



**HAL**  
open science

## Annotation d'éléments dialogiques de l'IHR

Cécile Cottier

► **To cite this version:**

Cécile Cottier. Annotation d'éléments dialogiques de l'IHR. Sciences de l'Homme et Société. 2015.  
dumas-01223974

**HAL Id: dumas-01223974**

**<https://dumas.ccsd.cnrs.fr/dumas-01223974>**

Submitted on 3 Nov 2015

**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.



# Annotation d'éléments dialogiques de l'IHR

**Nom : COTTIER**  
**Prénom : Cécile**

UFR LLASIC

---

Mémoire de master 2 professionnel – Sciences du langage

Spécialité : Industries de la langue

Sous la direction de Véronique AUBERGE et Solange ROSSATO

Année universitaire 2014-2015



## Remerciements

Je voudrais tout d'abord remercier mes tuteurs de stage, Véronique AUBERGE et Solange ROSSATO, ainsi que les membres de mon jury Georges ANTONIADIS et Frédéric AMAN pour m'avoir accompagnée tout au long de ce master et durant ce stage.

Je souhaiterai également remercier mes collègues de travail Liliya TSVETANOVA, Yuko Sasa, Frédéric AMAN, Romain MAGNANI, Marie-Claire KANKU, Leslie Guillaume, Jander NASCIMENTO, Maxence Girard-Rivier, Jean-Philippe GUILBAUD, qui ont fait de chaque journée un épanouissement professionnel et personnel.

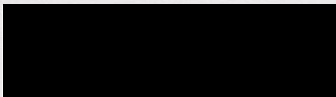
Un grand merci à tous mes camarades de master pour la robustesse de ces deux années d'apprentissages et l'élégance de cette amitié, et plus particulièrement à Noémie LAGIER pour avoir été un binôme fantastique et complémentaire.

Enfin je remercie ma famille et mes amis auvergnats pour leur soutien sans faille et leur aide de tous les jours.

**DECLARATION**

1. Ce travail est le fruit d'un travail personnel et constitue un document original.
2. Je sais que prétendre être l'auteur d'un travail écrit par une autre personne est une pratique sévèrement sanctionnée par la loi.
3. Personne d'autre que moi n'a le droit de faire valoir ce travail, en totalité ou en partie, comme le sien.
4. Les propos repris mot à mot à d'autres auteurs figurent entre guillemets (citations).
5. Les écrits sur lesquels je m'appuie dans ce mémoire sont systématiquement référencés selon un système de renvoi bibliographique clair et précis.

NOM : COTTIER ..... PRENOM : Cécile .....

DATE : 14-09-15 ..... SIGNATURE : 

# Sommaire

<b>PARTIE 1 CADRE DE TRAVAIL : INTERABOT</b> .....	<b>5</b>
CHAPITRE 1 – LE PROJET INTERABOT .....	6
CHAPITRE 2 – LES CORPUS SPONTANES .....	7
<i>EEE (Elderly Emoz Expressions)</i> .....	8
<i>GEE (Gestural Emoz Expressions)</i> .....	9
<b>PARTIE 2 BUT DE MON PROJET</b> .....	<b>10</b>
CHAPITRE 4 – MA PARTICIPATION .....	11
CHAPITRE 5 – TRAVAIL COLLABORATIF .....	11
CHAPITRE 6 – MENTIONS.....	11
<b>PARTIE 3 MA PARTICIPATION AUX DIFFERENTES ETAPES D’INTERABOT</b> .....	<b>13</b>
CHAPITRE 7 – OUTIL DE VISUALISATION .....	14
<i>Première version</i> .....	14
<i>Version actuelle</i> .....	15
CHAPITRE 8 – ALGORITHME DE PONDERATION.....	17
CHAPITRE 9 – TRAITEMENT DE DONNEES DU CORPUS DE PAROLE SPONTANEE .....	18
<i>Données lexicales et morphosyntaxiques</i> .....	20
<i>Données paralinguistiques</i> .....	20
<i>Auto-annotations des expériences du corpus GEE</i> .....	24
<i>Expérimentations pour le recueil du corpus de parole spontanée</i> .....	26

## **Introduction**

De nos jours, avec le vieillissement de la population, le secteur de l'aide à la personne âgée s'est beaucoup développé, offrant ainsi différentes solutions aux personnes fragilisées qui ne sont plus tout à fait autonomes. Les maisons médicalisées sont généralement vues comme un dernier recours et l'on préfère faire appel à des aides à domicile ou des auxiliaires de vie pour permettre à la personne âgée de rester dans son logement aussi longtemps que possible. Cette solution amène souvent les personnes âgées à l'état d'isolement social et affectif et provoque une fragilisation de leur santé physique et mentale ; on voit régulièrement cette fragilisation accompagnée d'une dégradation des compétences interactionnelles.

Palier à ces dégâts en ré-entraînant les personnes à la communication est devenu une des principales préoccupations de la Silver économie ; une telle solution serait également une aide précieuse pour les personnes atteintes d'Alzheimer ou d'autisme.

Une intuition naïve des chercheurs donne une ébauche de piste où commencer les recherches, mais c'est bien une méthodologie scientifique rigoureuse qui permettra d'établir une théorie fondée et probante.

## **Partie 1**

### **Cadre de travail : Interrobot**



## Chapitre 1 – Le projet Interabot

Le projet Interabot (Projet Investissement d'Avenir du Ministère de l'Industrie, porté par la Société de robotique Awabot, avec comme partenaires le LIG, le LIRIS à Lyon et la société Voxler) a pour objectif la création d'une nouvelle génération de robots de service, l'innovation étant les capacités interactionnelles avancées du robot avec l'utilisateur dont le but est de connaître la nature de l'attachement et en donner le contrôle à l'utilisateur. Interabot constitue un défi technologique en intelligence artificielle, en analyse et traitement de l'environnement visuel et sonore, et intègre les dernières innovations en matière d'interface homme-machine.

En donnant un rôle précis au robot Emox, celui de majordome domotique, une expérience écologique menée avec des personnes âgées isolées tente de démontrer que des expressions paralinguistiques seules peuvent permettre de co-construire un lien avec la personne fragilisée, ce lien est appelé la « glue socio-affective » [Aubergé et al, 2014]. Cette glu peut être positive ou négative à différents degrés, évolue au fil de l'interaction, et est un indicateur des caractéristiques que la personne prête au robot (compagnon, ami, simple aide).

L'équipe du LIG travaillant sur le projet a premièrement créé une voix pour Emox constituée de primitives du langage : éléments prosodiques non lexicaux, tels que des bruits de bouche et des imitations avec de la glu prosodique. Ces éléments sont porteurs d'informations pragmatiques telles que les émotions, intentions, attitudes et états mentaux d'une personne en interaction ; ils sont nommés Feeling of Thinking – FoT [Aubergé, 2012].

En faisant ces expériences, les chercheurs espèrent voir apparaître dans le discours de la personne âgée isolée des éléments saillants des changements de la glu. Ces éléments seraient ensuite rendus reconnaissables à Emox qui pourrait adapter son comportement en fonction des changements observés chez la personne âgée. Le robot, une fois accepté par la personne âgée, serait un outil pour ré-entraîner ses compétences interactionnelles et ainsi lui permettre de retrouver des rapports de qualité avec les personnes de son entourage.

Ce projet fait partie du programme d'investissements d'avenir. Les fonds de ce programme sont un soutien à la recherche et aux projets innovants, et permettent une collaboration entre l'industrie et les laboratoires de recherche autour du projet Interabot et ainsi de ne pas s'éloigner des enjeux réels de l'innovation.

Les technologies d'interaction que développe le LIG sont basées sur l'apprentissage machine ; les corpus occupent donc une place centrale dans la méthodologie de développement, en particulier la caractérisation (annotation fine) de ces corpus. Non seulement les corpus sont petits car très complexes à recueillir, mais de plus l'annotation (c'est-à-dire la modélisation de ce qui doit être étiqueté) est centrale. Les différentes étapes du projet pour la contribution du LIG sont la constitution des corpus, l'annotation et l'analyse de ceux-ci, l'utilisation des résultats des analyses dans un outil de visualisation et un algorithme de pondération, et enfin la création des briques de reconnaissance vocale et gestuelle.

Ces étapes sont réalisées en parallèle afin de créer des ébauches avec les premières données créées et voir les résultats et problèmes qui émergent. Ainsi certaines difficultés peuvent être évitées par la suite en modifiant certains points dans la manière de récolter et de traiter les données en amont.

Le recueil de corpus se fait en plusieurs phases. La première est l'expérimentation pendant laquelle les sujets sont observés et enregistrés, ce qui constitue la plus grande partie des données du corpus. Puis il y a une étape d'auto-annotation, pendant laquelle les mêmes sujets reviennent sur l'expérience pour donner leur point de vue et commentaires et ainsi enrichir les données premièrement acquises.

L'étape d'annotation, ensuite, inclut l'analyse des auto-annotations en termes de mots-clés pour annoter les données des expérimentations, puis l'annotation des corpus recueillis, soit par rapport à des critères lexicaux, morphosyntaxiques et paralinguistiques, soit suivant un classement en types de gestes – comme décrire un mouvement, désigner ou attirer l'attention.

## **Chapitre 2 – Les corpus spontanés**

Pour mener à bien ce projet un corpus est recueilli par l'équipe de la plateforme Domus, composé de parole spontanée de personnes âgées en état d'isolement social parlant à un robot-majordome. Ce corpus est constitué dans le but de déterminer quels sont les

composantes de la glue socio-affective et comment elles évoluent au cours de ce dialogue homme-machine.

Un second corpus est également recueilli auprès de personnes jeunes, guidant Emox et lui pointant des objets, afin de créer une brique de reconnaissance gestuelle pour la version finale du robot Emox.

### ***EEE (Elderly Emox Expressions)***

Le but de ce corpus de parole spontanée recueillie en situation écologie est de trouver les marqueurs discursifs qui permettent de mesurer le niveau de glu socioaffective établie entre une personne et un robot par interaction, ceci pour donner au robot une base d'apprentissage et lui permettre d'adapter ses réactions et agissements par rapport à la personne et à ce lien particulier.

Pour tenter d'identifier ces marqueurs, on demande à des personnes âgées isolées de se familiariser avec l'appartement connecté de la plateforme Domus grâce à une liste de commandes<sup>1</sup> vocales qu'ils doivent adresser à un robot majordome ; cette liste est composée de phrases à l'infinitif telles que « mettre la lumière ». Au fil de l'interaction ces personnes vont progressivement dévier de cette liste et produire des phrases telles que « tu nous mets la lumière Emox s'il te plaît ». Ces variations lexicales et morphosyntaxiques combinées aux variations observées sur la qualité de voix et les éléments paralinguistiques témoignent d'un changement d'attitude vis-à-vis du robot, conséquence de l'évolution du niveau de glu socio-affective liant la personne âgée et Emox.

