



HAL
open science

Modélisation de la coarticulation en Langue des Signes Française pour la diffusion automatique d'informations en gare ferroviaire à l'aide d'un signeur virtuel

Jérémie Segouat

► **To cite this version:**

Jérémie Segouat. Modélisation de la coarticulation en Langue des Signes Française pour la diffusion automatique d'informations en gare ferroviaire à l'aide d'un signeur virtuel. Informatique [cs]. Université Paris Sud - Paris XI, 2010. Français. NNT: . tel-00602117

HAL Id: tel-00602117

<https://theses.hal.science/tel-00602117>

Submitted on 21 Jun 2011

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.

N° d'ordre :

Université Paris Sud

Thèse

Pour obtenir le titre de

**Docteur de l'Université Paris Sud
Spécialité : Informatique**

Présentée par

Jérémie Segouat

**Modélisation de la coarticulation en Langue des Signes Française
pour la diffusion automatique d'informations en gare ferroviaire
à l'aide d'un signeur virtuel**

Soutenue publiquement le 15 décembre 2010 devant le jury composé de

Annelies Braffort	directrice
Christian Cuxac	rapporteur
Patrice Dalle	rapporteur
Julia Pelhate	examinatrice
Jean-Paul Sansonnet	examinateur

Préparée au LIMSI-CNRS

Dans le cadre de l'école doctorale d'informatique de l'université paris-sud

REMERCIEMENTS

« (...) Et encore, moi je faisais gaffe. J'étais un de ceux qui étudiait le moins. »

Coluche

A ma directrice de thèse, à l'origine de tout ce travail

Au LIMSIS et à WebSourd, pour m'avoir permis de faire cette thèse

A mes collègues du LIMSIS (du thème, du groupe, du labo) et à mes collègues de WebSourd, pour m'avoir supporté et surtout pour m'avoir supporté

A Annelies, mon couteau suisse

A mes parents et mes sœurs, mes fondations

A mes amis, qu'ils sachent que ce travail est la raison de mes indisponibilités ces dernières années

« L'homme de science le sait bien, lui, que seule la science, a pu, au fil des siècles, lui apporter l'horloge pointeuse et le parc-mètre automatique sans lesquels il n'est pas de bonheur terrestre possible. »

Desproges

0.

RESUME

Le cadre de nos recherches est la diffusion d'informations en Langue des Signes Française via un signeur virtuel, par combinaison de segments d'énoncés préenregistrés. Notre étude porte sur une proposition de modèle de coarticulation pour ce système de diffusion. Le phénomène de coarticulation est encore très peu étudié dans le domaine des langues des signes : en puisant dans différents domaines (langues vocales, gestes) nous proposons une définition de ce qu'est la coarticulation en langue des signes, et nous présentons une méthodologie d'analyse de ce phénomène, en nous focalisant sur les configurations des mains et la direction du regard. Nous détaillons les différents aspects de la création et de l'annotation de corpus, et de l'analyse de ces annotations. Des calculs statistiques quantitatifs et qualitatifs nous permettent de proposer un modèle de coarticulation, basé sur des relâchements et des tensions de configurations des mains. Nous proposons et mettons en œuvre une méthodologie d'évaluation de notre modèle. Enfin nous proposons des perspectives autour des utilisations potentielles de ce modèle pour des recherches en traitement d'image et en animation de personnages 3d s'exprimant en langue des signes française.

MOTS-CLES

Traitement automatique des langues, Annotation de corpus, Langue des signes française, Alignement d'annotations, Coarticulation, Evaluation, Animation de signeur virtuel

ABSTRACT

Our research lies in the wake of the efforts to build French Sign Language public announcement systems by virtual signers, using combined prerecorded chunks of utterances. Our study focuses on modelling coarticulation for it to be used in this system. The coarticulation phenomenon is still very little studied in the field of sign language. Using research results in different areas (vocal languages, gesture), we propose a definition of what coarticulation is in Sign Language, and we present a methodology for the analysis of this phenomenon by focusing on hand configuration and gaze direction. We detail the various aspects of corpus creation and annotation, and annotation analysis. Quantitative and qualitative statistics allows us to propose a coarticulation model, based on tensions and relaxations of hand configurations. We propose and implement a methodology for evaluating our model. Finally, we propose prospects for this model in image processing and 3d character animation in French Sign Language.

KEYWORDS

Natural language processing, Corpus annotation, French Sign Language, Annotations Alignment, Coarticulation, Evaluation, Virtual signer animation

TABLE DES MATIERES

TABLE DES MATIERES	1
CHAPITRE I. INTRODUCTION	7
I. Les langues des signes	7
II. Les théories linguistiques	10
III. Les recherches sur les LS.....	11
IV. Problématique et contexte.....	11
CHAPITRE II. LA COARTICULATION : DEFINITIONS, MODELES ET METHODES	14
I. Une définition unique ?.....	14
1. Langues vocales.....	14
2. Gestes.....	18
3. Têtes parlantes.....	22
4. Langues des signes	24
5. Aspects à retenir pour notre méthodologie.....	25
II. Modèles et mises en œuvre informatiques	26
1. Langues vocales.....	26
a. Synthèse.....	26
b. Reconnaissance.....	28
2. Gestes.....	29
3. Têtes parlantes.....	30
4. Langues des signes	31
5. Bilan.....	34
III. Méthodologies de recherche	34
1. Langues vocales.....	34
2. Gestes.....	36
3. Têtes parlantes.....	37

4.	Langues des signes	38
a.	Méthodologies	38
b.	Sélection des locuteurs	38
c.	Conception / réalisation du corpus.....	39
d.	Analyse des productions	40
e.	Evaluation	40
f.	Bilan	41
 CHAPITRE III. METHODOLOGIE : CORPUS, ANNOTATION ET ANALYSE		43
I.	Corpus	43
1.	Définition.....	43
a.	Langues vocales orales.....	44
b.	Langues vocales écrites.....	45
c.	Langues des signes.....	46
d.	Bilan	49
2.	Constitution de corpus.....	50
a.	Langues vocales orales.....	50
b.	Langues vocales écrites.....	50
c.	Langues des signes.....	51
d.	Contraintes de notre étude	52
II.	Annotation de corpus.....	52
1.	Annotations.....	52
a.	Définition	52
b.	Différents types.....	53
c.	Nos choix pour l'annotation	58
2.	Outils d'annotation	59
a.	Définition	59
b.	Nos besoins.....	59
c.	L'existant.....	60
d.	Notre choix : Anvil.....	62
3.	Synthèse des choix effectués	63
III.	Outils et méthodes d'analyses.....	64
1.	Besoins	64
2.	Outils et méthodes d'alignement.....	65
a.	Linguistique de corpus	65
b.	Bioinformatique	67

c.	La méthode retenue	71
IV.	Conclusion	72
V.	Mise en œuvre de la méthode	74
1.	Etape 1 : sélection du locuteur, conception et réalisation du corpus	74
a.	Sélection du locuteur	74
b.	Conception du corpus	75
c.	Réalisation du corpus	77
2.	Etape 2 : annotation du corpus	78
3.	Etape 3 : traitements des annotations	81
4.	Etape 4 : analyses et conception du modèle	86
5.	Synthèse	86
CHAPITRE IV.	RESULTATS	88
I.	Variations de valeurs	88
1.	Par composante et par type de variation	89
a.	Direction du regard	89
b.	Configuration de la main dominante (mD)	96
c.	Configurations de la main dominée (md)	108
2.	Par variations de segments d'énoncés	114
a.	Segments d'énoncés qui ne varient pas lors de leur mise en contexte	114
b.	Segments d'énoncés qui varient lors de leur mise en contexte	115
3.	Par catégories de configurations	121
II.	Variations temporelles	124
1.	Par énoncés complets	126
a.	Synthèse	132
2.	Par composante	133
a.	Synthèse	137
3.	Par segments d'énoncés	138
a.	Synthèse	139
III.	Synthèse générale des résultats d'analyse	139
1.	Par catégories de segments d'énoncés	139
a.	Annonce de voies (E1 à E11)	139
b.	Informations (E12 à E21)	140
c.	Motifs de perturbations (E23 à E31)	140

d.	Chiffres et heures.....	140
e.	Gares.....	141
2.	Par énoncés.....	141
3.	Par catégorie de configurations / de direction du regard.....	141
a.	Le regard.....	141
b.	Les configurations des mains.....	143
4.	Proposition d'un modèle de coarticulation dans le contexte de notre étude	145
CHAPITRE V. EVALUATION		148
I.	Méthodologies d'évaluation.....	148
1.	Têtes parlantes.....	148
2.	Agents Conversationnels Animés.....	150
3.	Gestes.....	152
4.	Langues des signes	152
II.	Notre méthodologie d'évaluation.....	155
1.	Sélection des sujets.....	158
2.	Déroulement des évaluations	159
3.	Questionnaires	161
III.	Résultats.....	161
a.	Evaluation du modèle de coarticulation	161
b.	Evaluation du SV et du système de diffusion en gare.....	164
c.	Synthèse.....	165
CHAPITRE VI. CONCLUSIONS, DISCUSSIONS ET PERSPECTIVES		167
I.	Conclusions.....	167
II.	Discussions	169
1.	Modèle de coarticulation	169
2.	Signeur Virtuel.....	169
3.	Système de diffusion en gare.....	170
III.	Perspectives.....	171
1.	Possibilités d'amélioration de nos résultats.....	171
a.	Le corpus.....	171
b.	Les annotations.....	172
c.	L'évaluation.....	172

2. Utilisations potentielles de nos résultats	173
BIBLIOGRAPHIE.....	174
ANNEXES	191
I. Énoncés et segments d'énoncés de notre corpus	191
Énoncés complets.....	191
Énoncés partiels.....	194
II. Correspondances valeurs / capture d'écrans / caractère ASCII	197
III. Variations des instances d'énoncés pour l'évaluation	206
ROYA	206
SOY	209
AS01.....	212
DOY.....	214
ILLUSTRATIONS.....	217
Table des tableaux.....	217
Table des figures.....	218
Table des graphiques	219

Chapitre I. INTRODUCTION

I. Les langues des signes

Les langues des signes (LS) sont les langues naturelles des personnes sourdes. Nous écrivons « les langues des signes » au pluriel, car contrairement à ce que pense la majorité des personnes, il n'existe pas une unique langue des signes, internationale, mais une langue des signes par pays où il y a des personnes sourdes. Cette affirmation est une approximation, puisqu'il arrive que dans un pays il y ait cohabitation entre plusieurs LS (par exemple en Afrique, au Mali, où il y a eu successivement influence de la langue des signes américaine (ASL) puis de la langue des signes française (LSF), avec pour résultat aujourd'hui une langue des signes qui emprunte aux deux, et toujours la présence des deux LS d'origine), et il y a de nombreuses variations (générationnelles, régionales, etc.) au sein d'un même pays (par exemple en France). Un des premiers témoignages à la fois de l'existence des personnes sourdes et du fait qu'ils communiquent dans une modalité visuo-gestuelle est dans « Le Cratyle » de Platon, où l'auteur relativise le fait que les sourds de son époque s'expriment avec autre chose que la voix : « Si nous n'avions point de voix, ni de langue et que nous voulussions nous montrer les choses les uns les autres, n'essaierions-nous pas, comme le font en effet les muets, de les indiquer avec les mains, la tête et le reste du corps ? ». Ainsi, il est considéré que les LS, bien que n'étant pas nommée de cette manière, ont toujours existé, du moins dès qu'il y a eu des personnes sourdes.

Sans chercher à être exhaustif (se référer à Encrevé08 [ENC08] pour plus de détails) nous souhaitons présenter rapidement l'histoire de la LSF et de la communauté sourde française pour mieux mettre en valeur les besoins actuelles autour de cette langue et de la communauté des personnes sourdes, tant au niveau sociétal que recherche.

Les personnes sourdes ont bénéficié de courtes périodes pendant lesquelles ils ont été intégrés dans la société de l'époque, mais la majeure partie du temps ils étaient considérés comme incapables de dire, de savoir, de faire.

Au 18^{ème} siècle, celui des Lumières, les personnes sourdes étaient intégrées dans la vie quotidienne, ils étaient menuisier, etc. L'abbé de l'Épée, vers 1760, suite à sa rencontre avec la langue des signes et les sourds signants (locuteurs de la LS), mit au point une méthode et en place

une école afin d'enseigner le français aux enfants sourds. A cette période, les personnes sourdes pouvaient être professeurs dans les instituts pour jeunes sourds, écrivains, et autres professions intellectuelles. Le congrès de Milan, en 1880 mit fin à cet âge d'or en préconisant l'interdiction de l'utilisation de la modalité gestuelle dans l'éducation des sourds en France (et dans la plupart des pays européens) au profit de la modalité vocale (donnant lieu à la notion d'« oralisme »). Cette préconisation n'est pas une loi, mais respecte l'avis de l'église (qui considérait que le lien avec Dieu est par la voix non par les gestes) et celui des riches (en coupant court à l'éducation gratuite mise en place par l'abbé de l'Epée), donc fait force de loi dans le système éducatif français (et européen, les américains ne suivant pas cette préconisation connaîtront un développement de l'ASL qui fait encore aujourd'hui exemple pour les autres pays du monde). Cette interdiction est levée par l'amendement Fabius de 1991¹, donnant aux parents le droit du libre choix entre une communication bilingue (LSF / français écrit) et une communication orale pour l'éducation de leurs enfants sourds. Pendant ces cent années d'interdiction de la LSF à l'école, l'enseignement portait sur l'apprentissage de la faculté de parler en français oral : les résultats sont que peu d'enfants sourds ont réussi dans cette voie, et que la très grande majorité s'avèrent aujourd'hui illettrés (selon le rapport Gillot [GIL98], le pourcentage d'illettrisme varie entre 60 et 80 % dans la population des adultes sourds). Il est important de noter que bien qu'il existe des systèmes de notation, de transcription, d'annotation, de description des langues des signes, il n'existe pas d'écriture, à l'instar du français écrit pour le français oral. Les personnes sourdes signantes n'ayant presque que la LSF pour communiquer et s'informer sont donc exclus du monde actuel où l'écrit est roi. En 2005, est votée la loi du 11 février² qui non seulement reconnaît la LSF comme une langue à part entière mais également oblige les établissements recevant du public à être accessibles entre autres aux personnes sourdes. Cette accessibilité peut être mise en place de différentes manières, notamment par l'embauche de personnes sourdes pour les accueils aux guichets, ou l'utilisation d'interprète LSF/français pour qu'une personne entendante et une personne sourde puissent communiquer. Avec les nouvelles technologies, de nouvelles possibilités de mise en accessibilité sont offertes : les interlocuteurs sourds et entendants peuvent communiquer à distance, via un interprète lui aussi à distance, des vidéos en LSF peuvent être enregistrées et diffusées à volonté sur des écrans. Une des toutes dernières avancées dans le domaine est l'utilisation de signeurs virtuels : personnages humanoïdes 3d apparaissant sur un écran (TV, ordinateur, téléphone portable, etc.) et s'exprimant en LSF (à l'instar

¹ Loi n° 91-73 du 18 janvier 1991 portant dispositions relatives à la santé publique et aux assurances sociales, titre III, article 33

² Loi n°2005-102 du 11 février 2005 pour l'égalité des droits et des chances, la participation et la citoyenneté des personnes handicapées

des agents animés qui accueillent sur le site Web d'une banque ou d'un magasin de bricolage). C'est dans ce mouvement actuel d'utilisation des nouvelles technologies de l'information et de la communication, grâce aux signeurs virtuels, que se situe notre étude (nous y revenons dans la présentation du projet dans lequel est notre recherche).

La LSF est donc une langue à part entière, au sens légal, mais aussi naturelle au sens linguistique du terme : elle n'a pas été fabriquée, et répond au concept de « double articulation » énoncé par Martinet [MAR49]. La LSF possède sa propre grammaire : une syntaxe propre à la modalité visuo gestuelle, et des structures linguistiques spécifiques. La double articulation a été mise en évidence par Stokoe [STO60], en 1960 : elle établit qu'il existe deux niveaux, dénommés, pour les langues vocales, phonème et morphème, qui sous-tendent les LS. Ainsi, pour Stokoe, et nombre de chercheurs après lui (cf. [BOU08] pour plus de détails), les morphèmes seraient les signes, et les phonèmes seraient les paramètres composant ces signes. Au début au nombre de quatre, puis cinq, ces paramètres sont : l'orientation, le mouvement, l'emplacement, la configuration de la main, et le regard. Il est maintenant communément admis que ce découpage est un peu trop simple par rapport à tout ce qui est engagé par le locuteur quand il s'exprime en LS : chaque élément de la partie supérieure de son corps (au dessus de la ceinture) est mobilisé et s'inscrit dans un espace dit de signation. Cet espace se situe face au signeur, et forme grosso modo un volume d'une profondeur, d'une largeur et d'une hauteur égale à la longueur des bras du locuteur (Figure 1).



Figure 1 : Représentation schématique de l'espace de signation [GUI04]

Chaque personne sourde signante a ce qui est communément appelé une main dominante et une main dominée. Selon Cuxac [CUX00] on appelle « main dominée » la main gauche d'un locuteur droitier, ou la main droite d'un locuteur gaucher, et « main dominante » la main droite d'un locuteur droitier ou la main gauche d'un locuteur gaucher. En ce qui concerne l'expression du lexique standard, si la réalisation d'un signe ne nécessite qu'une seule main, c'est généralement la main dominante qui est utilisée, et dans le cas des signes à deux mains, si une des deux doit faire un mouvement alors c'est la main dominante qui l'exécutera (il existe des exceptions dues au contexte

d'expression). Nous utiliserons ces dénominations, que nous abrègerons en « mD » pour la main dominante, et « md » pour la main dominée.

II. Les théories linguistiques

Actuellement la plupart des théories linguistiques sur les LS sont issues des études sur les LV, et reposent sur une approche générativiste qui suppose que toutes les langues sont descriptibles par un unique système [FIL08]. Selon Garcia [GAR10], la majorité des chercheurs qui étudient les LS ne considèrent pas comme faisant partie de la linguistique de ces LS les phénomènes d'iconicité : soit ils les excluent [MAC00], soit ils les repèrent comme posant un problème théorique sans pour pouvoir trancher sur leur nature exacte ([SLO03], [SCH03]), soit ils considèrent qu'elles résultent d'un mélange de linguistique et de gestes (par définition, non linguistiques) ([LID03], [JOH07]). Une minorité de chercheurs se basent sur la théorie de Cuxac ([CUX96], [CUX00]) qui place au cœur du système linguistique de la LS le mécanisme d'iconicité. Ceci pour deux raisons : d'une part les structures développées sont dues à la modalité visuo-gestuelle et à sa potentialité d'expression, d'autre part ces structures sont issues de l'expérience du locuteur avec son environnement et des exigences de la communication. Sans entrer plus avant dans ces détails, il nous semble important de souligner que ces structures (que nous allons détailler un peu plus par la suite) ne sont pas que des éléments de la langue, mais qu'il s'agit des outils grâce auxquels la langue vit, se développe et évolue au cours du temps.

Nous allons maintenant détailler la théorie de Cuxac [CUX00], puisque c'est dans ce cadre linguistique que se déroule notre étude. Selon Cuxac, comme cela a été dit précédemment, les signes sont issus d'une expérience de l'environnement : au départ construits sous une forme iconique (montrant visuellement ce qui est exprimé), les signes évoluent ensuite vers une forme dite « générique » perdant de son intention d'iconicité (l'intention disparaît, mais l'iconicité peut être conservée dans le signe). Ces signes génériques sont largement appelés « signes standards » (Cuxac [CUX10] préfère maintenant le terme de « signe-mot »). Ainsi, il y a deux manières de dire en LS : la langue est potentiellement exprimable suivant qu'on vise à dire sans montrer, ou dire en montrant. La première solution consiste à employer des signes-mots (faisant partie du lexique), plus ou moins iconiques, quand la seconde met en œuvre des structures dites de « grande iconicité » qui montrent à l'interlocuteur ce dont il est question dans la conversation. Le locuteur s'exprime dans son espace de signation, en alternant et mêlant les deux manières de s'exprimer.

Nous nous situons donc dans cette théorie linguistique de la LS qui inclut dans la langue la notion de grande iconicité : ainsi, au contraire d'autres études basées sur d'autres théories, nous ne

considérons pas uniquement les mains comme véhiculant le sens de la LS, mais l'ensemble des composants, ce qui implique les expressions du visage, la direction du regard, etc.

III. Les recherches sur les LS

Il y a actuellement des recherches sur les LS dans de nombreux domaines : histoire [ENC08], sociologie [MOT06], [DAL06], anthropologie (Schetritt, Schmitt, non publiés), etc. et les recherches en linguistique se sont considérablement développées. Les progrès technologiques, qui ont été bénéfiques à ces différents domaines, ont vu le jour dans le cadre du traitement automatique des LS (TALS) : ce domaine couvre les aspects de reconnaissance et de synthèse de LS, ainsi que de conception de grammaire (pour la génération de langue) et de traduction, et puise des connaissances dans les autres domaines, principalement linguistique. Plusieurs équipes de recherche travaillent sur l'un et/ou l'autre aspect, en France cela concerne l'IRIT à Toulouse (reconnaissance et modélisation de la grammaire de LSF [LEN05]), le Valoria à Vannes (synthèse de gestes et de LSF [GIB08b]) et le LIMSI à Orsay (synthèse et modélisation de la grammaire de LSF [BRA08]). Notre étude se situe donc dans le domaine du TALS, dans l'aspect synthèse de la LSF : notre but est en effet de traiter automatiquement un phénomène de la LSF, en l'occurrence la coarticulation, afin d'animer un signeur virtuel.

IV. Problématique et contexte

Notre problématique peut se subdiviser en deux questions :

- Qu'est-ce que la coarticulation en LS ?
- Comment étudier ce phénomène en LSF ?

La coarticulation est un phénomène linguistique complexe, dont les définitions convergent ainsi que les méthodologies d'analyse. Quelques études affirment que sans cette coarticulation (cf. chapitre suivant, dans les domaines des LV et des LS), la langue n'est perçue comme étant naturelle. Or notre objectif est de proposer des énoncés en LSF, pas du « français signé » (mélange de syntaxe de langue française avec des signes de LSF, qui n'est donc pas une langue).

Notre étude prend place dans un projet industriel de l'entreprise WebSourd, de diffusion d'information en LSF en direction du grand public, via un système de signeur virtuel. Ce projet est réalisé en lien avec la SNCF, dont la volonté est de produire en LSF l'équivalent des énoncés vocaux. Il ne s'agit pas de traduire les énoncés vocaux, mais bien de partir des mêmes principes : une base de données contient des morceaux d'énoncés (des groupes de mots, des mots, des chiffres ou des lettres, enregistrés vocalement par une personne humaine, de manière isolée) que le système va

mettre ensemble suivant un modèle syntaxique donné (celui du français). Dans le cas de la LSF, il s'agit bien sûr d'une base de données de signes isolés, et le système utilise les structures syntaxiques de la LSF. Notre recherche a pour but de proposer un modèle de coarticulation pour le système de LSF, à l'instar de celui qui existe pour le système vocal.

Nous pouvons déjà noter deux contraintes majeures dues au projet :

- le vocabulaire est restreint (au domaine de la SNCF),
- les règles grammaticales sont restreintes (au type d'énoncés de la SNCF)

Nous ne prétendons pas proposer un modèle de coarticulation qui soit applicable à l'ensemble de la langue, cependant notre objectif est que notre contribution améliorera nos connaissances de manière générale sur ce phénomène, et que nos résultats serviront de bases à des généralisations futures.

Nous présentons dans le chapitre suivant (chapitre 2) un état de l'art dans plusieurs domaines, afin de déterminer comment va se dérouler notre étude : nous définissons notre sujet d'étude, la coarticulation, puis nous présentons les méthodologies et outils employés pour étudier ce phénomène, ensuite nous définissons ce qui sera le terreau de nos analyses, le corpus, et nous proposons différents outils d'analyse de corpus, avant de présenter des outils statistiques issus du domaine de la bioinformatique et du TAL. Le chapitre 3 explicite la méthodologie employée : après avoir fait le tour des différents éléments à prendre en compte dans notre étude (la définition du phénomène, les outils et méthodes d'analyse, etc.) nous avons mis en place notre méthodologie d'étude du phénomène de coarticulation en LSF dans le cadre de notre projet d'animation d'un signeur virtuel. Dans le chapitre 4, nous exposons nos résultats et nos conclusions, ces dernières étant les bases de notre modèle de coarticulation. Nous évaluons le modèle dans le chapitre 5, et dressons finalement conclusions et perspectives.

Chapitre II. LA COARTICULATION :

DEFINITIONS, MODELES ET

METHODES

Dans ce chapitre nous présentons la notion de coarticulation et les différentes études qui ont été menées sur ce phénomène. Nous nous attachons dans une première partie à la définition de ce qu'est la coarticulation, dans une seconde partie nous présentons différents modèles de coarticulation existants dans plusieurs domaines, enfin nous détaillons les méthodes d'analyse et de conception de ces modèles.

I. Une définition unique ?

Nous avons cherché tout d'abord s'il existait une définition large du phénomène de coarticulation, qui pourrait s'appliquer quelque soit le domaine de recherche considéré. Nous avons trouvé des études et définitions dans le cadre de la production de gestes ou de langue naturelle (qu'elle soit vocale ou signée).

Le terme désignant ce phénomène est divisible en « co » qui vient de la préposition latine « cum », signifiant « réunion », « adjonction », et en « articulation », lui-même issu du latin « articulare », formé à partir du radical « artus » qui signifie « articulation, membre ». Ainsi, il s'agirait étymologiquement de la réalisation d'articulations en liens les unes avec les autres, ces articulations étant liées à des membres.

Au-delà de cette considération purement étymologique, différents domaines proposent des définitions en fonction des articulations et membres considérés.

1. Langues vocales

Dans le domaine de la parole, le phénomène de coarticulation est étudié selon Kühnert et Nolan (dans Hardcastle, [HAR99]) depuis les années 1850 d'après des expérimentations mises au point par Brueke et Bell aux Etats-Unis, mais le terme « coarticulation » date lui des années 1930.

Depuis, selon le texte de l'appel à communication de l'atelier AFCP 07 [AFC07], bien qu'il existe encore des débats sur des subtilités de définition, il semble qu'une base commune soit que la coarticulation est « une expansion de traits propres à un segment phonétique sur les segments adjacents ». Cette définition nous semble un peu réductrice puisqu'elle induit un sens unique d'influence, d'un segment vers les autres. Les auteurs du texte rappellent à ce sujet que les débats ne sont pas clos et que si certains considèrent qu'il y a un ajustement d'un seul segment dont les traces vont se retrouver dans l'autre segment, d'autres envisagent que cet ajustement s'applique aux deux segments (et incidemment qu'ils se retrouvent tous deux affectés l'un par l'autre).

La coarticulation est un phénomène envisagé au niveau de la phonétique articulatoire où il est question d'étudier les organes de la parole et la production des sons. Des auteurs comme Daniloff et Hammarberg [DAN73] ajoutent des considérations mécaniques, anatomiques et neurologiques pour caractériser ce phénomène.

L'existence du phénomène de coarticulation est justifiée par Deketelaere et coll. [DEK01] qui expliquent que la coarticulation est partie prenante de la parole et que générer de la parole naturelle requière de pouvoir générer de la parole continue et coarticulée : ainsi, sans coarticulation il n'y a pas de parole naturelle.

Certaines propriétés de ce phénomène sont mises en avant. Ainsi Deketelaere et coll. [DEK01] expliquent que ce phénomène a pour objectif de minimiser les efforts articulaires : l'objectif d'économie est physique, cela induit que pour appréhender le phénomène dans son intégralité il faut tenir compte des propriétés physiques des articulateurs qui sont coarticulés. Cette idée est assumée également par Ma [MA08], quand l'auteure établit que la coarticulation est « la conséquence de deux facteurs principaux : (1) la stratégie de planification régissant les gestes articulatoires de la parole, et (2) les caractéristiques du système physique de production mis en jeu lors de l'exécution des gestes de la parole. Comme d'autres mouvements humains, les mouvements de la parole sont planifiés selon des stratégies optimales de commandes motrices, en minimisant les « efforts » du locuteur dans le système moteur, tout en atteignant les objets de la tâche ». Sur un autre plan, Mannell [MAN08] explique que ce phénomène serait plus robuste (se produit de manière quasi identique à chaque occurrence) au cœur des syllabes qu'entre les syllabes.

Le phénomène de coarticulation est subdivisé en deux phénomènes : l'anticipation et la persévération. Perrier [PER04] définissent le phénomène d'anticipation ainsi : selon eux, l'étude de la

production de la parole s'appuie sur le postulat que la production d'une chaîne de phonèmes est fondamentalement une tâche séquentielle consistant à produire le phonème (i) avant le phonème (i+1) et après le phonème (i-1). Ils parlent alors d'anticipation lorsqu'on trouve soit dans les signaux articulatoires, soit dans le signal acoustique, des indices sur les caractéristiques articulatoires du phonème (i+1) avant qu'on ait atteint la réalisation du phonème (i). Gosy et coll. [GOS08] réalisent une mesure de l'influence par anticipation du phénomène de coarticulation, en langue hongroise. Ils se concentrent sur l'allongement en durée de la production d'une voyelle dans le cas où elle précède une consonne nasale. Il est également question du phénomène de « persévération » (ou carryover) : à l'inverse du phénomène d'anticipation, dans la persévération un phonème subit des influences des phonèmes suivants. Ma [MA08] souligne que ces deux processus ont des statuts différents : la persévération « serait plutôt vu comme une conséquence de la physique du système de production, le point de départ pour chaque phonème étant les positions atteintes par les articulateurs dans les phonèmes précédents. Dans ce cas, la configuration articulatoire atteinte pour le deuxième phonème est différente selon la configuration articulatoire atteinte lors du premier phonème. Par contre, l'anticipation a lieu seulement si le locuteur peut prévoir et anticiper les phonèmes à venir. Elle est donc vue comme le résultat de la stratégie de planification au niveau supérieur. »

Que ce soit en anticipation ou en persévération, des études ont montrées qu'il y avait une influence du phénomène de coarticulation bien au-delà des unités immédiatement voisines. Ainsi, Benguerel et coll. [BEN74] ont montré dans une étude sur le français que la protrusion des lèvres (le fait de les avancer) pouvait être anticipée (ce qu'ils nomment « forward coarticulation ») et commencer jusqu'à six segments phonétiques avant qu'elle soit nécessaire pour la prononciation. De la même manière les auteurs ont mis en évidence une influence étendue du carryover (nommé « backward coarticulation » dans leur étude). D'assez nombreuses études se sont par la suite intéressées à cet aspect de la coarticulation, et comme le souligne Grosvald¹⁰ [GRO10], qui propose un état de l'art assez conséquent sur le sujet (sur lequel nous revenons en fin de cette partie), cette influence dépend des voyelles et consonnes impliquées, des voyelles potentiellement utilisables dans la langue, de la prosodie, et des habitudes de chaque locuteur.

Quelques études présentent des aspects particuliers de la coarticulation. Nguyen et coll. [NGU01] expliquent que ce phénomène a un impact sur l'identification des mots : plutôt que de se baser sur une représentation phonétique abstraite des mots, l'auditeur peut être directement influencé par une grande variété de contrastes phonétiques extrêmement ténus. Lolive [LOL08] introduit la notion de « coarticulation nasale » qui selon lui permet de différencier les locuteurs : il

s'agit de la coarticulation entre des phones, consonnes ou voyelles, l'un des deux (ou les deux) étant nasal.

Des études ont également été menées sur différents phénomènes d'altérations phonétiques, dont fait partie la coarticulation. Elles sont réparties en quatre catégories, suivant qu'il s'agit d'une addition, suppression, substitution ou déplacement d'unités phonétiques. Nous en citerons une, qui s'est attachée à un phénomène qu'on retrouve dans les autres domaines d'études de la coarticulation (dans celui du geste et de la langue des signes en particulier) : l'épenthèse.

Cette altération phonétique est définie par Bibeau [BIB75] comme l'insertion d'une ou plusieurs unités sonores à l'intérieur d'un morphème ou entre deux morphèmes. Ce phénomène a plusieurs formes. Le premier concerne le passage d'un son de voyelle à un son de glide, par exemple la voyelle /i/ devenant /y/ dans « iode ». Ceci a pour effet de regrouper les morphèmes, qui passent dans notre exemple de deux à un. Il y a également l'épenthèse du /ɲ/ qui se retrouve dans la prononciation de « ours blanc » dit « ours/ɲ/blanc ». Enfin, il y a l'épenthèse dite « nasale », qui consiste en une modification de la prononciation d'une consonne : le « n » dans « grand » est prononcé différemment dans « grandir » (le second est nasal, du fait de la présence d'un contexte particulier à la fin du premier morphème et au début du second). L'épenthèse consiste donc en un ajout et/ou une modification de phonèmes (qui a des conséquences sur les morphèmes) dans un contexte particulier, définie précisément par l'auteur (mais dans lequel nous ne rentrerons pas en détail ici, se reporter à l'ouvrage).

Une sorte d'altération également possible est l'assimilation, définie par Mannell [MAN08] comme un relâchement dans l'expression de la langue : ce relâchement serait dû à des contraintes visant à maintenir un équilibre entre une perception distincte et une facilité d'articulation. Ainsi, selon l'auteur, l'assimilation peut être vue comme une version améliorée de la coarticulation.

Pour des définitions et exemples de coarticulation par anticipation et persévération, entre autres, voir Flechsig [FLE06], et pour une présentation complète des « origines de la coarticulation » en LV, se référer à Kuhnert [KUH99].

Ainsi, la coarticulation est définie en parole au niveau phonétique, articulaire comme acoustique. Ce phénomène intervient lors de la mise en articulation séquentielle de phonèmes, dans le cadre de l'expression vocale d'une langue naturelle, et est la modification des unités phonétiques composant

le phonème en fonction des unités précédentes et suivantes. Cette modification est décrite comme un recouvrement spatial et temporel des paramètres articulatoires des deux unités coarticulés.

2. Gestes

Sur le site du Centre national de ressources textuelles et lexicales (CNRTL³), la définition de « geste » est au sens propre la suivante : « Mouvement du corps, notamment du bras, de la main, de la tête. », ce terme étant étymologiquement « emprunté du latin *gestus* : attitude, mouvement du corps, mimique, jeu ». Cette définition large englobe presque tous les mouvements du corps.

Kendon [KEN04] définit le geste (« gesture ») comme une dénomination pour les actions qui ont pour objectif déclaré d'exprimer quelque chose⁴. Ainsi, pour l'interlocuteur, ces gestes sont des actions intentionnellement générées par le locuteur dans le but d'exprimer quelque chose, ou en tous cas pour aider à atteindre un objectif que s'est fixé le locuteur. La décision de l'intentionnalité d'un geste est laissée à l'appréciation de l'interlocuteur et est totalement subjective. L'auteur signale des divergences dans les différentes définitions de ce qu'est un geste et ce qui ne l'est pas, et malgré un état de l'art des différentes typologies proposées par différents auteurs, il arrive à la conclusion que vu que les gestes considérés sont de nature humaine, il ne peut être établi de manière pérenne de catégories de ces gestes. Il faut donc, toujours selon l'auteur, plutôt raisonner sur la base des différentes typologies proposées et s'en servir comme outils provisoires utiles dans certaines recherches, suivant certains buts. Mais il ne faut surtout pas considérer une classification comme étant universelle et applicable quelque soit le locuteur et quelques soient les conditions de réalisation.

Nous avons tenu à présenter ces deux définitions pour montrer deux aspects de la notion de geste : pour certains les gestes ne sont que des mouvements, et dans cette acceptation on retrouvera les « gestes vocaux » qui regroupent tous les mouvements des organes sollicités dans l'expression de la parole, quand pour d'autres les gestes sont uniquement présents dans le cadre de la transmission d'information. Par conséquent, nous pouvons dire que selon la conception de ce qu'est un geste, la notion de coarticulation pourra s'appliquer ou pas : si un geste est un mouvement non obligatoirement intentionnel et non communicatif, alors le phénomène de coarticulation tel que définit précédemment dans le domaine de la parole (s'appliquant à une langue naturelle) n'est pas pertinent.

³ <http://cnrtl.fr/definition/academie9/geste>

⁴ Citation originale : « 'Gesture' we suggest, then, is a label for actions that have the features of manifest deliberate expressiveness. »

Néanmoins, des recherches ont été menées, par exemple dans le domaine de la musique, et qui font référence à ce phénomène. Ainsi, Godoy [GOD09] définit la coarticulation comme étant la fusion d'événements distincts, qui sont à la fois des actions (i.e. des mouvements) et des sons, perçus dans une globalité, comme par exemple dans la fusion des syllabes en des mouvements articulatoires continus et des sons de la langue (en cela il rejoint la définition du domaine de la parole), ou dans la fusion de sons isolés en un son continu produisant des gestes et des mélodiques en musique⁵. L'auteur ajoute que c'est un sujet traité principalement dans les domaines de la linguistique [HAR99], des sciences du mouvement [ROS91], mais aussi dans le domaine de la musique, où quelques études existent, par exemple concernant le piano [ENG97] ou le violon [WIE06]. Godoy trouve une certaine beauté à la coarticulation dans le sens où elle peut fonctionner dans le futur et dans le passé, ce qui veut dire que les événements futurs sont influencés par les événements passés et inversement. Ceci rejoint pleinement la définition du domaine de la parole. Dans [JEN10], les auteurs établissent les principes de la coarticulation pour le domaine du geste : cela concerne des événements isolés mis en contexte, les événements passés influençant ceux à venir, et inversement, il semble que cela soit dû à une contrainte biomécanique, et que cela nécessite un contrôle au niveau moteur⁶. Ils rapportent une citation de Rosenbaum et coll. [ROS91] selon laquelle la coarticulation peut être vue comme un avantage : « Pensez à un homme qui écrit au clavier en utilisant un doigt à la fois. Sans la capacité de coarticulation des doigts, sa vitesse serait très lente. Des mouvements simultanés des doigts permettent une réalisation rapide, tout comme les mouvements de la langue, des lèvres et du voile du palais, permettent un débit de parole rapide. La coarticulation est une méthode efficace pour augmenter la vitesse de réalisation alors que les articulateurs individuels (parties du corps utilisées pour les mouvements) se meuvent relativement lentement »⁷.

⁵ Citation originale : « the fusion of otherwise distinct events, meaning both action events and sound events, into larger and holistically perceived chunks, e.g. as in the fusion of syllables into continuous articulatory movement and sound in language, or in the fusion of singular tone events into continuous sound producing gestures and melodic or textural patterns in music ».

⁶ Citation originale : « 1) Otherwise singular events embedded in a context 2) Past events influence present events, i.e. position and shape of effectors are determined by recent action 3) Future events influence present events, i.e. position and shape of effectors are determined by preparation for future actions (anticipatory movements) 4) Seems to be a biomechanical necessity 5) Seems to be a motor control necessity, i.e. anticipation in motor control ».

⁷ Citation originale : « coarticulation can be seen as an advantageous element: "...it is a blessing for us as behaving organisms. Think about a typist who could move only one finger at a time. Lacking the capacity for finger coarticulation, the person's typing speed would be very slow. Simultaneous movements of the fingers allow for rapid responding, just as concurrent movements of the tongue, lips and velum allow for rapid speech. Coarticulation is an effective method for increasing response speed given that individual effectors (body parts used for movement) may move relatively slowly." »

Parmi les études citées par [JEN10], Engel et coll. [ENG97] ont étudié les gestes des pianistes : leur conclusion est que les doigts se positionneraient de manière optimale avant (une à deux notes avant) la frappe de la touche, Baader et coll. [BAA05] ont analysé la coordination des mains de violonistes jouant une séquence musicale : leur conclusion est qu'il y a une anticipation du doigt qui va appuyer sur la corde à se placer correctement bien avant que la note ne soit jouée (en fait, dès que la note précédente a été jouée), et Dahl et coll. [DAH04] ont étudié les gestes des batteurs et observé une influence au niveau de l'accentuation de la frappe qui peut être anticipée plusieurs frappes en amont. Dans leur étude, Engel et coll. [ENG97] font référence explicitement à la notion de « coarticulation » par laquelle ils désignent la modification d'un geste répété, lors de la répétition d'un morceau. De manière générale, Jensenius et coll. [JEN10] présentent plusieurs études qui traitent du geste musical (piano, batterie, violon, etc.) et de sa modification en fonction de la répétition, selon les points de vue intra et inter personnels (sur un sujet qui répète le même geste, et par comparaison entre sujets experts et novices, ou experts entre eux).

Ces études considèrent donc la coarticulation comme un phénomène résultant d'un apprentissage, non pas comme dépendant d'un contexte de production mais plutôt d'une expérience de production.

Jensenius et coll. [JEN10] vont plus loin en conclusion de leur étude, puisqu'ils introduisent la notion de « Goal-Posture » En effet, suite à leur expérience (enregistrements par capture de mouvement et audio de morceaux joués par des pianistes), ils concluent sur le fait que la coarticulation implique le son mais aussi l'action qui amène à produire ce son, donc la perception et la production. Ces segments action-son sont caractérisés par des points de saillance dans la musique, comme le beat, les accents, des pics mélodiques. Ces points de saillance sont vus comme des « goal-postures » autant pour le mouvement qui produit le son que pour le son perçu⁸. Nous n'approfondirons pas cette étude et ces conclusions plus en détail, nous voulons souligner le fait que, dans le domaine musical, en ce qui concerne la relation entre les mouvements de production et la perception du résultat sonore, le phénomène de coarticulation connaît des résultats intéressants et, selon les auteurs, potentiellement applicables à la LS (en effet, la théorie linguistique de Johnson [JOH] postule un système de « posture-transition » comme fonctionnement de la LS). Cette dernière hypothèse nécessite d'être approfondie pour être confirmée.

⁸ Citation originale : « So far: We believe there are indications of coarticulation in sound-action chunks, both in trajectory and velocity data. And: We believe that coarticulation concerns both the sound and the sound-producing action, hence both perception and production. But we also believe these sound-action chunks are centered on certain salient points in the music such as downbeats, other accents, and melodic peaks. These salient points we consider as goal-postures both for the sound-producing movement and for the perceived sound »

Dans Sosnik et coll. [SOS04] les auteurs font état du fait que le phénomène de la coarticulation est très étudié dans le domaine de la concaténation de mouvement. Le terme réfère au fait que dans un système expert les unités de base sont influencées par anticipation par les unités adjacentes, ce qui résulte en des chevauchements spatiaux et temporels, créant une nouvelle entité différente de la somme des unités qui la composent. Cependant les auteurs insistent sur le fait qu'il ne faut pas cantonner le terme à la simple superposition de segments successifs d'une séquence de mouvements donnés. Leurs résultats montrent qu'effectivement la nouvelle unité est mémorisée et qu'elle ne correspond pas à la somme des deux unités qui ont engendré sa conception : il pourrait de même s'être créés des raccourcis dans le système nerveux pour effectuer ces nouveaux mouvements. Les auteurs concluent en proposant que la définition de la coarticulation soit élargie à la substitution de deux ou plus unités de mouvement de base par une autre unité du même type.

Dans le même ordre d'idée, Ansuini [ANS09], dans une étude portant sur la coarticulation et le contrôle moteur impliqué au niveau neuronal, déclare que ce phénomène fait référence à l'influence subie par les primitives de commande motrice lorsqu'elles s'enchaînent : il y a alors une anticipation de la primitive suivante. Cette anticipation se matérialise par un chevauchement spatial et temporel d'éléments de chaque primitive, et la nouvelle primitive résultante est différente de la simple addition des éléments des deux primitives adjacentes ([ENG97], [SOS04]).

Dans le domaine de l'étude du geste, il y a également des auteurs comme Gibet et coll. [GIB01] qui s'intéressent à la synthèse de gestes communicatifs, avec un modèle s'inspirant de la réalisation des signes de LS. Les auteurs utilisent une fonction pour simuler la coarticulation des gestes, qui calcule les coordonnées d'un angle (d'une articulation) à un instant donné t en fonction de ses coordonnées à l'instant précédent $t-1$ et de ses futures coordonnées à l'instant suivant $t+1$. Dans cette étude, le phénomène de coarticulation est explicitement décrit comme une anticipation des mouvements qui consiste à adapter le mouvement à réaliser en fonction du contexte de l'exécution, ce contexte étant composé du mouvement précédent et du mouvement suivant. Dans le cadre de cette étude, une solution proposée pour modéliser le phénomène est de prendre en compte les cibles les plus proches dans le passé et dans le futur (ces cibles étant des valeurs d'angle que les articulations doivent atteindre). Dans une autre étude, Gibet et coll. [GIB07] présentent un modèle de génération de gestes de la LS française. Ce modèle s'appuie sur une approche de modélisation semi-formelle des gestes, et sur un formalisme de spécification capable de traduire un énoncé en un flux continu de données, ceci permettant le contrôle du mouvement d'un signeur virtuel. Leur modèle bénéficie à la fois d'une connaissance des structures linguistiques propres à la LSF, et de résultats d'analyse issus de mouvements capturés (mouvements qui ne sont pas de la langue des

signes). Dans le cadre de cette étude, les auteurs font référence à la notion de coarticulation afin de réaliser l'enchaînement des gestes modélisés : il s'agit de pré-positionner les articulateurs en vue d'exécuter le mouvement à venir.

Il s'agit dans cette étude (de même que [GIB01]) d'une coarticulation par anticipation uniquement (il n'y a pas de prise en compte du fait qu'un geste puisse avoir une influence sur les mouvements à venir).

Ces études partagent une même base de définition : le phénomène de coarticulation correspond à une modification du mouvement, voire du geste tout entier, lors de la mise en contigüité de deux unités gestuelles. On peut noter que la définition porte sur deux aspects : le mouvement et l'emplacement des gestes et des effecteurs du geste (bras, mains, doigts).

Ces études sur le geste, que ce soit suivant la définition au sens large incluant les gestes de musique ou celle de Kendon incluant les gestes communicatifs, utilisent la notion de coarticulation quasiment dans le sens qu'elle a dans le domaine de la parole : une modification d'un élément lors de sa mise en contexte, les modifications étant dues à une influence des éléments précédents et suivants. Cependant, les gestes tels qu'envisagés par Kendon ne sont pas produits suivant des règles de langue mais des règles de communication. Ainsi, il est possible de réutiliser certaines notions et conclusions de ces études, tout en conservant à l'esprit qu'il ne faut pas envisager de les réutiliser simplement telles quelles dans le domaine des LS.

3. Têtes parlantes

Il est un domaine à mi chemin entre la synthèse vocale et gestuelle, c'est le domaine de l'animation d'agents conversationnels animés, et des têtes parlantes. En vue de l'animation d'une tête parlante virtuelle, Robert et coll. [ROB07] ont réalisé une étude destinée à prédire la coarticulation labiale en fonction des caractéristiques des phonèmes. Selon les auteurs, la parole ne correspond pas à une simple juxtaposition de phonèmes isolés : ils définissent le phénomène de coarticulation comme la perturbation résultant de l'influence des phonèmes entre eux, et insistent sur l'importance de la modélisation de ce phénomène pour rendre la lecture labiale possible. En parlant de coarticulation labiale et de coarticulation de phonèmes, les auteurs font là le lien entre la coarticulation des sons de la parole et la coarticulation des lèvres qui émettent ces sons. Ils axent leur recherche sur la coarticulation des lèvres vis-à-vis des sons et ont basé leur modèle de coarticulation sur trois paramètres labiaux : l'ouverture des lèvres, leur étirement et la protrusion. On peut en déduire que c'est, selon eux et dans le cadre de leur étude, des paramètres importants pour la réalisation de la coarticulation des lèvres vis-à-vis de la parole.

La définition de la coarticulation englobe dans ce domaine hybride les aspects gestuels et vocaux, en considérant à la fois la coarticulation des phonèmes vocaux et la coarticulation des lèvres vis-à-vis des ces phonèmes. Cette vision de la coarticulation n'est pas sans rappeler celle de [JEN10] (cf. partie « gestes ») quand il considère les mouvements du pianiste et le son résultant.

De même, la définition de la coarticulation donnée par Schroeder [SCH] est similaire à celles données précédemment, appliquée au visage, soit un phénomène ayant pour résultat la modification de la partie visuelle de la parole en fonction des éléments de parole précédents et suivants⁹. Dans son étude, l'auteur dresse un état de l'art dans le domaine de l'animation faciale, et présente plusieurs modèles d'animations qui ont pris en compte la notion de coarticulation. Les méthodes d'implémentations sont différentes, mais les caractéristiques modélisées sont celles précédemment évoquées (influence de segments les uns sur les autres en fonction du contexte, anticipation, carryover).

Certaines études s'attachent à créer une tête parlante naturelle dans l'objectif précis de permettre la lecture labiale. Ainsi Bevacqua et coll. [BEV03] prennent en compte le carryover et l'anticipation de la coarticulation pour les phonèmes, et les appliquent aux visèmes : afin de réaliser les mouvements des lèvres pour les consonnes, les auteurs analysent le contexte environnant ces consonnes et utilisent des règles de hiérarchie d'influence des voyelles sur les consonnes. L'influence d'une voyelle sur une suite de consonnes est fonction des autres voyelles présentes et de la distance temporelle entre cette voyelle et les différentes consonnes.

Pelachaud [PEL91] faisant référence à Brooke [BRO90] rapporte un aspect particulier à l'animation d'une tête parlante : des sons peuvent être identiques acoustiquement alors que le mouvement des lèvres réalisé pour le produire est différent. La définition que donne Pelachaud de la coarticulation est un chevauchement entre éléments pendant leur phase de production. Les frontières entre les phonèmes deviennent floues. Le phénomène peut être une anticipation du mouvement des lèvres sur la voyelle suivante, ou une même configuration des lèvres pendant une succession de phonèmes¹⁰ (ce qui pourrait être assimilé aux notions d'anticipation et de carryover citées précédemment).

Ces études à l'interface entre celles sur la LV et les gestes montrent que le phénomène de coarticulation a une base de définition qui est commune à ces domaines : il s'agit d'un phénomène

⁹ Citation originale : « Coarticulation refers to the way visual speech changes based on surrounding segments. »

¹⁰ Citation originale : « This problem occurs due to the overlap of units during their production. The boundaries among phonemic items are blurred. Sometimes, lips move in anticipation on the following vowel (for example, the `r' in `read' or in `rat' shows different patterns). In some other cases, lips keeps the same shape over a succession of segments (for example, the `t' in `rat' looks different from the `t' in `complete). »

de modification d'unités (phonèmes, mouvements de l'appareil articulatoire, mouvements du visage, et gestes de certains membres du corps) de par leur mise en articulation les unes avec les autres. Fort de cette constatation, nous allons maintenant nous intéresser à la situation dans le domaine des langues des signes.

4. Langues des signes

Dans le domaine des LS, les études se sont en grande partie inspirées des connaissances et techniques issues du domaine des LV. Ainsi, la définition du phénomène de coarticulation est étudiée au niveau phonétique. Le tout est de savoir ce qui est considéré comme élément phonétique en langue des signes. Même si beaucoup d'études ont uniquement considéré les mains (et plus particulièrement leurs emplacements dans les études sur l'épenthèse), les dernières recherches tendent à considérer un phénomène de coarticulation plus global et incluent la direction du regard, les expressions du visage et les mouvements du corps. Mais la coarticulation a en très grande majorité été considérée comme s'appliquant aux mouvements (par opposition l'emplacement, l'orientation, etc.)

Dans l'article de Holt et coll. [HOL06], au sein d'un état de l'art en reconnaissance automatique de langues des signes, les auteurs font le parallèle entre la coarticulation et la variation en ce sens que c'est également une modification de sons et de signes en une forme différente. Avec la coarticulation, les sons ou signes adjacents sont modifiés parce qu'ils se chevauchent, au point qu'ils sont parfois réalisés simultanément. Les auteurs définissent également le phénomène d'anticipation, où selon eux la main dominée se déplace pour se mettre en position pour un signe à deux mains. Si l'anticipation rejoint la définition des langues vocales, la coarticulation est ramenée à une unique forme de modification, l'assimilation.

Dans l'étude de Ojala et coll. [OJA09], les auteurs affirment qu'étudier la coarticulation est plus facile avec les signes puisque tous les articulateurs sont visibles en permanence, ce qui fait la différence avec la parole où les articulateurs sont en majorité cachés. Ils admettent que la coarticulation intervient aussi dans les expressions du visage et dans les gestes en général, mais ils ne s'y intéressent pas dans leur étude. En effet, ils se concentrent sur les mouvements des mains uniquement. L'objectif de la coarticulation est selon eux de rendre les transitions plus faciles : en parole, elle permet même de parler plus vite. Ses effets sont visibles sur les configurations des mains et la localisation dans l'espace de signation.

L'épenthèse est traitée de manière beaucoup plus riche dans le domaine, peut-être parce qu'elle est considérée comme n'impliquant que le mouvement des mains.

Dans une des premières études sur le sujet, Vogler [VOG99] définissent l'épenthèse comme des mouvements ajoutés entre la fin du premier signe et le début du second. Cependant, cette transition n'a pas de signification linguistique, et est due à la caractéristique visuo gestuelle de la langue des signes.

