



ROSETTA

accès multilingue

RObot de Sout-titrage Et Toute Traduction
Adaptés

Rapport final



Partenaire valideur non financé





Table des matières

1.	Rappel de l'organisation générale du projet	2
1.1	OBJECTIFS DU PROJET	2
1.2	PLANNING PREVISIONNEL GENERAL	2
1.3	PRINCIPAUX JALONS	3
2.	BILAN DE L'AVANCEMENT DU PROJET	4
2.1	BILAN DE LA PERIODE DE SUIVI TECHNIQUE N°1	4
2.2	BILAN DE LA PERIODE DE SUIVI TECHNIQUE N°2	5
2.3	BILAN DE LA PERIODE DE SUIVI TECHNIQUE N°3	6
3.	TABLEAU RECAPITULATIF DES TACHES DU PROJET	8
4.	TABLEAU RECAPITULATIF DES LIVRABLES DU PROJET	9
5.	DESCRIPTION DETAILLEE DES TRAVAUX	13
5.1	SP 1 : GESTION DE PROJET	13
5.2	SP 2 : COLLECTE ET PRODUCTION DE CORPUS	18
5.3	SP 3 : MODULE DE PRODUCTION DE SOUS-TITRES ADAPTES	26
5.4	SP 4 : MODULE DE TRADUCTION MULTILINGUE DE SOUS-TITRES ADAPTES	29
5.5	SP 5 : MODULE DE GENERATION DE LSF	33
5.6	SP 6 : ÉVALUATION DES MODULES	38
5.7	SP 7 : TESTS IN SITU - FOCUS GROUPS	45
5.8	SP 8 : VALORISATION	50



INTRODUCTION

Ce rapport est le rapport final du projet co-financé ROSETTA2 s'étant déroulé de juin 2018 à novembre 2021 (42 mois).

En premier lieu, un rappel est fait des objectifs du projet, du calendrier des 3 périodes de suivi technique et des étapes importantes ayant jalonné le projet.

En second lieu, ce rapport présente une synthèse des progrès et difficultés rencontrés sur les 3 périodes successives, et un récapitulatif de l'avancée cumulée des tâches et livrables au fil du temps.

Enfin, nous présentons par sous-projet (appelé également lot) un rappel des objectifs, l'état de livraison pour chaque livrable en fin de projet, puis une description qualitative des avancées réalisées remettant en contexte ces différents livrables et annonçant les perspectives possibles en fin de projet.

Les partenaires du projet sont désignés ci-dessous par les mentions suivantes :

- SYSTRAN
- France.tv access (FTVS), anciennement MFP
- LIMSI : CNRS-LIMSI, nouvellement LISN, au sein duquel coexistent plusieurs groupes :
 - TLP : Traitement du Langage Parlé
 - ILES : Information Langue Ecrite et Signée
- LUTIN : EPHE-CHArt-Lutin, Lutin-Userlab
- MOCAPLAB

Holken Consultants (HC) intervient également en sous-traitance pour les lots 7 et 8.

1. Rappel de l'organisation générale du projet

1.1 Objectifs du projet

Le projet ROSETTA, en visant à une meilleure accessibilité des contenus multimédia, poursuit 3 objectifs :

- Une chaîne de production automatisée de sous-titres adaptés et multilingues de contenus audiovisuels
- Un pas vers une automatisation de la traduction de contenus journalistiques du français vers la langue des signes française (LSF) via l'animation d'un avatar 3D
- Une intégration de ces technologies dans un démonstrateur évalué par les cibles du projet

1.2 Planning prévisionnel général

Suite à l'accord de Bpi France le 06/07/2020 pour un report de calendrier de 6 mois demandé en début de projet par le consortium, le projet d'initialement 36 mois a commencé en juin 2018 et se termine en novembre 2021 (42 mois). Le présent rapport suit le calendrier des périodes de suivi technique rappelé ci-dessous :



- Période n°1 : du 01/06/2018 au 31/08/2019 (15 mois)
- Période n°2 : du 01/09/2019 au 31/08/2020 (12 mois)
- Période n°3 : du 01/09/2020 au 30/11/2021 (15 mois)

	Période de suivi technique												Période n°1												Période n°2												Période n°3																							
	Nombre de mois												2018												2019												2020												2021											
	Date	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42																	
	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11																		
LOT1	Gestion du projet																																																											
LOT2	Collecte et production de corpus																																																											
LOT3	Module de production de sous-titres adaptés																																																											
LOT4	Module de traduction textuelle																																																											
LOT5	Module de génération de LSF																																																											
LOT6	Évaluation des modules																																																											
LOT7	Tests in situ – Focus groupes																																																											
LOT8	Valorisation																																																											

1.3 Principaux jalons

Plusieurs jalons clefs ont marqué le déroulement global du projet et la synchronisation des différentes tâches, et ont représenté les objectifs à atteindre collectivement pour chaque période :

- T0+15 (Septembre 2019) :
 - Accord de consortium finalisé
 - Finalisation de la préparation du premier lot des données d’entraînement (existantes) permettant l’entraînement de modèles de sous-titrage
 - Spécification du module de génération de LSF
 - Spécification des évaluations techniques et humaines
 - Recueil des préconisations des utilisateurs
 - Spécification du *benchmarking* du projet sur le plan européen et international
- T0+27 (Septembre 2020) :
 - Livraison de l’ensemble des données télévisuelles du projet
 - Livraison du premier corpus de “*motion capture*”
 - Premières versions d’outils pour la génération de sous-titres et animations LSF :
 - Outils d’alignement vidéo LSF – texte français
 - Modèles de génération de sous-titres français et multilingues
 - Génération d’animations en LSF pour du contenu synthétique
 - Livraison du premier rapport de *benchmark* et positionnement de ROSETTA sur le plan européen et international (recherches documentaires)
 - Identité visuelle et supports de communication en place
- T0+33 (Mars 2021) :
 - Livraison de l’ensemble des données d’entraînement du projet, dont livraison du second corpus de “*motion capture*”
 - Livraison de premiers fichiers générés par les modèles de sous-titrage français et multilingue pour évaluation
 - Spécification des tests utilisateurs
 - Lancement du site web et de vidéos de promotions
 - Livraison du second rapport de *benchmark* et positionnement de ROSETTA sur le plan européen et international (entretiens)



- T0+42 (Novembre 2021) :
 - Livraison de fichiers générés par des modèles ultérieurs de sous-titrage français et multilingue après premier feedback pour évaluation
 - Livraison d'animations en LSF pour des titres de journaux synthétiques
 - Intégration de l'ensemble des composants en un démonstrateur
 - Évaluation globale du point de vue utilisateur et métier
 - Plan d'exploitation et rapport de dissémination

2. Bilan de l'avancement du projet

2.1 Bilan de la période de suivi technique n°1

Lors des 15 premiers mois du projet, nous avons d'abord installé le **cadre** dans lequel les partenaires peuvent travailler ensemble, par le biais de l'accord de consortium et d'outils de communication.

Nous avons ainsi pu nous accorder sur le processus de **collecte, traitement et évaluation des données** vidéos et sous-titres existantes fournies par france.tv access, ainsi que sur le contenu d'un premier corpus de LSF enregistré par capture de mouvement (*mocap*) chez MOCAPLAB.

Ces données vont permettre de réaliser lors de la période n°2 diverses expérimentations pour la mise au point de modèles de génération automatique de sous-titres français, sous-titres multilingues et avatar LSF par SYSTRAN, le LIMSI et MOCAPLAB. Il est à noter que les expériences menées pour la génération automatique de LSF sont **centrées sur les actualités**, un domaine dont la couverture est éminemment demandée par la communauté sourde et malentendante, et, sans promettre la fluidité de leurs consœurs pour la génération automatique de texte, visent l'expression de structures particulièrement idiomatiques de la LSF, connues pour poser un défi de traduction à partir du français.

Le LUTIN et Holken Consultants ont recueilli les **préconisations de divers groupes d'utilisateurs**, dont la vocation est d'être intégrées le plus en amont possible dans la conception des solutions. Les méthodes d'évaluation techniques et humaines pour évaluer l'accessibilité, l'acceptabilité, l'utilité et l'utilisabilité du prototype, ainsi que la méthode de *benchmarking* pour le positionner dans l'offre européenne et internationale actuelle ont été spécifiées et leur mise en œuvre, plutôt que d'attendre la fin des expérimentations, va procéder par itération lors des périodes n°2 et n°3.

Les principales difficultés rencontrées se situent au niveau de la **livraison des éléments d'un lot d'un partenaire à un autre** : problème à délimiter avant de progresser, interface à installer avant de recevoir des données, données devant passer par des étapes de pré-traitement avant d'être mises à disposition, données et prototype devant être partagés pour être évalués, méthodes devant être spécifiées avant d'être implémentées. Ces blocages sont principalement levés par la communication directe entre les deux partenaires concernés, qui permet de connaître les contraintes techniques mutuelles, et construire une solution commune avec les forces de chacun pour les besoins spécifiques du projet.



2.2 Bilan de la période de suivi technique n°2

Après une première année de mise en place administrative et technique, la deuxième année du projet ROSETTA a été marquée par les **expérimentations** en œuvre chez les différents partenaires.

L'ensemble des **données existantes**, programmes télévisuels des chaînes de France Télévisions et sous-titres produits par les équipes de france.tv access, ont été **mises à disposition** des partenaires sur une plateforme sécurisée et traitées pour reconnaissance vocale automatique par le LIMSI.

De **nouvelles données ont également été créées** au titre du projet par MOCAPLAB, avec un premier corpus de *motion capture (mocap)*, qui a permis d'enregistrer avec précision les mouvements d'une journaliste signant en LSF des titres d'actualité de journaux télévisés, qui est le cas d'usage choisi pour le système de génération d'animations en LSF.

Ces données existantes et nouvelles ont été rendues accessibles aux partenaires qui ont pu réaliser des **premières expérimentations** en matière de génération de sous-titres français et multilingue au LIMSI et chez SYSTRAN, d'alignement vidéo LSF – texte français au LIMSI, ainsi que de génération d'animations en LSF pour du contenu synthétique (non déjà couvert dans le corpus) chez MOCAPLAB.

Les partenaires techniques (SYSTRAN, LIMSI, MOCAPLAB) sont désormais concentrés sur la **production d'outils utilisables en contexte**. Il est à noter que ces expériences s'inscrivent dans un état de l'art bien différent en ce qui concerne le sous-titrage et la génération de LSF, ainsi que l'ont montré les recherches documentaires effectuées pour le *benchmark* du projet.

Dans le cas du sous-titrage français et multilingue, le but pour SYSTRAN et le LIMSI est d'obtenir une **chaîne de traitement réalisant de bout en bout l'ensemble des opérations pour une mise en sous-titres automatique** sur des émissions entières dans divers domaines (documentaires, fictions, journaux télévisés) et de pouvoir mesurer à la fois les gains de temps qu'elle permet aux professionnels de réaliser chez france.tv access et l'acceptabilité de ses résultats par divers utilisateurs telle qu'évaluée par le LUTIN.

Dans le cas de la génération d'animation en LSF, il n'existe pas actuellement de solution automatique satisfaisante qui génère une langue des signes fluide et correcte à partir de contenu texte. Le travail commun du LIMSI, MOCAPLAB et SYSTRAN sur ce sujet vise alors à permettre sur de courts extraits la couverture par un avatar 3D de contenu nouveau appartenant au domaine précis des titres de journaux, ce qui représente une grande avancée par rapport à l'existant. Un deuxième corpus est en préparation afin de **pousser plus loin cette méthodologie**.

Des tests utilisateurs sont prévus par le LUTIN et HC afin d'**évaluer diverses dimensions des modules** du projet ROSETTA et d'aider à leur conception. Une concertation entre les partenaires est en cours car ces tests devront à la fois tenir compte de la maturité (émission entière/extrait d'émission), des délais de livraison des modules techniques (mars/mai 2021), mais aussi des différents publics (besoin d'interprètes) et des conditions sanitaires de l'année 2020-2021 (besoin de prévoir éventuellement des possibilités d'évaluation à distance).

En parallèle de ces expériences, le projet se dote actuellement grâce au LIMSI et HC d'une identité visuelle reconnaissable et de vidéos de promotion bientôt mises en ligne sur son site internet, qui souhaite démontrer toute l'**accessibilité vers laquelle est tendue le projet** : sous-titrage français et multilingue, interprète LSF.



2.3 Bilan de la période de suivi technique n°3

La dernière année du projet a permis de pousser plus loin les **développements** dans tous les lots techniques, les **évaluer** du point de vue des concepteurs, des professionnels, des utilisateurs, et d'ouvrir le projet **vers l'extérieur** en présentant objectifs, démonstrateurs et perspectives à un conseil consultatif.

Dans le domaine du **sous-titrage**, la mise à disposition d'émissions et de sous-titres de référence par l'équipe de france.tv access, les étapes de traitement préalables de transcription et alignement par le LISN et la définition commune de critères adaptés pour l'évaluation ont abouti au **développement et raffinement de modèles par expériences successives**, notamment spécialisés en fonction du domaine de l'émission (série, journal TV, jeu, documentaire, etc.).

Pour le **sous-titrage français**, le LISN montre que de meilleures notes sont obtenues pour les programmes de direct, où les sous-titres sont très **proches du verbatim** avec quelques compressions relativement systématiques, que pour les programmes de stock, où figurent des **choix beaucoup plus variés et personnels** du sous-titreur de reformulation, contextualisation, sélection. De façon intéressante, la génération de sous-titres sur les mêmes programmes à partir de transcriptions manuelles montre une amélioration mais qui reste légère : ainsi, la **qualité de la reconnaissance vocale**, la plupart du temps pointée du doigt par les experts dans les erreurs produites dans les sous-titres automatiques, ne peut forcément être toujours mise en cause et ces tests semblent plutôt révéler la difficulté pour les systèmes automatiques actuels à **maîtriser toutes les dimensions de la mise en sous-titres**.

Pour le **sous-titrage multilingue**, SYSTRAN s'est intéressé à la tâche de génération automatique sous deux angles, d'une part à partir des sous-titres de référence de france.tv access dans l'idée de rendre accessible en d'autres langues **des programmes déjà sous-titrés** en langue source, et d'autre part à partir des transcriptions de la bande son produites par le LISN, ce qui couvre un champ plus large car permet de produire des sous-titres pour **des programmes non préalablement sous-titrés**. La 2^e approche, le passage du transcript aux sous-titres en langue cible, ouvrant théoriquement plus de perspectives d'accessibilité mais ayant recueilli des retours métiers mitigés en l'état actuel, il a été choisi de concentrer les efforts sur la 1^{ère} approche, la **traduction des sous-titres** d'une langue source vers une langue cible, en particulier du français vers l'anglais, avec des **progrès mesurables** permettant d'envisager une exploitation.

Au vu du **niveau de qualité des sous-titres attendu pour la télévision** (charte CSA) et france.tv access s'intéressant principalement à l'automatisation de la chaîne de production de sous-titres pour des programmes de stock, les méthodes automatiques développées dans le projet **ne peuvent mener à une mise en production directe**.

En revanche, france.tv access a pu initier une communication auprès de ses équipes sur les méthodes d'intelligence artificielle arrivant dans leur domaine d'expertise, et la recherche de meilleurs systèmes se poursuit pour le LISN et SYSTRAN, en parallèle de l'exploration par SYSTRAN d'**outils d'intégration et d'édition de sous-titres automatiques (alignement temporel, post-édition facilitée)**, maillon concret sans lequel l'adoption de solutions automatiques pour le sous-titrage dans ce contexte ou d'autres restera limitée.

Dans le cas de la **génération d'animations LSF pour couvrir des titres journalistiques**, le traitement de la seconde partie du corpus par MocapLab, l'élaboration de nouvelles façons d'annoter linguistiquement et temporellement les données existantes et nouvellement créées autour de la représentation intermédiaire Azee par le LISN, le développement d'un algorithme de traduction par analogie par SYSTRAN et le LISN, et l'optimisation des méthodes de combinaison des blocs de



mouvement par MocapLab, ont été les briques essentielles de la dernière année d'**expériences de traduction du français vers la LSF**.

Fruit de ce travail, le démonstrateur intègre la **production d'animations LSF pour 7 exemples choisis de titres journalistiques**. On est donc loin d'un « produit fini » directement utilisable, mais le choix d'une représentation multilinéaire et fidèle à la langue cible, ainsi que l'enregistrement du mouvement avec une très grande précision par motion capture, offrent des résultats inédits pour la tâche. Contrairement à certains travaux récents dans la traduction vers des langues signées utilisant les réseaux de neurones, avec un fort pouvoir de généralisation mais peu de contrôle sur le résultat, les enjeux de l'utilisation des technologies choisies dans le projet sont plutôt le **passage à une échelle plus grande** et la **sélection du bon niveau d'informations**, nécessaires et suffisantes pour une traduction acceptable, dans des représentations de la langue et du mouvement aussi précises et exhaustives.