L'objet étudié est donc la commande adressée au robot, car elle peut être comparée à un élément invariant au contexte ou au sujet : la liste des commandes prédéfinies. On veut caractériser les formes produites et les comparer aux formes canoniques pour tenter de repérer à quel niveau de glu l'interaction se trouve. Ainsi le robot majordome pourra prendre des décisions d'actions en fonction de ce niveau. Dans un cas de mauvaise reconnaissance d'une commande, si le niveau de glu est plutôt bas le robot aura intérêt à demander à la personne de répéter sa commande pour être sûr d'effectuer l'action demandée et avoir des retours positifs de la part de l'utilisateur ; en revanche si le niveau de glu est assez élevé il pourra effectuer l'action correspondant à la commande reconnue, même si le taux de

---

<sup>1</sup> Voir annexe n°1

correspondance à la commande de référence n'est pas très bon, car l'utilisateur sera plus enclin à accepter une possible erreur et reformulera volontiers sa demande pour aider le robot dans sa tâche.

A partir de leurs observations, les chercheurs tentent de dégager des points de rupture et des éléments saillant, marqueurs de changements d'état de cette glu.

### ***GEE (Gestural Emoz Expressions)***

Le corpus de gestes, collecté en situation écologique dans le living lab Domus, recueille des gestes spontanés de personnes ayant entre 20 et 35 ans, de niveau bac+5 en général, toutefois l'objectif reste de recruter un panel plus large.

On souhaite que ces gestes soient le plus proches possible de ce qu'il se passerait dans la vie courante, pour cela on porte l'attention des sujets sur un point tout autre que celui étudié et ont créé un contexte propice à l'émergence de l'objet que l'on recherche. Pour cette expérience, les sujets sont orientés vers un rébus qu'ils doivent résoudre en soumettant des objets au robot Emox, qui est fait juge de cette tâche, en n'utilisant que les capteurs visuels de celui-ci. Sans leur dire comment procéder, et donc sans évoquer la gestualité, les sujets vont d'eux même utiliser leur corps pour se faire comprendre du robot.

Les gestes observés sont donc qualifiables d'authentiques, et seront utilisé comme apprentissage pour le logiciel de reconnaissance gestuelle qui sera installé sur la version finale du robot d'Awabot.

## **Partie 2**

### **But de mon projet**

## **Chapitre 4 – Ma participation**

Mon projet de stage portait sur une interface d’alignement optimal pour les éléments du dialogue d’IHR. Au fil des discussions et réflexions sur le sujet, nous avons convenu qu’une interface n’était pas nécessaire : nous ferions les modifications des poids directement dans la grammaire de l’algorithme de pondération et utiliserions seulement l’outil de visualisation – développé avec une camarade plus tôt dans l’année – pour analyser le résultat des modifications.

J’ai donc contribué à l’amélioration de l’outil de visualisation, aux annotations lexicales, morphosyntaxiques et paralinguistiques du corpus de parole spontanée, ainsi qu’à l’adaptation de l’algorithme de pondération.

## **Chapitre 5 – Travail collaboratif**

L’équipe de travail dans laquelle j’étais a un fonctionnement très collaboratif. Chaque personne a son domaine de compétences et ses propres tâches à effectuer, mais comme tout le monde travaille sur le même projet, il est important d’en connaître tous les aspects.

Le travail collaboratif n’est pas très courant dans le domaine de la recherche en informatique, mais il reste très cohérent. Cela implique que chaque membre de l’équipe connaisse le but final du projet, toutes les étapes intermédiaires, le travail effectué par chacun et sa place dans le déroulement du projet.

Dans cette optique de travail collaboratif, j’ai également participé à d’autres étapes du projet Interabot avec plusieurs membres de l’équipe et ai ainsi pu le découvrir dans son ensemble et sous ses différents aspects. J’ai en effet assisté à quelques expériences du corpus EEE et participé à la plupart des séances d’auto-annotation du corpus GEE ; j’ai également effectué une partie importante des transcriptions d’auto-annotations du corpus de gestes et une partie plus légère des transcriptions d’expérimentations du corpus de parole.

## **Chapitre 6 – Mentions**

Ma contribution au projet Interabot a été observable à deux reprises, lors de présentations du projet à deux assemblées distinctes.

Au cours de la journée de présentation de GETALP où chaque équipe faisait l'exposé de ses travaux réalisés (ou en cours de réalisation) de l'année écoulée, l'équipe du CTL travaillant sur le projet Interabot a constitué un seul groupe de parole et a présenté le projet et son avancement via le travail de chacun de ses membres. Mon travail sur les annotations de données prosodiques notamment faisait partie de cette présentation.

Lors du salon Innorobo dédié aux innovations en robotique, ayant eu lieu les 1, 2 et 3 juillet 2015 à Lyon, le corpus GEE a été présenté au workshop ARSO (Advanced Robotics and its Social Impacts) lors d'une session sur la robotique interactive et l'interaction humain-robot. Pour cette occasion un article a été rédigé [GUILLAUME L. et al, 2015] par toute l'équipe ayant collaboré à la construction de ce corpus. Ayant effectué avec Leslie Guillaume, la personne chargée de créer le corpus de gestes, la plupart des séances d'auto-annotation des sujets de son expérience, je suis co-auteur de cet article.

## **Partie 3**

### **Ma participation aux différentes étapes d'Interabot**



## Chapitre 7 – Outil de visualisation

### *Première version*

Durant cette deuxième année de master, il a fallu réaliser un projet professionnel. Cela se caractérisait par l'élaboration d'un cahier des charges pour répondre à une demande d'une entreprise ou d'une entité de l'université, et la mise en place de la solution imaginée. Mon binôme, Noémie LAGIER, et moi-même avons répondu à un besoin du projet Interabot en créant un outil de visualisation pour les éléments dialogiques de l'IHR.

Pour créer cet outil nous avons dû étudier les éléments lexicaux, morphosyntaxiques, prosodiques et paralinguistiques du corpus de parole spontanée pour trouver le meilleur moyen de les représenter graphiquement. Il est apparu qu'une étude beaucoup plus poussée devait être effectuée sur les éléments prosodiques du discours afin de les caractériser et de passer à l'écrit ces éléments oraux ; il en était de même pour les éléments paralinguistiques, tels que les bruits de bouches ou la présence de pauses dans le discours. Nous nous sommes donc concentrées sur les éléments lexicaux et morphosyntaxiques à notre disposition.

Nous avons établi une première ébauche de règles de pondération de ces objets. Les différents poids attribués aux différents éléments de chaque énoncé donnaient un poids global à la phrase, et ce dernier serait représenté par un point sur un graphique.

Nous avons ensuite créé l'outil de visualisation en PHP et avec la librairie pChart, s'alimentant d'une base de données contenant les commandes de référence, les énoncés produits par les sujets, les différents éléments lexicaux et morphosyntaxiques ainsi que leur poids.

En choisissant un sujet dans une liste déroulante, un premier graphique se générait et affichait les commandes énoncées des personnes âgées sous forme de points, dans l'ordre de production. Plus le point s'éloignait de l'axe des abscisses, plus le poids de la commande énoncée était élevé et donc plus cette commande variait de la commande de référence. Ce graphique permettait également de voir le time code de la commande produite, son numéro – pour pouvoir s'y référer plus facilement en changeant de graphique ou en observant la base de données, l'information indiquant si le sujet se trouvait seul sur la plateforme ou en présence d'un expérimentateur ou de son aide à domicile, et enfin, par survol de la commande énoncée, sa commande de référence.



Figure 1 : Aperçu du graphique général des commandes d'un sujet généré par la première version de l'outil de visualisation

Un second graphique était accessible par un clic sur un énoncé. On pouvait alors voir en détail la pondération de chaque élément de la phrase et si cet élément était considéré comme lexical ou morphosyntaxique.

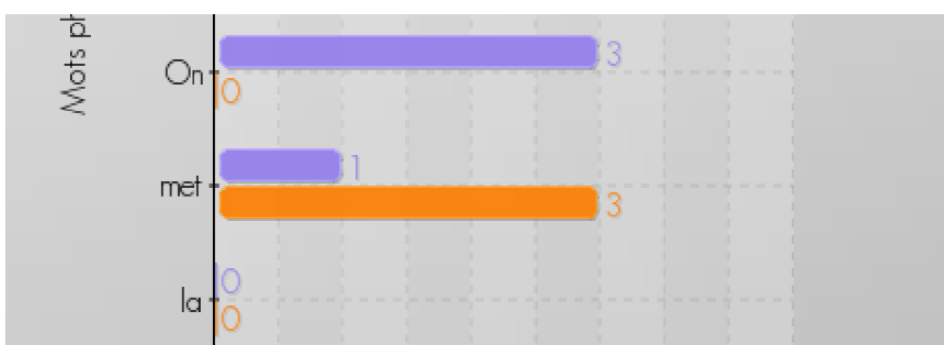


Figure 2 : Aperçu du graphique particulier à une commande d'un sujet généré par la première version de l'outil de visualisation

Plusieurs mois après la livraison, l'état des données du corpus ayant avancé, il a fallu adapter cet outil. Frédéric AMAN, qui travaille sur le module de reconnaissance de la parole des voix âgées et qui a pour tâche de créer la brique de reconnaissance/génération vocale de l'interaction finale avec le robot, a pris connaissance de notre outil de visualisation et l'a adapté au format des données qu'il utilise. J'ai apporté quelques modifications graphiques d'ordre plutôt ergonomique.

### ***Version actuelle***

Après une quinzaine d'heures de travail une seconde version<sup>2</sup> de l'outil a vu le jour. Cette version de la visualisation fait état de tout le discours prononcé par le sujet observé

<sup>2</sup> Voir annexe n°2

(une couleur permet de discriminer les commandes du sujet, les nouvelles commandes, les énoncés qui ne sont pas des commandes et le dialogue que la personne a avec l'expérimentateur) ainsi que des actions de Domus, les actions et productions vocales du robot Emox, et enfin une droite de régression représentant un niveau de glu théorique de l'interaction représentée. Le temps physique est cette fois-ci représenté par l'axe des abscisses dont l'unité est la seconde (et non plus le numéro de la commande) et l'on peut observer une proximité plus ou moins grande des phrases entre-elles selon le temps écoulé entre chaque production.



Figure 3 : Aperçu du graphique général du sujet S01 généré par la version actuelle de l'outil de visualisation

Pour plus de clarté, il est possible de générer un graphique ne comportant que les commandes reformulées par les sujets, toujours disposées dans l'ordre de production avec la représentation du temps physique, et en arrière-plan la droite de régression symbolisant le niveau de glu présumé.