Kelly et coll. [KEL09], Dreuw [DRE10] et Yang [YAN10] suivent la même définition que celle de Vogler ; Yang ajoutent que ces mouvements ne correspondent pas un signe et peuvent inclure la forme de la main, son mouvement, et peut avoir une certaine durée de réalisation dans le temps, parfois égale à celle de certains signes.

Cette définition de l'épenthèse est beaucoup moins approfondie que celle du domaine des LV, et ne concerne majoritairement que le mouvement des mains.

Il y a également une étude de Tyrone et coll. [TYR10] sur la réduction des signes comme une réduction phonétique en ASL. Ils expliquent que la réduction phonétique apparaît dans le cadre d'une expression langagière naturelle, quand au lieu de faire attention à l'articulation précise de chaque mot qu'on prononce, l'expression produite est une forme bien moins articulée. Ainsi, quand les signes sont exprimés en contexte par des locuteurs natifs d'une LS, ils diffèrent souvent de leur forme isolée. Dans certains cas, la réduction phonétique se manifeste par un signe réalisé en contexte à une localisation plus basse que dans la réalisation du signe isolé.

Ces auteurs ne nomment pas leur phénomène la coarticulation, mais cette réduction pourrait être un des aspects de la coarticulation dont la définition est donnée dans d'autres études.

5. Aspects à retenir pour notre méthodologie

Nous pouvons remarquer, dans ce tour d'horizon de la notion de la coarticulation, que les définitions sont assez semblables. Elles sont plus approfondies dans le domaine des LV et de l'animation de gestes et de têtes parlantes, mais c'est peut-être parce que ces domaines de recherche bénéficient de plus d'ancienneté.

Grosvald [GRO10] présente un état de l'art des différentes études réalisées dans les domaines des langues parlées et signées et concernant la coarticulation. Sa conclusion est que les facteurs impliqués dans la coarticulation vont des caractéristiques articulatoires des segments aux variations phonémiques inter locuteurs. L'auteur ajoute que si on considère que la parole et les signes sont deux manifestations d'une même source de production langagière, alors des études dans ces domaines sont indispensables si nous voulons comprendre le phénomène de coarticulation dans son entier.

En nous appuyant sur les différentes définitions que nous avons passées en revue, il nous semble important de considérer, pour notre étude, les points suivants.

La coarticulation :

- intervient au niveau phonétique,
- implique les gestes manuels et non manuels,
- (en lien avec les deux points suivants) concerne les différents paramètres articulatoires des gestes (mouvement, configuration, localisation, orientation),
- est une variation de segments phonétiques lors de leur mise en contexte,
- peut avoir une influence sur plusieurs segments avant (anticipation) et/ou après (persévération) la mise en contexte des segments,
- peut s'exercer entre deux segments phonétiques et au cœur d'un segment phonétique.

De plus, les paramètres concernés ne varient pas au même moment ni à la même intensité (certains peuvent potentiellement être modifiés de manière plus considérable et/ou sur une longue durée les uns par rapport aux autres).

Bien que nous ne nous soyons pas penchés en détail sur tous ces aspects, il est important selon nous de les présenter afin de délimiter le cadre théorique de notre étude, en ce qui concerne le phénomène de coarticulation.

II. Modèles et mises en œuvre informatiques

Dans cette partie nous présentons les modèles de coarticulation existants et leurs implémentations (traduction des modèles en programme informatique) dans les mêmes domaines que précédemment : les LV, les gestes, les têtes parlantes et les LS.

1. Langues vocales

Dans le domaine des LV, il existe des modèles pour la reconnaissance automatique et des modèles pour la synthèse de la parole. Nous présentons dans un premier temps les principes des modèles utilisés dans le cadre de la synthèse de parole, et dans un second temps le principe des systèmes de reconnaissance, où sont principalement utilisées des approches statistiques.

a. Synthèse

Depuis la machine parlante du baron Von Kempelen en 1730, en passant par le Bell Labs' Voder (premier synthétiseur électrique) de Homer Dudley en 1939, les techniques orientées sur la synthèse

par articulation ont bien évolué¹¹. Au début, seule la technique de synthèse par articulation avait été exploitée. Celle-ci consistait en la reproduction mécanique ou électrique du conduit vocal de l'être humain (les lèvres, la bouche et le larynx). Depuis la fin du 20ème siècle, la technologie prend ses fondements dans les secteurs de la linguistique, des mathématiques, de l'électronique et de l'informatique. L'auteur liste trois approches de synthèse de la parole :

- la synthèse par règles,
- la synthèse par concaténation de diphtongues,
- la synthèse par sélection d'unités dans une grande base de données.

La synthèse par règles se base sur l'idée que si un spécialiste aguerri de la phonétique est capable de lire la représentation graphique (le spectrogramme) d'une suite de paroles, il lui est possible de déduire et donc de produire les règles permettant de fabriquer artificiellement un tel spectrogramme pour une suite donnée de phonèmes. Lorsque cette représentation graphique est dessinée, il suffit alors de générer le signal électrique correspondant à l'aide de composantes électroniques. Cette méthode de génération de la parole correspond à la première génération de logiciels et décrit les règles régissant la coarticulation de la suite des phonèmes nécessaires. Les voix générées par le biais de cette technique étaient certes intelligibles, mais la parole métallique obtenue n'avait rien de naturel et était peu agréable à entendre.

La synthèse par concaténation de diphtongues (unité acoustique qui commence au milieu de la zone stable d'un phonème et se termine au milieu de la zone stable du phonème suivant) consiste à mettre bout à bout des segments acoustiques déjà coarticulés qui sont extraits d'une base de données de signaux de parole. Ces segments correspondent à des sons préenregistrés qui sont stockés dans une base de données et non pas fabriqués par un modèle expert. Comparativement à la technique de synthèse par règle, la synthèse par concaténation de diphtongues produit une parole plus fluide. Afin d'améliorer encore le naturel et l'agrément d'écoute, la concaténation est suivie par une phase de "lissage acoustique des discontinuités" : à chaque jonction entre diphtongues, des fonctions mathématiques sont appliquées afin de rendre la liaison plus douce. Bien que plus performante, cette seconde technologie n'est pas encore totalement satisfaisante sur le plan du caractère naturel de la voix produite. En effet, la voix synthétique semble souvent hyper-articulée et l'intonation paraît encore très artificielle.

La synthèse par sélection d'unités dans une grande base de données, récente, consiste essentiellement à multiplier les unités acoustiques utilisées lors de la concaténation. Plutôt que de garder un seul exemplaire de chaque diphtongue de la langue, on stocke plusieurs instances différentes de la même unité phonétique. Afin de compléter la base de données, le système extrait de plusieurs

¹¹ <http://www.awt.be/web/can/index.aspx?page=can,fr,voc,200,010>

heures de parole segmentée les unités phonétiques identiques, mais ayant des prosodies et des contextes phonétiques différents. La méthode de synthèse consiste à concaténer des unités phonétiques en sélectionnant les instances les plus adéquates au sein de la base de données. Les segments « plus adéquats » sont ceux dont :

- le contexte est le plus proche de la chaîne phonétique à synthétiser,
- la prosodie se rapproche le plus de la prosodie à produire.

Dans cette présentation des différentes possibilités de synthèse de la parole, on peut constater que le phénomène de coarticulation est pris en compte dans les systèmes les plus récents. Il s'agit de le traiter sous forme combinatoire.

b. Reconnaissance

Il existe différentes méthodes pour la reconnaissance automatique de la parole, selon Deketelaere [DEK01]. Cependant, un schéma général peut être esquissé sous la forme suivante (Figure 2) :

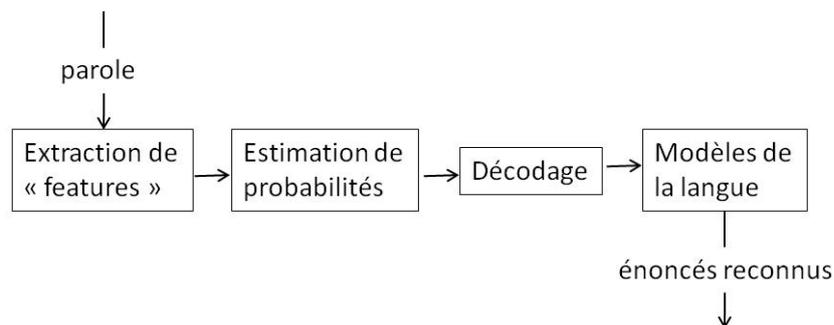


Figure 2 : Schéma synthétique d'un système de reconnaissance de la parole

Il y a tout d'abord captation de la parole (environnement acoustique plus la transduction opérée par le matériel d'enregistrement). Sur ce signal est réalisée une extraction de « features » (permet de régler les problèmes de bruits dus à l'environnement et au matériel, supprime les variations inutiles et repère les catégories de sons), puis une estimation des probabilités (comparaison du spectrogramme enregistré par rapport à des modèles connus et proposition de plusieurs réponses possibles). En fonction des implémentations, il y a une étape de décodage, qui permet de déterminer la réponse qui convient le mieux. Enfin, des modèles de la langue permettent de prendre en compte des effets connus comme le raccourcissement de voyelles en parole continue, pour aboutir au final à des énoncés reconnus.

Typiquement sont utilisés des HMM (modèles de Markov cachés) : un HMM est un automate à états finis construits d'après un ensemble finis d'états, chaque état de l'automate correspondant à une probabilité de transition. Il y a parfois utilisation de ANN (réseau de neurones artificiels), conjointement avec les HMM dans des systèmes hybrides. Ces réseaux ont la particularité de pouvoir avoir plusieurs entrées et peuvent recevoir des informations sur le contexte acoustique. Les modèles de langue sous-jacents doivent donc être capables de fournir ces informations au système.

Il s'agit d'une implémentation de solutions statistiques, qui ne prennent pas en compte un modèle de coarticulation : les données sont analysées par comparaisons avec une base de références, et la coarticulation est laissée de côté au profit de la reconnaissance des mots (plutôt que ce qu'il se passe entre deux mots).

Ainsi, s'il existe des études prenant en compte la coarticulation aussi bien en synthèse qu'en reconnaissance de parole, il n'existe de modèles que dans le cadre de la synthèse. Le phénomène est traité de manière combinatoire.

2. Gestes

Dans Gibet et coll. [GIB01] les auteurs expliquent que la synthèse d'un geste complexe peut être envisagée comme une séquence de cibles à atteindre, mais cela ne permet ni la fluidité de passage d'une cible à l'autre, ni la possibilité d'anticiper le mouvement précédent. Donc, il faut un système qui permet de tenir compte du geste précédent et du geste suivant pour garantir cette fluidité et plus de naturel dans le geste réalisé. Les auteurs utilisent une fonction mathématique (pour plus de détails sur la formule, nous renvoyons à l'article) qui prend en compte les gestes juste passés et juste à venir en leur attribuant une valeur d'importance. Ainsi, les articulateurs qui permettent de réaliser le geste peuvent être pré positionnés pour le réaliser, permettant une meilleure fluidité. Dans cette solution choisie par les auteurs, les mécanismes de transition entre gestes ne sont pas explicitement spécifiés : ils sont réalisés directement au niveau du moteur de génération. Ceci a été implémenté pour les signes de LSF par Gibet et coll. [GIB08a].

Nous avons cité plusieurs études traitant de la coarticulation de gestes dans le domaine musical. Il n'y a pas d'auteurs, à notre connaissance, qui ait réalisé une implémentation de leur modèle, cependant leurs analyses aboutissent à un ensemble de constatations qui est parfois dérivé en règles de coarticulation. Ainsi, le résultat de l'étude d'Engel [ENG97] montre selon les auteurs que le contrôle des mouvements de la main et des doigts a les mêmes caractéristiques que le contrôle de la parole : ils estiment avoir décrit l'équivalent pour les doigts du phénomène de coarticulation pour les doigts. Wiesendanger [WIE06] et Dahl [DAH04], respectivement pour le violon et la batterie, ont fait

les mêmes observations, sans pour autant proposer de modèle. Jensenius et coll. ([JEN10]) ne proposent pas non plus de modèle mais posent des bases théoriques nouvelles pour envisager le phénomène de coarticulation de gestes musicaux : selon les auteurs, il y a des éléments de coarticulation dans les unités son-action (qui composent le geste musical), tant dans la trajectoire que dans l'accélération, et la coarticulation concerne non seulement le son produit mais aussi les articulateurs mis en œuvre pour le produire.

Hormis dans l'étude de Gibet et coll. ([GIB01]), il n'y a pas à notre connaissance de modèle de coarticulation pour les gestes.

3. Têtes parlantes

Dans cette partie, nous ne citerons qu'une seule étude qui dresse un état de l'art sur les modèles utilisés pour l'animation des têtes parlantes. Robert et coll. [ROB08] présentent une approche qui combine deux étapes. Un algorithme qui permet de prédire l'existence de la coarticulation, qui exploite des représentations phonétique articulaires des sons de la parole. Puis un second algorithme réalise une concaténation et une complétion qui synthétise la dynamique des quatre paramètres articulatoires des fonctions entraînées sur un corpus audiovisuel.

Les auteurs passent tout d'abord en revue quelques modèles du phénomène d'anticipation :

- Le modèle « look-ahead » ne raisonne que sur un plan purement phonologique et par exemple, entraîne un mouvement de protrusion des lèvres sur une voyelle v_1 de part la présence de v_2 dans la séquence v_1ccv_2 (où « c » représente une consonne). Perrier et coll. [PER94] expliquent que les traits articulatoires caractérisant un phonème se rétro-propagent aux phonèmes précédents tant qu'ils sont compatibles avec les traits caractéristiques de ces derniers ; c'est donc un modèle où les caractéristiques propres du système de production sont résolument ignorées.
- Le modèle « time-locked » suit le principe de gestes qui se chevauchent : il met l'accent sur les contraintes dynamiques des articulateurs de la parole, et propose que la durée de l'anticipation soit déterminée par le temps de réponse des articulateurs à une commande gestuelle, et ceci afin que la configuration articulatoire requise soit atteinte au moment où le son associé est émis.
- Le modèle hybride est un mélange des deux modèles précédents, et décompose l'anticipation en deux phases : la première qui dépend de contraintes phonologiques, la seconde plus rapide, qui dépend de la dynamique des articulateurs impliqués dans la coarticulation.

- Le Modèle d'Expansion du Mouvement propose que, pour un geste articulatoire donné, la durée de l'anticipation augmente linéairement avec le temps séparant deux sons successifs pour lesquels ce geste articulatoire est pertinent.

En parallèle de ces études sur l'anticipation, d'autres modèles ont été mis au point afin de décrire le phénomène de coarticulation dans son ensemble et pas uniquement l'anticipation.

- Le modèle de Öhman ([OHM67]) modélise les mouvements de la langue : il part de l'hypothèse que lors de la prononciation de consonnes rapides, les mouvements se superposent aux mouvements nécessaires pour réaliser la voyelle suivante.
- Un autre modèle, celui de Cohen & Massaro ([COH93]) utilise des fonctions de dominance affectées à chaque articulateur. La somme de ces fonctions représente le profil du mouvement articulaire mis en œuvre. Cette approche est similaire à celle du modèle « time-locked ».
- Enfin, certains systèmes se basent sur des règles phonétiques et articulaires afin de prédire l'anticipation et/ou la rétention de gestes (ainsi que nous l'avons vu précédemment). Ces approches sont similaires au modèle « look-ahead ».

En conclusion de leur étude, Robert et coll. [ROB08] font état de leur modèle de prédiction de la coarticulation, qui se sert d'un corpus d'apprentissage pour extraire des informations pertinentes sur la coarticulation et permet de reconstruire les articulations des lèvres et de la mâchoire dans n'importe quelle séquence de phonèmes grâce à une concaténation et une complétion de ces articulations.

Tous les modèles cités précédemment sont issus des études en parole, il n'y a pas à notre connaissance de modèle spécifique issu directement du domaine des têtes parlantes.

4. Langues des signes

Pour les LS, nous pouvons également distinguer les études concernant la synthèse de celles sur la reconnaissance automatique.

Il n'existe à ce jour à notre connaissance aucun modèle de coarticulation pour la synthèse de LS, au-delà de l'interpolation (utilisée pour l'animation des signeurs virtuels par exemple dans les projets Visicat [VER01], eSign [ELL04], Tessa [COX02], Simon the signer [BAN00], SignSmith [SIG], etc.), qui consiste à considérer la fin du signe précédent et le début du signe suivant pour réaliser une liaison d'emplacement à emplacement plus ou moins fluide suivant l'algorithme utilisé.

En revanche, plusieurs auteurs se sont intéressés à ce phénomène, que ce soit pour en faire une analyse, ou pour le prendre en compte dans le cadre de systèmes de reconnaissance de LS. Selon Bhuyan et coll. [BHU06], les techniques développées à ce jour ne sont pas adaptées à un vocabulaire large (en termes de quantité d'éléments). Les systèmes proposés se sont rarement intéressés à la reconnaissance de signes exprimés de manière continue, et préfèrent se centrer sur les signes exprimés de manière isolée. Parmi les études qui traitent de reconnaissance de signes continus nous n'avons trouvé que celle de Vogler et coll. [VOG99], qui date maintenant de plus d'une dizaine d'années, qui propose une modélisation explicite non pas du phénomène de coarticulation mais de l'épenthèse (ajout d'un mouvement entre les signes). Depuis, les auteurs sont revenus sur cette proposition et ne modélisent plus explicitement ce phénomène [VOG03]. La prise en compte de l'épenthèse était réalisée par un HMM (modèle de Markov caché), modèle de traitement statistique, tout comme les autres signes : cette solution suppose un apprentissage préalable du système, il a donc fallu que les auteurs apprennent au système les différents mouvements d'épenthèse que le locuteur pouvait réaliser. Grâce à une optimisation, les auteurs n'ont pas eu à faire apprendre toutes les possibilités d'épenthèse mais se sont servis de catégories (un mouvement vertical entre le front et le torse était considéré comme identique à un mouvement vertical entre le torse et la taille, par exemple) afin de réduire le nombre de possibilités. Ils ont réalisés une évaluation avec un vocabulaire de 22 signes et obtiennent un taux de reconnaissance correct selon eux, identique à ceux qu'on peut obtenir avec les systèmes de reconnaissance vocaux. Depuis, comme nous l'avons souligné, ces auteurs sont revenus sur cette modélisation explicite du mouvement d'épenthèse et ont préféré évacuer ce phénomène (et la coarticulation dans son ensemble) en se focalisant sur les signes.

C'est d'ailleurs à notre connaissance dans cette voie que se sont engagés la totalité des auteurs qui travaillent sur la reconnaissance en LS : ils utilisent des modèles statistiques (HMM ou assimilés) pour reconnaître les signes, après une phase de segmentation des signes. Cette phase de segmentation peut être réalisée suivant différentes techniques. Par exemple, Bhuyan et coll. [BHU06] proposent de segmenter les signes en fonction de mouvements : les auteurs segmentent dans un premier temps la vidéo en gestes, suivant les pauses de déplacement de la main, puis chaque segment ainsi obtenu est analysé pour déterminer les phases de coarticulation et les éliminer grâce à un traitement par automate (proche du principe statistique des HMM) ou pas d'autres considérations de mouvements. Un autre exemple est l'étude de Yang et coll. [YAN06] qui suivent une approche en deux temps, similaire à la précédente : la première étape consiste à segmenter les signes, la seconde à les reconnaître. Pour la première étape, les auteurs utilisent de qu'ils nomment des CRF (Conditionnal Random Field) : ce sont des modèles qui permettent de s'affranchir des considérations du contexte et qui ne nécessitent pas autant de données d'apprentissage que les

HMM. Le système effectue d'abord une sélection d'images clés, pouvant être soit dans un signe soit dans une phase de coarticulation, puis les CRF sont utilisés pour déterminer, en fonction des images clés déjà déterminées et celles à venir, ce qui est de la coarticulation. Les signes ainsi repérés (parce qu'ils ne sont pas de la coarticulation) sont ensuite reconnus par le système d'après une base de données de signes pré enregistrés.

Citons également les travaux récents de Lefebvre sur la segmentation dans des énoncés vidéo de LSF. Dans [LEF08], les auteurs nous proposent une nouvelle approche de la segmentation entre signes, qui ne se base pas sur une connaissance a priori des signes. Ils se servent d'indices provenant du seul mouvement des mains, indices calculés à la fois automatiquement à l'aide d'un algorithme de programmation dynamique et fournis par un prétraitement humain qui permet de déterminer où sont les signes dans la vidéo (mais pas leur début ni leur fin de réalisation).

Nous n'irons pas plus avant dans le détail des différentes solutions existantes pour les systèmes de reconnaissance actuelle, nous souhaitons mettre en avant le fait que la tendance actuelle est à la non modélisation du phénomène de coarticulation et à une reconnaissance signe à signe de la vidéo après une phase de segmentation des signes.

Du point de vue de l'analyse du phénomène de coarticulation, il existe une certaine quantité d'études, qui font état des résultats de leur analyse sans proposer de modèle.

L'étude de Cheek [CHE01] porte sur la coarticulation de la forme de la main et plus précisément les variations des formes « 1 » et « 5 » (respectivement un index tendu, et les 5 doigts tendus et écartés). Les résultats montrent que le phénomène de coarticulation est bien présent, autant en anticipation qu'en persévération. Mauk [MAU03] s'intéresse lui à la coarticulation des emplacements des mains dans l'espace, et plus précisément la variation d'emplacement au niveau de l'espace neutre (utilisé pour exprimer des signes ne nécessitant pas d'emplacement particulier) dans un contexte où le signe précédent est réalisé au niveau du front. Il s'avère que l'emplacement au niveau du front a une influence sur l'espace neutre, qui est réalisé plus haut, mais l'influence n'est pas réciproque. Le signe réalisé au niveau du front étant un signe de contact, l'auteur conclue que les signes de ce type sont moins susceptibles d'être influencé par la coarticulation. Dans leur étude, Jerde et coll. [JER03] analysent la coarticulation du mouvement de la main dans une séquence de dactylologie réalisée par des interprètes de LS, afin de mettre à jour des effets d'anticipation et de persévération. Leurs conclusions sont qu'il y a effectivement de tels effets, et qu'ils pourraient être assimilés à respectivement de l'assimilation (qui aurait tendance à réduire les différences entre deux formes de mains) et de la dissimilation (qui aurait tendance à augmenter les différences entre deux formes de mains). Ils remarquent également qu'il y a des occurrences simultanées d'assimilation de

certaines articulations et de dissimilation pour d'autres, ce qui démontre selon eux un contrôle en parallèle d'articulateurs pour réaliser une tâche pourtant sérielle. Plus récemment, Ojala et coll. [OJA09] font une analyse eux aussi de la coarticulation en ce qui concerne les mouvements des mains et des doigts, mais sur toute la LS et pas seulement sur la dactylogogie. Leurs résultats montrent qu'il y a des schémas d'accélération et de décélération des articulateurs en LS qui sont similaires aux articulateurs en LV. Ils ajoutent que le rythme de base des plus hauts niveaux de LS (qu'ils ne définissent pas) montre clairement des actions et des pauses entre les signes lorsqu'ils sont en contexte ([LID89], [KIT98]). Enfin, Grosvald [GRO09] mène une des premières (la première selon l'auteur) étude sur la mesure de l'influence à long terme de la coarticulation entre emplacements. Il constate que les personnes sourdes sont sujettes à la coarticulation entre emplacements, mais selon lui dans un degré moindre à ce que les personnes entendant produisent en LV : cela pourrait s'expliquer, toujours selon l'auteur, par le fait que la LS de par sa modalité visuelle offre une meilleure perception des articulateurs que la LV, donc que les locuteurs de cette dernière améliorent leur perception par la présence de la coarticulation.

5. Bilan

Nous avons vu qu'il existe des modèles de gestes qui prennent en compte la coarticulation, dans le domaine des têtes parlantes, modèles qui sont issus de travaux sur les LV. Il n'existe pas de modèle de coarticulation pour les LS, bien qu'il y ait une première tentative par Gibet et coll. [GIB07] pour appliquer son modèle conçu pour les gestes.

III. Méthodologies de recherche

Nous nous intéressons maintenant aux méthodologies existantes et aux outils utilisés pour mettre au point les modèles et résultats présentés dans la partie précédente, toujours pour les domaines des langues vocales, des gestes, des animations du visage (gestes et voix), et des langues des signes.

1. Langues vocales

Deketelaere [DEK01] présente un état de l'art des méthodologies existantes dans les domaines de la reconnaissance, de la synthèse, et de l'analyse de LV. En reconnaissance, le problème principal est la différence d'expression entre deux locuteurs qui expriment pourtant les mêmes mots. Cette particularité est connue sous le nom de variabilité inter locuteurs. De plus, une même personne ne prononce pas le même mot de la même manière dans des contextes différents. Cette autre particularité est dénommée variabilité intra locuteur. Un humain ne sera pas perturbé par ces

variations, mais un système informatique éprouve de grandes difficultés. De plus, la qualité d'enregistrement est très importante dans le cas des systèmes de reconnaissance : le bruit et les distorsions du signal acoustique sont très délicats à prendre en compte, d'autant plus qu'il n'y a aucune connaissance a priori sur ce bruit ou ces distorsions. Nous avons vu dans la partie précédente le principe général de fonctionnement d'un système de reconnaissance (cf. Figure 2), et d'après ce schéma nous pouvons conclure qu'il n'y a pas de méthodologie spécifique pour établir un modèle de coarticulation puisque ce phénomène n'est pas modélisé : dans un système de reconnaissance, le signal est segmenté suivant ce qu'il est estimé être des unités (des mots ou des phonèmes) afin que celles-ci soient reconnues par comparaison avec des données préenregistrées.

L'auteur présente également une structure assez générique des systèmes de synthèse de LV : ils sont constitués d'un module de traitement de la langue, qui permet d'obtenir une transcription phonétique du signal d'entrée (texte par exemple), et d'un module de traitement du signal qui permet de transformer cette transcription phonétique en parole. La méthodologie de réalisation de ces systèmes n'est pas explicitée en tant que telle, mais suivant les modèles utilisés, il faut disposer de corpus de phonèmes ou de syllabes, représentatif de toutes les combinaisons possibles de la langue. Il y a donc en amont une analyse de la langue qui permet de concevoir des règles pour mettre en œuvre la coarticulation dans les systèmes de synthèse.

Enfin, concernant les modèles de langue, les recherches utilisent très fréquemment une modélisation ad hoc par apprentissage du système et utilisation de grand corpus. Ceci est du au paradigme empirique qu'une bonne estimation des probabilités d'un événement peut être obtenue par observation de cet événement dans des corpus de taille suffisamment grande. La méthodologie consiste donc à enregistrer, ou réutiliser s'ils existent déjà, de très grands corpus et à les utiliser comme base d'apprentissage par calculs statistiques d'un phénomène.

Dans le cas où un réel modèle est mis au point, et non pas seulement une utilisation de résultats empiriques de manière statistique comme nous venons de le présenter, la conception des modèles de coarticulation pour les LV suit une méthodologie générale 1) de réalisation d'un corpus avec des locuteurs donnés qui réalisent des tâches précises, 2) d'annotation de ce corpus, et 3) d'analyse de ces annotations pour aboutir au modèle. Les analyses sont réalisées grâce à des outils spécifiques de traitement du signal. Les analyses d'annotations sont faites par comparaison des durées des unités mesurées.

Ainsi, la méthodologie générale qui s'applique sur les modèles de langue parfois utilisés dans les systèmes de synthèse ou de reconnaissance est : constitution d'un corpus et analyse de ce corpus par annotations et comparaisons de durées et de valeurs.

2. Gestes

Dans l'antiquité les gestes sont considérés comme important car faisant partie de la rhétorique. Depuis les centres d'intérêts se sont multipliés. Les premières études du geste sont textuelles et dessinées ([KEN04]). Les enregistrements présentés dans cet ouvrage sont des vidéos de dialogues entre deux personnes sur des thèmes divers, menés sous la direction de Kendon ou d'autres chercheurs poursuivant les mêmes buts, les participants étant au courant qu'ils étaient filmés. Une transcription de la parole a été réalisée, ainsi que des annotations textuelles et des dessins faits à la main pour les informations sur les séquences de gestes et l'alignement avec la transcription. Il s'agit là de l'enregistrement de gestes « isolés », pour une comparaison de gestes issus de différentes cultures, cependant la méthodologie est similaire à ce que nous venons de voir pour les LV.

Avec les progrès technologiques, la vidéo a été bien souvent conservée mais s'est vu adjoindre les techniques de capture de mouvement. Cette technique consiste à enregistrer les mouvements d'une personne en fonction de capteurs placés sur cette personne (ceux-ci pouvant être sur n'importe quelle partie du corps et sur la face).

Ainsi, dans Soechting et coll. [SOE87] les auteurs conduisent une étude sur les mouvements du bras dans le but de montrer que les gestes compliqués sont en fait composés de plusieurs gestes de base, et que ces gestes de base ont une forme d'un arc d'ellipse. Des sujets droitiers, debout, doivent réaliser une figure simple, comme dessiner un « 8 », ou une étoile, ou un cube. La taille de la figure et le plan dans lequel elle doit être dessinée sont donnés. Il leur est demandé de réaliser la figure plusieurs fois à un rythme confortable. Les localisations du coude et du poignet sont enregistrées par ultrasons et la flexion/extension du coude mesuré goniométriquement. Des enregistrements électromyographiques de l'activité des muscles de l'épaule et du coude sont réalisés par des électrodes. Les technologies évoluant encore, Engel et coll. [ENG97] utilisent des enregistrements audio par ordinateur et des capteurs optoélectroniques (basés sur la réflexion de lumière) pour les mouvements. Leur étude porte sur la coarticulation chez les pianistes afin de déterminer si la cinématique des doigts varie en fonction de la note suivante. Les pianistes ont été sélectionnés selon leur niveau, et il leur a été demandé de jouer un ensemble de morceaux dont les premières notes étaient identiques. Les enregistrements sont analysés par des calculs de moyenne et de variance.

Dans Sosnik et coll. [SOS04], les auteurs mènent une étude sur la coarticulation de séquences de gestes. Ils montrent que l'entraînement intensif sur une séquence de mouvements de la main dans le plan, passant par plusieurs cibles, entraîne une coarticulation des composantes des mouvements et la génération de nouvelles composantes. La coarticulation est étudiée au travers une tâche de construction de trajectoires passant par plusieurs cibles. Il s'agissait de relier quatre points avec leur main dominante « aussi rapidement et précisément que possible », en démarrant au signal sonore.

Le suivi des trajectoires et l'enregistrement des données temporelles et spatiales ont été réalisés sur une tablette tactile Wacom.

Plus récemment, Godoy et coll. [GOD09] utilisent la capture de mouvements et la vidéo pour enregistrer les trajectoires de la main, du bras et du torse, et les comparer dans leur forme isolée et en contexte (c'est-à-dire dans une séquence de plusieurs gestes) afin d'étudier la coarticulation.

Enfin, des études dédiées aux gestes expressifs s'attachent à la segmentation et l'annotation de ces gestes. Ces études visent à caractériser la structure spatiale et temporelle des gestes, permettant ainsi le passage d'une description segmentale vers la synthèse du mouvement ([KIP04]).

Dans la totalité des études que nous venons de citer, la méthodologie d'analyse des gestes consiste en la réalisation de tâches par plusieurs sujets, l'enregistrement de données sous forme numérique (principalement à l'aide de capteurs de mouvements) et l'analyse statistique de ces données. Ces études, bien que portant sur l'analyse de mouvements balistiques, non linguistiques, montrent que la méthodologie reprise du domaine des LV semble bien adaptée au domaine du geste. Nous allons d'ailleurs voir dans la partie suivante que le domaine des têtes parlantes puise quelques inspirations de ces domaines, pour l'animation des lèvres et la synthèse de la voix coarticulée avec le mouvement des lèvres.

3. Têtes parlantes

Les méthodologies utilisées dans les études de la coarticulation dans le domaine des têtes parlantes sont issues des domaines de la parole principalement. Certaines recherches s'appuient sur des recherches sur l'expression d'émotions, mais il n'est pas question de coarticulation ou d'autre phénomène du même ordre dans ce cas là.

Au niveau technique, les données sont collectées par capture de mouvement ([SIF05], [DEN06], [BEV03]), séquentialisées et utilisées avec des modèles d'animation (nous ne rentrerons pas dans le détail de ces modèles, pour plus de détails se référer à [BEV03], [PER04] et [ROB08])

Il existe également des études utilisant la vidéo, telle celle de Buisine [BUI06] où, dans le but d'étudier les comportements émotionnels multimodaux de la vie de tous les jours, les auteurs ont collectés un corpus d'interviews télévisées riches émotionnellement. Plusieurs niveaux d'annotation ont été codés manuellement à l'aide d'un outil spécifique (Anvil, que nous détaillerons dans une partie suivante) par trois experts. Les expressions faciales sont là encore générées à partir d'un corpus (comme pour le cas de la capture de mouvements), mais les paramètres faciaux (mouvements des sourcils, direction du regard, tension de la bouche, etc.) sont spécifiés à partir des annotations manuelles de la vidéo initiale.

Nous n'avons pas trouvé dans nos recherches de travaux portant sur la définition d'une méthodologie d'étude de la coarticulation dans ce domaine, contrairement aux domaines des LV et des gestes. L'animation des têtes parlantes utilise les modèles de coarticulation mis au point dans le domaine de la parole, pour la synthèse de la voix, mais aussi pour l'animation des lèvres, puisque certaines études en LV portent sur les mouvements des articulateurs.

4. Langues des signes

Dans cette partie nous allons nous intéresser aux méthodologies d'analyse du phénomène de la coarticulation, et de la LS en général.

Nous réalisons une analyse de l'existant en plusieurs temps : tout d'abord, nous avons cherché à savoir quelles étaient les méthodologies existantes pour les LS, afin de déterminer celle qui serait la plus à même de servir notre étude. Ensuite, nous avons analysé en détail les différentes étapes composant ces méthodologies afin de vérifier leur pertinence pour notre recherche.

a. Méthodologies

Dans la plupart des études qu'il nous a été donné de lire, les méthodologies d'analyse de LS au sens large, c'est-à-dire pas uniquement sur la coarticulation, reposent sur les bases suivantes : sélection d'un panel de locuteurs de LS, conception du corpus par des tâches qu'aura à réaliser le panel, analyse des résultats par traitements statistiques des productions. D'un point de vue de la recherche en informatique, les productions sont directement les enregistrements du corpus, que ce soit de la vidéo ([OJA09], [KEL09], entre autres), ou de la capture de mouvements ([TYR10], [VOG99], [JER03], entre autres) ; en linguistique, il y a d'abord une étape d'annotation du corpus ([JOH08b], [SCH08], [CRA08], [SAL03], entre autres), avec un outil d'annotation, et les traitements s'appliquent sur les annotations. En TALS, puisque les recherches se basent sur des études linguistiques, il est également souvent fait recours à des annotations ([LSC02], [CHE10], entre autres), et de récentes recherches s'intéressent à une automatisation de ces annotations ([MLO10], non sur un corpus vidéo mais sur un corpus de capture de mouvements). Nous allons maintenant voir un peu plus dans le détail chacune de ces étapes, déjà succinctement présentées dans [SEG08].

b. Sélection des locuteurs

En fonction des études cette première étape est plus ou moins détaillée. Quand en informatique (sans prise en compte de notions d'autres domaines, comme la linguistique par exemple) il est assez peu fait état de la sélection des locuteurs ([VOG99]), en linguistique et en TALS cette étape est souvent jugée comme très importante ([JOH08b]). Ceci est selon nous dû à l'objectif des études qui dans le premier cas est d'implémenter et de valider un système informatique, un programme, dans le

second cas d'étudier la langue et de mettre au point des modèles, avec une implémentation en TALS, de la langue : il faut donc être sûr de la langue enregistrée pour les études, au sens où il faut pouvoir délimiter un cadre linguistique au contenu du corpus.

Les locuteurs sont sélectionnés suivant des critères variés (âge, sexe, langue maternelle, etc.) en fonction de l'objectif de l'étude. Les critères de sélection entrent en ligne de compte lors de l'analyse des résultats et des conclusions de l'étude.

c. Conception / réalisation du corpus

Cette seconde étape consiste à concevoir les tâches que devront réaliser les locuteurs et à enregistrer leurs productions en LS. Les tâches peuvent être de nature variée, allant de la dactylographie de lettres données ([JER03], [RIC09], [TYR10]) à l'enregistrement d'expressions complètement libres ([CRA08], [SCH08]), en passant par des expressions libres en réponse à une question ou sur un thème donné ([LSC02], [EFT09]). Les dispositifs d'enregistrement sont principalement de deux sortes : vidéo ou capture de mouvement. La seconde technologie a beaucoup évolué : alors qu'au départ il n'y avait que des enregistrements par gants numériques ([BRA96]), qui ne fournissaient d'informations que sur les paramètres des mains, quelques corpus sont aujourd'hui réalisés grâce à des capteurs placés sur le corps de la personne (pour la LS, les capteurs se limitent à la partie supérieure du corps).

Holt et coll. [HOL06] mentionnent que la capture de mouvements a l'avantage de fournir rapidement des informations sur la forme de la main, les mouvements et les emplacements dans l'espace, alors que dans le cas de la vidéo il faut appliquer des algorithmes de traitement d'images afin d'extraire de données 2d des données 3d (emplacements, mouvement, etc.). Les auteurs signalent que l'inconvénient est la non captation par capture de mouvement des paramètres non manuels (expressions du visage notamment). Et puisque les algorithmes de traitements d'images prennent de plus en plus en compte l'expression du locuteur dans sa totalité (et pas que les mains), la vidéo a encore un certain avantage pour l'étude des LS. Cependant, nous nuancerons ces propos puisqu'aujourd'hui, certes les algorithmes de traitements d'images ont bien évolués, mais les premières captations d'expressions du visage par capture de mouvements sont en cours et semblent prometteuses.

La vidéo a encore tout de même un avantage certain puisqu'il s'agit d'un dispositif non intrusif, dont on peut dire qu'il influe moins sur l'expression du locuteur que la capture de mouvements. Enfin, la captation vidéo reste beaucoup moins onéreuse que l'enregistrement par capture de mouvements.

d. Analyse des productions

Comme nous l'avons signalé, les recherches en informatique appliquent directement des algorithmes sur les données vidéo. En TALS et en linguistique il y a d'abord une étape d'annotation du corpus, ensuite des analyses sur ces annotations. Les annotations sont en fait présentes dans les études informatiques, mais elles restent sommaires et se cantonnent à labelliser les données vidéos dans le but d'apprentissage par des programmes. Mais d'autres annotations peuvent être réalisées, en fonction toujours de l'objectif poursuivi. Une annotation est, selon Boutora [BOU08], l'utilisation de symboles donnés pour labelliser un événement temporel sur une vidéo. Ces symboles peuvent être de différentes natures (cf. partie sur les annotations), mais sont toutes stockées sous forme informatique afin de pouvoir être traités par des programmes (statistiques pour la majorité des cas).

e. Evaluation

Une fois les conclusions des analyses dressées, il y a évaluation des résultats auprès d'un panel de locuteurs sélectionnés. Cette évaluation a lieu en laboratoire ou en situation « réelle » mais majoritairement sur des prototypes, ce qui influence le côté « réel » de la situation. Les locuteurs testeurs sont sélectionnés suivant des critères méthodologiques généraux de représentativité.

Nous ne connaissons pas à ce jour d'évaluations sur l'implémentation d'un modèle de coarticulation pour l'animation d'un signeur virtuel. Néanmoins, il y a quelques études sur le résultat de la synthèse de LS à l'aide d'un signeur virtuel.

Dans le cadre de la signeuse virtuelle TESSA, mise en place pour faciliter les interactions entre un usager sourd et le guichetier de la poste en « traduisant » les communications, Cox et coll. [COX02] expliquent qu'ils souhaitent évaluer trois aspects : la qualité des signes, la difficulté de la réalisation de la transaction, le ressenti des usagers sourds et des guichetiers de la poste. Ils réalisent une sélection d'évaluateurs (des personnes sourdes et des guichetiers), leur font remplir le début d'un questionnaire, demandent aux personnes de faire une tâche d'appariement entre des énoncés en BSL (LS anglaise) et des expressions typiques de la poste, font passer le test « réel » d'une transaction entre la personne sourde et le guichetier via le système, et enfin demandent à ce qu'ils remplissent la fin du questionnaire.

Plus récemment, Huenerfauth [HUE07] propose une méthodologie d'évaluation de la génération de l'ASL (LS américaine). La génération des énoncés est évaluée par le sujet au travers du choix d'une séquence animée traduisant visuellement ce qui est exprimé en LS. Le sujet visualise un énoncé en ASL où il est question de spatialisation de deux éléments (personnes, animaux, objets etc.) et de leur positionnement ainsi que leur interaction. Ensuite le sujet visualise trois propositions de traduction visuelle de ce qui a été exprimé en ASL : cette traduction visuelle utilise une représentation

graphique des éléments et les anime. Le sujet doit désigner la traduction qui lui semble la plus proche de ce qui a été signé précédemment. Il doit également répondre à plusieurs questions sur l'ergonomie, la compréhensibilité des énoncés générés etc.

Nous ne rentrerons pas ici plus avant dans le détail des évaluations citées, et n'en citerons pas d'autres, réservant ces informations pour le chapitre traitant de l'évaluation de notre modèle.

f. Bilan

Dans ce chapitre, nous avons constaté que le phénomène de coarticulation a des définitions assez semblables dans les différents domaines de recherche auxquels nous nous sommes intéressés. Elles sont plus approfondies dans le domaine des LV et de l'animation de gestes et de têtes parlantes, mais c'est peut-être parce que ces domaines de recherche bénéficient de plus d'ancienneté d'existence.

En nous appuyant sur les différentes définitions que nous avons passées en revue, il nous semble important de considérer, pour notre étude, les points suivants : la coarticulation

- intervient au niveau phonétique,
- implique les gestes manuels et non manuels,
- (en lien avec les deux points suivants) concerne les différents paramètres articulatoires des gestes (mouvement, configuration, localisation, orientation),
- est une variation de segments phonétiques lors de leur mise en contexte,
- peut avoir une influence sur plusieurs segments avant (anticipation) et/ou après (persévération) la mise en contexte des segments,
- peut s'exercer entre deux segments phonétiques et au cœur d'un segment phonétique.

De plus, les paramètres concernés ne varient pas au même moment ni à la même intensité (certains peuvent potentiellement être modifiés de manière plus considérable et/ou sur une longue durée les uns par rapport aux autres).

Bien que nous ne nous soyons pas penchés en détail sur tous ces aspects, il est important selon nous de les présenter afin de délimiter le cadre théorique de notre étude, en ce qui concerne le phénomène de coarticulation.

A l'issue de ce tour d'horizon, nous constatons que la méthodologie généralement utilisée dans les études correspond à nos besoins. Nous allons donc suivre les différentes étapes qui la constituent. En plus de la procédure globale de méthodologie, il nous semble important d'observer en détail ce qui a été fait dans les différentes étapes pour là encore nous en inspirer.

Nous avons choisi d'élaborer notre modèle de coarticulation à partir de données réelles. Ces données, comme nous l'avons vu, sont collectées dans le cadre de la constitution d'un corpus. La partie suivante se penche sur la notion de corpus et ses méthodes de conception.

Chapitre III. METHODOLOGIE :

CORPUS, ANNOTATION ET ANALYSE

Dans ce chapitre, nous explicitons dans une première partie la notion de corpus en langue des signes, en la positionnant par rapport aux corpus de langues vocales (LV), et en détaillant les différentes méthodologies aujourd’hui employées pour leur constitution. Ensuite, dans une seconde partie, nous dressons une liste des types d’annotations possibles pour l’étude de ces corpus, qui sont utilisées dans différents outils dont nous citons quelques exemples, et suivant certains critères que nous détaillons. Puis, dans une troisième partie, nous présentons des outils d’analyse des annotations de corpus multimodaux ou de LS. Enfin, dans une quatrième partie, nous présentons des outils et méthodes statistiques qui permettent l’analyse de données du type dont nous disposons.

I. Corpus

Dans cette partie, nous présentons les définitions de la notion de corpus dans les domaines des LV orales et écrites, et des LS, puis les méthodologies de constitution de corpus dans les domaines des LV, des gestes et des LS.

Nous souhaitons analyser un corpus, dans le sens où nous voulons repérer certains phénomènes : pour cela nous devons constituer ce corpus, puis nous devons l’« annoter » (cela revient à mettre des labels sur les phénomènes, pour pouvoir ensuite leur appliquer des traitements statistiques quantitatifs et qualitatifs). Nous nous intéresserons donc dans un premier temps à la définition et aux modalités de conception d’un corpus, puis à l’annotation de corpus et aux outils disponibles pour remplir cette tâche. Enfin, nous voulons traiter ces annotations, leur appliquer des calculs pour faire ressortir de la masse de données les informations que nous jugeons pertinentes (en l’occurrence, l’alignement entre les énoncés, et des statistiques sur ces alignements : pour finir, nous abordons rapidement le domaine de la bioinformatique d’où est issu notre algorithme d’alignement des énoncés.

1. Définition

La langue des signes est avant tout une langue, nous avons donc voulu nous inspirer des différents domaines de traitements des langues, écrites, vocales et signées, pour définir ce qu’est un

corpus, puisque nous souhaitons en concevoir un pour l'analyser. Nous introduisons tout d'abord la notion de corpus au sens général, indépendamment de la modalité et de la discipline dans lequel il est constitué. Ensuite, nous bordons le domaine des langues vocales dans leur modalité vocale, puis dans leur modalité écrite, et terminons par présenter la notion de corpus dans le domaine des LS.

Pour Mellet [MEL02], un corpus est « un recueil, formé d'un ensemble de données sélectionnées et rassemblées pour intéresser une même discipline ». Il n'y a pas de restriction quant à la modalité des données recueillies, qui peuvent donc être écrites, sonores ou visuelles (voire même odoriférantes ou tactiles). Il n'y a pas non plus de restriction quant aux éventuelles propriétés linguistiques que doivent vérifier les données d'un corpus. Les deux seules règles données dans cette définition sont qu'il y a une sélection du futur contenu du corpus, dans l'optique d'une utilisation par une même discipline.

Selon la définition de Lebart [LEB88], dans le domaine linguistique, un corpus est l' « ensemble limité des éléments (énoncés) sur lesquels se base l'étude d'un phénomène linguistique », et dans le domaine de la lexicométrie il s'agit d'un « ensemble de textes réunis à des fins de comparaison, servant de base à une étude quantitative ». La première définition semble applicable quelque soit la nature des éléments : du moment qu'il s'agit d'énoncés et que l'on souhaite étudier des phénomènes linguistiques, que les éléments soient sous forme graphique (écriture ou symboles), audio (sonore), ou visuelle, n'importe pas. Par contre, il faut que les éléments suivent des règles linguistiques : on ne peut donc pas considérer n'importe quel ensemble d'éléments, même s'ils peuvent être exprimés dans n'importe quelle modalité. Le domaine d'application de la seconde définition est celui de l'étude statistique du vocabulaire (définition de « lexicométrie » selon Dugast [DUG79] : « Étude statistique du vocabulaire; science qui étudie la répartition du vocabulaire dans le discours »). La définition de Lebart, bien que portant potentiellement sur une forme orale ou écrite de discours, spécifie que le corpus est, en lexicométrie, un ensemble de textes qui doivent pouvoir être comparés.

Bien que cela ne soit pas clairement formulé dans la définition de Lebart, il est également question d'une sélection de ce que contiendra le corpus, puisqu'il y a étude d'un phénomène sur un ensemble limité d'éléments : les éléments choisis pour le corpus seront donc fonction du phénomène à observer.

a. Langues vocales orales

Pour Baude [BAU06], dans le domaine des langues orales, le rôle d'un corpus est de répondre à un des deux objectifs suivants : être exhaustif (i.e. restituer « la totalité des données sonores

produites à l'occasion d'un événement, c'est-à-dire avec un commencement et une fin définis ») ou être représentatif (i.e. répondre « à la question de la représentativité : comment rendre compte d'une langue, d'un dialecte, d'un parler ? » en tant qu'« échantillon d'une langue »). La méthodologie mise en œuvre pour la constitution d'un corpus de langue orale découlera du choix de l'objectif ce qui intrinsèquement entraînera des contraintes quant aux données et à leur sélection.

D'autre part, l'auteur met l'accent sur quatre aspects des corpus oraux : « les types de données et de locuteurs, la dimension du corpus, les transcriptions ». Les données sont sollicitées, dans le sens où elles ne viennent pas toutes seules au chercheur, mais celui-ci doit créer les conditions dans lesquelles il va les recueillir. Les locuteurs sont choisis, ainsi que les situations d'enregistrement, en fonction des objectifs de départ. La taille du corpus et des éléments qui le composent sont eux aussi liés aux objectifs d'utilisation du corpus : le choix se limite généralement à l'exhaustivité ou à la représentativité. Les transcriptions effectuées sur les corpus sont de natures très différentes d'une étude à une autre, en fonction des objectifs mais aussi des domaines scientifiques d'utilisation du corpus : les transcriptions ne seront pas les mêmes en informatique, linguistique, sociologie, etc.

La constitution d'un corpus oral est ici décrite aussi bien lors de sa conception théorique qu'au moment de son traitement : en amont il s'agit de sélectionner les locuteurs, les données et une limite de taille, en aval il faut tenir compte du fait que les transcriptions font partie du corpus et les y inclure.

Dans le domaine des langues orales, Baude¹² définit succinctement le corpus oral comme un ensemble, non pas de textes et de documents écrits, mais d'enregistrements vocaux. Il est sous-entendu dans cette définition que les caractéristiques des corpus oraux sont les mêmes que celles des corpus écrits.

Abouda [ABO06] précise qu'un corpus oral est composé de deux types d'éléments : les « données primaires » que sont « les enregistrements de la parole », et les données secondaires » que sont « la transcription et les autres annotations éventuelles ». Il est ici explicitement question de parole, et pas d'oral. Cette définition sous-entend qu'un ensemble de données brutes ne constitue pas un corpus : il faut que ces données soient augmentées d'une transcription et d'éventuelles annotations.

b. Langues vocales écrites

Sinclair [SIN04], considère qu'un corpus est un ensemble de textes au format électronique, sélectionnés suivant des critères de représentation, autant que possible, d'une langue ou d'une

¹² <http://www.corpusdelaparole.culture.fr/spip.php?rubrique2>

variété de langue, et servant de données aux recherches linguistiques¹³. Il n'est donc plus question de sons ou de vidéos, mais uniquement de textes, regroupés selon des critères qui assurent que ce qui est rassemblé est bien une représentation, même partielle, de la langue, dans le but de réaliser des recherches linguistiques.

c. Langues des signes

Il n'existe pas de forme écrite de LS. Plus précisément, il n'existe pas pour les LS de « système graphique permettant de communiquer par écrit, c'est-à-dire de recevoir et de transmettre directement du sens sans transiter par la forme » oralisée « de la langue », selon Garcia [GAR06].

Il existe cependant plusieurs « système de notation » qui ont été développés suivant différents objectifs : pédagogique, linguistique, informatique. Ces systèmes de notation permettent de représenter par écrit (en deux dimensions) ce qui est observé dans l'espace de signation (en trois dimensions), selon le point de vue de l'étude.

Ainsi Bébien [BEB1825], en 1825, mettait au point un premier système de notations dénommé la « mimographie », avec un objectif pédagogique : constituer un recueil de signes sur lesquels élèves et professeurs pourraient s'appuyer. Un deuxième objectif sous-jacent est de normer les signes utilisés, afin d'éliminer les signes méthodiques créés par l'Abbé de l'Épée, et autres créations non naturelles ([BOU08]). Il s'agit pour Bébien de tenter « de réduire tout le langage d'action à un petit nombre d'éléments » en les notant au moyen de caractères qui représenteraient « l'organe qui agit », « le mouvement qui est exécuté », et « s'il y a lieu, l'expression de physionomie qui accompagne quelquefois le geste » ([BEB1825]).

Stokoe [STO60], en 1960, démontre que l'ASL est une langue à part entière. Sa démonstration s'appuie sur une description des signes, réalisée grâce à un système de notation particulier qui porte le nom de son auteur : notation Stokoe. Son système repose sur une approche paramétrique, à savoir qu'un signe est uniquement manuel et est décomposable en une configuration, un emplacement et un mouvement. Bien qu'enrichi d'un paramètre renseignant sur l'orientation de la main, puis d'un paramètre non manuel pour spécifier l'expression du visage, ce système repose toujours sur les paramètres choisis et sur l'inventaire des valeurs possibles pour chacun. Ce dernier dépend de la langue observée, et les critiques formulées sur ce système de notation y sont liées : les paramètres manuels choisis sont réalisés dans une temporalité continue et non discrète, ce qui implique qu'il est impossible de tenir compte de toutes les valeurs potentielles. Essayer de prendre

¹³ Citation originale : « A corpus is a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research. »

en compte le continu grâce à ce système de notation reviendrait à discrétiser la dynamique du signe, ce qui serait réducteur pour la LS.

Depuis 1960, les systèmes de notation développés l'ont été dans le cadre de recherches informatiques ou linguistiques (constitution de dictionnaire, transcription de données, traitements automatiques pour la reconnaissance, la synthèse de LS). Nous n'aborderons pas le détail de ces systèmes ici, le lecteur intéressé pourra se référer à Bergeron [BER06].