Un aspect important du travail effectué est toutefois d'avoir contribué à rendre **plus automatique** une chaîne de production des animations qui était jusqu'alors entièrement manuelle, tant du côté traduction que du côté génération. Les résultats permettent de valider des choix techniques et ouvrent de **nouveaux champs** à automatiser pour pouvoir couvrir un contenu plus large.

Ces travaux s'inscrivent dans une démarche plus large d'**outillage des langues signées**, qui bénéficient encore d'outils bien moins nombreux et aboutis que les langues écrites, et d'**implication de Sourds et professionnels de la LSF** à différents niveaux (enregistrement, annotations, évaluation). Le LISN, SYSTRAN et MocapLab préparent d'ores et déjà une publication commune sur le travail effectué, et sont ouverts à de futures opportunités de collaboration sur le sujet.

Deux phases d'**évaluation auprès des utilisateurs** des modules techniques arrivés à une maturité suffisante ont pu être menées par le LUTIN et Holken Consultants lors de l'année 2021, permettant de situer les développements réalisés par rapport aux outils existants et aux attentes de divers groupes d'utilisateurs. Ces évaluations ont été **fortement adaptées à un format distanciel** en raison du contexte sanitaire, ce qui a révélé certaines limites notamment lorsque le matériel consiste en vidéos dont on veut évaluer l'accessibilité mais a aussi ouvert de nouvelles pistes de recrutement et de recueil d'avis.

Basées sur la libre expression de ressentis face à l'innovation technologique présentée, ces évaluations donnent des **impressions globales informatives**. De futurs projets pourraient bénéficier d'évaluations plus spécifiques, à la fois aux groupes d'utilisateurs et aux modules évalués, afin de canaliser les retours **selon des dimensions précises**, permettant aux concepteurs de mieux les intégrer.

Les activités de dissémination du projet par HC ont aussi dû se reporter essentiellement sur le volet virtuel. Ont néanmoins eu lieu trois réunions, accessibles avec interprétariat LSF et transcription de la parole, réunissant en présentiel **les partenaires et les conseillers et observateurs**, membres d'associations, fédérations et institutions intéressées.

Tandis que pour les professionnels une évolution des outils implique une évolution importante des métiers, les concepteurs d'outils doivent garder à l'esprit de rester un **apport de qualité** et un **gain de temps pour ces métiers**, dans un contexte où du sous-titrage automatique de qualité variable et de nouveaux standards arrivent sur les plateformes de diffusion massive, mais aussi où les usages (smartphones, objets connectés, *zapping*, usage générationnel du code couleur...) et les contenus multimédias (émissions en replay, cours en ligne, vidéos courtes, tutoriels, vidéos générées par des indépendants sur les réseaux sociaux...) **évoluent à la demande des utilisateurs**.



3. Tableau récapitulatif des tâches du projet

Lot	Tâche	P1	P2	P3
1	Tâche 1.1 : Accord de consortium	100 %		
	Tâche 1.2 : Gestion de projet	33 %	64 %	100%
2	Tâche 2.1 : Identification et collecte de corpus existants	50 %	100 %	
	Tâche 2.2 : Outil logiciel d'aide à l'alignement	50 %	64 %	100 %
	Tâche 2.3 : Alignement, annotation et analyse des données	50 %	75 %	100%
3	Tâche 3.1 : Modèles neuronaux de production de sous-titres adaptés par domaine	20 %	50 %	100%
	Tâche 3.2 : Evaluation des modèles			100 %
4	Tâche 4.1 : Production de modèles de sous titres adaptés - EN, ES, IT, PT, ZH	20 %	50 %	100%
	Tâche 4.2 : Evaluation des modèles			100 %
5	Tâche 5.1 : Cahier des charges	100%		
	Tâche 5.2 : Conception et développement	33 %	64 %	100%
	Tâche 5.3 : Intégration et visualisation			100%
6	Tâche 6.1 : Spécifications	100 %		
	Tâche 6.2 : Production des corpus et outils d'évaluation	80 %	80 %	100%
	Tâche 6.3 : Évaluations	30 %	30 %	100 %
7	Tâche 7.1 : Phase préparatoire : caractérisation des situations d'accessibilité télévisuelles	100 %		
	Tâche 7.2 : Organisation du retour d'expérience	100 %		
	Tâche 7.3 : Tests d'utilisateurs in situ	40 %	60 %	100 %
	Tâche 7.4 : Etude comparative sur le plan international	30 %	50 %	100 %
8	Tâche 8.1 : Plan de dissémination et d'exploitation	En attente	25 %	100 %
	Tâche 8.2 : Matériels de communication	En attente	80 %	100 %
	Tâche 8.3 : Démonstrateur		20 %	100 %



4. Tableau récapitulatif des livrables du projet

#	Échéance	Intitulé livrable	Type	Responsable	État
1.1.1	T0+6	Accord de consortium	Document	SYSTRAN	Livré P1
1.2.1	T0+3	Rapport intermédiaire	Document	SYSTRAN	Livré P1
1.2.2	T0+6	Rapport intermédiaire	Document	SYSTRAN	Livré P1
1.2.3	T0+12	Rapport intermédiaire	Document	SYSTRAN	Livré P1
1.2.4	T0+21	Rapport intermédiaire	Document	SYSTRAN	Fusionné avec 1.2.8
1.2.5	T0+27	Rapport intermédiaire	Document	SYSTRAN	Fusionné avec 1.2.8
1.2.6	T0+33	Rapport intermédiaire	Document	SYSTRAN	Fusionné avec 1.2.9
1.2.7	T0+39	Rapport intermédiaire	Document	SYSTRAN	Fusionné avec 1.2.9
1.2.8	T0+21	Rapport de mi-parcours	Document	SYSTRAN	Livré P2
1.2.9	T0+39	Rapport Final	Document	SYSTRAN	Livré P3
2.1.1	T0+3	Cahier des charges pour la constitution des ressources langagières	Document	FTVS	Livré P1
2.1.2	T0+9	Pertinence extensionnelle et intensionnelle des corpus et recommandations	Document	LUTIN	Livré P2
2.2.1	T0+12	Outils logiciels d'aide à l'alignement et l'annotation des données 3D de LSF	Logiciel	LIMSI	Livré P3
2.3.1	T0+12	Corpus existants alignés, annotés et validés – lot1	Corpus	FTVS	Livré P1
2.3.2	T0+24	Corpus existants alignés, annotés et validés – lot2	Corpus	FTVS	Livré P2
2.3.3	T0+12	Nouveau corpus alignés, annotés et validés – lot1	Corpus	MOCAPLAB	Livré P2
2.3.4	T0+27	Nouveau corpus alignés, annotés et validés – lot2	Corpus	MOCAPLAB	Livré P3



3.1.1	T0+6	Préconisations utilisateurs pour le sous-titrage	Document	SYSTRAN	Remplacé par les rapports 7.1.1.
3.1.2	T0+30	Modèles neuronaux de production sous-titres adaptés aux journaux télévisés	Modèle	SYSTRAN	Livré P3
3.1.3	T0+30	Modèles neuronaux de production sous-titres adaptés aux débats documentaires	Modèle	SYSTRAN	Livré P3
3.1.4	T0+30	Modèles neuronaux de production sous-titres adaptés aux séries et films	Modèle	SYSTRAN	Livré P3
3.2.1	T0+33	Rapport d'évaluation des modèles de sous-titrage simplifié	Document	SYSTRAN	Livré P3
4.1.1	T0+12	Préconisations des utilisateurs pour le sous-titrage adapté multilingue	Document	SYSTRAN	Remplacé par les rapports 7.1.1.
4.1.2	T0+35	Rapport sur le développement des modèles de sous-titrages adaptés en anglais, espagnol, italien, portugais et chinois	Document	SYSTRAN	Fusionné avec 4.2.1
4.2.1	T0+37	Rapport de développement et d'évaluation	Document	SYSTRAN	Livré P3
5.1.1	T0+6	Cahier des charges du module	Document	LIMSI	Livré P1
5.1.2	T0+9	Recommandations utilisateurs pour l'animation LSF	Document	LIMSI	Remplacé par les rapports 7.1.1.
5.2.1	T0+12	Version 1 de la représentation et des deux systèmes	Logiciel	LIMSI	Livré P1
5.2.2	T0+27	Version 2 de la représentation et des deux systèmes	Logiciel	LIMSI	Livré P2
5.2.3	T0+39	Version Finale de la représentation et des deux systèmes	Logiciel	LIMSI	Livré P3
5.3.1	T0+39	Intégration des deux systèmes et restitution des animations dans un player 3D	Logiciel	LIMSI	Livré P3
5.3.2	T0+39	Rapport de développement des systèmes pour la génération automatique d'animations 3D en LSF à partir du français	Document	LIMSI	Livré P3



6.1.1	T0+3	Spécifications des évaluations	Document	LUTIN	Livré P1
6.2.1	T0+6	Méthodes d'évaluation : outils et procédure. Livraison intermédiaire	Outils / Procédures	LUTIN	Livré P1
6.2.2	T0+12	Méthodes d'évaluation : outils et procédure. Livraison intermédiaire	Outils / Procédures	LUTIN	Livré P2
6.2.3	T0+21	Méthodes d'évaluation : outils et procédure. Livraison intermédiaire - Méthodes d'évaluation <i>focus group</i>	Outils / Procédures	LUTIN	Livré P3
6.2.4	T0+27	Méthodes d'évaluation : outils et procédure. Livraison intermédiaire - Méthodes d'évaluation en ligne Partie 1 : Module LSF Partie 2 : Module sous-titrage	Outils / Procédures	LUTIN	Livré P3
6.2.5	T0+33	Méthodes d'évaluation : outils et procédure. Livraison intermédiaire - Méthodes pour sous-titres bruités	Outils / Procédures	LUTIN	Livré P3
6.2.6	T0+39	Méthodes d'évaluation : outils et procédure. Final - Synthèse	Outils / Procédures	LUTIN	Livré P3
6.3.1	T0+12	Résultats évaluation technique-1 - Recommandations	Document	LUTIN	Livré P3
6.3.2	T0+27	Résultats évaluation technique-2 – Questionnaires en ligne	Document	LUTIN	Livré P3
6.3.3	T0+39	Résultats évaluation technique-3 – Sous-titres bruités	Document	LUTIN	Livré P3
6.3.4	T0+39	Résultats de l'évaluation humaine - Synthèse	Document	LUTIN	Livré P3
7.1.1	T0+3	Rapport phase préparatoire intégrant les axes de décision	Document	LUTIN	Livré P1, P2
7.2.1	T0+6	Rapports sur les typologies et sur l'apport incrémental des phases de développement	Document	LUTIN	Livré P1
7.3.1	T0+12	Rapports sur les dimensions de l'utilisabilité	Document	LUTIN	Livré P3
7.3.2	T0+27	Rapports sur les dimensions de l'utilisabilité - 2	Document	LUTIN	Livré P3
7.4.1	T0+9	Rapport benchmark et positionnement de ROSETTA sur le plan européen et international	Document	LUTIN	Livré P1



7.4.2	T0+33	Mise à jour benchmark et positionnement de ROSETTA sur le plan européen et international	Document	LUTIN	Livré P3
8.1.1	T0+7	Plan de dissémination	Document	LIMSI	Livré P2
8.1.2	T0+9	Advisory Board (AB) en place et réunion 1	Meeting	LIMSI	12 juillet 2021, rapport livré P3
8.1.3	T0+21	Advisory Board (AB) en place et réunion 2	Meeting	LIMSI	29 sept 2021, rapport livré P3
8.1.4	T0+30	Advisory Board (AB) en place et réunion 3	Meeting	LIMSI	23 nov 2021, rapport livré P3
8.1.5	T0+39	Plan d'exploitation	Document	LIMSI	Livré P3
8.1.6	T0+39	Rapport de dissémination	Document	LIMSI	Livré P3
8.2.1	T0+9	Identité visuelle	Document	LIMSI	Livré P2
8.2.1	T0+12	Site web et réseaux sociaux	Meeting	LIMSI	Livré P3
8.2.3	T0+9	Matériels de communication en place	Meeting	LIMSI	Livré P2
8.3.1	T0+9	Spécification du démonstrateur	Document	SYSTRAN	Description intégrée dans le rapport 8.1.6
8.3.2	T0+27	Prototype du démonstrateur	Document	SYSTRAN	Description intégrée dans le rapport 8.1.6
8.3.3	T0+39	Version Finale du démonstrateur	Logiciel	SYSTRAN	Livré P3



5. Description détaillée des travaux

5.1 SP 1 : Gestion de projet

Objectif

L'objectif de ce sous projet est le pilotage et la gestion globale du projet.

Date de début : T0		Date de fin : T0+42			
Responsable : SYSTRAN		Participants : Tous			
	SYSTRAN	FTVS	MOCAPLAB	LIMSI	LUTIN
Charge (h*m)	6	3	6	4	4

Tâches et livrables sur la période

Tâche 1.1 : Accord de consortium

Tâche 1.2 : Gestion de projet

La gestion de projet comporte les activités suivantes :

- Planning des activités collaboratives
- Suivi des tâches et de leur coordination
- Rapport des réunions et décisions
- Suivi des livrables
- Analyse des risques et définition des plans éventuels pour y remédier

#	Échéance	Intitulé livrable	Type	Responsable	État
1.1.1	T0+6	Accord de consortium	Document	SYSTRAN	Livré P1
1.2.1	T0+3	Rapport intermédiaire	Document	SYSTRAN	Livré P1
1.2.2	T0+6	Rapport intermédiaire	Document	SYSTRAN	Livré P1
1.2.3	T0+12	Rapport intermédiaire	Document	SYSTRAN	Livré P1
1.2.4	T0+21	Rapport intermédiaire	Document	SYSTRAN	fusionné avec 1.2.8
1.2.5	T0+27	Rapport intermédiaire	Document	SYSTRAN	fusionné avec 1.2.8
1.2.6	T0+33	Rapport intermédiaire	Document	SYSTRAN	fusionné avec 1.2.9
1.2.7	T0+39	Rapport intermédiaire	Document	SYSTRAN	fusionné avec 1.2.9
1.2.8	T0+21	Rapport de mi-parcours	Document	SYSTRAN	Livré P2
1.2.9	T0+39	Rapport Final	Document	SYSTRAN	Livré P3



Avancées

En se réunissant pour un point téléphonique de suivi du projet toutes les deux semaines, les partenaires du consortium ont fait progresser le projet sur plusieurs dimensions en parallèle :

Période	Sous-titrage	Génération de LSF	Evaluations utilisateurs	Projet
Juin 2018 – Mars 2019 : Mise en place du cadre du projet et des étapes des chaînes de production	<p>Nov 2018 – Fev 2019 : Tests et discussions sur le format des données mises à disposition par FTVS</p> <p>Visite des locaux et démonstration des méthodes de France.TV access le 03/12/18</p> <p>Février 2019 : Réunion de travail LIMSI(TLP)-SYSTRAN (reconnaissance vocale, transcription, choix d'émissions)</p>	<p>Nov 2018 – Jan 2019 : Demandes de récupération de corpus LSF externes (Mediapi, INA, M6, Bête à Bon Dieu productions)</p> <p>Mars 2019 : Réunions de travail LIMSI(ILES)-SYSTRAN-MOCAPLAB (traduction FR > représentation LSF > corpus de <i>mocap</i>, choix d'émissions)</p>	<p>Phase préparatoire de brainstorming sur les attentes des partenaires le 27/02/19</p> <p>Mars 2019 : Réunion de travail LIMSI (ILES) et LUTIN (préparation pour <i>focus groups</i> et <i>brainstormings</i>)</p>	<p>Réunion de kick-off du projet le 17/10/18 dans les locaux de SYSTRAN</p> <p>Décembre 2018 : Installation des moyens de communication entre partenaires</p> <p>Réunions administratives et juridiques le 16/01/19 et le 19/03/19</p> <p>Mars 2019 : Signature d'un NDA</p>
Mars 2019 - Sept 2019 : Groupes de travail thématiques & spécifications	<p>Mai – Juil 2019 : Réunions de travail (sélection d'émissions pertinentes) -France.TV access et SYSTRAN -LIMSI(TLP) et SYSTRAN -France.TV access et LIMSI(TLP) -France.TV access et LUTIN</p> <p>Mai – Sept 2019 : Mise en place de l'interface définitive ATTESOR d'accès aux données chez FTVS</p>	<p>Avril – Juin 2019 : Tests avec données LSF internes (FTVS)</p> <p>Mars – Mai 2019 : Réunions de travail SYSTRAN, LIMSI(ILES) et MOCAPLAB (étapes de génération, contenu du 1er corpus)</p> <p>Tournage du 1er corpus de mocap le 05/09/19</p>	<p>Avril – Juil 2019 : Phase préparatoire de brainstorming et focus groups avec divers publics :</p> <p>-Experts en accessibilité le 10/04/19</p> <p>-Sourds et malentendants le 10/07/19</p> <p>-Etudiants étrangers le 26/07/19</p>	<p>Mai 2019 : Signature de l'accord de consortium</p>