Figure 4 : Aperçu du graphique général des commandes reformulées par le sujet S01 généré par la version actuelle de l'outil de visualisation

On a également le graphique particulier à une commande qui permet de voir le poids de chaque élément de l'énoncé, mis en regard avec les éléments de la commande de référence. Cette fois ci en revanche, l'algorithme de pondération ne permet pas une

séparation des poids attribués aux éléments lexicaux et aux éléments morphosyntaxiques, donc la représentation graphique n'en fait pas état.

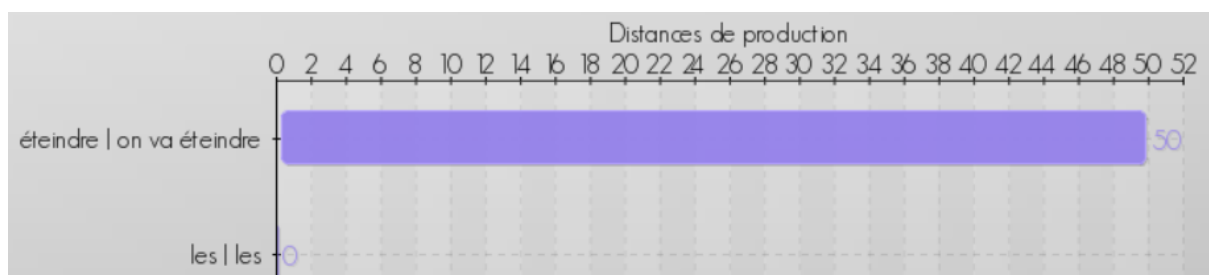


Figure 5 : Aperçu du graphique spécifique à une commande énoncée généré par la version actuelle de l'outil de visualisation

## Chapitre 8 – Algorithme de pondération

Lors de sa thèse rédigée en 1996, Nada Ghneim avait mis au point un algorithme d'alignement optimal (distance d'édition, accès tolérant) qui était utilisé dans le cadre d'une application de correction orthographique. Un utilisateur entrait un mot mal orthographié, l'algorithme comparait ce mot au dictionnaire et à l'index des mots erronés qu'on lui avait déjà soumis pour retrouver la bonne orthographe du mot donné. Cet algorithme fonctionnait avec une grammaire de règles qui apposaient un poids aux délétions, ajouts, inversions et remplacements selon le contexte de chaque lettre ; ce poids pouvait également varier selon des catégories définies et ainsi faire qu'un remplacement était estimé plus coûteux qu'un autre. Cet algorithme est très intéressant car il permet de comparer n'importe quelle entrée à un dictionnaire. L'utilisation d'une grammaire permet en effet le traitement d'une entrée sans le besoin qu'elle soit répertoriée auparavant.

J'étais chargée de refaire tourner ce programme (écrit dans une très ancienne version de Pascal sur un mac ancien) pour ensuite l'adapter à nos besoins. Inéluctablement, comme ce programme est vieux de 20 ans j'ai rencontré un certain nombre de problèmes de compatibilité. L'ordinateur devait être un macintosh assez vieux pour supporter le langage Pascal (qui n'est plus développé aujourd'hui) et tout de même récent dans une certaine mesure pour pouvoir faire le transfert du programme depuis une clé USB. Après avoir déniché et fait une série de tests sur un Macintosh Quadra 800 et un PowerBook G4, il s'est avéré que la version du programme de Nada Ghneim n'était pas complète, et donc n'était pas exécutable.

Avec ma camarade de master, Liliya TSVETANOVA – qui effectuait son stage de master de recherche sur la plateforme Domus également – nous avons tenté de trouver une autre solution que cet algorithme pour notre problématique. Liliya a tenté une approche par une grammaire ANTLR qui crée et analyse des arbres syntaxiques ou lexicaux, mais pour pouvoir attribuer des poids il fallait un vocabulaire fermé. Cette solution n’était pas envisageable dans le cadre d’une utilisation en contexte spontané. De mon côté j’ai exploré le TER (Translation Error Rate) et y ai consacré une centaine d’heures. Il était possible de faire des comparaisons mot par mot et non lettre par lettre, et d’attribuer un poids différent aux insertions, délétions, inversions et remplacements mais sans plus de détails, aucune sous-catégorisation n’était possible.

Nous avons cherché un peu partout et interrogé un certain nombre de collègues du domaine, mais cet algorithme n’avait jamais été repris ni amélioré. Il a donc fallu reprendre la version inachevée de l’algorithme en notre possession et effectuer un certain nombre de modifications : trouver la partie du code correspondant à l’algorithme dont nous avions besoin, rendre le code compatible avec Linux et FreePascal, le déboguer et le modifier pour l’utiliser avec nos données. Ce travail a été effectué par Frédéric AMAN car notre niveau en programmation ne nous le permettait pas.

## **Chapitre 9 – Traitement de données du corpus de parole spontanée**

Pour tenter de trouver les objets qui reflètent les variations du niveau de glu il faut annoter différents éléments du corpus et observer si leurs variations – combinées ou non – les révèlent pertinents.

Avant de faire ces annotations cependant, il faut sélectionner ce qui va être annoté. Cette tâche ne peut pas être confiée à un automate et doit être effectuée à la main par un annotateur expert. C’est un travail décisionnel important car parfois la frontière entre la commande et le discours de la personne n’est pas très évidente.

Prenons comme exemple la phrase « fermer les rideaux non non il a jamais voulu ». Quand le sujet prononce cette phrase, il est dans l’appartement avec l’expérimentatrice qui l’a rejoint et souhaite lui montrer comment fonctionne le petit majordome qu’il commande depuis quelques dizaines de minutes. La personne explique que la commande de fermer les rideaux dans le salon ne fonctionne pas et le démontre en prononçant la commande à

l'attention d'Emox. Ceci constitue la première partie de la phrase – « fermer les rideaux ». La seconde partie – « non non il a jamais voulu » – même si elle est prononcée dans le même souffle que les mots d'avant, est cette fois-ci à l'adresse de la personne à ses côtés. Dans ce contexte on peut donc séparer la commande à proprement parler du discours du sujet.

Quelquefois aucun des termes initiaux de la commande ne sont repris, comme dans le cas de la phrase « oui oui eh c'est pas le tout il faudrait l'allumer ». A cet instant, le sujet a demandé plusieurs fois au robot de « mettre la bouilloire » sans qu'il ne s'exécute, et commence à s'agacer. Le robot dit « ok » et la personne rétorque cette phrase. Dans le contexte, on voit que la personne s'impatiente mais persévère. La phrase de notre exemple est donc une nouvelle occurrence de « mettre la bouilloire », même si elle ne comporte aucun des termes de base.

On comprend donc bien l'importance qu'un humain expert effectue ce travail manuellement, cela demande une certaine réflexion et des capacités d'analyse du contexte.

J'ai travaillé sur cette tâche pour quatre sujets, généralement ceux pour lesquels j'avais fait la transcription de l'expérimentation car une connaissance préalable du déroulement de l'expérimentation facilite le traitement des données. Cette tâche a nécessité environ dix heures de travail, avec un nombre de commande allant de 30 à 150 selon les sujets.

Il y a donc un certain nombre de règles à respecter pour la séparation des commandes du reste du corpus, laisser ensemble des éléments qui sont séparés par une pause mais qui font partie de la même commande, regrouper une répétition avec la commande précédente quand on constate que la personne s'est en fait reprise et qu'il s'agit d'un ordre unique, etc. Il arrive que la décision soit trop difficile à prendre, dans ce cas il m'est arrivé de consulter le chercheur en charge du corpus pour qu'ensemble on arrive à faire un choix.

Au sortir de cette partie du traitement du corpus, nous avons constaté que la plupart des sujets s'étaient servi de commandes ne figurant pas sur la liste qui leur a été fournie, comme par exemple des commandes de déplacement – « suis moi Emox » – ou d'arrêt d'une manœuvre en cours – « arrêter les stores ». Nous avons donc créé une liste de 'nouvelles commandes' qui sera intégrée aux traitements suivants.

La première annotation a été d'indiquer si la commande comportait ce que nous nommons un introducteur – dans « alors mettre la lumière », « alors » est un introducteur à

la commande – ou un élément post-commande – comme « hein » dans « on ferme les rideaux hein ». Dans les ajouts, une catégorie « contexte » a été créée dans la grammaire de l'algorithme de pondération pour attribuer un poids unique à ce type d'ajout, et non un poids en fonction du nombre de mots ajoutés avant ou après la commande.

### ***Données lexicales et morphosyntaxiques***

Les éléments que l'on pense naïvement comme ceux les plus parlants sont ceux logés dans le discours même : les variations dans le lexique et la morphosyntaxe.

Chaque commande a donc ensuite été annotée sur ces points, en précisant à chaque fois si le verbe est à l'infinitif ou conjugué, à quel temps et quelle personne, si le verbe est le même que celui de la commande de référence ou s'il a été remplacé par un synonyme, et ce pour chaque terme de la commande prononcée.

La sortie du programme de pondération ne donnant pas le détail des différentes règles appliquées pour arriver au poids final, il était inutile ici de différencier les variations lexicales des variations morphosyntaxiques.

### ***Données paralinguistiques***

On peut imaginer que les variations de la glu peuvent aussi se refléter dans la modalité même de l'expression, c'est-à-dire la voix et la manière de parler. Nous avons donc choisi d'annoter un certain nombre d'éléments prosodiques ainsi que les bruits de bouche<sup>3</sup>.

#### **Annotations prosodiques**

On ne peut annoter la prosodie d'une personne d'un simple qualificatif. Les recherches à ce sujet sont toujours en cours, et pour l'instant il s'agit d'observer trois éléments : la qualité de voix, la vitesse, et la fréquence fondamentale.

J'ai annoté seule la qualité de voix, la vitesse et les pauses, et j'ai aidé Liliya sur la fréquence fondamentale ; ceci pour 5 sujets avec une moyenne de 80 commandes par sujets.