Le seul système qui ne suive pas exactement les traces de l'approche paramétrique est SignWriting : il s'agit un hybride entre la description paramétrique et le dessin. Ce système de notation, développé par Sutton [SUT75] au départ pour la danse, est graphique et relativement iconique dans le sens où il y a une réelle ressemblance entre ce qui est noté et ce qui est signé. Ce système est le plus proche de l'appellation « système d'écriture » puisque de nombreuses ambiguïtés sont compensées par la conscience qu'a le lecteur du contexte et de la sémantique de l'énoncé. Il est utilisé de manière pédagogique comme support écrit dans des classes accueillant des enfants sourds à travers le monde, et notamment en France comme expliqué dans Brugeille [BRU06], afin de faciliter l'acquisition de la notion d'écriture et de langue. Les recherches de Di Renzo [REN06] mettent en avant que SW permet de coucher par écrit ce qui est produit à l'oral : en ce sens, ce n'est pas une écriture. De plus, les auteurs mettent en avant qu'il est important de différencier la transcription de l'écriture, et qu'à ce jour SW ne remplit pas toutes les conditions pour satisfaire aux critères d'une écriture (voir leur article pour plus de détails).

Dernièrement, dans les années 1980, trois tomes d'un dictionnaire de LSF (en vérité, un lexique bilingue LSF-français) ont vu le jour en France. Les signes ont été couchés sur papier sous forme de dessins, moins finement détaillés que ceux de Bébien et ayant l'avantage de ne représenter que ce qui paraissait pertinent aux auteurs pour que le lecteur soit en mesure de reproduire le signe donné.

Ces premiers « corpus » sont au format papier, donc en deux dimensions. Qu'il s'agisse de système de notation, de système hybride dessin/notation ou de dessin pur, les notations qui en découlent peuvent être vues comme étant des corpus au sens que propose Mellet [MEL02]. Ces corpus ne sont cependant pas centrés sur la langue mais sur une description possible de cette langue : puisque les analyses sont faites sur des annotations et/ou des transcriptions de corpus, ce n'est pas la langue qui est étudiée mais une interprétation de cette langue.

Les corpus de la langue sont apparus avec l'avènement de la vidéo. Un parallèle peut être fait ici avec les langues orales vocales puisque Abouda [ABO06] considère que « la linguistique dispose de peu de corpus oraux » et que ceci s'explique par le fait que « la tradition littéraire est continue depuis

l'Antiquité quand les modes de conservation du son ont moins d'un siècle et demi d'existence ». Les premières caméras vidéo sont apparues dans les années 1950, ce qui fait que les « modes de conservation » de la vidéo ont seulement un demi-siècle d'existence. Les enregistrements vidéo ont permis de capturer directement la langue des signes à des fins de recherches, plutôt que d'en analyser la description (forcément réductrice par rapport à la langue exprimée oralement).

Un des premiers corpus vidéo de langue des signes en France a été celui sur lequel s'est appuyé Cuxac dans les années 80 pour construire sa théorie (signalons que Jouison [JOU96] l'a précédé en réalisant des enregistrements vidéo de LSF exprimées spontanément, tout comme le fera ensuite Cuxac, et non par élicitations comme cela se faisait majoritairement à l'époque). L'objectif était une description de cette langue d'un point de vue linguistique et cognitif afin de mettre en lumière le fonctionnement de la langue. Il ne s'agissait pas de proposer un système de notation, mais plus de décrire en interne la grammaire de la langue et ses soubassements cognitifs.

Depuis, plusieurs corpus vidéo de LS ont été créés suivants des objectifs de recherche principalement linguistiques, patrimoniaux, informatiques. Dans un cadre linguistique, il s'agit d'affiner les connaissances sur la langue, son fonctionnement, ses variantes, etc. (projet BSL Corpus [SCH08], projet NGT [CRA08]), alors que dans un cadre patrimonial, le but est de mettre en place des bases de données de la langue dans un contexte de sauvegarde (projet LaSiMa [NYS]). Dans un cadre informatique, les chercheurs effectuent des traitements automatiques pour faire du traitement d'images, ou effectuent des analyses « formelle » afin de proposer une génération ou une synthèse de la langue des signes (corpus de l'université de Boston [ATH10], projet LS-COLIN [LSC02]). Dernièrement a été créé un des premiers corpus de LS sous forme de « signeur virtuel » (dans le cadre de notre étude, mais aucune publication n'a encore eu lieu sur cet aspect). Grâce à une technique d'infographie qui permet une « recopie » d'un énoncé vidéo de LS sous la forme d'un personnage virtuel en trois dimensions, nous disposons d'un corpus d'énoncés et de signes du lexique issus du domaine des transports ferrés.

Il existe depuis peu dans la littérature des définitions de ce qu'est un corpus de LS.

Pour Thoutenhoofd et coll. [THO07], « un corpus de langue des signes est un ensemble multimédia contenant des vidéos de personnes Sourdes signantes, ainsi que des annotations qui facilitent les recherches sur les données » (citation originale : a sign language corpus is a multimedia repository that contains digital movies of signing Deaf people, along with annotations that facilitate searching through the data).

Efthimiou et coll. [EFT07] se basent sur la définition de Sinclair (donnée précédemment) et précise que dans le cadre de la constitution de son corpus de LS, le terme de corpus désigne un

ensemble d'extraits de langue sous forme électronique. L'auteur ajoute que la différence qui peut être faite entre un ensemble lambda de telles données de langue et un corpus, est que ce dernier est un ensemble fini constitué d'éléments sélectionnés selon des critères précis dans un but donné.

Plus récemment, Johnston ([JOH08a]) explique que la « lemmatisation », le fait d'appliquer des « lemmes » ou labels sur le contenu vidéo du corpus de LS est le cœur du processus de création du corpus, puisque les vidéos doivent au moins être stockées avec ces labels dans un ordinateur. Ceci s'applique également aux annotations qui doivent à un moment être stockées sous forme de caractères dans un ordinateur. Sans cette lemmatisation (qui doit entraîner une réflexion, selon l'auteur, et ne pas être considéré à la légère) un recueil de vidéos de LS ne constitue pas un corpus, puisqu'il ne peut alors pas être traité par des programmes statistiques ou autres.

d. Bilan

De ce bref historique des collections de données qui ont pu être réalisées sur la LS, il ressort que la notion de corpus dans le domaine de la LS suit peu ou prou les définitions existantes dans le domaine de l'écrit et dans celui de la parole.

Bien qu'il n'y ait pas une littérature aussi abondante sur les corpus de LS qu'il en existe sur les corpus de LV, les recherches sur les LS ont bénéficié des avancées des travaux de leurs collègues sur les LV. Ainsi les méthodes de constitution de corpus et la notion elle-même de corpus dans le domaine des LS ont leurs sources parmi les méthodes et notions issues du domaine des LV.

Un premier point commun est la définition même de ce qu'est un corpus : au vu des définitions citées précédemment, que ce soit pour les textes ou la parole, il est clair que la notion de corpus est formellement la même pour la LS. Cependant, nous pouvons noter une insistance dans la définition de Thoutenhoofd [THO07] quand il parle de personnes « Sourdes » (*Deaf*) : nous revenons plus tard sur ce point très important (cf. infra).

Un autre point commun qui peut être mis en avant est la monodisciplinarité d'utilisation des corpus de LS. Le fait que des disciplines différentes, majoritairement linguistique et informatique, utilisent les mêmes données pour travailler sur une langue n'est pas courant en LS, tout comme en LV. En 2002 a été créé le corpus LS-COLIN (cf. détails dans la partie suivante) dont un des objectifs étaient d'être utilisé par des linguistes et des informaticiens dans le cadre d'études de la langue, de formalisation de modèles linguistiques. Ce corpus pourrait également être utilisé par des interprètes, dans un cadre pédagogique. En 2005 a été conçu le corpus TALS¹⁴, dont le but était de servir de base de travail à des chercheurs de tous horizons, afin qu'ils appliquent leurs résultats ou qu'ils en extraient

¹⁴ <http://tals.limsi.fr/tals2005.html>

des validations de leurs travaux. Les différents chercheurs ont ensuite présentés leurs conclusions lors de la journée d'atelier TALS 2005. Cependant, hormis de rares exceptions comme le corpus LS-COLIN ou le corpus TALS 05, la plupart des corpus de LS ont été créés soit pour être analysés linguistiquement, soit pour être utilisé pour la reconnaissance ou la synthèse informatique, mais pas dans le but d'être utilisés par plusieurs disciplines. Ainsi, le corpus reste en LS un objet monodisciplinaire.

Nous avons mis en lumière le fait que la notion de corpus en LS rejoignait par certains aspects celle de corpus en LV. Nous allons maintenant essayer d'établir un autre parallèle, entre les méthodologies de constitution de corpus de LV ou de LS, afin de voir quelles sont les éventuelles spécificités des corpus de LS.

2. Constitution de corpus

Nous souhaitons ici faire ressortir des concepts généraux de constitution de corpus, pour connaître les grands principes qu'il faut respecter lors de la création d'un corpus en langue vocale, orale ou écrite, afin d'analyser ce qui peut être mis en parallèle dans le domaine des LS.

a. Langues vocales orales

Baude [BAU06] met l'accent sur quatre aspects des corpus oraux : « les types de données et de locuteurs, la dimension du corpus, les transcriptions ». Les données sont sollicitées, dans le sens où elles ne viennent pas toutes seules au chercheur, mais celui-ci doit créer les conditions dans lesquelles il va les recueillir. Les locuteurs sont choisis, ainsi que les situations d'enregistrement, en fonction des objectifs de départ. La taille du corpus et des éléments qui le compose sont eux aussi liés aux objectifs d'utilisation du corpus : le choix se limite généralement à l'exhaustivité ou à la représentativité. Les transcriptions effectuées sur les corpus sont de nature très différentes d'une étude à une autre, en fonction des objectifs mais aussi des domaines scientifiques d'utilisation du corpus : les transcriptions ne seront pas les mêmes en informatique, linguistique, sociologie, etc.

La constitution d'un corpus oral est ici décrite aussi bien lors de sa conception théorique qu'au moment de son traitement : en amont il s'agit de sélectionner les locuteurs, les données et une limite de taille, en aval il faut tenir compte du fait que les transcriptions font partie du corpus et les y inclure.

b. Langues vocales écrites

Dans la littérature, il est fait état de différents critères à respecter afin de constituer un ensemble de données qui puisse être dénommé corpus. Ainsi, dans le domaine des corpus de texte, Bommier-

Pincemin [BOM99], pose le principe que c'est le traitement sur le corpus qui va guider la conception du corpus. Il y a donc constitution d'un ensemble de textes suivant des critères de rapport entre quantité et qualité, de contraintes techniques (de formats ou de mémoire, entre autres), et d'adaptabilité aux traitements ultérieurs. Benzécri [BEN81] met en avant que bien qu'en théorie il faille établir une sélection des constituants du corpus, en pratique il s'agit de se contenter de ce qui est disponible.

O'Donnell [ODO08] préconise une réflexion en trois étapes lorsqu'il est question de construire un corpus. La première consiste à réutiliser l'existant, voire à combiner plusieurs corpus ou à n'en retenir qu'une partie. Si cela ne convient pas, il faut essayer de redéfinir l'existant suivant l'étude qu'on souhaite poursuivre. Enfin, si cela ne convient toujours pas, il faut construire son propre corpus. Cette réflexion met l'accent sur le fait que dans le domaine textuel il ne faut pas hésiter à utiliser les corpus existants avant de penser à en constituer un nouveau.

Il nous semble intéressant de noter dans les travaux précédemment cités que, d'une manière générale, même si la nature des données va dépendre de l'étude menée, la première question à se poser lors de la constitution du corpus est la suivante : « pour quels objectifs ces données sont-elles recueillies ? ». De la réponse à cette question découlent d'autres problématiques comme les contraintes en présence (techniques, par exemple), les différentes conditions (signifiante, acceptabilité, exploitabilité) à vérifier, l'optique choisie (quantité vs. qualité), etc.

En résumé, un corpus de textes est constitué dans un but précis, suivant des contraintes qui peuvent ne pas être liées au corpus lui-même (limite de mémoire informatique, par exemple), et doit vérifier des conditions a posteriori.

c. Langues des signes

Les premiers corpus de LS ont été conçus dans une optique de linguistique descriptive, pour mettre en lumière le fonctionnement de la langue : c'est la démarche de recherche dite « corpus-based », c'est-à-dire des études sans postulat théorique initial, qui analysent des corpus pour proposer des théories à partir des faits relevés, sans a priori. Les résultats de ces études visent parfois à être généralisables à un ensemble plus large que le cadre du corpus ; par exemple Cuxac [CUX96] a construit sa théorie dans cette démarche. Tout comme les corpus textuels et oraux, certaines recherches sur la LS considèrent le corpus comme « un observatoire d'une théorie a priori » [MAY05] : c'est la démarche dite « corpus-driven » où la théorie existe initialement et le corpus est utilisée pour la valider ; par exemple, Sallandre03 [] a démontré le bien fondé, qualitativement et quantitativement, de la théorie de Cuxac.

Par la suite, les corpus ont été utilisés et conçus par des informaticiens pour des recherches sur la reconnaissance automatique ou le traitement d'images. Plus récemment, les informaticiens ont

collaborés avec les linguistes pour constituer leurs corpus. Il est important de noter que les linguistes, et donc les informaticiens, se sont quasiment toujours assurés de la collaboration de la communauté sourde, afin de disposer de données valables.

De manière analogue aux corpus de langue vocale (que ce soit sous forme écrite ou sous forme orale), Thoutenhoofd [THO07] exprime l'idée générale que ce qu'il faut inclure dans les corpus de LS est fonction d'un équilibre entre représentativité et quantité, des ressources disponibles, de la possibilité ultérieure de comparer avec des données existantes, et des motifs linguistiques.

d. Contraintes de notre étude

Nous avons au départ plusieurs contraintes dues au contexte de notre recherche, en ce qui concerne la conception de notre corpus, qui portent sur la grammaire et le vocabulaire. Nous avons décidé de faire réaliser des tâches d'expressions d'énoncés dits complets et d'énoncés isolés, afin de les comparer et de tirer des conclusions quand à la variation de certains paramètres en contexte. Nous avons une locutrice, experte en LSF et habituée des productions filmées, et nous suivons le principe de corpus-based : nous ne cherchons pas à vérifier d'hypothèse, mais à faire ressortir des phénomènes qui nous permettront ensuite de proposer un modèle. La combinatoire du contenu du corpus, entre les énoncés complets et les morceaux d'énoncés qui les composent, nous permet une vision représentative des phénomènes de variations dans le domaine considéré (cf. partie « Mise en œuvre de la méthode » pour plus de détails).

II. Annotation de corpus

Dans cette partie nous présentons les différents types d'annotation qui sont effectuées sur les corpus, ainsi que les outils, plus ou moins spécifiques, qui permettent de faire ces annotations.

1. Annotations

a. Définition

Selon le « Linguistic annotation wiki »¹⁵ : les « annotations linguistiques » regroupent l'ensemble des notations descriptives ou analytiques s'appliquant à des données langagières brutes. Ces données peuvent être temporelles (comme les enregistrements audio, vidéo et/ou physiologiques) ou textuelles. D'autres notations additionnelles peuvent être sous formes de transcriptions de toutes natures (de phonétique à discursive), des labels sémantiques, des analyses syntaxiques, des

¹⁵ http://annotation.exmaralda.org/index.php/Linguistic_Annotation

repérages d'entité nommées, d'identification, d'annotation de coréférence, etc. La base de ces annotations est à la fois les outils qui sont largement utilisés pour créer les bases de données linguistiques, et le format standard de ces outils et bases de données.

b. Différents types

Il existe différents types de valeurs possibles pour annoter un corpus de LS. Selon Boutet [BOU], actuellement, aucune technologie graphique ne permet de remplir pour les LS les fonctions quotidiennes et primaires de l'écrit (trace, notation à la volée, lecture linéaire). Ainsi, l'auteur établit que pour annoter un corpus de LS il peut y avoir recours à une langue vocale (Français, Anglais, Allemand..) sous sa forme écrite, et des codages que la communauté des chercheurs des LS tend à partager. Ces codages peuvent être sous forme de caractères alphanumériques ou de symboles graphiques, et depuis peu de courbes issues de divers traitements.

Caractères

Les annotations peuvent être faites de caractères qui peuvent être alphabétiques ou numériques, et peuvent représenter plusieurs notions : des gloses, des traductions, des valeurs d'angle, etc. La très grande majorité des annotations effectuées en langues écrites, vocales et signées sont réalisées de cette façon. L'avantage de ce type d'annotation est qu'il est simple à utiliser, et manipulable directement dans des programmes informatiques. Le gros inconvénient est qu'il lie la LS à la forme écrite d'une LV, et cela introduit implicitement des biais de raisonnement : il est difficile de fait de faire abstraction du label écrit en y mettant le concept en LS qu'il désigne, d'autant plus qu'il est parfois délicat de mettre un label, vu que certaines structures ne sont pas dénommées en langue écrite.

Il est possible d'utiliser un code, basé sur des caractères alphanumériques, pour annoter un corpus : on enlève l'inconvénient de reposer sur une langue écrite, mais on s'appuie toujours sur les composants d'une langue écrite qui ne sont pas celles des LS (ces caractères ne représentant pas un alphabet propre aux LS).

Symboles graphiques

Les linguistes travaillant sur les aspects phonologiques, transcription ou écriture de la LS ont eu besoin de développer des systèmes de notation des signes. Certains sont utilisés dans des systèmes d'annotation informatisés. Nous en citerons deux, SignWriting et HamNoSys, selon nous les principaux, et renvoyons à Boutora [BOU08] pour des présentations détaillées et une liste plus complète.

Le système SignWriting, élaboré par Sutton [SUT] est un système d'écriture qui emploie des symboles graphiques pour représenter les formes de la main, les mouvements, et les expressions faciales des langues signées (Figure 3). C'est un « alphabet », une liste de symboles employés pour écrire n'importe quelle langue signée dans le monde. Il a été conçu à partir d'une notation pour les mouvements de la danse. La personne qui note du SignWriting et celle qui le lit doivent répéter mentalement les signes pour les écrire et les lire, car ce système ne permet pas d'accéder directement au sens au contraire de la plupart des systèmes d'écriture existants.



Figure 3 : Exemple de notation d'une histoire en SignWriting (début de « Blanche neige »)

Ce système de notation ne permet pas de déterminer tous les aspects de la LS et se contente de ceux jugés indispensables pour la compréhension du message par le lecteur. Ainsi, les symboles qui composent ce système représentent des ensembles de variations autour d'une valeur, jugées identiques : concrètement, pour la notation d'une configuration des mains par exemple, on considère un ensemble de variations autour de cette configuration comme identiques et notées par un même symbole. Pour cette raison, nous ne retiendrons pas SignWriting comme système d'annotation pour notre recherche.

Le système de notation de langue de signe de Hambourg (Hamburg Notation System, HamNoSys) [PRI89] est un système phonétique (Figure 4), dont les symboles sont en grande partie iconiques, et destiné à permettre la notation détaillée de toutes les LS ([BOU08]).

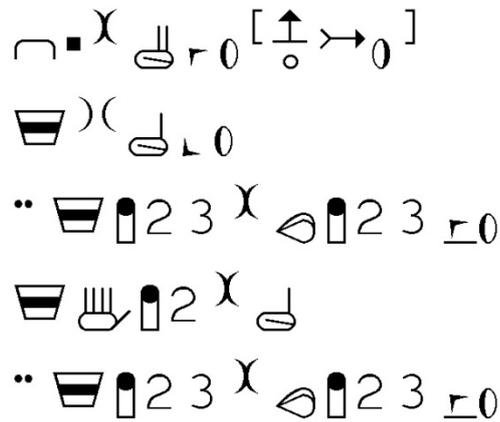


Figure 4 : Exemple de notation HamNoSys

Cependant, dans le cadre de notre étude, nous ne retiendrons pas non plus HamNosys pour la même raison que SignWriting : il pourrait selon nous induire une disparition de certaines variations. En effet, il y a un cadre qui fixe si tel articulateur avec telle valeur doit être représenté avec tel symbole : par extension, on voit bien qu'il se peut que différentes valeurs d'articulations soient notées par le même symbole, ce qui nous empêcherait de faire ressortir les variations. Or nous ne pouvons gérer nous même ce cadre, l'amender ou autre, sans remettre en cause le système en lui-même.

Dans Chetelat-Pele10 [CHE10], l'auteure a fait le choix de l'utilisation de symboles graphiques au lieu de mots afin d'aider à la perception visuelle des structures récurrentes à l'annotation avant tout traitement informatique. Dans son système, un symbole peut décrire plusieurs phénomènes, par exemple l'épaisseur et la couleur d'un symbole « flèche » varient en fonction de l'amplitude d'un mouvement, quand sa forme varie en fonction du plan du mouvement (pointue pour le plan horizontal et vertical, arrondie pour le plan sagittal) (Figure 5). Ce premier type de représentation des sourcils est particulièrement efficace pour l'analyse de par sa simplicité et la mise en évidence visuelle de phénomènes à l'aide de couleur.



Figure 5 : Visualisation aisée des phénomènes grâce à une annotation par symboles visuels

Dans Segouat [SEG03], nous nous étions intéressés au domaine de la création d'icône car les images que nous voulions concevoir devaient être le plus intuitives possibles et donc fortement iconiques. Nous avons arrêté notre choix sur une représentation graphique qui permettrait de mieux appréhender les différentes composantes de l'espace : nous avons essayé de dégager deux formalismes textuels et de créer un formalisme graphique afin de représenter plus intuitivement certains aspects de l'espace en LS. Le formalisme graphique proposé reprenait une segmentation spatiale en cube communément admise et proposait plusieurs versions pour la visualisation des phénomènes (Figure 6, la portion de l'espace symbolisée est la zone située au niveau des épaules, à une distance atteinte par les coudes du locuteur lorsqu'il a les bras tendus, respectivement sur la droite – vert –, devant – orange –, et à gauche – rouge – du locuteur).



Figure 6 : Différentes propositions de représentation de la segmentation de l'espace de signation

Des deux systèmes de symboles de notations présentés ci-dessus, seul le premier a connu une réelle mise en application, quand le second a constitué une première étape de réflexion sur une étude de segmentation de l'espace.

Courbes

Dans Chetelat-Pele [CHE10], l'auteure a réalisée une annotation des mouvements des sourcils par repérage directement sur la vidéo. Elle a noté, image par image, des positions précises sur les sourcils, ce qui a abouti à autant de courbes décrivant le mouvement que de points annotés (Figure 7). Les données issues de ces courbes ont été traitées au moyen d'un outil statistique (Scilab), mais le visuel dans le logiciel d'annotation permet déjà de se rendre compte de variations générales ou indépendantes de certaines courbes, et de faire des premières corrélations avec les autres valeurs (dessins ou caractères, repérés par des couleurs).

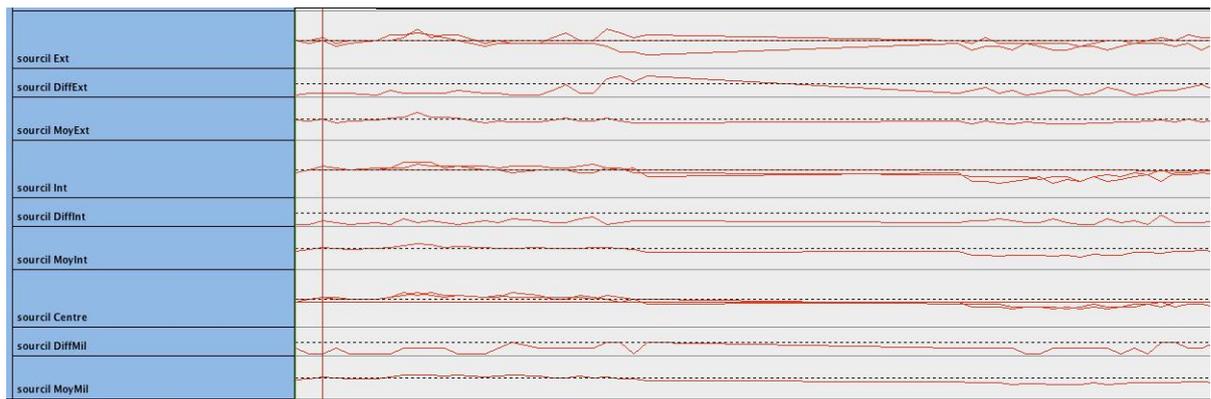


Figure 7 : Visualisation sous Anvil des courbes issues d'annotations sur la vidéo

Dans Segouat [SEG06], nous avons combiné deux approches : une annotation linguistique par caractères et une annotation informatique par courbes. Les courbes sont issues d'un traitement d'images réalisé par Cassel [CAS05], qui a utilisé des algorithmes de détection et suivi du mouvement mis au point pour le trampoliner. Les courbes obtenues représentent les positions (X,Y) et la taille (L,H) de la boîte dite « englobante » qui entoure le signeur (Figure 8). De la même manière que pour Chételat-Pele [CHE10], il est alors possible d'obtenir des premiers résultats visuellement en observant les variations des courbes et les variations de valeurs, pour peu que ces dernières soient identifiables aisément (donc en couleur et/ou en image, plutôt que sous forme de caractères).

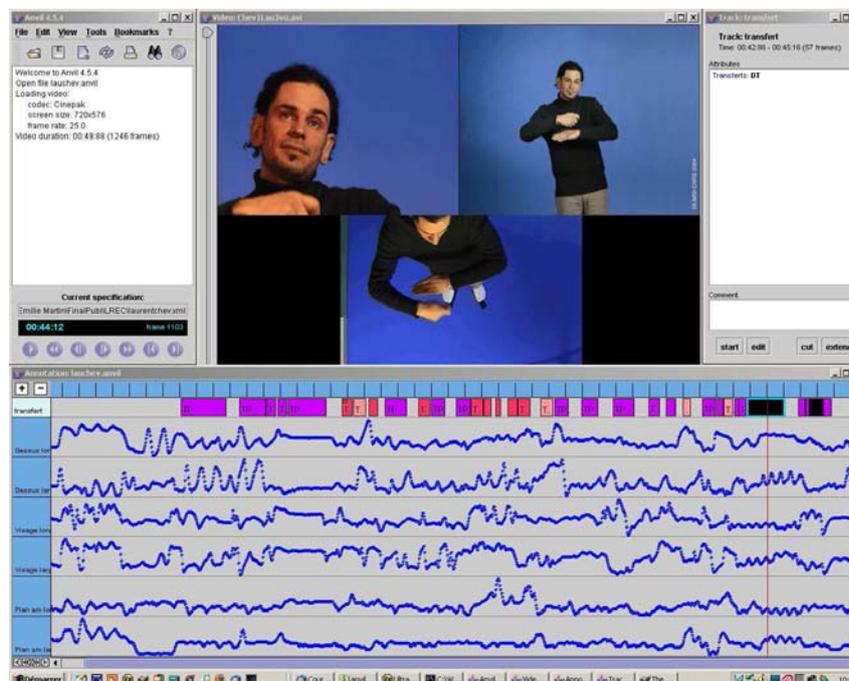


Figure 8 : Visualisation sous Anvil des courbes issues d'algorithmes de traitement d'images

Les solutions d'utilisation de courbes n'ont pas été retenues dans le cadre de la présente recherche puisque nous ne nous intéressons pas au mouvement. Cependant, il nous semble important de les mentionner puisque c'est un type d'annotation apparu récemment, mais qui va selon nous prendre de l'ampleur puisqu'il y a de plus en plus d'analyses de corpus de capture de mouvements dans le domaine des LS (ainsi ces courbes pourront être obtenues de manière automatique).

c. Nos choix pour l'annotation

Nous avons choisi de nous intéresser à la direction du regard et aux configurations des mains. En effet, vu le contexte de notre étude, à savoir un corpus construit suivant un ensemble fermé de règles grammaticales et de lexique, les paramètres de mouvement et d'emplacement ont été stéréotypés : puisque le but du projet dans lequel est inscrit notre recherche est de diffuser des informations au public, il y a eu un travail de formatage de cette information. Ce formatage a pris la forme d'un découpage de l'espace afin de toujours faire figurer, par exemple, les gares d'arrivée et de destination aux mêmes emplacements. De même, les mouvements des mains ont été formatés suivant les emplacements précédemment définis. Reste le paramètre de la configuration qui bien que lui aussi normalement formaté, est celui qui selon nous était le plus riche en informations de variations vu le nombre d'articulateurs internes (tous les doigts).

Dans la suite de ce manuscrit, nous désignerons par « composante » la configuration d'une main et la direction du regard, plutôt que « paramètre ».

Les systèmes de notation existants, et les possibilités offertes par les actuels outils d'annotation nous permettent de disposer d'un choix assez large pour choisir nos valeurs d'annotations. Nos besoins sont de pouvoir différencier des éléments suivant des critères de précision qui nous sont propres : étant donné que nous ne savons pas a priori quelles variations sont pertinentes et lesquelles ne le sont pas, nous souhaitons pouvoir annoter librement, grâce à des ensembles de valeur non prédéfinis, ou alors modifiables.

Ainsi, au vu de toutes ces considérations, nous avons décidé de constituer notre propre ensemble de valeurs afin de gérer directement le niveau de finesse de nos annotations pour ce qui est des configurations des mains et de la direction du regard, et d'utiliser un code couleur pour ce qui concerne la fermeture et l'ouverture des yeux (nous ne conservons pas ce dernier composant dans le cadre de notre étude, le réservant à des analyses ultérieures).

Les valeurs d'annotation de la direction du regard consistent en des icônes issues de capture d'écran permettant de visualiser le regard dans une certaine direction, quand celles des

configurations des mains sont des icones permettent de visualiser une configuration donnée. Nous avons au départ puisé dans la liste des configurations proposées par Cuxac [CUX00]. Rapidement nous avons opté pour nos propres valeurs, celles de l'étude de Cuxac ne permettant pas d'annoter toutes les configurations que nous avons rencontrées, ainsi que les différentes variations que nous avons besoin de mettre en exergue (plus de détails dans ce même chapitre, partie Mise en œuvre de la méthode).

Nous avons également dû déterminer ce qui représentait pour nous une valeur d'annotation, c'est-à-dire déterminer dans le continuum temporel le début et la fin d'une certaine valeur. Pour tous les composantes, nous avons considéré le moment où le regard se fixait sur un emplacement, ou que la configuration était statique (l'orientation, le mouvement et/ou l'emplacement pouvait être modifié) pour déterminer le début d'une valeur, la fin étant le déplacement du regard où le changement de la configuration.

Un cas particulier est le suivi des mains par le regard, où la direction absolue est modifiée (le regard peut alors traverser l'espace de signation) mais la direction relative n'est pas modifiée (le regard est toujours dirigé vers les mains). De même pour les configurations dynamiques, nous avons essayé d'isoler les parties statiques quand il y en avait, sinon nous avons annoté avec un symbole donc nous savons qu'il représente une configuration dynamique.

2. Outils d'annotation

a. Définition

Un outil d'annotation est un logiciel qui va permettre d'annoter, et de stocker ces annotations suivant une logique de stockage interne (comme dans une base de données relationnelle par exemple). Cet outil peut être générique, donc utilisé à d'autres fins beaucoup plus générales que celles de l'annotation, ou spécifique, souvent à un domaine donné ou à un ensemble défini de domaines. Cet outil doit permettre d'annoter un signal (de parole, de vidéo, etc.) avec au moins une modalité citée : caractères, images, courbes, etc. Il autorise enfin plus ou moins de partitionnement et de hiérarchie dans les différents ensembles d'annotation.

b. Nos besoins

Nous avons besoin d'un outil qui nous permette non seulement d'annoter le corpus réalisé, mais éventuellement de pouvoir traiter ces annotations (que ces traitements soient internes à l'outil, ou externes). Nous allons détailler plus avant ces besoins, puis nous passons brièvement en revue quelques outils avant de présenter plus spécifiquement celui sur lequel notre choix s'est porté.

Il faut que l'outil d'annotation que nous choisissons nous permette de :

- lire une vidéo,
- annoter d'après cette vidéo plusieurs composantes avec différentes valeurs :
 - o ces valeurs doivent pouvoir être du texte ou des symboles graphiques (dessins),
 - o ces annotations doivent pouvoir être visuellement analysables (c'est-à-dire qu'on puisse, en visualisant les annotations, avoir la possibilité de remarquer certains phénomènes, préalablement à un traitement statistique),
- pouvoir utiliser les annotations dans des programmes de calculs statistiques (ce qui implique que les traitements soient incorporés dans l'outil, ou qu'il y ait une possibilité d'export des annotations vers d'autres programmes).

Fort de la connaissance de ces besoins, nous avons donc orienté notre recherche d'un outil d'annotation dans les domaines des LS et de la multimodalité.

c. L'existant

Il existe quelques outils d'annotation qui ont été développés pour les études sur les LS. Ces outils s'inspirent des outils d'annotation de corpus multimodaux, développés pour les études sur le geste et la voix notamment. Parmi les outils existants, nous présentons tout d'abord un outil conçu dans le domaine des LS, puis deux outils issus du domaine multimodal. Nous ne prétendons pas être exhaustif, mais avons considéré iLex, du domaine des LS, parce qu'il a été conçu en vue d'études linguistiques sur corpus et lexicographiques (contrairement à d'autres qui sont plus anciens et qui manquent de possibilités, comme SignStream [NEI00]), et nous nous sommes intéressés à Elan et Anvil, du domaine multimodal, puisque ces deux outils (surtout le premier) sont utilisés dans le cadre de recherches sur les LS. Les trois outils présentés ont tous des principes de fonctionnement communs : des annotations sont placées sur une grille d'annotation suivant le temps de lecture d'une vidéo et/ou d'un signal audio. La grille d'annotation est composée de pistes, qui regroupent un ensemble d'annotations portant sur une même composante.

iLex, développé par Hanke [HAN08] est un outil construit sur une base de données qui a pour objectif initial l'analyse de vidéos de LS. La base de données consiste en un ensemble d'énoncés, chacun associé à une séquence vidéo et sa transcription. Cet outil permet notamment la recherche dans la base de données pour l'affichage en parallèle de différentes occurrences d'un même phénomène. Il a été développé dans le cadre d'un grand projet de recherche linguistique sur l'ASL, dans le but de créer une grande base de données d'énoncés ASL annotés. Les avantages de cet outil sont qu'il autorise l'annotation avec le système HamNoSys, qu'il permet d'associer des « tokens » (instances) à des « types » (formes génériques), qu'il facilite le travail collaboratif grâce à son

architecture client/serveur, et qu'il dispose d'une base de données (donc de tous les avantages de pouvoir faire des requêtes sur cette base plus facilement que si les données sont stockées sous forme de fichiers). Les inconvénients sont entre autres qu'il n'est pas possible d'annoter avec d'autres valeurs que des caractères ou des symboles HamNoSys (donc pas d'images), que le visuel n'est pas pratique parce que notamment le rendu d'un élément d'annotation n'est pas proportionnel au temps (un élément d'une durée courte aura le même visuel qu'un élément de durée longue, il faudra lire les temps de début et de fin de chaque élément pour se rendre compte de la différence), et enfin que certaines manipulations des annotations sont délicates (notamment les modifications).

Rohlfing [ROH06] propose un état de l'art des différents outils d'annotation multimodaux. Nous proposons au lecteur de se reporter à la référence précédente pour plus de détails, cependant que nous nous concentrerons sur la présentation des outils Elan et Anvil, puisqu'ils sont également utilisés dans le domaine des LS.

Elan est un outil développé par l'équipe du Max Planck Institute [CRA08]. Cet outil permet d'annoter avec des caractères alphanumériques ou des courbes (issues généralement d'un enregistrement audio). Les différentes pistes d'annotation peuvent être liées entre elles, tout comme les annotations. Les annotations sont stockées dans des fichiers au format XML. Les avantages de cet outil sont qu'il est largement répandu dans la communauté de la recherche en LS (particulièrement chez les linguistes), qu'il propose des outils statistiques intégrés, et qu'il est modifiable et adaptable en fonction des besoins. Un ses inconvénients est l'impossibilité d'annoter avec des images (seulement en caractères Unicode).

Anvil est un outil développé par Kipp [KIP04] dans le cadre de sa thèse et maintenu et enrichi depuis. Il est également multi partitions et offre une visualisation colorée des segments d'annotations : cette possibilité est offerte dans Elan, mais dans une modalité qui permet une moindre visualisation. Concrètement, dans Anvil une couleur peut être attribuée à une valeur, et chaque instance de cet élément (chaque label d'annotation avec cette valeur) sera coloré, alors que dans Elan la couleur est associée à une piste : dans Anvil apparaissent visuellement clairement des informations quantitatives et qualitatives sur les éléments colorés, alors qu'Elan ne permet que de visualiser clairement les pistes entre elles (qui sont déjà présentées séparément et de manière claire, d'où notre conclusion que la couleur n'apporte pas de réel plus).

Les annotations peuvent être de plusieurs natures : caractères alphanumériques, images, courbes issus de traitements d'images ou de paroles, ou d'annotations directement sur la vidéo. Il est

possible de faire des liens entre les pistes, et de manipuler des objets qui ne dépendent pas du temps. Au départ conçu pour des recherches sur les gestes, Anvil est maintenant utilisé dans de nombreux domaines tels que les interactions homme-machine, la linguistique, l'éthologie, l'anthropologie, la psychothérapie, les agents conversationnels, l'animation, l'océanographie, et bien sûr les LS. Les données sont stockées dans des fichiers au format XML. Les avantages d'Anvil sont : il permet l'annotation par des images et symboles graphiques, la visualisation est aisée et permet des analyses qualitatives préliminaires. L'inconvénient est qu'il ne dispose pas d'outil interne d'analyse (le développement des statistiques nécessite de maîtriser le langage informatique Java).

d. Notre choix : Anvil

Etant donné les besoins que nous avons et les possibilités offertes par les outils listés, notre choix s'est naturellement porté sur Anvil qui nous permet une plus grande souplesse dans nos valeurs d'annotations, puisque nous pouvons utiliser des images, et qu'il offre une possibilité de visualisation autorisant le repérage visuel de certains phénomènes (pour qu'ensuite nous les quantifions par traitement statistique). En outre, c'est un outil que nous connaissons bien pour l'avoir utilisé dans différentes études ([SEG03], [SEG06]) et pour lequel nous avons déjà mis au point des traitements statistiques externes (sans réutiliser intégralement ces traitements, nous bénéficions de cette expérience pour d'autres traitements des données d'annotation).

Anvil (Figure 9) est, comme nous l'avons vu, à la base un outil d'annotation multimodale qui a été conçu dans le cadre d'une thèse, puis enrichi au fur et à mesure des années, pour l'analyse des gestes co-verbaux. Il est constitué de 4 parties, le menu principal (A), la fenêtre de visualisation des valeurs des annotations (B), la fenêtre d'affichage vidéo (C) et le tableau d'annotation (D). Les deux parties qui nous intéressent le plus sont la vidéo et les annotations.

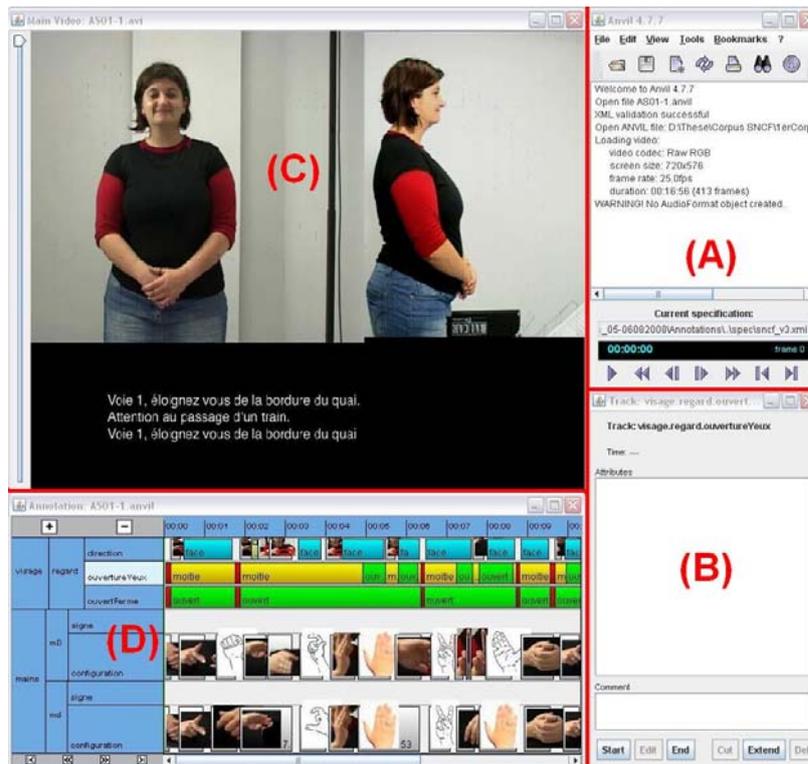


Figure 9 : Anvil, constitué de 4 parties (les 2 principales : (C) et (D))

Le tableau d'annotation illustre le système multipartition du logiciel : il est possible d'annoter sur plusieurs pistes, parallèles, qui peuvent être dépendantes les unes des autres, en relation de hiérarchie ou pas (nous n'avons cependant pas considéré la notion de relation entre les pistes proposée par le logiciel dans notre étude). Les annotations se positionnent à la main dans le tableau suivant un axe temporel gauche droite sur la durée de la vidéo. Elles sont stockées dans un fichier au format XML, dont la lecture est assez facilement accessible. La structure des annotations, ce qui définit ce qu'on annoté, par quoi et comment on l'annoté, est elle aussi stockée dans un fichier XML, (qui était à concevoir et écrire à la main). Nous avons opté pour des valeurs différentes pour le regard et les mains, ce choix étant du à la nature différentes des phénomènes annotés, cependant nous avons profité de la possibilité d'annoter avec des images sous Anvil ces composantes corporelles.

3. Synthèse des choix effectués

Nous avons décidé de constituer notre propre ensemble de valeurs sous forme d'images (captures d'écran) pour ce qui est des configurations des mains et de la direction du regard, et d'utiliser un code couleur pour ce qui concerne la fermeture et l'ouverture des yeux.

Pour toutes les composantes, nous avons considéré le moment où le regard se fixait sur un emplacement, ou que la configuration était statique (c'est-à-dire que l'orientation, le mouvement et/ou l'emplacement pouvait être modifié) pour déterminer le début d'une valeur, la fin étant le déplacement du regard où le changement de la configuration.

Etant donné les besoins que nous avons et les possibilités offertes par les outils listés, notre choix s'est porté sur Anvil (outil d'annotation multimodale) qui nous permet une plus grande souplesse dans nos valeurs d'annotations, puisque nous pouvons utiliser des images, et qu'il offre une possibilité de visualisation autorisant le repérage visuel de certains phénomènes (pour qu'ensuite nous les quantifions par traitement statistique). Anvil nous permet d'annoter suivant nos critères les vidéos de notre corpus.

III. Outils et méthodes d'analyses

Nous devons analyser les annotations effectuées précédemment, afin de proposer ensuite des pistes pour la conception d'un modèle de coarticulation. Concrètement, l'analyse porte sur les variations observées entre des composantes de LS selon que les signes sont contextualisés ou non. Nous souhaitons donc pouvoir faire des comparaisons entre les valeurs des composantes et leurs durées, puis quantifier et qualifier ces comparaisons. Afin de réaliser ces comparaisons, nous nous sommes intéressés aux algorithmes d'alignement du domaine de la linguistique de corpus et de la bioinformatique, et pour les statistiques nous nous sommes servis des calculs standards de moyenne et de pourcentages.

Nous précisons tout d'abord nos besoins, puis nous présentons différents algorithmes d'alignement dans les domaines cités ci-dessus, et explicitons enfin notre choix.

1. Besoins

Nous disposons d'un certain nombre d'annotations, stockées sous forme de chaînes de caractères dans un fichier spécifié au format XML. Il nous faut maintenant traiter ces données pour établir notre modèle suivant les résultats de ces analyses. Nous voulons comparer nos annotations suivant des critères de temps et de valeur. Il nous faut donc disposer d'un outil qui soit capable de réaliser cette comparaison et d'un outil qui nous permette de croiser différents points de vue sur cette comparaison. Pour cette deuxième étape, nous utilisons un tableur, pour des calculs de moyenne et de pourcentage sur les différents points de vue adoptés. Pour la première étape, il nous faut effectuer un alignement de nos annotations : le principe de l'alignement est de faire ressortir les points communs (et incidemment les différences) entre les deux éléments alignés. L'alignement est

au final réalisé sur deux ensembles d'éléments qui correspondent à la forme isolée et à la forme en contexte d'une composante (direction du regard ou configuration d'une main).

2. Outils et méthodes d'alignement

Il existe des outils d'alignement en linguistique de corpus qui permettent d'aligner des chaînes de caractères, que ce soient des mots, des phrases ou des textes. Cependant les algorithmes utilisés nécessitent des informations de nature linguistiques propres à la modalité écrite de la langue, et propre à la langue considérée. Il existe également des outils d'alignement dans le domaine de la bioinformatique, où il est question d'aligner des séquences de différentes natures (mais qui sont des chaînes de caractères à la base), avec des considérations qui n'ont rien de linguistique en soi. Enfin, il existe dans le domaine du geste une utilisation de l'algorithme DTW (« Dynamic Time Warping », déformation temporelle dynamique) sur la LSF : dans [HEL06] les auteurs présentent une analyse de la variabilité stylistique (à noter que pour cet algorithme en particulier, les échantillons des séquences doivent être équidistants en temps).

Ces domaines utilisent des algorithmes dont la base est la programmation dynamique : elle garantit un coût minimal pour un alignement optimal, ou un alignement optimal pour un coût minimal. Cette méthode est dite « de résolution ascendante » parce qu'elle détermine une solution optimale du problème à partir des solutions de tous les sous-problèmes (d'où le terme « ascendant ») et permet de résoudre de nombreux problèmes dont la solution directe n'est pas possible puisque de complexité exponentielle. Ce qui est souvent recherché avec les méthodes existantes est la plus longue sous chaîne commune aux deux chaînes. Pour cela, des coûts de similarité, d'insertion/suppression et de substitution sont affectés lors du déroulement de l'algorithme de comparaison. Ces coûts sont déterminés de manière ad hoc, et il n'existe à ce jour pas de méthodologie objective qui évalue la validité des coûts choisis, ou ce que les coûts choisis impliquent pour la relation entre les chaînes.

a. Linguistique de corpus

Définitions

Selon Allison [ALL89], la question de l'alignement est de savoir si deux chaînes de caractères sont alignées, et le cas échéant, de quelle manière. En effet, deux chaînes peuvent toujours avoir une relation de similarité selon le point de vue, mais c'est le modèle de relations qui va permettre de conclure sur la nature des relations entre ces deux chaînes.

Selon Ballim [BAL96], le travail d'un programme d'alignement est d'établir un isomorphisme entre deux textes à un certain niveau de détails, en d'autres termes, de déterminer des corrélations

entre deux textes. Dans cette étude, le programme prend en entrée des textes qui sont des traductions l'un de l'autre. Chaque texte contient un certain nombre de repères quant à la structure, par exemple des informations sur les limites de phrases et de paragraphes. Le programme fournit en résultat des ensembles de parties alignées entre les deux textes. Dans l'outil qu'il présente, Multext v2.0, l'auteur explique que l'algorithme consiste à aligner deux arbres : les documents sont donc considérés comme ayant une telle structure, et l'objectif est d'aligner les structures entre elles. Pour ensuite comparer les éléments, l'auteur fait référence à une méthode telle que celle proposée par Gale93 [] qui consiste à compter le nombre de caractères dans deux régions (i.e. des feuilles dans la structure d'arbre) et les aligner en s'appuyant sur des notions de fréquence et proportionnalité entre les textes en considérant une distribution probabiliste.

Mises en œuvre

Dans le domaine de l'alignement de corpus de textes parallèles, qui sont des textes traduits en une ou plusieurs langues dont on essaie le plus souvent d'extraire des traductions de langue à langue, il existe selon Véronis [VER00a] plusieurs niveaux d'alignement :

- les structures des textes,
- les phrases,
- les mots et les expressions.

Au niveau de la phrase, il existe deux grands axes de travail sur l'alignement, dont l'auteur cite selon lui les fondateurs (ces études portent toutes sur la langue américaine).

D'un côté Kay et coll. [KAY93] partent du principe que les mots doivent se correspondre. Leur algorithme débute en sélectionnant des phrases potentiellement alignées, en partant de l'hypothèse que les premières et dernières phrases ont de grandes chances de se correspondre, ainsi que celles du milieu (mais pas obligatoirement toutes celles du texte strictement, il est ici question de longs textes) : il y a alors une sorte de fractionnement du texte sur lequel se déroule la seconde étape de l'algorithme. Il s'agit alors de comparer la distribution des mots dans les phrases en partant de l'hypothèse qu'une distribution identique de deux mots dans deux phrases de deux langues implique une forte probabilité pour ces deux mots d'être une traduction l'un de l'autre.

D'un autre côté, Gale et coll. [GAL91] et Brown, Lai et Mercer 91 [BRO91] partent du principe que les phrases d'un texte qui est la traduction d'un autre auront la même longueur d'un texte à l'autre : les phrases courtes sont traduites par des phrases courtes, les phrases longues par des phrases longues. Ils considèrent aussi que le ratio du nombre de mots et du nombre de caractères dans les textes de chaque langue sont les mêmes : les textes français sont ainsi plus long (1,1 fois, selon Véronis et coll. [VER00b]) que leurs traductions en anglais. Est ensuite utilisé un modèle probabiliste qui utilise, dans Gale et Church, une mesure de la dissimilarité des phrases (le but étant qu'elle soit

minimale) basée sur le nombre des caractères, et dans Brown, Lai et Mercer, un modèle de Markov caché pour mesurer la longueur des phrases, basé sur la longueur des mots.

Au niveau des mots et des expressions, les algorithmes utilisés le sont souvent pour l'extraction de lexique. Il s'agit tout d'abord de détecter les mots et expressions dans les textes, ensuite de les aligner. Pour cela, sont utilisés des méthodes statistiques qui rencontrent des problèmes par rapport aux mots composés, et aux mots dits « fonction » (par opposition aux mots « contenus ») entre autres. A ces méthodes sont alors ajoutées des connaissances linguistiques sur par exemple la grammaire, mais cela complexifie et alourdit considérablement les traitements.

Au niveau de la structure des textes, Véronis [VER00a] fait état que presque toutes les méthodes utilisées font les hypothèses suivantes au niveau de la structure globale et des paragraphes dans les textes sources et cibles : l'ordre est le même, il y a peu d'addition ou d'omission, et les alignements sont en majorité un pour un (sinon, cela ne dépasse pas deux pour un).

Notre besoin en termes d'algorithme d'alignement ne peut pas tenir compte de connaissances a priori (linguistiques) ni poser de fortes hypothèses (structurelles ou probabilistes). Il nous faut donc voir ailleurs si d'autres algorithmes sont plus adaptés, ou a minima adaptables, notamment en bioinformatique.

b. Bioinformatique

Les outils développés dans le cadre de la bioinformatique l'ont été notamment pour aligner des séquences ensemble¹⁶. Ces outils utilisent différents algorithmes mais nous citerons un des plus utilisés : l'algorithme de programmation dynamique. Le principe consiste à découper le problème en sous problèmes plus simples à résoudre. Dans le cadre de l'alignement de séquences, il existe deux grands courants méthodologiques pour l'alignement avec cet algorithme : l'alignement local ou l'alignement global. Les alignements globaux sont plus souvent utilisés quand les séquences mises en jeu sont similaires et de taille égale. Une technique générale, appelée algorithme de Needleman-Wunsch est basée sur la programmation dynamique. Les alignements locaux sont plus souvent utilisés quand deux séquences dissemblables sont soupçonnées de posséder des motifs semblables malgré l'environnement. L'algorithme de Smith-Waterman est une méthode d'alignement local générale basée aussi sur la programmation dynamique.

¹⁶ Une liste des outils d'alignement de séquences est disponible ici : http://en.wikipedia.org/wiki/Sequence_alignment_software

Dans notre cas, nous nous intéressons aux alignements locaux puisque nous alignons des séquences d'éléments dont nous ne sommes pas certains que la taille et les ensembles de valeurs soient « similaires ». Nous détaillons maintenant l'algorithme de Smith-Waterman que nous avons choisi pour réaliser nos calculs d'alignement.

Comme nous l'avons signalé, cet algorithme est dit local car il ne va pas aligner les séquences données de manière globale, mais chercher à aligner toutes les subdivisions possibles des deux séquences entre elles et proposer le meilleur alignement selon un taux de similarité.