Oct 2019 - Mar 2020 : Premier bilan & mise à disposition des données	Décembre 2019 : Début de la transcription de programmes télévisuels Février 2020 : Réunion de travail SYSTRAN – LIMSI(TLP) (jeux de test, expériences, date possible des premières livraisons de fichiers)	Oct 2019 – Janv 2020 : Traitement et livraison complète du 1er corpus de mocap enregistré Déc 2019 - Mar 2020 : Réunions de travail SYSTRAN – LIMSI(ILES) - MOCAPLAB (étapes de traitement et d'annotation, points de blocages, planning des expériences à réaliser)	Focus group préparatoire additionnel : -Seniors le 31/01/20 Février 2020 : Réunion LUTIN – FTVS sur la pertinence extensionnelle des programmes choisis pour le système ROSETTA	Préparation le 01/10/19 + Réunion d'avancement n°1 le 21/11/19 Décembre 2019 : Appel d'offre LUTIN – HC validé Réunion de consultation avant interviews pour le <i>benchmark</i> le 03/12/19 Mars 2020 : Lettre de report de fin de projet envoyée au nom du consortium
Avril 2020 - Oct 2020 : Premiers résultats & début des actions de communication	Avril 2020 : Réunion de travail FTVS – LIMSI(TLP) - SYSTRAN (critères d'évaluation automatique des sous-titres) Avril-Juin 2020 : Interruption temporaire de la réception des données FTVS sur ATTESOR Août 2020 : 1ers résultats d'expériences de mise en sous-titre multilingue (Système V0, 5 langues) Oct 2020 : 1ere version des modèles de mise en sous-titres français (avec simplification)	Mai – Oct 2020 : Réunions de travail LIMSI(ILES) – SYSTRAN – MOCAPLAB (expériences avec 1er corpus de <i>mocap</i> , traduction texte vers représentation LSF, génération d'animations, préparation du 2e corpus de <i>mocap</i>) Mars – Juillet 2020 : Annotation linguistique pour le 1er corpus de <i>mocap</i> Septembre 2020 : Premiers résultats d'expériences pour la génération d'animation en LSF 2e tournage de mocap le 27/10/20	Octobre 2020 : Réunions de préparation des évaluations - LUTIN- LIMSI(ILES)- MOCAPLAB pour la LSF - LUTIN-LIMSI(TLP)- SYSTRAN pour le sous-titrage	Avril 2020 : Début des activités de communication Juillet 2020 : Report de la fin du projet confirmé Réunion d'avancement n°2 le 17/09/20 Tournages de vidéos de promotion le 17/09/20, 28/09/20 Octobre 2020 : Première version du site internet en ligne



<p>Nov 2020 - Avril 2021 : Améliorations techniques & premières évaluations métier et utilisateurs</p>	<p>Novembre 2020 : Développement finalisé sur ATTESOR pour l'évaluation métier</p> <p>Février 2021 : Réunion de travail LIMSI(TLP) - SYSTRAN - FTVS (méthodes d'évaluation)</p> <p>Février 2021 : Livraison de premiers fichiers de sous-titres français générés automatiquement par le LIMSI(TLP)</p> <p>Début de l'évaluation par FTVS pour le sous-titrage français à partir du 1er mars (50-100/semaine)</p> <p>Mars 2021 : Interruption de l'accès à ATTESOR suite à l'incendie du data center d'OVH le 10/03/21, puis reprise de l'accès fin mars</p>	<p>Nov 2020 – Mars 2021 : traitement du 2e corpus de mocap</p> <p>Fév - Mars 2021 : Réunions de travail SYSTRAN – LIMSI(ILES) – MOCAPLAB (alignements AZee, expériences de similarité, algorithme de traduction de français vers AZee)</p> <p>Janv – Avril 2021 : Annotation linguistique du 2e corpus de mocap</p> <p>Janv – Avril 2021 : Annotation AZee de corpus permettant l'alignement entre français, structure LSF et mouvement</p>	<p>Janv – Avr 2021 : 1ere phase de focus groups pour l'évaluation des 1eres versions produites par les lots techniques (en visio) :</p> <ul style="list-style-type: none">- Adultes tout venant le 07/01/21- Hispanophones apprenant le français le 28/01/21- Francophones apprenant l'anglais le 04/02/21- Seniors le 11/02/21- Sinophones apprenant le français le 04/03/21- Sourds et Malentendants le 25/03/21- Malentendants le 15/04/21- Francophones apprenant l'espagnol le 15/04/21	<p>Mars 2021 : Version finale du Benchmark International (phase d'entretiens)</p> <p>Mars 2021 : Appel d'offre LIMSI – HC finalisé</p> <p>Avril 2021 : Site internet français et anglais finalisé</p>
---	--	---	--	---



<p>Mai 2021 – Nov 2021 : Production et évaluation des derniers résultats & ouverture sur l'extérieur</p>	<p>Avr – Nov 2021 : Evaluation métier par FTVS</p> <p>Mai 2021 : Envoi de fichiers de sous-titres multilingues automatiques pour évaluations (Système V1, anglais & espagnol)</p> <p>Mai 2021 : Réunion de mise au point FTVS-LIMSI(TLP)-SYSTRAN</p> <p>Août 2021 : Envoi de fichiers de sous-titres multilingues automatiques pour évaluations (Système V2, anglais)</p> <p>Nov 2021 : Génération des fichiers de sous-titres pour le démonstrateur avec les meilleurs systèmes développés</p>	<p>Mai 2021 : Réunion de travail SYSTRAN-LIMSI(ILES)-MOCAPLAB</p> <p>Juin – Août 2021 : Génération d'animations en LSF avec les dernières méthodes pour le démonstrateur</p> <p>Juin – Nov 2021 : Rédaction d'un article et réflexion sur la présentation de la méthodologie de façon visuelle</p>	<p>Juillet 2021 : Focus group additionnel sur le matériel de la 1ere phase (en présentiel) :</p> <ul style="list-style-type: none">- Entendants locuteurs de la LSF le 05/07/21 <p>Sept-Oct 2021 : 2e phase de focus groups pour l'évaluation des versions suivantes produites par les lots techniques (en visio) :</p> <ul style="list-style-type: none">- Seniors le 16/09/21-Adultes tout venant le 23/09/21- Malentendants le 30/09/21- Hispanophones apprenant le français le 14/10/21- Francophones apprenant l'anglais le 21/10/21- Sourds le 26/10/21- Entendants locuteurs de la LSF le 28/10/21	<p>Advisory Board n°1 le 12/07/21 à la Cité des Sciences</p> <p>Réunion d'avancement n°3 le 14/09/21 à SYSTRAN</p> <p>Advisory Board n°2 le 29/09/21 à SYSTRAN</p> <p>Advisory Board n°3 le 23/11/21 à france.tv access</p> <p>Nov 2021 : Intégration du démonstrateur final sur le site internet</p>
---	---	---	---	--



5.2 SP 2 : Collecte et production de corpus

Objectif

L'objectif du sous-projet est de collecter et de produire des corpus qui seront ensuite utilisés pour l'apprentissage, la réalisation et l'évaluation des modèles de production automatique de sous-titres adaptés multilingues et de LSF.

Ce lot consiste à collecter les ressources langagières utilisées dans les lots 3, 4 et 5.

Les corpus doivent être pertinents en termes de représentativité et de recouvrement, en termes de redondance et de diversité et avoir une bonne validité écologique.

Date de début : T0		Date de fin : T0+33			
Responsable : France.TV access		Participants : Tous			
	SYSTRAN	FTVS	MOCAPLAB	LIMSI	LUTIN
Charge (h*m)	8	6	30	34	7

Tâches et livrables sur la période

Tâche 2.1 : Identification et collecte de corpus existants

Cette tâche a les objectifs suivants :

- Identification de ressources pour la constitution des corpus
- Collecte de corpus existants
- Validation des corpus

#	Échéance	Intitulé livrable	Type	Responsable	État
2.1.1	T0+3	Cahier des charges pour la constitution des ressources langagières	Document	FTVS	Livré P1
2.1.2	T0+9	Pertinence extensionnelle et intensionnelle des corpus et recommandations	Document	LUTIN	Livré P2



Tâche 2.2 : Outil logiciel d'aide à l'alignement

Cette tâche consiste en la création d'un outil permettant l'enregistrement en 3D (capture de mouvement) et le post-traitement d'un corpus en LSF afin de pouvoir créer des corpus alignés vidéos – transcription.

#	Échéance	Intitulé livrable	Type	Responsable	État
2.2.1	T0+12	Outils logiciels d'aide à l'alignement et l'annotation des données 3D de LSF	Logiciel	LIMSI	Livré P3

Avancées

Comme la constitution de logiciels de traitement automatiques dédiés aux langues des signes sont des travaux à long terme, nous avons lancé en parallèle trois études complémentaires qui ont porté sur la constitution d'outils autour des corpus de LS.

1. Logiciel de Traduction Assistée par Ordinateur français/LSF

La traduction automatique nécessite des données alignées et il en existe très peu en français/LSF. Un moyen d'en constituer plus est de développer des outils de type concordancier dans un environnement de traduction assistée par ordinateur (TAO), ce qui devrait permettre de collecter les alignements faits par les traducteurs professionnels. Ce type de logiciel n'existe pas encore pour les LS, donc nous avons lancé une étude (thèse de Marion Kaczmarek) sur ce sujet.

Afin de déterminer plus spécifiquement les étapes de la traduction français/LSF, deux études ont été menées avec des professionnels du métier : d'une part un brainstorming pour les amener à réfléchir et verbaliser leurs besoins et les problèmes rencontrés au quotidien dans leurs pratiques professionnelles, et d'autre part, des sessions où des binômes de traducteurs étaient filmés pendant qu'ils travaillaient sur la traduction de textes journalistiques. Cela a permis de dresser une liste des tâches inhérentes à la traduction en LS, mais également déterminer si elles étaient systématiques, et ordonnées.

A partir de ces observations, une liste de fonctionnalités a été élaborée pour un logiciel de TAO : traitement du texte source (par exemple repérage automatique des dates et des entités nommées), possibilité de modifier l'ordre des séquences, car cet ordre n'est pas identique en LS (qui place toujours le contexte avant) et dans les langues parlées, présence de modules d'aide (lexique, encyclopédie, carte géographique, etc.), mémoire de traduction et prompteur pour la réalisation finale.

Deux prototypes ont été réalisés. Une interface qui implémente les principes définis dans le cahier des charges (figure 1), ainsi qu'un concordancier bilingue (figure 2), basé sur un corpus parallèle de textes et de vidéos de traduction en LSF. Celui-ci est actuellement accessible en ligne sur une plateforme dédiée, et permet aux testeurs de faire des requêtes de mots ou ensemble de mots afin de pouvoir visualiser les extraits vidéo en LSF en contexte. Il reste à être intégré au logiciel de TAO.

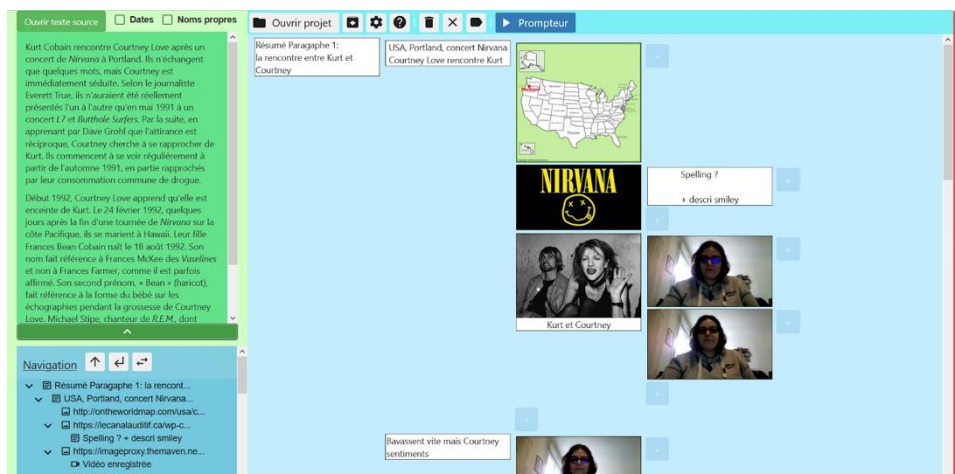


Figure 1 : Interface du prototype de logiciel de TAO texte vers LSF.

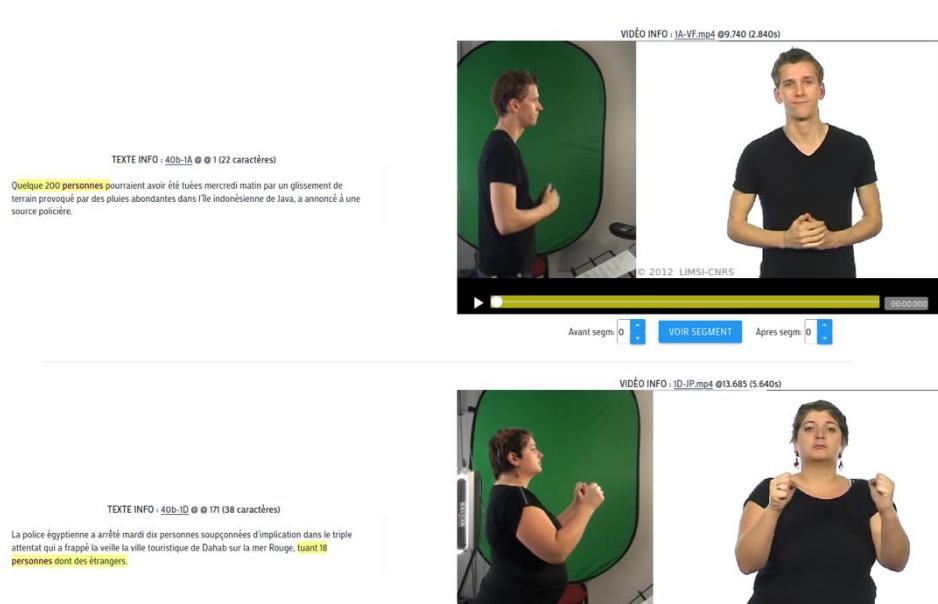


Figure 2 : Concordancier en ligne texte/LSF

Publications associées :

M. Kaczmarek, M. Filhol (2019), Computer-assisted Sign Language translation: a study of translators' practice to specify CAT software, 6th International Workshop on Sign Language Translation and Avatar Technology (SLTAT 2019), Hamburg, Germany.

M. Kaczmarek, M. Filhol (2019), Assisting Sign Language Translation: what interface given the lack of written form and the spatial grammar?, 41st Translating and the Computer (TC 2019), Londres, United Kingdom, 83-92.

M. Kaczmarek, M. Filhol (2020), Use cases for a Sign Language Concordancer, 9th Workshop on the Representation and Processing of Sign Languages (WRPSL@LREC 2020), 12th Int Conf on Language Resources and Evaluation (LREC2020), Marseille, France, 113-116.

M. Kaczmarek, M. Filhol (2020), Alignments Database for a Sign Language Concordancer, 12th Int Conf on Language Resources and Evaluation (LREC2020), Marseille, France, 6069-6072.

M. Kaczmarek, A. Larroque, M. Filhol. Logiciel de Traduction Assistée par Ordinateur des Langues des Signes. Colloque Handiversité 2021, Apr 2021, Orsay, France.



M. Kaczmarek, A. Larroque (2021), Traduction Assistée par Ordinateur des Langues des Signes : élaboration d'un premier prototype, 28ème conférence sur le Traitement Automatique des Langues Naturelles, 2021, Lille, France. pp.108-122.

M. Kaczmarek, M. Filhol. Computer-assisted sign language translation: a study of translators' practice to specify CAT software - 2021 version. Machine Translation, Springer Verlag, 2021, 35 (3), pp.305-322.