La vitesse d'élocution d'une personne se calcule en syllabes par secondes. Pour connaître le nombre de syllabes j'ai donc écouté chaque commande (quarante heures

---

<sup>3</sup> Voir annexe n°3

d'écoute) pour la transcrire en API (Alphabet Phonétique International). Avec cette transcription j'ai pu compter le nombre de syllabes, et ensuite le comparer à la durée de la commande calculée par soustraction du time code du début de la commande à celui de la fin de la prononciation de la même commande.

commande		nb syll	Temps		durée (s)	vitesse
production	tr API		Début	Fin		
arrêter là stop	avɛtɛlastɔp	5	00:45:18.155	00:45:19.805	1,65	3,03030303
mettre la bouilloire	mɛtɔɛlabujwɔɛ	5	00:45:26.687	00:45:27.930	1,243	4,022526146
mettre la télé	mɛtɔɛlatele	5	00:46:01.145	00:46:02.305	1,16	4,310344828
plus fort la télé	plyfɔɛlatele	5	00:46:26.390	00:46:27.525	1,135	4,405286344
changer la changer d(e) chaîne	ʃɛʒɛlafɛʒɛdʃɛn	6	00:47:19.153	00:47:21.360	2,207	2,718622565
éteindre [bb] la bouilloire	ɛtɛɛdɛ [bb] labujwɔɛ	7	00:49:44.700	00:49:47.035	2,335	2,997858672

Figure 6 : Aperçu de la transcription en API du sujet S04

La qualité de voix se qualifie par 3 critères : si la voix est plutôt modale (déclarative, sans émotion particulière), breathy (mot à mot avec de l'air dans la voix, une impression de voix douce et bienveillante) ou creaky (un peu chevrotante, qui n'exprime pas de sentiments très positifs) ; si l'énonciation est plutôt tense (ferme, attaques précises) ou lax (relâchée) ; et enfin par la hauteur de celle-ci et la courbe grossière de son évolution.

Ces critères ne sont pas booléens, il a fallu parfois nuancer l'annotation d'une commande clairement breathy mais moins breathy que d'autres, de même pour la tension de la voix. Concernant la hauteur de la voix, j'ai séparé ce critère en 3 segments pour les cas où la commande comportait un introducteur ou un élément post-commande. Ces annotations ont nécessité quarante-cinq heures d'écoute.

production	nb	qualité de voix
mettre la lumière	0	breathy-_tense_int0_F0Desc_postcmd0
éteindre la lumière	0	breathy-_tense_int0_F0AscH_postcmd0
[euh] beh on va mettre la bouilloire	1	breathy-_tense_intB_F0Asc_postcmd0
éteindre la lumière vite la bouilloire pardon	0	breathy-_tense_int0_F0AscH_postcmdB
on descendre les stores	0	breathy-_tense_int0_F0AscH_postcmd0
arrêter les st arrêtes les stores	0	breathy-_tense_int0_F0AscDescMoyen_postcmd0

Figure 7 : Aperçu de la base de données montrant la qualité de voix

La notation « int0 » indique qu'il n'y a pas d'introducteur à la commande, la notation F0AscH signifie que la voix est ascendante et qu'elle termine haut, la notation « postcmdB » décrit qu'il y a un élément après la commande et que cet élément est prononcé avec une voix plutôt basse.



Avec ces annotations de qualité de voix j'ai pu délimiter, pour chaque sujet, des périodes de comportement. Par exemple le sujet S03 débute l'expérimentation par des commandes avec une voix modale, puis les commandes 10 à 17 sont exprimées avec une voix breathy.

### **Pauses**

Ma tâche comportait également l'annotation des pauses observées pendant la prononciation des commandes. Un arrêt dans l'énonciation d'un propos peut avoir une signification et une importance, il est donc important de le notifier. Ces pauses sont parfois silencieuses, donc logiquement nommée pauses silencieuses, mais sont parfois vocalisées, on les nomme donc bruits de bouche.

### **Bruits de bouche**

La présence de bruits de bouche peut signifier que la personne a prononcé une longue voyelle – comme par exemple « euh », qu'elle a eu une inspiration (ou expiration) assez sonore, ou encore qu'elle a émis un bruit de succion dû à un appareillage dentaire.

### **Fréquence fondamentale**

Différents calculs de la fréquence fondamentale ont été effectués sur les enregistrements audio de chaque commande de chaque sujet. Ces calculs ont été effectués avec Praat, logiciel de traitement de son.

Praat a premièrement déterminé les valeurs pour la fréquence fondamentale initiale et finale (les premier et dernier sons) et minimale et maximale (le son le plus haut et le plus bas) de chaque commande pour ensuite calculer la moyenne, le range et le delta ; toutes ces valeurs sont exprimées en Hertz et en Semitone. Seulement le logiciel a parfois eu du mal à déterminer ces valeurs de fréquence fondamentale. On se retrouvait parfois avec un F0min de moins de 100 Hz, ce qui est très rare pour une voix de femme ; la détection était en général biaisée par le bruit ambiant.

J'ai donc parcouru la liste de valeurs de chaque sujet et repris les phrases qui me semblaient avoir des valeurs erronées : j'ai écouté chaque phrase en déterminant le son le plus grave et le plus aigu ainsi que le son initial et le son final, Praat m'en a donné les valeurs que j'ai replacées dans la base de données. Avec ces valeurs correctes obtenues en trente heures d'écoute, les calculs de fréquence moyenne, delta et F0 range ont été refait.

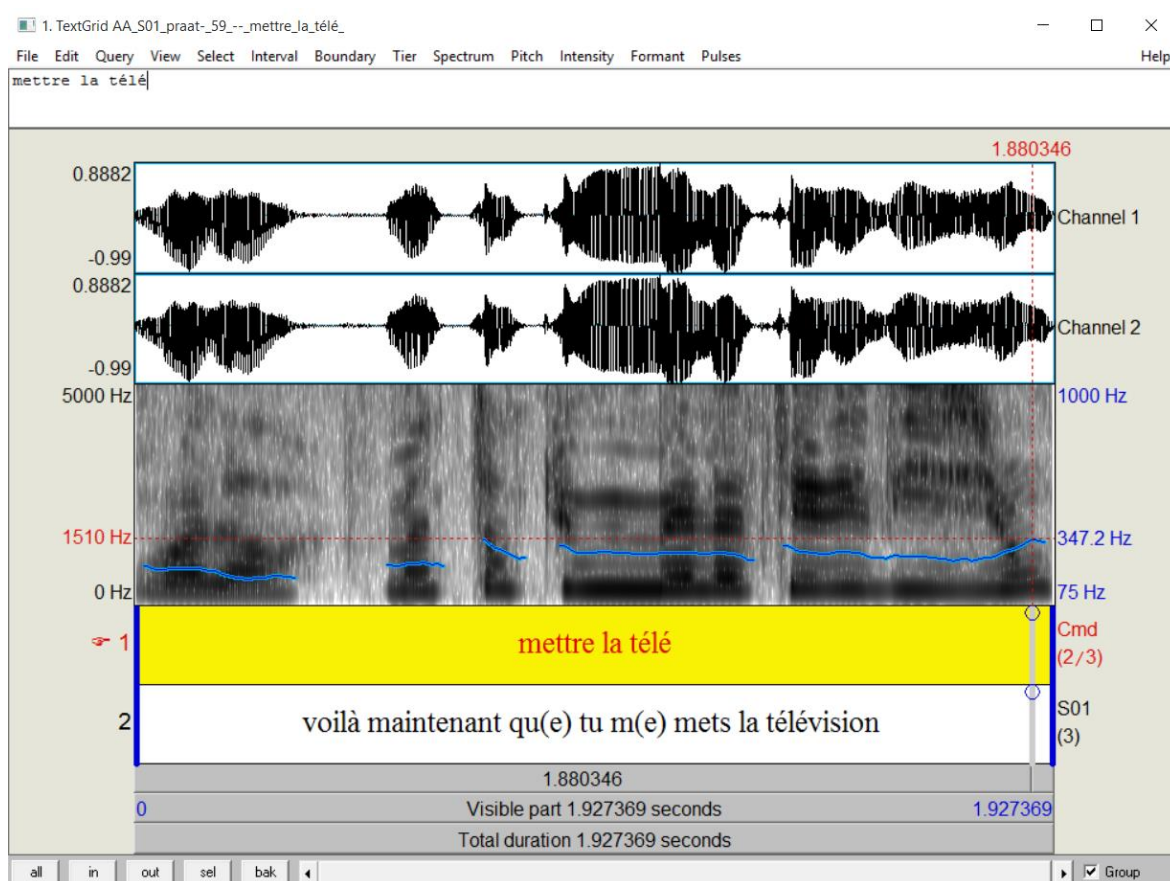


Figure 8 : Visualisation de la fréquence fondamentale en Hz de la commande 59 du sujet S01

Filename	Transcription	Segment label	F0max (semi)	F0max (Hz)	F0min (semi)	F0min (Hz)	deltaF0 (sen)	deltaF0 (Hz)
AA_S01_praat-01	alors mettre la lumière [-] s	mettre la lumière	23.52409607	389.1448822	13.83945790	222.4199988	9.684638167	166.7248834
AA_S01_praat-02	ah non non y a trop d(e) lun	éteindre la lumière	19.55837962	309.4840306	10.43128085	182.6745348	13.70096306	169.2226294
AA_S01_praat-03	on met la lumière	mettre la bouilloire	22.16631348	359.7992320	11.28829281	191.9447404	10.87802066	167.8544916
AA_S01_praat-04	alors on éteint la radio	éteindre la bouilloire	21.64121731	349.0502575	6.433818897	145.0076012	15.20739841	204.0426563
AA_S01_praat-05	on va mettre de l'eau dans l	éteindre la bouilloire	21.36063255	343.4387227	9.831155835	176.4512681	24.27530331	261.6475077
AA_S01_praat-06	on baisse les volets	descendre les stores	24.77536847	418.3219140	11.62844640	195.7533776	13.14692206	222.5685365
AA_S01_praat-07	voilà	descendre les stores	23.60127578	390.8921155	12.57975117	206.8108912	11.02152466	184.0812240
AA_S01_praat-08	et les lampes les lampes du	arrêter les stores	24.86993322	420.6132900	13.12484530	213.4261685	11.74508791	207.1871214
AA_S01_praat-09	tu veux éteindre la télé s'il t	arrêter les stores	23.97758924	399.4825365	9.522084445	173.3225105	14.45550475	226.1600260

Figure 9 : Aperçu du tableau récapitulatif des valeurs de fréquence fondamentale pour les commandes du sujet S01

Ces annotations prosodiques, la délimitation de séquences pour la qualité de voix et les calculs sur la fréquence fondamentale ont été analysés par Liliya TSVETANOVA lors de ses travaux pour son mémoire de recherche. Elle a pu trouver une certaine corrélation entre la qualité de voix et l'auto-annotation des sujets. Par exemple quand les personnes déclarent donner un ordre avec douceur ou adopter un ton professoral, leur voix est généralement breathy.

