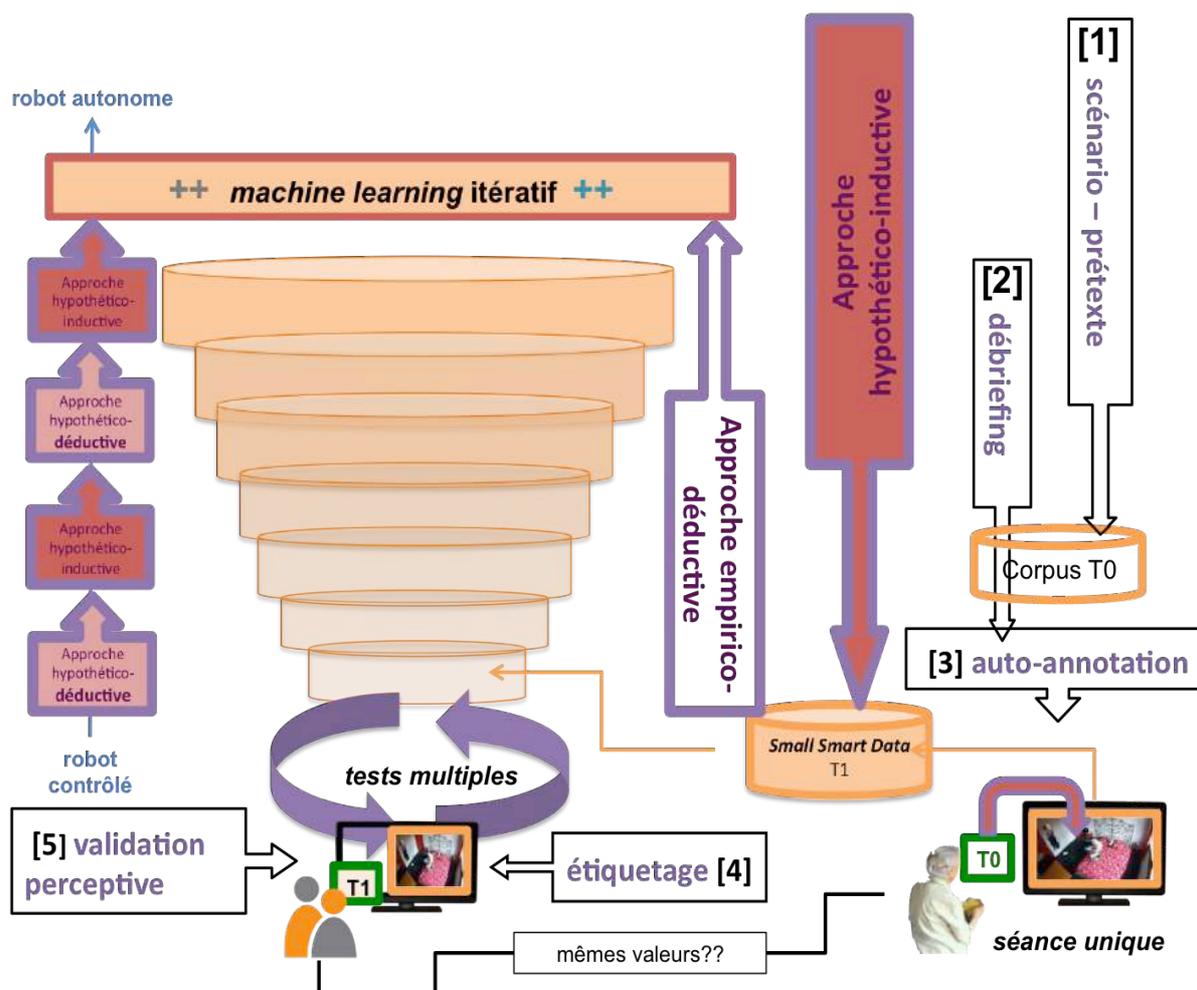


Figure 26 : Schéma global simplifié des approches scientifiques sous-jacentes aux étapes méthodologiques (sens de lecture de droite à gauche)

Dans l'absolu, le corpus obtenu ne représente qu'une petite quantité de données. En revanche, elle maximise les phénomènes à observer tout en tentant de les associer à des valeurs externes, formant ainsi un corpus *de Small Smart Data* (T1).

Cette démarche est réalisée parallèlement aux approches suivantes, augmentant petit à petit le corpus initial par les données des nouveaux sujets. Ces nouvelles données sont induites sur les mêmes protocoles que le premier cas puisque ces protocoles sont eux reproductibles, contrairement aux données qui dépendent d'un sujet évolutif.

À partir du moment où le corpus T1 existe, un étiquetage transversal est réalisé sur la base d'une approche hypothético-déductive, enrichissant le corpus d'observation descriptive, sur des critères les plus minimaux possible, sans passer par le filtre d'observation d'un modèle théorique particulier. Une fois que cet étiquetage (étape [4]) permet de cibler des événements spécifiques associés aux valeurs d'auto-annotations, une première validation perceptive est réalisée. Cette démarche détermine si des tiers totalement externes à l'expérience arrivent à discriminer les mêmes valeurs sur les événements relevés ou si ces dernières sont totalement idiosyncrasiques au sujet. Cette validation perceptive (étape [5]) s'appuie donc sur une très courte approche hypothético-déductive. Les paramètres décrivant des valeurs informatives et fonctionnelles relevées sont posés en hypothèses et leur comportement ainsi que leurs critères de fonctionnement sont déduits à travers le test perceptif. L'évaluation des déductions réalisées s'effectue alors par la reproduction courte de l'approche hypothético-inductive précédemment vue, augmentant encore la taille et la richesse du corpus. En revanche, alors que la première itération transversale des

expérimentations s'appuyait sur un système totalement contrôlé et non autonome, les nouvelles inductions itérées au cours de l'évaluation sont effectuées à travers un système qui permet de tester par couche d'acquisition d'autonomie successive, les modèles théoriques sous-tendant sa fabrication.

Ainsi sur le principe de boucles rapides et agiles, des hypothèses sont posées successivement, au fur et à mesure que l'étiquetage relève de nouveaux phénomènes, ce qui caractérise la dimension itérative. Dès lors que les hypothèses arrivent à déduire une partie du système interactionnel relevé, une courte phase d'approche hypothético-inductive est mise en place. La méthodologie s'appuie alors sur l'alternance entre les deux approches, en parallèle de l'étiquetage empirico-déductive longitudinale, qui elle-même est remaniée en fonction des validations et des évaluations concomitantes. Les cinq sous-sections qui suivent vont ainsi décrire ces différentes approches par les principes généraux illustrés à travers des étapes expérimentales numérotées ci-dessus, qui sont inhérentes à la méthodologie globale adoptée.

### **3.3.3.1. Approche hypothético-inductive : scénarii expérimentaux prétextés**

Le principe des prétextes est fondamental pour toute induction de comportement interactionnelle spontanée. Par exemple, dans le cadre de cette étude, le public visé devait faire l'objet d'un isolement constaté par les partenaires sociétaux et des critères formels pour cibler les sujets qui seront à l'origine des comportements observables dans le corpus récolté. En revanche, ce type de critère qui définit la spécificité du sujet ne doit en aucun cas être explicite, tant que le sujet est amené à produire des données à recueillir. De même, toute information ou tout fondement théorique support d'hypothèses qui motivent des objectifs de l'étude ne sont pas mentionnés tant qu'ils sont susceptibles d'influencer les productions du sujet. Cette condition implique de « masquer » ces informations, afin de détourner l'attention du sujet pour que ses comportements restent spontanés, non actés et non biaisés par l'envie de satisfaire aux objectifs de l'expérience ou pour toute autre raison l'incitant à changer son comportement, consciemment ou non. Un scénario expérimental détaillé est alors construit à l'appui de nombreux prétextes. Ces contraintes définissent tout d'abord un prétexte principal qui prend la forme d'une « fausse étude » scientifiquement réaliste et véritablement valorisable au sein de la communauté scientifique, même si elle n'intéresse pas directement les expérimentateurs. Ce prétexte principal s'accompagnera de prétextes connexes, qui auront le même but de détournement d'attention, mais qui porteront sur des aspects plus spécifiques pour satisfaire toutes les contraintes (théoriques et techniques) pour réaliser l'observation et la capture de données interactionnelles humaines.

Cette tâche prétexte principale sera celle qui est présentée lors de l'étape de recrutement du sujet. Elle ne transparaît jamais comme une manière d'évaluer les compétences du sujet ou elle ne présente pas explicitement les spécificités pour lesquelles les sujets ont été véritablement choisis. En effet, ces spécificités peuvent être rattachées à des jugements de valeur. Or, l'écologie expérimentale est la principale tare du paradoxe de l'observateur qui modifie le comportement authentique d'un humain. Ainsi, ce prétexte principal focalise d'autres caractéristiques, en modifiant la justification des raisons qui motivent le choix des profils des sujets recrutés. Pour augmenter la motivation à cette tâche qui est également un critère important, les stratégies sont multiples (cf. par exemple Joule et Beauvois, 2004). Les plus communes sont par exemple de valoriser ces autres caractéristiques, inversant l'effet de l'observation qui n'est plus dirigée sur le sujet, car il devient un facteur démontrant ou invalidant l'efficacité des propositions des expérimentateurs. Les sujets présentent ainsi une forme « d'expertise », « d'originalité » ou « de singularité » émanant de ses caractéristiques secondaires (dans notre cas ce serait l'âge de la personne qui se justifie dans une fausse étude gérontologique et les enjeux de maintien à domicile actuels). Ainsi les qualités évoquées sont

véritables, mais dévient la réalité de l'observation sur celle de l'expérience prétendue. La motivation à la tâche peut ainsi flatter l'*ego* des personnes, pour qu'elle puisse se sentir utile et qu'elle ait envie d'exploiter ces qualités, ou encore souligner ses caractéristiques secondaires comme très valorisables pour l'étude.

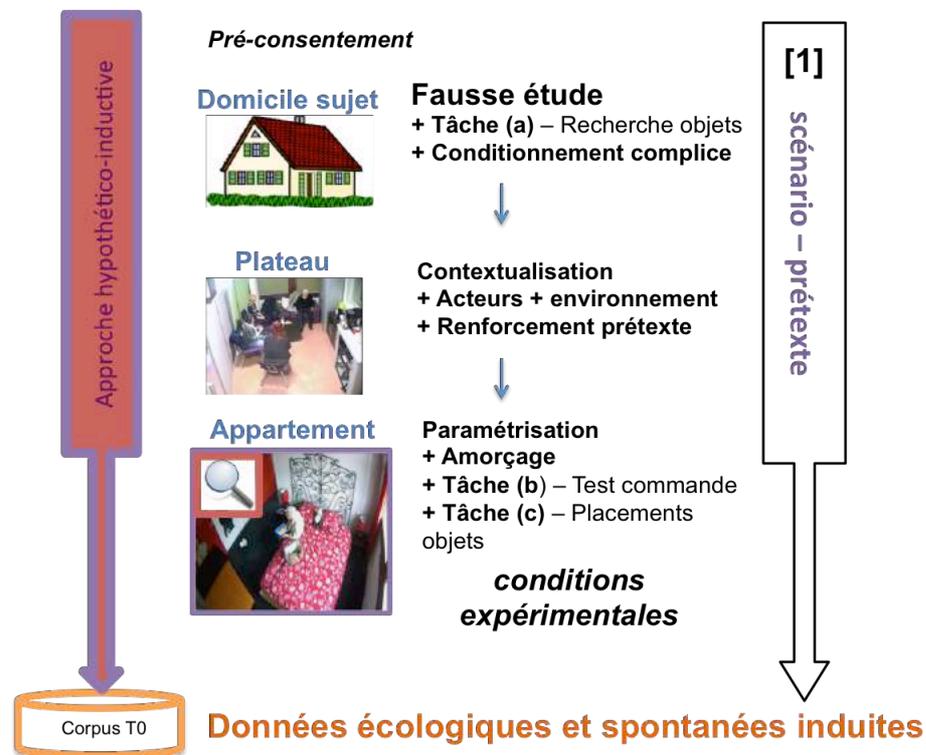


Figure 27 : Approche hypothético-déductive dans le recueil de corpus écologique et spontanée

Ce prétexte expérimental est donc un leurre, qui doit définir des tâches que le sujet doit réaliser au cours de l'expérience. Ces tâches, de manière détournée, doivent ainsi justifier le contexte expérimental d'une part, et d'autre part maximiser indirectement la production quantitative et qualitative des comportements qui sont véritablement visés et vont être induits par cette situation prétexte. Une fois cet objectif principal fixé, les approximations interexpérimentales doivent être limitées le plus possible. Les comportements seront donc induits selon un même protocole, décrit de manière très précise à travers des scénarii. Il y a alors deux types de scénarii qui interviennent :

- le **scénario scientifique** : il décrit à chaque étape de l'expérience les contraintes imposées par les hypothèses réelles de l'expérience (p. ex. pour notre cas, il s'agirait des conditions de présentation des stimuli vocaux du robot qui doivent suivre un ordre graduel spécifique – cf. Annexe 3)

- le **scénario de mise en écologie** : il se présente comme un script de théâtre où toutes les personnes participantes à l'expérience jouent des rôles précis, afin de recréer l'écologie d'interaction cadrée par le prétexte principal. Ainsi le discours et les actes de chacun sont « théâtralisés ». Les expérimentateurs deviennent des « acteurs » au sens littéral, mais ils ne sont pas forcés de suivre à la lettre le texte, pour éviter la perte de leur propre naturalité vis-à-vis du sujet. Cependant, ils doivent obligatoirement fournir certaines informations clés, de manière similaire à chaque expérience, et à des moments précisément déterminés dans ce scénario. Ce dernier doit être extrêmement bien contrôlé pour contextualiser et conditionner le sujet sans le faire douter sur le véritable objectif de l'observation. Ainsi, le timing la mise en scène (d'entrées/sorties des différentes pièces, les modalités d'accueil, le ton à avoir, la justification des outils utilisés etc.), où tout élément pouvant avoir un effet d'amorçage ou une influence quelconque sur le comportement du sujet doit être réfléchi et mentionné dans ce

scénario (cf. Annexe 4 pour l'exemplaire formalisé du scénario de notre expérimentation). Il s'agit donc d'un recueil d'indications pour tous les participants de l'expérience, à l'exception du sujet observé qui « vit le scénario en direct ». C'est notamment pour cette raison que cette étape fait l'objet d'une co-construction avec les partenaires sociétaux. En effet, ces partenaires étant proches des sujets, ils sont à même de pointer certaines incohérences du scénario qui peuvent aussi induire des comportements inhabituels chez le sujet. Ils peuvent donc être très actifs dans les expérimentations, en devenant eux aussi des acteurs de cette mise en scène et même conditionner le sujet en amont de l'expérience. La procédure appliquée pour l'étude de cette thèse est décrite section 0.

### **3.3.3.2. Transversalité de l'approche hypothético-inductive : « levée de voile » et continuité expérimentale**

Les expérimentations en sciences humaines présentent toujours, sauf exception, une étape de débriefing (étape 2 de la Figure 26). Cette étape permet de revenir sur les événements qui ont eu lieu au cours de l'expérimentation et de divulguer les objectifs cachés de l'étude en explicitant les prétextes. Il s'agit de la « levée de voile » du détournement d'attention des sujets. Cette étape peut également faire l'objet d'un protocole spécifique en plusieurs étapes. En effet, à l'issue d'un débriefing, il ne sera plus possible de réutiliser les prétextes montés précédemment, et le comportement du sujet ne pourra plus faire l'objet de la spontanéité attendue par rapport au moment où il était encore « non conscient » des conditions expérimentales. Cette authenticité transparait également dans la manière dont il donnera son avis sur l'expérimentation ou ses propres réactions au cours de cette dernière. En effet, un discours conscientisé ne reflète pas forcément la réalité objective ni perceptive du comportement. L'état de non-conscience du sujet est ainsi parfois prolongé jusqu'à une étape ultérieure de l'expérimentation qui peut être temporellement très éloignée. Ainsi, en fonction de la nécessité de recueil, le moment de la révélation est également à déterminer dès l'élaboration du protocole. Cette révélation peut être progressive aux différentes étapes, ou immédiate et totale à la fin de l'étude. De plus, il est possible d'intervertir les étapes 2 (débriefing) et 3 (auto-annotation) pour les mêmes raisons. En effet, le protocole peut impliquer et prévoir durant l'auto-annotation, des tâches qui ne peuvent être réalisées que si le sujet reste conditionné par le prétexte. La divulgation finale sera alors réalisée à la fin de l'auto-annotation, pour compléter un premier débriefing qui n'aurait alors été que partiel à la fin de l'expérimentation. L'expérimentateur doit parfois pouvoir recueillir de nouvelles données complémentaires ou « de contrôle » durant ces phases. En effet, l'induction de comportements s'appuie sur des hypothèses demandant parfois de comparer certains de ces comportements pour un même sujet. Cela permet de vérifier les effets d'une condition expérimentale par exemple, ou d'équiper par d'autres types d'appareil de mesure le sujet, pour compléter plus finement certains aspects des données recueillies (p. ex. des capteurs de sudation, des caméras go-pro, des micros à sensibilité variable etc.). L'utilisation de ces nouveaux instruments peut alors elle-même se justifier par des prétextes.

Dès lors que la révélation est réalisée, il est indispensable de s'assurer que l'ensemble du protocole soit bien intégré par le sujet. Le leurre et les prétextes devraient alors être bien compris et acceptés. Le temps que peut prendre cette étape de débriefing peut être long, mais il est incompressible. La procédure demandant de répéter oralement les conditions expérimentales par le sujet lui-même et avec ses propres mots, est notamment un moyen qui permet aux expérimentateurs de vérifier si ce dernier a été correctement informé des leures du format expérimental. Un consentement éclairé du sujet est alors possible à partir de ce moment. En revanche, des consentements sont donnés à différentes étapes de l'expérimentation, et sont au minimum de deux pour ce type d'expérience. Le premier est proposé en amont de l'expérience, et sollicite l'acceptation de participation à l'étude prétexte.

Le second est signé après la divulgation et le débriefing, en annulant le premier consentement par une mention explicite. Il s'accompagnera d'un document de droits d'image et d'utilisation des données récoltées. (cf. Annexe 5 pour le matériel d'expérimentation et documents utilisés dans ce cadre).

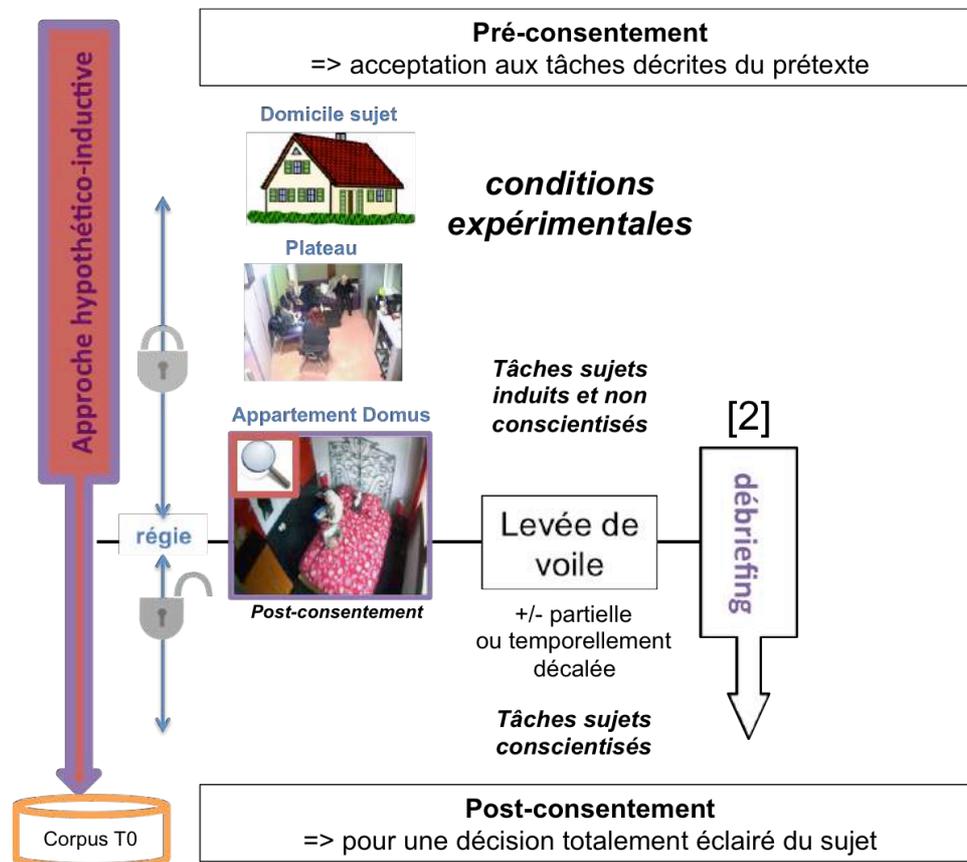


Figure 28 : Maintien méthodologique pour une approche hypothético-déductive transversale

Cette méthode prend le risque de devoir effacer les données du sujet, car les consentements sont à chaque fois donnés sur les informations que nous leur sont fournies au fur et à mesure de l'expérience. Les étapes de révélation annulent donc à chaque fois les consentements précédents, le dernier mentionnant bien qu'il s'agit d'un document terminal. Les sujets ont alors la possibilité de refuser l'exploitation des données recueillies, mais ce qui en général est très rare, car ils ont été sollicités sur la base d'une motivation à la tâche. En revanche, c'est un moyen d'assurer la collecte de données spontanées tout en préservant un minimum d'éthique. Enfin, lors de la révélation, tous les participants de l'expérience sont présentés (notamment ceux qui sont derrière le magicien d'Oz et que le sujet ne voit jamais pendant le protocole). En effet, qu'il s'agisse d'un simple observateur, d'un complice ou d'un expérimentateur-technique, cette révélation est nécessaire et elle facilite la compréhension de la situation expérimentale pour le sujet. Elle propose ainsi une transparence quant aux enjeux de l'étude. La procédure appliquée pour l'étude de cette thèse est décrite section 3.4.2.1.4.

### 3.3.3.3. Approches inspirées de l'ethnométhodologie : annotations externes des valeurs informatives

Dans un corpus de parole multimodale spontanée, l'étiquetage des données à différents niveaux d'observation linguistique et interactionnelle (acoustique, phonétique/phonologique,

morphologique, lexicologique, morphosyntaxique, gestuelle, proxémique etc.) permet de décrire des comportements observables et mesurables. Néanmoins, ces éléments véhiculent également des valeurs informatives et fonctionnelles, découlant de processus émotionnels, intentionnels, cognitifs ou socio-affectifs, qui dépendent d'un degré d'ambiguïté et de subjectivité important. De même que l'étiquetage, la méthodologie tente de faire émerger, sans s'appuyer sur modèle théorique préalable, des étiquettes de valeurs, tout en essayant de suivre la temporalité des changements qui sont induits. En effet, certaines études (Douglas-Cowie et coll., 2007) ont montré qu'il existait des méthodes permettant d'accéder à ce type de descripteurs. Elles sont à la base de nouvelles hypothèses et doivent être vérifiables et falsifiables (Popper, 1980) pour rester en adéquation avec la première approche hypothético-inductive de la méthodologie.

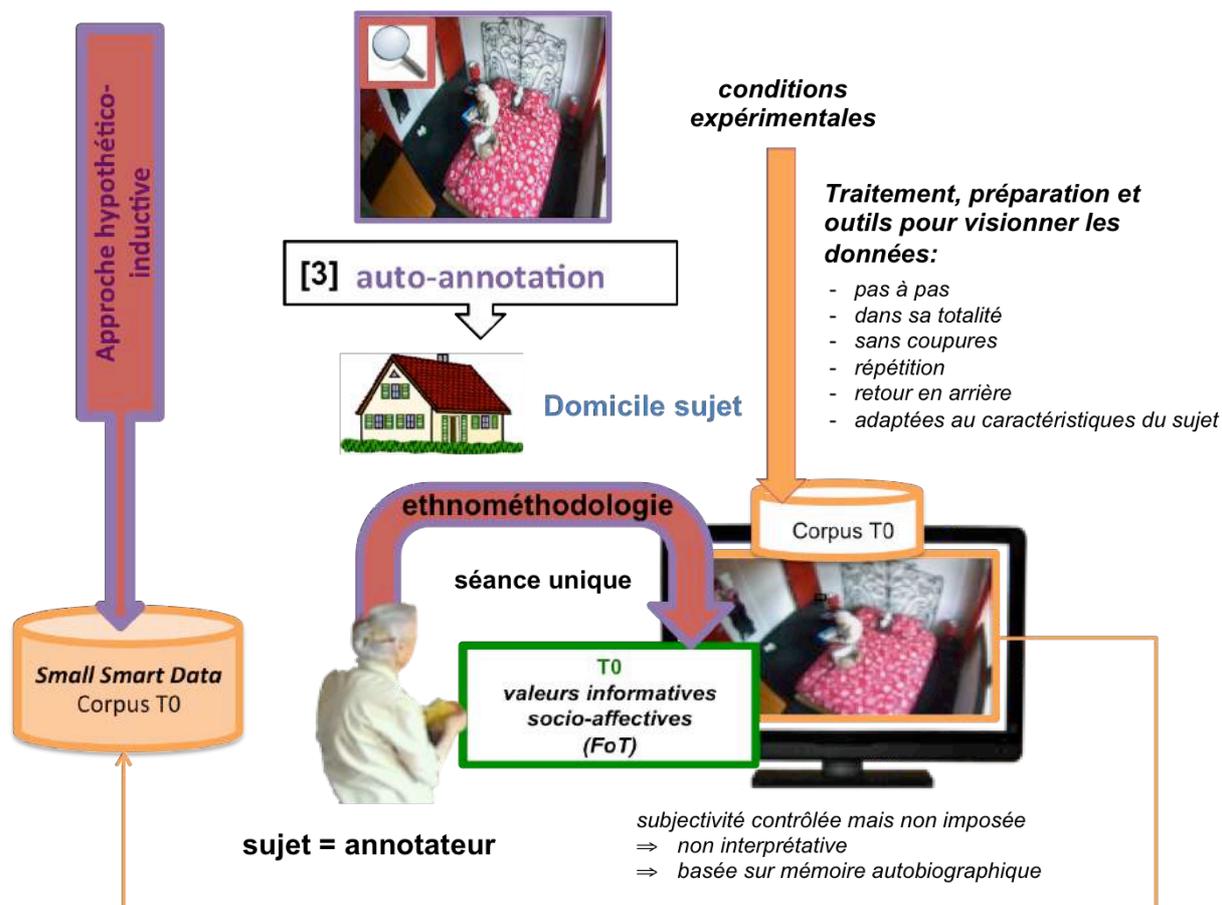


Figure 29 : Approches d'ethnométhodologie couplée aux approches hypothético-inductives

L'auto-annotation est une méthode proposée comme une alternative aux annotations expertes, initialement mise au point dans l'analyse du corpus E-Wiz (Aubergé et coll., 2005). Elle s'inspire des principes d'ethnométhodologie (Garfinkel, 1967/2005 ; Pléty, 1993 ; Tinbergen, 1963) qui ont été utilisés et adaptés par (Audibert et coll. 2006, Loyau et coll., 2006 ; Vanpé, 2011) sur des données induites d'interactions personne-machine. Dans des études interactions homme-robot, elle est à distinguer des méthodes telles que les questionnaires post-expérimentaux effectués par le sujet ou des tiers externes (Hall et coll., 2014 ; Ivaldi, Anzalone, Rousseau, Sigaud, et Chetouani, 2014), les annotations manuelles expertes (Rich, Ponsler, Holroyd, et Sidner, 2010 ; Sidner, Kidd, Lee, et Lesh, 2004 ; Sidner, Lee, Kidd, Lesh, et Rich, 2005), des annotations par le sujet, ou encore générés à partir de modèles théoriques émotionnelles dans le cadre de la robotique sociale (Devillers, Tahon et al. 2015). Pour ce dernier cas, la terminologie anglaise de *self-labelling* opposée à *auto-annotation* est certainement plus illustrative. En effet, le *self-labelling* est un moyen de

demander aux sujets d'annoter leur propre production, ce qui est un point commun avec notre méthodologie. En revanche, dans ces méthodes, les sujets sont amenés soit à suivre des grilles d'étiquettes, construites sur un modèle théorique, que les expérimentateurs proposent pour en faire une tâche d'association à leurs stimuli (Shibata, Ohkawa, et Tanie, 1996), soit faire l'objet d'une proposition libre et explicite sur la nature des données à annoter (Sanghvi et coll., 2011). Or des études comme (Girard-Rivier et coll., 2016) ont pu montrer qu'entre un acte consciemment reproduit et un acte spontané, la nature de l'acte peut grandement changer vis-à-vis d'une même étiquette, voire même générer des actes qui n'existent pas dans un corpus, alors que le sujet était convaincu de les avoir produit quelques minutes auparavant. Cette déviation dépendait de la condition où l'association demandée est explicite et consciente, ce qui motive le choix de notre démarche de ne pas faire conscientiser cette tâche à l'image de *self-labelling*, dont les mesures des degrés de confiance seraient également à préciser.

Par opposition aux précédentes méthodes, l'auto-annotation est une annotation externe qui tente de s'extraire de la dimension interprétative du corpus. Elle sollicite la réminiscence des souvenirs par les sujets, en faisant appel à leur mémoire autobiographique (Gentzler et Kerns, 2006 ; Healy et Williams, 1999). Cette dernière semble être un mécanisme relativement robuste pour des annotations de cette nature (Frijda, Kuipers, et Ter Schure, 1989), y compris au facteur temps : lorsqu'une personne se revoit en train d'effectuer une action qu'elle a vécue par le passé, ses états d'esprit, ses intentions, ses buts, ses motivations, ses émotions, ses socio-affects ou encore ces processus cognitifs (que nous décrivons pour rappel sous les termes de *Feelings of Thinking* ou *FoT*) refont surface. Cela peut être facilité par un visionnage pas à pas de l'expérience vécue. Durant ces sessions, les sujets sont amenés à utiliser leurs propres mots pour décrire leurs ressentis par l'émergence de leurs souvenirs, en évitant d'être confrontés à des suggestions de la part des expérimentateurs. En revanche, ils peuvent être recadrés s'ils tendent à effectuer des interprétations de ce qu'ils visualisent, car l'expérimentateur sait qu'elle est la nature de l'élément à recueillir, même s'il doit se forcer à rester aveugle à la valeur qu'il évite de suggérer. La procédure propose donc que les données expérimentales axées sur le recueil d'interactions soient entièrement visionnées (N.B. y compris les parties qui semblent passives, car nous pointons des processus cognitifs), à l'appui des temps de déroulement réels de l'expérience. Cela faciliterait l'accès à cette mémoire autobiographique, tout en réduisant le risque d'un biais lié à une perception temporelle faussée. Enfin, le visionnage des données se fait chronologiquement, avec la possibilité de revoir la partie qui est en cours de description, mais sans retour en arrière possible, une fois que la description de cette partie est terminée. En effet, cela minimise toute démarche interprétative qui serait favorisée par une vision plus globale du sujet sur l'ensemble de ses données.

Le protocole appliqué pour cette étude est précisé dans la section 3.4.3. Les observations qui en découlent font l'objet du Chapitre 5.

### **3.3.3.4. Approche empirico-déductive : étiquetage descriptif longitudinal des données**

Il s'agit d'une étape intrinsèque pour tout corpus, et qui représente l'un des enjeux du Traitement Automatique des Langues. Elle est évolutive et longitudinale, à travers les différentes boucles itératives. Chaque traitement aboutit à un nouvel état de corpus et permet de produire progressivement des données d'apprentissage machine ou de modéliser partiellement des briques de systèmes automatiques. Elle résulte de la dimension empirique du protocole, qui tente de décrire le plus finement possible les données, en les associant à des étiquettes purement descriptives, quantitativement mesurables. Cet étiquetage vise à être le plus complet possible en évitant de passer sous le filtre de grilles de lectures théoriques. En effet, elles pourront être confrontées à ces divers modèles, mais seulement à des étapes

ultérieures. L'enjeu de cette étape est de noter un maximum d'événements potentiellement significatifs, de manière exhaustive. Ce sont ces éléments qui pourraient éclairer les questions et hypothèses de l'étude. En effet, la précaution à prendre serait de ne pas rentrer dans un processus tautologique ou purement ontologique des données. Selon les principes de l'ethnométhodologie repris par (Vanpé, 2011 ; Loyau, 2005), appliqués aux données humaines visant à être utilisées par des systèmes informatiques automatisés, l'idée serait de constituer des éthogrammes des productions interactionnelles des sujets.

La finesse de la description, ainsi que la richesse de l'annotation est un enjeu important de cette méthodologie pluridisciplinaire. À partir d'un même objet d'étude, chaque approche d'observation est différente selon une discipline donnée, puisque notre système attentionnel n'est pas dirigé au même endroit. Ainsi, plus large est le spectre d'observation, plus on éviterait de « passer à côté » de certains phénomènes potentiellement significatifs. L'intelligence humaine présente l'atout et la perversion de nous aveugler devant certaines évidences par le tri sélectif informationnel qu'elle réalise pour percevoir le monde. Là où une discipline faillira à noter un phénomène, une autre pourrait donc la rattraper. Cet échange transdisciplinaire a donc avant tout pour objectif de repérer les éléments à observer, sans forcément appliquer la méthode d'observation de cette discipline à la lettre, sauf si elle décrit objectivement et finement les données. Mais en réalité, elle amène souvent à adapter de manière hybride ces méthodes dont elle fait un usage détourné, afin de contourner les limites des approches classiques si elles existent, tout en essayant de proposer de nouvelles manières d'adapter ces approches.

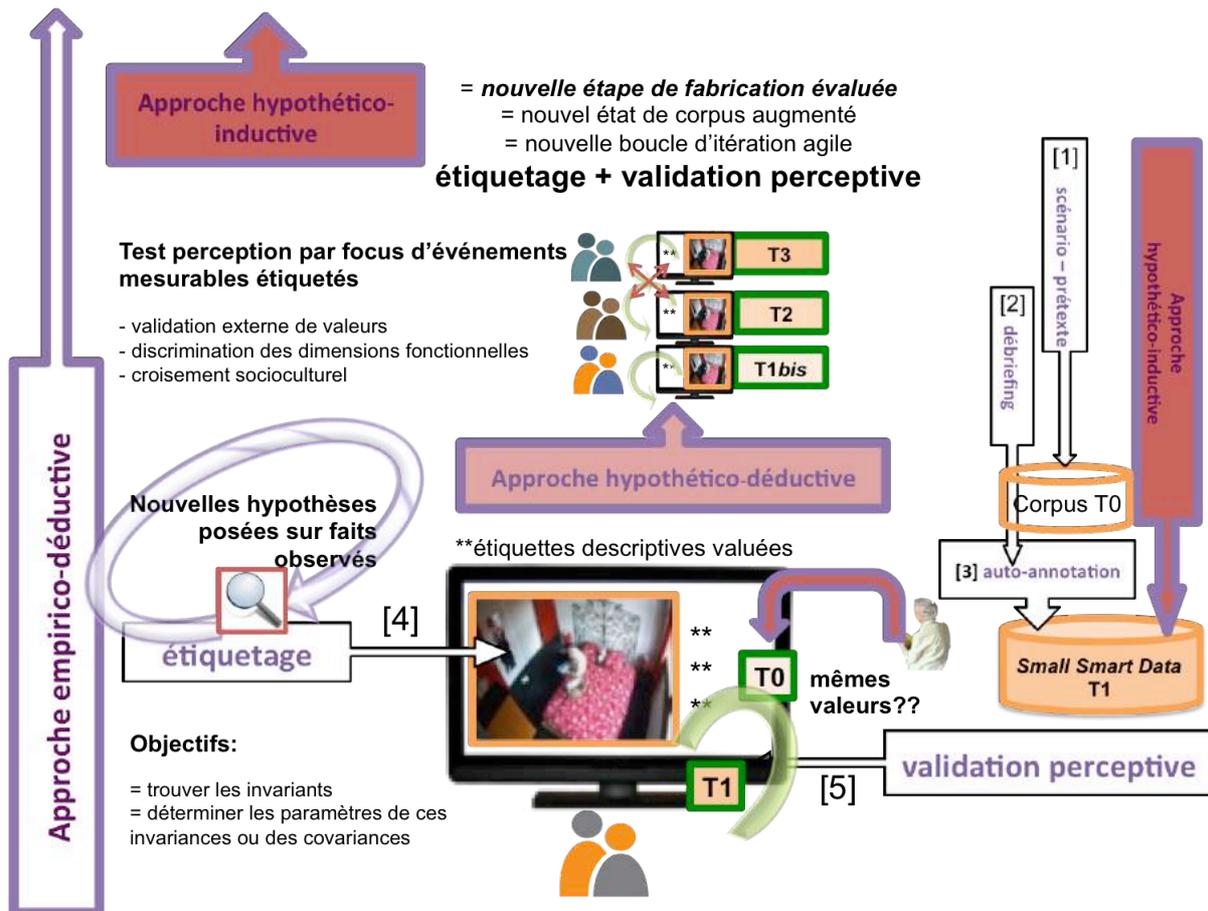


Figure 30 : Approches empirico-déductive et hypothético-déductive pour une fabrication évaluée par nouvelle approche hypothético-inductive courte

De plus, dans cet étiquetage, la rareté ou la finesse est aussi un possible indice explicatif dont certains détails peuvent être décrits (p. ex. les micro-expressions). Des règles et des nomenclatures seront alors définies pour chaque étiquetage, visant à faire une description objective la plus minimale possible et de faciliter le filtrage des données, pour en associer des étiquettes complémentaires. Ainsi, la valeur informative de nature socio-affective ne sera proposée qu'à partir des auto-annotations et non issue de l'étiquetage. Les démarches appliquées au corpus de notre étude font quant à elles l'objet de tout le Chapitre 4. Il est également à noter que l'étiquetage fonctionne de pair avec la validation perceptive, ce qui est illustrée également par la Figure 30 et la section suivante.

### 3.3.3.5. Approche hypothético-déductive : la validation perceptive

Il s'agit de méthodes expérimentales qui sont le propre de la psychologie de la perception, qui peut être de nature auditive, visuelle ou audiovisuelle, et dont les ouvrages ne manqueront pas. Ainsi cette section sera loin d'être exhaustive sur les méthodes et mécanismes impliqués dans ce type d'expérimentation. En revanche, elle essaiera d'introduire quelques indications susceptibles de faciliter l'entrée en matière de cette étape importante de la méthodologie globale adoptée, notamment si certains seraient également amenés à l'adopter dans le futur.

Cette validation perceptive est basée sur la présentation de stimuli sur des critères variables. Elle permet de valider ou non des valeurs associées à ces stimuli, par l'intermédiaire de juges totalement externes à l'expérimentation de recueil et d'annotation. Il s'agit ainsi de vérifier si les étiquettes d'auto-annotations, qui ne restent que des valeurs subjectives et individuelles sont viables. Elle vérifie notamment si ces valeurs appariées aux comportements, relevés par l'étiquetage, sont généralisables et donc potentiellement imputables à des processus cognitifs réels. Elle vérifie ainsi que les valeurs relevées ne dépendent pas uniquement d'une spécificité idiosyncrasique du sujet. Cette étape permet ainsi de cibler et produire de nouvelles hypothèses permettant de vérifier ces valeurs plus ou moins indépendantes, mais sur plusieurs expériences perceptives. Ainsi, chaque étude valide au fur et à mesure des dimensions spécifiques du corpus, aboutissant ainsi sur une version de corpus augmentée, à un instant donné (cf. Figure 25 et Figure 30). Dans cette approche, des étapes successives confirment ou non la validité des données relevées et définissent les invariants et leurs critères. Elle consiste :

- à commencer par effectuer des tests perceptifs associatifs avec des juges de caractéristiques semblables aux participants de l'étude (p. ex. de Biasi, 2012) afin de valider la portée des valeurs informatives et fonctionnelles des objets interactionnels dénotés par les sujets.

- puis à effectuer des croisements avec des populations de caractéristiques différentes sur paramètres socioculturels variables. La variation peut également porter sur la nature et les dimensions spécifiques qui différencient ou assimilent les stimuli jugés sur des critères définis. Les étiquettes seront alors composées de l'ensemble des valeurs portées par ces critères. Le nombre de prototypes qui sera alors testé devra être quantitativement suffisant et significatif pour expliquer la variation de ces dimensions observées. Si plusieurs paramètres semblent coévoluer, cette validation perceptive tente de mesurer le poids de chaque paramètre.

Le critère de variabilité choisi permettra alors de repérer les invariants et de mesurer la portée de leur universalité, notamment si cette validation perceptive s'appuie sur une étude interculturelle (Signorello et coll., 2010 ; Mac, Aubergé, Castelli, et Rilliard, 2012 ; Lu, Aubergé, Rilliard, et Gu, 2013) en adoptant des approches comparatives (p. ex. Sasa et coll., 2012 croisé sur de Biasi, 2012). Dans la conception du test, sachant que le critère de variation est associé à des stimuli avec de multiples descripteurs, les traces de réponses devront prévoir un maximum d'informations portant sur : 1) les spécificités des stimuli (c.-à-d. ce qui est

réalisé à travers l'étape de l'étiquetage), 2) les spécificités du sujet, et 3) toute information liées aux conditions expérimentales (p. ex. temps de réaction, l'ordre de présentation du stimulus, le degré de confiance de la réponse, actions effectuées sur les interfaces de réponses etc.). De plus l'ensemble de ces informations n'est pas reporté globalement, mais doit être redondant sur chaque enregistrement de réponse associé à un stimulus, afin de pouvoir isoler et filtrer les réponses en fonction des valeurs de critères associés (c.-à-d. créer une matrice organisée avec des critères organisés en colonne et les enregistrements présentés en ligne). Le test perceptif dépend fortement de l'interface de présentation et des consignes d'amorçage (p. ex. la nature multimodale du test qui impose un ordre de présentation de modalité ; l'instanciation d'un chronomètre pour induire des effets comme le stress etc.). Ces éléments devront être analysés en détail dans la conception (p. ex. taille/nature des boutons, couleurs de fond, positionnement des éléments etc.). Cette interface devrait notamment pouvoir faciliter la tâche, réduisant au maximum la charge cognitive des juges, afin de recueillir des réponses instinctives. En effet, les mécanismes automatiques et sous-jacents sont le propre de ces tests. Le choix est toujours motivé, qu'il soit conscient ou inconscient, malgré le caractère parfois déroutant qui donne l'impression de « donner des réponses au hasard » (N.B. cela sera particulièrement ressenti pour les stimuli de type micro-expressions ou les tests basés sur l'impressionnisme qui est à la base des théories gestaltistes). Enfin, il existe dans ce cadre plusieurs facteurs d'influence, tels que l'effet d'apprentissage dont découlera l'ordre de présentation des stimuli. Cet effet d'apprentissage sera consécutif à la nature répétitive de la tâche du test. Ainsi l'ordre de présentation des prototypes devra être randomisé (soit avoir plusieurs patterns de présentation, plus ou moins regroupés) afin de contourner ces effets.

Cette démarche est donc hypothético-déductive puisqu'à partir d'hypothèses portées sur les comportements relevés par l'étiquetage, mis en regard des auto-annotations, elle permet de valider ou d'invalider les critères de variation et de fonctionnement de dimensions communicatives. Elle est appliquée en parallèle de l'approche empirico-déductive d'étiquetage. Dans le cadre de la thèse, les travaux n'ont pas pu être atteindre cette étape méthodologique. En revanche des travaux comme ceux de (Magnani, Aubergé, Bayol, et Sasa, 2017) ont initié l'approche de cette méthodologie par étapes encapsulées, en s'attachant à la recherche de primitives d'animisme sur divers paramètres couplant des caractéristiques visuelles et vocales, ces dernières étant portées par des micro-expressions semblables à celles présentées dans notre étude, et contribuant ainsi indirectement à définir les fonctions des paramètres vocaux de ces stimuli sur le phénomène d'animisme.

### **3.3.3.6. Boucles itératives : alternance et transversalité des approches scientifiques**

À travers les étapes d'étiquetage et de validation perceptives basées d'une part sur une approche hypothético-déductive pour le premier, et d'autre part une approche empirico-déductive pour le second, la déduction permet de rentrer dans un processus itératif de modélisation, co-fabrication, et évaluation. En effet, à chaque nouvelle hypothèse posée sur un phénomène interactionnel mis en évidence par l'étiquetage, elle permet de déduire un fonctionnement de comportement qui pourra contribuer à modéliser une partie d'une brique technologique composant un système automatique global, comme celui du robot, un instrument qui permet de simuler le comportement humain modélisé. De cette modélisation, il est possible d'en fabriquer des prototypes, voire d'en modifier une brique existante sur la base de ces observations. Cette fabrication implique aussi bien l'équipe de recherche, aidée par les plateformes FabMSTIC et Amiqua4Home, ainsi que le partenaire industriel. Ce dernier pourra lui-même enclencher des processus de développement pour son produit. Cependant, cette co-fabrication se fait à l'appui d'évaluation courte, qui lui permettra de prendre des décisions par un éclairage que sa technologie pourrait avoir sur ses futurs usagers et les contraintes de son intégration sociétale, impliquant donc le partenaire social. Cette évaluation

est donc mise en situation dans une approche hypothético-inductive, où la nouvelle hypothèse devient celle qui sous-tend la modélisation. Le protocole est alors reproduit de la même manière que la toute première approche appliquée, en reproduisant dans les mêmes conditions expérimentales de nouvelles interactions, mais sur la base d'un prototype partiellement automatisé. Ainsi, si les comportements interactionnels sont reproductibles à travers la modélisation du prototype, elle est temporairement viable, jusqu'au moment où elle sera démontrée comme partiellement ou totalement falsifiée. Toute hypothèse étant, par nature scientifique, falsifiable, de nouvelles hypothèses seront alors proposées, tout en n'oubliant pas de vérifier si de potentiels paramètres ne peuvent pas expliquer les raisons de cette non-reproductibilité.

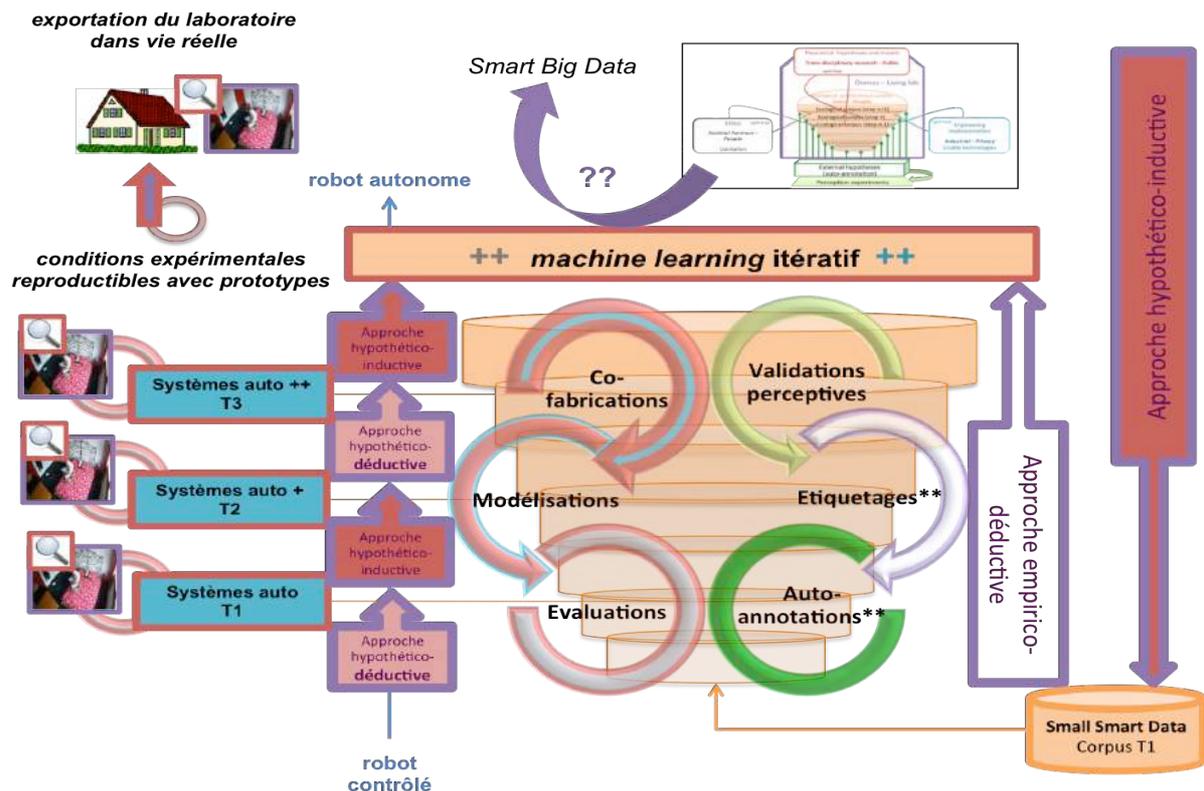


Figure 31 : Schéma de la dimension itérative des boucles agiles de la méthodologie globale

Cette itération est donc induite par l'alternance entre courtes approches hypothético-déductives et hypothético-inductive de fabrication/modélisation, en parallèle d'une approche empirico-déductive de recueil de données qui est longitudinale. En effet, l'étiquetage s'enrichit au fur et à mesure de ces itérations. Le corpus initial devient donc de plus en plus dense, aussi bien en information qu'en quantité, car le processus permet d'accumuler de nouvelles données à chaque nouvelle boucle. Ce corpus cumulatif sert d'autre part, à travers le filtrage rapide des étiquettes, à la constitution d'un corpus d'entraînement dans le cadre d'un apprentissage machine. D'autre part, elle souligne de potentiels mécanismes inhérents à l'évolution des phénomènes interactionnels. Elle contribue ainsi à modéliser le système sur lequel est basé le *machine learning*, qui d'ailleurs pourrait lui-même évoluer pour que la conception de l'apprentissage puisse devenir itérative à l'image de cette méthodologie.

Les étapes, qui suivent la logique de cette méthodologie, seraient ensuite de sortir de la condition laboratoire pour étendre cette méthodologie à la vie réelle. Elles exporteraient l'exécution de ces approches pour les appliquer des situations courantes, toujours à l'appui des mêmes principes sous-jacents à l'observation humaine. La question du passage à l'échelle interviendrait alors, interrogeant sur les méthodes qui permettraient de passer de Small Smart Data à des Smart Big Data. En effet, les Big Data s'appuient sur des méthodes mathématiques

qui généralisent et donnent une approximation des mécanismes cognitifs de l'humain par le guidage de l'intelligence humaine qu'il retrouve dans les données de manière implicite. Ainsi l'explicitation de ces mécanismes mis en œuvre est contournée et compensée à l'appui d'une importante quantité de données. En revanche, cela ne semble pas permettre d'un point de vue théorique de comprendre le comportement humain lui-même pour le modéliser. De fait le changement de granularité pour observer non plus la somme des comportements de chaque individu, mais la dynamique qui change un ensemble d'individus comme un système à part entière, et dont le comportement émane d'un système collectif plus ou moins implicite, mais explicative du phénomène interactionnel au passage de cette échelle. Les enjeux des futures recherches pourraient ainsi porter sur les méthodes qui permettraient ce changement de granularité.

### 3.4. Protocoles expérimentaux

Suite à l'explicitation de la méthodologie globale et des approches impliquées, cette section présentera les protocoles mis en œuvre pour leur application. L'étude *EmOz Elderly Expressions*, qui s'intéresse donc aux interactions entre des personnes âgées fragilisées par un isolement relationnel et un robot majordome d'habitat intelligent, suit des démarches successives d'induction de comportement et recueil d'un corpus nommé EEE, dont les étapes de collecte sont résumées dans la Figure 32 ci-dessous.

Pour aborder ces étapes produisant des données variées, les sections abordent successivement le processus de recrutement, la capture des données expérimentales, et des démarches d'auto-annotation.

#### 3.4.1. Processus de recrutement

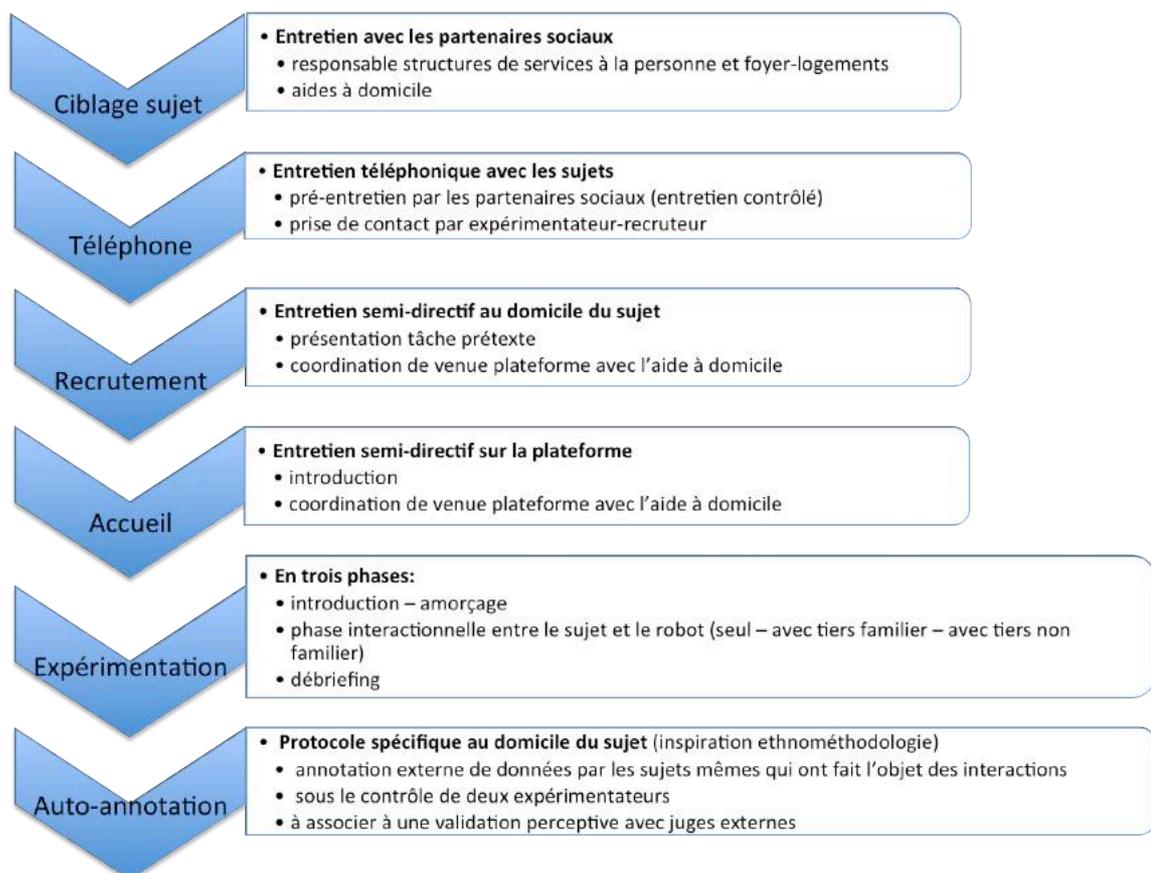


Figure 32 : Chronologie des démarches de l'étude EmOz Elderly Expressions

Le ciblage des sujets, fondé sur un choix méthodologique fort, posait comme conditions d'observations l'implication de personnes âgées fragilisées par un état d'isolement relationnel. En effet, cette condition permettrait d'observer la glu socio-affective, ou plus précisément de vérifier les effets des primitives posées en hypothèses sur l'état dégradé de la glu chez ces personnes. En revanche, ce type de public, comme cela est évoqué section 3.3.1.3, n'est pas formellement référencé ni facilement accessible. Le recrutement s'est donc effectué en plusieurs vagues et de manière longitudinale, sur deux ans, car la mise en place des prétextes et des collaborations est un processus progressif.

### 3.4.1.1. Sélection des sujets

Une première étape, avant consultation des sujets, consiste à exposer les motivations sociétales et expérimentales aux responsables des structures aidantes (services à la personne, foyer-logement pour personnes âgées). La seconde étape porte sur la réflexion commune pour cibler les personnes âgées fragilisées sur le critère d'isolement relationnel. Malgré la variabilité de la fragilité et du rythme de vieillissement de la personne, ce paramètre reste non définissable uniquement sur un critère d'âge. Les critères de sélection visaient préférentiellement :

1) Des personnes de plus de 75 ans, une tranche d'âge communément choisie pour des études gérontologiques, malgré les divergences et le recul de cet âge par rapport aux changements en faveur d'une qualité de vieillissement. La qualité du passage au « quatrième âge » dépendrait d'ailleurs des régions géographiques et de l'appartenance socioculturelle. Pour l'étude, nous nous limitons à Grenoble et ses proches agglomérations pour permettre les déplacements sur la plateforme expérimentale. L'origine socioculturelle des sujets est souhaitée variée. Ces critères ne sont pas absolus et restent indicatifs ;

2) Des personnes préalablement proposées et évaluées par les assistantes sociales des structures comme étant d'un niveau de GIR 6 et 5, sans aucune atteinte cognitive, afin que le déplacement soit rendu possible et que les personnes ne soient pas dans une dépendance physique trop importante. En fonction du profil de la personne, le critère d'âge a été un facteur de sélection secondaire, l'état de santé et d'isolement étant bien plus prioritaire.

## Méthodologie et protocoles expérimentaux

		GIR			6		
		Variables Discriminantesss			A	B	C
GIR 1	<ul style="list-style-type: none"> <li>Personne dont les fonctions physiques et mentales sont altérées, confinées au lit ou en fauteuil et qui a besoin d'une aide constante et continue.</li> <li>Personne en fin de vie.</li> </ul>	<b>1 Cohérence</b>	Converser et se comporter de façon sensée par rapport aux normes admises.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
GIR 2	<ul style="list-style-type: none"> <li>Personne confinée au lit ou en fauteuil, dont les capacités mentales ne sont pas totalement endommagées mais qui exige une assistance pour la plupart des gestes quotidiens.</li> <li>Personne dont les capacités physiques ne sont pas complètement altérées mais dont l'état mental nécessite une surveillance continue.</li> </ul>	<b>2 Orientation</b>	Se repérer dans le temps (jour et nuit, matin et soir), dans les lieux habituels ...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
				<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
GIR 3	<ul style="list-style-type: none"> <li>Personne autonome mentalement et partiellement sur le plan physique mais qui nécessite d'une aide quotidienne et plusieurs fois par jour pour les soins corporels.</li> </ul>	<b>3 Toilette</b>	Se laver seul, habituellement et correctement, le haut et le bas du corps.	Haut	<input type="radio"/>	<input type="radio"/>	
				Bas	<input type="radio"/>	<input type="radio"/>	
GIR 4	<ul style="list-style-type: none"> <li>Personne pouvant se déplacer à l'intérieur de son domicile mais qui ne peut pas assumer ses sorties et qui a besoin d'une assistance pour l'habillement et la toilette.</li> <li>Personne possédant une autonomie locomotrice mais qui requiert de l'aide pour les repas et les soins corporels.</li> </ul>	<b>4 Habillement</b>	S'habiller, se déshabiller, seul, totalement et correctement.	Haut	<input type="radio"/>	<input type="radio"/>	
				Moyen	<input type="radio"/>	<input type="radio"/>	
GIR 5	<ul style="list-style-type: none"> <li>Personne nécessitant d'une aide ponctuelle pour les activités quotidiennes (toilette, repas) ainsi que pour le ménage.</li> </ul>	<b>5 Alimentation</b>	Se servir et manger seul, correctement. <i>On considère que les aliments sont déjà préparés.</i>	Se servir	<input type="radio"/>	<input type="radio"/>	
				Manger	<input type="radio"/>	<input type="radio"/>	
GIR 6	<ul style="list-style-type: none"> <li>Personne autonome pour les gestes essentiels de la vie courante.</li> </ul>	<b>6 Elimination</b>	Assurer seul et correctement l'hygiène de l'élimination urinaire et anale.	Urinaire	<input type="radio"/>	<input type="radio"/>	
				Anale	<input type="radio"/>	<input type="radio"/>	
		<b>7 Transfert</b>	Pouvoir passer seul, d'une des 3 positions (debout, assis, couché) à une autre, dans les 2 sens.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
		<b>8 Déplacement intérieur</b>	Se déplacer seul à l'intérieur du domicile (même avec canne, déambulateur ou fauteuil roulant).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	

Figure 33 : Échelles GIR et modèle (gauche) d'évaluation de la grille AGGIR<sup>83</sup> (à droite)

Les demandes des personnes âgées, visant à bénéficier d'une aide à domicile, ou encore à être déplacées au sein de foyer-logement, sont des critères utilisés pour cibler un public qui a de plus fortes chances de vivre une situation d'isolement relationnel et pas seulement situationnel. Les critères définissant l'isolement ne font pas encore l'objet d'échelles consensuelles, en sociologie par exemple. Les partenaires sociaux listent donc dans un premier temps de potentiels sujets par leur observation empirique de terrain, en connaissance de nos motivations théoriques, ainsi que du protocole expérimental. Les caractéristiques de ces personnes sont également renseignées aux expérimentateurs, cas par cas lors de cette étape. Elles permettent aussi d'obtenir des informations nécessaires pour la prise de contact avec les personnes choisies. Ce processus est donc une première étape de co-construction expérimentale.

Des précautions sont adressées aux partenaires sociétaux qui établissent la mise en contact des expérimentateurs et des personnes âgées. Un entretien au préalable explicite quels sont les éléments que les partenaires ont le droit d'évoquer ou non avec les personnes. La consigne a ainsi été de dire qu'une étudiante fait des études sur des personnes âgées pour tester des habitats témoins et qu'elle aimerait avoir un entretien avec la personne pour discuter de cette étude. Cette demande de participation par le partenaire social se fait sous plusieurs conditions :

- une prise de contact par les responsables de la structure sociale avec cette formulation prédéterminée avec les expérimentateurs ;
- la non-divulgence des tâches et objectifs au sujet (prétextée ou réelle) ;

<sup>83</sup> Cette évaluation de la grille AGGIR fait l'objet d'une passation auprès d'institutions spécialisées ainsi que par le biais des assistants sociaux de structures de services à la personne. Le formulaire donné en exemple à titre indicatif est disponible sur le site de l'ASPA (aides sociales aux personnes âgées) : [http://www.aidesociale.com/Dossiers/calcul\\_du\\_gir.htm](http://www.aidesociale.com/Dossiers/calcul_du_gir.htm) (consulté le 13/11/2014)

L'échelle à 3 niveaux est définie comme suit : A : Fait spontanément seul, totalement, régulièrement et correctement. / B : Fait partiellement, incorrectement, irrégulièrement ou sur incitation. / C : Ne fait pas, ne peut pas ou ne veut pas le faire.

## Méthodologie et protocoles expérimentaux

- la non-stigmatisation sur le profil du sujet (notamment de la fragilisation et de leur situation d'isolement supposé) ;
- rassurer la personne pour les motiver à participer, mais sans faire d'amorçage excessif, qu'il soit positif ou négatif ;
- éluder toute question détaillée pour les renvoyer à l'expérimentateur ;

Les partenaires sociaux gardent ainsi une liste de personnes motivées et susceptibles de répondre positivement à la participation à l'expérience. Puis l'expérimentateur-recruteur contacte rapidement, mais dans un second temps, les personnes identifiées pour se présenter et obtenir un rendez-vous pour discuter du contenu de l'étude menée, directement au domicile de la personne. À cette étape, aucune information détaillée n'est fournie, l'étude est menée par une étudiante en gérontologie qui fait une étude sur les aspects de maintien à domicile par le biais d'habitats témoins.

Une seconde voie de recrutement visait également des sujets sans qui ne sollicitaient pas forcément d'aidants. Ce processus s'est effectué en parallèle, par le réseau de proximité, voisinage, clubs et porte-à-porte, ciblé sur les dires ou des connaissances. Cette communauté de résidence a permis d'identifier directement certaines personnes comme des personnes âgées isolées. Ces sujets ne bénéficient en général pas d'aide-externe et ont un réseau d'aidants familiaux restreint, voire inexistant. Ces personnes ont ainsi directement été recrutées sur entretien et n'étaient pas accompagnées d'aide à domicile.

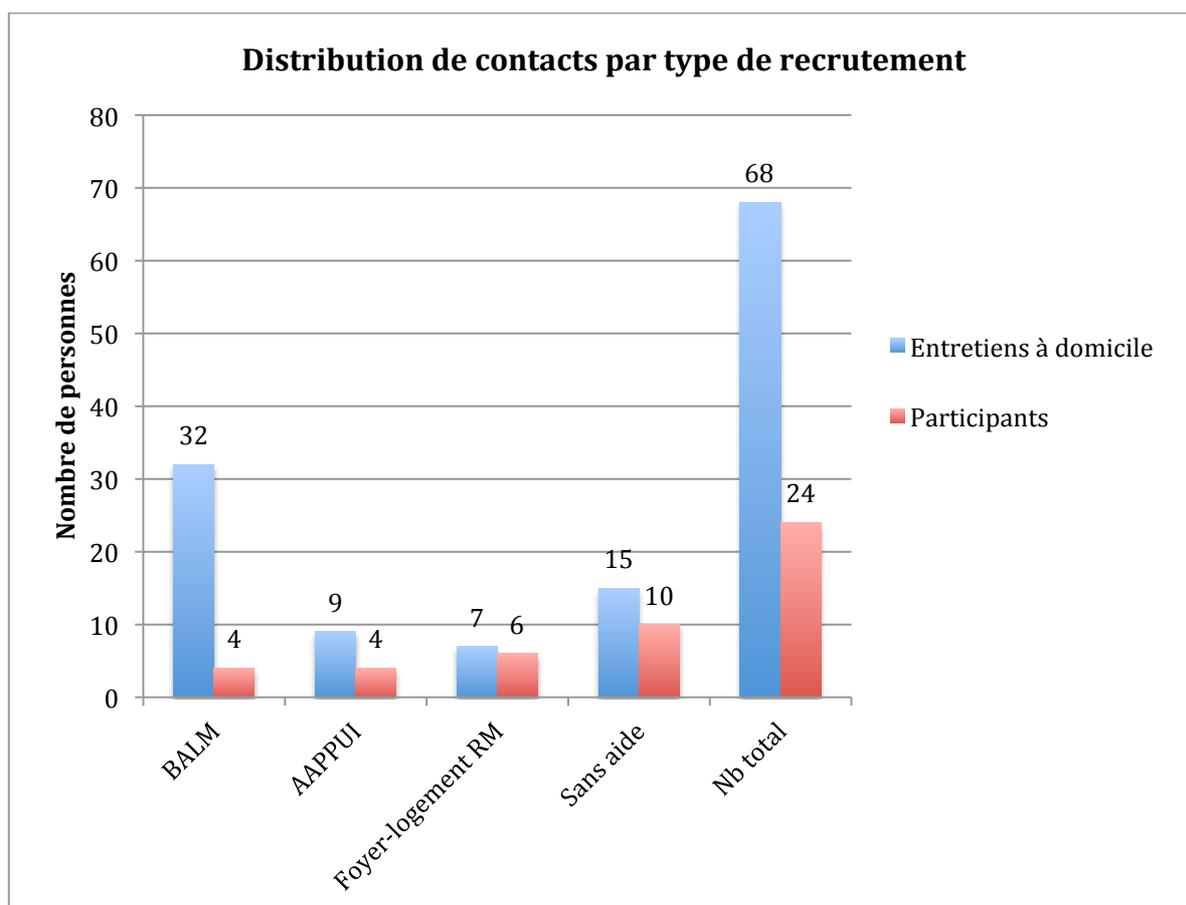


Figure 34 : Ratio de personnes entretenues/participant par origine de recrutement

Initialement, il était prévu d'avoir un panel de 30 à 50 sujets. Le recrutement a permis de voir différentes écologies et communautés, à une échelle très locale, mais de manière transversale. Il est passé à travers 173 contacts établis, dont 113 personnes âgées (soit par recrutements directs, soit par le biais des partenaires sociétaux). Parmi elles, 68 personnes ont été impliquées dans une démarche participative parmi lesquelles 24 ont effectivement pu participer aux expérimentations de la plateforme, mais seulement 20 ont suivi entièrement toutes les étapes de la méthodologie (dont l'étape d'auto-annotation) et 16 seront préférentiellement présentées dans les chapitres suivants. Cela laisse une cinquantaine de contacts représentant de potentiels partenaires sociaux locaux. Ces contacts sont des structures associatives et clubs (de personnes âgées, de retraités, de quartier, ou encore d'aidants – professionnels ou familiaux), les compagnies de services, les CCAS, les maisons des habitants, la CPA, les différents types de structures de prise en charge des personnes (de la traditionnelle maison de retraite, foyers-logements ou EPA jusqu'aux EPHAD) et enfin certains commerçants.

### 3.4.1.2. Gestion de la tâche prétexte dans les entretiens

Cette étape, réalisée sous forme d'entretien d'accueil (cf. Annexe 5), vise à : 1) comprendre la situation de la personne, sous la forme d'un récit de vie qui permet de se présenter mutuellement ; 2) à dévoiler une tâche prétexte et motiver la personne à participer à une étude factice qui l'oblige à se déplacer sur la plateforme expérimentale, pour enfin 3) organiser son déplacement sur la plateforme, en coordonnant la présence d'une aidante (l'aide à domicile ou les responsables foyers que les personnes âgées connaissent bien). La participation des aidants est entièrement bénévole et leurs implications font également l'objet d'un entretien soit individuel (dans le cadre d'une entreprise de services à la personne où les contraintes de planning individuel sont extrêmement importantes), soit collectif (dans le cadre associatif de services à la personne). Parfois, ce dernier critère de présence de l'aidant a été trop compliqué à mettre en place d'où la venue des personnes sans accompagnant. Par ailleurs, certains cas ont été repérés comme isolés, mais en couple. Notons que cette dernière situation implique une aggravation extrêmement importante de la fragilisation (p. ex. diminution des capacités immunitaires, changement des habitudes nutritionnelles) lors de la perte d'un des deux partenaires (Grimby, 1993 ; Rosenbloom et Whittington, 1993). Parmi les sujets, nous avons le cas d'un couple dont les partenaires sont venus ensemble sur la plateforme. Cependant, la majorité a fait l'objet du déplacement de l'une des deux personnes, en général l'aidant au sein du couple, dont l'isolement ressenti a également été perçu lors des entretiens.

Cet entretien de recrutement est généralement enregistré sur dictaphone, mis à part si la personne refuse de l'être explicitement. Ainsi 75 entretiens ont ainsi été réalisés dans ces conditions au domicile. Par défaut, les personnes refusant catégoriquement d'être enregistrées durant l'entretien ont été retirées de la liste des potentiels sujets, car le protocole de recueil de données ne pouvait vraisemblablement pas se faire au regard des contraintes imposées par le protocole. Ces données font partie d'un matériel exploitable pour une meilleure compréhension de l'isolement des personnes, mais très peu utilisé dans cette thèse.

Les sujets rencontrent ainsi, lors d'un entretien semi-directif, une « étudiante en gérontologie » qui fait une étude sur les pistes permettant de réduire les effets de désorientation d'une personne âgée qui est amenée à changer d'environnement de vie dans un contexte qui vont faciliter les actes de la vie quotidienne. En effet, la contextualisation englobe l'étude dans le contexte d'une favorisation de maintien à domicile par l'utilisation de technologies associées à l'habitat permettant d'aider la personne dans son quotidien. Sachant que la plupart des domiciles des personnes âgées sont des bâtiments trop anciens pour intégrer des technologies aidant le maintien à domicile. Ainsi les solutions actuelles, à des coûts plus réduits et sans surcharges exorbitantes, tendraient à construire des bâtiments entiers

initialement équipés des infrastructures nécessaires. Or dans certaines études montrent que le déplacement et le changement d'environnement est en soit un phénomène désorientant (Brittain, Corner, Robinson, et Bond, 2010 ; Burnette, 1986 ; Castle, 2001 ; Danermark et Ekstrom, 1990), remettant alors en question la possibilité de l'aide qui pourrait être apportée par les environnements dits intelligents. L'objectif prétexté de l'étude s'appuie donc sur une hypothèse plausible de réduction des effets de désorientation grâce des objets personnelles, telles que des objets transitionnels (Sherman et Newman, 1978 ; Wapner, Demick, et Redondo, 1990) à fortes valeurs affectives ou pratiques, que la personne choisirait elle-même et qu'elle viendrait placer dans cet environnement non familier. Cette fausse étude prétextée présenterait ainsi deux groupes dans le cadre de cette étude, l'un n'apportant pas d'objet, l'autre composé de personnes apportant des objets choisis, groupe dont font partie les sujets sélectionnés.

En termes de tâche à réaliser, les sujets âgés sont amenés à choisir une dizaine d'objets facilement transportables qu'ils listeront (en Annexe 5 la feuille support fourni aux sujets), et qu'ils rapporteront le jour de l'expérience pour les placer dans l'appartement intelligent témoin qu'ils utiliseront. Le temps de réflexion est de quelques jours à une semaine environ. Les emplacements où les personnes vont poser ces objets sont présentés comme importants pour l'étude et devront être rajoutés sur la liste des objets qu'ils auront remplis pour l'expérience. À cette étape, l'expérimentateur-recruteur doit « jouer un étudiant naïf ». Il prétend ne pas encore avoir eu l'occasion de visiter l'environnement que les personnes vont visiter. Il découvrira alors en même temps que le sujet les lieux de l'expérience. Les réservations de la plateforme ayant été réalisées par son encadrant, il ne connaîtrait pas non plus les ingénieurs responsables de la plateforme. Les expérimentateurs joueront donc ce jeu de la « première fois » pour chacun des sujets. De même, pour rassurer la personne, mais également pour des raisons sécuritaires, nous proposons que les personnes soient accompagnées de leur aide à domicile ou d'un aidant (non familial) si elle en bénéficie. Les aides à domicile accompagnantes peuvent s'associer au renforcement du prétexte durant la période de préparation, en aidant à choisir les objets et en augmentation la motivation à cette tâche. L'aidant peut également parler de sa sollicitation accrue en cette période, qui l'incite souvent à être appelé pour des cas d'urgence.

Par ailleurs, tout entretien de cette étude présente un risque non négligeable de confronter les personnes devant une réalité d'isolement. Celle-ci est plus ou moins conscientisée, voire pour les cas extrêmes, contournée sous la forme d'un déni. Or les compétences des recruteurs sont limitées pour prendre en main une telle situation, et ces derniers n'ont aucun droit ni légitimité d'attirer ni même soumettre cette situation. Pourtant, selon la capacité de sociabilisation des recruteurs, qui doivent tout de même réussir à établir un climat de confiance afin de solliciter la participation de ces sujets, un équilibre très délicat doit être géré. La confiance déliant les langues, les personnes qui se déclarent isolées commenceront à parler de leur situation spontanément à travers leur récit de vie. Les autres en parleront également à travers le vécu d'un autre que soi. Ces entretiens posent ainsi la question d'un encadrement et d'un approfondissement des méthodologies d'approche, du fait de la sollicitation de personnes fragilisées, non pathologiques, mais souvent confrontées à une grande souffrance.

La participation des personnes (âgées et aidantes) est entièrement bénévole. Ce critère est un moyen de maintenir un certain niveau de motivation sur la tâche et non par des formes de rémunération qui biaiserait la spontanéité du comportement des sujets et acteurs impliqués. La scénarisation des prétextes nécessite une mise en place spécifique, l'expérimentateur-recruteur véhicule l'aidant et la personne âgée à partir de son domicile jusque sur la plateforme expérimentale. Il gère également le retour du sujet (au domicile) et son accompagnant (lieu de travail) à la fin de l'expérience. En effet, cette modalité de déplacement est elle-même scénarisée pour avoir une entière maîtrise du contexte

expérimental. L'ensemble de ce matériel expérimental est disponible en Annexe 5, avec notamment : le formulaire de préconsentement, la fiche de rappel d'informations, la feuille de route pour les aidants, la fiche sujet de l'entretien, tous utilisés durant cette étape de recrutement.

### 3.4.2. Capture des données spontanées mais contrôlées

Les étapes qui suivent le processus de recrutement sont dédiées à la capture de données interactionnelles induites par le contrôle de l'écologie. Pour rappel, l'instrument choisi, pour générer les primitives langagières posées en hypothèse, est un robot – Emox. Par ailleurs, l'environnement d'interaction dans lequel va évoluer le sujet est la plateforme expérimentale, et plus particulièrement l'appartement intelligent à contrôle domotique. La première contrainte technique est ainsi de pouvoir contrôler aussi bien les primitives d'interactions du robot et celles de l'environnement. Un outil – EmOz – a donc été développé à travers un travail collaboratif conséquent, pour effectuer ce contrôle technique, dont la description fera l'objet de la section suivante.

#### 3.4.2.1. Plateforme EmOz

EmOz est une plateforme de magicien d'Oz adaptée, constituée d'une interface liant des actions décomposées d'événements domotiques de la plateforme Domus ainsi que ceux du robot (Aubergé et coll., 2014). Elle comporte d'une part ses actions vocales, d'autre part pour ses mouvements. Cet outil de contrôle est associé à un scénario motivé et orienté par une forte hypothèse scientifique afin de faire émerger une interaction spontanée et naturelle entre le robot Emox et ses interactants, en incluant spécifiquement les marqueurs permettant de mieux comprendre certains processus communicationnels qui sont la cible de l'observation. L'outil peut ainsi associer n'importe quelles primitives au robot. Les expériences pourraient même porter sur des stimuli audio d'autres langues à travers cette interface. En effet, EmOz a par exemple été utilisé sur des scénarii (*HireBot*) pour étudier les phénomènes de certitude ou incertitude en Brésilien (Antunes, Aubergé, et Sasa, 2014) ainsi qu'en Chinois<sup>84</sup>.

L'interface (cf. Figure 35), développée en Java, comporte un module de décomposition des primitives de communication (vocale et mouvements) du robot Emox, afin que ces derniers puissent être réassociés sans programmation, à travers des petits fichiers de description Excel. EmOz a ainsi été conçue pour permettre une utilisation par des non-programmeurs, pour faciliter l'autonomie et la collaboration avec de divers spécialistes les études pluridisciplinaires. Elle est également modulable et systémique pour que des informaticiens puissent reprendre certaines parties de programmation facilement afin de rajouter des dimensions de primitives qui ne seraient pas encore abordées. Elle fait le pont avec le *middle-ware* Open-hab<sup>85</sup>, en intégrant les protocoles récoltant les données de différents capteurs et activant des actionneurs domotiques de l'appartement intelligent. Tous ces éléments sont également contrôlables à partir de cette même interface. La fin de l'Annexe 3 présente l'architecture générale de l'outil, avec la visualisation des fichiers d'entrée et de sortie nécessaires pour sa génération.

Tous les boutons de l'interface peuvent être déplacés par simple drag-and-drop en mode inactif. Les coordonnées de ces boutons sont actualisées dans un fichier spécifique et

---

<sup>84</sup> Travaux de Lu Yan, qui est une version traduite au Chinois de (Batista Antunes, Aubergé, et Sasa, 2014) mais dont les travaux ne sont pas encore publiés.

<sup>85</sup> Présentation lors de la Conférence EclipseCon 2015 - Toulouse, (<https://www.eclipsecon.org/france2015/user/3223>) explicitant la partie technique de cette intégration : “*Ruling the experiment how openHAB really empowers Living Lab platforms*” par Nicolas Bonnefond, accès sur : <https://www.youtube.com/watch?v=pKI9lBiNXgw>

personnalisable, pour garder en mémoire une interface individualisée ou pour des usages d'analyses ergonomiques. Ces aspects de l'interface sont motivés par un besoin méthodologique essentiel : réduire au maximum les délais temporels liés au matériel technique afin que le temps de réaction du robot puisse se rapprocher au plus de la réaction résultant du temps cognitif humain. Il n'est tout de même pas possible d'égaliser ce temps du fait des contraintes techniques (c.-à-d. usage d'une interface avec de multiples boutons, utilisation d'un réseau pour l'envoi des commandes d'actionneurs dont la qualité peut être variable, connexion wifi du robot etc.). Les boutons sont pour la majorité reproduits par des fichiers de scripts (format intuitif nécessitant uniquement la saisie de paramètre pour créer une primitive), permettant de réaliser de manière simultanée (p. ex. un geste et un son sont produits en même temps sur le robot) ou décalée (p. ex. le mouvement du robot est réalisé avant le son qu'il émet) les primitives qui en font l'objet. L'Annexe 3 présente la nomenclature de ces scripts.

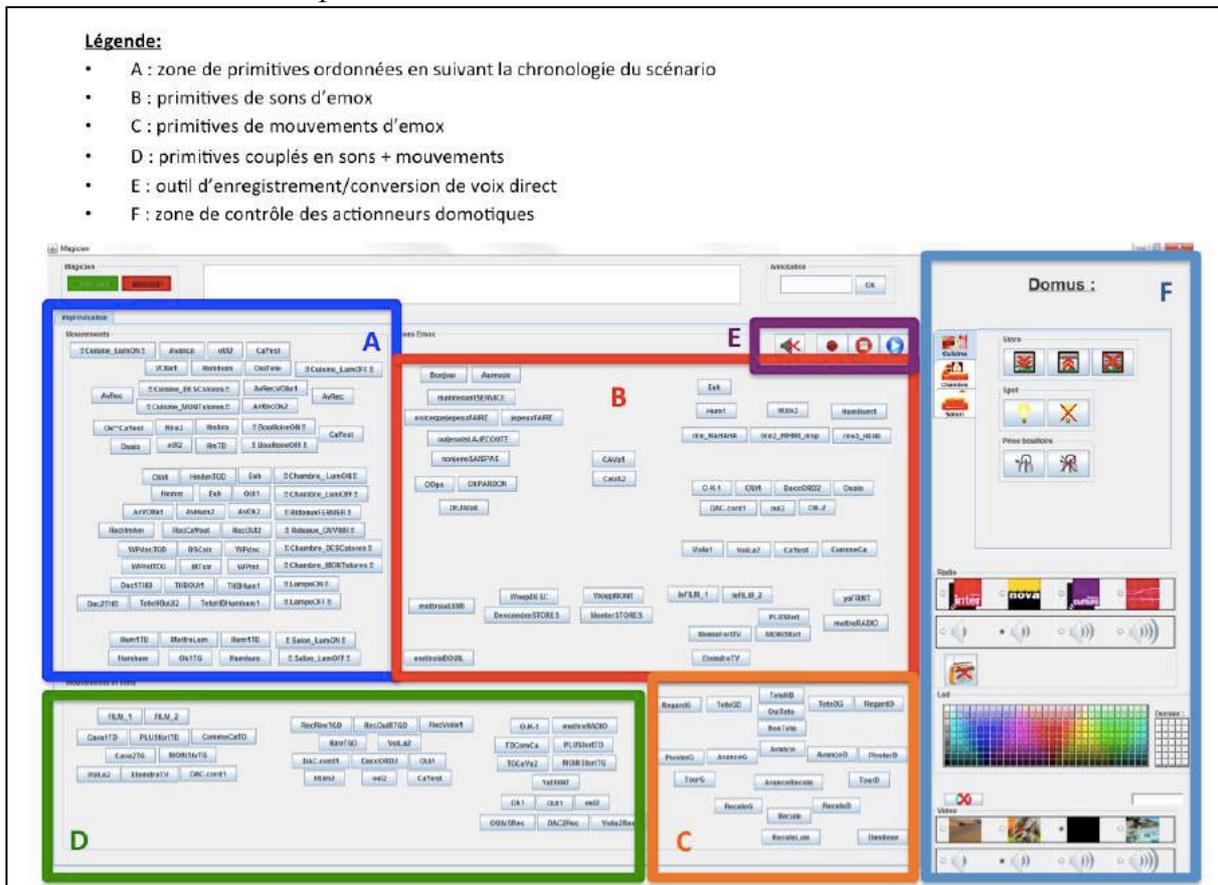


Figure 35 : Illustration de l'interface du magicien EmOz

La zone A propose des primitives simples (gestuelle et vocale du robot ou domotique) ou composées (une combinaison de deux ou trois primitives simples). Elles sont arrangées dans un ordre précis, afin de reproduire la succession de primitives graduelle à chaque expérience. Outre cette zone scénarisée, nous avons prévu des zones « d'improvisation » (B, C et D) permettant de s'adapter aux réactions du sujet afin d'assurer la continuité interactive si des événements imprévisibles survenaient. En revanche, « improvisation » n'est pas synonyme d'exécutions de réactions libres. La gradation de primitives étant l'hypothèse testée, celle-ci est respectée, quelle que soit la situation dans l'étude EEE (cf. section 3.4.2.1.3 pour les règles spécifiquement suivies dans le scénario).

Cette partie comporte des zones dont les règles d'organisation ne dépendent que du scénario expérimental, donc elles peuvent (y compris la zone A) être utilisées sans contraintes architecturales. L'improvisation comporte ainsi :

- les boutons de la zone B représentent l'ensemble des primitives sonores du robot. Ils sont générés automatiquement à partir d'une liste de sons compris dans une base de données de sons du robot, permettant l'ajout/suppression aisée de ces sons.
- les boutons de la zone C sont les primitives gestuelles du robot.
- les boutons de la zone D sont des primitives composées qui associent des mouvements et les sons du robot.
- la zone E présente un outil d'enregistrement de voix avec convertisseur (celle d'un des magiciens qui est initialement paramétré sur un rapport de pitch rendant la même esthétique vocale que les autres stimuli préenregistrés) et retransmission sur le robot ;
- enfin la zone F, iconisée, représente les primitives domotiques de l'habitat intelligent lançant les commandes aux actionneurs domotiques.

Les événements d'interaction et de contexte apparaissant sur cette interface sont marqués temporellement et produisent des traces d'interactions de l'interface conservées au format.csv. Toutes les sources d'entrées et sorties ne sont pas forcément installées sur le même serveur du fait des contraintes techniques et multisource, ainsi l'ensemble de ces données, y compris les sons et vidéos enregistrées, sont marqués par des temps de référence, pour une synchronisation par un alignement sur le *timestamp Unix* des différentes machines, dont l'heure est paramétrée sur un serveur unique. À noter, les ordinateurs ont un temps-machine propre à chaque outil. Cette question impliquant une dérive temporelle n'est pas anodine, car les données doivent pouvoir entièrement être alignées sur une même base temporelle, ce qui a fait l'objet d'un post-traitement pour être corrigé (cf. Annexe 5). Sans cette resynchronisation, les interactions ne sont pas abordables, car les contraintes d'analyse de la parole sur des événements précis descendent souvent à des ordres de grandeur entre des dizaines de millisecondes et la seconde.

La liste des stimuli de primitives par scripts sur l'interface et la description technique que la génération des primitives est présentée en Annexe 3. Le scénario scientifique de présentation de ces stimuli est quant à lui disponible Annexe 4. Scénario de capture du corpus EmOz Elderdy Expressions

L'outil technique n'a de sens que s'il est associé à des scénarios, comme la section 3.3.3.1 le présentait. Ainsi la proposition du scénario expérimental essaie de tirer partie d'une part de l'essor des réflexions autour des environnements intelligents, en gérontechnologie notamment. Elle contextualise l'expérience dans un Smart Home dans le cadre d'une réflexion générale de soin, en faveur d'un maintien à domicile idéalement prolongé, favorisant l'autonomie des personnes âgées. D'autre part, elle s'appuie sur l'émergence soutenue de la robotique sociale et de services dans l'optique d'utiliser un tel objet comme outil d'étude de phénomènes interactionnels et communicationnels dans ce type de situation. Le cadre expérimental propose le robot Emox doté de primitives graduelles et « socio-affectivement gluant » comme majordome d'un habitat intelligent où les commandes domotiques sont adressées vocalement au robot. Ce dernier donne ainsi l'illusion de les exécuter et il répond par des feedbacks de primitives après chaque commande. Les interactions sont observées auprès de personnes âgées de plus de 75 ans, en isolement relationnel et fortement motivées pour rester autonome chez elles, en retardant une éventuelle entrée en institution. Méthodologiquement, cela relève d'une mise en scène construite et préparée, dont la Figure 36 ci-dessous explicite les différentes étapes par lesquelles un sujet passe.

## Méthodologie et protocoles expérimentaux

### Légende :

**E1** : chercheur-ingénieur de plateforme

**E2** : chercheur-recruteur

**PA** : la personne âgée – sujet de l'expérience

**AD** : aide à domicile de la personne âgée

Dans les phases expérimentales, les astérisques (\*), la nature des contrôles de primitives du robot. Un code couleur indique d'une part le type de primitives pris en charge du côté expérimentateurs-régie, et il indique d'autre part la nature des stimuli autorisés par le scénario du côté du robot :

\* contrôle des déplacements du robot (contrôle manette)

\* contrôle des primitives du mouvement du robot (contrôle interface EmOz)

\* contrôle des primitives vocales du robot (contrôle interface EmOz)

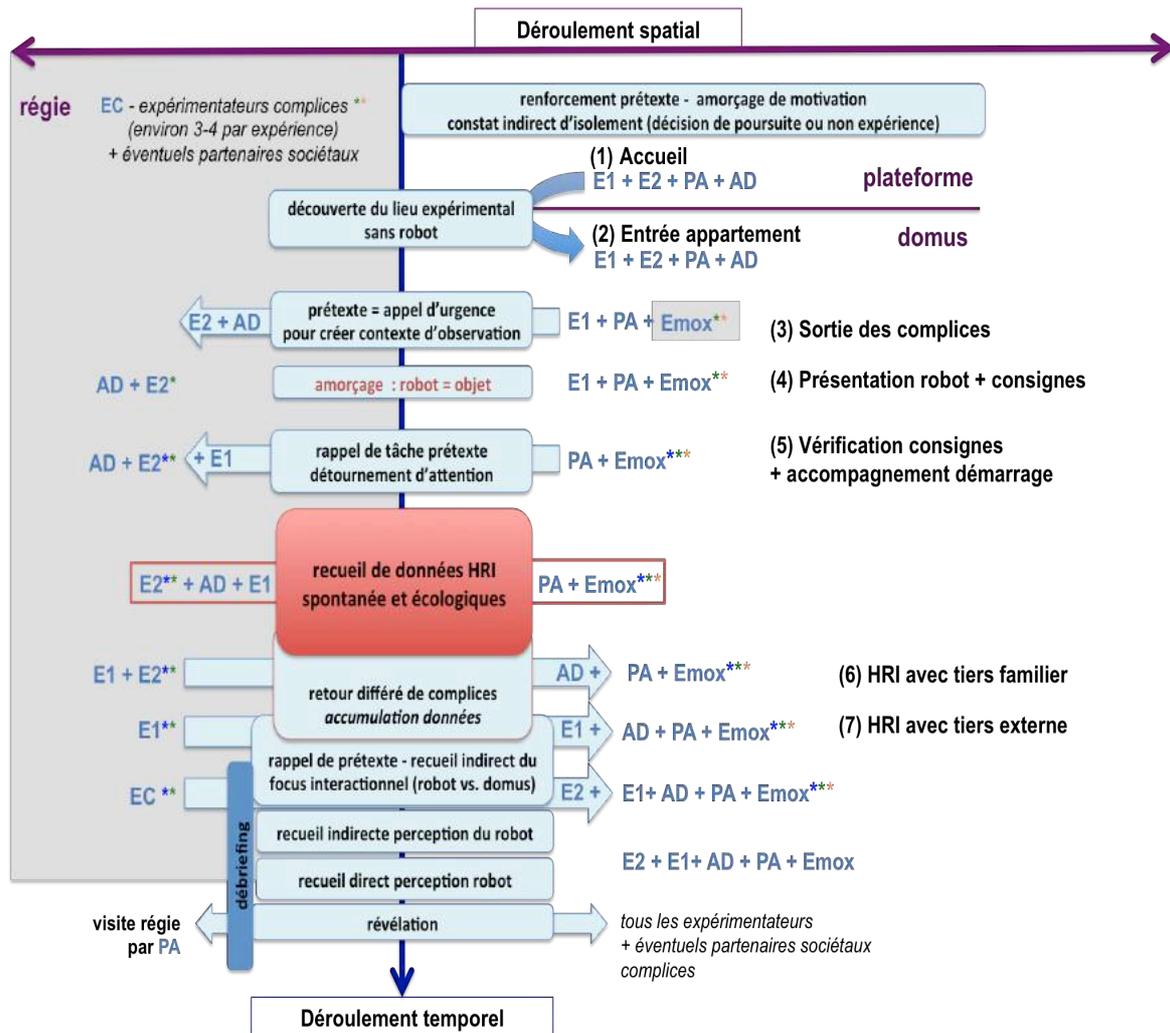


Figure 36 : Schématisation spatio-temporelle du scénario EEE

Le jour de l'expérimentation, le sujet (PA) est véhiculé par le recruteur (E2) en compagnie de son aide à domicile (AD). Ils arrivent sur les lieux de l'expérimentation où ils sont accueillis par un ingénieur responsable la plateforme (E1). Le scénario de mise en scène détaillée, qui décrit les parties suivantes de l'expérience, est présenté en Annexe 5. Les notions clés qui motivent les choix du format expérimental sont ensuite explicitées par différentes étapes à travers les quatre sous-sections suivantes.

### 3.4.2.1.1. Accueil : déviation d'attention et indices d'isolement

L'accueil est mené sous forme d'un entretien semi-directif introductif dirigé par E1. Il présente plusieurs objectifs : a) rassurer la personne âgée sur l'environnement, en contextualisant le cadre général de l'expérience ; b) faire un rappel et insister sur la tâche prétexte ; c) obtenir des informations complémentaires qui puissent aider à caractériser l'isolement du sujet de manière indirecte ; et d) potentiellement décider de la possibilité ou non de continuer l'expérience si des fragilisations trop importantes susceptibles de perturber la personne sont détectées durant cet entretien.

Avant l'entrée dans l'appartement, les sujets passent donc par cette phase d'accueil sur la partie plateau, qui s'inscrit dans le prolongement du recrutement. Il s'agit d'un *pitch* qui doit permettre au sujet de commencer l'expérience, avec une motivation liée à la tâche. L'entretien est menée de manière détournée par E1, qui feint une certaine « curiosité » qui le pousse à poser des questions sur l'objet de la fausse étude gérontologique. En effet, il prétend ne pas connaître les raisons pour laquelle sa plateforme a été réservée.

Cette phase fait l'objet d'un enregistrement vidéo et audio par une caméra de plateau ainsi que d'un complément d'enregistrement sur Tascam. Cet enregistrement est cependant masqué tout comme la suite de l'expérience, où les sujets ne savent pas qu'ils sont enregistrés jusqu'à l'étape du débriefing.

### 3.4.2.1.2. Entrée dans Domus : la question de l'amorçage

Il s'agit de l'entrée dans l'appartement intelligent. Elle est progressive et permet la mise en place du prétexte secondaire permettant de laisser la personne seule pour une interaction exclusive avec le robot. L'absence des accompagnants doit être introduite de manière réaliste. En effet, une fois dans l'appartement, E1 commence à présenter les différentes pièces sans faire mention des technologies existantes. Très rapidement, des expérimentateurs de régie appellent AD sur son téléphone portable en jouant le rôle de l'employeur d'AD tout en lui donnant la réplique, pour la solliciter sur une « fausse urgence ». Ces cas d'urgence, le sujet en avait entendu parler à travers le conditionnement que AD avait effectué entre le recrutement et le jour de l'expérimentation. Il est important qu'à ce niveau de l'expérience, E2 et AD n'aient pas vu le robot en fonctionnement pour pouvoir le faire à leur retour. Ainsi, comme c'est E2 qui a véhiculé AD, il n'a d'autre moyen que d'accompagner AD. En revanche, le sujet est prévenu que le retour des accompagnants sera rapide, et que cela n'empêche pas le début de la fausse expérience. E2 rappelle alors une tâche prétexte au sujet, celui de commencer à placer ses objets dans l'appartement, comme le voulait la première étape de la fausse expérience gérontologique. En sortant de l'appartement, AD et E2 rejoignent donc la régie et E1 enclenche une phase d'amorçage.

En effet, la manière dont le robot est présenté fait l'objet d'un effet d'amorçage ou de *priming*, qui est un phénomène étudié dans le cadre de l'attention sélective en particulier (Carlson, Charlin, et Miller, 1988 ; Neumann et DeSchepper, 1991 ; Joule et Beauvois, 2004 ; Mayr et Buchner, 2007). Ce phénomène va venir influencer les interactions que la personne va produire selon la nature de l'amorce. Dans le cadre de cette expérience, un réglage a été effectué, suite à l'observation d'effets que pouvait avoir le comportement de E1 vis-à-vis du robot sur l'interaction de la personne âgée avec le robot. En effet, dans le cas d'une amorce positive (focus personnalisé et positif sur le robot, modèle de forte attache relationnel) l'interaction progresse très vite et positivement, ce qui biaise l'effet des primitives langagières sur les personnes. Dans le cas d'une amorce négative (un focus sur l'aspect purement matériel et technologique sans intelligence avec adjectifs associés – humeur négative, sans aucune marque d'attention au niveau du regard), l'interaction ne progresse pas ou très peu et change la dynamique induite par les primitives du robot. L'amorce reproduite pour les expériences a donc été neutralisée au maximum en respectant les règles suivantes : a) présenter Emox

comme une télécommande vocale de l'appartement ; b) ne pas utiliser le terme de « robot », mais le désigner l'objet par « cette chose – ça » et n'évoquer son nom « Emox » que si le sujet le sollicite, mais présenter comme un mot-clé non affectif ; c) ne pas donner d'exemples de commandes à l'oral sur la manière dont on peut s'adresser au robot pour que les productions du sujet soient celles émergeant naturellement et spontanément. Durant cette phase, le robot n'émet encore aucun son, ses seules manifestations sont un déplacement limité, et très peu de mouvements.

### 3.4.2.1.3. Phase d'interaction personne-robot

Il s'agit de la partie principale composant le corpus d'analyse d'événements interactionnels qui se déroule au sein de l'appartement Domus. Elle commence avec l'amorçage qui introduit Emox. Cet outil est présenté par E1 comme la télécommande vocale de l'appartement qui lui permettra d'utiliser cet appartement, dont le fonctionnement est particulier, à savoir que les interrupteurs ne fonctionnent pas et que tout contrôle se fait par la voix. Le sujet est alors équipé d'un micro-casque, dont la distance à la bouche reste fixe pour pouvoir effectuer des mesures acoustiques fines comme l'amplitude ou de l'intensité du signal vocal du sujet, qui sont des paramètres descriptifs de la prosodie. En revanche, un nouveau prétexte est ici utilisé, pour justifier ce micro comme un moyen de communiquer avec le robot. En effet, il est prétendu que ce dernier est construit sur une technologie bon marché qui nécessite un paramétrage pour que le robot « apprenne » à reconnaître la voix du sujet, comme le proposait les anciens systèmes de dictée vocale par exemple. Ainsi, une tâche secondaire complémentaire se rajoute à la tâche prétexte du sujet. E1 demande au sujet de produire, au moins une fois, une liste de 30 commandes « test » qui équivaut à un mode d'emploi de l'appartement, et qui permettra ce paramétrage vocal. E1 reste avec le sujet pour l'exécution de la première commande pour s'assurer que le sujet a bien compris la consigne, puis il laisse le sujet en autonomie. En effet, le rôle de E1 est de mettre à disposition l'appartement, mais comme cela a été introduit durant la phase d'accueil, il est externe à cette étude gériologique et il a donc d'autres tâches à accomplir par ailleurs. E1 rassure néanmoins avant son départ le sujet, quant à la possibilité de le rappeler à n'importe quel moment, à travers le robot. Il rappelle également le prétexte en mentionnant qu'une fois que cet apprentissage est terminé, le sujet a la possibilité de disposer librement de cet appartement comme il le souhaite et donc de disposer les objets comme lui avait mentionné E2. La Figure 37 résume le schéma d'acquisition/contrôle des interactions qui suivent cette mise en situation.

Cette phase permet donc de dérouler le scénario scientifique des primitives vocales et comportementales du robot, puisque le magicien devait suivre un ordre précis de présentation des primitives pour la reproductibilité de l'expérience. En effet, c'est la gradualité du contrôle prosodique socio-affective qui est supposée manipuler l'hypothèse de glu testée (cf. Annexe 4). En revanche, suite à l'évolution et aux conditions de l'expérience, ce scénario scientifique a rapidement dévié au cours des expériences, du fait de la naturalité et la spontanéité du rythme des interactions, car elle implique un magicien humain. En effet, l'expérimentateur qui contrôle le robot fait l'objet d'une adaptation inévitable face au panel d'objets interactionnels mis à disposition. Sachant que ses traces d'actions sont conservées à travers EmOz, les influences du sujet sur l'expérimentateur lui-même ont donc été privilégiées, par opposition à un suivi stricto sensu du scénario scientifique. En effet, ce scénario aurait pour effet d'induire plus d'artefacts dans l'interaction observée, alors que l'expérimentateur est inclus dans ce système interactionnel à travers le robot.

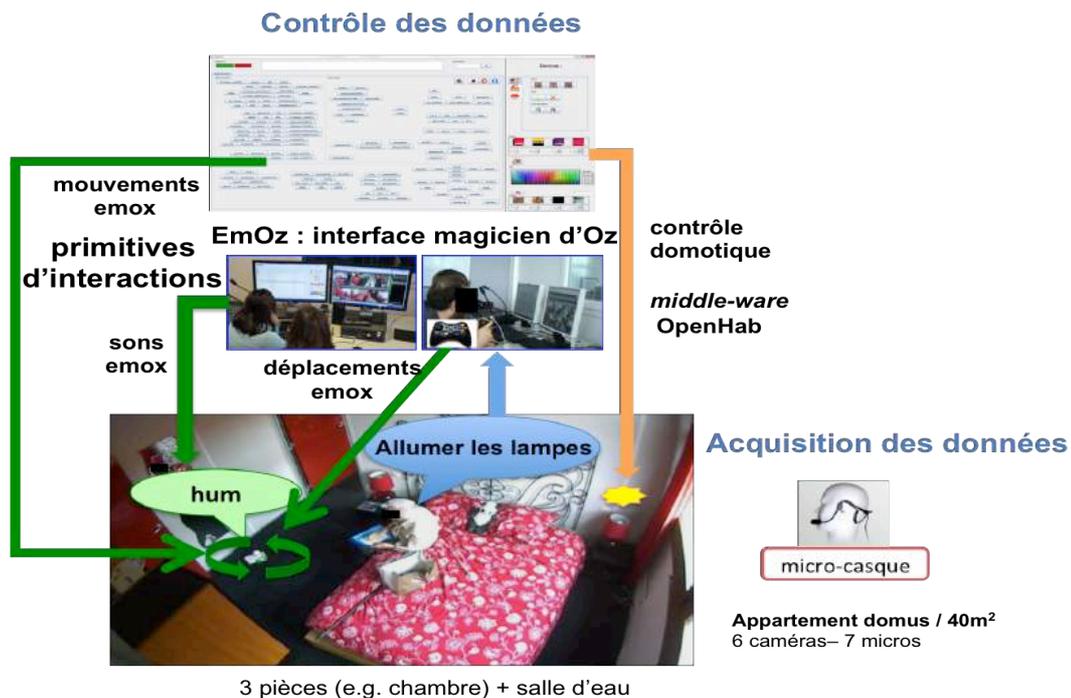


Figure 37 : Schéma de contrôle/acquisition des données induites à travers la plateforme EmOz

En revanche, pour conserver un minimum de cadrage, la nature des stimuli est présentée en suivant les règles ci-dessous :

- La gradation des différents niveaux de primitives doit être respectée, même si le rythme de présentation et la quantité de stimuli d'un niveau socio-affectif donné sont variables d'un sujet à l'autre. De même, pour l'ordre de présentation, si des stimuli font partie d'un même niveau de contrôle prosodique, ils peuvent s'interchanger (ceci à un effet de randomisation dans un même groupe de primitives). En revanche les stimuli, de différents niveaux de gradualité ne peuvent se remplacer mutuellement.
- Il est possible de raccourcir la phase d'un niveau pour passer au niveau de contrôle prosodique supérieur si un décrochage du sujet est perçu. Cela a d'ailleurs été un comportement qui s'est naturellement imposé au magicien, induit par la naturalité de l'interaction.
- Privilégier les stimuli de bas niveau et les stimuli de mouvements dont les valeurs sont moins explicites lorsque l'interaction est interrompue volontairement par le sujet. Ces primitives sont émises pour solliciter le sujet à nouveau. De plus, ne passer à des niveaux supérieurs que si aucun changement n'est visible et que les interactions stagnent. De même, n'utiliser que des éléments de bas niveau si le sujet tente de rentrer dans un échange conversationnel, plutôt que d'utiliser l'outil de conversion de voix pour rentrer dans ce dialogue.
- Ne pas utiliser l'outil de conversion de voix en direct, sauf si le contexte induit un besoin pratique réel ou une incompréhension de tâche par le sujet, pour éviter de couper l'interaction en cours. Si vraiment la condition génère un blocage, les magiciens simulent une « panne du système » qui permet à E1 d'intervenir. Ce dernier va prétexter que le système lui a renvoyé une alerte de panne et qu'il vient vérifier que tout fonctionne correctement. Cette situation peut-être sollicitée en cas de problème technique (p. ex. panne des actionneurs domotiques, déconnexion du robot etc.).

### 3.4.2.1.4. Retours des complices et débriefing en deux phases

Tant que l'interaction évolue, la phase HRI se poursuit. En revanche, dès que le prétexte d'urgence utilisé par AD vient à sa limite temporelle (en cohérence d'une situation réelle et en concertation avec le complice), ou que le sujet arrive à un point mort montrant des signes importants de lassitude, ou encore des patterns d'interaction similaires, l'expérience passe à une étape de « retour de complices ». En effet, le retour des complices continue de faire partie de la phase expérimentale et elle s'effectue en deux vagues par l'inclusion de tiers dans la situation d'interaction. Elle est initiée par AD. Cette personne représente le cas d'un tiers qui est une connaissance familière du sujet. AD demande alors si tout s'est bien passé, en rassurant le sujet, et il s'excuse de ce départ fortuit tout en faisant un rappel de la tâche prétexte, en demandant si le sujet a pu placer ses objets. Elle simule la découverte du robot sans insistance et demande au sujet de lui montrer ce qu'elle a fait durant son absence. Puis après un certain temps d'interactions, E1 revient, qui lui représente un tiers connu, mais non familier du sujet. Il fait son retour de la même manière que AD. Il justifie notamment le décalage de son retour par le fait qu'il ait été retenu par E2, un prétexte en cohérence avec les propos de AD au moment de son entrée dans l'appartement. Ce retour en deux vagues permet d'une part d'accumuler des données de type « commandes vocales », où la présence de tiers est susceptible de modifier les manières dont le sujet les effectue. Cette variabilité présente en soi un intérêt pour les corpus d'apprentissage des systèmes de reconnaissance automatique de parole qui doivent être robustes à ces phénomènes. D'autre part, elle permet de recueillir les indices potentiels pouvant illustrer l'état relationnel, soit la forme de la glu, qui s'est établie entre le robot et la personne âgée, en l'observant par le biais des interactions en présence de ces tiers.

L'expérience se termine par le débriefing. Elle est initiée par E2 qui commence par des questions générales sur l'appartement et sur le robot, sans focus spécifique sur ces éléments, juste par « curiosité ». Cependant, il revient très vite sur le prétexte des objets, cette transition donnant un repère pour le timing d'entrée de E1. Son retour permet alors de rentrer plus explicitement dans ce débriefing. Cette phase est menée sous la forme d'un entretien semi-directif. Elle est décomposée en deux temps : a) un questionnement, qui ne fait aucun focus sur le robot. Il révèle par un dernier prétexte que l'expérimentation a eu lieu pour tester l'appartement et sa praticité, ceci permettant d'induire un avis non conscientisé du robot si ce dernier est évoqué spontanément par le sujet ; b) un questionnement avec focus sur le robot. Il présente alors l'ensemble des prétextes y compris le dernier en clarifiant totalement les véritables objectifs de l'expérimentation. Il s'agit de la « levée de voile » introduite dans la section 3.3.3.2. L'isolement ne fait jamais partie des questions explicites. En revanche, la caractérisation de la situation d'isolement des sujets peut être complétée par les propos recueillis durant cette phase, à travers les projections évoquées sur son quotidien, ou par l'évocation de personnes à travers lesquelles le sujet pense ce robot utile. La feuille de route de cette étape se trouve en Annexe 6.