Il faut disposer initialement de deux séquences qu'on voudrait comparer, et de valeurs ou scores de similarité, suppression/insertion et de substitution des éléments de chaque séquence. Les valeurs de similarité sont regroupées au sein d'une matrice dite de similarité, qui permet de faire la correspondance entre deux éléments : lorsque l'algorithme a besoin de déterminer si un élément est identique à un autre, il va consulter la matrice de similarité qui lui fournit une valeur prédéterminée (cette valeur peut être égale à 0 si les éléments sont similaires, et 1 si les éléments sont non similaires, mais ces valeurs sont laissées au bon soin de la personne qui réalise ses alignements : en fonction de l'éclairage qu'elle voudrait apporter sur la similarité de certains éléments, et en fonction de sa définition d'éventuels degrés de similarité, elle définit ses propres valeurs). Les scores de suppression / insertion et de substitution correspondent à un « coût » : en bioinformatique on considère que l'insertion, la suppression et la substitution d'un élément par un autre (quelque soit l'élément) peut avoir pour conséquence une plus grande dissimilarité entre les séquences. Un coût est alors ajouté, ce coût étant une valeur négative elle est en fait retranchée, au score final de similarité entre les séquences. Au final, nous obtenons un score de similarité qui tient compte des éléments qui ont pu être alignés car considérés comme similaires, et des éléments qui n'ont pu être alignés car ajoutés, supprimés ou substitués.

Si nous disposons d'un ensemble d'éléments « A », « B », « C », « D » qui sont les constituants des séquences que nous souhaitons aligner, nous concevons une matrice de similarité qui donne une valeur de 0 pour toute comparaison entre deux éléments sauf avec lui-même. La matrice est donc de cette forme :

	A	B	C	D
A	1	0	0	0
B	0	1	0	0
C	0	0	1	0
D	0	0	0	1

Ce type de matrice, dont les valeurs dans la diagonale sont toutes égales à 1 et le reste égal à 0, est appelé matrice d'identité. Cette matrice est choisie généralement lors de la première tentative d'alignement. Une fois que des premiers résultats issus d'analyses des alignements sont vérifiés, une autre matrice de similarité est conçue, prenant en compte ces premiers résultats (et permettant de raffiner la notion de similarité et d'introduire des degrés de similarité en considérant plus que deux valeurs).

Le calcul d'alignement est réalisé par la création d'une matrice d'alignements : dans un premier temps tous les alignements possibles entre les éléments et groupes d'éléments des séquences sont déterminés dans cette matrice, puis le meilleur alignement, suivant le score obtenu, permet de désigner le meilleur alignement entre les deux séquences. Les valeurs de cette matrice d'alignement sont obtenues en tenant compte des valeurs de similarité des éléments comparés, ainsi que des éléments précédemment alignés. Les scores d'insertion, suppression et substitution sont utilisés lorsque tous les alignements potentiels sont réalisés, afin de déterminer celui qui aura le score le plus intéressant pour l'étude. C'est pour cela que l'algorithme est dit à résolution ascendante, puisque dans un premier temps il éclate le problème de l'alignement des deux séquences en plusieurs problèmes d'alignement de séquences plus simples (parce qu'uniques) et que dans un second temps il remonte au problème global en prenant en compte tous les calculs précédemment faits.

En détail, voici comment se déroule l'algorithme.

Admettons que nous souhaitons aligner les séquences « ABCBA » et « ABDB », que notre matrice de similarité soit la matrice identité avec une valeur de « 2 » dans la diagonale, et que nous ayons un score de « -1 » pour les suppressions, insertions et substitutions.

La matrice d'alignement sera initialement celle-ci :

	*	A	B	C	B	A
*	0	0	0	0	0	0
A	0					
B	0					
D	0					
B	0					

On constate que les deux séquences sont réparties horizontalement et verticalement, et que la première ligne et la première colonne sont remplies de 0, ceci pour pouvoir déterminer les valeurs dans les autres cases (en effet, pour déterminer la valeur d'une case il faut connaître la valeur des

cases précédentes dans le parcours de la matrice, soit les cases en diagonale en haut à gauche, à gauche, et en haut, par rapport à la case considérée).

Pour chaque case, on utilise la formule suivante, où « Valeur » correspond à la valeur de la case qu'on veut calculer :

$$\text{Valeur} = \max (0, \text{MatAlign}(i-1,j-1)+\text{MatSimil}(i,j), \text{MatAlign} (i-1,j)- \text{MatSimil} (i,*), \text{MatAlign} (i,j-1)- \text{MatSimil} (*,j))$$

où « i » est un indice de parcourt des colonnes, « j » est un indice de parcours des lignes, « MatAlign » est la matrice d'alignement (en cours de construction), « MatSimil » est la matrice de similarité, et « MatSimil (*,j) » ou « MatSimil (i,*) » est la valeur de pénalité en cas de décalage de l'alignement.

Au final, on obtient la matrice d'alignement complétée suivante :

	*	A	B	C	B	A
*	0	0	0	0	0	0
A	0	2	1	0	0	0
B	0	1	2	1	2	1
D	0	0	1	1	1	1
B	0	0	2	1	3	2

A partir de cette matrice complétée, l'algorithme consiste à partir de la valeur la plus haute et de parcourir les valeurs de la matrice, soit en remontant d'une case, soit en se décalant d'une case vers la gauche, soit d'une case en diagonale en remontant sur la gauche. La case est choisie entre les trois citées suivant la valeur la plus haute, en privilégiant la diagonale. Dans notre cas, le parcours sera suivant les valeurs 3 (valeur de départ), 1 (en diagonale), 2 (en diagonale), 2 (en diagonale). Au cours de ce parcours inverse (par rapport au sens de remplissage de la matrice), on considère qu'un passage d'une case à une autre en diagonal indique un alignement (similaire ou en substitution), d'une colonne à une autre (donc vers la gauche) indique une suppression, et d'une ligne à une autre (donc en remontant) indique une insertion. Dans notre exemple, on obtient l'alignement suivant « AB(D/C)B- », signifiant que les deux premiers éléments de chaque séquence sont alignés, que les troisièmes sont substitués, que les quatrièmes éléments sont alignés, et qu'il y a une insertion en fin de séquence (matérialisé par le symbole « - »).

Dans cet exemple, nous utilisons un vocabulaire simple limité à 4 caractères pour que notre explication soit la plus claire possible quand au fonctionnement de l'algorithme. Appliqué à notre étude, cet algorithme doit pouvoir traiter un vocabulaire d'entrée plus important et plus complexe qu'un simple caractère, vocabulaire constitué de nos différentes valeurs d'annotations.

c. La méthode retenue

L'algorithme présenté nous intéresse dans la mesure où il permet de réaliser des alignements avec toutes les parties d'une séquence, pas seulement avec la séquence elle-même. Ainsi, si les variations entre deux séquences sont importantes et qu'il n'y a qu'une petite partie d'éléments alignés, nous auront l'alignement en résultat du calcul.

Cependant nous allons devoir l'adapter à notre étude, dans la mesure où dans son implémentation d'origine il traite des caractères uniques (ceci est dû à la nature des données normalement traitées par l'algorithme, qui raisonne soit sur des caractères, soit sur des triplets de caractères, qui représentent des éléments du génome).

Ce qui est intéressant dans les méthodes et outils d'alignement de séquences en bioinformatique, et dans l'algorithme de Smith-Waterman présenté, c'est que l'alignement peut être en partie naïf et ne pas solliciter de relation, entre les éléments au contraire des méthodes d'alignement de textes. Notre choix se porte donc sur cet algorithme, avec la contrainte de pouvoir traiter nos annotations, qui sont des chaînes de caractères enregistrées dans des fichiers, plutôt que des caractères uniques (au sens de « un seul caractère »).

Nous avons récupéré une implémentation en langage Java de l'algorithme de Smith-Waterman [AHM], et plutôt que de modifier l'algorithme en lui-même, nous avons préféré modifier les données d'entrées pour faire en sorte de fournir à l'algorithme des caractères uniques. Pour ce faire, nous transposons les valeurs d'annotations sous forme de chaînes de caractères en caractère unique, en établissant une bijection entre l'ensemble des valeurs d'annotation et la norme de codage ASCII. Ainsi, nous obtenons un codage de chaque valeur d'annotation par un caractère unique, que nous insérons dans l'algorithme, et en sortie nous effectuons la correspondance inverse : depuis le caractère ASCII vers la valeur d'annotation réelle.

Le fait de ne pas avoir modifié l'algorithme pour ne prendre en compte que nos valeurs nous permet de pouvoir élargir l'ensemble des différentes valeurs d'annotation (i.e., si nous ajoutons des valeurs possibles, elles seront-elles aussi traitées dans l'algorithme sans qu'il y ait besoin de le modifier à chaque ajout), jusqu'à une limite qui est celle du nombre de caractère ASCII scriptables (au nombre de 94).

IV. Conclusion

Du bref historique des collections de données qui ont pu être réalisées sur la LS, il ressort que la notion de corpus dans le domaine de la LS suit peu ou prou les définitions existantes dans le domaine de l'écrit et dans celui de la parole. Il s'agit de collecter des données suivant un certain objectif de recherche, suivant des critères prédéfinis (de locuteur, de contenu, de moyen de captation, entre autres), en y ajoutant des annotations suivant la recherche poursuivie.

Nous avons au départ plusieurs contraintes dues au contexte de notre recherche, en ce qui concerne la conception de notre corpus, qui portent sur la grammaire et le vocabulaire. Nous avons décidé de faire réaliser des tâches d'expressions d'énoncés dits complets et d'énoncés isolés, afin de les comparer et de tirer des conclusions quant à la variation de certaines composantes en contexte. Nous suivons le principe de corpus-based : nous ne cherchons pas à vérifier d'hypothèse, mais à faire ressortir des phénomènes qui nous permettront ensuite de proposer un modèle.

Les systèmes de notation existants, et les possibilités offertes par les actuels outils d'annotation nous permettent de disposer d'un choix assez large pour choisir nos valeurs d'annotations. Etant donné que nous ne savons pas a priori quelles variations sont pertinentes et lesquelles ne le sont pas, nous souhaitons pouvoir annoter librement, grâce à des ensembles de valeur non prédéfinis, ou alors modifiables. Ainsi, nous avons décidé de constituer notre propre ensemble de valeurs afin de gérer directement le niveau de finesse de nos annotations pour ce qui est des configurations des mains et de la direction du regard, et d'utiliser un code couleur pour ce qui concerne la fermeture et l'ouverture des yeux.

Etant donné les besoins que nous avons et les possibilités offertes par les outils listés, notre choix s'est naturellement porté sur Anvil qui nous permet une plus grande souplesse dans nos valeurs d'annotations, puisque nous pouvons utiliser des images, et qu'il offre une possibilité de visualisation autorisant le repérage visuel de certains phénomènes (pour qu'ensuite nous les quantifions par traitement statistique). En outre, c'est un outil pour lequel nous avons déjà mis au point des traitements statistiques externes.

Pour l'analyse des annotations réalisées, nous avons opté pour l'algorithme de Smith-Waterman, utilisé pour l'alignement de séquences en bioinformatique. Cet algorithme nous permet de nous concentrer sur les valeurs des éléments et de comparer ces valeurs pour déterminer les variations entre la forme en contexte et la forme isolée d'un élément.

Notre choix méthodologique se base sur toutes ces constatations, sur les contraintes du projet SNCF, et sur l'expérience que nous avons des analyses de corpus. Nous souhaitons suivre la

méthodologie suivante : réaliser un corpus, annoter ce corpus, analyser ces annotations, et enfin tirer des conclusions qui nous permettront de bâtir notre modèle de la coarticulation en LSF pour le pilotage d'un signeur virtuel. Ces différentes étapes sont présentées dans le chapitre suivant.

V. Mise en œuvre de la méthode

Dans cette partie nous présentons les différentes étapes de la méthodologie que nous avons mise en place. La première étape consiste à concevoir le contenu du corpus, à partir des énoncés écrits et des ensembles de valeurs (noms, chiffres, etc.) fournis par la SNCF, et à réaliser ce corpus avec une locutrice prédéterminée. La seconde étape est celle de l'annotation du corpus, suivant des choix spécifiques, avec un outil d'annotation adapté. Le traitement statistique de ces analyses est la troisième et dernière étape de notre méthodologie, grâce à un algorithme basé sur l'alignement de séquences, issu du domaine de la bioinformatique.

1. Etape 1 : sélection du locuteur, conception et réalisation du corpus

La première étape de notre méthodologie consiste à déterminer qui sera notre « locuteur » de LSF, quelles données nous désirons avoir dans notre corpus, et comment réaliser notre corpus en fonction de l'objectif de notre étude.

a. Sélection du locuteur

L'objectif du projet dans lequel s'inscrit notre étude est de fournir des informations en LSF via une signeuse virtuelle. Cette signeuse virtuelle est créée par rotoscopie, qui est un processus qui à partir d'images clé d'une vidéo permet de recréer en animation 3d le contenu de la vidéo. Au départ du projet, il a fallu choisir la personne sourde qui servirait de modèle pour la signeuse virtuelle, avec la contrainte qu'au final le personnage 3d soit une femme, et que les informations soient données de manière claire et compréhensible dans la forme. Suivant la première contrainte, nous nous sommes posé la question de savoir si, pour que le résultat final soit une signeuse (et pas un signeur), il était obligatoire que le modèle servant de base pour le personnage soit une femme. Nous n'avons pas trouvé dans la littérature d'étude qui comparait deux signeurs virtuels de même sexe créés à partir de personnes de sexes différents, nous nous sommes donc basés sur les connaissances de l'infographiste du LIMSI-CNRS (qui avait déjà réalisé par le passé des animations en rotoscopie), les expériences des personnes sourdes qui ont eu l'occasion de voir les premières animations de la signeuse virtuelle, et le fait que potentiellement il y avait moins de risque pour le projet de choisir

une locutrice plutôt qu'un locuteur¹⁷. De plus, suivant la seconde contrainte, nous devons choisir parmi l'équipe de traducteurs de WebSourd qui sont des personnes qui quotidiennement présentent en vidéo sur le Web des actualités brèves issues de la traduction de dépêches AFP. En effet, ces personnes sont formées à la traduction et à l'expression de discours clairs face à une caméra. Pour toutes ces raisons, notre choix s'est porté sur une locutrice de LSF, traductrice et responsable du pôle traduction à WebSourd, qui est également notre experte LSF pour la partie de traduction du contenu du corpus, comme nous le présentons dans la sous-étape suivante.

b. Conception du corpus

Les données de notre corpus sont imposées par le projet et fournies par la SNCF. Ces données sous forme écrite sont des énoncés dans lesquels il y a des blancs pour laisser la place à des « variables ». Concrètement, l'énoncé « [appellation de voie], éloignez vous de la bordure du quai, [type de train], numéro [numéro de train] entre en gare » comprend trois variables qui sont l'appellation de la voie (qui peut être une ou des lettres ou un ou des chiffres), le type de train (TGV, Corail, etc.) et le numéro du train (chiffres).

Nous avons traduit, avec l'aide de notre locutrice LSF, également experte LSF, ces énoncés en LSF. Nous avons établi un ensemble de plusieurs exemplaires de chaque énoncé, en essayant d'insérer toutes les combinaisons de types de variables dont il nous semblait qu'elles pouvaient engendrer une traduction différente de l'énoncé. Cependant, quelque soit la variable employée, quelque soit l'énoncé, ces variations lexicales (combinaison des variables possibles) n'ont pas modifié la grammaire de la traduction en LSF. Nous obtenons donc une traduction LSF par énoncé complet écrit en français. Nous avons segmenté les énoncés complets en regroupant les parties communes à d'autres énoncés complets. Ce travail nous permet de disposer au final, non plus des énoncés complets initiaux, mais de segments d'énoncés et de variables, termes que nous regrouperons dorénavant sous l'appellation « segments d'énoncés », qui coarticulés ensemble permettent de produire les énoncés.

Nous nous sommes ensuite intéressés aux types des différents énoncés proposés, aussi bien au niveau discursif (le message est-il une information standard ou une alerte, une urgence ?) qu'au niveau linguistique (en essayant de regrouper les structures identiques). Nous avons regroupé les énoncés suivant cette classification, et sélectionné quelques un jugés représentatifs de la totalité, au moins un par type de structure et par type d'information. Enfin, nous avons élaboré plusieurs

¹⁷ Il nous semblerait cependant intéressant de réaliser une comparaison entre des signeurs virtuels de même sexe créés à partir de locuteurs de sexes différents, tout comme une comparaison entre des signeurs virtuels de sexes différents créés à partir de locuteurs de même sexe.

exemples d'énoncés, que nous appellerons « instance d'énoncé », avec différentes valeurs de variables.

Les énoncés qui ont été instanciés pour être enregistrés dans notre corpus sont au nombre de 7, et sont nommés « AS01 », « AS16 », « IT12AB », « DOY », « DT02 », « RT20A », « RT20B » d'après les références fournies par la SNCF. Nous avons créé 8 instances de AS01, 9 de AS16, et 3 pour chaque autre énoncé. Nous avons enregistré les différentes instances d'énoncés sélectionnés, ainsi que la majorité des segments d'énoncés qui les composent. Au total, notre corpus contient :

- 32 instances d'énoncés,
- 67 segments d'énoncés (dont 38 variables : 6 signes de villes, 23 nombres, 6 lettres, 3 horaires), qui composent les 32 instances d'énoncés.

Nous avons constitué notre corpus en fonction de notre objectif d'analyse de la coarticulation, nous allons maintenant détailler ce que nous avons voulu observer de par nos choix d'instances d'énoncés. Tout d'abord nous souhaitons observer la coarticulation des différentes variables mises en contexte : ceci concerne les chiffres, la dactylogogie, les horaires et les signes de gare. Ainsi, la coarticulation des chiffres peut être analysée dans les instances d'énoncés AS01, DOY, DT02, RT20A et RT20B, la coarticulation de la dactylogogie dans les instances de DOY et IT12AB, la coarticulation des signes de gare dans DOY, DT02 et IT12AB, et celle des horaires dans RT20A. Mais nous nous intéressons également à la coarticulation entre les différents segments d'énoncés. Ainsi, nous avons enregistré dans notre corpus plusieurs structures (i.e. construction) d'énoncés. Si nous considérons la composition d'un énoncé par des segments d'énoncés, en représentant chaque segment par une lettre majuscule de l'alphabet, nous pouvons symboliser les différentes structures d'énoncés ainsi :

- AS01 a comme structure « ABCAB »
 - o A = Voie ...
 - o B = éloignez vous de la bordure du quai
 - o C = attention au passage d'un train
- AS16 a comme structure « DEFG »
 - o D = votre attention s'il vous plait
 - o E = 'cause d'incident'
 - o F = la circulation des trains est interrompue
 - o G = nous vous donnerons des précisions dès que possible
- DOY a comme structure « AHI »
 - o H = le TGV ... à destination de ... va partir
 - o I = prenez garde à la fermeture des portes, attention au départ
- DT02 a comme structure « AKI »

- K = les TGV ... à destination de ... vont partir
- IT12AB a comme structure « ABL »
 - L = les TGV en provenance de ... et à destination de ... entrent en gare
- RT20A a comme structure « DAEKM »
 - M = arriveront à 'horaire'
- RT20B a comme structure « DAEKG »

Notons que nous avons représenté le même segment toujours par la même lettre majuscule. Ainsi, nous pouvons regrouper les énoncés dont la structure est similaire (les « ... » représentant une partie qui n'est pas commune aux deux énoncés) :

- AS01 et IT12AB ont en commun AB... ;
- AS16 et RT20A ont en commun D...E... ;
- AS16 et RT20B ont en commun D...E...G ;
- DOY et DT02 ont en commun A...I ;
- RT20A et RT20B ont en commun DAEK...

Ces regroupements vont nous permettre d'analyser la coarticulation entre les segments d'énoncés lorsqu'ils seront mis en contexte, et de croiser les résultats en fonction des contextes.

Les énoncés que nous avons choisis pour notre corpus peuvent enfin être regroupés suivant le type d'information qu'ils véhiculent. Ce regroupement n'est pas de nature à faire ressortir les effets de la coarticulation, mais plutôt de nous permettre de distinguer les variations dues à la contextualisation (d'un point de vue plutôt sémantique et pragmatique) et celles dues à la coarticulation.

Nous considérons deux types d'énoncés dans notre corpus : une information standard ou un énoncé prévenant d'un incident. Suivant cette classification, nous pouvons regrouper :

- AS01, IT12AB, DOY et DT02 qui sont des énoncés d'information,
- AS16, RT20A et RT20B qui sont des énoncés d'incident.

c. Réalisation du corpus

Le dispositif de captation du corpus vidéo a été mis en place à WebSourd, dans un studio d'enregistrement, avec des matériels prêtés par le LIMSI-CNRS. Nous avons filmé avec deux caméras, une de face en plan large (depuis la ceinture jusqu'à une cinquantaine de centimètres au dessus de la tête) et une de profil (Figure 10). Ces enregistrements ont ensuite été montés de manière synchrone au LIMSI-CNRS, et livré sous forme brut (non compressée) afin de garantir une qualité maximale.



Figure 10 : Instantané de notre corpus vidéo

2. Etape 2 : annotation du corpus

Pour annoter notre corpus nous avons utilisés des caractères alphanumériques, des couleurs, et des images.

Les caractères alphanumériques sont la base de toutes les annotations sous Anvil : les annotations sont enregistrées dans des fichiers sous cette forme, tout en autorisant diverses possibilités de visualisation via des courbes, des dessins et des couleurs. Les ensembles de valeurs créés nous ont permis d'annoter les configurations des mains, la direction du regard et l'ouverture et la fermeture des yeux.

Nous avons aussi utilisé le système de mise en valeur visuelle (Figure 11) disponible sous l'outil ANVIL : une couleur a été affectée pour chaque période de temps où les yeux sont ouverts ou fermés.



Figure 11 : Visualisation aisée sous Anvil grâce aux couleurs (pour l'ouverture (en vert) et la fermeture (rouge) des yeux)

Nous avons également utilisé des dessins et des images issues de captures d'écran pour annoter les configurations des mains et la direction du regard. Nous ne voulions pas décrire finement dès le départ toutes les configurations du corpus, estimant que ce serait bien trop fastidieux et que devant la somme de données résultante nous aurions des difficultés à tirer des conclusions. Nous avons donc utilisé la liste des configurations établie par Cuxac [CUX00], que nous avons numérisées pour les insérer dans ANVIL (Figure 12).

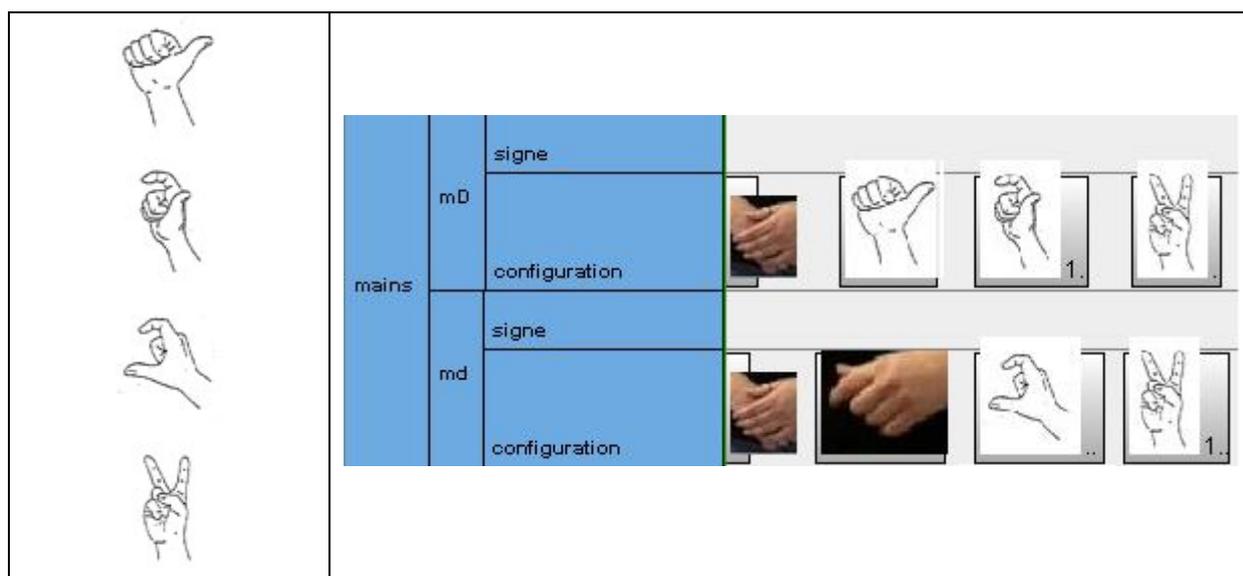


Figure 12 : Exemple de configurations (à gauche numérisées, à droite insérées dans Anvil)

Cette liste n'étant pas exhaustive, nous nous sommes retrouvés dans la situation où certaines configurations de notre corpus n'avaient pas de dessin associé. Nous avons décidé de réaliser des captures d'écran de ces configurations, en rassemblant parfois en une vignette les deux vues issues du corpus (Figure 13).



Figure 13 : Exemples de captures d'écrans pour les configurations des mains

L'avantage des captures d'écrans est qu'il n'y a pas de risque d'erreur dans la recopie ou dans la qualification de l'élément annoté : le label d'annotation est identique ou pas à ce qui est visualisé dans la vidéo. Nous avons de même rapidement utilisé des labels de capture d'écran pour la direction du regard, en repérant par une marque sur le label l'endroit de l'espace où selon nous le regard se portait (Figure 14).



Figure 14 : Exemples de capture d'écrans pour la direction du regard

Les différentes valeurs d'annotation du regard sont rassemblées sur le schéma suivant (Figure 15).



Figure 15 : Schéma récapitulatif des valeurs de direction du regard

Dans cette étape d'annotation du corpus s'est posée la question de la segmentation : comment déterminer le début et la fin du phénomène que nous voulons annoter ? En ce qui concerne l'ouverture et la fermeture des yeux, il s'agit de repérer le moment où les yeux sont clos : les yeux sont alors considérés comme fermés. Le reste du temps, y compris les images où on voit que l'œil est en train de se fermer mais qu'il ne l'est pas encore complètement, les yeux sont considérés comme étant ouverts. En ce qui concerne les configurations des mains, nous avons annoté les moments où la main, soit sa configuration, son orientation, son mouvement et son orientation dans l'espace, était dans une production linguistique, que ce soit dans le cadre d'un signe standard ou une reprise iconique de ce qui est exprimé (appelé aussi proforme). Enfin, pour la direction du regard, nous avons là aussi annoté les moments où il était fixé sur un emplacement de l'espace.

3. Etape 3 : traitements des annotations

Une fois les annotations d'ouverture des yeux, de configuration des mains et de direction du regard effectuées, il nous faut analyser ces annotations et mener des calculs statistiques afin de faire émerger des corrélations entre les valeurs de ces éléments en contexte et hors contexte.

Nous avons à notre disposition des outils statistiques adaptés au traitement des fichiers d'annotations issus d'ANVIL ([SEG03]). Ces outils permettent de calculer tout un ensemble de corrélations entre plusieurs pistes dans ANVIL. Ainsi, il est possible de déterminer si un élément donné sur une piste se produit en même temps, avant, ou après un autre élément donné sur une autre piste, avec la possibilité de faire varier les indices temporels. Concrètement, une de nos études précédentes ([SEG03]) portait sur une analyse des corrélations de manière naïve (sans considérations linguistiques) entre la direction du regard et l'emplacement des mains. Il s'agissait de déterminer si un des deux éléments (emplacement du regard ou de la main) précédait l'autre, et de combien de temps. Les corrélations ont permis de mettre à jour une anticipation du regard par rapport aux mains dans certaines situations. Cependant, nous avons besoin de faire des calculs entre plusieurs fichiers sur des valeurs identiques, là où les calculs existants nous permettent seulement de traiter un seul fichier, et nous avons besoin de traiter les données par des calculs en alignement, et non plus seulement en s'intéressant à chaque élément en détail. Nous avons donc conçu de nouveaux outils, sur la base de ceux existants, afin de pouvoir faire un alignement des annotations des signes en contexte et hors contexte.

Nous nous sommes appuyés sur une implémentation en Java de l'algorithme de Smith-Watermann (cf. chapitre précédent), développée par Moustafa Ahmed [AHM] et disponible avec son code source sur le Web. Nous avons choisi cette implémentation pour plusieurs raisons :

- Techniquement, elle a été réalisée en Java, langage dans lequel nous avons déjà développé des outils de statistiques. L'interfaçage avec l'existant en sera facilité.
- Anvil étant développé en Java, l'extension de notre développement également en java pourra donner lieu à un plug-in utilisable par d'autres utilisateurs d'Anvil.
- Le code source est disponible librement sous licence GPL, donc complètement modifiable, ce qui répond à notre besoin d'implémentation adaptée.

Les annotations sont stockées dans des fichiers XML, sous forme de chaînes de caractères. L'implémentation de l'algorithme de Smith-Waterman est prévue pour traiter des séquences de d'éléments sous formes de chaînes de caractères (de quelques centaines à des milliers).

L'implémentation que nous souhaitons utiliser nécessite en entrée des chaînes de caractère dont chaque caractère va être traité séparément et comparé aux autres.

Nous avons choisi de convertir nos données en caractères ASCII (American Standard Code for Information Interchange) qui est une norme de codage des caractères. Celle-ci nous garantit un ensemble unique de caractères. Pour chaque ensemble de valeurs d'annotation (direction du regard, configuration des mains, ouverture des yeux), nous avons donc converti chaque valeur d'annotation (chaîne de caractère) en caractère ASCII unique.

Nous avons tout d'abord effectué un repérage des parties d'énoncés en contexte correspondant aux parties hors contexte. Pour cela, nous avons procédé en plusieurs étapes :

- Nous avons déterminé le début et la fin temporels d'un bloc d'annotation d'un segment d'énoncé. Dans cette étape, nous n'avons pas considéré les positions de repos de début et de fin comme faisant partie du bloc.
- Nous avons cherché dans l'énoncé en contexte à quelle partie temporelle correspondait cet intervalle : si un élément en contexte débutait avant le début de la partie hors contexte mais terminait après ce début, alors il est considéré comme faisant partie du bloc à aligner. De même, si la fin d'un élément en contexte arrive après la fin de la partie hors contexte, mais qu'il a débuté avant, alors il est considéré comme faisant partie du bloc à aligner.
- Une fois ces deux blocs repérés, un hors contexte et un en contexte, nous avons transformé les valeurs des annotations suivant la conversion en caractères ASCII.
- Enfin, nous avons aligné les séquences résultantes grâce à l'algorithme de Smith-Waterman.

La matrice de similarité sur laquelle nous nous sommes basés est la matrice identité : celle-ci permet de spécifier qu'il n'y a aucune valeur identique à une autre, hormis elle-même. Ainsi, il n'y a une valeur minimale (« 1 ») que sur la diagonale, le reste étant constitué de « 0 » (signifiant qu'il n'y a pas égalité). L'utilisation d'une telle matrice est révélatrice d'alignements, bien qu'il semble plus intéressant de constater des distances de similarité entre certaines valeurs de composantes (par exemple entre certaines configurations de mains) et de réaliser les alignements en prenant en compte ces similarités. Cependant la matrice identité est une base qui nous permet de procéder à un premier alignement : dans une éventuelle prochaine étude, il sera possible de tenir compte de similarités à plusieurs niveaux : phonétique, sémantique, discursif, etc.

Nous avons toutefois modifié la matrice identité afin de respecter les similarités de configurations observées, et résultant du processus d'annotation : en effet, il est arrivé que nous

avons annoté par deux valeurs différentes un même phénomène d'une composante, sans forcément nous en rendre compte lors de l'annotation (Figure 16). Suite à cette constatation, plutôt que de modifier toutes nos annotations, nous avons simplement modifié la matrice de similarité afin que le calcul d'alignement considère les deux valeurs comme identiques.



Figure 16 : Configurations jugées similaires

Une fois ces traitements effectués, nous obtenons une sortie des résultats dans des fichiers texte, dans un format présentant clairement les informations. Nous obtenons un fichier par énoncé complet. Ce format permet de présenter plusieurs informations (Figure 17) :

```
Fichier : ResStat/5_ScoresAlign_IT12AB-1.anvil.txt a
Piste A = 3hghlm
Piste C = hghg    lg = 4 } b
    Bloc A      2 hgh      4 } c
                | | |
    Bloc X      1 hgh      3 }
f La durée (entre le début du premier élément et la fin du dernier élément) de la séquence A (0.44 s)
  est plus longue que cette durée pour la séquence C (0.16 s) : 0.28 seconde(s)
Les durées d'annotation des éléments alignés sont :
  plus longues en A en nombre : 33 %      g
  plus longues en A en temps : 0.12 seconde(s) h
i Les modifications ont lieu au DEBUT de la séquence A
e 50.0 % d'éléments de A sont alignés avec 75.0 % d'éléments de la séquence C
```

Figure 17 : Description du format de fichier de sortie des résultats (cf. points ci-dessous)

- le nom de la piste considérée (unique dans le fichier) (a)
- les deux séquences considérées (au format ASCII) (b)
- une visualisation graphique de l'alignement de ces séquences (c)
- des informations statistiques numériques
 - o le nombre d'éléments dans chaque bloc (d)
 - o le pourcentage d'éléments alignés dans chaque bloc (e)
 - o la différence de durée d'annotation globale des blocs (par addition de la durée d'annotation des éléments alignés dans chaque bloc) (f)

- la proportion de durées d’annotations des éléments alignés qui sont supérieures, inférieures ou identiques en temps dans les blocs (g)
- la différence de somme des durées d’annotations des éléments alignés (h)
- la position majoritaire des modifications constatées (en se reposant sur le calcul de l’alignement, les modifications ont lieu là où il n’y a pas alignement) (i)

La visualisation graphique permet de se rendre compte rapidement de certains phénomènes décrits numériquement ensuite, comme le rapport du nombre d’éléments de chaque bloc, le pourcentage d’éléments alignés, mais aussi d’erreurs dans l’alignement des séquences (Figure 18), principalement pour les valeurs de direction du regard (dans l’exemple cité, le programme prétend que c’est la première valeur de la piste B qui est aligné avec la seconde valeur de la piste A, alors qu’il s’agit dans la piste B de la seconde valeur : l’erreur est corrigée par analyse visuelle des annotations sous Anvil).

Piste A = \$\$\$''	lg = 5
Piste B = \$\$	lg = 2
Bloc A	2 \$ 2
Bloc X	1 \$ 1

Figure 18 : Exemple d’erreur de l’algorithme d’alignement

Le nombre d’éléments de chaque bloc permet de mettre en correspondance la taille des blocs et de constater que certains blocs sont plus courts, donc ont moins d’éléments, alors que la durée globale des blocs est peut-être la même. Cela signifierait qu’il y a moins de variation dans un des blocs, ou plus dans d’autres, par exemple.

Le pourcentage d’éléments alignés dans chaque bloc nous informe sur la quantité d’éléments identiques dans les deux blocs, selon l’algorithme d’alignement. Cela peut mériter vérification visuelle sur l’alignement présenté plus haut, afin de vérifier que les éléments alignés sont corrects. Cela apporte une information importante sur la variation.

La différence de durée d’annotation globale des blocs, est calculée d’abord par, dans chaque bloc, addition de la durée d’annotation des éléments alignés, puis par soustraction de ces sommes. La soustraction est toujours « somme des durées des éléments du bloc en contexte – somme des durées des éléments du bloc isolé », ce qui implique qu’une valeur positive signifie que la somme des durées des éléments du bloc en contexte est supérieure, donc grossièrement que le bloc en contexte a une durée plus importante que le bloc isolé.

Sur la base des éléments alignés, nous effectuons un comptage des durées par catégories : pour chaque élément de chaque bloc, nous comparons sa durée avec l'élément avec lequel il est aligné dans l'autre bloc. En fonction de la valeur du signe du résultat de la soustraction des durées (toujours bloc en contexte – bloc isolé), nous incrémentons le compteur de la catégorie « plus long » (si c'est positif), « plus court » (si c'est négatif), ou « équivalent ». Nous sélectionnons le plus grand nombre, et établissons son pourcentage par rapport à l'ensemble de comparaisons établies. Concrètement, si il y a trois éléments alignés, que dans le bloc en contexte ces éléments ont une durée pour deux supérieure et pour un inférieure à leurs alignements dans l'autre bloc, alors le résultat dans le fichier sera « Les durées d'annotation des éléments alignés sont :

supérieures en nombre : 66 % ».

Ceci est le résultat du calcul de la proportion de durées d'annotations des éléments alignés qui sont supérieures, inférieures ou identiques en temps dans les blocs, représente le pourcentage d'éléments. Noter que si le résultat est « inférieur », il y aura un signe négatif placé devant la valeur du pourcentage.

La différence de somme des durées d'annotations des éléments alignés nous permet de savoir si, outre le pourcentage d'éléments alignés, la différence des durées est grande ou non. Cela nous donne une première information quand à la relation entre la proportion précédente et la durée réelle constatée de cette proportion. On peut par exemple poser comme hypothèse que les éléments ont peu varié en durée quand la proportion est de 100% et que la durée cumulée est de 0.12 secondes. Bien sûr, cela est à mettre en parallèle avec d'autres données, notamment le nombre d'éléments alignés sur lequel reposent ces calculs.

Enfin, la position majoritaire des modifications constatées nous informe, selon les valeurs numériques, sur la position des plus importantes variations entre les blocs. Le résultat a ses limites, puisque si des variations ont lieu en même intensité au début et à la fin, le résultat est vide. Cependant, dans ce cas, un résultat vide invite le lecteur des résultats à se pencher sur l'alignement visuel et à faire sa propre conclusion sur l'emplacement des variations.

Ces différentes données sont ensuite transposées dans un tableur afin de pouvoir visualiser aisément à l'aide de graphiques les différentes informations recueillies.

Nous avons conçu un classeur de tableur avec plusieurs fichiers : un pour chaque énoncé (AS01, AS16, IT12AB, DOY, DT02, RT20A, RT20B), un pour chaque composante considérée (direction du regard, configuration main dominée (gauche), configuration main dominante (droite)), un regroupant tous les morceaux d'énoncés en contexte, et d'autres qui regroupent des calculs croisés entre les précédentes feuilles.

Nous avons réalisés des calculs de somme, de moyenne et de variance pour les différents énoncés et composantes, et nous avons produit des graphiques représentant les sommes sous forme d'histogramme empilé. Cette visualisation nous permet d'apprécier rapidement la valeur et la quantité d'un élément, ainsi que son pourcentage dans un ensemble. Nous avons préféré les tableaux aux graphiques pour les données temporelles.

4. Etape 4 : analyses et conception du modèle

Rassemblés dans un tableur et analysés au moyen de graphiques, les résultats obtenus nous permettent de définir des corrélations qui vont être les bases de notre modèle de coarticulation. Nous avons utilisé l'algorithme de Smith-Waterman, et nous avons réalisé un transcodage de nos valeurs d'annotations pour qu'elles puissent être traitées par cet algorithme. Au vu de ces alignements, et des statistiques sur ces résultats d'alignements, nous concevons des règles de coarticulation qui sont la base de notre modèle.

Nous présentons les résultats des analyses et nos conclusions dans le chapitre suivant.

5. Synthèse

Nous avons constaté que la méthodologie généralement utilisée dans les études correspond à nos besoins, nous allons donc suivre les différentes étapes qui la constituent.

Nous avons choisi d'élaborer notre modèle de coarticulation à partir de données réelles. Ces données, comme nous l'avons vu, sont collectées dans le cadre de la constitution d'un corpus. Dans cette première étape nous sélectionnons notre locutrice, le contenu et la technique de réalisation du corpus. Nous souhaitons analyser ce corpus, dans le sens où nous voulons repérer certains phénomènes : nous devons donc, dans une seconde étape, « annoter » ce corpus suivant des critères définis en amont, quant à ce qui est annoté (la direction du regard, l'ouverture et la fermeture des yeux, les configurations des mains) et comment c'est annoté (par des caractères, des couleurs et des images). L'étape suivante consiste en l'analyse de ces annotations à l'aide d'outils de traitements statistiques afin de faire ressortir de la masse de données récoltée les informations que nous jugeons pertinentes (en l'occurrence, l'alignement entre les énoncés, et des statistiques sur ces alignements). L'algorithme d'analyse que nous avons choisi nous permet de mettre en exergue les variations entre les productions dans leur forme isolée et dans leur forme en contexte, puis avec un traitement statistiques nous quantifions et qualifions ces alignements. Enfin, nous analysons ces résultats pour concevoir notre proposition de modèle de coarticulation.

Dans le chapitre suivant nous présentons et détaillons nos résultats, après les traitements statistiques, et proposons en conclusion un modèle de coarticulation.

Chapitre IV. RESULTATS

Nous allons maintenant présenter les résultats des différentes analyses que nous avons menées, et qui ont été détaillées dans la partie précédente. Nous avons étudié des variations de valeur et des variations temporelles : les valeurs sont les configurations des mains et l'emplacement de la direction du regard, les variations temporelles correspondent aux durées des segments d'annotation des configurations des mains et direction du regard. Nous présentons tout d'abord les variations en valeur, en les regroupant dans un premier temps par composante et par type de variation (modification, insertion, suppression), puis dans un second temps par énoncé. Dans un troisième temps nous proposons une présentation par catégories de configuration des mains, basée sur la classification de Boutora [BOU08]. Nous présentons ensuite les variations temporelles, c'est-à-dire les modifications de durée de chaque composante en comparant les éléments dont la valeur n'a pas été modifiée, c'est-à-dire qui n'ont pas subi de variation entre leur forme isolée et leur forme en contexte.

Nous précisons certaines notions que nous allons employer dans la suite de ce chapitre :

- Une **modification** est le changement en valeur d'une (ou plusieurs) configuration(s) ou d'une (ou plusieurs) direction(s) de regard. La modification n'ajoute pas ni n'enlève d'information (au sens sémantique du terme) à l'énoncé, sinon c'est une insertion ou une suppression.
- Une **insertion** correspond à l'insertion d'une (ou plusieurs) valeur(s) dans la forme en contexte par rapport à la forme isolée.
- Une **suppression** correspond à la suppression d'une (ou plusieurs) valeur(s) dans la forme en contexte par rapport à la forme isolée.
- Une **variation** représente un changement (que ce soit une modification, une insertion ou une suppression) entre la forme isolée et la forme en contexte d'un segment énoncé.

I. Variations de valeurs

Nous présentons les variations de valeurs que nous avons pu constater. Ces résultats sont directement issus du calcul d'alignement et des statistiques que nous avons réalisés sur les annotations. Nous nous intéressons uniquement aux valeurs, et aux séquences de valeurs qui représentent les segments d'énoncés dans leur forme isolée et dans leur forme en contexte.

Le symbole « -> » n'a pas de signification en tant que valeur d'annotation, mais permet de spécifier la variation lors de la mise en contexte : « X -> Y » signifie que X est modifié en Y, « X -> » signifie que X est supprimé, « -> Y » signifie que Y est ajouté.

1. Par composante et par type de variation

Nous présentons ici les variations de valeur pour chaque composante direction du regard, configuration de mD et configuration de md. Nous avons différencié les résultats suivant les types de variations (modifications, insertions et suppressions).

La présentation de cette partie suit le déroulé suivant : un graphique montre les variations dans leur ensemble pour la composante considérée, ensuite, par variation, un tableau détaille les principales valeurs, enfin nous dressons des conclusions. Au sein des graphiques, nous avons réalisé des regroupements de valeurs, par variation puis par similarité. Nous détaillons ces regroupements à la suite de chaque graphique.

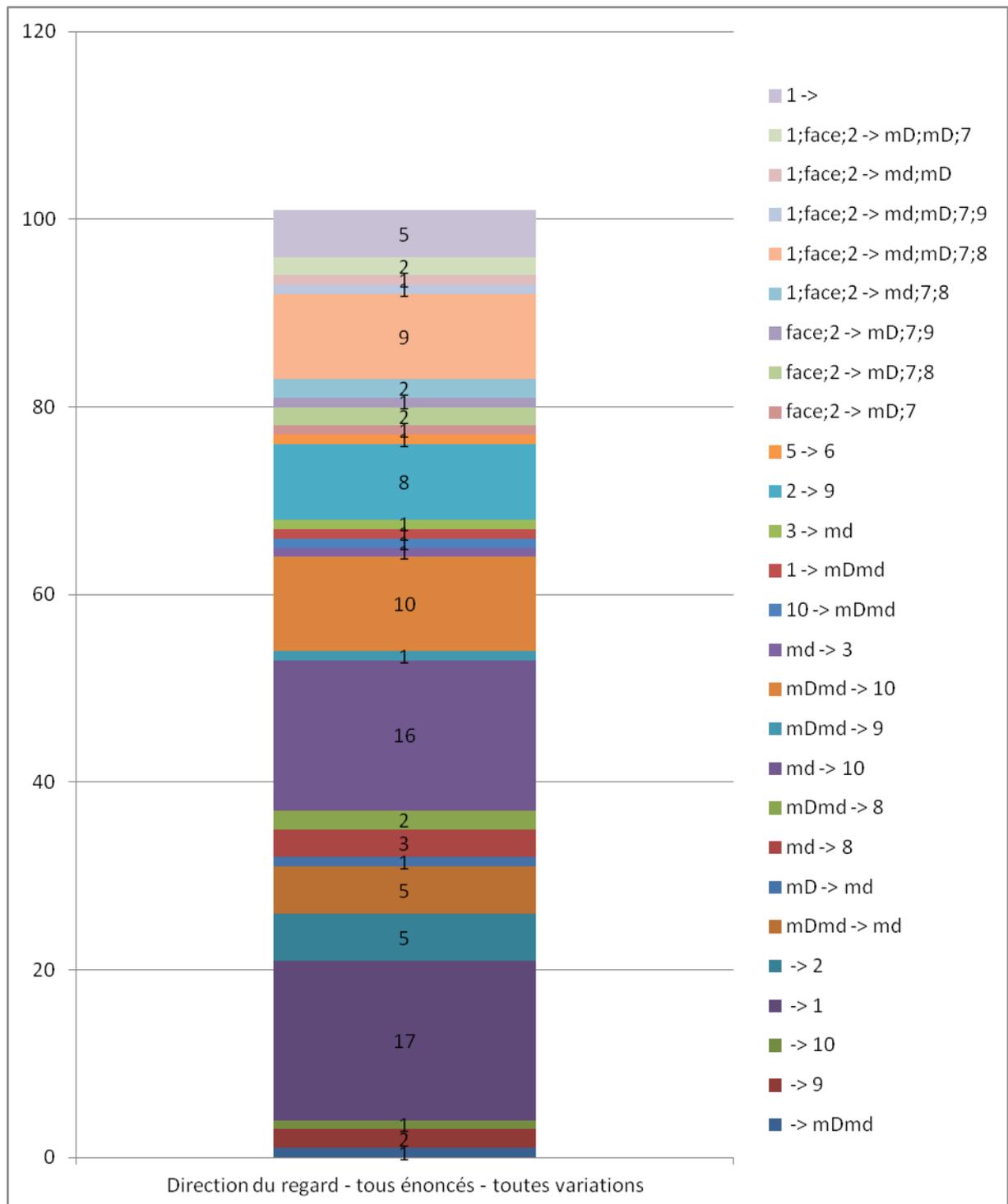
a. Direction du regard

Les valeurs qui concernent la direction du regard représentent les différents emplacements détaillés dans le chapitre précédent et repris dans le schéma suivant (Figure 19).



Figure 19 : Schéma récapitulatif des différents emplacements

Le graphique suivant présente la totalité des variations constatées de la direction du regard, tous énoncés confondus.



Graphique 1 : Direction du regard – tous énoncés – toutes variations

Les variations présentées sont regroupées, autant dans leur nature (modification, insertion ou suppression) que dans la quantité des éléments qu’ils impliquent : en effet, nous avons regroupé les variations consécutives d’un même segment d’énoncé, comme par exemple dans « 1 ;face ;2 -> mD ;mD ;7 ». Nous avons regroupé ces valeurs lorsque nous avons jugé qu’il n’était pas pertinent

d'essayer un appariement de valeur à valeur des variations constatées. Nous avons également rassemblé les variations en fonction des valeurs impliquées. Ainsi, nous observons :

- une suppression d'emplacement de l'espace, « 1 -> »
- des modifications d'ensembles de valeurs, de « 1 ;face ;2 -> mD ;mD ;7 » à « face ;2 -> mD ;7 »
- des modifications depuis et vers un emplacement de l'espace, de « 5 -> 6 » à « 2 -> 9 »
- des modifications depuis un emplacement de l'espace vers les mains, de « 3 -> md » à « 10 -> mDmd »
- des modifications depuis les mains vers un emplacement de l'espace, de « md -> 3 » à « md -> 8 »
- des modifications depuis les mains vers les mains, « mD -> md » et « mDmd -> md »
- des insertions vers l'emplacement de l'espace, de « -> 2 » à « -> 9 »
- une insertion vers les mains, « -> mDmd »

Les plus importantes modifications représentent 49 % des variations totales de la direction du regard, quand les plus importantes insertions représentent 22 % et les suppressions les plus importantes représentent 5 %.

Modifications

Forme isolée	Forme en contexte	Quantité (total des variations de direction du regard = 101)
1 ;face ;2	md ;mD ;7 ;8	9
2	9	8
mDmd	10	10
md	10	16
mDmd	md	17

Tableau 1 : Direction du regard – Modifications les plus importantes

Les modifications les plus importantes décrites ci-dessus représentent 60% des modifications de la direction du regard. Elles interviennent principalement dans l'énoncé AS01 (« XX, éloignez vous de la bordure du quai. Attention au passage d'un train. XX, éloignez vous de la bordure du quai. » où « XX » est une dénomination de voie) et dans une moindre mesure dans l'énoncé IT12AB (« XX,

éloignez-vous de la bordure du quai, les TGV en provenance de YY et à destination de Reims et Paris Est entrent en gare. », où « XX » est une dénomination de voie et « YY » est le nom d'une gare).

- 1 ;face ;2 -> md ;mD ;7 ;8

Cette modification n'apparaît que dans les énoncés AS01, au moment où est représenté le quai, dans le segment d'énoncé E12 « éloignez-vous de la bordure du quai ». Le regard de la locutrice suit les mains puis est dirigé vers l'espace plutôt que de fixer l'interlocuteur (« face » est remplacé par « mD » et « 7 »), et l'emplacement « 1 » est abandonné au profit du suivi de « md ». Concernant cette variation, il faut savoir que la main dominée « md » est à ce moment là à l'emplacement « 1 » mais se déplace légèrement et que le regard suit ce déplacement, ce qui n'est pas le cas dans la forme isolée, d'où le fait qu'en contexte nous considérons que le regard est plutôt dirigé vers md. Les emplacements « 7 » et « 8 » marquent le début et la fin de la voie dans la forme en contexte, quand seule la fin de la voie « 2 » est marquée dans la forme isolée ; de plus, l'emplacement « 2 » est plus excentré sur la gauche de la locutrice que « 8 ».

Nous remarquons une plus grande implication du regard de la locutrice dans ce qui est exprimé : selon nous, la locutrice représente la forme du quai dans l'espace tout en maintenant son regard dirigé vers ce qu'elle représente, afin d'insister sur le fait que la forme du quai est comme celle décrite, et qu'il n'est pas présenté une forme générique de quai. Nous notons également un rapprochement des directions de regard vers la locutrice avec la modification des directions depuis « 2 » vers « 8 ».

Nous faisons ici l'hypothèse que ces modifications interviennent en regard de l'information donnée par l'énoncé : il s'agit de prévenir les personnes de s'éloigner de la bordure du quai, le regard va donc insister davantage sur le quai, la foule, et se focaliser sur un espace plus proche pour concentrer le regard de l'interlocuteur.

- 2 -> 9

Cette modification n'apparaît également que dans les énoncés AS01 au moment de l'expression du fait qu'un train va passer (E13, « attention au passage d'un train »). Au lieu de fixer le quai, le regard se porte vers la fin du quai, voire la destination du train, dans tous les cas en direction du départ du train. Il y a une élévation de la direction du regard, ainsi qu'une fixation sur un point qui semble plus éloigné pour l'emplacement « 9 » que pour l'emplacement « 2 ».

Notre hypothèse est que la cette modification est due à une élévation générale de l'espace de signation, qui de fait entraîne une élévation des directions de regard.

- md -> 10

Cette modification a lieu dans la quasi-totalité (13 sur 16) des mises en contexte dans AS01, et dans la totalité des mises en contexte de IT12AB (3 sur 3) : il s'agit du même segment d'énoncé (E12, « éloignez-vous de la bordure du quai ») mis en contexte dans toutes les instances de AS01 et IT12AB. La modification apparaît avec le signe « foule », où en contexte le regard est plus marqué vers l'emplacement où a été réalisé le signe que vers la main qui réalise le signe.

- mDmd -> 10

Cette modification apparaît dans AS01 dans 8 occurrences sur 16. Parmi ces 16 occurrences, 8 sont en contexte au début des énoncés et 8 à la fin : les variations ont lieu au début et à la fin de 2 instances (AS01-2 et AS01-5), quand dans les autres instances elles sont soit à la fin soit au début. Cette modification apparaît également dans les 3 instances de IT12AB où apparaît le même segment d'énoncé que dans AS01 (E12). Il s'agit du même énoncé que pour la modification précédente, la foule ayant été positionnée à un emplacement, le regard se reporte à ce même emplacement pour la réalisation du signe signifiant « éloignez-vous ».

Des deux modifications précédentes, nous constatons que la direction vers un emplacement « 10 » est préférée à celle d'une main « md » dans le cas du signe « foule » ou des deux mains « mDmd » dans le cas du signe « éloignez-vous ».