Pour l’instant ces différents éléments ne sont pas intégrés au calcul de la pondération par l’algorithme car ils ne sont pas visibles dans la transcription écrite. Il faudrait créer un système de notation pour les symboliser et les intégrer à l’écrit de façon à ne pas créer de mots en plus et fausser les calculs de distance, ainsi qu’établir une référence pour la comparaison.

id	commande			nb syll	Temps		durée (s)	vitesse (syllabes)	pauses silencieuses	BB	qualité de voix	Hyperarticulation
	référence	production	tr API		Début	Fin						
1	mettre la lumière	mettre la lumière	metlalymjɛv	4	00:00:12	00:00:13	1,078	3,71058	0	0	breathy_tense_int0_F0Desc_postcmd0	0
2	éteindre la lumière	éteindre la lumière	etɛ̃dlalymjɛv	5	00:00:27	00:00:28	1,191	4,19815	0	0	breathy_tense_int0_F0AscH_postcmd0	0
3	mettre la bouilloire	[euh] beh on va m [ø] beʒvametlabuj		8	00:00:36	00:00:38	1,83	4,37158	0	1	breathy_tense_intB_F0Asc_postcmd0	0
4	éteindre la bouilloire	éteindre la lumière	etɛ̃dlalymjɛvɪlab	11	00:00:59	00:01:01	2,664	4,12913	0	0	breathy_tense_int0_F0AscH_postcmdB	0
5	descendre les stores	on descendre les s	ʒdesädlestɔv	5	00:01:16	00:01:17	1,335	3,74532	0	0	breathy_tense_int0_F0AscH_postcmd0	0
6	arrêter les stores	arrêter les st arrêt	avetelestavetlestɔ	8	00:01:23	00:01:25	1,795	4,45682	0	0	breathy_tense_int0_F0AscDescMoyen_postcmd0	0
7	monter les stores	[insp] c’est monte	[insp] semötelest:	5	00:02:02	00:02:04	1,663	3,00661	0	1	breathy_tense_int0_F0AscH_postcmd0	0
8	descendre les stores	alors descendons	aldesäd3lestɔv	7	00:02:06	00:02:08	1,423	4,91918	0	0	breathy_tense_intB_F0Desc_postcmd0	0
9	arrêter les stores	arrêter les stores	avetelestɔv	5	00:02:23	00:02:24	1,051	4,75737	0	0	breathy_tense_int0_F0Desc_postcmd0	0

Figure 10 : Aperçu de la base de données reprenant les données de la prosodie d’un sujet

### *Auto-annotations des expériences du corpus GEE*

Pour compléter les données audio et visuelle du corpus, les sujets sont amenés à auto-annoter l’expérience qu’ils ont passée sur la plateforme. J’ai personnellement participé aux séances d’auto-annotations dans le cadre du corpus de gestes, tour à tour à mener l’entretien et à prendre des notes. Ces entretiens ont duré entre 25 minutes (pour une expérience d’une dizaine de minutes) et 3 heures (pour une expérience d’une heure et demi).

### **Passations**

Pour les personnes âgées il est impératif que l’expérimentateur soit aux commandes du logiciel, la plupart auraient beaucoup trop de mal à utiliser un outil informatique seules. Pour le corpus de geste où les sujets sont beaucoup plus jeunes, on aurait pu utiliser un logiciel d’annotation et les laisser seuls avec une vidéo prédécoupée. Dans ce cas il aurait fallu être très directif pour être sûr que les sujets décrivent les bonnes données de la bonne manière.

Beaucoup de problèmes se posent alors. Les données subissent un premier traitement avec un certain nombre de décisions prises par le chercheur selon des critères plus ou moins arbitraires. Le découpage des gestes par exemple peut être facilement erroné : un mouvement interprété comme un geste à l’attention du robot, inversement un geste que le chercheur n’aurait pas perçu, un seul geste perçu comme deux, etc. Ces détails sont parfois ambigus et

seul le sujet sait ce qu'il a fait et voulu dire. L'auto-annotation est mise en place justement pour compléter le corpus de précisions données par le sujet lui-même.

Ainsi les données n'augmentent pas en quantité mais en qualité, ce qui permet ensuite une utilisation qui ne souffrira pas, ou beaucoup moins, de mauvaises interprétations.

L'expérimentateur montre au sujet la vidéo de son expérience en utilisant le logiciel Elan, découpée grossièrement en séquences de façon à pouvoir arrêter la vidéo et relire ces passages plus facilement si le sujet le souhaite. Que la séance ait lieu quelques semaines ou quelques mois après l'expérience, en se regardant lui-même la mémoire autobiographique de l'être humain est activée et il devient capable de se remémorer la plupart de ses gestes, pensées et sentiments comme si l'action avait eu lieu quelques instants plus tôt.

On demande au sujet d'exprimer ce qu'il fait, pourquoi et comment il le fait, comment il se sent à ce moment-là. Pendant le débriefing les sujets ont été mis au courant du but réel de l'expérience et du sujet sur lequel portait l'étude, à savoir créer un corpus de gestes qui servira à créer un logiciel de reconnaissance gestuelle. Au début de l'auto-annotation on leur explique que l'on voudrait qu'ils nous disent tout ce à quoi ils pensent en revoyant leurs images. Certains sujets ont à cœur d'être le plus précis possible et certains autres préfèrent aller droit au but et donner une seule information pertinente par geste, ainsi les séances peuvent durer entre vingt minutes et trois heures. J'ai personnellement passé environ quinze heures en séances d'auto-annotations.

Quand les commentaires sont substantiels, il y en a une bonne partie qui n'est pas congruente et qui sera mise de côté ; en revanche on est sûr de ne pas faire de mauvaise interprétation quant à la signification du geste du sujet et on peut choisir un terme clé qu'il aura énoncé et qui sera représentatif de sa pensée [Aubergé et al., 2006].

Au contraire quand les commentaires ne sont pas nombreux, que le sujet dit des choses comme « pareil que tout à l'heure » ou qu'il ne dit simplement rien car il a déjà donné une explication qui peut s'appliquer à une geste suivant, il faut arriver à relancer son propos afin qu'il donne plus de détails. Cet exercice n'est pas aisé, il s'agit de ne pas influencer le sujet dans sa réflexion ou ses propos et donc se contenter de recycler les termes qu'il a utilisés ou d'adopter une formulation plutôt évasives – « que faites-vous à cet instant », « dans quel état d'esprit êtes-vous à ce moment-là ». Grâce à cette demande de précisions on obtient en général des informations assez intéressantes : un geste de pointage peut se révéler être un

pointage agacé et impatient. L'analyse future du corpus révèlera si l'état émotionnel du sujet peut influencer son comportement gestuel, et si oui dans quelle mesure.

### **Transcription**

La transcription de ces auto-annotations est nécessaire afin de pouvoir les exploiter. Un deuxième chercheur accompagne l'expérimentateur pour faire la transcription de l'auto-annotation pendant qu'elle se déroule et peut ainsi noter tout ce qui semble pertinent. Le fait d'être sur place et d'utiliser le logiciel Elan permet également de prendre des notes directement dans la bonne séquence et ainsi d'éliminer – ou réduire – la tâche de remplacement des commentaires par rapport à la bonne séquence (comparé à une transcription faite avec un simple outil de traitement de texte).

Cependant ici encore on souhaite encrer tout ce que le sujet dit, donc la séance est enregistrée et les notes prises par le second chercheur sont complétées a posteriori.

Certains sujets ont une élocution assez fluide et une séance d'auto-annotation assez courte, leur transcription a alors nécessité environ 1h à 1h30 de travail, d'autres sujets ont une élocution plus haché et un discours difficile à suivre et la transcription de leur auto-annotation a parfois nécessité plusieurs jours de travail. Au total les transcriptions auront exigé plus de cent-vingt heures de travail.

### ***Expérimentations pour le recueil du corpus de parole spontanée***

#### **Passation**

C'est la doctorante Yuko SASA qui mène ces expérimentations. Elle choisit des personnes âgées en état d'isolement social pour les faire participer à une expérimentation dont elle a élaboré le scénario avec Véronique AUBERGE, chercheur au CNRS, scénario qui a été approuvé par un comité d'éthique.

Les personnes âgées croient se trouver seules dans l'appartement et interagir avec un robot fonctionnel, mais en fait l'expérience est menée en Magicien d'Oz, c'est-à-dire que la personne est observée par le biais de caméras et de micros disposés partout dans l'appartement, et des techniciens – appelés magiciens – manipulent le robot et font réagir l'appartement depuis la régie.

Pour ces expériences, il faut une personne qui dirige le robot (à l'aide d'une manette directionnelle) et au moins une personne qui manipule l'interface permettant de faire réagir les différents éléments de l'appartement et qui fait les « bruits » d'Emox. Pour la majeure

partie de l'expérience c'est Yuko SASA qui manipule cette interface, mais une autre personne doit s'en charger quand elle se trouve à l'intérieur de l'appartement avec le sujet en début et en fin d'expérience.

J'ai été amenée à manipuler cette interface lors d'une expérience. Il faut être très attentif et réagir rapidement pour que le sujet ait en face de lui un robot réactif.

## Transcription

Par la suite la transcription de ces expériences doit être effectuée. Pour ce faire il y a une étape de traitement audio et vidéo : le sujet évoluant dans les trois pièces de l'appartement Domus, il faut synchroniser les vidéos de toutes les caméras ainsi que l'audio des micros plafond et du micro-cravate. Ce traitement a deux finalités : premièrement il faut pouvoir visionner toute l'expérience sans avoir à jongler entre plusieurs vidéos donc créer un montage de compréhension aisée à utiliser pendant les séances d'auto-annotation avec les personnes âgées ; et deuxièmement appliquer la transcription et l'auto-annotation à un fichier unique.

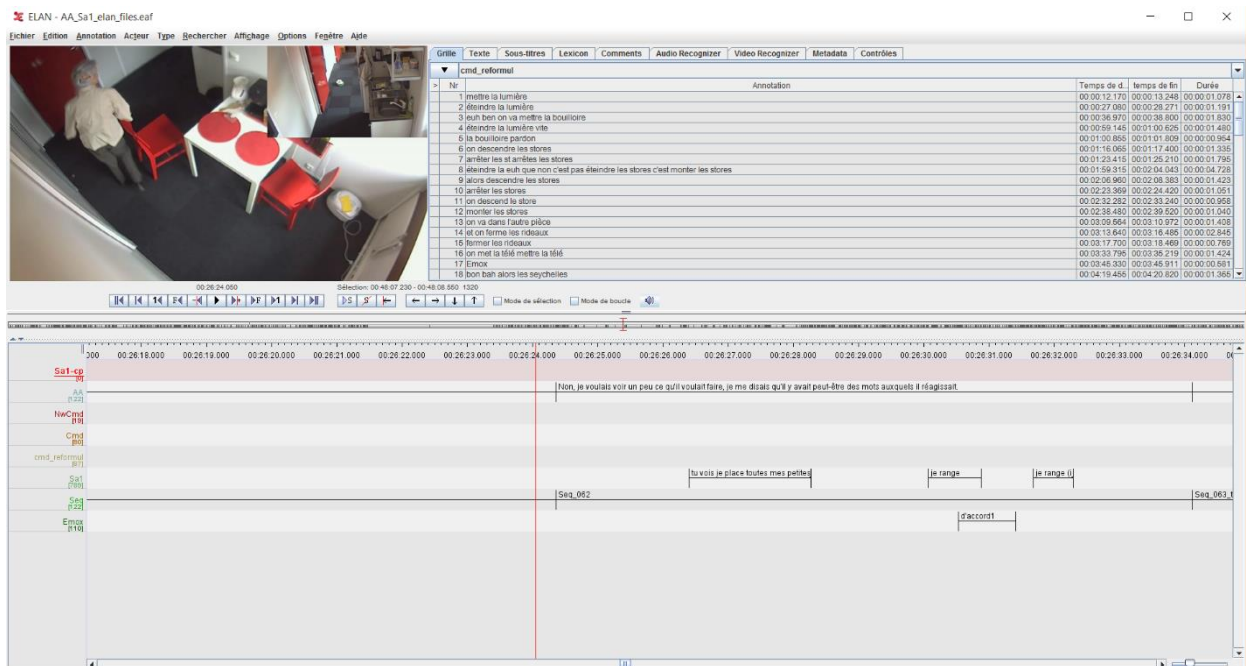


Figure 11 : Vue du logiciel Elan avec la vidéo du sujet Sa1 et les différentes transcriptions de parole (expérience et auto-annotation)<sup>4</sup>

Pour la tâche de transcription à proprement parler il s'agit d'écrire exactement tout ce que les personnes qui se trouvent autour du sujet disent pendant l'expérimentation.