### ***3.4.3. Auto-annotation : les contraintes d'adaptation pour la personne âgée***

#### **3.4.3.1. Choix et conditions de l'auto-annotation**

Le protocole d'une auto-annotation est spécifique à chaque cas de recherche, car elle dépend des hypothèses théoriques à tester qui permettent d'orienter et de contrôler la tâche du sujet. De plus, selon les caractéristiques des sujets, le format du protocole varie pour s'adapter aux personnes. Il essaie de réduire le surplus de charges cognitives associées, car l'annotation en soi demande un effort cognitif non négligeable d'accès à la mémoire et sur des séances relativement longues. De manière générale, elle s'effectue sous forme d'un entretien semi-directif outillé.

Le mode de recueil peut également varier selon l'âge des sujets ou encore la nature ou le degré de finesse attendu sur les étiquettes. Ainsi une réflexion est à mener sur : 1) les informations à recueillir, 2) le moment (c.-à-d. la durée entre le jour de l'expérience et le jour de l'auto-annotation) et le lieu de l'entretien (p. ex. au laboratoire, domicile ou autre environnement d'acquisition de données), 3) les outils à utiliser pour le visionnage (p. ex. taille de l'écran et formats de vidéo à montrer), 4) la modalité de transcription/enregistrement des données (p. ex. nombre d'annotateurs, enregistrement audio/vidéo de la séance, le matériel utilisé) avec d'éventuelles tâches supplémentaires (p. ex. reproduction de certains stimuli, imitation, schématisation, dessin... etc.).

Dans le cadre du corpus EEE, l'auto-annotation s'effectue en moyenne entre un à trois mois après le passage sur la plateforme expérimentale. L'expérience étant un changement notable dans le quotidien de la personne, cet intervalle de temps permet d'une part d'atténuer les émotions directes liées à la situation expérimentale, sujettes à favoriser des interprétations motivées. D'autre part, il favorise l'accès aux souvenirs les plus marquants pour la personne et donc supposés également les plus résistants au temps.

Une séance d'auto-annotation dure en moyenne entre deux et quatre heures pour la plus longue. Cette durée est variable en fonction bien sûr de la durée de l'expérience dans Domus mais aussi de la densité d'informations liée à la richesse des interactions du sujet avec le robot : le nombre d'interactions plus ou moins important, les multiples changements d'état du sujet, la présence d'éléments plus ou moins perturbateurs dans la situation d'interaction sur l'état du sujet (p. ex. pannes, appels téléphoniques... etc.), ou encore des spécificités de perception très personnelles pouvant donner lieu à des différences interpersonnelles marquées. Cette dernière variation est en revanche l'un des enjeux d'uniformisation des annotateurs. La séance s'est effectuée au domicile de la personne, préférentiellement avec deux annotateurs : un interrogateur et un transcripteur, mais nécessairement au moins un interrogateur. Toutes les séances sont enregistrées sur un dictaphone classique.

#### **3.4.3.2. Prétraitement des données**

Les sujets s'appuient sur les vidéos de l'expérience afin de pouvoir réaliser une auto-annotation, ce dernier se déplaçant dans l'appartement qui est doté de six caméras, d'angles de vue différents. Il est donc cognitivement lourd de se repérer sur l'ensemble de l'expérience, à travers ces six petits écrans. De plus, la taille de l'image est en soi problématique pour un public qui présente majoritairement une perte d'acuité visuelle, plus ou moins importante. Un premier traitement est donc d'avoir une seule vidéo zoomant sur les angles de vue où le sujet et le robot sont bien visibles, et agrandir les images au maximum sans perdre trop de qualité, pour faciliter le visionnage des données d'expérience.

Par ailleurs, le type des caméras de la plateforme et l'infrastructure technique impliquent que toutes les données des sujets soient capturées sur des canaux d'enregistrement

individuels et séparés. Les vidéos et les sons ne sont donc pas initialement couplés ensemble. Il est alors possible d'associer les éléments de deux manières :

- soit les vidéos sont associées avec les canaux de microphones de l'appartement. Chaque canal audio est alors couplé avec la vidéo de la caméra qui est le plus proche du microphone d'acquisition concerné.

- soit les vidéos sont associées avec le canal du micro-casque que portait le sujet sur lui, qui donne une bonne qualité de ce que la personne dit, en revanche étouffe énormément les bruits ambiants, dont les feedbacks d'Emox, alors qu'ils sont justement supposés être des déclencheurs de comportements interactionnels chez le sujet.

Les deux sont fondamentalement nécessaires pour l'analyse des données par la suite, mais pour l'auto-annotation, ce sont les données sur le premier couplage qui ont été choisies, ces derniers permettant d'entendre plus facilement les bruits environnants, favorisant la réminiscence des souvenirs à l'appui de la mémoire autobiographique des sujets (cf. Annexe 7 pour les détails de paramétrage).

De plus, l'acuité auditive suivant les mêmes contraintes que celles du visuel, il existe deux versions des vidéos associées aux canaux des micros ambiants : d'une part une version originale sans traitement qui pourrait être directement utilisée comme données d'entraînement pour un système de reconnaissance automatique et qui fonctionnerait par ces mêmes micros le cas échéant. D'autre part une version « nettoyée » qui consiste : 1) en un dé-bruitage partiel (p. ex. de bruits de fond de ventilations, grésillements fortuits), et 2) une amplification du son pour faciliter la séance d'auto-annotation (ceci est réalisé avec le logiciel Audacity). Malgré ces modifications, l'ajout d'enceintes et parfois même l'utilisation de casques a été nécessaire pour certains sujets.

D'autre part, une auto-annotation propose un visionnage pas-à-pas de la vidéo et elle nécessite de pouvoir répéter une séquence dans certaines conditions. Ces manipulations doivent donc être facilitées au maximum, sachant qu'il est ergonomiquement impossible de les effectuer sur les lecteurs multimédias classiques. Afin également d'éviter d'oublier des commandes, le prétraitement prévoit un séquençage, initialement arbitraire, d'intervalles qui englobent une commande avec ce qui la précède et ce qui la suit (cf. Chapitre 4, section 4.2). L'enchaînement ininterrompu des séquences permettra ainsi une présentation chronologique, en donnant la possibilité de revoir une même séquence plusieurs fois, tant que la description de cette dernière n'est pas terminée. Le protocole privilégiant le découpage temporel défini par le sujet, le séquençage est à modifier à travers cette procédure et ne représente qu'une aide provisoire. Enfin, la vidéo annotée concerne une partie qui a été nommée « Expé » dans le corpus EEE. À l'appui du scénario d'ensemble tel qu'il est illustré Figure 36, la vidéo commence par quelques minutes introductives (moins de deux minutes) montrant la fin de l'amorçage pour aider les sujets à se remémorer le contexte de l'expérience après leur entrée dans l'appartement, une fois que les accompagnants (AD et E2) sont sortis de l'appartement. En revanche, nous avons la présence de E1 afin de conserver la présentation du robot ainsi que la production de la première commande. Sans aucune coupure, cette vidéo illustre donc la situation dyadique entre le robot Emox et les personnes âgées. Elle se termine par le retour des complices (AD et E2) et s'arrête à l'entrée de E1 qui marque le début du débriefing.

### 3.4.3.3. Déroulement de la séance

Une feuille de route de séance dédiée aux expérimentateurs est disponible Annexe 7. En voici les points clés :

Le *transcripteur* se met en arrière-plan lors de la séance et guette les potentiels descripteurs associés aux séquences. Il essaie de transcrire les impressions du sujet avec les mots exacts que ce dernier emploie, en triant les informations utiles ou non. Il suit les séquences prédécoupées qui sont visionnées au pas-à-pas par le sujet sur son propre écran. Il aligne en direct les descriptions sur chaque séquence. La nature des informations fournit par

le sujet ne visant pas à suivre exactement les frontières de séquences, et va même potentiellement se chevaucher entre elles. Le découpage temporel est donc idéalement effectué en direct, comme il est décrit par le sujet, mais cette tâche reste complexe le transcripteur ne peut interrompre intempestivement la séance. Il tente donc de rajouter les informations marquant des événements précis au sein de chaque séquence pour associer les étiquettes aux bons endroits afin de marquer les frontières d'annotations, et cela, les plus précisément possible. Afin qu'il puisse revoir ou vérifier les séquences de son côté, il peut utiliser des écouteurs pour éviter de perturber la séance tout en lui facilitant cette tâche.

L'*interrogateur*, quant à lui, suit une procédure, quasi-maïeuticienne. Il présente les séquences une par une, en suivant le même écran que le sujet. Sa tâche est de faire émerger les descripteurs permettant de répondre aux questions de recherche de l'étude, sans jamais proposer de valeurs d'étiquettes. De cette manière, le sujet est invité dans une méta-tâche dans laquelle il essaie de se souvenir des *Feelings of Thinking* auxquels il a été confronté au moment même où il réalisait les actions ou non-actions qu'il visionne. L'interrogation n'est cependant jamais explicite quant à la nature ou la valeur du FoT. C'est à l'interrogateur de décider alors si les informations attendues ont été données par le sujet et de le solliciter répétitivement tant que les commentaires du sujet ne contiennent pas des descripteurs concordants à la nature des informations recherchées. L'Annexe 7 donne une feuille de route quant à la manière d'introduire une séance et les exemples types de questions qui peuvent être posées par un interrogateur pour faciliter la démarche.

Les annotateurs utilisent ainsi deux interfaces séparées sur le logiciel d'annotation vidéo Elan (Brugman, Russel, et Nijmegen, 2004). Une première, qui est présentée au sujet est simplifiée au maximum. Elle permet de revoir de manière simple et rapide la séquence en cours (notamment pour réduire le coût cognitif de la tâche). La seconde interface, dédiée à l'annotateur-scripteur, présente les mêmes séquences que celles de l'interface sujet, mais elle est indépendante de la première pour permettre d'afficher des tiers séparés, voire de modifier les frontières de séquence sans que cela ne perturbe le sujet. L'ergonomie de ces interfaces reste toutefois à retravailler. L'avantage de l'outil est qu'il est intuitif et ouvert à de nombreux formats de fichiers. Il présente également l'avantage d'une compatibilité de partage de fichiers avec le logiciel de manipulation de signal audio, Praat (Boersma, 2006).

### Interface sujet et interrogateur

### Interface transcripteur

visionnage facilité



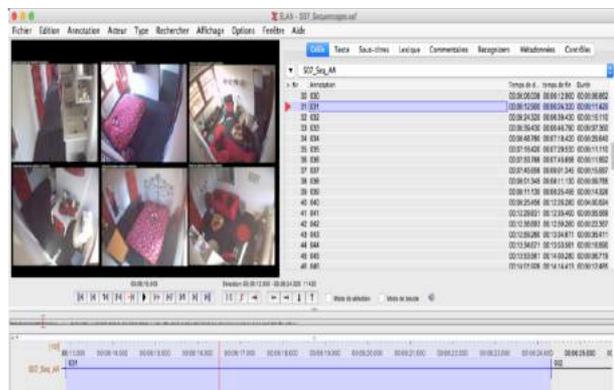
= même séquençage  
vidéos synchronisées



repérage d'événements



- microphones appartement - audio nettoyé
- vidéo 1 vue agrandie



- microphones appartement - audio non modifié
- vidéo 1 vue ou 6 vues au choix

Figure 38 : Interfaces d'auto-annotation

Pour le cas des personnes âgées, il existe par ailleurs un risque qui est celui de perturber la personne en lui montrant une vidéo qui lui renvoie une image différente de celle qu'elle imaginait, ce qui peut avoir des conséquences importantes en cas de fragilisation. En effet, par l'isolement, sa propre représentation sociale est déjà dégradée vis-à-vis de l'inconsistance de sa place sociale (Bennett, 1973 ; Cornwell et Waite, 2009 ; Pitaud, 2013). Or sa perception peut également être faussée du fait des déformations inhérentes au type de la caméra (c.-à-d. des vues plongeantes du plafond) ou de la qualité de l'enregistrement audio. En tant qu'expérimentateur, la précaution minimale est de pouvoir rassurer la personne âgée. Il lui rappelle que la perception que l'on peut avoir de sa voix et de son corps est toujours en décalage sur les enregistrements numériques. Il peut également effectuer des suivis après séance, en consultant les partenaires sociétaux pour éviter toute perturbation de la personne et les prévenir de cet état. Par ailleurs, les contraintes physiques des personnes sont parfois handicapantes pour une séance de travail longue. Cette dernière peut donc être ponctuée de petites pauses (notamment pour la réhydratation des sujets), qui soient choisies sur des moments avec moins de dynamisme interactionnel, en évitant toutefois de trop s'attarder. Les coupures doivent minimiser la perte du déroulement temporel qui facilite l'accès à la mémoire autobiographique. De plus, cette séance doit être unique et elle est nécessairement réalisée en une seule fois.



## Chapitre 4. Corpus EEE : vers une modélisation de l'interaction homme-robot

### 4.1. Caractéristiques générales du corpus

#### 4.1.1. Nature des données récoltées

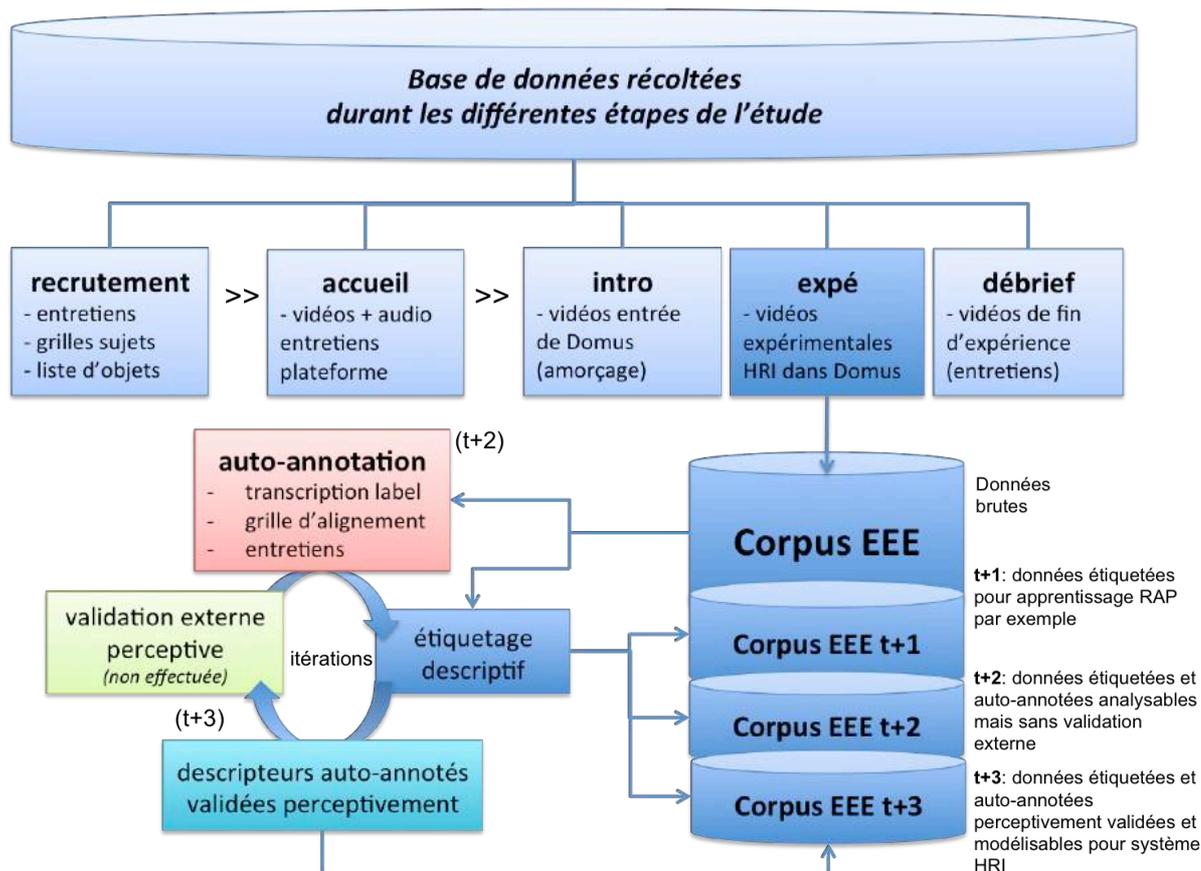


Figure 39 : Schéma récapitulatif des données et traitements appliqués au corpus EEE dans la méthodologie living-lab Domus-LIG

Le corpus *EmOz Elderly Expressions* représente 46 :22 :42 heures d'enregistrement à l'intérieur de l'appartement Domus, dont 14 :46 :23 heures concernées pour les parties auto-annotées et 16 :29 :07 heures de débriefing. Les données du corpus ont été recueillies aux différentes phases de prise de contact avec les sujets et sont de nature variable : questionnaire, entretien semi-directif et principalement bien sûr le recueil audiovisuel étiqueté et annoté sur des données multimodales et événementielles, associées à des valeurs socio-affectives, des interactions entre les sujets, Emox qui trace les événements des magiciens d'Oz. Les analyses, modélisations et applications de ce corpus pourront entrer dans des champs divers, psychosociologique par l'étude des caractéristiques du sentiment d'isolement relationnel dans le cadre gériologique, soit linguistique informatique, par des approches quantitatives et/ou qualitatives, dans un cadre d'interactions ancrées dans la temporalité de la relation. Ces données pourront également servir dans l'apprentissage pour des briques de traitement automatique de la parole ou dans le but de modéliser un système de dialogue dynamique. Dans cette section, les observations relevées à partir des phases de « recrutement »,

« accueil », « intro » et « débrief » seront brièvement résumées pour révéler les perspectives d'exploitation du corpus, qui n'ont pas pu être poussées dans le cadre de cette thèse et sur lesquelles nous ne reviendrons pas dans l'analyse.

Toutes les phases étant recueillies, le corpus peut être exploité pour ses différents aspects de manière plus spécifique en fonction de l'intérêt disciplinaire et applicatif. Par exemple si l'analyse de l'amorçage est supposée être déterminante pour l'explicitation d'un phénomène, elle peut être analysée avec des outils de psychologie en exploitant les données de « intro ». Sinon, si la question de la caractérisation de l'isolement veut être explorée, elle peut solliciter des démarches sociologiques sur l'analyse des phases d'entretien du débriefing ou encore les phases d'accueil voire même le matériel des phases de recrutement (pour celles qui sont disponibles, car les entretiens sont enregistrés uniquement sur accord explicite des sujets)<sup>86</sup>. Ce matériel n'est donc exploité que très partiellement dans le cadre de cette thèse, et de nombreuses pistes d'exploration restent possibles.

### 4.1.2. Profil des sujets retenus

Au total, 24 sujets sont venus sur la plateforme d'expérimentation dont 20 au final sont exploitables. En effet, quatre sujets sont seulement partiellement exploitables à la suite d'un défaut d'auto-annotation ou de problèmes purement techniques (p. ex. défauts sur l'exactitude de temporalité des traces ou qualité d'enregistrement). Certains traitements étant encore en cours de réalisation, les résultats de 16 sujets seront préférentiellement abordés dans le cadre des analyses de cette thèse. Les abandons de milieu de parcours s'expliquent notamment par le temps de post-traitement de données nécessaires pour préparer et les conditions externes du participant (p. ex. : maladies, déménagement, situation familiale changeante, hospitalisation, etc.).

Les informations exploitables sont ainsi :

- une fiche de caractérisation du sujet (dont les informations par sujets sont résumées Annexe 8).
- la liste des objets que les personnes remplissent pour leur prétexte
- les fichiers .wav ou .mp4 des entretiens sont obtenus avec un enregistreur TASCAM ou un dictaphone de téléphone portable.

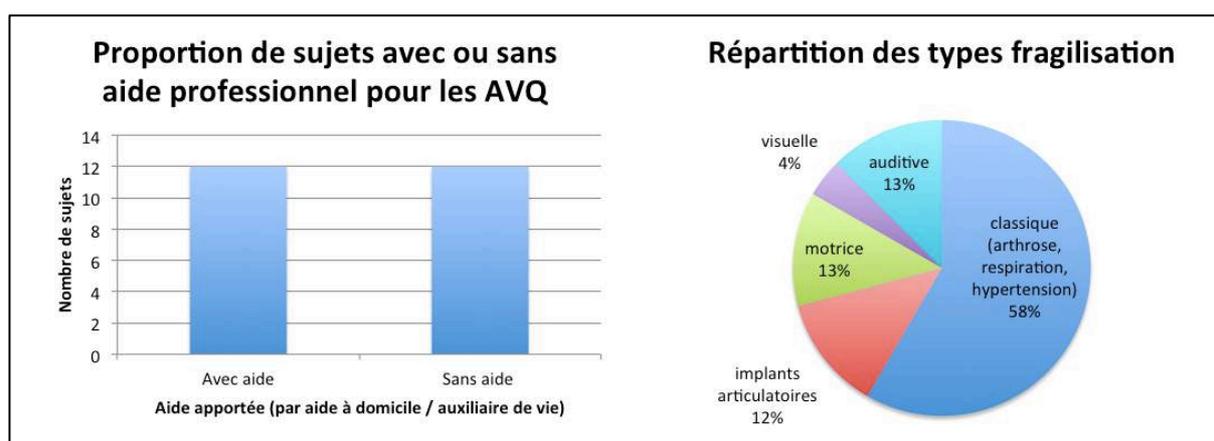


Figure 40 : Illustration des principales fragilisations physiques des personnes et aides aux Activités de la Vie Quotidienne (ou AVQ)

<sup>86</sup> Cf. Mémoire Master en gérontechnologie, par M-C. Dikebely, 2015 qui initie cette approche sociologique.

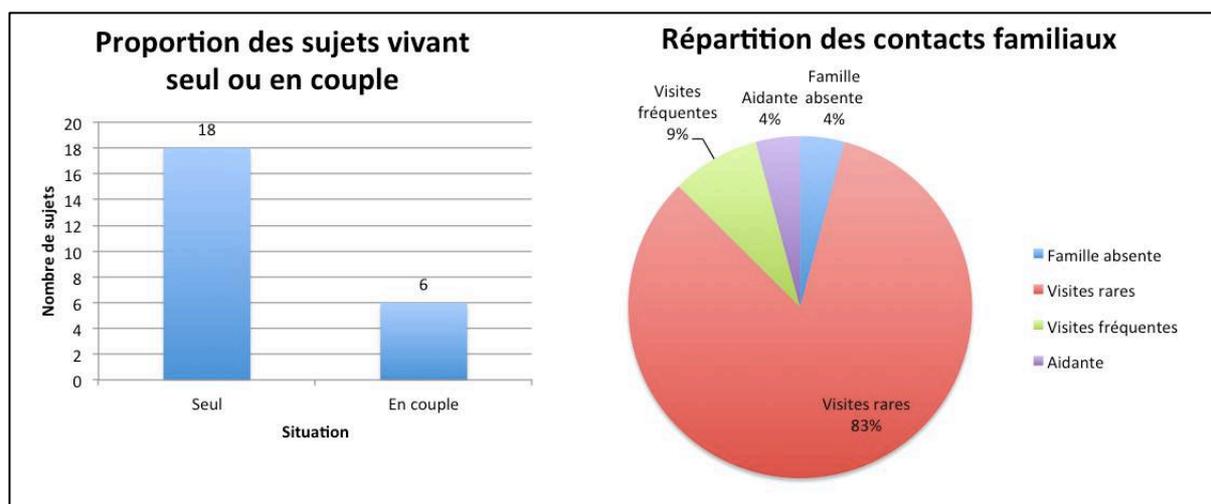


Figure 41 : Informations en lien avec la situation du réseau relationnel des sujets

À partir de ces données, non entièrement exploitées, la Figure 40 donne un aperçu, d'une part des fragilisations physiques des sujets sollicités, et d'autre part de l'état relationnel de la personne, permettant de déterminer le type de profil d'isolement situationnel des sujets, à défaut d'avoir des descripteurs de « sentiments d'isolement ».

La moyenne d'âge des 20 participants est de 80,4 ans pour des sujets âgés de 68 à 93 ans. Dans le groupe d'étude, les personnes sollicitant une aide à domicile et ceux qui n'en bénéficient pas sont équitablement réparties. Sur l'ensemble de ces sujets, la majorité des personnes présentent des fragilisations physiques dites classiques, ou elles sont équipées d'implants limitant certains de leurs mouvements. De plus, ces sujets ne présentent aucune déficience cognitive mais 13% d'entre eux présentent des faiblesses auditives importantes, formant un faible panel de contrôle pour évaluer la portée des primitives auditives testées et leurs effets sur la glu socio-affective.

Pour ce qui concerne les aspects relationnels, la majorité vit seule, ou pour les cas de couple, ces dernières sont généralement elles-mêmes aidantes de leur conjoint qui présente de plus importantes déficiences physiques et/ou mentales. 9% d'entre elles décrivaient recevoir de nombreuses visites, ce qui n'empêchait pas en revanche l'apparition de sentiment de solitude chez ces sujets. Mais la plus grande partie des sujets ne recevaient que de rares visites (soit moins d'une fois par semaine) dont certaines n'avaient plus de famille dans leur environnement proche.

#### 4.1.3. Panel des non participants entrés dans le protocole de sélection

Parmi les 113 contacts de personnes âgées (soit 53 à travers les partenaires sociaux, 60 par les autres structures du réseau ou directement), 68 personnes ont fait l'objet d'un entretien face-à-face. Parmi ces personnes, les 44 sujets n'ont pas donné suite à l'expérience pour les raisons et caractéristiques citées ci-dessous. Il est cependant important de noter que ces personnes composent également la population « d'isolés relationnels supposés » et que leurs conditions de vie contribuent plus ou moins à cultiver cet état isolement :

**Les sujets ont refusé d'être enregistrés pour l'entretien de recrutement :** en effet, une demande d'enregistrement explicite mettant en évidence un dérangement, un scénario en double consentement (avant et après expérience), avec une phase d'enregistrement validée post-expérimentalement, a été jugé comme non réalisable.

**Les sujets sont physiquement trop fragilisés :** ceci concerne plus particulièrement le déplacement où les personnes ont une trop grande réduction de mobilité pour envisager l'évolution au sein de la plateforme : ce sont les contraintes techniques pour leur prise en charge adaptée qui étaient trop contraignantes (p. ex. mobilisation de véhicules spécifiques pour le transport, aide à la levée nécessaire, branchement de respirateurs etc.).

**Des pathologies avérées entre le recrutement et l'expérimentation :** les cas sont rares, mais elles concernaient des diagnostics de la maladie d'Alzheimer et d'épilepsie. Sinon la majeure partie des cas concernent les hospitalisations par un déclencheur d'une maladie saisonnière (p. ex. grippe, bronchite, etc.) ou chutes dont les effets ont été amplifiés par l'état de fragilisation latent.

**Des sujets dont la logistique et l'emploi du temps sont complexes :** une grande partie des personnes âgées restent actives et mobiles, avec un flux de passage constant entre soignants, familles, aidants et les contacts de proximité tels que le voisinage. Certaines ont alors un planning extrêmement chargé en visite planifiée, ne pouvant mobiliser ne serait-ce qu'une après-midi pour l'expérimentation.

**Les annulations de participation de l'aide à domicile :** les difficultés des conditions de travail des aides à domicile sont souvent évoquées en focalisant un *turn-over* important. Pourtant la relation que ces professionnels développent auprès de leur bénéficiaire est conséquente, sachant que de nombreux sujets ont refusé de s'impliquer dans l'expérimentation suite au désistement de leur aide à domicile qui avait elle-même un planning très restrictif<sup>87</sup>.

**Les cas d'isolement extrêmement importants :** à l'image des bénéficiaires suivis par l'association EPPI par exemple, les intervenants qualifient les personnes qu'ils suivent de « personnes du quatrième âge ». Il s'agit d'un public en général de plus de 85 ans, mais qui ici concerne des personnes à grande fragilisation. Pour ces personnes, c'est la relation sociale elle-même qui induit volontairement un isolement situationnel de la personne, que l'on pourrait même qualifier de *hikkikomori* en éludant la question de l'âge. Cette situation démarre avec des événements déclencheurs variables (première chute, situation traumatisante lors d'une sortie, évitement de certaines personnes) ou s'installant par un désengagement progressif du fait notamment de la perte de sa place sociale, il s'agit de cas de personnes qui ne sortent plus de chez elles.

**Les sujets pour lesquels la famille a refusé la participation :** ultérieurement à l'entretien de recrutement, l'entourage familial a parfois eu son importance sur l'issue du recrutement de 3 manières. La première raison est basée sur un refus, car le proche considère que son parent âgé ne sera pas en mesure de réaliser l'expérience. La seconde est une initiative de participation demandée par la famille pour la personne âgée « pour la faire sortir » ou « lui donner envie de faire quelque chose » qui s'est manifestée par une opposition importante de la part des potentiels sujets. La troisième est un refus lié à la responsabilité vis-à-vis du parent âgé ou des conditions de participation incompatibles avec le contrôle des biais (p. ex. un accompagnant qui ne quitte pas la personne sous aucun prétexte, condition absolue d'une divulgation d'enregistrement des données au départ, etc.). Dans la majorité des cas, la personne âgée elle-même avait accepté la participation à l'expérience.

---

<sup>87</sup> Nous tenons à souligner que la participation des aides à domicile a été d'une grande qualité pour cette expérience, entièrement bénévole, aux dépens de leur temps de travail et donc aussi pour un temps de services que les personnes âgées n'ont pas non plus pu bénéficier sur ces après-midis.

**La difficulté de prise de contact :** l'isolement situationnel a parfois de vrais freins si la personne est socialement coupée au-delà de son cercle de proximité (p. ex. certaines personnes n'avaient pas le téléphone alors qu'elles étaient géographiquement éloignées). Le suivi étant parfois trop contraignant (p. ex. certains des sujets qui ne possédaient pas de téléphone ou partageaient plusieurs lieux de vie pour leur prise en charge).

**Les sujets âgés qui sont eux-mêmes aidants :** de nombreuses personnes vivent encore en couple ou avec des fratries à proximité. Il est fréquemment arrivé que le sujet contacté fût le sujet « aidé » avec un « aidant âgé » qui a été recruté à sa place. Malgré les apparences, ces aidants dits naturels ont souvent décrit un sentiment de solitude face à l'impuissance à laquelle ils faisaient face, et ont souvent dû renoncer aux expériences par l'aggravation de l'état de santé de la personne aidée.

## 4.2. Étiquetage de données multi-sources

L'étiquetage des données consiste à faire des annotations descriptives qui caractérisent les stimuli sur un niveau de granularité qui soit le plus fin possible. Les grilles d'alignement visent ensuite à être lues en regard des auto-annotations afin de déterminer quels objets interactionnels, et plus particulièrement quels changements de leur dynamique intrinsèque ou de leur dynamique d'enchaînement plus transversal, marquent des points de rupture décrivant le changement de l'état relationnel, soit l'état de la glu socio-affective.

### 4.2.1. *Matériel interactionnel sélectionné*

Cette étape d'étiquetage se focalise sur la partie nommée « Expe » du corpus. Pour cela, le corpus EEE a suivi un prétraitement de synchronisation et de découpage permettant des alignements sur des événements parallèles multiples, pouvant être traité puis regroupé par différents annotateurs, quelle que soit la source d'enregistrement de données. En d'autres termes, les différentes ontologies annotées ainsi que leurs différents niveaux de granularité sont alignés indépendamment de la source d'enregistrement donc l'étiquetage peut être partitionné et fusionné à n'importe quelle étape. En effet, la capture des données s'effectue à travers 6 caméras de surveillance intégrées au plafond (sans son), 6 micros plafonds (proches des caméras et enregistrant les sons d'environnement) et 1 micro-casque (pour la capture de qualité permettant de faire des mesures acoustiques et notamment prosodiques). Le son des micros du plafond permet de rassembler les données écologiques qui seront les potentielles entrées d'un système de reconnaissance vocale associé à l'habitat intelligent. Le son du micro-casque est dédié à la mesure des variations acoustiques plus fines (tels que l'étiquetage des bruits de bouche ou l'analyse prosodique). Une autre source intéressante aurait certainement été les inputs audio et vidéo directs du robot. En revanche pour contrebalancer les risques des pannes techniques liés à la multiplication des sources d'enregistrement, ces derniers n'ont pas été capturés dans le cadre de cette expérience, mais elle pourrait être envisagée pour de futurs travaux. Une seconde raison est également plus architecturale du système dans son ensemble. En effet, ici nous ne considérons pas deux technologies (habitat intelligent et robot) séparées, mais une seule qui intègre l'ensemble de l'écologie interactionnelle *via* le rôle attribué au robot qui est considéré comme la télécommande domotique vocale de l'appartement. Cette manière d'englober les deux serait d'ailleurs peut-être un gain de robustesse pour un système de reconnaissance automatique de la parole et un système de dialogue. Du fait des caractéristiques de ce robot, et plus particulièrement lors des déplacements du sujet, le robot est plus lent de par la force de ses servomoteurs et de sa taille, augmentant la distance homme-robot à certains moments. La qualité de la source sonore serait alors très largement diminuée si elle ne s'appuie que sur les capteurs du robot, ce qui aurait néanmoins une possible significativité si cette dégradation du signal était l'objet même de l'analyse du signal qui permettrait par exemple de trouver des indices acoustico-spatiaux de la proxémie. Cette dernière approche ne sera cependant pas adoptée dans le cadre de cette thèse, car la proxémie sera abordée, mais d'un point de vue plus visuel qu'acoustique dans le cadre de cette étude.

Les données « Expe » analysées vont donc alterner la manipulation de 4 sources principales :

- une vidéo multicam (avec les six angles de vues en même temps, mais charge cognitive plus lourde à analyser), qui est la même vue que celle des chercheurs « magiciens » au moment de l'expérimentation. Elle permet néanmoins de retrouver des indices sur l'ensemble des données visuelles.
- une vidéo en vue principale qui prend le(s) (deux) plan(s) de caméras où le sujet et Emox sont les plus visibles. Il s'agit de cet angle de vue qui est proposé aux sujets pour faciliter leur auto-annotation.

- le canal audio du micro-casque
- les 6 canaux audio multiplexés sur un seul enregistrement, sélectionnant le micro le plus proche du plan de vue principale courant. Des méthodes automatiques fondées sur le niveau sonore est explorée par (Vacher et coll., 2011) mais n'ont pas été exploitées dans le cadre de cette étude. Chacune des deux vidéos a ainsi une version synchronisée avec d'une part sur l'enregistrement du micro-casque et d'autre part sur l'enregistrement des micros du plafond. Le découpage des sources est identique aux millisecondes près pour permettre de leur associer les mêmes grilles d'alignement (.eaf Elan en audiovisuel, .TextGrid Praat en audio). Cette précaution a été prise afin de pouvoir faire des découpages automatisés des différentes ontologies et différents niveaux de granularités des annotations, en s'appuyant sur un alignement unique applicable à l'ensemble des sources, afin de filtrer et produire des sous-corpus d'apprentissage ou d'analyse sur des éléments spécifiques du corpus.

### 4.2.2. *Transcription des données*

La transcription concerne principalement la parole humaine, puisque les productions du robot peuvent être alignées automatiquement grâce aux traces du magicien d'Oz. La transcription concerne principalement la parole des sujets âgés, pour un gain de temps sur le reste du traitement, cependant 5 séances expérimentales ont également été transcrites avec la parole des expérimentateurs et des aides à domicile.

La parole des sujets âgés du corpus EEE est transcrite orthographiquement (cf. Annexe 9 pour les règles de transcription). Pour réaliser cette transcription, elle est alignée sur le signal acoustique à l'aide du logiciel Praat. Un premier préséquençage en IPU (pour *inter-phrastic units*) est réalisé à l'aide de SPASS (Bigi et Hirst, 2012) afin de faciliter la transcription. En effet, le choix a été d'effectuer une transcription alignée simultanément, car : 1) l'alignement forcé n'est pas réalisable sans texte de référence et elle apporte trop d'inexactitudes, car les productions sont spontanées, bruitées par les sons ambiants et multi-locuteurs. De plus, les voix principales qui sont celles des âgés posent plus de problèmes de reconnaissance ; 2) elle permet d'aligner précisément l'énoncé en se basant sur le support visuel du spectrogramme. Cela facilite notamment la transcription des micro-expressions vocales. En effet, dans notre hypothèse, nous supposons que les « bruits de bouche » sont d'importants marqueurs des fonctions interactionnelles.

Or dans une parole spontanée, et surtout avec des transcripateurs natifs, il semblerait que malgré la consigne d'une transcription exacte, son à son, qui soit la plus précise possible, au moins la moitié des micro-expressions vocales ne sont pas relevées. Cela rejoint les observations des études de psychologie qui montrent à travers les phénomènes d'attention et de mémoire que les événements faisant l'objet d'un automatisme socioculturel et linguistiquement assimilé rendent par exemple aveugles les natifs anglais à certaines lettres. En effet, des natifs, devant compter le nombre d'occurrences de certaines lettres dans un texte, telles que les « F », vont être quasi systématiquement oubliés dans les « *if* » et « *of* », alors qu'un non-natif ne les manquera pas<sup>88</sup>. Concernant toujours les micro-expressions vocales, un autre risque relevé dans les anciens travaux (p. ex. Vanpé, 2012) révèle qu'une transcription basée sur les stimuli audiovisuels, notamment faciaux, tend également à rajouter du bruit, car les annotateurs auraient tendance à créer des descripteurs acoustiques induits d'expressions seulement faciales, alors qu'aucun événement ne semblait visible sur le spectrogramme du signal acoustique.

Ainsi l'avantage de la méthode de transcription adoptée, qui reste néanmoins chronophage, permet de « forcer » le transcripateur à noter un maximum de micro-expressions sur un signal acoustique décelable. À travers le spectrogramme, le canal visuel lui facilite : 1) le repérage de ces signaux acoustiques sur lequel il réalise en direct l'alignement/transcription, 2) certaines transcriptions mal-audibles du fait des caractéristiques de la parole âgée. Ces micro-expressions produisent d'ailleurs du bruit pour un système de RAP puisqu'il s'agit d'un signal acoustique notable, mais encore très peu définies dans les modèles acoustiques ni dans les modèles de langage en TAL, puisqu'ils ne correspondent à aucune entrée (sauf dans le cadre d'interjections) du dictionnaire de lexique. Leur recensement peut donc ainsi contribuer à l'aide de la constitution du dictionnaire de lexique ou l'amélioration des modèles acoustiques (pour les objets prélexicaux et bruits de bouche), voire même faire l'objet de réflexion de nouveau moyen de modélisation pour augmenter l'architecture d'un système de RAP faisant partie intégrante d'un système de dialogue plus large.

---

<sup>88</sup> Un exemple de ces tests : <http://sharpbrains.com/blog/2009/04/06/brain-teaser-to-stimulate-your-concentration-skills/>

Enfin, sachant qu'il existait des différences de précisions inter-annotateurs assez importantes sur la transcription et notamment celle des micro-expressions, une normalisation de leur transcription est en cours. La nomenclature adoptée étant la suivante :

- Pour les micro-expressions de type « bruits de bouche » qui ne sont ni phonétiques ni phonologiques, la transcription est une description acoustique associée à un descripteur général qui peut servir de filtre. La nomenclature permet ainsi de considérer les différences de précision de description acoustique inter-annotateurs :

**[BB : descripteur – description acoustique]** ex : *[BB : clic – latéral, mouillé]*, *[BB : inspiration – nasale, courte]*, *[BB : inspiration – buccale, longue, bloqué en final]*

- Pour les micro-expressions de type « prélexicales » qui sont phonétiques, voire même phonologiques :

**[PreLex : /transcription phonémique en API]** ex : *[PreLex : /mwe/]*, *[PreLex : /o :f/]*

- Pour les micro-expressions de type « interjections », il existe en général une forme lexicalisée et commune dans le lexique de langue. Si celle-ci n'existe pas ou à du mal à être définie, elle change de catégorie et suit la règle de transcription des prélexicales.

**[Interj : transcription orthographique commune]** ex : *[Interj : voilà]*, *[Interj : oui]*

### 4.2.3. Outils et organisation générale

Les annotations vidéo sont réalisées à l'aide du logiciel ELAN pour sa simplicité ergonomique (du fait d'un usage multi-annotateurs pour réduire le temps du traitement chronophage) et sa compatibilité avec Praat. Il est ainsi plus aisé de faire des allers-retours entre des descriptions acoustiques fines et des descriptions audiovisuelles globales. La Figure 42 illustre l'interface sur laquelle les alignements sont effectués.

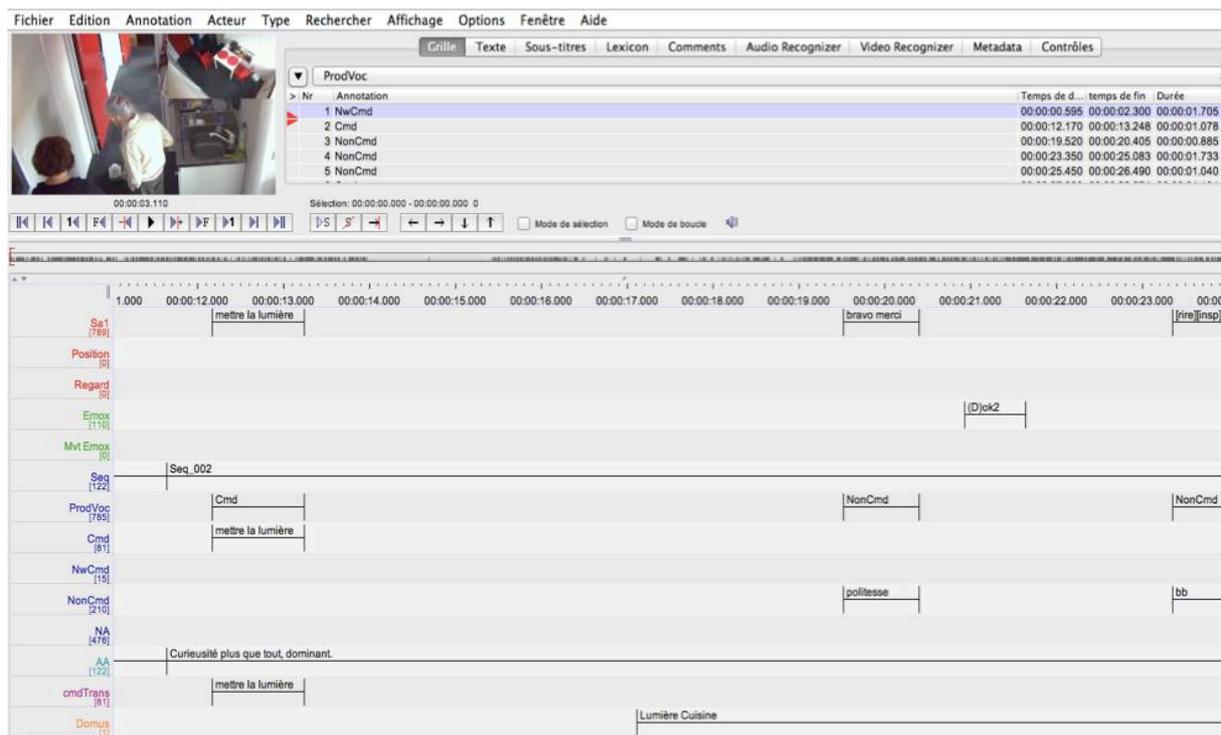


Figure 42 : Interface d'alignement du corpus EEE sous le logiciel ELAN

Les sections suivantes décriront les différents niveaux d'annotation des productions vocales du sujet, comme illustré en Figure 43. En parallèle des transcriptions, un premier séquençage est réalisé rapidement, afin de préparer la phase d'auto-annotation, tout en restant dans une boucle méthodologique agile et rapide. Ce séquençage est un découpage arbitraire des expérimentateurs dont le but est de pouvoir suivre commande par commande l'expérimentation. La définition d'une « commande » n'est pourtant pas chose aisée. En effet, une amorce de commande reconnaissable, mais abandonnée ou encore des itérations de même contenu lexical seraient-elles considérées comme une commande ou plusieurs ? Ou alors ne considérons-nous comme commande que ce qui se passe autour d'une action domotique effective ? Intrinsèquement, de manière sémantique ou pragmatique, l'acte d'effectuer une commande s'arrête-t-elle à la seule formulation de l'exécutable ou intègre-t-elle d'autres marqueurs accompagnant cet acte (p. ex. le suivi à travers le regard de l'actionneur en exécution, un feedback de type remerciement ou correction, etc.) ? Sachant que la démarche de collecte est inductive, il est *a priori* assez difficile de fixer des règles basées sur une théorie pragmatique par exemple, ou formaliser un tour de parole dans l'absolu. Ainsi le premier séquençage global « Séq » ici proposé tente de déterminer de petites unités pour le déroulement pas-à-pas de la vidéo, tout en minimisant les possibilités d'interprétations du sujet. Par exemple, s'il y a des itérations de contenu lexical, elles sont séparées dans la mesure du possible, car elles peuvent potentiellement représenter deux actes indépendants, sauf si le signal acoustique ne présente pas de pause entre les deux itérations. Cependant, elles peuvent être identiques tout comme elles peuvent être dans une continuité. En revanche, aucune décision ne peut être prise *a priori* (cf. Annexe 10).

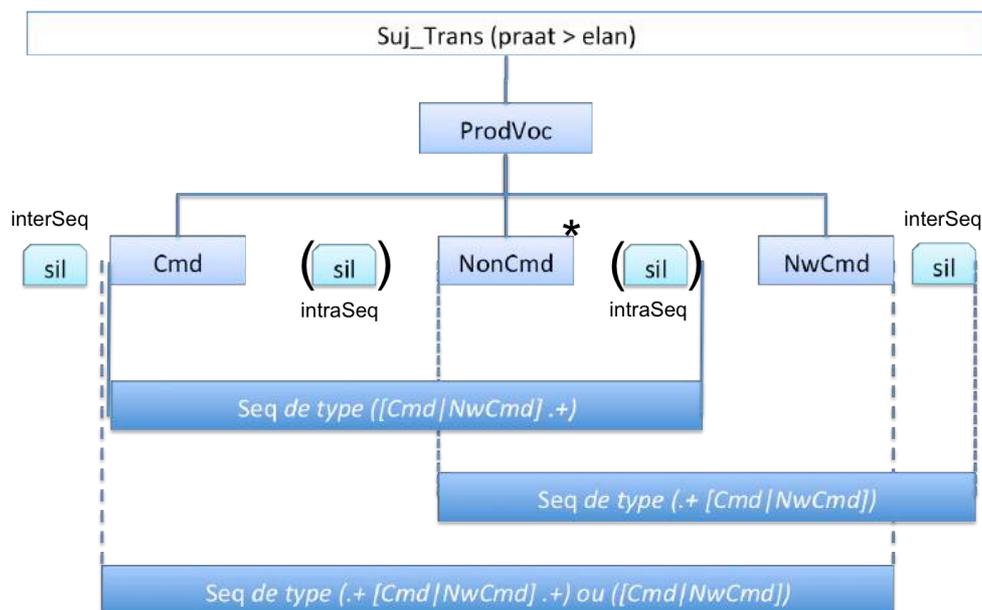


Figure 43 : Hiérarchie d'étiquetage des productions vocales des sujets

Globalement, l'intention de ce séquençage a été de capturer et de rassembler tout ce qui se passe autour d'une « commande » dans une même séquence. L'hypothèse posée est que vocalement, il peut y avoir des éléments avant et après le contenu lexical porteur du vocabulaire de la commande. Ainsi chaque séquence commence au début du premier élément « satellite » précédent la commande (p. ex. les interjections de type « euh », « alors », etc.), sachant qu'il peut potentiellement y en avoir plusieurs. Il se termine ensuite au début du silence suivant le dernier élément « satellite » la commande (p. ex. les interjections de type « merci » ou des micro-expressions telles que des rires) et précédant le passage à une nouvelle commande. Ce dernier silence qui est à la transition entre séquences est plus ou moins long. Il est ainsi soit le point de transition avec la commande suivante ou peut également être un

passage sur une « pause » interactionnelle, voire un « moment » cognitif très intense (découpé en différents temps, pouvant ou non avoir lieu en parallèle). Ainsi les frontières de la séquence sont totalement arbitraires, mais systématiques, et l'un des buts de l'auto-annotation sera également de mettre en évidence la nature informationnelle de ces « silences » pour suivre la dynamique de l'interaction par la mise en évidence des transitions cognitives interactionnelle, non interactionnelle ou passive par exemple. Ces frontières établies *a priori* n'ont aucune valeur fonctionnelle et vont même pouvoir être comparées aux découpages temporels que les sujets réalisent dans l'auto-annotation. Ce deuxième découpage est supposé faire émerger des marqueurs plus plausibles de séquences interactionnelles. En revanche, ce séquençage *a priori* peut néanmoins donner un aperçu de la rythmicité du signal de parole sans prise en compte des fonctions communicationnelles.

Le diagramme de la Figure 43 suit la nomenclature classique d'expressions régulières. L'astérisque « \* » représente la possibilité de zéro à plusieurs occurrences du type NonCmd, les parenthèses « () » illustrent le caractère facultatif de l'élément, le « + » désigne les occurrences multiples à partir de 1, le point « . » représente n'importe quel type d'élément et la barre verticale « | » désigne une disjonction entre les éléments définis à l'intérieur de crochets « [] ».

#### 4.2.4. *Étiquetage des productions vocales des sujets*

Comme le montre la Figure 43, une fois que la transcription est réalisée, sous Praat une tier de productions vocales nommées ProdSuj est étiquetée selon 3 catégories pour faciliter l'extraction des différents événements afin de réaliser par une réimportation dans le logiciel Elan. Cet étiquetage permet : 1) une analyse structurelle des objets vocaux du sujet et 2) une analyse de leur organisation temporelle dans l'interaction globale. Un premier marquage grossier a donc été réalisé semi-automatiquement, selon trois catégories présentées dans les sous-sections suivantes.

##### 4.2.4.1. **Commandes vocales – Cmd**

Les **Cmd** (les « commandes ») : il s'agit des productions correspondant aux exécutable de la liste des « commandes à tester » pour que le robot puisse apprendre la voix du sujet au début de l'expérience. Il s'agit donc du « vocabulaire » de base du système domotique. L'enjeu principal de ces objets est de pouvoir extraire l'évolution de leur forme au cours de l'interaction, supposé être l'un des indices illustrant l'existence et la dynamique de la glu socio-affective. L'analyse portera ainsi sur la forme même de l'énoncé, et notamment sur ses dimensions morphosyntaxiques et prosodiques (cf. Chapitre 5 et section 6.1). Puis dans un second temps, ils seront utilisés comme repère autour duquel gravitent les micro-expressions pour en analyser la nature et les évolutions. Enfin, une dernière exploitation possible de ces éléments du corpus est de recueillir les variantes multiples produites spontanément par les sujets. Elle permettrait alors un apprentissage spécialisé pour une RAP (cf. section 6.2), par l'amélioration des modèles de langage, dans le cadre restreint de ce contexte interactionnel réduit.

##### 4.2.4.2. **Nouvelles commandes – NwCmd**

Les **NwCmd** (les « nouvelles commandes ») : ce sont des productions spontanées qui n'étaient pas proposées aux sujets explicitement sur la liste, mais que les sujets ont essayé de réaliser en suivant le même type de format de productions vocales que les Cmd. Si certains patterns de fonctionnement émergent de ces objets (p. ex. savoir à quel « moment » ces éléments sont susceptibles d'être produits, ou en repérant les fonctions communes aux Cmd à travers des marqueurs objectifs (p. ex. prosodie) de ces événements, il serait potentiellement possible de modéliser certains points transitionnels pour un système de dialogue adaptatif incrémental. Certaines propositions des sujets sont également des possibilités d'actions réalisables, mais qui ne sont pas encore formalisées dans le système et n'apparaissent donc pas dans la liste proposée aux sujets comme commandes possibles. Elles peuvent donc augmenter le panel de commande domotique (ou robotique) du système initial.

##### 4.2.4.3. **Non-commandes – NonCmd**

Les **NonCmd** (« les non commandes ») : il s'agit des productions spontanées et naturelles qui véhiculent d'autres informations que les exécutable domotiques ou robotiques. Elles représentent l'un des défis majeurs des systèmes de TAP s'appuyant sur un modèle de langage générique, ou à l'origine d'erreurs dans les systèmes à vocabulaire restreint, car il s'agit de la parole naturellement produite, mais non prévisible, qui constitue l'ambiguïté et la complexité du langage spontané. Leur analyse et la compréhension de leur organisation fonctionnelle, notamment au niveau temporel, pourraient être un moyen de passer au-dessus d'un verrou théorique et technologique, comme celui de pouvoir discriminer les commandes adressées au système par rapport au reste de la parole. Les éléments composant ces objets étant nombreux, la section suivante développera les sous-catégories de descripteurs qui ont été associés à ce type de productions.

## 4.2.5. Sous-étiquetage : focus sur les événements hors commandes

### 4.2.5.1. Principes de segmentation des NonCmd

Le sous-étiquetage des Cmd et des NwCmd est relativement simpliste, il s'agit surtout de pouvoir repérer les formes canoniques communes afin de pouvoir accéder à la commande domotique/robotique associée pour l'exécution d'une action sur un contrôleur. Ce filtrage permet notamment de faciliter le repérage des évolutions des formes morphosyntaxiques d'une même commande pour faciliter la création du lexique d'un RAP par exemple. En termes d'analyse, il facilite également : la comparaison des commandes entre les différents sujets. Les listes des formes canoniques se trouvent en Annexe 5.

La partie la plus fine concerne le sous-étiquetage des productions de type NonCmd. Il s'agit ainsi de repérer tout d'abord deux grandes catégories d'éléments : les micro-expressions vocales (notées ME) et les formes complexes du langage (notées FC). La Figure 44 ci-dessous présente les détails des sous-étiquetages des NonCmd.

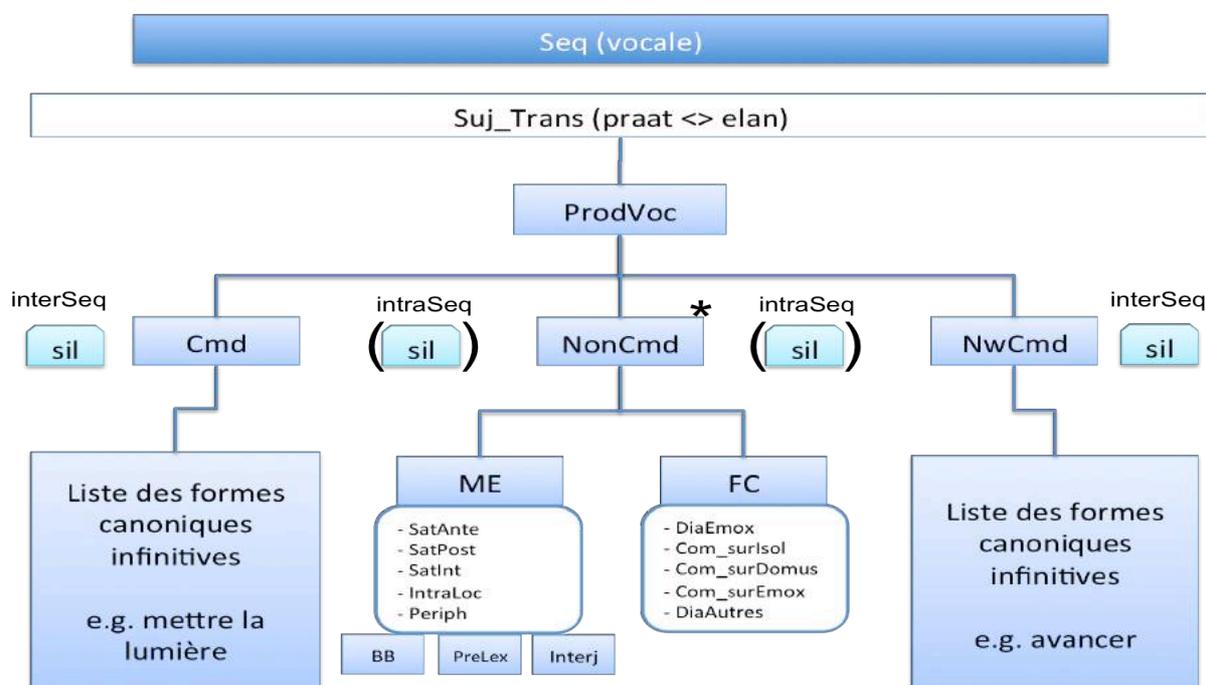


Figure 44 : Diagramme des productions vocales – granularités minimales

### 4.2.5.2. Micro-expressions vocales de sujets – ME

Les ME représentent les micro-expressions vocales des sujets, des éléments qui ont été transcrits avec les notations [BB], [PreLex], [Interj]. De la même manière que le vocabulaire de primitives qui est utilisé par le robot Emox, les sujets produisent ce type de productions dans des proportions non négligeables dans leur parole spontanée. L'évolution de leur nature et la variation de leur fréquence sont ainsi supposées être des marqueurs de changements de l'état de la glu socio-affective. En effet, l'hypothèse principale est construite sur l'idée que des productions de « prosodie pure », que l'on retrouve en l'état dans ces ME est porteuse d'informations socio-affectives fonctionnelles qui permettent la manipulation de la glu. Cette dernière est donc valable pour la parole des sujets. Dans la nomenclature d'étiquetage, ces différents types de ME sont organisés par rapport à la position de la commande domotique (ou robotique).

En effet, cet étiquetage des ME suppose que leur organisation rythmique, ainsi que leur distribution en fonction de leur pattern prosodique intrinsèque pourraient être des indicateurs de l'état de glu. Ces éléments sont ainsi pris dans leur globalité, mais aussi à des degrés de granularité plus fins en termes de nature et de position. Concernant la position, la proposition qui est faite ici, mais qui reste totalement arbitraire propose de différencier : 1) les éléments, qui gravitent dans la proximité immédiate d'une commande (la notion d'immédiateté est précisée dans la section suivante) et semblent être inclus dans le mouvement prosodique de la commande, sont considérés comme des « satellites » (notés Sat). Leur emplacement (soit avant ou après) est précisé par rapport à la commande, mais ils peuvent être considérés comme des composants internes « à la commande » ; 2) les éléments plus isolés, qui gravitent autour des commandes, mais ne rentrent pas dans le même signal acoustique que la commande ou ceux qui entrecoupent des silences entre deux commandes sont définis comme des éléments périphériques (notés Periph). En revanche, les ME qui se retrouvent dans le flux de parole conversationnelle entre humains (au retour des complices) ne sont pas associées à ce type de méta-étiquette. Elles précisent uniquement s'il s'agit de [BB], [PreLex] ou de [Interj], si elles sont relevées. En effet, cette partie de discours rendant beaucoup plus difficile l'étiquetage de ces expressions (nombreuses et amalgamées), les occurrences ne sont peut-être pas aussi bien relevées que la partie interactionnelle dyadique sujet-robot).

### 4.2.5.2.1. Micro-expressions « satellites » de commandes

La section précédente décrivait ces éléments comme étant dans la proximité immédiate des commandes. Pour préciser, l'étape de transcription de la commande et l'automatisation du séquençage permet de séparer des IPU – ce qui serait similaire au séquençage automatique d'un système RAP par exemple). Cette dernière permet de discriminer des séquences de « parole » et « non-parole ». Parmi les frontières associées, certaines productions interactionnelles font partie de la même unité phrastique (c.-à-d. le même intervalle de signal de parole) que celle portant le vocabulaire associé pour désigner une action d'exécutable de la commande vocale. Parmi ces éléments qui sont inclus dans cet intervalle, des ME apparaissent. Ce sont ces éléments qui ont été provisoirement appelés des « satellites ». Ces satellites font ainsi partie du même bloc de signal que le « noyau » de la commande, par opposition à d'autres qui sont détectés sur des séquences de signal indépendantes, qui ont été définis comme des éléments « périphériques ». Parmi les éléments « satellites », nous avons des éléments qui apparaissent avant le noyau de la commande (notés SatAnte), ceux qui apparaissent après (**SatPost**) et d'autres qui sont à l'intérieur même d'une commande (**SatInt**). Ces étiquettes sont associées pour faciliter le filtrage et avoir la possibilité de les filtrer ou d'en étudier la distribution le cas échéant, ce qui peut être utile pour la constitution intégrée ou non intégrée de ces micro-expressions dans les corpus d'apprentissage des systèmes de reconnaissance de parole par exemple.

**SatAnte** : ce sont les ME que l'on retrouve devant une commande. Ces éléments sont supposés révéler les transitions de l'état de la glu, et peuvent potentiellement varier en durée, en délai d'attaque, en nature, etc. Autant d'indices qui pourraient expliquer l'état de la glu.

*Ex : [BB : hum] mets la lumière ; [PreLex : euh] mettre la lumière ; [Interj : alors] tu mets la lumière*

**SatPost** : ce sont les ME que l'on retrouve après une commande, et sont les éléments les plus proches des feedbacks du robot. Pour le moment nous nous limitons à noter sa position par rapport à la commande mais il est possible d'imaginer des degrés encore plus fins quant à savoir par rapport aux feedbacks suivant les commandes (si avant le feedback du robot, si après le feedback du sujet en réaction au feedback du robot, etc.).

*Ex : allez mettre en jaune – rire1(emox) – [BB : rire] ; mettre la lumière [PreLex : oh] ; tu me mets la lumière [sil] [Interj : s'il te plaît], [Interj : voilà]*

**SatInt** : parfois la commande est difficile à définir parce qu'elle n'a pas une forme unifiée (la parole peut être saccadée, comme dans le cas de faux départ, d'hésitations ou même prendre des formes plus complexes). Ainsi, dans cette « commande », on retrouve de manière plus ou moins récurrente des ME à l'intérieur même de la commande, d'où la proposition de ce label.

*Ex : tu veux bien me mettre [PreLex : euh] la lumière.*

### 4.2.5.2.2. Micro-expressions « périphériques » de commandes

**Periph** : il existe des formes de ME qui peuvent plus ou moins être proches de la commande, mais toujours entrecoupées de silences ou des occurrences de parole qui surviennent entre les commandes. Ces intervalles comportant ces productions « indépendantes » sont ceux qui sont provisoirement appelés des éléments « périphériques ». Cette nomenclature est issue d'une vision centrée sur la commande et elle n'est donc donnée que pour faciliter certaines tâches de tri. La différence peut sembler parfois ambiguë avec les « satellites », notamment lorsque les silences qui précèdent ou suivent ces éléments sont courts. Mais ces silences acoustiques existent toujours. Par ailleurs, au regard des auto-annotations, la dimension fonctionnelle des valeurs associées semblait différente alors que certaines formes acoustico-lexicales étaient identiques, ce qui a motivé la distinction entre satellites vs périphériques. (cf. section 4.3 pour les détails de nomenclature des auto-annotations mentionnées en exemple).

#### Exemples :

- *Ex1 : transcription : [BB : inspiration] mettre la lumière => SatAnte  
auto-annotation : FoT\_ProdSuj : j'articule particulièrement bien parce que je m'applique, je veux qu'il m'entende bien*
- *Ex2 : transcription : [BB : inspiration] (pause) mettre la lumière => Periph  
auto-annotations : FoT\_Gn : je commence à trouver le temps long – FoT\_ProdSuj : là c'est un peu la lassitude parce que je sais plus quoi faire*
- *Ex3 : transcription : on met la lumière [Interj : voilà] => SatPost  
auto-annotation : FoT\_ProdSuj : intérieurement je me suis dit « bon, ça y est, j'ai passé un cap »*
- *Ex4 : transcription : mets la lumière (pause) [Interj : voilà] => Periph  
auto-annotations : FoT\_ProdSuj : c'était l'ordre et voir si l'ordre est bien exécuté – FoT\_ProdSuj : c'est pour le récompenser – je le félicite.*

Pour les satellites, l'auto-annotation associée au bruit de bouche concerne la même étiquette que celle de la « commande », soit des annotations de type *FoT\_ProdSuj* alors que les périphériques sont généralement associées à une étiquette d'auto-annotation qui est différente. La difficulté ici est qu'il n'était pas pertinent de fixer *a priori* une valeur seuil de la durée du silence qui permettent de distinguer une opposition entre satellites et périphériques. En effet, le temps cognitif interactionnel des sujets étant différent, déterminer une valeur absolue de cette durée ne semblait pas pertinent. Or ces premières observations suscitaient une réflexion autour de ces deux types d'événements, quant à leur appartenance de groupe prosodique. En effet, pouvons-nous considérer comme équivalents des objets qui semblent appartenir à des groupes prosodiques différents ? L'étiquetage sur cet aspect fait encore l'objet d'interrogations.

De plus, précisons les exemples 3 et 4. Si la référence est le signal acoustique, nous avons une « pause », soit un silence marqué comme sur Ex2 et Ex4. Or la séquence en condition multimodale montre que le « voilà » de Ex4 est produit après la mise en route de l'actionneur domotique, alors que dans le cas Ex3, il est produit avant la réalisation des feedbacks (robotiques ou domotiques). Maintenant, si l'on suppose que c'est au niveau des

feedbacks qu'il est possible de discriminer ces phénomènes, ce n'est pas non plus si simple, puisqu'un « périphérique » n'est pas systématiquement produit après un feedback. En effet, de la même manière que les « voilà » nous avons des exemples de « s'il te plaît » fonctionnant de la même manière. En revanche, il y en a certains, dont des périphériques, qui sont produits avant feedbacks. Ces derniers sont associés à des étiquettes de type « *ça a été un peu long alors je me dis il a pas compris ou il m'a pas entendu* ». Il semble donc exister des processus spécifiques autour de ces phénomènes qui ne sont pas encore clairs, mais qui n'auraient pas pu être remarqués si l'étiquetage se limitait à la transcription [BB], [PreLex] et [Interj].

Avec ce marquage, une autre motivation était également de pouvoir observer si les mêmes objets sont plus ou moins mobiles, et ce que le changement de position peut avoir un effet. Plus les ME ont un niveau de contrôle élevé, moins ils semblent mobiles du fait des contraintes morphosyntaxiques de la langue. En revanche, les focus qui semblent beaucoup plus libres seraient susceptibles d'apparaître plus spécifiquement sur ces éléments.

#### **4.2.5.3. Formes d'expressions complexes et dialogiques – FC**

Les **FC** sont des formes lexicalisées d'unités plus complexes qui sont structurées dans un regroupement syntaxique. Elles rassemblent des mots avec une sémantique autosuffisante et désignent des éléments composés organisés morpho-syntaxiquement. Ces éléments sont plus ou moins ambigus et représentent du bruit pour le système, introduit par la parole spontanée. Or il s'agira de productions courantes qui peuvent survenir dans des environnements intérieurs, non prévisibles, mais qui ne font pas l'objet des cibles d'une commande qui peut être adressée. L'un des enjeux des systèmes de RAP serait ainsi de pouvoir s'extraire de ces productions, non pas en les reconnaissant pour leur contenu, mais pour leur caractéristique fonctionnelle résidant dans le fait qu'ils ne soient pas des commandes. À but d'analyse, ces parties ont ici été étiquetées avec des descripteurs permettant de définir des parties plus ou moins dialogiques (dans le sens conversationnel), s'adressant aux robots, ou à l'entourage humain (incluant ou non le robot). Cette dernière est focalisée ou non, sur une parole orientée vers « soi-même ». Elles représentent aussi des productions « oralisées » dont la cible n'est pas explicite ou ambiguë. Les « commentaires » (notés **Comm**) ne sont que d'infimes recueils d'informations qualitatives de ce que les personnes révèlent au cours de l'expérience. Elles sont à compléter et à croiser avec les recueils de recrutement et de débriefing et sont ici secondaires, mais leurs exploitations peuvent être pertinentes selon le but de modélisation/interprétation ultérieure.

##### **4.2.5.3.1. Filtrage rapide de commentaires**

Les étiquettes qui sont des « ontologies » de FC ont été choisies de manière arbitraire pour favoriser l'extraction d'information potentiellement utile pour répondre à la problématique de recherche, ainsi que la forme associée. La forme est en réalité une locution complexe lexicalisée dans une structure syntaxique, que tous les systèmes de reconnaissance automatique de la langue naturelle essaient de désambiguïser. Le sous-type a surtout été attribué aux « commentaires » des sujets, émis durant la phase d'interactions HRI, qui pouvaient donner des informations exploitables pour une analyse qualitative.

**Comm\_surIsol** : cette étiquette est associée aux commentaires sur l'état d'isolement que les sujets évoquent à leur propos ou associent à d'autres. Ce sont de potentiels indices supplémentaires pour caractériser l'état d'isolement des sujets ou des récits de vécus, de ressentis, d'opinions sur le sujet. Une analyse approfondie, notamment psychosociologique, pourrait être travaillée avec les partenaires sociétaux et les aidants familiaux pour une meilleure compréhension collective de phénomène.

**Comm\_surDomus** : il peut s'agir de commentaires principalement ergonomiques de l'environnement, tout comme des perspectives d'amélioration, de focus de besoin, de mise en évidence de problèmes dans leur propre quotidien ou des indices relevant des améliorations du protocole. À noter qu'un affinement de ces productions peut notamment filtrer le même vocabulaire que celui des commandes (voire exactement les mêmes formulations, mais dans une énonciation modale qui n'a pas vocation à être exécutée) et qui pourraient donc être utiles de comparer avec les caractéristiques acoustiques, et particulièrement prosodiques, des « commandes à exécuter », révélant ainsi indirectement les indices attentionnels des sujets vis-à-vis du robot.

**Comm\_surEmox** : ils sont associés aux commentaires sur Emox, qu'il s'agisse de sa forme, son esthétisme, sa voix, sa fonctionnalité ou ses caractéristiques relationnels et socio-affectives. Parfois la frontière est tenue entre cette étiquette et « DiaEmox » puisque la cible des commentaires est ambiguë. L'intérêt est de pouvoir ici avoir le recueil des impressions et les effets que le robot a suscités auprès des personnes. Ces commentaires et la manière dont ils ont évoqué les caractéristiques du robot, pourraient être comparés en fonction des différentes phases de l'expérience, en fonction des personnes présentes dans le contexte interactionnel (en présence de l'aide à domicile, des expérimentateurs, dans le débriefing). Ils peuvent également être mis en regard des auto-annotations pour voir dans quels buts ou intentions ces commentaires ont été produits ou encore voir si cette perception a évolué au cours du temps, après l'expérience.

#### **4.2.5.3.2. Énoncés dialogiques complexes**

**DiaEmox** : ce sont des productions qui sont directement adressées au robot, mais qui sortent du cadre strict des commandes. Ils sont porteurs de marqueurs d'attachement/intimité/*care* et semblent tout au moins être des indices de continuité interactionnelle. Notons notamment que ces éléments sont également ceux qui peuvent traditionnellement être associés aux « fonctions phatiques » dans les théories de la communication, mais il s'agit également de traces interactionnelles comme des imitations ou encore des écholalies. Nous supposons qu'en plus des objets eux-mêmes, l'organisation rythmique globale de ces éléments (au même titre que les ME), pourrait illustrer des formes d'adaptation et de synchronisation associées à des comportements non vocaux (visibles notamment dans les indices de proxémie) adoptés par le sujet.

**DiaAutres** : cette catégorie représente les productions conversationnelles les plus ambiguës (y compris ceux tournés vers soi), dont la cible et le contenu sont encore bien trop complexes pour l'exploiter de manière fonctionnelle. Ces éléments font partie du « bruit » auquel le système doit se rendre aveugle pour gagner en robustesse.

## 4.2.6. *Étiquetage multimodal des productions non vocales des sujets*

### 4.2.6.1. Nomenclature d'étiquetage manuelle : les limites

À l'heure actuelle, il y a 6 sujets qui ont été entièrement étiquetés manuellement sur les productions qui ne sont pas vocales. Du fait de l'approche inductive, nous n'avons pas d'ontologie de description préalable afin d'éviter de passer à côté d'événements rares, mais significatifs. En revanche, les étiquetages des indices multimodaux dans l'interaction sont nombreux, avec même l'existence de normes pour le codage des actes de dialogue par exemple définis par la norme DIAMl ISO- 24617-2. Cette dernière est notamment proposée par le développeur du logiciel d'annotateur vidéo Anvil (Kipp, 2001). Notons que de nombreux travaux s'intéressent à l'étude de la multimodalité en robotique sociale, comme par exemple le projet JOKER (Devillers, Rosset et al., 2015). Le choix de ne pas directement appliquer des normes usuelles est toutefois double : 1) les actes sont codés de manière à étudier l'interaction entre humains et non dans un contexte IHM ou même HRI. Or la spécificité du robot en lui-même pose des contraintes ; 2) cette norme propose un codage en deux zones, une dédiée à la description de chaque acte et l'autre à leurs fonctions interactionnelles, un format que nous conserverons. En revanche, la nature et l'ontologie des fonctions communicationnelles sont fondées sur des interactions prototypiques de contextes comme le style journalistique ou les débats télévisés dont l'écologie diffère fortement de notre contexte et qui sont loin d'être généralisables du fait des rôles sociaux incarnés par chacun. À défaut, voici une proposition d'étiquetage adaptée à la nature et à la qualité des données recueillies<sup>89</sup>.

**DéplacementSuj** : cet alignement marque les phases de déplacement du sujet selon les modalités suivantes :

- devant – attente : le sujet passe devant Emox mais l'attend
- devant – sans attente : le sujet passe devant Emox mais sans l'attendre
- derrière : le sujet suit Emox
- à côté : le sujet passe à côté d'Emox
- enjambement Emox : le sujet enjambe le robot

**Position** : il s'agit de l'emplacement du sujet et du robot l'un par rapport à l'autre à l'état stationnaire (hors déplacement) :

- derrière : le robot est derrière le sujet qui est de dos par rapport au robot
- devant : le robot est devant le sujet lui tournant le dos
- face – prêt : le sujet et le robot sont face-à-face très près (le robot est au pied du sujet)
- face – 1 carreau : le sujet et le robot sont proches face-à-face (distance d'un carreau de moquette)
- face – plus 2 carreaux : le sujet et le robot sont face-à-face relativement éloignés (à plus de 2 carreaux de moquette)
- autre pièce : le robot et le sujet sont dans une autre pièce
- côté – près : le sujet et le robot sont face-à-face très près (le robot est au pied du sujet)
- côté – 1 carreau : le sujet et le robot sont proches face-à-face (distance d'un carreau de moquette)

---

<sup>89</sup> Cf. article (Tsvetanova et coll., 2016) concernant les observations préliminaires de l'analyse multimodale.

- côté – plus 2 carreaux : le sujet et le robot sont face-à-face relativement éloignés (à plus de 2 carreaux de moquette)

**PostureSuj** : il s'agit de la posture générale du sujet

- debout
- assis
- accroupi
- debout penché
- assis penché

**RegardSuj** : idéalement cette annotation devrait s'appuyer sur des mesures oculométriques ou utiliser un programme d'évaluation de l'attention visuelle afin de déterminer avec précision la cible sur laquelle le regard du sujet est porté. De manière générale nous relevons 7 étiquettes pour cette catégorie (annotée uniquement lorsque l'angle de vue de la caméra nous permet de le faire) :

- Emox
- Feuille de commande
- L'objet de la commande domotique
- l'expérimentateur E1
- l'expérimentateur E2
- l'aide à domicile
- l'environnement

Les caméras du plafond utilisées n'étant pas adaptées pour un suivi fin des comportements, les enjeux de l'analyse non vocale sont une part importante de l'évolution méthodologique de la plateforme en elle-même. Le défi serait ainsi de pouvoir maximiser les séquençages sur des bases de paramètres externes et objectifs, à partir de l'automatisation que l'on pourrait obtenir par de traces de capteurs suffisamment robustes et fins, notamment pour la mesurer la proxémie. Il peut aussi affiner le repérage de la distance/position homme-robot qui semble porter également de nombreux indices de l'évolution de la glu socio-affective.

#### **4.2.6.2. Ciblage de la multimodalité par les auto-annotations**

L'étiquetage multimodal proposé reste bien trop chronophage pour l'étendre à un corpus étoffé. De fait, dans le cadre de cette thèse, une autre approche beaucoup plus approximative fut également réfléchi. Elle consisterait à relever les points saillants de la multimodalité en s'appuyant sur les éléments pointés dans le cadre de l'auto-annotation. En effet, sans faire de focus spécifiques sur ces aspects, les sujets évoquent des « moments » dont ils se souviennent particulièrement, qui sont relativement courts et ponctuels (cf. section 5.3.5) par rapport à d'autres moments. Ils mentionnent ainsi ce qui est visible dans leur propre comportement (*p. ex.* « *là je regarde [...]* », « *je m'accroupis pour [...]* » ), celui du robot (*p. ex.* « *quand il tournait comme ça je trouvais que [...]* », « *c'est à ce moment précis quand il a [...]* ») ou encore des éléments contextuels (*p. ex.* « *j'attendais que les stores se ferment complètement pour faire le suivant* »). L'auto-annotation pointe ainsi, soit une mention directe du comportement, soit un marquage « temporel » englobant ce type d'événements permettant de relever les éléments multimodaux potentiellement significatifs. Ces éléments peuvent être des marqueurs de points de rupture/transition interactionnels et relationnels. En effet, le marquage d'un « moment » de l'auto-annotation permet de repérer un endroit où « il se passe potentiellement quelque chose ». Tous ces éléments ont ainsi été dupliqués de la tier d'auto-annotation nommée FoT\_IPC dans une nouvelle tier nommée « **Events** », mentionnant sur ces points saillants en mentionnant la nature du trait de comportement évoqué en suivant la nomenclature suivante :

*[natureComportement : description] / ex : [déplacement : enjambement du robot]*

Ce repérage multimodal est donc aligné au FoT, marqué par le sujet à ces moments-clés. L'étiquetage est encore en cours de réalisation pour les sujets qui n'ont pas été traités par la nomenclature d'étiquetage manuelle. En revanche, une fois l'étape d'auto-annotation terminée, il s'agirait peut-être d'un moyen moins chronophage et plus rapide pour aborder cette dimension multimodale.

#### **4.2.7. Étiquetage des stimuli contrôlés par EmOz**

##### **4.2.7.1. Des traces de l'interaction face-à-face aux traces du magicien**

Les stimuli produits par le robot sont construits sur une base de données de sons, de scripts de mouvements et de déplacements/mouvements sur une manette. Les traces des sons et des scripts de mouvements sont conservées durant l'expérience. La trace de la manette quant à elle n'a pu faire l'objet de sauvegarde pour ces expérimentations, mais pourrait être envisagée dans le développement futur de la plateforme Domus-LIG. De même, les traces des activations domotiques sont également conservées. Pour les productions à fort potentiel communicatif, les marquages sont doubles. Elles sont composées d'une part des événements relevés dans les traces, alignées automatiquement. Elles illustrent les temps de réaction approximatifs des magiciens d'Oz pour exécuter les commandes (notés **Woz**). D'autre part, elles représentent ces mêmes événements observables directement, à l'intérieur de l'appartement, en contexte face-à-face avec les sujets (notés **FtF**). Pour les traces Woz, l'alignement est automatisé, en conservant le temps de départ. Par défaut, toutes les traces alignées automatiquement sont associées à un intervalle de 40msec (sous Elan) car le PointTiers n'existe pas sous ce logiciel. Cela évite de générer les chevauchements liés à l'état de plusieurs éléments du système domotique en se concentrant sur les temps de déclenchement des commandes, notamment pour le calcul de la rythmicité et de la cadence des événements. Dans ce type d'étude HRI, et plus généralement pour toute récolte/observation de comportements naturels/spontanés en sciences humaines, l'expérimentation compose une écologie propre inhérente, dans laquelle les expérimentateurs/observateurs sont eux-mêmes intégrés dans le contexte d'observation. L'étiquetage tente ainsi de relever les informations de manière à pouvoir appréhender cette dimension. Pour faciliter l'alignement automatique des traces, une nomenclature spécifique de nommage des scripts a été adoptée pour catégoriser le type de la primitive et retrouver sa valeur associée de manière systématique (cf. Annexe 3 pour les règles de cette nomenclature, valables pour tous types de scripts utilisés par EmOz).

##### **4.2.7.2. Alignements des commandes domotiques**

**DomusWoz** : il s'agit des traces des clics de déclenchement d'une commande domotique à travers l'interface EmOz par un magicien. À noter que les valeurs restent approximatives en fonction de la qualité variable du réseau et que des perspectives de gain de précision sur la récupération de ces traces sont à réfléchir au niveau de la plateforme elle-même.

**DomusFtF** : ils marquent le moment où les actions domotiques sont enclenchées au sein de l'appartement. Le focus de l'étiquetage est porté sur le début de l'action et par défaut, la valeur de l'intervalle de 40msec est conservée. En revanche, quand la réalisation de la commande domotique n'est pas fluide par rapport à l'exécution par défaut (p. ex. en cas de clignotement de LED, des lumières qui ne s'allument que progressivement, etc.), cette durée est reportée en augmentant la taille par défaut de l'intervalle.

#### 4.2.7.3. Alignements des commandes robotiques scriptées

**EmoxWoz\_son** : il s'agit de l'alignement de la production vocale du robot, soit les primitives testées, dont le temps de déclenchement s'appuie sur celui du clic du magicien qui contrôle les stimuli vocaux d'Emox. Les traces conservées sont celles des scripts ou des bases de sons qui sont produits sur l'interface EmOz, et dont les alignements sont automatisés. Chaque bouton de l'interface correspond à ce type de stimuli dont la trace porte un nom unique pour permettre un repérage précis du bouton déclenché. Ce repérage avait initialement été envisagé pour l'étude de l'ergonomie de l'interface lors des premiers tests, dans le but de réduire les délais du temps interactionnel pour se rapprocher au maximum du temps cognitif du magicien humain.

**EmoxWoz\_mvt** : il s'agit de l'alignement des mouvements interactionnels du robot, codés sous forme de petits scripts d'enchaînement plus ou moins longs de mouvements. L'idée était ici de pouvoir associer de petits mouvements primitifs (cf. Annexe 3). D'une part, cela permettait de pouvoir effectuer un feedback visuel au début de l'expérimentation, sans que le robot produise de son, et d'autre part de conserver ces mouvements pour maintenir un degré d'animicité comme initialement proposé, afin de voir les ruptures de perception animiste du sujet vis-à-vis du robot. Il s'agit là aussi des traces alignées sur les clics du magicien.

**EmoxFtF\_son** : il s'agit de l'alignement de la production vocale du robot qui est audible dans le contexte FtF, soit avec le sujet à l'intérieur de Domus. Si nous avons une plateforme de magicien d'Oz d'une robustesse infaillible, le nombre de stimuli de ce tiers serait identique au nombre relevé dans la trace. Nous verrons que l'interface de notre contexte expérimental génère une écologie propre avec ses imprécisions intrinsèques (cf. Tableau 2 et commentaire infra) mais qui montre aussi la possibilité de perspectives d'analyse de synchronie sur différents niveaux.

**EmoxFtF\_mvt\_script** : parmi les mouvements du robot, certains sont effectués avec une intention d'interaction forte à travers des primitives de mouvements codés (p. ex. avancer, reculer, faire des tours, dandiner, etc.). Ce sont ces mouvements pour lesquels les traces du magicien sont conservées. Ces éléments pourraient gagner en rigueur scientifique avec une validation perceptive, ce qui n'a pu être réalisé en amont de l'expérimentation puisque nous n'avons aucune présomption de valeurs à associer. En revanche, ces mouvements sont désormais alignés sur des valeurs d'auto-annotations qui pourraient permettre de vérifier l'existence d'une sémantique intrinsèque tout en ayant conscience qu'elle est susceptible de changer en fonction de l'état relationnel et du degré d'animisme associé au robot. Un test de perception spécifique à ce type de stimuli peut donc être envisagé sur une perspective à court terme.

#### 4.2.7.4. Alignements manuels des mouvements robotiques à la manette

**EmoxFtF\_mvt\_manette** : il s'agit de l'alignement des mouvements du robot (en particulier les déplacements) réalisés à travers une manette. Certaines primitives comme des hochements de têtes sont également associées à certains boutons de cette manette. Dans l'ensemble, le contrôle du mouvement effectué donne un degré de liberté relativement important au magicien qui le contrôle par ce biais. En revanche, pour ce contrôle, la trace du magicien n'existe pas encore dans le système EmOz actuel, mais techniquement il pourrait être intégré. Dans le contexte FtF, ces mouvements plus libres sont contrôlés, mais restent tout de même à l'origine des nombreux effets non prévisibles, pouvant fortement marquer l'interaction d'un point de vue communicationnel. La variation de la proxémie émane par

exemple de l'intelligence socioculturelle et humaine du magicien, qui guide son adaptation naturelle de la distance homme-robot. Cette distance est d'ailleurs tributaire de l'angle de vue des caméras que regarde ce magicien. En effet, ce dernier a la possibilité de regarder d'une part la caméra du robot, renforçant alors la dimension immersive du magicien comme dans le cadre de la téléprésence, et d'autre part s'appuyer sur la caméra du plafond, en déconnectant fortement ce dernier de la situation d'interaction face-à-face. Ainsi, les paramètres comme la modalité de déplacement, la variation de la vitesse, le positionnement naturel du robot par rapport aux sujets, etc., sont autant d'éléments qui sont laissés ouverts au magicien.

**EmoxFtF\_mvt** : il s'agit de l'alignement qui regroupe l'ensemble des étiquettes *EmoxFtF\_mvt\_script* et *EmoxFtF\_mvt\_manette* afin de récupérer l'ensemble des temps de déclenchement où le robot bouge. En effet, c'est une dimension qui est continue, sans distinction pour le sujet âgé qui interagit avec le robot. En outre, il met en évidence les chevauchements entre ces deux alignements dont il dépend, donnant indirectement des informations sur le niveau de synchronie du magicien-manette et du magicien-interface par exemple.

### **4.3. Auto-annotations et marquage des « temps cognitifs »**

Dans le cadre du corpus EEE, une première constatation est celle de la variabilité et l'étendue de la nature des *Feelings of Thinking* (FoT) recueillis (cf. Chapitre 1 – 1.3.3 pour l'explicitation de cette notion). En effet, le protocole de recueil contrôle de façon scientifiquement motivée l'attention pour accéder aux processus cognitifs communicationnels des personnes et relever « tout ce qui s'est passé dans leur tête » au moment de l'expérience. Or dans une interaction spontanée, ce flux d'informations cognitives marque des temps plus ou moins longs, et véhicule une quantité de valeurs fluctuantes et de différentes natures. Certaines informations semblent perdurer, comme un état latent d'arrière-plan alors que d'autres sont associés à des actions. Elles devraient même certainement les précéder parfois, bien qu'il ne soit pas réellement possible d'obtenir un degré de finesse sur le moment exact de déclenchement par cette méthode qui ne reste qu'une approximation à l'appui d'événements vécus. L'annotation essaie néanmoins de rester au plus près des changements pour voir les frontières de ces différents états. Ainsi dans le corpus, les principales informations ont été arbitrairement catégorisées en quatre groupes. Ces groupes, qui mélangent des informations de même nature, permettent de gérer les chevauchements d'étiquettes semblant se dérouler sur une base temporelle différente. Ils seraient ainsi presque assimilables à des « temps cognitifs » différents du sujet. Cet étiquetage ne vise pas à catégoriser les émotions, les socio-affects, les intentions, etc. mais à marquer les changements d'état de la situation ou de la relation, et indirectement aussi l'« intensité cognitive » qui peut être plus ou moins forte dans l'implication interactionnelle (des sujets et des magiciens d'Oz). L'une des observations possibles par ce biais est par exemple de pouvoir caractériser plus précisément les processus cognitifs en cours durant les silences par exemple. En effet, un silence vocal peut sembler « passif » *a priori*, mais il peut se révéler très actif sur le plan interactionnel. L'activité cognitive associée peut : soit illustrer une continuité interactionnelle (p. ex. réfléchir à ce que l'on veut dire et comment on veut le dire, dans quel but) ; soit s'apparenter à un « décrochage » qui l'oriente vers autre chose (p. ex. le sujet ne dirige plus son attention sur l'interaction menée avec le robot, mais il s'intéresse à un aspect spécifique de l'appartement ou à la tâche prétexte). Ce décrochage peut aboutir à des formes d'états relativement passifs. Alors que l'auto-annotation ne demandait pas explicitement à fournir des étiquettes portant sur ces groupes en particulier, le protocole a induit l'émergence des données récurrentes chez chaque sujet. L'alignement s'appuie donc sur ces éléments récurrents formalisés par les regroupements proposés et codés ci-dessous. Chaque groupe est ainsi numéroté chronologiquement par une nomenclature propre à chacun. L'ensemble des étiquettes regroupées par sujet par regroupement est disponible en Annexe 14. Chaque étiquette n'est pas forcément associée à un seul événement, il peut être répété ou repris ultérieurement sur l'ensemble de l'expérience d'un même sujet.

#### **4.3.1. Intentions à travers les productions de sujets : *FoT\_ProdSuj***

**AA\_FoT\_ProdSuj\_(1-...)** : il s'agit des *FoT* majoritairement intentionnels et motivationnels des sujets, associés aux productions, aussi bien vocales que multimodales. Elles ne se limitent pas en revanche à ces natures puisque les sujets peuvent évoquer d'autres processus ou états pour informer de la valeur socio-affective fonctionnelle associée à ces comportements, partageant cette même base temporelle. En effet, l'auto-annotation focalise les productions vocales (principalement les commandes incluant les autres productions qui gravitent autour) en essayant de faire émerger des *FoT* informant sur « la manière » dont les sujets produisent ces éléments, ce qui fait principalement l'objet de la prosodie socio-affective associée. Le repérage des fonctions communicatives associées sera ainsi un moyen de repérer

les potentiels changements de l'état relationnel des sujets avec le robot. Les valeurs d'étiquettes sont de type : « là c'était un ordre » ; « je teste que tout fonctionne » ; « là je prenais des pincettes pour lui dire » ; « c'est le maître d'école qui parle à son élève » ; « là c'est plus cordial, c'est gentil » ; etc.

#### **4.3.2. Temps latents généraux : FoT\_Gn**

**AA\_FoT\_Gn\_(A-...)** : ces *FoT* sont de natures plus variables, mais donnent une information temporelle plus longue, illustrant notamment l'état d'esprit ou des émotions de la personne au cours de ces processus d'interaction, mais il peut s'agir d'autres processus ou états plus longs, en arrière-plan durant l'interaction. Lors de l'annotation, l'apparition de l'étiquette est nette, puisque les sujets se souviennent de ces états généraux (p. ex. « je suis stressée de savoir si je vais y arriver » ; « c'est la curiosité de découvrir » ; « je suis en train de m'amuser avec le lui, je le prends comme un jeu » ; etc.). Ce sont des états qui disparaissent rapidement de leur discours d'annotation, mais qui lorsqu'ils sont repris, montrent qu'ils restent encore « en cours » au moment où d'autres étiquettes de type *FoT\_ProdSuj* sont mentionnées. L'expérimentateur-interrogateur reposait fréquemment la question de savoir si le sujet était toujours dans cet état pour s'assurer de sa persistance. Ainsi, dès lors qu'une étiquette de cette nature apparaissait, le protocole d'étiquetage imposait de maintenir cette étiquette jusqu'à l'apparition d'une nouvelle étiquette générale, et ce tant que le sujet confirmait être dans cet état. Les indices semblent ainsi relever des aspects plus situationnels et contextuels, ancrés sur de plus longues périodes, tout en ayant des passages parfois plus brefs d'états de temps à autre. Ils restent tout de même toujours plus longs ou de longueur équivalente aux productions du sujet, mais avec des valeurs informatives de nature plus contextuelle qu'effective.

#### **4.3.3. Indices ponctuels communicationnels : FoT\_IPC**

**AA\_FoT\_IPC\_(a-...)** : ces étiquettes sont des « indices ponctuels communicationnels ». Les natures de *FoT* associée sont variables, mais elles portent sur les effets d'événements précis et généralement plus brefs dans le temps, quasiment du même ordre que les productions des sujets. En revanche, ces événements interactionnels co-occurrent parfois avec un comportement du sujet, tout en le modifiant en même temps. Ce regroupement est donc également associé à une modification perceptive du sujet, dont les origines sont multiples : à savoir le robot, l'appartement ou un élément du contexte qui provoque un effet sur le sujet. Il peut également être un acte interactionnel initié par le sujet, portant sur des éléments autres que des commandes, associés aux productions multimodales qu'il n'était pas forcément évident de repérer ou qui sont co-occurrents à des productions vocales (p. ex. « le fait qu'il me fixe comme ça, ça me mettait un peu mal à l'aise » ; « j'avais l'impression qu'il était content quand il tournait comme ça » ; « je me dis, il me pose une question, je dois répondre » ; « là je regarde pour que ça se ferme complètement avant de lui faire la commande suivante »). En effet, des événements co-occurrents avec valeurs informatives différentes ne pouvant être annotés sur une même ligne temporelle, le choix a été de créer ce nouveau groupe d'alignement.

Par ailleurs, à défaut de ne pouvoir étiqueter toutes les productions multimodales, et cela de manière plus efficace, systématisée et automatisée (cf. section 4.2.6), les intervalles *AA\_FoT\_IPC* permettent de pointer les événements dont les sujets se souviennent plus particulièrement et qui sont donc susceptibles d'être des points de rupture de la dynamique communicationnelle.

### 4.3.4. Impressions perceptives d'Emox : FoT\_PE

AA\_FoT\_PE\_(I...): il s'agit de la perception très globale que le sujet évoque pour décrire sa relation au robot. La question n'est pas posée véritablement explicitement, elle s'appuie sur des propos accompagnant leur manière de faire, en repérant des mots clés qu'ils ont associés aux trois groupes précédents. En revanche, dans les transcriptions des auto-annotations, ces éléments faisaient l'objet de nombreuses répétitions de ces mots-clés qui montre de manière indirecte la perception que le sujet a du robot. Parfois les sujets eux-mêmes explicitent cette information qui est ainsi exploitée pour l'alignement. Ces étiquettes décrites spontanément par le sujet surviennent à des moments précis ou dans leur déviation de l'étiquetage de la commande, en utilisant les termes de type « je m'adresse comme si c'était un... » ou « là pour moi c'est un... ». De cette manière, nous pouvons avoir une manière approximative et simpliste d'illustrer l'état de la glu à travers la perception que le sujet a du robot, mais dont chaque changement est déclenché sur la base d'événements interactionnels. Il ne s'agit donc pas d'un recueil d'avis générique conscient. Il s'agit de la base temporelle la plus sous-jacente avec les phases les moins changeantes et généralement plus longues.

### 4.3.5. Alignements multiples d'auto-annotations chevauchées

Tous les éléments ambigus ou trop imprécis des auto-annotations n'ont pas été conservés pour éviter les faux positifs d'information. Chacun des groupes temporels s'aligne donc sur des événements ou des états à durée variable. Ces étiquettes peuvent donc se chevaucher entre elles et sont marquées au sein de différents tiers Elan, comme le montre le schéma simplifié en Figure 45.

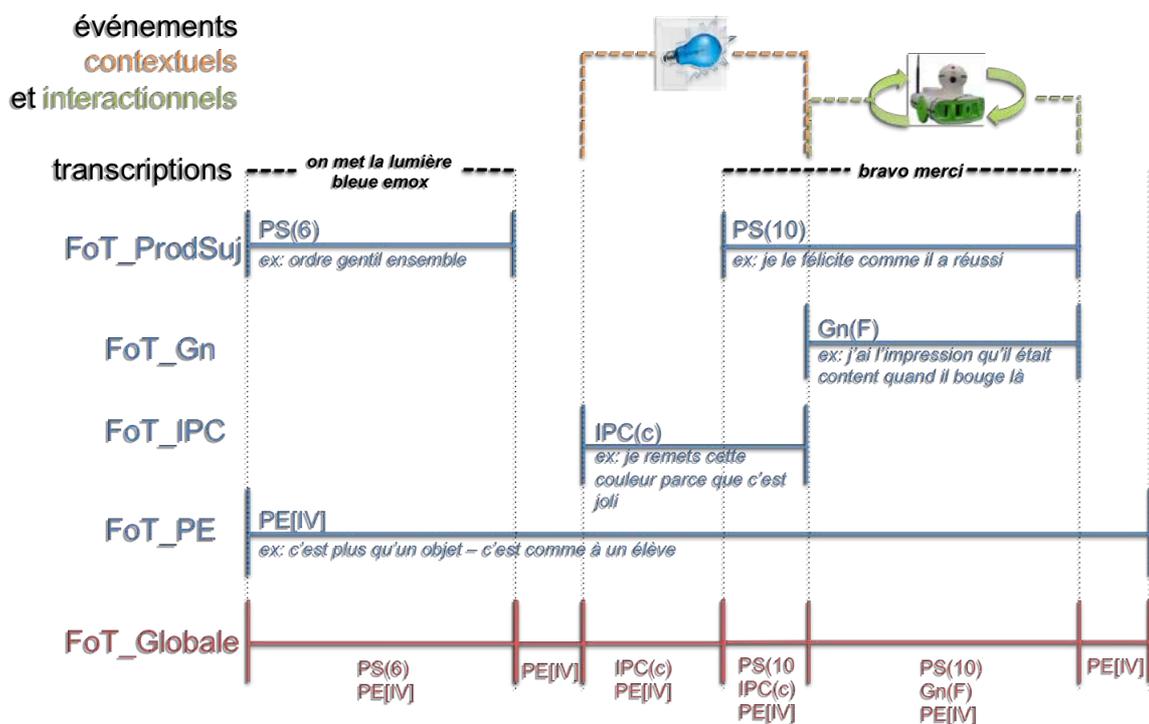


Figure 45 : Exemple simplifié d'alignement d'auto-annotations

Le schéma illustre la manière dont l'alignement de ces étiquettes est effectué pour un sujet donné sur une très courte séquence interactionnelle (ici elles sont représentées sur les intervalles en bleu). Cet alignement est reproduit sur l'ensemble de la phase expérimentale de chacun des sujets. L'ensemble des étiquettes fournies sont présentées en Annexe 14.

Certaines étiquettes n'ont été associées qu'une seule fois et constituent des événements rares. Cette rareté peut parfois être un indice qui pourrait potentiellement faire basculer la relation et la progression de la glu (ex : sur un échange précis, une étiquette « ça c'est extraordinaire parce que je l'ai imité et pas du tout parce qu'il l'a fait » qu'on ne retrouve pas ailleurs et qui coïncide avec un changement d'étiquette de FoT\_PE). Inversement, d'autres événements sont associés à une même étiquette d'auto-annotation, soit parce qu'une suite de commandes est effectuée avec une même intention/motivation par le sujet ; soit du fait de retours en arrière ou de répétitions d'un contexte ou d'événements interactionnels qui privilégient une étiquette. Le chapitre suivant essaiera d'ailleurs d'illustrer ces changements de dynamique temporelle.

Enfin, bien que les étiquettes d'auto-annotation soient propres à chaque sujet, nous retrouvons une similitude d'étiquettes entre celles que certains sujets donnent, soulignant notamment une évolution globale, qui renforce positivement ou négativement la relation et donc l'état de la glu. Une perspective possible cet alignement parallèle serait notamment d'associer et de systématiser une nomenclature d'étiquette comparable sur des objets interactionnels dont le fonctionnement serait semblable, mais elles devront être soumises à validation perceptive.

Par ailleurs, un travail qui n'est pas encore terminé, mais mentionné ici en rouge sur la Figure 45, illustre un groupe nommé *FoT\_Global*. Ce dernier groupe montre que si les chevauchements des quatre autres groupes temporels sont regroupés sur une même base temporelle, elles peuvent aussi faire apparaître un séquençage des interactions, mais avec des densités d'activité et des valeurs informationnelles différentes. Si le passage d'un groupe temporel nécessite un changement d'échelle de visualisation, cette dernière représentation permettrait de rester sur une même échelle tout en conservant les valeurs associées à ces temps. Elle s'opposerait ainsi d'une certaine manière au séquençage arbitraire « Séq » initialement réalisée pour l'auto-annotation. Elle en serait ainsi une version dont les frontières interactionnelles seraient « corrigées ».



## Chapitre 5. Analyses des productions vocales : vers une prosodie du dialogue ?

### 5.1. Aperçu quantitatif des données

#### 5.1.1. Proportion générale des productions vocales

Les analyses quantitatives sont réalisées sur les 16 sujets dont les étapes d'auto-annotation, d'extraction des étiquettes et de leurs alignements temporels sont exploitables. Le Tableau 1 ci-dessous récapitule le nombre d'occurrences des productions vocales (ProdVoc) annotées sur la partie « Expe » de chaque sujet. Il présente également la proportion des Cmd, NwCmd, NonCmd par rapport aux occurrences de ProdVoc.

Tableau 1 : Répartition des différents types de productions vocales des sujets âgés

SUJETS	ProdVoc	Cmd	NwCmd	Non Cmd
aSa1	785	10,32%	1,91%	87,77%
S01	347	21,04%	6,92%	72,05%
S03	430	33,02%	0,93%	66,05%
S04	492	17,68%	2,03%	80,28%
S05	540	10,19%	0,74%	89,07%
S06	477	27,46%	0,84%	71,70%
S07	623	7,87%	2,73%	89,41%
S10	786	5,85%	4,83%	89,31%
S11	573	6,11%	2,79%	91,10%
S12	313	15,97%	0,96%	83,07%
S14	278	15,83%	0,72%	83,45%
S15	670	10,75%	3,13%	86,12%
S16	411	19,71%	1,46%	78,83%
S17	224	7,14%	0,89%	91,96%
S18	248	17,34%	1,61%	81,05%
S19	354	8,47%	0,85%	90,68%
<b>Total</b>	<b>7551</b>	<b>1035</b>	<b>173</b>	<b>6343</b>
<b>Moyenne</b>	<b>471,94</b>	<b>64,69</b>	<b>10,81</b>	<b>396,44</b>
<b>Ecart-type</b>	<b>146,31</b>	<b>26,77</b>	<b>8,27</b>	<b>143,67</b>
<b>Min</b>	<b>224</b>	<b>16</b>	<b>2</b>	<b>201</b>
<b>Max</b>	<b>786</b>	<b>142</b>	<b>38</b>	<b>702</b>

Sur un total de 7551 productions vocales soit un peu moins de 500 productions par sujet, nous constatons que ces derniers produisent en moyenne environ 65 commandes par expérience, mais avec un grand écart-type dont la production varie de 16 à 142 commandes. Une dizaine de NwCmd est produite à chaque expérience, et au regard de l'Annexe 11, nous constatons que ces dernières correspondent principalement à du guidage spatial (p. ex. « aller au salon », « me suivre ») pour les changements de pièce ou pour indiquer un emplacement afin que le robot s'approche de cette cible (p. ex. aller près de, avancer, reculer etc.). Les

autres sollicitations concernent les appareils électriques (p. ex. micro-onde, ordinateur, aspirateur), les parties sanitaires (p. ex. faire couler l'eau, le savon, baisser la cuvette des toilettes, ouvrir les portes de la douche etc.). Enfin, environ 84% des productions concernent les NonCmd, dont la nature sera explicitée dans la section suivante. L'enjeu est donc extrêmement important, car en supposant que cette technologie passe dans une écologie réelle d'usage mettant en jeu ce système (robot contrôleur d'appartement intelligent), seulement 20% de la parole (alors que ce scénario maximise d'ailleurs la production des commandes) ne relève que du vocabulaire d'un système RAP dédié à la tâche et qui doit être filtré à travers ce flux du langage naturel.

### 5.1.2. Proportion des actions domotiques et robotiques

Tableau 2 : Répartition des stimuli produits du robot Emox (Woz vs. FtF)

SUJETS	EmoxWoz son	EmoxFtF son	EmoxWoz mvt	Proportion approx. mvt manette	EmoxFtF mvt	EmoxFtF
Sa1	-	37,85%	0	62,15%	62,15%	288
S01	57	42,64%	9	50,39%	57,36%	129
S03	118	53,95%	27	33,49%	46,05%	215
S04	-	35,22%	-	-	64,78%	247
S05	90	34,80%	47	46,40%	65,20%	250
S06	187	36,18%	34	55,28%	63,82%	398
S07	123	30,32%	62	47,29%	69,68%	277
S10	178	62,37%	51	19,35%	37,63%	279
S11	108	44,49%	22	46,19%	55,51%	236
S12	82	43,58%	16	47,49%	56,42%	179
S14	87	38,21%	68	29,72%	61,79%	212
S15	140	57,81%	46	22,78%	42,19%	237
S16	142	44,34%	136	12,89%	55,66%	318
S17	28	17,39%	33	62,11%	82,61%	161
S18	76	31,20%	82	33,76%	68,80%	234
S19	35	23,85%	68	23,85%	76,15%	130
<b>Total</b>	<b>1451</b>	<b>1530</b>	<b>701</b>	<b>1399</b>	<b>2260</b>	<b>3790</b>
<b>Moyenne</b>	<b>103,64</b>	<b>95,63</b>	<b>46,73</b>	<b>93,27</b>	<b>141,25</b>	<b>236,88</b>
<b>Ecart-type</b>	<b>38,64</b>	<b>32,08</b>	<b>24,92</b>	<b>39,39</b>	<b>37,28</b>	<b>49,88</b>
<b>Min</b>	<b>28</b>	<b>28</b>	<b>0</b>	<b>31</b>	<b>74</b>	<b>129</b>
<b>Max</b>	<b>187</b>	<b>174</b>	<b>136</b>	<b>220</b>	<b>254</b>	<b>398</b>

En moyenne, parmi les 3790 productions robotiques, un peu plus d'une centaine de feedbacks vocaux est produite par expérience. Cela représente environ le tiers de toutes les productions du robot. Le total des traces FtF est ici plus important, car parfois, des problèmes techniques sont apparus lors de certaines expériences (notamment pour S04 et aSa1), et donc de manière générale, les traces Woz sont plus nombreuses que les traces FtF pour chaque sujet. En effet, ceci est lié à l'évolution technique de l'interface EmOz avec l'intégration du middleware OpenHab (avec gain de finesse de certaines traces telles que le volume sonore des enceintes). Mais en réalité, il reflète surtout l'état interactionnel du magicien, illustrant notamment le décalage de sa propre synchronie avec le sujet qui représente son propre interagissant à travers le robot. Ainsi, les itérations de clics sur un même bouton peuvent, par exemple, illustrer une frustration que le stimulus du robot n'arrive pas suffisamment rapidement dans le contexte FtF. L'évolution de ces clics montre parfois des anticipations (c.-à-d. les traces des magiciens montrent que le déclenchement d'une commande précède parfois la production vocale de cette commande par le sujet), soulignant l'existence d'une inter-

synchronie illustrant les caractéristiques d'un « *embodied* » magicien. En effet, le magicien expérimentateur s'adapte progressivement au rythme du sujet, mais à travers un « corps technologique » qu'il incarne de plus en plus (pas seulement robotique, mais aussi domotique, puisqu'il s'agit d'une fonction intégrée à celle du robot). La complexité est ici de voir que d'une part le magicien évolue à travers cette technologie, mais que d'autre part, cette évolution ne peut s'appréhender qu'à travers son interaction avec le sujet, qui lui-même évolue aussi vis-à-vis du système. Le tout conduit par le changement graduel des stimuli testés, modifie ou non l'état relationnel. L'enjeu d'un développement méthodologique futur serait ainsi de conserver les données interactionnelles de la régie pour que les magiciens puissent réaliser eux aussi des auto-annotations en tant qu'acteurs du système, afin de pouvoir exploiter au mieux les traces qu'ils laissent sur EmOz et de pousser les réflexions sur la corporéité. Plus généralement, cela illustre que dans une expérience, en particulier celle qui implique des sujets humains, l'expérimentateur fait intrinsèquement partie de cette expérience. En effet, même s'il la déclenche et qu'il la contrôle, en partie sur les contraintes des hypothèses théoriques la sous-tendant, son intelligence naïve est également l'un des processus qui doit être observé. De fait, ses données doivent également faire partie intégrante des systèmes produits à partir des traces/données.

Concernant les mouvements, ils représentent les deux tiers des productions du robot, ce qui montre une part relativement importante de l'interaction spatiale. Cela correspond à environ une cinquantaine de stimuli scriptés (c.-à-d. des actions à but explicitement interactionnel) par expérience. La proportion de ce type de mouvements par rapport à ceux de la manette est relativement variable selon les expériences. Nous supposons que la variabilité de cette dynamique peut aussi être liée à la nature de l'isolement du sujet. Certains sujets sont également beaucoup plus mobiles que d'autres en fonction de leur degré de fragilisation physique. Mais plus intéressant, au sein d'une même expérience, la répartition/rythmicité et la coordination des mouvements-manette et des mouvements-scripts semblent également varier, supposant relever de la variation de la synchronie entre les magiciens. La difficulté de visualisation de cette donnée relève du « séquençage ». Pour relever la proportion de chacun de ces éléments, il faudrait pouvoir visualiser ces informations en regard de la répartition des objets interactionnels avant et après chaque point de rupture potentiel.