2. Alignement automatique sous-titrage/LSF

La deuxième étude a consisté à explorer si de tels alignements peuvent être réalisés de manière automatique. Une étude (thèse de Hannah Bull) a été consacrée à l'alignement automatique entre les sous-titres et des segments de LSF avec une approche par apprentissage. Pour ce faire, nous avons développé un corpus parallèle de près de 30h, basé sur le site web de Media'Pi !, déjà partenaire du projet pour la constitution du corpus Rosetta (voir tâche 2.3). Une première version du système a été basée sur des éléments prosodiques tels que les pauses dans le mouvement, ou des postures spécifiques. Le travail actuel consiste à repérer dans la vidéo des segments qui correspondent à des « unités de sous-titrages », par une approche par apprentissage illustrée figure 3.

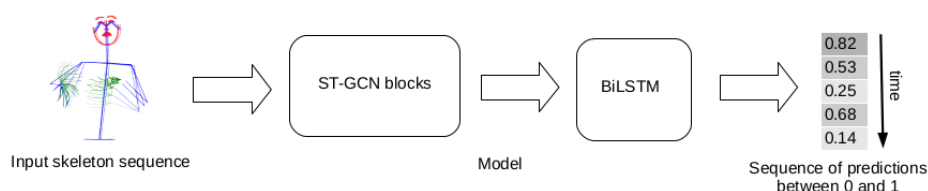


Figure 3 : Architecture du système de segmentation de vidéos de LSF en unités de type phrases.

Publications associées :

H. Bull, A. Braffort, M. Gouiffès (2020), MEDI-API-SKEL - A 2D-Skeleton Video Database of French Sign Language With Aligned French Subtitles, 12th Int Conf on Language Resources and Evaluation (LREC2020), Marseille, France, 6063-6068.

H. Bull, A. Braffort, M. Gouiffès (2020), Corpus Mediapi-skel sur Ortolang : <https://www.ortolang.fr/market/corpora/mediapi-skel> (56 téléchargements)

H. Bull, A. Braffort, M. Gouiffès (2020), Automatic Segmentation of Sign Language into Subtitle-Units, Sign Language Recognition, Translation & Production workshop (SLRTP20), 16th European conference on computer vision (ECCV'20) – Best paper award.

3. Anonymisation des animations

Enfin, une troisième étude (thèse de Félix Bigand) a porté sur la nature des données collectées dans le projet. Comme la génération des animations est réalisée à partir de données captées à l'aide d'un système de capture de mouvement très précis, il est tout à fait possible de reconnaître la personne qui a été « *mocapée* », même si les traits de l'avatar sont très différents de ceux de la personne enregistrée, ce qui peut être problématique dans certains contextes. L'objectif est d'identifier quels sont les paramètres du mouvement qui définissent le style individuel des signeurs, de manière à pouvoir jouer sur ces paramètres pour rendre la LSF enregistrée anonyme, tout en la gardant parfaitement compréhensible.



Après des études sur la structure complexe des mouvements en LSF, le travail s'est centré sur la problématique de l'identification des locuteurs par leur mouvement, avec 3 autres contributions :

- Une étude de perception visuelle qui a permis de démontrer que les sujets étaient capables de reconnaître la personne qui s'exprimait à partir d'un simple extrait court sous forme de point lumineux en mouvement (figure 4).
- Un modèle par apprentissage basé sur les données statistiques issues du corpus MOCAP1 pour chaque signeur (figure 5). Les paramètres de la signature cinématique de l'identité des signeurs a pu être définie grâce à ce modèle.
- Un algorithme de synthèse qui permet de modifier les paramètres définissant la signature cinématique des signeurs, qui peut être appliqué à l'anonymisation (figure 6).

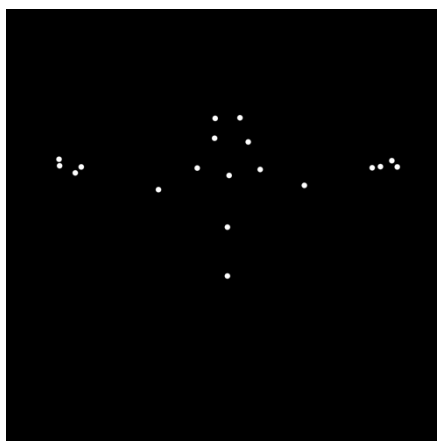


Figure 4 : Exemple d'affichage par points lumineux utilisé dans l'expérimentation.

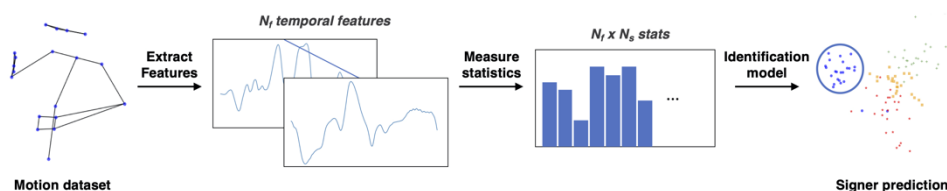


Figure 5 : Schéma des étapes utilisées dans le modèle d'apprentissage pour l'identification.

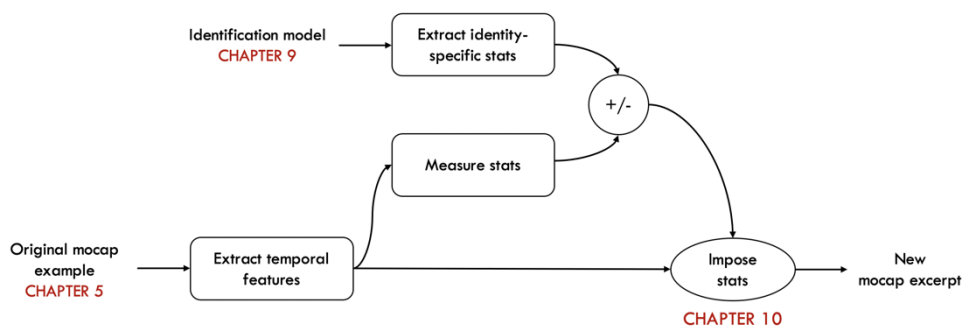


Figure 6 : Schéma des étapes de l'algorithme de synthèse pour le contrôle cinématique de l'identité.



Publications associées :

F. Bigand, A. Braffort, E. Prigent, B. Berret (2019), Signing Avatar Motion: Combining Naturality and Anonymity, 6th International Workshop on Sign Language Translation and Avatar Technology (SLTAT 2019), Hambourg, Germany.

F. Bigand, E. Prigent, A. Braffort (2019), Animating virtual signers: the issue of gestural anonymization, 19th International Conference on Intelligent Virtual Agents (IVA 2019), Paris, France.

F. Bigand, E. Prigent, A. Braffort (2019), Retrieving Human Traits from Gesture in Sign Language: The Example of Gestural Identity, 6th International Symposium on Movement and Computing (MOCO 2019), Tempe, United States.

F. Bigand, E. Prigent, A. Braffort, B. Berret (2019), Animer un signeur virtuel par mocap : le problème de l'identité gestuelle, 5è Journée de la Fédération Demeny-Vaucanson (FéDeV 2019), Palaiseau, France.

F. Bigand, E. Prigent, A. Braffort (2020), Person Identification Based On Sign Language Motion: Insights From Human Perception And Computational Modeling, 7th International Conference on Movement and Computing (MOCO'20).

F. Bigand, E. Prigent, B. Berret, A. Braffort. How Fast Is Sign Language? A Reevaluation of the Kinematic Bandwidth Using Motion Capture. 29th European Signal Processing Conference (EUSIPCO 2021), 2021

F. Bigand, E. Prigent, B. Berret, A. Braffort. Machine Learning of Motion Statistics Reveals the Kinematic Signature of the Identity of a Person in Sign Language. *Frontiers in Bioengineering and Biotechnology*, Frontiers, 2021, 9,

F. Bigand, E. Prigent, B. Berret, A. Braffort. Decomposing spontaneous sign language into elementary movements: A principal component analysis-based approach. *PLoS ONE*, Public Library of Science, 2021, 16 (10),

Tâche 2.3 : Alignement, annotation et analyse des données

Cette tâche consiste en la création de corpus suite aux tâches 2.1 et 2.2.

#	Échéance	Intitulé livrable	Type	Responsable	État
2.3.1	T0+12	Corpus existants alignés, annotés et validés – lot1	Corpus	FTVS	Livré P1
2.3.2	T0+24	Corpus existants alignés, annotés et validés – lot2	Corpus	FTVS	Livré P2
2.3.3	T0+12	Nouveau corpus alignés, annotés et validés – lot1	Corpus	MOCAPLAB	Livré P2
2.3.4	T0+27	Nouveau corpus alignés, annotés et validés – lot2	Corpus	MOCAPLAB	Livré P3





Avancées

Sous-titrage : Corpus existants alignés (2.3.1 & 2.3.2)

Pour la mise en place d'une chaîne de production automatique de sous-titres, les données sont passées par deux premières étapes de mise à disposition et traitement :

1. Mise à disposition par france.tv access de 17 953 émissions et sous-titres de qualité humaine créés par les équipes de france.tv access sur une plateforme dédiée et sécurisée nommée ATTESOR
2. Transcription par le LISN de la bande son de 3 984 émissions par un outil de reconnaissance vocale, aidée par le texte de référence des sous-titres, et segmentation et alignement de la transcription avec les sous-titres de référence au niveau de la phrase

Les données obtenues sont alors en quantité et qualité suffisante pour l'entraînement de modèles :

- de génération de sous-titres en français (transcriptions > sous-titres)
- ou de bruitage correspondant à la reconnaissance vocale (sous-titres > transcriptions), permettant d'étendre l'apprentissage à d'autres données.

Afin de tester les performances de ces modèles, ont aussi été partagés durant le projet :

- 20 h de transcription manuelle commandées en sous-traitance, pour la comparaison avec la transcription automatique
- Des fichiers pouvant servir de jeux de test dans divers domaines : MOOCs, séries, JT.

Motion capture : Nouveaux corpus annotés (2.3.3 & 2.3.4)

Pour la réalisation du système de traduction texte/LSF, un nouveau corpus nommé ROSETTA-LSF a été créé. Nous avons enregistré les productions en LSF d'une personne sourde sélectionnée sur la base de son expérience dans la production d'actualités en LSF pour les médias. Elle est la directrice de Media'Pi! (<https://media-pi.fr/>), un site d'information indépendant, bilingue en français et en LSF, et produit chaque semaine des contenus journalistiques pour le site. Deux journées de captations ont été réalisées, avec différentes tâches, en particulier la traduction de titres d'actualités. Nous avons choisi ce type d'énoncé comme étude de cas pour notre projet, car il s'agit d'un contenu journalistique, qui peut donc traiter de n'importe quel sujet, et qui est de taille variable mais jamais très long (30 mots maximum).

Tournage 1 (le 05/09/19) :

- Traduction en LSF de 114 titres d'actualités extraits de brèves AFP
- Description en LSF de 29 photos
- Répétition de 18 vidéos LSF
- 800 signes isolés

Tournage 2 (le 27/10/20) :

- Traduction en LSF de 50 titres d'actualités extraits de brèves AFP
- Description du projet Rosetta (basée sur une traduction LSF réalisée en amont)
- Traduction de phrases déjà générés via l'animation d'avatar pour comparaison
- Traduction en LSF de 30 phrases avec des structures proches de celles réalisées lors du 1er tournage.
- Traduction de 13 phrases issues de contenu non journalistique 400 signes isolés

Cela représente en tout, 5h45 min de données 3D enregistrées et 51 minutes de rendu vidéos 2D réalisés par MocapLab pour annotation, expériences, évaluation ou pour démonstration.

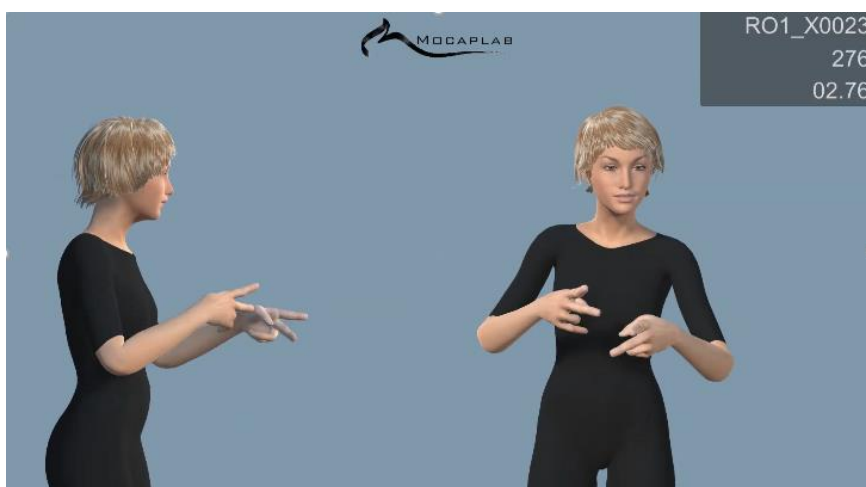


Ce corpus a été annoté en sous-traitance auprès d'un expert en vue de son utilisation pour la traduction texte/LSF. Le schéma d'annotation a été conçu pour permettre l'extraction d'éléments pertinents pour le processus de génération. De manière classique, les unités manuelles (c'est-à-dire l'activité des mains et des bras) ont été segmentées sur la ligne de temps. Ensuite, chaque unité a été annotée sur deux pistes (droite et gauche si nécessaire, la locutrice enregistrée étant droitère) avec un attribut qui permet de récupérer les éléments à réutiliser dans la base de données de *mocap* et un autre qui décrit la relation entre les deux mains, le cas échéant. A ce jour, les 194 phrases sur la thématique journalistique ont été annotées.

L'ensemble des vidéos sera mis à disposition pour la recherche sur le site Ortolang (<https://www.ortolang.fr/>) avec les annotations réalisées.

Livraison des vidéos d'avatar signant issues du 1er tournage :

Exemple : « " Gilets jaunes " : Avignon se barricade et interdit la manifestation de samedi »



Livraison des vidéos d'avatar signant issues du 2e tournage :

Exemple : « Une mesure inédite annoncée par le président Macron. »





5.3 SP 3 : Module de production de sous-titres adaptés

Objectif

L'objectif de ce lot est de produire des modules de production de sous-titres adaptés au public des sourds et malentendants. Ce lot s'appuiera sur les corpus acquis et développés dans le lot 2 pour entraîner des modèles neuronaux de production de sous-titres adaptés.

Date de début : T0		Date de fin : T0+33				
Responsable : SYSTRAN		Participants : FTVS, LIMSI, LUTIN				
	SYSTRAN	FTVS	MOCAPLAB	LIMSI	LUTIN	
Charge (h*m)	22	6	0	34	7	

Tâches et livrables

Tâche 3.1 : modèles neuronaux de production de sous-titres adaptés par domaine

Cette tâche consiste à expérimenter sur les architectures de réseaux de neurones permettant une bonne gestion du flux vidéo et du flux textuel et de réaliser les entraînements successifs des modèles de production de sous titres.

La première étape sera de mettre en place le cadre d'entraînement général, et en s'aidant des critères d'évaluation définis en 3.2, la recherche de meilleurs hyper-paramètres et configuration des réseaux de neurones sera effectuée.