- mDmd -> md

Cette modification a lieu dans AS01 à 5 reprises sur 16 mises en contexte de E12. Elle est à rapprocher de la modification précédente, puisque quand le regard n'est pas dirigé vers l'emplacement « 10 » il est dirigé principalement vers « md ». Nous soulignons qu'au début de la réalisation du signe concerné (« quai »), « md » est positionné à l'emplacement « 10 » mais qu'ensuite le regard suit la main plutôt que de rester focalisé sur « 10 ».

Notre constat est qu'il y a une plus grande importance accordée à l'emplacement qu'au suivi des mains, puis aux deux mains. Ceci est du, selon nous et notre locutrice/experte LSF, à l'importance accordée à l'emplacement de ce qui est signé plutôt qu'à ce qui est signifié : il est jugé plus important que les personnes comprennent qu'il faut qu'elles se déplacent de la bordure du quai (en LSF,

l'expression de ces informations est inversés, d'où le fait qu'il y ait un regard d'abord vers le quai puis vers les mains) plutôt que l'endroit où elles se trouvent.

Insertions

Forme en contexte	Quantité (total des variations de direction du regard = 101)
2	5
1	17

Tableau 2 : Direction du regard – Insertions les plus importantes

Les insertions les plus importantes décrites ci-dessus représentent 85% des insertions de la direction du regard.

Ces insertions des emplacements « 1 » et « 2 » ont lieu dans différentes instances des énoncés AS01, RT20A, RT20B, et DT02.

Elles ont lieu dans toutes les instances de AS01, lors de la mise en contexte du segment d'énoncé d'annonce de la voie (soient E1 à E11). Ce segment est présent à 2 reprises dans chaque instance, au début et à la fin de l'énoncé. Pour 4 instances d'énoncé (AS01-1, AS01-2, AS01-4, AS01-5) il y a ajout d'une direction de regard vers un emplacement de l'espace dans les deux occurrences du segment d'énoncé. Pour 3 instances d'énoncés (AS01-3, AS01-7, AS01-8) il n'y a pas d'ajout d'une direction de regard vers un emplacement de l'espace dans la 2^{nde} occurrence du segment d'énoncé alors qu'il y en a dans la première ; pour une instance d'énoncé (AS01-6) il n'y a pas d'ajout d'une direction du regard vers un emplacement de l'espace dans la 1^{ère} occurrence du segment d'énoncé alors qu'il y en a dans la 2^{nde}.

Dans une instance de RT20A et une instance de RT20B, de la même manière que pour les instances de AS01, on retrouve une insertion dans le segment d'énoncé (E6) lorsqu'il est mis en contexte.

Cette insertion se retrouve également lors de la mise en contexte du segment d'énoncé E25 (« en raison de travaux sur les voies »), mais pour seulement la moitié des instances d'énoncé où E25 apparaît (dans RT20B-2 et RT20B-3).

Enfin il y a des insertions d'emplacement « 1 » qui se retrouvent dans les 3 instances d'énoncés DT02, lors du segment d'annonce de la voie (E1, « voie A »).

Nous formulons l’hypothèse que les insertions considérées interviennent, comme dans le cas des modifications, pour focaliser l’intérêt de l’interlocuteur sur une partie précise du message : le quai, dont il faut s’éloigner de la bordure, que ce soit pour signaler qu’il y a des travaux sur les voies (qui n’est certes pas la même chose que le quai, mais les deux sont liés de fait par la même dénomination), ou pour signaler que le train part.

Suppressions

Forme isolée	Quantité (total des variations de direction du regard = 101)
1	5

Tableau 3 : Direction du regard – Suppression la plus importante

La suppression décrite ci-dessus est la seule suppression concernant la direction du regard.

Il s’agit de la suppression d’un regard dirigé vers un emplacement, lors de la mise en contexte de E25 (« en raison de travaux sur les voies ») dans les instances d’énoncés AS16-3, RT20A-3, RT20B-2, RT20B-3, et lors de la mise en contexte de E29 (« en raison d’un incident affectant la voie ») dans l’instance d’énoncé AS16-7.

Dans les instances AS16-3, RT20B-2, RT20B-3, il s’agit de la suppression du signe « situation », il nous semble donc logique que la direction de regard associée soit supprimée. Dans AS16-7, le regard est dirigé vers l’emplacement de réalisation du signe « voie », dans la forme isolée du segment d’énoncé correspondant il y a ensuite un regard porté à nouveau vers cet emplacement lors de la réalisation du signe « problème », là où en contexte il y a un pointage du doigt. Nous formulons l’hypothèse que le pointage du regard a été remplacé par un pointage du doigt. En lien avec ce que nous avons formulé comme hypothèse plus avant, il se pourrait que le délai de temps soit trop court pour qu’il y ait de nouveau un regard dans la direction du signe. Dans RT20A-3 la direction du regard est conservée lors de la réalisation du signe « situation », cependant qu’il n’est plus pour le signe « voie ». Nous formulons ici deux hypothèses : il pourrait y avoir un délai minimal en temps entre deux directions de regard vers un même emplacement (délai que nous tentons de déterminer dans la partie suivante concernant les analyses de durées), ou bien il s’agit de mettre l’accent sur la situation

(des travaux sur la voie) plutôt que sur la voie en elle-même (qui a été dénommé précédemment, avec le regard dirigé vers le signe « voie » à ce moment là).

Synthèse

Nous constatons des variations principalement dans les énoncés E12 et E13.

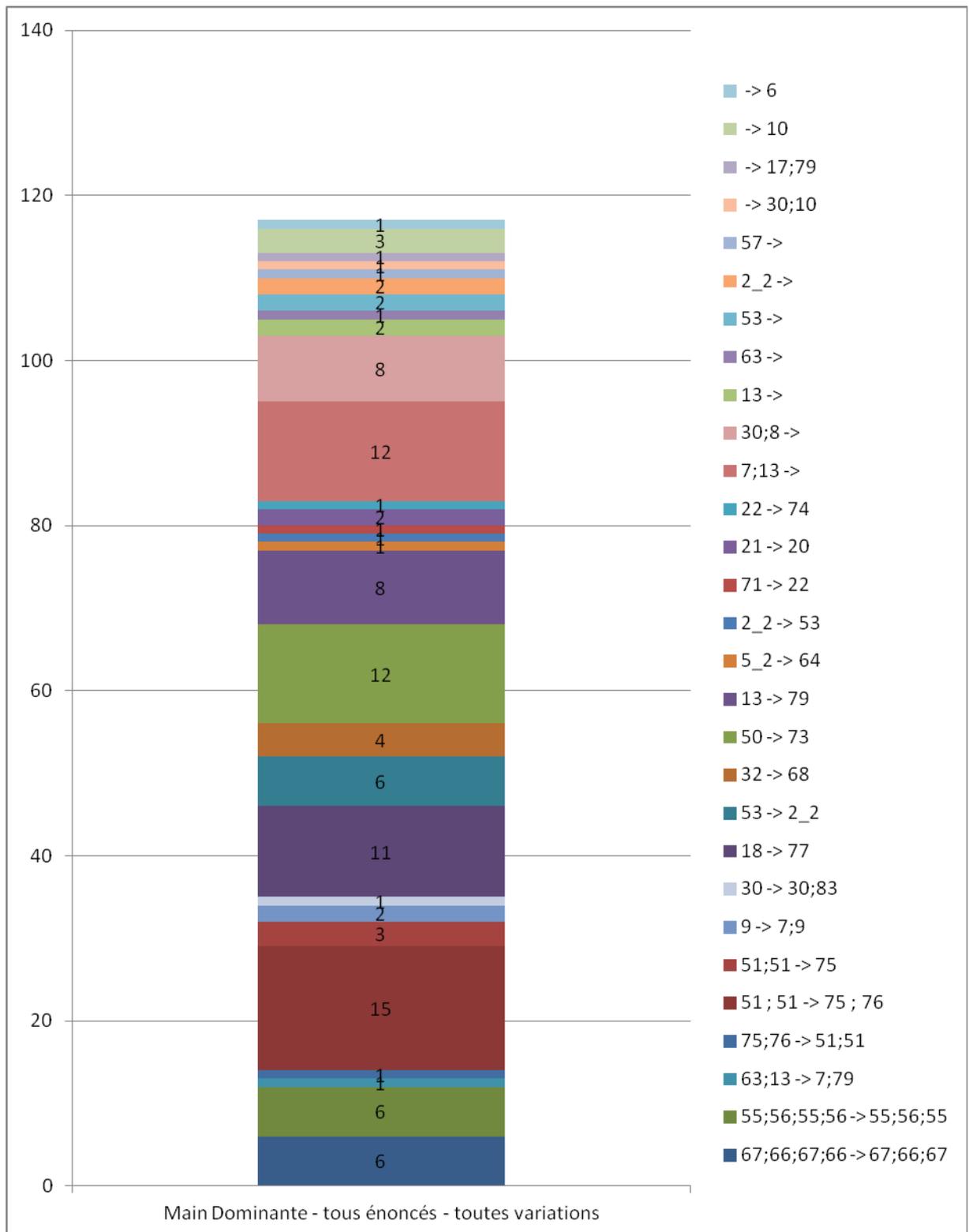
Les variations de direction du regard se retrouvent principalement dans AS01, un peu dans IT12AB, et quasiment pas dans les autres énoncés. Les modifications interviennent en regard de l'information donnée par l'énoncé : il s'agit de prévenir les personnes de s'éloigner de la bordure du quai, le regard va donc insister davantage sur le quai, la foule, et se focaliser sur un espace plus proche pour concentrer le regard de l'interlocuteur. Ceci est dû, selon nous et notre locutrice/experte LSF, à l'importance accordée à l'emplacement de ce qui est signé plutôt qu'à ce qui est signé : il est jugé plus important que les personnes comprennent qu'il faut qu'elles se déplacent de la bordure du quai (en LSF, l'expression de ces informations est inversés, d'où le fait qu'il y ait un regard d'abord vers le quai puis vers les mains) plutôt que l'endroit où elles se trouvent. Les variations de direction de regard sont dues à une élévation générale de l'espace de signation, qui de fait entraîne une élévation des directions de regard. Ces variations sont selon nous dues au contexte sémantique plus qu'au contexte phonétique (donc pas dues à la coarticulation).

Les suppressions de direction de regard sont dues d'une part à la suppression des signes qui impliquent leur réalisation, et soit à la proximité d'une direction de regard équivalente dans un intervalle de temps court (un intervalle de deux configurations de mains ou une durée de moins d'une seconde) soit au sens de l'énoncé (dans ce dernier cas, le regard se focalise sur l'information importante de l'énoncé).

Ces variations sont dues en grande majorité au contexte sémantique plutôt qu'au contexte phonétique, il n'y a donc pas, hormis pour la suppression de regard, de variations dues à la coarticulation.

b. Configuration de la main dominante (mD)

Le graphique suivant présente la totalité des variations constatées de la main dominante (mD), tous énoncés confondus.



Graphique 2 : Main dominante – tous énoncés – toutes variations

De la même façon que pour le graphique précédent, les variations présentées sont regroupées, autant dans leur nature (modification, insertion ou suppression) que dans la quantité des éléments

qu'ils impliquent. Nous avons regroupé les variations consécutives concernant un même signe, comme par exemple dans « 67 ;66 ;67 ;66 -> 67 ;66 ;67 » où il s'agit du signe « Strasbourg ». Ainsi, nous observons :

- des insertions de configuration unique, « -> 6 » et « -> 10 »
- des insertions de plusieurs configurations, « -> 17 ;79 » et « -> 30 ;10 »
- des suppressions de configuration unique, de « 57 -> » à « 13 -> »
- des suppressions de plusieurs configurations, « 30 ;8 -> » et « 7 ;13 -> »
- des modifications de configuration unique à configuration unique, de « 22 -> 74 » à « 18 -> 77 »
- des modifications de configuration unique en plusieurs configurations, « 30 -> 30 ;83 » et « 9 -> 7 ;9 »
- une modification de plusieurs configurations en configuration unique, « 51 ;51 -> 75 »
- des modifications de plusieurs configurations en plusieurs configurations, de « 51 ;51 -> 75 ;76 » à « 67 ;66 ;67 ;66 -> 67 ;66 ;67 »

Les plus importantes modifications représentent 55 % des variations totales de la configuration de la main dominante (mD), quand les plus importantes insertions représentent 3 % et les suppressions les plus importantes représentent 17 %.

Modifications

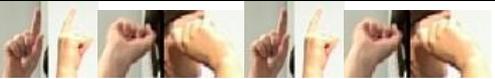
Forme isolée	Forme en contexte	Quantité (total des variations de mD = 117)
		8
		12
		6
		11
		6
		15
		6

Tableau 4 : Main dominante – Modifications les plus importantes

Les principales modifications décrites ci-dessus représentent 78 % des modifications concernant la configuration de mD.



Cette configuration est réalisée en contexte de manière plus relâchée que dans sa forme isolée.

Dans l'instance AS16-8, il s'agit d'une modification de la réalisation du signe « j'ai » lors de la mise en contexte du segment d'énoncé E30 « en raison de la présence de manifestants sur les voies ». Dans toutes les instances de RT20A et RT20B, 6 en tout, il s'agit de la modification de la réalisation du signe « information » lors de la mise en contexte du segment d'énoncé E17 « votre attention s'il vous plaît », et de la modification de la réalisation du signe « dix » lors d'une mise en

contexte (sur deux) du segment d'énoncé E6 « voie dix » (dans RT20B-1 il y a modification, pas dans RT20A-1).

Dans AS16-8, cette modification intervient en contexte entre  et , alors que dans la forme isolée du segment d'énoncé E30, la configuration considérée est entre  et la configuration de repos .

Le relâchement de la configuration précédente en contexte nous laisse penser qu'il pourrait y avoir une influence sur le relâchement de la configuration considérée, d'autant plus vu la nature tendue de la configuration précédente dans la forme isolée ; de plus, la configuration suivante en contexte qui implique un repli des doigts, peut être anticipée dans la configuration considérée ; enfin, le fait que dans la forme isolée la configuration considérée soit la dernière produite peut avoir pour effet une plus grande précision de la configuration, ce qui n'est pas le cas dans la forme en contexte.

Dans les instances de RT20A et RT20B, cette modification intervient en contexte, pour le segment E17, entre  et , alors que dans la forme isolée du segment d'énoncé E17, la configuration considérée est entre  et .

De même que précédemment, nous pensons que le fait que la configuration considérée soit la dernière produite dans la forme isolée amène à une plus grande tension ; cela dit, le fait que la configuration suivante dans la forme en contexte soit relâchée peut avoir une influence sur la configuration considérée.

Dans RT20B-1, cette modification intervient en contexte, pour le segment E6, entre  et , alors que dans la forme isolée du segment d'énoncé E6, la configuration considérée est entre  et .

Le relâchement de la configuration suivante dans la forme en contexte (les index et pouces sont plus arrondis) peut avoir eu une influence sur la configuration considérée, la configuration précédente étant la même nous ne supposons pas d'influence particulière.



Cette configuration est réalisée en contexte de manière plus tendue que dans sa forme isolée.

Cette modification apparaît à 12 reprises dans les instances des énoncés AS16 et IT12AB, dans la mise en contexte de V1 (« Paris Est ») : ce signe ne se fait plus avec mD seule en l'air, mais en utilisant les deux mains. La forme de la configuration en contexte est plus tendue.

Cette modification se retrouvant dans tous les contextes, nous supposons qu'elle est due au signe en lui-même qui est réalisé de manière différente en contexte et dans sa forme isolée.



Cette configuration est réalisée en contexte de manière plus tendue que dans sa forme isolée.

Le pouce est plus écarté de la main (il est quasiment à 90 degrés par rapport aux autres doigts, tout en restant dans le même plan). Il y a 5 occurrences de cette modification dans les instances de AS01 et 1 dans l'instance RT20A-3, lors de la mise en contexte respectivement des segments d'énoncés E13 (« attention au passage d'un train »), et E25 (« en raison de travaux sur les voies »).

Dans les instances de AS01, cette modification intervient en contexte entre  et la configuration de repos , alors que dans la forme isolée du segment d'énoncé E13, la configuration considérée est entre  et la configuration de repos .

La modification intervient en contexte suite à un raccourcissement de la configuration précédente (cf. dans cette partie, plus bas), mais dans la forme isolée comme en contexte, avant une configuration de repos. Bien que la plus grande tension constatée en contexte ne soit pas présente dans chaque instance de AS01 (5 sur 8 seulement), nous pensons que la présence d'une configuration précédente avec les doigts tendus influe la réalisation de la présente configuration. D'autant plus que lorsque le signe précédent n'est pas raccourci, et se termine par une configuration avec les doigts légèrement repliés, comme dans la forme isolée, alors il n'y a pas de tension plus grande de la configuration considérée. Ainsi, notre hypothèse est que la réalisation précédente influe sur la réalisation de la configuration considérée.

Dans RT20A-3, cette modification intervient en contexte entre les configurations  et



, alors que dans la forme isolée du segment d'énoncé E25, la configuration considérée est

entre la configuration de repos  et la même configuration qu'en contexte .

Le segment d'énoncé E25 est également mis en contexte dans les instances RT20B-2, RT20B-3 et AS16-3. Dans RT20B-2 et RT20B-3, nous constatons une modification générale de l'expression lors de la mise en contexte : la configuration considérée disparaît et une autre construction syntaxique

apparaît. Dans RT20B-2, la voie est matérialisée, nommée, puis pointée par , et enfin il est dit qu'il y a des travaux. Dans RT20B-3, le signe est modifié en « j'ai », et la construction est la même que

RT20B-2. Dans AS16-3, il y a ajout d'un pointage , mais la construction reste la même et la configuration n'est pas modifiée.

Au vu de toutes ces variations, il ne nous semble pas possible de proposer une conclusion quand à l'origine de la modification lors de la mise en contexte de cette configuration.



Cette configuration est réalisée en contexte de manière plus relâchée que dans sa forme isolée.

L'index et le majeur ne sont plus dans le même plan mais légèrement décalés en ce sens que l'angle entre le majeur et la paume de la main est inférieur à l'angle entre l'index et la paume de la main, et le pouce n'est pas refermé sur les doigts repliés dans la paume (auriculaire et annulaire, eux-mêmes un peu moins repliés dans la paume) mais légèrement écarté.

Il y a 6 occurrences de cette modification dans les instances AS01, qui concernent la réalisation du signe « train » lors de la mise en contexte du segment d'énoncé E13 ; 4 dans les instances de AS16 dont 3 concernent la réalisation du signe « train » lors de la mise en contexte des segments d'énoncés E26 (« en raison d'un accident de personne ») et E27 (« afin d'assurer les correspondances »), et 1 concerne la réalisation du signe « technique » lors de la mise en contexte du segment d'énoncé E23 (« en raison d'un incident technique ») ; 1 dans l'instance d'énoncé RT20A-1, qui concerne la réalisation du signe « technique » lors de la mise en contexte de l'énoncé E23.

Autrement dit : 2 modifications apparaissent lors de la mise en contexte du signe « technique » (dans une instance de AS16 et dans une instance de RT20A), et 9 modifications apparaissent lors de la mise en contexte du signe « train » (dans trois instances de AS16 et dans six instances de AS01).

Dans les instances de AS01, cette modification intervient en contexte entre les configurations



et



, alors que dans la forme isolée du segment d'énoncé E13, la configuration

considérée est entre



et



Le relâchement de la configuration précédente en contexte, par rapport à une configuration plus tendue dans la forme isolée, nous laisse penser que la configuration précédente en contexte a une influence sur la configuration considérée.

Dans AS16-4, cette modification intervient en contexte entre



et



, de même que dans la forme isolée du segment d'énoncé E26.

Nous ne nous risquons pas à avancer de conclusion dans ce cas, bien que le contexte sémantique semble être la raison de ce relâchement : l'accent est mis sur le fait qu'il y a eu un accident de personne, pas spécifiquement sur le signe train, d'ailleurs repris dans le discours ultérieurement et réalisé par une autre configuration (un proforme, cf. partie suivante sur les catégories de configurations) cette fois tendue.

Dans l'instance AS16-5, cette modification intervient en contexte entre les configurations



et



, et entre



et



, quand dans la forme isolée du segment d'énoncé E27, la

configuration considérée est tout d'abord entre



et



, puis de même qu'en contexte,

entre



et



.

Le contexte des configurations plus relâchée lors de la première occurrence nous laisse à penser qu'au moins la configuration précédente a influé sur la réalisation de la configuration considérée, cependant que le contexte des configurations lors de la seconde occurrence identique à la forme isolée nous laisse supposer que la réalisation de la première occurrence a conditionnée la réalisation de la seconde.

Dans AS16-1 et dans RT20A-1, cette modification intervient en contexte entre  et la configuration de repos , alors que dans la forme isolée du segment d'énoncé E23, la configuration considérée est entre  et la configuration de repos .

Au vu des contextes dans la forme en contexte et dans la forme isolée, il nous est difficile de conclure. Sachant que ces instances d'énoncés n'ont pas été les premiers enregistrés, il se pourrait que machinalement la locutrice ait continué à réaliser les mêmes modifications alors que le contexte ne s'y prêtait pas obligatoirement.



Il s'agit de la suppression de la dernière configuration lors de la réalisation du signe « vaVa » signifiant que quelque chose qu'on vient de citer va arriver : dans notre cas il s'agit d'un train qui va passer en gare, ou qui va partir. Cette modification survient dans 6 instances sur 8 de AS01. La

configuration suivant étant , notre hypothèse est qu'il y a une assimilation entre cette configuration et .



Il s'agit de la suppression de la dernière configuration lors de la réalisation du signe « Strasbourg » : tout comme le signe présenté juste avant (« vava ») c'est un signe comportant une répétition dont la répétition est raccourcie. Sur les 10 occurrences de ce signe dans notre corpus, 7

sont suivies du signe « Reims » qui débute par la configuration , 1 est suivie par le signe

« vaVa » (cf. ci-dessus), et 2 par la configuration  : le raccourcissement est constaté respectivement 5 fois sur 7, 0 fois sur 1, et 1 fois sur 2.

Notre hypothèse est qu'il y a une assimilation avec la configuration suivante dans le cas des

configurations  et .



Cette configuration est réalisée en contexte de manière plus relâchée que dans sa forme isolée.

Il s'agit dans le cas de la modification en « 75 » d'un relâchement des doigts et d'un positionnement du pouce moins dans la paume de la main, et dans la modification en « 76 » du même relâchement des doigts mais cette fois le pouce n'est plus dans la paume de la main mais parallèle à l'index. Toutes les occurrences de cette modification se retrouvent dans les instances de AS01. L'énoncé concerné est toujours E12 et la variation intervient lors de la réalisation du signe « quai ».

Dans les instances de AS01, cette modification intervient en contexte entre différentes configurations (de numéro de voie) et , alors que dans la forme isolée du segment d'énoncé E12, la configuration considérée est entre la configuration de repos  et la même configuration qu'en contexte .

Les configurations de numéros de voie qui précèdent les configurations considérées sont exprimées de manière claire et précise, sans modification entre leur forme isolée et leur forme en contexte (sauf pour E6 « voie dix » où il y a un relâchement de configuration), et la configuration suivante est identique en contexte ou dans la forme isolée. Nous supposons ici que le quai n'était pas l'information principale (pas en soi, c'est la foule qui est dessus qui est importante puisqu'elle doit ne pas se trouver sur la bordure de ce quai), donc il est exprimé de manière plus relâchée en contexte : nous pensons que c'est là un effet dû au contexte sémantique plus qu'à la coarticulation.

Insertions

Forme en contexte	Quantité (total des variations de mD = 117)
	3

Tableau 5 : Main dominante – Insertion la plus importante

L'insertion principale décrite ci-dessus représente 50 % des insertions concernant la configuration de mD.

Il s'agit d'une configuration de pointage. On retrouve 2 occurrences de cet ajout dans les instances de l'énoncé AS16 (lors de la mise en contexte des segments d'énoncés E25 « en raison de travaux sur les voies » dans AS16-3 et E29 « en raison d'un incident affectant la voie » dans AS16-7, où il y a respectivement un pointage vers la voie qui est en travaux et un pointage vers la voie où il y a un incident), et 1 occurrence dans DOY (suite à la mise en contexte du segment d'énoncé V15 « Reims » dans DOY-1, il y a pointage vers le train qui va partir). Dans l'ordre des instances

présentées, on retrouve ce pointage respectivement entre les configurations  et , entre



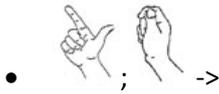
et , entre  et . Si les deux premiers contextes nous laissent à penser qu'il ne s'agit pas d'un phénomène de coarticulation mais plus d'une insistance due au contexte, la troisième, d'autant plus de par sa courte durée (cf. partie suivante sur les analyses de durées), pourrait être due à la coarticulation.

Suppressions

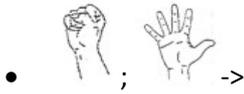
Forme isolée	Quantité (total des variations de mD = 117)
	8
	12

Tableau 6 : Main dominante – Suppressions les plus importantes

Les suppressions principales relevées représentent 17 % des variations totales (modifications et insertions comprises), et 71% des suppressions concernant les configurations de mD.



Cette suppression intervient dans toutes les instances de AS01 : il s'agit du signe « attention » qui disparaît lors de la mise en contexte du segment d'énoncé E13 « attention au passage d'un train ».



Il y a 10 occurrences de cette suppression dans les instances de AS16 et 2 occurrences dans RT20AB : il s'agit du signe « information » qui disparaît lors de la mise en contexte du segment d'énoncé E17 « votre attention s'il vous plait ».

Ces suppressions ne sont pas dues à la coarticulation mais au contexte sémantique, puisque les signes de ces configurations ne sont plus réalisés lors de la mise en contexte du segment d'énoncé correspondant.

Synthèse

Nous constatons que pour la main dominante, la majorité des variations sont des modifications, qu'il y a une quantité moindre d'insertions, et peu de suppressions. Le phénomène de coarticulation ne se retrouve que dans les modifications de configuration, pas dans les insertions ni les suppressions.

Nous constatons que dans un premier temps il y a une influence du contexte sémantique, qui impose des configurations tendues (identiques à la forme dans les segments d'énoncés) pour certaines catégories de signes, tels les chiffres et les gares, ainsi que lors de la réalisation des informations importantes, et des configurations plus relâchées pour les informations moins importantes, ne nécessitant pas forcément d'être réalisées en détail mais plutôt dans une certaine forme (au sens forme de la main) générale. Ensuite, il y a une influence du contexte phonétique, qui s'applique par persévération : lors de la mise en contexte, si la configuration précédente est tendue ou relâchée, alors la configuration mise en contexte sera réalisée de manière plus tendue ou plus relâchée que dans sa forme isolée. Nous remarquons également que lors de la répétition de segments d'énoncés en contexte (par exemple les 2 occurrences des segments E1 à E11 dans AS01), la seconde occurrence subit les mêmes variations que la première : il y a donc une influence là encore par persévération.

De la même façon que pour les graphiques précédents, les variations présentées sont regroupées, autant dans leur nature (modification, insertion ou suppression) que dans la quantité des éléments qu'ils impliquent. Nous avons réalisé les mêmes types de regroupement que pour la configuration de mD, et observons :

- des insertions de configuration unique, « -> 53 » et « -> 6 »
- des suppressions de configuration unique, de « 2_2 -> » à « 7 -> »
- des suppressions de plusieurs configurations, « 30 ; 8 -> » et « 7 ;13 -> »
- des modifications de configuration unique à configuration unique, de « 57 -> 80 » à « 51 -> 76 »
- des modifications de configuration unique en plusieurs configurations, « 53 -> 2_2 ; 53 » et « 53 -> 53 ; 54 »
- une modification de plusieurs configurations en plusieurs configurations, « 63 ; 13 -> 7 ; 79 »

Les plus importantes modifications représentent 35 % des variations totales de la configuration de la main dominée (md), quand les plus importantes insertions représentent 16.25 % et les suppressions les plus importantes représentent 25 %.

Modifications

Forme isolée	Forme en contexte	Quantité (total des variations de md = 80)
		9
		19

Tableau 7 : Main dominée – Modifications les plus importantes

Les principales modifications décrites ci-dessus représentent 68.3 % des modifications de la configuration de la main dominée (md).



Cette configuration est réalisée en contexte de manière plus relâchée que dans sa forme isolée. On retrouve ce relâchement au niveau de mD.

L'index et le majeur ne sont plus dans le même plan mais légèrement décalés en ce sens que l'angle entre le majeur et la paume de la main est inférieur à l'angle entre l'index et la paume de la main, et le pouce n'est pas refermé sur les doigts repliés dans la paume (auriculaire et annulaire, eux-mêmes un peu moins repliés dans la paume) mais légèrement écarté. Il y a 6 occurrences de cette modification dans les instances de AS01 lors de la mise en contexte de E13 « attention au passage d'un train », et 3 occurrences dans les instances de AS16 (1 lors de la mise en contexte de E26 « en raison d'un accident de personne », 2 lors de la mise en contexte de E27 « afin d'assurer les correspondances »).

Dans les instances de AS01, cette modification intervient en contexte entre les configurations  et une configuration de repos () , alors que dans la forme isolée du segment d'énoncé

E13, la configuration considérée est entre  et la même configuration de repos .

Dans la forme en contexte, la configuration suivante est la même que dans la forme isolée, mais la configuration précédente est plus relâchée quand dans la forme isolée cette dernière est tendue : il pourrait y avoir une influence de la part de la configuration précédente lors de la mise en contexte.

Dans l'instance AS16-4 C, cette modification intervient en contexte entre la configuration de repos  et  , quand dans la forme isolée du segment d'énoncé E26, la configuration considérée est entre une autre configuration de repos  et la même configuration qu'en contexte .

Dans la forme isolée comme dans la forme en contexte, la position qui précède est une position de repos et la position qui suit est la même : la configuration de mD qui est elle aussi réalisée de manière plus relâchée à ce moment pourrait être à l'origine de la modification constatée pour md.

Dans l'instance AS16-5, cette modification intervient en contexte entre les configurations entre  et  , et entre  et  , quand dans la forme isolée du segment d'énoncé E27, la

configuration considérée est tout d'abord entre la configuration de repos  et , puis

entre  et .

Là encore, vu les configurations environnantes en contexte et dans la forme isolée, il semblerait que ce soit la modification de la configuration de mD qui entraîne une modification de la configuration de md. Il se pourrait également que le contexte sémantique y soit pour quelque chose, dans la mesure où la focalisation du message n'est pas sur le train en lui-même, mais sur le fait qu'il y ait une correspondance entre les trains (donc l'accent serait plutôt mis sur la correspondance).



Cette configuration est réalisée en contexte de manière plus relâchée que dans sa forme isolée. Là encore, on retrouve ce relâchement au niveau de mD.

Il s'agit d'un relâchement des doigts et d'un positionnement du pouce moins dans la paume de la main. Il y a 16 occurrences de cette modification sur 16 mises en contexte de E12 « éloignez-vous de la bordure du quai » dans les instances de AS01, et 3 occurrences sur 3 dans IT12AB là encore lors de la mise en contexte de E12.

Dans les instances de AS01, il y a une configuration différente qui précède celle considérée (puisque dans chaque instance le numéro de la voie est différent), et dans les instances de AS16 la configuration qui précède est une configuration de repos () , mais dans toutes ces instances, la

configuration qui suit est toujours .

Dans sa forme isolée, la configuration considérée est entre une configuration de repos ()

et .

Le contexte des configurations est donc le même en contexte et dans la forme isolée, ce qui nous laisse supposer que cette modification n'est pas due à la coarticulation des configurations immédiatement voisines, voire due au contexte sémantique.

Forme en contexte	Quantité (total des variations de md = 80)
	12
	1

Tableau 8 : Main dominée – Insertions les plus importantes

Les principales insertions sont les seules insertions de la configuration de la main dominée (md).



Cette insertion est due au fait que le signe V1 « Paris Est » n'est pas réalisé de la même manière en contexte que sous sa forme isolée. La main dominée sert de base à une configuration mD en forme de « P », plutôt que le signe soit réalisé avec cette même configuration « P » (ou approchant) sans l'aide de md.

Cette insertion se retrouvant dans tous les contextes où apparaît le segment d'énoncé v1, nous supposons qu'elle est due au signe en lui-même qui est réalisé de manière différente en contexte et dans sa forme isolée.



Cette insertion unique intervient dans une instance de RT20A, et est due à l'ajout du signe « problème » lors de la mise en contexte de « en raison d'une panne » (le même ajout se produit pour mD, ce signe se réalisant avec les deux mains).

Cette insertion n'est pas due à la coarticulation, mais à l'ajout d'un signe qui n'était pas présent dans la forme isolée.

Suppressions

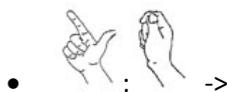
Forme isolée	Quantité (total des variations de md = 80)
	9
	8
	3

Tableau 9 : Main dominée – Suppressions les plus importantes

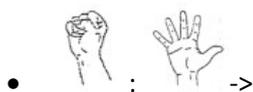
Les suppressions principales décrites ci-dessus représentent 77% des suppressions concernant la configuration de md.



Il s'agit d'une suppression du signe « information » (qui se retrouve aussi au niveau de mD, lors de la suppression des configurations « 7 ;13 »), qui intervient dans toutes les instances de AS16 lors de la mise en contexte de l'énoncé E17.



Il s'agit du signe « attention » qui est supprimé (les mêmes configurations sont supprimées pour mD) dans tous les instances de AS01, lors de la mise en contexte de l'énoncé E13 « attention au passage d'un train ».



Il s'agit d'une suppression du signe « ça arrive » (la même suppression des mêmes configurations intervient pour mD). Il n'y a que dans une instance de AS16 qu'on peut noter cette suppression, car la configuration 13 est transformée en 53 dans les autres énoncés où le signe « ça arrive » est mis en contexte : il n'y a donc pas suppression de la configuration mais modification. Il y a également 2

occurrences de cette suppression dans les instances de RT20AB, lors de la mise en contexte de l'énoncé E23, dues à la suppression de ce même signe.

Ces suppressions ne sont pas dues à la coarticulation, mais à l'ajout d'un signe qui n'était pas présent dans la forme isolée des segments considérés.

Synthèse

Nous constatons que, comme pour mD, il y a une majorité de modifications par rapport aux insertions et suppressions, et que le phénomène de coarticulation n'intervient que pour les modifications.

Tout comme mD, seuls les énoncés E12 et 13 subissent des effets de la coarticulation.

Nous remarquons également une influence de la réalisation de mD sur md.

Nous rajoutons donc une étape par rapport à ce que nous avons élaboré précédemment : il y a tout d'abord une mise en contexte sémantique, qui implique certaines variations, puis une influence de mD sur md, ensuite une mise en contexte phonétique, et une influence de mD sur md.

2. Par variations de segments d'énoncés

Dans cette partie nous analysons différents cas de variation des segments d'énoncés mis en contextes, entre ceux qui ne varient pas lors de leur mise en contexte, quelque soit le contexte, et ceux qui subissent des variations. Pour ces derniers, nous avons également différencié les segments d'énoncés variant suivant le contexte de ceux dont les variations ne dépendent pas du contexte.

a. Segments d'énoncés qui ne varient pas lors de leur mise en contexte

Tout d'abord nous présentons les segments d'énoncés qui ne subissent aucune variation lors de leur mise en contexte, et ce quelque soit le contexte. Il y a assez peu de segments d'énoncés dans cette situation : il s'agit principalement des gares et des heures, et également de 2 signes de gare et d'une numérotation de voie (E3 « voie C »).

Gares

Les signes pour « Metz » et « Colmar », qui sont respectivement mis en contexte dans 1 et 3 énoncés, ne varient pas.

Ces signes de gares n'ont pas de spécificité particulière qui ferait qu'ils ne varieraient pas en contexte. Si nous reprenons la liste des gares contenues dans notre corpus, nous constatons que « Paris Est » et « Mulhouse » ont une forme en contexte différente de leur forme isolée car le signe en lui-même est différent (cf. parties suivantes), que « Strasbourg » subit un raccourcissement

de sa seconde répétition (cf. partie précédente), et que « Reims » ne subit qu'une modification de la direction du regard, mais qui n'est pas due au signe en lui-même ou à ses configurations, mais plus au contexte dans lequel il est placé (cf. par la suite).

Nous en concluons que selon nous au niveau de la configuration des mains au moins il n'y a pas d'effet de la coarticulation constaté.

Chiffres et heures

Les chiffres « 1 », « 21 », « 200 », « 400 », « 500 », « 1000 », « 4000 » et « 7000 » ne varient pas malgré la diversité des énoncés dans lesquels ils sont mis en contexte (respectivement dans 2, 2, 7, 2, 6, 8, 7 et 3 contextes différents) et leur place dans la succession de chiffres. De même pour « 12 », « 33 », « 300 », « 900 », « 8000 » à la différence que ces chiffres ne sont mis que dans un unique contexte.

La réalisation des heures « 13h », « 12h », « minuit » ne subit pas non plus de variation.

Comme nous allons le voir dans la suite, les seuls chiffres à subir des variations sont « 2000 » et « 3000 ». Dans le cas du second il s'agit d'une réalisation différente en contexte, pas forcément due à la coarticulation, quand dans le premier cas il y a un phénomène de coarticulation.

Comme pour les signes de gare, nous estimons que la coarticulation n'a pas de réel effet sur la réalisation des chiffres. Ceci est d'autant plus compréhensible que les noms des gares et les chiffres donnés (pour les numéros de train ou pour les horaires) doivent être précis afin de ne pas laisser la place à une ambiguïté : l'effet de la coarticulation est ainsi annihilé par la locutrice qui « s'applique » lors de la réalisation de ces informations.

Autre

Le segment d'énoncé E3 « voie C » ne varie pas lors de son unique mise en contexte.

Comme nous allons le voir dans la suite, il semblerait qu'ici il aurait dû y avoir ajout d'une direction du regard en direction de la voie en question. Le fait que ce soit le seul segment d'énoncé de ce type à ne pas subir de variation nous laisse penser à un phénomène de variabilité intra locutrice.

b. Segments d'énoncés qui varient lors de leur mise en contexte

Dans un premier temps nous présentons les segments d'énoncés qui varient toujours de la même manière quelque soit le contexte dans lequel ils sont placés. Ensuite nous présentons les segments d'énoncés qui subissent ou non des variations en fonction de leur mise en contexte, mais s'ils varient alors c'est toujours de la même manière quelque soit le contexte, sinon ils ne varient pas. Puis nous présentons les segments d'énoncés qui subissent des variations lors de leur mise en

contexte qui se retrouvent dans les différentes mises en contexte, bien que certaines diffèrent (très) peu d'un contexte à l'autre. Enfin, nous présentons un segment d'énoncé dont les variations sont très dépendantes du contexte dans lequel il est placé.

Les tableaux suivants présentent différents segments d'énoncés selon leur référence dans notre corpus (dans la colonne référence), les variations subies par ces segments d'énoncés, en nature (la composante qui est affectée) et en valeur (les valeurs qui varient), et selon le nombre de contextes dans lesquels on retrouve ces variations par rapport au nombre total d'énoncés dans lesquels sont insérés ces segments d'énoncés.

Pour faciliter la lecture de ces tableaux, nous ne faisons pas figurer comme précédemment les icônes dont nous nous sommes servis pour annoter, mais la chaîne de caractère les représentants dans les fichiers d'annotation (pour une correspondance avec les icônes d'annotation, cf. Annexe II).

De plus, nous utilisons les raccourcis suivants : « regdir » désigne la direction du regard, « md » la main dominée et « mD » la main dominante.

Segment d'énoncé dont un ou plusieurs composantes varient toujours de la même façon quelque soit le contexte

Référence	Variations		Nombre de contextes	
	Nature	Valeur	Variation	Total
E1	regdir	-> 1	3	3
E2	regdir	-> 1	2	2
E5	regdir	-> 2	2	2
E7	regdir	-> 1	2	2
E8	regdir	-> 2	2	2
E11	regdir	-> 1	2	2
V1	mD	50 -> 73	12	12
	md	-> 53		
V20	mD	9 -> 7 ;9	2	2
2000	mD	30 -> 30 ;83	1	1
	md	2_2 -> 53		
3000	mD	71 -> 22	1	1

Tableau 10 : Variations des segments d'énoncés qui varient toujours de la même façon

E1, E2, E5, E7, E8 et E11 sont des énoncés de type « voie ... » (respectivement, « voie A », « voie B », « voie 6 », « voie 17 », « voie 21 » et « voie 76 »). Ces énoncés ne subissent qu'une seule variation au niveau de la direction du regard : il y a ajout d'un regard vers l'espace au début de la réalisation en contexte, concrètement pour positionner tout d'abord du regard le quai dans l'espace, qui est ensuite positionné au même endroit de l'espace avec les mains. Ces variations ne sont pas dues à la coarticulation mais à la sémantique du contexte dans lequel ils sont insérés.

V1 « Paris Est » et V20 « Mulhouse » subissent des variations dues au fait que le signe en lui-même change. Ce ne sont donc pas là des variations dues au phénomène de coarticulation.

Les chiffres 2000 et 3000 sont réalisés de manière différente en contexte et isolés : la configuration change pour la réalisation de « 2000 » en contexte (il s'agit plus d'une modification que de l'ajout d'une configuration), alors que la forme isolée de « 3000 » est réalisée en un seul signe quand la forme en contexte nécessite la réalisation de deux signes : « 3 » puis « mille ». La modification en question pour la réalisation en contexte de « 2000 » est selon nous due à un phénomène de coarticulation.

Segments d'énoncé qui varient ou pas, mais si ils varient alors c'est toujours de la même manière quelque soit le contexte

Référence	Variations		Nombre de contextes	
	Nature	Valeur	Variation	Total
E9	regdir	-> 1	1	2
E10	regdir	-> 1	1	2
E21	md	53 -> 2_2	2	6

Tableau 11 : Variations des segments d'énoncés qui s'ils varient le font toujours de la même façon

E9, et E10 sont des segments d'énoncés de type « voie ... » (respectivement, « voie 27 » et voie 71 »). Ils varient en contexte mais une fois sur deux : lorsqu'ils sont placés dans une instance de AS01, ils apparaissent au début et à la fin de l'instance, tout comme les énoncés E1 à E11. E9 se voit ajouter une direction du regard vers l'espace (la voie dont il est question dans l'énoncé) à sa première occurrence mais pas à la seconde, quand E10 se voit ajouter la même direction du regard à sa seconde occurrence et pas à sa première.

Même si ces énoncés ne rencontrent pas toujours la même variation, nous pensons qu'ils sont à regrouper avec les autres énoncés du même type vus précédemment (de E1 à E11) : il s'agit d'une variation due au contexte sémantique, pas à la coarticulation.

E21 « prenez garde à la fermeture des portes » est modifié dans un tiers des contextes où il apparaît (dans 2 instances de DT02) : il s'agit de md qui lors de la réalisation du signe signifiant « prenez garde » adopte une configuration avec le pouce plus écarté des autres doigts de la main que dans la configuration réalisée de manière isolée. Cette variation pourrait être due à la coarticulation, mais ce résultat n'est pas forcément significatif puisque 2 instances sur 3 de l'énoncé DT02 varient, mais 3 instances sur 3 de l'énoncé DOY ne varient pas

Segments d'énoncé qui varient de façon «quasi» identique quelque soient les contextes

Un segment d'énoncé dont les variations sont considérées comme étant « quasi identiques » dans les différentes mises en contexte est un segment dont la grande majorité des variations se retrouvent dans la majorité des mises en contexte de ce segment.

Référence	Modifications		Nombre de contextes	
	Nature	Valeur	Variation	Total
E6	regdir	-> 1	3	4
	mD	32 -> 68	4	
	md	32 -> 68	4	
E12	regdir	1 ;face ;2 -> ...	15	19
	regdir	md ->... (10 / 8)	19	
	regdir	mDmd -> ...	18	
	mD	51 ;51 -> 75 ;76	15	
	md	51 -> 75	18	
E13	regdir	2 -> 9	8	8
	mD	30 ;8 ->	8	
	mD	55;56;55;56 -> 55;56;55	6	
	mD	53 -> 2_2	5	
	mD	18 -> 77	6	
	md	18 -> 77	6	
E23	mD	7 ;13 ->	3	3
	md	7 ;13 ->	3	
V15	regdir	-> 10	1	11
	mD	-> 10	1	
V22	mD	67;66;67;66 -> 67;66;67	6	10

Tableau 12 : Segments d'énoncés qui varient de manière « quasi » identique quelque soit le contexte

E6 « voie 10 » varie presque tout le temps de la même manière, sauf dans une instance où il n'y a pas d'ajout de direction du regard vers un emplacement de l'espace lors de la mise en contexte de l'énoncé. Dans la même logique que précédemment, nous considérons cette variation comme celle subie par les énoncés E1 à E11, non due au phénomène de coarticulation mais plutôt au contexte sémantique dans lequel ces segments sont insérés.

E12 « éloignez-vous de la bordure du quai » subit globalement les mêmes variations au niveau du regard (vers l'espace plutôt que vers l'interlocuteur ou les mains) et des configurations des mains (configurations plus relâchées) quelque soit le contexte. Il en va de même pour E13 « attention au passage d'un train » : élévation de la direction du regard, suppression et raccourcissement de signes, relâchement ou tension des configurations des mains. De ce point de vue, nous pourrions penser que

ces variations sont dues au contexte sémantique, mais les résultats de la partie précédente (par composantes) nous permettent d'avancer que certaines variations (notamment raccourcissement des signes – disparition d'une configuration dans un ensemble comportant une répétition, un doublon de configuration – et relâchement ou tension de la configuration des mains) relèvent de la coarticulation.

E23 subit une suppression de manière systématique (signe « information »), ainsi que d'autres modifications qui ne sont pas dépendantes du contexte : une modification apparaît dans une instance de RT20A mais pas dans une instance de RT20B alors qu'on la retrouve dans une instance de AS16.

V15 « Reims » varie en configuration et direction du regard (pour la direction du regard, il apparaît que c'est surtout un pointage ajouté entre deux énoncés coarticulés, E15 et E21, plus qu'une variation de l'énoncé en lui-même).

V22 « Strasbourg » varie en configuration (raccourcissement du signe par omission de la dernière configuration).

En ce qui concerne le segment d'énoncé E23 et ces signes de gare, comme nous l'avons vu précédemment, nous considérons que ces variations ne relèvent pas de la coarticulation.

Segment d'énoncé dépendant du contexte

Référence	Modifications		Nombre de contextes	
	Nature	Valeur	Variation	Total
E17	mD	13 -> 79	6	6 (RT20AB)
	mD	7 ;13 ->	9	9 (AS16)
	md	7 ->	9	
	md	53 -> 2_2	8	8 (AS16)

Tableau 13 : Variations du segment d'énoncé dépendant du contexte

Dans le segment d'énoncé E17 (« votre attention s'il vous plait »), il apparaît clairement que les variations sont dues au contexte dans lequel il va être inséré. En effet, dans les deux contextes où il est présent (les instances de RT20AB et celles de AS16), il subit au minimum soit une modification de la configuration de mD, soit une suppression de configuration de mD et de md. Ce segment subit également d'autres variations que celles relevées dans le tableau mais elles ne sont pas systématiques dans les différentes instances (un exemple indiqué ici est la modification de la configuration de md de « 53 » en « 2_2 » qui se retrouve dans toutes les instances de AS16 sauf une).

Comme nous l'avons vu précédemment, ce segment d'énoncé semble subir des effets de la coarticulation, due au contexte dans lequel il est inséré (en fonction des configurations voisines).

Autres segments d'énoncé

Le segment d'énoncé E4 est présent dans 2 contextes d'une instance (à l'instar des E1 à E11), et se voit ajouter une direction du regard vers l'espace (la voie considérée dans l'énoncé) à chacune de ses mises en contexte, mais le regard n'est pas dirigé exactement vers le même emplacement à chaque mise en contexte.

Les segments d'énoncé E24 à E31 subissent diverses variations, dans plus ou moins de contextes différents. Il n'est pas possible d'établir une généralité sur ces variations, sauf à dire que la direction du regard est modifiée (de plusieurs manières différentes) et que les configurations des mains, mD ou md, sont réalisées de manière plus relâchée en contexte que sous forme isolée.

Synthèse

Nous constatons que les signes des chiffres et des gares ne subissent aucune variation due au contexte phonétique, mais varient suivant le contexte sémantique. Ceci est du selon nous à l'importance qu'ont ces informations dans le domaine de la SNCF.

Les segments d'énoncés E1 à E11 varient par insertion d'une direction du regard, qui selon nous est due au contexte sémantique, pas au contexte phonétique.

Les segments E12 et E13 subissent des effets des deux contextes, de même que l'énoncé E17.

Les segments E23 à E31 subissent des effets de la mise en contexte sémantique, qui sont des relâchements de certaines configurations.

3. Par catégories de configurations

Nous reprenons ici les catégories proposées par Boutora [BOU08], qui s'appuient sur les configurations relevées dans Cuxac [CUX00] et la classification proposée par Dubuisson [DUB99]. Cette classification est représentée sur le schéma suivant (Figure 20).

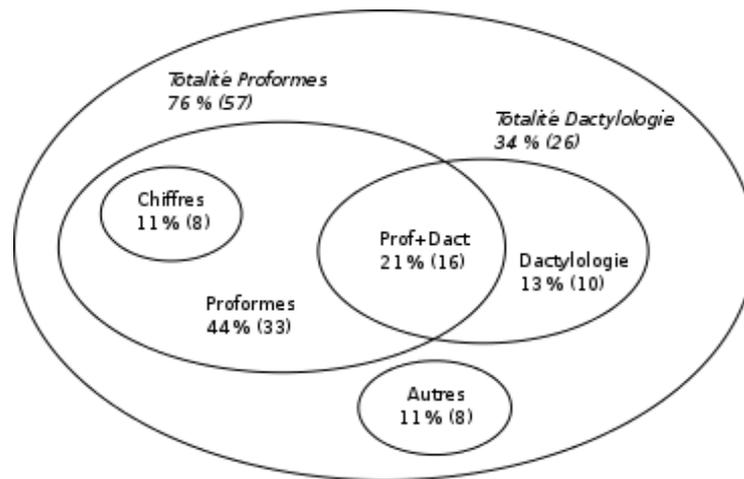


Figure 20 : « Distribution des 75 configurations manuelles de la LSF »

Reprenant la notion de catégorie telle que présentée, nous avons classé nos configurations selon qu'elles sont utilisées pour réaliser de la dactylogogie, des chiffres, des pointages, des signes du lexique et/ou des proformes. Nous définissons ces catégories ainsi :

- La dactylogogie est la réalisation de lettres de l'alphabet de la langue française en LSF.
- Les chiffres sont la réalisation de données numériques.
- Les pointages servent à désigner des endroits de l'espace de signation.
- Les signes du lexique sont des signes qu'on trouve dans les dictionnaires existants.
- Les proformes sont « des configurations manuelles qui entretiennent un rapport iconique avec leur référent » (Boutora [BOU08]).

Les proformes sont des configurations qu'on peut retrouver dans des signes du lexique, et également dans les expressions iconiques non lexicalisées, de même que lors de l'expression de dactylogogie ou de chiffres.

Pour information, voici les regroupements que nous obtenons, visualisés avec les symboles alphanumériques d'annotations plutôt que les images pour une meilleure visualisation quantitative.

Dactylogogie (D)	2_3, 5_2, 17, 22, 23_2, 38, 51, 73
Chiffres (C)	4_4_2, 12, 13, 14, 17, 22, 23, 24, 27, 30, 32, 51, 57, 60, 68, 70, 71, 74
Pointages (Po)	10, 29
Signes du lexique (S)	2_2, 2_3, 5_2, 6, 7, 8, 9, 10, 11, 12, 13, 16, 17, 18, 19_2, 20, 21, 26, 30, 36, 39_2, 39_2_1, 39_2_2, 38, 50, 51, 53, 55, 56, 57, 58, 59, 60, 61, 62, 63, 65, 66, 67, 73, 76, 77, 78, 79, 82, 83
Proformes (Pr)	2_2, 5_2, 12, 13, 19_2, 27, 51, 52, 54, 64, 75, 76, 80, 81

Tableau 14 : Regroupement des configurations relevées, par catégories

Si nous nous concentrons sur les principales variations citées en partie précédente, et que nous ne retenons que celles qui selon nous sont dues à la coarticulation, nous remarquons qu'il n'y a aucune configuration concernée en suppression ou en insertion. Par contre, il y en a quelques unes pour les modifications, que nous listons ci-après (nous présentons l'annotation en image de la configuration pour plus de clarté visuelle, puis entre parenthèse l'annotation en symbole pour pouvoir faire le lien avec le tableau ci-dessus, enfin la ou les catégories auxquelles appartiennent ces configurations d'après le tableau ci-dessus) :

- mD

-  (13) ->  (79) : S
-  (53) ->  (2_2) : S
-  (18) ->  (77) : S
-  ;  ;  ;  (55 ;56 ;55 ;56) ->  ;  ;  (55 ;56 ;55) : S
-  ;  ;  ;  (66 ;67 ;66 ;67) ->  ;  ;  (66 ;67 ;66) : S

- md

-  (18) -> (79) : S

Nous remarquons que seules des configurations d'expression de signes du lexique sont affectées par le phénomène de coarticulation. Pour mD, il y a deux relâchements de configurations (13 -> 79 et 18 -> 77), une tension (53 -> 2_2), et deux raccourcissements, quant pour md il y a un relâchement.