### ***5.1.3. Proportion des étiquettes d'auto-annotation***

Les occurrences d'intervalles alignées sur les auto-annotations, soit le nombre de phases temporelles, ont été également comptabilisées. D'après le Tableau 3, les niveaux d'annotations les plus étendus sont la FoT\_PE qui informe du niveau de perception générale d'Emox sur une moyenne de 7 états par sujet, puis la FoT\_Gn qui présente une quinzaine d'états généraux du sujet au cours de l'expérience. Le nombre de FoT\_ProdSuj qui rapporte les valeurs informatives communicationnelles des productions du sujet est très variable. Il en est de même pour le nombre de FoT\_IPC illustrant les « moments » ponctuels spécifiques d'interaction dont les sujets se sont particulièrement souvenus. En effet, cette variation dépend de la quantité de commandes et de la précision des souvenirs du sujet (cf. section 5.3.5.2).

Tableau 3 : Tableau d'occurrences d'auto-annotation par type de FoT

SUJETS	FoT_ProdSuj	FoT_Gn	FoT_IPC	FoT_PE	FoT_Global
aSa1	109	35	7	10	253
S01	69	39	8	16	183
S03	139	12	46	7	-
S04	81	39	9	9	198
S05	55	21	32	8	-
S06	131	8	47	4	-
S07	49	26	32	8	-
S10	48	6	54	5	-
S11	33	9	27	3	-
S12	50	12	28	9	-
S14	44	12	36	4	-
S15	96	12	43	12	-
S16	100	6	19	5	-
S17	17	8	15	3	-
S18	53	8	15	5	-
S19	36	15	21	4	-
<b>Total</b>	<b>1110</b>	<b>268</b>	<b>439</b>	<b>112</b>	<b>634</b>
<b>Moyenne</b>	<b>69,38</b>	<b>16,75</b>	<b>27,44</b>	<b>7,00</b>	<b>211,33</b>
<b>Ecart-type</b>	<b>29,97</b>	<b>9,53</b>	<b>12,31</b>	<b>2,88</b>	<b>27,78</b>
<b>Min</b>	<b>17</b>	<b>6</b>	<b>7</b>	<b>3</b>	<b>183</b>
<b>Max</b>	<b>139</b>	<b>39</b>	<b>54</b>	<b>16</b>	<b>253</b>

Une piste exploitable pour l'analyse de ces auto-annotations serait un découpage temporel dont les frontières dépendraient de la densité et de la nature des *FoT*, soit le *FoT\_Global*. Cela pourrait potentiellement donner lieu à des temps interactionnels qui ne sont pas seulement basés sur l'alternance des productions vocales et des silences, mais qui s'appuieraient sur l'ensemble des événements pouvant venir modifier le système global de communication auquel le sujet et le robot appartiennent. Ceci permettrait d'avoir une idée de la cadence du rythme interactionnel. La motivation est la même que lors de la création de l'alignement « Seq » (cf. Chapitre 4 – 4.2). Il s'agirait de pouvoir marquer des « moments d'interaction », mais ici la frontière n'est plus arbitraire, elle suit le fil des transitions pointées par les sujets lors de l'auto-annotation.

Sur le séquençage de *FoT\_Globale* réalisé pour trois sujets, nous remarquons que cette dernière n'équivaut pas à la somme d'étiquettes des quatre autres catégories de *FoT*. Elle donne un nombre plus important de séquences que cette somme, notamment du fait de la nature des combinaisons issues des chevauchements d'étiquettes. De plus, en croisant succinctement le Tableau 1 et le Tableau 2, le nombre de de séquences proposé par le sujet lui-même au cours de l'auto-annotation est moins important que le nombre de séquences défini *a priori* sur la base des événements interactionnels considérés séparément. Ce regroupement permet ainsi de rassembler des ensembles de tours de parole qui s'enchaîneraient sur un même temps interactionnel marqué par une certaine valeur communicationnelle. Sa temporalité reste néanmoins imprécise, car nous associons tout de même ces étiquettes à des événements physiques, alors que l'information produite fait certainement l'objet de phénomènes d'anticipation ou autres processus cognitifs dont la durée et les points d'attaque sont difficilement accessibles par les méthodes adoptées.

## 5.2. Des marqueurs de changements de glu

### 5.2.1. Modification des formes morphosyntaxiques des commandes

#### 5.2.1.1. Patterns de paraphrasage des commandes

<b>Forme imposée</b>
<ul style="list-style-type: none"><li>• forme infinitive de la commande comme apparaissant sur la liste imposée dans les consignes</li></ul>
<b>Forme au présent de l'impératif (vous)</b>
<ul style="list-style-type: none"><li>• uniquement avec la seconde personne du pluriel</li></ul>
<b>On va</b>
<ul style="list-style-type: none"><li>• associé aux premiers changements de qualité de voix (breathiness) et des marques d'attention (regard) qui passe de la feuille de commande à Emox</li></ul>
<b>Forme au présent de l'impératif (tu/nous)</b>
<ul style="list-style-type: none"><li>• majoritairement à la deuxième personne du singulier, mais également des occurrences à la première personne du pluriel</li></ul>
<b>(Il) vaut mieux - (Il) faut</b>
<ul style="list-style-type: none"><li>• - formes impersonnelles avec des semi-auxiliaires avec une idée de devoir/obligation</li></ul>
<b>On au présent de l'indicatif</b>
<ul style="list-style-type: none"><li>• on + : la forme la plus commune, associée généralement à un pitch globalement plus haut que les formes précédentes</li><li>• on a dit : souvent associé à la reprise d'erreur</li><li>• on peut</li></ul>
<b>Tu – avec focus prosodiques / inférences pronominales</b>
<ul style="list-style-type: none"><li>• présent de l'indicatif : souvent associé à l'apparition de nominatif (Emox)</li></ul>
<b>Je – le sujet donne son opinion avec justifications</b>
<ul style="list-style-type: none"><li>• « je trouve que tu pourrais »</li><li>• « je vais te demander de »</li><li>• « je voudrais / j'aimerais »</li></ul>
<b>Tu - interrogatif</b>
<ul style="list-style-type: none"><li>• « tu peux »</li><li>• « tu veux »</li><li>• =&gt; d'abord avec "pouvoir" ensuite avec "vouloir". Ces formes sont soit effectuées avec une prosodie montante en fin d'énoncé, soit produites avec une inversion sujet-verbe sous la forme interrogative (parfois plus explicite avec des formes comme "est-ce que ")</li></ul>
<b>Tu nous / tu me – liberté de la forme des commandes</b>
<ul style="list-style-type: none"><li>• au présent de l'indicatif, ce sont généralement des formes pronominales qui peuvent s'accompagner d'inférences et d'inversions, avec également une cooccurrence d'itérations enchaînées ex: tu nous mets en bleu, tu mets en bleu la lumière</li></ul>

Figure 46 : Etapes d'évolution morphosyntaxique des commandes<sup>90</sup>

<sup>90</sup> Ce travail a fait l'une des principales analyses du mémoire de recherche IdL de L. Tsvetanova (2015) réalisées d'abord sur 5 sujets, étendues à 12 dans le cadre du projet Interabot. La dernière révision réalisée dans le cadre de cette thèse résume ces évolutions sur les observations des 16 sujets d'analyse retenus.

L'Annexe 12 illustre l'évolution de la forme morphosyntaxique des commandes vocales des sujets, dont les ordres d'apparition sont illustrés chronologiquement. Chaque forme est associée au nombre d'occurrences dans cette annexe. Il s'agit d'une approximation obtenue par filtrage des commandes. Globalement, deux grands comportements sont visibles : 1) les formes morphosyntaxiques des commandes varient graduellement ; 2) la forme des commandes reste inchangée ou très peu variable, car certains sujets ne s'écartent pas des consignes imposées.

Les modifications morphosyntaxiques, lorsqu'elles ont lieu, semblent suivre une gradualité et une transversalité similaires chez les sujets. De manière générale, l'évolution observée passe par les 10 étapes présentées Figure 46. L'observation systématique ne signifie pas que tous les sujets passent par toutes les formes morphosyntaxiques présentées. En effet, les personnes « sautent » certaines étapes d'évolution. En moyenne, les sujets ne présentent que 5 à 6 formes lorsque leurs commandes évoluent. En revanche, le recoupement transversal des 16 sujets analysés a permis de voir que l'ordre d'apparition des formes morphosyntaxiques reste toujours globalement identique. C'est de ce constat que la Figure 46 est produite et résume ces formes dans leur ordre d'apparition chronologique. Par ailleurs, les sujets qui suivent cette évolution, de manière plus ou moins partielle, sont néanmoins ceux qui sont décrits comme « glués » positivement avec le robot (cf. section 5.3.2.3).

En effet, ce constat repose sur une dimension mentionnée également dans la section 5.3.2.3, qui a aussi été montrée à travers des travaux réalisés parallèlement en 2015. Ces études<sup>91</sup> ont convergé de manière fortuite sur des analyses à orientations linguistique-informatique et sociologique. L'analyse des modifications morphosyntaxiques des commandes semblait faire apparaître des groupes de sujets dont les comportements d'évolution étaient semblables. En parallèle, des travaux sociologiques qui tentaient de caractériser le type d'isolement des sujets à partir d'indices totalement externes ont vu émerger les mêmes groupes de sujets. Ainsi, la tendance constatée était que les personnes, dont la perception d'isolement était moindre, avaient beaucoup moins tendance à paraphraser les commandes, alors que plus les personnes se sentaient isolées et plus elles avaient des tendances de paraphrases marquées. Ici, ce qui est important, c'est de bien constater qu'il s'agit de la perception/sentiment d'isolement et non uniquement du contexte géographique, sociodémographique ou socioculturel qui donne lieu à ces observations. Le comportement interactionnel des personnes semble donc assez fortement dépendre de ce sentiment d'isolement de la personne.

### **5.2.1.2. Cas de non-paraphrasage : indices équivalents**

L'observation sur l'isolement en lien avec l'évolution morphosyntaxique des commandes est un premier fait observable, mais qui reste à affiner, car le comportement interactionnel semble encore plus subtil. En effet, pour les sujets qui montrent peu de changements morphosyntaxiques sur leurs commandes, il existe des variations associées à d'autres dimensions que la morphosyntaxe. Ainsi, des indices se retrouvent au niveau prosodique ou dans une ou plusieurs autres modalités de communication. En effet, si nous regardons les auto-annotations, les personnes avec peu de variations sur cette dimension ne sont pas forcément seulement les personnes qui « n'ont pas glué ». Une évolution positive des auto-annotations, avec d'importantes activités cognitives de nature interactionnelle, a également été observée auprès de ces personnes. En particulier, au début, nous avons des

---

<sup>91</sup> Les travaux concernés sont d'une part le mémoire de linguistique informatique (L. Tsvetanova, 2015) et d'autre part le mémoire de gérontechnologie (M.C. Dikebelay, 2015). Les travaux de Dikebelay consistaient en une approche sociologique par entretiens semi-directifs type « récit de vie » interrogeant les mêmes sujets que ceux du corpus EEE pour une tentative de précision et de caractérisation du sentiment d'isolement des personnes.

## Analyses des productions vocales : vers une prosodie du dialogue ?

---

étiquettes décrivant « l'application » du sujet qui s'accompagne d'une hyperarticulation. Puis, sans changements morphosyntaxiques, cette hyperarticulation s'amoindrit, en donnant lieu à :

- des élisions de mots ou des changements de verbes notés comme « variantes » (p. ex. « mettre la lumière » qui devient « allumer ») ;
- des marques d'inférences pronominalisées (« les allumer ») ;
- des ajouts adverbiaux (« maintenant », « ici », « là ») ;
- l'apparition de micro-expressions telles que des rires ou des imitations (qui semblent d'ailleurs plus nombreuses chez ces sujets n'effectuant pas de changement morphosyntaxique).

Les auto-annotations sont quant à elles : 1) soit empreintes d'une dimension de *care* qui justifie la non-variation de la forme morphosyntaxique (p. ex. « *je fais attention à ce que je dis* », « *je faisais mon apprentissage, je voulais qu'il comprenne ma voix* »), 2) soit illustratives d'une volonté de continuité communicative (p. ex. « *je réfléchis à ce que je pourrais lui dire d'autres* », « *là je voulais le voir répondre* », « *dialogue d'entente gentille* », « *je lui parle, il me répond, c'est une communication* »). Ces auto-annotations sont différentes pour « les personnes qui gluent moins ». En effet, elles décrivent le même type de formes, mais avec un désintérêt (p. ex. « *je fais passer le temps* », « *je sais plus quoi faire* », « *c'est mécanique ma manière de faire là* »).

Par ailleurs, même si la forme de la commande ne change pas en soi, les ME « périphériques » apparaissent progressivement et sont associées à des étiquettes de type : « *je lui commente pour essayer de l'aider de comprendre ce qu'on fait – pour moi c'était le langage des personnes futures* ». De plus, pour ces personnes, la variation morphosyntaxique semble visible pour les productions de type « NwCmd ». Ces commandes produites spontanément, à l'initiative du sujet, sont notamment observables lors des changements de pièce, où même les sujets qui ne présentent pas de variations morphosyntaxiques des commandes utilisent des énoncés de type « *allez on va dans le salon* », « *tu me suis Emox* » etc. qui reflètent bien les évolutions pronominales observées. Il existerait donc des sujets présentant peu de variations morphosyntaxiques sur la commande, mais qui seraient tout de même « glués » au robot. De même, il existerait des comportements dans d'autres modalités qui évolueraient de la même manière que la morphosyntaxe, en étant chargés d'information communicationnelle évoluant de manière semblable.

## 5.2.2. Nature et organisation des « micro-expressions vocales »

### 5.2.2.1. Proportion générale des événements hors commandes

Même s'il existe une consigne explicite de commandes vocales à effectuer, les sujets produisent un nombre important d'objets périphériques hors commandes. Le Tableau 4 récapitule la proportion de ces objets qui gravitent autour de la commande.

Le Tableau 4 est scindé en 3 objets : les SatAnte, les SatPost et les périphériques présentés dans la section 4.2.5.2 du Chapitre 4. Ces éléments représentent des objets étiquetés comme des NonCmd. Une distinction est faite entre ces éléments en fonction de la durée des silences et la position des objets par rapport au noyau de la commande.

Tableau 4 : Répartition des objets vocaux interactionnels hors commandes

	ME	FC
<b>Nb SatAnte</b>	117	26
<b>Proportion</b>	1,72%	0,38%
<b>Nb SatPost</b>	47	51
<b>Proportion</b>	0,69%	0,75%
<b>Périphériques</b>	3044	2331
<b>Proportion</b>	44,71%	34,24%

La première observation notable est que le nombre d'éléments périphériques est beaucoup plus important que les éléments satellites. Ce qui signifie pour un système RAP que la désambiguïsation ne serait pas à faire majoritairement sur la cible du système (soit la commande vocale), car le nombre de satellites (3,19% de satellites par rapport à l'ensemble des productions vocales annotées) est moindre par rapport au reste (40,3% de micro-expressions et 30,8% de formes complexes, soit 71,2% d'objets périphériques sur l'ensemble des productions vocales annotées). Cela signifie également que sur les 7551 productions vocales annotées sur ce corpus, environ 40% d'objets interactionnels ne seraient pas traités par les systèmes RAP (ils seraient soit considérés comme du bruit, soit non détectés). Or, ces éléments portent des valeurs non négligeables d'un point de vue communicationnel au regard des auto-annotations. De plus, environ 30% des occurrences de parole, représentant des formes complexes (notés FC), bruyeraient le système, puisqu'il s'agit de parole non-dédiée à être reconnue. (NB : ici nous parlons bien d'occurrences de segments de parole et non de la durée en pourcentage. En effet, ce dernier chiffre serait bien plus important, puisque ces formes complexes sont beaucoup plus longues que les commandes, car elles désignent des commentaires et des énoncés conversationnels du sujet).

Pour les éléments périphériques, la distinction entre éléments antérieurs/postérieurs n'est pas réalisée, car la discrimination est impossible ou serait totalement interprétative. Le Tableau 5 illustre que la majorité de ces productions concernent des micro-expressions vocales de type « bruits de bouche » (notés BB). Or, la valeur informative, et donc l'emplacement vis-à-vis de la commande, n'est pas directement définissable. Le contexte ne renseigne pas si ce bruit est un élément qui introduit la commande, s'il apparaît comme un élément de continuité de la commande, ou s'il est encore totalement indépendant de cette commande. Seule l'auto-annotation donnerait cette information. Cependant, sur l'ensemble du corpus, les objets périphériques semblent majoritairement précéder ou suivre les feedbacks d'Emox et apparaissent après la commande. La dépendance à la commande et l'apparition de ces éléments semblent être déterminées par : d'une part, la durée du « silence » du système (soit en termes de feedback vocal d'Emox, soit au niveau de l'exécution de l'action

domotique), d'autre part, la nature du bruit émis par Emox et le niveau de glu atteint à cet instant.

### 5.2.2.2. Observations sur les micro-expressions « satellites »

Parmi les satellites, les micro-expressions qui précèdent la commande sont plus nombreuses que celles qui sont postérieures. Pour les formes complexes, la tendance est inversée. D'un point de vue descriptif, elle montre que les sujets produisent plus de bruits courts avant la commande, alors que ce qui suit la commande présente une forme plus complexe. La répartition détaillée du type d'éléments satellites produits (cf. Tableau 5), montre que ces objets sont majoritairement des interjections. L'intérêt est que ces éléments pourraient donc être potentiellement rajoutés dans le lexique et le modèle de langage d'un système RAP. Cela serait moins évident pour les BB définis par des descriptions acoustiques.

De même, la dernière colonne « Pas de sat » détermine le nombre de commandes qui ne sont pas associées à des éléments satellites. Cela représente 902 à 948 occurrences sur 1035 commandes annotées). Par rapport au nombre de satellites, cette répartition n'est pas forcément étendue sur l'ensemble. Ces objets ne concernent qu'un nombre réduit de commandes qui ont été observées à un moment relativement avancé de l'expérience et non au début, sauf pour certaines interjections (p. ex. « alors ») ou certains événements prélexicaux (p. ex. « euh »).

**Tableau 5 : Répartition des différents types de satellites observés**

	Nominatif	PreLex	Interj	BB	Précision	Autres FC	Pas de sat
<b>Nb SatAnte</b>	4	23	73	21	9	17	902
<b>Proportion</b>	0,38%	2,19%	6,96%	2,00%	0,86%	1,62%	85,99%
<b>Nb SatPost</b>	7	10	13	24	37	14	948
<b>Proportion</b>	0,67%	0,95%	1,24%	2,29%	3,53%	1,33%	90,37%

Parmi les satellites, les antérieurs sont principalement des interjections, alors que les postérieurs sont des éléments qui précisent la commande. Il s'agit majoritairement de productions adverbiales du type « *de moitié* », « *un peu* », « *dans le salon* », « *près du lit* » etc. Par ailleurs, l'étiquette « nominatif » est utilisée lorsque les sujets utilisent une dénomination pour désigner le robot. Il s'agit soit de celui que les expérimentateurs ont introduit, « Emox », soit d'une variante p. ex. « *Emos* », « *Emma* », « *mon Jules* » etc. Ce type de production rentre cependant davantage dans le cadre des productions périphériques (cf. Tableau 6). Néanmoins, la différence de regroupement de ce « nominatif » ou plus généralement des micro-expressions ne semble pas anodine, car elle pourrait changer la fonction communicative qui leur est associée. En effet, à titre d'exemple sur le « nominatif », les étiquettes d'auto-annotation montrent qu'une locution « Emox » des sujets a plutôt été un marqueur de *care* (p. ex. « *je lui dis gentiment* »). À l'opposé, un « Emox » périphérique semble être associé à une interpellation, une demande d'attention (p. ex. « *c'est un ordre, pour qu'il exécute* » ou « *je lui redis parce que j'ai l'impression qu'il a pas entendu* »). L'analyse détaillée de la nature des micro-expressions n'a pas encore pu être réalisée en détails, mais la caractérisation de ces objets en fonction de leur nature, leur position, leur prosodie intrinsèque ainsi que la durée des silences qui les entourent pourrait être envisagée dans de futurs travaux en perspective de cette thèse.

### 5.2.2.3. Observations sur les micro-expressions « périphériques »

Le Tableau 6 est un récapitulatif plus détaillé de la nature des périphériques annotés. Parmi ces productions, les micro-expressions vocales sont majoritaires. À noter qu'en fonction du niveau d'expertise des annotateurs, nous avons une imprécision sur l'étiquette BB qui regroupe à la fois de vrais « bruits de bouche » non phonétiques et non phonologiques et des éléments pré-lexicaux (notés PreLex), d'où le nombre réduit de bruits de type PreLex à l'état actuel du corpus.

Certains phénomènes ont également fait l'objet d'un étiquetage supplémentaire. Il s'agit :

- de vocalisations mélodiques « chantées » ou « sifflées »,
- d'écholalies,
- de répétitions de stimuli Emox,
- d'imitations d'Emox,
- d'interpellation par des « nominatifs » du robot,
- des formes de politesse,
- des compliments,
- des énoncés qui précisent la commande,
- des justifications de commande,
- des confirmations de la commande,
- des formes de *care*/attention vis-à-vis du robot,
- des reprises d'erreurs.

Ces événements ont également fait l'objet d'un étiquetage, afin d'en voir tout d'abord la proportion, puis la distribution. Les proportions de ces événements sont résumées Tableau 6.

Tableau 6 : Tableau de répartition des objets vocaux périphériques

	Forme	Occurrences	Proportion dans groupe	Proportion dans corpus total
<b>ME</b>	BB	2621	83,98%	45,51%
	Son chanté/sifflé	12	0,38%	0,21%
	Echolalie	3	0,10%	0,05%
	Répétition	18	0,58%	0,31%
	Imitation	10	0,32%	0,17%
	PreLex	67	2,15%	1,16%
	Interjection	313	10,03%	5,43%
	Nominatif	77	2,47%	1,34%
	<b>Sous-total ME</b>	<b>3121</b>	<b>100,00%</b>	<b>54,19%</b>
<b>FC interactionnels</b>	Politesse	136	3,59%	2,36%
	Compliment	138	3,65%	2,40%
	Précision	109	2,88%	1,89%
	Justification	21	0,55%	0,36%
	Confirmation	61	1,61%	1,06%
	Care / attention pr Emox	29	0,77%	0,50%
	Interactions diverses	126	3,33%	2,19%
	Reprise d'erreurs	44	1,16%	0,76%
	<b>Sous-total FC</b>	<b>3785</b>	<b>100,00%</b>	<b>65,72%</b>
<b>Commentaires</b>	Sujet vis-à-vis d'Emox	40	2,03%	0,69%
	Description Emox	78	3,95%	1,35%
	Commentaires divers	1856	94,02%	32,23%
	<b>Sous-total Commentaires</b>	<b>1974</b>	<b>100,00%</b>	<b>34,28%</b>
<b>Total</b>		<b>5759</b>	-	<b>100,00%</b>

Le dénombrement des trois phénomènes relève d'un problème inter-annotateur, qui est en cours de correction. En effet, le premier est que les micro-expressions du Tableau 6 portent principalement sur l'écoute prosodique de l'élément ainsi que sur la production du robot qui précède. Or, une règle d'annotation, qui consistait à noter les phénomènes de *chameleon effects*, a été très peu respectée ou oubliée. Ainsi, nous avons actuellement un nombre important d'écholalies, répétitions et imitations réparties dans les étiquettes de type BB, PreLex, Interjection, mais également certaines formes de FC, car les productions du robot imitées ou répétées sont de cette nature. Nous n'avons donc pas ici un dénombrement exact. En revanche, le nombre de ces événements est plutôt de l'ordre d'une centaine sur l'ensemble des sujets. Ces phénomènes sont également abordés dans la section 5.2.4.3.

### 5.2.3. Marqueurs remarquables de l'interaction socio-affective « gluante »

#### 5.2.3.1. Compliments et politesse

Concernant les formes plus complexes (notées FC), le Tableau 6 a mis en évidence des objets « interactionnels ». Ces événements sont supposés être des marqueurs d'évolution de la glu. En effet, nous pouvons observer un nombre non-négligeable de formes de politesse (p. ex. « merci », « s'il te plaît ») ou de compliments (p. ex. « c'est très bien », « tu es un gentil robot ») dont les timings d'apparition semblent, là encore, être à l'origine de fonctions communicatives différentes. En effet, un « merci » peut être étiqueté comme « *je demande pourquoi je l'ai dit, c'est qu'un objet mais c'est sorti comme ça* », « *je suis poli, je lui réponds c'est dans ma nature, je le dis automatiquement* », « *je suis vraiment gentil avec le petit robot et très poli avec lui* », « *je me sentais obligée de lui dire merci, parce que c'était tellement bien* », « *c'est pour le récompenser – je le félicite* ». Les étiquettes reflètent différents effets du robot résultant sur le même objet lexical « merci ». Elles montrent l'influence des comportements du robot sur la personne. Le robot semble être graduellement considéré comme un objet qui permet de redonner une certaine envie et entraîne la personne vers une dynamique de communication. Là encore, cet effet est fortement en lien avec le timing de feedback du système (robotique ou domotique), puisque les étiquettes associées à « s'il te plaît » peuvent être similaires à celles du « merci », mais elles peuvent aussi être liées à des auto-annotations telles que « *je lui répète deux fois, j'ai pris l'habitude – ça va pas assez vite* » qui montrent explicitement le moment où il y a eu des défauts dans le timing de réponse du système et qui illustrent la modification du comportement du sujet, d'où l'importance de mesurer cet effet.

#### 5.2.3.2. Justification et marqueurs du *care*

Par ailleurs, plus l'interaction avance et plus des formes variées apparaissent. Parmi elles, nous avons les justifications de commande (p.ex. : « *il y a trop de lumière* », « *c'est pas la peine* », « *ça sert à rien de la garder allumé, il faut économiser* ») qui apparaissent soit de manière concomitante aux formes de commande avec « je » qui exprime une opinion personnelle qu'on retrouve avec des objets annotés « commentaires » ou « sujet vis-à-vis d'Emox », soit avec des formes en « on » avec une dimension d'accompagnement, presque pédagogique comme relevé dans des auto-annotations : « *je lui explique comme à un gosse* ». Ces événements sont également notés avec des éléments notés « confirmation » (p. ex. « *c'est ça, tu as bien compris* ») ou « compliments » (p. ex. « *c'est bien tu es un gentil robot* » ; « *bravo* »), dont les occurrences sont relativement variables en fonction des sujets. Puis, sur les phases les plus avancées de l'expérience, avant le retour de l'aide à domicile, nous voyons apparaître des énoncés du sujet empreints de *care*, dont les manifestations se retrouvent notamment dans des énoncés demandant des changements de position/emplacement du robot (p. ex. « *ne reste pas derrière la porte, avance, à gauche, encore, encore plus à gauche, voilà pour pas prendre si elle s'ouvre* » ; « *viens à côté de moi, hein, on reste ensemble au salon* »). Celles-ci s'accompagnent également d'énoncés interactionnels divers (p. ex. « *coucou, tu me vois?* », « *est-ce que tu m'entends?* », « *tu t'amuses hein* », « *eh oui tu rigoles* ») associés à une volonté de continuité communicative qui est définie dans les auto-annotations de type « *en fait j'étais un peu tenté de lui parler à lui inconsciemment mais en même temps je me disais c'est une chose* », « *je lui réponds – je communique avec lui, y a vraiment une communication* », « *j'essaie d'établir un dialogue mais aussi de jouer* », « *je m'amuse, je rigole, pour lui faire voir que c'est pas qu'un objet, on parle* », « *c'est marrant quand il rigole – c'est rigolo – on ne s'ennuie pas parce qu'il y a un échange* ».

### **5.2.3.3. Phénomène de récupération d'erreurs**

Enfin, un dernier point d'observation est celui du phénomène de récupération d'erreurs. En effet, l'écologie de l'expérimentation a initialement généré des erreurs d'exécution des commandes domotiques engendrant différentes réactions. Ainsi, comme la génération d'erreurs semblait être un point d'observation remarquable, elle a été intégrée dans le protocole pour volontairement produire des erreurs à différents temps de l'interaction. Le phénomène alors observé est celui de la récupération d'erreurs. Au niveau expérimental, l'échec d'un actionneur ou les faux positifs amènent à faire adopter deux comportements principaux :

- le premier cas s'illustre par l'abandon de la commande. Ces événements sont couplés à de nombreux marqueurs de micro-expressions de nature expiratoire telles que les soupirs. Ces dernières sont associées à des auto-annotations comme « *la première fois j'étais pas arrivée mais en fait je crois avoir compris que c'est parce qu'il faut du temps – là je m'aperçois qu'il fallait lui laisser du temps au système – je m'attends toujours à ce que ça aille plus vite* », « *je trouve tout ça complètement absurde, pour moi c'est impossible de vivre comme ça aujourd'hui – c'est tellement bête* ».

- le second cas de comportement s'observe lorsque la personne reprend le robot en le corrigeant jusqu'à ce qu'il y arrive. Les sujets décrivent alors des auto-annotations du type : « *Il y avait toujours un certain ralentissement comme l'ordinateur, il y a des pauses – c'est l'habitude quand on se sert de matériel – on s'habitue – quand on parle ça va plus vite* », « *j'essayais de voir si on peut pas y arriver, comme en première intention ça marchait pas* », « *je lui commente pour essayer de l'aider de comprendre de ce qu'on fait* », « *[je] dis comme à des enfants, j'en ai qui viennent prendre des cours, c'est un peu ma manière de m'adresser à eux* », « *j'ai dit c'est la question qu'il fallait pas lui poser* ». Les sujets vont parfois même jusqu'à se reprendre en pensant « *je l'ai pas compris, c'était de ma faute* ». Là encore, les feedbacks du robot peuvent engendrer une correction et une adaptation du comportement des personnes.

De manière générale, l'effet de récupération va dans le sens d'un renforcement de la glu. Elle s'accompagne du phénomène de paraphrasage de la commande vocale ou d'une modification de la dynamique prosodique (la plupart avec une finale montante), apportant une dimension d'accompagnement qui renforce la dynamique interactionnelle. Par rapport à ces deux observations, il semble exister un point à partir duquel la récupération d'erreurs devient systématique. Ainsi, si ce seuil d'état de glu est atteint, la personne semble rentrer dans ce mécanisme de récupération, alors qu'en deçà, elle a tendance à rejeter la technologie et le processus de « gluage » semble se ralentir. Ces seuils et ces rythmicités seront plus amplement explicités dans le Chapitre 5.

### 5.2.4. Observations acoustiques et prosodiques générales

Cette partie d'analyse est une première approche très globale, car les mesures acoustiques liées à la qualité de voix restent encore assez imprécises. Elle est d'ailleurs rendue davantage complexe par les caractéristiques vocales de la personne âgée pour laquelle, par exemple, la voix est naturellement plus soufflée ou *creaky*, diminuant également la robustesse des systèmes de reconnaissance de la parole (cf. section 6.2.1). Les observations principales de cette section s'appuient principalement sur une écoute experte des stimuli pour la qualité de voix. En revanche, nous avons également recensé 999 énoncés produits par les 16 sujets analysés (filtrés sur les commandes vocales) pour effectuer des mesures automatisées, notamment pour le pitch.

#### 5.2.4.1. Augmentation générale du pitch

Un premier constat général est une augmentation de la hauteur de la voix des sujets quand ceux-ci s'adressent au robot. Ici, nous insistons sur le terme « s'adresser », car durant les premières phases de l'expérience, les auto-annotations sont de nature « *je teste l'appartement* » ou « *je lis les commandes* ». La hauteur de la voix des sujets semble plus basse que celle des commandes paraphrasées. Un second constat est celui d'une augmentation croissante de la *breathiness* au cours de l'expérience.



Figure 47 : Tendence d'évolution des pitches moyens des commandes pour des sujets prototypiques

En revanche, qu'il s'agisse de la hauteur de voix ou de la *breathiness*, l'augmentation n'est pas linéairement croissante puisqu'elle semble suivre les « fluctuations de la glu » (cf. section 5.3), dont les valeurs locales peuvent faire l'objet de retours en arrière sur des états rencontrés précédemment. En revanche, si nous regardons la tendance globale sur l'ensemble de l'expérience, nous voyons que pour les sujets « glués », la hauteur de voix est ascendante, alors qu'elle est descendante pour les sujets « non glués ». C'est notamment ce que la courbe de tendance noire montre sur la Figure 47.

La Figure 47 représente l'évolution prototypique des *pitchs* moyens des commandes sur l'ensemble des productions effectuées par les sujets. La mesure ayant été réalisée de manière automatique sur 999 énoncés de commandes (cf. Annexe 10), elles pourront parfois manquer de précision. Sur les trois premiers graphiques les données illustrent l'évolution des sujets « glués » homme (p. ex. S11) et femmes (p. ex. Sa1 et S07). Pour ces cas, les personnes ont soit glué avec de nombreuses formes de paraphrase (p. ex. Sa1), soit avec très peu de variations morphosyntaxiques sur la commande (p. ex. S07). Au contraire, les deux dernières figures (homme – S19 et femme – S17) sont des cas de sujets « non glués » (cf. section 5.3.2.3 qui montrent le cas de S19 avec déficience auditive et S17 un cas difficile en phase dépressive). Ces derniers montrent une tendance descendante pour la hauteur de leur voix. Ici, les mesures ont été réalisées en Hertz (valeurs absolues) et semi-tons re100Hz (valeurs relatives). Les figures ont ici conservé l'échelle en Hertz puisque l'objectif n'était pas de faire une comparaison intersujets, mais plutôt de montrer la tendance intrinsèque à chaque sujet, dont les variations étaient plus visibles sur cette échelle.

Notons également que sur certaines commandes, nous avons des pics plus importants de ces *pitchs*, un phénomène qui serait à mettre en parallèle avec le niveau de projection de la voix. En effet, le sujet se déplaçant dans l'appartement et l'environnement étant naturellement bruité à l'exécution des commandes domotiques, nous avons remarqué des pics d'intensité en lien avec ces pics de fréquence fondamentale (particulièrement visible pour S19 dont la déficience auditive caractérise ces variations). En effet, d'après les normes usuelles<sup>92</sup>, ces pics seraient liés à la production par voix projetées ou par voix d'appel. En effet, les occurrences relevées semblent observables à des moments plus bruités, mais également lorsque nous observons une variation de proxémie entre le sujet et le robot suite/au cours d'un déplacement, voire lorsque nous observons une erreur d'exécution de la commande domotique. Les ruptures de variations globales du *pitch* des commandes pourraient ainsi être une piste de réflexion qui ferait converger des dimensions vocales à celles de la proxémie.

Concernant les mesures d'intensité, nous ne nous attarderons pas dessus, car la variation d'intensité est délicate du fait du mouvement du micro qui a servi au recueil. Le choix d'un micro-casque très léger a été réalisé pour garder une distance à la bouche fixe, et permettre de prendre ces mesures. Sur des phases stables, les mesures restent très largement exploitables, mais les sujets, portant souvent des lunettes, ils ont fréquemment touché et modifié la configuration du micro. Un tri des données serait donc nécessaire avant de pouvoir exploiter correctement cette dimension.

---

<sup>92</sup> Nous notons généralement dans les manuels d'orthophonie et de phonétique acoustique que la voix conversationnelle est comprise entre 110-165Hz pour un homme, 120-330Hz pour une femme ; la voix projetée de 120Hz-185Hz pour l'homme, 245Hz-370Hz pour la femme ; et enfin la voix d'appelle de 220Hz-330Hz pour l'homme, 440Hz-660Hz pour la femme.

#### 5.2.4.2. La *breathiness* : dynamique subtile de la dimension « altruiste » ?

Actuellement, la qualité de voix est un champ d'investigation important. Certains auteurs s'intéressent par exemple à sa classification automatique, en leur associant des dimensions socio-affectives (Gobl & Chasaide, 2003 ; Scherer, Kane, Gobl, & Schwenker, 2013). Elle fait également l'objet d'extraction automatique, notamment des micro-expressions vocales (Ishi, Ishiguro, & Hagita, 2008 ; Ishi, Ishiguro, & Hagita, 2010), cependant ces mesures restent encore exploratoires. Parmi les qualités de voix, la *breathiness* a été étudiée (Frohlich, Michaelis, & Strube, 1998) comme un paramètre acoustique déterminant le profil d'une voix pathologique, où (Samlan, Story, & Bunton, 2013) ont eu une approche plus physiologique dans le cadre d'une modélisation informatique. Par ailleurs, hors voix pathologique, la voix *breathy* a été montrée comme celle qui accompagne l'intimité, et plus spécifiquement le *care* ou la politesse, fortement marquée chez les locuteurs masculins en japonais notamment (Ito, 2004) et est associée à des dimensions d'intimité et de familiarité par (Laver, 1980).

Dans le cadre de cette thèse, nous n'avons malheureusement pas pu mettre en place de mesure par EGG (électroglottographie) du fait de la complexité du scénario, et de plus la pose de ce type d'appareil est complexe lorsque le sujet bouge (cf. Audibert 2008). En effet, une justification aurait été nécessaire par un nouveau prétexte pour inclure ce matériel dont il aurait fallu détourner l'attention. Par ailleurs, les analyses possibles à partir du signal acoustique auraient, par exemple, été celle du NAQ (*Normalized Amplitude Quotient*) (Alku, Bäckström, & Vilkman, 2002) qui semble être l'une des plus vérifiées à l'heure actuelle, bien qu'elle présente des limites très fortes (Gobl & Chasaide, 2003 ; Campbell & Mokhtari, 2003), en particulier si on veut l'utiliser en suivi de signal, puisque la variation de la voyelle fait varier le NAQ avec la même amplitude que la variation de laxité des cordes vocales (Audibert, Rossato, Aubergé, 2008). Les autres mesures indirectes de la *breathiness* s'appuient sur la différence d'amplitude des deux premiers formants H1-H2 (Ladefoged & Maddieson, 1998), mais celles-ci sont largement critiquées pour leur imprécision (Henrich, d'Alessandro, & Doval, 2001 ; Hillenbrand, Cleveland, & Erickson, 1994) et notamment pour leur faiblesse de robustesse en présence de nasalité (Simpson, 2012). La tendance actuelle s'appuie ainsi sur des mesures autour des HNR (*harmonics-to-noise ratio*), NHR (*noise-to-harmonics ratio*), and SNR (*signal-to-noise ratio*) à l'instar des travaux de voix pathologiques marquées par la *breathiness* (p. ex. Martin, Fitch, & Wolfe 1995 ; de Krom, 1995 ; Heman-Ackah et al., 2003 ; Schindler et al., 2008 ; Cantarella et al., 2010) ou encore sur les tilts spectraux sur bandes fréquentielles B0-B2 (Hillenbrand et al., 1994 ; Kreiman, Gerratt, & Antonanzas-Barroso, 2007).

Dans notre observation, nous nous sommes appuyés sur une perception acoustique experte de cette dernière<sup>93</sup>. L'annotation a montré qu'il existe une augmentation générale de la *breathiness* suivant le même type de variation que le *pitch*, globalement renforcée pour les sujets qui ont glué. Au contraire, pour les sujets non glués, la voix devenait moins laxa et plus tendue, notamment sur les deuxièmes occurrences d'une même commande répétée.

L'évolution de cette *breathiness* a vraisemblablement un lien avec une dimension de *care* qui s'observe notamment dans les parlars maternels (*motherese*) ou ceux des aidants (*maidese*). (Nwokah, 1987) qui s'est intéressée à la comparaison de ces deux styles de langage dans l'interaction adulte-enfant qui montrait notamment l'usage commun de cette *breathiness*. En parallèle de cette qualité de voix, elle met également en évidence l'apparition d'auxiliaires modaux comme *shall* ou *will*, ce qui rejoint l'apparition de « pouvoir » et

---

<sup>93</sup> Les premières annotations ont été marquées dans le cadre du Master professionnel IdL de (C. Cottier, 2015) suivi par (L. Tsvetanova, 2015) dans le cadre de son mémoire de recherche Master IdL, sous l'encadrement de V. Aubergé et Yuko Sasa.

« vouloir » associés aux paraphrases avec le pronom « tu » de notre corpus, sur des phases avancées de l'expérience (cf. section 5.2.1.1). En revanche, l'une des différences notées par l'auteure porte notamment sur les durées des énoncés entre le *motherese* et le *maidese*. Dans les études de (Nwokah, 1987) ainsi que celles rapportées par (Saint-Georges et al., 2013), le *motherese* présentait une durée d'énoncé globalement plus importante que le *maidese* ou autres parlars associés à l'enfant (p. ex. celui des grand-parents à leurs petits-enfants). De même, la distribution quantitative du nombre de composants d'énoncé semble plus importante dans le *motherese*. De plus, il semble également que la durée des énoncés est également différente, plus courte dans la cadre du parler maternel. Ces subtilités sont également des pistes d'exploration, non pas sur la durée absolue, mais sur le débit de parole. À noter que pour les énoncés de notre corpus, une automatisisation par détection de noyaux vocaliques pour compter le nombre de voyelles et une segmentation automatisée avec des outils tels que le plug-in EasyAlign (Goldman, 2011) ont été testées. Cependant, elles ont été abandonnées dans le cadre de cette thèse du fait des imprécisions engendrées. En effet, la première difficulté relève tout d'abord du phénomène de paraphrasage des commandes, du fait de la nature spontanée de la parole enregistrée. L'étiquetage permet de filtrer commande par commande pour effectuer d'éventuelles comparaisons, mais les variations de formes morphosyntaxiques changent également le nombre de syllabes, en particulier avec les occurrences des micro-expressions « satellites » (cf. section 4.2.5.2.1). Lorsque ces derniers sont internes à la commande, ils augmentent le nombre d'erreurs du traitement automatique, nécessitant alors une correction manuelle. Une exploration plus approfondie serait donc nécessaire sur ces aspects.

Un autre point, qui converge également avec les évolutions d'énoncés observés, est la nature de la forme interrogative. Le *maidese* serait basé à des *WH-questions* soit des réponses complexes attendues, alors que le *motherese* est plutôt sujet à des interrogations de réponses oui/non (Olsen-Fulero, 1983). Dans le corpus EEE, les personnes qui gluent rapidement semblent présenter plus rapidement des formes similaires au *maidese*, alors que, lorsque l'évolution est plus progressive, les sujets interrogeaient le robot sur un format plus proche à celui du *motherese*, ou tout au moins un degré de *care* renforcé. De plus, (Cross, Johnson-Morris, & Nienhuys, 1980), qui ont fait des observations comparatives avec des enfants sourds, montrent que cette tendance de *motherese* qui interroge par des réponses oui/non était davantage remarquable pour les enfants sourds. Les auteurs rapportaient qu'il s'agissait d'un moyen d'interpréter et de récolter plus d'indices de comportements par la mère, qui avait ainsi tendance à associer le oui ou le non de l'enfant à des comportements non lexicalisés ainsi qu'à des micro-expressions vocales. Cette subtilité associée à la *breathiness* est en réalité proche d'un qualificatif de *softness/mildness* ou de dimensions de *care* sensiblement différentes. Il s'agit peut-être d'une dimension sur laquelle ledit « altruisme » de la glu socio-affective se développe.

Enfin, cette dimension, qui semble être une forme de dynamique prosodique associée vocalement à la *breathiness* est extensible à des dynamiques semblables qui se retranscrivent dans d'autres modalités. Les travaux de (Herrera, Reissland, & Shepherd, 2004) font par exemple le pont entre le *motherese* vocal (associé à la notion de *closeness* ou proximité entre mère et enfant) et le comportement tactile pour une étude comparative avec des mères en dépression postnatale. En effet, les auteurs évoquent une différence du *motherese*<sup>94</sup> des mères dépressives, qui se remarquerait surtout sur la complexification des formes interrogatives, des énoncés moins cadencés et un manque de considération d'*agency* ou d'agentivité (c.-à-d.

---

<sup>94</sup> Ce parler est le plus souvent caractérisé comme des courts énoncés, répétés, cadencés, avec de nombreuses formes impératives et interrogatives, mais peu directives, avec un contrôle sur ses états affectifs et des corrections en s'appuyant sur l'expérience de l'enfant (Snow & Ferguson, 1977 ; Murray & Trevarthen, 1986).

capacité de pouvoir agir sur le monde) de l'enfant par la mère (Cox, Puckering, Pound, & Mills, 1987 ; Murray, Kempton, Woolgar, & Hooper, 1993). Le lien avec le toucher montre que la différence de dynamique se retrouve également au niveau de la fréquence des touchers des mères et de leur impact sur les comportements autocentrés de l'enfant. En effet, la diminution des touchers chez les mères dépressives semble induire des formes de compensation qui augmentent le nombre de touchers autocentrés de l'enfant. Ce type d'attitude est également semblable à la tendance de sollicitations accrues des magiciens. Ces derniers cherchaient parfois à attirer l'attention du sujet trop souvent, durant certaines phases moins actives des expériences, notamment pour les sujets qui ont eu une attitude type *maidese*. Cette transposition de dynamique sur une modalité différente, dont les observations sont moindres pour le toucher dans le cadre de nos expériences robotiques, semble également visible dans le regard, par la variation de la fréquence et la direction du regard (Tsvetanova, Aubergé, & Sasa, 2016). Les sujets centrent ainsi leur regard sur soi (à travers la feuille de commandes), puis sur le robot. Les sujets à comportements *maidese*, présentent une progression de la cible du regard qui semble davantage se porter sur l'actionneur domotique que sur le robot. D'ailleurs, ce second profil semble basculer progressivement vers un profil *motherese* au fil de l'expérience. Nous observons également ce type de changement de dynamique subtile sur la gestualité avec ce même robot (Girard-Rivier et al., 2016). L'idée du toucher amène d'ailleurs des réflexions avec développement de la notion de *Social Touch* en robotique (notion développée par Aubergé et Maisonnasse depuis 2015) augmentant le robot, dont les trois points de contact saillants existants (vocal, tactile, regard)<sup>95</sup> tentent d'augmenter l'immersion interactionnelle dans le cadre de la téléprésence, afin de donner la possibilité de faire manier cette subtilité de sociabilisation par les utilisateurs. Ces points de contact sont également évoqués par (Tisseron, 2015) dans son exploration de la robotique.

### 5.2.4.3. Bref aperçu des profils prosodiques remarquables

Les analyses prosodiques entamées ont été localement réalisées sur les commandes. Elles sont soit manuellement réalisées, soit semi-automatiquement, mais cela demande encore des corrections. Le choix de travail a été orienté sur le logiciel Praat (Boersma & Weenink, 1992) du fait des aspects d'automatisation facilitée par des scripts au vu de la quantité de données, mais sachant que les manipulations concernent avant tout la F0, des outils tels que WinPitch (Martin, 1996) auraient pu être plus précis sur ce type de mesure et donc restent à explorer.

La mesure phonétique s'est effectuée à partir des cibles initiales et finales sur le point le plus stable de la voyelle. Ainsi, une première approche consistait à voir si le mouvement était globalement ascendant ou descendant en mesurant le début de la voyelle ascendante ou la fin de la voyelle descendante. Cette mesure est effectuée sur soit les 3 premiers cycles, soit avant les 3 derniers cycles de la voyelle, afin d'éviter les effets micromélodiques influencés par la consonne précédente ou suivante. Cette mesure s'accompagne également du maximum et du minimum de la F0 sur l'ensemble de la commande, à partir des points les plus hauts ou les plus bas dans la voyelle, en excluant les 3 premiers ou derniers cycles vocaliques. Un survol global des différentes occurrences a mis en évidence 9 profils de contours prosodiques récurrents auprès de différents sujets<sup>96</sup> comme présenté Figure 35. En réalité, les profils prosodiques sont bien plus nombreux, dont certains plus idiosyncrasiques ou plus rares. La

---

<sup>95</sup> Ces travaux sur le *Social Touch* sont développés dans le cadre du développement de robot de téléprésence fabriqué en collaboration avec le Fablab MSTIC du LIG ainsi que la plateforme Domus-LIG, avec en particulier la thèse de A. Davat (débutée en 2016) qui propose une exploration sur la dimension de toucher vocal autour de la distance spatiale et la distance sociale.

<sup>96</sup> Le travail manuel, qui a fait émerger ces premiers patterns prosodiques récurrents, a été réalisé en collaboration avec L. Antunes, chercheur en linguistique et prosodie en visite au Laboratoire d'Informatique de Grenoble courant 2013-2014.

## Analyses des productions vocales : vers une prosodie du dialogue ?

rareté n'est pas mise de côté également, car elle peut être très significative dans l'observation. Le repérage effectué par les FoT\_IPC permettrait notamment de cibler sur ceux qui sont remarquables dans l'évolution interactionnelle.

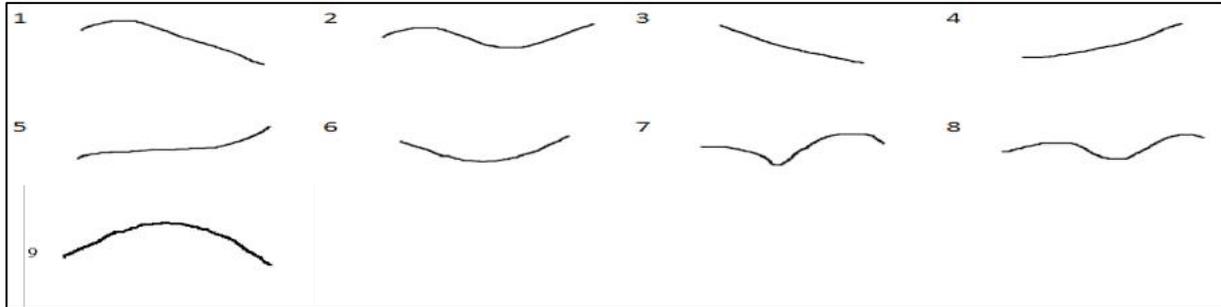


Figure 48 : Patterns de contours prosodiques récurrents

De manière générale, les contours à finales ascendantes sont ceux qui accompagnent l'apparition des formes interrogatives (soit explicite avec inversion auxiliaire-sujet, soit intonative comme le contour 4 qui est co-occurent avec les formes en « on »). Les contours à finales descendantes sont quant à eux plutôt émergents aux phases initiales de l'expérience, où le sujet réalise des « lectures » de commande en voix modale ou sur les énoncés impératifs voire ceux avec le pronom « tu ». En revanche, l'auxiliaire « pouvoir », associé au « tu », semble être produit sur des intonations ascendantes, alors que le « vouloir » est plus souvent corrélé à une intonation descendante. La Figure 49 donne un exemple de quelques-uns de ces profils de contours prosodiques sur spectrogrammes.

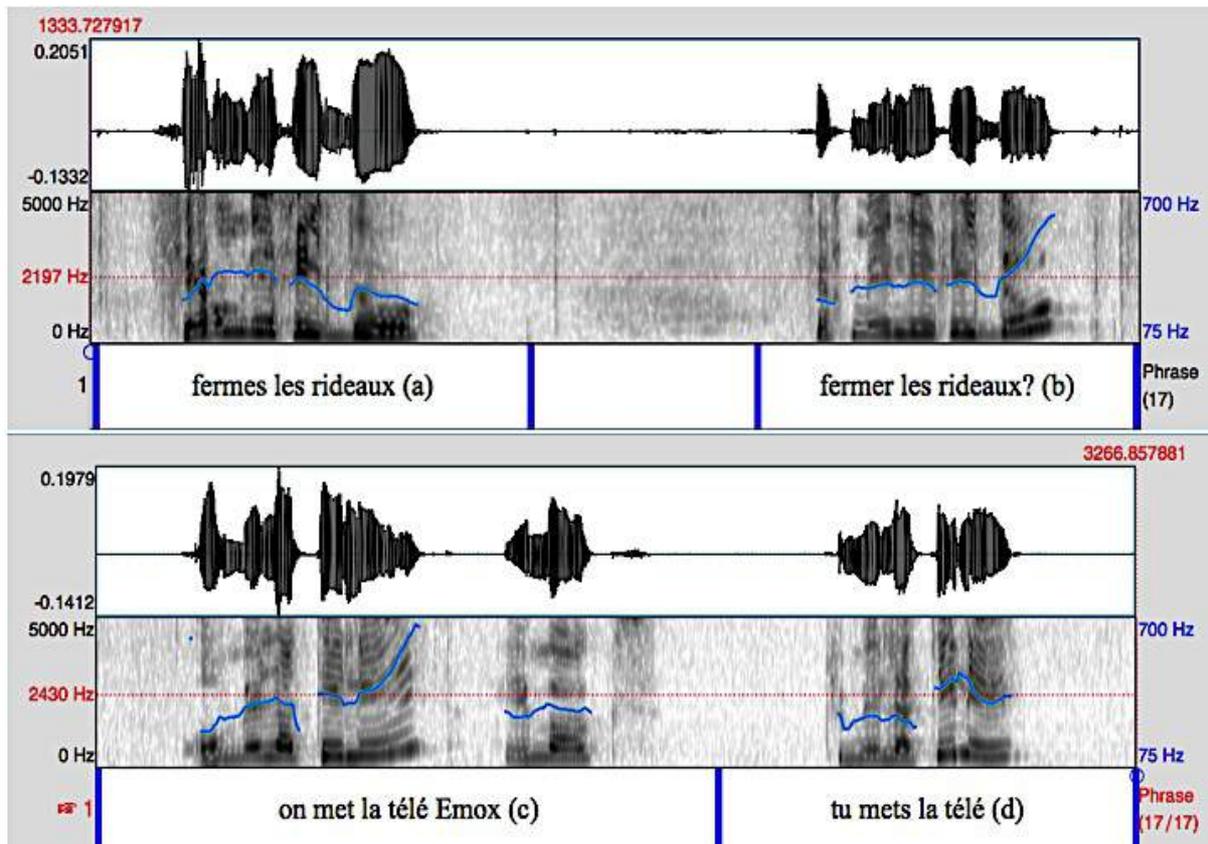


Figure 49 : Spectrogrammes d'énoncés produits à 4 temps distincts de l'expérience (sujet Sa1)  
Les lettres indiquent l'ordre chronologique d'apparition après les premières paraphrases du sujet  
(a= première forme – d= dernière forme)

Par rapport aux événements acoustiques remarquables, nous observons des productions de type « écholalie » (copie prosodique d'un son d'Emox sur un autre support lexical) et des imitations (production lexicale et prosodique identique à celles du robot) ainsi que des répétitions (production lexicale identique, mais prosodie différente). Ces éléments qui sont rares sur l'ensemble du corpus semblent pourtant être des marqueurs dans la caractérisation de la glu, car il s'agit de manifestations vocales du *chameleon effect* qui augmenterait l'empathie entre les sujets en désamorçant les postures de dominance (cf. Chapitre 1). Ces objets apparaissent-ils alors comme prédicteur d'un état de glu ou comme résultante de formes interactionnelles spécifiques ? Ils pourraient ainsi être envisagés dans une modélisation comme marqueurs de rupture déterminant des niveaux de glu spécifiques. Pour le moment, nous remarquons uniquement que leur timing d'apparition n'est pas parallèle à celui des premiers bruits de bouche du robot, mais qu'ils apparaissent un peu plus tard dans l'interaction, d'où cette hypothèse de potentiels marqueurs de points de rupture de l'interaction, mais à un niveau avancé de projection empathique.

### **5.3. Evolutions temporelles des marqueurs interactionnels**

Les sections précédentes ont surtout montré des distributions quantitatives globales, bien qu'avec quelques observations des dynamiques vocales, nous sommes vite rattrapés par la nécessité de voir l'évolution temporelle des données, le suivi de la dynamique relationnelle. La présente section va donc axer les observations sur ces dimensions temporelles, en faisant un bref focus sur la rythmicité des événements interactionnels. Puis, nous aborderons la progression chronologique des auto-annotations. En effet, l'enjeu est ici de savoir comment naviguer à travers la progression de l'interaction, en se demandant où apparaissent les changements et quels événements sont susceptibles d'expliquer ces changements dans la communication.

#### ***5.3.1. Chronogrammes des événements interactionnels***

Un enjeu des analyses de comportements, comme défini par exemple dans les pratiques de l'éthologie, est l'organisation temporelle des données et la rythmicité qui en ressort. Ceci reviendrait à déterminer une « prosodie interactionnelle » globale sur l'ensemble des comportements réalisés sur une certaine période. Dans cet ordre d'idées, les chronogrammes vont pouvoir mettre en évidence des cyclicités de certains événements (Graff, 2008 in Vanpé, 2011). Pour cela, la durée intrinsèque de l'événement est réduite, pour uniquement considérer la survenue de l'événement comme une occurrence (sur base binaire). Ainsi, tout événement équivaudrait à un autre sans distinction. Dans la thèse de (Vanpé, 2011), l'étude des micro-expressions vocales recueillies sur méthodologies similaires mentionnait en perspectives l'usage de ces chronogrammes comme possibles outils d'analyse de la rythmicité interactionnelle.

Ainsi, à partir des différents événements alignés sur le corpus EEE, nous avons recensé et filtré parmi les événements, ceux qui interviennent directement dans l'interaction face à face, à savoir les commandes produites par les sujets (incluant les nouvelles commandes), les autres types de productions vocales (qui peuvent être sous-filtrés sur les micro-expressions par exemple), les stimuli d'Emox en situation face à face avec le sujet (mouvements et sons, filtrable sur l'un ou l'autre) et les actions domotiques ayant effectivement eu lieu dans l'appartement. Ces événements, dont les quantités sont rappelées pour chaque sujet Tableau 7 sont ainsi tous visualisables sur un chronogramme. Par rapport au temps total de l'expérience, chacun de ces événements est très court. Leur durée devient donc négligeable à l'échelle globale. Nous nous sommes ainsi intéressés aux IOI (*Inter-Onsets-Intervals*), prenant pour mesure le début des événements où, quelle que soit la valeur, ces événements sont supposés comme équivalents. Sur ce type de figure, les médianes représentent la cadence des

## Analyses des productions vocales : vers une prosodie du dialogue ?

événements (à savoir si la cyclicité des comportements est lente ou rapide) et les écarts-types permettent de déterminer si le comportement est régulier ou non.

Tableau 7 : Récapitulatif des événements interactionnels sur le corpus EEE sur 16 sujets

	Cmd	NonCmd	Emox	Domus	Total
<b>aSa1</b>	96	210	288	46	640
<b>S01</b>	98	203	129	29	459
<b>S03</b>	146	248	211	122	727
<b>S04</b>	97	394	248	60	799
<b>S05</b>	59	481	250	25	815
<b>S06</b>	135	342	398	55	930
<b>S07</b>	66	557	277	41	941
<b>S10</b>	87	699	279	45	1110
<b>S11</b>	51	522	236	30	839
<b>S12</b>	53	253	179	37	522
<b>S14</b>	47	232	212	43	534
<b>S15</b>	94	580	237	70	981
<b>S16</b>	87	323	318	75	803
<b>S17</b>	18	206	161	15	400
<b>S18</b>	47	201	234	39	521
<b>S19</b>	33	319	130	26	508
<b>TOTAL</b>	<i>1214</i>	<i>5770</i>	<i>3787</i>	<i>758</i>	<b>11529</b>

La Figure 50 illustre l'exemple du chronogramme du sujet Sa1, où en abscisse est représenté le temps chronologique de l'expérience et en ordonnée la durée des délais entre deux événements, rapportés sur une échelle logarithmique pour une meilleure visualisation, sachant que la variation des délais extrêmes est importante. Pour les événements, nous avons ici pris sans distinctions tous les événements interactionnels, à savoir les 640 stimuli de Sa1 représentés sur le total de la première ligne du Tableau 7.

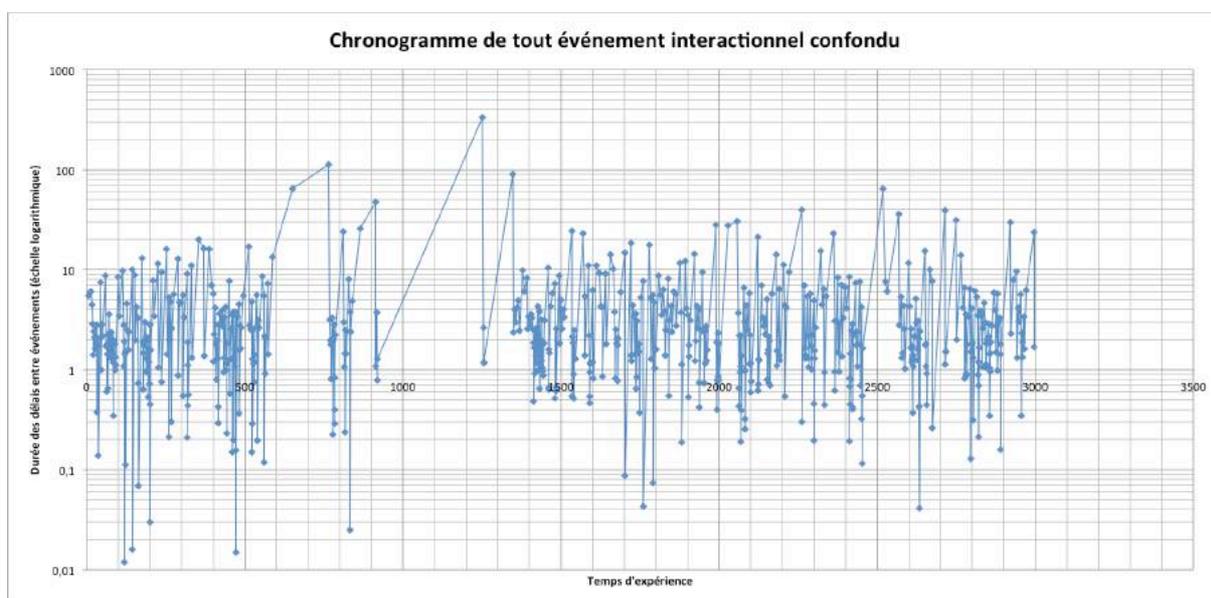


Figure 50 : Exemple de chronogramme des événements interactionnels (sujet Sa1)

Pour ce sujet Sa1, deux écarts importants sont remarquables. Ils correspondent à deux temps non interactionnels (l'un étant un passage aux toilettes ; l'autre la phase d'installation du sujet qui place ses objets dans l'appartement sans interagir avec le robot). Puis, la Figure illustre différentes périodes de phases régulières, avec des moments où les événements sont effectués sur des cadences rapides (les zones denses à faible amplitude du graphique), puis reviennent sur une cadence plus lente. Des alternances s'observent donc entre événements à cadence rapide et événements à cadence lente. Le problème de cette représentation rassemblant tous les événements sans distinction est la perte de la nature de l'événement. Cette distinction des événements par type (chaque colonne du Tableau 7) est également réalisée en Annexe 13 avec des versions par catégorie d'événements ou par leur superposition.

De plus, certains couples d'événements sont particulièrement intéressants pour voir différentes dynamiques de rythmicité entre :

- l'exécution des actions domotiques après la commande produite par le sujet,
- les micro-expressions vocales des sujets par rapport à leurs occurrences de commandes,
- les *feedbacks* (surtout vocaux) d'Emox par rapport aux occurrences de commandes,
- les micro-expressions vocales des sujets au regard des stimuli d'Emox ;

En revanche, regarder ces couplages seuls n'aurait pas de sens. En effet, le premier constat est qu'il existe bien des zones de rythmicités variables, qui, si nous zoomons dessus, pourraient être des temps d'interactions permettant de comprendre le passage ou le maintien d'un état de glu socio-affective donné. Les enjeux de visualisation, permettant ces zooms et dé-zooms sur les transitions, sont importants et ils sont à observer à travers le séquençage des auto-annotations.

### 5.3.2. *Evolution de la perception d'Emox*

Chaque auto-annotation porte des valeurs informatives différentes, mais également des temporalités distinctes qui nous permettraient d'orienter notre observation. Les sections qui vont suivre au sein de cette sous-partie vont mettre en évidence ces variations temporelles qui se produisent en parallèle sur des temps cognitifs variables. Les axes temporels présentés dans les graphiques préservent une échelle proportionnellement relative au temps interactionnel intrinsèque de chaque sujet, le but étant de voir les changements de dynamiques globales et non de faire une comparaison en valeurs absolues temporelles. En effet, la durée des expériences et la vitesse d'interaction des sujets étant variables, l'échelle temporelle n'est pas normalisée. Parmi les différents niveaux d'auto-annotations, nous avons pu voir que celles associées aux étiquettes *FoT\_PE* rassemblaient des *Feeling of Thinking* illustrant la perception globale que les sujets avaient vis-à-vis du robot. Ainsi, chaque sujet passerait par environ 5 à 15 états de *FoT\_PE* par expérimentation (cf. Tableau 3).

#### 5.3.2.1. **Description de « niveaux » de glu**

Une analyse transversale sur les 16 sujets analysés a permis de récapituler par niveau des étiquettes communes (N.B. : elles sont arbitrairement définies au regard des valeurs d'auto-annotations observées). Elles sont représentées sur un axe qui pourrait résumer l'état le plus global de la glu socio-affective (soit approximativement la force ou le niveau de la glu) évoluant sur une dimension animiste socio-affective altruiste. Sachant que la glu peut aussi bien évoluer positivement que négativement, nous avons fixé un ordre 0 qui reflète le même état que le niveau d'amorçage neutre de l'expérimentation (robot = une chose, un objet), avec

## Analyses des productions vocales : vers une prosodie du dialogue ?

une échelle qui évolue positivement ou négativement. La formalisation retenue est présentée dans le Tableau 8.

**Tableau 8 : Pondération arbitraire systématisée résumant les valeurs de FoT\_PE des sujets âgés**

Niveau	Description
-2	animé – qualificatif négatif
-1	objet artificiel – qualificatif négatif
0	objet – truc – chose – ça
1	objet utile / possession d’une chose
2	vivant – animal – chose qui bouge
3	vivant – chose intelligente
4	qualificatif de <i>pet</i> – animal domestique (dimension affective) / personne humaine / objet sensitif
5	ça tient compagnie – présence – quelqu’un/humain qui m’aide ( <i>care</i> reçu) – familial/partenaire
6	comme un enfant/bébé/gosse ( <i>care</i> produit /protection)
7	possessif (mon/ma) + descripteur affectif / p. ex. compagnon-bonhomme-ami etc.

### 5.3.2.2. Visualisation temporelle des niveaux de glu mesurés

À partir de ce tableau, ces niveaux ont été attribués aux différentes étiquettes d’auto-annotations correspondantes, c’est-à-dire une annotation externe aux données ici analysées. Puis une visualisation temporelle des changements de niveau basé sur le temps de début de chaque étiquette est produite comme sur le modèle de la Figure 51 ci-dessous.

En abscisse est représenté l’axe temporel de l’expérience, en ordonnée, le niveau de « glu » approximé sur les annotations *FoT\_PE*. Il s’agit donc là de la visualisation d’une « courbe mesurée » de glu socio-affective. Sur chaque graphique, deux seuils sont affichés : un premier seuil d’animisme s’accompagnant des premières occurrences de paraphrases de commandes, et un second seuil typé « compagnon » où les personnes commencent à considérer le robot comme « *une présence* », « *quelque chose qui tient compagnie* » et où la nature des événements interactionnels commence à varier énormément. Ce deuxième seuil semble surtout coïncider avec un seuil à partir duquel l’induction d’erreur du système (p. ex. faux positifs d’actions domotiques) donne lieu au phénomène de récupération d’erreur relevé dans la section 5.2.3.3, et où l’on pourrait dire que les personnes ont commencé à « gluer » avec le robot (axe vertical représenté en pointillé au premier point de dépassement de ce seuil). En deçà de ce seuil, les sujets abandonnent la commande non réalisée.

La Figure 51 illustre l’exemple du sujet S07. Ici nous voyons que la courbe est graduellement ascendante donc évolue dans un sens positif où le robot se renforce en perception animiste. Ces montées s’effectuent par palier et le dépassement du premier seuil montre que la personne a « glué » relativement rapidement par rapport à la ligne temporelle de l’expérience. Nous avons donc là une visualisation d’un pattern global d’un niveau de glu. Y aurait-il des mouvements systématiques visibles ou des dimensions de cette courbe qui peuvent être exploitables en vue d’une modélisation de la glu ou explicitant certains comportements humains ?

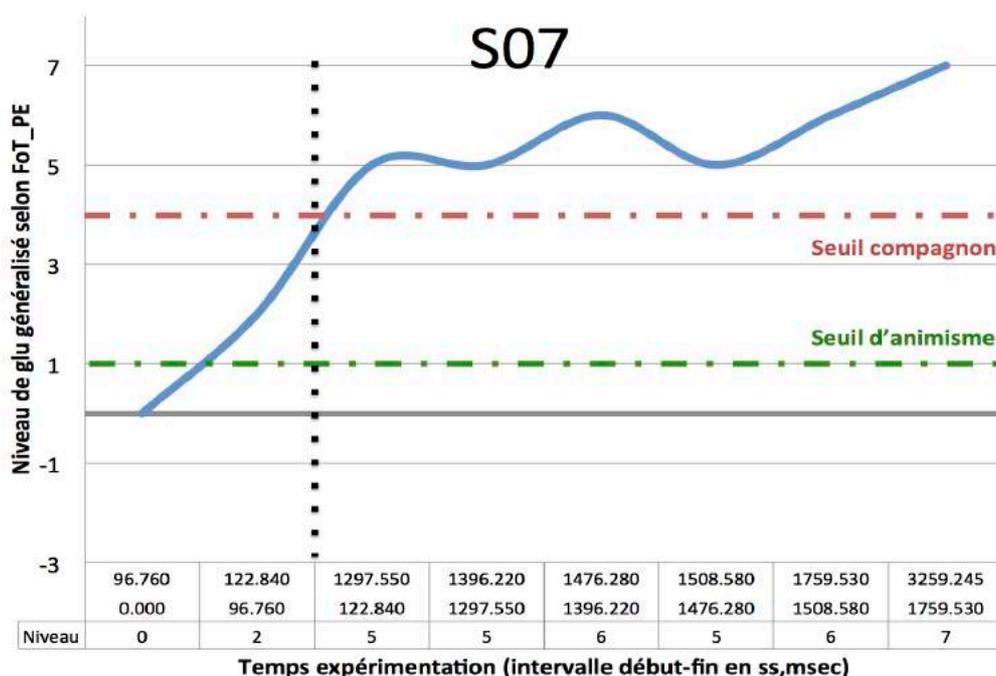


Figure 51 : Évolution du niveau de glu basée sur FoT\_PE – exemple du sujet S07

### 5.3.2.3. Patterns d'évolution de la glu des sujets

La courbe de visualisation appliquée aux 16 sujets donne les résultats de la Figure 52 ci-dessus. Nous constatons alors des patterns variables, mais aussi certaines formes remarquables, qui à l'instar de l'évolution morphosyntaxique (cf. section 5.2.1) semblent également illustrer un comportement en lien avec l'état d'isolement des personnes.

**(S01, Sa1)** : les deux premiers sujets de la première ligne vivent seuls. Leur particularité commune est qu'ils sont encore socialement actifs, mais qu'ils font face à un début de fragilisation accentuée (c.-à-d. problèmes de santé répétés) qui a tendance à les écarter de cette émulation sociale qu'ils tendaient à maintenir. Leur évolution est progressive, avec des oscillations ascendantes.

**(S05, S07, S15, S18, S16)** : les 5 sujets suivants sont ceux qui ont été décrits comme des personnes isolées par leur entourage aidant, mais dont l'isolement ne transparaissait pas directement dans leur discours. Selon ces sujets, l'isolement est le problème d'un autre, mais la description donnée est très exemplifiée et précise. De fait, certains cas extrêmes laisseraient même à penser qu'il y a peut-être une forme de déni d'isolement par certains sujets. Cela rajoute une difficulté méthodologique et une réflexion éthique à mener, avec la nécessité d'être dans une dynamique de co-construction, car il existe un réel risque de perturbation de la personne en la confrontant à son isolement. Pour ces personnes, le pattern est vraiment systématique, à savoir une croissance régulière et progressive. Par ailleurs, il semblerait également que la vitesse de progression (visible par rapport à la pente de la courbe) est en lien avec le degré d'isolement de la personne, ou en d'autres termes, l'impact que ce dernier a pu avoir sur la dégradation de la capacité relationnelle de la personne. En effet, les personnes qui sont vernaculièrement perçues comme « sociables » ont tendance à passer le seuil de « glu » relativement rapidement (p. ex. S07 qui est une personne d'apparence très sociable). A contrario, cette progression semble plus lente pour les personnes qui souffrent davantage de leur état d'isolement et qui sont perçues comme « moins sociables ». Ces dernières présentent

un plus grand risque d'être mises en échec en situation de communication (p. ex. S16 qui a beaucoup de mal à rester concentrée et « radote pas mal » selon les aides à domicile).

**(S12, S11) :** ce sont des cas de personnes, qui ne souffrent pas à proprement parler d'un véritable sentiment d'isolement. Elles sont encore actives, plus isolées d'un point de vue « géographique » ou « situationnel », mais ce sont des personnes qui ne présentent pas de fragilisation physique apparente ou handicapante. Pour ces personnes, le processus est extrêmement rapide. De même, la perception d'Emox passe par de petites phases d'oscillations régulières à un haut niveau, au-dessus du seuil de gluage « compagnon ». Dans le même profil, nous avons également (S10) annoncé comme une personne « très sociable », mais avec une courbe de glu qui n'oscille pas. La seule différence, avec les deux autres sujets du même profil réside dans le fait que ce sujet S10 avait récemment été très affecté et fragilisé par un deuil, et que malgré les apparences, dans l'intimité des entretiens au domicile, une souffrance de solitude est très clairement apparue dans ses propos. Ainsi, une hypothèse liée au processus de glu socio-affective est peut-être la perte de cette dynamique de passage d'état socialement haut, mais non équivalent, avec un aplatissement de la courbe, ce qui rejoint la dynamique des sujets du groupe 2.

**(S04, S06) :** ce sont les deux personnes pour qui les conditions expérimentales ont été plus compliquées, ponctuées de panne domotique ou robotique. Il est intéressant, tout de même, de voir qu'il est possible de retrouver des patterns qui tendent à décrire le contexte interactionnel. La particularité observable est une oscillation très importante en amplitude, qui tantôt dépasse le seuil « compagnon » ou de glu, tantôt descend sous le seuil d'animisme. C'est notamment ces premières observations qui nous ont permis de faire le constat de la récupération d'erreur. Mais, une autre interprétation possible à tester serait de voir si ce profil de courbe avec des oscillations amples ne dépend pas aussi du profil d'isolement de ces personnes. En effet, ces dernières sont peut-être les cas les plus typiques d'isolement qui étaient recherchés par l'expérimentation : des personnes vivant seules, qui étaient connues pour être des personnes « très sociables », mais qui commencent à souffrir d'une fragilisation physique engendrée et/ou engendrant un sentiment isolement (oralisé ou non mais constatable). Elles semblent d'ailleurs présenter quelques ratés communicationnels et relationnels dans leur quotidien. Ainsi, pour avoir été dans le rôle du magicien, il est difficile de dire si la panne du système a été fortuite ou si elle a été inconsciemment provoquée par le magicien qui essayait de se raccrocher à cette interaction complexe. Ses comportements auraient ainsi pu « forcer » des erreurs du système du fait de sa nature humaine qui faisait des actions extrêmes pour l'outil EmOz. Les traces des magiciens sont ainsi des éléments importants pour voir si, par exemple, la fréquence ou la cadence, voire le nombre de répétitions d'un même stimulus n'est pas en augmentation, ce qui donnerait une plus grande probabilité à la survenue de la seconde interprétation.

**(S03, S14, S19, S17) :** ce sont quatre sujets que l'on pourrait qualifier de « non glués », voire presque négativement glués pour le cas de S17. Parmi ces sujets, nous avons S19 et S14, qui avaient des troubles auditifs importants et qui ont déclaré ne pas entendre le robot, et plus précisément entendre un son, un « borborigme », mais pas la « mélodie » qui est produite<sup>97</sup>. Cette surdité à la prosodie nous permet d'avoir des sujets contrôles<sup>98</sup>. Et, en effet, l'observation d'un non-gluage dans cette situation tend vers la confirmation de notre hypothèse de primitives permettant la construction de la glu. Les deux autres personnes sont

---

<sup>97</sup> Cela a été testé à nouveau lors de la phase de débriefing en demandant si la nature d'un stimuli produit par un expérimentateur est identique ou pas à celle produite par le robot, et les personnes n'arrivaient effectivement pas à discriminer les stimuli de même nature prosodique).

<sup>98</sup> Malheureusement la difficulté de recrutement fait que nous n'avons pas pu avoir suffisamment de sujets pour se permettre d'avoir un nombre plus important de situations contrôles qui sont pourtant nombreuses.

## Analyses des productions vocales : vers une prosodie du dialogue ?

des cas relativement atypiques parmi le panel des sujets. S03 est une personne relativement jeune (68 ans), mise en retraite anticipée du fait de forte fragilisation physique, et que les partenaires sociétaux nous ont pointée suspectant une forme d'isolement dans « sa manière d'être » avec les aides à domicile. Pour ces trois personnes, la technologie en tant que telle pourrait être utile et intéressante, mais uniquement pour des handicapés moteurs et ne reste qu'un outil et rien de plus. Ainsi, leur courbe de glu gravite autour du seuil d'animisme, sans jamais atteindre le second seuil de glu. S17, quant à elle, est un cas qui a été suivi par de nombreux acteurs durant l'expérience, dont la psychologue de l'association d'aide à la personne. Elle était notamment dans une phase de dépression importante, et elle était extrêmement opposée à ce qu'une technologie comme la robotique rentre dans son environnement social. Sa courte interaction avec le robot a eu pour effet un renforcement de glu, mais négativement.

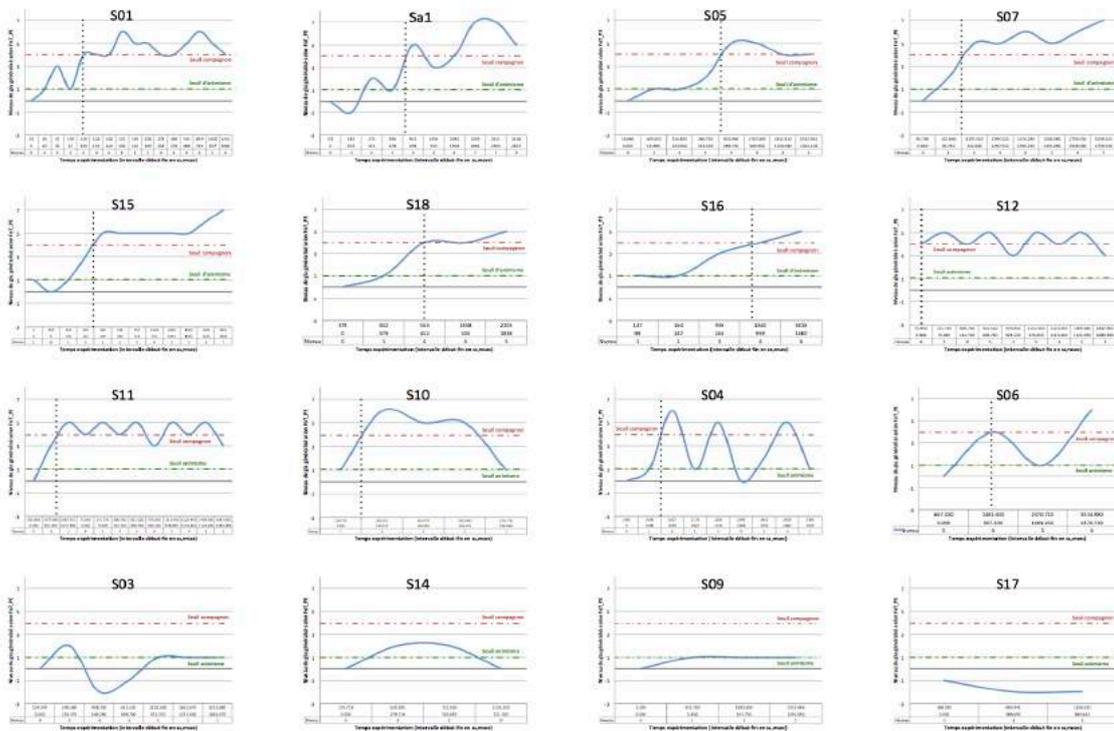


Figure 52 : Évolution des FoT\_PE nivelés – patterns pour 16 sujets

### 5.3.3. Evolution de « gluage » des commandes

Les sujets ont principalement été sollicités pour annoter leurs propres productions, ce qui transparaît dans la *FoT\_ProdSuj*. Chaque étiquette d’auto-annotation a été numérotée chronologiquement selon son ordre d’apparition, un indice qui est exploité pour produire des courbes de visualisation, non plus perceptives, mais sur une dimension plus effective. Nous présenterons donc ici la manière dont les sujets ont par eux-mêmes glué, en regardant s’ils restent sur un mode d’exécution des commandes semblable ou différent par rapport aux productions précédentes. Cela reflétera donc si l’information socio-affective portée dans la prosodie de la commande (mais plus globalement celle du comportement général de la personne) reste la même ou change au cours de l’expérience.

#### 5.3.3.1. Visualisation temporelle des productions interactionnelles

L’indice des étiquettes de *FoT\_ProdSuj*, associé au temps où débute cette étiquette, permet d’obtenir une courbe telle que présentée en Figure 53. La courbe illustre plus particulièrement l’évolution de « la manière de faire la commande ». En effet, l’un des buts de l’auto-annotation est de relever la valeur socio-affectivement fonctionnelle des énoncés produits par le sujet, afin de pouvoir mettre une étiquette sur la manière dont a été effectué cet énoncé (p. ex. « *je suis à la découverte, je teste en lisant* », « *c’est un ordre, je veux qu’il m’écoute, je sais qu’il va le faire* », « *c’est un ton professoral* », « *je prends des pincettes pour le dire* » etc.). Ainsi, cette courbe ne dit pas si l’état est positif ou négatif, et ne s’intéresse pas à la valeur-même de cette information, mais illustre l’évolution temporelle de chaque « fonction communicationnelle » par laquelle passe un sujet. En effet, chaque étiquette de l’auto-annotation a été chronologiquement incrémentée à chaque passage de nouvel état. Ainsi, la courbe nous permet de voir si un sujet est resté sur une fonction donnée ou s’il a changé rapidement d’état, et à quel moment. Sur cet exemple du sujet 16, nous observons ainsi qu’il existe des moments où l’interaction évolue rapidement quand la pente est raide, et qu’elle reste sur une même fonction au niveau des paliers. Nous voyons également qu’à certains moments de l’interaction, le sujet revient sur des « manières de faire » antérieures, au niveau des pics descendants.

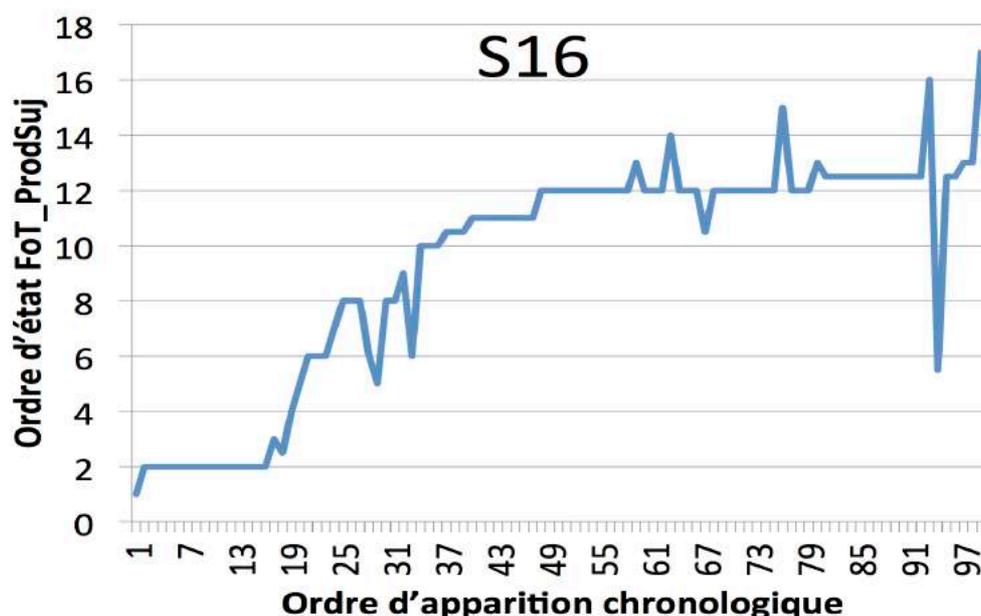


Figure 53 : Évolution temporelle des états fonctionnels des commandes du sujet basée sur *FoT\_ProdSuj*

### 5.3.3.2. Patterns d'évolution des « processus de gluage » des sujets

La production de la courbe sur l'ensemble des 16 sujets nous montre à nouveau des patterns d'évolution variables. Un premier point observable est que la courbe d'états de production est globalement ascendante pour tout sujet, ce qui revient à dire très simplement que la manière dont les sujets parlent évolue au cours du temps. De manière générale, ces courbes semblent également être cohérentes par rapport aux courbes *FoT\_PE* des sujets, avec des « retours en arrière » qui coïncideraient avec une descente de la courbe d'oscillations perceptives d'Emox. Pour les personnes qui ont « glué progressivement » ou qui n'ont pas glué, nous voyons aussi que les paliers sont plus marqués et plus longs (p. ex. S14, S15, S16, S17).

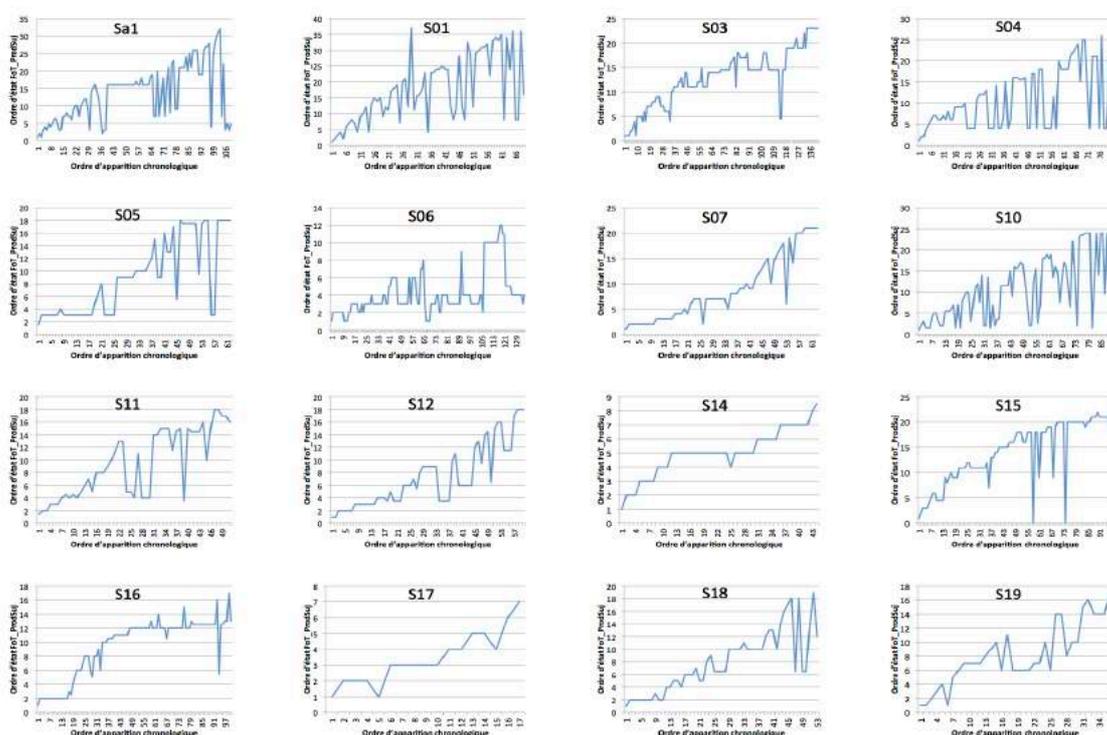


Figure 54 : Patterns d'évolution des valeurs socio-affectives fonctionnelles de 16 sujets à la base de leur *FoT\_ProdSuj*

Cette représentation permet donc de repérer globalement l'état d'évolution de la commande, pas à pas, ou commande à commande. Cette simplification a le défaut de faire perdre la notion de durée de chaque processus. En effet, même si nous savons que la manière de faire du sujet change très vite, nous ne pouvons pas savoir si ce processus est fait de manière étendue et progressive sur un temps long ou s'il est abrupt, car réalisé sur un temps très court. En effet, l'interrogation porterait sur le fait de savoir si c'est l'itération répétée d'un type d'interaction qui est à l'origine du changement de niveau de glu ou si c'est un événement fort qui en est le déclencheur, voire une combinaison des deux. Dans le dernier cas, est-ce que l'itération est uniquement facilitatrice ou serait-elle indispensable au changement ? Nous allons essayer de nous intéresser à cette dimension de durée à travers les deux sections suivantes.

### 5.3.4. Evolution d'états latents du sujet

Le *FoT\_Gn* permet de repérer globalement des états généraux et latents des sujets par rapport à l'ensemble du contexte communicationnel. Ainsi, nous repérons de grandes « séquences » qui peuvent comporter plusieurs commandes, mais dont le nombre est variable, et qui sont associées à un état d'esprit plus global de la personne. Par exemple, le sujet peut successivement passer par des états tels que « *plutôt à l'aise – je pensais que c'était pour tester l'appart, j'étais prête* », « *il fallait faire un apprentissage, je le faisais sagement – je m'oblige* », « *là je commence à lui parler* » (...) « *maintenant je trouvais qu'il obéissait bien, c'était son boulot, je ne suis plus émerveillée* » (...) « *je me dis je pense que je l'ai adopté, il faut vivre avec lui dans l'appart et bah ça se passe plutôt bien entre nous* » etc.

Ces états latents permettent donc d'avoir une indication de la dynamique interactionnelle qui rajoute une dimension de durée de chaque phase d'interaction.

#### 5.3.4.1. Visualisation temporelle des durées de phases latentes

Les durées des séquences *FoT\_Gn* sont chronologiquement représentées sur la ligne temporelle expérimentale de chaque sujet afin de visualiser leur variation. La courbe obtenue pour cette visualisation est disponible Figure 55, en prenant pour exemple le sujet S01. En noir est représentée la courbe de tendance générale qui est calculée sur les durées consécutives. À partir de cette dernière, le sujet semble intercaler des séquences courtes et plus longues en début et en fin d'expérience. Cependant, dans les phases intermédiaires, durant le travail interactionnel du sujet avec le robot, les états semblent très courts, avec des durées similaires.

Remarquons que l'ordre de grandeur de la durée est propre à chaque personne et n'est pas normalisé sur ces figures, tout comme précédemment. En effet, le temps cognitif et interactionnel est propre à chacun. Si nous rapportons l'échelle à un standard identique à chaque sujet, il serait plus facile de comparer entre les sujets ceux qui ont des temps d'interaction du même ordre, ce qui en soi pourrait potentiellement être un indicateur sur l'état de la personne. En revanche, le panel de sujets est ici trop faible pour que ce type de comparaison soit véritablement exploitable.

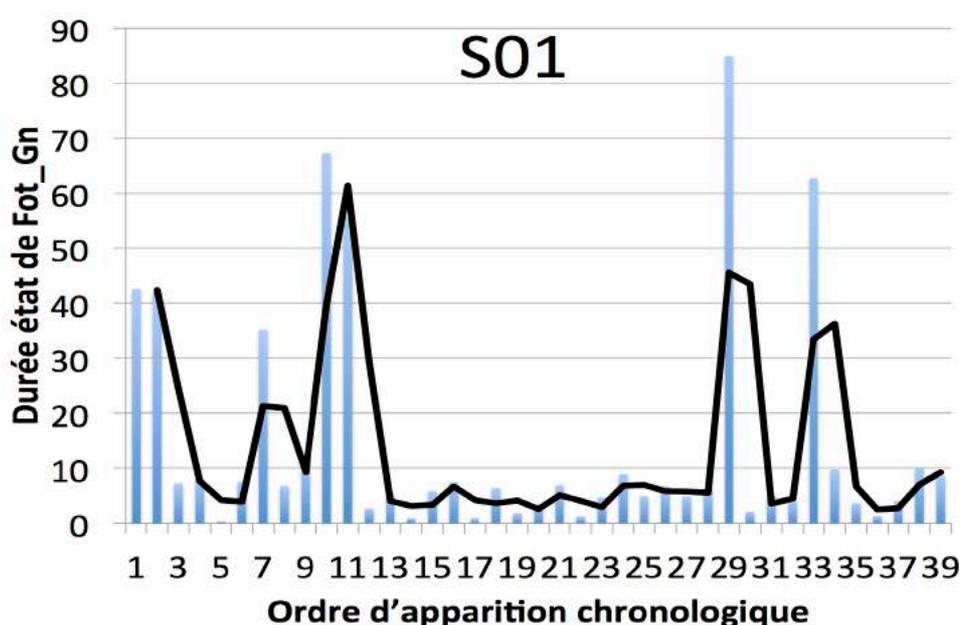


Figure 55 : Évolution des durées des séquences *FoT\_Gn* – Exemple de S01

### 5.3.4.2. Patterns d'évolution des états latents des sujets

L'application de la visualisation de la durée des phases latentes sur les 16 sujets est représentée Figure 56. La dynamique interactionnelle semble propre à chaque individu, mais il existe également des patterns communs. Nous avons des sujets tels que Sa1 et S04 pour qui alternent des temps longs et des temps courts sur l'ensemble de l'expérience ; des sujets comme S07, S11, S17 pour qui les différents états sont relativement courts en milieu d'expérience, et plus ou moins long en début ou en fin d'expérience ; et enfin d'autres pour qui nous avons des augmentations et diminutions progressives des durées (p. ex. S16, S18).

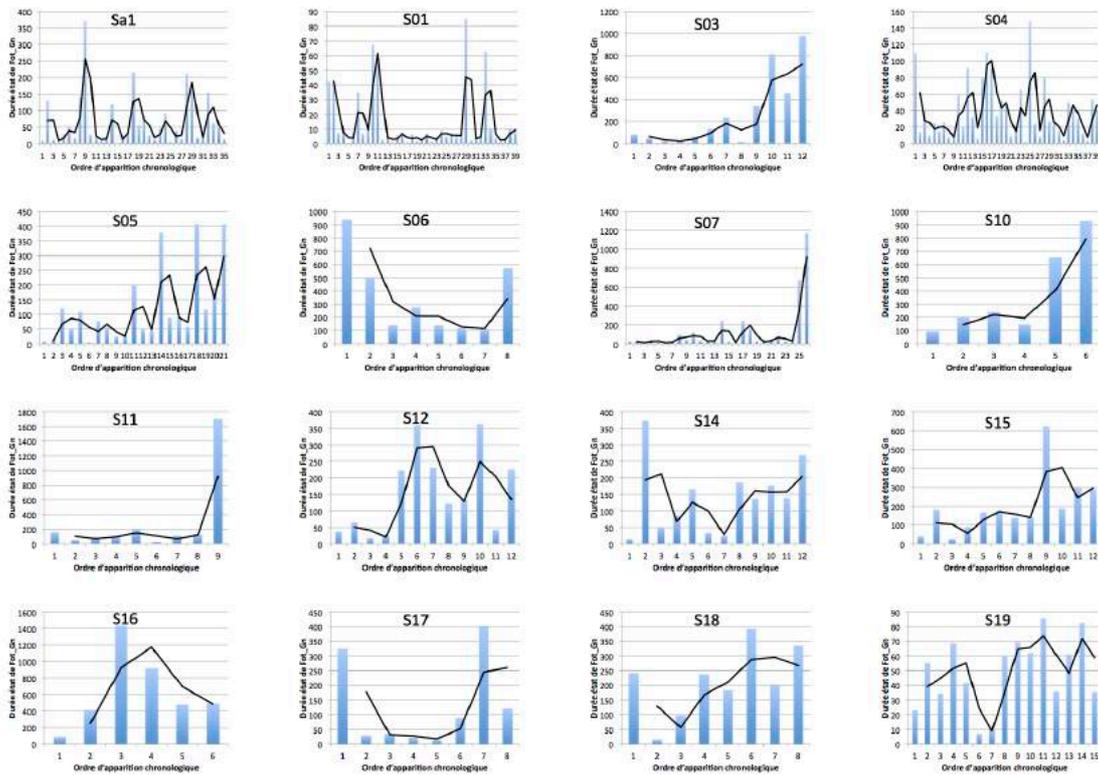


Figure 56 : Évolutions des durées des séquences pour les 16 sujets retenus

Ainsi, en couplant cette courbe avec celle de la FoT\_Gn, mais sur un alignement non pas séquentiel, mais temporel, nous aurons la possibilité de voir si les changements d'état « de la manière de faire des sujets » sont denses ou non sur une séquence, pour voir si le rythme de variation ou la valeur des fonctions expliquent ou non un changement d'état plus global.

Le défaut de cette représentation est de ne pas pouvoir correctement accéder aux moments décisifs de l'interaction, à savoir le timing temporel qui explique les transitions et les changements d'état ou de séquence. Ainsi, la section suivante va essayer de mettre en évidence les éléments qui permettraient de repérer les points de rupture de l'interaction.

### 5.3.5. Détection de temps interactionnels forts ponctuels

Les deux sections précédentes se sont particulièrement intéressées au sujet, mais l'interaction ne résulte pas seulement de lui, mais de l'ensemble des autres sous-systèmes avec lesquels est construit le système interactionnel global. Différentes sources de stimuli vont donc venir le modifier, de la même manière qu'il les modifie à l'image des *FoT\_ProdSuj*. Parmi la quantité d'objets interactionnels que le sujet gère naturellement, certains semblent l'avoir marqué plus particulièrement, ces événements portant des valeurs socio-affectives variables. C'est la tentative qui a été à l'initiative de l'auto-annotation *FoT\_IPC* qui essaie de pointer les « moments » où il se passe quelque chose (c.-à-d. réactions du robot, réponse domotique, un événement de contexte) qui a marqué le sujet, associant la valeur informative qu'il attribue à cet événement. Ces auto-annotations sont représentées par étiquettes de type : « *je me suis dit, mais il se déplace – c'était un peu étonnant* » (accompagnant le moment du premier déplacement d'Emox), « *c'est un peu l'étonnement parce qu'il répond, parle* » (à l'émission du premier bruit du robot) ou encore « *il m'a amusé à ce moment-là* » (associé à un mouvement spécifique du robot réalisé à la manette), d'autres mouvements pouvant aussi être décrits par « *qu'il bouge, qu'il se tourne un peu vers moi, je me suis demandé ce qu'il attendait, mais là je me dis non, mais attends il attend que je parle euh... qu'est-ce que je peux faire* ».

#### 5.3.5.1. Distribution et durées des événements remarquables

Chaque occurrence d'événement interactionnel remarquable est indiquée par un point bleu. Le graphique présente en abscisse l'avancement temporel de l'expérience, et en ordonnée, la durée (en secondes) du moment « où il se passe quelque chose » d'après l'auto-annotation. À nouveau, cette durée est relative à chaque sujet et n'a pas été normalisée, nous restons sur une durée relative, courte ou longue par rapport au temps d'interaction intrinsèque du sujet. De manière générale, il s'agit de moments relativement courts de moins d'une minute, comme cela est illustré en Figure 57 pour le sujet S03.

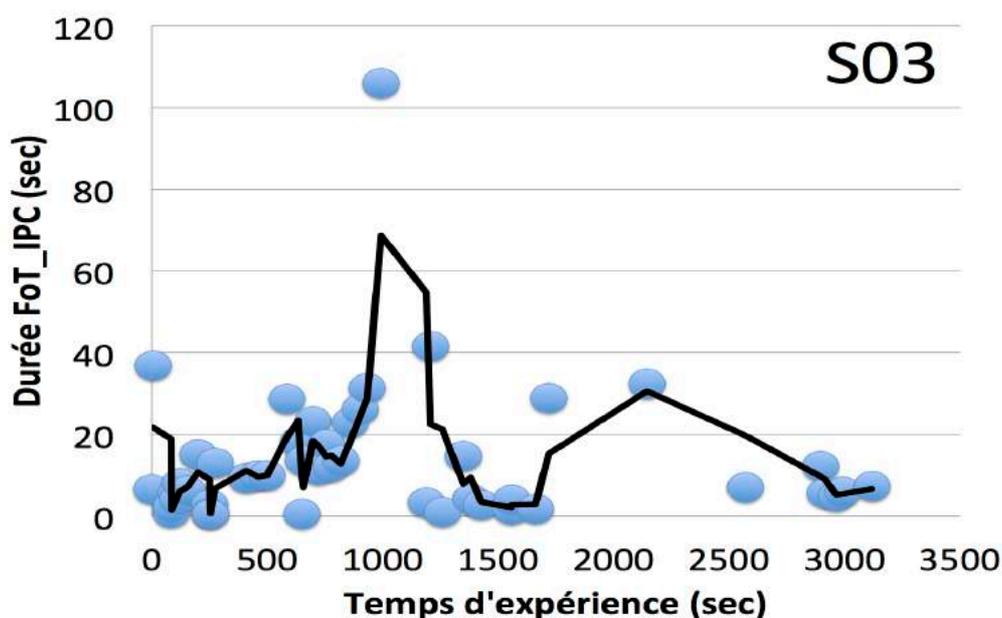


Figure 57 : Marquage temporel des événements et tendance d'évolution de leur durée

À l'appui de la distribution de ces événements remarquables, les occurrences semblent être le plus notables dans la première moitié de l'expérience, avec des moments associés relativement courts (une vingtaine de secondes en moyenne, même si sur certains événements, il s'agit de quelques secondes, jusqu'à une minute pour d'autres). Ils sont notamment un peu moins fréquents dans le dernier tiers de l'expérience. Dans le premier quart de l'expérience, nous remarquons aussi des événements relativement plus longs. Ces événements sont également plus rares que les autres et dont la durée pourrait être un indicateur de changement. En effet, ce typ d'événements peuvent potentiellement modifier l'état relationnel entre le sujet âgé et le robot, mais cela ne reste qu'une supposition. En réalité, chaque point pouvant être chacun un possible point de rupture ou un point de renforcement interactionnel, une perspective possible d'exploitation de ces données serait de pouvoir les utiliser comme des « marqueurs » sur lesquels nous pourrions zoomer pour voir quels objets interactionnels sont observables durant ce temps et de les superposer aux autres courbes, afin de regarder si ces événements ont effectivement eu un effet sur les états relationnels latents des interactants.

### 5.3.5.2. Comparaison inter-sujets des événements remarquables

Sur la Figure 58, récapitulative pour les 16 sujets retenus, la distribution et la durée sont extrêmement variables, ce qui est cohérent, car chaque expérience est unique dans son déroulement. Il n'est pas possible de remarquer sur cette Figure si les moments portent sur le même type d'événements et si les valeurs sont communes. En revanche, à partir de ces figures il serait possible d'extraire plus facilement des stimuli à choisir pour une validation perceptive par exemple, car ils ont plus de chance d'avoir un effet communicationnel important.

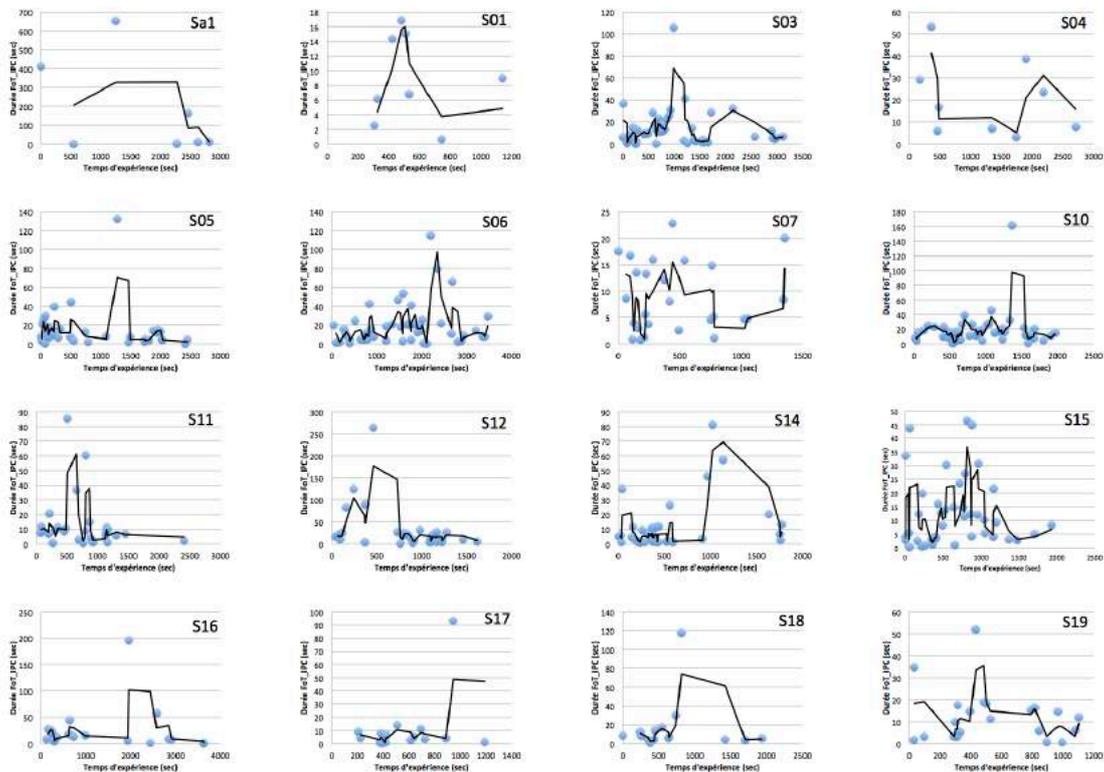


Figure 58 : Patterns de distributions des événements interactionnels ponctuels avec l'évolution de leur durée sur les 16 sujets retenus

En revanche, ce qu'il est possible de voir est la conséquence du délai temporel entre la réalisation de la séance d'auto-annotation et l'expérience. En effet, pour Sa1 et S01, S04 dont les auto-annotations ont été réalisées six mois après l'expérience<sup>99</sup>, le nombre d'occurrences de moments relevés est beaucoup moins important, que pour d'autres sujets pour lesquels l'auto-annotation a été réalisée en moyenne dans les deux mois après l'expérience. Ce délai d'exécution de la séance d'auto-annotation avait volontairement été rallongé dans le protocole pour essayer de conserver l'annotation des événements les plus marquants pour le sujet. En revanche, une analyse plus poussée pourrait donner lieu à une formalisation méthodologique encore plus fine par l'évaluation de la finesse et de la pertinence des auto-annotations obtenues en fonction de ces délais. En effet, nous pourrions définir plus précisément la grandeur de délai nécessaire pour obtenir un certain niveau de précision des étiquettes et d'un filtrage du « bruit », engendré par le caractère exceptionnel venant perturber le quotidien de la personne. Ce délai n'est cependant pas le seul facteur expliquant la quantité de moments remarquables, également influencés par la mémoire et l'état général de la personne le jour de l'auto-annotation ou encore son profil (p. ex. si elle est de nature « bavarde » ou non). Cependant, ces derniers paramètres devraient être relativement contrôlés, car le rôle de l'expérimentateur est de pouvoir modérer et/ou solliciter les personnes durant cette séance en fonction de son comportement et donc d'uniformiser la qualité des étiquettes attendues.

---

<sup>99</sup> Le délai plus important est surtout lié à la fragilisation physique des personnes qui ont été malades et hospitalisées peu après l'expérience.

## Chapitre 6. Système SASI : *Socio-Affective Speech Interaction*

### 6.1. Modélisation et calculs de niveaux de glu socio-affective

Des propositions explorant les fonctions à modéliser seront abordées dans ce chapitre. Elles contribuent notamment à la robustesse des technologies et des systèmes d'interaction. Mais avant cela, le début de ce chapitre rappellera et situera dans un bref état des lieux, notre positionnement amont dans la représentation :

- du processus interactionnel,
- de ses motivations,
- de ses conséquences,
- de son essence même.

En effet, cette prise de position débouchera sur une modélisation, qui soulignons-le, ne s'inscrit peut-être pas dans les voies les plus classiquement empruntées en traitement automatique des langues. Cependant, elle aura une incidence sur nos choix de modélisation informatique, s'étendant bien au-delà de cette thèse. Une conséquence dérivant de nos observations expérimentales réside dans le fait que notre système s'écarte de l'objectif standard d'un système de reconnaissance automatique de la parole ; soit de discriminer le contenu du message parlé, ce qui est couramment recherché afin d'assurer la continuité du dialogue. En effet, notre proposition ne vise pas à reconnaître les constituants de parole dont l'assemblage donnerait une valeur communicative donnée. Il s'agirait plutôt de reconnaître la glu socio-affective qui découle de l'interaction à l'appui des constituants structurels de parole permettant de reconnaître les propriétés irréductibles de la glu elle-même.