<sup>4</sup> Voir annexe n°4

Depuis la phase d'accueil sur la plateforme jusqu'au débriefing, tout ce qui est prononcé, même les phrases non finies ou les choses qui paraissent incompréhensibles sont notées. Le logiciel utilisé est Elan, chaque locuteur est noté sur un tiers et il est ainsi aisé de transcrire la parole de plusieurs personnes à la fois. Ce système de tiers permet également au chercheur d'en créer de nouveaux à chaque nouvelle information qu'il souhaite annoter. Toutes les informations pertinentes concernant un sujet se trouvent sur ce fichier, et elles peuvent être facilement réutilisées grâce à la possibilité de sélection des tiers à exporter sous le format souhaité.

## Conclusion et perspectives

Ma participation aux différentes étapes du projet a contribué à son avancement.

Aujourd'hui, on sait que les recherches se dirigent dans la bonne direction car l'outil de visualisation permet de voir que la courbe des commandes spontanées retrace globalement celle du niveau attendu de la glue.

L'algorithme de pondération est aujourd'hui opérationnel et permet pour l'instant de fournir les données de l'outil de visualisation.

L'analyse effectuée sur les auto-annotations et les éléments prosodiques démontre qu'il y a des marqueurs non lexicaux de l'interaction qui relatent des variations du niveau de glu socio-affective.

Les auto-annotations des expérimentations pour la constitution du corpus de gestes spontanés ont permis une classification des gestes assez nette et aujourd'hui le corpus d'apprentissage pour la reconnaissance gestuelle est en cours de construction dans les locaux du LIRIS à Lyon. Un sujet de thèse va également porter sur l'analyse des gestes et les informations sur la relation humain-robot qu'ils peuvent véhiculer.

Le protocole d'expérimentations pour le corpus de parole spontanée va être légèrement complété pour les prochaines expériences. L'analyse du corpus montre qu'un amorçage négatif va avoir une influence sur le comportement du sujet vis-à-vis du robot. Les expériences seront maintenant menées avec un amorçage positif, négatif, ou neutre.

Une ouverture à d'autres disciplines, comme la sociologie ou la psychologie, sur l'analyse du corpus de parole est également attendue. Une étude sur les actions des expérimentateurs travaillant en régie pendant les expériences face au comportement des sujets observés peut faire émerger des informations intéressantes et pertinentes.

Ce stage m'a également beaucoup apporté en termes de formation et d'ouverture d'esprit.

J'ai été immergée dans le monde de la recherche, en TAL et en « robotique sociale », au sein d'une équipe pluridisciplinaire de travail collaboratif. En conversant avec plusieurs



personnes d'autres équipes du LIG, en participant à la première édition de PersyCup<sup>5</sup> et au salon Innorobo j'ai découvert différentes facettes de la formation, de la recherche et de l'innovation technologique en robotique.

J'ai acquis une rigueur et de la méthode dans mon travail.

Enfin j'ai été sensibilisée aux enjeux éthiques de la robotique sociale et des « robots compagnons ». J'ai ainsi pu me faire une opinion sur ce qu'il ne faut pas faire, ce qui peut être dangereux pour la société, et sur la ligne de conduite à adopter vis-à-vis de ces nouvelles technologies.

Lors de mon stage j'ai parfois regretté un manque d'encadrement et de directives claires à suivre, mais j'ai eu l'opportunité de m'appuyer sur mes collègues pour palier à ce déséquilibre. J'ai toutefois été conviée à réfléchir à certaines problématiques et j'ai pu donner mon point de vue sur des questions pratiques, ce qui m'a permis d'aller au-delà du travail d'étudiant en parcours professionnel.

---

<sup>5</sup> Challenge robotique, regroupant les élèves du master informatique de l'UJF et des amateurs. Ils développent un robot et s'affrontent pour déterminer la meilleure équipe de programmeurs.

## **Bibliographie**

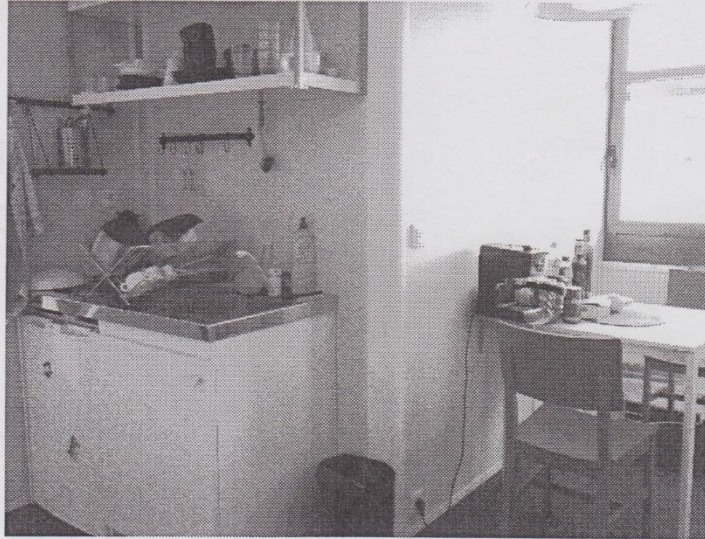
- AUBERGE V., AUDIBERT N., RILLIARD A. (2006). Auto-annotation : an alternative method to label expressive corpora. 1st International Workshop on Corpora for Research on Emotion and Affect, Genova, Italy, pp. 45-46.
- AUBERGE V. (2012). Attitude vs. Emotion: A Question of Voluntary vs. Involuntary Control. In GSCP. Belo Horizonte, Brazil.
- AUBERGE V., SASA Y., BONNEFOND N., et al. (2014). The EEE corpus: socio-affective "glue" cues in elderly-robot interactions in a Smart Home with the EmOz platform. 5th International Workshop on EMOTION, SOCIAL SIGNALS, SENTIMENT & LINKED OPEN DATA, Reykjavik, Iceland. pp.LREC.
- GUILLAUME L., AUBERGE V., MAGNANI R., et al. (2015). HRI in an Ecological Dynamic Experiment: The GEE Corpus Based Approach for the Emox Robot. ARSO 2015, IEEE International Workshop on Advanced Robotics and its Social Impacts, Lyon.

## Table des annexes

ANNEXE 1 LISTE DES COMMANDES DE REFERENCE .....	33
ANNEXE 2 VERSION ACTUELLE DE L'OUTIL DE VISUALISATION .....	36
ANNEXE 3 APERÇU DU TABLEAU DES ANNOTATIONS PROSODIQUES DES COMMANDES DU SUJET S03 .....	38
ANNEXE 4 VUE DU LOGICIEL ÉLAN AVEC LA VIDEO DU SUJET ET LES DIFFERENTES TRANSCRIPTIONS DE PAROLE (EXPERIENCE ET AUTO-ANNOTATION).....	39

**Annexe 1**  
**Liste des commandes de référence**

**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**

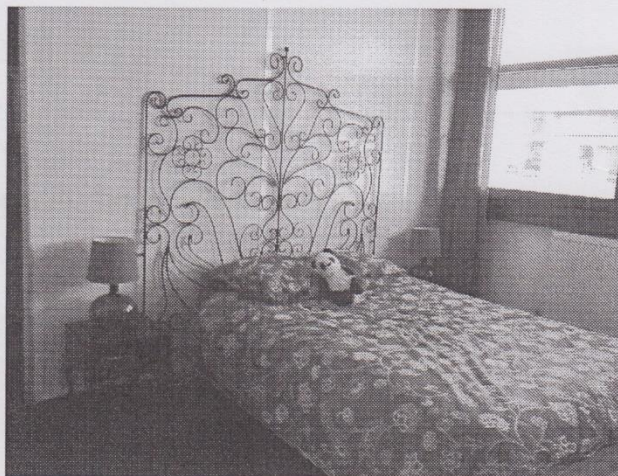
- « Mettre les lampes »
- « Eteindre les lampes »

### **la TV**

- « Mettre la télé »
- « Plus fort la télé »
- « Moins fort la télé »
- « Plus fort la radio »
- « Eteindre la télé »

### **Les lumières colorées au plafond**

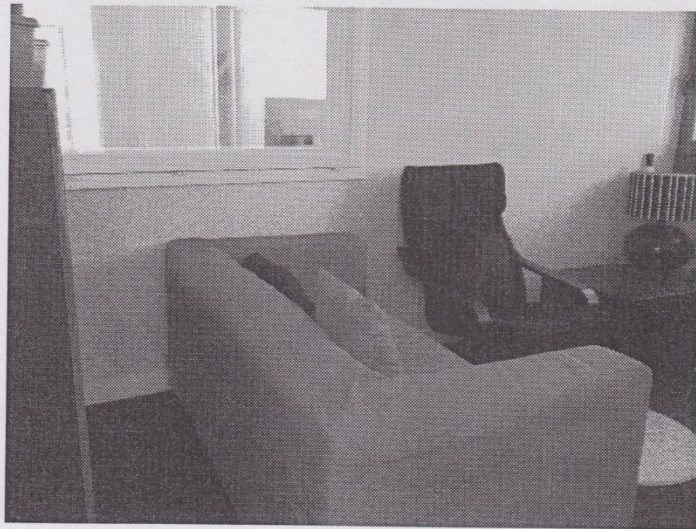
- « Mettre la lumière jaune »





- « Mettre en bleu »
- « Mettre en vert »
- « Eteindre cette lumière »