La configuration 13 se retrouve dans d'autres catégories que celles des signes lexicaux : elle est utilisée pour les chiffres et en tant que proforme, cependant ses modifications n'ont lieu que lors de son utilisation au sein d'une expression de signes lexical.

Synthèse

Ceci nous amène à penser que la coarticulation n'affecte dans notre domaine que des signes du lexique. Cela rejoint l'idée émise précédemment que pour les chiffres et la dactylologie il est important d'être précis sur la configuration en elle-même, donc que la forme en contexte subit peu d'effet de coarticulation (même si la présence de coarticulation peut aider à une meilleure perception du sens, elle n'aide pas à la perception d'un élément en particulier).

Nous remarquons également que les configurations initialement réalisées de manière tendues dans les segments d'énoncés deviennent plus relâchées si dans le contexte sémantique ce ne sont pas les informations importantes de l'énoncé final. Les configurations répétées raccourcissent lorsqu'elles sont suivies de configurations phonétiquement proches (quand par exemple une configuration répétée se terminant par un index tendu précède une configuration avec l'index et le majeur tendus et croisés, alors la dernière configuration index tenu disparaît).

II. Variations temporelles

Nous présentons maintenant des comparaisons de durées des différentes composantes lorsqu'elles sont réalisées dans les segments d'énoncés (leur forme « isolée ») et dans les énoncés (leur forme « en contexte »). Ces comparaisons ont été réalisées sur les composantes dont les valeurs n'ont pas variées lors de la mise en contexte : en effet, il nous semble délicat de comparer des éléments dont la valeur a changé, ce n'est alors parfois plus le même élément linguistique et il ne peut pas systématiquement être comparé à une réalisation isolée.

Nous avons réalisé différentes comparaisons, par énoncés complets et par segments d'énoncés. Nous avons considéré plusieurs seuils de variation de la durée d'une valeur d'une composante : 0%, 5%, 10% et 20%. En cela nous avons considéré que la variation de la durée d'une valeur d'une composante, entre sa forme isolée et en contexte, pouvait être prise en compte à partir du moment où elle était réduite ou augmentée de respectivement 0%, 5%, 10% ou 20%.

Par la suite, nous désignons par le terme « élément » une direction de regard ou une configuration de main.

Pour faciliter la lecture des tableaux suivants, nous proposons de détailler le premier tableau afin d'explicitier clairement ce que nous indiquons comme information.

Les colonnes « Seuil » indiquent qu'à partir d'une certaine valeur les données ont été prises en compte. Ainsi, « 0 % » signifie que toutes les valeurs ont été prises en compte, « 5% » signifie que les valeurs ayant variées de 5% ou plus sont prises en compte, etc. La variation considérée est un pourcentage de variation de durée entre la réalisation d'un élément sous forme isolée et sa réalisation en contexte : pour un seuil de 10%, n'ont été pris en compte que les éléments dont la durée en contexte était de 10% (ou plus) supérieure ou inférieure à la durée isolée.

La première ligne « Quantité d'éléments dont la valeur est identique isolée et en contexte mais dont la durée varie » indique le nombre d'éléments qui ont été considérés dans les calculs, c'est-à-dire ceux dont la valeur n'a pas varié entre la réalisation isolée et la réalisation en contexte, mais dont la variation de durée est supérieure ou égale au seuil.

La seconde ligne « % d'éléments dont la durée varie (par rapport à la quantité totale des éléments de l'énoncé) » représente le pourcentage d'éléments dont la durée a varié, mais pas la valeur, et qui sont considérés par rapport au nombre d'éléments total de l'énoncé dont la durée a varié mais pas la valeur : pour le seuil « 0% », la valeur est logiquement toujours égale à 100%.

La troisième ligne « Moyenne de la variation des durées » indique la moyenne de la variation des durées pour l'ensemble des éléments considérés, donc en fonction du seuil donné. Un résultat de cette moyenne positif signifie que la durée moyenne des éléments en contexte est supérieure à celle des éléments dans les segments d'énoncés, un résultat négatif signifie le contraire.

La dernière ligne « Quantité d'éléments dont la durée a » indique en deux lignes le nombre d'éléments dont la durée a augmentée et le nombre d'éléments dont la durée a diminuée, toujours en fonction du seuil donné.

1. Par énoncés complets

AS01		Seuil			
		0 %	5%	10%	20%
Quantité d'éléments dont la valeur est identique isolé et en contexte mais dont la durée varie		241	206	162	103
% d'éléments dont la durée varie (par rapport à la quantité totale des éléments de l'énoncé)		100 %	85 %	67 %	43 %
Moyenne de la variation des durées		+ 0,7 %	+ 0,83 %	+ 1,14 %	+ 5,56 %
Quantité d'éléments dont la durée a	augmenté	114	87	67	51
	diminué	127	119	95	52

Tableau 15 : AS01 - Variations en durée, suivant plusieurs seuils, des éléments dont la valeur ne varie pas

Plus de la moitié (57%, valeur complémentaire des 43% de la dernière colonne) des éléments varient en durée de moins de 20% lors de leur mise en contexte, par rapport à leur réalisation isolée.

Pour les seuils de 0 à 10%, la majorité des éléments a une durée qui diminue, cependant en moyenne la durée de l'ensemble des éléments reste la même : les éléments dont la durée augmente ont donc une variation de leur durée plus importante que la variation de la durée des éléments qui diminuent. Si on considère le seuil de 20%, le nombre d'éléments dont la durée augmente est égal au nombre d'éléments dont la durée diminue cependant la durée moyenne des éléments augmente : là encore, la variation de durée des éléments dont la durée augmente est supérieure à la variation des éléments dont la durée diminue.

La majorité des éléments dont la durée augmente (67 sur 114) ou diminue (95 sur 127) ont des variations de 10% ou moins.

AS16		Seuil			
		0 %	5%	10%	20%
Quantité d'éléments dont la valeur est identique isolé et en contexte mais dont la durée varie		83	73	68	46
% d'éléments dont la durée varie (par rapport à la quantité totale des éléments de l'énoncé)		100 %	88 %	82 %	55 %
Moyenne de la variation des durées		+ 6,99 %	+ 7,91 %	+ 8,38 %	+ 5,94 %
Quantité d'éléments dont la durée a	augmenté	38	29	26	20
	diminué	45	44	42	26

Tableau 16 : AS16 - Variations en durée, suivant plusieurs seuils, des éléments dont la valeur ne varie pas

La moitié (45%, valeur complémentaire des 55% de la dernière colonne) des éléments varient en durée de moins de 20% lors de leur mise en contexte, par rapport à leur réalisation isolée.

Quelque soit le seuil considéré, la majorité des éléments a une durée qui diminue, cependant en moyenne la durée de l'ensemble des éléments augmente : les éléments dont la durée augmente ont donc une variation de leur durée beaucoup plus importante que la variation de la durée des éléments qui diminuent.

La majorité des éléments dont la durée augmente (26 sur 38) ou diminue (42 sur 45) ont des variations de 10% ou moins.

IT12AB		Seuil			
		0 %	5%	10%	20%
Quantité d'éléments dont la valeur est identique isolé et en contexte mais dont la durée varie		40	33	31	21
% d'éléments dont la durée varie (par rapport à la quantité totale des éléments de l'énoncé)		100 %	82 %	77 %	52 %
Moyenne de la variation des durées		- 7.53%	- 9.13 %	- 10.13 %	- 19.29 %
Quantité d'éléments dont la durée a	augmenté	23	16	14	6
	diminué	17	17	17	15

Tableau 17 : IT12AB - Variations en durée, suivant plusieurs seuils, des éléments dont la valeur ne varie pas

La moitié (48%, valeur complémentaire des 52% de la dernière colonne) des éléments varient en durée de moins de 20% lors de leur mise en contexte, par rapport à leur réalisation isolée.

Si on ne considère aucun seuil (0%), la majorité des éléments a une durée qui augmente ; pour les seuils de 5% et 10%, il y a autant d'éléments dont la durée augmente que d'éléments dont la durée diminue ; si on considère le seuil de 20%, la grande majorité des éléments a une durée qui augmente. Quelque soit le seuil considéré, la moyenne des durées de l'ensemble des éléments diminue.

Les deux tiers (14 sur 23) des éléments dont la durée augmente ont des variations de 10% ou moins et un quart (6 sur 23) a des variations de 20% ou moins, alors que quasiment tous les éléments dont la durée diminue ont des variations de 20% ou moins (15 sur 17).

DOY		Seuil			
		0 %	5%	10%	20%
Quantité d'éléments dont la valeur est identique isolé et en contexte mais dont la durée varie		67	56	54	37
% d'éléments dont la durée varie (par rapport à la quantité totale des éléments de l'énoncé)		100 %	84 %	81 %	55 %
Moyenne de la variation des durées		- 11.09 %	- 13.22 %	- 13.43 %	- 18.78 %
Quantité d'éléments dont la durée a	augmentée	21	13	13	6
	diminuée	46	43	41	31

Tableau 18 : DOY - Variations en durée, suivant plusieurs seuils, des éléments dont la valeur ne varie pas

La moitié (45%, valeur complémentaire des 55% de la dernière colonne) des éléments varient en durée de moins de 20% lors de leur mise en contexte, par rapport à leur réalisation isolée.

Quelque soit le seuil considéré, la grande majorité des éléments a une durée qui diminue, de même que la moyenne des durées de l'ensemble des éléments.

La moitié (13 sur 21) des éléments dont la durée augmente a des variations de 10% ou moins et le quart (6 sur 21) a des variations de 20% ou moins, alors que quasiment tous les éléments (41 sur 46) dont la durée diminue ont des variations de 10% ou moins, et les deux tiers (31 sur 46) ont des variations de 20% ou moins.

DT02		Seuil			
		0 %	5%	10%	20%
Quantité d'éléments dont la valeur est identique isolé et en contexte mais dont la durée varie		70	61	54	41
% d'éléments dont la durée varie (par rapport à la quantité totale des éléments de l'énoncé)		100 %	87 %	77 %	59 %
Moyenne de la variation des durées		- 9,4 %	- 10.68 %	- 11.98 %	- 13.56 %
Quantité d'éléments dont la durée a	augmenté	26	19	16	13
	diminué	44	42	38	28

Tableau 19 : DT02 - Variations en durée, suivant plusieurs seuils, des éléments dont la valeur ne varie pas

Plus du tiers (41%, valeur complémentaire des 59% de la dernière colonne) des éléments varient en durée de moins de 20% lors de leur mise en contexte, par rapport à leur réalisation isolée.

Quelque soit le seuil considéré, la grande majorité des éléments a une durée qui diminue, de même que la moyenne des durées de l'ensemble des éléments.

La moitié (13 sur 26) des éléments dont la durée augmente a des variations de 20% ou moins et les deux tiers (16 sur 26) ont des variations de 10% ou moins, alors que les deux tiers (28 sur 44) des éléments dont la durée diminue ont des variations de 20% ou moins.

RT20A		Seuil			
		0 %	5%	10%	20%
Quantité d'éléments dont la valeur est identique isolé et en contexte mais dont la durée varie		67	55	53	43
% d'éléments dont la durée varie (par rapport à la quantité totale des éléments de l'énoncé)		100 %	82 %	79 %	64 %
Moyenne de la variation des durées		- 16.6 %	- 20.23 %	- 21.03 %	- 24.99 %
Quantité d'éléments dont la durée a	augmenté	25	13	12	8
	diminué	42	42	41	35

Tableau 20 : RT20A - Variations en durée, suivant plusieurs seuils, des éléments dont la valeur ne varie pas

Un tiers (36%, valeur complémentaire des 64% de la dernière colonne) des éléments varient en durée de moins de 20% lors de leur mise en contexte, par rapport à leur réalisation isolée.

Quelque soit le seuil considéré, la grande majorité des éléments a une durée qui diminue, de même que la moyenne des durées de l'ensemble des éléments.

La moitié des éléments dont la durée augmente (12 sur 25) ont des variations de 10% ou moins, alors que la majorité des éléments dont la durée diminue (35 sur 42) ont des variations de 10% ou moins.

RT20B		Seuil			
		0 %	5%	10%	20%
Quantité d'éléments dont la valeur est identique isolé et en contexte mais dont la durée varie		56	48	48	35
% d'éléments dont la durée varie (par rapport à la quantité totale des éléments de l'énoncé)		100 %	86%	86%	62%
Moyenne de la variation des durées		- 17.49 %	- 20.33 %	- 20.33 %	- 24.9 %
Quantité d'éléments dont la durée a	augmenté	21	14	14	11
	diminué	35	34	34	24

Tableau 21 : RT20B - Variations en durée, suivant plusieurs seuils, des éléments dont la valeur ne varie pas

Un tiers (38%, valeur complémentaire des 62% de la dernière colonne) des éléments varient en durée de moins de 20% lors de leur mise en contexte, par rapport à leur réalisation isolée.

Quelque soit le seuil considéré, la grande majorité des éléments a une durée qui diminue, de même que la moyenne des durées de l'ensemble des éléments.

Les deux tiers des éléments dont la durée augmente ont des variations de 10% ou moins et la moitié ont des variations de 20% ou moins, alors que quasiment tous les éléments dont la durée diminue ont des variations de 10% ou moins, et les deux tiers de 20% ou moins.

a. Synthèse

La durée des éléments (ceux qui sont identiques en contexte et dans leur forme isolée) des énoncés est moindre quand ils sont mis en contexte que quand ils sont dans les segments d'énoncés, sauf pour AS16 où cette durée est plus grande en contexte que dans les segments.

Dans les énoncés AS16, les éléments dont la durée augmente en contexte ont des variations de durées supérieures à 20%. Dans les énoncés DT02, les éléments dont la durée augmente en contexte ont des variations de durées variables. Dans les énoncés AS01, DOY et RT20A, les éléments dont la durée augmente en contexte ont des variations de durées inférieures à 10%. Dans les énoncés IT12AB et RT20B, les éléments dont la durée augmente en contexte ont des variations de durées inférieures à 20%.

Dans les énoncés AS01, AS16, DOY et DT02, les éléments dont la durée diminue en contexte ont des variations de durées supérieures à 10%. Dans les énoncés IT12AB, RT20A et RT20B, les éléments dont la durée diminue en contexte ont des variations de durées supérieures à 20%.

Pour rappel, voici les différentes structures d'énoncés de notre corpus :

- AS01 et IT12AB ont en commun AB... ;
- AS16 et RT20A ont en commun D...E... ;
- AS16 et RT20B ont en commun D...E...G ;
- DOY et DT02 ont en commun A...I ;
- RT20A et RT20B ont en commun DAEK...

Toujours pour rappel, voici les types d'énoncés de notre corpus :

- AS01, IT12AB, DOY et DT02 sont des énoncés d'information,
- AS16, RT20A et RT20B sont des énoncés d'incident.

Pour les énoncés AS16, nous disposons de peu de comparaisons possibles, nous préférons donc les laisser de côté dans cette partie de nos analyses.

Pour les autres énoncés, nous pouvons conclure que la durée des éléments (ceux qui sont identiques en contexte et dans leur forme isolée) des énoncés est moindre quand ils sont mis en contexte que quand ils sont dans les segments d'énoncés.

Par extrapolation, pour les énoncés d'information, les éléments dont la durée diminue en contexte ont des variations de durées supérieures à 10%. Pour les énoncés d'incident, les éléments dont la durée diminue en contexte ont des variations de durées supérieures à 20%.

2. Par composante

Nous avons souhaité mettre en parallèle les variations de durées d'une même composante au travers des différentes instances d'énoncés de notre corpus. Nous avons considéré toutes les variations de durée, sans aucun seuil minimal. Nous présentons les différentes variations dans des tableaux synthétiques reprenant les mêmes informations que celles des tableaux ci-dessus (mais cette fois regroupées par énoncé et plus par seuil).

Direction du regard		AS01	AS16	IT12AB	DOY	DT02	RT20A	RT20B
Quantité d'éléments dont la valeur est identique isolé et en contexte mais dont la durée varie		52	25	6	9	7	5	7
% (par rapport au total des éléments de chaque énoncé, dont la durée varie mais pas la valeur)		21,58	30,12	15,00	13,43	10,00	7,46	12,50
Moyenne de variation des durées		- 0,5%	- 2,2%	+7,2%	+24,2%	+16,5%	+8,9%	-12,4%
Quantité d'éléments dont la durée a	augmenté	25	11	5	7	4	2	1
	diminué	27	14	1	2	3	3	6

Tableau 22 : Direction du regard – Comparaisons inter-énoncés des variations relevées

Dans les instances d'énoncés AS01 et AS16, il y a quasiment autant d'éléments qui voient leur durée augmenter en contexte que d'éléments dont la durée diminue, et en moyenne la durée des éléments en contexte est la même que la durée des mêmes éléments isolés.

Dans les instances d'énoncés IT12AB et RT20A, la durée moyenne des éléments en contexte a augmenté d'un peu moins de 10%. Dans le cas des instances d'énoncés RT20A, il y a quasi autant d'éléments dont la durée augmente que d'éléments dont la durée diminue, ce qui implique que les éléments dont la durée augmente connaissent une variation plus importante que la variation de la durée des éléments dont la durée diminue.

Dans les instances d'énoncés DOY et DT02, en moyenne la durée des éléments augmente beaucoup (respectivement de 25 et 15%). Dans le cas des instances d'énoncés DT02, il y a quasi autant d'éléments dont la durée augmente que d'éléments dont la durée diminue, ce qui implique que les éléments dont la durée augmente connaissent une variation beaucoup plus importante que la variation de la durée des éléments dont la durée diminue.

Dans les instances d'énoncés RT20B, en moyenne la durée des éléments diminue d'un peu moins de 15% lors de leur mise en contexte.

Configuration de mD		AS01	AS16	IT12AB	DOY	DT02	RT20A	RT20B
Quantité d'éléments dont la valeur est identique isolé et en contexte mais dont la durée varie		111	42	26	41	42	48	35
% (par rapport au total des éléments de chaque énoncé, dont la durée varie mais pas la valeur)		46,06	50,60	65,00	61,19	60,00	71,64	62,50
Moyenne de variation des durées		+1,2%	+12,1%	-16,1%	-16%	-15,3%	-16,6%	-18,4%
Quantité d'éléments dont la durée a	augmenté	54	16	12	11	11	18	15
	diminué	57	26	14	30	31	30	20

Tableau 23 : Main dominante – Comparaisons inter-énoncés des variations relevées

Dans les instances d'énoncés AS01, il y a quasiment autant d'éléments qui voient leur durée augmenter en contexte que d'éléments dont la durée diminue, et en moyenne la durée des éléments en contexte est la même que la durée des mêmes éléments isolés.

Dans les instances d'énoncés AS16, en moyenne la durée des éléments augmente de plus de 10% lors de leur mise en contexte. Il y a beaucoup plus d'éléments (deux tiers) dont la durée diminue que d'éléments dont la durée augmente, ce qui implique que les éléments dont la durée augmente connaissent une variation beaucoup plus importante que la variation de la durée des éléments dont la durée diminue.

Dans les instances d'énoncés IT12AB, DOY, DT02, RT20A et RT20B, en moyenne la durée des éléments diminue de beaucoup (entre 15 et 20% environ) lors de leur mise en contexte. Dans le cas des instances d'énoncés IT12AB, il y a quasiment autant d'éléments qui voient leur durée augmenter en contexte que d'éléments dont la durée diminue, ce qui implique que les éléments dont la durée diminue connaissent une variation beaucoup plus importante que la variation de la durée des éléments dont la durée augmente. Dans le cas des instances d'énoncés DOY et DT02, il y a trois fois moins d'éléments qui voient leur durée augmenter que d'éléments dont la durée diminue en contexte, et pour les instances d'énoncés RT20A et RT20B, il y en a deux fois moins.

Nous pouvons remarquer que suite à ces constatations, les énoncés DOY et DT02 peuvent être regroupés, de même que les énoncés RT20A et RT20B.

Configuration de md		AS01	AS16	IT12AB	DOY	DT02	RT20A	RT20B
Quantité d'éléments dont la valeur est identique isolé et en contexte mais dont la durée varie		78	16	8	17	21	14	14
% (par rapport au total des éléments de chaque énoncé, dont la durée varie mais pas la valeur)		32,37	19,28	20,00	25,37	30,00	20,90	25,00
Moyenne de variation des durées		+0,5%	+7,9%	+9,3%	-17,9%	-6,2%	-25,8%	-17,6%
Quantité d'éléments dont la durée a	augmenté	35	11	6	3	11	5	5
	diminué	43	5	2	14	10	9	9

Tableau 24 : Main dominée – Comparaisons inter-énoncés des variations relevées

Dans les instances d'énoncés AS01, en moyenne la durée des éléments en contexte est la même que la durée des mêmes éléments isolés, mais il y a plus d'éléments dont la durée diminue que d'éléments dont la durée augmente en contexte : les éléments dont la durée augmente connaissent donc une variation beaucoup plus importante que la variation de la durée des éléments dont la durée diminue.

Dans les instances d'énoncés AS16 et IT12AB, en moyenne la durée des éléments augmente de presque 10% lors de leur mise en contexte, et il y a également plus d'éléments (respectivement deux et trois fois plus) qui voient leur durée augmenter que d'éléments dont la durée diminue en contexte.

Dans les instances d'énoncé DT02, en moyenne la durée des éléments diminue de plus de 5% lors de leur mise en contexte, alors qu'il y a quasi autant d'éléments dont la durée augmente que d'éléments dont la durée diminue : les éléments dont la durée diminue connaissent donc une variation beaucoup plus importante que la variation de la durée des éléments dont la durée augmente.

Dans les instances d'énoncé DOY, RT20A et RT20B, en moyenne la durée des éléments diminue de presque 20% (pour DOY et RT20B) et 25% (pour RT20A) lors de leur mise en contexte. Il y a beaucoup plus (deux fois plus pour RT20A et RT20B, 5 fois plus pour DOY) d'éléments qui voient leur durée diminuer que d'éléments dont la durée augmente en contexte.

Nous pouvons remarquer que suite à ces constatations, les énoncés RT20A et RT20B peuvent être regroupés (même si en moyenne leur variation de durée des éléments les composant diffère quelque peu).

a. Synthèse

D'une manière générale, nous pouvons conclure qu'il y a un raccourcissement des durées de configurations, que ce soit mD ou md, et un allongement des durées de regard.

Concernant la direction du regard, on peut remarquer de grandes disparités dans les divers énoncés. Il n'y a pas de regroupement possible qui nous permette de dire que dans tel contexte d'énoncé la direction du regard subira tel effet quant à la durée de ses éléments.

Concernant la configuration de la main dominante (mD), on peut remarquer également de grandes disparités, mais des regroupements sont possibles. Ainsi, on constate que le nombre d'éléments dont la durée diminue en contexte est toujours supérieure au nombre d'éléments dont la durée augmente. Comme cela a été signalé, il est possible de regrouper les énoncés DOY et DT02 d'une part et RT20A et RT20B d'autre part : au-delà du regroupement basé sur ces données numériques, il s'avère que ces énoncés partagent les mêmes structures de construction et la même sémantique d'information (le même type, cf. rappel dans la partie précédente et détails dans le chapitre méthodologie). Il se pourrait que d'autres énoncés ayant une même structure et/ou une même sémantique soient affectés de la même manière lors de leur construction à partir d'éléments isolés.

Concernant la configuration de la main dominée (md), on peut remarquer là encore de grandes disparités entre les énoncés. Là également, comme cela a été signalé, il est possible de regrouper les énoncés RT20A et RT20B qui partagent une structure et une sémantique communes.

Si nous essayons de corrélérer les différentes composantes entre eux au travers des énoncés, nous remarquons qu'hormis AS01 et RT20B, la variation de durée d'une composante n'est pas en lien direct avec la variation d'un autre, c'est-à-dire que ce n'est pas parce que la direction du regard varie en augmentation que mD et/ou md également, et réciproquement. Pour AS01, les composantes ne varient globalement pas, alors que pour RT20B les trois composantes subissent des diminutions de durées. Cela ne nous permet pas de conclure sur une éventuelle influence d'une composante sur les autres.

Nous remarquons que le contexte sémantique semble avoir une influence sur la durée des éléments quand la coarticulation n'en a pas (puisque les éléments considérés dans cette partie sont identiques en valeur entre leur forme isolée et leur forme en contexte). Les durées varient en fonction de certaines structures d'énoncés, cependant nous remarquons de telles variations dans

certain énoncés et pas dans d'autres pourtant de mêmes structures : la généralisation sur ce point, à ce stade de l'analyse de nos résultats, semble donc délicate.

3. Par segments d'énoncés

Suite aux constats faits dans la partie précédente, nous avons décidé de conserver le seuil de 10% minimum de modifications pour évaluer au sein des segments d'énoncés mis en contexte comment s'effectuent les variations de durées : c'est-à-dire, est-ce que le début de l'énoncé est raccourci et la fin augmentée, ou l'inverse, ou autre ?

Pour répondre à cette question, nous avons comparé les durées des éléments dont la valeur ne varie pas sous forme isolée et en contexte. Nous avons traité ces éléments, segment d'énoncé par segment d'énoncé, afin de déterminer si certaines variations de durées étaient dues au contexte ou non.

Note : dans cette partie nous parlerons simplement d'augmentation ou de diminution pour ne pas surcharger notre présentation, mais il faut bien sûr comprendre « augmentation de la durée des éléments en contexte par rapport à leur durée sous forme isolée » et « diminution de la durée des éléments en contexte par rapport à leur durée sous forme isolée ».

Dans le cas des segments d'énoncés E1 à E11, qui apparaissent chacun à deux reprises dans les instances d'énoncés AS01, en première et avant dernière position, on peut émettre l'hypothèse que la direction du regard varie à cause du contexte sémantique de l'énoncé final. La configuration de mD semble subir tout d'abord une augmentation de durée puis une diminution, quand la configuration de md semble subir une diminution. Dans le cas du segment d'énoncé E6 en particulier, qu'on retrouve dans beaucoup plus de contextes différents que les autres, il semble qu'il y ait une diminution de la durée des configurations que ce soit pour mD ou md.

Dans le cas du segment d'énoncé E12, la direction du regard semble subir une augmentation, quand les configurations de mD et md suivent deux schémas : diminution/augmentation/diminution ou augmentation/diminution. Le premier schéma s'applique lorsque E12 est mis en contexte au milieu ou à la fin de l'énoncé complet, quand le premier s'applique lorsque E12 est mis en contexte au début de l'énoncé complet.

Dans le cas de E13, les configurations de mD et md subissent aussi bien des diminutions que des augmentations donc il n'est pas possible de conclure, cependant que la direction du regard subit une diminution.

Dans le cas de E17, on constate une diminution et ce dans deux contextes différents, quand il est inséré au début de l'énoncé complet. Au début du segment d'énoncé E17, il y a suppression d'un

signe lors de sa mise en contexte : on peut imaginer que la réalisation de ce signe suivrait un schéma augmentation/diminution si ce signe avait été conservé (ce qui serait dans la même logique que l'énoncé E12).

Pour tous les autres énoncés, soient les chiffres et les gares, il y a clairement une diminution en ce qui concerne les configurations de mD et md.

a. Synthèse

Des constatations de cette partie, nous pouvons conclure que pour les segments d'énoncés E1 à E11 il y a une influence du contexte sémantique uniquement, et seulement pour la direction du regard. Pour les segments représentant les chiffres et les gares, il y a une influence du contexte phonétique qui transparait au niveau de la durée des éléments : nous formulons l'hypothèse que ces segments ne subissent pas de variation de valeur puisqu'il faut selon nous qu'ils soient réalisés en contexte de manière très précise (et qu'ils le sont déjà dans les segments d'énoncé), de ce fait la durée de leur réalisation est raccourcie puisqu'ils sont précis (et donc perçus de manière claire).

III. Synthèse générale des résultats d'analyse

1. Par catégories de segments d'énoncés

Cette partie regroupe toutes les synthèses données dans les parties précédentes afin d'en déduire les aspects à mettre en œuvre dans le modèle de coarticulation.

a. Annonce de voies (E1 à E11)

Ces segments d'énoncés varient très largement par un ajout de direction du regard vers un emplacement de l'espace au début de l'énoncé. Il y a également un relâchement de certaines configurations, qu'on peut retrouver dans d'autres segments d'énoncés. Ceci entraîne une modification des durées des éléments pour la direction du regard : au lieu de regarder l'interlocuteur, donc face à elle, la locutrice regarde un emplacement, ce qui fait que l'élément « regard de face » est raccourci en temps, pour laisser de la place à l'ajout de direction du regard initiale. Les configurations subissent également des variations dans leur durée : mD a des valeurs de configuration dont la durée augmente puis diminue, quand md diminue.

Ces segments d'énoncés E1 à E11 varient donc par insertion d'une direction du regard, qui selon nous est due au contexte sémantique, pas au contexte phonétique.

b. Informations (E12 à E21)

Les segments d'énoncé E12, E13 et E17 subissent des variations dues à la contextualisation sémantique et phonétique.

Le reste des segments d'énoncés ne subissent pas de variations qui permettent de conclure sur quelques corrélations entre eux. On peut seulement souligner une modification de signe, ainsi qu'un relâchement des configurations.

c. Motifs de perturbations (E23 à E31)

Dans le cas des segments d'énoncés E23 à E31, les variations sont très différentes d'un segment d'énoncé à un autre, et les mises en contextes ne sont pas assez nombreuses pour que nous puissions conclure. Ces segments d'énoncés subissent aussi parfois une suppression de signe, et plus généralement un relâchement ou une tension de certaines configurations. La direction du regard suit plusieurs schémas différents quant à sa variation en durée, et le point commun de ces schémas est une diminution de la durée des directions de regard à la fin de l'énoncé. Les durées des configurations des mains, dominante (mD) ou dominée (md), étant quant à elles en diminution tout au long de l'énoncé.

d. Chiffres et heures

Nous estimons que la coarticulation en valeur (celle qui est responsable de la variation des valeurs des composants) n'a pas de réel effet sur la réalisation des chiffres. Ceci est d'autant plus compréhensible que les noms des gares et les chiffres donnés (pour les numéros de train ou pour les horaires) doivent être précis afin de ne pas laisser la place à une ambiguïté : l'effet de la coarticulation en valeur est ainsi annihilé par la locutrice qui « s'applique » lors de la réalisation de ces informations. Concernant la coarticulation en durée (celle qui est responsable de la variation des durées des composants), les configurations des mains, mD et md, connaissent une diminution lors de leur mise en contexte.

Nous soulignons que les nombres « 400 » et « 500 », sont réalisés avec une répétition dans leur forme isolée, mais le sont en contexte sans la répétition, et la réalisation du signe pour les milliers change parfois dans le signe en lui-même (au lieu d'un seul signe il y a réalisation du signe du millier puis de « mille »).

e. Gares

Comme pour les signes de chiffres, nous constatons que les signes de gares ne subissent aucune variation due au contexte phonétique, mais varient suivant le contexte sémantique. Ceci est dû selon nous à l'importance qu'ont ces informations dans le domaine de la SNCF.

Au niveau des durées, là encore il y a une diminution de la durée des configurations de mD et md lors de leur mise en contexte. Il y a une influence du contexte phonétique qui transparaît au niveau de la durée des éléments : nous formulons l'hypothèse que ces segments ne subissent pas de variation de valeur puisqu'il faut qu'ils soient réalisés en contexte de manière très précise (et qu'ils le sont déjà dans les segments d'énoncé), de ce fait la durée de leur réalisation peut être raccourcie puisqu'ils sont précis (avec l'objectif d'être perçus par l'interlocuteur de manière claire).

2. Par énoncés

Pour les énoncés AS16, nous disposons de peu de comparaisons possibles, nous préférons donc les laisser de côté dans cette partie de nos conclusions.

Pour les autres énoncés, nous pouvons conclure que la durée des éléments (ceux qui sont identiques en contexte et dans leur forme isolée) des énoncés est moindre quand ils sont mis en contexte que quand ils sont dans les segments d'énoncés.

Par extrapolation, pour les énoncés d'information, les éléments dont la durée diminue en contexte ont des variations de durées supérieures à 10%. Pour les énoncés d'incident, les éléments dont la durée diminue en contexte ont des variations de durées supérieures à 20%.

3. Par catégorie de configurations / de direction du regard

Au niveau des durées, d'une manière générale, nous pouvons conclure qu'il y a un raccourcissement des durées de configurations, que ce soit mD ou md, et un allongement des durées de regard.

a. Le regard

Concernant les durées de direction du regard, on peut remarquer de grandes disparités dans les divers énoncés. Il y a cependant un phénomène remarquable : s'il y a ajout d'une direction de regard, il y a alors raccourcissement de la durée de la direction de regard initiale. Ainsi, l'insertion provoque une modification en durée de la direction de regard initiale (réalisée dans la forme isolée).

Les modifications de direction du regard interviennent fréquemment pour s'orienter vers un emplacement de l'espace de signation plutôt que de se focaliser sur les mains comme cela se passe dans les segments d'énoncés. Il y a également une élévation notable de l'espace de signation selon le plan vertical, puisque les emplacements en contexte sont « 7 », « 8 » ou « 9 », qui sont plus hauts dans l'espace que les emplacements « 1 » et « 2 » qu'ils remplacent. Enfin, on constate que le regard a tendance à suivre les mains là où sous la forme isolée il était dirigé vers l'interlocuteur.

Les suppressions de direction de regard interviennent peu et, dans le cas qui nous intéresse, lorsqu'une direction de regard vers un emplacement de l'espace en contexte a eu lieu peu de temps auparavant (estimé à un intervalle d'environ une seconde ; valeur demandant à être affinée sur d'autres analyses), ce qui signifierait qu'il faut considérer un laps de temps minimal entre deux directions de regard vers le même espace, même si c'est pour le même référent.

Nous avons également analysé la fréquence des clignements des yeux. Il nous a semblé de prime abord que les clignements des yeux en contexte correspondait à certains segments d'énoncés, c'est-à-dire qu'à chaque début et chaque fin du segment d'énoncé il y avait un clignement des yeux lors de sa mise en contexte. Nous avons observé que les choses ne sont pas aussi simples. Dans tous les énoncés, les clignements des yeux correspondent à des morceaux d'énoncés (pas forcément strictement aux segments d'énoncés) sémantiquement viables (c'est-à-dire qu'ils pourraient être exprimés seuls, sans ambiguïté sur le sens, ce qui n'est pas le cas des segments d'énoncés mis en contexte, puisqu'ils subissent des suppressions de morceaux entiers dus à leur mise en contexte).

Les variations de direction du regard se retrouvent principalement dans AS01, un peu dans IT12AB, et quasiment pas dans les autres énoncés. Les modifications interviennent en regard de l'information donnée par l'énoncé : il s'agit de prévenir les personnes de s'éloigner de la bordure du quai, le regard va donc insister davantage sur le quai, la foule, et se focaliser sur un espace plus proche pour attirer le regard de l'interlocuteur. Ceci est dû, selon nous et notre locutrice/experte LSF, à l'importance accordée à l'emplacement de ce qui est signé plutôt qu'à ce qui est signifié : il est jugé plus important que les personnes comprennent qu'il faut qu'elles se déplacent de la bordure du quai (en LSF, l'expression de ces informations est inversée, d'où le fait qu'il y ait un regard d'abord vers le quai puis vers les mains) plutôt que l'endroit où elles se trouvent. Les variations de direction de regard sont dues à une élévation générale de l'espace de signation, qui de fait entraîne une élévation des directions de regard. Ces variations sont selon nous dues au contexte sémantique plus qu'au contexte phonétique (donc pas dues à la coarticulation).

Les suppressions de direction de regard sont dues à la suppression des signes qui impliquent leur réalisation, et soit à la proximité d'une direction de regard équivalente dans un intervalle de temps court (un intervalle de deux configurations de mains ou une durée de moins d'une seconde), soit au sens de l'énoncé (dans ce dernier cas, le regard se focalise sur l'information importante de l'énoncé).

Les variations relevées sont dues en grande majorité au contexte sémantique plutôt qu'au contexte phonétique, il n'y a donc pas, hormis pour la suppression de regard, de variations de la direction du regard dues à la coarticulation.

b. Les configurations des mains

Selon nous, la coarticulation dans notre domaine n'affecte que des signes du lexique. Cela rejoint l'idée émise précédemment que pour les chiffres et la dactylologie il est important d'être précis sur la configuration en elle-même, donc que la forme en contexte subit peu d'effet de coarticulation (même si la présence de coarticulation peut aider à une meilleure perception globale, elle n'aide pas à la perception d'un élément en particulier).

Nous remarquons également que les configurations initialement réalisées de manière tendues dans les segments d'énoncés deviennent plus relâchées si dans le contexte sémantique ce ne sont pas les informations importantes de l'énoncé final. Les configurations répétées raccourcissent lorsqu'elles sont suivies de configurations phonétiquement proches (quand par exemple une configuration répétée se terminant par un index tendu précède une configuration avec l'index et le majeur tendus et croisés, alors la dernière configuration index tenu disparaît).

Concernant la configuration de la main dominante (mD), on constate que le nombre d'éléments dont la durée diminue en contexte est toujours supérieure au nombre d'éléments dont la durée augmente. Concernant la configuration de la main dominée (md), on peut remarquer de grandes disparités entre les énoncés.

Nous remarquons que le contexte sémantique semble avoir une influence sur la durée des éléments quand la coarticulation n'en a pas (puisque les éléments considérés dans cette partie sont identiques en valeur entre leur forme isolée et leur forme en contexte). Les durées varient en fonction de certaines structures d'énoncés.

Si nous essayons de corrélérer les différentes composantes entre eux au travers des énoncés, nous remarquons qu'hormis AS01 et RT20B, la variation de durée d'une composante n'est pas en lien direct avec la variation d'un autre, c'est-à-dire que ce n'est pas parce que la direction du regard varie

en augmentation que mD et/ou md également, et réciproquement. Cela ne nous permet pas de conclure sur une éventuelle influence d'une composante sur les autres au niveau de la durée des éléments.

Il y a une assimilation entre les configurations de signes comportant des répétitions en fonction de la configuration suivante, suivant la proximité phonétique que ces configurations entretiennent. Cette assimilation montre selon nous un effet de coarticulation par anticipation.

Nous constatons que dans un premier temps il y a une influence du contexte sémantique, qui impose des configurations tendues (identiques à la forme dans les segments d'énoncés) pour certaines catégories de signes, tels les chiffres et les gares, ainsi que lors de la réalisation des informations importantes, et des configurations plus relâchées pour les informations moins importantes, ne nécessitant pas forcément d'être réalisées en détail mais plutôt dans une certaine forme (au sens forme de la main) générale. Ensuite, il y a une influence du contexte phonétique, qui s'applique par persévération : lors de la mise en contexte, si la configuration précédente est tendue ou relâchée, alors la configuration mise en contexte sera réalisée de manière plus tendue ou plus relâchée que dans sa forme isolée. Nous remarquons également que lors de la répétition de segments d'énoncés en contexte (par exemple les 2 occurrences des segments E1 à E11 dans AS01), la seconde occurrence subit les mêmes variations que la première : il y a donc une influence là encore par persévération.

Nous rajoutons donc une étape par rapport à ce que nous avons élaboré précédemment : il y a tout d'abord une mise en contexte sémantique, qui implique certaines variations, puis une influence de mD sur md, ensuite une mise en contexte phonétique, et une influence de mD sur md.

Raccourcissement de signes

Les raccourcissements de signes constatés, par suppression de la dernière configuration, sont selon nous dus à la configuration suivante : il s'agit d'une coarticulation par assimilation, par anticipation de la configuration suivante.

Les coarticulations constatées de fait

Au-delà des analyses présentées tout au long de cette partie, nous avons également relevé certaines coarticulations sans avoir eu recours à l'alignement des annotations, tout simplement parce que nous ne pouvions aligner ces configurations en contexte avec une forme isolée (puisque nous n'avions pas connaissance de cette dernière).

Dans AS16, il y a une coarticulation des configurations  et  lors de la réunion des signes « attention » et « situation » au tout début de l'énoncé (sauf avant le signe « panne »), une configuration plus relâchée dans AS16-4 pour le signe « caArrive » qui est précédé d'une configuration relâchée ( isolé et en contexte).

Dans AS16-7, une configuration de md () se rajoute lors du signe « j'ai » effectué normalement seulement par mD avec la configuration  . Ceci dénote une influence de la main dominante (mD) sur la main dominée (md).

Dans AS16-9, il y a une coarticulation entre les signes « jeNeSaisPas » et « qui » (configurations respectives  et ). Dans DT02, lors de la réalisation de « 4000 », md prend déjà la configuration qui va suivre (au lieu d'être en configuration  , elle devient ). Ces coarticulations sont des phénomènes d'assimilation, où les configurations qui se succèdent prennent chacune une ou plusieurs caractéristiques l'une de l'autre.

4. Proposition d'un modèle de coarticulation dans le contexte de notre étude

Au vu de toutes ces constatations, interprétations des résultats et hypothèses, nous souhaitons évaluer un premier modèle de coarticulation des signes de LSF pour l'animation d'un signeur virtuel dans le domaine de l'information en gare ferroviaire. Nous précisons ce contexte pour rappeler que nous n'avons pas l'ambition de proposer un modèle s'appliquant à toute la langue, dans sa complexité grammaticale et sa variété lexicale.

Nous avons constaté des variations au niveau de la direction du regard, en valeur, dues au contexte sémantique, et en durée dues au contexte phonétique. Nous avons constaté des variations au niveau des configurations des mains, tant en valeurs qu'en durées dues soit au contexte sémantique soit au contexte phonétique.

La mise en œuvre de l'animation du signeur virtuel doit prendre en compte ces deux aspects afin que la mise en contexte des segments d'énoncés paraisse la plus « naturelle » possible.

Il faut tout d'abord analyser la sémantique de l'énoncé final pour déterminer quelle est l'information importante. Cette information sera réalisée de manière claire et précise, les configurations initialement relâchées seront tendues, le regard venant appuyer ces configurations.

Il faut ensuite considérer la réalisation des informations importantes suivant le domaine : les noms de gares, les chiffres (de train, de voie, etc.) et les horaires, qui doivent être réalisés de manière claire et précise, donc a priori identiquement à la forme isolée. Il faut cependant tenir compte des éventuelles modifications de signes pour certaines gares.

Puis, il y a une première considération du contexte phonétique, une fois que certaines configurations ont été décidées : celles-ci auront une influence sur celles restant à réaliser, en fonction de leurs positions (avant ou après) relatives à ces dernières.

Enfin, il faut tenir compte de l'influence de la main dominante (mD) sur la main dominée (md).

Voici les différents aspects pour les mises en contexte sémantique et phonétique que nous considérons dans notre modèle pour le pilotage de l'animation d'un signeur virtuel :

- Contextualisation sémantique :
 - o Elévation de la direction du regard
 - o Ajout de directions du regard pour insister sur certains emplacements (en anticipation de la réalisation de signes par les mains)
 - o Suppressions de signes
- Coarticulation (contextualisation phonétique) :
 - o Configurations des mains
 - Assimilation
 - Influence par persévération
 - Influence de mD vers md
 - o Direction du regard
 - Délai minimal entre 2 directions du regard identiques

Nous devons maintenant évaluer les principes de notre modèle. Pour cela, nous présentons dans le chapitre suivant la méthodologie que nous avons adoptée, et les résultats de notre évaluation. Nous avons demandé à plusieurs personnes sourdes locutrices de LSF d'une part de donner leur avis selon plusieurs critères (compréhension, naturel, adéquation avec le domaine de l'information en gare) sur le résultat de l'implémentation de notre modèle, d'autre part de comparer différents types d'expression (concaténation de segments d'énoncés sans modèle de coarticulation, concaténation de segments d'énoncés avec modèle de coarticulation, énoncés complets – i.e. non issus de la concaténation de segments d'énoncés).

Chapitre V. EVALUATION

Une fois notre modèle de coarticulation mis en place, nous l'avons évalué auprès de différents locuteurs de LSF, afin de déterminer si les aspects que nous avons pris en compte sont bien pertinents dans le cadre de notre recherche, qui est de fournir des informations en LSF via un signeur virtuel aux usagers sourds dans les gares ferroviaires.

Dans ce chapitre nous dressons un état de l'art des différentes méthodologies d'évaluation des modèles de coarticulation dans les domaines des têtes parlantes, des agents conversationnels animés (ACA) et des LS (via des signeurs virtuels). Ensuite, nous explicitons, globalement puis en détail, la méthodologie d'évaluation que nous avons choisie. Enfin nous présentons les résultats de notre évaluation, au niveau du modèle de coarticulation proposé, mais également sur le signeur virtuel et le dispositif d'information en gare.

I. Méthodologies d'évaluation

Nous n'avons pas trouvé de référence sur une évaluation de la coarticulation implémentée dans le système de synthèse vocale utilisée par la SNCF dans ses gares, de même que nous n'avons pas trouvé de référence sur une évaluation générale de ce système.

Au vu des spécificités des LS par rapport aux LV, notamment en ce qui concerne la modalité visuelle, nous nous sommes intéressés aux évaluations des têtes parlantes, des personnages animés nommés ACA et des signeurs virtuels. Nous faisons également une présentation de la seule étude à notre connaissance sur l'évaluation de la coarticulation des gestes dans le domaine de la synthèse. Nous présentons dans un premier temps des considérations générales sur l'évaluation, puis nous détaillons quelques exemples de méthodologies utilisées dans ces domaines.

1. Têtes parlantes

Les têtes parlantes sont définies largement comme des visages dont les mouvements des lèvres sont synthétisés en lien avec les propos prononcés par les visages. Dans le domaine de l'animation de ces têtes parlantes, il existe deux types de tests d'évaluation, selon Govokhina [GOV08] : les tests objectifs (quantitatifs) et subjectifs (qualitatifs). Les premiers portent sur la différence entre la réalisation en synthèse et la réalisation réelle, par des mesures ne nécessitant pas de participants. Les tests subjectifs portent sur trois aspects : le réalisme (mesure de la distance entre la synthèse et

le réel), l'intelligibilité (compréhension de la synthèse), l'acceptabilité (dans certaines situation la synthèse peut être jugée acceptable quand dans d'autres elle ne sera pas jugée suffisante). Selon le même auteur, la question de l'évaluation des résultats de la synthèse audiovisuelle est très importante. Ce sont les tests d'évaluation objective et subjective qui permettent de dire comment les systèmes de synthèse répondent au cahier des charges prévu. Dans beaucoup de cas, l'évaluation subjective des systèmes donne des résultats très pauvres par rapport au bon réalisme obtenu. Cette contradiction peut être expliquée par la complexité accumulée à chaque module de la chaîne de synthèse audiovisuelle. Les mouvements corrects peuvent être jugés inadéquats si on a un mauvais modèle d'apparence. De même, un mauvais modèle de contrôle engendre des mouvements qui peuvent être jugés comme non acceptables. Govokhina ([GOV08]) conclut qu'il est difficile de déterminer le meilleur modèle de contrôle d'animation lié à la parole à partir des tests d'évaluations disponibles dans la littérature. L'auteur souligne tout de même l'importance des tests subjectifs, car ce sont eux qui permettent d'évaluer le résultat final d'un système de synthèse pour une application donnée. L'auteure détaille sa méthodologie d'évaluation : il s'agit tout d'abord de réaliser une sélection de sujets suivant des critères d'âge, de sexe, de contexte socioprofessionnel varié, et de connaissances dans le domaine de l'expérience (l'animation graphique). Il y a ensuite une visualisation de séquences par les sujets : les sujets ont droit à une série d'entraînement/habitude, puis sont confrontés à la série test. Les séquences ne sont jouées qu'une seule fois. L'évaluation de l'animation est réalisée par réponse, selon une échelle à 5 valeurs (Très insuffisant, Insuffisant, Moyen, Bon, Très bon), à la question « Les mouvements du visage sont calculés par ordinateur à partir du son et utilisés pour animer un visage de synthèse. Sont-ils bien en cohérence avec la phrase prononcée ? ».

Pour évaluer la performance de leur système d'animation d'expression du visage, Deng et coll. ([DEN06]) ont conçu deux tests : le premier pour évaluer la synthèse produite par leur système, à partir de données nouvelles, le second pour comparer des expressions synthétisées avec les mêmes expressions réalisées par des locuteurs réels. Dans le second cas, les auteurs ont comparé leur modèle de coarticulation aux données réelles en analysant les différences de production synthèse/réel image par image. Ils n'impliquent pas de sujets extérieurs dans cette évaluation, se basant sur leurs mesures de similarité pour établir leurs conclusions. Ils se situent donc dans une évaluation purement objective de leur modèle.

Ces deux exemples illustrent les deux types de tests, objectif et subjectif, et mettent en avant le fait qu'il est possible de ne faire que des tests objectifs, mais que les deux types de test sont

importants : en effet, s'agissant de système d'interaction avec l'humain, il faut une évaluation humaine pour valider le système (compréhension, acceptabilité, etc.).

2. Agents Conversationnels Animés

Un ACA est une « entité logicielle plus ou moins autonome et intelligente, dotée d'une personnification et utilisée pour communiquer avec un utilisateur » selon la définition de Ruttkay ([RUT04]). Apparus dans les années 90, ce sont généralement des représentations graphiques 2d ou 3d, animées ou fixes, d'apparence humanoïde ou fantaisiste, de plein pied ou tronquées, capables ou non de communiquer de façon verbale et non-verbale ([DEV09]). L'ACA peut jouer plusieurs rôles : celui d'assistant (exemple du trombone de Microsoft), de tuteur (dans des applications pédagogiques) ou de représentation de l'utilisateur dans un environnement virtuel. L'objectif de la recherche sur les ACA est d'obtenir des agents intelligents capables de certains comportements. Leurs comportements doivent renforcer l'illusion qu'ils sont des entités sociales.

Ruttkay ([RUT04]) définit deux objectifs de l'évaluation des ACA :

- déterminer l'effet d'un ou de plusieurs paramètres de l'ACA sur la perception et la performance de l'utilisateur en évaluant l'ACA en lui-même.
- de déterminer l'apport de l'ACA à l'application, en évaluant l'ACA en temps qu'interface d'évaluation.

Selon Devos ([DEV09]), il existe autant de définitions d'un ACA que de champs d'études (psychologie, sociologie, ergonomie ou informatique). Ainsi, chaque communauté de recherche a une idée précise de ce qu'est un ACA et des critères qu'il est pertinent d'évaluer. Une taxonomie a donc été proposée pour aider à la création de critères d'évaluation des ACA. Cette taxonomie met en avant quatre critères de recherche sur les ACA : la crédibilité (believability), la sociabilité, les domaines d'application et les questions sur l'aspect et la performance de calcul (« agency and computational issues »). La conception d'un ACA s'intéresse à des aspects physiques et des aspects mentaux. Les premiers ont trait à l'aspect du corps (look) et aux modalités de communication (caractéristiques dynamiques des mouvements du visage et du corps). Les seconds regroupent les caractéristiques conversationnelles et sociales ainsi que la personnalité de l'ACA. Ce sont ces critères de recherche et ces aspects physiques et mentaux qui sont évalués.

Bailly ([BAI01]) explique que l'évaluation des ACA porte sur 3 aspects : l'ACA en lui-même (intelligibilité), l'interaction entre l'ACA et l'interface dans laquelle il est intégré (l'intégration multimodale) et l'impact sur l'utilisateur (charge cognitive). L'auteur cite l'expérience de Pandzic et coll. ([PAN99]) qui procèdent par étapes (en trois expériences) pour évaluer la compréhension de l'utilisateur, sa satisfaction et sa préférence (entre plusieurs propositions d'implémentation d'ACA).

Ces auteurs ont proposé des hypothèses et les ont évaluées de manière objective (par évaluation entre leur solution et des données réelles) et subjective (en demandant leur avis aux sujets, notamment en questionnant leur satisfaction). Dans ce dernier cas, un questionnaire post-évaluation a été proposé aux sujets, où les réponses étaient choisies sur une échelle numérique de 1 à 6. Les auteurs ont demandé non seulement aux sujets leur avis dans l'absolu, mais aussi en leur demandant d'interagir avec l'ACA dans un objectif donné : ceci permet de déterminer si le résultat a été atteint ou pas (mesure objective) et ce qu'en ont pensé les sujets (mesure subjective). Les réponses ont été récoltées et analysées par traitements statistiques (ANOVA).

Dans l'étude de Buisine ([BUI06]), où il est question de proposer des ACA avec des émotions, l'objectif est double : tester si les sujets perçoivent des émotions mélangées dans les animations de l'ACA de la même manière que dans les vidéos d'origine et comparer deux approches pour la reproduction d'émotions mélangées. Les auteurs procèdent par sélection des sujets suivant des critères de sexe et d'âge et par proposition de tâches de comparaison entre la vidéo et l'ACA, et deux implémentations d'ACA (qui ont été conçues suivant deux approches différentes). Les sujets devaient faire une opération de classement d'animations suivant leur similarité avec la vidéo d'origine, puis annoter les émotions qu'ils percevaient dans l'animation la mieux classée (par sélection d'un ou plusieurs labels émotionnels dans une liste). Pour traiter les résultats, les auteurs ont calculé le nombre de fois où chaque animation a été classée comme la plus similaire à la vidéo. Ils ont également mené une analyse de variance sur deux conditions d'expérimentations pour tester la variabilité intra sujet. Le classement des animations a été converti en scores de similarité (le 1er rang est devenu un score de similarité de 3 points, le dernier et 4ème rang un score de similarité de 0 point).