Cette perspective fait écho à la maxime d'Aristote « le tout est plus que la somme des parties » qui renvoie à la notion d'émergence structurelle. L'émergence, comme décrite par la philosophie du XIX<sup>e</sup> siècle est à la base de nombreuses réflexions théoriques (cf. Sciences et avenir – hors série *L'énigme de l'émergence*, 2005). Elle est considérablement visible depuis les années 70, d'une part à l'appui des modélisations en mécanique quantique ou en biologie moléculaire, et, d'autre part à travers les réflexions des sciences cognitives. Ces dernières accompagnent le développement des technologies de l'information et de la communication qui interrogent à la fois la complexité des systèmes et les réflexions décrivant l'esprit et la conscience. Or ce concept, qui est motivé par la physicalité de propriétés observables, est abordé par de multiples sciences du vivant. Par exemple, en biochimie (comme les phénomènes enzymatiques apparaissant par la modification de configuration de protéines), la botanique (p. ex. l'apparition de configuration spatiale de certaines brousses selon des périodicités données), en zoologie (p. ex. les comportements auto-organisés chez les insectes) mais aussi en sciences humaines, au sein desquelles le langage est l'exemple typique décrivant cette perspective. Dans chacun des phénomènes cités, les constituants forment un système complexe d'où « émergent », des propriétés authentiques et irréductibles, qui influencent elles-mêmes de manière causale l'organisation de ces composants. En d'autres termes, la parole est composée de multiples constituants interactionnels de nature et de granularité variables (p. ex. phonétiques, phonémiques, morphosyntaxiques, prosodiques, gestuels, regards, etc.) organisés en système complexe. La configuration globale et évolutive de ces éléments déboucherait sur une glu qui deviendrait ainsi une propriété émergente de ce système (cf. chapitre 1 section 1.2). Cependant, cette notion d'émergence structurelle pose encore un problème mathématique et cognitif sur le plan de la modélisation. Et ce sont en particulier ces notions holistiques qui sont abordées par les propositions diverses de systèmes

complexes<sup>100</sup> depuis le début de l'intelligence artificielle. En effet, dès lors que la glu socio-affective s'illustre par des marqueurs variés se modifiant au cours du temps, la tentation serait de calculer le poids des paramètres communicationnels de ces marqueurs qui influencent la manipulation de la glu. Or, selon Aubergé, (cf. *workshop: Emotions, Attitudes, Illocutions, Modality* – présentation GSCP Aubergé, 2012) : « la valeur, le niveau, les dimensionnalités de la glu socio-affective, et notamment son indice d'évolution, est un facteur émergent global dont la nature est intrinsèquement multiparamétrique, et la calculabilité intrinsèquement non modularisable ». Notre approche va ainsi considérer que l'essentiel du dialogue est capturable dans l'évolution de la glu. En d'autres termes, la trace tangible dans les signaux sociaux marque l'évolution de la relation, et les actes interactionnels eux-mêmes pourraient être considérés comme des sous-systèmes d'un système global émergent dont la surface serait palpable par la glu. Nous chercherons donc à reconnaître la glu qui par son tracé de la relation est un indicateur aussi bien de l'état socio-affectif, d'une part de celui qui à un temps donné produit l'acte (l'humain dans notre application), et, d'autre part de celui qui le perçoit (reconnaissance de l'acte vocal par le robot). De cette valeur de glu, nous déduirons quelles formes interactionnelles et quels contenus sont produits par le sujet humain afin d'orienter la reconnaissance vocale en tant que telle. Et dans ce même processus, les énoncés du robot seront également produits selon l'évolution relationnelle que le système doit pouvoir manipuler en suivant l'hypothèse de la glu altruiste dans notre première approche.

Une possibilité, s'appuyant sur les méthodes informatiques développées depuis l'augmentation des méthodes calculatoires, serait d'utiliser des algorithmes statistiques de *mapping* entre différentes dimensions langagières de la glu. De ces dimensions, nous pourrions définir des poids (c.-à-d. proposer des cartographies arborescentes où chaque nœud représente un constituant associé à son poids dans le système). En revanche, comme tout processus stochastiques de « boîte noire », cette pondération serait définie de manière calculatoire en fonction : 1) de la proximité des formes d'états proches ou d'états consécutifs, 2) de la charge des liens associés à un chemin stochastique. Or aucun de ces critères ne permettent de véritablement comprendre le choix d'une valeur (c.-à-d. que chaque branche de l'arbre serait empruntée aléatoirement en fonction de la probabilité contextuelle de passer sur un nœud donné). Et cela n'expliquerait pas non plus la raison de ce choix, si ce n'est qu'il est statistiquement le plus probable, alors que nous ne connaissons pas *a priori* les lois qui déterminent le comportement sous-jacent de la glu, ni celles de la distribution des objets langagiers. Rappelons que les données expérimentales recueillies dans cette thèse ne découlent pas d'un *bootstrap* de contenus structurels imposés (c.-à-d. que l'hypothèse ne porte pas sur la probabilité d'apparition d'objets langagiers selon des conditions contextuelles et interactionnelles définies). La capture spontanée des données a bien été orientée par des hypothèses posées *a priori* sur la nature de la glu et notamment sur les éléments primitifs déclencheurs d'un attachement d'une intensité forte, positive et hors dominance (ce sont ces éléments recueillis, mesurés et évalués entre 2004 et 2010 par Aubergé, Audibert, et Rilliard, 2004; à Signorello, Aubergé, Vanpé, Granjon, et Audibert, 2010). Ces hypothèses portaient ainsi sur des principes théoriques majeurs et non sur l'apparition de *patterns* produits en réponse aux objets primitifs. Or comme le suggérait des études antérieures (Aubergé et coll., 2004), de telles hypothèses nécessitent méthodologiquement d'être confrontées aux résultats d'analyses des données observées.

Il s'agirait alors de comparer les représentations de :

- la glu à « valeur mesurée » de manière externe avec l'auto-annotation. Il s'agirait de mettre en évidence les objets concomitants aux stades de glu qui permettent de retracer les prémisses structurelles holistiques (c.-à-d. les résultats d'analyses

---

<sup>100</sup> Billet de blog sur le sujet écrit par Delahaye, 2003 accessible sur: <http://www.scilogs.fr/complexites/le-tout-est-il-plus-que-la-somme-des-parties/> (consulté le 28 juin 2014).

- issus de l'étiquetage *smart data* – cf. Figure 29 et Figure 30 section s 3.3.3.3 et 3.3.3.4) ;
- avec la glu à « valeur calculée ». À partir de ces mêmes objets interactionnels, il s'agirait de définir le tracé de la glu à l'aide d'un indicateur sous la forme d'un score global et unique (c.-à-d. des hypothèses). Dans ce second cas, nous initions le calcul statistique après que les données brutes aient été orientées, afin de proposer une modélisation du fonctionnement qui sera soit continu, soit organisable en catégories de temps cognitifs relationnels (c.-à-d. traçables par des points de rupture).

L'idée est donc de pouvoir générer un coût global d'une séquence interactionnelle, mais calculable à partir des divers champs existants, et pour chacun desquels un coût expert réglable sera donné. Ce coût global est l'optimisation minimale des poids relatifs donnés à l'ensemble des champs interactionnels mesurés : nous sommes donc ici dans un problème typique des familles de méthodes de calcul de distance optimale d'édition ou d'accès tolérant au lexique.

Ce « réglage expert » permet ainsi :

1) à court terme, de tester les hypothèses sur les structures langagières qui sous-tendent ce réglage. La précaution : à travers des boucles agiles, vérifier la modélisation qui permet l'automatisation du comportement posé en hypothèse ;

2) d'affiner les entrées d'un système de RAP qui soit au cœur du système technologique. La méthode : formaliser l'apparition des micro-expressions vocales, notamment en intégrant des effets de variations prosodiques ;

3) et à plus long terme, de trouver des solutions de modélisation d'un système de dialogue hybride. L'objectif : élaborer un système en mesure de combiner la nature multimodale des données à travers le score de glu socio-affective.

En particulier, cette « agilité » dans la méthodologie et le réglage est cadrée par la référence permanente aux données externes (c.-à-d. les valeurs de l'auto-annotation).

### ***6.1.1. Algorithme Damereau-Levenshtein modifié***

La proposition ci-dessous est une tentative de pondération des coûts minimaux évoqués, mais qui se limite pour le moment à une seule dimension. Il s'agit de celle qui se trouve sur l'axe syntagmatique, soit la plus facilement appréhendable et observable dans la morphosyntaxe des commandes. L'idée est ainsi de pouvoir suivre l'évolution lexico-morphosyntaxique des commandes produites par les sujets au cours de l'expérience. En effet, ces changements coïncident aux stades de glu que les personnes âgées elles-mêmes autoannotent de manière externe. L'observation est alors opérée par rapport à la référence imposée (c.-à-d. qui correspond au temps cognitif de la condition initiale de la glu). Cette forme initiale se présente comme un énoncé à choix lexical et morphosyntaxique imposé, qui est toujours construite au sein d'une forme verbale infinitive (cf. liste des commandes imposées aux sujets au début de l'expérience – Annexe 5). Dès lors, la pondération de distance pourrait modéliser cette évolution linguistique de la commande en accord avec les niveaux de glu autoannotés. La distance d'édition est typiquement le degré de différence entre les deux formes comparées. La méthode de calcul de cette distance d'édition la plus utilisée est celle de (Levenshtein, 1966), dont un algorithme a été développé par (Wagner et Lowrance, 1975) et qui attribue un poids au type d'opération (insertion, effacement et substitution d'un élément de la chaîne considérée, soit ici la commande vocale). Ce calcul est modifié dans l'algorithme Damereau-Levenshtein qui associe une quatrième opération de « transposition » donnant la possibilité de noter en une seule opération le couplage délétion/insertion. En traitement automatique des langues, ce sont ces calculs associés à la précision qui permettent de calculer le WER (*Word Error Rate*) ou le TER (*Traduction Error Rate*) pour démontrer la robustesse des règles linguistiques associées à cet algorithme permettant de calculer les taux de différence entre

chaînes. Cependant, l'objectif n'est pas de réduire ce taux, il est d'exploiter ce taux afin de niveler le degré de glu.

Dans les algorithmes classiques de Levenshtein ou de Damereau-Levenshtein, chaque opération a un coût semblable, mais il existe des variantes qui associent la possibilité d'un coût différent par type d'opération : ces coûts sont réglables pour chaque opération. Cependant, une variante algorithmique qui proposerait des valeurs de coûts spécifiques aux éléments opérés ou à des catégories d'éléments, et non pas aux fonctions d'opération, ne semble pas s'être développée dans la communauté taliste. Dès lors, la seule variante que nous ayons trouvée est celle proposée par Ghneim et Aubergé (en 1997) qui propose de régler des coûts spécifiques sur les opérands dans une grammaire en entrée qui peut également organiser les éléments opérés en classe. Ainsi, les coûts minimaux associés doivent être relatifs à la nature des outils langagiers mis en jeu, soit par rapport à la valeur socio-affective, où les coûts les plus importants seraient associés aux transformations/opérands ayant le plus grand effet, autrement dit, aboutissant à un changement d'état de la glu socio-affective.

Par exemple, considérons les énoncés suivants :

Commande : mettre la lumière

1) allumer la lumière du salon

2) mettons la lumière Emox

Dans les énoncés proposés, deux opérations de même nature sont visibles, mais l'effet de l'information portée par chaque opérande n'aura pas les mêmes impacts vis-à-vis de la glu socio-affective dans les opérations suivantes :

- une substitution de « mettre » : ici, nous voyons qu'« allumer » en 1) est une variante sémantique qui garde la même modalité syntaxique que la référence ; alors que « mettons » en 2), construit sur le même objet verbal, est morphologiquement transformé par un changement de personne, qui apporte une dimension inclusive (c.-à-d. faire ensemble) du fait de l'usage de la première personne du pluriel. Ainsi, le poids de l'élément 1) est moindre par rapport à celui du 2) sur la dimension de « nivelage » de glu.

- une insertion « du salon » et « Emox » : l'attention peut porter sur deux choses. Tout d'abord, l'opérande n'est pas forcément équivalent à une unité syntaxique minimale, mais peut faire l'objet d'une expression plus ou moins longue (p. ex. « du salon » = 2 unités syntaxiques minimales, mais considérées comme un seul « marqueur » et opérande). La seconde est à nouveau liée à la nature de l'élément inséré. Dans le cas 1), l'élément est un élément étiqueté « précision » dans le corpus d'éléments satellites (soit directement concomitant à la commande), soit une précision désambiguïsant l'actionneur de la commande domotique à activer. Dans le cas 2), il s'agit d'un élément noté « nominatif » utilisé pour s'adresser au robot par son nom. Cet élément pourrait aussi être un marqueur de politesse comme « s'il te plaît » ou encore des micro-expressions telles que des rires avec un coût de « renforcement » de glu plus conséquent.

Nous avons ainsi utilisé cette variante algorithmique proposée par (Ghneim et Aubergé, 1997) qui permet également d'établir des grammaires de pondération des opérands du calcul des distances d'édition. Cet algorithme a été développé pour des éléments de l'écrit orthographique et avait pour but général d'augmenter la dimension pédagogique du correcteur orthographique par une approche plus globale de la faute dans un énoncé. À l'appui du corpus ORTHOTEL composé d'énoncés, formant un concordancier de fautes orthographiques, cet algorithme avait pour particularité de pouvoir donner des coûts à des ensembles morphosyntaxiques pour se défaire du traitement caractère à caractère ou mot à mot pour apporter un coût relatif minimal des fautes plus cohérentes. Cette segmentation « par ensemble » est ce que nous souhaitons transposer à notre contexte, en supposant que nous avons une forme canonique de commande qui subit des transformations par ensemble, dont la nature porte des coûts minimaux en cohérence avec l'évolution de la glu socio-affective. Le

passage du traitement oral au traitement écrit s'adapterait ainsi au niveau du système RAP, qui prend en entrée un signal acoustique, puis lui associe une forme écrite reconnue en sortie. C'est cet élément décodé qui ferait ainsi l'objet du traitement algorithmique pour déterminer le niveau de glu reconnu à partir de l'énoncé de commande vocale domotique de la personne âgée (cf. partie gauche de la Figure 61). Cela orienterait l'intelligence des *feedbacks* du robot, qui ferait suite à la partie de reconnaissance vocale du système.

### 6.1.2. Grammaire de coûts des variations de structures langagières

Une version générique<sup>101</sup>, actualisée de l'algorithme Ghneim-Aubergé a été redéveloppée en parallèle des travaux d'étiquetage et d'auto-annotation de cette thèse. Cet algorithme de calcul de distance d'édition, associé à cette segmentation d'ensemble (sur une fenêtre de 1 à 5 unités lexicales), est construit sur une grammaire de coûts qui intègrent différentes variations langagières illustrant sur l'ensemble de l'expérience l'état de la glu, en s'appuyant sur une mesure observée (l'auto-annotation) et prédite (les hypothèses vocales qui sont produites sur le robot). Ainsi, le coût de chaque transformation est décrit dans une base de règles dont une première version se présente comme décrite en Figure 59. Les règles sont exprimées dans la syntaxe :

#### (X) --> (Y) Coût

Les formes X représentent une chaîne de caractères de la commande de référence, les Y décrivent une chaîne de caractères de la commande observée.

<p>❑ Effacement (un seul coût)</p> <p>(Infinitif) -&gt; ( ) 80 (Complément) -&gt; ( ) 80</p>	<p>Infinitif = {mettre, éteindre, descendre, monter, ...} Complément = {la lumière, les stores, la télé, ...}</p>	<p>//ex. La lumière //ex. Allez mets</p>
<p>❑ Insertion (un seul coût)</p> <p>( ) -&gt; (Pronobjme) 2 ( ) -&gt; (Pronobjnous) 3</p>	<p>Pronobjme = {me, m, moi} Pronobjnous = {nous}</p>	<p>//ex. Mets-moi la lumière //ex. Mets nous la lumière</p>
<p>❑ Substitution (plusieurs coûts selon catégorie)</p> <p>(Infinitif) -&gt; (Infinitif) 0 (Infinitif) -&gt; (ImperatifSg) 10 (Infinitif) -&gt; (ImperatifPl) 15 (Infinitif) -&gt; (VerbOn) 20 (Infinitif) -&gt; (VerbOnPouvoir) 25 (Infinitif) -&gt; (VerbTuDevoir) 35 (Infinitif) -&gt; (VerbTu) 40 (Infinitif) -&gt; (VerbTuPouvoir) 45 (Infinitif) -&gt; (VerbTuVouloir) 50 (Infinitif) -&gt; (VerbJeVouloir) 55</p>	<p>Infinitif = {mettre, éteindre, descendre, monter, ...} ImperatifSg = {mets, éteins, descends, monte, ...} ImperatifPl = {mettons, éteignons, descendons, montons, ...} VerbOn = {on met, on éteint, on descend, on monte, ...} VerbOnPouvoir = {on peut mettre, on peut éteindre, ...} VerbTuDevoir = {tu dois mettre, tu dois éteindre, ...} VerbTu = {tu mets, tu éteins, tu descends, tu montes, ...} VerbTuPouvoir = {tu peux mettre, tu peux éteindre, ...} VerbTuVouloir = {tu veux mettre, tu veux éteindre, ...} VerbJeVouloir = {je veux mettre, je veux éteindre, ...}</p>	<p>//ex. Mettre -&gt; Mettre OUI Mettre -&gt; Allumer //ex. Mettre -&gt; Mets OUI Mettre -&gt; Allume //ex. Mettre -&gt; Mettons OUI Mettre -&gt; Allumons //ex. Mettre -&gt; On met OUI Mettre -&gt; On allume //ex. Mettre -&gt; On peut mettre OUI Mettre -&gt; On peut allumer //ex. Mettre -&gt; Tu dois mettre OUI Mettre -&gt; Tu dois allumer //ex. Mettre -&gt; Tu mets OUI Mettre -&gt; Tu allumes //ex. Mettre -&gt; Tu peux mettre OUI Mettre -&gt; Tu peux allumer //ex. Mettre -&gt; Tu veux mettre OUI Mettre -&gt; Tu veux allumer //ex. Mettre -&gt; Je veux mettre OUI Mettre -&gt; Je veux allumer</p>

Figure 59 : Grammaire de coûts : présentation simplifiée des règles pour le calcul d'une valeur intégratrice pour la commande émise par les sujets<sup>102</sup>  
(proposée par Tsvetanova et Aman, rapport interne Interabot)

Cette première tentative de grammaire intermédiaire restera à enrichir à travers les boucles agiles de méthodologie qui devront faire suite à cette thèse. En effet, l'ensemble des travaux de labellisation/alignement présentés dans le Chapitre 4, ainsi que les observations du Chapitre 5 visent à repérer les formes remarquables pour tenter d'associer des coûts minimaux aux opérantes les plus pertinents sur le plan communicationnel. Ainsi, nous

<sup>101</sup> Cette adaptation (programmée en Pascal), a été effectuée par F. Aman (postdoc sur Projet Investissement d'Avenir Interabot). Un outil permettant de manipuler l'algorithme et les grammaires est disponible en open-source sur : <https://github.com/FredericAMAN/GrammaticalDamerouLevenshtein.git>

<sup>102</sup> Cette première version de grammaire de règles est proposée par L. Tsvetanova (2015) dans le cadre de son master et du projet investissement d'Avenir Interabot. La grammaire actuelle s'appuie sur les données de 7 à 12 sujets du corpus EEE et peut être augmentée à 26 sujets.

voyons par l'extension des observations sur un nombre plus grand de sujets que la variation morphosyntaxique est un premier point d'entrée, mais qu'elle reste limitée. Les sujets qui ont fortement « glué » peuvent aboutir à ces états sans forcément laisser d'indices dans la morphosyntaxe des commandes. En revanche, ils peuvent avoir produit un nombre de micro-expressions beaucoup plus important, ou encore présenter des variations qui semblent apparaître sur d'autres dimensions telles que la prosodie (p. ex. à travers des écholalies appliquées sur des commandes de référence donc de morphosyntaxe inchangée) ou d'autres marqueurs interactionnels multimodaux. De fait, cet algorithme qui se limite à un paramètre peut être une brique qui centralise l'ensemble des paramètres technologiquement détectables, dans la perspective d'un système de dialogue hybride. Les réflexions qui se poseraient alors au niveau de la modélisation porteraient sur les opérateurs qui interviennent entre les différentes dimensions des signaux sociaux échangés. En effet, les effets des différentes modalités se multiplient-ils ? Ou, sont-elles complémentaires, s'annihilent-ils ? Est-ce toujours une question d'intermodalité ou de dimensions communes entre les différentes modalités qui interagissent entre elles pour donner un effet communicationnel ? L'idée finale est que la glu soit pondérée par un score final unique qui soit représentatif du résultat des processus communicationnels à un temps donné, afin de pouvoir suivre son évolution à travers une courbe comme celle illustrée en section 5.3.

## **6.2. Prototype de reconnaissance d'états de glu-parole**

La solution technologie proposée à la personne âgée étant un robot majordome d'un habitat intelligent qui se commande vocalement, la première brique importante du système sera une solution de RAP qui soit capable de reconnaître la commande de l'appartement. Les productions vocales des sujets âgés sont globalement définies en deux parties : d'une part, les ordres domotiques adressés au robot qui sont les premiers éléments à faire reconnaître dans un système de RAP, d'autre part, les productions hors commandes. Ces productions hors commandes ont différentes perspectives d'approche.

### **6.2.1. Reconnaissance de la parole et des caractéristiques des voix âgées**

L'un des problèmes génériques des systèmes de RAP est la difficulté de reconnaissance de la voix âgée. En effet (Baba, Yoshizawa, Yamada, Lee, et Shikano, 2004 ; Aman et coll., 2012 ; Aman, Vacher, Rossato, et Portet, 2013) rapportent que les modifications physiologiques du vieillissement amènent à transformer les tissus du tractus vocal, modifiant ainsi la nature acoustique des sons émis par les personnes âgées, générant des chutes des taux de reconnaissance par les systèmes. Cette baisse de performance s'illustrerait par trois à quatre fois plus d'erreurs comptabilisées sur la parole âgée. Ces erreurs sont notamment générées par des spécificités des modèles acoustiques, problématiques notamment pour certaines consonnes fricatives et occlusives (Aman et coll. 2012). La parole spontanée de la personne âgée sera par ailleurs sujette à de nombreux bruits de bouche liés à la présence de dentier, d'essoufflement, de raclements de gorge, de clics etc. dont les traces sont également nombreuses dans le corpus EEE. Ces sons, généralement non-référencés dans le vocabulaire du système ni dans les modèles de langage, sont donc un facteur bruit important par rapport au système classique. Ainsi, les méthodes d'apprentissage spécifique permettent d'améliorer ces taux de reconnaissance. Cependant, dès lors qu'elle est exposée à une « parole émotionnelle » (simulée ou non), la reconnaissance est à nouveau dégradée. La gestion de la parole âgée spontanée est donc un défi en soi. L'avantage du corpus EEE est qu'il est une première source de données pour composer un corpus d'apprentissage (bien que très restreint dans la perspective *Big Data*), mais qui peut néanmoins contribuer à l'amélioration des modèles de langage. En effet, il a permis de recueillir de nombreuses variations de formes de la commande domotique dans une production spontanée susceptible d'apparaître en contexte réel. De plus, l'étiquetage des micro-expressions vocales et les transitions entre parole adressée au robot et non-adressée au robot sont potentiellement exploitables pour trouver de nouveaux modes d'apprentissage machine, pour reconnaître les éléments utiles à reconnaître et ceux à « ignorer ». Ci-dessous, un synoptique simplifié des étapes successives du système RAP est présenté en Figure 60.

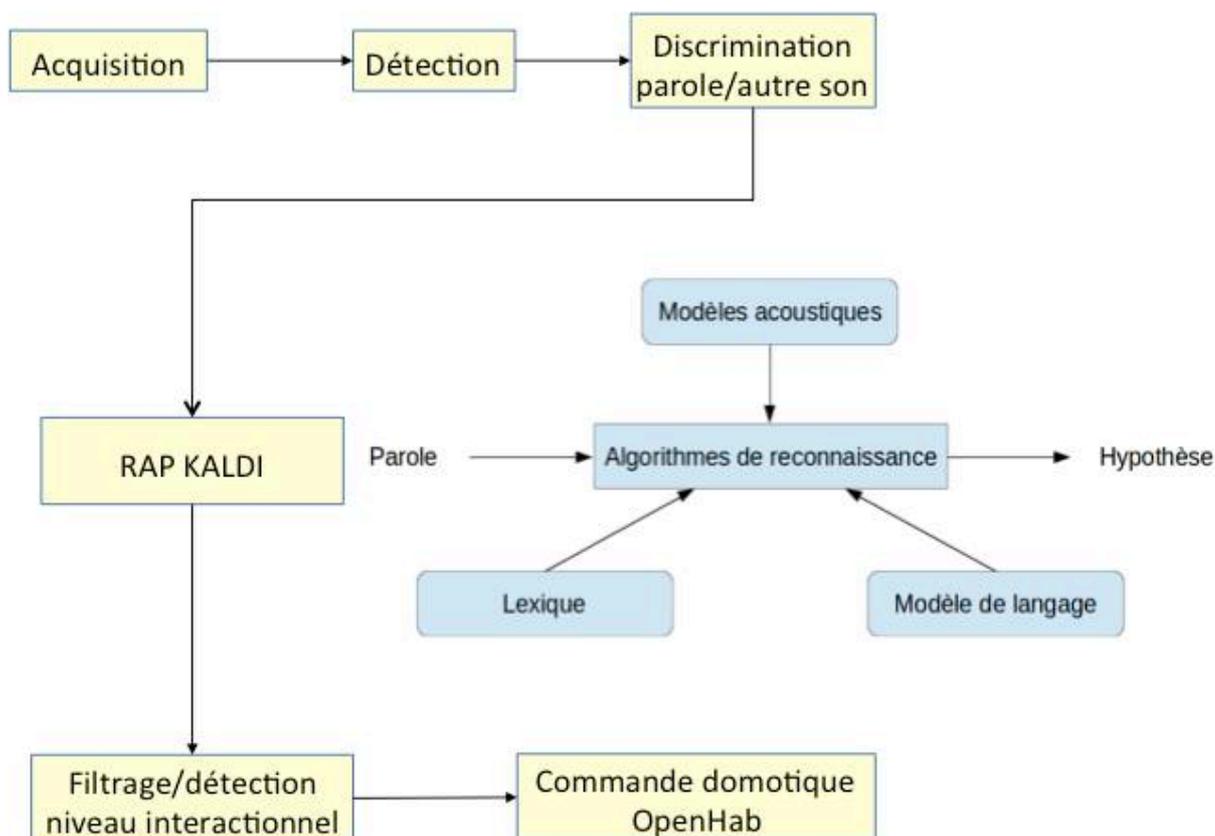


Figure 60 : Schéma global des entrées/sorties successives d'un système RAP Cirdox (adaptée dans projet Interabot depuis Aman, Vacher, Portet, Duclot, et Lecouteux, 2016)

À l'heure actuelle, nous avons pris appui sur un système d'acquisition de données AuditHis (Vacher, Fleury, Portet, Serignat, et Noury, 2010) développé dans le cadre des projets ANR Cirdo et Sweet-Home (Vacher, Bouakaz, et coll., 2014 ; Vacher, Lecouteux, et coll., 2014) et une version d'un système RAP associé, CirdoX, adapté pour la voix âgée (Aman et coll., 2016). Les apprentissages du système sont renforcés par des corpus externes (Gigaword, BREF120 et ERES38 pour les spécificités de la voix âgées). Le système CirdoX comporte notamment une brique permettant de discriminer les séquences de parole et de non-parole, il rentre ensuite dans une brique construite à partir de la boîte à outils Kaldi (Povey et coll., 2011), permettant d'effectuer des apprentissages par réseaux de neurones (DNN), adapté pour effectuer des expérimentations temps réel et hors temps réel<sup>103</sup>.

<sup>103</sup> La mise en œuvre technique de la brique de reconnaissance vocale a fait l'objet du travail postdoc de F. Aman, dans le cadre du projet Investissement d'Avenir Interabot, à partir des idées et des observations développées dans le cadre de cette thèse et les apports théoriques de V. Aubergé.

Pour le prototypage et les mesures évaluatives, F. Aman a utilisé les conditions suivantes pour effectuer les expérimentations exploratoires du système sur les données du corpus EEE.

- L'entraînement de ce DNN a été effectué à partir du corpus BREF120 pour effectuer un décodage en temps réel. Les paramètres en entrée des données continues sont des coefficients de type *log-mel filterbank, raw* (non-adaptés et non-normalisés) et de dimension 40, calculés sur une fenêtre de 7 trames, auxquels un i-vector de dimension 600, représentant les caractéristiques du locuteur a été rajouté. Ce réseau de neurones entraîné est composé de 5 couches cachées, avec une p-norme de dimension 400 en entrée et 4000 en sortie, et 12000 sous-classes.

- Pour les expérimentations *offline*, un modèle acoustique GMM a été entraîné sur le corpus BREF120, de type triphone, et avec adaptation de type LDA, MLLT, SAT et fMLLR.

### 6.2.2. Vers une solution de détection de « niveaux de glu »

Nous avons exploité cet outil CirdoX pour détourner l'usage classique d'un système de RAP. Classiquement, l'objectif d'un tel système est d'avoir le meilleur taux de reconnaissance possible. Ce taux est évalué notamment sur le WER, soit la distance entre l'énoncé produit et l'énoncé reconnu sous forme de sortie textuelle, à partir d'une entrée sonore (le schéma général d'un système RAP est représenté sur la partie gauche de la Figure 61. En revanche, nous souhaitons ici exploiter ce WER différemment, non pas pour reconnaître le contenu lexical de la commande (qui reste en soi un objectif, mais plutôt secondaire dans cette approche), mais dans le but de reconnaître le niveau de glu associé à la morphosyntaxe de la commande comme nous l'avons précédemment vu avec l'algorithme Damereau-Levenshtein.

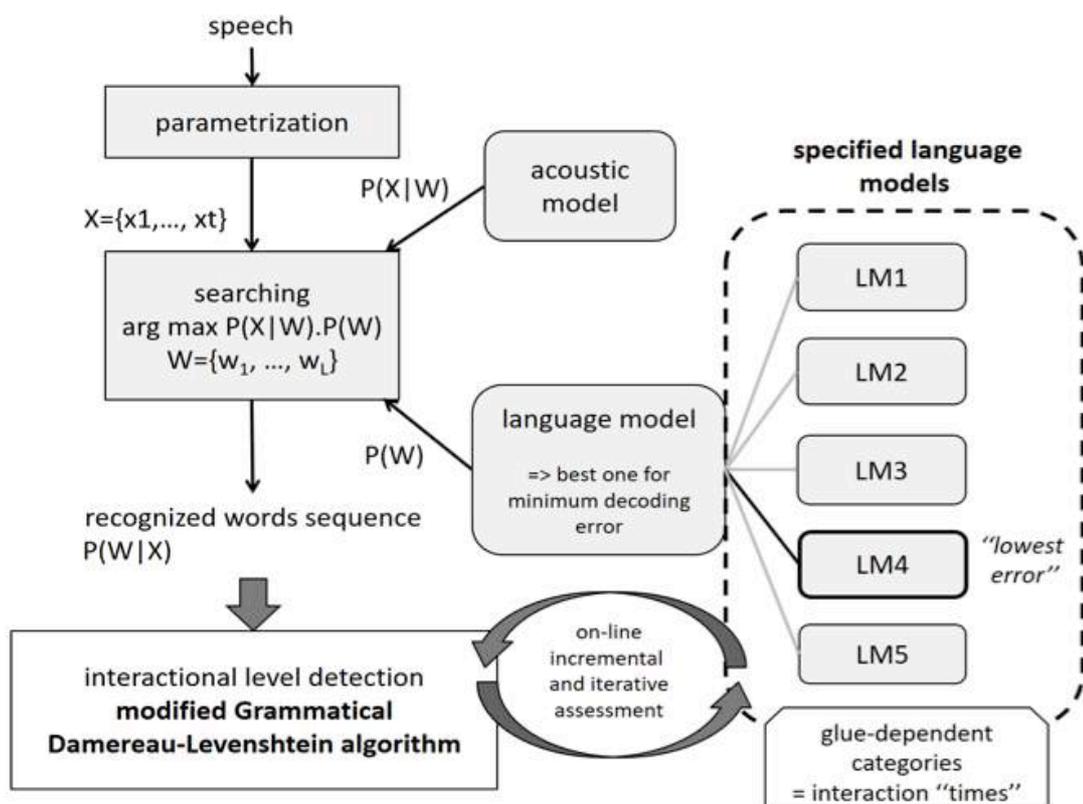


Figure 61 : Intégration l'algorithme de distance d'édition modifié dans le synopsis d'un système RAP

Les caractéristiques morphosyntaxiques ont été regroupées en suivant les variations d'auto-annotations indiquant un renforcement de glu. En rouge, sur la **Erreur ! Source du renvoi introuvable.**, est indiquée la répartition du nombre d'occurrences d'énoncés ayant servi pour construire le modèle dont les formes morphosyntaxiques principales sont rappelées. Pour apporter une représentation plus globale, la Figure illustre à titre indicatif les changements d'autres modalités ayant lieu en parallèle. Ceci récapitule l'ensemble des variations observables à travers les évolutions de niveaux de glu. Ces autres modalités peuvent faire l'objet d'autres briques de traitement technologique contribuant au calcul du score global de glu.

Pour le système RAP, un prototype intermédiaire a été développé sur ces principes, en s'appuyant sur la mise en place de 5 modèles de langage regroupés selon des caractéristiques morphosyntaxiques des commandes observées sur les données d'une dizaine de sujets.

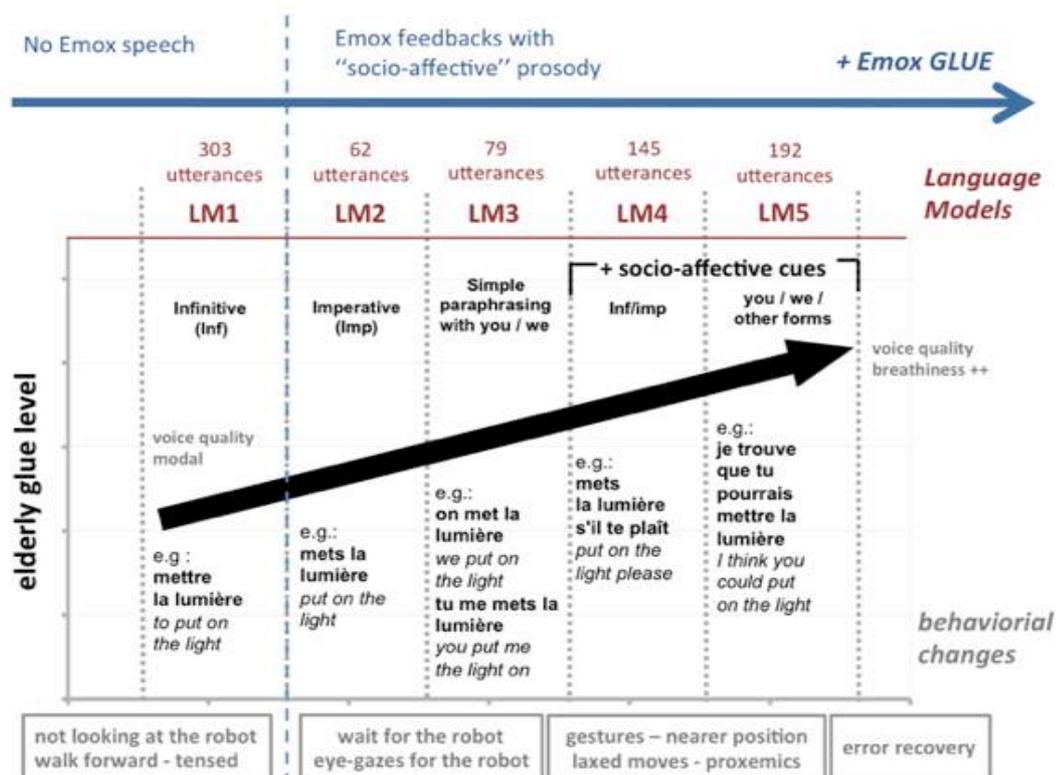


Figure 62 : Récapitulatif des variations langagières observées pour la définition de 5 niveaux de modèles de langage pour le système RAP

L'approche adoptée pour cela a été de restreindre la description des variations langagières dans de très petits modèles de langage, où chaque modèle ne décrirait les variations correspondantes qu'à un seul « niveau de glu ». Ensuite, après l'étape de décodage acoustique, les données seraient traitées en parallèle par tous les modèles de langage définis, et chacun donnera une sortie en fonction de ses règles et un WER associé. Ainsi, le WER n'aura pas pour objectif de mesurer la qualité du contenu lexical reconnu, mais il sera utilisé comme un déterminant de glu, soit le modèle avec le meilleur taux de reconnaissance (c.-à-d. celui ayant le WER le plus bas), comme celui décrivant le mieux le niveau de glu rencontré à un instant donné. L'algorithme ne se base, en réalité, pas sur le WER, mais plutôt sur le score de confiance associé à la sortie du modèle. Cependant, les scores de confiance des RAP sont toujours assez délicats à manipuler, car ils sont définis sur des intervalles aléatoires ou proportionnels par rapport à la répartition des populations observées. Or, les patterns de variations de la glu sont loin d'être définis ni appréhendés à travers les données du corpus EEE qui sont bien trop restreintes. Cette approche est donc une proposition temporaire, qui reste une piste de réflexion importante pour la modélisation d'un futur système de dialogue.

### 6.2.3. Proposition d'un algorithme de détection de glu

La Figure 63 représente un premier algorithme de reconnaissance de commande « nivelée en glu », associé au prototype du système RAP actuel.

À la sortie du décodage opéré par Kaldi (comme illustré en Figure 60), le système propose une hypothèse de l'énoncé testé (notée HYP). Cette hypothèse est l'entrée de l'algorithme, associée à un score de confiance (SC) et à la commande domotique effective à exécuter qui en découle (CMD). De plus, un score de glu calculé à partir de l'algorithme Damereau-Levenshtein modifié (D) est défini en donnant une hypothèse de glu. Ici, l'objectif est de mesurer la pertinence d'utilisation de modèles de langage multiples et de vérifier s'il est possible d'associer un niveau de glu à partir de ces petits modèles de langage supposés comme illustratifs de différents états de la glu. Les modèles sont ainsi ordonnés et testés à partir d'une variable itérative illustrant l'ordre du modèle de langage (noté i) permettant de tester par une boucle incrémentale l'ensemble des modèles. En sortie, nous souhaitons que le système fournisse les scores de confiance (SC) associés à chaque modèle, le meilleur déterminant l'ordre du modèle associé à un « niveau de glu » (GLU) théorique. En sortie, le système fournit également la (CMD) domotique à exécuter. En l'état, le (SC) permet de vérifier approximativement le modèle le plus efficace, mais ce paramètre n'est sans doute pas le plus pertinent et reste un objet de réflexion. La quantité de données est loin d'être suffisante pour un véritable apprentissage, en revanche, les modèles utilisés actuellement s'appuient soit strictement sur une partie du corpus EEE (noté E), soit sur un modèle plus générique entraîné également sur le corpus EEE, mais augmenté de Gigaword<sup>104</sup> (E+G). Par défaut, nous avons fixé un seuil arbitraire, permettant de générer une branche qui assure un filtre sur les données complexes (c.-à-d. distance d'édition trop importante). Ces données seraient difficilement associées à un niveau de glu ou de langage trop générique et ouvert pour la prise en charge efficace par un système RAP. À terme, cette branche pourrait être augmentée et affinée, pour discriminer le « bruit » (p. ex. parole hors commande) et les commandes du système.

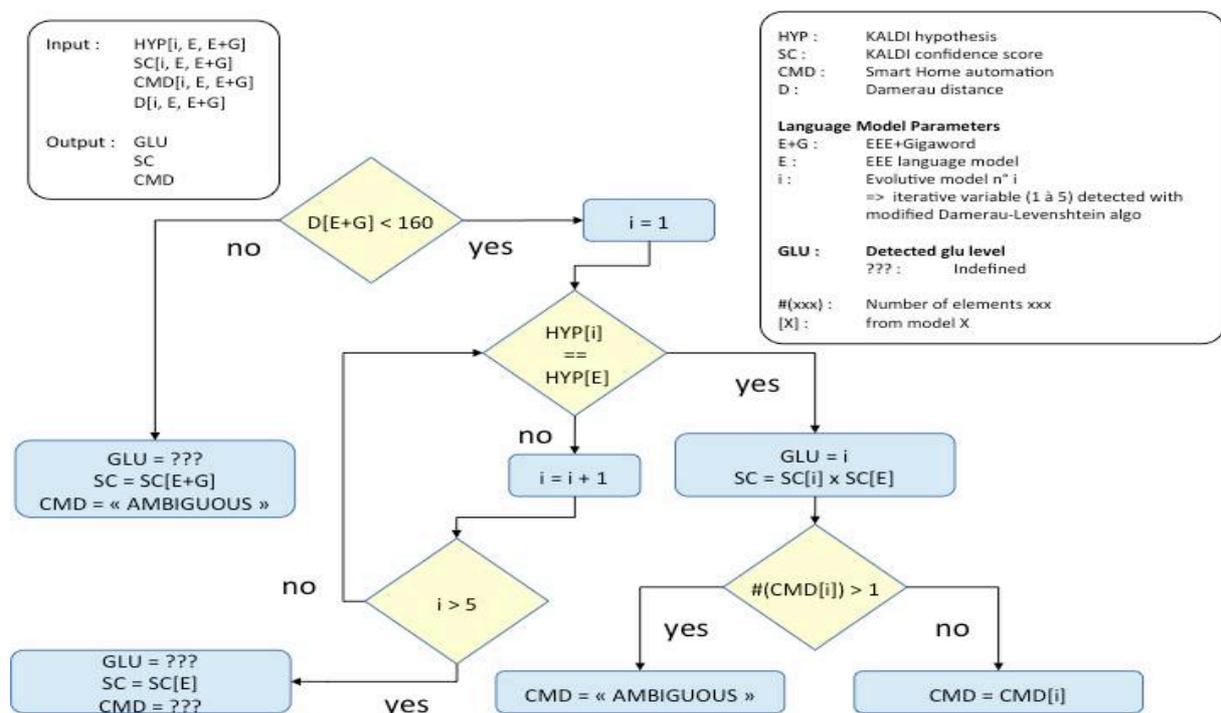


Figure 63 : Algorithme de reconnaissance des commandes (proposé avec Aman et Aubergé dans rapport interne Interabot)

<sup>104</sup> <https://catalog.ldc.upenn.edu/LDC2011T10>

Si l'hypothèse est analysable en rentrant dans une fourchette de distance d'édition manipulable, une boucle incrémentale est ouverte, en testant l'hypothèse sur l'ensemble des modèles proposés. Elle donne en paramètres de sortie les scores de confiance et un niveau de glu. L'algorithme est encore incomplet, mais permet de poser plusieurs questions de perspectives par rapport aux sorties illustrées. En effet, dans cette version, il existe des ambiguïtés, soit pour le niveau de glu à associer, soit pour la commande à exécuter. Il y aurait donc une partie complémentaire à cet algorithme qui serait associée à la décision du feedback du robot dans la suite de l'interaction, ainsi que la décision de la commande à exécuter même dans les cas ambigus. En effet, le système visant à être évolutif, nous pouvons poser deux pistes de réflexion qui influenceront le choix du feedback à émettre :

- la première est celle de la mémoire des états précédents. En effet, un niveau de glu est par définition ambigu, puisqu'il s'agit d'un mouvement fluctuant et approximatif par nature, mais la gestion de cette ambiguïté est un enjeu important. La question serait de savoir s'il est possible de rester dans le niveau courant de glu, sans apporter de perturbations dangereuses pour la personne. Elle consiste aussi à réfléchir si l'état directement précédent, voire le mouvement dynamique qui précède l'état courant, permet de passer à des niveaux supérieurs de glu, voire s'orienter vers une direction spécifique pour un apport utile à la personne.
- La seconde dimension, qui est intrinsèquement liée à la première, est l'intelligence perçue du système par la proposition de son niveau de robustesse sur l'activation de l'actionneur. Cette réflexion découle de l'observation du phénomène de dialogue mis en place par les sujets « glués », en particulier la manière dont les sujets accompagnent la gestion des erreurs, volontairement ou involontairement produits par le système. Objectivement, dans la perspective d'un système automatisé, « l'adaptation » de la production par les personnes âgées serait encore plus difficile à reconnaître. En effet, même si les reprises restent proches, les variations prosodiques associées ainsi que les ajouts de micro-expressions vocales bruyeraient grandement le signal de parole. En revanche, ce « bruit » porte des caractéristiques fortement marquées socio-affectivement car les personnes avaient pour intention que la commande est reconnue plus facilement. Il existe donc ici une information remarquable à fortes valeurs interactionnelles, qui si elle arrive à être gérée par le système, augmenterait la perception de la capacité communicationnelle et relationnelle du robot. Par exemple, la section 5.3.2.3 rapportait qu'un sujet, dès lors qu'il atteint un certain niveau de glu, était plus regardant face aux faux positifs d'actionneurs domotiques. Ces derniers ont pour effet d'engendrer des attitudes du type « c'est pas grave, on recommence » avec des indices interactionnels relevant d'un potentiel de glu élevé. Si ce faux positif a lieu avant ce niveau, le robot est totalement rejeté après quelques tentatives, qui elles sont beaucoup plus proches des premières commandes émises et qui respectent la forme morphosyntaxique donnée en consigne. Les personnes perçoivent alors une intelligence diminuée pour le robot. Ainsi, si l'algorithme détermine dans les états précédents les indices de niveau de glu, il peut varier sa réponse domotique pour éviter la production de faux positifs qui aboutirait à l'abandon d'usage de la technologie.

Dans le cadre méthodologique, la décision découle des analyses théoriques sur le constat des effets que les primitives interactionnelles ont pu avoir à différents niveaux de glu. Cependant, la décision du développement du système automatisé doit être issue d'une décision collective pour assurer la dimension éthique du comportement du système, qui, si

elle est choisie par l'industriel, se fera par la mesure du risque pris, éclairé en concertation avec les partenaires. Ces derniers auront, eux aussi, une idée plus précise de l'impact de l'interaction qu'aura cette technologie sur la personne âgée elle-même, mais aussi sur son écologie sociale et aidante qui influencera cette décision.

### 6.2.4. Modèles multiples de langage en fonctionnement parallèle

Le protocole de test de l'algorithme décrit dans la section précédente fait l'objet de validations croisées effectuées sur une partie des données du corpus EEE et dans différentes conditions. Cette section propose de s'intéresser à un résultat en particulier, portant sur la comparaison du WER entre les différents modèles spécialisés et un modèle global qui serait théoriquement le type de modèle de langage qui serait à la base de tout système de RAP. Ces résultats sont rassemblés dans le Tableau 9.

Afin d'effectuer ces mesures, une partie réduite du corpus EEE (sur 12 sujets) a été divisée en 2 groupes, k1 et k2. Les apprentissages sont dans un premier temps effectués sur k1 et les tests sur k2, puis croisés en apprenant sur k2 et testés sur k1. L'annotation gx représente des énoncés associés à de différents niveaux de glu (de 1 à 5), illustrés par les modèles de langage spécialisés « évolutifs ». La notation kxgall est associée aux énoncés du modèle générique prenant en compte l'ensemble des phrases du groupe x. Les modèles spécialisés kxg1 à kxg5 ont été interpolés avec le modèle kxgall du corpus EEE avec un poids de 90% donné aux probabilités des modèles spécialisés kxgy, et de 10% à celles du modèle kxgall. Enfin, les modèles de langage kxgy et kxgall ont également été interpolés avec un modèle appris sur les phrases du corpus OpenSubtitles<sup>105</sup>, dans le but de couvrir un vocabulaire plus important. Nous avons ainsi les modèles de langage spécifiés en niveaux de glu définis par la notation kxgy\_OpenSubtitles et un modèle générique kxgall\_OpenSubtitles. Tous ces modèles sont trigrammes et ont été appris avec l'outil SRILM.

Tableau 9 : Comparaison des WER basés sur les modèles de langage spécialisés et le modèle de langage générique

Test \ LMs	g <sub>1</sub> _OpenSg	g <sub>2</sub> _OpenSg	g <sub>3</sub> _OpenSg	g <sub>4</sub> _OpenSg	g <sub>5</sub> _OpenSg	g <sub>all</sub> _OpenSg
g <sub>1</sub>	6,70%	50,19%	30,10%	14,40%	23,25%	9,39%
g <sub>2</sub>	56,25%	25,57%	57,95%	44,32%	56,25%	35,23%
g <sub>3</sub>	44,16%	48,72%	13,68%	37,89%	25,64%	16,81%
g <sub>4</sub>	44,61%	60,56%	57,54%	22,84%	50,22%	25,75%
g <sub>5</sub>	56,22%	60,21%	52,22%	51,62%	18,31%	19,37%
					15,56%	18,11%

Le Tableau 9 illustre donc des tests effectués sur des phrases associées à plusieurs niveaux de glu (définis en ligne) avec le WER résultant de chaque modèle de langage (les colonnes) où la dernière colonne représente le modèle générique. Pour les phrases testées, le meilleur score de reconnaissance est associé au modèle de langage spécifique du même niveau de glu (en vert). Ce qui est plus intéressant de noter est que la généralisation du modèle, alors qu'il intègre les mêmes phrases d'apprentissage que les modèles spécifiques, augmente le taux de WER et est donc sensiblement moins efficace (comparaison case verte et bleu sur chaque ligne). Ceci a donc deux avantages, d'une part, nous augmentons la chance

<sup>105</sup> Source : <https://www.opensubtitles.org/fr>

d'avoir la reconnaissance de la bonne commande domotique à exécuter, d'autre part, nous avons l'avantage de pouvoir prédire de manière relativement robuste le niveau de glu, dont nous n'avons aucune idée si nous passons par le modèle générique.

Ainsi, l'apport détourné de cette méthode proposant de multiples modèles de langage est surtout lié au fait que sans passer par un algorithme descriptif de glu à appliquer, dont il serait complexe de définir les règles sur un modèle de langage générique, la simple discrimination du modèle spécifique le plus efficace permet de savoir à quel niveau de glu l'interaction se situe. Cette première solution simple semble donc une perspective intéressante pour la discrimination de l'état de la glu socio-affective et mérite certainement de pouvoir être développée.

### 6.3. Positionnement du système dans le contexte actuel

#### 6.3.1. Quelles évolutions à partir des systèmes de dialogue actuels ?

Le dialogue homme-machine (Jurafsky, 2000 ; Landragin, 2013) est généralement créé sur un schéma modulaire, présenté en trois temps (compréhension, gestion, génération), comme il est représenté Figure 64.

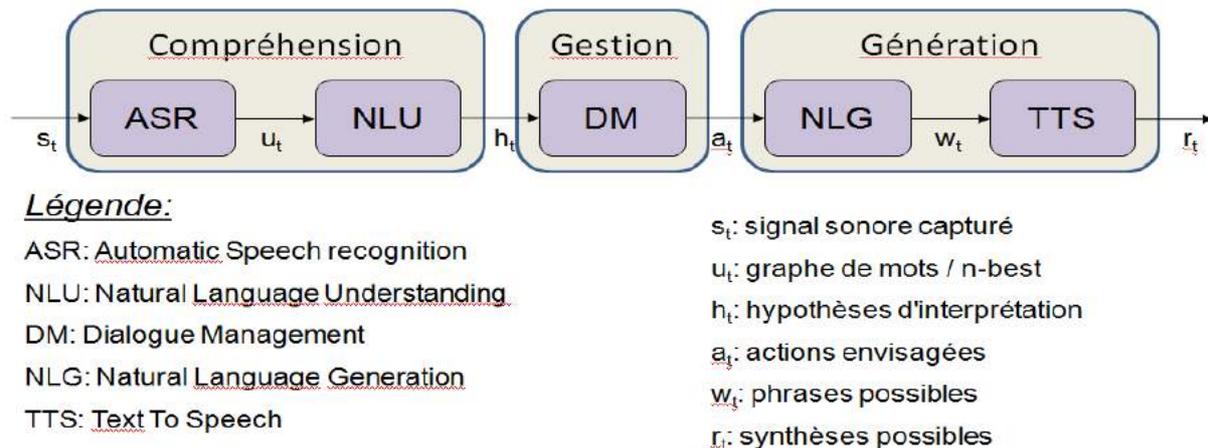


Figure 64 : Synopsis général d'un système de dialogue<sup>106</sup>

Sur le versant industriel, des outils de développement basés sur ce principe sont proposés comme des solutions robustes pour fabriquer des briques spécifiques à chaque système comme par exemple l'ODS – Orange Dialogue Studio (développé par OrangeLabs). Avec l'émergence des assistants personnels (Siri, Amazon Echo, Cortana, Google Now), de nombreuses applications tendent vers une intégration d'un système de reconnaissance vocale associé à un système d'interrogation de base de connaissances par contenu désambiguïté. Le dialogue, dans ces perspectives, n'est donc défini qu'à travers la valeur sémantique et son interprétation/exécution est liée à une résolution d'un problème émanant du contenu lexical. C'est notamment l'approche des systèmes tels que Watson (IBM) qui a remporté le Jeopardy face à l'humain, fondé sur le principe d'interrogation de bases de données répondant à des requêtes utilisateurs (Ferrucci, 2012 ; Laroche, 2015). L'assistant personnel est considéré dans une perspective multitâche et serait donc un système de résolution de tâches (Gašić et coll., 2013 ; Planells, Hurtado, Segarra, et Sanchis, 2013). L'intérêt accru pour les IoT (ou objets connectés) tend également vers des réflexions sur l'adaptation de l'assistant pour une personnalisation à l'utilisateur (Casanueva, Hain, Christensen, Marxer, et Green, 2015 ; Genevay et Laroche, 2016).

Les modalités d'interaction sont alors majoritairement orientées sur deux approches : une vision conversationnelle qui sous-tend une « métaphore humaine » et une vision orientée commande portant sur une « métaphore d'interface » (Edlund, Gustafson, Heldner, et Hjalmarsson, 2008). La première suggère la modélisation des variations de l'acte langagier dans un contexte restreint, mais visant une interface parlée relativement universelle (Rosenfeld et Harris, 2004). La seconde focalise sur le contenu même de ce qui est produit à partir d'un vocabulaire restreint défini. La plupart des modèles suivent l'approche des actes de dialogue (Austin, 1975 ; Searle, 1979), que l'on retrouve dans le codage des corpus annotés à travers des schémas du type DAMSL (Core et Allen, 1997). Ceci a pour conséquence, dans les années 70 jusqu'au début des années 2000, le développement des systèmes basés sur la

<sup>106</sup> Schéma tiré de l'article : <https://recherche.orange.com/dialoguer-avec-des-machines-entre-mythes-et-realite/> (consulté le 02/02/2017)

prise en charge de « paires adjacentes », soit des va-et-vient de tour de parole comme défini dans les approches de (Schegloff et Sacks, 1973), ne permettant qu'une prise en charge séquentielle de la parole échangée.

En revanche, depuis moins d'une dizaine d'années, nous voyons l'émergence de système de dialogue incrémental (Schlangen et Skantze, 2009 ; Skantze et Schlangen, 2009 ; Raux et Eskenazi, 2012 ; Khouzaimi, Laroche, et Lefevre, 2016) développé pour des agents tels que Eve, les systèmes comme NUMBERS ou encore le Pentonimo du projet Inpro. Alors que les anciens systèmes devaient attendre qu'un tour de parole soit explicitement terminé, ces systèmes permettent une prise en charge plus globale, autorisant le traitement de chevauchements de parole ou d'actes simultanés en prenant en considération une « chaîne de dialogue » de bout en bout. Le principe vise à modéliser l'adaptation globale et continue liée à une analyse d'un *self-monitoring*, à l'origine de l'ajustement de nos comportements communicationnels aussi bien en situation de *feedback* que de *backchannel* (Schlangen et Skantze, 2009).

Sans rentrer dans le détail des différentes méthodes d'apprentissage machine, de nombreuses se sont très largement développées autour des approches de type *Deep Learning* dans les systèmes de traitement de parole (Hinton et coll., 2012), les systèmes incrémentaux adoptant majoritairement deux approches : elles peuvent faire l'objet d'un apprentissage « artisanal », plus ou moins supervisé ; ou encore suivre des méthodes dites par renforcement (Lemon et Pietquin, 2007 ; Sutton et Barto, 1998 ; Young, Gašić, Thomson, et Williams, 2013). La seconde approche relève de test de résultats dans des conditions « de vie réelle » par des approches en temps réel, basés sur des méthodes stochastiques (Levin et Pieraccini, 1997). La première approche nécessite une quantité importante de données en entrée, avec un coût important de prétraitement, ou de manière alternative, travailler sur un nombre important de données simulées (Pietquin et Hastie, 2013 ; Schatzmann, Weilhammer, Stuttle, et Young, 2006), posant les limites actuelles des systèmes. Parmi tous ces systèmes, nous voyons tout de même que depuis les années 2000, ce sont les apprentissages par renforcement qui sont majoritairement discutés (Levin et Pieraccini, 1997 ; Singh, Barnett, et Singh, 1997). Il s'agit de méthodes qui s'appuient principalement sur des principes de psychologie comportementale (Skinner, 1992) en appliquant soit des récompenses (renforcement positif), soit des sanctions (renforcement négatif) en fonction du comportement du système. Le système calcule ainsi les coûts des récompenses maximales qu'il peut obtenir en adoptant un certain comportement dans un état courant, qui est récompensé si l'action effectuée est celle qui va dans le sens du modèle attendu. Parmi ce type de méthodes (Williams et Young, 2007 ; Gašić et coll., 2010 ; Young et coll., 2013) qui sont aussi des points d'entrée de l'intelligence artificielle, des start-ups tels que VocalIQ semblent donner des résultats qui attirent l'attention des GAFAMI, notamment Apple.

Cette problématique de dialogue est alors associée à une dimension évaluative (Dybkjær, Bernsen, et Minker, 2004). Si nous considérons par exemple l'apprentissage par renforcement, les problèmes consisteraient à définir les critères d'attribution des récompenses et des sanctions. En effet, si le comportement est « autoalimenté » comme dans certains systèmes entièrement automatisés en s'appuyant sur des méthodes de régression (El Asri, Lemonnier, Laroche, Pietquin, et Khouzaimi, 2014 ; Walker, Kamm, et Litman, 2000), un certain état peut définir le coût à associer. Or, si cet état est par exemple un arrêt d'input de la part de l'humain, les incertitudes persistent sur les raisons de cet état. En effet, si l'humain a eu une réponse à sa demande qui le satisfait et il ne poursuit pas son interaction, le système devrait théoriquement avoir un coût de récompense. En revanche, si cette non-réponse est uniquement liée au fait que le sujet abandonne toute tentative, car il n'arrive pas à obtenir le résultat escompté, le système est censé recevoir une sanction. Or, du point de vue de ce système, les deux états sont identiques. Ainsi, comme toute autre discipline basée sur corpus, les méthodes d'évaluation subjective sont également proposées, soit a posteriori par les

utilisateurs du système (Walker, Litman, Kamm, et Abella, 1997), soit par des annotateurs experts (Evanini et coll., 2008).

Enfin la terminologie « système de dialogue » commence également à émerger pour désigner des systèmes d'analyse de conversations entre humains (Barlier 2015). Ainsi, nous voyons apparaître des systèmes d'écoute tels que le Mindmeld (Expect Labs), une application proposant du contenu contextuel en se fondant sur des conversations téléphoniques.

Ainsi, l'interaction communicationnelle dans son ensemble ne semble pas rentrer dans le paradigme de ces réflexions. En revanche, la continuité interactionnelle est, quant à elle, omniprésente. Cette dernière serait dotée d'une dynamique du dialogue d'autant plus marquée dans le dialogue spontané (Luzzati, 1995, 2007). Mais, si cette dynamique peut être observée sur un objet résultant d'un échange qui s'ajuste au fil du temps par divers marqueurs significatifs, ne peut-elle pas être considérée comme un élément qui sous-tend le dialogue ?

### 6.3.2. Vers une technologie hybride de dialogue évolutif et incrémental

La première tentative de prototypage du système RAP reconnaissant la glu est au coeur des réflexions pour le développement d'un futur système de dialogue hybride, incrémental et évolutif : SASI (*Socio-Affective Speech Interaction*). Une tentative de schématisation est représentée Figure 65.

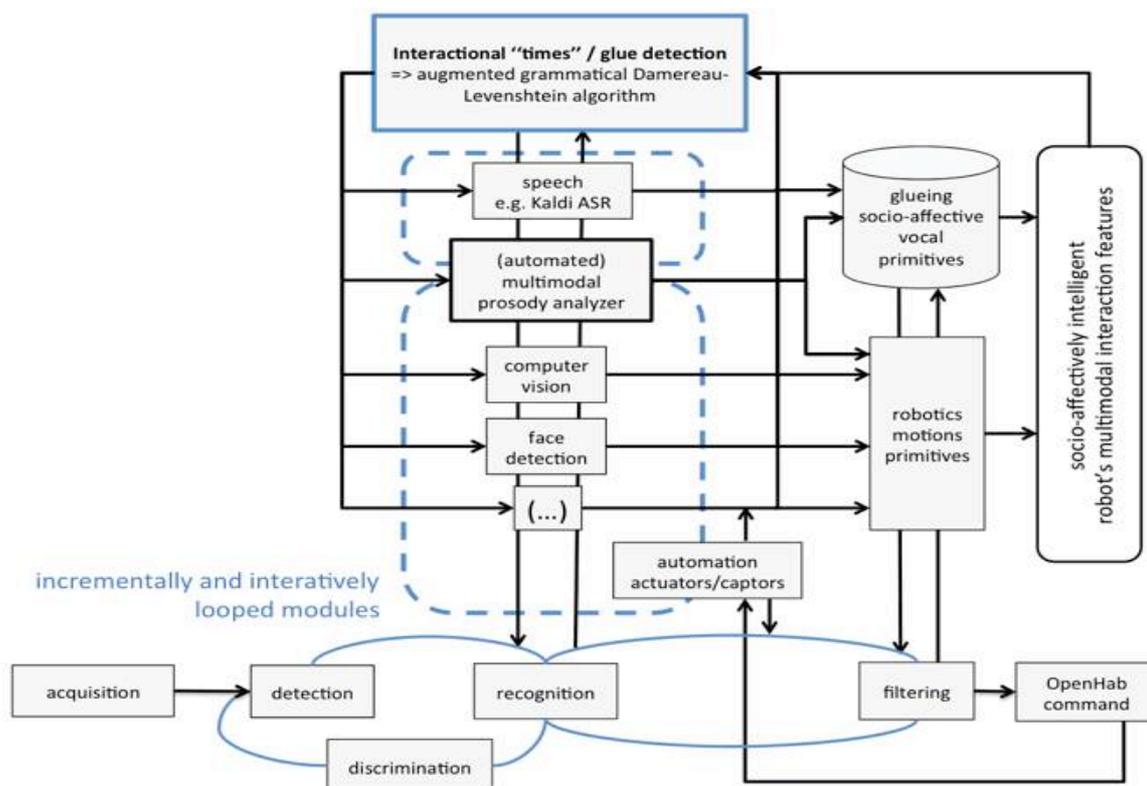


Figure 65 : Tentative de schématisation du système SASI

Ce système introduit trois enjeux. Il propose des réflexions transdisciplinaires autour de nouveaux paradigmes émergents, où les méthodologies de l'apprentissage machine pourraient être revisités dans de nouvelles perspectives, moins séquentielles et plus holistiques, tout en restant temporellement ancrées.

- système hybride : en effet, la complexité de la technologie robotique amène à avoir plusieurs sources d'analyses, qui sont à l'heure actuelle considérées comme des modules séparés et complémentaires. Cette multimodalité « analysée » semble être une évidence dans

le cadre d'une interaction incarnée (ou *embodied*) comme en robotique. En revanche, elle ne reste qu'une augmentation d'interface comme dans le cadre du Smart Home, où les moyens d'interactions tactiles ou haptiques font l'objet de ce qui est appelé le dialogue multimodal (p. ex. Bellik, 1995 ; Bui, 2006 ; Oviatt, 2003). Ces problèmes visent alors plutôt la coordination des entrées/sorties du système pour la gestion de ces interfaces, par la fusion ou la fission d'une modalité par rapport à une autre. Ceci émane aussi de l'idée que chaque canal de la multimodalité « s'additionne » pour enrichir l'information communiquée. Or les systèmes complexes tels que les systèmes multi-agents (cf. p. ex. chapitre 1- section 1.2), ont quant à eux abordé de front la problématique du calcul holistique émergent d'un point de vue mathématique. Par exemple, dans le modèle cognitif du traitement interactionnel (C-Clone) proposé par Aubergé en 2001 : « l'architecture est décrite en modules autonomes dans leur morphologie intrinsèque, mais ils sont totalement dépendants et inscrits dans un même processus dynamique d'encapsulation de temps cognitifs de la relation ». Reprenant ce type de structuration, nous proposons ici une brique de reconnaissance où une modalité donnée n'est le reflet que d'un aspect d'un tout, qui aura un effet global portant une fonction communicative. Si nous supposons alors que le « tout » obtenu est un état de glu socio-affective comme résultante de chaque interaction de sous-systèmes, il faudrait pouvoir définir quelle influence il a sur chaque sous-système pour modifier l'ensemble. Il s'agirait ainsi de trouver non plus les opérandes, mais les « opérateurs » qui aboutissent à la production des opérandes qui ont notamment été étiquetés dans le corpus EEE. Ainsi, la proposition d'usage détournée de l'algorithme de Damereau-Levenshtein qui associe un score unique de glu à la commande est ici limitée. Elle pourrait néanmoins dans de futurs développements être le point de convergence entre les différentes briques de traitement associées aux différentes natures de signaux.

- incrémental : cette dimension remet en question le traitement linéaire et séquentiel, fortement rattaché à la dimension hybride du système. En effet, l'information permettant de refléter le niveau de glu est transversale à plusieurs modalités qui peuvent expliquer l'état du système. Pour la parole spontanée, la dimension incrémentale est une piste exploratoire qui n'est pas négligeable puisqu'elle ne serait pas entièrement prédictible, mais construite au fur et à mesure, au plus près des données.

- dynamique et évolutif : puisque la glu est la résultante d'un processus dynamique, son état est en perpétuel changement. De plus, la complexité de l'humain étant par définition imprévisible, l'évolution prendra des directions variées au fur et à mesure du dialogue. De fait, la compréhension de l'évolution de la glu sur un plus long terme est un enjeu important pour la modélisation de la mémoire du système. En effet, les observations de cette thèse laissent à penser qu'il existerait peut-être des paliers dans les différents états de la glu, qui, quand ils sont franchis, ne donnent pas forcément lieu à des retours à un état antérieur. Les hypothèses d'apparition de ces paliers sont multiples : la fréquence d'apparition d'un pattern, un point de rupture brutale, une question de durée, des aller-retours entre des états socio-affectifs spécifiques, etc. Ainsi, si ces mécanismes peuvent trouver des pistes de réflexion, la nature de l'évolution ne sera certainement pas prédictible, mais le changement ou la nécessité de création d'un nouvel état de glu pourrait être modélisé (p. ex. par la création d'un nouveau modèle de langage ou la transformation d'un état existant). Le système pourrait alors développer une dimension adaptative.



## Chapitre 7. Discussions et Perspectives

Pour ce travail, nous avons recueilli de nombreuses informations et données, de contenus et de formes multiples, dont toutes n'ont pas été utilisables ou nécessaires pour les buts de modélisation et de prototypage applicatif que nous nous étions fixés. Ces données encore en friche représentent cependant certainement des ressources exploitables pour des travaux continuatifs à court et moyen terme. C'est pourquoi nous en décrivons ci-après les caractéristiques pour référencement.

### 7.1. Pistes d'exploitation de données non traitées

#### *7.1.1. Entretiens : caractérisation de l'isolement et « entraînement » communicationnel*

Les entretiens de recrutement, de débriefing ou d'auto-annotation, ayant eu lieu avant et après chaque expérience, ont duré, entre une heure à plus de 5h chacun. La prolongation de la durée des entretiens était notamment liée à des cas de fortes fragilisations associées à :

- la réminiscence de souvenirs nostalgiques ou affectivement marqués;
- des interventions de professionnels socio-médicaux ou médicaux interrompant l'entretien;
- des visites (de famille, amis, et/ou de voisins), mais qui laissaient les sujets dans un état fragile, donc nécessitant de les rassurer pour reprendre un contexte d'entretien stable. Elle est également liée à la fluidité de l'entretien elle-même, du fait de sujets de caractéristiques communicatives hésitantes ou saccadées.

Sachant que leur transcription est extrêmement chronophage, l'exploitation de ce matériel est très incomplète, mais compose un enjeu important pour affiner la caractérisation de l'isolement. En effet, au cours de ces entretiens, nous avons observé que le robot a un potentiel « pouvoir d'entraînement à communiquer ». Cette idée revient également dans certaines étiquettes telles que « *en fait j'étais un peu tenté de lui parler à lui inconsciemment, mais en même temps je me disais c'est une chose* ». Puis, la parole des sujets évolue de plus en plus vers des étiquettes de type « en fait là j'arrivais à lui parler comme on parle avec une personne ». Les personnes isolées en particulier rapportaient ainsi « *j'arrive quand même à discuter, mais ça je ne m'en rendais pas compte* », « *je réponds quand même – au fond ce petit truc ça vous oblige à parler* ». De plus, les patterns d'interaction, comme ont pu le montrer les figures d'évolution temporelle (cf. section 5.3), permettent de faire presque ressortir des situations d'isolement et des niveaux de fragilisation qui motivent le développement de tels outils. Il a été même montré un franc enthousiasme des ergothérapeutes et des acteurs de gérontechnologie<sup>107</sup>. Or, cet engouement est aussi à freiner, car les observations ne peuvent avoir aucune valeur de diagnostic d'un quelconque isolement ni avoir un but thérapeutique. En revanche, nous voyons que la problématique principale porte bel et bien sur ce sentiment d'isolement qui impacte notre manière d'interagir et qu'il y a une tendance de changement de l'usage du robot. La technologie d'abord présentée comme une aide « physique » par l'apport domotique, est transformée à travers son rôle et son interaction secondaire. Elle devient « *une présence* » qui cultive l'envie et les compétences relationnelles de la personne. De fait, il nous semble important de commencer à intégrer des acteurs du domaine de la santé, mais toujours

---

<sup>107</sup> Des retours notables au cours de la présentation des travaux dans l'édition 2014 et 2015 du Salon de robotique Innorobo et à la conférence ISG 2016.

dans cette approche collaborative de co-construction. Une perspective plutôt de bien-être, c'est-à-dire non médicale, serait alors conservée, car l'objet robotique ne reste qu'une aide potentielle, mais dont le développement sociétal commence à être trop important pour l'ignorer.

### *7.1.2. Expériences annexes : améliorations de protocole*

D'autre part, l'expérimentation, en guise de prétest, a été réalisée avec de jeunes adultes gérontologues (infirmiers, chefs d'entreprise pour technologies pour seniors, aides à domicile) à travers un scénario adapté pour le prétexte, mais fonctionnant exactement sur le même principe (apporter des objets à placer dans l'appartement). La rythmicité des échanges est très clairement beaucoup plus rapide, et la technophilie semble avoir de fortes influences sur les comportements, mais dans l'ensemble, le même type d'évolution paraphrasée semble observable à la manière des personnes non isolées du corpus EEE. Ce qui est en revanche notable est l'atteinte rapide d'un certain palier du fait de la limite du rôle du robot, car les jeunes adultes ne présentent aucune fragilisation. La comparaison intergénérationnelle et l'exploration d'autres rôles sociaux restent donc à tester pour le robot.

Par ailleurs, la validation perceptive n'a malheureusement pas pu être réalisée dans le cadre de cette thèse donc les valeurs d'auto-annotation récupérées n'ont aucune certitude d'être robustes et partagées de manière externe à cette étude. Il en existe d'ailleurs un niveau d'imprécision, difficile à contrôler, car l'auto-annotation reste basée sur la subjectivité du sujet, sans mesure de la croyance perceptive partageable par d'autres, et bien que contrôlée dans la mesure du possible. De plus, aucune validation avant l'expérience n'a été réalisée sur les stimuli de mouvements d'Emox, à défaut d'étiquettes et des caractéristiques spécifiques liées aux possibilités de mouvance du robot. Elle pourrait néanmoins être envisagée à partir des données directement recueillies dans le corpus. L'enjeu de ces futurs tests perceptifs, surtout pour les productions des sujets, n'est cependant pas orienté uniquement sur l'appariement d'un stimulus avec l'étiquette, mais il est également lié à la progression de la dynamique prosodique. En effet, même avec une étiquette fonctionnelle semblable, la glu s'illustre par une « prosodie dialogique » plus générale, qui change subtilement la dynamique d'un événement effectué, y compris celui effectué avant et après. Cette prosodie serait la même que celle observée au niveau local des composants interactionnels, qui s'étendrait ensuite sur de plus grandes échelles d'interactions. Elle passerait ainsi d'une prosodie « pure » et minimale, à une « macro-prosodie » (portant sur des blocs interactionnels) jusqu'à celle du dialogue qui porte les traces de l'évolution relationnelle entre individus. De plus, dans ce dialogue, chaque bloc devient un événement, à l'image par exemple du modèle granulaire superpositionnel proposé initialement dans la thèse de (Aubergé, 1991). Les premières traces d'évaluation de ce modèle sont données par Audibert 2006, Loyau 2007 et Vanpé 2012. Dans cette proposition, la prosodie du signal d'un acte de parole est modélisée par superposition multiparamétrique d'entités encapsulées :

- au niveau de granularité du dialogue, chaque acte est un événement. Dans notre approche (qui réactualise cette proposition en mettant en avant la dynamique interactionnelle), l'événement est la valeur de la glu qui contient la trace émergente de l'interaction. La prosodie du dialogue représente alors la courbe d'évolution de ces événements ;
- au niveau de granularité de la relation interpersonnelle, chaque dialogue est un événement et la prosodie de la relation est la courbe de l'évolution sur ces événements dialogues.

Le même principe de superposition s'applique sur ces niveaux, chaque niveau se décrivant sur plusieurs modalités, et s'organisant morphologiquement selon les différents temps cognitifs. Ce changement est principalement visible dans la qualité de voix, avec par

exemple une *breathiness* (marqueur du *care* selon Campbell ou Wichman) qui augmente au fur et à mesure de l'expérience, aussi bien à l'échelle du bruit de bouche que sur le dialogue. Là encore, ce sont les auto-annotations qui motivent cette réflexion. Sur une même étiquette (c.-à-d. définie par un même indice), les sujets peuvent faire des retours en arrière, mais au moment où ils décrivent l'étiquette, ils nous disent « c'est comme avant (sous-entendu la même fonction, la même intention, le même ton), mais il est un peu différent ». Ainsi, à chacune de ces mentions, l'étiquette d'auto-annotation a été associée au même indice, mais avec des apostrophes, en fonction de l'évolution subtile de cette commande. De fait, un affinement des analyses prosodiques devrait être effectué en parallèle. Déterminer les profils prosodiques est une chose, mais des changements de dynamiques d'un même profil pourraient suggérer les dimensions sur lesquelles intervient le changement de glu. Plus encore, ces dimensions seraient certainement transposables sinon analogues sur d'autres modalités (p. ex. une dynamique d'amplitude vocale qui correspond à une dynamique d'amplitude gestuelle ou encore associée à la direction et/ou le changement de la fréquence du regard) justifiant alors la définition d'une « prosodie multimodale ». La perception humaine pourrait alors être testée sur ces changements de dynamique (qui peuvent s'observer pour différentes fonctions) en choisissant des stimuli à différents moments de l'interaction globale.

L'une des raisons expliquant les limites descriptives de l'analyse prosodique des sujets est notamment liée aux spécificités physiologiques induites par le vieillissement. Quelles que soient les personnes, une *breathiness* et une *creakiness* plus ou moins importantes sont notables, et les variations de *jitters* (les microperturbations locales de la fréquence fondamentale du signal de parole) et de *shimmers* (les microperturbations locales de l'amplitude du signal de parole), notamment utilisées comme descripteurs de la qualité de voix, deviennent rapidement difficiles à interpréter. De même, ce vieillissement a aussi eu un impact sur l'adaptation du protocole d'auto-annotation de certains sujets avec une importante déficience visuelle, auditive ou physique (nécessitant des changements de posture, l'alitement etc.) qui a amené à l'abandon du protocole pour certains. De fait, cela pointe une autre limite de l'étude qui est celle du manque de sujets-contrôles, dont les profils pourraient être étendus et systématisés. La difficulté dans ce type de protocole réside dans le fait de savoir comment mieux intégrer le paradoxe l'observateur de Labov en continuant à détourner l'attention. En effet, ne serait-ce que pour la déficience acoustique ou visuelle, elle est parfois restée inaperçue même du côté des assistants à domicile, non pas du fait d'un manque de professionnalisme, mais parce que c'est bien la personne âgée elle-même qui décide de montrer ou non certaines faiblesses.

## 7.2. Acceptance technologique et éthique

L'éthique est une dimension intrinsèque et fondamentale, que ce soit dans la manière :

- de mener l'expérience,
- dont les technologies construites par la méthode 3P agiront relationnellement sur l'humain,
- dont ces technologies bousculeront le rôle ou l'emploi de l'humain.

L'expérimentation, qu'il s'agisse du format du protocole ou des stimuli interactionnels directs du robot, amène à faire évoluer et à modifier l'état des sujets.

Nous avons cru remarquer que l'expérimentation a pu confronter les personnes « non gluantes » devant certains faits de leur situation (p. ex. isolement latent lié à une situation complexe d'aidant, la confrontation à une fragilisation physique handicapante, un très fort sentiment de mal-être etc.). Cela aurait pu également contribuer à l'effet de rejet de la technologie. Les effets de l'expérience-même sont donc à discuter continuellement avec l'ensemble des partenaires. Par exemple, S19 rapportait sur l'expérience, dont il disait avoir apprécié la participation, qu'elle l'avait peut-être indirectement « chamboulé » ou « désorienté », qu'il n'en arrivait pas à spécifier exactement les raisons mais a constaté que ses proches lui disaient aussi « qu'il avait changé, qu'il n'était pas comme d'habitude ». C'est là une dimension éthique importante à réfléchir, qui pose la question de la manière d'intégrer cette technologie et comment accompagner ces changements à travers des protocoles. Dans ces situations, la visée directe de la personne pose peut-être déjà un problème du fait du risque de confrontation de la personne à sa propre situation. Les alternatives seraient peut-être, comme le suggérerait certaines aides à domicile, de proposer cette technologie « indirectement », de manière à ce qu'elle serve en premier lieu l'aide à domicile dans ses tâches. Le contact progressif à cet objet serait alors le point d'entrée pour un « entraînement » des mécanismes enroués, pour tendre vers une envie de communication qui soit consécutivement plus fluide.

Une hypothèse qui est également motivée par la nature des auto-annotations est que la direction finale de la courbe de glu (globalement ascendante ou descendante) pourrait illustrer un aspect relationnel important de la personne. En effet, dans le cadre expérimental précis où les complices reviennent sur les derniers temps de l'expérience, une courbe ascendante illustre une continuité d'évolution positive de la glu, alors que chez certains cette courbe est descendante sur la fin de l'expérience (avant débriefing). En effet, cette différence dans l'auto-annotation illustre d'une certaine façon le focus et le degré d'attention portée sur le robot dans un contexte interactionnel où le sujet est en présence d'autres personnes. En effet, nous remarquons ce contraste dans les étiquettes du type « *c'est certainement une fierté très mal placée, mais j'étais contente de montrer ce que j'avais découvert avec lui* », « *on était deux il me suivait, maintenant que vous (E2) êtes là il veut plus me suivre* » produites par les personnes dont le mouvement final est ascendant. Au contraire, les personnes qui ont « positivement glué », mais avec une courbe finale descendante tiennent des discours tels que « *je m'en fous, je suis de marbre, je m'en contre fous c'est vous (E2) mon interlocutrice* », « *je le félicite, sous-entendu que les personnes qui ont travaillé dessus et bien je reconnais que c'est bien – en lui parlant à lui je parle à vous – le compliment c'est pas tant pour lui c'est pour vous* ». Sur la perception d'Emox cela se traduisait sur la différence que les personnes évoquaient. Nous distinguons alors les premières personnes qui restaient sur une dimension « *ça tient compagnie, c'est une présence* », et celles qui nuancent en disant que « *c'est bien, on se sent moins seul, mais ça reste un robot* ». Ainsi, de la même manière qu'il existerait une subtilité de l'effet du sentiment d'isolement, le besoin ressenti du robot est lui aussi à garder à l'esprit dans la considération éthique que ce système peut avoir sur la personne. La décision de la prise de risque doit donc être envisagée de manière collective. De plus, la réflexion vers

## **Discussions et Perspectives**

---

de nouvelles étapes méthodologiques permettant l'intégration de population externe à relativement assez grande échelle devrait être envisagée. Cette intégration ne devra pas être une simple demande d'avis sur une prise de décision, mais devrait faire l'objet d'un protocole participatif avec des stratégies de détournement de l'attention qui puissent mettre collectivement ces personnes externes de la société en méta-tâche, afin de les confronter par la suite à des choix réellement éclairés pour prendre des décisions à l'échelle de la société.

### **7.3. Visualisation de données : comparer courbe « calculée » et courbe « mesurée » de glu**

À l'état actuel, l'une des limites de l'utilisation de l'algorithme d'édition de la glu relève en partie d'un problème de visualisation. En effet, la section 5.3 montre qu'il est potentiellement possible de suivre la courbe d'évolution de la glu socio-affective en s'appuyant sur l'évolution des auto-annotations. Nous avons ainsi une « courbe mesurée » faisant état de la relation socio-affective entre le robot et le sujet. En parallèle, dans le contexte expérimental en magicien d'Oz proposé, la prédiction de cette construction relationnelle s'appuyait sur les hypothèses graduelles de primitives vocales du robot, qui étaient supposées générer cette glu. En revanche, maintenant que nous avons la mesure de l'effet réel de ces stimuli (mais aussi leur modalité/timing d'exécution interactionnelle), l'algorithme de distance d'édition est un moyen de pondérer l'effet que chaque changement interactionnel génère. Ainsi, en modifiant manuellement les coûts minimaux des opérands interactionnels, il serait possible de visualiser une « courbe calculée » de glu à partir des coûts définis dans l'algorithme. Les deux courbes seraient ainsi visualisables sur le même graphe temporel. La courbe calculée serait produite par modification du niveau de glu associé en partant de la forme de référence. L'ajustement des coûts se faisant manuellement, une actualisation en direct sur l'ensemble des stimuli concernés par une règle pourrait alors changer de valeur et donc la tendance d'évolution de cette courbe calculée visualisée. Ainsi, en alignant la courbe calculée et la courbe mesurée sur un même graphe, la courbe « mesurée » servirait de modèle à la « courbe calculée ». Cette dernière pourrait alors suivre au plus près l'évolution de la première. L'ajustement manuel des coûts sert ainsi à trouver le bon paramétrage pour réduire au maximum la distance en termes de « niveau de glu socio-affective » entre la commande de référence modifiée et la commande réellement produite. Ce procédé serait ainsi un moyen de régler empiriquement, par tâtonnement, ce que les décisions de la « boîte noire stochastique » donnent, mais avec pour différence la possibilité d'explicitier quel facteur a été plus influent qu'un autre puisque nous sommes à l'origine des modifications des règles expliquant la variation humaine du phénomène. La modélisation qui en suivrait serait alors empirico-déductive.

Une fois que la courbe calculée est correctement ajustée, nous obtiendrions une paramétrisation qui est une possible illustration de la modélisation théorique du poids des dimensions interactionnelles qui pourront alors faire éventuellement l'objet d'une analyse statistique pour déterminer si des règles de distribution mathématiquement prédictibles s'appliquent pour décrire la répartition des coûts trouvés. Enfin, pour vérifier la modélisation et les hypothèses théoriques associées, le système devra être automatisé. Cela consisterait à avoir une génération automatique des feedbacks du robot qui suivent les prédictions de la courbe calculée. Si la technologie est mise en échec et que la glu ne prend pas, le modèle et la théorie pourront alors être repris, générant ainsi les essais-échecs des boucles agiles permettant de comprendre globalement comment les sous-systèmes, du plus large système interactionnel, interagissent entre eux.

À l'heure actuelle, quelques travaux de visualisation<sup>108</sup> ont été effectués, dans la limite d'une représentation figée de la glu, prenant en compte des pondérations associées à une base de données, sans avoir la possibilité de modifier rapidement et simplement les coûts de l'algorithme de distance d'édition sur l'interface de visualisation. Idéalement, il faudrait donc pouvoir observer rapidement les changements des pondérations. De même, si les règles de la grammaire doivent être remises en question, elles aussi devraient pouvoir être modifiées

---

<sup>108</sup> Cf. mémoires de master IdL professionnel N. Lagier et C. Cottier (2015), mémoire de master IdL recherche N. Borel (2016).

## **Discussions et Perspectives**

---

facilement à travers cette interface de visualisation. Enfin, les états des itérations devraient être conservés en mémoire pour faciliter les aller-retours entre les différentes étapes des règles de pondération.

## 7.4. Pistes de réflexion : de l'étiquetage à la synchronie

Les éléments étiquetés, tant du côté du sujet, du robot, de l'appartenance que des magiciens, ouvrent ici des pistes de réflexion sur l'organisation des différents objets interactionnels relevés. Sur l'analyse, notamment celle associée à la synchronie, l'intérêt serait de pouvoir observer en parallèle certains sous-systèmes, par exemple :

**l'inter-synchronie magicien-technologie.** Nous avons une première dimension découlant de l'écologie expérimentale propre, à savoir la maîtrise des briques de technologie multiples que le magicien doit gérer afin d'appréhender un « nouveau corps » qui augmente l'humain-magicien. Elle s'apparenterait à la dimension d'appréhension/adaptation à cet autre corps, incluant des capacités d'anticipation pour pouvoir intégrer le système interactionnel global. Cette dimension de la « prise en main » technologique par l'humain sera une dimension d'autant plus importante dans le cadre de la robotique de téléprésence où ce contrôle est explicite dans la situation d'interaction. À partir de la notion d'*embodiment*, où s'arrêterait-il le soi et serait-il d'ailleurs possiblement prolongé à travers la relation que nous tissons avec autrui ?

**l'inter-synchronie synchronie entre magiciens à plusieurs niveaux.** Nous avons un premier niveau entre celui qui manipule « sons et mouvements scriptés » et celui qui est aux commandes de la manette. De plus, il serait intéressant de repérer les changements entre les magiciens. En effet, certains magiciens sont « acteurs » dans le contexte FtF avec le sujet, et laissent donc la main à d'autres magiciens quand ils doivent quitter la régie. Les perspectives d'exploration sont alors multiples : la signature personnelle (Chetouani, 2014) des magiciens ou des sujets permettrait-elle de dessiner des profils de « magicien gluant » vs « magicien moins gluant », en fonction des profils d'isolement de sujets ? Les aidants professionnels, les marionnettistes ou les animateurs seraient-ils alors de meilleurs magiciens ? De futurs protocoles pourraient alors soit inverser la cible d'induction du comportement sur les magiciens, soit idéalement induire en double-aveugle les comportements interactionnels. Les magiciens visés seraient ainsi ceux qui auront développé une expertise naturelle d'une intelligence relationnelle construite empiriquement sur leurs pratiques quotidiennes, et dont les manifestations spontanées seraient la cible de ces expériences.

**l'évolution de l'inter-synchronie des magiciens et des sujets.** Il s'agit là de la perspective principale d'analyse qui nous intéresse, puisqu'elle serait une mesure qui viendrait valider de manière externe l'état de la relation entre le robot et le sujet. Cela permettrait ainsi de vérifier une potentielle pondération qui caractériserait l'état de la glu socio-affective dans le cadre d'une modélisation de l'évolution de la glu. D'ailleurs, à travers ces approches, y aurait-il une évolution des marqueurs de la signature personnelle inhérente au magicien, inhérente au robot, un effacement de l'un au profit de l'autre ou l'apparition d'artefacts, une convergence des entités ? Si la possibilité d'efficacité de manipulation de la glu devient l'un des objectifs de perspectives, nous pouvons également penser à des protocoles intégrant des magiciens spécifiques, manifestant une « expertise » vernaculaire et empiriquement acquise dans ce travail relationnel, comme les aides à domicile, dont le corps de métier dépend fortement de leur capacité relationnelle dans la dimension du *care*.

## 7.5. Quelques réflexions sur le « bruit » du dialogue

Les « bruits » du dialogue pourraient être pertinents de reconnaître pour assurer la continuité interactionnelle dans le cadre d'un système de dialogue (p. ex. s'ils sont des marqueurs de l'état de la glu socio-affective entre les interactants). En revanche, ils pourraient aussi être des informations bruitant l'interaction (p. ex. une parole non-adressée au système) qui fait partie du flux acoustique au même titre que ce qui est fonctionnellement adressé. S'il est alors possible de trouver les caractéristiques de ce qui est « bruit » et « non bruit » pour le système RAP, une perspective intéressante du corpus EEE serait de se « rendre aveugle » aux données non-pertinentes pour l'interaction tout en utilisant celles qui peuvent l'être, mais qui ne relèvent pas d'ordres domotiques directs. En effet, toute la difficulté réside dans le fait que le contenu linguistique ne pourra pas désambiguïser l'interaction. Si l'utilisateur du produit final lit tout simplement le mode d'emploi à haute voix ou explique à un tiers comment son système fonctionne sans l'actionner, le lexique et les formes morphosyntaxiques seraient forcément identiques aux éléments adressés au robot. En revanche, un humain arrive à faire cette distinction, donc il existerait des indices, notamment prosodiques, qui sont porteurs de cette caractéristique langagière. Ils permettraient de repérer la cible à laquelle s'adresse une production vocale, notamment si un acte est explicitement attendu par l'émission de ce signal oral. À l'image de l'état de l'art, la question de l'attention, notamment en HRI, (cf. section 1.4) a principalement été étudiée à travers le regard, mais beaucoup moins sur les indices acoustiques. Or, quand l'entité a des caractéristiques très marquées comme le robot Enox (petite taille, voix très aigüe, tête ronde, blanc... etc.), des discriminants semblent exister et apparaître dans la voix (p. ex. augmentation généralisée du pitch). Plus encore, ces caractéristiques interactionnelles (que ce soit le vocal ou le regard) évoluent au fil du temps, mais surtout en fonction du niveau de construction de la glu. Ainsi, la section 4.2.6 a montré que la cible du regard tend à évoluer, tantôt vers la liste des commandes, tantôt vers l'actionneur domotique ou le robot. En parallèle, nous avons une évolution de la qualité de voix des personnes, qui devient de plus en plus *breathy* dans les interactions avec le robot. Comment l'attention se retranscrit-elle dans ces phénomènes, sachant notamment que les commandes sont toujours adressées au système robotique, mais avec des variations des indices attentionnels ?

En supposant que nous serions dans la capacité de discriminer la parole adressée au robot à celle qui ne l'est pas, un autre verrou, serait de faire la part de ce qui est adressé comme commande ou comme interaction prétexte, sans attente d'exécution domotique. À l'heure actuelle, la robotique ou les technologies d'assistance vocale utilisent des *triggers* ou « une gâchette » de commande, soit par des mots clés (p. ex. très souvent le nom de l'IoT, comme le robot pot de fleur BILLY-BILLY), soit par un actionneur physique (p. ex. le bouton d'activation de Siri, Cortana, Alexa d'Amazon Echo etc.). Or, dans le cadre de l'annotation, le problème serait de savoir sur quelles parties du discours placer ces étiquettes, et surtout à quel niveau de granularité. En effet, dans le corpus, nous avons un label NA (« non adressé ») qui reste une décision à prendre par rapport aux stratégies à adopter en fonction du modèle de *machine learning* qui découlerait des deux problèmes cités précédemment. En effet, avec un but de filtrage pur, une étiquette NA incluse dans le tiers NonCmd serait suffisante pour former un corpus d'apprentissage très global du « bruit », en opposition aux « commandes ». Ceci permettrait notamment de différencier les productions telles que les micro-expressions vocales ainsi que des formes plus complexes illustrant des formes d'attachement, d'intimité, ou de *care*. En effet, ces éléments qui gravitent autour des commandes ou qui apparaissent de manière plus isolée dans l'interaction pourraient marquer de potentiels points de rupture dans l'évolution de la glu socio-affective. En revanche, si l'on veut affiner cette problématique, ce label NA pourrait aussi être associé au même niveau de granularité que Cmd, NonCmd, NwCmd et avoir plus de précision quant à la nature des

## **Discussions et Perspectives**

---

énoncés plus indirects, mais produits aussi en présence du robot. Une possibilité présentée serait alors d'associer un tiers NA, avec des labels décrivant la cible du message (locuteur spécifique ou toute personne du contexte) ou le sujet du message (notamment, s'ils portent sur certains aspects du robot ou de l'appartement). En effet, dans une approche pragmatique, un discours pouvant faire l'objet d'un acte indirect, un message non-ciblé au robot peut potentiellement lui être adressé, si bien entendu, la forme de la relation a évolué de manière à ce que la perception du sujet change pour considérer le robot comme un « autre » qui entend ce discours détourné. Une seconde motivation serait de pouvoir filtrer rapidement les discours des sujets échangés autour de l'expérience, afin d'en faciliter les analyses qualitatives.

### 7.6. Pistes d'implémentation du prototype SASI

Le prototype de reconnaissance vocale sensible à la glu est à l'heure actuelle une version non-interfacée et non encore physiquement intégrée à l'appartement Domus contenant Emox. Ainsi, l'une des perspectives immédiates serait de pouvoir implémenter ce prototype au système, comme le pont créé avec EmOz, afin de pouvoir tester en boucles itératives les changements des modèles. Afin d'augmenter la rapidité de cette implémentation, une idée possible serait d'interfacer ce système au système de visualisation. Ainsi, nous pourrions garder des versions de grammaires et patterns de pondération basée sur le travail de l'interface de visualisation, en proposant la génération automatique des fichiers supports pour chacune de ces configurations. Ceci permettrait véritablement de rentrer dans le processus de boucle agile avec une fabrication véritablement parallèle à l'évaluation.

L'avantage du schéma expérimental général adopté dans cette étude est que le système automatisé peut être testé au fur et à mesure, dans exactement les mêmes conditions que la version en magicien d'Oz. Il permettrait ainsi une reproductibilité de contexte interactionnel non-négligeable pour l'évaluation du système.

De plus, un aspect qui a été réfléchi, mais non mis en place dans cette thèse est l'exploitation de l'équipex Amigual4Home et du FabLab MSTIC. Ces deux plateformes sont complémentaires à Domus-LIG. En effet, dans la continuité de l'évaluation du processus de co-fabrication, les perspectives étaient : d'une part, d'effectuer des expérimentations avec le robot-appartement équipé du prototype de reconnaissance automatique de parole gluée (soit une brique de SASI) ; d'autre part, il s'agissait de sortir Emox de l'appartement pour rentrer dans l'écologie quotidienne de la personne âgée. En effet, à la vue du nombre de non-participants et du profil d'isolement de ces personnes qui ne sortent pas de chez elles, l'usage de mini-kits tels que le propose Amigual4Home, est une possibilité technique extrêmement utile. En revanche, n'ayant plus l'infrastructure d'observation, ni la raison d'usage de l'appartement domotisé, il est difficile de trouver un nouveau rôle au robot qui puisse incarner une utilité qui ne soit pas exécutable par un humain et qui réponde à un réel besoin. Ainsi, des pistes telles que le contrôle de la malnutrition ou de la déshydratation ont été, par exemple, abordées<sup>109</sup>. L'intégration d'un objet connecté permettrait alors de faire l'acquisition de données interactionnelles, telles que des fontaines intelligentes ou tout autre IoT utile. Ce matériel serait doté d'outils de capture de données et des effecteurs dont le traitement et le contrôle passerait par le robot.

De même, il existe aussi des pistes d'amélioration pour la plateforme Domus-LIG et ses outils d'analyses méthodologiques, notamment celles qui permettraient d'augmenter la finesse de l'analyse interactionnelle d'une part, et d'autre part, des solutions qui permettraient d'automatiser l'étiquetage et le post-traitement des données acquises dans Domus. Cela permettrait de réduire la dimension chronophage de l'expertise et du traitement manuel des données, mais dont les traitements fins sont nécessaires et irréductibles dans la méthodologie adoptée.

---

<sup>109</sup> La déshydratation semble être un problème récurrent ayant un impact particulier sur la prise de médicaments qui amènent une forme de toxicité en situation de déshydratation d'après les échanges avec les partenaires médico-sociaux.



## Conclusion générale

Une hypothèse majeure sous-tendant ce travail est le processus de co-construction d'une glu-socio-affective qui instancierait l'évolution de la relation inter-personnelle. Cette idée construite antérieurement à cette thèse s'est progressivement formalisée par l'apport des travaux successifs depuis (Morlec, Bailly, et Aubergé, 2001; Aubergé, Audibert, & et Rilliard, 2006; Audibert, 2008; Loyau et Aubergé, 2006; Vanpé et Aubergé, 2012; Signorello, Aubergé, Vanpé, Granjon, et Audibert, 2010; Biasi, Aubergé, et Granjon, 2012, etc.). Dans les premiers chapitres états de l'art, nous avons présenté ces travaux et d'autres propositions théoriques, de différents domaines, qui étayaient la viabilité de cette hypothèse et qui permettent de situer notre travail dans une large perspective inter-disciplinaire.

Les processus cognitifs et empathiques communicationnels engendrant cette glu sont supposés intégrer une dimension altruiste, qui serait un axe en soi orthogonale à la dimension de dominance (Aubergé 2015, Aubergé 2017)<sup>110</sup>, communément utilisée pour la modélisation des émotions supposées sous-tendre les interactions humaines. Une seconde hypothèse proposait que la glu puisse être modifiée à travers des primitives langagières, représentées par des formes de « prosodie socio-affective pure », dont le contrôle progressif a un effet graduel sur la modification de la glu permettant d'assurer la continuité interactionnelle (Audibert, 2008; Vanpé, 2011; Sasa, 2012). Ces aspects ont ainsi contribué à déterminer les notions et les facteurs d'influence afin d'établir les contraintes méthodologiques dans l'approche globale d'observation et de recueil de données que nous souhaitons suivre. Dans le contexte sociétal actuel, ces choix se justifient par l'induction d'interactions spontanées et écologiques de personnes âgées en isolement relationnel avec un robot dont le rôle social attribué, indispensable à l'élaboration de la glu, est celui de majordome d'un habitat intelligent. Le seul véhicule anthropomorphe de cet instrument robotique, simulant l'activation des actionneurs domotiques, réside dans la production contrôlée des primitives vocales. Ces dernières sont produites sous forme de feedbacks répondant aux commandes vocales des sujets âgés. Leur isolement est considéré comme une situation d'observation contrastive d'un état de glu dégradé, par la détérioration des processus communicationnels permettant son maintien. Les approches scientifiques mises en place pour l'expérimentation sont dès lors héritées de pratiques antérieures d'induction de comportements humains par magicien d'Oz, ainsi que des apports des méthodes living-lab. Ces approches sont présentées dans une méthodologie expérimentale globale au sein de la plateforme Domus-LIG, et elles sont construites sur des boucles agiles et itératives de co-construction 3P (cf. *workshop: Emotions, Attitudes, Illocutions, Modality* – présentation GSCP Aubergé, 2012).

De fait, une part importante de cette étude pluridisciplinaire est axée sur la mise en place méthodologique, l'expertise et l'étiquetage fin des données en vue d'une modélisation de l'interaction homme-robot, avec toute une réflexion sur ce que cette technologie implique pour la communication humaine. Cette réflexion s'appuie sur une forte hypothèse de manipulation d'une « glu socio-affective » dont l'existence a pu être avérée grâce aux expériences menées et analysées dans cette thèse. Les expérimentations sont outillées à travers la plateforme de magicien d'Oz EmOz ainsi qu'un scénario expérimental complexe. Suite à un long processus contrôlé de recrutement, passant par une centaine de potentiels sujets, 24 sujets se sont déplacés pour l'expérience avec le robot, 20 sujets ont pu réaliser

---

<sup>110</sup> Exemples de références vulgarisées consultables sur: <http://www.18h39.fr/articles/veronique-auberge-en-quete-d-une-robotique-humaine.html> (septembre 2015) et <https://lejournal.cnrs.fr/articles/comment-les-humains-sattachent-aux-robots> (mars 2017).

l'auto-annotation et les productions de 16 d'entre eux ont été transversalement analysées de manière systématique.

Les observations du corpus EEE ainsi construit illustre des progressions systématiques des productions des sujets, qui semblent suivre la gradualité dans les primitives langagières testés sur le robot. Les stimuli vocaux en particulier, qui représentent l'unique anthropomorphisme visé n'invalident pas l'hypothèse de manipulation de la glu socio-affective qui se modifierait à travers le renforcement prosodique qui est proposé. Parmi les changements des productions du sujet, nous observons principalement l'apparition de paraphrases des commandes vocales émises par les sujets, portant sur des variations morphosyntaxiques globalement systématiques. Cette modification s'accompagnerait d'une augmentation généralisée du pitch vocal, d'une *breathiness* de plus en plus marquée, de focus prosodiques ponctués de micro-expressions vocales dont l'organisation et la nature changent au cours de l'expérience. Parmi les productions, nous notons également des marques de *care*, d'attention, de politesse, des compliments ainsi que des phénomènes illustratifs de *chameleon effects* tels que les écholalies, les répétitions ou les imitations des feedbacks du robot par les sujets. À partir d'un certain stade de glu atteint, nous observons également l'apparition de phénomènes de récupération d'erreurs remplaçant les attitudes d'abandon ou de rejet de système, notables dans les stades antérieurs de glu. Ainsi, l'intelligence socio-affective du robot contribuerait à renforcer l'intelligence ou la robustesse du système, en favorisant une continuité communicative.

Une seconde partie importante de l'étude porte sur la valeur informative des événements interactionnels ainsi que sur l'appréhension de leur évolution temporelle à travers les auto-annotations. Suite à cette étape chronophage nécessitant un alignement manuel, des « temps cognitifs » séparés ont pu être mis en évidence, permettant d'associer différents niveaux de *Feeling of Thinking* dont de nouvelles valeurs ont émergé. Ces différents temps interactionnels se chevauchent pour donner une vision globale des transformations de l'état de la glu socio-affective à travers: 1) des éléments liés à la perception d'Emox par le sujet, 2) la variation des intentions et des fonctions communicatives associées aux productions des sujets, 3) la survenue d'événements interactionnels remarquables et contextuels, contribuant à la modification 4) d'états plus généraux latents. Ces temps sont influencés et influencent eux-mêmes les effets des outils langagiers qui modifient l'état de la glu. Cet ensemble donne ainsi une idée de la rythmicité globale de l'interaction qui propose l'existence d'une « prosodie dialogique » dérivée des variations micro-prosodiques locales, dont le plus petit niveau de granularité est représenté par les primitives langagières testées. Il s'agit alors d'observer une courbe d'évolution mesurée de glu socio-affective.

Ensuite, un travail conjoint rassemblant de nombreuses contributions a permis de réfléchir sur la modélisation et le prototypage d'un système de reconnaissance vocale ainsi qu'un algorithme de pondération utilisant la distance d'édition des commandes. L'approche adoptée pour la reconnaissance de parole détourne les usages habituels de *WER* et des scores de confiance des systèmes courants, basés notamment sur un modèle de langage unique. Dans le cadre de cette thèse, nous avons proposé le principe de modèles de langage multiples définissant différents « niveaux de glu » de manière spécifique. Ces modèles, lancés parallèlement à chaque décodage d'énoncé, sont dédiés à reconnaître avant tout le niveau d'évolution de la glu. Ainsi les premières observations sur prototype, entraîné sur données réduites, montre qu'en s'appuyant sur la discrimination du « meilleur » modèle de langage, il serait possible de définir le niveau de glu socio-affective, sans passer par un modèle complexe qui décrirait spécifiquement et uniquement cette glu. Cette approche semble par ailleurs plus robuste qu'un modèle de langage unique et générique, à partir duquel il ne serait pas possible

de décrire l'évolution de la glu et qui engendre également plus de bruit dans la reconnaissance de la commande vocale des sujets.

Le second aspect de développement repose quant à lui sur la pondération des dimensions d'interaction. L'approche consiste en une adaptation détournée de la distance d'édition calculée par l'algorithme de Damereau-Levenshtein. Cette adaptation est inspirée d'un ancien algorithme basé sur l'écrit permettant une segmentation par blocs de l'énoncé d'entrée. Il permet ainsi de pondérer les opérandes et non plus les opérations de transformation pour calculer la distance et donc la différence entre un énoncé modèle (la commande vocale de référence) et l'énoncé produit (la commande vocale « gluée »). Cela reviendrait ainsi à pondérer les changements qui déterminent un niveau de glu à travers l'application d'une grammaire intermédiaire de coûts. Cet algorithme n'est pour le moment appliqué que sur la dimension morphosyntaxique des commandes, mais dès lors que l'interaction est multimodale, une pondération similaire pourrait être appliquée à l'ensemble des dimensions de différents canaux de modalité. Ces canaux dépendent de dimensions prosodiques, de proxémie, de regard, de gestualité, d'expressions faciales, de posture, de position ou encore de déplacement. Ainsi le score unique de « niveau de glu » obtenu à travers l'application de cet algorithme permettrait de faire le pont entre le prototype du système de reconnaissance de parole gluée, et d'autres briques de traitement. Ces briques composeraient ainsi des sous-systèmes qui interagissent entre eux dans un large système de dialogue hybride, incrémental de type *long-life* – SASI – *Socio-Affective Speech Interaction*. L'algorithme donne ainsi une courbe calculée de glu socio-affective qui permettrait de lier l'ensemble des sous-systèmes fonctionnant conjointement pour aboutir à résultante interactionnelle commune en perpétuelle évolution.

La visualisation couplée de la courbe de glu prédite par calcul et de la courbe de glu mesurée permettrait alors d'appréhender par des essais de pondération manuelle les effets interactionnels des différentes dimensions de l'interaction. Ces sous-systèmes intégreraient des dimensions de modalités distinctes, en modifiant l'ensemble du système interactionnel dont font partie l'humain et le robot eux-mêmes, sous-systèmes de cet ensemble. Cette manière, qui semble au premier abord chronophage, serait notamment le moyen de simuler en quelque sorte la boîte noire des systèmes d'apprentissage stochastiques, permettant d'explicitier le poids d'un effet interactionnel sans se focaliser uniquement sur sa résultante. Le développement itératif d'un tel système dans le contexte interactionnel donnerait ainsi lieu à de véritables boucles agiles de co-fabrication/évaluation où l'automatisation par petits essais du robot fondé sur ce principe d'intelligence socio-affective. Ainsi cela permettrait peut-être d'appréhender un nouveau paradigme de modélisation informatique passant des sciences affectives informatiques (ou *Affective Computing*) à une modélisation relationnelle informatique soit tendre vers des tentatives de *Relational Computing*. En effet, les sciences affectives ont largement contribué à introduire une dimension socio-affective à la modélisation de l'interaction. En revanche, elles représentent des états latents de moments interactionnels et font face au verrou théorique de la prise de décision du comportement à adopter pour assurer la continuité communicative. En effet, que dire à quels moments sachant cet état socio-affectif ? La modélisation de la relation tend quant à elle à décrire les résultantes des processus communicationnels socio-affectifs. Elle permettrait de définir, des choix collectifs de développement quant à la prise de risques qu'induirait le comportement à associer au robot. Ces prises de décision représentent ainsi l'un des enjeux éthiques majeurs associés au développement de la robotique sociale, que seules des méthodologies de co-construction peuvent garantir. Cette dimension est d'autant plus marquée pour les futurs usagers fragilisés par des formes d'isolement. À travers le public des personnes âgées isolées, nous avons pu voir que les primitives langagières de formes de « prosodie pure » peuvent suffire à rentrer dans des processus de manipulation relationnelle à l'appui d'importants

## **Discussions et Perspectives**

---

mécanismes empathiques. Ces dernières peuvent contribuer, dans le meilleur des cas, à ré-entraîner les « rouages rouillés » de la dynamique communicationnelle, et dans le pire désorienter la personne. Un robot ne naît donc pas compagnon, il le devient à travers des processus interactionnels contribuant à renforcer et à modifier la glu socio-affective qui soutendrait l'ensemble des processus langagiers qui contribuent à l'esquisse, au maintien et la modification de l'identité portée par le rôle social de l'humain.