Tâche 3.2 : Evaluation des modèles

#	Échéance	Intitulé livrable	Type	Responsable	État
3.1.1	T0+6	Préconisations utilisateurs pour le sous-titrage	Document	SYSTRAN	Remplacé par les rapports 7.1.1.
3.1.2	T0+30	Modèles neuronaux de production sous-titres adaptés aux journaux télévisés	Modèle	SYSTRAN	Livré P3
3.1.3	T0+30	Modèles neuronaux de production sous-titres adaptés aux débats documentaires	Modèle	SYSTRAN	Livré P3
3.1.4	T0+30	Modèles neuronaux de production sous-titres adaptés aux séries et films	Modèle	SYSTRAN	Livré P3
3.2.1	T0+33	Rapport d'évaluation des modèles de sous-titrage simplifié	Document	SYSTRAN	Livré P3

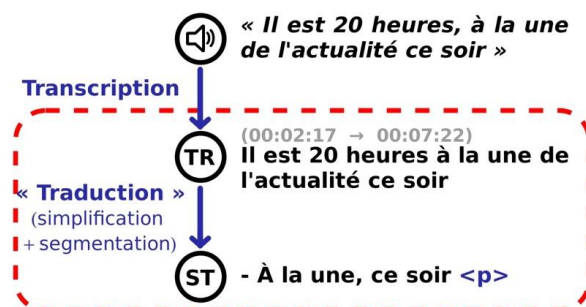


Avancées

La production de sous-titres monolingues destinés en particulier à les rendre accessibles au public sourd ou malentendant est devenue une obligation légale pour toutes les émissions télédiffusées depuis 2005, conduisant à une augmentation considérable du nombre d'heures sous-titrées. Cette étape est souvent la première marche vers la production de sous-titres en langue étrangère, qui permettront de démultiplier l'audience des émissions francophones. La production de sous-titres intervient souvent en bout de chaîne du processus de diffusion, et est principalement réalisée selon deux modalités très différentes : le sous-titrage en direct, pour les journaux d'information, les émissions de plateau ou les événements retransmis en direct (par exemple, les émissions politiques ou les retransmissions de compétitions sportives) ; le sous-titrage en différé pour les émissions de jeux, les documentaires et les fictions. Dans le premier cas, des contraintes de temps réels sont critiques, et le sous-titreur doit s'adapter à la spontanéité des prises de parole et plus généralement aux aléas du direct ; dans le second cas, il faut potentiellement faire face à une plus grande variété des événements sonores à traiter et retranscrire : chansons, rires, bruits d'ambiance, interventions en langue étrangère, etc.



Dans cette étude, nous examinons la possibilité de concevoir des chaînes de traitement automatisées des flux audio d'émissions télévisuelles en nous appuyant sur des méthodes modernes d'intelligence artificielle (IA), dont les avancées récentes ont permis d'améliorer considérablement la qualité d'applications telles que la transcription vocale et la traduction automatique. Deux questions sont principalement abordées : (a) est-il possible de parvenir à automatiser intégralement la production de sous-titres pour les émissions télévisuelles ? (b) La réponse est-elle la même pour tous les types et tous les genres d'émissions, ou bien certains sont-ils plus difficiles à traiter que d'autres ?





Pour essayer de répondre à ces questions, nous avons développé des algorithmes et des systèmes de sous-titrage entièrement automatisés. Partant d'une retranscription automatique de la bande son sous une forme textuelle, ces algorithmes utilisent des méthodes inspirées de la traduction automatique pour compresser les énoncés et calculer leur disposition à l'écran, afin de garantir qu'ils respecteront les standards les plus stricts d'affichage et de lisibilité. Dans un second temps, ces algorithmes ont été spécialisés et adaptés aux différents types d'émission, qui correspondent à des usages et à des publics variés : on peut vouloir regarder une émission pour s'informer, pour se divertir, ou encore pour apprendre de nouveaux savoir-faire, et les sous-titres ne remplissent pas exactement le même rôle dans ces trois situations.



- TR** Suite à votre passage dans l'émission comment les élèves à l'époque **avait réagi**.
- ST** A votre passage dans l'émission, **
** comment les élèves **avaient réagi** ? **<p>**

Des évaluations systématiques de ces algorithmes ont été mises en place ; d'un côté sur la base d'un jeu de test incluant une diversité d'émissions ; d'un autre côté sur la base d'une analyse par les spécialistes de france.tv access. Ces évaluations ont mis en évidence la faisabilité de systèmes de sous-titrage fonctionnant de bout en bout et adaptés aux types d'émissions, mais également les limites des systèmes actuels : si le niveau de qualité s'approche, pour certaines émissions de direct, du niveau d'exigence souhaité, la qualité moyenne reste très en deçà de la cible, et est inexploitable en particulier pour de nombreuses émissions sous-titrées en différées. Une raison de cette qualité médiocre provient des erreurs de retranscription, qui restent trop élevées ; une autre raison provient des systèmes de sous-titrage eux-mêmes, qui échouent à distinguer dans le flux audio les éléments de signification qui doivent être impérativement préservés.

Publications associées :

[F. Buet \(2020\), Analyse de la régulation de la longueur dans un système neuronal de compression de phrase : une étude du modèle LenInIt, Actes de la conférence des jeunes chercheurs en traitement automatique des langues \(RECITAL 2020\).](#)

[François Buet, François Yvon. Toward Genre Adapted Closed Captioning. Interspeech 2021, Aug 2021, Brno \(virtual\), Czech Republic. pp.4403-4407.](#)

[François Buet, François Yvon. Vers la production automatique de sous-titres adaptés à l'affichage. Traitement Automatique des Langues Naturelles, 2021, Lille, France. pp.91-104.](#)



5.4 SP 4 : Module de traduction multilingue de sous-titres adaptés

Objectif

Ce lot a pour objectif de produire des sous-titres multilingues en plusieurs langues dont l'anglais, l'espagnol, le chinois, l'italien et le portugais.

Date de début : T0		Date de fin : T0+33			
Responsable : SYSTRAN		Participants : FTVS, LUTIN			
	SYSTRAN	FTVS	MOCAPLAB	LIMSI	LUTIN
Charge (h*m)	22	6	0	0	6

Tâches et livrables sur la période

Tâche 4.1 : Production de modèles de sous titres adaptés - EN, ES, IT, PT, ZH

L'objectif de la tâche est de produire les modèles pour les 2 options envisagées :

- des modèles de traduction spécialisés pour les sous-titres adaptés
- des modèles de production de sous titres adaptés en langues étrangères (Easy-To-Read)

Tâche 4.2 : Evaluation des modèles

#	Échéance	Intitulé livrable	Type	Responsable	État
4.1.1	T0+12	Préconisations des utilisateurs pour le sous-titrage adapté multilingue	Document	SYSTRAN	Remplacé par les rapports 7.1.1.
4.1.2	T0+35	Rapport sur le développement des modèles de sous-titrages adaptés en anglais, espagnol, italien, portugais et chinois	Document	SYSTRAN	Fusionné avec 4.2.1
4.2.1	T0+37	Rapport de développement et d'évaluation	Document	SYSTRAN	Livré P3

Avancées

SYSTRAN a travaillé sur **2 variations de la tâche** de génération de sous-titres automatiques multilingues :

- En partant du sous-titre de référence français de france.tv access (traduction de sous-titres, nécessite que la vidéo soit déjà sous-titrée dans une langue de départ)
- En partant de la transcription alignée du LISN (génération de sous-titres à partir de la bande son, peut concerner n'importe quelle vidéo)



3 générations de modèles ont été développées et évaluées au fil du projet :

V0 - Août 2020 - Preuve de concept - Modèle de traduction de sous-titres

- Système unique multilingue, du français vers toutes les autres langues
- Quantité de données d'entraînement : 135 millions de lignes parallèles (0.74% annotées avec une segmentation propre aux sous-titres)
- Prend comme source le sous-titre français provenant de france.tv access
- Intégration : développement d'un premier code, du ttml source au ttml cible
- Evaluation : interne (toutes langues)

V1 – Avril 2021 - Modèle de génération de sous-titres à partir de la reconnaissance vocale

- Système unique multilingue, du français et anglais vers toutes les autres langues
- Quantité de données d'entraînement : 14,3 millions de lignes parallèles (20% annotées avec une segmentation propre aux sous-titres)
- Prend comme source la reconnaissance vocale segmentée générée par le LISN : tâche plus intéressante car permet de générer des sous-titres pour une émission qui n'en aurait pas d'existants
- Intégration : ajout des couleurs signalant le type de prise de parole (personne à l'écran, hors écran, voix off, langue étrangère, chanson...)
- Evaluation : interne (toutes langues), métier (anglais, espagnol), utilisateurs (espagnol)

V2 – Juillet 2021 - Modèle de traduction de sous-titres aidée par le contexte

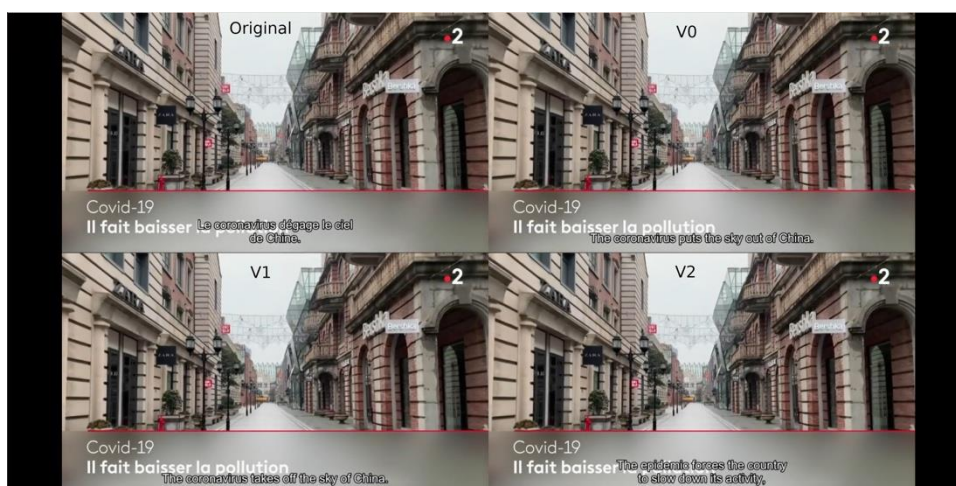
- Système monolingue, du français vers anglais, standard en traduction automatique
- Quantité de données d'entraînement : 1,2 million de lignes parallèles (22% annotées avec une segmentation propre aux sous-titres, 79% contextualisées avec la source et la cible précédentes concaténées à l'exemple présent)
- Prend comme source le sous-titre français provenant de france.tv access, pour évaluer la qualité de la traduction et segmentation sans effets de bord introduits par la reconnaissance vocale
- Prise en compte du contexte précédent en source et cible
- Intégration : amélioration des couleurs et introduction des entêtes (tirets, nom du locuteur) provenant du français, *timecodes* ajustés pour éviter des sous-titres trop courts (durée minimale 19 images = 760ms) ou trop rapprochés (intervalle minimal 8 images = 320ms). Les *timecodes* utilisés sont ceux des sous-titres français pour toutes les émissions de stock, et ceux de la reconnaissance vocale - plus proches du temps réel - pour toutes les émissions en direct.
- Evaluation : interne (anglais), métier (anglais), utilisateurs (anglais)



Systèmes de sous-titrage V0, V1, V2 en action sur un MOOC (apprendre)



Systèmes de sous-titrage V0, V1, V2 en action sur une série (se divertir)



Systèmes de sous-titrage V0, V1, V2 en action sur un journal TV (s'informer)

Les évaluations internes et métier, complétées par celles des utilisateurs finaux de sous-titres menées par le LUTIN et HC, concordent à montrer des progrès mesurables au fil du projet, notamment en anglais entre la première version du système (développement en avril 2021, évaluation en mai-juillet 2021) et la seconde version du système (développement en juillet 2021, évaluation en août-novembre 2021), et en particulier dans les émissions de stock tels que les documentaires et



magazines. **Les avancées sont perceptibles tant sur le contenu que son intégration en sous-titres** (synchronisation, affichage des couleurs, locuteurs), qui a une forte importance pour l'expérience utilisateur.

Au vu de ses critères de qualité, **france.tv access conclut que les sorties des modèles automatiques développés ne sont pas directement utilisables pour une diffusion à l'antenne. En revanche, l'utilité de ces sorties automatiques reste considérée par SYSTRAN dans d'autres contextes** – contextes où les moyens humains spécialisés sont moins importants pour faire face au besoin de traduction, ce qui a pour résultat actuel le non-sous-titrage et la non-accessibilité du contenu - et dans d'autres chaînes de production (avec post-édition, formation, information sur l'émission, etc.).

De plus, le travail mené sur ce thème au long du projet a révélé des **axes de recherche** à étudier pour faire passer les sous-titres automatiques à un niveau supérieur :

- Penser la traduction au niveau de l'émission (voire de la saison, la série)
- Séparer le son (musiques, *jingles*, bruits de fond) de la parole dans le cas de séries, jeux, dessins animés, si cela reste impossible d'accéder à la version avant mixage
- Détecter des informations dans l'image (positionnement, couleurs, *timecodes*, plans)

Tandis que des **pistes d'amélioration** se dessinent pour le montage de futurs projets :

- Trouver des façons d'avoir accès à plus de données multilingues de qualité professionnelle
- Accompagner les résultats automatiques d'une interface pour une interaction avec l'humain facilitée
- Etablir une juste comparaison avec d'autres outils automatiques du marché et avec les méthodes existantes



5.5 SP 5 : Module de génération de LSF

Objectif

Ce lot a pour objectif de produire un module de sous-titrages en LSF sous forme d'animations 3D.

Il consiste à élaborer une représentation de la LSF, deux systèmes de traduction et de génération d'animations, qui utiliseront cette représentation, ainsi qu'un *player* 3D qui permettra d'afficher le contenu en LSF par l'animation d'un signeur virtuel (avatar). Ce lot sera guidé par les recommandations des utilisateurs.

La "représentation-LSF" sera conçue à partir de l'analyse des corpus constitués dans le lot 2. Elle servira de pont entre le système de traduction et le système de génération. L'intégration des deux systèmes et du *player* dans un module servira à la réalisation du démonstrateur du lot 8.

Date de début : T0		Date de fin : T0+36				
Responsable : LIMSI		Participants : Tous				
	SYSTRAN	FTVS	MOCAPLAB	LIMSI	LUTIN	
Charge (h*m)	22	6	12	42	12	

Tâches et livrables sur la période

Tâche 5.1 : Cahier des charges

La première tâche consiste à élaborer le cahier des charges et les besoins fonctionnels pour ce module, en prenant en compte les recommandations des futurs utilisateurs.

Tâche 5.2 : Conception et développement

Cette tâche consiste en la conception et le développement :

- d'une représentation intermédiaire entre le système d'apprentissage automatique et le système de génération d'animations
- d'un système de traduction français / représentation LSF
- d'un système de génération d'animations à partir de la représentation LSF

Tâche 5.3 : Intégration et visualisation

#	Échéance	Intitulé livrable	Type	Responsable	État
5.1.1	T0+6	Cahier des charges du module	Document	LIMSI	Livré P1
5.1.2	T0+9	Recommandations utilisateurs pour l'animation LSF	Document	LIMSI	Remplacé par les rapports 7.1.1.
5.2.1	T0+12	Version 1 de la représentation et des deux systèmes	Logiciel	LIMSI	Livré P1



5.2.2	T0+27	Version 2 de la représentation et des deux systèmes	Logiciel	LIMSI	Livré P2
5.2.3	T0+39	Version Finale de la représentation et des deux systèmes	Logiciel	LIMSI	Livré P3
5.3.1	T0+39	Intégration des deux systèmes et restitution des animations dans un <i>player</i> 3D	Logiciel	LIMSI	Livré P3
5.3.2	T0+39	Rapport de développement des systèmes pour la génération automatique d'animations 3D en LSF à partir du français	Document	LIMSI	Livré P3

Avancées

L'une des spécificités de la traduction depuis un texte vers une langue des signes (LS) est la différence de canal. En effet, les LS n'ont pas de forme écrite et il faut donc produire du contenu sous forme visuo-gestuelle. La solution consiste à passer par une représentation intermédiaire, et procéder en deux étapes :

1. Traduire le texte en français sous cette représentation intermédiaire
2. Utiliser cette représentation comme entrée pour un système d'animation qui permet d'afficher un signeur virtuel animé

Les choix retenus pour le projet ont été les suivants :

- L'utilisation d'AZee, développé au LISN, comme représentation intermédiaire, représentation qui a déjà fait ses preuves pour la génération et qui tire pleinement parti des correspondances entre message et mouvement contenues dans les données. En étant basée sur la langue d'arrivée, la LSF, et non la langue de départ, le français, cette représentation est aussi plus proche d'une formulation naturelle en langue cible, ce qui facilite la compréhension du message et la génération.
- Une traduction du texte vers AZee par analogie, en reconstituant de nouvelles phrases à partir d'exemples disponibles, qui nous permet de pointer précisément dans le format cible (la représentation intermédiaire) ce qui doit être extrait et combiné : l'enjeu devient alors l'identification dans les exemples disponibles des constituants correspondants à la phrase que l'on souhaite traduire, la traduction des nouveaux constituants et la combinaison des résultats.
- La création d'un corpus parallèle français/LSF enregistré à l'aide d'un système de capture de mouvement par Mocaplab et qui doit permettre de générer une animation à partir de segments d'animations présents dans le corpus.