Chittarro et coll. ([CHI06]) soulignent l'importance d'une évaluation des ACA en faisant référence aux arguments de Ruttkay ([RUT04]). Dans leur étude, les auteurs évaluent l'application Mage-AniM qu'ils ont développée. Cette application permet à un utilisateur de créer des animations. Ils demandent à des utilisateurs, sélectionnés pour leur naïveté dans ce domaine, d'atteindre un certain objectif (réaliser une animation grâce à leur outil à partir de photos). Les sujets s'entraînent à manipuler l'interface puis tentent de réaliser la tâche demandée. Trois groupes d'utilisateurs sont constitués, pour réaliser trois tâches différentes : des animations d'ACA, une évaluation d'animations de LS, des animations de fitness. Dans le cas de la LS, les utilisateurs (locuteurs de LS) ont eu à juger de la compréhensibilité des animations, mais ils n'ont pas eu à les créer eux-mêmes. Ils ont donné leur avis de manière libre quant aux améliorations possibles des animations produites (notamment

ils ont souligné l'importance d'ajouter des expressions du visage, inexistantes à l'heure de l'évaluation dans le logiciel d'animation).

3. Gestes

Dans le domaine de l'animation de gestes, nous présentons une étude, avec des considérations issues d'une LS. Gibet ([GIB01]) ont pour objectif de synthétiser des mouvements communicatifs, à partir de considérations issues de la LSF. Les auteurs réalisent deux évaluations : la première par comparaison des gestes synthétisés avec des données réelles, et la seconde par vérification de la présence dans les gestes synthétisés de règles définies à partir d'analyses du réel. La comparaison porte sur les données chiffrées et les courbes de mouvements : il n'y a pas de participation de sujets à cette évaluation. Un corpus de gestes est construit à partir d'enregistrements de deux locuteurs via un dispositif de capture de mouvement, et les données d'accélération et de type de courbes pour le mouvement sont comparées avec celles produites par l'outil d'animation mis au point par les auteurs.

4. Langues des signes

Ces dernières années ont vu émerger des évaluations de signeurs virtuels dans le domaine des LS. Devos ([DEV09]) fournit un état de l'art assez complet sur ce sujet, dont nous citerons les grands points. Selon l'auteure, contrairement aux recherches menées sur les ACA, la quantité de travaux sur les SV est beaucoup plus restreinte puisque ce domaine est plus récent.

Dans le domaine de la reconnaissance de signes, Vogler ([VOG03]) évalue son système de reconnaissance en s'enregistrant lui-même (l'auteur est sourd, d'origine allemande, et locuteur de LS américaine au moment de son étude) et en testant le taux de reconnaissance de son système sur les données enregistrées. Il considère trois différentes erreurs que peut faire le système : des substitutions (au lieu de « papa lit livre » le système reconnaît « maman lit livre »), des suppressions (« papa livre » au lieu de « papa lit livre »), ou des insertions (« papa lit donne livre » au lieu de « papa lit livre »). L'auteur réalise un alignement entre les données reconnues et celles qui devaient être reconnues, afin de déterminer une distance entre ce qu'a reconnu le système et ce qu'il aurait du reconnaître. Afin de mesurer la prise en compte de l'épenthèse (ajout d'un mouvement entre deux signes, selon la définition de l'auteur), il a réalisé 5 expérimentations avec différents modèles de traitement de l'épenthèse (basés principalement sur des modèles de Markov cachés).

Dans le même domaine, Aran ([ARA08]) évalue son interface destinée à l'apprentissage de LS : des signes sont visualisés sur l'écran, le sujet doit les reproduire, et sa production est analysée par le système de reconnaissance. Son évaluation consiste tout d'abord à sélectionner ses sujets : suivant

leur sexe, leur participation à un cours d'introduction à la LS turque, leur implication dans des études scientifiques ou de langues étrangères, et leur grande motivation à participer à cette expérimentation. Ensuite, deux tâches leur sont proposées, avec à chaque fois des explications avant la réalisation de la tâche et un questionnaire après la réalisation de la tâche. Le temps de réponse des sujets a été enregistré tout au long de la tâche qui consistait à reproduire trois signes isolés. Les réponses aux questions devaient être fournies à l'aide d'une échelle de valeur (de 1 à 5).

Dans le domaine de la synthèse de LS, nous pouvons citer le projet européen ViSiCAST ([VER01]) et son successeur eSIGN ([ELL04]), qui ont eu pour objectif de fournir des informations en LS au moyen de signeurs virtuels. ViSiCAST propose des SV pour une utilisation dans le cadre de prévisions météorologiques, des transactions à la poste, et des applications multimédia et Web. eSIGN propose l'utilisation de la langue des signes sur des sites Web au moyen de SV.

Cox et coll. ([COX02]) présentent l'évaluation du SV Tessa, du projet ViSiCAST, dans le cadre de l'utilisation au guichet d'une poste : le SV est présent pour aider à la communication entre un guichetier et un client sourd, le premier utilisant une interface pour choisir quelles réponses à donner au client, le second s'exprimant en LS. Pour la communication depuis la personne sourde vers le guichetier, nous n'avons pas de détail : il n'y a pas de système de reconnaissance de LS implémenté, nous ignorons donc comment la demande est reçue et comprise par le guichetier. L'évaluation porte sur plusieurs aspects : sur la transaction en elle-même et sur le SV. Pour ce dernier, deux critères ont été évalués : la compréhensibilité de ce que le SV signait, et l'acceptabilité des signes réalisés. Ces critères sont différents dans le sens où, comme le souligne les auteurs, un parallèle peut être fait avec la synthèse vocale qui peut être comprise mais non appréciée (acceptée pour l'expérimentation, mais pressentie comme non acceptable dans une situation réelle) par les sujets. Les sujets sont sélectionnés suivant des critères de langue : ce sont des sujets dont la surdité a été décelée avant le stade pré-lingual, qui ont pour langue première la BSL (LS anglaise), recrutés via des réseaux sociaux de personnes sourdes. Les sujets ont évalué la compréhensibilité en visionnant des énoncés donnés par le SV et en couchant par écrit ce qu'ils avaient compris de ce qui avait été signé. Par comparaison entre ce qu'ils ont écrits et ce qui était signé, les auteurs de l'étude ont pu déceler certains passages du message compris alors que d'autres en étaient pas. Les sujets ont également eu à s'exprimer, sur une échelle de 1 à 5 sur la difficulté de comprendre le message. Enfin, à l'aide d'une échelle de valeur, de 1 à 3 cette fois (« peu acceptable » à « très acceptable »), les sujets se sont exprimés sur l'acceptabilité de ce qui était signé par rapport à la BSL naturelle. Une discussion a eu lieu à l'issue de

l'évaluation pour recueillir des informations plus détaillées de la part des sujets sourds sur le pourquoi de la non compréhension ou la non acceptabilité de certains messages.

Toujours dans le domaine de la synthèse, et plus particulièrement de l'évaluation de la synthèse de LS par SV, une étude remarquable est celle de Huenerfauth ([HUE06]) : l'auteur décrit plusieurs aspects de la méthodologie d'évaluation de la synthèse de LS par des sujets, depuis des considérations sur la sélection des sujets jusqu'aux modalités de réponse, en passant par le dispositif qui doit être mis en place. Nous détaillons maintenant son étude, première à notre connaissance à approfondir autant la recherche sur ces aspects.

L'auteur explique dans un premier temps que les méthodes d'évaluation automatique utilisées dans le cadre de la génération de langues dans leur forme écrite ne peuvent être appliquées aux LS : n'ayant pas de forme écrite, il n'est pas possible de faire tourner les algorithmes de calcul utilisés pour quantifier la distance entre le résultat du système et ce qu'il aurait dû produire. L'auteur préconise donc d'utiliser plutôt une évaluation avec des sujets, ce qui permet en outre de ne pas s'attacher uniquement à des résultats numériques et de pouvoir laisser à l'être humain le fait de considérer ce qui est bon ou pas, au-delà de critères figés et chiffrés. De plus, il est possible de proposer deux types de tâches aux sujets : une subjective dans laquelle ils estiment par exemple leur compréhension, et une objective dans laquelle est mesurée leur compréhension (ce qui permet de mettre en lumière des fausses impressions de la part des sujets). L'auteur a donc choisi de réaliser une expérimentation avec des sujets.

Il choisit ses sujets en fonction de leur qualité de « native speaker » (locuteur natif, en traduction littérale) : selon Huenerfauth, un « native speaker » d'ASL est une personne qui a appris l'ASL dans son enfance au travers d'interactions avec des membres sourds de sa famille, ou à l'internat pour enfants sourds dans son parcours éducatif. Afin de s'assurer du respect de ce critère, l'auteur ne pose pas directement la question aux sujets, mais leur demande si un membre de leur famille est sourd, s'ils ont utilisé l'ASL dès leur enfance, etc.

L'auteur préconise de fournir un environnement confortable pour le sujet, c'est-à-dire tout en ASL. Les interactions doivent se faire dans cette langue, idéalement par un autre « native speaker » afin de ne pas influencer l'expression linguistique du sujet. L'utilisation d'un interprète peut être envisagée, mais il faut que celui-ci s'exprime en ASL, et pas en américain signé (expression de signes d'ASL suivant la grammaire de l'américain), et qu'il soit « near native speaker » d'ASL (expression pour laquelle nous ne connaissons pas de définition). Il n'est pas recommandé que ce soit le chercheur qui mène l'étude qui interagisse avec le sujet lors de l'expérimentation, dans la mesure où cela peut influencer le sujet sur ce qu'il a envie d'exprimer : le sujet devine que le chercheur présente

est à l'origine du système qu'il est en train d'évaluer et ne sera pas aussi objectif que si c'est une autre personne (qui, dans l'esprit du sujet, ne serait pas associée à l'étude) qui assiste le sujet tout au long de l'expérimentation et lui pose les questions à la fin.

Enfin, concernant les questionnaires finaux, Huenerfauth préconise d'avoir le moins possible le recours à l'écrit : des groupes de mots sont préférables aux longues phrases, et des QCM ainsi que des échelles de valeur (de 1 à 10) sont proposés pour l'évaluation des animations. Les critères retenus pour l'évaluation des énoncés exprimés par le SV sont : la compréhensibilité (de « facile à comprendre » à « confus »), le naturel (de « fluide, comme une personne réelle » à « comme un robot ») et la grammaticalité (de « grammaire ASL parfaite » à « mélangé » ou « comme la grammaire de LV anglaise »). Les sujets ont aussi la possibilité de s'exprimer librement sur la vitesse de l'animation, la visibilité des mains du SV, la couleur et les lumières choisies, l'acceptabilité des mouvements des mains, des expressions du visage, et de la direction du regard.

II. Notre méthodologie d'évaluation

Nous souhaitons évaluer notre modèle de coarticulation dans le même contexte dans lequel nous avons réalisé notre analyse : celui des messages diffusés en gare par SV. Nous avons établi notre modèle à partir d'énoncés, il nous faut donc reprendre nos résultats et les appliquer à des énoncés, pour ensuite faire évaluer ces énoncés par des sujets. Nous nous orientons donc vers une évaluation subjective, avec des sujets dont le rôle sera d'évaluer le SV et son expression en LSF. Nous évaluons notre modèle en comparant les résultats d'évaluation donnés par les sujets sur des énoncés coarticulés et les mêmes énoncés non coarticulés.

Notre intention est de suivre les grandes étapes des études citées : sélection des sujets, expérimentations avec enregistrements vidéo des sujets, et questionnaires post expérimentations. Globalement, le protocole d'évaluation se déroule comme suit :

- 1) Demande d'informations sur le sujet.
- 2) Explications sur le déroulement général :
 - Visualisation de plusieurs animations (8 en tout), avec possibilité de revoir une animation un maximum de deux fois (donc 3 visionnages d'une animation).
 - Plusieurs questions sont posées après chaque animation, sur l'animation qui vient d'être visualisée.
 - Quant les huit animations ont été visualisées, une série de questions est posée sur le SV, puis sur le système de diffusion en gare.
- 3) Explications pratiques :
 - L'expression spontanée est encouragée, car tout ce qu'exprimera le sujet est filmé.

- Seul l'expérimentateur manipule le dispositif, le sujet est passif.
- 4) Passage de l'évaluation.
- 5) Explications finales sur le dispositif, mon étude en particulier, et discussion libre sur les thèmes intéressants le sujet.

Nous avons sélectionné quatre énoncés, deux de type « information » et deux de type « incident », dont nous avons décliné trois instances :

- Information

○ Enoncé AS01

- AS01-1 « Voie A éloignez-vous de la bordure du quai. Attention au passage d'un train. Voie A éloignez vous de la bordure du quai. »
- AS01-2 « Voie 12 éloignez-vous de la bordure du quai. Attention au passage d'un train. Voie 12 éloignez vous de la bordure du quai. »
- AS01-3 « Voie X éloignez-vous de la bordure du quai. Attention au passage d'un train. Voie X éloignez vous de la bordure du quai. »

○ Enoncé D0Y

- D0Y-1 « Voie B, le TGV numéro 4567 à destination de Colmar va partir. Prenez garde à la fermeture des portes. »
- D0Y-2 « Voie 12, le TER numéro 12 047 à destination de Strasbourg va partir. Prenez garde à la fermeture des portes. »
- D0Y-3 « Voie C, le Corail numéro 88 676 à destination de Paris Est va partir. Prenez garde à la fermeture des portes. »

- Incident

○ Enoncé S0Y

- S0Y-1 « Votre attention s'il vous plait, voie A, en raison de mouvements sociaux, le TER numéro 88 634 à destination de Reims est supprimé. La SNCF vous présente ses excuses. »
- S0Y-2 « Votre attention s'il vous plait, voie B, en raison de travaux sur les voies, le TER numéro 88 676 à destination de Strasbourg est supprimé. La SNCF vous présente ses excuses. »
- S0Y-3 « attention s'il vous plait, voie C, en raison d'un incident technique, le TGV numéro 2047 à destination de Paris Est est supprimé. La SNCF vous présente ses excuses. »

○ Enoncé R0YA

- ROYA-1 « Votre attention s’il vous plait, voie A, en raison de travaux sur les voies, le TGV, numéro 8634 à destination de Reims et Strasbourg partira à 8h34. »
- ROYA-2 « Votre attention s’il vous plait, voie B, en raison de travaux sur les voies, le TER, numéro 1234 à destination de Colmar et Reims partira à 8h12. »
- ROYA-3 « Votre attention s’il vous plait, voie X, en raison d’un incident technique, le TER, numéro 8634 à destination de Strasbourg et Reims partira à 17h34. »

Nous avons choisi ces instances parce qu’elles sont des deux types, « information » et « incident », que deux d’entre elles font partie de notre corpus (donc nous voudrions vérifier que notre proposition de variation est acceptée sur ces énoncés en particulier), et surtout qu’elles nous permettent d’évaluer notre modèle sur des segments d’énoncés dans lesquels nous avons relevé d’importantes variations en valeur.

Nous avons présenté les différentes instances à nos sujets dans l’ordre décrit dans le tableau 24. Nous avons choisi de présenter successivement une instance d’énoncé réalisée selon la concaténation de base (ce sont les noms d’instances qui terminent par « _base »), et l’instance de ce même énoncé réalisée avec le modèle de coarticulation que nous proposons (noms se terminant par « _modifiée »). Ceci permet aux sujets de comparer directement les deux réalisations et de pouvoir relever les différences (en mieux ou en moins bien) qu’ils ont perçues. Nous varions l’ordre de présentation afin que le sujet ne soit pas influencé par le fait qu’on présenterait toujours une même version d’une instance (de base ou modifiée) avant l’autre. Pour le détail des animations qui ont été modifiées selon les règles de notre modèle, se reporter en annexe III.

Ordre de présentation	Sujet 1	Sujet 2	Sujet 3
1	AS01-1_base	SOY-3_base	SOY-2_base
2	AS01-1_modifie	SOY-3_modifie	SOY-2_modifie
3	DOY-2_modifie	DOY-1_modifie	ROYA-3_modifie
4	DOY-2_base	DOY-1_base	ROYA-3_base
5	SOY-1_base	ROYA-1_modifie	AS01-2_base
6	SOY-1_modifie	ROYA-1_base	AS01-2_modifie
7	ROYA-2_modifie	AS01-3_base	DOY-3_modifie
8	ROYA-2_base	AS01-3_modifie	DOY-3_base

Tableau 25 : Ordre de présentation des instances aux trois sujets

Pour la création des animations modifiées, qui composent les instances d'énoncés modifiées, nous avons récupéré les animations 3d réalisées par les infographistes de Websourd, et les avons modifiées manuellement (dans le logiciel 3DSMax) suivant les règles de notre modèle de coarticulation. L'interpolation entre les animations a été réalisée grâce à une interface développée pour notre étude par le LIMSI-CNRS, qui permet de faire le lien entre une posture de fin de signe (le terme « posture » regroupant l'ensemble des caractéristiques du SV à un instant donné) et la posture de début du signe suivant.

1. Sélection des sujets

Les sujets n'ont pas été sélectionnés suivant la caractéristique de « native speaker » : nous laissons ce terme entre guillemets, car il n'est pas défini de manière identique par tous les auteurs qui l'emploient, du moins dans le domaine des LS. En effet, qu'est-ce qu'un locuteur natif, étant donné la diversité des personnes locutrices de la LSF à ce jour :

- Est-ce un enfant sourd, né dans une famille sourde ? Mais si ces sourds oralisent, alors la LSF peut avoir été apprise sur le tard, du coup le terme « natif » nous paraît inadapté.
- Est-ce un enfant sourd, né dans une famille où est présente la LSF ? Mais si les parents sont entendants, il est très probable qu'ils apprennent en même temps que leur enfant cette langue : peut-on qualifier la langue de l'enfant comme étant « native » dans le sens où c'est une langue seconde voire troisième pour les parents ?

- Est-ce un enfant, né dans une famille sourde où la LSF est présente ? Mais si l'enfant est entendant, alors sa langue première sera le français oral (appris à l'école) quand sa langue maternelle sera la LSF : quid de sa langue dite « native » ?

Nous ne nous étendrons pas plus sur le sujet, il y a encore de nombreux cas qui peuvent être envisagés, quand on ajoute à ces questions des considérations de transmissions intergénérationnelles (une langue n'est pas qu'une grammaire et un vocabulaire, c'est aussi une culture, et dans le cas de la LSF ce sont des modalités différentes pour s'exprimer, ce qui rajoute encore de la complexité à la définition de « natif » puisqu'il faut peut-être séparer les personnes issues de générations de familles sourdes et celles issues de familles entendants bilingues), de variations régionales, de situations sociétales (l'interdiction de la LSF dans l'enseignement de 1880 à 1991 a influencé nombre de locuteurs de LSF, et cela a encore des répercussions aujourd'hui notamment vis-à-vis de l'influence du français dans la LSF), d'éducation (quelques classes dites bilingues français écrit/LSF existent aujourd'hui, et si nous laissons de côté la définition encore ambiguë de « bilingue », il nous faut décider si seuls les enfants ayant suivi un tel parcours sont « natifs » ou si d'autres peuvent être considérés comme tel), etc. Nos sujets sont donc des personnes sourdes locutrices de la LSF, issus de régions différentes (sud et nord de la France), utilisant la LSF dans leur vie quotidienne, professionnelle comme publique, peu importe leur contexte familial ou éducatif : par ce choix, nous estimons pouvoir rassembler des locuteurs de LSF, ceux qui seront plus tard utilisateurs du système de diffusion en gare (et pas une niche de locuteurs « idéaux », pour reprendre l'expression de Huenerfauth).

Nous avons recruté plusieurs locuteurs de LSF, en prenant soin de noter certaines informations qui peuvent nous permettre de mettre en perspectives les résultats des évaluations. Ces informations sont les suivantes :

- sexe (H ou F) ;
- tranche d'âge (18 à 30, 31 à 40, 41 à 50, 50 et plus) ;
- fréquence d'utilisation des transports ferrés, surtout train (en moyenne sur l'année : « plusieurs fois par semaine », « une fois par semaine », « une fois par mois », « rarement », « jamais ») ;
- niveau de connaissance des nouvelles technologies, et des signeurs virtuels en particulier (« je connais très bien », « je connais un peu », « je ne connais pas du tout »).

2. Déroulement des évaluations

Nous menons personnellement les expérimentations, bien qu'Huenerfauth préconise que ce soit un « native speaker » qui mène les échanges. Cette préconisation est là encore délicate à mettre en œuvre : la LSF étant diverse en fonction des personnes, de leur parcours, de leur militantisme, de

leur style d'expression, il est difficile de garantir qu'un « native speaker » n'influencera pas le sujet. Par exemple, si le sujet est quelqu'un qui s'exprime beaucoup en signes lexicalisés et peu en utilisant des structures de grande iconicité (cf. la théorie de Cuxac dans le chapitre Etat de l'art) et que le « native speaker » censé le mettre dans une situation confortable est dans une expression différente (beaucoup de structures de grande iconicité, peu de signes lexicalisés) il peut y avoir une influence sur le sujet. Hunerfauth préconise également l'utilisation d'un interprète seulement si celui-ci s'exprime comme un « near-native » (là aussi la définition est floue, celle-ci n'étant pas définie par l'auteur, et n'étant pas utilisée par d'autres auteurs), en insistant sur le fait qu'il faut lui demander de s'exprimer en ASL et pas en américain signé : les interprètes sont pourtant censé être bilingues, c'est-à-dire capables de s'exprimer en ASL et en américain, et ne pas avoir à s'adapter à leur interlocuteur, donc nous ne comprenons pas le pourquoi de cette insistance (d'autant plus que le sujet est censé être un « native speaker » donc même s'il y a adaptation de l'interprète ce vers une version « native » de l'ASL).

Etant interprète (donc bilingue), nous n'interagissons pas via une autre personne mais en direct avec le sujet, ce qui permet d'éliminer le biais de l'interprétation et d'assurer une communication sûre entre le sujet et nous-mêmes. Reste le biais du chercheur présent qui peut influencer le sujet quant à l'objectivité de ses réponses : nous ne pouvons pas prétendre éliminer ce biais, cependant les personnes sourdes sélectionnées sont averties qu'il faut qu'elles soient franches si elles veulent disposer d'un résultat acceptable en gare.

Nous interrogeons tout d'abord le sujet pour obtenir les informations sur sa tranche d'âge, ses connaissances des SV et sa fréquence d'utilisation des transports ferrés. Ensuite nous réalisons l'évaluation. Enfin, nous discutons librement avec le sujet sur son ressenti, nous répondons aux questions qu'il a formulées lors de l'évaluation (auxquelles nous n'avons pas répondu pour ne pas influencer son évaluation), et sur la finalité de notre étude dans le cadre du projet de diffusion d'information par SV en gare.

L'expérimentation consiste en la présentation d'animations :

- sans modèle de coarticulation (simple concaténation des signes pour former l'énoncé) ;
- avec notre modèle de coarticulation.

Chaque animation est vue au maximum trois fois de suite, puisque lorsque le système sera en place dans les gares ferroviaires, le message sera diffusé trois fois sur les écrans.

Les remarques du sujet sont enregistrées en vidéo tout au long de l'expérimentation, afin qu'il s'exprime le plus librement possible en LSF.

3. Questionnaires

Les questionnaires post-expérimentaux portent sur trois aspects : le SV, la compréhension du message, le système d'information. Concernant ce dernier point, il est tout d'abord demandé de déterminer sur une échelle de 1 à 10 si selon le sujet le message a été bien compris, si la compréhension a été aisée, et si l'expression en LSF est acceptable. Pour l'évaluation du SV, nous nous intéressons à la compréhension qu'en ont les sujets, à leur ressenti quant à l'animation de différentes parties du corps comme les mains, le regard, les expressions du visage et les mouvements du corps, et à leur avis global sur l'expression LSF su SV. Enfin, une évaluation de la pertinence du système est réalisée au travers de questions portant sur la pertinence d'avoir un SV dans les gares, la pertinence de diffuser ces informations (et pas d'autres), la pertinence du dispositif d'écrans (plutôt que l'envoi sur mobile, par exemple) dans les gares.

III. Résultats

Nous avons demandé à trois sujets de participer à notre évaluation. Ce sont trois hommes, de la tranche d'âge 31 à 40 pour l'un et 41 à 50 pour les deux autres. Nos sujets ont une connaissance de base sur les signeurs virtuels, c'est-à-dire qu'ils en connaissent l'existence, qu'ils ont vus le signeur virtuel de démonstration d'une gare ferroviaire donner des informations (il s'agit d'énoncés enregistrés intégralement, qui ne sont pas générés par concaténation de différents segments), mais qu'ils n'en savent pas plus quant à leur conception ou leur contrôle. Deux sujets prennent le train régulièrement (plusieurs fois par mois) et le troisième ne le prend qu'une fois tous les deux mois.

Nous présentons dans un premier temps les résultats de l'évaluation de notre modèle de coarticulation, puis des résultats sur le signeur virtuel en lui-même et le système de diffusion en gare.

a. Evaluation du modèle de coarticulation

Compréhension

Globalement, les sujets ont assez peu compris les informations données par le signeur virtuel. Ils soulignent un manque de fluidité, des signes qu'ils ne connaissent pas (ce qui implique soit une incompréhension de l'énoncé soit un malentendu), une réalisation de la dactylogogie et des chiffres beaucoup trop rapide. Un des sujets n'a rien compris du tout : il a bien deviné plusieurs informations, mais quelques signes lui échappaient, qui auraient pu apporter un sens nouveau à ce qu'il avait déjà compris, donc il ne pouvait prétendre avoir compris. Les sujets ont dans l'ensemble pointé le manque de contexte, le fait de ne pas avoir pour objectif de prendre un train, avec son billet à la main : selon eux, cela aurait aidé à mieux comprendre les informations données, surtout si elles se rapportent à

leur train. Même si la compréhension globale a été nulle dans un cas, dans tous les cas les sujets font part du bénéfice de la répétition des énoncés, qui leurs donnent l'habitude des spécificités d'expression du SV, qui rend les dernières instances d'énoncés plus compréhensibles que les premières. Enfin, les sujets ont tous relevés une certaine difficulté à comprendre le message donné : il leur a fallu se concentrer, et tous ont répété les signes réalisés pour mieux se les approprier, les mémoriser, et essayer d'extraire du sens.

Comparaison par énoncés

Les instances dont la réalisation a été acceptée par les sujets sont celles des énoncés AS01 et SOY, tandis que celles des énoncés DOY et ROYA ont été complètement rejetées par tous les sujets. Ce score correspond à la compréhension moyenne pour chaque instance d'énoncé exprimée par les sujets. Nous pouvons remarquer que les sujets acceptent certains énoncés sans les comprendre en détail, mais soulignent eux-mêmes que leur acceptation n'est pas complète et qu'elle pourrait l'être si certaines modifications étaient apportées (au niveau des choix de traduction surtout, pour une expression plus iconique).

Nous ne pouvons faire de conclusion quant à un type d'énoncé, information ou incident, en particulier, puisqu'un énoncé de chaque type se retrouve soit compris et accepté soit incompris et rejeté.

Comparaison entre instances sans coarticulation (base) et avec coarticulation (modifiée)

Le fait que l'instance soit modifiée par le modèle de coarticulation ne facilite pas la compréhension des sujets, de même qu'il ne la pénalise pas. La facilité de compréhension est elle aussi jugée identique, que soit présentée une « instance coarticulée » (cela correspond aux instances que dont nous avons modifié des segments) ou non. Pour un des sujets il semblerait que les instances modifiées soit plus acceptables, mais cela ne ressort pas de manière assez flagrante pour que nous puissions être plus affirmatif. Cependant, il s'avère que le nombre de répétitions nécessaires aux sujets pour se faire une idée de l'énoncé (considérer qu'ils ont compris ou que de toutes façons ils ne comprendront pas plus) est inférieur en moyenne dans le cas des énoncés coarticulés que pour ceux qui ne le sont pas : il y a entre deux et trois répétitions pour les énoncés non coarticulés, quand il y en a entre une et deux pour les énoncés coarticulés.

Nous ne relevons donc pas de différence très significative entre la version d'une instance avec une interpolation et celle proposée avec les variations de notre modèle de coarticulation, au-delà du nombre de répétitions nécessaire qui est légèrement différent.

Bilan

Nous n'avons pas poursuivi notre évaluation au-delà de trois sujets, et n'avons pas approfondi les questions pour focaliser l'attention des sujets sur des points particuliers de notre modèle. En effet, au vu des résultats quasi identiques entre les sujets, nous pressentons que d'autres sujets ne feraient que confirmer ce que nous constatons : au vu des problèmes de compréhensibilité, notre évaluation ne peut pas nous apporter d'information sur la validité de notre modèle de coarticulation. D'autre part, les sujets ont été tellement dans la recherche de sens, dans la compréhension des signes, qu'il n'est pas possible de leur demander de faire abstraction de ces incompréhensions pour ne se focaliser que sur le regard ou la configuration des mains : sans sens, que la configuration des mains ou la direction du regard varie, cela n'a aucune pertinence et ne peut être qualifié de « mieux » ou « moins bien » par le sujet.

Nous relevons des problèmes à plusieurs niveaux.

Tout d'abord au niveau de la traduction des énoncés initiaux venant de la SNCF, qui sont au final perçus comme n'étant pas exprimés en LSF, mais plutôt dans une syntaxe de français signé : partir d'une idée de modularité finale des segments d'énoncés entre eux ne pouvaient qu'induire le biais de traduire certains énoncés de la même manière afin de réduire leur nombre et la complexité finale des animations à composer. Toujours au niveau de la traduction, il aurait fallu, selon les sujets, privilégier la grande iconicité aux signes lexicalisés : ceci a pour inconvénient de multiplier les segments d'énoncés, puisque la grande iconicité implique plus de relations entre signes réalisés dans l'espace de signation par rapport à la situation réelle (il existe beaucoup de situations différentes dans les différentes gares, par exemple les voies les unes par rapport aux autres). Enfin, certains signes ne peuvent être modifiés en structure de grande iconicité, par exemple les noms des villes : ces signes ont été choisis parmi plusieurs existants suite à une expertise menée par notre experte LSF. Il s'avère que d'autres consultations doivent être menées afin de proposer via le SV des signes acceptés par une plus grande majorité (par exemple pour le signe de « Strasbourg »).

Ensuite au niveau de la réalisation des animations 3d, nous relevons des problèmes de rythme, de configurations, d'expressions du visage qui sont absentes, etc. Ces malfaçons entraînent de fait des incompréhensions plus ou moins totales de l'énoncé.

Enfin, le logiciel de contrôle des animations, qui nous a permis de réaliser nos évaluations, propose une interpolation des emplacements entre les segments d'énoncés (puisque nous ne traitons pas les variations au niveau des emplacements dans notre modèle) : cette interpolation se réalise par troncature des animations, la première se voyant rognée à la fin, la suivante étant rognée au début. Du fait que ces animations sont réalisées avec des temps de pause différents au départ, leur interpolation ensuite peut très bien se passer, mais peut aussi entraîner des saccades entre les

différents signes. Ces biais influent très fortement sur la compréhension du message par le sujet, et implique que le sujet se focalise sur le sens plutôt que sur des aspects plus fins comme la coarticulation. Une dernière inadéquation vient de notre modèle de coarticulation en lui-même, qui propose des variations que nous qualifierons de « fines » : nous ne proposons pas de modifier intégralement certains signes ou l'ordre de certains segments d'énoncé, ce qui aurait été sûrement noté par les sujets. Ces fines variations, déjà initialement peu visibles, mais dont nous pensions que la présence allait faciliter l'aisance à la compréhension, et l'absence la pénaliser, sont d'autant plus difficiles à appréhender si le sens du message vous échappe, et encore plus difficiles si vous vous focalisez alors uniquement sur ce sens à extraire (ce qu'ont fait les trois sujets).

Au vu de ce bilan, nous proposons une autre méthode d'évaluation, dite « objective » : il s'agit de comparer nos résultats avec les résultats d'analyse issus de corpus avec d'autres locuteurs. Nous revenons sur ce point dans le chapitre suivant.

b. Evaluation du SV et du système de diffusion en gare

Nous avons également réalisé une évaluation plus globale du signeur virtuel et du système de diffusion d'information en gare, afin d'avoir une idée de leur perception et de leur acceptation par les sujets.

Le signeur virtuel

Les commentaires des sujets sur le SV, au-delà des questions que nous avons posées, font remonter un manque de visibilité de certaines configurations ou parties du corps dû à un manque de contraste entre les couleurs de peau de la main et du visage, et des doigts entre eux. Le rythme est également pointé du doigt comme n'étant pas satisfaisant, tantôt trop rapide, tantôt trop lent, voire saccadé, au sein même des signes.

Le SV est globalement jugé peu compréhensible par les sujets, sans expression du visage ni direction de regard changeante. Deux sujets ont souligné qu'ils n'ont pas vu les expressions du visage et/ou différentes directions de regard parce qu'ils se sont focalisés sur les mains afin d'en extraire du sens. Le sujet qui estime y avoir fait attention juge les directions de regard et les expressions du visage peu satisfaisantes.

Le système de diffusion en gare

Les sujets ont apprécié les informations données et le principe de diffusion d'annonces en gare par signeur virtuel en LSF. Cependant, nous rappelons que la compréhension des informations données n'est pas satisfaisante : les sujets ont d'une part pressenti que les informations données (surtout celles traitant d'incidents, moins celles d'informations générales) sont utiles et qu'elles

doivent leur être accessibles, et d'autre part mettent en avant le côté militant d'un tel dispositif, qui permet enfin d'afficher la LSF en public afin de sensibiliser le grand public (à la LSF mais aussi au fait que les sourds existent, pour répondre à Mottez [MOT06]). Les sujets ont tous une lecture aisée en français, ce qui implique qu'ils prennent déjà le train sans problème : l'information en LSF leur serait très utile en train, que ce soit par écran personnel (comme dans les avions) ou sur téléphone portable. En effet, un sujet a proposé de pouvoir recevoir les informations concernant son train directement sur son téléphone portable, et un autre a estimé qu'il serait bien de pouvoir recevoir les informations données sur les écrans en gare en même temps sur son téléphone portable.

c. Synthèse

Nous avons réalisé une évaluation en partant de l'hypothèse implicite que les animations réalisées par rotoscopie étaient acceptées et fidèles au corpus vidéo source. Or le procédé de rotoscopie implique un choix d'images à utiliser pour l'animation 3d, ce qui entraîne de fait une animation dont la fidélité à la réalité dépend de l'expertise et du talent de l'infographiste, et leur réalisation actuelle n'est pas acceptée par les sujets auxquels nous les avons montrés. Ainsi, nous constatons que les sujets ne peuvent arriver à une évaluation de notre modèle, celui-ci intervenant à un niveau plus fin que les problèmes relevés. Cette évaluation nous permet cependant d'avoir des retours sur le signeur virtuel et le système de diffusion en gare, qui bien que ne servant pas directement dans le cadre de notre étude de la coarticulation, permettrait une meilleure acceptation du dispositif et du SV, ce qui serait un pas vers une évaluation possible de notre modèle.

Nous avons fait état dans ce chapitre des résultats que nous avons obtenus suite à l'évaluation subjective de notre modèle auprès de différents sujets. Nous présentons dans le chapitre suivant une conclusion générale de notre travail, puis nous ouvrons des discussions quant à certains points de notre méthodologie, de notre modèle de coarticulation, et à notre évaluation, et enfin nous proposons plusieurs perspectives à notre travail.

Chapitre VI. CONCLUSIONS, DISCUSSIONS ET PERSPECTIVES

Dans ce dernier chapitre, nous dressons le bilan de notre étude, en rappelant les différentes propositions méthodologiques que nous avons faites. Puis nous discutons de quelques aspects de notre recherche et du projet dans lequel elle s'est tenue. Enfin, nous présentons quelques perspectives à notre étude, soit pour l'approfondir, soit pour l'utiliser dans d'autres domaines.

I. Conclusions

Notre étude a pour objectif la conception et l'implémentation d'un modèle de coarticulation pour l'animation d'un signeur virtuel, ce dernier étant utilisé dans un dispositif de diffusion d'information en LSF dans les gares ferroviaires. Nous ne prétendons pas proposer un modèle s'appliquant à la langue dans son ensemble, mais à un domaine linguistique particulier (les informations diffusées en gare ferroviaire), bien que nous pensions que certaines variations relevées pourraient être des pistes de recherche pour des études sur d'autres domaines.

Nous avons proposé une méthodologie d'analyse en plusieurs étapes, issue d'une analyse des méthodologies existantes dans différents domaines : les langues vocales orales et écrites, les gestes, les animations de têtes parlantes, et les langues des signes. Notre objectif étant la conception d'un modèle, nous avons opté pour une analyse « corpus-driven » où nous ne posons pas d'hypothèse mais les formulons d'après le corpus. La première étape de notre méthodologie consiste en la sélection du contenu de ce corpus, et de la forme de ce contenu (au niveau technique pour la réalisation du corpus, mais aussi au niveau du choix du locuteur du corpus). Ensuite, vient l'étape d'annotation du corpus, pour laquelle nous avons fait des choix de valeurs d'éléments d'annotation (utilisation de captures d'écrans comme valeurs visuelles d'annotation), et des choix de segmentation de ces éléments (détermination du début de la fin d'une valeur d'élément). Enfin, nous avons conçu et implémenté un outil d'analyse par alignement (issu du domaine de la bioinformatique) des annotations, afin mettre en exergue les variations dues à la coarticulation dans notre corpus. Ces alignements ont été traités puis mis en forme dans un tableur, pour nous permettre d'en tirer des résultats. Ces résultats constituent la base de notre modèle de coarticulation.

Notre modèle consiste en un ensemble de règles qui s'appliquent en deux étapes, la première n'étant pas vraiment de la coarticulation mais étant nécessaire pour la mise en œuvre de la coarticulation qui vient en seconde étape. Sa description synthétique est la suivante :

1. Contextualisation sémantique :
 - Elévation de la direction du regard
 - Ajout de directions du regard pour insister sur certains emplacements (en anticipation de la réalisation de signes par les mains)
 - Suppressions de signes
2. Coarticulation (contextualisation phonétique) :
 - Configurations des mains
 - Assimilation
 - Influence par persévération
 - Influence de mD vers md
 - Direction du regard
 - Délai minimal entre 2 directions du regard identiques

Nous avons également proposé une méthodologie d'évaluation de notre modèle. Cependant il ressort que des biais à plusieurs niveaux (principalement linguistique et animation 3d) ne nous permettent pas de pouvoir conclure quoi que ce soit quant à notre proposition de modèle de coarticulation. Nous ouvrons la discussion dans la partie suivante sur d'éventuelles possibilités d'autres évaluations pour notre modèle, et précisons une d'entre elles en perspectives.

Les apports de notre recherche sont multiples. Nous proposons une méthodologie complète d'analyse de la coarticulation en LSF, depuis la conception du corpus avec le choix du fond et de la forme, jusqu'à l'analyse des annotations, en passant par le choix de ces annotations. Nous proposons également une méthodologie d'évaluation, qui n'est pas encore utilisable aujourd'hui au vu des biais répertoriés, mais qui pourra être utilisée lorsque les biais auront été corrigés (ce qui est envisageable à court terme). Nous avons de plus implémenté un algorithme d'alignement qui peut être utilisé pour d'autres études sur les variations : l'intérêt de notre implémentation est qu'elle permet de prendre n'importe quelles annotations et de les formater pour être adaptées à l'algorithme d'alignement. Ainsi, ce ne sont pas seulement les variations phonétiques mais des variations d'annotations qui peuvent être analysées, quelque soit la nature des annotations. Enfin, nous avons proposé un état de l'art issu de différents domaines qui peut servir de point de départ solide pour d'autres études dans ce domaine.

II. Discussions

1. Modèle de coarticulation

Nous n'avons pas trouvé de référence sur le sujet, mais il nous semble qu'il y a un « fait exprès » de la qualité moyenne de la coarticulation pour la synthèse vocale. Ceci serait dû au fait qu'il ne faut pas que la voix synthétisée soit trop réaliste sinon les usagers n'y prêteront pas plus attention qu'aux voix des autres usagers présents dans la gare. Transposé au système de synthèse de LSF, il nous paraît intéressant de considérer que la coarticulation doit être présente pour que les usagers ne peinent pas à comprendre l'information, mais qu'il est peut-être judicieux, dans le domaine des gares ferroviaires, que la coarticulation des signes ne soit pas complètement naturelle afin que les usagers prennent en compte le message. A la différence près que pour les usagers entendants, l'information leur parvient de quelque endroit de la gare qu'ils soient, alors que les usagers sourds sont obligés de se focaliser sur les écrans : ainsi, une coarticulation « naturelle » pourrait ne pas être gênante au même titre que dans le système en LV. Ce point de vue, à savoir s'il faut de la coarticulation (donc au niveau phonétique) pour le système de diffusion en gare, et le cas échéant s'il faut un modèle complet ou de grandes règles générales, serait intéressant à creuser dans une future étude, où il serait question de comparer la synthèse sans coarticulation, avec une « simple » coarticulation (il faudra définir précisément ce qu'est une « simple » coarticulation), et avec une coarticulation complète. Ceci ne permettrait pas de valider le modèle de coarticulation, mais de valider le fait qu'il faut qu'il y en ait un pour le système de diffusion en gare, et par là même valider le fait que s'il en faut un, alors une analyse telle que nous l'avons proposée est pertinente.

2. Signeur Virtuel

Le projet WebSourd-SNCF vise à fournir en LSF les mêmes informations que celles données par le système de synthèse vocale. Cependant, toutes les informations ne sont pas jugées comme pertinentes par les sujets : il ressort qu'a priori les informations d'incident seraient souhaitées, mais pas les informations générales. Ceci est à mettre en lien avec les considérations d'Abrahams¹⁸ lorsqu'il fait un tour d'horizon des pour et contre l'utilisation d'un signeur virtuel. Selon cet auteur, qui résume des échanges entre chercheurs créateurs de signeurs virtuels en BSL (LS anglaise) et des personnes sourdes, les arguments contre les signeurs virtuels sont :

- La mauvaise qualité des signeurs virtuel actuels.

¹⁸ <http://www.it-director.com/business/compliance/content.php?cid=10365>

- L'utilisation d'un signeur virtuel implique une certaine normalisation des signes, qui risque d'entraîner un appauvrissement de la langue.
- L'argent utilisé pour la conception des signeurs virtuels devrait plutôt être utilisée pour former des interprètes.
- Une vidéo d'une personne sourde serait beaucoup plus utile.
- Les personnes sourdes savent lire, même si c'est de façon un peu limitée, et l'utilisation de signeurs virtuels pour les messages courts n'est pas pertinente.

Les arguments en faveur des signeurs virtuels sont :

- Nous en sommes au début des recherches, la technologie va s'améliorer (et devenir acceptable).
- Les enfants et les jeunes sont beaucoup exposés aux mondes virtuels et dessinés, ils seront surement plus ouverts aux signeurs virtuels que les adultes d'aujourd'hui.
- Dans un contexte particulier de diffusion d'information il est utile de disposer d'un signeur virtuel qui permet de modifier aisément les messages produits, ce qui n'est pas le cas avec la vidéo.

L'auteur ajoute que les personnes sourdes ont souligné l'intérêt de disposer d'un signeur virtuel notamment pour la diffusion d'information dans les gares, mais que les informations écrites étaient à ce jour moins chères, suffisantes, et accessibles à la majorité des personnes sourdes.

Dans le cadre de notre projet il ne s'agit pas de traduire en LSF les informations écrites, mais bien de donner une version LSF des informations vocales : nous sommes donc en adéquation avec les remarques ci-dessus, et nos sujets ont de leur côté précisé le type de message qu'il faut diffuser (des messages d'incident plutôt que des informations générales), tout en reconnaissant que pour les informations générales, la forme écrite actuelle (sur les tableaux d'affichage) pouvait tout à fait convenir. Bien sûr, ce type de conclusion demande à être vérifiée sur un plus grand nombre de sujets.

3. Système de diffusion en gare

Se pose également la question de l'emplacement de l'affichage : dans le hall de la gare c'est un minimum, mais il faut aussi que soit diffusée cette information sur les quais, et idéalement dans les trains (pour que les passagers soient informés des problèmes survenant au train dans lequel ils sont). Pour ce faire, plutôt que des écrans, il pourrait être intéressant de proposer l'envoi de l'information vidéo sur le mobile de l'utilisateur (il faut en amont réfléchir à quelles informations diffuser, pour cela une inscription de l'utilisateur pourrait se faire en amont, afin de déterminer s'il faut lui annoncer que tel train a du retard ou pas). A cette question de l'affichage en gare se greffe le problème de l'information comme quoi un message va être diffusé ou est en train d'être diffusé : pour les

messages vocaux, un jingle est joué, puis le message est donné. Pour les messages en LSF, qui ne seront perçus par les personnes sourdes qu'à la condition qu'elles regardent l'écran, il faudra imaginer un système d'avertissement pour prévenir qu'un message va être délivré. En effet, il n'est pas possible de demander aux personnes sourdes de fixer continuellement l'écran d'affichage, pour être sûr qu'elles ne ratent pas une information les concernant. Le recours à l'information sur téléphone portable permet de s'affranchir de cette contrainte.

III. Perspectives

Nous présentons dans cette partie plusieurs possibilités de poursuite de notre étude, à plusieurs niveaux : le corpus, l'annotation, l'analyse, l'évaluation, et terminons par une présentation succincte de l'utilisation potentielle de notre modèle avec d'autres résultats ou études.

1. Possibilités d'amélioration de nos résultats

a. Le corpus

Nous avons utilisé un corpus vidéo, enregistré avec des caméras numériques. Au début de notre étude, la seule réalisation possible connue du SV était la méthode de rotoscopie, qui implique de disposer de vidéos de référence pour la création des animations. De plus, les techniques de capture de mouvement n'étaient pas assez bien maîtrisées, pour ce qui concerne l'enregistrement de LSF, pour être utilisées dans le cadre de notre étude. Depuis, des avancées ont été faites, et des tests menés par Websourd sur certains énoncés enregistrés par capture de mouvement se sont révélés prometteurs. Une possibilité d'amélioration de nos résultats serait d'utiliser un corpus de capture de mouvement, ce qui permettrait de raisonner pour l'analyse sur les mêmes données que celles qui seront utilisées lors de la diffusion des informations (et non pas comme aujourd'hui de réaliser notre analyse sur des vidéos et d'appliquer ces résultats sur des animations 3d différentes des vidéos initiales). Cependant, cela entraîne d'autres contraintes, comme de redéfinir les annotations, puisqu'elles s'appliqueraient directement sur le résultat de la capture de mouvement et concerneraient une réduction des données enregistrées : par exemple, l'annotation de la configuration des mains se ferait d'après un ensemble restreint de points et non pas d'après une vision globale de la configuration (donc, le positionnement des capteurs sur les mains et leur nombre limité réduit la vision de l'annotateur à des informations choisies en amont, pas au moment de l'annotation). Notons qu'il existe des études sur le positionnement des capteurs dans le cas d'enregistrement de LS par capture de mouvement, entre autres celles de Mauk [MAU08] (où les auteurs explicitent les choix faits pour le positionnement des capteurs) et Chetelat-Pele [CHE08] (où

les auteurs explicitent leurs choix de positionnement de points d'annotation qui peuvent tout à fait être assimilés à des capteurs de mouvement, même s'ils sont disposés manuellement). Enfin, notons également que l'outil d'annotation Anvil permet depuis peu d'intégrer directement des données issues de capture de mouvement, de visualiser un squelette 3d qui rejoue ces enregistrements, et d'enregistrer ces données en XML (tout comme le sont les annotations). Ce qui nous ouvre des possibilités d'analyse de corpus de capture de mouvements de LSF avec cet outil, vu les outils d'analyse que nous avons développé et dont nous disposons aujourd'hui.

b. Les annotations

Nous avons décidé de concentrer notre étude sur l'annotation de la direction du regard et la configuration des mains. Au vu des dernières avancées dans le domaine du suivi de mouvement en LSF ([LEF10]), nous pourrions envisager d'ajouter une annotation automatique de l'emplacement des mains, ce qui apporterait des informations supplémentaires sans demander beaucoup d'effort de la part de l'annotateur. Cependant, comme nous l'avons signalé lors de nos choix d'annotation, l'expression initiale est très contrôlée, ce qui implique que la plupart du temps les emplacements ne suivent pas une règle « naturelle » mais plutôt « éditoriale ». Une analyse de la variation des emplacements permettraient au moins de déterminer si les emplacements relèvent effectivement d'un schéma prédéfini, ou si, malgré les choix initiaux, des variations peuvent tout de même être constatées : dans ce dernier cas, ce serait alors très intéressant de comparer les résultats avec d'autres études, comme celles de Grosvald ([GRO09], [GRO10]).

c. L'évaluation

Notre méthodologie d'évaluation nous semble pertinente, mais non adaptée au dispositif actuel. Cependant, nous pouvons proposer deux perspectives : travailler sur l'amélioration technique et linguistique des animations 3d, et envisager une évaluation objective, par comparaison avec d'autres résultats.

Pour la première solution, il est possible d'agir sur plusieurs niveaux. Nous présentons ces niveaux depuis la solution la plus rapide à mettre en œuvre, à celle qui va nécessiter plus de temps pour être réalisée. Dans un premier temps il faudrait corriger les animations 3d afin qu'elles soient acceptées par les sujets. Ensuite, il faudrait corriger la syntaxe des énoncés. Enfin, il faudrait revoir en amont les traductions proposées et s'orienter plutôt vers une diffusion d'information en LSF à l'aide de structures de grande iconicité, afin d'être mieux compris et par le plus grand nombre (sourds étrangers et entendants compris). Ces améliorations peuvent être réalisées à court terme, d'ici une année. Il serait d'ailleurs intéressant de ne pas corriger les animations déjà créées, mais d'enregistrer

un corpus de capture de mouvement avec les mêmes animations et de réaliser une nouvelle évaluation avec les données enregistrées en capture de mouvement.

Pour la seconde solution, l'évaluation objective, nous proposons de comparer les résultats que nous avons obtenus sur d'autres corpus où le contenu serait identique, mais les locuteurs différents. Cette seconde solution est peut-être plus rapide à mettre en œuvre que la première, mais elle ne permet pas de disposer au final d'animations 3d de qualité : ce sera une évaluation uniquement du modèle de coarticulation.

2. Utilisations potentielles de nos résultats

Nos résultats, s'ils sont confirmés lors d'une prochaine évaluation, pourraient selon nous être utiles à deux catégories d'étude : dans le domaine du traitement d'images, celles sur le suivi et la dynamique du mouvement, et dans le domaine de la génération automatique (qui consiste à générer de la langue à partir d'un modèle), celles sur l'animation de signeur virtuel par génération. Dans le premier cas, le modèle permettrait de préciser quelles sont les variations subies à cause de la coarticulation, ce qui aiderait à déterminer dans les systèmes de suivi du mouvement quand le signe débute et s'arrête (puisqu'on pourrait repérer les variations subies et qu'on disposerait de la forme initiale du signe). Dans le second cas, le modèle de coarticulation pourrait être couplé avec un modèle de gestion du confort pour l'animation d'un signeur virtuel, et permettrait de réaliser des animations encore plus réalistes que celles existantes (où pour la LSF, à notre connaissance, il n'est pas encore question de notions ergonomiques comme le confort, ou linguistiques comme la coarticulation pour l'animation des signeurs virtuels).

BIBLIOGRAPHIE

ABO06 ABOUDA L., et BAUDE O., 2006, « Constituer et exploiter un grand corpus oral : choix et enjeux théoriques. Le cas des ESLO », Corpus en Lettres et Sciences sociales : des documents numériques à l'interprétation, Actes du colloque international d'Albi, RASTIER, François, BALLABRIGA, Michel (dir.), Paris.

AFC07 Appel à communication, 2007, Séminaire AFCP Traitement automatique du langage parlé et langues peu dotées, Grenoble, 21 juin 2007.

AHM AHMED M., JAligner: Open source Java implementation of Smith-Waterman, (<http://jaligner.sourceforge.net>)

ALL89 ALLISON L., WALLACE C. S., et YEE C. N., 1990, « When is a String Like a String? », International Symposium on Artificial Intelligence and Mathematics (AIM), Ft. Lauderdale, Floride, du 3 au 5 Janvier.

ANS09 ANSUINI C., GRIGIS K., MASSECCESI S., et CASTELLIO U., 2009, « Breaking the flow of an action. », Exp Brain Res. Jan;192(2):287-92.

ARA08 ARAN, O., 2008, « Vision based sign language recognition : modeling and recognizing isolated signs with manual and non manual components », Thèse de doctorat, université de Bgazici, Turquie.

ATH10 ATHISTOS V., et coll., 2010, « Large Lexicon Project: American Sign Language Video Corpus and Sign Language Indexing/Retrieval Algorithms », 4th Workshop on the Representation and Processing of Sign Languages : Corpora and Sign Language Technologies, LREC, Malte.

BAA05 BAADER A. P., KAZENNIKOV B. O., WIESENDANGER M., 2005, « Coordination of bowing and fingering in violin playing », Cognitive Brain Research 23, 436–443.

BAI01 BAILLY G., présentation orale, « Synthèse audiovisuelle », 2004.