## Bibliographie

Adams, S., Arel, I., Bach, J., Coop, R., Furlan, R., Goertzel, B., ... et coll. (2012). *Mapping the landscape of human-level artificial general intelligence*. *AI magazine*, 33(1), 25–42.

Adolphs, R. (1999). *Social cognition and the human brain*. *Trends in cognitive sciences*, 3(12), 469–479.

Ainsworth, M. D., et Wittig, B. (1969). *Attachment, exploration, and separation: illustrated by the behavior of one-year-olds in a strange situation*. *Determinants of infant behaviour*, 4, 113–136.

Ainsworth, M. S., et Bowlby, J. (1991). *An ethological approach to personality development*. *American psychologist*, 46(4), 333.

Ajili, M., Bonastre, J.-F., Kahn, J., Rossato, S., et Bernard, G. (2016). *FABIOLE, a Speech Database for Forensic Speaker Comparison*. In LREC. Consulté à l'adresse : [http://www.lrec-conf.org/proceedings/lrec2016/pdf/35\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2016/pdf/35_Paper.pdf)

Akita, K. (2009). *A grammar of sound-symbolic words in Japanese: Theoretical approaches to iconic and lexical properties of mimetics*. Kobe University.

Albrecht, K. (2006). *Social Intelligence: The New Science of Success; Beyond IQ, Beyond EI, Applying Multiple Intelligence Theory to Human Interaction*. San Francisco, CA: Jossey-Bass.

Alku, P., Bäckström, T., et Vilkman, E. (2002). *Normalized amplitude quotient for parametrization of the glottal flow*. *The Journal of the Acoustical Society of America*, 112(2), 701–710. <https://doi.org/10.1121/1.1490365>

Allen, K. M., Blascovich, J., Tomaka, J., et Kelsey, R. M. (1991). *Presence of human friends and pet dogs as moderators of autonomic responses to stress in women*. *Journal of personality and social psychology*, 61(4), 582.

Allport, G. W., et Odbert, H. S. (1936). *Trait-names: A psycho-lexical study*. *Psychological monographs*, 47(1), i.

Allwood, J. (1995). *An activity based approach to pragmatics*. *Gothenburg papers in theoretical linguistics*, (76), 1–38.

Aman, F., Vacher, M., Portet, F., Duclot, W., et Lecouteux, B. (2016). *CirDoX: an On/Off-line Multisource Speech and Sound Analysis Software*. In *Language Resources and Evaluation Conference* (p. 1978–1985). Consulté à l'adresse : <https://hal.archives-ouvertes.fr/hal-01323596/>

Aman, F., Vacher, M., Rossato, S., Dugheanu, R., Portet, F., le Grand, J., et Sasa, Y. (2012). *Étude de la performance des modèles acoustiques pour des voix de personnes âgées en vue de l'adaptation des systèmes de RAP*. In *Actes de la conférence conjointe JEP-TALN-RECITAL 2012* (Vol. 1, p. 707-714). Grenoble, France: Association for Computational Linguistics. Consulté à l'adresse : <http://aclweb.org/anthology//F/F12/F12-1089.pdf>

Aman, F., Vacher, M., Rossato, S., et Portet, F. (2013). *Speech recognition of aged voice in the aal context: Detection of distress sentences*. In *2013 7th Conference on Speech Technology and Human – Computer Dialogue (SpeD)* (p. 1–8). Cluj-Napoca, Romania: IEEE. <https://doi.org/10.1109/SpeD.2013.6682669>

Ambady, N., et Rosenthal, R. (1992). *Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis*. *American Psychological Association. Psychological Bulletin*, 111(2), 256–274. <http://dx.doi.org/10.1037/0033-2909.111.2.256>

Ameka, F. (1992). *Interjections: The universal yet neglected part of speech*. *Journal of Pragmatics*, 18(2), 101–118.

- Anzalone, S. M., Boucenna, S., Ivaldi, S., et Chetouani, M. (2015). *Evaluating the Engagement with Social Robots*. International Journal of Social Robotics, 1–14.
- Argyle, M. (1994). *The Psychology of Interpersonal Behaviour*. Penguin UK.
- Argyle, M. (2013). *Bodily communication*. Routledge.
- Arnold, M. B. (1960). *Emotion and personality*. Consulté à l'adresse : <http://psycnet.apa.org/psycinfo/1960-35012-000>
- Asch, S. E. (1946). *Forming impressions of personality*. The Journal of Abnormal and Social Psychology, 41(3), 258.
- Attfield, S., Kazai, G., Lalmas, M., et Piwowarski, B. (2011). *Towards a science of user engagement* (position paper). In WSDM workshop on user modelling for Web applications (p. 9–12). Consulté à l'adresse : <http://www.dcs.gla.ac.uk/~mounia/Papers/engagement.pdf>
- Aubergé, V. (1991). *La synthèse de la parole : des règles aux lexiques*. Grenoble 2. Consulté à l'adresse : <http://www.theses.fr/1991GRE29041>
- Aubergé, V. (2002). *A gestalt morphology of prosody directed by functions: the example of a step by step model developed at ICP*. In Speech Prosody 2002, International Conference. Consulté à l'adresse : [http://www.isca-speech.org/archive\\_open/sp2002/sp02\\_151.html](http://www.isca-speech.org/archive_open/sp2002/sp02_151.html)
- Aubergé, V. (2016, février). *Illusion empathique du robot social : réparation de « l'autre » fantôme*. Présenté à 7e rencontres du Pôle Grenoble Cognition, Grenoble, France.
- Aubergé, V., Audibert, N., et Rilliard, A. (2004). *E-Wiz: a Trapper Protocol for Hunting the Expressive Speech Corpora in Lab*. In LREC. European Language Resources Association.
- Aubergé, V., Audibert, N., et Rilliard, A. (2006). *De E-Wiz à C-Clone: recueil, modélisation et synthèse d'expressions authentiques*. Consulté à l'adresse : <https://hal.archives-ouvertes.fr/hal-00364495>
- Aubergé, V., et Cathiard, M. (2003). *Can we hear the prosody of smile?* Speech Communication, 40(1), 87–97.
- Aubergé, V., Rilliard, A., et Audibert, N. (2005). *De E-Wiz à E-clone: méthodologie expérimentale pour la modélisation des émotions et affects authentiques*. In Actes du Premier Workshop francophone sur les Agents Conversationnels Animés (p. 125-134). Consulté à l'adresse : <http://hal.archives-ouvertes.fr/hal-00364524>
- Aubergé, V., Sasa, Y., Bonnefond, N., Meillon, B., Robert, T., Rey-Gorrez, J., ... Nebout, F. (2014). *The EEE corpus: socio-affective « glue » cues in elderly-robot interactions in a Smart Home with the Emoz platform*. In 5th International Workshop on Emotion, Social Signals, Sentiment and Linked Oen Data, Reykjavik, Iceland.
- Aubergé, V., Sasa, Y., Robert, T., Bonnefond, N., et Meillon, B. (2013). *Emoz: a wizard of Oz for emerging the socio-affective glue with a non humanoid companion robot*. Présenté à WASSS – Workshop on Affective Social Speech Signals, Grenoble, France.
- Audibert, N. (2008, décembre 2). *Prosodie de la parole expressive : dimensionnalité d'énoncés méthodologiquement contrôlés authentiques et actés* (phdthesis). Institut National Polytechnique de Grenoble – INPG. Consulté à l'adresse : <https://tel.archives-ouvertes.fr/tel-00489924/document>
- Austin, J. L. (1975). *How to do things with words*. Oxford university press.
- Avril, M., Chetouani, M., et Sabouret, N. (2013). *Étude d'une collaboration socio-affective entre une personne et le robot jazz*. Revue Interfaces Numériques, 2(1), 57–76.
- Baba, A., Yoshizawa, S., Yamada, M., Lee, A., et Shikano, K. (2001). *Elderly acoustic model for large vocabulary continuous speech recognition*. <https://doi.org/10.1002/ecjb.20101>
- Bain, D. E. (1989). *An evaluation of evolutionary processes: studies of natural selection, dispersal, and cultural evolution in killer whales (Orcinus orca)*.

- Baltus, G., Fox, D., Gemperle, F., Goetz, J., Hirsch, T., Magaritis, D., ... coll. (2000). *Towards personal service robots for the elderly. In the Workshop on Interactive Robotics and Entertainment*. In WIRE-2000. Consulté à l'adresse : <http://www.fore.robot.cc/papers/thrun.nursebot-early.pdf>
- Barbulescu, A., Hueber, T., Bailly, G., Ronfard, R., et undefined, et coll. (2013). *Audio-Visual Speaker Conversion using Prosody Features. In International Conference on Auditory-Visual Speech Processing*. Consulté à l'adresse : <http://hal.archives-ouvertes.fr/hal-00842928/>
- Bargh, J. A., Chen, M., et Burrows, L. (1996). *Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action*. *Journal of Personality and Social Psychology*, 71(2), 230-244. <https://doi.org/10.1037/0022-3514.71.2.230>
- Barlow, J., Roehrich, J. K., et Wright, S. (2010). *De facto privatization or a renewed role for the EU? Paying for Europe's healthcare infrastructure in a recession*. *Journal of the Royal Society of Medicine*, 103(2), 51–55.
- Baron-Cohen, S. (1995). *The eye direction detector (EDD) and the shared attention mechanism (SAM): Two cases for evolutionary psychology*. In Portions of this paper were presented at the Society for Research in Child Development Conference, New Orleans, Mar 1993; the British Psychological Society, Welsh Branch, "Faces" Conference, U Wales Coll of Cardiff, Sep 1993; and the British Society for the Philosophy of Science "Roots of Joint Reference" Conference, U Bristol, Nov 1993. Lawrence Erlbaum Associates, Inc. Consulté à l'adresse : <http://doi.apa.org/psycinfo/1995-97586-002>
- Barrett-Lennard, G. T. (1981). *The empathy cycle: Refinement of a nuclear concept*. *Journal of counseling psychology*, 28(2), 91.
- Barthe, J.-F., Clément, S., et Drulhe, M. (1990). *Vieillesse ou vieillissement? Les processus d'organisation des modes de vie chez les personnes âgées*. *International Review of Community Development/Revue internationale d'action communautaire*, (23), 35–46.
- Batista Antunes, L., Aubergé, V., et Sasa, Y. (2014). *Certainty and uncertainty in Brazilian Portuguese: methodology of spontaneous corpus collection and data analysis*. In *Speech Prosody 7*, Dublin, Ireland.
- Batliner, A., Hacker, C., Steidl, S., Nöth, E., D'Arcy, S., Russell, M. J., et Wong, M. (2004). *"You Stupid Tin Box" -Children Interacting with the AIBO Robot: A Cross-linguistic Emotional Speech Corpus*. In LREC. Consulté à l'adresse : <http://hmk.ffzg.hr/bibl/lrec2004/pdf/317.pdf>
- Batliner, A., Kießling, A., Burger, S., et Nöth, E. (2011). *Filled pauses in spontaneous speech*. Technical Report 88, Verbmobil Project. also in ICPhS'95.
- Batson, C. D. (1991). *The Altruism Question: Toward A Social-psychological Answer*. Hillsdale, NJ: Psychology Press.
- Batson, C. D. (2009). *These things called empathy: eight related but distinct phenomena*. In J. Decety & W. Ickes (Eds.), *Social neuroscience. The social neuroscience of empathy* (pp. 3-15). <http://dx.doi.org/10.7551/mitpress/9780262012973.003.0002>
- Baudonnière, P.-M. (1997). *Le mimétisme et l'imitation: un exposé pour comprendre, un essai pour réfléchir*. Flammarion.
- Bauer, A., Wollherr, D., et Buss, M. (2008). *Human-robot collaboration: a survey*. *International Journal of Humanoid Robotics*, 5(1), 47–66.
- Baum, L. F. (1987). *The Wonderful Wizard of Oz*. 1900. New York: Books of Wonder.
- Beauvois, J.-L. (1982). *Théories implicites de la personnalité, évaluation et reproduction idéologique*. *L'année psychologique*, 82(2), 513-536. <https://doi.org/10.3406/psy.1982.28434>
- Beck, A., Edwards, N., Friedman, B., et Khan, P. (2003). *Robotic pets and the elderly*. Project overview. Consulté à l'adresse : <http://www.ischool.washington.edu/robotpets/elderly/>
- Bellangeon, M. (1984). *Problèmes posés au vétérinaire par la consultation de ces Nouveaux Animaux de Compagnie (NAC)*. *Sci. Vét. Méd. Comp*, 86(3), 77–78.

- Bellik, Y. (1995). *Interfaces multimodales: concepts, modèles et architectures*. Thèse de doctorat. Paris 11.
- Belmin, J., Auffray, J.-C., Berbezier, C., Boirin, P., Mercier, S., de Reviers, B., et Golmard, J.-L. (2007). *Level of dependency: a simple marker associated with mortality during the 2003 heatwave among French dependent elderly people living in the community or in institutions*. *Age and ageing*, 36(3), 298–303.
- Bemelmans, R., Gelderblom, G. J., Jonker, P., et de Witte, L. (2012). *Socially Assistive Robots in Elderly Care: A Systematic Review into Effects and Effectiveness*. *Journal of the American Medical Directors Association*, 13(2), 114-120.e1. <https://doi.org/10.1016/j.jamda.2010.10.002>
- Bennett, R. (1973). *Living Conditions and Everyday Needs of the Elderly with Particular Reference to Social Isolation*. *The International Journal of Aging and Human Development*, 4(3), 179-198. <https://doi.org/10.2190/Y1HA-CG77-WCT5-URGV>
- Benoît, C., et Le Goff, B. (1998). *Audio-visual speech synthesis from French text: Eight years of models, designs and evaluation at the ICP*. *Speech Communication*, 26(1), 117–129.
- Bergvall-Kareborn, B., Hoist, M., et Stahlbrost, A. (2009). *Concept design with a living lab approach*. In *System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference on* (p. 1–10). IEEE.
- Berkman, L. F. (1983). *The assessment of social networks and social support in the elderly*. *Journal of the American Geriatrics Society*, 31(12), 743-749.
- Bernhardt, B. C., et Singer, T. (2012). *The neural basis of empathy*. *Annual Review of Neuroscience*, 35, 1-23. <https://doi.org/10.1146/annurev-neuro-062111-150536>
- Besacier, L., Barnard, E., Karpov, A., et Schultz, T. (2014). *Automatic speech recognition for under-resourced languages: A survey*. *Speech Communication*, 56, 85–100.
- Beskow, J. (1997). *Animation of talking agents*. In *Audio-Visual Speech Processing: Computational et Cognitive Science Approaches*. Consulté à l'adresse : [http://www.isca-speech.org/archive\\_open/avsp97/av97\\_149.html](http://www.isca-speech.org/archive_open/avsp97/av97_149.html)
- Biasi, G. D., Auberge, V., et Granjon, L. (2012). *Perception of social affects from non lexical sounds*. In *GSCP*. Consulté à l'adresse : <http://hal.archives-ouvertes.fr/hal-00959138>
- Bickerton, D. (1992). *Language and Species*. University of Chicago Press.
- Bigi, B., et Hirst, D. (2012). *SPeech Phonetization Alignment and Syllabification (SPPAS): a tool for the automatic analysis of speech prosody*. In *Speech Prosody* (p. 1–4). Consulté à l'adresse : <https://hal.archives-ouvertes.fr/hal-00983699/>
- Billard, A., et Dautenhahn, K. (1998). *Grounding communication in autonomous robots: an experimental study*. *Robotics and Autonomous Systems*, 24(1-2), 71–79.
- Billings, D. R., Schaefer, K. E., Chen, J. Y., et Hancock, P. A. (2012). *Human-robot interaction: developing trust in robots*. In *Human-Robot Interaction (HRI), 2012 7th ACM/IEEE International Conference on* (p. 109–110). IEEE.
- Bird - David, N. (1999). « *Animism* » *Revisited: Personhood, Environment, and Relational Epistemology*. *Current Anthropology*, 40(S1), S67-S91. <https://doi.org/10.1086/200061>
- Birdwhistell, R. L. (2010). *Kinesics and Context: Essays on Body Motion Communication*. University of Pennsylvania Press.
- Bocquet, H., et Andrieu, S. (1999). « *Le burden* »: *Un indicateur spécifique pour les aidants familiaux: Le vécu relationnel*. *Gérontologie et société*, (89), 155–166.
- Boden, M., Bryson, J., Caldwell, D., Dautenhahn, K., Edwards, L., Kember, S., ... coll. (2017). *Principles of robotics: regulating robots in the real world*. *Connection Science*, 29(2), 124–129.
- Müller C. Cienki A. Fricke E. Ladewig S. McNeill D. Tessedorf S. (n.d.). *Body – Language – Communication*. An International Handbook on Multimodality in Human Interaction. Berlin, Boston: De Gruyter Mouton.

- Boersma, P. (2006). *Praat: doing phonetics by computer*. Outil en ligne (première version 1992) : <http://www.praat.org/>. Consulté à l'adresse : <http://ci.nii.ac.jp/naid/10017594077/>
- Bonnell, B. (2010). *Viva la Robolution !* Paris: JC Lattès.
- Borovoy, A. (2008). *Japan's hidden youths: mainstreaming the emotionally distressed in Japan*. *Culture, Medicine, and Psychiatry*, 32(4), 552–576.
- Bouma, H., Fozard, J. L., Bouwhuis, D. G., et Taipale, V. T. (2007). *Gerontechnology in perspective*. *Gerontechnology*, 6(4), 190–216.
- Bowlby, J., et Ainsworth, M. (2013). *The origins of attachment theory*. *Attachment Theory: Social, Developmental, and Clinical Perspectives*, 45.
- Bowlby, J., Robertson, J., et Rosenbluth, D. (1952). *A two-year-old goes to hospital*. *The Psychoanalytic Study of the Child*, 7(1), 82–94.
- Bradley, N., et Poppen, W. (2003). *Assistive technology, computers and Internet may decrease sense of isolation for homebound elderly and disabled persons*. *Technology and Disability*, 15(1), 19-25.
- Breazeal, C. (2003). *Toward sociable robots*. *Robotics and Autonomous Systems*, 42(3–4), 167-175. [https://doi.org/10.1016/S0921-8890\(02\)00373-1](https://doi.org/10.1016/S0921-8890(02)00373-1)
- Breazeal, C., Kidd, C. D., Thomaz, A. L., Hoffman, G., et Berlin, M. (2005). *Effects of nonverbal communication on efficiency and robustness in human-robot teamwork*. In *Intelligent Robots and Systems, 2005.(IROS 2005)*. 2005 IEEE/RSJ International Conference on (p. 708–713). IEEE.
- Brennan, S. E., et Williams, M. (1995). *The feeling of Another's Knowing: prosody and filled pauses as cues to listeners about the metacognitive states of speakers*. *Journal of memory and language*, 34(3), 383–398.
- Brenton, H., Gillies, M., Ballin, D., et Chatting, D. (2005). *The Uncanny Valley: does it exist and is it related to presence*. *Presence Connect*. Consulté à l'adresse : [http://www.bespokevr.com/pdfs/brenton\\_UncannyValley\\_2005.pdf](http://www.bespokevr.com/pdfs/brenton_UncannyValley_2005.pdf)
- Brittain, K., Corner, L., Robinson, L., et Bond, J. (2010). *Ageing in place and technologies of place: the lived experience of people with dementia in changing social, physical and technological environments*. *Sociology of health et illness*, 32(2), 272–287.
- Broekens, J., Heerink, M., et Rosendal, H. (2009). *Assistive social robots in elderly care: a review*. *Gerontechnology*, 8(2). <https://doi.org/10.4017/gt.2009.08.02.002.00>
- Brooks, R. A., et coll. (1991). *Intelligence without reason*. *Artificial intelligence: critical concepts*, 3, 107–63.
- Brugman, H., Russel, A., et Nijmegen, X. (2004). *Annotating Multi-media/Multi-modal Resources with ELAN*. In LREC. Consulté à l'adresse : <https://pdfs.semanticscholar.org/3dac/d66abb9a06950017e057a14c911b592b1809.pdf>
- Bryson, A. E., Ho, Y. C., et Hestenes, M. R. (1969). *Techniques for improved convergence in Neighboring optimum guidance*. *Methods*, 3(4). Consulté à l'adresse : <https://arc.aiaa.org/doi/pdfplus/10.2514/3.6093>
- Buchanan, R. (1992). *Wicked problems in design thinking*. *Design issues*, 8(2), 5–21.
- Bui Huu Trung, B. H. T. (2006). *Multimodal Dialogue Management – State of the art*. (CTIT Technical Report Series; No. 06-01). Enschede: Centre for Telematics and Information Technology University of Twente.
- Burgar, C. G., Lum, P. S., Shor, P. C., et Van der Loos, H. M. (2000). *Development of robots for rehabilitation therapy: The Palo Alto VA/Stanford experience*. *Journal of rehabilitation research and development*, 37(6), 663–674.
- Burgoon, J. K., Stern, L. A., et Dillman, L. (1995). *Interpersonal adaptation: Dyadic patterns of interaction*. New York: Cambridge University Press.
- Burlingame, H. J. (1891). *Leaves from Conjurers' Scrap Books, Or, Modern Magicians and Their Works*. Donohue, Henneberry et Company.

- Burnette, K. (1986). *Relocation and the Elderly Changing Perspectives*. Journal of Gerontological Nursing, 12(10), 6–9.
- Butterworth, G. (1991). *The ontogeny and phylogeny of joint visual attention*. In A. Whiten (Ed.), *Natural theories of mind: Évolution, development and simulation of everyday mindreading* (pp. 223–232). Cambridge, MA: Basil Blackwell.
- Byrne, R. W., et Whiten, A. (1997). *Machiavellian intelligence. Machiavellian intelligence II: Extensions and evaluations*, 1–23.
- Cacioppo, J. T., et Cacioppo, S. (2014). *Social relationships and health: The toxic effects of perceived social isolation*. Social and personality psychology compass, 8(2), 58–72.
- Cacioppo, J. T., Hawkey, L. C., Crawford, L. E., Ernst, J. M., Burlinson, M. H., Kowalewski, R. B., ... Berntson, G. G. (2002). *Loneliness and health: Potential mechanisms*. Psychosomatic Medicine, 64(3), 407–417.
- Cambria, E., Rajagopal, D., Olsher, D., et Das, D. (2013). *Big social data analysis*. Big data computing, 401–414.
- Campbell, N. (2004). *Speech et Expression; the Value of a Longitudinal Corpus*. In LREC. Consulté à l'adresse : <http://www.speech-data.jp/nick/feast/pubs/lrec04.pdf>
- Campbell, N. (2005). *Developments in corpus-based speech synthesis: Approaching natural conversational speech*. IEICE transactions on information and systems, 88(3), 376–383.
- Campbell, N. (2007). *On the use of nonverbal speech sounds in human communication*. In *Verbal and nonverbal communication behaviours* (p. 117–128). Springer.
- Campbell, N., et Mokhtari, P. (2003). *Voice quality: the 4th prosodic dimension*. In 15th ICPhS (p. 2417–2420). Consulté à l'adresse : [https://www.researchgate.net/profile/Parham\\_Mokhtari/publication/228720666\\_Voice\\_quality\\_The\\_4th\\_prosodic\\_dimension/links/0c960518075455b448000000/Voice-quality-The-4th-prosodic-dimension.pdf](https://www.researchgate.net/profile/Parham_Mokhtari/publication/228720666_Voice_quality_The_4th_prosodic_dimension/links/0c960518075455b448000000/Voice-quality-The-4th-prosodic-dimension.pdf)
- Campos, J. J., Stenberg, C., et coll. (1981). *Perception, appraisal, and emotion: The onset of social referencing*. Infant social cognition: Empirical and theoretical considerations, 273, 314.
- Candea, M., Vasilescu, I., et Adda-Decker, M. (2005). *Inter-and intra-language acoustic analysis of autonomous fillers*. In DISS 05, Disfluency in Spontaneous Speech Workshop (p. 47–52). Consulté à l'adresse : <https://halshs.archives-ouvertes.fr/halshs-00321914/>
- Carlson, M., Charlin, V., et Miller, N. (1988). *Positive mood and helping behavior: a test of six hypotheses*. Journal of Personality and Social Psychology, 55(2), 211–229.
- Carston, R. (2008). *Thoughts and utterances: The pragmatics of explicit communication*. John Wiley et Sons.
- Casanueva, I., Hain, T., Christensen, H., Marxer, R., et Green, P. D. (2015). *Knowledge transfer between speakers for personalised dialogue management*. In SIGDIAL Conference (p. 12–21). Consulté à l'adresse : <http://www.anthology.aclweb.org/W/W15/W15-46.pdf#page=32>
- Cassel, J. (1976). *The Contribution of the Social Environment to Host Resistance the Fourth Wade Hampton Frost Lecture*. American journal of epidemiology, 104(2), 107–123.
- Cassidy, R. E. (1995). *The Art of Mentalism*. Dream Crow Press. Consulté à l'adresse : <http://vincent.mucchielli.free.fr/PUA/Skills/Magic%20and%20Card%20Tricks/Bob%20Cassidy/Art%20of%20Mentalism%202.pdf>
- Castle, N. G. (2001). *Relocation of the elderly*. Medical care research and review, 58(3), 291–333.
- Challoner, J., et Haond, P. (2003). *L'Intelligence artificielle*. Paris: Dorling Kindersley.
- Chaminade, T., Hodgins, J., et Kawato, M. (2007). *Anthropomorphism influences perception of computer-animated characters' actions*. Social Cognitive and Affective Neuroscience, 2(3), 206–216. <https://doi.org/10.1093/scan/nsm017>

- Charles, G. (1981). *Conversational organization: Interaction between speakers and hearers*. New York, Academic Press.
- Charpentier, F., et Stella, M. (1986). *Diphone synthesis using an overlap-add technique for speech waveforms concatenation*. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'86*. (Vol. 11, p. 2015–2018). IEEE.
- Chartrand, T. L., et Bargh, J. A. (1999). *The chameleon effect: the perception–behavior link and social interaction*. *Journal of personality and social psychology*, 76(6), 893.
- Chemero, A. (2003). *An outline of a theory of affordances*. *Ecological psychology*, 15(2), 181–195.
- Cheney, D. L., et Seyfarth, R. M. (1992). *How monkeys see the world: Inside the mind of another species*. University of Chicago Press.
- Cheng, Y., Chen, C., Lin, C.-P., Chou, K.-H., et Decety, J. (2010). *Love hurts: an fMRI study*. *NeuroImage*, 51(2), 923-929. <https://doi.org/10.1016/j.neuroimage.2010.02.047>
- Chesnais, J.-C. (1990). *Demographic Transition Patterns and Their Impact on the Age Structure*. *Population and Development Review*, 16(2), 327-336. <https://doi.org/10.2307/1971593>
- Chetouani, M. (2014). *Role of inter-personal synchrony in extracting social signatures: some case studies*. In *Proceedings of the 2014 Workshop on Roadmapping the Future of Multimodal Interaction Research including Business Opportunities and Challenges* (p. 9–12). Istanbul, Turkey: ACM. <https://doi.org/10.1145/2666253.2666263>
- Chomsky, N. (2005). *Three factors in language design*. *Linguistic inquiry*, 36(1), 1–22.
- Christensen, W., Sutton, J., et McIlwain, D. J. (2016). *Cognition in skilled action: Meshed control and the varieties of skill experience*. *Mind et Language*, 31(1), 37–66.
- Cialdini, R. B., Baumann, D. J., et Kenrick, D. T. (1981). *Insights from sadness: A three-step model of the development of altruism as hedonism*. *Developmental Review*, 1(3), 207-223. [https://doi.org/10.1016/0273-2297\(81\)90018-6](https://doi.org/10.1016/0273-2297(81)90018-6)
- Clark, H. H., et Krych, M. A. (2004). *Speaking while monitoring addressees for understanding*. *Journal of memory and language*, 50(1), 62–81.
- Cobb, S. (1976). *Social support as a moderator of life stress*. *Psychosomatic medicine*, 38(5), 300–314.
- Colletta, J.-M. (2007). *Signaux corporels et acquisition du langage: des relations constantes et étroites*. *Langage et pratiques*, 39, 20–33.
- Condon, W. S., et Ogston, W. D. (1966). *Sound film analysis of normal and pathological behavior patterns*. *The Journal of nervous and mental disease*, 143(4), 338–347.
- Conein, B. (1998). *Les sens sociaux: coordination de l'attention et interaction sociale*. *Intellectica*, 1(2), 26–27.
- Contini, M. (1989). *L'interjection en sarde. Une approche linguistique*. *Espaces romans. études de dialectologie et de géolinguistique offertes à Gaston Tuaille*, 320–329.
- Conway, J. H. (1971). *Regular Algebra and Finite Mechanics*. London: Chapman and Hall.
- Cooley, C. H. (1992). *Human nature and the social order*. Transaction Publishers.
- Copans, J. (2008). *L'enquête et ses méthodes: l'enquête ethnologique de terrain*. Armand Colin.
- Core, M. G., et Allen, J. (1997). *Coding dialogs with the DAMSL annotation scheme*. In *AAAI fall symposium on communicative action in humans and machines* (Vol. 56). Boston, MA. Consulté à l'adresse : <http://www.cis.udel.edu/~carberry/CIS-885/Papers/Core-Allen-Coding-Damsl.pdf>
- Cornelius, R. R. (2000). *Theoretical approaches to emotion*. In *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion*. Consulté à l'adresse : [http://www.isca-speech.org/archive\\_open/archive\\_papers/speech\\_emotion/spem\\_003.pdf](http://www.isca-speech.org/archive_open/archive_papers/speech_emotion/spem_003.pdf)

- Cornwell, E. Y., et Waite, L. J. (2009). *Social Disconnectedness, Perceived Isolation, and Health among Older Adults*. *Journal of Health and Social Behavior*, 50(1), 31-48. <https://doi.org/10.1177/002214650905000103>
- Corrigan, L. J., Peters, C., Castellano, G., Papadopoulos, F., Jones, A., Bhargava, S., ... Aylett, R. (2013). *Social-task engagement: Striking a balance between the robot and the task*. In *Embodied Commun. Goals Intentions Workshop ICSR* (Vol. 13, p. 1–7). Consulté à l'adresse : <https://pdfs.semanticscholar.org/e697/142d11aee850a988775c0d64e0c2b6cbe040.pdf>
- Cosnier, J. (1927-) A. du texte. (1994). *Psychologie des émotions et des sentiments*. Paris: Retz. Consulté à l'adresse : <http://gallica.bnf.fr/ark:/12148/bpt6k48001944>
- Cosnier, J. (1996). *Les gestes du dialogue, la communication non verbale*. *Psychologie de la motivation*, 21, 129–138.
- Coudin, G. (2004). *La réticence des aidants familiaux à recourir aux services gérontologiques: une approche psychosociale*. *Psychologie et NeuroPsychiatrie du vieillissement*, 2(4), 285–296.
- Cowie, R. (2009). *Perceiving emotion: towards a realistic understanding of the task*. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535), 3515-3525. <https://doi.org/10.1098/rstb.2009.0139>
- Cox, A. D., Puckering, C., Pound, A., et Mills, M. (1987). *The impact of maternal depression in young children*. *Journal of Child Psychology and Psychiatry*, 28(6), 917–928.
- Cross, T. G., Johnson-Morris, J. E., et Nienhuys, T. G. (1980). *Linguistic feedback and maternal speech: Comparisons of mcoll. addressing hearing and hearing-impaired children*. *First Language*, 1(3), 163–189.
- Crystal, D. (1976). *Prosodic systems and intonation in English* (Vol. 1). CUP Archive.
- Crystal, D., et Quirk, R. (1964). *Systems of prosodic and paralinguistic features in English*. Walter De Gruyter Inc.
- Cunningham, P., et Cunningham, M. (2008). *Collaboration and the knowledge economy: issues, applications, case studies*. IOS Press.
- Cutler, A., Dahan, D., et Van Donselaar, W. (1997). *Prosody in the comprehension of spoken language: A literature review*. *Language and speech*, 40(2), 141–201.
- Cutler, A., et Pearson, M. (1985). *On the analysis of prosodic turn-taking cues*. Consulté à l'adresse : <http://repository.uibn.ru.nl/bitstream/handle/2066/15702/6098.pdf>
- Cutting, J., et coll. (2005). *Pragmatics and discourse: A resource book for students*. Routledge.
- Dahlbäck, N., Jönsson, A., et Ahrenberg, L. (1993). *Wizard of Oz studies—why and how*. *Knowledge-based systems*, 6(4), 258–266.
- Damasio, A. (2003). *Looking for Spinoza: Joy, Sorrow, and the Feeling Brain* (1st éd.). Orlando Toronto London: Mariner Books.
- Damasio, A. (2010). *L'erreur de Descartes : La raison des émotions*. Paris: Odile Jacob.
- Damasio, A. R. (1999). *The feeling of what happens: body and emotion in the making of consciousness*. Harcourt Brace, 386, ISBN: 0156010755, 9780156010757.
- Danermark, B., et Ekstrom, M. (1990). *Relocation and Health Effects on the Elderly-A Commented Research Review*. *J. Soc. and Soc. Welfare*, 17, 25.
- Darling, K. (2012). *Extending Legal Rights to Social Robots* (SSRN Scholarly Paper No. ID 2044797). Rochester, NY: Social Science Research Network. Consulté à l'adresse : [http://robots.law.miami.edu/wp-content/uploads/2012/04/Darling\\_Extending-Legal-Rights-to-Social-Robots-v2.pdf](http://robots.law.miami.edu/wp-content/uploads/2012/04/Darling_Extending-Legal-Rights-to-Social-Robots-v2.pdf)
- Darwall, S. (1998). *Empathy, sympathy, care*. *Philosophical Studies*, 89(2), 261–282.
- Dautenhahn, K. (1999). *Embodiment and interaction in socially intelligent life-like agents*. *Computation for metaphors, analogy, and agents*, 102–141.

Dautenhahn, K., Ogden, B., et Quick, T. (2002). *From Embodied to Socially Embedded Agents – Implications for Interaction-aware Robots*. *Cogn. Syst. Res.*, 3(3), 397–428.  
[https://doi.org/10.1016/S1389-0417\(02\)00050-5](https://doi.org/10.1016/S1389-0417(02)00050-5)

Davidson, R. J., Sherer, K. R., et Goldsmith, H. H. (2009). *Handbook of affective sciences*. Oxford University Press.

Dawkins, R., et Krebs, J. R. (1978). *Animal signals: information or manipulation?* In *Behavioural ecology: An evolutionary approach* (Vol. 2, p. 282-309).

De Ruyter, B., Saini, P., Markopoulos, P., et Van Breemen, A. (2005). *Assessing the effects of building social intelligence in a robotic interface for the home*. *Interacting with computers*, 17(5), 522–541.

de Waal, F. B. M. (2008). *Putting the altruism back into altruism: the evolution of empathy*. *Annual Review of Psychology*, 59, 279-300.  
<https://doi.org/10.1146/annurev.psych.59.103006.093625>

Decety, J. (2010). *Mécanismes neurophysiologiques impliqués dans l'empathie et la sympathie*. *Revue de neuropsychologie*, Volume 2(2), 133-144.  
<https://doi.org/10.3917/rne.022.0133>

Decety, J., Echols, S., et Correll, J. (2010). *The blame game: the effect of responsibility and social stigma on empathy for pain*. *Journal of Cognitive Neuroscience*, 22(5), 985-997.  
<https://doi.org/10.1162/jocn.2009.21266>

Decety, J., et Chaminade, T. (2003). *Neural correlates of feeling sympathy*. *Neuropsychologia*, 41(2), 127–138.

Decety, J., et Jackson, P. L. (2004). *The functional architecture of human empathy*. *Behavioral and cognitive neuroscience reviews*, 3(2), 71–100.

Decety, J., et Lamm, C. (2006). *Human Empathy Through the Lens of Social Neuroscience*. *The Scientific World Journal*, 6, 1146-1163.  
<https://doi.org/10.1100/tsw.2006.221>

Delaborde, A., & Devillers, L. (2010, October). *Use of nonverbal speech cues in social interaction between human and robot: emotional and interactional markers*. In *Proceedings of the 3rd international workshop on Affective interaction in natural environments* (pp. 75-80). ACM.

Delaherche, E., Chetouani, M., Mahdhaoui, A., Saint-Georges, C., Viaux, S., et Cohen, D. (2012). *Interpersonal Synchrony : A Survey of Evaluation Methods across Disciplines*. In *IEEE Transactions on Affective Computing*.

Delattre, P. (1966). *Les dix intonations de base du français*. *French review*, 1–14.

Demuth, K., et Morgan, J. (1996). *The prosodic structure of early words*. *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, 171, 184.

Deng, L., Yu, D., et coll. (2014). *Deep learning: methods and applications*. *Foundations and Trends® in Signal Processing*, 7(3–4), 197–387.

Descola, P., et Pálsson, G. (1996). *Nature and society: anthropological perspectives*. Taylor et Francis.

Dessalles, J.-L. (2000). *Language and hominid politics*, na. Consulté à l'adresse : <http://groups.lis.illinois.edu/amag/langev/localcopy/pdf/dessalles00languageAnd.pdf>

Devillers, L., Rosset, S., Duplessis, G. D., Sehili, M. A., Béchade, L., Delaborde, A., ... & Türker, B. B. (2015, September). *Multimodal data collection of human-robot humorous interactions in the joker project*. In *Affective Computing and Intelligent Interaction (ACII), 2015 International Conference on* (pp. 348-354). IEEE.

Devillers, L., Tahon, M., Sehili, M. A., & Delaborde, A. (2015). *Inference of human beings' emotional states from speech in human–robot interactions*. *International Journal of Social Robotics*, 7(4), 451-463.

Devillers, L. (2017). *Des robots et des hommes*. Plon.

Di Cristo, A. (1985). *De la microprosodie à l'intonosyntaxe* (Vol. 1). Publications, Université de Provence.

- Digard, J.-P. (2004). *La construction sociale d'un animal domestique: le pitbull*. *Anthropozoologica*, 39(1), 17–26.
- Dijkstra, A., Buist, G. A., Dassen, T. W., van den Heuvel, W. J., et voor Gezondheidsvraagstukken, *Care Dependency Scale (CDS)*. N. C. (s. d.). Consulté à l'adresse : <http://www.umcg.nl/SiteCollectionDocuments/research/institutes/SHARE/assessment%20tools/CDS%20manual%20english.pdf>
- Dijkstra, C., Kraemer, E., et Swerts, M. (2006). *Manipulating uncertainty: The contribution of different audiovisual prosodic cues to the perception of confidence*. In *Proceedings of the Speech Prosody Conference*. Citeseer. Consulté à l'adresse : <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.536.2623>
- Dimberg, U., Thunberg, M., et Elmehed, K. (2000). *Unconscious facial reactions to emotional facial expressions*. *Psychological science*, 11(1), 86–89.
- Douglas-Cowie, E., Cowie, R., Sneddon, I., Cox, C., Lowry, O., Mcrorie, M., ... et coll. (2007). *The HUMAINE database: addressing the collection and annotation of naturalistic and induced emotional data*. In *Affective computing and intelligent interaction* (p. 488–500). Springer.
- Duffield, A. M., Robertson, A. V., Djerassi, C., Buchanan, B. G., Sutherland, G. L., Feigenbaum, E. A., et Lederberg, J. (1969). *Applications of artificial intelligence for chemical inference. II. Interpretation of low-resolution mass spectra of ketones*. *Journal of the American Chemical Society*, 91(11), 2977–2981.
- Duffy, B. R. (2003). *Anthropomorphism and the social robot*. *Robotics and autonomous systems*, 42(3), 177–190.
- Dunbar, P. R. (2004). *Grooming, Gossip and the Évolution of Language* (New ed edition). London: Faber and Faber.
- Dunbar, R. I. M. (1988). *Primate social systems: studies in behavioural adaptation*. Croom Helm, London.
- Duncan, S. (1972). *Some signals and rules for taking speaking turns in conversations*. *Journal of personality and social psychology*, 23(2), 283.
- Durkheim, E. (1951). *Suicide: a study in sociology* [1897]. Translated by JA Spaulding and G. Simpson (Glencoe, Illinois: The Free Press, 1951).
- Dutoit, T. (1997). *An introduction to text-to-speech synthesis* (Vol. 3). Springer Science et Business Media.
- Dybkjær, L., Bernsen, N. O., et Minker, W. (2004). *Evaluation and usability of multimodal spoken language dialogue systems*. *Speech Communication*, 43(1), 33–54. <https://doi.org/10.1016/j.specom.2004.02.001>
- Edlund, J., Gustafson, J., Heldner, M., et Hjalmarsson, A. (2008). *Towards human-like spoken dialogue systems*. *Speech communication*, 50(8), 630–645.
- Eisenberg, N., et Strayer, J. (1990). *Empathy and its development*. CUP Archive.
- Ekman, P. (1992). *An argument for basic emotions*. *Cognition et emotion*, 6(3-4), 169–200.
- Ekman, P., et Friesen, W. V. (1969). *The repertoire of nonverbal behavior: Categories, origins, usage, and coding*. *Semiotica*, 1(1), 49–98.
- Ekman, P., et Friesen, W. V. (1971). *Constants across cultures in the face and emotion*. *Journal of personality and social psychology*, 17(2), 124.
- Ekman, P., et Friesen, W. V. (1975). *Unmasking the face: A guide to recognizing emotions from facial cues*. Englewood Cliffs, NJ: Prentice Hall.
- El Asri, L., Lemonnier, R., Laroche, R., Pietquin, O., et Khouzaimi, H. (2014). *NASTIA: Negotiating Appointment Setting Interface*. In *LREC* (p. 266–271). Consulté à l'adresse : [http://www.lrec-conf.org/proceedings/lrec2014/pdf/575\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2014/pdf/575_Paper.pdf)

- Elger, G., et Furugren, B. (1998). *SmartBo-an ICT and computer-based demonstration home for disabled people*. In Proceedings of the 3rd TIDE Congress: Technology for Inclusive Design and Equality Improving the Quality of Life for the European Citizen. Helsinki, Finland June.
- Enard, W., Przeworski, M., Fisher, S. E., Lai, C. S. L., Wiebe, V., Kitano, T., ... Pääbo, S. (2002). *Molecular evolution of FOXP2, a gene involved in speech and language*. *Nature*, 418(6900), 869-872. <https://doi.org/10.1038/nature01025>
- Enninger, W. (1987). *What interactants do with non-talk across cultures*. *Analyzing Intercultural Communication*. Berlin: Mouton de Gruyter, 269–302.
- Erickson, D. (2005). *Expressive speech: Production, perception and application to speech synthesis*. *Acoustical Science and Technology*, 26(4), 317–325.
- Eriksson, J., Mataric, M. J., et Winstein, C. J. (2005). *Hands-off assistive robotics for post-stroke arm rehabilitation*. In *Rehabilitation Robotics, 2005. ICORR 2005. 9th International Conference on* (p. 21–24). IEEE.
- Eriksson, M., Niitamo, V.-P., et Kulkki, S. (2005). *State-of-the-art in utilizing Living Labs approach to user-centric ICT innovation-a European approach*. Lulea: Center for Distance-spanning Technology. Lulea University of Technology Sweden: Lulea. Consulté à l'adresse : <https://pdfs.semanticscholar.org/2edd/5e0fef9f7f9fd0262dea937cb997b3ab8d5f.pdf>
- Eslinger, P. J. (1998). *Neurological and neuropsychological bases of empathy*. *European Neurology*, 39(4), 193-199.
- Evanini, K., Hunter, P., Liscombe, J., Suendermann, D., Dayanidhi, K., et Pieraccini, R. (2008). *Caller experience: a method for evaluating dialog systems and its automatic prediction*. In *Spoken Language Technology Workshop, 2008. SLT 2008. IEEE* (p. 129–132). IEEE.
- Fansten, M., Figueiredo, C., et Pionnié-Dax, N. (2014). *Hikikomori, ces adolescents en retrait*. Armand Colin.
- Fausto, C. (2011). *Le masque de l'animiste*. *Gradhiva. Revue d'anthropologie et d'histoire des arts*, (13), 48-67. <https://doi.org/10.4000/gradhiva.2030>
- Feil-Seifer, D., et Mataric, M. J. (2005). *Defining socially assistive robotics*. In *Rehabilitation Robotics, 2005. ICORR 2005. 9th International Conference on* (p. 465–468). IEEE.
- Feil-Seifer, D., et Mataric, M. J. (2011). *Socially assistive robotics*. *IEEE Robotics et Automation Magazine*, 18(1), 24–31.
- Ferré, G. (2012). *Critères de segmentation de la gestualité co-verbale*. In *JEP-TALN-RECITAL 2012, Atelier DEGELS 2012: Défi GEst Langue des Signes* (p. 9–21). Consulté à l'adresse : <https://hal.inria.fr/hal-00707594/>
- Ferrucci, D. A. (2012). *Introduction to "This is Watson"*. *IBM Journal of Research and Development*, 56(3.4), 1:1-1:15. <https://doi.org/10.1147/JRD.2012.2184356>
- Festinger, L. (1954). *A Theory of Social Comparison Processes*. *Human Relations*, 7(2), 117-140. <https://doi.org/10.1177/001872675400700202>
- Fitch, K. (1990). *A ritual for attempting leave-taking in Colombia*. *Research on Language et Social Interaction*, 24(1-4), 209–224.
- Fónagy, I. (1983). *La vive voix: essais de psycho-phonétique* (Vol. 20). Payot.
- Fong, T., Nourbakhsh, I., et Dautenhahn, K. (2003). *A survey of socially interactive robots*. *Robotics and Autonomous Systems*, 42(3–4), 143-166. [https://doi.org/10.1016/S0921-8890\(02\)00372-X](https://doi.org/10.1016/S0921-8890(02)00372-X)
- Fontaine, J. R., Scherer, K. R., Roesch, E. B., et Ellsworth, P. C. (2007). *The world of emotions is not two-dimensional*. *Psychological science*, 18(12), 1050–1057.

- Forsberg, M. (2003). *Why is speech recognition difficult*. Chalmers University of Technology. Consulté à l'adresse : [https://www.researchgate.net/profile/Markus\\_Forsberg/publication/228763868\\_Why\\_is\\_speech\\_recognition\\_difficult/links/0deec53229ad44345e000000.pdf](https://www.researchgate.net/profile/Markus_Forsberg/publication/228763868_Why_is_speech_recognition_difficult/links/0deec53229ad44345e000000.pdf)
- Fraser, N. M., et Gilbert, G. N. (1991). *Simulating speech systems*. *Computer Speech et Language*, 5(1), 81–99.
- Freedly, E., DeVisser, E., Weltman, G., et Coeyman, N. (2007). *Measurement of trust in human-robot collaboration*. In *International Symposium on Collaborative Technologies and Systems*, 2007. CTS 2007 (p. 106-114). <https://doi.org/10.1109/CTS.2007.4621745>
- Freud, S. (2014). *Inhibitions, symptoms and anxiety*. Read Books Ltd.
- Frey, S., Hirsbrunner, H. P., Florin, A., Daw, W., & Crawford, R. (1983). *A unified approach to the investigation of nonverbal and verbal behavior in communication research*. In W. Doise & S. Moscovici (Eds.), *Current issues in European social psychology* (pp. 143-199). Cambridge: Cambridge University Press.
- Frijda, N. H., Kuipers, P., et Ter Schure, E. (1989). Relations among emotion, appraisal, and emotional action readiness. *Journal of personality and social psychology*, 57(2), 212.
- Frisch, K. von. (1993). *The dance language and orientation of bees*. Harvard University Press.
- Frohlich, M., Michaelis, D., et Strube, H. W. (1998). *Acoustic “breathiness measures” in the description of pathologic voices*. In *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing*, 1998 (Vol. 2, p. 937-940 vol.2). <https://doi.org/10.1109/ICASSP.1998.675420>
- Fujino, Y., et Yamada, T. (2006). *Development of a “Japanese Sports Onomatopoeias” Computerized Dictionary for Understanding of Subtle Movement*. In *EdMedia: World Conference on Educational Media and Technology* (p. 1034–1041). Association for the Advancement of Computing in Education (AACE).
- Fujisaki, H. (1988). *A note on the physiological and physical basis for the phrase and accent components in the voice fundamental frequency contour*. *Vocal physiology: Voice production, mechanisms and functions*, 347–355.
- Fujisaki, H. (1997). *Prosody, models, and spontaneous speech*. In *Computing Prosody* (p. 27–42). Springer.
- Fujita, C. (1971). *“Hitomishiri” as a Japanese Concept of Stranger Anxiety and Anthropophobia*. *Psychiatry and Clinical Neurosciences*, 25(1), 1–8.
- Fultz, J., et Cialdini, R. B. (1991). *Situational and personality determinants of the quantity and quality of helping*. *Cooperation and prosocial behavior*, 135–146.
- Furlong, A. (2008). *The Japanese hikikomori phenomenon: acute social withdrawal among young people*. *The sociological review*, 56(2), 309–325.
- Gadet, F. (2003). *Derrière les problèmes méthodologiques du recueil des données*. *Texto! 2003 pour l'édition électronique*. [http://icar.univ-lyon2.fr/ecole\\_thematique/idocora/documents/Gadet\\_Derriere\\_le\\_probleme\\_methodologique.pdf](http://icar.univ-lyon2.fr/ecole_thematique/idocora/documents/Gadet_Derriere_le_probleme_methodologique.pdf)
- Gaikwad, S. K., Gawali, B. W., et Yannawar, P. (2010). *A review on speech recognition technique*. *International Journal of Computer Applications*, 10(3), 16–24.
- Gallese, V., et Goldman, A. (1998). *Mirror neurons and the simulation theory of mind-reading*. *Trends in Cognitive Sciences*, 2(12), 493-501.
- Gamez, D. (2008). *Progress in machine consciousness*. *Consciousness and cognition*, 17(3), 887–910.
- Garfinkel, H. (2005). *Ethnomethodological studies of work*. Routledge.
- Garnier, C. (1991). *Le Corps rassemblé : pour une perspective interdisciplinaire et culturelle de la corporéité*. Montréal: Éditions agence d'arc.

- Garnier, M., Dubois, D., Poitevineau, J., Henrich, N., et Castellengo, M. (2004). *Perception et description verbale de la qualité vocale dans le chant lyrique: une approche cognitive*. JEP. Consulté à l'adresse : <http://www.afpc-evta-france.com/wp-content/uploads/sites/10/2015/02/99-Perception-et-description-verbale-de-la-qualit%C3%A9.pdf>
- Gašić, M., Breslin, C., Henderson, M., Kim, D., Szummer, M., Thomson, B., ... Young, S. (2013). *POMDP-based dialogue manager adaptation to extended domains*. In Proceedings of SIGDIAL. Consulté à l'adresse : [http://mi.eng.cam.ac.uk/~mg436/papers/gp\\_adapt\\_revised.pdf](http://mi.eng.cam.ac.uk/~mg436/papers/gp_adapt_revised.pdf)
- Gašić, M., Jurčićek, F., Keizer, S., Mairesse, F., Thomson, B., Yu, K., et Young, S. (2010). *Gaussian processes for fast policy optimisation of POMDP-based dialogue managers*. In Proceedings of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue (p. 201–204). Association for Computational Linguistics.
- Gatica-Perez, D. (2009). *Automatic nonverbal analysis of social interaction in small groups: A review*. Image and vision computing, 27(12), 1775–1787.
- Gee, F. C., Browne, W. N., et Kawamura, K. (2005). *Uncanny valley revisited*. In Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on (p. 151–157). IEEE.
- Genevay, A., et Laroche, R. (2016). *Transfer learning for user adaptation in spoken dialogue systems*. In Proceedings of the 2016 International Conference on Autonomous Agents et Multiagent Systems (p. 975–983). International Foundation for Autonomous Agents and Multiagent Systems.
- Gentzler, A., et Kerns, K. (2006). *Adult attachment and memory of emotional reactions to negative and positive events*. Cognition et Emotion, 20(1), 20–42.
- Gerin, W., Pieper, C., Levy, R., et Pickering, T. G. (1992). *Social support in social interaction: a moderator of cardiovascular reactivity*. Psychosomatic medicine, 54(3), 324–336.
- Gibson, J. J. (1977). *The theory of affordances*. Hilldale, USA.
- Giese, M. A., et Poggio, T. (2003). *Neural mechanisms for the recognition of biological movements*. Nature Reviews. Neuroscience, 4(3), 179–192. <https://doi.org/10.1038/nrn1057>
- Girard-Rivier, M., Magnani, R., Aubergé, V., Sasa, Y., Tsvetanova, L., Aman, F., et Bayol, C. (2016). *Ecological Gestures for HRI: the GEE Corpus*. In LREC. Consulté à l'adresse : [http://www.lrec-conf.org/proceedings/lrec2016/pdf/1040\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2016/pdf/1040_Paper.pdf)
- Gobl, C., et Chasaide, A. N. (2000). *Testing affective correlates of voice quality through analysis and resynthesis*. In ISCA tutorial and research workshop (ITRW) on Speech and Emotion. Consulté à l'adresse : [http://www.isca-speech.org/archive\\_open/speech\\_emotion/spem\\_178.html](http://www.isca-speech.org/archive_open/speech_emotion/spem_178.html)
- Gobl, C., et Chasaide, A. N. (2003). *Amplitude-based source parameters for measuring voice quality*. In ISCA Tutorial and Research Workshop on Voice Quality: Functions, Analysis and Synthesis. Consulté à l'adresse : [http://www.isca-speech.org/archive\\_open/voqual03/voq3\\_151.html](http://www.isca-speech.org/archive_open/voqual03/voq3_151.html)
- Goertzel, B. (2012). *What counts as a thinking machine – and when will we meet one?* New Scientist, 215(2881), 18. [https://doi.org/10.1016/S0262-4079\(12\)62309-4](https://doi.org/10.1016/S0262-4079(12)62309-4)
- Goertzel, B., Pennachin, C., et Geisweiller, N. (2014). *A Preschool-Based Roadmap to Advanced AGI*. In Engineering General Intelligence, Part 1 (p. 355–364). Atlantis Press, Paris. [https://doi.org/10.2991/978-94-6239-027-0\\_18](https://doi.org/10.2991/978-94-6239-027-0_18)
- Goffman, E. (1961). *Encounters: Two studies in the sociology of interaction*. Oxford, England: Bobbs-Merrill.
- Goffman, E. (1981). *Forms of talk*. University of Pennsylvania Press.
- Goffman, E. (2002). *The presentation of self in everyday life*. 1959. Garden City, NY.
- Goffman, E. (2008). *Behavior in public places*. Simon and Schuster.

- Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning* (1st éd.). Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc.
- Goldman, J.-P. (2011) *EasyAlign: an automatic phonetic alignment tool under Praat*. In: Interspeech'11, 12th Annual Conference of the International Speech Communication Association. Firenze (Italy). Consulté à l'adresse : <https://archive-ouverte.unige.ch/unige:18188>
- Goldstein, E. B. (1981). *The ecology of JJ Gibson's perception*. *Leonardo*, 14(3), 191–195.
- Gómez, J. C. (1991). *Visual behaviour as a window for reading the mind of others in primates*. In A. Whiten (Ed.), *Natural theories of mind: Évolution, development and simulation of everyday mindreading* (pp. 195-207). Cambridge, MA: Basil Blackwell. <http://doi.apa.org/psycinfo/1991-97348-012>
- Goodrich, M. A., et Schultz, A. C. (2007). Human-robot Interaction: A Survey. *Found. Trends Hum.-Comput. Interact.*, 1(3), 203–275. <https://doi.org/10.1561/11000000005>
- Goodwin, C. (2003). The body in action. In *Discourse, the body, and identity* (p. 19–42). Springer.
- Gordon, R. M. (1995). *Sympathy, Simulation, and the Impartial Spectator*. *Ethics*, 105(4), 727–742.
- Goudbeek, M., et Scherer, K. (2010). *Beyond arousal: Valence and potency/control cues in the vocal expression of emotion*. *The Journal of the Acoustical Society of America*, 128(3), 1322–1336.
- Graf, H. P., Cosatto, E., Strom, V., et Huang, F. J. (2002). *Visual prosody: Facial movements accompanying speech*. In *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on* (p. 396–401). IEEE.
- Gravano, A., et Hirschberg, J. (2009). *Backchannel-inviting cues in task-oriented dialogue*. In *Tenth Annual Conference of the International Speech Communication Association*. Consulté à l'adresse : [http://habla.dc.uba.ar/gravano/files/gravano\\_hirschberg\\_2009a.pdf](http://habla.dc.uba.ar/gravano/files/gravano_hirschberg_2009a.pdf)
- Gravano, A., et Hirschberg, J. (2011). *Turn-taking cues in task-oriented dialogue*. *Computer Speech et Language*, 25(3), 601–634.
- Green, A., Huttenrauch, H., et Eklundh, K. S. (2004). *Applying the Wizard-of-Oz framework to cooperative service discovery and configuration*. In *Robot and Human Interactive Communication, 2004. ROMAN 2004. 13th IEEE International Workshop on* (p. 575–580). IEEE.
- Green, S. A., Billingham, M., Chen, X., et Chase, J. G. (2008). *Human-robot collaboration: A literature review and augmented reality approach in design*. *International Journal of Advanced Robotic Systems*, 5(1), 1.
- Greenberg, Y., Shibuya, N., Tsuzaki, M., Kato, H., et Sagisaka, Y. (2006). *A trial of communicative prosody generation based on control characteristic of one word utterance observed in real conversational speech*. In *Proc. Speech Prosody* (p. 37–40). Consulté à l'adresse : [http://20.210-193-52.unknown.qala.com.sg/archive/sp2006/papers/sp06\\_037.pdf](http://20.210-193-52.unknown.qala.com.sg/archive/sp2006/papers/sp06_037.pdf)
- Grice, H. P. (1991). *Studies in the Way of Words*. Harvard University Press.
- Grice, M., Reyelt, M., Benzmueller, R., Mayer, J., et Batliner, A. (1996). *Consistency in transcription and labelling of German intonation with GToBI*. In *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on* (Vol. 3, p. 1716–1719). IEEE.
- Grimaud, E., et Paré, Z. (2011). *Le jour où les robots mangeront des pommes: conversations avec un Geminoid*. Editions Pétra.
- Grimby, A. (1993). *Bereavement among elderly people: grief reactions, post-bereavement hallucinations and quality of life*. *Acta Psychiatrica Scandinavica*, 87(1), 72–80.
- Gross, J. J. (1998). *The emerging field of emotion regulation: An integrative review*. *Review of general psychology*, 2(3), 271.

Gross, J. J., et Barrett, L. F. (2011). *Emotion Generation and Emotion Regulation: One or Two Depends on Your Point of View*. *Emotion review*, 3(1), 8-16.  
<https://doi.org/10.1177/1754073910380974>

Gubrud, M. A. (1997). *Nanotechnology and international security*. In Fifth Foresight Conference on Molecular Nanotechnology (Vol. 1).

Gudeman, S. (1986). *Economics as culture: models and metaphors of livelihood*. Routledge.

Guichet, J.-L. (2011). *L'animal familial aujourd'hui: la réduction du domestique à l'appivoisé*. *Le divan familial*, (1), 13–26.

Guidetti, M., Turquois, L., Adrien, J.-L., Barthélémy, C., et Bernard, J.-L. (2004). *Aspects pragmatiques de la communication et du langage chez des enfants typiques et des enfants ultérieurement diagnostiqués autistes*. *Psychologie Française*, 49(2), 131-144.  
<https://doi.org/10.1016/j.psfr.2004.04.001>

Guillaume, L., Aubergé, V., Magnani, R., Aman, F., Cottier, C., Sasa, Y., ... Girard-Rivier, M. (July 1st – July 3rd). *HRI in an ecological dynamic experiment: the GEE corpus based approach for the Emox robot*. The 2015 IEEE International Workshop on Advanced Robotics and its Social Impacts. IEEE.

Gupta, V., et Korte, C. (1994). *The effects of a confidant and a peer group on the well-being of single elders*. *The International Journal of Aging and Human Development*, 39(4), 293–302.

Gussenhoven, C. (2002). *Intonation and interpretation: Phonetics and phonology*. In *Speech Prosody 2002*, International Conference. Consulté à l'adresse : [http://www.isca-speech.org/archive\\_open/sp2002/sp02\\_047.html](http://www.isca-speech.org/archive_open/sp2002/sp02_047.html)

Gussenhoven, C. (2005). *Transcription of Dutch intonation*. *Prosodic typology: The phonology of intonation and phrasing*, 118–145.

Guthrie, S. E. (1995). *Faces in the clouds: A new theory of religion*. Oxford University Press.

Guyomard, M., et Siroux, J. (1986). *Suggestive and Corrective Answers: a Single Mechanism. Structure of Multimodal Dialogues Including Voice*. In NATO Workshop, Sept (p. 1–5).

Haag, G. (1996). *Réflexions sur quelques particularités des émergences de langage chez les enfants autistes*. *Journal de pédiatrie et de puériculture*, 9(5), 261–264.

Haddington, P., Mondada, L., et Nevile, M. (2013). *Interaction and mobility: Language and the body in motion* (Vol. 20). Walter de Gruyter.

Hagège, C. (2009). *Dictionnaire amoureux des Langues*. Paris: Plon.

Halimi-Falkowicz, S., Quaderi, A., et Joule, R.-V. (2007). *La soumission librement consentie au service de la socialisation de l'Âgé*. *Pratiques psychologiques*, 13(2), 213–225.

Hall Edward, T. (1976). *Beyond culture*. Garden City, NY: Doubleday et Company.

Hall, B., et Noguchi, M. (1995). *Engaging in Kenson: An extended case study of one form of "common" sense*. *Human Relations*, 48(10), 1129–1147.

Hall, E. T. (1959). *The silent language*. Garden City, N.Y.: Doubleday.

Hall, E. T. (1983). *The Dance of Life: The Other Dimension of Time*. Anchor Press.

Hall, E. T. (1990). *The Hidden Dimension* (27th printing edition). New York, NY: Anchor.

Hall, E. T., Birdwhistell, R. L., Bock, B., Bohannon, P., Diebold Jr, A. R., Durbin, M., ... coll. (1968). *Proxemics [and comments and replies]*. *Current anthropology*, 9(2/3), 83–108.

Hall, J., Tritton, T., Rowe, A., Pipe, A., Melhuish, C., et Leonards, U. (2014). *Perception of own and robot engagement in human–robot interactions and their dependence on robotics knowledge*. *Robotics and Autonomous Systems*, 62(3), 392–399.

- Harcourt, A. H. (1988). *Alliances in contests and social intelligence*. In R. W. Byrne & A. Whiten (Eds.), *Machiavellian intelligence: Social expertise and the evolution of intellect in monkeys, apes, and humans* (pp. 132-152). New York: Clarendon Press/Oxford University Press.
- Harlow, H. F. (1958). *The nature of love*. *American psychologist*, 13(12), 673.
- Harris, M. (1983). *Cultural anthropology*. New York: Harper and Row Publishers.
- Hart, J. T. (1965). *Memory and the feeling-of-knowing experience*. *Journal of educational psychology*, 56(4), 208.
- Hasegawa, T., et Gudykunst, W. B. (1998). *Silence in Japan and the United States*. *Journal of Cross-Cultural Psychology*, 29(5), 668–684.
- Hatfield, E., Cacioppo, J. T., et Rapson, R. L. (1993). *Emotional Contagion*. Cambridge: Cambridge University Press.
- Haugeland, J. (Éd.). (1997). *Mind Design II: Philosophy, Psychology, and Artificial Intelligence* (2nd Revised, Enlarged ed. edition). Cambridge, Mass: A Bradford Book.
- Hawley, L. C., Burleson, M. H., Berntson, G. G., et Cacioppo, J. T. (2003). *Loneliness in everyday life: cardiovascular activity, psychosocial context, and health behaviors*. *Journal of personality and social psychology*, 85(1), 105.
- Healy, H., et Williams, J. M. G. (1999). *Autobiographical memory*. *Handbook of cognition and emotion*, 229–242.
- Heath, C. (1986). *Body movement and speech in medical interaction*. Cambridge University Press.
- Hebb, D. O. (2005). *The organization of behavior: A neuropsychological theory*. Psychology Press.
- Heerink, M., Kroese, B., Evers, V., et Wielinga, B. (2006). The influence of a robot's social abilities on acceptance by elderly users. In *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on* (p. 521–526). IEEE.
- Heidegger, M. (1992). *Être et Temps*. Paris: Gallimard.
- Heider, F. (1958). *The Psychology of Interpersonal Relations*. Psychology Press.
- Henrich, Nathalie. (2001) *Etude de la source glottique en voix parlée et chantée : modélisation et estimation, mesures acoustiques et électroglottographiques, perception*. Phdthesis, Université Pierre et Marie Curie – Paris VI, 2001. <https://tel.archives-ouvertes.fr/tel-00123133/document>.
- Henrich, N., d'Alessandro, C., et Doval, B. (2001). *Spectral correlates of voice open quotient and glottal flow asymmetry: theory, limits and experimental data*. In *Seventh European Conference on Speech Communication and Technology*. Consulté à l'adresse : <http://www.ee.columbia.edu/~dpwe/papers/HenAD01-glottal.pdf>
- Herrera, E., Reissland, N., et Shepherd, J. (2004). *Maternal touch and maternal child-directed speech: effects of depressed mood in the postnatal period*. *Journal of affective disorders*, 81(1), 29–39.
- Heyes, C. M. (1998). *Theory of mind in nonhuman primates*. *The Behavioral and Brain Sciences*, 21(1), 101-114-148.
- Heylen, D., et coll. (2007). *Computing backchannel distributions in multi-party conversations*. In *Proceedings of the Workshop on Embodied Language Processing* (p. 17–24). Association for Computational Linguistics.
- Hillenbrand, J., Cleveland, R. A., et Erickson, R. L. (1994). *Acoustic correlates of breathy vocal quality*. *Journal of Speech, Language, and Hearing Research*, 37(4), 769–778.
- Himbert, M.-N. (2012). *Le robot pensant*. Paris: Editions du Moment.
- Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A., Jaitly, N., ... et coll. (2012). *Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups*. *IEEE Signal Processing Magazine*, 29(6), 82–97.

- Hodges, S. D., & Wegner, D. M. (1997). *Automatic and controlled empathy*. In W. J. Ickes (Ed.), *Empathic accuracy* (pp. 311-339). New York: Guilford Press.
- Hoffman, G. (2013). *Evaluating fluency in human-robot collaboration*. In International conference on human-robot interaction (HRI), workshop on human robot collaboration (Vol. 381, p. 1–8).
- Hoffman, G., et Breazeal, C. (2007). *Effects of anticipatory action on human-robot teamwork: Efficiency, fluency, and perception of team*. In 2007 2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI) (p. 1-8).  
<https://doi.org/10.1145/1228716.1228718>
- Hoffman, M. L. (2008). *Empathy and prosocial behavior*. *Handbook of emotions*, 3, 440–455.
- Hoffman, M. L. (2010). *Empathy and Moral Development: Implications for Caring and Justice*. (New Ed). Cambridge: Cambridge University Press.
- Hoffmann, P. (1982). *Modèle mécaniste et modèle animiste*. De quelques aspects de la représentation du vivant chez Descartes, Borelli et Stahl in *La machine dans l’imaginaire (1650-1800)*. *Revue des Sciences Humaines Lille*, (186-187), 199-211.
- Hojo, H., Nomachi, N., Tomoto, Y., Nakamura, T., Kanoh, M., et Yamada, K. (2014). *Fundamental Study for Verbalization of Embodied Expertise based on Pattern Recognition*. In *Advanced Applied Informatics (IIAIAAI)*, 2014 IIAI 3rd International Conference on (p. 925–929). IEEE.
- Hölldobler, B., et Wilson, E. O. (1994). *Journey to the ants: a story of scientific exploration*. Edward Osborne Wilson – Belknap Press., 304. ISBN 9780674485266.
- Holmes, D. S. (1968). *Dimensions of projection*. *Psychological bulletin*, 69(4), 248.
- Hombert, J.-M., Bergounioux, G., Bocquet-Appel, J.-P., Coupé, C., et coll. (2005). *Aux origines des langues et du langage*. Paris: Fayard.
- Hornak, J., Bramham, J., Rolls, E. T., Morris, R. G., O’Doherty, J., Bullock, P. R., et Polkey, C. E. (2003). *Changes in emotion after circumscribed surgical lesions of the orbitofrontal and cingulate cortices*. *Brain*, 126(7), 1691–1712.
- House, J. S., Landis, K. R., Umberson, D., et coll. (1988). *Social relationships and health*. *Science*, 241(4865), 540–545.
- Howe, J. (2006). *The rise of crowdsourcing*. *Wired magazine*, 14(6), 1–4.
- Humphrey, N. K. (1976). *The social function of intellect*. In *Growing points in ethology* (p. 303–317). Cambridge University Press. Consulté à l’adresse :  
<http://cogprints.org/1737/2/SocialFunction.pdf>
- Hunter, D. E., et Whitten, P. (1976). *The Study of Anthropology*. Harper et Row.
- Hymes, D. (1962). *The ethnography of speaking*. *Anthropology and human behavior*, 13(53), 11–74.
- Ickes, W. (1993). *Empathic Accuracy*. *Journal of Personality*, 61(4), 587-610.  
<https://doi.org/10.1111/j.1467-6494.1993.tb00783.x>
- Inoue, K., Wada, K., et Uehara, R. (2011). *How effective is robot therapy?: Paro and people with dementia*. In 5th European Conference of the International Federation for Medical and Biological Engineering (p. 784–787). Springer.
- Inui, A. (2005). *Why freeter and NEET are misunderstood: recognizing the new precarious conditions of Japanese youth*. *Social Work et Society*, 3(2), 244–251.
- Ishi, C. T., Ishiguro, H., et Hagita, N. (2008). *Automatic extraction of paralinguistic information using prosodic features related to F0, duration and voice quality*. *Speech Communication*, 50(6), 531-543. <https://doi.org/10.1016/j.specom.2008.03.009>
- Ishi, C., Ishiguro, H., et Hagita, N. (2006). *Using prosodic and voice quality features for paralinguistic information extraction*. In *Proc. of Speech Prosody* (p. 883–886). Consulté à l’adresse : <http://www.irc.atr.jp/~carlos/pdf/carlos-sp2006.pdf>

Ishi, C., Ishiguro, H., et Hagita, N. (2010). *Analysis of the Roles and the Dynamics of Breathy and Whispery Voice Qualities in Dialogue Speech*. EURASIP Journal on Audio, Speech, and Music Processing, 2010(1), 528193. <https://doi.org/10.1155/2010/528193>

Ishihara, K., Komatani, K., Ogata, T., et Okuno, H. G. (2005). *Sound-imitation word recognition for environmental sounds*. Transactions of the Japanese Society for Artificial Intelligence, 20, 229–236.

Istrate, D., Castelli, E., Vacher, M., Besacier, L., et Serignat, J.-F. (2006). *Information extraction from sound for medical telemonitoring*. IEEE Transactions on Information Technology in Biomedicine, 10(2), 264–274.

Ito, M. (2004). *Politeness and voice quality-the alternative method to measure aspiration noise*. In Speech Prosody 2004, International Conference. Consulté à l'adresse : [http://www.isca-speech.org/archive\\_open/sp2004/sp04\\_213.html](http://www.isca-speech.org/archive_open/sp2004/sp04_213.html)

Ivaldi, S., Anzalone, S. M., Rousseau, W., Sigaud, O., et Chetouani, M. (2014). *Robot initiative in a team learning task increases the rhythm of interaction but not the perceived engagement*. Frontiers in neurorobotics, 8.

Iwamura, Y., Shiomi, M., Kanda, T., Ishiguro, H., et Hagita, N. (2011). *Do elderly people prefer a conversational humanoid as a shopping assistant partner in supermarkets?* In Human-Robot Interaction (HRI), 2011 6th ACM/IEEE International Conference on (p. 449–457). IEEE.

Jackendoff, R. (1994). *Is there a faculty of social cognition?* Noam Chomsky: Critical assessments. L, 3, 629–643.

Jacobson, S. W. (1979). *Matching behavior in the young infant*. Child development, 425–430.

Jambon, F., Meillon, B., Perrot, C., Caelen, J., et Harsché, M. (2009). *Création d'une plate-forme d'utilisabilité/usage: objectifs, réalisation, usages et retours d'expérience*. IHM.

James, W. (1884). *What is an emotion?* Mind, 9(34), 188–205.

Jaworski, A. (1992). *The power of silence: Social and pragmatic perspectives* (Vol. 1). Sage Publications.

Jee, E.-S., Jeong, Y.-J., Kim, C. H., et Kobayashi, H. (2010). *Sound design for emotion and intention expression of socially interactive robots*. Intelligent Service Robotics, 3(3), 199-206. <https://doi.org/10.1007/s11370-010-0070-7>

Jenkins, O. C., Nicolescu, M. N., et Mataric, M. J. (2004). *Autonomy and supervision for robot skills and tasks learned from demonstration*. In AAAI-04 Workshop on Supervisory Control of Learning and Adaptive Systems. Consulté à l'adresse : <https://www.cse.unr.edu/~monica/Research/Publications/aaai-sc04.pdf>

Jimenez, F., Kanoh, M., Yoshikawa, T., Furuhashi, T., et Nakamura, T. (2014). *Effect of robot utterances using onomatopoeia on collaborative learning*. Journal of Artificial Intelligence and Soft Computing Research, 4(2), 125–131.

Johnstone, T., et Scherer, K. R. (1999). *The effects of emotions on voice quality*. In Proceedings of the XIVth International Congress of Phonetic Sciences (p. 2029–2032). University of California, Berkeley San Francisco. Consulté à l'adresse : [https://www.researchgate.net/profile/Tom\\_Johnstone/publication/239727245\\_The\\_effects\\_of\\_emotions\\_on\\_voice\\_quality/links/0a85e52d655578f3dd000000.pdf](https://www.researchgate.net/profile/Tom_Johnstone/publication/239727245_The_effects_of_emotions_on_voice_quality/links/0a85e52d655578f3dd000000.pdf)

Jones, E. E., Nisbett, R. E., et coll. (1971). *The actor and the observer: Divergent perceptions of the causes of behavior*. Attribution: Perceiving the causes of behavior In E.E. Jones, D. E. Kanouse, H. H. Kelly, R.E. Nisbett, S. Valins, & B. Weiner, (Eds.). Consulté à l'adresse : [http://isites.harvard.edu/fs/docs/icb.topic628923.files/D\\_jones\\_nisbett1971pp79-94.pdf](http://isites.harvard.edu/fs/docs/icb.topic628923.files/D_jones_nisbett1971pp79-94.pdf)

Jonze, S. (2014). Her. Consulté à l'adresse : [http://www.allocine.fr/film/fichefilm\\_gen\\_cfilm=206799.html](http://www.allocine.fr/film/fichefilm_gen_cfilm=206799.html)

Joule, R.-V., et Azdia, T. (2004). Dissonance cognitive et engagement: lorsque deux comportements de soumission sont en contradiction avec la même attitude. *Les Cahiers internationaux de psychologie sociale*, (4), 5–11.

Joule, R.-V., et Beauvois, J.-L. (2004). *Petit traité de manipulation à l'usage des honnêtes gens* (Édition : Nouv. version). Grenoble: Presses Universitaires de Grenoble.

Journet, N. (2010). *L'animisme est-il une religion ?* *Les Grands Dossiers des Sciences Humaines*, (5), 4-4.

Jurafsky, D. (2000). *Speech and language processing: An introduction to natural language processing*. Computational linguistics, and speech recognition. Consulté à l'adresse : <http://ci.nii.ac.jp/naid/10013460372/>

Juslin, P. N., et Scherer, K. R. (2005). *Vocal expression of affect*. *The new handbook of methods in nonverbal behavior research*, 65–135.

Kaerlein, T. (2015). *The Social Robot as Fetish? Conceptual Affordances and Risks of Neo-Animistic Theory*. *International Journal of Social Robotics*, 7(3), 361-370. <https://doi.org/10.1007/s12369-014-0275-6>

Kagitani, T., Goto, M., Watanbe, J., et Sakamoto, M. (2014). *Sound Symbolic Relationship between Onomatopoeia and Emotional Evaluations in Taste*. In *Proceedings of the Cognitive Science Society* (Vol. 36). Consulté à l'adresse : <http://escholarship.org/uc/item/9bd4h4t2.pdf>

Kahn, J. (2011). *Parole de locuteur: performance et confiance en identification biométrique vocale*. Thèse. Avignon. Consulté à l'adresse : <http://www.theses.fr/2011AVIG0187>

Kaiser, S., Wehrle, T., et Schenkel, K. (2009). *Expression faciale des émotions*. *Traité de psychologie des émotions*. Paris: Dunod, 109–157.

Kaiser, S., Wehrle, T., et Schenkel, K. (2014). *Chapitre 3. Expression faciale*. In *Traité de psychologie des émotions* (p. 87–119). Dunod.

Kamarck, T. W., Manuck, S. B., et Jennings, J. R. (1990). *Social support reduces cardiovascular reactivity to psychological challenge: a laboratory model*. *Psychosomatic medicine*, 52(1), 42–58.

Kanamori, M., Suzuki, M., et Tanaka, M. (2002). *Maintenance and improvement of quality of life among elderly patients using a pet-type robot*. *Nihon Ronen Igakkai Zasshi. Japanese Journal of Geriatrics*, 39(2), 214-218.

Kaneko, S. (2006). *Japan's 'Socially Withdrawn Youths' and Time Constraints in Japanese Society: Management and conceptualization of time in a support group for 'hikikomori'*. *Time et Society*, 15(2-3), 233–249. <https://doi.org/DOI:10.1177/0961463X06067034>

Kaneshige, Y., Nihei, M., et Fujie, M. G. (2006). *Development of new mobility assistive robot for elderly people with body functional control*. In *Biomedical Robotics and Biomechatronics, 2006. BioRob 2006. The First IEEE/RAS-EMBS International Conference on* (p. 118–123). IEEE.

Kang, K. I., Freedman, S., Mataric, M. J., Cunningham, M. J., et Lopez, B. (2005). *A hands-off physical therapy assistance robot for cardiac patients*. In *Rehabilitation Robotics, 2005. ICORR 2005. 9th International Conference on* (p. 337–340). IEEE.

Kaplan, G. A., Salonen, J. T., Cohen, R. D., Brand, R. J., Leonard Syme, S., et Puska, P. (1988). *Social connections and mortality from all causes and from cardiovascular disease: prospective evidence from eastern Finland*. *American Journal of Epidemiology*, 128(2), 370–380.

Kato, T. A., Shinfuku, N., Sartorius, N., et Kanba, S. (2011). *Are Japan's hikikomori and depression in young people spreading abroad?* *Elsevier*, 378(9796), 1049-1116. [https://doi.org/http://dx.doi.org/10.1016/S0140-6736\(11\)61475-X](https://doi.org/http://dx.doi.org/10.1016/S0140-6736(11)61475-X)

Kato, T. A., Shinfuku, N., Sartorius, N., et Kanba, S. (2017). *Loneliness and Single-Person Households: Issues of Kodoku-Shi and Hikikomori in Japan*. In N. Okkels, C. B. Kristiansen, et P. Munk-Jorgensen (Éd.), *Mental Health and Illness in the City* (p. 1-15). Springer Singapore. [https://doi.org/10.1007/978-981-10-0752-1\\_9-1](https://doi.org/10.1007/978-981-10-0752-1_9-1)

Kelley, J. F. (1984). *An iterative design methodology for user-friendly natural language office information applications*. *ACM Transactions on Information Systems (TOIS)*, 2(1), 26–41.

Kendon, A., Harris, R. M., et Key, M. R. (1975). *Organization of Behavior in Face-to-Face Interaction*. Walter de Gruyter.

Kerbrat-Orecchioni, C. (1998). *Les interactions verbales* (3ème édition). Paris: Armand Colin.

Khouzaimi, H., Laroche, R., et Lefevre, F. (2016). *Reinforcement Learning for Turn-Taking Management in Incremental Spoken Dialogue Systems*. In *Proceedings of International Joint Conference on Artificial Intelligence (IJCAI)*. Consulté à l'adresse : <https://pdfs.semanticscholar.org/0e50/8c6b9a9df45691346981dbb481b40c934082.pdf>

Kidd, C. D., et Breazeal, C. (2004). *Effect of a robot on user perceptions*. In *Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on* (Vol. 4, p. 3559–3564). IEEE.

Kidd, C. D., Taggart, W., et Turkle, S. (2006). *A sociable robot to encourage social interaction among the elderly*. In *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on* (p. 3972–3976). IEEE.

Kidd, R. F., et Berkowitz, L. (1976). *Effect of dissonance arousal on helpfulness*. *Journal of Personality and Social Psychology*, 33(5), 613.

Kientz, J. A., Patel, S. N., Jones, B., Price, E. D., Mynatt, E. D., et Abowd, G. D. (2008). *The georgia tech aware home*. In *CHI'08 extended abstracts on Human factors in computing systems* (p. 3675–3680). ACM.

Kipp, M. (2001). *Anvil-a generic annotation tool for multimodal dialogue*. In *Seventh European Conference on Speech Communication and Technology*. Consulté à l'adresse : [http://www.mirlab.org/conference\\_papers/International\\_Conference/Eurospeech%202001/papers/page1367.pdf](http://www.mirlab.org/conference_papers/International_Conference/Eurospeech%202001/papers/page1367.pdf)

Kiratli, J. (2006). *Telehealth technologies for monitoring adherence and performance of home exercise programs for persons with spinal cord injury: Tele-exercise*. *Exercise and Recrea.*

Kirouac, G. (2004). *Cognition et émotions*. Imprensa da Universidade de Coimbra, Coimbra University Press.

Kita, S. (2009). *Two-dimensional semantic analysis of Japanese mimetics*. *Linguistics*, 35(2), pp. 379-416. Retrieved 27 Nov. 2017, from doi:10.1515/ling.1997.35.2.379

Kitano, N. (2006). « *Rinri* »: *An Incitement towards the Existence of Robots in Japanese Society*. *Ethics in Robotics*, 6, 78.

Kivik, P.-K. (1998). *What silence says: Communicative style and identity*. *Trames*, 2(52/47), 1.

Klewitz, G., et Couper-Kuhlen, E. (1999). *Quote-Unquote?: The role of prosody in the contextualization of reported speech sequences*. *Pragmatics* 9 (4): 459–485. DOI: 10.1075/prag.9.4.03kle.

Knapp, M. L., Hall, J. A., et Horgan, T. G. (2013). *Nonverbal communication in human interaction*. Cengage Learning.

Köhler, W. (1929). *Gestalt psychology*. New York: H. Liveright.

Kosloff, S., et Greenberg, J. (2006). *Android science by all means, but let's be canny about It!* *Interaction Studiesinteraction Studies Social Behaviour and Communication in Biological and Artificial Systems* 7 (3):343-346. DOI: 10.1075/is.7.3.06kos.

- Krahmer, E., et Swerts, M. (2009). *Audiovisual prosody—introduction to the special issue*. SAGE Publications Sage UK: London, England. Consulté à l'adresse : <http://journals.sagepub.com/doi/pdf/10.1177/0023830909103164>
- Krebs, D. (1975). *Empathy and altruism*. *Journal of personality and social psychology*, 32(6), 1134.
- Krebs, D. L. (1970). *Altruism: An examination of the concept and a review of the literature*. *Psychological bulletin*, 73(4), 258.
- Krebs, H. I., Volpe, B. T., Ferraro, M., Fasoli, S., Palazzolo, J., Rohrer, B., ... Hogan, N. (2002). *Robot-aided neurorehabilitation: from evidence-based to science-based rehabilitation*. *Topics in stroke rehabilitation*, 8(4), 54–70.
- Krebs, J. R. (1984). *Animal signals: mind-reading and manipulation*. In *Behavioural Ecology: an evolutionary approach* (p. 380-402).
- Kreiman, J., et Sidtis, D. (2011). *Foundations of voice studies: An interdisciplinary approach to voice production and perception*. John Wiley et Sons.
- Kreiman, J., Gerratt, B. R., et Antonanzas-Barroso, N. (2007). *Measures of the glottal source spectrum*. *Journal of speech, language, and hearing research*, 50(3), 595–610.
- Kuhn, T. S., et Meyer, L. (2008). *La structure des révolutions scientifiques*. Paris: Flammarion.
- Kummer, H. (1967). *Tripartite relations in hamadryas baboons*. *Social communication among primates*, 63–71.
- Kurzweil, R. (2005). *Human life: the next generation*. *New Scientist*, 24, 32–7.
- Kurzweil, R. (2013). *How to create a mind: The secret of human thought revealed*. Penguin.
- L'énigme de l'émergence: comment comprendre l'apparition spontanée de formes naturelles sans invoquer un ordre caché ou une force occulte ?* (2005, août). *Sciences et avenir*. Hors-série, (143), 83.
- Labov, W. (1972). *Some principles of linguistic methodology*. *Language in society*, 1(1), 97–120.
- Labov, W. (1973). *The linguistic consequences of being a lame*. *Language in society*, 2(1), 81–115.
- Lacheret-Dujour, A., et Beaugendre, F. (1998). *La Prosodie du français*. Paris: CNRS Editions.
- Ladd, D. R. (1983). *Phonological features of intonational peaks*. *Language*, 721–759.
- Ladd, D. R., Silverman, K. E., Tolkmitt, F., Bergmann, G., et Scherer, K. R. (1985). *Evidence for the independent function of intonation contour type, voice quality, and F 0 range in signaling speaker affect*. *The Journal of the Acoustical Society of America*, 78(2), 435–444.
- Ladefoged, P., et Maddieson, I. (1998). *The sounds of the world's languages*. *Language*, 74(2), 374–376.
- Landragin, F. (2013). *Dialogue homme-machine: conception et enjeux*. Hermes science publications. Consulté à l'adresse : [http://fred.landragin.free.fr/publi/13\\_HERMES.pdf](http://fred.landragin.free.fr/publi/13_HERMES.pdf)
- Laroche, R. (2015). *Content finder AssistanT*. In 2015 18th International Conference on Intelligence in Next Generation Networks (p. 231-238). <https://doi.org/10.1109/ICIN.2015.7073836>
- Lasseter, J. (2001). *Tricks to Animating Characters with a Computer*. *SIGGRAPH Comput. Graph.*, 35(2), 45–47. <https://doi.org/10.1145/563693.563706>
- Laver, J. (1980). *The phonetic description of voice quality*. *Cambridge Studies in Linguistics* London, 31, 1–186.
- Laver, J. (1991). *The gift of speech: papers in the analysis of speech and voice*. Columbia University Press.
- Lazarus, R.S. (1966). *Psychological stress and the coping process*. New York: McGraw-Hill.

- Lazarus, R. S. (1991). *Emotion and adaptation*. Oxford University Press on Demand.
- Lecouteux, B., Vacher, M., et Portet, F. (2011). *Distant speech recognition in a smart home: Comparison of several multisource ASRs in realistic conditions*. In *Interspeech 2011 Florence* (p. 2273–2276). Consulté à l'adresse : <https://hal.archives-ouvertes.fr/hal-00642306/>
- Legerstee, M. (1990). *Infants use multimodal information to imitate speech sounds*. *Infant behavior and development*, 13(3), 343–354.
- Lehmann, J., Lalmas, M., Yom-Tov, E., et Dupret, G. (2012). *Models of user engagement*. *User Modeling, Adaptation, and Personalization*, 164–175.
- Leibniz, G. W. (1973). *Bref essai sur l'origine des peuples déduite principalement des indications fournies par les langues*. Trad. de J. Sukada, in *JACOB 1973*, texte 1, p. 46-62.
- Lemak, A. (2012). *Silence, intercultural conversation, and miscommunication*. Consulté à l'adresse : <https://tspace.library.utoronto.ca/handle/1807/33661>
- Lemon, O., et Pietquin, O. (2007). *Machine learning for spoken dialogue systems*. In *European Conference on Speech Communication and Technologies (Interspeech'07)* (p. 2685–2688). Consulté à l'adresse : <https://hal.archives-ouvertes.fr/hal-00216035/>
- Léon, P. (1993). *Précis de phonostylistique: parole et expressivité*, coll. Fac Linguistique, Paris: Nathan.
- Les tics de langage, ces petits mots-moteurs*. (2011, novembre 2). Consulté 26 juillet 2017, à l'adresse [http://www.francetvinfo.fr/replay-radio/modes-de-vie/les-tics-de-langage-ces-petits-mots-moteurs\\_1723295.html](http://www.francetvinfo.fr/replay-radio/modes-de-vie/les-tics-de-langage-ces-petits-mots-moteurs_1723295.html)
- Leslie, A. M. (1995). *A theory of agency*. *Causal cognition: A multidisciplinary debate*, 121–141.
- Lestel, D. (1994). *Symbols of discord: are apes that talk trivia more interesting than apes that don't talk at all?* *Social Science Information*, 33(2), 335-369.  
<https://doi.org/10.1177/053901894033002010>
- Levenshtein, V. I. (1966). *Binary codes capable of correcting deletions, insertions, and reversals*. In *Soviet physics doklady* (Vol. 10, p. 707–710). Consulté à l'adresse : <https://nymity.ch/sybilhunting/pdf/Levenshtein1966a.pdf>
- Lévi-Strauss, C. (1983). *Histoire et ethnologie*. *Annales*, 1217–1231.
- Lévi-Strauss, C. (2002). *Le totémisme aujourd'hui*. Presses universitaires de France.
- Levin, E., et Pieraccini, R. (1997). *A stochastic model of computer-human interaction for learning dialogue strategies*. In *Eurospeech* (Vol. 97, p. 1883–1886). Consulté à l'adresse : [https://www.researchgate.net/profile/Esther\\_Levin/publication/2581863\\_A\\_Stochastic\\_Model\\_Of\\_Computer-Human\\_Interaction\\_For\\_Learning\\_Dialogue\\_Strategies/links/0c960528a5bae6e272000000.pdf](https://www.researchgate.net/profile/Esther_Levin/publication/2581863_A_Stochastic_Model_Of_Computer-Human_Interaction_For_Learning_Dialogue_Strategies/links/0c960528a5bae6e272000000.pdf)
- Lewis, M., Haviland-Jones, J. M., et Barrett, L. F. (2010). *Handbook of emotions*. Guilford Press.
- Lheureux-Davidse, C. (s. d.). *Émergences du langage verbal chez des enfants autistes*. *Perspectives Psy*, 45(3), 226-230.
- Li, C. (2005). *A cognitive-pragmatic account of interjections*. *US-China Foreign Language*, 3(9), 65–70.
- Lorenz, K. (1985). *My family and other animals*. *Leaders in the study of animal behavior: Autobiographical perspectives*, 259–87.
- Lorenz, K., et Jospin, L. (2007). *Évolution et modification du comportement : L'inné et l'acquis*. Paris: Payot.
- Loyau, F., Abergé, V., et Vanpé, A. (2006). *Expressions hors des tours de parole: éthogrammes du «feeling of thinking»*. In *Proceedings of JEP*. Consulté à l'adresse : <http://jep2006.irisa.fr/openconf/author/final/final-168.pdf>

- Loyau, F., et Aubergé, V. (2006). *Expressions outside the talk turn: ethograms of the feeling of thinking*. In 5th LREC (p. 47–50). Consulté à l'adresse : <http://lrec.elra.info/proceedings/lrec2006/workshops/W09/Emotion-proceeding.pdf#page=56>
- Lu, Y., Aubergé, V., Rilliard, A., et Gu, W. (2013). *Perceptual Study for Mandarin Attitudinal Speech*. Journal of School of Chinese Language and Culture Nanjing Normal University, 3, 39.
- Luo, Y., Hawkey, L. C., Waite, L. J., et Cacioppo, J. T. (2012). *Loneliness, health, and mortality in old age: A national longitudinal study*. Social Science et Medicine, 74(6), 907-914. <https://doi.org/10.1016/j.socscimed.2011.11.028>
- Luzzati, D. (1995). *De l'erreur en dialogue homme-machine*. Cahiers de linguistique française, (16), 175–191.
- Luzzati, D. (2007). *Le dialogue oral spontané: quels objets pour quels corpora*. Revue d'Interaction Homme-Machine, 8(2). Consulté à l'adresse : <http://www.europia.org/RIHM/V8N2/2-RIHM-Luzzati%20PDF.pdf>
- Mac, D.-K., Aubergé, V., Castelli, E., et Rilliard, A. (2012). *Local vs. global prosodic cues: effect of tones on attitudinal prosody in cross-perception of Vietnamese by French*. In Speech Prosody 2012. Consulté à l'adresse : <https://pdfs.semanticscholar.org/503a/2ed3ced4f3e8515b0ebfb5f5af4342a2df6b.pdf>
- MacDorman, K. F., et Ishiguro, H. (2006). *The uncanny advantage of using androids in cognitive and social science research*. Interaction Studies, 7(3), 297–337.
- Maclay, H., et Osgood, C. E. (1959). *Hesitation phenomena in spontaneous English speech*. Word, 15(1), 19–44.
- Maekawa, K., Kikuchi, H., Igarashi, Y., et Venditti, J. J. (2002). *X-JToBI: an extended j-toBI for spontaneous speech*. In INTERSPEECH. Consulté à l'adresse : <https://pdfs.semanticscholar.org/75ff/6f614a4ae4f45b3ec75f7c1cb822319246ff.pdf>
- Magnani, R., Aubergé, V., Bayol, C., et Sasa, Y. (2017). *Bases of Empathic Animism Illusion: audio-visual perception of an object devoted to becoming perceived as a subject for HRI*. In VIHAR 2017, the 1st International Workshop on Vocal Interactivity in-and-between Humans, Animals and Robots, Stockholm, Sweden.
- Mairesse, F., Walker, M. A., Mehl, M. R., et Moore, R. K. (2007). *Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text*. J. Artif. Intell. Res.(JAIR), 30, 457–500.
- Malle, B. F., et Pearce, G. E. (2001). *Attention to behavioral events during interaction: Two actor-observer gaps and three attempts to close them*. Journal of personality and social psychology, 81(2), 278–294.
- Marie Chan, E. C. (2005). *Assessment of activity of elderly people using a home monitoring system*. International journal of rehabilitation research. Internationale Zeitschrift für Rehabilitationsforschung. Revue internationale de recherches de réadaptation, 28(1), 69-76. <https://doi.org/10.1097/00004356-200503000-00010>
- Marr, D., et Vision, A. (1982). *A computational investigation into the human representation and processing of visual information*. WH San Francisco: Freeman and Company, 1(2). Consulté à l'adresse : <https://mitpress.mit.edu/books/978-0-262-29037-1>
- Marti, P., Bacigalupo, M., Giusti, L., Mennecozzi, C., et Shibata, T. (2006). *Socially assistive robotics in the treatment of behavioural and psychological symptoms of dementia*. In Biomedical Robotics and Biomechatronics, 2006. BioRob 2006. The First IEEE/RAS-EMBS International Conference on (p. 483–488). IEEE.
- Martin, G. B., et Clark, R. D. (1982). *Distress crying in neonates: Species and peer specificity*. Developmental psychology, 18(1), 3.
- Martin, P. (1996). *WinPitch: un logiciel d'analyse temps réel de la fréquence fondamentale fonctionnant sous Windows*. In XXI Journées d'Etude sur la Parole (p. 224–227). Consulté à l'adresse : <https://halshs.archives-ouvertes.fr/halshs-01450486/>

- Massaro, D. W., et Egan, P. B. (1996). *Perceiving affect from the voice and the face*. *Psychonomic Bulletin et Review*, 3(2), 215–221.
- Matsumoto, D. (1990). *Cultural similarities and differences in display rules*. *Motivation and emotion*, 14(3), 195–214.
- Matsumoto, D. (2006). *Are cultural differences in emotion regulation mediated by personality traits?* *Journal of Cross-Cultural Psychology*, 37(4), 421–437.
- Matsumoto, D. (2007). *Culture, context, and behavior*. *Journal of personality*, 75(6), 1285–1320.
- Matsuyama, Y., Akiba, I., Fujie, S., et Kobayashi, T. (2015). *Four-participant group conversation: A facilitation robot controlling engagement density as the fourth participant*. *Computer Speech et Language*, 33(1), 1-24. <https://doi.org/10.1016/j.csl.2014.12.001>
- Mayr, S., et Buchner, A. (2007). *Negative priming as a memory phenomenon*. *Journal of Psychology*, 215(1), 35–51.
- McCrae, R. R., et John, O. P. (1992). *An introduction to the five-factor model and its applications*. *Journal of personality*, 60(2), 175–215.
- McCulloch, W. S., et Pitts, W. (1943). *A logical calculus of the ideas immanent in nervous activity*. *The bulletin of mathematical biophysics*, 5(4), 115–133.
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago press.
- Mead, G. H. (1934). *Mind, self and society* (Vol. 111). Chicago University of Chicago Press.
- Mehrabian, A., et Ferris, S. R. (1967). *Inference of attitudes from nonverbal communication in two channels*. *Journal of consulting psychology*, 31(3), 248.
- Meltzoff, A. N., et Moore, M. K. (1997). *Explaining facial imitation: A theoretical model*. *Early development et parenting*, 6(3-4), 179.
- Meltzoff, A. N., et Moore, M. K. (2005). *Imitation et développement humain : les premiers temps de la vie*. *Terrain. Anthropologie et sciences humaines*, (44), 71-90. <https://doi.org/10.4000/terrain.2455>
- Meltzoff, A. N., et Prinz, W. (2002). *The imitative mind: Development, evolution and brain bases* (Vol. 6). Cambridge University Press.
- Merleau-Ponty, M. (1976). *Phénoménologie de la perception*. Paris: Gallimard.
- Mesman, J., van IJzendoorn, M. H., et Bakermans-Kranenburg, M. J. (2009). *The many faces of the Still-Face Paradigm: A review and meta-analysis*. *Developmental Review*, 29(2), 120–162.
- Michalowski, M. P., Sabanovic, S., et Simmons, R. (2006). *A spatial model of engagement for a social robot*. In *Advanced Motion Control, 2006. 9th IEEE International Workshop on* (p. 762–767). IEEE.
- Mikhak, B., Lyon, C., Gorton, T., Gershenfeld, N., McEnnis, C., et Taylor, J. (2002). *Fab Lab: an alternate model of ICT for development*. In *2nd international conference on open collaborative design for sustainable innovation*. Consulté à l'adresse : <http://18.85.8.56/events/03.05.fablab/fablab-dyd02.pdf>
- Miller, K. W. (2010). *It's not nice to fool humans*. *IT Professional Magazine*, 12(1), 51.
- Millet, J., Digard, J.-P., et Mezger, A.-C. (1989). *Une enquête sur les nouvelles sensibilités à l'égard des animaux domestiques*. *Anthropozoologica*, (10), 31–47.
- Minsky, M. L., et Papert, S. A. (1969). *Perceptrons*. MIT Press, Cambridge, MA.
- Mischel, W., Shoda, Y., et Rodriguez, M. L. (1989). *Delay of gratification in children*. *Science*, 244(4907), 933–938.
- Missonnier, S. (2014). *De l'Einführung à l'énonciation métaphorisante, dans les consultations thérapeutiques de Serge Lebovici*. *Journal de la psychanalyse de l'enfant*, 3(2), 85-130.
- Mitchell, G., et O'Donnell, H. (2013). *The therapeutic use of doll therapy in dementia*. *British Journal of Nursing* (Mark Allen Publishing), 22(6), 329-334.

- Miyata, M., Ogata, H., Kondo, T., et Yano, Y. (2008). *JAMIOLAS 2.0: Supporting to learn Japanese mimetic words and onomatopoeia with wireless sensor networks*. In Proceedings of the International Conference on Computer in Education, Taipei, Taiwan (p. 643–650). Consulté à l'adresse : <http://www.academia.edu/download/30837499/ICCE2008-paper127.pdf>
- Mondada, L. (2005). *La constitution de l'origo déictique comme travail interactionnel des participants : une approche praxéologique de la spatialité*. *Intellectica – La revue de l'Association pour la Recherche sur les sciences de la Cognition (ARCo)*, 41-42(2/3), 75-100.
- Mondada, L. (2007). *Multimodal resources for turn-taking: Pointing and the emergence of possible next speakers*. *Discourse studies*, 9(2), 194–225.
- Mondada, L. (2009). *Emergent focused interactions in public places: A systematic analysis of the multimodal achievement of a common interactional space*. *Journal of Pragmatics*, 41(10), 1977-1997. <https://doi.org/10.1016/j.pragma.2008.09.019>
- Mondada, L. (2012). *Organisation multimodale de la parole-en-interaction : pratiques incarnées d'introduction des référents, The multimodal organization of talk-in-interaction : embodied practices for the introduction of new referents*. *Langue française*, (175), 129-147.
- Moore, C., et Barresi, J. (1993). *Knowledge of the psychological states of self and others is not only theory-laden but also data-driven*. *Behavioral and Brain Sciences*, 16(1), 61–62.
- Moore, R. K. (2012). *A Bayesian explanation of the 'Uncanny Valley' effect and related psychological phenomena*. *Scientific reports*, 2, 864.
- Morency, L. (2010). *Modeling human communication dynamics*. *Signal Processing Magazine, IEEE*, 27(5), 112–116.
- Morlec, Y., Bailly, G., et Aubergé, V. (2001). *Generating prosodic attitudes in French: data, model and evaluation*. *Speech Communication*, 33(4), 357–371.
- Moscovici, S. (2015). *Psychologie sociale des relations à autrui*. Nathan.
- Munhall, K. G., Jones, J. A., Callan, D. E., Kuratate, T., et Vatikiotis-Bateson, E. (2004). *Visual prosody and speech intelligibility: Head movement improves auditory speech perception*. *Psychological science*, 15(2), 133–137.
- Murphy, F. C., Nimmo-Smith, I. A. N., et Lawrence, A. D. (2003). *Functional neuroanatomy of emotions: a meta-analysis*. *Cognitive, Affective, et Behavioral Neuroscience*, 3(3), 207–233.
- Murray, L., et Trevarthen, C. (1986). *The infant's role in mother–infant communications*. *Journal of child language*, 13(1), 15–29.
- Murray, L., Kempton, C., Woolgar, M., et Hooper, R. (1993). *Depressed mothers' speech to their infants and its relation to infant gender and cognitive development*. *Journal of Child Psychology and Psychiatry*, 34(7), 1083–1101.
- Murstein, B. I., et Pryer, R. S. (1959). *The concept of projection: A review*. *Psychological Bulletin*, 56(5), 353.
- Myers, D. G., et Diener, E. (1995). *Who is happy?* *Psychological science*, 6(1), 10–19.
- Myers, I. B., McCaulley, M. H., et Most, R. (1985). *Manual, a guide to the development and use of the Myers-Briggs type indicator*. Consulting Psychologists Press.
- Nadel, J., et Decety, J. (2002). *Imiter pour découvrir l'humain*. Paris (6, avenue Reille 75685): Presses Universitaires de France – PUF.
- Nakane, I. (2006). *Silence and politeness in intercultural communication in university seminars*. *Journal of pragmatics*, 38(11), 1811–1835.
- Nass, C., et Brave, S. (2007). *Wired for Speech – How Voice Activates and Advances the Human–Computer Relationship* (New Ed). Cambridge Mass.: MIT Press.
- Neri, A. L., Yassuda, M. S., Fortes-Burgos, A. C. G., Mantovani, E. P., Arbex, F. S., de Souza Torres, S. V., ... Guariento, M. E. (2012). *Relationships between gender, age, family conditions, physical and mental health, and social isolation of elderly caregivers*. *International Psychogeriatrics*, 24(3), 472–483.

- Neumann, E., et DeSchepper, B. G. (1991). *Costs and benefits of target activation and distractor inhibition in selective attention*. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(6), 1136.
- Newell, A. F. (1987). *Speech simulation studies—performance and dialogue specification*. In Unicorn Seminar, London.
- Newell, A., et Simon, H. (1956). *The logic theory machine—A complex information processing system*. *IRE Transactions on information theory*, 2(3), 61–79.
- Newell, A., Shaw, J. C., et Simon, H. A. (1959). *A general problem-solving program for a computer*. *Computers and Automation*, 8(7), 10–16.
- Newman, J. D., et Symmes, D. (1982). *Inheritance and experience in the acquisition of primate acoustic behavior*. *Primate communication*, 259-278.
- Nielsen, J. (1994). *Usability laboratories*. *Behaviour et Information Technology*, 13(1-2), 3–8.
- Nilsson, N. J. (2005). *Human-level artificial intelligence? Be serious!* *AI magazine*, 26(4), 68.
- Norman, D. (2004). *Affordances and design*. Unpublished article, available online at: <http://www.jnd.org/dn.mss/affordances-and-design.html>.
- Noury, N., Virone, G., Barralon, P., Rialle, V., et Demongeot, J. (2004). *Maisons intelligentes pour personnes âgées: technologies de l'information intégrées au service des soins à domicile*. *J3eA*, 3, 20.
- Nummenmaa, L., et Calder, A. J. (2009). *Neural mechanisms of social attention*. *Trends in Cognitive Sciences*, 13(3), 135-143. <https://doi.org/10.1016/j.tics.2008.12.006>
- Nwokah, E. E. (1987). *Maidese Versus Motherese—Is The Language Input of Child and Adult Caregivers Similar?* *Language and Speech*, 30(3), 213–237.
- Ogawa, N., et Retherford, R. D. (1993). *Care of the elderly in Japan: Changing norms and expectations*. *Journal of Marriage and the Family*, 585–597.
- Ohala, J. J. (1994). *The frequency code underlies the sound symbolic use of voice pitch*. *Sound symbolism*, ed. by Leanne Hinton, Johanna Nichols, and John J. Ohala, 325–47. Cambridge: Cambridge University Press.
- Ohala, J. J. (1996). *Ethological theory and the expression of emotion in the voice*. In *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on (Vol. 3, p. 1812–1815)*. IEEE.
- Olsen-Fulero, L. (1983). *Informational functions of mother–child discourse: Knowing them when we see them*. *Journal of Child Language*, 10(1), 223-229. <http://dx.doi.org/10.1017/S0305000900005250>
- Orth-Gomer, K., et Johnson, J. V. (1987). *Social network interaction and mortality: a six year follow-up study of a random sample of the Swedish population*. *Journal of chronic diseases*, 40(10), 949–957.
- Osgood, C. E., May, W. H., et Miron, M. S. (1975). *Cross-cultural universals of affective meaning (Vol. 1)*. University of Illinois Press.
- Oviatt, S. (2003). *Multimodal interfaces*. *The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications*, 14, 286–304.
- Pacherie, E. (2004). *L'empathie et ses degrés*. Editions Odile Jacob. Consulté à l'adresse : [https://hal-ens.archives-ouvertes.fr/ijn\\_00353957/](https://hal-ens.archives-ouvertes.fr/ijn_00353957/)
- Pan Ké Shon, J. L. (2003). *Isolement relationnel et mal-être*. INSEE première, 931, 1–4.
- Pantelis, P. C., et Feldman, J. (2012). *Exploring the mental space of autonomous intentional agents*. *Attention, Perception, et Psychophysics*, 74(1), 239-249. <https://doi.org/10.3758/s13414-011-0215-6>
- Patterson, M. (1968). *Spatial factors in social interactions*. *Human Relations*, 21(4), 351–361.