## Les commandes du salon



### **les spots au plafond**

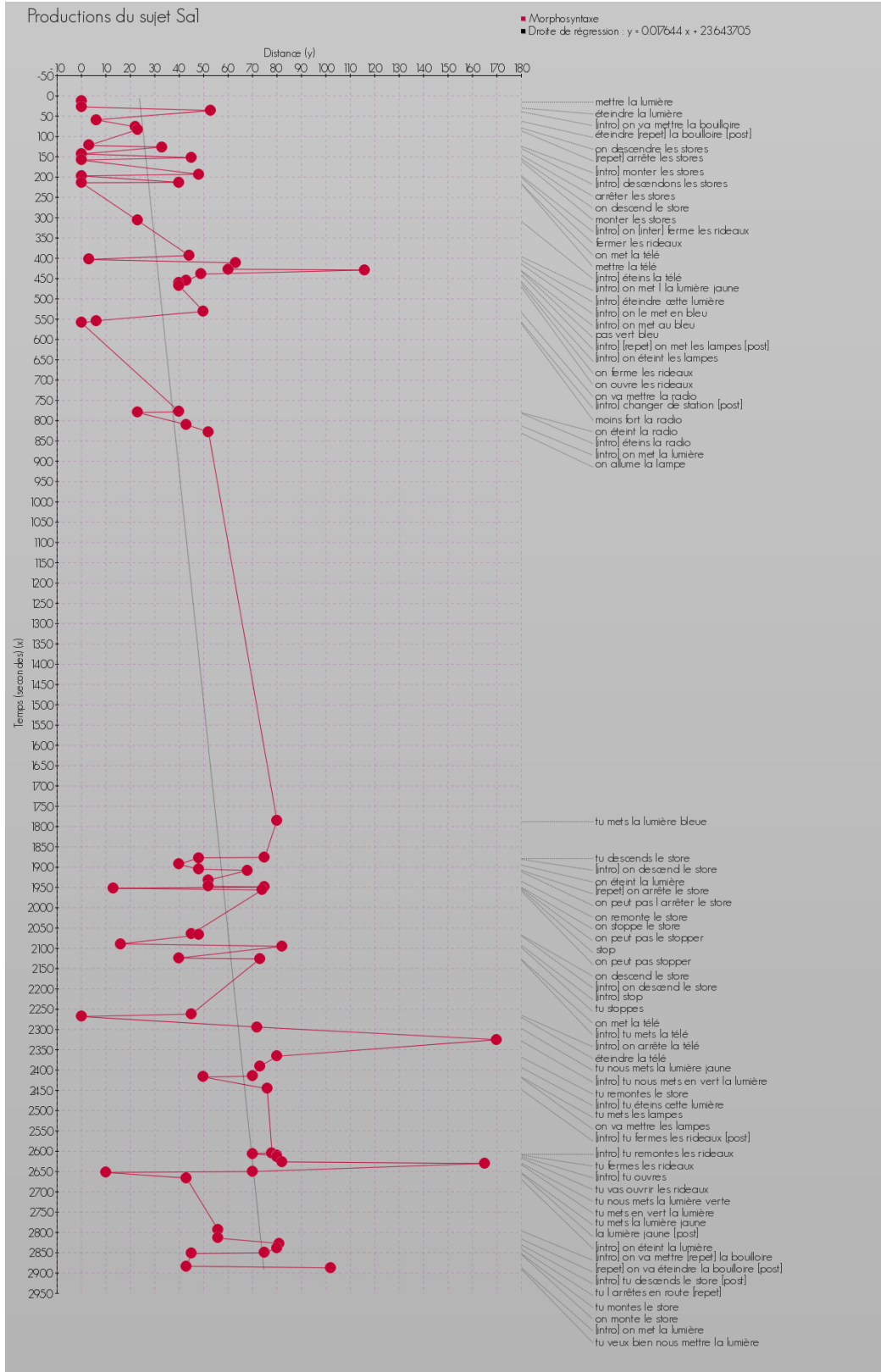
- « Mettre la lumière »
- « Eteindre la lumière »

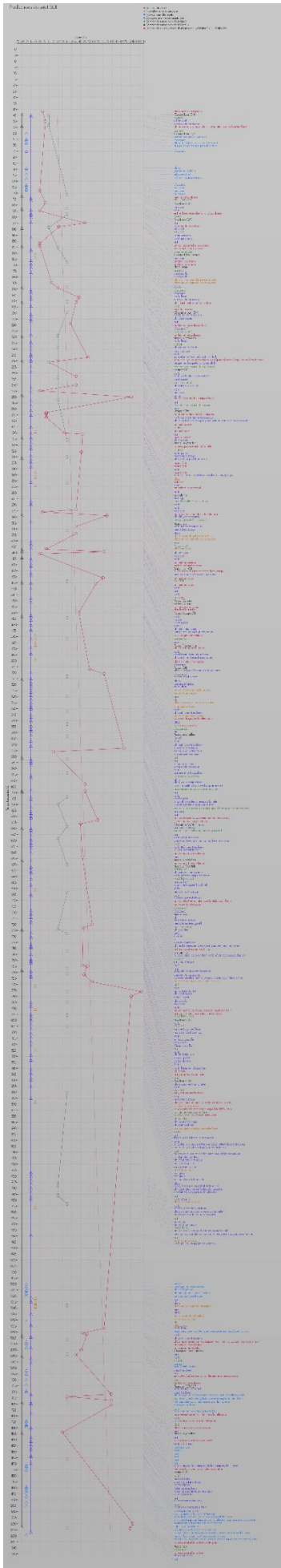
### **la radio**

- « Mettre la radio »
- « Plus fort la radio »
- « Moins fort la radio »
- « Changer de station »
- « Eteindre la radio »

## Annexe 2

### Version actuelle de l'outil de visualisation





Le format de ce graphique ne permet pas la lecture des énoncés mais permet la visualisation de la courbe dans son ensemble.  
 La droite de régression en rose correspond à la courbe attendue de la glue.  
 La courbe rouge relie les points correspondant aux commandes émises par le sujet.



### Annexe 3

## Aperçu du tableau des annotations prosodiques des commandes du sujet S03

id	commande			nb syll	Temps		durée (fin-début)	vitesse	pauses silencieus	BB	qualité de voix	Hyperarticulation
	référence	production	tr API		Début	Fin						
1	mettre la lumière	mettre la lumière	mɛʁalalyɱjɛ	5	00:03:02.483	00:03:03.587	1,104	4,528985507	0	0	modale_tense_int0_F0AscDesc_postcmd0	1
2	éteindre la lumière	éteindre la lumière	eʔɛdʁalalyɱjɛ	6	00:03:07.987	00:03:09.104	1,117	5,371530886	0	0	modale_tense_int0_F0AscDescB_postcmd0	1
3	descendre les stores	descendre les stor	desãdʁələstɔʁ	5	00:03:35.195	00:03:36.208	1,013	4,935834156	0	0	modale_tense+_int0_F0AscDesc_postcmd0	1
4	monter les stores	monter les stores	mɔ̃tələstɔʁ	4	00:03:52.402	00:03:53.285	0,883	4,530011325	0	0	modale_tense_int0_F0AscDesc_postcmd0	1
5	mettre la bouilloire	mettre la bouilloire	mɛʁalabujwɛ	5	00:04:03.896	00:04:05.000	1,104	4,528985507	0	0	modale_tense_int0_F0AscDesc_postcmd0	1
6	éteindre la bouilloire	éteindre la bouilloi	eʔɛdʁalabujwɛ	6	00:04:14.370	00:04:15.539	1,169	5,132591959	0	0	modale_tense_int0_F0AscDescB_postcmd0	1
7	mettre la lumière	mettre la lumière	mɛʁalalyɱjɛ	5	00:04:45.717	00:04:46.834	1,117	4,476275739	0	0	modale_tense_int0_F0HDesc_postcmd0	1
8	éteindre la lumière	éteindre la lumière	eʔɛdʁalalyɱjɛ	6	00:04:52.029	00:04:53.237	1,208	4,966887417	0	0	modale_tense_int0_F0DescB_postcmd0	1
9	fermer les rideaux	fermer les rideaux	fɛʁmɛʁləido	5	00:04:57.119	00:04:58.054	0,935	5,347593583	0	0	modale_tense_int0_F0AscDesc_postcmd0	1
10	ouvrir les rideaux	ouvrir les rideaux	uvʁiʁləido	5	00:05:12.915	00:05:13.941	1,026	4,873294347	0	0	breathy-_tense_int0_F0AscDesc_postcmd0	1
11	descendre les stores	descendre les stor	desãdʁələstɔʁ	5	00:05:27.448	00:05:28.500	1,052	4,752851711	0	0	breathy-_tense-_int0_F0AscDesc_postcmd0	1
12	arrêter les stores	arrêter les stores	asɛtələstɔʁ	5	00:05:46.030	00:05:47.056	1,026	4,873294347	0	0	breathy-_tense-_int0_F0AscDesc_postcmd0	1
13	monter les stores	monter les stores	mɔ̃tələstɔʁ	4	00:05:50.899	00:05:51.847	0,948	4,219409283	0	0	breathy-_tense_int0_F0AscDesc_postcmd0	1
14	mettre les lampes	mettre les lampes	mɛʁələlãp	4	00:06:16.334	00:06:17.178	0,844	4,739336493	0	0	breathy-_tense-_int0_F0HDesc_postcmd0	1
15	éteindre les lampes	éteindre les lampe	eʔɛdʁələlãp	5	00:06:25.406	00:06:26.419	1,013	4,935834156	0	0	modale_tense_int0_F0Desc_postcmd0	1
16	mettre la télé	mettre la télé	mɛʁələtele	5	00:06:35.604	00:06:36.538	0,934	5,353319058	0	0	breathy-_tense_int0_F0AscDesc_postcmd0	1
17	mettre la télé	sur la [-] télé	sɥʁla [-] tele	4	00:06:59.265	00:07:01.239	1,974	2,026342452	1	0	breathy-_tense-_int0_F0Desc_postcmd0	1
18	moins fort la télé	moins fort la télé	mɔ̃fɔʁlətele	5	00:07:31.818	00:07:32.818	1	5	0	0	modale_tense-_int0_F0Desc_postcmd0	1
19	éteindre la télé	éteindre la télé	eʔɛdʁələtele	6	00:07:38.513	00:07:39.565	1,052	5,703422053	0	0	modale_tense-_int0_F0Desc_postcmd0	1
20	mettre la lumière jaune	mettre la lumière	mɛʁalalyɱjɛ	6	00:07:47.234	00:07:48.779	1,545	3,883495146	0	0	breathy-_tense-_int0_F0DescAscDesc_postcmd0	1
21	mettre en bleu	mettre en bleu	mɛʁəblø	3	00:07:53.822	00:07:54.640	0,818	3,667481663	0	0	modale_tense_int0_F0DescMoyen_postcmd0	1
22	mettre en vert	mettre en vert	mɛʁəvɛʁ	3	00:07:59.717	00:08:00.574	0,857	3,500583431	0	0	modale_tense_int0_F0Desc_postcmd0	1
23	éteindre cette lumière	éteindre cette lumi	eʔɛdʁəsɛʔlyɱ	6	00:08:05.620	00:08:06.867	1,247	4,811547715	0	0	breathy-_tense_int0_F0Desc_postcmd0	1
24	mettre la lumière	mettre la lumière	mɛʁalalyɱjɛ	5	00:08:29.626	00:08:30.859	1,233	4,055150041	0	0	modale_tense_int0_F0Desc_postcmd0	1
25	éteindre la lumière	éteindre la lumière	eʔɛdʁalalyɱjɛ	6	00:08:41.890	00:08:43.110	1,22	4,918032787	0	0	breathy-_tense_int0_F0DescB_postcmd0	1
26	mettre la radio	mettre la radio	mɛʁələʁadjo	5	00:08:49.186	00:08:50.225	1,039	4,812319538	0	0	breathy-_tense_int0_F0HDesc_postcmd0	1
27	plus fort la radio	plus fort la radio	plyfɔʁləʁadjo	5	00:09:17.991	00:09:19.225	1,234	4,051863857	0	0	breathy-_tense_int0_F0HDesc_postcmd0	1
28	moins fort la radio	moins fort la radio	mɔ̃fɔʁləʁadjo	5	00:09:24.068	00:09:25.185	1,117	4,476275739	0	0	breathy-_tense_int0_F0Desc_postcmd0	1
29	changer de station	changer d(e) static	fãʒdestasiõ	4	00:09:31.386	00:09:32.672	1,286	3,110419907	0	0	modale_tense_int0_F0Desc_postcmd0	0
30	éteindre la radio	éteindre la radio	eʔɛdʁələʁadjo	6	00:09:42.974	00:09:44.155	1,181	5,080440305	0	0	modale_tense_int0_F0Desc_postcmd0	1
31	mettre la lumière	mettre la lumière	mɛʁalalyɱjɛ	5	00:10:01.446	00:10:02.666	1,22	4,098360656	0	0	breathy_tense-_int0_F0Desc_postcmd0	1
32	éteindre la lumière	éteindre [-] éteindi	eʔɛd [-] eʔɛdɛ	8	00:10:08.951	00:10:12.555	3,604	2,219755827	1	0	&&_laxe-int0_F0DescB_postcmd0	1
33	mettre la lumière	mettre la lumière	mɛʁalalyɱjɛ	5	00:10:44.269	00:10:45.334	1,065	4,634835681	0	0	breathy-_tense_int0_F0Desc_postcmd0	1
34	éteindre la lumière	éteindre la lumière	eʔɛdʁalalyɱjɛ	6	00:10:54.507	00:10:55.831	1,324	4,531722054	0	0	breathy-_tense-_int0_F0Desc_postcmd0	1
35	descendre les stores	descendre les stor	desãdʁələstɔʁ	5	00:11:13.556	00:11:14.802	1,246	4,012841091	0	0	breathy-_tense-_int0_F0AscDesc_postcmd0	1
36	arrêter les stores	arrêter les stores	asɛtələstɔʁ	5	00:11:40.091	00:11:41.208	1,117	4,476275739	0	0	breathy_tense_int0_F0AscHDesc_postcmd0	1
37	monter les stores	monter les stores	mɔ̃tələstɔʁ	4	00:11:47.832	00:11:48.936	1,104	3,623188406	0	0	breathy-_tense-_int0_F0AscDesc_postcmd0	1
38	mettre la bouilloire	mettre la bouilloire	mɛʁalabujwɛ	5	00:12:24.874	00:12:26.042	1,168	4,280821918	0	0	breathy_laxe-_int0_F0Desc_postcmd0	1
39	éteindre la bouilloire	éteindre la bouilloi	eʔɛdʁalabujwɛ	6	00:12:37.087	00:12:38.385	1,298	4,622496148	0	0	breathy-_tense_int0_F0Desc_postcmd0	1
40	mettre les lampes	mettre les lampes	mɛʁələlãp	4	00:13:31.801	00:13:32.684	0,883	4,530011325	0	0	breathy-_tense_int0_F0HDesc_postcmd0	1