- BAL96** BALLIM A., 1996, Deliverable 2.5.3 Aligner v2.0.
- BAN00** BANGHAM J.A., COX S.J., LINCOLN M., MARSHALL I., TUTT M., WELLS, M., 2000, université de East Anglia, Norwich, dans Speech and Language Processing for Disabled and Elderly People (Ref. No. 2000/025), IEEE Seminar 2000, 4/1 - 4/5, London , UK.
- BAU06** BAUDE O., 2006, « Corpus oraux, Guide des bonnes pratiques 2006 », Paris et Orléans , CNRS éditions et PUO.
- BEB1825** BEBIAN R-A., 1825, « Mimographie ou essai d'écriture mimique, propre à régulariser le langage des sourds-muets », L. Colas, Paris.
- BEN74** BENGUEREL A.-P., 1974, « Nasal Airflow Patterns and Velar Coarticulation in French, Speech Communication Seminar », Stockholm, pp. 105-112.
- BEN81** BENZECRI J.-P. et coll., 1981, « Pratique de l'analyse des données », Linguistique et lexicologie, Paris, Dunod.
- BER06** BERGERON L.-F., 2006, « Typologie des systèmes écrits pour les langues signées », Revue des étudiants en linguistique du Québec, vol. 1, no 2.
- BEV03** BEVACQUA E. et PELACHAUD C., 2003, « Triphone-based Coarticulation Model », AVSP, du 5 au 7 Septembre.
- BHU06** BHUYAN M. K., GHOSH D., et BORA P. K., 2006, « Continuous Hand Gesture Segmentation and Co-articulation Detection. », ICVGIP.
- BIB75** BIBEAU G., 1975, « Introduction à la phonologie générative du français », Montréal, Paris, Bruxelles, Didier, (« Studia Phonetica », vol 9).
- BOM99** BOMMIER-PINCEMIN B., 1999, « Diffusion ciblée automatique d'informations : conception et mise en oeuvre d'une linguistique textuelle pour la caractérisation des destinataires et des documents », Thèse de doctorat, université Paris IV (Sorbonne).

BOU <http://corpusdelap parole.in2p3.fr/spip.php?article117>

BOU08 BOUTORA L., 2008, « Fondements historiques et implications théoriques d'une phonologie des langues des signes. Etude de la perception catégorielle des configurations manuelles en LSF et réflexion sur la transcription des langues des signes », Thèse de doctorat, université Paris 8.

BRO91 BROWN P., LAI J., MERCER R., 1991, « Aligning Sentences in Parallel Corpora », 29th Annual Meeting of the Association for Computational Linguistics, Berkeley, Californie.

BRU06 BRUGEILLE J.-L., 2006, « Une expérience d'utilisation de forme graphique dans la scolarité des enfants sourds : Méthodes de travail et premières observations », colloque Syntaxe, interprétation, lexicque des langues signées, Université Lille 3 Villeneuve d'Ascq, 1^{er} et 2 juin.

BRA96 BRAFFORT A., 1996, « Reconnaissance et compréhension de gestes — Application à la langue des signes », Thèse de doctorat, université Paris 11.

BRA08 BRAFFORT A., DALLE, P., 2008, « Sign language applications: preliminary modeling », dans : Universal Access in the Information Society, Springer-Verlag, Numéro spécial Emerging Technologies for Deaf Accessibility in the Information Society, Vol. Hors-série N. 4, p. 393-404.

BRO90 BROOKE N.M., 1990, « Computer Graphics Synthesis of Talking Faces », ESCA Workshop on Speech Synthesis, ESCA. Autrans.

BUI06 BUISINE S., ABRILIAN S., NIEWIADOMSKI R., MARTIN J.C., DEVILLERS L., et PELACHAUD C., 2006, « Perception d'émotions mélangées: Du corpus vidéo à l'agent expressif », WACA'06 second Workshop francophone sur les Agents Conversationnels Animés, pp. 83-91.

CAS05 CASSEL R., 2005, « Analyse du Mouvement Humain par un Système de Vision - Une approche globale pour l'analyse et la reconnaissance en temps réel de mouvements acrobatiques », Thèse de doctorat, université Paris Sud Orsay.

CHE01 CHEEK D. A., 2001, « The phonetics and phonology of handshape in American Sign Language », Thèse de doctorat, université de Texas à Austin, USA.

CHE08 CHETELAT-PELE E., BRAFFORT A., 2008, « Description des mouvements des sourcils pour la génération automatique », Actes de la conférence Traitement Automatique des Langues Naturelles (TALN), Traitement Automatique des Langues des Signes (TALS), Université d'Avignon, du 9 au 13 juin.

CHE10 CHETELAT E., 2010, « Les Gestes Non Manuels en langue des signes française ; Annotation, analyse et formalisation : Application aux mouvements des sourcils et aux clignements des yeux », Thèse de doctorat, université de Provence, Aix-en-Provence.

CHI06 CHITTARO L., BUTTUSSI F., et NADALUTTI D., 2006, « MAge-AniM: a System for Visual Modeling of Embodied Agent Animations and their Replay on Mobile Devices », AVI 2006: 8th International Conference on Advanced Visual Interfaces, ACM Press, New York.

COH93 COHEN M. M., et MASSARO D. W., 1993, « Modeling coarticulation in synthetic visual speech », dans N. M. Thalmann & D. Thalmann (Eds.) Models and Techniques in Computer Animation, Tokyo : Springer-Verlag.

COX02 COX S. J. et coll., 2002, « TESSA, a system to aid communication with deaf people », ASSETS 2002, Edinburgh, Scotland, pp. 205-212.

CRA08 CRASBORN O., ZWITSERLOOD, I., 2008, « The Corpus NGT: an online corpus for professionals and laymen », Construction and Exploitation of Sign Language Corpora, 3rd Workshop on the Representation and Processing of Sign Languages, Crasborn, Hanke, Efthimiou, Zwitserlood et Thoutenhoofd, eds. ELDA, Paris, pp 44-49.

COX02 COX S.J., LINCOLN M., TRYGGVASON J., NAKISA M., WELLS M., TUTT M. et ABBOTT S., 2002, « TESSA, a system to aid communication with deaf people », ASSETS Edinburgh, Ecosse, pp. 205-212.

CRA08 CRASBORN O., SLOETJES H., 2008, « Enhanced ELAN functionality for sign language Corpora », Construction and Exploitation of Sign Language Corpora, 3rd Workshop on the Representation and Processing of Sign Languages, Crasborn, Hanke, Efthimiou, Zwitserlood et Thoutenhoofd, eds. ELDA, Paris.

CRA08 CRASBORN O., ZWITSERLOOD I., 2008, « The Corpus NGT: an online corpus for professionals and laymen », Construction and Exploitation of Sign Language Corpora, 3rd Workshop on the Representation and Processing of Sign Languages, Crasborn, Hanke, Efthimiou, Zwitserlood & Thoutenhoofd, eds. ELDA, Paris. pp 44-49.

CUX96 CUXAC C., 1996, « Fonctions de l'iconicité », dans B. Virole (dir.), Psychologie de la Surdit , De Boeck Universit , Paris-Bruxelles, pp. 189-198.

CUX00 CUXAC C., 2000, « La langue des signes fran aise. Les voies de l'iconicit  », Faits de Langues 15/16, Paris,  ditions Ophrys, 391 pages.

CUX10 CUXAC C., 2010, Langage et soci t  2010/1 (n  131),  mergence, norme et variation dans les langues des signes : vers une red finition notionnelle.

DAH04 DAHL S., 2004, « Playing the accent - comparing striking velocity and timing in an ostinato rhythm performed by four drummers », Acta Acustica dans Acustica Vol. 90(4), 762-776.

DAL06 DALLE-NAZEBI S., 2006, « Chercheurs, Sourds et Langue des Signes. Le travail d'un objet et de rep res linguistiques (en France, du 17 me au 21 me si cle) », Th se de doctorat, universit  Toulouse 2, Toulouse.

DAN73 DANILOFF R., et HAMMARBERG R., 1973, « On defining coarticulation », Journal of Phonetics 1, 239-48.

DEK01 DEKETELAERE S., DUTOIT T., et DEROO O., 2001, « Speech Processing for Communications : what's new? », Revue HF, Mars 2001, pp. 5-24.

DEN06 DENG Z., NEUMANN U., LEWIS J.P., KIM T.-Y., BULUT M., et SHRIKANTH N., 2006, « Expressive Facial Animation Synthesis by Learning Speech Coarticulation and Expression Spaces », Visualization and Computer Graphics, IEEE Transactions, Volume: 12 Issue: 6, 1523-1534, Nov.-Dec..

DEV09 DEVOS N., 2009, « Evaluation de signeurs virtuels sur Internet », m moire de Master, universit  Paris 8,.

DRE10 DREUW P. et coll., 2010, « SignSpeak - Understanding, Recognition, and Translation of Sign Languages », 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies, La Valette, Malte.

DUB99 DUBUISSON C., LELIÈVRE L. et MILLER C., 1999, « Grammaire descriptive de la langue des signes québécoise : Tome 1, Le comportement manuel et le comportement »non manuel. Tome 2, Le lexique. édition revue et augmentée ». Montréal : UQAM.

DUG79 DUGAST D., 1979, « Vocabulaire et discours : fragments de lexicologie quantitative : essai de lexicométrie organisationnelle », Genève-Paris, Slatkine-Champion.

EFT07 EFTHIMIOU E., et FOTINEA, S.-E., 2007, « GSLC: Creation and annotation of a Greek Sign Language Corpus for HCI », Lecture Notes in Computer Science (LNCS), C. Stephanidis (Ed.): Universal Access in HCI, vol 1, pp. 657–666.

EFT09 EFTHIMIOU E. et coll., 2009, « Sign Language Recognition, Generation, and Modelling: A Research Effort with Applications in Deaf Communication », présentation à UAHCI/HCI, San Diego, CA, 24 juillet.

ELL04 ELLIOTT R., GLAUERT J., KENNAWAY R., MARSHALL I., et SAFR E., 2004, « Linguistic modelling and languageprocessing technologies for Avatar-based sign language presentation », Universal Access in the Information Society (UAIS), Special issue 6/4. Springer.

<http://www.visicast.cmp.uea.ac.uk/eSIGN/index.html>

ENC08 ENCREVÉ F., 2008, « Sourds et société française au XIXe siècle (1830-1905) », Thèse de doctorat, université de Paris 8 - Vincennes Saint-Denis.

ENG97 ENGEL K. C., FLANDERS M., et SOECHTING J. F., 1997, « Anticipatory and sequential motor control in piano playing », *Experimental Brain Research*, 113:189-199.

FIL08 FILHOL M., 2008, « Modèle descriptif des signes pour un traitement automatique des langues des signes », Thèse de doctorat, Université Paris-11 (Paris sud), Orsay.

FLE06 FLECHSIG I, 2006, « An investigation of coarticulation in rare and frequent syllables with regard to (the existence of) a mental syllabary », Thèse de doctorat, Institut für Maschinelle Sprachverarbeitung, Stuttgart, Allemagne

GAL91 GALE W., et CHURCH K., 1991, « A Program for Aligning Sentences in Bilingual Corpora », 29th Annual Meeting of the Association for Computational Linguistics, Berkeley, Californie.

GAL93 GALE W., et CHURCH K., 1993, « A Program for Aligning Sentences in Bilingual Corpora », Computational Linguistics, 19(2): 75-102.

GAR06 GARCIA, B., 2006, « The methodological, linguistic and semiological Bases for the elaboration of a written form of LSF (French Sign Language) », Second Workshop on the Representation and Processing of Dign Languages : Lexicographic Matters and Didactic Scenarios, 5th International Conference on Language Resources and Evaluation LREC 28 may 2006, Genoa, Italy.

GAR10 GARCIA B., et PERINI M., 2010, « Normes en jeu et jeu des normes dans les deux langues en présence chez les sourds locuteurs de la Langue des Signes Française », dans Langage et Société, vol1 / n° 131, Ed Maison des sciences de l'homme.

GIB01 GIBET S., LEBOURQUE T., et MARTEAU P.F., 2001, « High level Specification and Animation of Communicative Gestures », Journal of Visual Languages and Computing, 12, 657-687.

GIB08a GIBET S., et MARTEAU P.F., 2008, « Approximation of Curvature and Velocity using Adaptive Sampling Representations - Application to Hand Gesture Analysis and Segmentation », Actes du Gesture Workshop 2007, Gesture-Based Human-Computer Interaction and Simulation, Lecture Notes in Computer Science (Revised Selected papers), Vol. 5085, Berlin:Springer.

GIB08b GIBET S., HELOIR A., 2008, « Formalisme de description des gestes de la langue des signes française pour la génération du mouvement de signeurs virtuels », TAL, Special issue on Modeling and processing of sign languages, 3(48), pp. 111-145.

GIL98 GILLOT D., 1998, « Le droit des sourds », rapport remis à Lionel Jospin (voir, pour l'ensemble de la documentation relative à ce rapport, le site de l'Association nationale des parents d'enfants sourds, l'ANPES : <http://membres.lycos.fr/anpes/>)

GOD09 GODOY R. I., JENSENIUS A. R., et NYMOEN K., 2009, « Chunking by coarticulation in music-related gestures ».

GOS08 GOSY M., BEKE A., 2008, « Anticipatory Coarticulation in a Specific Context », ISSP 08, Strasbourg, France, du 8 au 12 décembre 2008.

GOV08 GOVOKHINA O., 2008, « Modèles de génération de trajectoires pour l'animation de visages parlants », Thèse de doctorat, Institut nationale polytechnique de Grenoble.

GRO09 GROSVOLD M. A., 2009, « Long-Distance Coarticulation: A Production and Perception Study of English and American Sign Language », Thèse de doctorat, université de Californie, Davis.

GRO10 GROSVLAD M. A., 2010, « Long-Distance Coarticulation in Spoken and Signed Language: An Overview », université de Californie, Davis, Vol. 4.

GUI04 GUITTENY P., LEGOUIS P., et VERLAINE L., 2004, « La langue des signes », Centre d'Information sur la Surdit  d'Aquitaine.

HAN08 HANKE T., STORZ J., 2008, « iLex – A database tool integrating sign language corpus linguistics and sign language lexicography », Construction and Exploitation of Sign Language Corpora, 3rd Workshop on the Representation and Processing of Sign Languages, Crasborn, Hanke, Efthimiou, Zwitserlood & Thoutenhoofd, eds. ELDA, Paris.

HAR99 HARDCASTLE W. J., HEWLETT N., 1999, « Coarticulation. Theory, Data and Techniques », Edited by William J. Hardcastle , Nigel Hewlett, Cambridge Studies in Speech Science and Communication.

HEL06 HELOIR A., COURTY N., GIBET S., MULTON F., 2006, « Temporal alignment of communicative gesture sequences », Computer Animation and Virtual Worlds (selected best papers from CASA'06), 17:347-357, Juillet.

HOL06 HOLT G. T., HENDRIKS P., et ANDRINGA T., 2006, « Why Don't You See What I Mean? Prospects and Limitations of Current Automatic Sign Recognition Research », *Sign Language Studies*, 6(4).

HUE06 HUENERFAUTH M., 2006, « Generating American Sign Language Classifier Predicates For English-To-ASL Machine Translation », Thèse de doctorat, université de Pennsylvanie.

HUE07 HUENERFAUTH M., ZHOU L., GU E. et ALLBECK J., 2007, « Evaluating American Sign Language Generation Through the Participation of Native ASL Signers » 9th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS-2007. Tempe, Arizona, USA.

JEN10 JENSENIUS A. R. et coll., 2010 « Sensing Music-Related Actions », fourMs, University of Oslo – Rapport de laboratoire.

JER03 JERDE T.E., SOECHTING J.F., FLANDERS M., 2003, Coarticulation in fluent finger spelling », *Journal of Neuroscience*, 23:2383–2393.

JOH JOHNSON R. E., LIDDELL, K. S., en préparation, « Sign language phonetics: architecture and description ».

JOH07 JOHNSTON T., SCHEMBRI A. C., 2007, « Australian Sign Language (Auslan): An Introduction to Sign Language Linguistics », Cambridge University Press.

JOH08a JOHNSTON T., 2008, « Corpus linguistics & signed languages: no lemmata, no corpus », *Construction and Exploitation of Sign Language Corpora. 3rd Workshop on the Representation and Processing of Sign Languages*, ELDA, Paris. Pp . 82-87.

JOH08b JOHNSTON T., 2008, « Creating a corpus of Auslan within an Australian National Corpus », *HSCNet Workshop Proceedings: Designing the Australian National Corpus*, UNSW, Sydney, 4-5 December 2008

JOU96 JOUISSON P., 1996, « Ecrits sur la langue des signes française », sous la dir. de B. Garcia, Editions L' Harmattan, 250 pages.

- KAY93** KAY M. et ROSCHEISEN M., 1993, « Text-Translation Alignment. Computational Linguistics », Vol. 19, No. 1, pp.121-142.
- KEL09** KELLY D. et coll., 2009, « A framework for continuous multimodal sign language recognition », actes de International conference on Multimodal interfaces, pp. 351-358, Cambridge, Massachusetts, USA.
- KEN04** KENDON A., 2004, « Gesture: Visible Action as Utterance », Cambridge University Press.
- KIT98** KITA S. et coll, 1998, « Movement phases in signs and cospeech gestures, and their transcription by human coders », Lecture Notes in Computer Science, Gesture and Sign Language in Human- Computer Interaction: International Gesture Workshop, Bielefeld, Germany, September 1997. Springer Verlag: Heidelberg. Volume 1371
- KIP04** KIPP M., 2004, « Gesture Generation by Imitation - From Human Behavior to Computer Character Animation », Thèse de doctorat, Boca Raton, Florida.
- KUH99** KÜHNERT B. et NOLAN F., 1999, « The origins of coarticulation », dans W. Hardcastle & N. Hewlett (eds.), Coarticulation: Theory, Data and Techniques. Cambridge: CUP, 7-30.
- LEB98** LEBART L., et SALEM A., 1988, « Analyse statistique des données textuelles », Mathématiques et sciences humaines, vol 105.
- LEF08** LEFEBVRE-ALBARET F., DALLE P., 2009, « Une approche de segmentation de la LSF », dans : Traitement Automatique des Langues Naturelles (TALN 2008), Avignon, du 9 au 13 juin, Association pour le Traitement Automatique des Langues (ATALA).
- LEF10** LEFEBVRE-ALBARET, F., 2010, « Traitement automatique de vidéos en LSF. Modélisation et exploitation des contraintes phonologiques du mouvement », Thèse de doctorat, Université Paul Sabatier, Toulouse.

LEN05 LENSEIGNE B., DALLE P., 2005, « Using signing space as a representation for sign language processing », Gesture Workshop, Berder, France, du 18-20 mai, dans LNAI n°3881. Springer, Berlin, 25-36.

LID89 LIDDELL S. K., JOHNSON R. E., 1989, « American Sign Language : the phonological base », Sign Language Studies, vol. 64. Gallaudet University Press.

LID03 LIDDELL S., 2003, « Grammar, Gestures, and Meaning in American Sign Language », Cambridge University Press, pp. 384.

LOL08 LOLIVE D., 2008, « Transformation de l'intonation : application à la synthèse de la parole et à la transformation de voix », Thèse de doctorat, IRISA – Université de Rennes 1.

LSC02 Rédaction collective, 2002, « Projet LS-COLIN. Rapport de fin de recherche », Ministère de la recherche, Programme Cognitique 2000.

MA08 MA L., 2008, « La coarticulation en français et en chinois : Étude expérimentale et modélisation », Thèse de doctorat, université de Provence Aix-Marseille I

MAC00 MACLAUGHLIN D., NEIDLE C., BAHAN B., et LEE R. G., 2000, « Morphological Inflections and Syntactic Representations of Person and Number in ASL », dans M. Blondel and L. Tuller (eds), Recherches linguistiques de Vincennes 29 : Langage et surdit , pp. 73-100.

MAN08 MANNELL R., 2008, « Coarticulation and Assimilation », Macquarie University, disponible ici : <http://clas.mq.edu.au/phonetics/phonetics/coarticulation/index.html>

MAR49 MARTINET A., 1949, « La double articulation linguistique ». TCLC, 5, 30–37.

MAU03 MAUK C. E., 2003, « Undershoot in two modalities: Evidence from fast speech and fast signing », Thèse de doctorat, universit  du Texas   Austin.

MAY05 MAYAFFRE D., 2005, « R le et place des corpus en linguistique : r flexions introductives », Toulouse,  dit  par P. VERGELY, 5-18

- MEL02** MELLET S., 2002, « Corpus et recherches linguistiques », Corpus [En ligne], vol 1.
- MLO10** BEN MLOUKA M., LEFEBVRE-ALBARET F., DALLE J., DALLE P., 2010, « Annotation automatique d'une vidéo en LSF à partir de données de capture de mouvement », Actes de la conférence Traitement Automatique des Langues Naturelles (TALN), Traitement Automatique des Langues des Signes (TALS), Montréal, 23 juillet.
- MOT06** MOTTEZ B., 2006, « Les Sourds existent-ils ? Textes réunis et présentés par Andrea Benvenuto », Paris, L'Harmattan.
- NEI00** NEIDLE C., 2000, « SignStream(TM): A Database Tool for Research on Visual-Gestural Language », American Sign Language Linguistic Research Project, Report No. 10, Université de Boston, USA.
- NGU01** NGUYEN N., 2001, « Rôle de la coarticulation dans la reconnaissance des mots », dans L'année psychologique, vol. 101, n°1, pp. 125-154.
- NYS** NYST V.A.S., « Sign languages in West Africa », dans Brentari, D., Sign Languages, Cambridge University Press
- ODO08** O'DONNELL M., 2008, « Principles of corpus construction », présentation au Corpus Linguistics Summer Institute, Liverpool, du 30 juin au 3 juillet.
- OHM67** ÖHMAN S., 1967, « Numerical model of coarticulation », Journal of the Acoustical Society of America, 41, pp. 310-320.
- OJA09** OJALA S., SALAKOSKI T., et AALTONEN O., 2009, « Coarticulation in sign and speech », actes de NODALIDA 2009 workshop Multimodal Communication, from Human Behaviour to Computational Models, Costanza Navarretta, Patrizia Paggio, Jens Allwood, Elisabeth Alsén and Yasuhiro Katagiri (Eds). NEALT Proceedings Series, Vol. 6 (2009), 21-24.
- PAN99** PANDZIC I., OSTERMANN J., et MILLEN D., 1999, « User evaluation: Synthetic talking faces for interactive services », The visual computer, Vol. 15, Nr. 7, pp. 330 -340.

PEL91 PELACHAUD C., 1991, « Communication and Coarticulation in Facial Animation », Thèse de doctorat, n° MS-CIS-91-77, Université de Pennsylvanie.

PER04 PERRIER P., PAYAN Y., et MARRET R., 2004, « Modéliser le physique pour comprendre le contrôle : le cas de l'anticipation en production de parole », dans L'anticipation à l'horizon du présent (R. Sock & B. Vaxelaire, eds). Sprimont, Belgique : Pierre Mardaga, pages 159-177.

PRI89 PRILLWITZ S. et coll, 1989, « HamNoSys. Version 2.0, Hamburg Notation System for Sign Languages. An introductory guide », International Studies on Sign Language and Communication of the Deaf; 5, Hamburg Signum - 46 p.

REN06 DI RENZO A., LAMANO L., LUCIOLI T., PENNACCHI B., PONZO L., 2006, « Italian Sign Language: Can we write it and transcribe it with Sign Writing ? », dans C. Vettori (ed.), Actes du 2^{ème} atelier sur la représentation et le traitement des langues des signes (Second Workshop on the Representation and Processing of Sign Languages), LREC (International Conference on Language Resources and Evaluation), Gènes, 28 mai.

RIC09 RICCO S. et TOMASI C., 2009, « Fingerspelling recognition through classification of letter-to-letter transitions », 9ème Asian Conference on Computer Vision (ACCV 2009), Xi'an, Chine, Septembre 2009.

ROB07 ROBERT V., BONNEAU A., WROBEL B. et LAPRIE Y., 2007, « Prédiction phonétique de la coarticulation labiale », dans Perturbations et réajustements, Rudolph Stock, (éd.) Editions de l'Université Marc Bloch de Strasbourg.

ROB08 ROBERT V., FELDMAR J. et LAPRIE Y., 2008, « Comparison between two predicting methods of labial coarticulation », The eighth International Seminar on Speech Production - ISSP'08, Strasbourg : France

<http://hal.inria.fr/inria-00336382/fr/>

ROH06 ROHLFING K. et coll., 2006, « Comparison of multimodal annotation tools ». Tech. rep. Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion.

<http://www.gespraechsforschung-ozs.de/heft2006/tb-rohlfing.pdf>.

- ROS91** ROSENBAUM D. A., 1991, « Human Motor Control », San Diego: Academic Press, Inc.
- RUT04** RUTTKAY Z. et PELACHAUD C., 2004, « From Brows to Trust. Evaluating Embodied Conversational Agents », Preface, Kluwer Academic Publishers, Norwell, MA.
- SAL03** SALLANDRE M.-A., 2003, « Les unités du discours en Langue des Signes Française. Tentative de catégorisation dans le cadre d'une grammaire de l'iconicité », Thèse de doctorat, université Paris VIII, Saint-Denis.
- SCH** <http://www.cse.ohio-state.edu/~parent/classes/888/Character/facialAnimation.pdf>
- SCH03** SCHEMBRI A., 2003, « Rethinking "classifiers" in signed languages », dans Perspectives on classifier constructions in sign languages, Emmorey, K. (Ed.), pp.3-34, Mahwah, NJ: Lawrence Erlbaum Associates.
- SCH08** SCHEMBRI A., 2008, « British Sign Language Corpus Project: Open access archives and the Observer's Paradox », dans les actes de Sixth International Conference on Language Resources and Evaluation, du 26 mai au 1^{er} juin 2008, Marrakech, Morocco. European Language Resources Association.
- SEG03** SEGOUAT J., 2003, «Étude et modélisation de certains aspects de la structure des énoncés en Langue des Signes Française», mémoire de DEA science-cognitives, université de Paris-Sud, Orsay.
- SEG06** SEGOUAT J., BRAFFORT A. et MARTIN E., «Sign Language corpus analysis: Synchronisation of linguistic annotation and numerical data», dans les actes de 2nd Workshop on the Representation and Processing of Sign Languages. 5th International Conference on Language Resources and Evaluation, LREC 2006, Gènes, Italie.
- SEG08** SEGOUAT J. et BRAFFORT A., 2010, « Toward Modeling Sign Language Coarticulation », Springer LNCS/LNAI, volume "Gesture in Embodied Communication and Human-Computer Interaction".

SIF05 SIFAKIS E., NEVEROV I., et FEDKIW R., 2005, « Automatic determination of facial muscle activations from sparse motion capture marker data », ACM Transactions on Graphics 24, 3, 417–425.

SIG <http://www.vcom3d.com/signsmith.php>

SIN04 SINCLAIR J., 2004, « Trust the text », Language, Corpus and Discourse, Londres, Routledge.

SLO03 SLOBIN D. I. et coll., 2003, « Acognitive/functional perspective on the acquisition of "classifiers" », dans K.Emmorey (Ed.), Perspectives on classifier constructions in sign languages (pp.271-296). Mahwah, NJ: Lawrence Erlbaum Associates.

SOE87 SOECHTING J. F. et TERZUOLO C. A., 1987, « Organization of arm movements. Motion is segmented », Neuroscience 23:39-51.

SOS04 SOSNIK R., HAUPTMANN B., KARNI A., FLASH T., 2004, « When practice leads to coarticulation: the evolution of geometrically defined movement primitives », Exp Brain Res 156:422–438.

STO60 STOKOE W., 1960, « Sign Language Structure: An Outline of the Visual Communication System of the American Deaf », Studies in Linguistics, New York.

SUT SUTTON V. (1974-2008). : <http://www.signwriting.org>.

SUT75 SUTTON V., 1975, « SignWriting: Read, write, type Sign Languages », <http://www.signwriting.org/>

THO07 THOUTENHOOFD E. et CRASBORN O., 2007, « Sign language corpus creation and interface concerns », The Emergence of Corpus Sign Language Linguistics, 40èmes Rencontres Annuelles de l'Association de Linguistique Appliquée (BAAL), du 6 au 8 Septembre, Edimbourg.

TYR10 TYRONE M. E. et MAUK C. E., 2010, « Sign lowering and phonetic reduction in American Sign Language », Journal of Phonetics, Volume 38, Issue 2, April 2010, Pages 317-328.

VER00a VERONIS, J. (Ed.), 2000, « Parallel Text Processing: Alignment and use of translation corpora ». Dordrecht: Kluwer Academic Publishers.

VER00b VERONIS J., LANGLAIS P., 2000, « Evaluation of parallel text alignment systems – The ARCADE project », dans VÉRONIS J. (ed.) Parallel Text Processing (Dordrecht: Kluwer Academic Publishers), 369-388.

VER01 VERLINDEN M., TIJSELING C., et FROWEIN H., 2001, "A Signing Avatar on the WWW", International Gesture Workshop, City University, London.

<http://www.visicast.cmp.uea.ac.uk/>

VOG99 VOGLER C. et METAXAS D., 1999, « Toward scalability in ASL recognition: Breaking down signs into phonemes. », dans les actes de Gesture Workshop (Gif-sur-Yvette, France), volume 1739, pages 211-224. Springer.

VOG03 VOGLER C., 2003, American sign language recognition : reducing the complexity of the task with phoneme-based modeling and parallel hidden Markov models », Thèse de doctorat, université de Pennsylvanie.

WIE06 WIESENDANGER M., BAADER A., et KAZENNIKOV O., 2006, « Fingering and Bowing in Violinists: A Motor Control Approach », dans Altenmüller, E., Wiesendanger, M., and Kesselring, J., editors, Music, Motor Control and the Brain, pages 109-123. Oxford: Oxford University Press.

YAN06 YANG R. et SARKAR S., 2006, « Detecting Coarticulation in Sign Language using Conditional Random Fields », ICPR (2)'2006. pp.108-112.

YAN10 YANG R., SARKAR S., et LOEDING B. L., 2010, « Handling Movement Epenthesis and Hand Segmentation Ambiguities in Continuous Sign Language Recognition Using Nested Dynamic Programming », IEEE Trans. Pattern Anal. Mach. Intell. 32(3): 462-477.

ANNEXES

I. Énoncés et segments d'énoncés de notre corpus

Énoncés complets

Nom de séquence	Énoncé
AS01-1	Voie 1, éloignez vous de la bordure du quai. Attention au passage d'un train. Voie 1, éloignez vous de la bordure du quai.
AS01-2	Voie 6, éloignez vous de la bordure du quai. Attention au passage d'un train. Voie 6, éloignez vous de la bordure du quai.
AS01-3	Voie 10, éloignez vous de la bordure du quai. Attention au passage d'un train. Voie 10, éloignez vous de la bordure du quai.
AS01-4	Voie 17, éloignez vous de la bordure du quai. Attention au passage d'un train. Voie 17, éloignez vous de la bordure du quai.
AS01-5	Voie 21, éloignez vous de la bordure du quai. Attention au passage d'un train. Voie 21, éloignez vous de la bordure du quai.
AS01-6	Voie 27, éloignez vous de la bordure du quai. Attention au passage d'un train. Voie 27, éloignez vous de la bordure du quai.
AS01-7	Voie 71, éloignez vous de la bordure du quai. Attention au passage d'un train. Voie 71, éloignez vous de la bordure du quai.
AS01-8	Voie 76, éloignez vous de la bordure du quai. Attention au passage d'un train. Voie 76, éloignez vous de la bordure du quai.
AS02	Voie A, éloignez vous de la bordure du quai. Attention à la manoeuvre du train. Voie A, éloignez vous de la bordure du quai.
AS03	Voie A, éloignez vous de la bordure du quai. Attention à la manoeuvre du train. Voie A, attendez l'arrêt complet avant de monter.

AS04	Mesdames, Messieurs, nous vous rappelons qu'il est interdit de fumer dans l'enceinte de la gare.
AS05	Mesdames, Messieurs, nous vous rappelons qu'il est interdit de fumer dans l'enceinte de la gare et sur la totalité des quais.
AS06	Mesdames, Messieurs, nous vous rappelons qu'un arrêté préfectoral interdit de fumer sur les quais de la gare. Merci.
AS07	Afin d'améliorer la tranquillité de votre voyage, nous vous demandons de ne pas laisser vos bagages sans surveillance et nous vous prions de nous signaler tout colis ou bagage qui vous paraîtrait abandonné.
AS13	Votre attention s'il vous plaît. Demain, en raison de mouvements sociaux, la circulation des trains sera perturbée. Un affichage en gare vous indiquera les trains qui circuleront. Cette information sera également disponible sur le site Internet SNCF.com
AS14	Votre attention s'il vous plaît. Demain, en raison de mouvements sociaux, la circulation des trains sera perturbée. Un affichage en gare vous indique les trains qui circuleront. Cette information est également disponible sur le site Internet SNCF.com
AS15	Votre attention s'il vous plaît. En raison de mouvements sociaux, la circulation des trains est perturbée. Un affichage en gare vous indique les trains qui circulent. Cette information est également disponible sur le site Internet SNCF.com
AS16-1	Votre attention s'il vous plaît, en raison d'un incident technique, la circulation des trains au départ et à l'arrivée Paris Gare de l'Est est interrompue. Nous donnerons des précisions dès que possible
AS16-2	Votre attention s'il vous plaît, en raison d'une panne, la circulation des trains au départ et à l'arrivée Paris Gare de l'Est est interrompue. Nous donnerons des précisions dès que possible.
AS16-3	Votre attention s'il vous plaît, en raison de travaux sur les voies, la circulation des trains au départ et à l'arrivée Paris Gare de l'Est est interrompue. Nous donnerons des précisions dès que possible
AS16-4	Votre attention s'il vous plaît, en raison d'un accident de personne, la circulation des trains au départ et à l'arrivée Paris Gare de l'Est est interrompue. Nous donnerons des précisions dès que possible.
AS16-5	Votre attention s'il vous plaît, afin d'assurer les correspondances, la circulation des trains au départ et à l'arrivée Paris Gare de l'Est est interrompue. Nous donnerons des précisions dès

	que possible.
AS16-6	Votre attention s'il vous plaît, en raison de mouvements sociaux, la circulation des trains au départ et à l'arrivée Paris Gare de l'Est est interrompue. Nous donnerons des précisions dès que possible.
AS16-7	Votre attention s'il vous plaît, en raison d'un incident affectant la voie, la circulation des trains au départ et à l'arrivée Paris Gare de l'Est est interrompue. Nous donnerons des précisions dès que possible.
AS16-8	Votre attention s'il vous plaît, en raison de la présence de manifestants sur les voies, la circulation des trains au départ et à l'arrivée Paris Gare de l'Est est interrompue. Nous donnerons des précisions dès que possible.
AS16-9	Votre attention s'il vous plaît, en raison de la présence d'un bagage non identifié, la circulation des trains au départ et à l'arrivée Paris Gare de l'Est est interrompue. Nous donnerons des précisions dès que possible.
D0Y-1	Voie B, le TGV 8335 à destination de Reims va partir. Prenez garde à la fermeture des portes. Attention au départ.
D0Y-2	Voie C, le TGV 7421 à destination de Strasbourg va partir. Prenez garde à la fermeture des portes. Attention au départ.
D0Y-3	Voie B, le TGV 7421 à destination de Metz va partir. Prenez garde à la fermeture des portes. Attention au départ.
DT02-1	Voie A, les TGV numéro 1233 et 4078, à destination de Reims et Strasbourg vont partir. Prenez garde à la fermeture des portes. Attention au départ.
DT02-2	Voie A, les TGV numéro 7001 et 3001, à destination de Colmar et Mulhouse vont partir. Prenez garde à la fermeture des portes. Attention au départ.
DT02-3	Voie A, les TGV numéro 2012 et 1998, à destination de Colmar et Mulhouse vont partir. Prenez garde à la fermeture des portes. Attention au départ.
IT12A/B-1	Voie J, éloignez-vous de la bordure du quai, les TGV en provenance de Strasbourg et à destination de Reims et Paris Est entrent en gare.
IT12A/B-2	Voie P, éloignez-vous de la bordure du quai, les TGV en provenance de Colmar et à destination de Reims et Paris Est entrent en gare.
IT12A/B-3	Voie X, éloignez-vous de la bordure du quai, les TGV en provenance de Strasbourg et à

	destination de Reims et Paris Est entrent en gare.
RT20A-1	Votre attention sil vous plait. Voie 10, en raison d'un incident technique, les TGV numéro 1234 et 4567 en provenance de Strasbourg et Reims arriveront à 13h30.
RT20A-2	Votre attention sil vous plait. Voie 22, en raison de travaux sur les voies, les TGV numéro 1234 et 4567 en provenance de Strasbourg et Reims arriveront à 12h57.
RT20A-3	Votre attention sil vous plait. Voie 66, en raison d'une panne, les TGV numéro 1234 et 4567 en provenance de Strasbourg et Reims arriveront à 00h15.
RT20B-1	Votre attention sil vous plait. Voie 10, en raison d'un incident technique, les TGV numéro 1234 et 4567 en provenance de Strasbourg et Reims arriveront à une heure que nous ne pouvons évaluer pour le moment. Nous vous renseignerons dès que possible.
RT20B-2	Votre attention sil vous plait. Voie 22, en raison de travaux sur les voies, les TGV numéro 1234 et 4567 en provenance de Strasbourg et Reims arriveront à une heure que nous ne pouvons évaluer pour le moment. Nous vous renseignerons dès que possible.
RT20B-3	Votre attention sil vous plait. Voie 66, en raison d'une panne, les TGV numéro 1234 et 4567 en provenance de Strasbourg et Reims arriveront à une heure que nous ne pouvons évaluer pour le moment. Nous vous renseignerons dès que possible.

Énoncés partiels

Nom de séquence	Énoncé	Prise
E1	Voie A	
E2	Voie B	
E3	Voie C	
E4	Voie 1	
E5	Voie 6	
E6	Voie 10	
E7	Voie 17	
E8	Voie 21	
E9	Voie 27	

E10	Voie 71	
E11	Voie 76	
E12	éloignez-vous de la bordure du quai	
E13	Attention au passage d'un train	
E14	Attention à la manoeuvre du train	
E15	attendez l'arrêt complet avant de monter	
E16	Mesdames, Messieurs	
E17	Votre attention s'il vous plaît	
E18	Un affichage en gare vous indiquera les trains qui circuleront	
E19	Un affichage en gare vous indique les trains qui circuleront	
E20	Un affichage en gare vous indique les trains qui circulent	
E21	Prenez garde à la fermeture des portes	
E22	Attention au départ	
E23	En raison d'un incident technique	
E24	En raison d'une panne	
E25	En raison de travaux sur les voies	
E26	En raison d'un accident de personne	
E27	Afin d'assurer les correspondances	
E28	En raison de mouvements sociaux	
E29	En raison d'un incident affectant la voie	
E30	En raison de la présence de manifestants sur les voies	
E31	En raison de la présence d'un bagage non identifié	

Lettres	A → Z	
Lettres doubles	aBBaye, aCCent, crEEr, oFFiciel, eLLe, soMMe, eNNui, zOO, aPPorter, aRReter, aSSurer, aTTendre, suGGestion	

Nombres	0 → 20 , 21, 22, 26, 30 , 31, 33, 36, 40 , 41, 44, 46, 50 , 51, 55, 56, 60 , 61, 66, 70 , 71, 76, 77, 80 , 81, 86, 88, 90 , 91, 96, 99, 100 , 101, 102, 110, 120, 160, 200 , 201, 202, 210, 220, 260, 300 , 400 , 500 , 600 , 601, 602, 610, 620, 660, 670, 700 , 701, 702, 710, 720, 760, 770, 800 , 900 , 1000 , 1001, 1002, 1010, 1020, 1060, 1100, 1101, 1110, 1160, 1200, 1600, 2000 , 2001, 2002, 2010, 2020, 2060, 2100, 2101, 2110, 2160, 2200, 2600, 3000 , 4000 , 5000 , 6000 , 7000 , 8000 , 9000 , 10000, 60000
Heures	1h → 10h, 11h (x3), midi, minuit, 12h → 23h (x2), matin, soir
Jours	Lundi, Mardi (x2), Mercredi (x2), Jeudi, Vendredi, Samedi, Dimanche
Mois	Janvier, Février, Mars, Avril (x2), Mai, Juin, Juillet (x2), Aout, Septembre, Octobre (x2), Novembre (x2), Décembre (x2)
Villes	Paris Est, Nancy, Epinal, Lorraine, Thionville (x2), Metz, Bar le Duc, Champagne Ardennes, Reims, Sedan, Mulhouse, Colmar, Strasbourg

II. Correspondances valeurs / capture d'écrans / caractère ASCII

Tableaux de correspondance entre la valeur numérique des annotations (« Valeurs »), la capture d'écran pour l'annotation visuelle (« Visuel »), et le code ASCII correspondant.

Direction du regard		
Valeurs	Visuel	ASCII
md	n/a	!
mD	n/a	"
mDmd	n/a	#
face	n/a	\$
1		%
2		&
7		'
8		(
9	)
10		*
11		+
3		,
4		-

5		.
6		/
autre	n/a	0

Ouverture des yeux		
Anciennes valeurs	Visuel	ASCII
ouvert	n/a	!
ferme	n/a	"
moitie	n/a	#

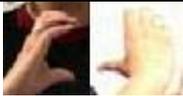
Configurations des mains		
Anciennes valeurs	Visuel	ASCII
repos_1		!
repos_3		"
repos_4		#
repos_5		\$
repos_6		%
1		&
2_2		'
2_3		(
3_4	)
4_4_1		*
4_4_2		+

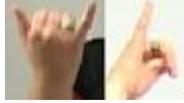
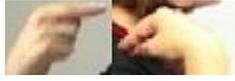
5		,
5_2		-
6		.
7		/
8		0
9		1
10		2
11		3
12		4
13		5
14		6
15		7

15_2		8
16		9
17		:
18		;
19_2		<
20		=
21		>
22		?
23		@
23_2		A
24		B
25_2		C

26		D
27		E
28		F
28_2		G
29		H
30		I
31		J
32		K
33		L
34		M
35		N
36		O

37		P
38		Q
39_1		R
39_2		S
39_2_1		T
39_2_2		U
39_3		V
50		W
51		X
52		Y
53		Z
54		[
55		\

56		j
57		^
58		-
59		\'
60		a
61		b
62		c
63		d
64		e
65		f
66		g
67		h
68		i
70		j

71		k
72		l
73		m
74		n
75		o
76		p
77		q
78		r
79		s
80		t
81		u
82		v

III. Variations des instances d'énoncés pour l'évaluation

ROYA

Instance ROYA-1 : Votre attention s'il vous plait, voie A, en raison de travaux sur les voies, le TGV, numéro 8634 à destination de Reims et Strasbourg partira à 8h34.

ROYA-1				
Code	Variations			Autres
	Regard	mD	md	
E5				
Voie	Insertion	Relachement	Relachement	
A		Tension	Tension	
Cr11				
TGV				
8		Tension	Tension	
1000		Tension	Tension	
600		Tension	Tension	Une seule fois
3		Tension		Décalage 1
4		Tension		Décalage 2
E28				Tension par persévération
Reims				
Strasbourg				
E53				
8h		Tension	Tension	
3		Tension		Décalage 1
4		Tension		Décalage 2

Instance ROYA-2 : Votre attention s'il vous plaît, voie B, en raison de travaux sur les voies, le TER, numéro 1234 à destination de Colmar et Reims partira à 8h12.

ROYA-2				
Code	Variations			Autres
	Regard	mD	md	
E5				
Voie	Insertion	Relachement	Relachement	
B		Tension		
Cr11				
TGV				
1000		Tension	Tension	
200		Tension		Une seule fois
3		Tension		Décalage 1
4		Tension		Décalage 2
E28				Tension par persévération
Colmar				
Reims				Pouce tendu par persévération
E53				
8h		Tension	Tension	
12		Tension		

Instance ROYA-3 : Votre attention s'il vous plait, voie X, en raison d'un incident technique, le TER, numéro 8634 à destination de Strasbourg et Reims partira à 17h34.

ROYA-3				
Code	Variations			Autres
	Regard	mD	md	
E5				
Voie	Insertion	Relachement	Relachement	
C		Tension		
Cr04				
TER		Tension		
8		Tension	Tension	
1000		Tension	Tension	
600		Tension	Tension	Une seule fois
3		Tension		Décalage 1
4		Tension		Décalage 2
E28				Tension par persévération
Strasbourg				Raccourcissement du signe
Reims				
E53				
17h		Tension	Tension	
3		Tension		Décalage 1
4		Tension		Décalage 2

SOY

Instance SOY-1 : Votre attention s'il vous plait, voie A, en raison de mouvements sociaux, le TER numéro 88 634 à destination de Reims est supprimé. La SNCF vous présente ses excuses.

SOY-1				
Code	Variations			Autres
	Regard	mD	md	
E5				
Voie	Insertion	Relachement	Relachement	
A		Tension	Tension	
Cr26				
TER				
Numéro				
8		Tension	Tension	
8		Tension	Tension	
1000		Tension	Tension	
600		Tension		
3		Tension	Tension	Décalage 1
4		Tension	Tension	Décalage 2
E28				
Reims				
E63		Tension	Tension	
E64				

Instance SOY-2 : Votre attention s'il vous plait, voie B, en raison de travaux sur les voies, le TER numéro 88 676 à destination de Strasbourg est supprimé. La SNCF vous présente ses excuses.

SOY-2				
Code	Variations			Autres
	Regard	mD	md	
E5				
Voie	Insertion	Relachement	Relachement	
B		Tension	Tension	
Cr11				
TER				
Numéro				
8		Tension	Tension	
8		Tension	Tension	
1000		Tension	Tension	
600		Tension		
7		Tension	Tension	Décalage 1
6		Tension	Tension	Décalage 2
E28				
Strasbourg				
E63		Tension	Tension	
E64				

Instance SOY-3 : Votre attention s'il vous plait, voie C, en raison d'un incident technique, le TGV numéro 2047 à destination de Paris Est est supprimé. La SNCF vous présente ses excuses.

SOY-3				
Code	Variations			Autres
	Regard	mD	md	
E5				
Voie	Insertion	Relachement	Relachement	
C		Tension	Tension	
Cr04				
TGV				
Numéro				
2		Tension	Tension	
1000		Tension	Tension	
4		Tension		Décalage 1
7		Tension	Tension	Décalage 2
E28				
Paris Est				
E63		Tension	Tension	
E64				

AS01

Instance AS01-1 : Voie A éloignez-vous de la bordure du quai. Attention au passage d'un train.

Voie A éloignez vous de la bordure du quai.

AS01-1				
Code	Variations			Autres
	Regard	mD	md	
Voie	Insertion	Relachement	Relachement	
A				
E51				
E72		Relachement	Relachement	
Voie	Insertion	Relachement	Relachement	
A				
E51				

Instance AS01-2 : Voie 12 éloignez-vous de la bordure du quai. Attention au passage d'un train.

Voie 12 éloignez vous de la bordure du quai.

AS01-2				
Code	Variations			Autres
	Regard	mD	md	
Voie	Insertion	Relachement	Relachement	
12				
E51				
E72		Relachement	Relachement	
Voie	Insertion	Relachement	Relachement	
12				
E51				

Instance AS01-3 : Voie X éloignez-vous de la bordure du quai. Attention au passage d'un train.

Voie X éloignez vous de la bordure du quai.

AS01-3				
Code	Variations			Autres
	Regard	mD	md	
Voie	Insertion	Relachement	Relachement	
X				
E51				
E72		Relachement	Relachement	
Voie	Insertion	Relachement	Relachement	
X				
E51				

D0Y

Instance D0Y-1 : Voie B, le TGV numéro 4567 à destination de Colmar va partir. Prenez garde à la fermeture des portes.

D0Y-1				
Code	Variations			Autres
	Regard	mD	md	
Voie	Insertion	Relachement	Relachement	
B				
TGV				
Numéro				
4		Tension	Tension	
1000		Tension	Tension	
6		Tension	Tension	Décalage 1
7		Tension	Tension	Décalage 2
E28				
Colmar				
E48				
E49				

Instance D0Y-2 : Voie 12, le TER numéro 12 047 à destination de Strasbourg va partir. Prenez garde à la fermeture des portes.

D0Y-2				
Code	Variations			Autres
	Regard	mD	md	
Voie	Insertion	Relachement	Relachement	
12				
TER				
Numéro				
12		Tension	Tension	
1000		Tension	Tension	
4		Tension	Tension	Décalage 1

7		Tension	Tension	Décalage 2
E28				
Strasbourg				
E48				
E49				

Instance DOY-3 : Voie C, le Corail numéro 88 676 à destination de Paris Est va partir. Prenez garde à la fermeture des portes.

DOY-3				
Code	Variations			Autres
	Regard	mD	md	
Voie	Insertion	Relachement	Relachement	
C				
Corail				
Numéro				
8		Tension	Tension	
8		Tension	Tension	
1000		Tension	Tension	
600		Tension	Tension	
7		Tension	Tension	Décalage 1
6		Tension	Tension	Décalage 2
E28				
Paris Est				
E48				
E49				

ILLUSTRATIONS

Table des tableaux

Tableau 1 : Direction du regard – Modifications les plus importantes	91
Tableau 2 : Direction du regard – Insertions les plus importantes.....	94
Tableau 3 : Direction du regard – Suppression la plus importante	95
Tableau 4 : Main dominante – Modifications les plus importantes	99
Tableau 5 : Main dominante – Insertion la plus importante	105
Tableau 6 : Main dominante – Suppressions les plus importantes	106
Tableau 7 : Main dominée – Modifications les plus importantes	109
Tableau 8 : Main dominée – Insertions les plus importantes.....	112
Tableau 9 : Main dominée – Suppressions les plus importantes	113
Tableau 10 : Variations des segments d'énoncés qui varient toujours de la même façon	116
Tableau 11 : Variations des segments d'énoncés qui s'ils varient le font toujours de la même façon	117
Tableau 12 : Segments d'énoncés qui varient de manière « quasi » identique quelque soit le contexte.....	119
Tableau 13 : Variations du segment d'énoncé dépendant du contexte.....	120
Tableau 14 : Regroupement des configurations relevées, par catégories	123
Tableau 15 : AS01 - Variations en durée, suivant plusieurs seuils, des éléments dont la valeur ne varie pas	126
Tableau 16 : AS16 - Variations en durée, suivant plusieurs seuils, des éléments dont la valeur ne varie pas	127
Tableau 17 : IT12AB - Variations en durée, suivant plusieurs seuils, des éléments dont la valeur ne varie pas	128
Tableau 18 : D0Y - Variations en durée, suivant plusieurs seuils, des éléments dont la valeur ne varie pas	129
Tableau 19 : DT02 - Variations en durée, suivant plusieurs seuils, des éléments dont la valeur ne varie pas	130

Tableau 20 : RT20A - Variations en durée, suivant plusieurs seuils, des éléments dont la valeur ne varie pas	131
Tableau 21 : RT20B - Variations en durée, suivant plusieurs seuils, des éléments dont la valeur ne varie pas	132
Tableau 22 : Direction du regard – Comparaisons inter-énoncés des variations relevées	134
Tableau 23 : Main dominante – Comparaisons inter-énoncés des variations relevées.....	135
Tableau 24 : Main dominée – Comparaisons inter-énoncés des variations relevées	136
Tableau 25 : Ordre de présentation des instances aux trois sujets.....	158

Table des figures

Figure 1 : Représentation schématique de l'espace de signation [GUI04].....	9
Figure 2 : Schéma synthétique d'un système de reconnaissance de la parole	28
Figure 3 : Exemple de notation d'une histoire en SignWriting (début de « Blanche neige »).....	54
Figure 4 : Exemple de notation HamNoSys	55
Figure 5 : Visualisation aisée des phénomènes grâce à une annotation par symboles visuels.....	55
Figure 6 : Différentes propositions de représentation de la segmentation de l'espace de signation	56
Figure 7 : Visualisation sous Anvil des courbes issues d'annotations sur la vidéo	57
Figure 8 : Visualisation sous Anvil des courbes issues d'algorithmes de traitement d'images.....	57
Figure 9 : Anvil, constitué de 4 parties (les 2 principales : (C) et (D)).....	63
Figure 10 : Instantané de notre corpus vidéo.....	78
Figure 11 : Visualisation aisée sous Anvil grâce aux couleurs (pour l'ouverture (en vert) et la fermeture (rouge) des yeux)	78
Figure 12 : Exemple de configurations (à gauche numérisées, à droite insérées dans Anvil).....	79
Figure 13 : Exemples de captures d'écrans pour les configurations des mains	79
Figure 14 : Exemples de capture d'écrans pour la direction du regard.....	80
Figure 15 : Schéma récapitulatif des valeurs de direction du regard	80
Figure 16 : Configurations jugées similaires	83
Figure 17 : Description du format de fichier de sortie des résultats (cf. points ci-dessous)	83
Figure 18 : Exemple d'erreur de l'algorithme d'alignement.....	84
Figure 19 : Schéma récapitulatif des différents emplacements	89
Figure 20 : « Distribution des 75 configurations manuelles de la LSF »	122

Table des graphiques

Graphique 1 : Direction du regard – tous énoncés – toutes variations.....	90
Graphique 2 : Main dominante – tous énoncés – toutes variations	97
Graphique 3 : Main dominée – tous énoncés – toutes variations.....	108