ROSETTA-LSF : corpus bilingue français/LSF

Pour traduire du texte vers AZee, il nous faut une banque d'alignements entre chaque segment de texte et des expressions AZee, chacune représentant par définition une traduction possible pour le segment textuel. Puisque les expressions AZee sont ensuite utilisées pour animer l'avatar à partir de données *mocap*, il nous faut également un alignement entre les données enregistrées sur de vraies productions de référence et les expressions AZee qui les représentent. Cela nous a amené à créer le



corpus de *mocap* ROSETTA-LSF et à réaliser des alignements français/AZee/LSF afin de pouvoir travailler sur les deux fronts, de part et d'autre de ce pivot qu'est AZee.

Afin de pouvoir constituer une banque de données *mocap* réutilisables et combinables, la 1ère étape a consisté à définir un schéma d'annotation du corpus qui permet d'ajouter des indications nécessaires et suffisantes sur les aspects linguistiques pour pouvoir à la fois sélectionner les segments utiles et les adapter dans un nouveau contexte pour la génération.

La constitution et l'annotation de ce corpus ont été décrits dans le lot 2 (tâches **2.3.3** et **2.3.4**).

Alignements avec français/AZee et AZee/LSF

Le travail suivant a consisté à réaliser les alignements texte/AZee/LSF. Le système AZee préexistait au projet ROSETTA. Le projet nous a permis de l'ajuster et de le stabiliser, par un travail d'alignement AZee/LSF, dans un premier temps sur un corpus préexistant (corpus 40 brèves), puis sur le corpus ROSETTA-LSF quand il a été disponible. Cet alignement a consisté à produire des expressions AZee (AZeefication) pour chaque vidéo de LSF. Il s'agit de repérer les règles de production AZee et de les imbriquer entre elles, par l'observation des formes (impliquant des articulateurs manuels et/ou non-manuels) produites par le signeur et l'interprétation du sens qui leur est associé. Ce travail, réalisé à la main par des experts en linguistique de la LSF, a nécessité une vérification constante à tous les niveaux de l'expression : arbres et sous-arbres AZee, règles de production. Une banque de données de 180 expressions AZee de discours a ainsi été constituée sur le corpus ROSETTA-LSF.

Ensuite, nous avons réalisé des alignements AZee/texte par la recherche de correspondances entre tout ou partie d'une expression AZee et tout ou partie du texte. Chaque expression AZee correspondant à un texte, de premiers alignements AZee-texte sont ainsi formés. Étant donné que chaque expression AZee est composée de sous-expressions et que chaque titre d'actualité peut être décomposé en plusieurs segments, d'autres alignements viennent compléter la liste par affinage successif. Dans le cas des segments impliqués dans les phrases à traduire retenues, un dernier alignement manuel des expressions AZee avec les *timecodes* des enregistrements de mocap a été effectué pour la manipulation et combinaison effectives des blocs de mouvement.

Modules de traduction texte/AZee et génération AZee/LSF

Finalement, nous avons réalisé les modules de traitement automatique nous permettant de traduire une phrase en français écrit en LSF via l'animation de signeur virtuel.

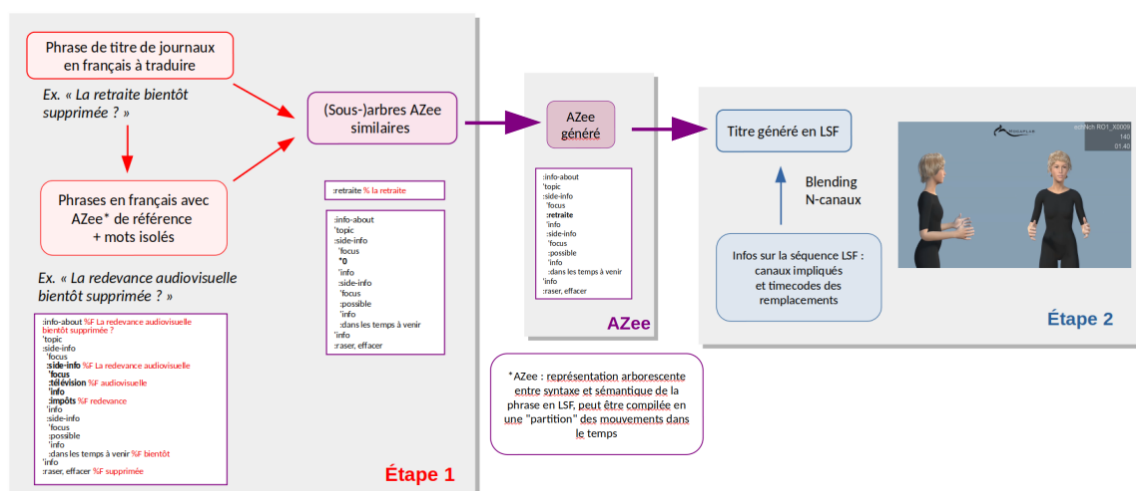
Pour une phrase à traduire (dans notre cas un titre de journal), nous cherchons à associer un ensemble de traductions possibles en description AZee, en procédant par recherche d'analogies dans notre corpus d'alignements. L'approche retenue est de procéder en tentatives successives sur trois niveaux différents : traduction 1) par match exact, 2) par anti-match, 3) par partition. Chaque niveau produit des traductions de qualité inférieure au niveau précédent, mais ce niveau n'est tenté que si le précédent n'a pas abouti. Cette première étape nous permet d'obtenir une description de la phrase en AZee, qui est ensuite utilisée pour produire le message en LSF.

Au niveau de la génération de l'animation, la structure récursive de AZee implique la formation de blocs de mouvement, chaque bloc étant éventuellement composé de sous-blocs définis par les sous-expressions AZee. Le principe général de la méthodologie de combinaison, appelée "multi-canaux", repose sur le fait de gérer le mouvement de manière synchrone sur plusieurs canaux d'animation. Chaque canal correspondant à un ensemble d'os représentant des effecteurs tels que les mouvements



du bras droit, du bras gauche, du tronc, de la tête et du reste du corps. Ainsi, la génération de nouvelles animations peut être constituée de segments incluant tous les éléments corporels ou de segments ne portant que sur un ou plusieurs éléments (par exemple uniquement la tête).

Le schéma suivant illustre l'approche retenue.



Par exemple, pour générer la traduction en LSF de la phrase :

Alsace : de grands chefs ont vendu leur vaisselle pour les plus modestes dans la banlieue de Gerstheim

Le système a utilisé les trois phrases suivantes présentes dans le corpus ROSETTA-LSF :

- Samedi 30 et dimanche 31 mars, de grands chefs ont vendu leur vaisselle en Alsace, à Gerstheim
- Moins de TVA pour les plus modestes : "Il ne faut pas traiter ça par le mépris", lance Xavier Bertrand au gouvernement
- Le superéthanol n'est proposé que dans 1 000 stations-service en France, comme ici dans la banlieue de Bordeaux

Et a généré une description en AZee utilisée elle-même pour générer une traduction en LSF via l'animation du signeur virtuel.

Résultats et perspectives

L'ensemble des algorithmes a été implémenté et testé sur des exemples représentatifs afin de disposer d'une preuve de concept de l'approche retenue. Il s'agit de 7 titres non présents dans le corpus, qui ont été sélectionnés pour leur représentativité des phénomènes à prendre en compte dans le cadre de ce projet.

Les recherches sur le traitement automatique des langues des signes est très récent : au début du projet, il n'existait pas de système de traduction automatique français/LSF, et s'il existait des systèmes de générations d'animations en LSF, ils étaient généralement basés sur la concaténation d'animations "monocanal", cela limitant la qualité linguistique des productions ainsi que leur réalisme physiologique.

Côté traduction, nous avons expérimenté une approche de traduction par analogie, avec une représentation intermédiaire, AZee, qui permet une bien meilleure précision sur les aspects linguistiques. Côté génération d'animations, nous avons expérimenté une approche "multicanaux" qui



permet de bénéficier de cette meilleure précision linguistique et de produire des animations plus réalistes. Les évaluations utilisateurs ont attesté la supériorité de cette méthode par rapport aux préexistantes, en montrant de meilleurs scores sur les aspects effort cognitif, compréhension et satisfaction.

Ainsi, s'il reste encore beaucoup de travail de recherche et développement à mener avant d'aboutir à un produit commercialisable, nous avons pu réaliser une preuve de concept opérationnelle qui nous a permis de valider nos choix et nous offre des perspectives de progression pour les prochaines années.



5.6 SP 6 : Évaluation des modules

Objectif

Ce lot a pour objectif d'évaluer la qualité des modules de sous-titrages multilingues et LSF.

La validation des modules de sous-titrages multilingues et LSF de ROSETTA sera faite auprès des utilisateurs concernés à partir :

- de la mesure de leur degré d'utilité et d'utilisabilité pour chacune des déficiences,
- de la mesure du contenu auquel l'utilisateur a pu accéder (à quel point l'information a été interprétée et comprise)
- de la mesure du degré de satisfaction des utilisateurs (quantifiée sur une échelle) en considérant les différentes dimensions de l'usage de ROSETTA et les différents types de contenus,
- de la comparaison avec un interprète humain pour les langues étrangères et en langue des signes.

Pour cela, un protocole classique sera utilisé qui consiste à évaluer l'apport de la nouvelle technique par rapport à l'existant.

En amont et conjointement aux validations des utilisateurs, nous procéderons, par itérations, à diverses validations techniques en laboratoire et in situ sur différentes chaînes, de bout-en bout allant de la captation *live* et/ou de services *catch-up*, jusqu'à la réception sur divers périphériques et cela à travers différents types de réseaux de distribution.

Date de début : T0		Date de fin : T0+42				
Responsable : LUTIN		Participants : Tous				
	SYSTRAN	FTVS	MOCAPLAB	LIMSI	LUTIN	
Charge (h*m)	4	2	2	2	12	

Tâches et livrables sur la période

Tâche 6.1 : Spécifications

Cette tâche consiste en la spécification des évaluations et des protocoles utilisés.

#	Échéance	Intitulé livrable	Type	Responsable	État
6.1.1	T0+3	Spécifications des évaluations	Document	LUTIN	Livré P1

Avancées

Pour préparer l'évaluation de la qualité des modules de sous-titrages multilingues et LSF, le LUTIN a rassemblé les différentes méthodes susceptibles de permettre l'évaluation et le recueil de recommandations des utilisateurs pour des modules de sous-titrage et de LSF. Nous avons distingué, d'une part, des méthodes qualitatives axées autour de la « verbalisation » des recommandations utilisateurs et, d'autre part, des méthodes dynamiques portées sur des expérimentations dans l'usage.



En prenant en compte les spécificités liées à chaque catégorie d'utilisateurs, le livrable 6.1.1 vise la spécification des différentes techniques d'évaluation, susceptibles d'être utilisées dans le cadre du projet ROSETTA afin de mesurer l'utilisabilité, l'utilité, l'acceptabilité et l'accessibilité des prototypes produits.

Tâche 6.2 : Production des corpus et outils d'évaluation

La tâche 6.2 spécifie les méthodes présentées en 6.1.1 avec les objectifs de présenter :

- les techniques d'évaluation développées par le Lutin-Userlab,
- les outils qui ont été développés à partir des techniques pour correspondre aux méthodes,
- les procédures d'évaluation qui utilisent ces outils.

Des corpus de référence source et de corpus à évaluer sont constitués pour instancier les procédures et outils relatifs aux méthodes.

#	Échéance	Intitulé livrable	Type	Responsable	État
6.2.1	T0+6	Méthodes d'évaluation : outils et procédure. Livraison intermédiaire	Outils / Procédures	LUTIN	Livré P1
6.2.2	T0+12	Méthodes d'évaluation : outils et procédure. Livraison intermédiaire	Outils / Procédures	LUTIN	Livré P2
6.2.3	T0+21	Méthodes d'évaluation : outils et procédure. Livraison intermédiaire - Méthodes de <i>focus group</i>	Outils / Procédures	LUTIN	Livré P3
6.2.4	T0+27	Méthodes d'évaluation : outils et procédure. Livraison intermédiaire - Méthodes d'évaluation en ligne Partie 1 : Module LSF Partie 2 : Module sous-titrage	Outils / Procédures	LUTIN	Livré P3
6.2.5	T0+33	Méthodes d'évaluation : outils et procédure. Livraison intermédiaire - Méthodes pour sous-titres bruités	Outils / Procédures	LUTIN	Livré P3
6.2.6	T0+39	Méthodes d'évaluation : outils et procédure. Final - Synthèse	Outils / Procédures	LUTIN	Livré P3

Avancées

Les livrables 6.2. (1, 2, 3, 4, 5 et 6) correspondent aux méthodes / outils d'évaluations qui sont appliqués aux modules ST (sous-titrage) et LSF.

Le livrable **6.2.1** renvoie au travail mené sur le recueil des avis utilisateurs dans le cadre de brainstorming sur leur "idéal de l'accessibilité". Ce procédé a permis d'administrer une méthode d'évaluation des attentes utilisateurs. Dans une démarche centrée utilisateur, il était important d'entamer le travail d'évaluation par les besoins et les attentes utilisateurs pour chaque public ciblé par le projet. Le livrable 6.2.1 décrit la méthode des *brainstorming* et *focus groups* réalisés *in situ* au début du projet.



Le livrable **6.2.2** renvoie au travail mené sur la construction d'une grille d'évaluation experte avec les recommandations utilisateurs publiées dans la littérature (guides, chartes, normes). Le travail a eu pour but d'évaluer les recommandations existantes pour le sous-titrage et la LSF par des utilisateurs afin d'établir une liste priorisée de critères ergonomiques. La meilleure façon de procéder pour recueillir l'avis des utilisateurs cibles via des questionnaires à distance a fait l'objet d'une réflexion particulière. Ce travail a été réalisé dans l'esprit d'un livre blanc de recommandations pour le sous-titrage et la LSF dans les programmes audiovisuels.

Le livrable **6.2.3** a détaillé les méthodes spécifiques d'entretiens collectifs (*focus groups*) spécifiées dans le livrable 6.2.1. Ce livrable a eu pour objectif d'adapter ces méthodes au contexte des évaluations à distance, ainsi que leur adaptation aux publics cibles (Sourds, personnes âgées, anglophones, hispanophones et sinophones). Ces méthodes concernent dans un premier temps les axes d'évaluations suivants :

1. Évaluation ergonomique : utilité, utilisabilité, lisibilité, compréhensibilité et acceptabilité
2. Évaluation Satisfaction : selon les usages (exemple : Divertissement, Information, Apprentissage)
3. Évaluation Satisfaction : types de contenus (exemple : Divertissement, Journal, MOOC)

Ce livrable présente les corpus retenus pour les évaluations utilisateurs ainsi que les différents logiciels sélectionnés. En effet, la prévision de *focus groups* à distance a posé quelques défis méthodologiques. La sélection des outils adaptés a donc fait l'objet d'une veille scientifique approfondie des outils existants pour l'annotation simplifiée et la visioconférence.



Le livrable **6.2.4** a répondu aux mêmes objectifs méthodologiques que le 6.2.3 par l'administration de méthodes quantitatives complémentaires. En revanche les axes ont porté sur :

4. Évaluation Compréhensibilité ST : lecture et compréhension
5. Évaluation Compréhensibilité LSF : décodage et compréhension
6. Évaluation fiabilité des modules ST et LSF : comparaison avec l'existant

La mise à disposition des premiers prototypes ont permis de définir des échelles ergonomiques spécifiques aux modules et le matériel à évaluer.



Sur une échelle de type Likert à 6 points (1=pas du tout d'accord, 6= tout à fait d'accord), nous avons évalué 7 composantes ergonomiques :

- L'effort cognitif *Est-ce que lire les ST est fatiguant ?*
- La compréhension *Est-ce que le message a été compris ?*
- La comprenabilité *Est-ce que le message était clair ?*
- L'acceptabilité *Est-ce que le ST-LSF est acceptable ?*
- L'utilisabilité *Est-ce que le ST/LSF est utilisable ?*
- L'utilité *Est-ce que le ST/LSF est utile ?*
- La satisfaction *Est-ce que le ST /LSF est satisfaisant ?*

L'analyse des données quantitatives recueillies vise à vérifier, infirmer ou consolider les résultats des *focus groups* menés lors des premières phases de tests. La méthodologie de l'évaluation est celle des dimensions de l'expérience utilisateur que sont l'utilisabilité, la comprenabilité, l'acceptabilité, la satisfaction et l'utilité, spécifiées pour le module de sous-titrage. L'outil de recueil de cette évaluation est un questionnaire en ligne spécialement conçu pour l'interface ROSETTA. Il correspond aux questionnaires d'utilisabilité standards éprouvés dans le cadre d'évaluations des interfaces homme-machine.