- Pelachaud, C. (2005). *Multimodal expressive embodied conversational agents*. In Proceedings of the 13th annual ACM international conference on Multimedia (p. 683–689). ACM.
- Pelachaud, C., Badler, N. I., et Steedman, M. (1996). *Generating facial expressions for speech*. *Cognitive science*, 20(1), 1–46.
- Pellissier, J. (2003). *La nuit, tous les vieux sont gris: la société contre la vieillesse*. Bibliophane-Daniel Radford.
- Pentland, A. (2005). *Socially aware, computation and communication*. *Computer*, 38(3), 33–40.
- Pettersson, J. S., et Wik, M. (2014). *Perspectives on Ozlab in the cloud: A literature review of tools supporting Wizard-of-Oz experimentation, including an historical overview of 1971-2013 and notes on methodological issues and supporting generic tools*. Karlstads universitet.
- Pfeifer, R. (2002). *On the role of embodiment in the emergence of cognition: Grey walter's turtles and beyond*. In Proc. of the Workshop The Legacy of Grey Walter. Consulté à l'adresse : [https://www.researchgate.net/profile/Rolf\\_Pfeifer/publication/228784049\\_On\\_the\\_role\\_of\\_embodiment\\_in\\_the\\_emergence\\_of\\_cognition\\_Grey\\_walter%27s\\_turtles\\_and\\_beyond/links/570e20ff08aec783ddd0916f/On-the-role-of-embodiment-in-the-emergence-of-cognition-Grey-walters-turtles-and-beyond.pdf](https://www.researchgate.net/profile/Rolf_Pfeifer/publication/228784049_On_the_role_of_embodiment_in_the_emergence_of_cognition_Grey_walter%27s_turtles_and_beyond/links/570e20ff08aec783ddd0916f/On-the-role-of-embodiment-in-the-emergence-of-cognition-Grey-walters-turtles-and-beyond.pdf)
- Phan, K. L., Wager, T., Taylor, S. F., et Liberzon, I. (2002). Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI. *Neuroimage*, 16(2), 331–348.
- Piaget, J. (2013). *The construction of reality in the child* (Vol. 82). Routledge.
- Piaget, J., et Cook, M. (1952). *The origins of intelligence in children* (Vol. 8). International Universities Press New York. Consulté à l'adresse : [http://www.bxscience.edu/ourpages/auto/2014/11/16/50007779/Piaget%20When%20Thinkin%20Begins10272012\\_0000.pdf](http://www.bxscience.edu/ourpages/auto/2014/11/16/50007779/Piaget%20When%20Thinkin%20Begins10272012_0000.pdf)
- Picard, D. (1948-). (1983). *Du code au désir : le corps dans la relation sociale*. Dunod. [Paris].
- Picard, R. W., et Picard, R. (1997). *Affective computing* (Vol. 252). MIT press Cambridge. Consulté à l'adresse : [http://www.cell.com/trends/cognitive-sciences/pdf/S1364-6613\(98\)01190-5.pdf](http://www.cell.com/trends/cognitive-sciences/pdf/S1364-6613(98)01190-5.pdf)
- Pietquin, O., et Hastie, H. (2013). *A survey on metrics for the evaluation of user simulations*. *The knowledge engineering review*, 28(1), 59–73.
- Piliavin, J. A., et Piliavin, I. M. (1972). *Effect of blood on reactions to a victim*. *Journal of Personality and Social Psychology*, 23(3), 353-361. <http://dx.doi.org/10.1037/h0033166>
- Pino, M., Faucounau, V., Wu, Y. H., Boulay, M., Riguet, M., Sant'Anna, M. de, ... et coll. (2009). *The LUSAGE Usability laboratory for elderly people with cognitive impairment*. *Gerontechnology*, 8(3), 185.
- Pinquart, M., et Sorensen, S. (2001). *Influences on loneliness in older adults: A meta-analysis*. *Basic and applied social psychology*, 23(4), 245–266.
- Pitaud, P. (2013). *Solitude et isolement des personnes âgées*. Eres.
- Pittenger, R. E., et Smith Jr, H. L. (1957). *A basis for some contributions of linguistics to psychiatry*. *Psychiatry*, 20(1), 61–78.
- Planells, J., Hurtado, L. F., Segarra, E., et Sanchis, E. (2013). *A multi-domain dialog system to integrate heterogeneous spoken dialog systems*. In INTERSPEECH (p. 1891–1895). Consulté à l'adresse : [http://users.dsic.upv.es/~lhurtado/papers/pdfs/2013\\_planells13\\_interspeech.pdf](http://users.dsic.upv.es/~lhurtado/papers/pdfs/2013_planells13_interspeech.pdf)
- Pléty, R. (1993). *Ethologie des communications humaines: aide-mémoire méthodologique*. Presses Universitaires Lyon.

- Plutchik, R. (1980). *A general psychoevolutionary theory of emotion*. Theories of emotion, 1(3-31), 4.
- Poggi, I. (2001). *Mind markers*. The Semantics and Pragmatics of Everyday Gestures. Berlin Verlag Arno Spitz. Consulté à l'adresse : <http://host.uniroma3.it/docenti/poggi/cursitopdf/MIMPORTODEF.pdf>
- Poggi, I. (2009). *The language of interjections*. In Multimodal Signals: Cognitive and Algorithmic Issues (p. 170–186). Springer.
- Pollack, M. E. (2005). *Intelligent technology for an aging population: The use of AI to assist elders with cognitive impairment*. AI magazine, 26(2), 9.
- Pollick, F. E. (2009). *In search of the uncanny valley*. In International Conference on User Centric Media (p. 69–78). Springer.
- Pooley, T. (1996). *Chtimi: The urban vernaculars of northern France* (Vol. 2). Multilingual Matters.
- Popper, K. R. (1980). *La connaissance objective*. Consulté à l'adresse : <https://philpapers.org/rec/POPLCO>
- Portet, F., Vacher, M., et Rossato, S. (2012). *Les technologies de la parole et du TALN pour l'assistance à domicile des personnes âgées: un rapide tour d'horizon*. In JEP-TALN-RECITAL 2012, Atelier ILADI 2012: Interactions Langagieres pour personnes Agées Dans les habitats Intelligents (p. 3–16). Consulté à l'adresse : <https://hal.inria.fr/hal-00953514/>
- Portet, F., Vacher, M., Golanski, C., Roux, C., et Meillon, B. (2013). *Design and evaluation of a smart home voice interface for the elderly: acceptability and objection aspects*. Personal and Ubiquitous Computing, 17(1), 127–144.
- Posner, M. I. (Éd.). (1993). *The Foundations of Cognitive Science*. Cambridge, Mass: A Bradford Book.
- Post, B. M. B. (2000). *Tonal and phrasal structures in French intonation* (Vol. 34). Thesus The Hague. Consulté à l'adresse : [http://www.lotpublications.nl/Documents/34\\_cover.pdf](http://www.lotpublications.nl/Documents/34_cover.pdf)
- Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., ... coll. (2011). *The Kaldi speech recognition toolkit*. In IEEE 2011 workshop on automatic speech recognition and understanding. IEEE Signal Processing Society. Consulté à l'adresse : <http://infoscience.epfl.ch/record/192584>
- Pratt, J., Radulescu, P. V., Guo, R. M., et Abrams, R. A. (2010). *It's Alive!: Animate Motion Captures Visual Attention*. Psychological Science, 21(11), 1724-1730. <https://doi.org/10.1177/0956797610387440>
- Premack, D. (1990). The infant's theory of self-propelled objects. Cognition, 36(1), 1-16. [https://doi.org/10.1016/0010-0277\(90\)90051-K](https://doi.org/10.1016/0010-0277(90)90051-K)
- Preston, S. D., et de Waal, F. B. M. (2002). *Empathy: Its ultimate and proximate bases*. Behavioral and Brain Sciences, 25(1), 1-20. <https://doi.org/10.1017/S0140525X02000018>
- Psathas, G. (1995). *Conversation analysis: The study of talk-in-interaction* (Vol. 35). Sage.
- Raatikainen, P. (2013). *Gödel's Incompleteness Theorems*. Consulté à l'adresse : <https://stanford.library.sydney.edu.au/entries/goedel-incompleteness/>
- Ramachandran, V. S., et Hubbard, E. M. (2001). *Synaesthesia—a window into perception, thought and language*. Journal of consciousness studies, 8(12), 3–34.
- Ramey, C. H. (2005). *The uncanny valley of similarities concerning abortion, baldness, heaps of sand, and humanlike robots*. In Proceedings of views of the uncanny valley workshop: IEEE-RAS international conference on humanoid robots (p. 8–13).
- Rauscher, W. V. (2002). *Mind Readers: Masters of Deception* (First edition. edition). Mystic Light Press.

- Rauterberg, M., Dätwyler, M., et Sperisen, M. (1995). *From competition to collaboration through a shared social space*. In Proc. of East–West Intern. Conf. on Human–Computer Interaction (EWHCI'95) (p. 94–101). Consulté à l'adresse : [https://www.researchgate.net/profile/Matthias\\_Rauterberg/publication/243777334\\_From\\_Competition\\_to\\_Collaboration\\_through\\_a\\_Shared\\_Social\\_Space/links/55b6570108ae092e9655d949.pdf](https://www.researchgate.net/profile/Matthias_Rauterberg/publication/243777334_From_Competition_to_Collaboration_through_a_Shared_Social_Space/links/55b6570108ae092e9655d949.pdf)
- Raux, A., et Eskenazi, M. (2012). *Optimizing the turn-taking behavior of task-oriented spoken dialog systems*. ACM Transactions on Speech and Language Processing (TSLP), 9(1), 1.
- Reggia, J. A. (2013). *The rise of machine consciousness: Studying consciousness with computational models*. Neural Networks, 44, 112–131.
- Relieu, M. (1999). Parler en marchant. *Pour une écologie dynamique des échanges de paroles*. Langage et société, 89(1), 37-67. <https://doi.org/10.3406/lsoc.1999.2883>
- Renneville, M. (2000). Le langage des crânes. Empêcheurs de penser en rond. Consulté à l'adresse : <https://halshs.archives-ouvertes.fr/halshs-00007229>
- Rich, C., Ponsler, B., Holroyd, A., et Sidner, C. L. (2010). *Recognizing engagement in human-robot interaction*. In Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on (p. 375–382). IEEE.
- Richmond, V. P., McCroskey, J. C., et Payne, S. K. (1991). *Nonverbal behavior in interpersonal relations*. Prentice Hall Englewood Cliffs, NJ. Consulté à l'adresse : [http://amberton.mylifeblue.com/media/Syllabi/Fall%202014/Undergraduate/COM4407\\_01.pdf](http://amberton.mylifeblue.com/media/Syllabi/Fall%202014/Undergraduate/COM4407_01.pdf)
- Riek, L. D. (2012). *Wizard of oz studies in hri: a systematic review and new reporting guidelines*. Journal of Human-Robot Interaction, 1(1).
- Rigaud, A.-S., Pino, M., Wu, Y.-H., De Rotrou, J., Boulay, M., Seux, M.-L., ... et coll. (2011). *L'aide aux personnes souffrant de maladie d'Alzheimer et à leurs aidants par les gérontechnologies*. Gériatrie et Psychologie Neuropsychiatrie du Vieillissement, 9(1), 91–100.
- Rizzolatti, G., Fadiga, L., Fogassi, L., et Gallese, V. (2002). *From mirror neurons to imitation: facts and speculations*. The imitative mind: Development, evolution, and brain bases, 6, 14, 247–266.
- Rizzolatti, G., Fogassi, L., et Gallese, V. (2001). *Neurophysiological mechanisms underlying the understanding and imitation of action*. Nature reviews neuroscience, 2(9), 661–670.
- Roach, P. (2000). *Techniques for the phonetic description of emotional speech*. In ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion. Consulté à l'adresse : [http://www.isca-speech.org/archive\\_open/speech\\_emotion/spem\\_053.html](http://www.isca-speech.org/archive_open/speech_emotion/spem_053.html)
- Rosenberg, A. (2010). *AutoBI-a tool for automatic toBI annotation*. In Interspeech (p. 146–149). Consulté à l'adresse : <http://eniac.cs.qc.cuny.edu/andrew/papers/autobi-is10.pdf>
- Rosenblatt, F. (1958). *The perceptron: A probabilistic model for information storage and organization in the brain*. Psychological review, 65(6), 386.
- Rosenbloom, C. A., et Whittington, F. J. (1993). *The effects of bereavement on eating behaviors and nutrient intakes in elderly widowed persons*. Journal of Gerontology, 48(4), S223–S229.
- Rosenfeld, R., et Harris, T. (2004). *A Universal Speech Interface for Appliances*. Computer Science Department. Consulté à l'adresse : <http://repository.cmu.edu/compsci/1301>
- Rosenthal-von der Pütten, A. M., Krämer, N. C., Hoffmann, L., Sobieraj, S., et Eimler, S. C. (2012). *An Experimental Study on Emotional Reactions Towards a Robot*. International Journal of Social Robotics, 5(1), 17-34. <https://doi.org/10.1007/s12369-012-0173-8>
- Ross, L., Greene, D., et House, P. (1977). *The “false consensus effect”: An egocentric bias in social perception and attribution processes*. Journal of experimental social psychology, 13(3), 279–301.

- Rossano, M. J. (2012). *The essential role of ritual in the transmission and reinforcement of social norms*. *Psychological bulletin*, 138(3), 529.
- Rossi, M. (2000). *Intonation: past, present, future*. In *Intonation* (p. 13–52). Springer.
- Ruby, P., et Decety, J. (2004). *How would you feel versus how do you think she would feel? A neuroimaging study of perspective-taking with social emotions*. *Journal of cognitive neuroscience*, 16(6), 988–999.
- Ruesch, J., Bateson, G., Pinsky, E. C., et Combs, G. (2006). *Communication: The Social Matrix of Psychiatry* (1re éd.). New Brunswick: Transaction Publishers.
- Rushton, J. P., Chrisjohn, R. D., et Fekken, G. C. (1981). *The altruistic personality and the self-report altruism scale*. *Personality and individual differences*, 2(4), 293–302.
- Rusko, M., et Hamar, J. (2006). *Character identity expression in vocal performance of traditional puppeteers*. In *Text, Speech and Dialogue* (p. 509–516). Springer.
- Russell, J. A. (2003). *Core affect and the psychological construction of emotion*. *Psychological review*, 110(1), 145.
- Russell, S., et Norvig, P. (2009). *Artificial Intelligence: A Modern Approach* (3 édition). Upper Saddle River, NJ: Pearson.
- Sadanobu, T. (2004). *A natural history of Japanese pressed voice*. *Journal of the Phonetic Society of Japan*, 8(1), 29–44.
- Sagisaka, Y. (1992). ATR v-talk speech synthesis system. In *Proc. ICSLP, 1992*.
- Sagisaka, Y. (2003). *Modeling and perception of temporal characteristics in speech*. In *Proc. ICPhS (Vol. 3, p. 1–6)*. Consulté à l'adresse : [https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2003/papers/p15\\_0001.pdf](https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2003/papers/p15_0001.pdf)
- Sagisaka, Y., Campbell, N., et Higuchi, N. (2012). *Computing prosody: computational models for processing spontaneous speech*. Springer Science et Business Media.
- Sagisaka, Y., et Greenberg, Y. (2015). *Communicative Speech Synthesis as Pan-Linguistic Prosody Control*. In *Speech Prosody in Speech Synthesis: Modeling and generation of prosody for high quality and flexible speech synthesis* (p. 73–82). Springer.
- Sahlins, M. (2005). *The original affluent society*. In M. Sahlins, *Stone Age Economics* Saint-Georges, C., Chetouani, M., Cassel, R., Apicella, F., Mahdhaoui, A., Muratori, F., ... Cohen, D. (2013). *Motherese in Interaction: At the Cross-Road of Emotion and Cognition? (A Systematic Review)*. *PLOS ONE*, 8(10), e78103. <https://doi.org/10.1371/journal.pone.0078103>
- Saito, D. N., Tanabe, H. C., Izuma, K., Hayashi, M. J., Morito, Y., Komeda, H., ... et coll. (2010). *“Stay tuned” : inter-individual neural synchronization during mutual gaze and joint attention*. *Frontiers in integrative neuroscience*, 4.
- Saitō, T. (1998). *Social Withdrawal: Adolescence without End*.
- Samlan, R. A., Story, B. H., et Bunton, K. (2013). *Relation of perceived breathiness to laryngeal kinematics and acoustic measures based on computational modeling*. *Journal of speech, language, and hearing research : JSLHR*, 56(4), 1209-1223. [https://doi.org/10.1044/1092-4388\(2012/12-0194\)](https://doi.org/10.1044/1092-4388(2012/12-0194))
- Sanghvi, J., Castellano, G., Leite, I., Pereira, A., McOwan, P. W., et Paiva, A. (2011). *Automatic analysis of affective postures and body motion to detect engagement with a game companion*. In *Human-Robot Interaction (HRI), 2011 6th ACM/IEEE International Conference on* (p. 305–311). IEEE.
- Sapolsky, R. M., Alberts, S. C., et Altmann, J. (1997). *Hypercortisolism associated with social subordination or social isolation among wild baboons*. *Archives of General Psychiatry*, 54(12), 1137–1143.
- Sasa, Y. (2012). *Micro-expressions audio-visuelles dans la communication expressive: enjeux pluri-culturels (Mémoire de recherche Master 2)* (p. 119). Grenoble, France: Université Stendhal – Grenoble 3. Consulté à l'adresse : <http://dumas.ccsd.cnrs.fr/dumas-00709396>

- Sasa, Y. (2013b). *Audio-Visual micro-expressions within Japanese-French contrast*. WASSS – Workshop on Affective Social Speech Signals, Grenoble, France.
- Sasa, Y. (2013). *EmOz : magicien d'Oz pour les interactions entre personnes âgées un robot compagnon dans l'habitat intelligent Domus* (Mémoire de recherche Master 2) (p. 123). Grenoble, France: Université Stendhal – Grenoble 3.
- Sasa, Y., Aubergé, V., Franck, P., Guillaume, L., et Moujtahid, S. (2012). *Des micro-expressions au service de la macro-communication pour le robot compagnon Emox*. In WACAI 2012 Workshop Affect, Compagnon Artificiel, Interaction (p. 37-45). Grenoble, France. Consulté à l'adresse : [http://rr.liglab.fr/research\\_report/RR-LIG-039\\_orig.pdf](http://rr.liglab.fr/research_report/RR-LIG-039_orig.pdf)
- Sasamoto, R., et Jackson, R. (2016). *Onomatopoeia—Showing-word or Saying-word? Relevance Theory, lexis, and the communication of impressions*. *Lingua*, 175, 36–53.
- Sauter, D. A., Eisner, F., Ekman, P., et Scott, S. K. (2010). *Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations*. *Proceedings of the National Academy of Sciences*, 107(6), 2408–2412.
- Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., et Frith, C. (2012). *The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions*. *Social Cognitive and Affective Neuroscience*, 7(4), 413-422.  
<https://doi.org/10.1093/scan/nsr025>
- Schatzmann, J., Weilhammer, K., Stuttle, M., et Young, S. (2006). *A survey of statistical user simulation techniques for reinforcement-learning of dialogue management strategies*. *The knowledge engineering review*, 21(2), 97–126.
- Schegloff, E. A., et Sacks, H. (1973). *Opening up closings*. *Semiotica*, 8(4), 289–327.
- Scherer, K. R. (1984). *Emotion as a multicomponent process: A model and some cross-cultural data*. *Review of Personality et Social Psychology*, 5, 37–63.
- Scherer, K. R. (1986). *Vocal affect expression: a review and a model for future research*. *Psychological bulletin*, 99(2), 143.
- Scherer, K. R. (1994). *Affect bursts*. *Emotions: Essays on emotion theory*, 161–196.
- Scherer, K. R., Schorr, A., et Johnstone, T. (2001). *Appraisal processes in emotion: Theory, methods, research*. Oxford University Press.
- Scherer, S., Kane, J., Gobl, C., et Schwenker, F. (2013). *Investigating fuzzy-input fuzzy-output support vector machines for robust voice quality classification*. *Computer Speech et Language*, 27(1), 263–287.
- Schlangen, D., et Skantze, G. (2009). *A general, abstract model of incremental dialogue processing*. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics* (p. 710–718). Association for Computational Linguistics.
- Schlegoff, E. A. (1984). *On some gestures' relation to talk* (Cambridge University Press, Vol. Structures of Social Action-(Chapter 12), p. 266-296). J. Maxwell AtkinsonCambridge Core.
- Schlosberg, H. (1954). *Three dimensions of emotion*. *Psychological review*, 61(2), 81.
- Schoenbach, V. J., Kaplan, B. H., Fredman, L., et Kleinbaum, D. G. (1986). *Social ties and mortality in Evans County, Georgia*. *American journal of epidemiology*, 123(4), 577–591.
- Schröder, M. (2003). *Experimental study of affect bursts*. *Speech communication*, 40(1), 99–116.
- Schröder, M., et Cowie, R. (2006). *Issues in emotion-oriented computing-towards a shared understanding*. In *Workshop on Emotion and Computing at KI*. Consulté à l'adresse : <https://pdfs.semanticscholar.org/d971/c0ef272f9718c37f08f5e8f662906e623830.pdf>
- Schröder, M., Heylen, D. K. J., et Poggi, I. (2006). *Perception of non-verbal emotional listener feedback*.
- Schuller, B., et Batliner, A. (2013). *Computational paralinguistics: emotion, affect and personality in speech and language processing*. United Kingdom: Wiley.

- Schumacher, J., et Feurstein, K. (2007). *Living Labs-the user as co-creator*. In Technology Management Conference (ICE), 2007 IEEE International (p. 1–6). IEEE.
- Searle, J. R. (1979). *What Is an Intentional State?* *Mind*, 88(349), 74-92.
- Searle, J. R. (1980). *Minds, brains, and programs*. *Behavioral and brain sciences*, 3(3), 417–424.
- Seibert, J. M., Hogan, A. E., et Mundy, P. C. (1982). *Assessing interactional competencies: The early social-communication scales*. *Infant Mental Health Journal*, 3(4), 244-258. [https://doi.org/10.1002/1097-0355\(198224\)3:4<244::AID-IMHJ2280030406>3.0.CO;2-R](https://doi.org/10.1002/1097-0355(198224)3:4<244::AID-IMHJ2280030406>3.0.CO;2-R)
- Severinson-Eklundh, K., Green, A., et Hüttenrauch, H. (2003). *Social and collaborative aspects of interaction with a service robot*. *Robotics and Autonomous Systems*, 42(3–4), 223-234. [https://doi.org/10.1016/S0921-8890\(02\)00377-9](https://doi.org/10.1016/S0921-8890(02)00377-9)
- Seyfarth, R. M., et Cheney, D. L. (1997). *Some general features of vocal development in nonhuman primates*. *Social influences on vocal development*, 13, 249.
- Shankar, A., McMunn, A., Banks, J., et Steptoe, A. (2011). *Loneliness, social isolation, and behavioral and biological health indicators in older adults*. *Health Psychology*, 30(4), 377.
- Sherman, E., et Newman, E. S. (1978). *The meaning of cherished personal possessions for the elderly*. *The International Journal of Aging and Human Development*, 8(2), 181–192.
- Shermer, M. (2008). *Patternicity: Finding Meaningful Patterns in Meaningless Noise*. *Scientific American*, 299, 48.
- Shibata, T., Ohkawa, K., et Tanie, K. (1996). *Spontaneous behavior of robots for cooperation. Emotionally intelligent robot system*. In *Robotics and Automation, 1996. Proceedings., 1996 IEEE International Conference on* (Vol. 3, p. 2426–2431). IEEE. Consulté à l'adresse : [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=506527](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=506527)
- Shibata, T., Wada, K., Saito, T., et Tanie, K. (2005). *Human interactive robot for psychological enrichment and therapy*. In *Proc. AISB* (Vol. 5, p. 98–109). Citeseer.
- Shochi, T. (2008, avril 21). *Prosodie des affects socioculturels en japonais, et anglais: à la recherche des vrais et faux-amis pour le parcours de l'apprenant*. Université Stendhal – Grenoble III. Consulté à l'adresse : <http://tel.archives-ouvertes.fr/tel-00366612>
- Shortliffe, E. H. (1976). *MYCIN: Computer-based medical consultations*. Elsevier, New York.
- Shriberg, E., Stolcke, A., et Baron, D. (2001). *Observations on overlap: findings and implications for automatic processing of multi-party conversation*. In *Interspeech* (p. 1359–1362). Consulté à l'adresse : <http://www.academia.edu/download/43762532/euro2001-overlap.pdf>
- Sidner, C. L., Kidd, C. D., Lee, C., et Lesh, N. (2004). *Where to look: a study of human-robot engagement*. In *Proceedings of the 9th international conference on Intelligent user interfaces* (p. 78–84). ACM. Consulté à l'adresse : <http://dl.acm.org/citation.cfm?id=964458>
- Sidner, C. L., Lee, C., Kidd, C. D., Lesh, N., et Rich, C. (2005). *Explorations in engagement for humans and robots*. *Artificial Intelligence*, 166(1), 140–164.
- Signorello, R., Aubergé, V., Vanpé, A., Granjon, L., et Audibert, N. (2010). *À la recherche d'indices de culture et/ou de langue dans les micro-événements audio-visuels de l'interaction face à face*. In *WACA 2010* (p. 69–76). Consulté à l'adresse : <https://hal.archives-ouvertes.fr/hal-00542004/>
- Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., ... Hirschberg, J. (1992). *ToBI: A standard for labeling English prosody*. In *Second International Conference on Spoken Language Processing*. Consulté à l'adresse : [http://www1.cs.columbia.edu/~julia/papers/TOBI\\_i92\\_0867.pdf](http://www1.cs.columbia.edu/~julia/papers/TOBI_i92_0867.pdf)
- Simner, M. L. (1971). *Newborn's response to the cry of another infant*. *Developmental psychology*, 5(1), 136.

- Simpson, A. P. (2012). *The first and second harmonics should not be used to measure breathiness in male and female voices*. *Journal of Phonetics*, 40(3), 477-490.  
<https://doi.org/10.1016/j.wocn.2012.02.001>
- Singer, T., Seymour, B., O’doherly, J., Kaube, H., Dolan, R. J., et Frith, C. D. (2004). *Empathy for pain involves the affective but not sensory components of pain*. *Science*, 303(5661), 1157–1162.
- Singh, M., Barnett, J., et Singh, M. P. (1997). *Automating spoken dialogue systems*. In *International Symposium on Methodologies for Intelligent Systems* (p. 207–216). Springer. Consulté à l’adresse : [http://link.springer.com/chapter/10.1007/3-540-63614-5\\_20](http://link.springer.com/chapter/10.1007/3-540-63614-5_20)
- Skantze, G., et Schlangen, D. (2009). *Incremental dialogue processing in a micro-domain*. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics* (p. 745–753). Association for Computational Linguistics.
- Skinner, B. F. (1992). “*Superstition*” in the pigeon. *Journal of Experimental Psychology: General*, 121(3), 273.
- Slade, A., Fear, J., et Tennant, A. (2006). *Identifying patients at risk of nursing home admission: The Leeds Elderly Assessment Dependency Screening tool (LEADS)*. *BMC Health Services Research*, 6, 31. <https://doi.org/10.1186/1472-6963-6-31>
- Slifka, J. (2007). *Irregular phonation and its preferred role as cue to silence in phonological systems*. In *XVIth International Congress of Phonetic Sciences* (p. 229–232). Consulté à l’adresse : <http://www.icphs2007.de/conference/Papers/1767/1767.pdf>
- Smith-Lovin, L., et Brody, C. (1989). *Interruptions in group discussions: The effects of gender and group composition*. *American Sociological Review*, 424–435.
- Snow, C. E., et Ferguson, C. A. (1977). *Talking to Children: Language Input and Acquisition*. Cambridge University Press.
- Sober, E., et Wilson, D. S. (1999). *The Évolution and Psychology of Unselfish Behavior*. Harvard University Press.
- Sorkin, D., Rook, K. S., et Lu, J. L. (2002). *Loneliness, lack of emotional support, lack of companionship, and the likelihood of having a heart condition in an elderly sample*. *Annals of Behavioral Medicine*, 24(4), 290–298.
- Sperber, D., et Wilson, D. (1987). *Precis of relevance: Communication and cognition*. *Behavioral and brain sciences*, 10(4), 697–710.
- Sperber, D., et Wilson, D. (1991). *La Pertinence, communication et cognition*. *Revue de Métaphysique et de Morale*, 96(3), 430.
- Sponselee, A., Schouten, B., Bouwhuis, D., et Willems, C. (2007). *Smart Home Technology for the Elderly: Perceptions of Multidisciplinary Stakeholders*. In M. Mühlhäuser, A. Ferscha, et E. Aitenbichler (Éd.), *Constructing Ambient Intelligence* (p. 314-326). Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-540-85379-4\\_37](https://doi.org/10.1007/978-3-540-85379-4_37)
- Standage, T. (2003). *Programmer Bookshelf -The Turk: The Life and Times of the Famous Eighteenth-Century Chess-Playing Machine*. *Dr Dobb’s Journal-Software Tools for the Professional Programmer*, 75.
- Stanton, M. E., Patterson, J. M., et Levine, S. (1985). *Social influences on conditioned cortisol secretion in the squirrel monkey*. *Psychoneuroendocrinology*, 10(2), 125–134.
- Steels, L. (2001). *Language games for autonomous robots*. *IEEE Intelligent systems*, 16(5), 16–22.
- Steels, L., et Kaplan, F. (2000). *AIBO’s first words: The social learning of language and meaning*. *Évolution of communication*, 4(1), 3–32.
- Steinfeld, A., Jenkins, O. C., et Scassellati, B. (2009). *The oz of wizard: simulating the human for interaction research*. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction* (p. 101–108). ACM.
- Step toe, A., Shankar, A., Demakakos, P., et Wardle, J. (2013). *Social isolation, loneliness, and all-cause mortality in older men and women*. *Proceedings of the National Academy of Sciences*, 110(15), 5797–5801.

- Stotland, E. (1969). *Exploratory investigations of empathy*. Advances in experimental social psychology, 4, 271–314.
- Strivay, L. (2004). *Enfants-loups, enfant-mouton, enfants-ours, enfants seuls*. Communications, 76(1), 41-57. <https://doi.org/10.3406/comm.2004.2158>
- Strum, S. C., Forster, D., et Hutchins, E. (1997). *Why Machiavellian intelligence may not be Machiavellian*. Machiavellian intelligence II: Extensions and evaluations, 50–85.
- Susini, P., Misdariis, N., Lemaitre, G., Houix, O., Rocchesso, D., Polotti, P., ... coll. (2006). *Closing the loop of sound evaluation and design*. Perceptual Quality of Systems, 2(4). Consulté à l'adresse : [http://www.researchgate.net/profile/Davide\\_Rocchesso/publication/224927745\\_Closing\\_the\\_Loop\\_of\\_Sound\\_Evaluation\\_and\\_Design/links/09e41508b8bcf8006c000000.pdf](http://www.researchgate.net/profile/Davide_Rocchesso/publication/224927745_Closing_the_Loop_of_Sound_Evaluation_and_Design/links/09e41508b8bcf8006c000000.pdf)
- Sutton, R. S., et Barto, A. G. (1998). *Reinforcement learning: An introduction* (Vol. 1). MIT press Cambridge. Consulté à l'adresse : [http://www.cell.com/trends/cognitive-sciences/pdf/S1364-6613\(99\)01331-5.pdf](http://www.cell.com/trends/cognitive-sciences/pdf/S1364-6613(99)01331-5.pdf)
- Suwa, M., et Suzuki, K. (2001). *Psychopathological features of« primary social withdrawal*. Seishin shinkeigaku zasshi= Psychiatria et neurologia Japonica, 104(12), 1228–1241.
- Suzuki, Y., Galli, L., Ikeda, A., Itakura, S., et Kitazaki, M. (2015). *Measuring empathy for human and robot hand pain using electroencephalography*. Scientific Reports, 5, 15924. <https://doi.org/10.1038/srep15924>
- Swerts, M., et Kraemer, E. (2005). *Audiovisual prosody and feeling of knowing*. Journal of Memory and Language, 53(1), 81–94.
- Taggart, W., Turkle, S., et Kidd, C. D. (2005). *An Interactive Robot in a Nursing Home: Preliminary Remarks*. In CogSci Workshop on Android Science (p. 56-61). Stresa, Italy.
- Tamura, T., Yonemitsu, S., Itoh, A., Oikawa, D., Kawakami, A., Higashi, Y., ... Nakajima, K. (2004). *Is an entertainment robot useful in the care of elderly people with severe dementia?* The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences, 59(1), 83-85.
- Tannen, D., et Saville-Troike, M. (1985). Perspectives on silence. Praeger Pub Text.
- Tapus, A., et Mataric, M. J. (2006). *Towards socially assistive robotics*. Journal of the Robotics Society of Japan, 24(5), 576–578.
- Tapus, A., Maja, M., et Scassellatti, B. (2007). *The grand challenges in socially assistive robotics*. IEEE Robotics and Automation Magazine, 14(1), N–A.
- Thomas, F., Johnston, O., et Thomas, F. (1995). *The illusion of life Disney animation* (1st Hyperion ed). New York Hyperion.
- Thórisson, K. R. (2002). *Natural turn-taking needs no manual: Computational theory and model, from perception to action*. In Multimodality in language and speech systems (p. 173–207). Springer. Consulté à l'adresse : [http://link.springer.com/chapter/10.1007/978-94-017-2367-1\\_8](http://link.springer.com/chapter/10.1007/978-94-017-2367-1_8)
- Thorndike, E. L. (1920, janvier). *Intelligence and Its Uses*. The Harpers Monthly.
- Tinbergen, N. (1963). *On aims and methods of ethology*. Ethology, 20(4), 410–433.
- Tisseron, S. (2013). *Real Humans : Robots ou humains?* Cerveau et Psycho, 60, 14-19.
- Tisseron, S. (2015). *Le jour où mon robot m'aimera – Vers l'empathie artificielle*. Paris: Albin Michel.
- Titchener, E. B. (1914). *An Historical Note on the James-Lange Theory of Emotion*. The American Journal of Psychology, 25(3), 427-447. <https://doi.org/10.2307/1412861>
- Tomaka, J., Thompson, S., et Palacios, R. (2006). *The Relation of Social Isolation, Loneliness, and Social Support to Disease Outcomes Among the Elderly*. Journal of Aging and Health, 18(3), 359-384. <https://doi.org/10.1177/0898264305280993>
- Tomasello, M. (2009). *The cultural origins of human cognition*. Harvard university press.

- Tomimoto, J., et Nishioka, A. (2013). *Idéophones et mots rigolos*. *Rencontres*, (27), 113–117.
- Toratani, K. (2005). *A cognitive approach to mimetic aspect in Japanese*. In Annual Meeting of the Berkeley Linguistics Society (Vol. 31, p. 335–346). Consulté à l'adresse : <http://journals.linguisticsociety.org/proceedings/index.php/BLS/article/download/856/639>
- Toratani, K. (2015). *Iconicity in the syntax and lexical semantics of sound-symbolic words in Japanese*. *Iconicity: East meets West*, 14, 125.
- Torrey, C., Fussell, S. R., et Kiesler, S. (2008). *Trying to be helpful: Social challenges for smart robots*. In Workshop-Proceedings of ACM/IEEE Human-Robot Interaction Conference (HRI2008), Amsterdam (p. 23–26). Consulté à l'adresse : [https://www.researchgate.net/profile/Cristen\\_Torrey/publication/228418579\\_Trying\\_to\\_be\\_Helpful\\_Social\\_Challenges\\_for\\_Smart\\_Robots/links/53fbb0940cf2dca8fffea600/Trying-to-be-Helpful-Social-Challenges-for-Smart-Robots.pdf](https://www.researchgate.net/profile/Cristen_Torrey/publication/228418579_Trying_to_be_Helpful_Social_Challenges_for_Smart_Robots/links/53fbb0940cf2dca8fffea600/Trying-to-be-Helpful-Social-Challenges-for-Smart-Robots.pdf)
- Trager, G. L. (1958). *Paralanguage: A first approximation*. *Stud. Linguist*, 13, 1–12.
- Tranter, S. E., et Reynolds, D. A. (2006). *An overview of automatic speaker diarization systems*. *IEEE Transactions on audio, speech, and language processing*, 14(5), 1557–1565.
- Trevarthen, C. (1998). *The concept and foundations of infant intersubjectivity*. *Intersubjective communication and emotion in early ontogeny*, 15–46.
- Trivers, R. (2004). *Mutual benefits at all levels of life*. *Science*, 304(5673), 964–966.
- Tronick, E., Als, H., Adamson, L., Wise, S., et Brazelton, T. B. (1978). *The infant's response to entrapment between contradictory messages in face-to-face interaction*. *Journal of the American Academy of Child psychiatry*, 17(1), 1–13.
- Tryggvason, M.-T. (2006). *Communicative behavior in family conversation: Comparison of amount of talk in Finnish, SwedishFinnish, and Swedish families*. *Journal of Pragmatics*, 38(11), 1795–1810.
- Tsujimura, N. (2001). *Revisiting the two-dimensional approach to mimetics: Reply to Kita (1997)*. *Linguistics*, 39(2; ISSU 372), 409–418.
- Tsvetanova, L., Aubergé, V., et Sasa, Y. (2016). *Peut-on caractériser globalement une «qualité d'acte expressif»: de "breathy voice" à "breathy turn taking" dans la glu socio-affective de l'interaction humain-robot?* In JEP 2016, Paris. Consulté à l'adresse : <https://jep-taln2016.limsi.fr/actes/Actes%20JTR-2016/Papers/J79.pdf>
- Turchet, P. (2000). *La synergologie*. Éditions de l'Homme. Consulté à l'adresse : <http://dnspf.com/actu-politique-et-societe/la-synergologie.pdf>
- Turing, A. M. (1950). *Computing machinery and intelligence*. *Mind*, 59(236), 433–460.
- Turkle, S. (2006). *A Nascent Robotics Culture: New Complicities for Companionship*. AAAI Technical Report Series.
- Turkle, S. (2012). *Alone Together: Why We Expect More from Technology and Less from Each Other*. New York, NY: Basic Books.
- Tylor, E. B. (1920). *Primitive culture : researches into the development of mythology, philosophy, religion, language, art, and custom*. London : Murray. Consulté à l'adresse : <http://archive.org/details/primitiveculture01tylouoft>
- Uchino, B. N., Cacioppo, J. T., et Kiecolt-Glaser, J. K. (1996). *The relationship between social support and physiological processes: A review with emphasis on underlying mechanisms and implications for health*. *Psychological Bulletin*, 119(3), 488–531. <https://doi.org/10.1037/0033-2909.119.3.488>
- Vacher, M., Bouakaz, S., Bobillier-Chaumon, M.-E., Portet, F., Aman, F., Rossato, S., et Guillou, E. (2014). *Le projet CIRDO d'assistance aux personnes âgées isolées à domicile*. In Congrès Handicap 2014-8ème Édition-Les technologies d'assistance: de la compensation à l'autonomie (p. 229–235). Consulté à l'adresse : <https://hal.archives-ouvertes.fr/hal-00990201/>

Vacher, M., Fleury, A., Portet, F., Serignat, J.-F., et Noury, N. (2009). *Reconnaissance des sons et de la parole dans un Habitat Intelligent pour la Santé: expérimentations en situation non contrôlée*. In GRETSI 2009 (p. ID456). Consulté à l'adresse : <https://hal.archives-ouvertes.fr/hal-00422561/>

Vacher, M., Fleury, A., Portet, F., Serignat, J.-F., et Noury, N. (2010). *Complete sound and speech recognition system for health smart homes: application to the recognition of activities of daily living*. In-Tech. Consulté à l'adresse : <https://hal.archives-ouvertes.fr/hal-00422576/document>

Vacher, M., Istrate, D., Portet, F., Joubert, T., Chevalier, T., Smidtas, S., ... et .coll. (2011). *The sweet-home project: Audio technology in smart homes to improve well-being and reliance*. In Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE (p. 5291–5294). IEEE.

Vacher, M., Lecouteux, B., Chahuara, P., Portet, F., Meillon, B., et Bonnefond, N. (2014). *The Sweet-Home speech and multimodal corpus for home automation interaction*. In The 9th edition of the Language Resources and Evaluation Conference (LREC) (p. 4499–4506). Consulté à l'adresse : <https://hal.archives-ouvertes.fr/hal-00953006/>

van Alphen, S. P. J., Oei, T. I., et Herzog-Evans, M. (2010). *Elderly people: The increase of over-60 crime in The Netherlands: An exploration*. Transnational Criminology Manual, Volume 2 (Crime), 491–496.

Vanpé, A. (2011). *Expressions et micro-expressions spontanées de la face et de la voix en Interaction Homme-Machine : esquisse d'un modèle du "Feeling of Thinking"*. Université Stendhal – Grenoble III. Consulté à l'adresse : <http://tel.archives-ouvertes.fr/tel-00625714>

Vanpé, A., et Aubergé, V. (2006). *Pertinence perceptive d'Icônes Gestuelles du Feeling of Thinking*. In Waca (p. 55–59). Consulté à l'adresse : <https://hal.archives-ouvertes.fr/hal-00364573/>

Vanpé, A., et Aubergé, V. (2012). *Early meaning before the phonemes concatenation? Prosodic cues for Feeling of Thinking*. GSCP, Belo Horizonte, Bresil.

Varela, F. (1989). *Autonomie et connaissance : Essai sur le vivant* (Seuil). Paris: Seuil.

Varela, F. (1996). *Invitation aux sciences cognitives* (Nouv. éd). Paris: Seuil.

Vasilescu, I., Adda-Decker, M., et Nemoto, R. (2008). *Caractéristiques acoustiques et prosodiques des hésitations vocaliques dans trois langues*. Traitement automatique des langues, 199–298.

Venkatesh, V., Morris, M. G., Davis, G. B., et Davis, F. D. (2003). *User acceptance of information technology: Toward a unified view*. MIS quarterly, 425–478.

Vergnes, P. (2015, août 22). *Empathie, conscience morale et psychopathie – le développement moral*. Consulté 24 juillet 2017, à l'adresse <http://4emesinge.com/empathie-conscience-morale-et-psychopathie-le-developpement-moral-partie-13/>

Vernier, P. (2005). *Évolution du cerveau et émergence du langage*. Consulté à l'adresse : <https://hal.archives-ouvertes.fr/hal-00383724>

Vertut, J. (2013). *Teleoperation and robotics: applications and technology* (Vol. 3). Springer Science et Business Media.

Victorri, B. (2000). *The role of narration in the emergence of human language*. In Proceedings of the Third International Conference on the Évolution of Language.

Vidal, D., et Gaussier, P. (2014). *Un robot comme personne*. Terrain. Anthropologie et sciences humaines, (62), 152-165. <https://doi.org/10.4000/terrain.15396>

Vigne, J.-D., et Guilaine, J. (2004). *Les premiers animaux de compagnie 8500 ans avant notre ère? ... ou comment j'ai mangé mon chat, mon chien et mon renard*. Anthropozoologica, 39(1), 249–273.

Vinciarelli, A., Pantic, M., Bourlard, H., et Pentland, A. (2008). *Social signal processing: state-of-the-art and future perspectives of an emerging domain*. In Proceedings of the 16th ACM international conference on Multimedia (p. 1061–1070). ACM. Consulté à l'adresse : <http://dl.acm.org/citation.cfm?id=1459573>

- Vinciarelli, A., Pantic, M., et Bourlard, H. (2009). *Social signal processing: Survey of an emerging domain*. *Image and Vision Computing*, 27(12), 1743–1759.
- Wada, K., et Shibata, T. (2007). *Robot therapy in a care house-change of relationship among the residents and seal robot during a 2-month long study*. In *Robot and Human interactive Communication*, 2007. RO-MAN 2007. The 16th IEEE International Symposium on (p. 107–112). IEEE.
- Wagner, R. A., et Fisher, M. J. (1975). *The String-to-String Correction Problem*. *Journal of the ACM (JACM)*, 21(1), 168–173. <https://doi.org/10.1145/321796.321811>
- Walker, M. A., Litman, D. J., Kamm, C. A., et Abella, A. (1997). *PARADISE: A framework for evaluating spoken dialogue agents*. In *Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics (p. 271–280)*. Association for Computational Linguistics.
- Walker, M., Kamm, C., et Litman, D. (2000). *Towards developing general models of usability with PARADISE*. *Natural Language Engineering*, 6(3-4), 363–377.
- Wallon, H. (2015). *Les origines du caractère chez l'enfant* (7e édition). Paris: Presses Universitaires De France – PUF.
- Walters, M. L., Dautenhahn, K., Te Boekhorst, R., Koay, K. L., Kaouri, C., Woods, S., ... Werry, I. (2005). *The influence of subjects' personality traits on personal spatial zones in a human-robot interaction experiment*. In *Robot and Human Interactive Communication*, 2005. ROMAN 2005. IEEE International Workshop on (p. 347–352). IEEE.
- Wapner, S., Demick, J., et Redondo, J. P. (1990). *Cherished possessions and adaptation of older people to nursing homes*. *The International Journal of Aging and Human Development*, 31(3), 219–235.
- Ward, J., et Simner, J. (2003). *Lexical-gustatory synaesthesia: linguistic and conceptual factors*. *Cognition*, 89(3), 237–261.
- Ward, J., Huckstep, B., et Tsakanikos, E. (2006). *Sound-colour synaesthesia: to what extent does it use cross-modal mechanisms common to us all?* *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, 42(2), 264–280.
- Ward, N. (2000). *The challenge of non-lexical speech sounds*. In *Sixth International Conference on Spoken Language Processing*. Consulté à l'adresse : <https://pdfs.semanticscholar.org/b77a/91afe0ad9f8e9a013539550a93c314b53000.pdf>
- Ward, N. (2004). *Pragmatic functions of prosodic features in non-lexical utterances*. In *Speech Prosody 2004, International Conference*. Consulté à l'adresse : [http://www.isca-speech.org/archive\\_open/sp2004/sp04\\_325.html](http://www.isca-speech.org/archive_open/sp2004/sp04_325.html)
- Watzlawick, P., Beavin, J. H., et Jackson, D. D. (2007). *Some tentative axioms of communication*. *Theorizing Communication: Readings Across Traditions*, 275.
- Weber, J. (2005). *Helpless machines and true loving care givers: a feminist critique of recent trends in human-robot interaction*. *Journal of Information, Communication and Ethics in Society*, 3(4), 209–218.
- Weiss, A., Bernhaupt, R., Lankes, M., & Tscheligi, M. (2010). *The USUS evaluation framework for human-robot interaction*. In *AISB2009: proceedings of the symposium on new frontiers in human-robot interaction*, 2009, April (Vol. 4, pp. 11–26).
- Weizenbaum, J. (1966). *ELIZA—a computer program for the study of natural language communication between man and machine*. *Communications of the ACM*, 9(1), 36–45.
- Werner, H., et Kaplan, B. (2014). *Symbol formation*. Psychology Press.
- Weyant, J. M. (1978). *Effects of mood states, costs, and benefits on helping*. *Journal of Personality and Social Psychology*, 36(10), 1169.
- Wharton, T. (2003). *Interjections, language, and the 'showing/saying' continuum*. *Pragmatics et Cognition*, 11(1), 39–91.
- Whiten, A., et Byrne, R. W. (1988). *Tactical deception in primates*. *Behavioral and brain sciences*, 11(2), 233–244.
- Whyte, W. F. (1984). *Learning from the field: A guide from experience*. Sage.

- Wichmann, A. (2002). *Attitudinal intonation and the inferential process*. In Speech Prosody 2002, International Conference. Consulté à l'adresse : [http://www.isca-speech.org/archive\\_open/sp2002/sp02\\_011.html](http://www.isca-speech.org/archive_open/sp2002/sp02_011.html)
- Wilkins, D. P. (1992). *Interjections as deictics*. *Journal of Pragmatics*, 18(2-3), 119–158.
- Wilkinson, C., Bultitude, K., et Dawson, E. (2011). *Oh yes, robots! People like robots; the robot people should do something: perspectives and prospects in public engagement with robotics*. *Science Communication*, 33(3), 367–397.
- Williams, J. D., et Young, S. (2007). *Partially observable Markov decision processes for spoken dialog systems*. *Computer Speech et Language*, 21(2), 393–422.
- Wilson, J. (1987). *The sociolinguistic paradox: data as a methodological product*. *Language et Communication*, 7(2), 161–177.
- Winningham, R. G., et Pike, N. L. (2007). *A cognitive intervention to enhance institutionalized older adults' social support networks and decrease loneliness*. *Aging and mental health*, 11(6), 716–721.
- Winograd, T. (1971). *Procedures as a representation for data in a computer program for understanding natural language*. Massachusetts Institute of Tech. Cambridge Project Mac.
- Wisner Fries, A. B., et Pollak, S. D. (2017). *The role of learning in social development: Illustrations from neglected children*. *Developmental science*, 20(2). Consulté à l'adresse : <http://onlinelibrary.wiley.com/doi/10.1111/desc.12431/full>
- Wispé, L. (1986). *The distinction between sympathy and empathy: To call forth a concept, a word is needed*. *Journal of personality and social psychology*, 50(2), 314.
- Wood, G., et Marty, S. (2005). *Le rêve de l'homme-machine : De l'automate à l'androïde*. Paris: Editions Autrement.
- Wrangham, R. W. (1987). *The significance of African apes for reconstructing human social evolution*. *The evolution of human behavior: Primate models*, 51–71.
- Wundt, W. M. (1907). *Outlines of psychology*. W. Engelmann.
- Yngve, V. H. (1970). *On getting a word in edgewise*. In Chicago Linguistics Society, 6th Meeting (p. 567–578). Consulté à l'adresse : <http://ci.nii.ac.jp/naid/10009705656/>
- Young, S., Gašić, M., Thomson, B., et Williams, J. D. (2013). *Pomdp-based statistical spoken dialog systems: A review*. *Proceedings of the IEEE*, 101(5), 1160–1179.
- Zahn-Waxler, C., Robinson, J. L., et Emde, R. N. (1992). *The development of empathy in twins*. *Developmental psychology*, 28(6), 1038.
- Zaslavsky, A., et Jayaraman, P. P. (2015). *Discovery in the Internet of Things: The Internet of Things (Ubiquity Symposium)*. *Ubiquity*, 2015(October), 2:1–2:10. <https://doi.org/10.1145/2822529>
- Zellner, B. (1994). *Pauses and the temporal structure of speech*. In Zellner, B. (1994). *Pauses and the temporal structure of speech*, in E. Keller (Ed.) *Fundamentals of speech synthesis and speech recognition*. (pp. 41-62). Chichester: John Wiley. (p. 41–62). John Wiley.
- Zimmermann, D. H., et West, C. (1996). *Sex roles, interruptions and silences in conversation*. *Amsterdam Studies in the Theory and History of Linguistic Science Series 4*, 211–236.

## Publications personnelles

- Y. Sasa, V. Aubergé, *SASI: perspectives for a socio-affectively intelligent HRI dialog system*. In Proc. of the BER Workshop on Behavior, Emotion and Representation: Building Blocks of Interaction, 5th International Conference on Human-Agent Interaction (HAI 2017), Bielefeld, Allemagne, 17th October, 2017.
- L. Tsvetanova, V. Auberge and Y. Sasa, *Multimodal breathiness in interaction: from breathy voice quality to global breathy “body behavior quality”*. In Proc. of the 1st International Workshop on Vocal Interactivity in-and-between Humans, Animals and Robots – VIHAR 2017, Stockholm, Suède, 25-26th August, 2017.
- R. Magnani, V. Aubergé, C. Bayol and Y. Sasa, *Bases of Empathic Animism Illusion: audio-visual perception of an object devoted to becoming perceived as a subject for HRI*. In Proc. of the 1st International Workshop on Vocal Interactivity in-and-between Humans, Animals and Robots – VIHAR 2017, Stockholm, Suède, 25-26th August, 2017.
- Y. Sasa and V. Aubergé. *Perceived isolation and elderly boundaries in EEE corpus : appeal to communication dynamics with a socio-affectively gluing robot in a Smart Home*. In Gerontechnology, 10th World Conference of Gerontechnology (ISG), Nice, France, 28-30th September, 2016
- L. Tsvetanova, V. Auberge, Y. Sasa, *Peut-on caractériser globalement une « qualité d’acte expressif » : de “breathy voice” à “breathy turn taking” dans la glu socio-affective de l’interaction humain-robot ?* In Actes des Journées d’Étude sur la Parole (JEP), pp. 509-517, Paris, 4-8th July 2016.
- M. Girard-Rivier, R. Magnani, V. Auberge, Y. Sasa and L. Tsvetanova, *Ecological gestures for HRI: the GEE corpus*. In Proc. of 10th Edition of Language Resources and Evaluation Conference (LREC), pp.1481-1484, Portorož, Slovenia, 23rd-28th May, 2016.
- Y. Sasa, R. Magnani, S. Castellanos-Paez, A. Arora, L. Tsvetanova, J. Maisonnasse, A. Alzhouri-Alyafi, *A Reflective, Adaptive & Communicative Ubiquitous Robot*. Selected for the – Robot Design Competition – of International Conference in Social Robotics (ICSR), Paris, France, 26-30th October, 2015.
- L. Guillaume, V. Aubergé, R. Magnani, F. Aman, C. Cottier, Y. Sasa, C. Wolf, F. Nebout, N. Neverova, N. Bonnefond, A. Nègre, L. Tsvetanova, and M. Girard-Rivier, *HRI in an ecological dynamic experiment: the GEE corpus based approach for the Emox robot*. IEEE International Workshop on Advanced Robotics and its Social Impacts, Lyon, France, July 1st – July 3rd, 2015.
- V. Aubergé and Y. Sasa, *Comment l’interaction rend le robot compagnon en augmentant la réalité sociale : indissociabilité du rôle social et de la glu socio-affective*. In Journées Nationales de la Robotique Interactive (JNRI), Toulouse, France, 2014.
- Y. Sasa and V. Aubergé, *Socio-affective interactions between a companion robot and elderly in a Smart Home context: prosody as the main vector of the “socio-affective glue*. In Proc. Speech Prosody 7, Dublin, Ireland, 2014.
- L. Batista Antunes, V. Aubergé, and Y. Sasa, *Certainty and uncertainty in Brazilian Portuguese: methodology of spontaneous corpus collection and data analysis*. In Proc. of the Speech Prosody 7, Dublin, Ireland, 2014.
- V. Aubergé, Y. Sasa, N. Bonnefond, B. Meillon, T. Robert, J. Rey-Gorrez, A. Schwartz, L. Batista Antunes, G. De Biasi, S. Caffiau, and F. Nebout, *The EEE corpus: socio-affective ‘glue’ cues in elderly-robot interactions in a Smart Home with the Emoz platform*. In

- Proc of the 5th International Workshop on Emotion, Social Signals, Sentiments and Linked Open Data, Reykjavik, Iceland, 2014. (**BEST PAPER AWARD**)
- Y. Sasa and V. Aubergé, *Caractéristiques prosodiques de la « glu socio-affective » de l'interaction face à face : un robot-compagnon médiateur d'un habitat intelligent pour personnes âgées*. In *Les Cahiers de Linguistiques Françaises* N°31/2014, the 3rd SWIP – Swiss Workshop on Prosody, Genève, Suisse, 2014.
- L. Guillaume, V. Aubergé, R. Magnani, F. Aman, C. Cottier, Y. Sasa, C. Wolf, F. Nebout, N. Neverova, N. Bonnefond, A. Nègre, L. Tsvetanova, and M. Girard-Rivier, *HRI in an ecological dynamic experiment: the GEE corpus based approach for the Emox robot*. In Proc. of the IEEE 2015 Int. Workshop Adv. Robot. Its Soc. Impacts, Lyon, France, July 1st – July 3rd, 2015.
- Y. Sasa, V. Aubergé, A. Rilliard, *Audio-Visual micro-expressions within Japanese-French contrast*. In WASSS, Grenoble, France, 2013.
- Y. Sasa, V. Aubergé, P. Franck, L. Guillaume, S. Moujtahid, *Des micro-expressions au service de la macro-communication pour le robot compagnon Emox*, In Actes du WACAI 2012, Grenoble, 54-59, 2012.
- F. Aman, M. Vacher, S. Rossato, R. Dugheanu, F. Portet, J. Legrand, Y. Sasa, *Caractérisation des différences phonétiques entre personnes âgées et non-âgées en vue de l'adaptation des modèles acoustiques des systèmes de RAP*. In Actes de JEP-TALN-RECITAL 2012, Grenoble, France, 707-714, 2012.

## Table des illustrations

### Figures

Figure 1 : Schéma simplifié des croisements de notions pluridisciplinaires de l'état de l'art . 15	
Figure 2 : Cartographie des sciences cognitives avec les champs de réflexion disciplinaire et les principaux auteurs (proposé par Varela 1989/1996 p.212) la flèche indique la position de l'auteur .....	29
Figure 3 : Continuum des quatre grandes approches des théories des émotions avec les contributeurs principaux (proposition de Barrett-Feldman et Gross, 2011) .....	37
Figure 4 : Nature des comportements non verbaux, codes et fonctions influençant la perception sociale de l'humain vis-à-vis d'un autre (d'après Vinciarelli et coll., 2008) .	42
Figure 5 : continuum du <i>showing-saying</i> de Wharton (extrait de Wharton, 2003).....	48
Figure 6 : Les paramètres acoustiques décrivant la prosodie (d'après Goudbeek et Scherer, 2010).....	53
Figure 7 : Timing des fonctions de la prosodie (in Audibert 2008, d'après Aubergé, 2002) ..	86
Figure 8 : Hypothèse de primitives vocales de <i>feedbacks/backchannels</i> à effet socio-affectif graduel dans le processus de construction de glu relationnelle en interaction face à face .....	87
Figure 9 : Principe général simplifié du <i>time stretching</i> .....	90
Figure 10: les changements de processus de socialisation inspirés de (Miura, 2001) (schémas tirés de Suzuki, Chapitre 3 in Fansten et coll., 2014).....	94
Figure 11: Pyramide des âges en France (a), au Japon (b) et en Europe (c) avec projections décennies à venir .....	102
Figure 12: évolution de l'espérance de vie et fragilisation physique des personnes (source : Rapport OMS de 2015) .....	102
Figure 13: robots d'assistance développés dans le cadre de projets gérontechnologiques ....	107
Figure 14: robotique de réhabilitation et d'assistance aux aidants et aux personnes âgées ...	107
Figure 15: robots domestiques et professionnels gérontechnologiques d'assistance sociale (Japon).....	108
Figure 16: Exemples de robots sociaux sur un continuum d'anthropomorphisme d'apparence .....	115
Figure 17: robot Emox (Awabot) (à gauche version sur laquelle les expériences de la thèse sont réalisées – à droite son nouveau design) .....	115
Figure 18 Conception de la combinaison Wizard/Oz selon (Steinfeld et coll., 2009).....	121
Figure 19 : Discontinuité perceptive de l'application robotique, continuité de la construction technologique (Aubergé, 2014).....	123
Figure 20 : Adaptation de l'approche 3P à la plateforme Living-Lab Domus-LIG.....	129
Figure 21 : Hypothèse de primitives vocales graduelle pour l'établissement d'une glu socio-affective .....	132
Figure 22 : Entreprise robotique Awabot (B. Bonnell, Villeurbanne – France) .....	133
Figure 23 : Identification des partenaires sociétaux – à gauche Bien à la Maison – au centre l'association AAPPUI – à droite le foyer-logement Roger Meffreys de la commune de Gières .....	137
Figure 24 : Plateforme expérimentale Domus du LIG et son appartement intelligent .....	138
Figure 25 : Schéma de la méthodologie Living-Lab expérimental de Domus-LIG (Aubergé, Sasa, Maisonnasse, 2015).....	140
Figure 26 : Schéma global simplifié des approches scientifiques sous-jacentes aux étapes méthodologiques (sens de lecture de droite à gauche).....	142

Figure 27 : Approche hypothético-déductive dans le recueil de corpus écologique et spontanée	144
Figure 28 : Maintien méthodologique pour une approche hypothético-déductive transversale	146
Figure 29 : Approches d’ethnométhodologie couplée aux approches hypothético-inductives	147
Figure 30 : Approches empirico-déductive et hypothético-déductive pour une fabrication évaluée par nouvelle approche hypothétique-inductive courte	149
Figure 31 : Schéma de la dimension itérative des boucles agiles de la méthodologie globale	152
Figure 32 : Chronologie des démarches de l’étude EmOz Elderly Expressions	153
Figure 33 : Échelles GIR et modèle (gauche) d’évaluation de la grille AGGIR (à droite)	155
Figure 34 : Ratio de personnes entretenues/participantes par origine de recrutement	156
Figure 35 : Illustration de l’interface du magicien EmOz	160
Figure 36 : Schématisation spatio-temporelle du scénario EEE	162
Figure 37 : Schéma de contrôle/acquisition des données induites à travers la plateforme EmOz	165
Figure 38 : Interfaces d’auto-annotation	169
Figure 39 : Schéma récapitulatif des données et traitements appliqués au corpus EEE dans la méthodologie living-lab Domus-LIG	172
Figure 40 : Illustration des principales fragilisations physiques des personnes et aides aux Activités de la Vie Quotidienne (ou AVQ)	173
Figure 41 : Informations en lien avec la situation du réseau relationnel des sujets	174
Figure 42 : Interface d’alignement du corpus EEE sous le logiciel ELAN	180
Figure 43 : Hiérarchie d’étiquetage des productions vocales des sujets	181
Figure 44 : Diagramme des productions vocales – granularités minimales	184
Figure 45 : Exemple simplifié d’alignement d’auto-annotations	196
Figure 46 : Etapes d’évolution morphosyntaxique des commandes	203
Figure 47 : Tendances d’évolution des pitches moyens des commandes pour des sujets prototypiques	212
Figure 48 : Patterns de contours prosodiques récurrents	217
Figure 49 : Spectrogrammes d’énoncés produits à 4 temps distincts de l’expérience (sujet Sa1) Les lettres indiquent l’ordre chronologique d’apparition après les premières paraphrases du sujet (a= première forme – d= dernière forme)	217
Figure 50 : Exemple de chronogramme des événements interactionnels (sujet Sa1)	219
Figure 51 : Évolution du niveau de glu basée sur FoT_PE – exemple du sujet S07	222
Figure 52 : Évolution des FoT_PE nivelés – patterns pour 16 sujets	224
Figure 53 : Évolution temporelle des états fonctionnels des commandes du sujet basée sur FoT_ProdSuj	225
Figure 54 : Patterns d’évolution des valeurs socio-affectives fonctionnelles de 16 sujets à la base de leur FoT_ProdSuj	226
Figure 55 : Évolution des durées des séquences FoT_Gn – Exemple de S01	227
Figure 56 : Évolutions des durées des séquences pour les 16 sujets retenus	228
Figure 57 : Marquage temporel des événements et tendance d’évolution de leur durée	229
Figure 58 : Patterns de distributions des événements interactionnels ponctuels avec l’évolution de leur durée sur les 16 sujets retenus	230
Figure 59 : Grammaire de coûts : présentation simplifiée des règles pour le calcul d’une valeur intégratrice pour la commande émise par les sujets (proposée par Tsvetanova et Aman, rapport interne Interabot)	236
Figure 60 : Schéma global des entrées/sorties successives d’un système RAP Cirdox (adaptée dans projet Interabot depuis Aman, Vacher, Portet, Duclot, et Lecouteux, 2016)	239

## Table des illustrations

---

Figure 61 : Intégration l’algorithme de distance d’édition modifié dans le synopsis d’un système RAP .....	240
Figure 62 : Récapitulatif des variations langagières observées pour la définition de 5 niveaux de modèles de langage pour le système RAP .....	241
Figure 63 : Algorithme de reconnaissance des commandes (proposé avec Aman et Aubergé dans rapport interne Interabot) .....	242
Figure 64 : Synopsis général d’un système de dialogue .....	246
Figure 65 : Tentative de schématisation du système SASI .....	248

## Tableaux

Tableau 1 : Répartition des différents types de productions vocales des sujets âgés.....	199
Tableau 2 : Répartition des stimuli produits du robot Emox (Woz vs. FtF).....	200
Tableau 3 : Tableau d'occurrences d'auto-annotation par type de FoT.....	202
Tableau 4 : Répartition des objets vocaux interactionnels hors commandes.....	206
Tableau 5 : Répartition des différents types de satellites observés.....	207
Tableau 6 : Tableau de répartition des objets vocaux périphériques.....	209
Tableau 7 : Récapitulatif des événements interactionnels sur le corpus EEE sur 16 sujets...	219
Tableau 8 : Pondération arbitraire systématisée résumant les valeurs de FoT_PE des sujets âgés.....	221
Tableau 9 : Comparaison des WER basés sur les modèles de langage spécialisés et le modèle de langage générique.....	244

## Table des annexes

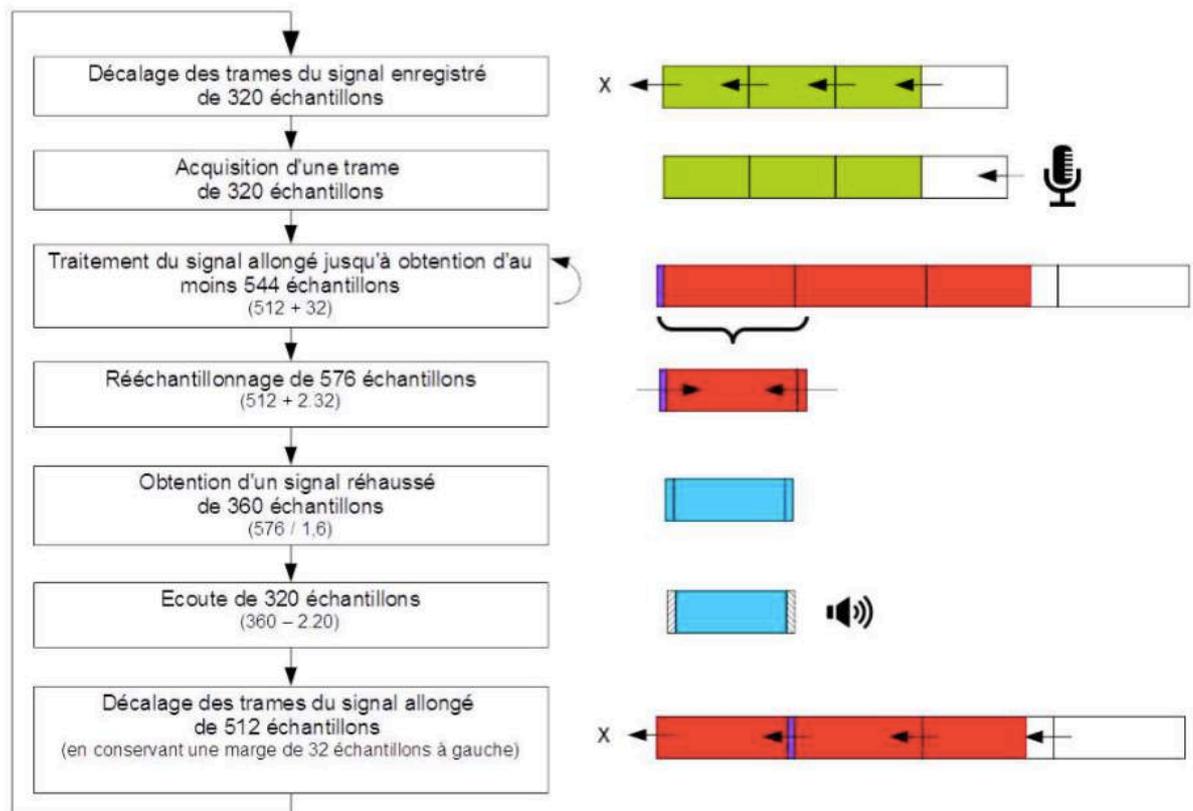
Annexe 1.	Outils de manipulation de la voix du robot .....	313
Annexe 2.	Approches scientifiques sous-jacentes la méthodologie globale Domus-LIG .....	317
Annexe 3.	Listes des primitives du robot et le principe de leur couplage en scripts .....	318
Annexe 4.	Scénario des stimuli du robot .....	329
Annexe 5.	Matériel et scénario du prétexte expérimental.....	332
Annexe 6.	Grille d'entretien de débriefing .....	358
Annexe 7.	Feuille de route de l'auto-annotation.....	361
Annexe 8.	Résumé des fiches sujets .....	365
Annexe 9.	Procédures et règles de transcription .....	368
Annexe 10.	Procédures de traitement du corpus EEE.....	371
Annexe 11.	Liste des commandes non-référencées produites par les sujets.....	383
Annexe 12.	Distribution des variations morphosyntaxiques par sujet.....	385
Annexe 13.	Chronogrammes des événements interactionnels .....	380
Annexe 14.	Étiquettes d'auto-annotation relevées par groupe temporel et par sujet.....	383



## Annexes

## Annexe 1. Outils de manipulation de la voix du robot

Le principe général du programme de conversion de voix pour la création des stimuli du robot s'appuie sur le déroulement algorithmique suivant (proposition de A. Davat et F. Gang dont l'illustration ci-dessous est tiré du rapport de stage de Davat, 2015):



La Figure décrit le fonctionnement général de l'algorithme du programme en temps réel, les premières étapes en vert étant évitées dans le cadre du traitement hors temps réel. Ce mode hors temps réel a fait l'objet de manipulation dans cette thèse, à travers le code ci-après. Ce programme tourne sur MatLab et prend en entrée des fichiers sons dont on veut modifier l'esthétique en s'appuyant sur le principe du *time stretching*, en déterminant des rapports de modification du pitch du signal.

```

%% Version finale hors temps-réel du programme de time-stretching
%% trame par trame
% rapport de transformation du pitch et taille de trame variables

close all
clear all
clc;

val_rapport = 1.5; % *modifiable
saut1 = 250; % longueur d'une trame de sig *modifiable
saut2 = saut1*val_rapport ; % longueur d'une trame de sig_modifie
long_lu = saut1*ceil((saut2+240)/saut1); % longueur du signal lu chaque
fois
% (pour ne pas avoir de trous dans sig_modifie, il faut que la trame
lue
% positionnée en pointeur2+k_pos permette d'atteindre pointeur2 +
saut2 +
% 119

SIG = audioread('ag01.wav');
LONG_SIG = length(SIG);
pSIG = 0;
pSIG_H = 0;
√©
% Témoin (ces variables n'interviennent pas dans les calculs mais
permettent
% de conserver le signal allongé par time-stretching)
SIG_MODIF = floor(LONG_SIG*1.6);
pMODIF = 0;

% Trames du signal sur lesquelles on travaille
nb_trames_sig = ceil(((saut2 + 40 + 119 + 120) +long_lu)/saut2);
long_sig = nb_trames_sig*saut1 ; % taille de la zone de traitement
if long_sig == long_lu,
    long_sig = long_sig + saut1;
    nb_trames_sig = nb_trames_sig +1;
end
sig = zeros(long_sig,1);

taille_win = 256/2; % taille de l'OLA
win0 = hanning(taille_win*4-1); win = zeros(taille_win*2,1);
for k=1:taille_win*2, win(k) = win0(k*2-1); end;
win_G = win(1 : taille_win);
win_D = win(taille_win+1 : taille_win*2);

k = [-120 : 120]'; pond = ones(241,1); % pondération de l'intercorr
pond(1:120) = 1 - abs(k(1:120))/120*0.5;
pond(122:241) = 1 - abs(k(122:241))/120*0.5;

pointeur1 = saut1;
pointeur2 = 0;

k_pos = 0;

nb_trames_sig_modifie = ceil(( saut2 + 40 + 240 + long_lu +
119)/saut2);
long_sig_modifie = nb_trames_sig_modifie*saut2;
sig_modifie = zeros(long_sig_modifie,1);

% Correction de resample

```

```

[N,D]=rat(val_rapport);
i = 0;
while (N*i<28)
    i = i+1;
end

marge_signal_allonge = N*i;
marge_signal_rehausse_long = floor(((marge_signal_allonge*2 +
saut2)/val_rapport - saut1)/2);
% Matlab ne reconnaît pas toujours que cette marge est un entier
% le floor ne sert qu'à s'assurer que ce sera le cas

marge_g = zeros(marge_signal_allonge,1);
signal_ecoute = zeros(saut2+2*marge_signal_allonge,1);
signal_hausse_long = zeros(saut1+2*marge_signal_rehausse_long,1);
signal_hausse = zeros(LONG_SIG,1); %signal réchantillonné en prenant
des marges

% Initialisation du signal sig

sig = SIG(pSIG + 1 : pSIG + nb_trames_sig*saut1);
pSIG = pSIG + nb_trames_sig*saut1 ;

% Construction du signal allongé

while (pSIG + long_lu <= LONG_SIG),

%   disp('enregistrement');

    for i=1:nb_trames_sig-1,
        sig((i-1)*saut1+1:i*saut1) = sig(i*saut1+1:(i+1)*saut1);
    end
    sig((nb_trames_sig-1)*saut1+1:end) = SIG(pSIG + 1 : pSIG + saut1);
    pSIG = pSIG + saut1 ;

    pointeur1 = pointeur1 - saut1;

    while (pointeur2 - 120 < saut2 + marge_signal_allonge + 119)

%       disp('time-stretching');

        sig_trame = sig(pointeur1+1 : pointeur1+long_lu);
        intensite = 10*log10(sum(sig_trame.^2)+1e-30);

        if(intensite > -45),

            if(pointeur1 > 0),
                ic = xcorr(sig_trame_precedente, sig_trame(1:saut1));
                lic = min(length(sig_trame_precedente),length(sig_trame));
                intercorr = ic(lic+1:lic+241) .* pond; %
                [Y, I] = max(intercorr);
                k_pos = I-121;
            end;
            else
                k_pos = 0;
            end;

% reconstruction du signal allongé
%   disp('extrapolation');
        if((pointeur1==0)&&(pointeur2==0)),
            sig_modifie(k_pos+1 : k_pos+long_lu) = sig_trame;

```

```

else
    sig_modifie(pointeur2+k_pos+taille_win+1 : pointeur2+k_pos+long_lu)
=...
    sig_trame(taille_win+1:long_lu);
    sig_modifie(pointeur2+k_pos+1 : pointeur2+k_pos+taille_win) =...
    sig_modifie(pointeur2+k_pos+1 :
pointeur2+k_pos+taille_win).*win_D...
    + sig_trame(1:taille_win).*win_G;
end;

%   disp('sauvegarde trame_precedente');
    if (pointeur2+saut2+1-120>=0)
        sig_trame_precedente = sig_modifie(pointeur2+saut2+1-120 :
pointeur2+saut2+120+saut1);
    else
        sig_trame_precedente(1:120) = marge_sig_trame_precedente ;
        sig_trame_precedente(121:saut1+240) =
sig_modifie(pointeur2+saut2+1 : pointeur2+saut2+120+saut1);
    end

    pointeur1 = pointeur1 + saut1;
    pointeur2 = pointeur2 + saut2;
end

%   disp('écoute');
    signal_ecoute(1:marge_signal_allonge) = marge_g;
    signal_ecoute(marge_signal_allonge+1:end) = sig_modifie(1 : saut2 +
marge_signal_allonge);
    [signal_hausse_long,rep] = resample(signal_ecoute,10,10*val_rapport);
    signal_hausse(pSIG_H+1 : pSIG_H + saut1) =
signal_hausse_long(marge_signal_rehausse_long + 1 : saut1 +
marge_signal_rehausse_long);
    pSIG_H = pSIG_H + saut1 ;

%   disp('décalage trames sig_modifie');
    marge_g = sig_modifie(saut2-marge_signal_allonge+1:saut2);
    marge_sig_trame_precedente = sig_modifie(saut2+saut2-
120+1:saut2+saut2);

    SIG_MODIF(pMODIF +1 : pMODIF + saut2) = sig_modifie(1:saut2);

    for i=1:nb_trames_sig_modifie-1,
        sig_modifie((i-1)*saut2+1:i*saut2) =
sig_modifie(i*saut2+1:(i+1)*saut2);
    end

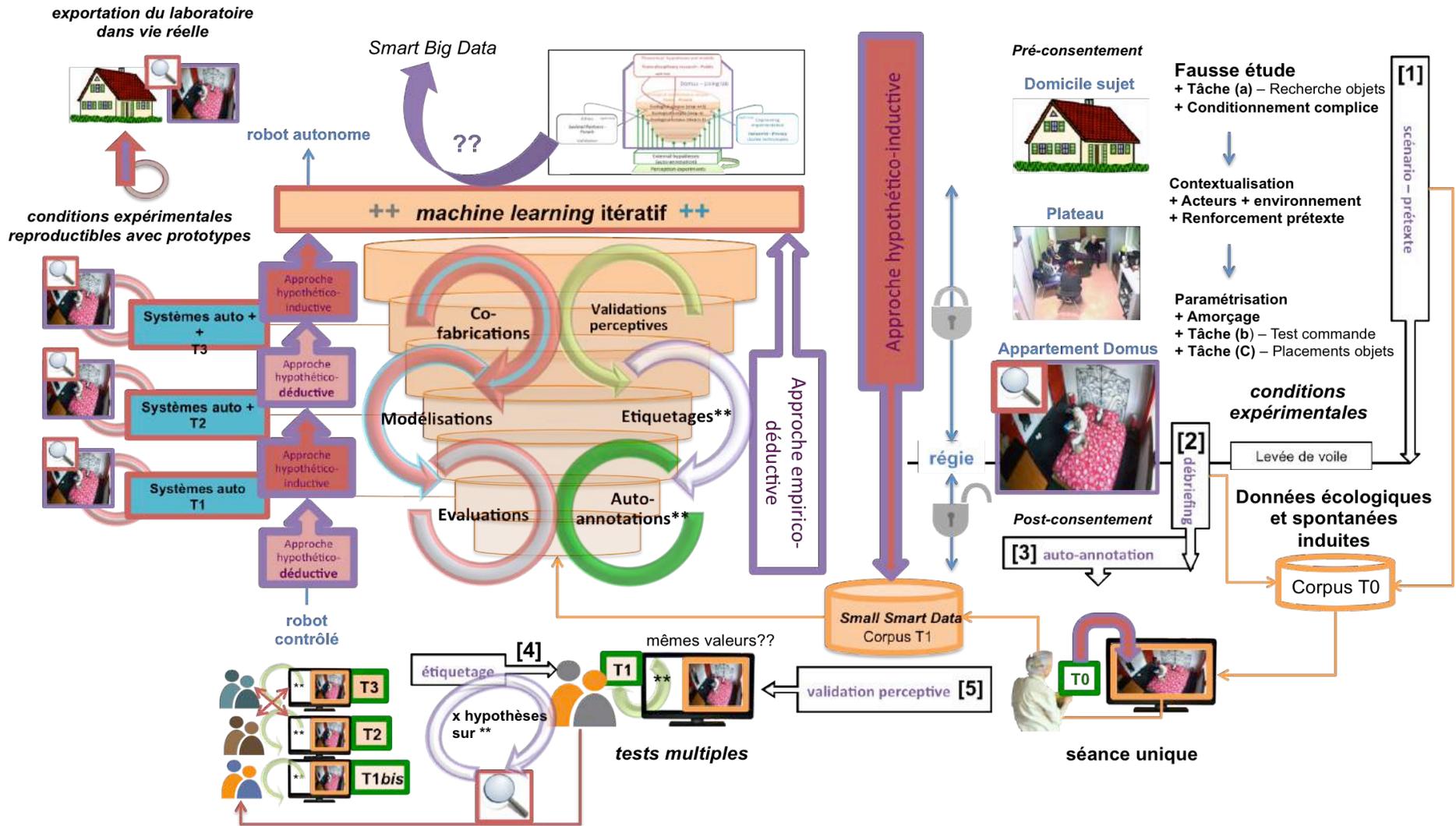
    pointeur2 = pointeur2 - saut2 ;
    pMODIF = pMODIF +saut2;

end

sound(signal_hausse,16000)

```

## Annexe 2. Approches scientifiques sous-jacentes la méthodologie globale Domus-LIG



### Annexe 3. Listes des primitives du robot et le principe de leur couplage en scripts

#### PRIMITIVES VOCALES DU ROBOT

Le tableau ci-dessous présente en deuxième colonne le nom des fichiers .wav associés à chaque stimulus. En troisième colonne nous avons une transcription orthographique pour les sons lexicalisés type interjection et imitation, une description acoustique entre [] pour les bruits de bouche et des transcriptions phonologiques en API (Alphabet Phonétique International) entre // pour les sons prélexicaux. La quatrième colonne décrit la nature du stimulus définissant la gradualité de contrôle de prosodie socio-affective posée en hypothèse. Enfin, la cinquième colonne indique le niveau de cette gradualité de contrôle (1= son primitif – 9 = sons d’imitation lexicalisés avec prosodie prononcée / x = complexe sortant du cadre d’hypothèse). Les stimuli notés « annexe » sont des éléments complexes qui sortent de l’hypothèse de gradualité, mais nécessaire pour une adaptation à l’écologie de l’interaction (p. ex. proposition de la station de radio disponible pour garder la cohérence du rôle majordome de l’habitat intelligent – ils sont notés « systématique »). Les sons « non systématiques » sont des énoncés enregistrés suite aux prétests pour débloquer certaines situations complexes d’interaction (p. ex. passivité du sujet, rassurer le sujet) – ce sont des stimuli utilisés le moins possible.

N°	Nom du stimulus	Transcription	Nature stimulus	Degré
1	3-Hum1.wav	[vocalisation – bouche fermée – monovocalique]	bruit de bouche	1
2	3-HUm2.wav	[vocalisation – bouche fermée – monovocalique]	bruit de bouche	1
3	3-HumHum1.wav	[vocalisation – bouche fermée – bivocalique]	bruit de bouche	1
4	rire_HAHAHA.wav	[rire – /hahaha/ – trivocalique]	bruit de bouche	1
5	rire2_HIHIHI_insp.wav	[rire – /hi:hihi/[inspiration] – trivocalique]	bruit de bouche	1
6	rire3_HEHE.wav	[rire – /hehe/ – bivocalique]	bruit de bouche	1
7	3-Euh.wav	/ø:/	son pré-lexical	2
8	4-OOps.wav	/ups/	son pré-lexical	2
9	53-WoupDESC.wav	/vup/	son pré-lexical	2
10	53-WoupMONT.wav	/vu::p/	son pré-lexical	2

## Annexes

11	1-O-K-1.wav	/o:ke:/	pré-lexical – interjection	3
12	1-OK-2.wav	/oke:/	pré-lexical – interjection	3
13	1-Ouais.wav	/uwε/	pré-lexical – interjection	3
14	4-OhPARDON.wav	/o:/ pardon	son pré-lexical + interjection	3
15	1-DAC-cord1.wav	d'accord	interjection	4
16	1-DaccORD2.wav	d'accord	interjection	4
17	1-OUi1.wav	oui	interjection	4
18	1-ouI2.wav	oui	interjection	4
19	2-Voila1.wav	voilà	interjection	4
20	2-VoiLa2.wav	voilà	interjection	4
21	2-CaYest.wav	ça y est	interjection	5
22	2-CommeCa.wav	comme ça	interjection	5
23	0-CAVa1.wav	ça va	interjection+	6
24	0-CaVA2.wav	ça va	interjection+	6
25	52-MoinsFortTV.wav	moins fort la télé	imitation	7
26	53-DescendreSTORES.wav	descendre les stores /vu:p/	imitation + pré-lexical	8
27	53-MonterSTORES.wav	monter les stores /wu:p/	imitation + pré-lexical	8

## Annexes

28	52-EteindreTV.wav	éteindre la télé voilà	imitation + interjection	9
29	5-mettrelaBOUIL.wav	mettre la bouilloire voilà	imitation + interjection	9
30	5-mettrelaLUMI.wav	mettre la lumière d'accord	imitation + interjection	9
31	5-mettreRADIO.wav	mettre la radio d'accord	imitation + interjection	9
32	51-MOINSfort.wav	moins fort comme ça	imitation + interjection	9
33	51-PLUSfort.wav	plus fort comme ça	imitation + interjection	9
34	52-EteindreTV.wav	éteindre la télé voilà	imitation + interjection	9
35	plusfort.wav	plus fort comme ça	imitation + interjection	9
45	0-Aurevoir.wav	au revoir (phatique pour rassurer)	annexe – non systématique	x
46	0-Bonjour.wav	bonjour je suis Emox (phatique pour rassurer)	annexe – non systématique	x
36	0-estcequejepeuxFAIRE.wav	est-ce que je peux faire quelque chose pour votre service (pour sollicitation)	annexe – non systématique	x
37	0-jepeuxFAIRE.wav	je peux faire quelque chose (pour sollicitation)	annexe – non systématique	x
43	4-DEJAFait.wav	mais c'est déjà fait (si cas d'erreur)	annexe – non systématique	x
44	4-nonjeneSAISPAS.wav	non je ne sais pas le faire (si cas imprévu)	annexe – non systématique)	x
38	leFILM_1.wav	le film il y a sur le désert ou bien sur la savane ou bien sur les Seychelles	annexe – systématique	x

Annexes

39	leFILM_2.wav	le film sur le désert ou bien sur la savane ou bien sur les seychelles	annexe – systématique	x
40	0-ouijesuisLAJECOUTE.wav	oui je suis là j'écoute	annexe – systématique	x
41	yaFRINT.wav	y a France Inter ou bien radio Nova ou alors Radio Culture	annexe – systématique	x
42	0-maintenantSERVICE.wav	maintenant je suis à votre service	annexe – systématique	x

### PRIMITIVES DE MOUVEMENT DU ROBOT

Le tableau ci-dessous récapitule les mouvements principaux développés. Ils sont émis en parallèle des primitives de sons, notamment pour induire les premières interactions de l'expérimentation, sans sons, puis maintenir le degré d'animisme généré tout au long de l'expérience avec un lot de mouvement semblable. Pour éviter les sollicitations par langage complexe, ces primitives permettent d'attirer l'attention du sujet. Ils complètent les mouvements et les déplacements effectués à la manette.

N°	NomScript	Descriptif
1	z2_Avance_M.xls	Mouvement d'avancée du robot
2	z2_AvanceD_M.xls	Le robot avance par une trajectoire courbe légère à droite
3	z2_AvanceG_M.xls	Le robot avance par une trajectoire courbe légère à gauche
4	z2_AvanceRecul_M.xls	Petits mouvements où le robot avance et recule de manière répétée
5	z2_Recul_M.xls	Petit mouvement de recul du robot
6	z2_Reculloin_M.xls	Grand mouvement de recul du robot
7	z2_ReculD_M.xls	Mouvement de recul vers la droite
8	z2_ReculG_M.xls	Mouvement de recul vers la gauche
9	z2_RegardD_M.xls	Mouvement de tête rapide pour orienter la tête du robot vers la droite
10	z2_RegardG_M.xls	Mouvement de tête rapide pour orienter la tête du robot vers la gauche
11	z2_TeteDG_M.xls	Mouvement de tête du robot de la droite vers la gauche
12	z2_TeteGD_M.xls	Mouvement de tête du robot de la gauche vers la droite
13	z2_TeteHB_M.xls	Mouvement de tête du robot de haut en bas

Annexes

14	z2_TeteNON_M.xls	Mouvements de tête répétés réguliers de gauche à droite
15	z2_TeteOUI_M.xls	Mouvements de tête répétés réguliers de droite à gauche
16	z2_PivoterD_M.xls	Mouvement sur un petit angle vers la droite (pivotement)
17	z2_PivoterG_M.xls	Mouvement sur un petit angle vers la gauche (pivotement)
18	z2_TourD_M.xls	Tour du robot sur axe fixe avec les roues avant bloquées vers la droite
19	z2_TourG_M.xls	Tour du robot sur axe fixe avec les roues avant bloquées vers la gauche
20	z2_Dandiner_M.xls	Petits mouvements irréguliers alternés à droite et à gauche avec changement d'angle de pivotement
21	Reinitialiser.xls	Couplage de scripts permettant de repositionner sur dimension centrale toutes les dimensions
22	Demarrage.xls	Couplage de script permettant d'initialiser l'état du robot (script réinitialiser.xls) et de l'appartement dans les conditions initiales de l'expérimentation (rideaux et volets ouverts, lumières éteintes, enceintes allumées, mais sur volume réglé, la télévision sur film « noir » mimant l'arrêt, arrêt des prises murales, lampes de chevet éteintes et LED de l'appartement éteinte)

## CONSTRUCTION DES SCRIPTS DE PRIMITIVES (ROBOTIQUES ET DOMOTIQUES)

Voici l'API (*Application Programming Interface*) des scripts en collaboration avec T. Robert, J. Rey-Gorrez en stage supinfo (2013-2014) sous la direction de N. Bonnefond, V. Aubergé et B. Meillon permettant de créer les primitives d'interactions de l'interface EmOz à travers des fichiers Excel. (1) API des scripts robotiques; (2) API des scripts domotiques; (3) *template* de fichier Excel de scripts.

(1)

Action	Paramètre 1	Paramètre2	Paramètre 3	Description	Restriction(s)
moveFront	<i>vitesse (en cm/s)</i>	<i>distance (en cm)</i>	/	Avance en ligne droite	Vitesse entre 5 et 25
moveBack	<i>vitesse (en cm/s)</i>	<i>distance (en cm)</i>	/	Reculé en ligne droite	Vitesse entre 5 et 25
spinLeft	<i>vitesse (en cm/s)</i>	<i>angle (en degrés)</i>	/	Tourne à gauche sur lui-même	Vitesse entre 5 et 25
spinRight	<i>vitesse (en cm/s)</i>	<i>angle (en degrés)</i>	/	Tourne à droite sur lui-même	Vitesse entre 5 et 25
turnLeft	<i>vitesse (en cm/s)</i>	<i>angle (en degrés)</i>	<i>diamètre (en cm)</i>	Tourne à gauche sur un cercle de diamètre spécifié	Vitesse entre 5 et 25
turnRight	<i>vitesse (en cm/s)</i>	<i>angle (en degrés)</i>	<i>diamètre (en cm)</i>	Tourne à droite sur un cercle de diamètre spécifié	Vitesse entre 5 et 25
turnBackLeft	<i>vitesse (en cm/s)</i>	<i>angle (en degrés)</i>	<i>diamètre (en cm)</i>	Tourne à gauche sur un cercle de diamètre spécifié, en marche arrière	Vitesse entre 5 et 25
turnBackRight	<i>vitesse (en cm/s)</i>	<i>angle (en degrés)</i>	<i>diamètre (en cm)</i>	Tourne à droite sur un cercle de diamètre spécifié, en marche arrière	Vitesse entre 5 et 25
lookLeft	<i>amplitude (en %)</i>	<i>temps (en ms)</i>	/	Regarde à gauche	Amplitude entre 0 et 100
lookRight	<i>amplitude (en %)</i>	<i>temps (en ms)</i>	/	Regarde à droite	Amplitude entre 0 et 100
lookUp	<i>amplitude (en %)</i>	<i>temps (en ms)</i>	/	Regarde en haut	Amplitude entre 0 et 100
lookDown	<i>amplitude (en %)</i>	<i>temps (en ms)</i>	/	Regarde en bas	Amplitude entre 0 et 100
speaker	<i>nom du son</i>	/	/	Joue un son	Nom du fichier sans l'extension
wait	<i>temps (en ms)</i>	/	/	Attend un temps donné	
resetHead	/	/	/	Repositionne la tête au centre	Dure 1s
stopWheels	/	/	/	Arrête les moteurs des roues	
script	<i>nom du script</i>	/	/	Insère un script	Nom du fichier sans l'extension

(2)

Pièce	Objet	Etat
<b>chambre</b>		
	spot	<i>on / off</i>
	store	<i>up / down / stop</i>
	rideaux	<i>up / down</i>
	chevet	<i>on / off</i>
<b>cuisine</b>		
	spot	<i>on / off</i>
	store	<i>up / down / stop</i>
	prise	<i>on / off</i>
<b>salon</b>		
	spot	<i>on / off</i>
	store	<i>up / down / stop</i>
<b>radio</b>		
	volume	<i>1 / 2 / 3 / 4</i>
	chaine	<i>radio_nova / radio_culture / france_inter / radio_classique</i>
	off	
<b>led</b>		
	color	<i>Rouge*Vert*Bleu (ex : 255*123*200)</i>
	off	

**Remarque:** un script de primitives simples peut se combiner avec des scripts d'autre nature (p. ex. sons du robot, mouvements du robot et domotiques, tous mélangés). Ainsi environ 126 scripts sont utilisés pour générer l'interface EmOz permettant de coupler les primitives de base du robot (sons et mouvements) et les primitives domotiques. Cela permet ainsi d'avoir des effets combinés de sons et mouvements des robots plus synchrones et également de réduire les délais d'exécution du magicien à la réalisation de la commande pour se rapprocher au plus près du temps cognitif du magicien.

(3)

	A	B	C	D	E
1	<b>Auteur</b>				
2	...				
3	<b>Description</b>				
4	...				
5					
6	<b>Paramètre</b>	<b>Valeur</b>			
7	AFFICHER	0 / 1			
8	ZONE	1 / 2			
9					
10	<b>Groupe</b>	<b>Action</b>	<b>P1</b>	<b>P2</b>	<b>P3</b>
11	...	...	...	...	...
12					
13					

**Remarque:** en complément, il est possible de personnaliser l’affichage du bouton d’interface en case A9. Il est à noter que la nomenclature des noms des scripts est extrêmement importante à formaliser à la base pour que chaque occurrence soit unique et parce que cette trace va être conservée dans les traces temporelles du magicien. Ainsi plus le script porte des informations utiles, plus le post-traitement des données sera facilité. Voici donc ci-dessous la nomenclature des noms de scripts qui fait partie des préparations pour le filtrage rapide des traces:

## Nomenclature nom de scripts

### 1° - Le type de script

On rajoute une extension à la fin du nom de fichier pour savoir quel est le type du script.

S = son

M = mouvement

D = domus

Action S : uniquement de son(s)

ex : *Oops\_S*

Action M : uniquement de mouvement(s)

ex : *Non\_M*

Action D : pour des actions de DOMUS

ex : *Lumcuisson\_D*

Action SM : son(s) + mouvement(s)

ex : *Oui2\_Tourg\_SM*

Action SD : son(s) + DOMUS

ex : *Ok1\_Storechboff\_SD*

Action M : mouvement(s) + DOMUS

ex : *Avancerecule\_*

Action SMD : son(s) + mouvement(s) + DOMUS

ex :

*Rire1TourdLumcuis\_SMD*

### 2° - Enchaînements d'actions dans un script

\*\* Si on a plusieurs actions simultanées dans un script (les actions appartiennent à un même groupe), on écrit ces actions en bloc, les uns à la suite des autres, chaque action commence par une majuscule.

Action1Action2Action3\_Typescript : enchaînement de 3 actions simultanées

ex : *Rire1TourdLumcuis\_SMD*

=> Emox rit en tournant vers la droite tout en allumant la lumière de la cuisine. Type du script : son + mouvement + DOMUS

\*\* Si les actions se font dans un ordre précis, on utilise le séparateur “\_” et on écrit les actions dans l'ordre.

Action1Action1bis\_Action2\_Typescript : enchaînement de 2 actions simultanées

(action du même groupe) puis une action qui se fait après.

ex : *Rire1Tourd\_Lumcuis\_SMD*

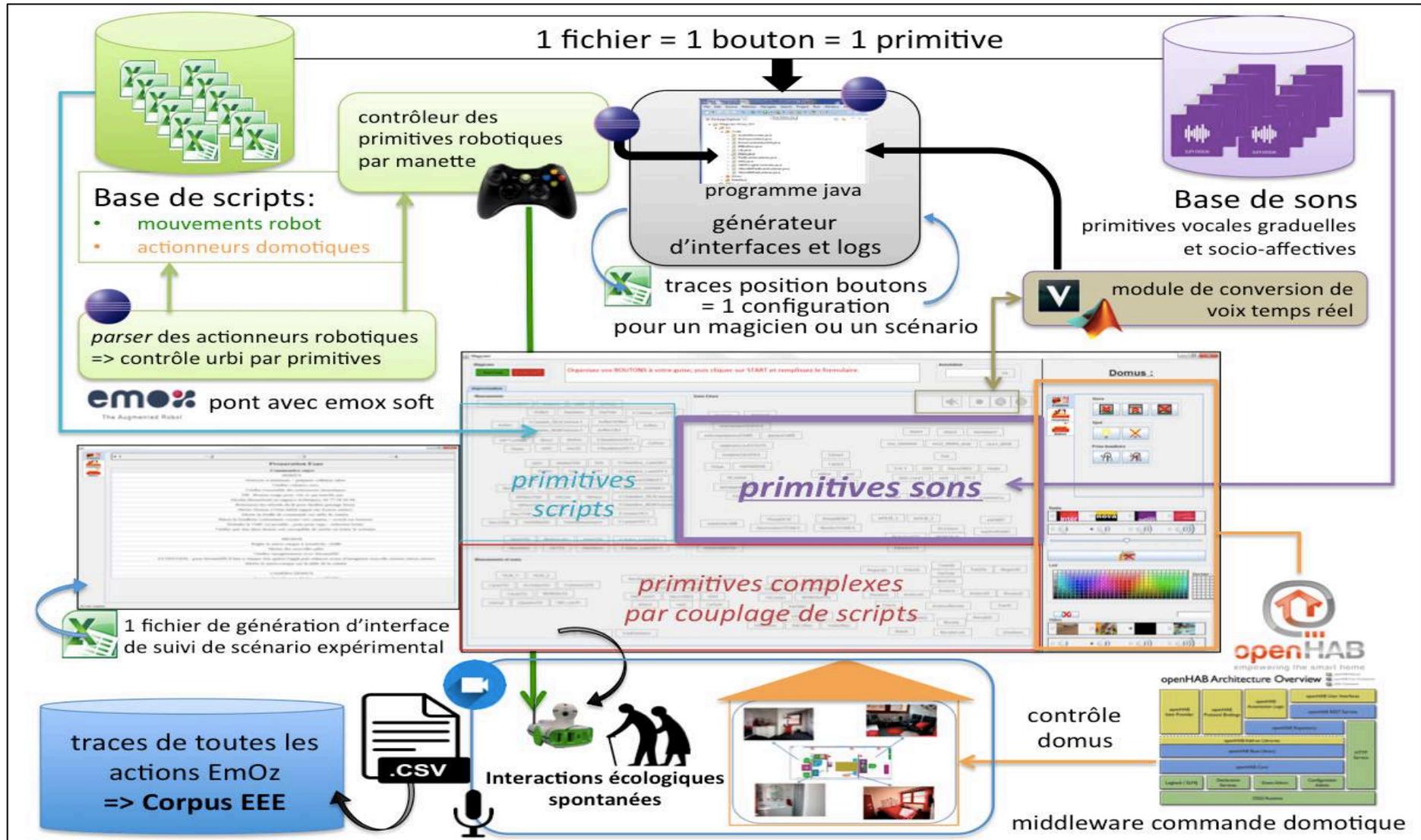
=> Emox rit en tournant vers la droite puis il allume la lumière de la cuisine. Type du script : son + mouvement + DOMUS

On peut également avoir des scripts avec une alternance entre des blocs d'actions qui se font simultanément, et des actions qui s'enchaînent les uns à la suite des autres.

ex : *Avance\_Lumcuisson\_CayestTourd\_SMD*

=> Emox avance, puis il allume la lumière de la cuisine, puis il fait un tour vers la droite tout en disant « ça y est ».

Rappel d'architecture technique de la plateforme de magicien d'oz EmOz dans lequel s'intègre ces scripts



### Annexe 4. Scénario des stimuli du robot

S : Sujet Exp1 (recruteur)

Exp2 (simule l'ingénieur DOMUS)

1/2 – mmTps : ordre commandes/réaction

	<b>Commandes</b>	<b>Suj (+Exp2)</b>	<b>Suj refait seul les commandes Phase 3 + 4 (=impro)</b>	<b>Suj (+Exp1)</b>
<b>Dans la cuisine</b>	« Mettre la lumière » 1	Phase2 Avance 2	Oui2	Cayest
	« Eteindre la lumière » 2	Oui de la tête 1	Humhum	Voila1
	« Descendre les stores » mmTps	Avance recule mmTps	Avance recule Voila1	Avance recule
	« Monter les stores » mmTps	Avance recule mmTps	Avance recule Ok2	Avance recule
	« Mettre la bouilloire » 2	Humhum 1	Rire2	Ok1 Cayest
	« Eteindre la bouilloire » 2	Hum1 Tourne 1	Oui2	ouais
<b>Dans la chambre</b>	« Mettre la lumière » 2	Euh 1	Humhum Tourne tete GD	Oui1
	« Eteindre la lumière » 2	Oui1 1	Oui2	Humhum
	« Fermer les rideaux » mmTps	Avance Ok2 mmTps	Avance Hum2	Avance Voila1
	« Ouvrir les rideaux » mmTps	Reculé Oui2 mmTps	Reculé Cayest	Reculé Humhum
	« Descendre les stores »	woupdesc Tête GD	Descendresstoreswoup	woupdesc

	2	1		
	« Monter les stores »	Woupmont Tête DG	monterstores	woupmont
	2	1		
	« Mettre les lampes »	Bouge la tête HB hum1	bouge la tête HB oui1	Bouge la tête HB
	2	1		
	« Eteindre les lampes »	Bouge la tête HB humhum	bouge la tête HB oui2	Bouge la tête HB
	« Mettre la télé »	Lefilm1 1 <i>Si pas compris</i> Lefilm2	Lefilm1  <i>Si pas compris</i> Lefilm2	Lefilm1  <i>Si pas compris</i> Lefilm2
<b>Télévision</b>	« plus fort la télé »	Cava1 Tourne	plusfortTV Tourne	Commecca Tourne
	1	2		
	« moins fort la télé »	Cava2 Tourne autre sens	moinsfortTV tourne autre sens	moinsfortTV Tourne autre sens
	1	2		
	« Eteindre la télé »	Voila2	eteindreTV	Daccord1
	1	2		
	« Mettre la lumière jaune »	Reculer loin Rire1 Tourne tete GD	Reculer loin Oui1 Rire1 Tourne tete GD	Reculer loin Voila1
	1	2		
<b>Utilisation des LED</b>	« mettre en bleu »	Rire2 Tourne tete GD	Rire2 Tourne tete GD	Voila2
	1	2		
	« mettre en vert »	Daccord1	Daccord2	Oui1

	2	1		
	« Eteindre cette lumière » 2	Hum2 1	Oui2	Cayest1
<b>Dans le salon</b>	« Mettre la lumière » 2	Hum1 Tourne 1	Mettrelalumi Tourne	Hum1 Tourne
	« Eteindre la lumière » 2	Humhum 1	Ok1	Humhum
<b>Radio</b>	« Mettre la radio » 2	Ok1 1	mettrradio	mettrradio
	« Plus fort la radio » 1	Tourne comme 2	plusfort Tourne	Plusfort Tourne
	« Moins fort la radio » 1	Tourne autre sens Cava2 2	Moinsfort Tourne autre sens	Moinsfort Tourne autre sens
	« Changer de station » 2	yaFRINT ou ok1 1	yaFRINT ou oui1	yaFRINT ou oui2
	« Eteindre la radio » 2	Ouais Tourne 1	Voila2 Tourne	Daccord2 Tourne

## Annexe 5. Matériel et scénario du prétexte expérimental

Voici les *pitchs* du prétexte constituant le scénario expérimental ainsi que le script des étapes pour réguler les différents effets tels que l'amorçage dans les étapes du scénario.

NB: il existe des variantes adaptées du scénario pour les cas où les personnes ne sont pas accompagnées par exemple – ce dernier a fait l'objet d'ajustements.

### Protocole expérimental EmOz – EEE

#### Cadre et prétexte d'étude

Dans cette étude, nous sollicitons des personnes âgées qui ont préservé un bon degré d'autonomie, elles peuvent faire l'objet de petites désorientations ou présenter de légères « fragilités » qui ne sont pas incapacitantes, nécessitant généralement un suivi par un proche (famille ou ami) ou un professionnel du domaine médico-social (plus particulièrement aide à domicile ou auxiliaire de vie) qui sera sollicité dans le cadre de cette recherche. Cet accompagnant sera complice des expérimentateurs et il aura connaissance des objectifs réels des tâches.

Lieu : DOMUS. Il s'agit d'une plateforme expérimentale prenant la forme d'un appartement intelligent pour la santé (HIS) équipé d'un coin cuisine / salle à manger, une chambre, une salle d'eau et un salon, le tout entièrement meublé et équipé.

À l'heure actuelle, nous ne sommes pas encore en mesure de mettre en place des installations technologiques (comme complément de la vie quotidienne) directement au sein du domicile des personnes. Cela nécessite donc de faire déplacer les sujets en dehors de leur environnement familial, pour qu'ils puissent accéder à des lieux équipés visant à répondre à leurs besoins de santé<sup>111</sup>. L'une de ces infrastructures est donc le HIS, qui sera ici doté de multiples capteurs et actionneurs dont les commandes se feront par le biais de la parole.

Ce changement de lieu n'est pas sans conséquence, amplifiant dans un premier temps la désorientation des personnes, similaire à celui que nous rencontrons lors d'un déménagement vers de nouveaux environnements de vie (ex : institutions, aménagement du domicile avec équipements spécifiques type HAD<sup>112</sup>...etc.).

Le prétexte d'expérimentation est donc le suivant : observer si l'insertion d'« objets repères »<sup>113</sup>, en possession des sujets, favorise l'habitation à un nouvel environnement spécialement aménagé pour une personne âgée « fragilisée ». Nous demandons donc pour cela aux sujets âgés de venir dans cet appartement avec une dizaine de ces « objets repères » (ou qui peuvent être considérés comme tels par la personne) et d'aménager ce cadre de vie à leur façon pour ensuite y passer un petit peu de temps. Une liste de ces objets sera faite durant les entretiens de recrutement pour éviter de les perdre durant l'expérimentation et permettre leur rendu sans oubli.

L'étude consisterait alors à comparer le niveau d'habitation à cet appartement entre deux groupes de personnes âgées : l'un qui apporte des « objets repères », l'autre venant sans objets. Cette évaluation serait faite *via* un questionnaire factice dont les réponses ne nous intéressent pas directement, mais dont les questions seront un prétexte pour faire expliquer à la personne âgée, par ses propres mots, comment fonctionne DOMUS et voir sous quelles formes communicatives elle s'adresse à Emox. Cependant, les sujets ne seront pas au courant que les commandes des actionneurs de l'habitat ne se font pas de façon ubiquitaire par la régie. Elles passent par le biais du robot Emox avec

<sup>111</sup> Santé au sens de l'OMS : « état de complet bien-être physique, mental et social, et ne consiste pas seulement en une absence de maladie ou d'infirmité ». Source : Préambule à la Constitution de l'Organisation mondiale de la Santé, tel qu'adopté par la Conférence internationale sur la Santé, New York, 19-22 juin 1946; signé le 22 juillet 1946 par les représentants de 61 États. 1946; (Actes officiels de l'Organisation mondiale de la Santé, n° 2, p. 100) et entré en vigueur le 7 avril 1948.

<sup>112</sup> HAD : hospitalisation à domicile

<sup>113</sup> « Objets repères » : à l'image des objets de vie courante ou décoratifs servant souvent de points d'appui aux malades d'Alzheimer souvent utilisé comme aide-mémoire pour des tâches, des mots, de localisation...

lequel ils seront amenés à interagir (l'observation de cette interaction étant le véritable but de l'expérience).

Un micro-cravate recueillera l'ensemble des productions vocales de la personne âgée et les scènes d'expérimentation seront enregistrées *via* les caméras de DOMUS.

Les commandes domotiques de DOMUS ainsi que le robot EMOX seront commandés par un protocole de type magicien d'Oz.

Un consentement pré-expérimental est proposé en suivant le prétexte de l'expérimentation et sera par la suite complété d'un consentement post-expérimental divulguant l'ensemble des modalités du protocole.