## Annexe 4

### Vue du logiciel Elan avec la vidéo du sujet et les différentes transcriptions de parole (expérience et auto-annotation)

ELAN - AA\_Sa1\_elan\_files.eaf

Fichier Edition Annotation Acjreur Type Rechercher Affichage Options Fenêtre Aide

Grille Texte Sous-titres Lexicon Commentaires Audio Recognizer Video Recognizer Metadata Contrôles

▼ cmd\_reformul

Nr	Annotation	Temps de d.	temps de fin	Durée
1	mettre la lumière	00:00:12.170	00:00:13.248	00:00:01.078
2	éteindre la lumière	00:00:27.080	00:00:28.271	00:00:01.191
3	euh ben on va mettre la bouilloire	00:00:36.970	00:00:38.800	00:00:01.830
4	éteindre la lumière vite	00:00:59.145	00:01:00.625	00:00:01.480
5	la bouilloire pardon	00:01:00.855	00:01:01.809	00:00:00.954
6	on descendre les stores	00:01:16.065	00:01:17.400	00:00:01.335
7	l'arrêter les st arrêtes les stores	00:01:23.415	00:01:25.210	00:00:01.795
8	éteindre la euh que non c'est pas éteindre les stores c'est monter les stores	00:01:59.315	00:02:04.043	00:00:04.728
9	alors descendre les stores	00:02:06.960	00:02:08.383	00:00:01.423
10	arrêter les stores	00:02:23.369	00:02:24.420	00:00:01.051
11	on descend le store	00:02:32.282	00:02:33.240	00:00:00.958
12	monter les stores	00:02:38.480	00:02:39.520	00:00:01.040
13	on va dans l'autre pièce	00:03:09.564	00:03:10.972	00:00:01.408
14	et on ferme les rideaux	00:03:13.640	00:03:16.485	00:00:02.845
15	fermer les rideaux	00:03:17.700	00:03:18.469	00:00:00.769
16	on met la télé mettre la télé	00:03:33.795	00:03:35.219	00:00:01.424
17	Emox	00:03:45.330	00:03:45.911	00:00:00.581
18	bon bah alors les seychelles	00:04:19.455	00:04:20.820	00:00:01.365

00:26:24.050 Sélection: 00:48:07.230 - 00:48:08.550 1320

Mode de sélection Mode de boucle

300 00:26:18.000 00:26:19.000 00:26:20.000 00:26:21.000 00:26:22.000 00:26:23.000 00:26:24.000 00:26:25.000 00:26:26.000 00:26:27.000 00:26:28.000 00:26:29.000 00:26:30.000 00:26:31.000 00:26:32.000 00:26:33.000 00:26:34.000 0

Sa1-cp (1) Non, je voulais voir un peu ce qu'il voulait faire, je me disais qu'il y avait peut-être des mots auxquels il réagissait.

AA (1:22)

NwCmd (1:9)

Cmd (6:9)

cmd\_reformul (6:7)

Sa1 (7:89)

Sa1 (1:22)

Emox (1:0)

tu vois je place toutes mes petites | le range | le range (i)

Seq\_062 | d'accord1 | Seq\_063 t

# Table des matières

<b>REMERCIEMENTS.....</b>	<b>3</b>
<b>SOMMAIRE.....</b>	<b>3</b>
<b>INTRODUCTION.....</b>	<b>4</b>
PARTIE 1 CADRE DE TRAVAIL : INTERABOT.....	5
<i>Chapitre 1 – Le projet Interabot.....</i>	6
<i>Chapitre 2 – Les corpus spontanés.....</i>	7
<i>EEE (Elderly Emoz Expressions).....</i>	8
<i>GEE (Gestural Emoz Expressions).....</i>	9
PARTIE 2 BUT DE MON PROJET.....	10
<i>Chapitre 4 – Ma participation.....</i>	11
<i>Chapitre 5 – Travail collaboratif.....</i>	11
<i>Chapitre 6 – Mentions.....</i>	11
PARTIE 3 MA PARTICIPATION AUX DIFFERENTES ETAPES D’INTERABOT.....	13
<i>Chapitre 7 – Outil de visualisation.....</i>	14
<i>Première version.....</i>	14
<i>Version actuelle.....</i>	15
<i>Chapitre 8 – Algorithme de pondération.....</i>	17
<i>Chapitre 9 – Traitement de données du corpus de parole spontanée.....</i>	18
<i>Données lexicales et morphosyntaxiques.....</i>	20
<i>Données paralinguistiques.....</i>	20
<i>Annotations prosodiques.....</i>	20
<i>Pauses.....</i>	22
<i>Bruits de bouche.....</i>	22
<i>Fréquence fondamentale.....</i>	22
<i>Auto-annotations des expériences du corpus GEE.....</i>	24
<i>Passations.....</i>	24
<i>Transcription.....</i>	26
<i>Expérimentations pour le recueil du corpus de parole spontanée.....</i>	26
<i>Passation.....</i>	26
<i>Transcription.....</i>	27
<b>CONCLUSION ET PERSPECTIVES.....</b>	<b>29</b>
<b>BIBLIOGRAPHIE.....</b>	<b>31</b>
<b>TABLE DES ANNEXES.....</b>	<b>32</b>
<b>TABLE DES MATIERES.....</b>	<b>40</b>

**MOTS-CLÉS** : corpus spontané, « glu » socio-affective, transcription, auto-annotation, robot majordome

## **RÉSUMÉ**

L'industrie des robots compagnon est actuellement en expansion et la demande ne cesse d'augmenter, mais les questions éthiques sont trop souvent éludées pour répondre rapidement à cette demande. La société doit pouvoir jouir des nouvelles technologies sans qu'elle en soit mise en danger. La recherche a pour but d'étudier les usages et effets d'une technologie ayant la capacité de s'adapter à l'utilisateur et à la relation créée avec cet utilisateur. Le robot majordome co-créé par ce laboratoire et une entreprise doit pouvoir être utilisé comme divertissement mais aussi comme outil de réadaptation à la société pour les personnes isolées socio-affectivement.

**KEYWORDS** : spontaneous corpora, socio-affective « glue », transcription, auto-annotation, butler robot

## **ABSTRACT**

The companion robot industry is nowadays expanding and the demand still grows, but ethical questions are often skipped in order to be more efficient. Society must be able to enjoy new technologies without being put in danger. Searches aims to explore uses and effects of a technology capable of adaptation to the user and the relation created with him. The butler robot created by this laboratory and an industry must be used for entertainment but also as a society rehabilitation tool for socio-isolated people.