Pour le sous-titrage, la méthode a permis d'évaluer les différences d'appréciations utilisateurs selon :

- Le type de contenu évalué (journalistique ou non journalistique)
- Le mode de génération des sous-titres (Humain, ROSETTA et YouTube)

Pour la LSF, la méthode définie dans ce livrable a permis d'évaluer les différences d'appréciations utilisateurs selon :

- Le type de contenu évalué (journalistique ou non journalistique)
- Le mode de génération automatique (monocanal ou multicanaux)
- Le mode de présentation (Signeur Humain ou Signeur automatique)

Le livrable **6.2.5** a eu pour objectif de déterminer une méthode d'évaluation de la qualité du sous-titrage à partir du seuil d'accessibilité du contenu ; la qualité du sous-titrage pouvant être mesurée par sa résistance au bruit.

Pour cela, la méthode a consisté à recenser des programmes issus des chaînes du groupe France télévisions et à recueillir les vidéos des émissions. Ces vidéos sont alors réparties en cinq catégories, en fonction de leur type de contenu (codé de t1 à t5). Parmi ces types d'émissions, on trouve t1 : Multi-usages ; t2 : Journal télévisé ; t3 : Divertissement ; t4 : Apprentissage ; t5 : Indéterminé (émission de débats « C dans l'air »). 3 juges ont évalué le degré de lisibilité, d'acceptabilité, de compréhension et de rappel du contenu vidéo selon le niveau de bruit en aveugle.

Le livrable **6.2.6** enfin livre une synthèse des différentes méthodes présentées dans les livrables précédents.



Tâche 6.3 : Évaluations

Des évaluations techniques ont été réalisées périodiquement, et une évaluation humaine a été réalisée en fin de projet.

#	Échéance	Intitulé livrable	Type	Responsable	État
6.3.1	T0+12	Résultats évaluation technique-1 - Recommandations	Document	LUTIN	Livré P3
6.3.2	T0+27	Résultats évaluation technique-2 – Questionnaires en ligne	Document	LUTIN	Livré P3
6.3.3	T0+39	Résultats évaluation technique-3 – Sous-titres bruités	Document	LUTIN	Livré P3
6.3.4	T0+39	Résultats de l'évaluation humaine - Synthèse	Document	LUTIN	Livré P3

Avancées

Les livrables 6.3.1, 6.3.2 et 6.3.3 ont été produits en écho aux méthodes développées dans les livrables 6.2 puis appliquées aux prototypes.

Le livrable **6.3.1** présente les résultats obtenus sur la base des méthodes élaborées sur le 6.2.2, notamment la liste priorisée des recommandations établie sur la base des évaluations utilisateurs. Ce livrable présente les résultats pour chacun des trois groupes de recommandations-questionnaires. Ces résultats sont de deux types : d'une part, l'analyse des dimensions d'accessibilité privilégiées par chacun des publics interrogés et, d'autre part, l'établissement d'un coefficient pour chaque recommandation, au regard de son importance ; ce coefficient servant à pondérer les évaluations de futurs prototypes.

Les principaux résultats indiquent que, pour les professionnels du sous-titrage, la dimension rythme des sous-titres (synchronisations, césures...) prime sur la qualité de la langue (orthographe, grammaire, ...) tandis que, pour les utilisateurs tout-venant, - qui sont les utilisateurs finaux-, c'est la qualité de la langue qui est considérée comme la dimension la plus importante ; ceci quel que soit l'âge. Toutefois, les seniors montrent une attention plus prononcée à l'ergonomie des sous-titres (taille des sous-titres, contrastes, polices, couleurs...) qu'à l'accessibilité de l'interface (personnalisation, synchronisation...) ; ceci contrairement aux moins de 65 ans qui montrent une préférence inverse, probablement due à une différence de préoccupation et de besoins en matière d'accessibilité. Par ailleurs, les pratiquants de la LSF estiment que la qualité de l'interprétation et la qualité de la langue LSF priment sur les autres recommandations d'accessibilité. L'ensemble de ces conclusions constitue des recommandations en faveur d'une accessibilité plurielle adaptée aux publics cibles.

Le livrable **6.3.2** fait écho aux résultats à obtenus par questionnaires en ligne sur la base des méthodes d'évaluations du 6.2.4.

Ce livrable résultat présente les différentes analyses menées sur les prototypes ROSETTA. Ces analyses ont investigué le module sous-titre et le module LSF. Les résultats ont concerné les réponses obtenues aux deux échelles ergonomiques élaborées dans le cadre du projet. Ces échelles ont été décrites dans le livrable méthodes 6.2.4 en deux parties et proposent d'évaluer 7 composantes ergonomiques qui favorisent l'utilisabilité d'un système. Chaque module a donc été évalué selon ces composantes par



des utilisateurs en ligne et les résultats ont été présentés. Deux objectifs principaux sont associés à ce livrable :

- Positionner ROSETTA du point de vue utilisateur par rapport à l'existant.
- Indiquer les composantes ergonomiques sur lesquelles ROSETTA peut être amélioré.

En phase 1, l'objectif a été d'évaluer l'échelle de mesure élaborée pour les évaluations utilisateurs en ligne. Les analyses ont révélé une bonne fiabilité de l'échelle de mesure à 13 items. Le second objectif a été pour le module LSF de distinguer les apports de la méthode multicanaux par rapport à d'autres méthodes de génération et également par rapport à un signeur virtuel exemplaire : l'humain. Les vidéos du ¹ ont présenté un score plus élevé dans toutes les qualités ergonomiques (effort cognitif, compréhension, compréhensibilité, acceptabilité, utilisabilité, satisfaction et utilité) comparé aux méthodes de génération automatiques. De même parmi les vidéos générées automatiquement, la méthode multicanaux a présenté de meilleurs scores ergonomiques. La préférence des utilisateurs pour la méthode multicanaux élaborée dans le cadre du projet est un argument supplémentaire en faveur de ce type de méthode de génération pour la production de LSF automatisée. Toutefois, les performances étant plus faibles qu'un signeur virtuel humain, ROSETTA pourra être amélioré dans le futur.

Le troisième objectif en phase 1 a été également d'évaluer le module sous-titrage de ROSETTA par rapport à d'autres modes de sous-titrage existants. Le modèle ROSETTA version française (124) a présenté des qualités ergonomiques pauvres (utilisabilité, acceptabilité, utilité...), similaires au sous-titrage automatique de YouTube. Le sous-titrage traditionnel humain présentait de bien meilleures qualités ergonomiques selon les utilisateurs.

À la suite des résultats obtenus en phase 1, l'objectif a été de dupliquer les questionnaires en ligne sur de nouvelles versions de ROSETTA plus abouties. Les résultats ont montré que le signeur virtuel humain présentait des scores toujours plus élevés que la méthode multicanaux contextualisée. Les analyses n'ont pas permis de mettre en évidence des différences de scores entre la version multicanaux 1 (non contextualisée) et la version 2, finale (contextualisée). Des analyses approfondies devraient être menées pour déterminer les axes d'amélioration prioritaires.

Par ailleurs dans le cadre du module sous-titrage, les modèles de ROSETTA basés sur l'intelligence artificielle s'étant améliorés depuis la phase 1, nous avons donc évalué leurs performances. Ainsi en phase 2, nous avons comparé les plus récentes versions de ROSETTA à 2 modes de sous-titrage existant. De même, nous avons comparé les nouvelles versions de ROSETTA aux plus anciennes pour mettre en évidence des signes d'amélioration. Les résultats ont montré que le sous-titrage traditionnel (humain) présentait des scores plus élevés que les autres modes de sous-titrage. Seul le sous-titrage ROSETTA multilingue (anglais) présentait des scores satisfaisants comparé au sous-titrage YouTube. Toutefois, nous avons relevé un biais du niveau de compétence qui semble avoir affecté les réponses utilisateurs mais demeure un biais qui pourrait bénéficier au projet dans les perspectives d'usage de ROSETTA. Nous avons observé que plus le niveau des participants était faible dans la langue de sous-titrage, plus ils estimaient les sous-titres de bonne qualité. En somme, des apprenants d'une langue étrangère seraient susceptibles de préférer ROSETTA pour sous-titrer du contenu à visée pédagogique.

Le livrable **6.3.3** a présenté les résultats des évaluations de sous-titres bruités par la méthode des juges selon le type d'émission et le niveau de bruit. Les principaux résultats montrent que l'effet du niveau

¹ Les vidéos du signeur virtuel humain sont des vidéos d'un humain signant des phrases et montré sous la forme d'un avatar. Ces vidéos ont été enregistrées avec de la *motion capture* via le dispositif utilisé dans l'élaboration du corpus enregistré mentionné dans le lot 2.



de bruit dépend du type d'émission. Enfin, le degré de corrélation des évaluations attribuées aux trois dimensions augmente avec le niveau de bruit.

Le livrable **6.3.4** enfin livre une synthèse des différents résultats présentés dans les livrables précédents.



5.7 SP 7 : Tests in situ - Focus groups

Objectif

Dans une approche LIVING LAB, notamment avec les utilisateurs concernés, ce lot vise à anticiper la réussite de la réalisation du projet ROSETTA en considérant que ROSETTA doit d'abord être utile, accepté, accessible, apprenable et utilisable. Les résultats des travaux avaient pour objectif l'amélioration des fonctionnalités à partir de la combinaison d'une approche raisonnée (ontologique) des pratiques télévisuelles et des retours d'expérience. Le Lutin Userlab, responsable de ce lot a travaillé étroitement avec le sous-traitant Holken Consultants & Partners (HC) pour la réalisation de l'ensemble des livrables du lot 7.

Date de début : T0		Date de fin : T0+42			
Responsable : LUTIN		Participants : Tous			
	SYSTRAN	FTVS	MOCAPLAB	LIMSI	LUTIN
Charge (h*m)	4	2	2	2	12

Tâches et livrables sur la période

Tâche 7.1 : Phase préparatoire : caractérisation des situations d'accessibilité télévisuelles

Au démarrage du projet, nous avons procédé à une phase préparatoire de caractérisation des besoins des utilisateurs d'accessibilité pour que les intentions des partenaires y correspondent, surtout pour la phase de la constitution de la maquette et des nouveaux services, et nous y avons proposé des éléments structurants pour optimiser l'expérimentation. Il s'agissait de préfigurer puis de tester des services et contenus projetés.

#	Échéance	Intitulé livrable	Type	Responsable	État
7.1.1	T0+3	Rapport phase préparatoire intégrant les axes de décision	Document	HC/LUTIN	Livré P1, P2

Avancées

La version finale du livrable **7.1.1** est une synthèse actualisée de 4 sessions de *brainstorming/focus groups* qui ont permis de mettre en exergue différents idéaux de l'accessibilité auprès de différents publics :

- Allophones : Faciliter l'apprentissage du français
- Seniors (60 à 87 ans) : Faciliter l'utilisation des différentes options d'accessibilité (dont une télécommande unique)
- Sourds, malentendants (22 à 69 ans) : Penser l'ergonomie du sous-titrage et de l'avatar signant : personnalisation de taille, couleurs, emplacement, etc.
- Experts de la LSF : Accroître la place des Sourds et malentendants dans l'industrie audiovisuelle, pour la création de contenus spécifiques et l'inclusion dans le processus de production d'accessibilité



Tâche 7.2 - Organisation du retour d'expérience

Dans la première partie du projet nous avons organisé des expérimentations basées sur des prédictions et servant non pas seulement à repérer ce qui ne va pas mais à connaître ce qui favorise les dimensions de l'usage que sont l'accessibilité, l'acceptabilité, l'utilité et l'utilisabilité.

#	Échéance	Intitulé livrable	Type	Responsable	État
7.2.1	T0+6	Rapports sur les typologies et sur l'apport incrémental des phases de développement	Document	HC/LUTIN	Livré P1

Avancées

Le Livrable 7.2.1, est dans la continuité des livrables 6.1.1 et 7.1.1. Après avoir relevé les besoins, demandes et recommandations des utilisateurs, il s'agissait de définir comment ceux-ci peuvent être traduits en critères d'accessibilité et d'organiser des expérimentations pour tester ce qui favorise les dimensions de l'usage que sont l'accessibilité, l'acceptabilité, l'utilité et l'utilisabilité.

Tâche 7.3 - Tests d'utilisateurs *in situ*

En fonction des résultats, et dès que les maquettes l'ont permis, des tests *in situ* ont été organisés. Ils se réalisés au fur et à mesure des avancées du projet, de manière itérative et en préfigurant l'étape suivante.

Ce procédé a permis de finaliser la maquette de ROSETTA en vue de la tester à plus grande échelle.

#	Échéance	Intitulé livrable	Type	Responsable	État
7.3.a	T0+12	Rapports sur les dimensions de l'utilisabilité (1ere phase de <i>focus groups</i> avec 1ers résultats du projet), dont : 7.3.1 FG Tout venant 7.3.2 FG Hispanophones 7.3.3 FG Francophones apprenant l'anglais 7.3.4 FG Seniors 7.3.5 FG Sinophones apprenant le français 7.3.6 FG Sourds - LSF 7.3.7 FG Sourds - sous-titrage français 7.3.8 FG Francophones apprenant l'espagnol 7.3.9 FG Entendants LSF Synthèse transversale des <i>focus groups</i> 1ère phase Rapport complémentaire –Objectivation des données ST français	Document	HC/LUTIN	Livré P3
7.3.b	T0+27	Rapports sur les dimensions de l'utilisabilité (2e phase de <i>focus groups</i> avec derniers résultats du projet) – synthèse des <i>focus groups</i> 2e phase	Document	HC/LUTIN	Livré P3



Avancées

Les *focus groups* ont donné lieu à la rédaction des rapports **7.3**. Ces entretiens collectifs ont tenu compte des délais de livraison et des différences de maturité des modules techniques, ainsi que des besoins spécifiques des différents publics et des conditions sanitaires (recueil des avis utilisateurs à distance notamment).

Une première phase de tests a recensé les dimensions de l'utilisabilité en réunissant des utilisateurs d'horizons différents pour des entretiens collectifs (*focus groups*). Chaque *focus group* avait donné lieu à la production d'un document de synthèse qui inclut la méthode et les résultats obtenus par *focus group*. Chaque rapport a illustré les points forts et ceux à améliorer des systèmes existants, les besoins, hypothèses et recommandations des profils d'utilisateurs cibles du projet ROSETTA. Cette première phase s'est terminée par un rapport de synthèse de l'ensemble des rapports individuels (livrable **7.3a**).

Une deuxième phase de tests a eu lieu vers la fin du projet avec les démonstrateurs les plus aboutis des partenaires du projet. Les mêmes profils d'utilisateurs qu'auparavant étaient conviés à participer à nouveau à ces *focus groups* pour exprimer leurs avis dans le but d'identifier d'éventuelles améliorations concernant l'utilisabilité des outils. Un rapport de synthèse dédié (livrable **7.3b**) dresse la synthèse de l'ensemble des données recueillies lors de la phase 2, met en évidence les besoins et points d'intérêt des utilisateurs-types interrogés, compare les résultats de cette phase 2 avec précédente et formule les recommandations pour améliorer les outils d'accessibilité audiovisuels développés par les partenaires du projet ROSETTA.

Les 40 utilisateurs ayant testé du matériel lors de la première vague (sous-titré en français, sous-titré en version multilingue ou des animations en LSF présentées sous forme d'un signeur virtuel) ont perçu spontanément quatre dimensions d'accessibilité : la compréhensibilité, la fidélité puis la visibilité et enfin la lisibilité. Les participants de ces *focus groups* ont accordé le plus d'importance au fait de comprendre ce qui est dit (31,5% des avis) et à ce que la restitution sous forme de sous-titres ou d'un signeur virtuel soit fidèle au contenu (29,5%). Venaient ensuite les observations faites par rapport à la visibilité (23,3 % des avis) et la lisibilité (15,8 %). Le fait que les sous-titres soient lisibles et bien visibles, ou que les signes du signeur virtuel soit clairs, fluides et bien placés. Les participants de cette première phase de *focus groups* ont accordé plus d'importance à la **transcription exacte du message, des émotions qu'à l'esthétisme de la solution d'accessibilité**.

Lors de la deuxième vague 24 utilisateurs ont testé le matériel ROSETTA le plus abouti en matière « d'automatisation », qu'il s'agisse du matériel sous-titré en français, sous-titré en version multilingue ou des animations en LSF et présentées sous forme d'un signeur virtuel. La plupart des testeurs avaient aussi participé à la première phase.