### **Scénario**

*(\*) Le jour de l'expérimentation, le sujet et l'accompagnant arrivent ensemble sur les lieux de l'expérimentation. L'expérimentateur 2, qui aura recruté les personnes, se chargera d'amener les personnes, mais il prétendra ne pas connaître les lieux de l'expérimentation, c'est son responsable de stage qui aurait planifié le passage des personnes dans DOMUS et que c'est donc la première fois qu'il vient.*

#### **Expérimentateur 2 (à l'interphone)**

« Bonjour, je suis à la recherche de Mme VVV (Expérimentateur 1). On a dû réserver une plateforme, un appartement pour une expérience avec des personnes âgées. »

*(\*) L'expérimentateur 2 qui prétend être l'ingénieur en charge de l'appartement accueille le sujet âgé, son accompagnant et l'expérimentateur 1.*

#### **Expérimentateur 1**

« Ah bonjour, Mlle YYY? (Expérimentateur 2) ? Je suis Mme VVV, l'ingénieur de la plateforme. »

#### **Expérimentateur 2**

« Bonjour, Mme VVV (Expérimentateur 1). Je suis Mlle YYY, c'est pour mes études qu'on vous a contacté. Je vous présente Mme/M. XXX notre participant et Mme AAA son aide à domicile. »

#### **Expérimentateur 1**

« Votre encadrant m'a contacté pour me dire que vous aviez besoin de la partie appartement. Venez, c'est par ici, je vais vous montrer, mais avant posons-nous un instant que je vous explique, vous voulez un café ou autre chose? On va faire le point quelques minutes donc mettez-vous à l'aise. »

*(\*) La montée aux étages se fait par l'ascenseur et on fait attention aux rebords de porte pour éviter les chutes de la personne.*

*(\*) L'expérimentateur 1 explique à quoi sert ce lieu et demande à l'expérimentateur 2 en quoi consiste son étude pour focaliser les gens sur la tâche prétexte qui est de mesurer le degré d'habitation des personnes à un habitat inconnu quand ce dernier apporte et dispose des objets qui lui sont familiers dans un lieu inconnu.*

#### **Expérimentateur 1**

« Ici on fait toute sorte d'études, et on travaille souvent avec des personnes âgées aussi, on a une partie appartement vous verrez, qui a été conçue pour ça. Maintenant, c'est vrai qu'on fait des aménagements, mais on ne peut pas encore équiper les appartements privés des personnes, en général c'est des vieux bâtiments et on ne peut pas. Donc on a des lieux comme ça qui permettent de tester les aménagements et ensuite les personnes pourront déménager pour bénéficier de ces nouveaux équipements. Parce qu'équiper au fur et à mesure c'est cher et de toute façon on ne peut pas équiper les appartements avec les choses qu'on propose. Mais, du coup, je ne sais pas trop que vous allez faire, on ne m'a pas expliqué. »

#### **Expérimentateur 2**

« Et bien justement, c'est ce sur quoi je travaille ces questions de déplacement des personnes. En fait je suis étudiante en gérontologie et dans le cadre de mon stage je fais une étude sur les effets de désorientation qu'un déménagement ou un changement de lieu de vie cause, parce qu'à un moment ou un autre on est amené à déplacer les personnes. »

**Expérimentateur 1**

« Ah oui c'est vrai, on le voit souvent ça. Mais du coup, qu'est-ce que vous allez faire ici, c'est des choses qui m'intéresse ça. »

**Expérimentateur 2**

« En fait on a fait des études et on a une hypothèse qui porte sur des objets. On pense que dans notre quotidien il y a des objets auxquels on tient, et qui peuvent nous aider à réduire ces effets de désorientation. Comme on ne sait pas lesquels, on essaie de trouver. Donc on demande aux personnes qu'elles sélectionnent elles-mêmes une dizaine d'objets dans leur quotidien, facilement transportables, et on leur demande de les placer dans les environnements typiques dans lesquels on peut déplacer les personnes pour faciliter leur quotidien, on m'a dit qu'ici on propose ce type d'environnement »

**Expérimentateur 2**

« Oui oui c'est bien ça, vous allez voir, je vais vous montrer après. »

*(au sujet âgé)* « Donc c'est vous qui avez choisi ces objets. Et donc comment vous avez choisi ? »

*(à expérimentateur2)* « Bon moi, je ne sais pas du tout sur quels critères vous faites ça, mais bon... »

**Expérimentateur 1**

« Vous avez certainement dû recevoir des réservations pour un collègue, mais l'idée c'est qu'on compare des personnes qui apportent ce type d'objets et d'autres qui n'en apportent pas pour voir les différences. Du coup ma collègue est en charge des personnes qui n'en apportent pas et moi de celles qui en apportent. »

**Expérimentateur 2**

« Ah d'accord je comprends mieux, c'est pour ça qu'il y a eu autant de réservations. »

*(\*) L'expérimentateur2 fait le curieux pour préciser la nature de l'isolement de la personne, sans être explicite et en posant des questions d'ordre général qui aboutissent sur des mini-récits de vie.*

**Expérimentateur 2**

*(à l'aide à domicile)* « Et donc vous, vous accompagnez Mme/Mr XXX c'est ça? Ça fait longtemps que vous intervenez chez Mme/Mr XXX? Comment ça se passe chez vous? Comme on pense à des aménagements, je me permets d'être un peu curieuse. »

*[données d'entretien]*

**Expérimentateur 2**

« Bon bah allez, et bien le mieux c'est que je vous montre comme c'est fait ici. Venez, je vais vous montrer l'appartement. »

*(\*) Le sujet, l'accompagnant et les expérimentateurs 1 et 2 rentrent dans DOMUS. L'expérimentateur2 commence à faire la visite de l'appartement.*

*(\*) En régie d'autres expérimentateurs suivaient les échanges et s'assuraient du bon fonctionnement du robot jusqu'à l'entrée dans l'appartement.*

**Expérimentateur 2**

« Voilà donc cet appartement. Alors, vous avez la salle à manger avec son petit coin-cuisine, la chambre avec sa télévision, et au fond une salle d'eau. Par là, vous avez un petit salon tout équipé. »

*(\*) L'accompagnant reçoit un « faux appel d'urgence » sur son téléphone portable professionnel à ce moment-là. Il s'agit du prétexte qui lui permet de quitter DOMUS. L'aide à domicile a été prévenue qu'elle recevrait cet appel au moment de rentrer dans l'appartement. Les expérimentateurs de régie réexpliquent brièvement les consignes et donnent la réplique à l'accompagnant.*

**Expérimentateur de régie**

« Bonjour Mme AAA (accompagnant), je suis le collègue de YYY (expérimentateur2). »  
« Il faut que vous partiez de l'appartement maintenant avec YYY (expérimentateur2). »  
« On va prétendre que je suis votre responsable et que je vous appelle pour une urgence, pas très loin du campus. »  
« C'est pas long, mais il faut vous faire accompagner par YYY (expérimentateur2), car vous ne savez pas où vous êtes et vous n'avez pas de voiture. Ne vous inquiétez pas, YYY (expérimentateur2) et VVV (expérimentateur1) vous donneront la réplique. Dites juste que c'est [votre association/entreprise/foyer] et que vous devez intervenir pour une urgence. »

**Expérimentateur1 ou 2**

« Qu'est-ce qu'il se passe, est-ce que tout va bien? »

**Accompagnant**

« Non en fait c'était ma chef [association/entreprise/foyer de l'accompagnant]. Ils m'appellent pour une urgence pas très loin d'ici, ils ont personne d'autre, il faut vraiment que j'y aille, c'est à LLL pas très loin d'ici... »

**Expérimentateur1**

« Mais comment vous allez faire ? C'est quoi le problème ? C'est grave ? »

**Expérimentateur2**

« Ah mais oui, c'est vrai c'est moi qui vous ai emmené en voiture, comment on va faire ? Vous devez vraiment partir ? »

**Accompagnant**

« Bah oui là c'est vraiment urgent, j'ai pas vraiment le choix... »

*l'expérimentateur 2 est embêté, l'expérimentateur 1 est un peu agacé de la situation.*

**Expérimentateur2**

« Si c'est vraiment une urgence... comment on peut faire ? Parce qu'il faudrait que je vous accompagne non ? »

**Expérimentateur1**

« Ah oui oui, ici c'est complètement isolé. Mais à LLL, c'est pas loin, mais il faut y aller en voiture. »

**Expérimentateur2**

« Est-ce que ce serait possible que je vous laisse le temps d'accompagner Mme AAA et qu'on revienne le plus vite possible ? Ce n'est pas pour longtemps c'est ça ? De toute façon la première étape de l'expérience il fallait que je laisse seul(e) Mme/Mr XXX (le sujet) seul pour qu'il/elle dispose les objets dans l'appartement. »

*(au sujet)* « Est-ce qu'on peut faire ça, Mme/Mr XXX (le sujet)? Ça a l'air urgent et on n'en aurait vraiment pas pour longtemps ? »

*(à expérimentateur1)* « Est-ce je peux laisser Mme/Mr XXX ? Le temps d'aller et revenir ? »

**Expérimentateur2**

« Bah moi j'ai des consignes pour vous montrer le fonctionnement de l'appartement. Après il est réservé toute l'après-midi pour vous donc vous faites ce que vous voulez, mais moi je peux pas rester, ce n'était pas prévu. Je devais juste vous montrer comment tout ça fonctionne. Donc oui, mais il faut que je montre à Mme/Mr XXX (le sujet) »

*(\*) après accord du sujet et expérimentateur1, l'expérimentateur2 se propose d'emmener l'accompagnant si c'est vraiment une urgence. Ceux-ci disent revenir le plus rapidement possible,*

*l'expérimentateur2 rappelant à la personne âgée les consignes pour ses objets personnels avant de partir.*

**Expérimentateur2**

*(au sujet) « Du coup, est-ce que ça vous dérangerait de commencer sans moi ? Il faudrait en attendant qu'on revienne que vous sortiez vos objets et que vous les disposiez dans l'appartement. Regardez, vous pouvez mettre ça ici, sur la liste que vous avez peut-être commencé à remplir avec les objets que vous avez apportés (feuille de support). Si ce n'est pas fait, il faudrait aussi remplir cette colonne »*

**Accompagnant**

*« Bon là il faut vraiment qu'on y aille... »*

*(\*) L'expérimentateur2 et l'accompagnant partent donc en régie, laissant l'expérimentateur1 et le sujet âgé terminer leur visite. L'expérimentateur 1 et l'accompagnant n'ont alors pas eu le temps de voir le fonctionnement spécifique de l'appartement. C'est l'expérimentateur1 qui prend les commandes de l'interface EmOz pour dérouler le scénario scientifique des sons.*

*(\*) L'expérimentateur1 explique ensuite que l'appartement fonctionne par commandes vocales, mais de façon spécifique. Il se tourne vers la chambre où se trouve le robot EMOX et l'appelle. Le robot était disposé de telle manière que ni l'expérimentateur1 ni l'accompagnant ne l'ont vu lorsqu'ils sont rentrés dans DOMUS. EMOX rejoint l'expérimentateur 2 et le sujet âgé.*

**Expérimentateur 1**

*« Cet appartement est un peu spécial, les interrupteurs et l'électricité... il y a pas, tout ça ne marche pas comme d'habitude, il y a un moyen...»*

*(\*) L'expérimentateur1 fait semblant de chercher et prend le micro sur la table de la cuisine. Quand il s'adresse au robot, il parle dans le micro.*

**Expérimentateur 1**

*« Euh il est où, ah oui voilà (en trouvant micro) EMOX ! (dans le micro). Ah le voilà.»*

*(\*) EMOX qui était caché dans un coin de la cuisine avance et apparaît devant le sujet pour la première fois. Il n'émet pas de sons et bouge très peu, uniquement pour donner un feedback qu'il a bien entendu.*

*(\*) Le nom EMOX ne sera donné que s'il est demandé par le sujet ou il peut être introduit de manière neutre, sans focus la première fois qu'on le sollicite.*

**Expérimentateur 1**

*« Voilà c'est cette chose, cet objet, qui pourra vous aider à faire les choses pour vous dans cet appartement et tout passe par la voix. Vous avez qu'à lui dire et c'est lui qui fait.*

*(\*) L'expérimentateur 1 sort une liste « mode d'emploi » des commandes vocales et explique qu'il suffit de dire des phrases de cette liste à EMOX. Il ne donnera en revanche pas d'exemples sur la manière de faire une commande. Il dit aussi que la commande a besoin de passer par une phase d'apprentissage et le port d'un micro-casque.*

**Expérimentateur 1**

*« Par contre, comme on a mis une techno pas trop chère, dessus, il faut d'abord qu'il apprenne votre voix. La mienne il la connaît, mais du coup, pour l'utiliser il faudrait que vous mettiez ça (en désignant le micro-casque et la place sur le sujet, en n'oubliant pas d'allumer le récepteur). Ensuite il faut faire au moins une fois toutes ces commandes, et après c'est bon, tout marchera »*

*(\*) Les expérimentateurs de régie vérifient que le récepteur du micro est bien allumé. S'il ne l'est pas, ils ne font pas réagir le robot ni les actionneurs domotiques. Si problème technique du robot ou de l'appartement, le signal est la lampe de LED de l'appartement qu'on passe au rouge. Si vraiment le message n'est pas compris des expérimentateurs on utilise le module de conversion de voix du robot et si cela ne marche pas non plus, un expérimentateur va jouer le rôle d'un ingénieur-étudiant (autre que l'expérimentateur2) et va demander à son chef (expérimentateur1) en le sollicitant pour un problème en simulant que ça rien à voir avec l'expérience, mais qu'il a besoin de l'expérimentateur2 pour un problème urgent dans une salle à côté. Les précautions sont prises pour que le sujet n'ait pas accès aux explications de ce problème technique.*

#### **Expérimentateur 1**

« Voilà, maintenant attendez je vais lui dire – EMOX début de l'apprentissage – voilà, là c'est bon il va apprendre votre voix. »

*(\*) L'expérimentateur1 regarde que le robot a bien réagi pour poursuivre la consigne où il fait attention de ne pas faire lui-même une commande qui puisse influencer le format de l'énonciation de la commande par le sujet.*

#### **Expérimentateur 1**

« Donc maintenant si vous voulez allumer la lumière par exemple, il faut lui dire ce qui est marqué là. Allez-y, essayez avec la première. »

*(\*) Le sujet dit la première commande, la lumière s'allume, le robot bouge, mais n'émet pas de son). L'expérimentateur1 s'assure que la réaction de Domus est normale, que le sujet a compris.*

#### **Expérimentateur 1**

« Voilà ça marche. Et si vous voulez éteindre, vous faites la suivante (la lumière s'éteint) et vous faites toute la liste et c'est bon. Bon, je vous laisse parce que j'ai un autre travail qui m'attend. Faites juste toutes les commandes. L'appartement est à vous donc posez vos objets où vous voulez, tout est à votre disposition, donc utilisez ce que vous voulez, faites comme chez vous. Il y a des gâteaux, du thé... N'hésitez pas à vous servir, c'est pour vous. Je repasserai plus tard, mais Mme AAA et Melle YYY ne devraient pas tarder, je pense.»

*(\*) L'expérimentateur 2 repart en régie.*

*(\*) Nous laissons la personne âgée évoluer seule dans son nouvel environnement entre 15 à 30 minutes environ, puis nous demandons à l'accompagnant de rejoindre le sujet. L'accompagnant sera amené à interroger la personne âgée sur l'appartement, son fonctionnement. Il essaiera de faire reproduire au sujet l'ensemble des commandes qui se trouvent sur la liste des commandes.*

#### **Accompagnant**

« Ah, me voilà Mme/M. XXX L'expérimentateur2 arrive, il a été retenu par VVV. Alors tout s'est bien passé ? Vous avez pu voir comment fonctionne cet appartement ? Vous avez posé vos objets ? »

*(\*) L'accompagnant demande à la personne de lui montrer comment tout ça marche. Le sujet est donc amené à produire à nouveau les commandes vocales, au maximum. Il n'oublie pas le rappel des objets.*

*(\*) Après un certain temps, l'expérimentateur2 retourne dans Domus avec son questionnaire d'entretien qu'il avait convenu de remplir avec la personne lorsqu'il avait fait son recrutement. Il laisse les commandes du robot à l'expérimentateur1.*

*(\*) L'expérimentateur 2 prétend ne pas connaître le robot EMOX et que dans les études précédentes il fallait juste parler à haute voix à l'appartement pour faire toutes les commandes. Il demande donc à la personne âgée de lui montrer comment tout ça fonctionne.*

#### **Expérimentateur 2**

« Me voilà Mme/M. XXX. Alors comment ça va ? Est-ce que tout s'est bien passé avec l'ingénieur ? Je suis vraiment désolée pour ce contretemps, je devais régler quelques petites choses. Alors comment est-ce que vous trouvez cet appartement ? Comment ça marche ? Est-ce que vous avez eu le temps de poser tous vos objets ?

*(\*) Le sujet, l'accompagnant et l'expérimentateur1 échangent, au bout d'un moment EMOX se manifeste. L'expérimentateur 1 remarque le robot et demande ce que c'est. Il demande ensuite comme l'accompagnant comment tout ça fonctionne. Il demande « par curiosité » comment le sujet trouve cet environnement et le robot.*

#### **Expérimentateur 2**

« Tiens, qu'est-ce que c'est ? Comment ça marche cette chose ? Vous me montrez, je ne connais pas du tout ? »

#### **Expérimentateur 2**

« Bon bah je vous propose de nous mettre dans le salon et qu'on s'assoie pour que vous m'expliquiez un peu comment vous avez placé vos objets et comment vous les avez choisis. On peut prendre un thé si vous voulez. »

*(\*) Ces paroles de l'expérimentateur2 indiquent que l'expérimentateur1 est arrivé au bout du protocole pour rentrer dans le débriefing. En fonction des événements, l'expérimentateur1 peut revenir plus vite dans DOMUS. Le contrôle du robot est laissé aux expérimentateurs de régie jusqu'aux révélations.*

#### **Expérimentateur 1**

« Je viens comment ça se passe, ça va tout s'est bien passé ? »

XXXXXXXX

#### **Entrée de débriefing**

L'expérimentateur1 rentre alors dans DOMUS et entame le débriefing en deux étapes. Il se joint aux personnes pour savoir si l'appartement a bien fonctionné. Il commence par faire une « fausse révélation en disant que le but de l'expérience était de savoir comment le sujet trouvait l'appartement et de voir la praticité de la commande domotique vocale, mais sans faire de focus sur le robot. Après explicitation non consciente du sujet, la véritable révélation est réalisée.

Nous expliquerons qu'en réalité toute l'expérience a été une mise en scène. Nous proposons donc au sujet et à son accompagnant de prendre le temps d'expliquer les véritables buts de l'expérimentation soit dans l'appartement, soit dans une salle à part. Puis, un véritable entretien a lieu sur le ressenti de l'expérience et sur l'interaction avec Emox. Le sujet récupère ses affaires personnelles (en vérifiant la liste de *check-up*).

Nous prévenons le sujet que nous avons enregistré ses productions orales, que des caméras filmaient l'appartement et que toutes les actions du robot, ainsi que celles de l'appartement étaient simulées en régie. Nous expliquerons que nous avons dû utiliser cette méthode pour des raisons scientifiques afin de ne pas influencer les comportements du sujet et le recueil de données spontanées. Nous redemandons alors l'accord de la personne pour l'utilisation des données enregistrées à des fins de recherche uniquement par l'intermédiaire d'un deuxième document de consentement post-expérimental. Toute l'équipe est ici présentée pour des raisons de transparence totale à la divulgation des prétextes.

Nous demanderons par ailleurs aux personnes d'auto-annoter leurs propres productions orales ainsi que leurs comportements en visualisant les données récoltées et cela en présence de l'expérimentateur qui servira d'aide technique pour la visualisation de ces données. Cela est réalisé durant un entretien semi-directif après coup, plus tard chez la personne, à un moment où la mémoire du sujet est encore bien reliée à son vécu (cette dernière information n'est pas explicitée au sujet). Les accompagnants

peuvent eux aussi faire l'objet d'un entretien semi-directif autour du vécu de l'expérience et sur leur avis sur ce type de technologies s'ils en ont le temps.

Ceci est la feuille de route remplie lors de chaque entretien de recrutement pour chaque sujet.

### **Fiche sujet**

#### **Introduction**

*Je vous remercie d'avoir accepté de vous entretenir avec moi. En tout cas si vous acceptez de participer à notre expérience, sachez que cela va contribuer à l'enrichissement de la recherche pour l'amélioration de la qualité de vie des seniors, et en particulier de deux thèses. Le sujet qui nous intéresse s'inscrit dans le cadre de travail d'une recherche autour des technologies qui permettent de faciliter le quotidien des personnes de plus de 70 ans. Mais avant de commencer j'ai besoin de vous connaître un peu plus pour pouvoir anonymiser les réponses que vous allez nous donner.*

#### **Informations personnelles**

Sexe :  Femme  Homme      Age : .....

Vivez-vous :       Seul(e)       En couple       En famille

Le dernier diplôme obtenu : .....

La profession ou activité que vous avez exercée le plus longtemps :

.....

Quelle est votre langue maternelle ? Celle de vos parents ?

.....

Où avez-vous vécu le plus longtemps (origines culturelles de la personne)?

.....

Langues / accents ? .....

#### **Transition**

*Maintenant on va passer au vif du sujet. Il n'y a pas de bonnes ou mauvaises réponses aux questions que je vais vous poser donc laissez libre cours à votre imagination, c'est votre avis qui nous intéresse.*

#### **Rapport à la technologie**

Qu'est-ce que les technologies pour vous ?

Etes-vous intéressé(e) par les technologies ? Placez votre degré d'intérêt sur cette échelle de 1 à 10 :

1   2   3   4   5   6   7   8   9   10

Lesquelles vous intéressent-elles pour vous aider dans votre quotidien ?

Utilisez-vous ces technologies chez vous ? Lesquelles ?

Dans quel contexte ou avec quel but les utilisez-vous ?

Que vous permettent-elles de faire ?  
Y-a-t-il des technologies que vous n'utilisez pas ? Lesquelles ?  
Pourquoi n'utilisez-vous pas celles-ci ?

**Explication de la tâche prétexte**

*En fait, nous vous avons posé ces questions sur ces technologies car nous avons justement créé un habitat intelligent qui est équipé de multiples technologies pour faciliter le quotidien des personnes âgées. A l'heure actuelle nous ne sommes pas encore capables d'équiper sur place les logements des personnes avec ces technologies de maintien à domicile, par contre il est envisagé de pouvoir faire déménager les personnes dans des appartements spécifiquement équipés pour qu'ils puissent y rester autonomes. Cela par contre pose le problème de la désorientation car l'habitué des personnes à des lieux de vie nouveaux ne semble pas facile. En parallèle, certaines études suggèrent que les personnes associent leur lieu de vie familier à certains objets auxquels ils sont particulièrement attachés. Ainsi nous souhaitons tester cette hypothèse en vérifiant si venir vivre dans un habitat technologisé nouveau est plus facile si les personnes viennent y disposer leurs objets les plus familiers. Si vous acceptez de participer à notre étude, vous viendrez tester l'appartement pilote technologisé en y apportant vos objets personnels que vous aurez sélectionnés, facilement transportables en venant vivre une ou deux heures dans notre appartement pilote. Ensuite, nous ferons le point ensemble sur vos impressions, et nous comparerons vos avis avec ceux d'un autre groupe de personnes venues aussi dans cet habitat pilote, mais sans aucun objet familier.*

**Prise de rendez-vous avec les personnes**



## Notice d'information

**Chercheur(s) titulaire(s) responsable(s) scientifique(s) du projet :** *Nom Prénom*

**Période de recrutement prévisionnelle :** *Date*

**Lieu de recherche :** LIG – Laboratoire d'Informatique de Grenoble – bâtiment CTL

Vous trouverez le détail de la localisation à : [http://www.liglab.fr/spip.php?page=mot&id\\_mot=586](http://www.liglab.fr/spip.php?page=mot&id_mot=586)

### **But du projet de recherche :**

Le projet vise à recueillir des avis de personnes travaillant sur les problématiques gérontologiques concernant un environnement domotique spécifiquement développé pour des personnes âgées. L'étude veut évaluer à terme, le degré d'habitation des personnes âgées à un environnement aménagé technologiquement et qui leur est non familier. L'expérience qui vous est proposée vérifiera la pertinence d'un de nos critères pour mesurer ce degré d'habitation en vous demandant de suivre le même protocole que ce que nous envisageons avec les personnes âgées.

En effet, de précédentes études ont montré que les nouvelles technologies de l'information et de la communication (NTIC) dédiées à des personnes âgées peuvent générer d'importantes perturbations et des difficultés d'habitation, en particulier si ces technologies font partie d'un environnement non familier. Par exemple, une nouvelle technologie de téléalarme domotique mise en œuvre dans une structure de type EPHAD ou dans des résidences aménagées a posé des problèmes d'usage pour les personnes âgées, et cela de façon encore plus importante lorsque ces personnes venaient d'emménager depuis peu de temps dans ce nouveau lieu de vie.

### **Ce que l'on attend de vous :**

Si vous acceptez de participer à cette étude, vous effectuerez un séjour d'une heure ou deux dans un prototype d'appartement équipé spécifiquement pour des personnes âgées. Vous devrez également apporter une dizaine de petits objets personnels facilement transportables (objets décoratifs, livres...etc.) que vous disposerez à votre guise dans l'appartement expérimental pour vous aider à prendre vos marques dans ce lieu (cela fait partie de nos conditions pour tester l'habitation des personnes à un nouvel environnement).

Un entretien avec un/des expérimentateur(s) sera effectué en début (environ 10 minutes) et en fin d'expérience (environ 30 minutes) pour recueillir vos avis.

### **Contact :**

Pour confirmer votre participation à l'étude et prendre un rendez-vous, merci de contacter par mail ou par téléphone :

*Nom Prénom - Statut*

Mail :

Téléphone :

*Vous pouvez également poser des questions au sujet de la recherche en tout temps en communiquant avec le responsable du projet de recherche *Nom Prénom* par courrier électronique à : *contact**

Nous vous remercions d'avance de votre participation !

**Remarque:** Ceci est le modèle de la notice d'information rappelant le prétexte aux sujets avec les contacts des expérimentateurs. Ils ont également une feuille de pré-consentement à participation expérimentale ainsi qu'une grille à pré-remplir ou à remplir lors de l'expérience pour compléter la liste des objets qu'ils apportent.

Ce document est un modèle du formulaire de pré-consentement expérimental rempli par les sujets suivant les contenus de prétexte qui leur ont été présentés.

### Formulaire de consentement pré-expérimental

**Titre du projet de recherche :** Plateforme expérimentale EmOz

**Investigateurs principaux :**

**Investigateurs associés**

**Responsable scientifique :**

**Lieu de recherche :** Laboratoire d'Informatique de Grenoble - Bâtiment CTL, plateforme Domus

**But du projet de recherche**

Ce projet vise à développer et améliorer les technologies facilitant la vie des personnes âgées fragilisées dans le cadre de leur maintien à domicile.

**Ce que l'on attend de vous (méthodologie)**

Si vous acceptez de participer à cette étude, vous effectuerez un séjour (1h30 environ) dans un appartement expérimental en amenant quelques objets familiers. Un entretien avec un expérimentateur sera effectué en début (15 minutes) et en fin d'expérience (30 minutes). Vous remplirez un questionnaire d'évaluation de cet habitat (environ 20 minutes à une heure). Votre participation requiert ainsi une demi-journée.

**Vos droits à la confidentialité**

Les données obtenues seront traitées avec la plus entière confidentialité. On voilera votre identité à l'aide d'un numéro séquentiel attribué. Aucun renseignement qui pourrait révéler votre identité ne sera dévoilé. Toutes les données seront gardées dans un endroit sécuritaire et seul le Responsable Scientifique y aura accès.

*Le Responsable Scientifique, [Nom Prénom](#) sera la personne auprès de laquelle vous pourrez exercer votre droit de consultation, de modification ou d'opposition des données vous concernant.*

**Vos droits de vous retirer de la recherche en tout temps**

Votre participation à cette recherche est volontaire. Vous pouvez vous en retirer ou cesser votre participation en tout temps, et vous pouvez demander que vos données soient détruites, sans conséquence.

**Avantages**

Les avantages attendus de cette recherche sont d'obtenir une meilleure compréhension des facteurs qui influencent l'habitation des personnes à un environnement qui leur est inconnu. Une meilleure compréhension de ces facteurs pourra entre autres contribuer à améliorer les outils servant au maintien à domicile des personnes âgées.

**Risques possibles**

À notre connaissance, cette recherche n'implique aucun risque ou inconfort autres que ceux de la vie quotidienne. Vous pourrez éventuellement être confronté à des désagréments techniques mineurs qui n'auront aucune conséquence sur vos actions, ne serait-ce que sur la lenteur d'exécution ou l'efficacité de certaines tâches que vous serez amené à exécuter.

**Diffusion**

Les résultats de cette recherche seront diffusés dans des colloques scientifiques et seront publiés dans des actes de colloque et des articles de revue académique.

**Vos droits de poser des questions en tout temps**

Vous pouvez poser des questions au sujet de la recherche en tout temps en communiquant avec le responsable du projet de recherche [Nom Prénom](#) par courrier électronique à : [contact](#)

**Consentement à l'utilisation des données pour la recherche**

En signant le formulaire de consentement, vous certifiez que vous avez lu et compris les renseignements ci-dessus, que nous avons répondu à vos questions de façon satisfaisante.

Participant : J'ai lu et compris les renseignements ci-dessus et j'accepte de plein gré l'utilisation de toutes les données produites par l'expérimentation et me concernant à des fins de recherche uniquement.

*Ce formulaire est signé en deux exemplaires dont une copie sera conservée par le Responsable Scientifique et l'autre copie fournie au sujet.*

Date et lieu :

Nom – signature :

Ceci est une grille de support fournie aux sujets afin qu'ils puissent réaliser leur tâche prétexte autour des objets. Ils peuvent la compléter partiellement avant l'expérience et finir de la remplir durant l'expérience, notamment lorsqu'ils ont fini de tester les commandes domotiques.

Nom :

Prénom :

Coordonnées :

	<b>Objet choisi</b>	<b>Emplacement</b>
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		

Voici le « mode d'emploi » de l'appartement que les personnes âgées sont amenées à suivre pour faire fonctionner l'appartement Domus correctement.

## **Les commandes de la cuisine**



### ***les spots au plafond***

- « Mettre la lumière »
- « Eteindre la lumière »

### ***les stores***

- « Descendre les stores »
- « Monter les stores »

### ***la bouilloire***

- « Mettre la bouilloire »
- « Eteindre la bouilloire »

## Les commandes de la chambre

### *les spots au plafond*

- « Mettre la lumière »
- « Eteindre la lumière »

### *les stores*

- « Descendre les stores »
- « Monter les stores »

### *les rideaux*

- « Fermer les rideaux »
- « Ouvrir les rideaux »

### *les lampes de chevet*

- « Mettre les lampes »
- « Eteindre les lampes »

### *la TV*

- « Mettre la télé »
- « Plus fort la télé »
- « Moins fort la télé »
  
- « Eteindre la télé »

### *Les lumières colorées au plafond*

- « Mettre la lumière jaune »
- « Mettre en bleu »
- « Mettre en vert »
- « Eteindre cette lumière »



## Les commandes du salon



### *les spots au plafond*

- « Mettre la lumière »
- « Eteindre la lumière »

### *la radio*

- « Mettre la radio »
- « Plus fort la radio »
- « Moins fort la radio »
- « Changer de station »
- « Eteindre la radio »

Ceci est le formulaire de consentement post-expérimental qui est signé par les sujets une fois le débriefing terminé, à la levée de l'ensemble des prétextes de l'expérience. Les sujets rapportent ce document pour le lire tranquillement chez eux, le signent et le rendent lors de la séance d'auto-annotation. Si le sujet abandonne l'expérience sans faire l'auto-annotation, l'expérimentateur doit se déplacer pour aller le récupérer afin de pouvoir conserver et traiter les données expérimentales.

**Formulaire de consentement post-expérimental**

**Titre du projet de recherche :** Plateforme expérimentale EmOz

**Investigateurs principaux**

**Investigateurs associés :**

**Responsable scientifique :**

**Lieu de recherche :** LIG– Bâtiment CTL, plate-forme Domus

**But du projet de recherche**

Ce projet vise à concevoir et observer comment un robot compagnon, médiateur de commandes de l'habitat, pourrait créer un lien socio-affectif et ainsi participer au maintien socio-relational des personnes, typiquement le cas des personnes âgées fragilisées, dans le cadre de leur maintien à domicile.

L'expérience a permis de constituer un corpus multi-sources en récoltant des données orales et audiovisuelles de vos interactions avec un robot-compagnon au sein d'un habitat intelligent commandé vocalement.

**Description de l'expérimentation**

Le prétexte que nous avons évoqué pour vous faire passer cette expérience était un leurre : l'apport d'outils familiers dans l'habitat n'est pas la motivation de cette étude, c'est-à-dire que nous n'évaluons donc pas votre degré d'habitation à l'appartement. Les expérimentateurs et votre aide à domicile vous ont joué la comédie, l'appel d'urgence était un faux appel qui était en réalité un prétexte pour vous laisser poursuivre l'expérience seule en compagnie d'Emox, puis avec Emox et votre aide à domicile, puis avec en plus l'expérimentateur recruteur. Notre but était d'observer comment vous vous adressez à Emox pour passer les commandes, et quel ressenti vous avez pour ce robot compagnon.

Nous avons en fait simulé et contrôlé toutes les commandes de l'appartement et du robot afin de pouvoir interagir avec vous à distance. Vos paroles ont été enregistrées et vos déplacements ont été filmés.

**Justification de la méthodologie**

Puisque nous voulions savoir si Emox était acceptable et comment il était perçu, il était essentiel que vous ne sachiez à l'avance que c'était notre intention, car cela aurait modifié

vos rapports avec Emox, d'autant plus si vous aviez su qu'en réalité Emox était téléguidé par nous et ne parlait pas ni ne bougeait pas de lui-même.

### **Vos droits à la confidentialité**

Les données obtenues seront traitées avec la plus entière confidentialité. On voilera votre identité à l'aide d'un numéro aléatoire. Aucun renseignement qui pourrait révéler votre identité ne sera dévoilé. Toutes les données seront gardées dans un endroit sécuritaire et seul le Responsable Scientifique y aura accès.

Conformément à la loi informatique et libertés, vous disposerez d'un droit d'accès, de rectification et d'opposition aux données vous concernant. Cette étude suit la procédure simplifiée de la loi informatique et libertés (loi n° 78-17 du 6 janvier 1978, relative à l'informatique, aux fichiers et aux libertés, modifiée en 2004) de la CNIL (Commission nationale de l'informatique et des libertés).

*Le Responsable Scientifique, **Nom Prénom** sera la personne auprès de laquelle vous pourrez exercer votre droit de consultation, de modification ou d'opposition des données vous concernant.*

### **Avantages**

Les avantages attendus de cette recherche sont de comprendre comment une aide par un robot compagnon peut d'une part rendre plus naturelle la commande des technologies d'un habitat quand il en est le médiateur, et d'autre part observer si un attachement peut se créer avec cet objet s'il devient intelligent et doté d'émotions, et si cet attachement peut être un confort supplémentaire dans le maintien à domicile.

### **Risques possibles**

À notre connaissance, cette recherche n'implique aucun risque ou inconfort autres que ceux de la vie quotidienne.

### **Diffusion**

Cette recherche sera diffusée dans des colloques scientifiques et elle sera publiée dans des actes de colloque et des articles de revue académique. Conformément à la loi du 09 août 2004, à la fin de l'étude, je peux demander à l'investigateur une synthèse des résultats globaux de la recherche. Nous vous proposons de nous laisser des coordonnées mail auxquelles nous vous ferons parvenir toutes les publications relatives à votre participation, accompagnées d'un texte de description vulgarisée spécialement dédié aux volontaires non scientifiques participant à cette expérience.

**Vos droits de poser des questions en tout temps**

Vous pouvez poser des questions au sujet de la recherche en tout temps en communiquant avec le responsable du projet de recherche **Nom Prénom** par courrier électronique à : [contact](#)

**Consentement à la participation**

En signant le formulaire de consentement, vous certifiez que vous avez lu et compris les renseignements ci-dessus et que nous avons répondu à vos questions de façon satisfaisante.

Je soussigné(e) Nom..... Prénom.....

déclare avoir compris le but et les modalités de l'étude qui m'ont été pleinement expliqués par **Noms des expérimentateurs**. En particulier j'ai bien compris que la tâche prétexte des objets familiers était un leurre et que le but de l'expérience était d'observer l'interaction avec le robot Enox.

Les informations relatives au principe de cette étude et de son intérêt m'ont été communiquées. Des réponses ont été apportées à mes questions. J'ai disposé d'un délai de réflexion avant de prendre ma décision.

J'ai reçu une information et un consentement spécifiques pour la vidéo dans le cadre du droit à l'image.

J'ai lu et compris les renseignements ci-dessus et j'accepte de plein gré que l'on utilise les données récoltées durant ma participation à cette expérimentation, et cela, uniquement à des fins de recherche.

*Ce formulaire est signé en deux exemplaires dont une copie sera conservée par le Responsable Scientifique et l'autre copie fournie au sujet.*

**Date et lieu :**

**Nom de l'intéressé(e) :**

**Date et lieu :**

**Nom de l'investigateur :**

**Signature précédée de lu et approuvé**

**Signature précédée de lu et approuvé**

Cette fiche est fournie en guise de pense-bête aux aides à domicile et accompagnants complices de l'expérience suite à l'entretien qui précède le recrutement des personnes âgées. Elle est fournie lors des premières phases du processus de recrutement où les partenaires sociaux et les aides à domicile sont consultés pour le recrutement des sujets, pour également prendre en notes des précautions à prendre auprès du potentiel futur sujet.

### **Consignes accompagnant**

=> Le but pour nous est de faire interagir le plus possible la personne avec le robot. Lister les commandes est un prétexte pour que la personne s'adresse au robot.

Donc on essaie de faire dire ces commandes d'abord avec l'expérimentateur 2 puis avec vous, puis avec l'expérimentateur 1, mais tout ça dans un scénario ou ça devrait paraître naturel, sans qu'elle se rende vraiment compte qu'on essaie de lui faire répéter ces commandes !!!!

#### **1 : votre arrivée dans l'appartement**

Dès l'arrivée sur le lieu de l'expérimentation, au 3e étage du bâtiment CTL, nous entrerons tous ensemble (vous, la personne âgée, l'expérimentateur 1 recruteur (Yuko) et l'expérimentateur 2 (Véronique, simulant être l'ingénieur de l'appartement) dans l'appartement Domus.

#### **2 : vous devez partir en urgence**

D'abord quelques échanges où l'expérimentateur expliquera que Domus ne peut pas fonctionner avec des interrupteurs et des télécommandes classiques, mais par commande vocale, qu'il est nécessaire de vous montrer à tous comment ça marche avant de vous laisser seuls. Mais elle n'a pas le temps de commencer la démonstration, car vous allez alors recevoir un appel, officiellement de votre responsable Mme/M. TTT. qui a besoin de vous en urgence sur place à l'agence, mais en réalité c'est un chercheur caché de l'équipe qui a fait sonner votre téléphone et vous a eu au bout du fil en simulant vous parler comme votre responsable Mme/M. TTT. Vous direz alors que vous avez absolument besoin de quelqu'un pour vous emmener en voiture à cause de l'urgence, mais que par contre ça devrait être réglé très vite, en ¾ d'heures, maximums. L'expérimentateur 2 recruteur se proposera alors de vous emmener et vous partirez donc ensemble. En réalité, vous irez toutes les deux dans la régie, la personne âgée restant avec l'expérimentateur 1 qui va lui expliquer le fonctionnement de l'habitat, c'est-à-dire du robot en lui faisant dire toutes les commandes, pour un (faux) apprentissage pour que le robot la reconnaisse. ATTENTION dans ce prétexte le robot n'obéit qu'à la personne qui lui a demandé de lui obéir et qui vient de lui apprendre sa voix grâce à la liste des commandes – le robot répondant en réalité par un téléguidage depuis la régie.

#### **3 : attente et préparation de votre prochaine intervention**

En régie vous pourrez suivre l'expérience et vous assisterez à la suite du scénario. Nous vous donnerons la liste de ces commandes, car dans une phase suivante, vous aurez donc, comme l'expérimentateur 2 à vous les faire démontrer par la personne.

#### **4 : votre retour dans l'appartement**

Quand nous vous en donnons la consigne, vous irez rejoindre la personne âgée dans l'appartement. Souvenez-vous que vous devez faire semblant de revenir de votre urgence !!!! Vous direz que l'expérimentateur qui vous a emmené arrive, qu'elle a été retenue par l'un des ingénieurs pour régler de petits problèmes techniques.

Vous lui demanderez alors si elle a pu voir comment disposer les objets. Vous en discutez un moment. Si la personne ne vous parle pas spontanément du robot, vous demandez comment marche cet habitat, qu'est-ce que lui a expliqué l'ingénieur. Si elle vous propose de faire apprendre votre voix au robot, dites-lui que vous préférez qu'elle vous montre elle, car ça peut être un truc sympa pour aider les personnes âgées et que vous aimeriez bien voir si un robot sait se mettre au service d'une personne âgée. Si elle vous montre spontanément le robot, c'est impeccable. Là, il faudra essayer de lui faire dire le plus de commandes possible, donc vous pouvez lui demander la feuille des commandes et lui demander de les faire exécuter, car ça vous amuse de voir le robot obéir et commander tous ces objets dans l'habitat, en prenant l'air surpris qu'on allume une TV, change de chaîne, change la couleur des lumières, etc.!! Essayez de faire produire un maximum de commandes vocales à la personne en lui demandant de vous montrer comment les différents éléments de l'appartement marchent, sachant que vous n'êtes censé(e) ne rien connaître, ni des commandes vocales, ni de l'appartement, ni du robot.

Enfin, proposez de prendre un thé à la personne en attendant que l'expérimentateur arrive. Tout sera dans les placards, mais pour allumer la bouilloire il faudra que la personne appelle Emox et fasse une commande vocale.

#### **5 : entretien à trois**

Enfin l'expérimentateur arrive enfin, boit le thé avec vous, discute à bâtons rompus de l'habitat et surtout bien sûr des objets, car c'est le but officiel. Notons qu'elle n'est PAS CENSEE savoir qu'un robot commande l'habitat. Elle simule croire que l'habitat est à commande vocale « en l'air » sans robot.

Elle commence alors un entretien de ressenti sur les objets familiers dans cet appartement inconnu. À un moment de ce questionnaire, noté dans le texte du questionnaire (que nous vous joignons aussi à ce mail), soit la personne a déjà parlé du robot (et on aura interrompu le questionnaire) soit c'est l'expérimentateur qui fait semblant de découvrir le robot (grâce aussi au questionnaire qui amène la question). Et là, comme vous, l'expérimentateur prend l'air surpris et intéressé, et demande une démo, vous pouvez aussi en rajouter et dire que vous avez vu la démo et c'est vraiment étonnant !!

À la fin on dit que c'était une caméra cachée et l'expérimentateur 1 arrive et explique la supercherie !!!

Et là alors on fait le vrai questionnaire qui est à la fin du fichier joint ☺

Ceci est le formulaire de consentement que chaque accompagnant de sujet signe en tant que complices des expérimentateurs, différent de celui du sujet. Au moment de l'acceptation du protocole par le sujet, à la levée des prétextes.

### Formulaire de consentement

**Titre du projet de recherche :** Plateforme expérimentale EmOz

**Investigateurs principaux**

**Investigateurs associés :**

**Responsable scientifique :**

Lieu de recherche : Laboratoire d'Informatique de Grenoble – Bâtiment CTL, plate-forme Domus

**But du projet de recherche**

Ce projet vise à concevoir et observer comment un robot compagnon, médiateur de commandes de l'habitat, pourrait créer un lien socio-affectif et ainsi participer au maintien socio-relationnel des personnes, typiquement le cas des personnes âgées fragilisées, dans le cadre de leur maintien à domicile.

**Ce que l'on attend de vous (méthodologie)**

Si vous acceptez de participer à cette étude, vous accompagnerez une personne âgée auprès de laquelle vous intervenez dans votre activité professionnelle, pendant 1 heure 30 dans un appartement "expérimental", au sens où une régie permet d'observer de l'extérieur un appartement de 3 pièces dans lequel boutons et télécommandes sont remplacés par des commandes vocales données à un petit robot compagnon. C'est le lien entre la personne que vous accompagnez et le petit robot qui sera observé. Mais cette personne ne sait pas que c'est ce lien qui intéresse les expérimentateurs, qui lui font croire que son séjour dans Domus sert à tester comment on s'habitue plus facilement à un habitat nouveau quand on apporte avec soi des objets familiers. Votre rôle sera ainsi d'être complice des expérimentateurs dans ce "leurre". Ils vous expliqueront comment vous devez intervenir au fur et à mesure de l'expérimentation. Un entretien avec un expérimentateur sera effectué en début (15-20 minutes) et en fin d'expérience (30-45 minutes) ou lors d'un rendez-vous ultérieur.

A la fin de l'expérience, vous pourrez solliciter toutes les questions qui vous seront venues pendant l'expérimentation. Vous pourrez si vous le souhaitez revenir sur votre consentement

après avoir été informée à nouveau des objectifs de l'étude, de la méthodologie, des contraintes et des éventuels risques prévisibles.

### **Vos droits à la confidentialité**

Les données obtenues seront traitées avec la plus entière confidentialité. Nous masquerons votre identité à l'aide d'un numéro aléatoire. Aucun renseignement qui pourrait révéler votre identité ne sera dévoilé. Toutes les données seront gardées dans un endroit sécuritaire et seuls le chercheur principal et les chercheur(e)s adjoint(e)s y auront accès.

Conformément à la loi informatique et libertés, vous disposerez d'un droit d'accès, de rectification et d'opposition aux données vous concernant. Cette étude suit la procédure simplifiée de la loi informatique et libertés (loi n° 78-17 du 6 janvier 1978, relative à l'informatique, aux fichiers et aux libertés, modifiée en 2004) de la CNIL (Commission nationale de l'informatique et des libertés).

*Nom de l'investigateur principal sera la personne auprès de laquelle vous pourrez exercer votre droit de consultation, de modification ou d'opposition des données vous concernant.*

### **Vos droits de vous retirer de la recherche en tout temps**

Votre participation à cette recherche est volontaire. Vous pouvez vous en retirer ou cesser votre participation en tout temps, et vous pouvez demander que vos données soient détruites, sans conséquence.

### **Avantages**

Les avantages attendus de cette recherche sont d'obtenir une meilleure compréhension des facteurs qui influencent l'habitation des personnes à un environnement qui lui est inconnu. Une meilleure compréhension de ces facteurs pourra entre autres contribuer à améliorer les outils servant au maintien à domicile des personnes âgées.

### **Risques possibles**

À notre connaissance, cette recherche n'implique aucun risque ou inconfort spécifique, outre les risques habituels de la vie quotidienne dans son propre habitat.

### **Diffusion**

Cette recherche sera diffusée dans des colloques scientifiques et elle sera publiée dans des actes de colloque et des articles de revue académique. Nous vous proposons de nous laisser des coordonnées mail auxquelles nous vous feront parvenir toutes les publications relatives à

vos participation, accompagnées d'un texte de description vulgarisée spécialement dédié aux volontaires non scientifiques participant à cette expérience.

**Vos droits de poser des questions en tout temps**

Vous pouvez poser des questions au sujet de la recherche en tout temps en communiquant avec le responsable du projet de recherche **Nom Prénom** par courrier électronique à : [contact](#)

**Consentement à l'utilisation des données pour la recherche et à la participation**

En signant le formulaire de consentement, vous certifiez que vous avez lu et compris les renseignements ci-dessus et que nous avons répondu à vos questions de façon satisfaisante.

Je soussigné(e) Nom..... Prénom.....

Déclare avoir compris le but et les modalités de l'étude qui m'ont été pleinement expliqués par **Noms des expérimentateurs**. En particulier j'ai bien compris que la tâche prétexte des objets familiers était un leurre et que le but de l'expérience était d'observer l'interaction de la personne, dont je suis aide à domicile, avec le robot Emox.

Les informations relatives au principe de cette étude et de son intérêt m'ont été communiquées. Des réponses ont été apportées à mes questions. J'ai disposé d'un délai de réflexion avant de prendre ma décision.

J'ai reçu une information et un consentement spécifiques pour la vidéo dans le cadre du droit à l'image.

J'ai lu et compris les renseignements ci-dessus et j'accepte de plein gré que l'on utilise les données récoltées durant ma participation à cette expérimentation, et cela, uniquement à des fins de recherche.

*Ce formulaire est signé en deux exemplaires dont une copie sera conservée par le Responsable Scientifique et l'autre copie fournie au sujet.*

**Date et lieu :**

**Nom de l'intéressé(e) :**

**Nom de l'investigateur :**

**Signature précédée de lu et approuvé**

**Signature précédée de lu et approuvé**

**Date et lieu :**

Ceci est le modèle de document de droit à l'image pour l'exploitation des données anonymisées des sujets. Si l'un l'un des consentements parmi les trois proposés ainsi que ce droit à l'image viennent à manquer, les données sont effacées et non traitées par le laboratoire.



Madame, Monsieur

Vous allez participer à une expérimentation pour une recherche universitaire menée par l'équipe Getalp et la plateforme Domus - CTL du laboratoire LIG, dans le cadre d'un projet de recherche, et nous vous en remercions. Pour garantir vos droits à la vie privée, votre droit à l'image et au son, nous vous demandons de bien vouloir donner votre consentement explicite.

Pour préserver ces droits, l'équipe de recherche s'engage à assurer la confidentialité des données filmées vous concernant. La conservation des données sera réalisée dans le laboratoire LIG dans les meilleures conditions de sécurité.

- Acceptez-vous que ces données soient utilisées par l'équipe de recherche ?  
(cochez la case correspondante)

OUI

NON

- Acceptez-vous que vos données soient réutilisées pour tout projet dont la finalité est une finalité de recherche à l'exclusion de toute exploitation à des fins commerciales ?  
(cochez la case correspondante)

OUI

NON

Je reconnais avoir donné mon consentement le

Nom et prénom :

Signature :

*Réservé*

## Annexe 6. Grille d'entretien de débriefing

Voici la feuille de route pour les expérimentateurs afin d'effectuer la phase de débriefing en deux temps.

### Grille d'entretien semi-directif

#### *Partie personne âgée*

***AVANT le DÉBRIEFING et AVANT que l'expérimentateur1 demande à voir Emox***

#### **Objets personnels et représentations associées**

Pouvez-vous me parler des objets que vous avez apportés ?

Que représentent-ils pour vous ?

Pour quelles raisons est-ce que vous les avez choisis ?

#### **Lieu de vie**

D'habitude, où est-ce que se trouvent ces objets chez vous ?

Pouvez-vous me parler de votre lieu de vie ?

Comment est-ce que vous avez disposé les objets ici (dans DOMUS) ?

Pourquoi ?

Et notre appartement, comment l'avez-vous trouvé ?

Avez-vous l'impression qu'avoir disposé vous-mêmes des objets que vous avez choisis peut vous aider à vous familiariser plus vite avec cet appartement qui vous est étranger ?

#### **Rapport à l'accompagnant**

Depuis combien de temps avez-vous une aide à domicile ?

Depuis combien de temps connaissez-vous Mme \*\*, votre aide à domicile actuelle ?

Pour quelles raisons avez-vous eu recours à ces services ?

Comment décririez-vous son travail ? Que fait-elle chez vous ?

Comment qualifieriez-vous votre relation avec Mme \*\* ?

Comment vivez-vous le fait d'avoir partagé cette expérimentation avec elle aujourd'hui ?

#### **Rapport aux technologies**

Comment est-ce que vous décririez cet appartement (DOMUS) ?

Est-ce que vous utilisez de nouvelles technologies au quotidien ?

Que pensez-vous de ces commandes de l'appartement par la parole ?

Ça marche bien ?

Ça ne vous paraît pas trop étrange de parler l'air à un appartement ?? Moi j'avais testé l'appartement au début, j'ai eu du mal à m'habituer à parler en l'air, au plafond ou aux murs !!! Etc.

*Est-ce que vous pouvez me montrer comment ça marche ? (...)*

**ATTENTION : APRÈS Emox**

Sans lire la feuille

Je ne connaissais pas du tout, comment vous le trouvez ?

Je me demande ce qu'il sait faire, s'il est intelligent ou pas ???

Je ne sais pas si j'aimerais en avoir un chez moi, au Japon y a plein de robots partout, mais c'est qu'en France je n'en vois jamais, vous, ça vous plairait ??

Il est trop marrant, je trouve qu'il ressemble... ben je ne sais pas à quoi en fait, à quoi il ressemble...

Il a l'air plutôt sympa, mais y a peut être des gens qui ne le trouvent pas sympa, enfin moi je viens de le voir, vous du coup vous êtes avec lui depuis 2h, ça vous a plu ??

**DÉBRIEFING**

**ATTENTION : APRÈS le DÉBRIEFING**

Comment ressentez-vous cette farce qu'on vous a jouée ?

Vous comprenez j'espère qu'on ne pouvait pas vous dire qu'on était derrière le robot sinon ça marchait pas !!!

Alors finalement maintenant que vous savez que c'est votre avis sur le petit robot qui nous intéresse, comment vous l'avez trouvé ?

Ça vous a paru naturel d'interagir avec lui ou difficile ?

Je sais que ce n'est pas toujours facile de le comprendre, c'est quelque chose qu'on doit améliorer, mais qu'est-ce que vous diriez sur sa façon de s'exprimer ?

Si vous aviez dû commander l'appartement par des paroles en l'air dans la pièce, est-ce que vous auriez trouvé ça plus ou moins facile, plus ou moins agréable ?

Qu'avez-vous aimé chez lui ? Qu'avez-vous moins apprécié ?

On appelle ces robots des robots-compagnons, vous trouvez que ça peut tenir un peu compagnie ?

Aimeriez-vous en avoir un chez vous ? Le même ou différent ?? Pourquoi ?

*Pour finir, qu'est-ce que le progrès pour vous ? (fin sur note positive)*

**Partie accompagnant**

**Profession d'aide à domicile**

Comment décririez-vous votre travail ?

Comment est-ce que s'organise votre temps de travail ? Vos déplacements ?

Combien de personnes est-ce que vous prenez en charge ?

Qu'est-ce que vous appréciez dans votre métier ? Qu'est-ce que vous appréciez moins ?

**Rapport à la personne âgée**

Depuis combien de temps vous aidez la personne ?

Comment est-ce que vous définiriez votre relation avec cette personne ?

Comment vivez-vous le fait d'avoir partagé l'expérimentation avec elle aujourd'hui ?

**Rapport aux technologies**

Est-ce que vous utilisez de nouvelles technologies au quotidien ?

Comment est-ce que vous décririez cet appartement (DOMUS) ?

Que pensez-vous des commandes que l'on doit faire vocalement ?

Comment avez-vous trouvé ce fonctionnement ?

Comment trouvez-vous ce robot ? Qu'en pensez-vous ?

Est-ce que vous le verriez dans le quotidien d'une personne âgée ou votre propre quotidien ?

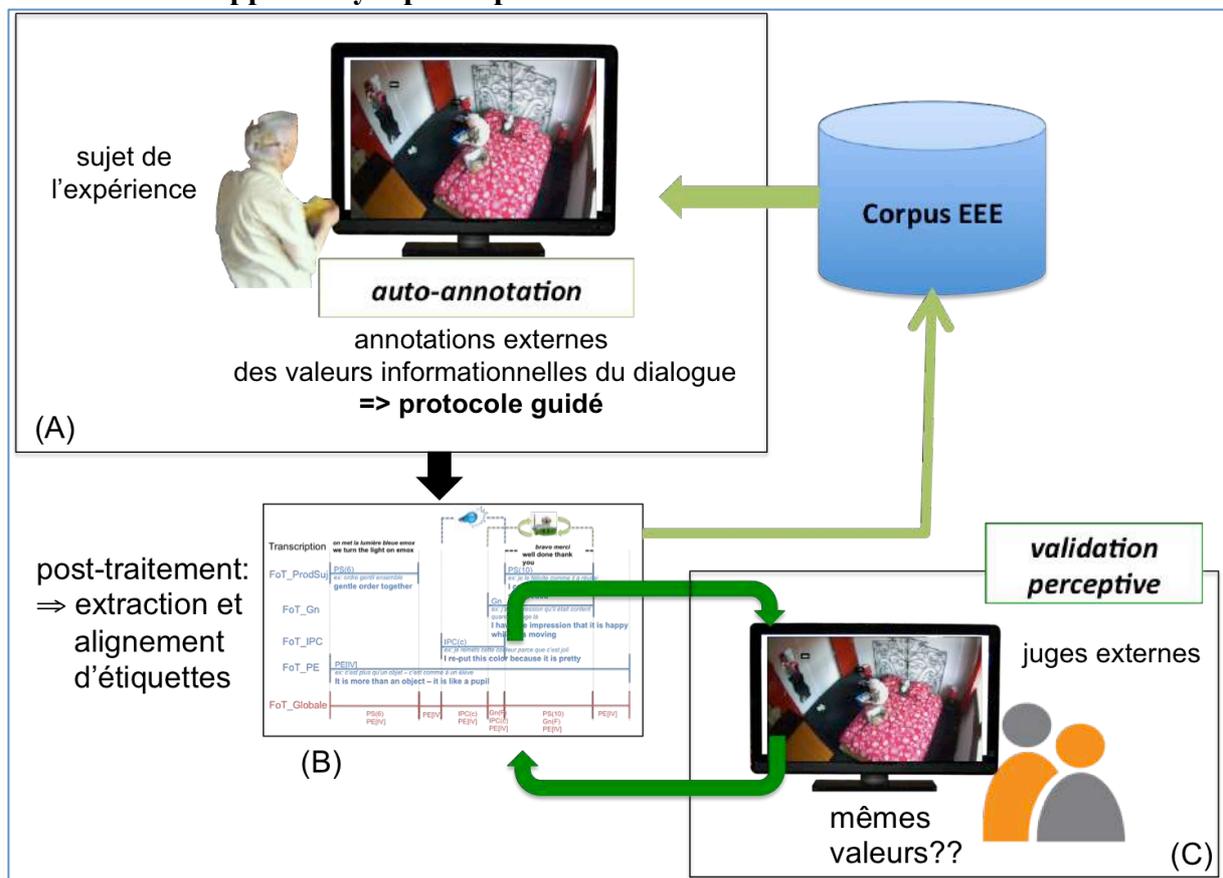
Quelle place pensez-vous que ce type de technologie pourrait avoir vis-à-vis de votre profession ?

## Annexe 7. Feuille de route de l'auto-annotation

Voici une feuille de route pour l'expérimentateur afin de pouvoir effectuer l'auto-annotation. Chaque séance est enregistrée, soit en audio, soit en audiovisuel, en fonction du focus expérimental qui est fait pour chaque séance. La feuille de route de la présente annexe permet la réalisation de l'étape (A) schématisée dans la Figure ci-dessous. Elle ne peut être réalisée que si le sujet expérimental annote lui-même les données qu'il a produites dans un protocole qui induit des comportements écologiques et spontanés par rapport à des hypothèses spécifiques.

L'auto-annotation est effectuée par deux annotateurs partageant la même interface sur deux ordinateurs séparés, mais avec le même pré-séquençage (un transcripneur et un interrogateur). Les données doivent être prétraitées en amont. Ici les séances ont été réalisées avec le logiciel Elan.

### Rappel du synopsis expérimental autour de l'auto-annotation



### **Introduction de séance:**

=> ce temps doit être réservé à définir quel est l'état de la personne au moment de la séance, pour penser à prévoir de proposer des pauses en fonction ou de faire attention à sa posture, son état de déshydratation, de fatigue etc.

=> Elle permet également d'introduire les annotateurs, notamment le transcripateur que le sujet âgé ne connaît pas, dans le but de la rassurer. En revanche, il ne faut pas faire trop de focus sur cette personne, et lui permettre de s'installer discrètement, en évitant de rentrer dans l'angle de vue de la personne (en général elle se place sur le côté du second d'annotateur, mais pas derrière ou en face), pour éviter son effet d'observateur.

Pour cette partie, l'annotateur principal (qui est le recruteur du sujet et qui a suivi la personne tout au long de l'expérience) introduit la séance de la manière suivante.

\*\* Vérifier si la personne a bien prévu un temps suffisamment long pour cette séance (en général il faut prévoir une après-midi entière de séance – une séance ne pouvant être réalisée sur plusieurs jours, mais obligatoirement en une seule fois).

\*\* Demander comment va le sujet âgé, son état physique notamment (il ne faudra pas hésiter à effectuer des micropauses en essayant de viser des parties de vidéo plus passive d'un point de vue interactionnel).

\*\* Interroger sur ce qu'il s'est passé depuis le passage de la personne dans l'appartement expérimental (cela peut également être un point qui peut être développé en fin de séance en fonction de la réponse, car la personne peut évoquer les effets longitudinaux de l'expérience).

### **Pitch de consignes:**

Vous avez participé à cette expérience dans l'appartement intelligent. Notre objectif est de pouvoir comprendre ce qui s'est passé tout au long de l'expérience, mais de manière très détaillée et on va donc vous faire voir la vidéo de l'expérience qui a été sauvegardée quand vous êtes venu sur notre plateforme, mais progressivement, étape par étape, pas à pas. Alors il ne va pas tout y avoir, juste la partie que vous avez passée dans l'appartement, mais dans tout ce qu'on va vous montrer, on n'a rien coupé, c'est vraiment tout ce qui s'est passé. En fait, on veut vraiment que vous vous souveniez, pas à pas, ce que vous avez vécu et on va s'appuyer sur ce qu'on appelle votre mémoire autobiographique. En réalité, même s'il s'est passé un certain temps depuis que vous êtes venu dans cet appartement, vous avez gardé en mémoire tout ce qui s'est passé. Lorsqu'une personne se revoit en train de faire une action, elle arrive à se souvenir exactement et de manière précise dans quel état elle était à ce moment-là. C'est comme si vous étiez en train de revivre l'expérience et vous allez essayer de vous souvenir des pensées, des choses qui sont passées par votre tête au fur et à mesure, mais très précisément à chaque moment de l'expérience. On va donc s'arrêter après chaque chose que vous dites pour que vous nous disiez ce qui s'est passé. Si vous voyez ou vous entendez des choses qui vous font vous souvenir de quelque chose que vous vous étiez dit à vous-même, arrêtez-nous et essayez de nous décrire ce qui s'est passé.

***Remarque:*** *L'auto-annotation est une manière d'accéder à des représentations mentales de diverses Feeling of Thinking. Elle n'est donc parfois pas facilement verbalisable et peu donc faire l'objet de dessin, de mouvement, de gestuel, d'oralisation spécifique. En fonction de la nature de l'expérience, il peut donc être intéressant d'avoir des consignes particulières telles que la reproduction de gestualité ou de productions vocales etc. Dans quel cas, ces parties doivent être formalisées à la fin de la séance d'auto-annotation avec un enregistrement vidéo de la séance (ici ce n'était pas le cas pour la personne âgée). Dans le*

*cas où la séance d'auto-annotation récolte ce mini-corpus, la levée du prétexte expérimental peut être décalée à la fin de l'auto-annotation et non avant l'auto-annotation pour le maintien du détournement d'attention. En effet, une auto-annotation ne peut être réalisée qu'une seule fois, car elle modifie le sujet donc le protocole doit bien être formalisé avant chaque séance.*

### **Conditionnement initial:**

Avant la toute première commande, une petite partie de l'amorçage est visionnée pour faciliter l'accès à la mémoire autobiographique. La vidéo commence au moment où les complices sont sortis et que la personne se retrouve seule avec l'ingénieur de la plateforme et le robot. L'annotateur demande si le sujet arrive à se souvenir de cette partie pour essayer d'accéder à l'état d'esprit général dans lequel était le sujet pour conditionner ce dernier à la suite de l'annotation. On s'arrête avant l'émission de la première commande.

### **Interrogations itératives d'étiquettes:**

La séance se déroule par itération successive d'interrogations pour extraire l'étiquette d'auto-annotation par une méthode proche de la maïeutique. Une séquence comportant une commande vocale est montrée au sujet et l'on demande par des questions de type:

**\*\*** Qu'est-ce qu'il s'est passé à ce moment-là? Qu'est-ce que vous vous êtes dit?

**\*\*** Comment est-ce que vous faites? Si vous devez mettre des mots sur la manière dont vous dites la commande, comment vous diriez que vous faites?

=> L'annotateur sait que ce qu'il veut récupérer sont les FoT du sujet, donc tant que le sujet fait de l'interprétation générique et qu'il ne révèle pas une étiquette basée sur un souvenir décrivant correctement l'intention communicative de sa commande et son état d'esprit, on continue sur la même commande en rejouant la séquence s'il le faut. Si une étiquette ne vient pas on peut l'aider par les stratégies suivantes:

**\*\*** Est-ce que c'est comme (*une étiquette qu'il aurait dite précédemment, en réutilisant les mots que le sujet a utilisés*) ou c'est autre chose, c'est différent?

=> L'interrogateur peut prendre des notes au fur et à mesure pour garder en tête les étiquettes qui sont produites par le sujet pour pouvoir les réutiliser.

=> Le transcripteur peut réviser les frontières de séquence en fonction de ce que dit la personne et de noter les indices permettant de préciser le moment où intervient une étiquette.

**\*\*** Écoutez et regardez bien comment vous faites. Nous on ne sait pas comment vous fonctionnez d'habitude, mais vous vous connaissez bien. En vous regardant est-ce que ça vous rappelle une manière de faire, une façon de dire quelque chose? À qui et dans quelle condition vous feriez ça? À quoi est-ce que ce serait similaire?

*NB:*

=> *il ne faut en aucun cas donner des exemples d'étiquettes, l'ensemble des auto-annotations doit être composé du vocabulaire du sujet.*

=> *On insiste jusqu'à l'étiquette satisfaisante, en rejouant la séquence autant de fois que nécessaire, mais une fois qu'on est passé à la séquence suivante, on ne revient jamais sur la précédente séquence. Aucune coupure ou saut de séquence n'est possible même si elle semble passive d'un point de vue interactionnel.*

### **Fin de séance:**

**\*\*** Demander s'il y a des remarques générales ou commentaires du sujet sur l'ensemble de l'expérience.

**\*\* Récupérer les consentements expérimentaux et droits à l'image (toujours avoir des copies en réserve, car la personne peut les avoir perdus entre-temps). Revoir le contenu des documents rapidement s'il le faut.**

**Remarque:**

Même si les annotations principales portent sur les FoT associées aux commandes, le sujet décrit des états cognitifs latents, l'impression sur le robot et les événements remarquables, avec les effets qu'ils ont pu avoir sur lui. Il faut donc également repérer ces éléments et pour les éléments latents, demander si cette étiquette est toujours valide ou si elle a changé de temps à autre pour vérifier l'étendue de cette dernière. À défaut, quand nous avons une étiquette de même nature consécutive, l'alignement de l'étiquette précédente est prolongé jusqu'à la survenue de l'étiquette suivante de même nature et de même ordre de grandeur temporel. Il y a ainsi un gros travail de post-traitement pour réaligner et extraire les étiquettes à l'issue de la séance (étape (B) de la première figure). Les étiquettes récoltées ne seront en définitive valides que si elles sont perceptivement validées par un nombre conséquent de juges externes sur des occurrences d'événements interactionnels filtrés et présentés avec un panel d'étiquettes à associer (étape (C) de la première figure).

### Annexe 8. Résumé des fiches sujets

<b>informations sujets</b>	<b>Sexe</b>	F (féminin) /M (masculin)
	<b>Age</b>	valeur (en année)
	<b>Situation</b>	seul(e) / en couple
	<b>Diplôme</b>	valeur descriptive
	<b>Profession</b>	valeur descriptive
	<b>Lieu de vie principal</b>	1 : sans diplôme / sans profession 2 : sans diplôme / profession artisanale/manuelle 3 : diplôme général profession artisanale/manuelle 4 : diplôme général profession intellectuelle 5 : diplôme spécialisé profession artisanale/manuelle 6 : diplôme spécialisé profession intellectuelle
	<b>Langues /accents</b>	valeur descriptive
	<b>Fragilisation physique</b>	valeur descriptive
	<b>Type fragilisation physique</b>	classiques (arthrose, respiration, hypertension) / implants articulaires / motrice / visuelle / auditive
	<b>AD (0/1)</b>	0 : sans aide / 1: avec aide
	<b>Famille</b>	absente / visites rares (moins d'une fois par semaine) / visites fréquentes (+ d'une fois par semaine) / aidante
	<b>Type logement</b>	domicile / foyer
	<b>Infos Supplémentaires</b>	valeur descriptive

## Nomenclature du codage des informations des sujets

Informations personnelles								
Sexe	Age	Situation	Niveau CSP	Fragilisation physique	AD (0/1)	Famille	Domicile / foyer	Infos Supplémentaires (fragilisation/activité)
F	91	seule	6	implants articulaires	1	visites rares	foyer	plusieurs chutes, fracture de hanche, hospitalisations multiples
F	83	en couple	2	auditive	0	visites rares	domicile	vision très faible en couple (mais isolée) avec Sa2F
H	85	en couple	2	visuelle	0	visites rares	domicile	audio très faible, en couple (mais isolé) avec Sa2H
F	68	en couple	4	classique	1	visites fréquentes	domicile	mari très malade, personne âgée elle-même aidante
F	85	seule	6	classique	0	visites rares	foyer	très active, multiples activités, sans attache à la communauté âgée
F	84	seule	4	classique	0	visites rares	foyer	très active – a du mal avec les personnes âgées communautaires
F	83	seule	5	classique	1	visites rares	domicile	donne des cours chez elle, mais en dehors semble relativement seule en journée – compagnie de chat
F	83	seule	4	motrice	1	visites rares	domicile	fragilité de déplacement (déambulateur), mais sous gaz à oxygène pour problèmes respiratoires – très alerte
F	68	seule	4	implants articulaires	1	visites rares	domicile	gros problèmes physiques d'où retraite anticipée mais confronté aux mêmes problèmes d'isolement/fragilisation malgré âge
F	73	seule	6	implants articulaires	1	absente	domicile	très alerte, activité silencieuse (peinture), peu/pas de visites – quelques amis – gros problèmes de physique au niveau dos
F	78	seule	6	classique	1	visites rares	domicile	très active, multiples activités sans trop d'attache à la communauté âgée
F	78	seule	1	classique	0	aidante	domicile	problème général physique, principalement rhumatisme douloureux et très affectée par l'isolement/solitude (phase

								de deuil mari)
F	85	seule	4	classique	0	visites rares	foyer	relation principale = le fils, assez isolée dans ses activités quotidiennes routinières, épisode de solitude ressenti
F	93	seule	3	classique	0	visites fréquentes	foyer	environnement géographique isolé mais très bien entourée, mais début de contraintes d'isolement ressenti
F	77	seule	3	classique	0	visites rares	foyer	alerte mais quelques réticences pour activité intergénérationnelle, bien physiquement et autonome mais épisode de solitude ressenti
H	72	seule	3	classique	0	visites rares	foyer	très à l'aise en communauté et très alerte mais situation familiale un peu difficile au niveau relationnel
H	81	en couple	4	classique	0	visites rares	domicile	en phase de deuil, très alerte et sans problème physique apparent, mais assez fortement affectée par solitude et relationnel familial
F	74	seule	4	classique	0	visites rares	domicile	active dans activité bénévole pour enfants malades, isolement géographique contraignant
F	81	en couple	6	auditive	0	visites rares	foyer	assez critique envers la communauté âgée et jeune, sans gros problème physique, mais relationnel un peu difficile, car audition faible
F	78	seule	4	classique	1	visites rares	foyer	arrivée récente en foyer logement, alerte et sans problème apparent mais sujette au stress et épisode de solitude ressenti
F	90	seule	4	motrice	1	visites rares	domicile	phase de fragilisation transitionnelle (en cours de recalcul de GIR) – fragilisation physique et mobilité, alerte, mais situation d'isolement subi et ressenti
F	81	seule	4	classique	1	visites rares	domicile	en phase de dépression, en deuil d'époux avec apparition de propos suicidaires dans discours
F	73	seule	6	motrice	1	visites rares	domicile	problème neuro-moteur mais très alerte, handicapant pour la mobilité
H	85	en couple	5	auditive	1	visites rares	domicile	audition faible, alerte et sans trop de handicap physique, mais relationnel familial et communication difficiles avec épouse sourde suite à maladie

## Annexe 9. Procédures et règles de transcription

Voici la feuille de route des règles de transcription de la parole du corpus.

### **Introduction générale**

La transcription orthographique de la parole semble être une étape classique, mais non anodine. Ainsi, certaines précautions ont été prises dans le cadre de cette recherche où les productions de type micro-expressions vocales des sujets prennent une place d'importance. De fait, les méthodes semi-automatisées avec la transcription orthographique (permettant un gain de temps en transcription par de nombreux non-spécialistes) puis par un alignement forcé sur le signal n'a pas été des plus efficaces puisque : 1) les transcrip-teurs passent à côté de nombreuses micro-expressions même s'ils essaient de focaliser leur attention dessus puisque les outils les plus primitifs et culturellement persistants sont des phénomènes automatisés et intégrés qui shuntent les boucles attentionnelles de notre conscience; 2) un pré-découpage ou un alignement non temporalisé sont impossibles à réaliser, car le signal est très bruité du fait de la collecte en condition écologique et multi-locuteurs ; 3) les caractéristiques de la voix et l'articulation de la parole de la personne âgée imposent des contraintes qui complexifient autant le travail du transcrip-teur que les systèmes de reconnaissance vocale.

La solution retenue la plus efficace, mais qui reste tout de même chronophage est donc une transcription de la production vocale des sujets âgés, directement alignée avec le logiciel Praat. L'avantage de cet outil est la possibilité de visualiser le signal acoustique de manière précise avec son spectrogramme, permettant alors de vérifier visuellement qu'un événement acoustique a bien été produit (au minimum pour un non-spécialiste) et d'en observer la nature acoustique (pour un spécialiste). Ainsi, les micro-expressions vocales (bruits de bouche, rires, soupirs, vocalisations, inspirations et expirations marquées, mais également certaines interjections qui sont souvent oubliées par les transcrip-teurs) peuvent être correctement déterminées et transcrites, avec comme contrainte de n'étiqueter que les productions visibles sur le spectrogramme, puisqu'il peut s'agir d'un événement potentiellement détectable par un futur système de reconnaissance automatique.

### **Outils exploratoires utilisés :**

**\*\* Exploitation d'outils existants**

- SPPASS (Brigitte Bigi)

- WinPitch (Philippe Martin)

**\*\* Diarisation**

### **\*\* Contraintes techniques liées aux micro-expressions vocales**

*Du point de vue de l'ingénierie, sachant que les sources d'enregistrement sont variables (micro dans l'environnement, au plafond, sur un objet fixe ou mobile, porté par la personne) la sensibilité des outils de captures de ces productions ainsi que les caractéristiques acoustiques recueillies sont également variables. Cela suggère donc que certaines informations potentiellement pertinentes d'un point de vue communicatif peuvent être non détectées selon les cas, ajoutant des contraintes techniques qui doivent être considérées dans la cohérence du système exploitant ces données. Ainsi des comparaisons des signaux acquis par différentes sources pourraient préciser la nature des données exploitables. Si le système envisage de s'appuyer sur des procédés d'apprentissage machine, la construction de modèles intégrant ces micro-expressions vocales devrait considérer cette dimension.*

### **Segmentation (non obligatoire, mais facilite l'étape de transcription)**

Elle a la particularité d'être précisément alignée au niveau acoustique, notamment pour pouvoir exploiter proprement les durées des silences et des IPU (*Inter-Phrasal Units* =

zones sans signal acoustique). Elles contiennent également la production des micro-expressions vocales, orthographiées si possible comme dans le cas des interjections ou de sons phonétiques et/ou phonologiques ; alignées avec un descripteur acoustique s'il s'agit de sons plus primitifs, non phonologiques, ni phonétiques.

### ***Recommandations d'utilisation de SPPAS par Brigitte Bigi***

\*\*\*\*\*

Vous pouvez télécharger SPPAS : <http://www.lpl-aix.fr/~bigi/sppas/>

Lisez le document « SPPAS for dummies ».

1. Lancez SPPAS, ajoutez votre fichier audio dans la liste
2. Cliquez sur le texte « IPU segmentation », fixer les valeurs :
  - volume : 0 (=automatique)
  - durée minimale de parole : 0.300 (ou plus, selon le type de corpus)
  - durée minimale de silence : 0.200 (si c'est du français)
  - décalage aux frontières : 0.020
3. Cliquez sur la « baguette magique ». Si votre fichier audio est long, ça va prendre du temps, patience.
4. Une fenêtre va s'ouvrir avec un compte-rendu. Notez la valeur de volume qui a été choisie.
5. Ouvrez le fichier audio + le TextGrid créé avec Praat, et vérifiez que les segments de parole sont corrects (les '#' indiquent les silences).  
Si les segments sont vraiment trop nombreux ou pas assez nombreux, effacez le fichier Textgrid, et retournez à l'étape 2, mais cette fois, fixez la valeur de volume manuellement (par exemple le double ou la moitié).
6. Quand vous avez obtenu une segmentation en IPUs acceptable, copiez-collez la transcription orthographique dans les blocs de parole.
7. Lorsque vous avez l'audio + le TextGrid transcrit, il vous suffit de les ajouter dans SPPAS et de cliquer sur « tokenization », « phonetization », « alignment » et « syllabation » pour obtenir les segmentations en mots, phonèmes et syllabes.
8. Vous pouvez exporter le TextGrid en fichier « eaf », pour Elan.  
Si vous travaillez avec Elan plutôt qu'Anvil, vous pourrez ensuite utiliser les outils de Statistiques et de Requêtes/Filtrage de SPPAS.

\*\*\*\*\*

### ***Règles de transcription***

- \*\* la frontière de segmentation doit être vérifiée avant chaque transcription de séquence, à savoir elle regroupe un bloc de signal acoustique
- \*\* la transcription de la parole de chaque locuteur doit se faire sur un tiers distinct
- \*\* la transcription est orthographique quand elle a une forme lexicalisable (p. ex. les interjections), elle se fait entre [] pour les bruits de bouche. Les prélexicaux doivent prioritairement être annotés entre // avec une écriture phonétique de type SAMPA (cf. infra). Si l'annotateur n'est pas habitué à ces écritures phonétiques, il doit transcrire sous forme orthographique le bruit prélexical (p. ex. « euh » qui a plus ou moins une forme lexicalisée).
- \*\* pas de majuscules en début d'énoncé
- \*\* pas de ponctuations dans le corpus
- \*\* le bruit de bouche doit être décrit avec une description acoustique (codifiée au besoin) et dont le code doit être fourni pour chaque annotateur (p. ex. [insp] pour un bruit d'inspiration [occ bilabiale] pour une occlusive bilabiale etc.). Si l'annotateur n'est pas habitué à la description acoustique, il doit mentionner l'occurrence du bruit de bouche par le code [bb].

**Nomenclature finale normalisée des micro-expressions vocales (en cours de normalisation):**

**[BB: descripteur – description acoustique]** ex: [BB: clic – latéral, mouillé]

**[PreLex: /transcription API/]** ex: [PreLex: /muw/] [PreLex: /o:f/]

**[Interj: transcription orthographique commune]** ex: [Interj: voilà]

**Sous-segmentation et transcription phonétique (non-terminée pour le corpus actuel)**

Les éléments orthographiés sont également associés à leur réalisation phonétique en suivant la norme SAMPA (*Speech Assessment Methods Phonetic Alphabet*). Cette transcription phonétique est segmentée en plusieurs niveaux de granularité, dont les séquences sont générées automatiquement en utilisant le plugin EsayAlign du logiciel Praat. Deux des trois passes disponibles sur cet outil ont été appliquées :

- la **conversion de graphème à phonème** : elle s'appuie sur la transcription orthographique réalisée au préalable, afin d'en obtenir une transcription phonétique en norme SAMPA. Cette étape est complétée par une correction manuelle intermédiaire avant l'application de la seconde passe afin de corriger les problèmes de conversion phonétique. Elle est également précédée d'un nettoyage automatisé des micro-expressions vocales non-orthographiées.

- la **micro-segmentation** : elle réalise une segmentation de la transcription phonétique en trois niveaux, a) en mots, b) en syllabes, c) en phones.

## Annexe 10. Procédures de traitement du corpus EEE

Voici un extrait de la feuille de route permettant de réaliser le post-traitement pour la synchronisation des données multimodales multi-sources pour l'auto-annotation et l'étiquetage des données.

### 1. SYNCHRONISATION

Les données vidéo et audio de DOMUS sont enregistrées séparément, notamment pour permettre l'acquisition de sons de bonne qualité par du matériel spécifique (changement de sensibilité, fréquence d'échantillonnage...etc.). Dans le cas d'une collecte de corpus audiovisuel, il est nécessaire de synchroniser le son et la vidéo. Tout logiciel de traitement vidéo permet d'effectuer cette modification. Voici quelques outils gratuits utiles pour le traitement vidéo :

- **iMovie** (Mac / application installée par défaut, sinon à charger gratuitement sur appStore)
- **VirtualDub** (Windows) : <http://www.virtualdub-fr.org/>
- **FFmpeg** (multi-plateforme / utilisation en ligne de commande et s'intègre bien dans un programme perl) : <https://ffmpeg.org/>
- **MPEG StreamClip** (multi-plateforme / pour conversion de fichier vidéo et changement codecs) : <http://www.squared5.com/>

**Remarque** : les enregistrements via AXIS conservent le timestamp de début d'enregistrement de chaque caméra. Pensez à conserver les vidéos entières en cas de besoin de synchronisation et alignement multimodale et de ne faire que des découpages à posteriori afin d'éviter toute perte d'information utile.

#### 1.1.1. Association son / video

##### 1.1.1.1. Correspondances son/video

Fond de la cuisine	CF	canal1
Entrée de la cuisine	CE	canal2
Fond de la chambre	ChF	canal4
Entrée de la chambre	ChE	canal5
Fond du salon	SF	canal6
Entrée du salon	SE	canal7

**Remarque** : le canal3 correspond au micro salle de bain, le canal8 au micro porté par les sujets

##### 1.1.1.2. Sous iMovie

1° Créer un nouvel événement (la gestion est plus simple si un événement correspond à 1 sujet) et le renommer.

2° Importer les vidéos .asf de DOMUS ainsi que les 6 canaux audio .wav StreamHIS associés pour pouvoir les associer.

3° Créer un projet (type film) par vidéo à créer, soit six dont chacune portera un nom explicite.

4° Dans chaque projet, drag droper la vidéo et le son qui vont ensemble. Idéalement, faire la synchronisation à cette étape afin de régler le problème du décalage constant entre vidéo et son.

**Remarque** : Normalement, tout de suite après le lancement des enregistrements, vous avez dû faire des claps devant chacune des caméras de DOMUS. Le burst sonore doit correspondre au contact visuel du clap correspondant (utiliser le zoom sur la séquence pour être le plus précis possible).

5° Faire les améliorations de contraste et de luminosité pour augmenter la qualité de la vidéo (utiliser l'icône de la baguette magique) et corriger au besoin le cadrage automatique de la vidéo en se mettant sur l'option « adapter » pour conserver une vidéo cadrée sans troncation automatique.

6° Exporter la vidéo

### 1.1.1.3. Quelques commandes ffmpeg utiles

---

#### Augmenter le volume d'un fichier vidéo

```
ffmpeg -i file.mpg -vcodec copy -vol 1024 file_louder.mpg
```

-i : fichier en entrée à modifier

-vcodec copy: laisser la partie vidéo inchangée

-vol 1024: multiplie le volume par 4 ( la valeur de base est -vol 256)

---

#### Compression de la vidéo (la compression marche pour de nombreux formats)

```
ffmpeg -i input.mp4 -vcodec libx264 -crf 20 output.mp4
```

---

#### Découpage de vidéo (marche pour de nombreux formats et s'appuie sur début et fin de séquence)

```
ffmpeg -i input.avi -vcodec copy -acodec copy -ss 00:00:00 -t 00:30:00 output1.avi
```

-ss : le temps de début de séquence à découper

-t : le temps de fin de séquence

---

## 2. CORRECTION DES DERIVES TEMPORELLES – VIRTUALDUB

Cette correction permet de travailler sur des données multimodales qui sont correctement synchronisées. Les flux sont enregistrés par les caméras du plafond de Domus (AXIS) et les canaux de sons sont capturés par des micros (StreamHIS). Il faut pouvoir corriger les *dérives temporelles* et le *décalage entre les sons et les vidéos* qui sont inhérents aux temps machines qui enregistrent les données. L'enregistrement passant par le réseau, les dérives sont susceptibles de varier en fonction de la qualité de ce réseau. De plus, plus la durée d'enregistrement est longue, plus la dérive entre le son et la vidéo peut être marquée. Il existe ainsi deux types de décalages :

- **constant** : qui est lié au problème de synchronisation donc du début de lancement des enregistrements (il ne doit pas exister si la vidéo a correctement été synchronisée sur les claps)
- **progressif** : qui est lié au problème de dérive temporelle dû à la différence des temps machines. Dans le cas de DOMUS, mieux vaut faire les traitements sur les vidéos plutôt que les sons afin d'éviter de dénaturer la qualité et de permettre de faire en parallèle de ces post-traitements les alignements sur le son. L'idée est de changer l'échantillonnage du *framerate* des vidéos à un point maximal de décalage

Pour le traitement du corpus, l'étiquetage est réfléchi de manière à faciliter les filtrages d'informations à granularité différente. Elle permet ainsi la génération automatique d'alignements composants des données d'entrée à un système automatique (données d'apprentissage pour la reconnaissance de parole). Tous les traitements effectués ne peuvent pas être décrits ici. En revanche, un extrait de documentation créée pour les annotateurs, afin de rendre systématiques certaines procédures, est présenté ci-dessous. Ces traitements ont l'avantage de pouvoir être exécutés simultanément sur l'ensemble des fichiers (soit tous les sujets en même temps) si les noms des tiers traités (ou acteurs Elan) sont rigoureusement identiques.

### **Préparation – méthodologie d'étiquetage sous Elan:**

Sachant que beaucoup d'éléments sont alignés nous pouvons recombinaison et filtrer facilement les événements pour récupérer un alignement propre des différents silences.

#### **Quelques bases d'Elan :**

##### **\*\* Acteur > Create annotations from gaps**

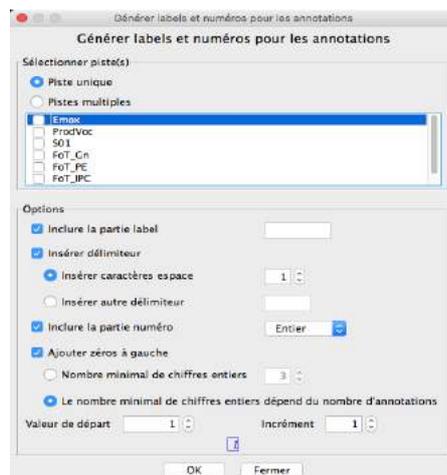
C'est l'option principale à utiliser pour annoter les silences. Cette option permet de sélectionner un tiers d'événements et de recréer un nouveau tiers en recréant son miroir, à savoir en annotant par des barres d'événements les endroits où il n'y a pas



**Remarque :** l'utilisation de cette option ne se fait qu'à la toute fin, quand l'alignement des événements est propre !

##### **\*\* Acteur > Labéliser et numéroté les annotations**

C'est une option qui est très utile pour codifier les éléments.



\*\* Avec la première option, il est possible soit de labéliser un seul tiers, soit de labéliser plusieurs tiers de même type (cocher le(s) tier(s) à labéliser).

\*\* Puis insérer ou non la partie labélisée

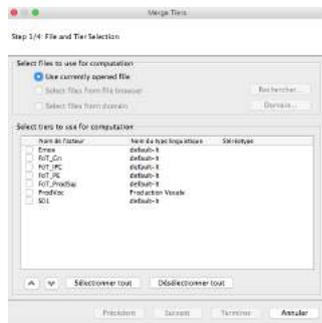
\*\* Rajouter un délimiteur (en général une partie du code associée aux étiquettes)

\*\* Mettre ou non une numérotation (celle-ci est automatique)

**Remarque :** ne pas hésiter à utiliser cette option dans des étapes intermédiaires pour faciliter les filtrages des données !!

On peut aussi faire des tiers avec juste le code sur l'ensemble et faire un merge après avec des tiers qui comportent des valeurs pour avoir des labels propres.

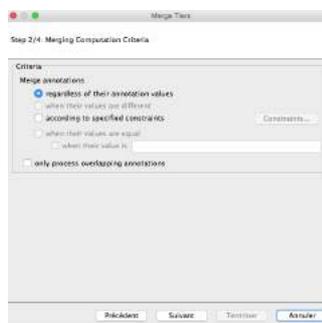
**\*\* Acteur > Merge Tiers**



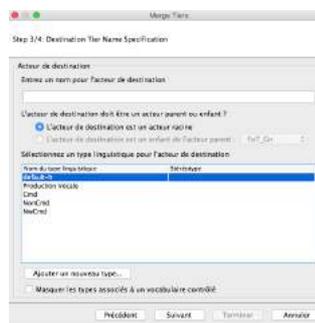
**Etape 1 :** Choisir les Tiers à rassembler dans un seul et même tiers. (Attention ce n'est pas la même chose que les chevauchements)



**Astuces :** ordonner durant cette étape les labels à émerger avec les flèches, cet ordre sera reporté directement à l'étape 4



**Etape 2 :**  
on choisit si on associe les tiers indépendamment de leur valeur ou uniquement s'ils se chevauchent



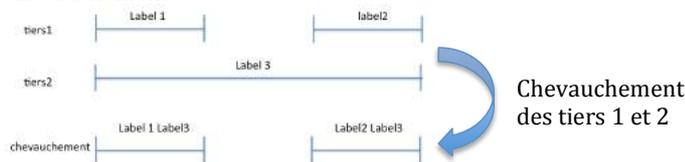
**Etape 3 :**  
on rentre le nom et le type linguistique à associer.  
**Remarques :** c'est intéressant de définir un type linguistique spécifique si les tiers mergés sont de même type, les données seront ainsi plus propres, en revanche si on merge des tiers avec des infos différentes, il vaut mieux définir la valeur « default ».



**Etape 4 :**  
on choisit si on associe les tiers selon leur ordre chronologique, selon des valeurs spécifiques, ne reporter que la valeur d'un tiers ou selon un ordre de tiers qu'on a défini (ce qui est reporté de l'étape 1, sinon il faut redéfinir avec les flèches cet ordre mais l'interface de cet étape 4 bug souvent)

**\*\* Acteur > Créer des annotations depuis le chevauchement**

Cette option permet de repérer les endroits où il y a des chevauchements et de créer des événements dans un nouveau tiers avec uniquement les zones où il y avait du chevauchement

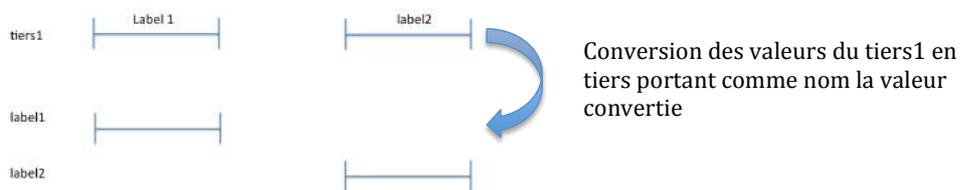


Chevauchement des tiers 1 et 2

**\*\* Acteur > Convert Annotation value to**

Permet de définir des valeurs d'un tiers

tiers  
comme des tiers



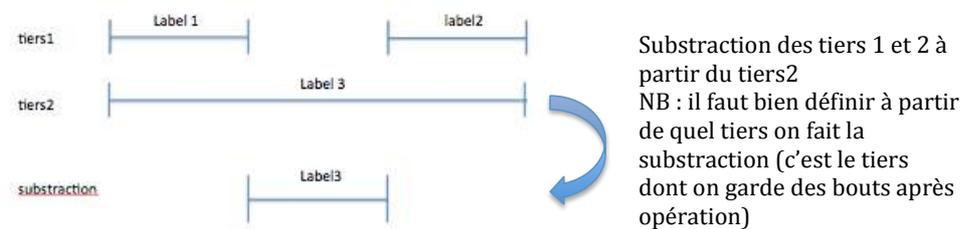
Conversion des valeurs du tiers1 en tiers portant comme nom la valeur convertie

**Remarque :** il s'agit d'une options très utile mais si les données ne sont pas propres (ex : espaces) ça créera autant de tiers que de différentes étiquettes.

A n'appliquer que sur les tiers avec un nombre de valeurs précis et défini (de préférence dans un vocabulaire contrôlé)

**\*\* Acteur > Create annotations for substraction**

**chevauchement vide**



Substraction des tiers 1 et 2 à partir du tiers2  
NB : il faut bien définir à partir de quel tiers on fait la substraction (c'est le tiers dont on garde des bouts après opération)

**\*\* Acteur > Copier un acteur + Acteur > Remove**

Il est utile de pouvoir copier ces 2 opérations pour faire du nettoyage de tiers déjà existants.

**Attention avant toute manipulation de tiers, toujours penser à faires des copies des tiers utilisés ou une copie de fichiers Elan !!**

Une fois que les tiers filtrés ont été créés, il est également possible d'utiliser des scripts de découpage (parallèlement sur sources audio et vidéo alignés) ou d'annotation automatiques dont un nombre important a été développé, adapté et réutilisés L'ensemble des scripts utilisés ne peut être présenté et ne présente pas forcément de nouveauté scientifique majeure en termes de traitement informatique, mais facilite la manipulation de grande quantité de données. Ici, l'intérêt n'est pas le code en lui-même, mais la philosophie de l'approche permettant de créer de multiples scripts simples pour réaliser de multiples micro-tâches de traitement du corpus. Voici un exemple de script perl permettant la génération automatique d'un ensemble de données filtrées pour créer le corpus d'entrée du système de reconnaissance de parole:

```
# Auteur Yuko SASA - version 01/04/2015
#####
#
# Descriptif
#####
# Ce script permet de générer des fichiers de type "idSuj_pNum.mp4" et "idSuj_pNum.wav" à partir d'une longue video "fileName.mp4" (ex: la video d'une expe)
# associée à un fichier "idSuj_TC.txt" (fichier type shoebox produit sous Elan) avec les timestamp de séquences transcrites (ex: des commandes).
# Les videos sont ainsi découpées en fonction des timestamps des séquences et sont placées dans un dossier généré nommé "idSuj_videoCmd"
# Les fichiers sons sont extraits des fichiers vidéos générés et sont placés dans un dossier généré nommé "idSuj_audioCmd"
# Un fichier de transcription de séquence de type "idSuj_cmd.txt" où chaque ligne = une transcription de séquence
# Pour l'utiliser placer dans un dossier avec "fileName.mp4" à traiter avec "idSuj_TC.txt" et lancer le script en suivant les indications à l'écran
# Pour lancer le script dans un terminal (ou console) :
# cd cheminAccesVersScript
# perl IFRecogSys.pl
#####

#!/usr/bin/perl
use strict;
use warnings;

### Paramétrage ###
print "Identifiant du sujet:";      # Interface terminal : demande de l'id sujet
my $idSuj = <>;                  # Récupération de l'ID du sujet
chomp $idSuj;                     # Retrait du saut de ligne
#print "idSuj: $idSuj\n";
print "Nom du fichier video à traiter:"; # Interface terminal : demande de la video à traiter
my $videoFile = <>;              # Récupération du nom fichier vidéo à traiter
chomp $videoFile;                # Retrait du saut de ligne
#print "videoFile: $videoFile\n";

### Création dossiers de sortie ###
my $outputFolder = $idSuj."_videoCmd"; # Nommage du dossier de sortie avec les videos des commandes
#print "outputFolder: $outputFolder\n";
mkdir $outputFolder;                 # Création du dossier de sortie avec les videos des commandes
print "$outputFolder créé\n";        # Message de confirmation de création du dossier videos cmd
my $outputFoldAudio = $idSuj."_audioCmd"; # Nommage du dossier de sortie avec les sons .wav des commandes
#print "outputFoldAudio: $outputFoldAudio\n";
mkdir $outputFoldAudio;             # Création du dossier de sortie avec les sons .wav des commandes
print "$outputFoldAudio créé\n";    # Message de confirmation de création du dossier sons cmd
```

```

### Découpage du fichier vidéo entrée suivant les timecodes ###
my $fileTC = $idSuj."_TC_Cmd.txt";           # Définition du nom du fichier de timestamps (pour découpage)
my $fileCmdClean = $idSuj."_cmd.txt";        # Définition du nom du fichier de transcription de commandes
my $nbTC = 0;                               # Initialisation du nb cmd utilisé pour générer les fichiers
my $TC = 0;                                  # Timestamp de début de production
my $output = $idSuj."_p";                    # Définition du format de nom des fichiers .wav de commandes
my $command;                                 # Création de la variable commande système d'exécution ffmpeg
my $regex = "\\.". $idSuj."_CmdValue#Cmd(\\s*)(.*) (\\r|\\n)";
open FILE, $fileTC or die $!;                # Ouverture du fichier timestamps
open FILECLEAN, '>', $fileCmdClean or die $!; # Création du fichier de sortie de transcriptions de cmd ligne par ligne
while (my $line = <FILE>) {                  # Lecture du fichier timestamp ligne par ligne
    if ($nbTC < 1) {                          # Boucle pour le premier traitement
        #print "line: $line";
        if ($line =~ m/\\ELANBegin(\\s*)(\\d\\d:\\d\\d:\\d\\d\\.\\d\\d\\d)\\s/) { # Récupération du timestamp de début de commande
            $command = "ffmpeg -i $videoFile -c copy -ss $2"; # Remplissage de commande ffmpeg par insertion de timestamp
            #print "!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!$2";
            $nbTC++; # Actualisation du compteur de cmd
            $TC=$2;
            #print "commande : $command\\n";
        }
        elsif ($line =~ m/\\ELANEnd(\\s*)(\\d\\d:\\d\\d:\\d\\d\\.\\d\\d\\d)\\s/) { # Récupération des timestamps de fin de séquence
            $nbTC--; # Initialisation de compteur de cmd
            $command = $command."-to $2 $outputFolder/$output$nbTC.mp4"; # Commande ffmpeg de découpage video générant la video d'une commande
            #print "commande : $command\\n";
            $nbTC++;
        }
        elsif ($line =~ m/$regex/){ # Récupération des transcriptions orthographiques exactes de la commande
            chomp $2; # Retrait du saut de ligne de la transcription de commande
            $nbTC++;
            print "!!!!!!!!!!!!!!!!!!!!!!".$TC;
            print FILECLEAN $output.$nbTC."\\t".$2."\\t".$TC."\\n"; # Ajout de la commande dans le fichier de transcriptions listées
            $nbTC--;
        }
        else {
            #print "pas de timecode ni de transcription sur la ligne\\n";
        }
    }
}
elseif ($nbTC >= 1){ # Boucle sur les autres lignes de traitement
    if ($nbTC == 1) { # Vérification du traitement de la première ligne
        #print "TC déjà pris\\n";
        $nbTC++; # Actualisation du compteur de cmd
    }
    if ($line =~ m/\\ELANBegin(\\s*)(\\d\\d:\\d\\d:\\d\\d\\.\\d\\d\\d)\\s/) { # Récupération du timestamp de début de commande
        $command = $command." -c copy -ss $2"; # Remplissage de commande ffmpeg par insertion de timestamp
        $TC=$2;
        $nbTC++; # Actualisation du compteur de cmd
        #print "commande : $command\\n";
    }
}

```

```

    elif ($line =~ m/\\ELANEnd(\\s*)(\\d\\d:\\d\\d:\\d\\d.\\d\\d\\d)\\s/) {          # Récupération du timestamp de fin de commande
        $nbTC--;                                                                    # Actualisation du compteur de cmd
        $command = $command." -to $2 $outputFolder/$output$nbTC.mp4";              # Remplissage de commande ffmpeg par insertion de nom de fichier video de sortie
        $nbTC++;                                                                    # Actualisation du compteur de cmd
        #print "commande : $command\\n";
    }
    elif($line =~ m/$regexp/){                                                       # Récupération des transcriptions orthographiques exactes de la commande
        chomp $2;                                                                    # Retrait du saut de ligne de la transcription de commande
        print FILECLEAN $output.$nbTC."\\t".$2."\\t".$TC."\\n";                    # Ajout de la commande dans le fichier de transcriptions listées
    }
}
}
close FILE;                               # Fermeture fichier timestamp
#print "commande : $command\\n";
system $command;                           # Lancement de la procédure de création de videos

my $commandAudio;                          # Création de la variable de commande de création de fichiers audio
my $inputFile;                              # Création de la variable du fichier d'entrée pour création de .wav de commande (= videos créées par la précédente procédure)

# Ouverture du dossier des sons générés
opendir (inputFolder,$outputFolder) or die "Couldn't open dir : $!"; # ouverture du dossier de videos de cmd créées

my $saut=1;                                # Initialisation du nombre de lignes à ajuster pour récupérer
my $nb=1;                                   # Initialisation du nombre de fichiers sons générés
while (my $ligne = readdir(inputFolder)){    # Récupération une par une du nom des fichiers vidéos créés en ouvrant leur dossier
    #print $ligne."\\n";
    $inputFile = $ligne;                    # Définition du nom de fichier d'entrée (nom du fichier video d'une commande donnée)
    if ($saut<=2){                          # Condition pour commencer la création de fichier audio
        #print "pas bon";
    }
    else{
        # Création de la commande ffmpeg permettant d'extraire des fichiers .wav à partir des videos de cmd .mp4
        $commandAudio = $commandAudio." ffmpeg -i $outputFolder/$inputFile -vn -acodec pcm_s16le -ar 16000 -ac 1 $outputFolder/$output$nb.wav \\n";
        $nb++;                               # Actualisation du compteur de fichiers audio/video
    }
    $saut++;                                 # Actualisation du compteur pour décaler le démarrage de création de fichier audio (. et .. du dossier)
}

closedir inputFolder;                       # Fermeture du dossier de videos

#print $commandAudio;
system $commandAudio;                       # Lancement de la procédure de création de sons

"/>

```

Voici un exemple de fichier Praat adapté pour effectuer des mesures automatisées de paramètres acoustiques généraux:

```
# This script goes through sound and TextGrid files in a directory,
# opens each pair of Sound and TextGrid, calculates the pitch maximum
# of each labeled interval, and saves results to a text file.
# To make some other or additional analyses, you can modify the script
# yourself... it should be reasonably well commented! ;)
#
# This script is distributed under the GNU General Public License.
# Copyright 4.7.2003 Mietta Lennes
# Script adapted by Yuko Sasa (23.05.2015)

form Analyze pitch maxima from labeled segments in files
  comment Directory of sound files
  text sound_directory
/Users/yukosasa/OneDrive/Yuko_Perso/0_TODAY_0/Final_TTT_EEE/20170718/ana_acoustics/All_split_sounds/
  sentence Sound_file_extension .wav
  comment Directory of TextGrid files
  text textGrid_directory
/Users/yukosasa/OneDrive/Yuko_Perso/0_TODAY_0/Final_TTT_EEE/20170718/ana_acoustics/All_split_textGrid/
  sentence TextGrid_file_extension .TextGrid
  comment Full path of the resulting text file:
  text resultfile
/Users/yukosasa/OneDrive/Yuko_Perso/0_TODAY_0/Final_TTT_EEE/20170718/ana_acoustics/pitchresults_test.txt
  comment Which tier do you want to analyze?
  sentence Tier Cmd_trans_net
  comment Pitch analysis parameters
  positive Time_step 0.01
  positive Minimum_pitch_(Hz) 70
  positive Maximum_pitch_(Hz) 600
endform

# Here, you make a listing of all the sound files in a directory.
# The example gets file names ending with ".wav" from /Users/yukosasa/Desktop/

Create Strings as file list... list 'sound_directory$'*'sound_file_extension$'
numberOfFiles = Get number of strings
```

```
# Check if the result file exists:
if fileReadable (resultfile$)
    pause The result file 'resultfile$' already exists! Do you want to overwrite it?
    filedelete 'resultfile$'
endif

# Write a row with column titles to the result file:
# (remember to edit this if you add or change the analyses!)

titleline$ = "Filename Segment label 'tab$' F0max (semitone re 100Hz) 'tab$' F0max (Hz) 'tab$' F0min (semitone re
100Hz) 'tab$' F0min (Hz) 'tab$' deltaF0 (semitone re 100Hz) 'tab$' deltaF0 (Hz) 'tab$' F0init (semitone re 100Hz) 'tab$'
F0init (Hz) 'tab$' F0final (semitone re 100Hz) 'tab$' F0final (Hz) 'tab$' F0moy (semitone re 100Hz) 'tab$' F0moy (Hz)
'tab$' F0range (semitone re 100Hz) 'tab$' Duree (msec) 'tab$' StandardDeviation (Hz) 'newline$"
fileappend "'resultfile$" 'titleline$'

# Go through all the sound files, one by one:

for ifile to numberOfFiles
    filename$ = Get string... ifile
    # A sound file is opened from the listing:
    Read from file... 'sound_directory$'filename$'
    # Starting from here, you can add everything that should be
    # repeated for every sound file that was opened:
    soundname$ = selected$ ("Sound", 1)
    To Pitch... time_step minimum_pitch maximum_pitch
    # Open a TextGrid by the same name:
    gridfile$ = "'textGrid_directory$'soundname$'textGrid_file_extension$"
    if fileReadable (gridfile$)
        Read from file... 'gridfile$'
        # Find the tier number that has the label given in the form:
        call GetTier 'tier$' tier
        numberOfIntervals = Get number of intervals... tier
        # Pass through all intervals in the selected tier:
        for interval to numberOfIntervals
            label$ = Get label of interval... tier interval
            if label$ <> ""
                # if the interval has an unempty label, get its start and end:
                start = Get starting point... tier interval+0.01
            
```

```

end = Get end point... tier interval-0.01
# get the Pitch maximum at that interval
select Pitch 'soundname$'
pitchminhz = Get minimum: start, end, "Hertz", "Parabolic"
pitchmaxhz = Get maximum: start, end, "Hertz", "Parabolic"
pitchminst = Get minimum: start, end, "semitones re 100 Hz", "Parabolic"
pitchmaxst = Get maximum: start, end, "semitones re 100 Hz", "Parabolic"
deltahz = pitchmaxhz-pitchminhz
deltast = pitchmaxst-pitchminst
pitchinithz = Get value at time: start+0.03, "Hertz", "Linear"
pitchfinalhz = Get value at time: end-0.03, "Hertz", "Linear"
pitchinitst = Get value at time: start+0.05, "semitones re 100 Hz", "Linear"
pitchfinalst = Get value at time: end-0.05, "semitones re 100 Hz", "Linear"
pitchmeanhz = Get mean: start, end, "Hertz"
pitchmeanst = 12*log2(pitchmeanhz/1)
pitchrangest = 12*log2(pitchmaxhz/pitchminhz)
durationsec = end - start
durationmsec = (end - start)*1000
stdev = Get standard deviation: start, end, "Hertz"
printline 'pitchmaxst' 'pitchmaxhz' 'pitchminst' 'pitchminhz' 'deltast' 'deltahz' 'pitchinitst'
'pitchinithz' 'pitchfinalst' 'pitchfinalhz' 'pitchmeanst' 'pitchmeanhz' 'pitchrangest' 'durationmsec' 'stdev'
# Save result to text file:
resultline$ = "'soundname$' 'tab$' 'label$' 'tab$' 'pitchmaxst' 'tab$' 'pitchmaxhz' 'tab$'
'pitchminst' 'tab$' 'pitchminhz' 'tab$' 'deltast' 'tab$' 'deltahz' 'tab$' 'pitchinitst' 'tab$' 'pitchinithz' 'tab$'
'pitchfinalst' 'tab$' 'pitchfinalhz' 'tab$' 'pitchmeanst' 'tab$' 'pitchmeanhz' 'tab$' 'pitchrangest' 'tab$'
'durationmsec' 'tab$' 'stdev' 'newline$'"
fileappend "'resultfile$'" 'resultline$'
select TextGrid 'soundname$'
endif
endif
# Remove the TextGrid object from the object list
select TextGrid 'soundname$'
Remove
endif
# Remove the temporary objects from the object list
select Sound 'soundname$'
plus Pitch 'soundname$'
Remove

```

```
        select Strings list
        # and go on with the next sound file!
    endfor

Remove

#-----
# This procedure finds the number of a tier that has a given label.

procedure GetTier name$ variable$
    numberOfTiers = Get number of tiers
    itier = 1
    repeat
        tier$ = Get tier name... itier
        itier = itier + 1
    until tier$ = name$ or itier > numberOfTiers
    if tier$ <> name$
        'variable$' = 0
    else
        'variable$' = itier - 1
    endif

    if 'variable$' = 0
        exit The tier called 'name$' is missing from the file 'soundname$'!
    endif

endproc
```

## Annexe 11. Liste des commandes non-référencées produites par les sujets

Cette liste correspond aux principales valeurs qui se trouvent sous la catégorie d'étiquette « NwCmd ».

<b>Etiquette descripteur</b>	<b>Description</b>
Plus vite	Variantes associées à des actions de déplacement du robot ou à des commandes domotiques
Se déplacer (chambre)	Variantes demandant au robot d'aller dans la chambre
Se déplacer (cuisine)	Variantes demandant au robot d'aller dans la cuisine
Se déplacer (salon)	Variantes demandant au robot d'aller dans le salon
Allumer micro-ondes	Le sujet demande au robot d'allumer le micro-onde qui se trouve à la cuisine salon (parfois confondu avec une télévision)
Eteindre micro-ondes	Le sujet demande au robot d'éteindre le micro-onde qui se trouve à la cuisine salon (parfois confondu avec une télévision)
Diriger	Le sujet donne des indications sur la direction spatiale au robot pour lui donner une certaine position
Attendre	Le sujet demande au robot de ne rien effectuer mais de rester attentif
Accompagner	Le sujet demande au robot de venir avec lui, de le suivre
Ton interlocuteur	Le sujet donne le contrôle d'Emox à la personne désignée (p. ex. à l'aide à domicile)
Allumer l'ordinateur	Le sujet demande au robot d'allumer l'ordinateur qui se trouve à l'intérieur du salon
Répéter	Le sujet demande au robot de répéter ce qu'il vient de dire
Arrêter action courante	Le sujet demande au robot d'arrêter l'(les) actions courantes domotiques

## Annexes

Passer devant	Le sujet demande au robot de passer devant lui, de le précéder pour un déplacement
Réinitialiser Domus	Le sujet demande au robot de remettre dans la position initiale l'ensemble des actionneurs de Domus
Dire quelque chose / Réagir	Variantes de sollicitations pour faire parler ou réagir le robot
Prendre téléphone	Le sujet demande d'attraper (d'apporter = déplacement d'objet) le téléphone
Parler plus fort	Le sujet demande au robot de parler plus fort
Attirer attention Emox	Variantes que le sujet effectue pour capter l'attention d'Emox (notamment après un temps sans interaction)
Commande impossible	Variantes de commandes non réalisables (p. ex. faire couler l'eau, le savon, composer un numéro...etc.). Ce sont des commandes qui seraient néanmoins techniquement envisageables.
Pas bouger	Le sujet demande au robot de ne pas bouger
Appeler X	Le sujet demande d'appeler quelqu'un (notamment les expérimentateurs)

## Annexe 12. Distribution des variations morphosyntaxiques par sujet

### Nomenclature des variantes de formes morphosyntaxiques

**temps.mode** = le point indique si le couple temps.mode du verbe de l'énoncé

**prs** = présent

**ind** = indicatif

**imp** = impératif

**inf** = infinitif

**cond**=contitionnel

**Compo**= temps composé

**pron** = tournure pronominalisée

**adv** = tournure adverbiale

**neg** = insertion d'une négation

**interro** = forme interrogative

**[tournure]** = indique une tournure spécifique

**[ME-int]** = indique une micro-expression vocale qui se trouve à l'intérieur de l'énoncé de commande

**-\$** = marque une interruption de l'énoncé

**imposee** = tournure d'une commande comme donnée initialement dans le mode d'emploi de l'appartement (tournure infinitive ou nominale)

**[variante]** = indique une variante du verbe principal de l'énoncé (p. ex. « allumer » à la place de « mettre »).

SUJET	Forme Cmd	Nb d'occurrence
aSa1	imposee	13
	on + [aller] prs.ind + inf	4
	sans verbe	4
	on + prs.ind	25
	imposee-\$	1
	inf – inférence[pron]	1
	nous-prs.imp	2
	[on a dit] + inf	1
	on + inférence[pron]	1
	on + prs.ind + inférence[pron]	1
	tu + prs.ind	12
	on + [pouvoir neg] + inférence[pron]	2

	on + [variante] prs.ind	2
	tu-prs.imp	2
	on + [pouvoir neg] + [variante] inf	1
	tu nous + prs.imp + inférence[pron] + focus[c'est]	1
	tu nous+ prs.ind	1
	tu nous+ prs.ind + inférence[pron]	1
	tu + [variante] prs.ind	2
	tu nous + prs.ind	1
	tu nous + prs.ind + inférence[pron]	1
	tu + prs.ind + inférence[pron]	1
	tu + [variante] prs.ind + inférence[pron] +tu-prs.imp	1
	tu nous + [vouloir adv-bien] prs.ind + inf	1
S01	imposee	10
	on + [aller] prs.ind + inf	5
	tu-prs.imp + moi	1
	[il faut] + [variante] inf	2
	tu-prs.imp	1
	on + prs.ind	11
	je + [vouloir adv-bien] prs.cond + [variante] inf	1
	sans verbe	3
	[vaut mieux] + inf	1
	tu me + [vouloir pas] prs.ind	1
	imposee-\$	3
	tu-prs.imp [variante] -\$	1
	on + [variante] prs.ind	7
	tu me + prs.ind	5
	je + [trouve que] prs.ind + tu + [pouvoir] prs.cond + inf	1

## Annexes

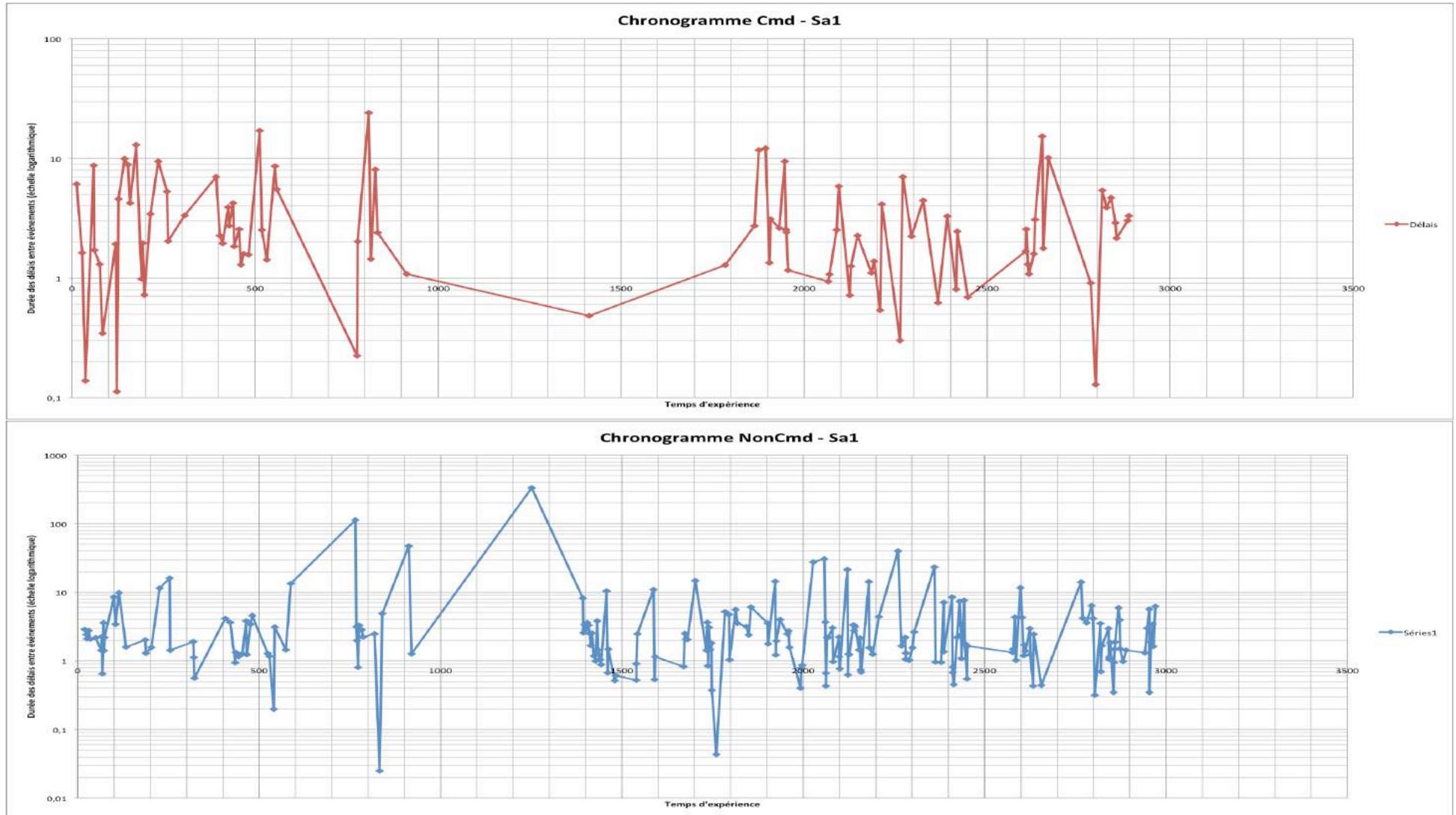
	on + [aller] prs.ind + [peut-être] + inf + focus	1
	je + [vouloir] prs.cond + inférence[pron]	1
	on + [aller] prs.ind + [variante] inf-\$	1
	on + [aller] prs.ind + [variante] inf	1
	je + [vouloir] prs.cond bien + [variante] inf	1
	je + [vouloir adv-bien] prs.cond bien + [variante] inf	1
	tu + [vouloir] prs.ind + inf	2
S03	imposee	131
	sans verbe	2
	imposee-\$	3
S04	imposee	68
	[variante] inf	11
	[faut] + inf	1
	on + [aller] prs.ind + inf	1
	sans verbe	3
	[variante] sans verbe	1
	je te + [dire de] passeCompo.ind + inf	1
	imposee-[ME interne]	1
S05	imposee	50
	tu + [vouloir neg] prs.ind + inférence[pron]	1
	sans verbe	1
	on + [pouvoir] prs.ind + inf	1
	[variante] sans verbe	1
	[variante] inf	1
S06	imposee	35
	[variante] sans verbe	5
	[variante] inf	3

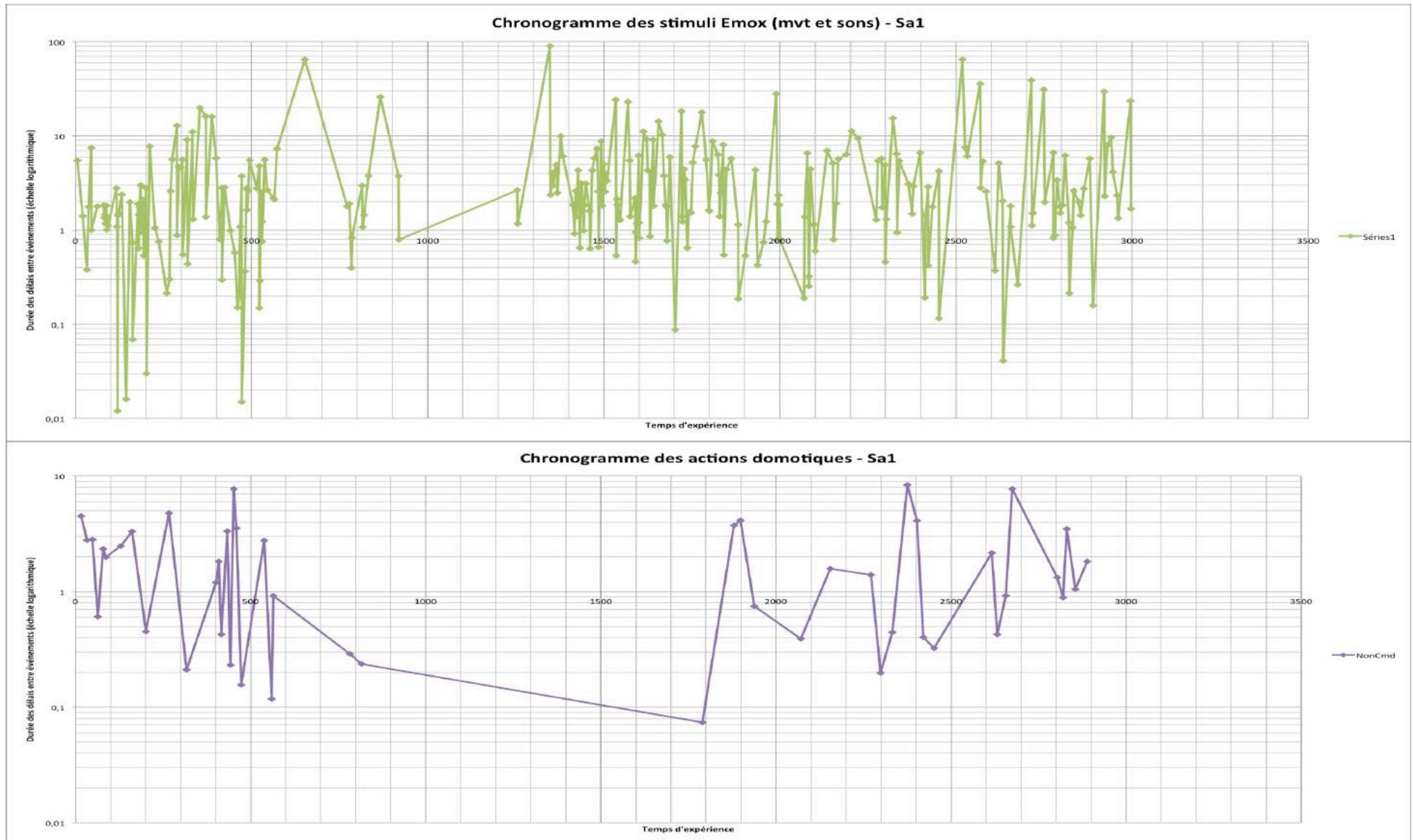
	on + [pouvoir] prs.ind – forme interrogative + [variante] inf	1
	tu + [pouvoir] prs.ind + [variante] inf	7
	tu + [vouloir] prs.ind + inf	4
	tu + [vouloir] prs.ind + [variante] inf	4
	tu-prs.imp	41
	sans verbe	17
	tu + [pouvoir] prs.ind – forme interrogative [est-ce que] + [variante] inf	1
	tu + [pouvoir] prs.ind + inf	2
	tu + prs.ind	1
	tu-prs.imp [variante]	4
	[variante]-intération	1
	je + [vouloir] + sans verbe	3
	tu + [variante] prs.ind	1
	tu + [vouloir] prs.ind + [variante] inf	1
	je + [dire] passeCompo.ind + inf	1
S07	imposee	37
	[variante] inf	10
	on + [aller] prs.ind + inf + focus	1
	je + [vouloir] prs.cond + sans verbe	1
S10	imposee	38
	[variante] inf	4
	tu + [pouvoir] prs.ind + inf	1
	sans verbe	3
S11	tu-prs.imp	11
	tu-prs.imp [variante]	10
	tu + [pouvoir] prs.ind + inf + focus	1

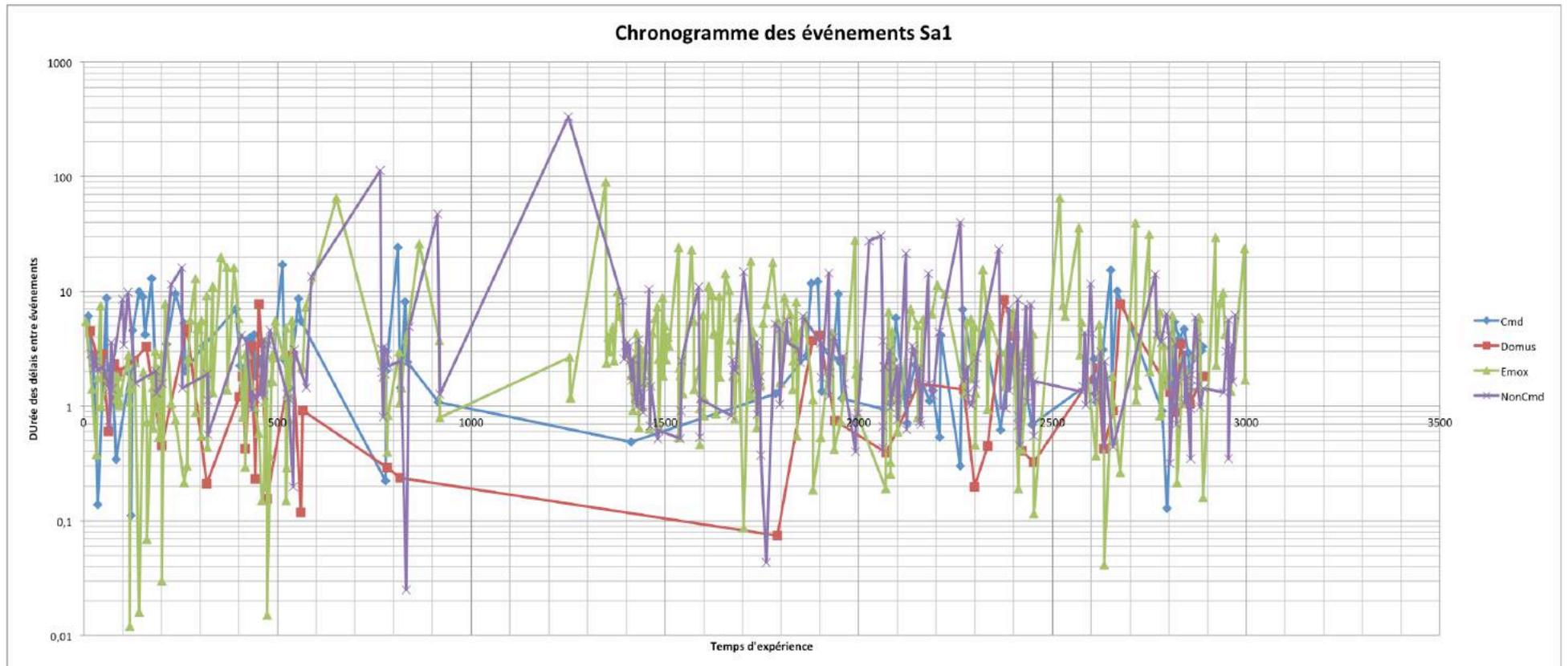
	[vas] prs.ind + [variante] inf	1
	tu-prs.imp [variante + adv]	2
	tu + [pouvoir] prs.ind + inf	4
	tu + [pouvoir] prs.ind -interro[inversion] + [variante] inf	2
	sans verbe	2
	tu + [pouvoir] prs.ind + [variante] inf	2
S12	imposee	35
	sans verbe	4
	[variante + adv]	4
	[variante] inf	3
	imposee [ME-int]	1
	[il faut] + [variante] inf	1
	tu-prs.imp [variante]	2
	tu-prs.imp	1
S14	imposee	44
	[variante] inf	1
S15	imposee	30
	sans verbe	12
	on + [pouvoir] prs.ind + inf + [Emoz-int]	1
	tu + [vouloir] prs.ind – interro[inversion] + inf	2
	tu + [pouvoir] prs.ind – interro[inversion] + inf	6
	[variante adv]	1
	tu-prs.imp	4
	tu + [vouloir] prs.ind + inf + inférence[pron]	2
	tu + [pouvoir] prs.ind – interro[inversion] + [variante] inf	1
	tu-prs.imp [variante]	2
	[il faut] + inf	1

	tu-prs.imp [variante] + inférence[pron]	3
	je vais te demander de] + inf + inférence[pron]	1
	tu + [pouvoir] prs.ind – interro[inversion] + inf + inférence[pron]	1
	tu + [pouvoir] prs.ind + inf + inférence[pron]	2
	tu + [pouvoir] prs.ind + [variante] inf	1
	tu + [pouvoir] prs.ind – interro[inversion] + inf	1
S16	imposee	58
	sans verbe	12
	[variante] inf	5
	[vous voulez qu'on] + inf	1
	sans verbe – itération	1
	vous-prs.imp	2
	[j'aimerai] + inf	1
	[variante] inf + itération-\$	1
S17	imposee	13
	vous-prs.imp	2
	[il faut] + inf	1
	[variante] inf	2
S18	imposee	35
	sans verbe	4
	sans verbe – itération	1
	sans verbe – itération – [ME-int]	1
	[variante] inf	2
S19	imposee	27
	[variante] inf	1
	sans verbe	2

### Annexe 13. Chronogrammes des événements interactionnels







## Annexe 14.      Etiquettes d'auto-annotation relevées par groupe temporel et par sujet

**FoT\_Gn** : états mentaux sur temps général latent

**FoT\_PE**: sur la perception d'Emox latente

**FoT\_ProdSuj**: sur production du sujet

**FoT\_IPC**: sur interactions ponctuelles communicationnelles remarquables

**Grille AA – aSa1**

FoT_Gn	FoT_ProdSuj	FoT_PE	FoT_IPC
A : je me souviens pas	1 : curiosité – dominant	[I] : je me demandais à quoi ça servait	a) il y avait toujours un certain ralentissement comme l'ordinateur, il y a des pauses – c'est l'habitude quand on se sert de matériel [délai à respecter, mais pas gênant] – on s'habitue – quand on parle ça va plus vite
B : je me demandais à quoi ça servait – je ne sens du tout en train de faire une expérience – c'est surtout un rappel des choses	2 : je ne le prends pas au sérieux 2' : là je me suis un peu moquée	[II] : ça me semble un peu artificiel	b) je déteste être agressée par le son
C : j'ai plus confiance que ça marche	3 : ton interrogatif – « est-ce que ça va marcher » 3' : interrogatif plus qu'ordre 3'' : question « est-ce que tu sais ? »	[III] : est-ce que ça marche mon truc	c) là j'ai apprivoisé l'entourage
D : là ça devient amusant D' : ça m'a fait rire – j'étais très contente	4 : je suis sûre que ça va marcher	[IV] : c'est comme un petit animal qui bouge quoi	d) ça c'est les explications
E : on va voir comment ça marche	5 : coopératif – on va faire 5' : on est d'accord pour le faire, on le fait ensemble	[V] : je lui parle comme à un petit compagnon – un petit élève	e) j'aimais bien la lumière verte, j'aime bien la couleur
F : je ne comprends pas	6 : c'est devenu un peu répétitif – j'ai pris un rythme de croisière	[VI] : relation avec un petit automate qui a l'air sympa quand même, qui me cause aucun ennui	

## Annexes

G : c'est pas encore au point	7 : c'est ma manière de donner l'ordre – des ordres sur le mode collaboratif 7' : l'ordre, mais sur le ton « on le fait » ou alors « on va voir » 7'' c'est ma façon de dire il faut le faire 7''' : je garde mon ton parce que j'ai vu que c'est pas la peine de donner un ordre... de manière impérative	[VII] : « je ne chante jamais devant quelqu'un » (mais devant cet objet oui)	
H : j'ai l'impression de connaître	8 : c'est bien gentil – c'est bien mignon	[VIII] : c'est mon compagnon là, on a été ensemble	
I : on était dans le rôle, je jouais le rôle I : là je cherche sa compagnie, je joue le rôle	9 : j'encourage – on y est arrivé c'est bien – encouragement	[IX] : là je l'appelle par son nom hein, je le nomme	
J : j'essaie d'être le plus disponible sans interpréter ni intellectualiser J' : je pense qu'il était toujours là si on lui demandait quelque chose	10 : j'ai essayé de le dire de manière plus simple	[X] : je le personnifiais un peu, je m'adresse comme à quelqu'un avec qui j'ai un lien c'est quelqu'un qu'il faut garder dans la conversation	
K : après avoir joué un rôle, on a besoin de se retrouver, sans rôle à jouer et vivre la relation avec la personne qui est là	11 : je l'aide, je veux l'aider plutôt		
L : là c'est la détente	12 : bah ça va pas vite, il bouge pas assez – alors ça vient ?		
M : j'avais l'impression qu'il fallait le tenir au courant M' : j'ai pris l'habitude de dire ce que je faisais	13 : j'attendais que ça se mette en route		
N: je me disais alors on va vivre avec un petit robot, comment on va avoir une relation avec lui	14 : ça ne me plaisait pas – je voulais changer		

## Annexes

N' : j'ai eu un petit dialogue avec lui			
O : je ne me concentrais sur rien d'autre que ça	15 : je l'ai entendu parler donc je lui fais un retour 15' : je devine plutôt que j'entends		
P : ça marche pas – je me suis dit tiens est-ce que c'est le ton que j'avais ou parce qu'il ne pouvait vraiment pas répondre	16 : je voulais essayer de voir s'il avait quelque chose à dire 16' : c'est une question qui n'est pas du fonctionnement, est-ce qu'il va me répondre quelque chose 16'' j'essayais de voir s'il connaissait des phrases 16''' : je me disais qu'il avait peut-être des mots auxquels il réagissait ou des attitudes		
Q : j'essaie de donner une image positive de ce qui est autour Q' : je montre qu'il sait faire beaucoup de choses	17 : ça sert à garder le lien comme ça 17' : là c'est pour partager 17'' : il faut quand même du lien		
R : là pour moi c'était fini – pour moi le... c'était terminé R' : c'est déjà fini le travail, je laisse aller	18 : je cherche ce qu'il sait – est-ce que tu vas savoir maintenant – tout à l'heure tu ne savais pas 18' : je cherche son répertoire		
S : là j'en ai marre	19 : c'est pour voir s'il répète		
	20 : je cherche à lui faire faire quelque chose 20' : il faut bien que j'en sorte quelque chose hein 20'' : je savais pas trop quoi lui dire d'autre 20''' : j'employais le temps		
	21 : pour le montrer 21 : je <u>devais</u> lui montrer		

## Annexes

	<p>22 : j'essayais de voir si on peut pas y arriver, comme en première intention ça marchait pas</p> <p>22' : on va voir si par hasard ça va marcher</p>		
	<p>23 : je double beaucoup mes ordres</p> <p>23' : première incitation et deuxième pour voir si c'est compris</p> <p>23'' : manière de fonctionner pour être sure qu'on m'a bien comprise</p> <p>23''' : j'attire l'attention et je vois si c'est... arrive</p>		
	<p>24 : ça va on conclut, on a vu ce qu'il y avait à voir</p> <p>24' : c'est l'au revoir quoi</p> <p>24'' : on part, on s'en va (donc on éteint)</p>		
	<p>25 : ah ça j'y tiens, c'est une manière de revoir</p>		
	<p>26 : ça va, c'est comme on dit bon, voilà ça y est, autre chose</p> <p>26' : ça va hein, c'est devenu accessoire</p>		
	<p>27 : interrogation – tu sais bien le faire quoi</p>		
	<p>28 : là j'ai un doute, très nettement</p>		
	<p>29 : si ça marche pas c'est pas grave (contrasté avec 10 &gt; sans chercher à savoir... quelle est la phrase la plus simple, alors qu'avant je cherche la phrase la plus simple)</p>		
	<p>30 : c'est un ton familier</p>		
	<p>31 : c'est machinal dans le sens que bah y a un certain accord de fond</p>		
	<p>32 : je parle aussi un peu à AD</p>		

## Grille AA S01

FoT_ProdSuj	FoT_Gn	FoT_PE	FoT_IPC
1 : ordre dit gentiment	A : je suis très copine avec ce robot – je joue le jeu	[I] : ce n'est pas un être – c'est comme dialoguer avec quelqu'un qui n'a ni sentiments, ni culture, ni langue, mais allait m'obéir par la voix  ** comparatif petit prince de Saint Exupéry : le petit robot c'est comme la rose qui... dans la vie d'une personne	(a) : ça c'est très joli
2 : ordre anonyme qui n'est pas à contester 2' : c'est un acte obligatoire	B : pour moi ça avait été très confus	[II] : c'est un objet qui n'a pas forcément une âme ou une intelligence humaine – c'est un personnage	(b) : il faut accepter tout ce qui est fantasque, l'irréalité, le rêve
3 : ordre fait avec tout le respect qu'on a pour quelqu'un	C : c'était ma vie qui continuait	[III] : c'est pas seulement un jouet	(c) : c'est pas du tout la même atmosphère
4 : ordre qui n'est pas dit durement – c'est un ordre gentil presque affectueux 4' : à mon service, mais dans la gentillesse 4'' : ordre de connivence – de gentillesse	D : c'est une séquence entourée de gentillesse et de douceur	[IV] : c'est toujours un robot, mais c'est un robot sensitif – un objet un peu surnaturel	(d) : extraordinaire parce que je l'ai imité et pas du tout parce qu'il l'a fait
5 : je le félicite parce qu'il a répondu	E : connivence qui se fait entre la personne physique et ce petit robot	[V] : c'est comme un être vivant – plus petit que moi, du point de vue mental et du point de vue réflexion	(e) : on ne comprend pas toujours les paroles qu'il dit, pas toujours, mais ça n'a aucune importance. C'est la modulation qui fait que « ah il est content »
6 : on a correspondu	F : on va être deux à le faire – je l'associe à mon travail, mais je ne lui demande pas son avis	[VI] : comme un petit enfant ou un objet que par ma question j'ai transformé en un être	(f) : les lampes de chevet c'est plus intime
7 : ordre commun/collectif – travail commun – regarde on va le modifier	G : il fait partie de ma vie	[VII] : un petit serviteur dont la grosseur et l'inefficacité me fait li donner de l'affection	(g) : je les découvre (les lumières colorées) et je me dis ah ça c'est drôle, on n'est pas

## Annexes

			dans une maison habituelle
8 : ordre que pour lui – autoritaire 8' : impératif – commandement – t'as pas à dire non	H : on commence à avoir une compagnie intéressante et gentille	[VIII] : petit bonhomme	(h) : (la télé) elle est fatigante
9 : tu as fait – maintenant on passe / transition / collectivité	I : je suis contente – il y a de l'affection parce qu'il est petit	[IX] : un être sourd-muet, mais sensitif	(i) : c'est juste parce que je l'ai vu passer pour partir
10 : c'est une demande et on va vous servir	J : je l'attends comme un personnage important dans l'histoire, il a toute sa place	[X] : il devient un peu un maître d'œuvre – tout tourne autour de lui tout à coup	
11 : c'est la constatation que l'ordre a bien été suivi 11' : approbation d'un ordre que j'ai déjà donné – c'est bien le résultat que j'attendais	K : je suis contente – il a fait ce qu'on voulait / il a réussi K' : appréciation d'un objet qui change au fur et à mesure où ce que l'on souhaite arrive	[XI] : on parle comme ça à un bébé qui a 8 mois – c'est un petit ce robot – un tout petit enfant	
12 : suggestion	L : il a compris comme un enfant comprend – c'est très affectif	[XII] : l'habitué des lieux – la personne qui vous a aidé – c'est le maître de maison – c'est lui le dominant	
13 : souhait personnel – je lui raconte une histoire	M : si le robot veut pas on pourra pas le faire – toujours avec l'accord du robot		
14 : je le félicite parce qu'il a compris	N : il a eu pitié de moi		
15 : regarde comme c'est bien – je lui fais part de mon admiration	O : je vois qu'il est content de le faire		
16 : ordre où on a l'impression qu'on parle à un jeune enfant – ton de gentillesse qui sous-entend qu'on serait content qu'il fasse ça	P : le petit robot n'avait pas l'air intéressé par ce qu'il avait mis		
17 : interrogation directe à ce robot	Q : on est d'accord tous les deux – c'est un jeu à deux		

## Annexes

18 : un côté fâcherie à l'intérieur – ouh mes oreilles	R : c'est lui qui va décider		
19 : voilà, maintenant c'est bien	S : là je sens que c'est lui, le petit robot qui a envie de...		
20 : un ordre volontaire parce que je ne suis pas contente	T : on a la même voix		
21 : l'ordre à soi-même	U : elle se demande s'il a compris de changer de sujet comme ça brutalement		
22 : dialogue d'entente gentille 22' on a parlé tous les deux sur le même ton – le ton de la joie d'être ensemble et de faire quelque chose 22'' : complicité	V : ras le bol, elle ne sert à rien V' : elle en a marre quoi		
23 : je lui demande son avis, mais c'est impératif – tu vois ça vaut pas la peine	W : on abandonne un peu l'idée de la télévision		
24 : demande d'avis – questions où s'il n'est pas d'accord on ne le ferait pas	X : c'est la découverte de l'inconnu – c'est lui qui me montre le chemin		
25 : je le suis – c'est lui le patron – maintenant c'est moi qui le suit	Y : la personne attend que le petit robot qui ne parle pas donne quand même un signe de son accord ou de son désaccord.		
26 : dérivatif à une idée, à une constatation immédiate	Z : elle a besoin de savoir où il veut aller – elle est invitée		
27 : je trouve qu'il fait pas très clair, qu'est-ce que tu en penses : d'abord suggestion puis l'ordre	AA : je ne me suis pas déplacée pour rien, je profite.		

## Annexes

28 : pensée de travail à faire parce qu'on a tout réussi – fin d'un épisode	AB : vous me prêtez-là un caractère un peu léger et opiniâtre		
29 : c'est un affirmatif	AC : on finit par se demander si la personne avec qui on a conversé euh... s'y retrouve		
30 : définition totale – c'est la solution de fin d'un travail	AD : la personne qui parle se rend compte qu'en fait c'est un être qui n'a pas pu me dire		
31 : besoin – demande de service	AE : maintenant il s'agit de faire des choses avec ce robot		
32 : lui faire comprendre – c'est une suggestion, mais il y a un peu d'ordre dedans	AF : on a besoin de quelqu'un pour le faire		
33 : on a fait notre travail – on change d'atmosphère – on passe à autre chose – un point virgule – une proposition			
34 : c'est le point d'orgue – toute l'histoire a été dite			
35 : c'est une proposition			
36 : ton intérieur d'une pensée qu'elle dit 36' : ton intérieur – dialogue personnel – monologue			
37 : allez c'est le moment			

## Grille AA S03

FoT_Gn	FoT_ProdSuj	FoT_IPC	FoT_PE
A : j'étais dans l'amusement	1 : c'est pas sous la forme d'un ordre, j'avais plus envie de paroles sympas – je lui parle plus gentiment	(a) : je me suis dit instinctivement « il parle » donc je peux parler aussi	[I] : c'est quelque chose de pas courant, c'est une découverte et ça m'amusait ce petit truc – c'était un jouet
B : je me sentais à l'aise	2 : plus interrogatif quand même, « qu'est-ce que t'es capable de faire d'autre », on va voir quoi	(b) : j'écoutais ce que la personne me disait et c'est tout	[II] : je me disais qu'il pouvait être une souris ça aurait été pareil, c'est un petit robot, il suit comme une souris – je le vois moins comme un jouet
C : je commençais à m'habituer, j'étais moins dans la découverte	3 : dans la découverte, à essayer de ce qu'il est capable de faire	(c) : j'ai mordu à l'hameçon – je me suis dit bon bah je vais faire mon expérience toute seule, c'est pas plus mal – je vais m'exercer avec lui	[III] : mais c'est pas une personne qui parle, c'est un robot – y a un côté humain qui sort et c'est dérangeant
D : j'ai commencé à me poser des questions en me disant, mais jusqu'où il peut aller	4 : je lui ai parlé gentiment, mais pas comme au départ – c'est amusant et ça reste ludique 4' : j'ai trouvé que vu qu'il avait parlé gentiment, je lui ai répondu gentiment	(d) : je me suis dit « mais il se déplace » – c'était un peu étonnant (d') : c'est un peu l'étonnement parce qu'il répond, parle (d'') : j'ai été étonnée de la longueur de sa phrase (d''bis) : même effet que la première longue phrase et là ça me dérange – c'est surprenant qu'il puisse donner les programmes comme ça (d''') : il a pas employé le même mot donc ça m'étonne un peu	[IV] : je me suis un peu désintéressé de lui – j'ai une gêne au fon de moi par sa présence
E : je suis très étonnée, c'est bizarre, étrange, ça va trop vite pour moi, très surprenant – dérangeant	5 : ça devient plus interrogatif, je cherche un peu plus à comprendre, « comment je dois me comporter avec lui » – dans l'étonnement	(e) : je me suis dit, mais il sait pas s'il y a de l'eau dedans	[V] : c'est curieux parce que parfois on se dit c'est un objet, mais il est capable de rendre service, et des moments on sait plus trop où on en est
F : j'ai fait le tour des questions, je sais plus trop quoi lui demander	6 : je suis dans les essais, pour voir jusqu'où il peut aller – je voulais le tester	(f) : j'ai plus pris conscience de lui à ce moment-là – je l'ai un peu plus personnalisé	[VI] : je lui dis plus merci, j'ai pas envie de le personnaliser, de l'humaniser – c'est pas un personnage,

## Annexes

			c'est un engin – c'est un bout de plastique quoi – j'ai lutté contre ça en fait
G : j'ai fait le tour et j'aurais bien aimé avoir votre retour, votre contact – j'avais des questions à poser	7 : un autre type d'interrogation, je me suis dit c'est pas possible, il analyse pas, enfin il va pas aller aussi loin – y a quelque chose que je pige pas	(g) : il m'a amusé à ce moment-là	[VII] : là je l'avais remis en l'état de robot, mais c'est vrai qu'esthétiquement il est mignon
H : pourquoi elle l'emmène ? là je comprenais pas trop ce qui se passait	8 : je vérifie quand même, on reprend le cours des commandes	(h) : j'ai eu l'impression qu'il me reprenait quoi en quelque sorte, « c'est pas bien ce que t'as dit c'est pas comme ça »	
I : j'en ai un peu marre – j'ai fait le tour, j'ai plus envie – je trouve le temps long	9 : je continue, il n'y a plus l'étonnement du début, y a moins de surprise, mais on continue, c'est toujours intéressant – on s'habitue trop vite qu'on lui donne des ordres qu'il exécute	(i) : c'est à ce moment-là que je réalise qu'il obéit à des commandes et pas à moi	
J : j'attends vraiment avec impatience le retour parce que je comprends plus pourquoi ça dure si longtemps	10 : je lui réponds parce qu'il m'a posé la question – c'est pour faire quelque chose quoi, c'est une réponse un peu banale	(j) : j'ai commencé à me dire « ouille, il va me parler sans arrêt », il va me suivre tout le temps, et ça m'a dérangé (j') : je sais pas, mais qu'il me suive pas comme un petit chien quoi, ça ça m'a dérangé quoi. Non à la limite qu'il puisse répondre, mais de loin, et qu'il se déplace à ce moment-là, mais pas qu'il suive comme ça (j'') : je n'aime pas ça, j'ai envie de lui dire fiche moi la paix, je t'appellerai quand j'aurai besoin de toi	
K : c'est vraiment là que j'ai commencé à regarder l'agencement plus que le robot	11 : je suis toute seule, il faut que je m'occupe – je sais plus quoi faire	(k) : des moments où j'ai décroché – je suis partie regarder l'appartement parce que moi je veux aussi déménager dans un appartement plus petit – j'ai été très intéressée par l'aménagement de l'appartement	
	12 : pour voir s'il devait se déplacer ou si de plus loin il pouvait répondre à ce qu'on lui demandait – je le teste	(l) : je me dis, mais qui tu es toi, qu'est-ce que tu es – beaucoup d'interrogations	

## Annexes

	13 : je lui parle comme à un petit chien, assis, couché, c'est bête hein, mais c'est réflexe	(m) : ça me dérange (parce que je l'appelle et il me répond) – c'est curieux comme sensation (m') : c'est toujours un peu ce côté humain qui me dérange, mais le rire ne m'a pas trop dérangé	
	14 : là c'est un peu la lassitude parce que je sais plus quoi faire 14' : je sais plus quoi faire – lassitude et ennui aussi	(n) : là quand il bouge, on a l'impression qu'il est impatient	
	15 : pour voir s'il re-fonctionne	(o) : qu'il bouge, qu'il se tourne un peu vers moi, je me suis demandé ce qu'il attendait – « non, mais attends il attend que je parle euh... qu'est-ce que je peux faire » (o') : on a l'impression qu'il danse, il suit la cadence – c'est cette impression de le voir à l'affût, un peu face à moi	
	16 : on le sent dans ma voix que j'en ai marre	(p) : il vient pas, il marche plus – je me suis surprise moi-même, je m'inquiétais de, pourquoi il répondait pas	
	17 : ça va bien comme ça – j'épuise les ordres	(q) : je me suis dit « ouais il comprend bien, ça fonctionne »	
	18 : là je pense « occupe-toi » – ton désabusé	(r) : qu'il répète ça, ça m'a lassé quoi	
		(s) : à partir du moment où ces phrases ont été prononcées, ça m'a titillé	
	19 : je veux montrer à AD ce qu'on peut faire avec ce petit bout	(t) : je veux pas me retrouver dans le noir	
	20 : là c'est parce que je voulais le faire (parce que je voulais quitter la pièce)	(u) : je me suis dit encore 10 minutes quoi – en fait quand la personne est partie je me suis dit, mais pourquoi je lui ai pas demandé ce que je faisais là quoi	

## Annexes

---

	21 : en fait j'étais un peu tenté de lui parler à lui inconsciemment, mais en même temps je me disais c'est une chose	(v) : bah là j'étais contente parce que là je me suis dit je vais enfin pouvoir parler avec elle pourquoi je suis venue et puis parler d'Emox	
	22 : il est rigolo et sympa donc on a envie de l'imiter	(w) : mais il a pas de réflexion, comment il peut comprendre ce que je lui dis ?	
	23 : c'est une démonstration pour E2 de ce que j'ai découvert	(x) : c'est certainement une fierté très mal placée, mais j'étais contente de montrer ce que j'avais découvert avec lui	

Grille AA S04

FoT_Gn	FoT_ProdSuj	FoT_PE	FoT_IPC
A : je me demande dans quoi je m'engageais – qu'est-ce qu'on attend de moi ? / je suis tout ouïe pour comprendre / il faut déjà découvrir si ça marche	1 : pour essayer de comprendre comment ça marche – j'essaie successivement	[I] : un petit machin – truc	a) c'est un appartement hyper lumineux – c'était éblouissant parce que moi je crains beaucoup la lumière
B : je voulais une question, mais j'ai pas pu	2 : pour moi c'était un jeu	[II] : objet utile – un outil qui dans la mesure où il tomberait en panne je serai dans le beau drap	b) c'est pas beau de se voir hein ** pas véritablement une étiquette d'auto-annotation car constat au visionnage, mais relevé car conséquences sur la perception du sujet sur sa propre personne
C : je suis sidérée – j'ai trouvé ça magique / émerveillement	3 : j'essaie d'établir un dialogue, mais aussi de jouer	[III] : comme un enfant	c) je me suis dit bon ça fait un moment que tu piétines et que t'es debout, assieds-toi
D : questionnement et dynamique de « et si j'étais dans un fauteuil est-ce que je peux commander – est-ce que ça va marcher »	4 : j'expérimente – je teste pour voir – je veux tout essayer en fait 4bis (version itérative) : je ré-expérimente – je re-teste – je redémarre – je réessaie	[IV] : est-ce que ça peut être une compagnie ?	d) je me suis même demandé s'il fallait que je rappelle la dame (E1)
E : mais il me suit partout	5 : intérieurement je me suis dit « bon, ça y est, j'ai passé un cap »	[V] : tu n'es jamais qu'un objet – t'es ni enfant – ni animal – ni rien	e) je me suis demandé est-ce que tu pourrais t'allonger et puis j'ai dit non il faut que t'enlèves tes godasses
F : j'ai rien compris au fonctionnement	6 : mais est-ce qu'il marche pas ce truc ?		f) la télé c'est un support de vie important pour moi
G : il veut pas m'obéir hein	7 : c'est pas le tout tu fais coucou, mais il faudrait m'obéir		g) la lumière bleue c'est finalement ce que je préfère – c'est nettement mieux
H : tu commences à me faire chier – ça m'a	8 : je lui communique mes pensées		h) il se met de nouveau à dérailler

## Annexes

énervé			
I : là je suis en colère – ça pinaille là ce truc	9 : mettre en route		i ) faut quand même qu'ils mettent une voix un peu mieux que cette espèce de miaulement comme un petit animal
J : laisse tomber ça marche pas, on va aller à côté	10 : j'insiste		
K : on se comprend pas	11 : pourquoi ça a été arrêté ?		
L : on se comprend de nouveau	12 : j'essaie de voir si elle l'a réparé		
M : je suis quand même perplexe hein – je trouvais ça bizarre – je me suis questionnée	13 : est-ce que j'ai eu deux idées : éteindre les la lumière et fermer les rideaux ?		
N : j'ai l'impression que ça répond à nouveau – maintenant ça y est ça marche	14 : y a un peu d'humour		
O : t'as pas dû demander ce qu'il fallait	15 : comme on parle à un enfant – hé dis donc faut pas t'endormir 15' : des enfants j'en ai qui viennent prendre des cours, c'est un peu ma manière de m'adresser à eux		
P : on retrouve notre langage commun	16 : j'ai fait le tout		
Q : je vois pas trop ce que je peux faire dans cette pièce	17 : un peu sec, mais il faut qu'il m'entende, j'articule vraiment		
R : on dirait qu'il veut communiquer avec moi	18 : je lui réponds – je communique avec lui, y a vraiment une communication		
S : il faut pas te laisser piéger comme si c'est un animal ou un enfant	19 : c'est logique j'ai fait ça donc après je fais ça		

## Annexes

---

T : il m'agace – bon c'est pas lui qui va me tenir compagnie quoi – il m'énerve	20 : je vois pas trop ce que je peux faire		
U : je trouve le temps long	21 : je lui répète deux fois, j'ai pris l'habitude – ça va pas assez vite		
V : je viens de vous retrouver (humain) c'est plus intéressant que cet objet	22 : je te comprends pas de toute façon, j'ai plus intéressant – j'ai mieux à faire		
W : ça me contrarie, j'ai pas pu tester comme chez moi	23 : je vous commande ce qui s'est passé – je vous fais un compte rendu		
	24 : rester dans des codes de langage		
	25 : je suis impérative – c'est les ordres		
	26 : je vous montre que ça marche pas		

## Grille AA S05

FoT_Gn	FoT_ProdSuj	FoT_IPC	FoT_PE
A : je m'appliquais à écouter	1 : aucune initiative	(a) : j'ai pas fantasmé	I : je n'avais pas vraiment d'a priori parce que je n'en avais jamais vu
B: plutôt à l'aise – je pensais que c'était pour tester l'appart, j'étais prête	2 : ton assez professoral	(b) tout ce qui a de plus naturel – je prenais contact – aucun <i>a priori</i>	II : je devais penser qu'il obéissait bien
C : il fallait faire un apprentissage, je le faisais sagement – je m'oblige	3 : j'articule particulièrement bien parce que je m'applique, je veux qu'il m'entende bien – j'articule bien c'est un ton neutre	(c) : je vérifie qu'il obéit (c') : je vérifie que tout marche bien	III : je suis assez étonnée qu'il réalise tout quoi
D : là je commence à lui parler	4 : le ton de voix a changé, hein, c'est plus... cordial je veux dire	(d) : là je pense que je le laisse descendre complètement (le store) parce que je veux voir s'il s'arrête tout seul	IV : le robot c'est un peu de la magie, toute la techno pour moi c'est un peu de la magie – pas tellement émerveillée par le robot que par la personne qui a fabriqué le robot
E : je m'occupe pas beaucoup de lui en fait, je vérifie simplement que l'ordre soit exécuté	5 : j'ai répété l'ordre parce qu'il l'a pas mise – plus autoritaire et moins aimable 5' : il a pas obéi donc je le rappelle à l'ordre	(e) : [regard-déplacement] j'ai compris qu'il me suivait et que c'était nécessaire qu'il me suive	V : je commence à le considérer comme de euh... un partenaire, comme s'il avait une volonté
F : j'comprends rien de ce qu'il me raconte, allez j'arrête ce truc	6 : j'arrivais pas alors je lui demandais – c'est plutôt sympa	(f) : je me suis demandée s'il y avait de l'eau dans la bouilloire	VI : j'avais compris que ce robot s'occupe de moi – maintenant quand il me parle je lui réponds toujours – il a une fonction de surveillance
G : maintenant je trouvais qu'il obéissait bien, c'était son boulot, je ne suis plus	7 : c'est un petit ton comme si c'était un gamin qui voulait pas m'obéir quoi	(g) : [geste] – un peu de perplexité, c'est pour ça que j'ai arrêté assez vite	VII : je lui parlais comme à un animal et tant mieux parce que j'avais peur de l'avoir gadgétisé

## Annexes

émerveillée		(g') : c'est de la perplexité, elle est par rapport à l'appartement	
H : je suis là, je prends connaissance des lieux, je vois que je tournicote un peu	8 : allez hop c'est un truc dont je me passerai	(h) : j'ai attendu qu'il arrive pour lui donner l'ordre – je pensais qu'il fallait qu'il soit là	VIII : même si je lui accordais un petit rôle d'une présence à partir du moment où il y a un être humain, il fait son « micmac » et c'est tout
I : je commence à échanger avec lui I' : quand je lui parle pas normalement, il établit, il établit le contact	9 : je me suis habituée – il me parle, je lui réponds 9' : c'est devenu un jeu entre nous, il me parle, je lui réponds	(i) : ça m'a amusé (i') : je me suis dit « tiens, bon c'est marrant ce truc-là » (i'') : je pense que ça m'amuse quelque part, ça m'amuse de communiquer avec ce petit robot	
J : pour préparer à passer le temps (manip avec le sac)	10 : je n'ai fait que donner les consignes	(j) : je me souviens d'avoir été assez émerveillée	
K : je m'applique à faire les consignes	11 : il me pose une question, je suis bien élevée donc je lui réponds	(k) : j'ai été assez étonnée	
L : perplexité pas par rapport au robot, mais à l'appart et aux objets à placer	12 : il est gentil, il est pas du tout autoritaire, ça veut dire, ça va, il y a pas de problème	(l) : j'étais étonné qu'il ait pas compris mon premier ordre, j'aime être obéi du premier coup,	
M : il faut que j'occupe le temps, ça me semble un peu artificiel donc assez vite ça me barbe	13 : t'inquiète pas, j'suis pas morte, j'suis là, je m'occupe et je bouge pas	(m) : là je crois que j'avais mal au dos	
N : j'essaie de voir comment je peux m'approprier cet appart	14 : attends, je vais changer d'activité	(n) : je pense que c'est la première fois que je lui parle directement	
O : je me souviens de m'avoir plongée dans l'article, c'était Camus O' : plongée dans la lecture,	15 : je veux pas le contrarier	(o) : je trouvais que c'était assez joli, que c'était joliment fait	

## Annexes

mais ne regarde même plus le robot et ne fais plus attention, j'ai perdu le sens du temps			
P : je me dis je pense que je l'ai adopté, il faut vivre avec lui dans l'appart et bah ça se passe plutôt bien entre nous	16 : je sature	(p) : je sais que je me suis dit « tiens, quand je reste un certain temps sans lui donner d'ordres, il vérifie que je suis toujours bien »  (p') : si je trouve que ça fait un certain temps que je lui ai rien dit, je me dis bon il va s'inquiéter donc je le tranquillise	
Q : je m'en fous, je suis de marbre, je m'en contrefous c'est vous (E2) mon interlocutrice	17 : c'est pour vous montrer ce qu'il peut faire 17' : j'essaie de montrer à quel point j'ai compris ma leçon 17'' : là c'est pour vous montrer les possibilités de l'appartement 17''' : là c'est pour vous montrer que dès que je m'assois un moment sans bouger il intervient	(q) : j'ai presque l'air d'un automate aussi saccadé que le robot	
	18 : je voulais que la commande s'arrête pour qu'on puisse dialoguer (avec E2)	(r) : ça ne me déplaisait pas du tout, je pensais que ça faisait partie de sa fonction de tenir compagnie	
		(s) : la radio ça devait me barber	
		(t) : très contente que vous veniez me libérer	
		(u) : la première fois j'étais pas arrivée, mais en fait je crois avoir compris que c'est parce qu'il faut du temps – là je m'aperçois qu'il fallait lui laisser du temps au système  (u') : je m'attends toujours à ce que ça aille plus vite	
		(v) : j'ai pris conscience que j'ai eu tendance à lui parler à lui	

## Annexes

---

		alors que c'était là (au micro) que tout se jouait	
		(w) : c'est pas forcément le plus beau des compliments, je... qu'il ricane et je trouve pas qu'il rit – c'est un peu systématique, c'est pas un être humain  (w') : quand il rit comme ça sans qu'il y ait d'échange il ricane alors que quand on est dans une conversation je trouve que c'est plus cohérent	
		(x) : j'avais l'impression qu'il allait me dire quelque chose	

## Grille AA S06

FoT_Gn	FoT_ProdSuj	FoT_IPC	FoT_PE
A : j'avais un peu peur de l'enfermement – inquiétude	(1) : j'étais plutôt curieuse là de savoir comment ça marchait	(a) : Je trouvais que ça marchait comme elle m'avait expliqué	[I] : ils arrivent à faire des choses quand même pas mal – contente qu'il y ait du progrès dans la vie
B : je commençais à avoir peur, je paniquais	(2) : je lui fais la commande – je pensais qu'elle le ferait, que le robot ferait ça, j'attendais ça de lui	(b) : je dis que ça marchait pas (b') : je dis rien ne marche – mais comment ça se fait	[II] : comme à une personne
C : je commençais à sentir le temps long – mais où ils sont passés	(3) : je dis encore, j'insistais – je suis têtue j'insiste	(c): bah j'ai dit il avait compris qu'il fallait changer la chaîne, mais il m'a répondu que ça marchait pas – c'est pas qu'il voulait pas le faire, c'est que ça marchait pas (c'): je pensais pas que c'était à cause du robot (c''): le pauvre, lui il n'y pouvait rien	[III] : je trouve que c'est un objet il est pratique, mais là c'est un compagnon pour moi c'est comme une présence
D : là c'est la peur	(4) : je lui demandais de faire – je lui demandais pour tester	(d) : il y avait un feuilleton qui était pas mal et je le suivais tout le temps	[IV] : c'est parler comme à un enfant
	(5) : je le disais comme je pourrai, comme je le dirai à une personne – c'est sans agressivité	(e): je voulais voir s'il pouvait aller plus loin	
	(6) : en fait moi je cherchais ça et pas autre chose	(f) : je me suis dit dès fois elle va marcher, c'est peut-être que c'est momentané	
	(7) : c'est trop fort	(g) : je me disais c'est pas vrai	
	(8) : en fait il les allumait toutes et il faisait déjà clair donc je dis	(h) : j'ai dit plutôt qu'il marchait pas le robot	
	(9) : là je dis gentiment	(i) : Là je me dis peut-être qu'ils vont arriver et m'expliquer pourquoi il marchait pas	

## Annexes

	(10) : là j'ai commencé à lui parler parce que je me disais que comme ça j'aurais moins peur	(j) : ça avait allumé la lumière sans que j'avais demandé (j') : elle s'est éteinte toute seule la télé ?	
	(11) : là je l'ai pas vraiment dit c'est parce que c'était marqué (11') : j'ai fait ce qu'on m'a dit de lui demander	(k) : quand vous êtes seule on aime bien avoir la télé	
	(12) : je peux le faire s'il y avait quelqu'un avec moi	(l) : je regardais comme ça et je me disais c'est bien fait	
		(m) : là je me dis c'est la chaîne qui marchait pas	
		(n) : moi ce que je voulais c'est qu'on vienne (n') : je me demandais est-ce qu'ils vont venir – peut-être ils vont pas venir	
		(o) : je pense que je pourrais pas rester comme ça toute seule s'il y avait pas le robot (o') : je me disais heureusement que le robot était là	
		(p) : parce que je me sentais pas seule je le trouvais gentil (p') : quand il me parlait des choses je me sentais mieux	
		(q) : je me suis dit c'était pas mal, que c'était confortable	
		(r) : étant donné que la télé ne marche pas je vais mettre la radio je me suis dit	
		(s) : je me disais qu'il y avait des choses qui étaient	

## Annexes

---

		trop hautes pour moi	
		(t) : je me suis dit je ne sais pas ce que ça va m'arriver, si quelqu'un rentre, que je connais pas... enfin je pensais à des choses – même si c'est pas vrai je pensais pas dans ma tête	
		(u) : je pense que j'ai dû chercher l'adresse d'une amie – je pensais à mon amie	
		(v) : c'est parce que je trouvais pas – j'étais embêtée de ne pas trouver	
		(w) : j'étais contente de vous voir	
		(x) : j'ai toujours eu peur qu'un robot puisse me faire ça, de l'eau et tout ça parce que je me dis que ça peut brûler – j'avais pas confiance au robot pour ça	
		(y) : c'est marrant je le trouvais sympa	

Grille AA S07

FoT_Gn	FoT_ProdSuj	FoT_IPC	FoT_PE
A : un peu impressionnée – je suis raide au début	1 : je parle doucement / je parlais j’entendais à peine ce que je disais	a : « bon bah je vais le faire »	I : je trouvais tout ça encore naturel / jusqu’à présent je ne m’occupe pas tellement du robot
B : étonnée que ça marche si bien / je suis étonnée de tout ce qui se passe	2 : c’est ma façon de parler pour qu’il comprenne bien / c’était des ordres que je donnais	b : je me dis comment ça va se faire ?	II : je me rends compte qu’il avait quelque chose quand même d’un peu extraordinaire
C : j’avais tellement peur qu’il ne me comprenne pas	3 : je lis euh... je suivais la liste / c’était des ordres	c : le fait qu’il me suive ça m’a perturbé	III : j’avais l’impression d’avoir quelqu’un en face de moi à qui je parlais et qui me répondait
D : j’ai commencé à me dire c’est formidable quoi	4 : je demande quelque chose, je ne commande plus, je demande	d : ça c’est son premier son et là j’ai été... il parlait pas là, il a fait un gémissement, je peux pas dire exactement ce que c’est, mais ça m’a fait quelque chose	IV : on était devenu copain quoi
E : ça m’a surprise	5 : c’est un compliment que je lui fais	e : j’ai pas compris (ce qu’il m’a dit) e’ : je l’ai pas compris, c’était de ma faute	V : c’est comme s’il y avait un petit enfant à côté de moi qui me tire la blouse
F : j’étais occupée par le fait de lui dire ce qu’il faut faire / j’étais toujours obnubilée par la liste	6 : entre les deux, on m’a demandé de lire la liste donc j’obéis à ce qu’on m’a dit et je dis je vais lui faire tout, je vais lui demander ça	f : bon bah qu’est-ce que ça va donner à la fin, comment ça va se passer	VI : un robot domestique – je trouvais ça tellement bien ce petit appareil qui pouvait nous rendre de tels services – appeler quelqu’un ou donner un signal, quelque chose pour qu’on vienne à mon secours
G : maintenant je sais que tout ce que je vais lui demander il le fait / je peux tout lui demander / j’avais aucun doute	7 : c’est des ordres (pas comme au début, plus une demande) et il fait ce que lui demande	g : je l’ai regardé en disant il réagit quoi	VII : il est craquant ce robot

## Annexes

H : bah moi je me disais il va jamais y arriver c'est pas possible	8 : j'essaie de lui demander d'autre chose	h : alors là j'étais soufflée	VIII : c'était peut-être un peu plus qu'un copain, oui c'était euh... la famille quelqu'un de la famille
I: je suis rentrée dedans là quoi en plein là / je commence à trouver ça normal alors qu'au début je comprenais pas je disais que c'est pas possible que ce petit robot il puisse faire tout ça I' : je suis moins étonnée, je dis il peut tout faire	9 : là je lui ai demandé (quelque chose qui n'est pas sur la liste), parce que je pensais qu'il pourrait le faire	i : je commence à lui parler aussi – je me suis habituée au fait qu'il me parle et je lui parle	
J : au début je tâtonnais	10 : je le fais rire là	j : j'ai dit je vais essayer	
K : elle était finie la liste là, et ben je pensais que c'était formidable d'avoir un petit robot comme ça qui puisse faire tout ce qu'on lui demande	11 : voyez là je chante presque	k : je cherche la radio oui, et pourtant la musique était là, mais je savais pas d'où ça venait	
L : j'avais plus de liste, j'ai essayé de faire d'autres choses L' : je dis je vais essayer	12 : au début j'aurais pas eu l'idée de lui dire « merci », mais on se parlait presque quoi, le merci est venu naturellement	l : [réaction Emox] alors là j'ai été... j'ai senti quelque chose là quand même ah oui – je me disais c'est formidable l' : [réaction Emox] j'ai été étonnée et je me suis dit ben il s'amuse	
M : j'attendais que quelqu'un vienne	13 : c'est déjà plus cool, j'ai assimilé le fait que ce petit robot ben il est là et j'ai qu'à lui demander pour qu'il fasse ce que je veux	m : là je me dis « tiens, mais j'ai pas rangé mes affaires »	
N : je voulais surtout pas l'écraser	14: en fait là j'arrivais à lui parler comme on parle avec une personne 14' : j'arrive à lui parler normalement, comme à un être normal	n : je lui ai demandé, mais je savais qu'il le ferait pas, mais je voulais voir sa réaction	
O : je me demande ce que je vais faire	15: il me répondait, je lui parlais, il me	o : je me demande ce que je vais faire	

## Annexes

	répondait, c'était magique		
P : je regardais, je pensais pas, enfin je pensais surtout à ce que je voyais P' : j'étais toujours dans mon esprit de dire ce qui me convenait pas dans l'appartement	16: je lui demande parce que je ne sais plus quoi faire là	p : il m'a fait rire p' : c'est marrant là	
Q : j'étais étonnée, j'étais émerveillée et j'ai dit bon on peut tout, on peut discuter avec lui	17: ah là j'aimerais qu'on continue, pour voir si j'ai pas rêvé, je lui pose la question	q : ça je m'y attendais pas (qu'il voit les images de la télé)	
R : j'étais déçue qu'ils me répondent pas parce que j'aurais voulu qu'ils l'entendent (détail : j'étais en colère, j'ai dit c'est pas possible, vous avez toujours le téléphone dans la main à chaque fois que je vous appelle et vous répondez pas) [commentaire associé à un appel téléphonique d'un proche]	18: je lui parle comme à quelqu'un que je connais, qui vit avec moi, qui comprend tout ce que je lui dis	r : [position debout] je me sentais pas chez moi là quand même	
S : je suis rassurée, je sais que ça existe, que ça peut se faire, c'est une découverte pour moi	19: je me sentais obligée de lui dire merci, parce que c'était tellement bien	s : pour moi c'est... je... j'avais fait mon travail	
T : j'avais fini ce qu'on m'avait demandé de faire, je commence un petit peu de me dire ça fait un moment que je suis toute seule	20: je commence à m'habituer à lui		
U : je cherche ce que je peux lui demander de faire	21: je suis vraiment une grand-mère là dans mon ton		
V : là j'étais complètement avec lui (Emox)			
W : c'était pas désagréable de sentir qu'il était toujours derrière moi, j'étais contente de voir qu'il me suivait partout – et je lui parle, il y a rien à faire hein, chaque fois que je lui ai parlé il a toujours compris			
X : c'était trop intéressant ce que je faisais là			

## Annexes

Y : dans mon esprit j'étais toute seule quand j'ai fait ça			
Z : je préfère qu'il me suive plutôt qu'il soit devant, il est toujours derrière moi			

Grille AA S10

FoT_Gn	FoT_ProdSuj	FoT_IPC	FoT_PE
A : c'est un jeu sans être un jeu, je me dis tiens je vais m'amuser avec lui	1 : un commandement, voyons s'il le fait 1' : c'est un petit commandement, c'était pour lui demander s'il pouvait faire – c'est pour voir s'il sait le faire	(a) : je dis tiens il comprend ce que je lui dis	[I] : c'est marrant un robot qui comprend tout comme ça
B : c'est un peu un contrôle, c'est un peu aussi la curiosité de voir ce qui faisait, ce qui faisait pas	2 : je lui parle comme si c'était un gosse 2' : je lui explique comme à un gosse	(b) : je me suis dit mince j'ai oublié de lui préciser « allumer » mais après j'ai rectifié	[II] : pour moi c'est un robot mais c'était une pers... je lui parle comme si c'était une personne euh... vivante – c'est comme à un gamin, un gosse
C : l'esprit c'est je dialogue avec quelque chose	3 : je me suis dit je vais le faire travailler 3' : il faut que je le fasse travailler, il faut bien qu'il fasse quelque chose	(c) : j'ai fait tiens on va voir s'il va me suivre	[III] : pour moi c'est pas tout à fait une qu'un objet, c'est quelqu'un qui parle – c'est un objet, un robot, c'est pas un humain mais c'est quelqu'un
D : bon là je m'installe	4 : ah tiens je vais lui faire un piège, voyons s'il sait faire	(d) : je me suis dit à moi-même, ah bah il me suit, c'est bon, mais je lui ai pas dit de me suivre	[IV] : on se sent pas seul, je ne me suis pas senti seul, il est avec moi, on est tous les deux, ça me fait une compagnie
E : là c'est de la curiosité pour m'amuser avec lui, de voir ce qu'il va faire	5 : ça reste un jeu et je l'ai testé 5' : c'est un jeu, je me suis amusé	(e) : je me dis là il a pas compris (e') : ça a été un peu long alors je me dis il a pas compris ou il m'a pas entendu	[V] : comme un gadget
F : là c'est pour faire passer le temps parce que je me dis qu'il faut que je m'occupe donc je reprenais le dialogue	6 : pour lui c'est la routine, il sait le faire, il est habitué à faire ça	(f) : aussi que je voulais voir s'il pouvait baisser un peu plus	

## Annexes

pour faire passer le temps	6' : je relis la liste – comme il avait déjà fait une première fois je me dis il sait le faire		
	7 : comme on répond à un gosse, il me parle à sa façon je lui réponds de manière de dire j'ai compris ce que tu as dit 7' : je sais pas ce qu'il me dit mais je réponds	(g) : là je réfléchissais à qu'est-ce que j'allais lui faire, j'étais en train de réfléchir – j'étais en train de calculer là, oui je réfléchis là	
	8 : je sais qu'il a pas fait de suite c'est pour ça que je lui redis	(h) : je réfléchis là, lui faire faire quelque chose en dehors du programme	
	9 : on fait la pause et moi aussi – on va se poser 5 minutes 9' : c'est tu te reposes, t'as bien travaillé, c'est pas qu'on est fatigué mais souffle, mais c'est pas un commandement	(i) : je dis il a fait à peu près tout ce que je voulais donc maintenant on va s'asseoir un moment	
	10 : je réponds à son rire	(j) : c'est une jolie musique alors dans mon esprit je me dis il a bien choisi	
	11 : là c'était pour l'arrêter, on va faire autre chose 11' : non là elle était un peu forte alors je lui ai dit de baisser 11'' : on va pas laisser la lumière donc je lui dis éteins 11''' : pour ne pas être sans bruit parce que sans son ça fait triste donc comme ça on n'est pas tout seul – pour lui expliquer qu'il faut qu'il rallume 11'''' : je lui fais rouvrir pour ne pas rester dans le sombre 11''''' : je me dis pas la peine de laisser trop	(k) : je me disais voyons s'il va faire quelque chose sans qu'on lui dise ou s'il attend que je lui dise quelque chose	

## Annexes

	longtemps, j'ai vu que ça fonctionnait, je lui demande d'éteindre		
	12 : je faisais un commentaire	(l) : je réfléchissais là en pensant à E2	
	13 : c'était pour lui faire voir que j'étais là, que je l'oubliais pas 13' : là je l'appelle pour lui faire voir que je suis toujours là 13'' : je lui redis on va aller, on repart pour mettre ça	(m) : c'est marrant quand il rigole – c'est rigolo – on ne s'ennuie pas parce qu'il y a un échange	
	14 : il me pose la question donc je lui réponds	(n) : je pensais que vous n'alliez pas tarder à revenir, on arrive à la fin de l'expérience là (n') : j'attendais qu'on me dise que c'est fini	
	15 : il faut lui faire voir que c'est bien 15' : là il m'a bien entendu, il a bien compris	(o) : je commençais à avoir un peu chaud	
	16 : il n'avait pas eu le commandement donc il n'avait pas à faire ce que j'ai pas dit	(p) : j'ai vu qu'il bougeait mais il m'a pas parlé alors moi non plus, je me dis il change de position	
	17 : je m'amuse, je rigole, pour lui faire voir que c'est pas qu'un objet, on parle	(q) : dans ma tête je me dis mais je lui ai pas dit de partir – je me dis il fait des choses hum... j'ai trouvé ça bizarre – mais où c'est qu'il va ?	
	18 : je lui parle, il répond, je suis pas tout seul	(r) : je me dis tiens, il voit que je bouge pas, il me fait bouger	
	19 : c'était pour voir jusqu'à quel point il pouvait me répondre	(s) : là j'ai voulu le piéger un peu	
	20 : je relis la liste	(t) : là je réfléchis, y a quelque chose qui me	

## Annexes

		turlupine, je commence à soupçonner quelque chose (t') : il est bien plus programmé que ce qu'on dit	
	21 : à la personne qui a tapé je dis oui	(u) : c'est un objet qu'on va mettre dans le garage, il partait comme si il allait se garer, on est tous les deux, il peut rester là parce qu'il n'a pas à partir	
	22 : ça c'est parce qu'on allait partir, je laisse les choses comme elles étaient	(v) : il y a plein de choses qui m'intriguent donc je regarde, mais là j'ai rien vu	
	23 : c'est pour vous faire voir que j'ai compris qu'il était programmé 23' : je vous fais voir qu'il écoute, qu'il fait bien ce que je lui dis, il est bien réglé 23'' : je vous montre ce que j'ai fait faire 23''' : vous faire voir qu'il sait comment il s'appelle, qu'il a su me répondre	(w) : il bouge tout seul, je me dis pourquoi	
	24 : je le félicite, sous-entendu que les personnes qui ont travaillé dessus et bien je reconnais que c'est bien – en lui parlant à lui je parle à vous – le compliment c'est pas tant pour lui c'est pour vous	(x) : il sait peut-être pas tout faire, je me dis le robot il est gentil mais il peut peut-être pas – mais sinon tant pis	
		(y) : qu'est-ce qu'il fait, qu'est-ce qu'il veut me dire ?	
		(z) : je le vois plus donc ça m'intrigue, je me dis où il est	
		(ab) : le doute m'était oublié, je pensais plus au doute, c'est passé j'étais avec vous	

Grille AA – S11

FoT_Gn	FoT_ProdSuj	FoT_IPC	FoT_PE
A : la surprise, c'est plus une découverte, on a l'impression d'être chez soi mais chez quelqu'un d'autre	1 : je suis poli, pour faire connaissance, c'est une entrée en matière	(a) : tiens c'est bien ça répond au doigt et à l'œil quoi, presque	[I] : j'aime la technique, je trouvais qu'il était relativement petit, j'essayais de découvrir la chose
B : je suis curieux je découvre tout B' : c'est la découverte, l'évolution – je scanne un peu la maison quand même [regard] B'' : je regarde ce qu'on peut faire	2 : on va bien voir ce qui se passe	(b) : j'essaie de voir quel rapport on pouvait avoir de se trouver devant un robot, si l'eau c'était automatique	[II] : je donne un petit côté humain à la mécanique
C : j'étais bien concentré	3 : on donne un ordre 3' : plus pour pouvoir donner un ordre	(c) : tiens ça fonctionne bien – ça correspondait à ce (aux choses pratiques) que je m'attendais	[III] : j'oublie le côté robot – c'est un compagnon d'une personne dans le besoin
D : je regardais un peu quel genre de chose on peut lui dire, je dois fouiner pour regarder un peu tout, essayer de comprendre	4 : est-ce que c'est possible – c'est un test 4' : c'est pour voir si c'était lié à quelque chose 4'' : j'ai dit peut-être qu'il y a une manœuvre	(d) : la réponse n'était pas toujours compréhensible (d') : j'ai dit bon y a des intonations ou des choses qui sont pas bien perçues quoi	
E : on parle et on a une réponse, j'ai établi un dialogue	5 : je lui commente pour essayer de l'aider de comprendre de ce qu'on fait – pour moi c'était le langage des personnes futures	(e) : le petit bruit j'en ai conclu que c'est quelque chose qui l'a dépassé c'est un petit « ouf » du robot	
F : maintenant je fais quoi	6 : je sais que la radio est télécommandée c'est plus facile de le mettre en route	(f) : j'ai dit c'est la question qu'il fallait pas lui poser	
	7 : j'avais l'impression que ça devrait pouvoir se réaliser	(g) : tout n'est pas possible	
	8 : je voulais me rendre compte que j'ai bien	(h) : il était pas concerné	

## Annexes

	donné l'ordre, le bon ordre		
	9 : je voulais éteindre mais je ne trouvais pas le... c'est rendre plus humain qu'un ordre	(i) : je dis « il s'est trompé »	
	10 : je lui donne des phrases de passage	(j) : il est parti tout seul avant moi	
	11 : c'est dans la suite à une question 11' : c'est une suite logique	(k) : j'ai regardé l'agencement (de la salle d'eau) et je me suis dit bon c'est un truc provisoire c'est pas non plus un...	
	12 : gentil, gentil petit robot	(l) : quoi faire ben je sais pas donc j'ai essayé de réfléchir et c'est tombé sur (la télé)	
	13 : c'est ce que je veux	(m) : il a dû me poser la question « est-ce que je fais quelques chose »/ « qu'est-ce qu'on fait maintenant »	
	14 : voilà les actes que j'ai fait et qu'il avait conduit convenablement 14' : c'est pour vous montrer qu'il est fonctionnel qu'il connaît bien les choses 14'' : je veux vous montrer que j'ai posé ces questions	(n) : je regarde la télé maintenant, laisse-moi tranquille 5 minutes	
	15 : je suis vraiment gentil avec le petit robot et très poli avec lui	(o) : je me dis il cherche quelque chose	
	16 : on vérifie si ça marche	(p) : j'ai dit bon sûrement que le manque de silence a dû le perturber, le manque d'ordre plutôt	
	17 : j'ai dit pourquoi il m'a laissé tomber, petit robot infidèle	(q) : tiens cette pièce, est-ce qu'il répond toujours	

## Annexes

---

	18 : c'est pour arrêter enfin, pour finir, terminer	(r) : comme de temps en temps le robot dit « je peux faire quelque chose » bah ça évite qu'il pose la question trop souvent	
		(s) : je me dis eh pourquoi qu'il a bougé	
		(t) : j'ai dit oui, il tient pas seul quoi	
		(u) : j'ai dit ben c'est bon je vais le faire – je ne voulais pas qu'il insiste trop pour chercher à ouvrir une porte ou fermer le pauvre petit robot	
		(v) : je n'avais pas de réponse je dis ben y a un problème de radio transmission	
		(w) : on était deux il me suivait, maintenant que vous êtes là (E2) il veut plus me suivre	

Grille AA S12

FoT_Gn	FoT_ProdSuj	FoT_IPC	FoT_PE
A : j'ai trouvé l'appartement très beau, c'était un vrai appartement d'habitation	(1) : je faisais mon apprentissage, je voulais qu'il comprenne ma voix (1') : je pensais qu'il était en train d'enregistrer ma voix donc j'ai voulu tout dire	(a) : ça m'a trop fait rire, c'est trop rigolo	[I] : c'était impressionnant, j'aimais bien parce que je trouvais que ça ressemble à un petit chien à cause de sa taille
B : je fais pas trop attention, je faisais plus attention à l'ordre, enfin quand je voyais le dessin des rideaux (feuille de commande) je voyais plus ça (le robot)	(2) : c'était l'ordre et voir si l'ordre est bien exécuté	(b) : j'aimais bien (le bb Emox) et j'avais bien compris que c'était des bruits différents selon ce qu'il voulait répondre ou pas	[II] : là je le prends pour une personne déjà – ça fait comme une vraie personne
C : j'ai commencé à être vraiment dans les ordres pour la vie de tous les jours	(3) : je voulais tout lire – je suis en train de lire (3') : je lisais pour bien mettre la bonne formule	(c) : (quand il tournait comme ça) je croyais qu'il était content – je voyais qu'il était content (c') : quand il fait des tours, qu'il dit des petites phrases, je trouve qu'il a de l'humour quoi	[III] : on dirait qu'il a un cerveau [III'] : j'oublie quand même pas que c'est un robot mais je pensais pas qu'il avait une intelligence mais presque
D : parce qu'il bougeait plus, j'attendais que vous veniez	(4) : j'ai pas fait comme je voulais, il fallait tout que je répète	(d) : ça marchait pas, je pensais que c'était un blocage	[IV] : j'ai l'impression que c'est... pas un duo mais un partenaire
E : j'étais très fière de faire cette expérience, moi j'étais contente – j'attendais que vous le répariez	(5) : j'ai compris que celui-là (5') : j'ai mis ça parce que j'ai dit « bon il va comprendre »	(e) : là je fais la curieuse	[V] : j'ai trouvé que c'était comme un petit animal, un vrai petit animal de compagnie
F : je pensais que vous arriviez pas à réparer quoi	(6) : je m'amuse un peu quand même, c'est plus trop les textes purs (6') : je me suis amusée et défoulée – c'est pour voir s'il comprenait, je disais plein de choses exprès pour voir s'il comprenait toutes les nuances	(f) : je trouvais très joli comme c'était installé	[VI] : je le prends comme une personne qui a des sentiments
G : je pensais qu'il fallait tout que je répète	(7) : mais j'étais très distraite, parce que je le filmais et	(g) : il met du temps à être réparé le	

## Annexes

	dès fois je le loupais	petit chien	
H : j'ai vu que ça marchait, j'étais ébahie quand même de voir que tout marchait	(8) : c'était un ordre (parce que j'entendais pas assez	(h) : je suis même patiente je trouve, j'ai dit « ben c'est l'expérience »	
I : je suis habituée, on se familiarise	(9) : maintenant je brode un peu	(i) : je vais voir qu'il me suit et j'étais trop contente – j'étais surprise de voir qu'il me suivait	
J : je me suis beaucoup amusée, je me suis défoulée	(10) : je pense que c'est un dialogue, je fais toujours consciencieusement ma lecture mais je pense que j'ai pris du plaisir, c'est autre chose	(j) : c'était un petit peu compliqué, je savais pas trop faire, j'avais pas compris	
K : j'ai un esprit, pas obéissant, c'est un petit peu sans réfléchir mais quand on me dit de faire je fais	(11) : c'est un peu plus un jeu (11') : je joue là, je pense que je m'amuse	(l) : il comprenait bien ce que je disais (l') : ça me disait oui j'ai bien compris	
L : eh ben encore en panne, j'ai pensé, comme la dernière fois	(12) : c'est pour le récompenser – je le félicite	(m) : je trouvais que c'était rigolo, je suis quand même admirative	
	(13) : c'est moi qui ai inventé la phrase, parce que je voulais faire chauffer mais j'ai oublié comment faire pour lui demander	(n) : pour mieux l'entendre parce que je trouvais que ça résonnait un peu, je pensais que de plus près j'entendrais mieux	
	(14) : là je voulais le voir répondre (14') : c'est aussi pour le filmer	(o) : je voulais aussi voir s'il comprenait (o') : je voulais voir s'il allait me suivre, pour voir s'il comprenait ce que je faisais	
	(15) : c'est aussi pour le tester, pour le tester aussi pour voir ce qu'il peut répondre	(p) : je revoyais ce que j'avais filmé	
	(16) : s'il parle je pense qu'il faut que je lui parle, comme	(q) : j'ai dû proposer une action qu'il	

## Annexes

---

	s'il y a un invité, avec un invité on peut pas rester muet	ne comprenait pas	
	(17) : parce que je voulais boire mon thé	(r): je me dis « tu boudes? »	
	(18) : pour tester ce que je n'avais pas fait là – je recommence		
	(19) : est-ce que tu boudes ?		

Grille AA S14

FoT_Gn	FoT_ProdSuj	FoT_IPC	FoT_PE
A : j'étais attentive à ce qu'on me disait	(1) : j'étais dans l'expectative de ce qui allait suivre	(a): c'était pour faire fonctionner donc j'attendais à ce que ça marche	[I] : cet appareil
B : je regardais un peu l'équipement, j'étais un peu à côté de la plaque	(2) : j'ai donné un ordre	(b) : j'ai dit bah on va voir	[II] : qu'est-ce que c'est que ce petit chien qui m'a suivi
C : je me suis dit je vais continuer à voir quelles instructions j'ai à ma disposition et je vais essayer	(3) : j'ai vu que ça marche et je continue mes instructions – expectative	(c): j'ai dit c'est normal parce que c'est fait pour	[III] : la présence de ce truc qui se ballade, qui n'est pas humain, qui n'est même pas animal, un objet étranger qui bouge
D : je me disais je vais pas le suivre tout le temps, il va où il veut, je sais qu'il revient vers moi quand je lui parle	(4) : je recommence	(d) : je me suis dit que ma foi que lui donner un ordre c'était guère plus simple que d'appuyer sur un bouton	[IV] : c'est un truc que je mettrais au placard
E : je me disais que j'avais épuisé les possibilités de l'expérience – j'arrive au bout du temps	(5) : je teste pour voir ce qu'il fait	(e): qu'est-ce que je vais lui demander encore (e') : à quoi je peux l'utiliser maintenant	
F : je commençais à trouver le temps long	(6) : c'est parce que j'avais pas réussi au premier coup	(f) : je me disais c'est pas rapide hein	
G : là je commençais à me désintéresser du robot – je commençais à l'ignorer	(7) : je voulais vous montrer qu'il ne fonctionnait pas	(g) : je dis mais alors là je m'en vais et il me suit	
H : c'est fini pour moi, attendre que la porte s'ouvre – le robot j'ai plus envie de l'utiliser moi	(8) : ça me paraissait normal que ça marche (8') : je dis sans me soucier d'autres choses, persuadé qu'il va exécuter	(h) : j'ai dû me dire pouf voyons voir si ça marche deux coups de suite	
I : tant qu'à faire je visitais tout		(i) : là je savais que ça allait réussir	
		(j) : il m'a parlé, je sais pas ce qu'il m'a dit	

## Annexes

		(j') : j'entendais des sons mais j'entendais pas ce qu'il me disait, mais ça m'arrive souvent aussi avec des personnes, la situation est analogue	
		(k) : je me dis ça y est c'est fait	
		(l) : je me suis dit chais pas pourquoi tu me réponds, tu ferais mieux de te taire (l') : il me parle cet imbécile qu'est-ce qu'il a besoin qu'il me parle	
		(m) : je me suis pas rendu compte qu'il y avait une rampe lumineuse, je me suis dit c'est peut-être quelque chose qui ne marche pas	
		(n) : (sur le geste d'appui) je m'appuis là parce que j'ai plutôt le cou raide	
		(o) : il m'a dit coucou ou je sais pas quoi (o') : c'est un petit aboiement là	
		(p) : je me suis dit je vais voir quelles commandes sont à ma disposition dans cette pièce et je vais essayer	
		(q) : je me suis demandé pourquoi il bouge tant que ça	
		(r) : je dirais que moins il parle, mieux je me porte, je m'en laisserai à la longue	
		(s) : je me disais, je rate une partie de l'expérience parce que j'entends pas ce qu'il disait	
		(t) : (en prenant sac d'objet) : ah oui j'avais apporté ça mais qu'est-ce que je vais en faire je me dis	

## Annexes

---

		(u) : je sais pas où poser, je suis hésitante	
		(v) : oh bah c'est temps qu'elles rentrent parce que moi j'avais plus rien à expérimenter là-dedans	
		(w) : je me suis dit avant il répondait maintenant il répond plus, qu'est-ce qu'il se passe, y a plus de batterie, il marche plus ?	
		(x) : c'est soulagement parce que s'il répond à la commande, il n'a pas besoin de le dire	
		(y) : je me suis dit j'ai levé le nez assez bien mais une lumière elle avait pas de sens là-haut au plafond	

## Annexes

Grille AA S15

FoT_Gn	FoT_ProdSuj	FoT_IPC	FoT_PE
A : je me revois pas très rassurée, je vais être toute seule A' : malgré que j'étais craintive je n'avais pas peur	1 : j'ai un petit peu escamoté ma façon de me convenir	(a) : qu'est-ce qui va se passer, est-ce que je vais pouvoir me débrouiller (a') : je me suis dit dans ma tête comment ça allait fonctionner	[I] : le petit robot j'avais compris que c'était une aide donc pour les personnes qui avaient des problèmes de mouvance
B : j'avais quand même pris de l'assurance, j'avais compris le système donc automatiquement je m'étais euh... détendue – là je me sens déjà moins stressée – on sent qu'il y a une rétention	2 : je voulais voir si à ma voix il allait obéir à ce que j'allais dire 2' : pour voir si tout ce que je lui demandais il faisait – voir si tout fonctionnait	(b) : je me sentais un peu rassurée parce qu'elle (E1) m'avait dit tiens, vous verrez tout va bien se passer	[II] : pensé à un objet téléguidé d'enfant, comme des voitures d'enfants
C : ici je n'avais plus de stress	3 : comme si j'étais à la place d'une personne qui a besoin – bah je vais voir si ça va marcher	(c) : j'étais un petit peu troublée... peur de ne pas y arriver	[III] : le comme je jouet est parti là, là c'est comme un outil
D : j'arrive à me décontracter un peu	4 : j'étais stressée, c'était pas naturel, c'est un peu par à-coup – j'étais contrôlée	(d) : j'ai compris qu'il avait compris	[IV] : c'est pas une personne, c'est un robot mais il remplace une personne
E : là je me sentais rassurée, j'avais pas le trac du départ	5 : c'est fonctionnel et différent	(e) : qu'est-ce que je peux lui commander de faire, je réfléchissais à qu'est-ce que je peux lui demander	[V] : comme quelqu'un qui vient m'aider comme une aide ménagère – comme si je m'adressais à quelqu'un de vivant
F : bah là c'est la sécurité qui est arrivée là voyez, j'avais compris très très bien le mécanisme, j'ai été vraiment très très rassurée – il s'est créé un climat de confiance, un climat d'amitié	6 : c'était un ordre qui était donné mais c'était aussi un ordre qui était fait	(f) : c'est le fonctionnement, tout ça est électrique mais j'ai dit mais si jamais il y avait une passe, si jamais ça ne marche pas, qu'est-ce qu'on va faire, c'est un peu le stress qui s'empare	[VI] : c'est un robot je pense mais tout en étant un compagnon
G : j'étais en train de me	7 : je lui posais comme si je m'adressais à une	(g) : ce qui m'a impressionné c'est quand il répond,	[VII] : c'est pas un ami mais c'est comme si je

## Annexes

poser la question si j'étais pas reliée quelque part avec un poste ou quelque chose, je me suis dit bon l'examen est fini, je devrais voir apparaître quelqu'un	personne	faire parler un objet je dis c'est fabuleux ça, mais comment font-ils pour faire parler un robot, c'est ça qui m'intrigue	m'adressai à une connaissance, une connaissance sympathique, quelqu'un qui est venu me voir, quelqu'un qui m'a fait une visite sympathique
H : j'avais vu l'ensemble de ce qui était prioritaire, bon là j'attendais patiemment	8 : ça m'a coupée, j'ai été bluffée, je ne savais plus que dire	(h) : il remplace une personne, autant le commander comme si c'était une personne	[VIII] : c'est une deuxième personne, une deuxième personne qui m'accompagnait – c'est plus un objet c'est un compagnon
I : je dois commencer à trouver le temps long	9 : c'est dans ma nature de demander quelque chose et de demander poliment et de remercier – poser les questions raisonnablement	(i) : comme ça m'a surprise, ça m'en a coupé la parole	[IX] : il vient m'aider pour ce que je ne peux pas faire, pour moi c'est un compagnon, une personne, une aide familiale, une aide, quelqu'un qui vient à la maison – je ne peux pas me permettre de tutoyer, c'est quelqu'un qui m'aide
J : j'étais rassurée c'est presque moi qui fait le cours là	10 : elle communique là, c'est une communication	(j) : j'en étais émerveillée des lumières (j') : j'ai été émerveillée dans l'appartement, ça se ressent	[X] : je le considère vraiment comme un compagnon hein, comme une présence – un accompagnateur
K : je suis complètement décontractée, c'est comme si j'habitais l'appartement – j'téais à l'aise, j'étais très bien	11 : pousser ma curiosité un peu plus loin pour voir comme ça allait se réaliser	(k) : j'osais pas trop au départ m'asseoir	[XI] : là je suis au tutoiement, ça m'arrive parce que pour moi c'est devenu quelque chose de familier pour moi, maintenant je le connais, la personne je la connais, on s'est arrangé pour se tutoyer, on se connaît – c'est une personne que je connais depuis pas mal de temps, que j'ai adoptée
L : là je me suis libérée, parce que j'avais fait quelque chose de très difficile pour moi – je suis contente	12 : ça allait toujours plus loin mais je savais que ça fonctionnait	(l) : que le robot réponde, ah ça ça m'a bluffée, quand je l'entendais c'est comme si il me disait voilà qui est fait ou quelque chose comme ça	[XII] : mon bonhomme
	13 : c'était pour voir, pour bien me rassurer qu'il euh... qu'il avait très bien compris le parcours de l'appartement	(m) : ce son là et bien écoutez c'est rassurant et en plus de ça on n'est pas seul	

## Annexes

	14 : ça je le partageais au robot	(n) : au départ j'avais pas compris	
	15 : la conversation était plus naturelle, c'est ma personne, c'est ma façon de réagir	(o) j'ai compris qu'il y avait radio lola	
	16 : il faut qu'il bouge donc je vais lui demander	(p) : c'est surtout pour la fonctionnalité, il m'avait bien suivi partout, il m'avait suivi de partout, il faisait partout ce que je lui demandais	
	17 : ça m'a intrigué	(q) : je regarde la disposition, j'ai pensé c'est très bien aménagé	
	18 : rapport qui devait être fait c'est « enregistrer sur le cd »	(r) : c'est pour voir si j'avais tout regardé, si j'avais posé toutes les questions, si j'avais rien oublié	
	19 : je lui (AD) retransmets ce que j'ai fait – je lui remontre dans le même but et de voir que ça se réalise à mon souhait 19'' : c'est à vous (E2) que je le redis pour vous montrer ce qu'on peut faire avec le robot	(s) : j'étais absorbée surtout pour comprendre tout le questionnaire	
	20 : c'est qui m'aide et de passer un moment chaleureux pour discuter – je redis de faire mais appliqué dans un climat détendu	(t) : je suis quand même intriguée de le voir bouger, mais pourquoi gesticule-t-il ? (t') : c'est aussi de voir ce que je peux faire ou je pense quand il fait ces mouvements j'avais très bien compris qu'il n'allait pas me poser de questions	
	21 : c'est naturel, comme si j'avais fait ça toute ma vie – il est intégré avec moi dans mon appartement, je suis chez moi, il est intégré	(u) : je me suis dit il est derrière la porte, si on ouvre la porte ça va le bousculer	
	22 : c'était une intimité qui s'était créé, nous étions d'un commun accord, il était là à mon service pour m'aider et moi je trouvais tout à fait naturel et je le remerciais de son aide	(v) : curiosité de regarder de partout (v') : curiosité de voir, après il n'y est pas arrivé, c'est pas grave	

## Annexes

	23 : c'est une récompense en quelque sorte	(w) : du fait qu'il parlait, j'ai compris qu'il devait avoir quelque chose qu'il puisse s'exprimer et enregistrer ce que je lui dis pour la personne qui va décortiquer	
		(x) : je pensais que c'était terminé mais j'ai pensé qu'après y aurait un questionnaire alors je me dis après j'attends les consignes (x') : je réfléchissais le cas échéant si on allait me poser des questions comme j'allais euh... répéter tout ça, retracer tout ça étape par étape	
		(y) : je m'attendais à ce que le robot me dise « et bien c'est terminé »	
		(z) : du fait que je ne disais rien, que je ne bougeais pas et que je ne lui demandais plus rien donc le robot devait s'intriguer de voir est-ce que c'était fini ou pas (z') : je pense qu'il est en admiration et qu'il doit se poser des questions	
		(az) : alors après je me suis dit au cas où il va te poser des questions, de toute façon tu seras réaliste, c'est ce que tu vois, c'est que tu sens	
		(ab) : j'étais soulagée franchement	
		(ac) : j'étais satisfaite d'avoir redis à AD lui faire voir un petit peu – c'est une grande satisfaction et surtout j'ai apprécié énormément mon bonhomme	

Grille AA S16

FoT_Gn	FoT_ProdSuj	FoT_IPC	FoT_PE
<p>A : il fallait que je sache ce que j'avais à faire, j'étais pas étonnée pour autant, j'attendais</p> <p>A' : je me suis pas posée de questions parce que je sais que les robots on pouvait donc parler avec lui, mais j'avais pas tellement de choses à dire – j'attendais puisqu'il fallait que je lise</p>	<p>1 : c'était une découverte donc j'attendais ce qui allait se produire</p>	<p>(a) : je pensais qu'il fallait que je le manipule pour qu'il me suive et tout ça, j'étais pas inquiète</p>	<p>[I] : j'avais acheté un aspirateur qui circule tout seul pour m'aider, donc j'avais un peu l'habitude de ce type de chose qui se promenait – c'était pas quelque chose de nouveau</p>
<p>B : j'étais étonnée de voir tout parce que j'avais encore pas vu</p>	<p>2 : là je sais que ce que je vais dire on va le faire, je sais qu'il faut que je lise</p> <p>2' : je suis studieuse, j'apprends bien ce qu'on me dit, ici j'étais trop occupée à lire ce qu'il y avait sur le papier</p>	<p>(b) : j'attendais que le store remonte</p>	<p>[II] : je pense que si vous avez un robot qui vous parle, vous répondez si vous avez personne, s'il avait discuté vous discuteriez avec le robot</p>
<p>C : là j'étais dans les mots croisés</p>	<p>3 : là je suis tellement occupée par ce que je vais lire que je vais voir que je ne parle pas avec, au robot, de ma discussion que je pourrais avoir avec lui</p>	<p>(c) : je m'apprêtais à lire le suivant, ce qu'il fallait que je lise</p>	<p>[III] : je trouvais que c'était mignon ce petit truc qui me parlait – ça aurait été un petit j'aurai fait pareil</p>
<p>D : je trouve que c'est très bien, que tout se fait autour, c'est très agréable</p>	<p>4 : merci c'est parce que je le fais systématiquement, moi je vois une femme de ménage, je suis toujours en train de lui dire merci</p>	<p>(d) : là je vais à côté dans la chambre où je me dis que je ne pourrais jamais me coucher, il est trop bas le lit</p>	<p>[IV] : quand vous discutez avec un chat ou un chien, il vous répond pas mais il vient là, ça fait un peu cet effet là</p>
<p>E : j'ai pensé que c'était fini, que c'était la fin</p>	<p>5 : là je l'ai fait parce que je lui ai fait baisser, elle était trop forte</p> <p>5' : là je voulais qu'on puisse ouvrir pour qu'il fasse jour</p>	<p>(e) ça m'amuse qu'il réponde quand je parle</p>	<p>[V] : voyez finalement je discute comme si c'était une personne de – on discute avec le robot sans s'en rendre compte</p>
	<p>6 : là je voulais voir – si on exécute ce qu'on est en train de dire, j'attends</p>	<p>(f) : j'ai pas compris donc je peux pas répondre</p>	

## Annexes

	7 : allez la suite	(g) : j'ai pensé tout de suite à mon handicap	
	8 : je regarde, je faisais que contrôler – je faisais tout ça en voulant le contrôler	(h) : là je ne me suis pas branchée sur le robot, j'ai pas tellement fixé mon attention sur le robot	
	9 : je m'amuse de lui dire merci	(**) : et là je me suis dit que j'allais plus me relever et ravie d'avoir apporté ma pince et tout ce que j'avais	
	10 : j'ai quand même répondu au robot, parce que je n'ai pas réfléchi mais j'attendais pas une réponse – il m'a parlé, je réponds 10' : j'arrive quand même à discuter, mais ça je ne m'en rendais pas compte 10'' : je réponds quand même – au fond ce petit truc ça vous oblige à parler	(i) : j'étais en train de peser que celui qui avait soif et tout ça... mais je ne me vois pas faire ça dans un appartement que je connais pas	
	11 : c'était pas à lui, c'était une réflexion que je faisais	(j) : là je veux pouvoir me lever, c'est la première des choses – je suis délivrée, j'étais prisonnière de mon fauteuil	
	12 : c'est pour montrer à AD – je voulais lui montrer tout ce qui se passait 12' : je vous montrais et vous expliquais (à E2) 12'' : là je suis en train de faire des démonstrations	(k) : je pense que j'avais pas dit entièrement la phrase qu'il fallait donc ça n'avait pas suivi mon instruction mon indication	
	13 : je répète comme s'il avait pas entendu 13' : je répète je pensais qu'il n'avait pas		

## Annexes

---

	compris		
	14 : je lui dis de faire mais on l'a pas fait		
	15 : ça voulait dire que ça allait bien, que c'est ce qu'il fallait faire		
	16 : je fais attention à ce que je dis		
	17 : je l'ai mis en route mais je sais pas si ça marchait parce qu'on est pas dans la pièce		

Grille AA – S17

FoT_Gn	FoT_ProdSuj	FoT_IPC	FoT_PE
A : je trouve tout ça complètement absurde, pour moi c'est impossible de vivre comme ça aujourd'hui – c'est tellement bête, idiot que vous voulez agir de la sorte aujourd'hui	1 : absurde, moi mettre la lumière j'appuie sur un bouton un point c'est tout/ c'est même pas la peine, y a un bouton c'est tout	(a) : comment on met la lumière ?	[I] : moi un truc comme ça c'est inadmissible
B : pour être en sécurité parce qu'aujourd'hui il y a tellement de choses absurdes et puis de mauvaises gens	2 : ce qui m'importait c'était de l'arrêter	(b) : alors le bruit là c'est autre chose, comme j'ai des acouphènes alors bon	[II] : cet espèce d'oiseau, l'autre guignol
C : j'essaie de comprendre si on me met dans une cabane comme ça	3 : je fais ça euh justement je dirai par gentillesse – je fais des efforts	(c) : alors du coup je cherche la bouilloire	[III] : le zig, le charlot
D : pour moi c'est plus que négatif, c'est angoissant	4 : ce qui m'intéressait c'était de voir un petit peu... quand vous dites quelque chose si ça marchait	(d) : quand j'entends déjà le robot qui parle ça me suffit, par contre je me rappelais pas cette voix idiote	
E : ce qui m'intéressait c'était de voir un peu... quand vous dites quelque chose si ça marchait quoi	5 : c'est pour essayer de faire quelque chose	(e) : je me rappelais pas que j'avais tenu le coup si longtemps	
F : d'un autre côté je voulais voir où ça pouvait aller	6 : je m'amusais à vérifier si vous voyez en parlant ça s'arrêtait	(f) : j'ai recommencé je sais pas pourquoi	
G : c'est pour m'occuper, j'en ai marre	7 : nous avons toujours été habitué à ne pas tutoyer comme ça se fait aujourd'hui et de dire merci c'est la moindre des choses	(g) : j'entends de moins en moins bien	

## Annexes

---

	8 : je voulais prouver aux dames qui sont venues que j'ai fait certaine chose	(h) : j'en avais marre, ça commençait à suffire quoi	
		(i) maintenant si malheureusement je devais me contenter d'un robot un jour, c'est pas la peine de lui dire merci parce qu'il ressent pas ces choses	
		(j) : non là je comprenais rien	
		(k) : j'ai plus rien à voir	
		(l) : bah je voulais pas lui dire, lui raconter non, non moi je parle pas au robot, je vais pas lui répondre	
		(m) : je regarde si ça peut s'ouvrir et je regarde si quelqu'un arrive	
		(n) : ça m'exaspère ce truc	

Grille AA S18

FoT_Gn	FoT_ProdSuj	FoT_IPC	FoT_PE
A : ça me paraissait pas un appartement fait pour des handicapés – ça a été ma première surprise	(1) : les côtés manipulation comme ça, ça me barbe et ça m'intéresse pas	(a) : je me souviens qu'il fallait placer les objets – je pensais que c'était impossible à faire et puis après j'ai trouvé assez vite	[I] : je me souviens plus du robot pas du tout du robot
B : je me souviens pas tellement rien de particulier	(2) : j'ai testé le robot	(b) : j'étais étonnée que ça se passe de cette façon	[II] : le robot marchait c'était pas du pipeau
C : j'ai trouvé ça un peu bizarre	(3) : c'est vachement autoritaire	(c) : je me disais « mais qu'est-ce que je vais faire s'il se passe quelque chose »	[III] : on aurait dit un petit chat
D : je suis pas très intéressée par ces côtés de la vie, enfin je m'en fous un peu	(4) : c'est plus amical	(d) : j'étais pas étonnée par les cris du robot, on aurait dit un chat qui miaulait	[IV] : je change un peu de position envers le robot, la méfiance a disparu
E : j'ai un peu considéré ça comme un jeu – je m'amuse à faire ça pour faire marcher le robot – c'est rigolo de faire marcher le robot	(5) : plus neutre mais différent des premiers	(e) : je me suis rendu compte que peut-être le robot il fallait parler, j'ai réalisé ça et je me suis dit que ça avait peut-être un sens	[V] : je pense qu'il m'est devenu familier
F : je suis rentrée complètement dans le truc, c'est plus un jeu, c'est une habitude, je fais partie du truc quoi	(6) : je suis devenue plus familière et je m'amuse (6') : je suis devenue très familière avec lui quoi	(f) je pense que c'est là suite à un petit bruit qu'il a fait, je me suis dit qu'il fallait le remercier	
G : l'expérience là elle était finie quoi et ça s'était pas mal passé	(7) : j'ai réalisé sur le coup ce qu'il fallait faire	(g) : j'ai pas bien compris ce qu'il me disait	
H : je commençais à trouver un peu le temps long	(8) : c'est un peu exaspéré	(h) : c'est une bonne surprise	
I : j'avais décidé de jouer le jeu jusqu'au bout – attente	(9) : tout est rentré dans l'ordre	(i) : je voulais expérimenter toutes les lumières	
	(10) : je me suis habituée, c'est une consigne	(j) : c'était pas au point	

## Annexes

		(k) : je me demandais pourquoi ces objets, je me suis demandé si j'avais bien fait d'apporter tout ça	
	(11) : là c'est vraiment un ordre	(l) : il a fait ce que je demandais	
	(12) : ça y est c'est fini	(m) : je me dis c'est bientôt fini	
	(13) : allez ça suffit comme ça	(n) : sur le coup je me suis dit que je me suis fait avoir – ça fait intrusif ouais, je supporte pas	
	(14) : je trouve que c'est assez désinvolte		
	(15) : c'est une constatation et à la limite, j'ai plus besoin de toi		
	(16) : c'est parce que j'ai besoin (de lire)		
	(17) : c'est juste gentil		
	(18) : pour montrer à AD ce qu'il est capable de faire (18') : montrer tous les talents qu'il a		
	(19) : c'est une révision		

Grille AA S19

FoT_Gn	FoT_ProdSuj	FoT_IPC	FoT_PE
A : j'étais en équilibre, j'étais détendu, j'étais bénévole, sans préjugés, j'étais ouvert	1 : je lis le mode d'emploi, j'applique les consignes – c'est très mécanique	(a) : alors là je me dis il faut que je comprenne ce qu'il se passe, qu'est-ce qu'on attend de moi ?	[I] : un jouet, un petit truc, une petite poupée, une petite mécanique
B : j'avais déjà utilisé des logiciels de dictaphone donc c'est une petite pratique, rien d'étonnant, je rentrais dans un jeu connu – je rentre dans quelque chose que je connais	2 : c'est très mécanique il faut apprendre le vocabulaire	(b) : je lis toute la page et les suivantes pour avoir une idée de la tâche	[II] : ce type de technologie, de logiciel de voix de dictaphone [II] la bête – un logiciel de parole
C : les choix d'objets, je réfléchis à ça, j'étais branché sur cette démarche, c'était quelque chose d'indéfini et j'attendais de comprendre plus tard	3 : je commence à dialoguer avec lui	(c) : j'ai eu une découverte olfactive qui m'a rapporté des souvenirs	[III] : mon jouet
D : déconnecté de ma tâche puisqu'il avait disparu (...) et parce que j'étais arrivé au bout des surfaces et apparemment avec cet échantillonnage là, ça allait suffire (en parlant des commandes)	4 : là c'est une hésitation, je cherche la suite	(d) : les choses que j'avais apportées, j'avais envie de les... de les préparer	[IV] : peut-être que ça me sera utile
E : j'avais l'impression d'avoir terminé ma tâche	5 : je suis le plan d'apprentissage de système de voix	(e) : le robot avait disparu, où est-ce qu'il était parti ? Pourquoi il me suit pas ?	
F : je passe dans une phase d'incompréhension parce que je me replonge dans mon livre – quand je me mets dans le livre	6 : là c'est un petit jeu pour moi 6' : là je m'amusais beaucoup parce que je devenais magicien	(f) : et alors là qu'est-ce que je dois faire ? je continue à crier s'il me suit pas ? Je m'interroge.	

## Annexes

c'est une phase d'incertitude			
G : là je vois que j'ai des hésitations, je sais plus où j'en suis – je cherchais dans mon cerveau quelle était la logique, enfin qu'est-ce que je devrais faire pour continuer	7 : je voulais voir s'il y avait de la précision 7' : je voulais voir jusqu'où pouvait aller le dispositif	(g) : je suis tombé sur les bouquins, un sur la domotique qui m'a attiré, et c'est une démarche de découverte – je plonge dans ce livre qui me passionne plus que le robot	
H : j'étais en équilibre, j'avais ma canne, j'avais tout ce qu'il me fallait alors qu'est-ce qu'on fait maintenant ?	8 : c'est mon invention, c'était pas écrit – j'ai sorti des choses comme ça, sans même m'en apercevoir parce que j'étais dans la logique du logiciel en formation	(h) : de temps en temps je jette un regard à la bête – les borborygmes... je savais pas les interpréter – je jette un coup d'œil en me demandant ce qu'il veut	
I : elle m'a relancé, voilà donc là j'ai compris que c'était insuffisant et elle m'a relancé / elle me réaligne, elle me recolle au procédé	9 : c'était une façon de tirer un trait en changeant de page	(i) : je vois qu'il bouge et il a attiré mon attention, mais je ne comprends pas tout ça	
J : là je commence à l'observer (le robot)	10 : c'était de la répétition	(j) : c'est pour tout mettre au même endroit, j'avais peur de perdre quelque chose, je rassemble mes meubles à moi	
K : je suis dans le registre d'apprentissage du logiciel, j'étais dans un mécanisme où j'étais <i>teacher</i>	11 : je commande mon jouet	(k) : je regarde la nourriture parce que bon on trouve pas toujours, enfin il y a parfois des chambres d'hôtel qui sont bien équipées avec des frigos, mais...	
L : je suis dans cette zone de questionnement – pour voir si je pouvais commander avec ma voix les processus progressifs (domotiques)	12 : là j'ai acquiescé et si j'ai répondu c'est que j'avais compris et je lui ai obéi parce qu'il m'a demandé de faire quelque chose et c'est pour ça que j'ai dit oui	(l) : elle m'a donné une autre règle que ce que j'avais interp... comprise, donc je dis ok, je continue	
M : je suis un peu désorienté, un peu fatigué	13 : il y avait une multiplicité d'ordre pour commander, qui m'intriguait	(m) : là je vérifie qu'il est à côté de moi et j'essaie de voir ses... s'il y a une relation de sa distance par	

## Annexes

---

		rapport à moi, si elle est la même tout le temps...	
	14 : je commence à explorer le clavier (des commandes)	(n) : je trouvais ça chouette, je parle et tout s'agite autour de moi	
	15 : là j'ai pas l'air de savoir quoi faire	(o) : je regardais aussi s'il y avait des rhéostats, si c'est un peu nuancé	
	16 : là j'ai retrouvé le fil	(p) : je sais plus quoi faire	
	17 : oui là je voyais suffisamment qu'il pouvait changer donc bon ça c'était très rudimentaire après	(q) : mais après comment je pouvais faire ? je sais plus là j'étais un perdu	
		(*) : cette réponse-là, c'est la seule, elle était claire et nette – quand il disait ok j'étais content parce que j'avais l'impression que la relation était établie	

---

## Résumé vulgarisé

Comment les humains rentrent-ils en relation communicante ? Certains « bruits de bouche » porteraient l'essence de mélodies qui constitueraient des primitives : ce sont outils porteurs de socio-affects intentionnels capables de lier les personnes par une « glu socio-affective ». Ce lien serait comme un canal, qui s'il n'existe pas, ne peut pas véhiculer les éléments du langage (les mots, la voix, le regard, les expressions faciales, la posture, etc.). Dans cette étude, le robot Emox est utilisé comme un moyen d'observation, de capture et de simulation de la glu, pour instrumenter ce phénomène. Dans le rôle de majordome domotique d'un habitat intelligent, Emox est introduit auprès d'une trentaine de personnes âgées, fragilisées par un isolement relationnel. À travers les commandes vocales, les interactions, qui sont contrôlées pour le robot, mais spontanées pour les personnes, changent progressivement. Leurs variations multimodales d'expressions communicantes, et l'évolution de la glu socio-affective sont ainsi mesurées, analysées et modélisées dans un système de dialogue qui s'appuie sur la courbe de glu : le robot ne naît pas compagnon, mais il le devient au fil de ses interactions avec l'humain. À terme, ce système est dédié à évoluer sur long terme, dans notre quotidien où la communication humaine fabrique cette glu. Mais l'éthique de son développement ne peut se construire que collectivement et collaborativement.

## *Lay Summary*

*How do humans enter into a communicational relation? Some “mouth noises” carry the essence of melodies composing primitives: they are tools of intentional socio-affects able to link people thanks to a “socio-affective glue.” This bond is like a channel, and if it does not exist it does not succeed in delivering the language tools (e.g., words, voice, gaze, facial expressions, posture, etc.). In this study, the Emox robot is used as a mean to observe, to capture and to simulate the glue to instrument this phenomenon. In the role of a Smart Home automation butler, Emox addresses to frail elderly within a relational isolation. Across the vocal commands, the interactions, which are on the one hand controlled for the robot, and on the other hand spontaneous for the persons, change progressively. The multimodal variations of communication expressions as well as the socio-affective glue progression are measured, analysed and modelled within a dialogue system based on the glue's curve: the robot is not born as a companion, it becomes one over interactions. On end, the system is dedicated to advance in our every day within long-life, where the human communication builds this glue. However, the only way to construct its development's ethics resides in a joint and collaborative effort.*