Les quatre (mêmes !) dimensions d'accessibilité ont été perçues spontanément par l'ensemble des testeurs : la compréhensibilité, la fidélité puis la visibilité et enfin la lisibilité, mais dans un ordre différent de la première vague. Au final, ces participants ont accordé le plus d'importance au fait d'avoir la possibilité **de saisir clairement le message**, sans confusion, sans trop d'effort (109 commentaires) mais aussi de comprendre le message (113 commentaires). Bien que l'approche critique étaient recherchée, les participants ont apporté quelques réactions très positives quant à l'évolution de l'automatisation des sous-titres notamment, mais aussi de l'évolution de la génération automatique de la langue des signes, en dépit du fait que pour cette dernière de nombreuses années de recherches sont encore nécessaires :

- Côté sous-titrage multilingue, c'est le groupe des « apprenants anglais » qui a été le plus enjoué face au prototype, notamment dans le cadre du visionnage de la seconde vidéo de TEDTalk. A l'image de ce participant qui expliquait : « *c'est impressionnant de voir à quel point*



le robot peut transcrire. J'ai tout compris ce qui était transcrit donc c'est un bon point. C'est dur à savoir que c'est un robot. Il arrive à savoir quand on parle de « biberon » entre parenthèses, il arrive à mettre les parenthèses tout seul. » (3 avis – apprenants anglais).

- Côté LSF, les participants du groupe « entendants » pratiquant la langue des signes notamment ont été particulièrement satisfaits du prototype présenté. Les trois participants ont déclaré : « *Je suis encore plus surpris(e) sur cette 2eme phase. Il y a un vrai pas qui a été fait entre les anciennes vidéos et celles-ci. On arrive à un bon niveau. J'imagine que ce type de techno a des débouchés. Je suis content(e) qu'il puisse se développer des avatars de bonne qualité en LSF. » (3 avis – entendants).*

En comparaison avec la première phase de *focus groups*, les principaux points améliorés se situent au niveau de la fidélité, où les progrès étaient bien perçus. Les participants ont exprimé davantage de satisfaction par rapport à cette dimension dans la phase 2, dont notamment :

- Le respect du sens des phrases,
- La transmission des intonations dans les solutions d'accessibilité proposées (dans le cadre du matériel sous-titré en version multilingue).

Si cette méthode d'expression libre a porté ses fruits, nous recommandons pour de futures passations de tests de réaliser en supplément des entretiens (collectifs et individuels) *semi-directifs*, permettant au fur et à mesure des avancées technologiques de cerner et traiter davantage de points précis par rapport aux rendus et aux perceptions et attentes potentielles tant côté fournisseurs/chercheurs que côté utilisateurs.

Dans ce sens, il pourrait être intéressant de poser aux participants des questions semi-directives permettant d'affiner des retours concernant des catégories telles l'accessibilité, l'acceptabilité, l'utilité et l'utilisabilité. Cela pourrait s'exprimer via des questions comme « *dans quelle mesure les sous-titres ainsi présentés vous semblent-ils acceptables ? Si non, auriez-vous une/des suggestions pour qu'ils le soient ?* ».

Tâche 7.4 - Étude comparative sur le plan international

Afin de relever les autres pratiques ou projets en Europe et à l'international, il convenait de mener une étude qualitative dans le but d'apprendre des meilleures pratiques et/ou d'apporter les savoir-faire et connaissances acquises ailleurs dans le monde (*benchmark*). Ceci également dans le but de conquérir de nouveaux marchés, liant l'inclusion, l'innovation sociale et sociétale et la haute performance technologique de l'intelligence artificielle.

#	Échéance	Intitulé livrable	Type	Responsable	État
7.4.1	T0+9	Rapport benchmark et positionnement de ROSETTA sur le plan européen et international	Document	HC/LUTIN	Livré P1
7.4.2	T0+33	Mise à jour benchmark et positionnement de ROSETTA sur le plan européen et international	Document	HC/LUTIN	Livré P3



Avancées

- Un premier rapport de benchmark et positionnement de ROSETTA a été réalisé via des recherches documentaires sur le plan international (livrable **7.4.1**), puis complété et enrichi des retours des partenaires.
- Une mise à jour a été faite (livrable **7.4.2**) au travers une phase d'entretiens sur le plan international (25 interviews à partir de 320 contacts identifiés) pour confronter ces recherches aux besoins et pratiques de professionnels :
 - Producteurs/diffuseurs,
 - Professionnels du ST,
 - Entreprises, administrations,
 - Plateformes d'enseignement,
 - Associations d'utilisateurs.

La recherche documentaire sur l'état de l'art international et mise en accessibilité automatique (L7.4.1) a mis en évidence que la digitalisation globale et croissante peut être au service d'une accessibilité pour tous. Elle s'est intéressée à l'existant dans le domaine de l'automatisation des outils numériques d'accessibilité audiovisuelle dans le domaine des sous-titres et de la génération de contenu en langue des signes (LS).

Dans la continuité de la littérature et des projets dédiés à ce sujet, la phase interviews (L7.4.2) a pu examiner de façon empirique les pratiques, avancées et besoins des acteurs de l'audiovisuel en matière de sous-titrage adapté, sous-titrage multilingue et en langue des signes.

L'analyse met en évidence que les pratiques-clés d'accessibilité des contenus vidéos varient fortement d'une structure à l'autre, et que certaines **inquiétudes**, voire des **obstacles techniques** ou l'absence de corpus important, notamment pour la génération automatique de la langue des signes, mais pas seulement, doivent être levées pour automatiser les services d'accès personnalisés dans ce domaine, évolution qui serait alors **favorablement accueillie**, tant par les bénéficiaires que les entreprises impliquées dans les domaines de la production-diffusion vidéo. **Quatre principaux constats** sont relevés à l'évocation des solutions automatiques pour le sous-titrage et la génération automatique de langue des signes par les organisations qui produisent et/ou diffusent du contenu vidéo :

- **une forte hétérogénéité dans les pratiques et les besoins,**
- **des inquiétudes face à l'automatisation,**
- **l'attente d'une contrepartie économique,**
- **des critères de qualité indispensables.**



5.8 SP 8 : Valorisation

Objectif

Ce lot 8 consistait à valoriser le projet ROSETTA notamment par la dissémination (communication et exploitation). L'ensemble des partenaires y ont participé. Ce lot a été mis en place par le sous-traitant Holken Consultants, ayant l'habitude de piloter la dissémination dans des projets collaboratifs public-privés et notamment dans le domaine du numérique et de l'accessibilité. Chaque partenaire y a contribué.

Date de début : T0		Date de fin : T0+42				
Responsable : LIMSI		Participants : Tous				
	SYSTRAN	FTVS	MOCAPLAB	LIMSI	LUTIN	
Charge (h*m)	2	3	2	2	2	

Tâches et livrables sur la période

Tâche 8.1 : Plan de dissémination et d'exploitation

Cette tâche visait à définir

- Le plan de dissémination et outil de collecte des événements
- Le plan d'exploitation

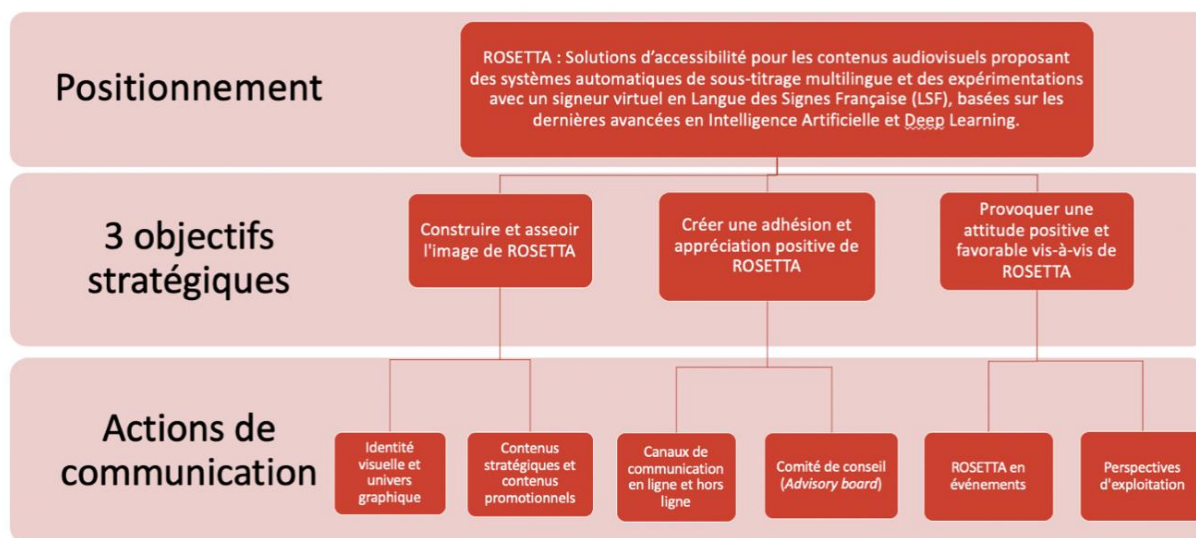
Un *Advisory Board* a été mis en place pour assurer le suivi.

#	Échéance	Intitulé livrable	Type	Responsable	État
8.1.1	T0+7	Plan de dissémination	Document	HC/LIMSI	Livré P2
8.1.2	T0+9	Advisory Board (AB) en place et réunion 1	Meeting	HC/LIMSI	12 juillet 2021, rapport livré P3
8.1.3	T0+21	Advisory Board (AB) en place et réunion 2	Meeting	HC/LIMSI	29 sept 2021, rapport livré P3
8.1.4	T0+30	Advisory Board (AB) en place et réunion 3	Meeting	HC/LIMSI	23 nov 2021, rapport livré P3
8.1.5	T0+39	Plan d'exploitation	Document	HC/LIMSI	Livré P3
8.1.6	T0+39	Rapport de dissémination	Document	HC/LIMSI	Livré P3



Avancées

Cette dernière année a été riche en actions de valorisation. Elle a vu la finalisation de l'ensemble des actions prévues initialement dans la stratégie de dissémination (**L8.1.1** Stratégie de dissémination) selon le plan suivant :



Phase opérationnelle

T0

T+24 mois dissémination
T+36 mois projet

- Advisory Board

HC a travaillé, en sous-traitance pour le LISN et en consultation avec l'ensemble des partenaires, sur la mise en place d'un comité de conseillers (*Advisory Board*) issu de la chaîne de valeur de l'accessibilité. Les membres ont été informés des avancées du projet et sont des relais d'information potentiels visant à asseoir dans la mesure du possible la phase déploiement/exploitation du projet.

La mise en place du comité des conseillers s'est réalisée durant la dernière année du projet. 3 réunions entre les membres du consortium et les membres du comité de conseil (ATAA, UNANIMES, BUCODES SurdiFrance, FICAM, FNSF, CSA) ont eu lieu :

- Une première présentation du projet ROSETTA auprès des conseillers. (**L8.1.2**)
- Une deuxième réunion afin de démontrer les avancées techniques de chaque partenaire du projet suivies de réflexions communes sur les enjeux et applications de chaque module (**L8.1.3**)
- La troisième et dernière réunion afin de présenter les résultats du projet et le démonstrateur final, ainsi que pour dresser des perspectives, tant du côté sous-titrage que LSF (**L8.1.4**)

Chaque réunion a donné lieu à un compte-rendu et les résultats et les conclusions des conseillers font l'objet d'une partie dédiée dans le rapport de dissémination.

Si les résultats du projet ne sont pas encore immédiatement exploitables, ce sont notamment les représentants des communautés des personnes sourdes et malentendantes qui sont très en attente et **ouverts à de nouveaux outils d'accessibilité**, tout en voulant maintenir le cap de la **qualité** préconisée dans la charte du CSA pour les sous-titres adaptés.

Concernant les sous-titres multilingues, les communautés de sous-titres, traducteurs et adaptateurs se positionnent comme **observateurs des changements** induits par l'intelligence artificielle et, **inquiets**



sur le devenir de leurs professions, ont besoin d'être rassurés quant à l'utilité de l'automatisation de la production de sous-titres pour leurs professions, qu'ils ne souhaitent pas voir réduites à la post-édition.

Concernant la génération de la langue des signes via un signeur virtuel, dont la recherche nécessite encore plusieurs années, il serait utile d'**intégrer les personnes** sourdes, donc utilisatrices de nouveaux outils, **dès la rédaction de réponses à des appels à projets**. Ceci est valable aussi pour le sous-titrage automatique, qu'il s'agisse d'utilisateurs potentiels (personnes sourdes, malentendantes, seniors, allophones) ou de professionnels du sous-titrage.

- Collecte des événements

La participation événementielle a dû être modifiée au cours du projet en raison de la crise sanitaire, que ce soit au plan national, européen ou international. La préparation à des événements nationaux et internationaux se trouvait dans un calendrier événementiel régulièrement mis à jour. Mais la plupart des événements ont été annulés ou proposés virtuellement dans un deuxième temps. Ce calendrier est présenté dans le rapport de dissémination et montre l'ensemble des événements sélectionnés et proposés aux partenaires. De facto, la partie événementielle n'a pas porté beaucoup de fruits, les partenaires s'étant plutôt concentrés sur la rédaction d'articles par exemple.

- Le plan d'exploitation

Le plan d'exploitation livré en toute fin du projet (**L8.1.5**) a été alimenté par l'ensemble des partenaires du projet. Chacun y explique les avancées et les difficultés qui restent encore à résoudre pour potentiellement devenir opérationnel avec les outils ROSETTA et dresse les futures perspectives en fonction de l'état des résultats et de son champ d'action.

- Rapport de dissémination

Ce rapport (**L8.1.6**) dresse la liste des activités de dissémination durant la période du projet (juin 2018 - novembre 2021) et donne également un aperçu concernant l'après-projet.

Il informe de manière détaillée sur les procédures et les actions menées pour la valorisation du projet (la communication interne et externe du projet, la constitution et la coopération avec le comité des conseillers (*Advisory Board*) et l'accompagnement dans la phase exploitation des réalisations du projet ROSETTA).

En conclusion, l'ensemble des livrables de dissémination ont été finalisés pour la clôture du projet.



Tâche 8.2 : Matériels de communication

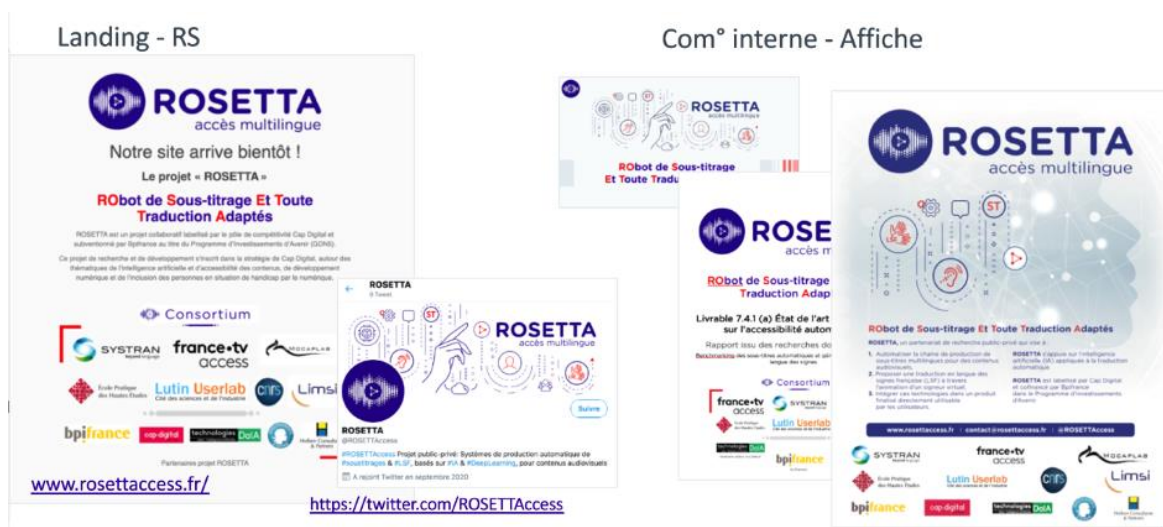
Cette tâche regroupe :

- La création de l'identité visuelle de ROSETTA
- La création du site avec la gestion des contenus (online/offline)
- Création de matériel de communication

#	Échéance	Intitulé livrable	Type	Responsable	État
8.2.1	T0+9	Identité visuelle : 8.2.1a Cahier des charges du site web 8.2.1b Contenu du site web	Document	HC/LIMSI	Livré P2
8.2.2	T0+12	Site web et réseaux sociaux	Meeting	HC/LIMSI	Livré P3
8.2.3	T0+9	Matériels de communication en place	Meeting	HC/LIMSI	Livré P2

Avancées

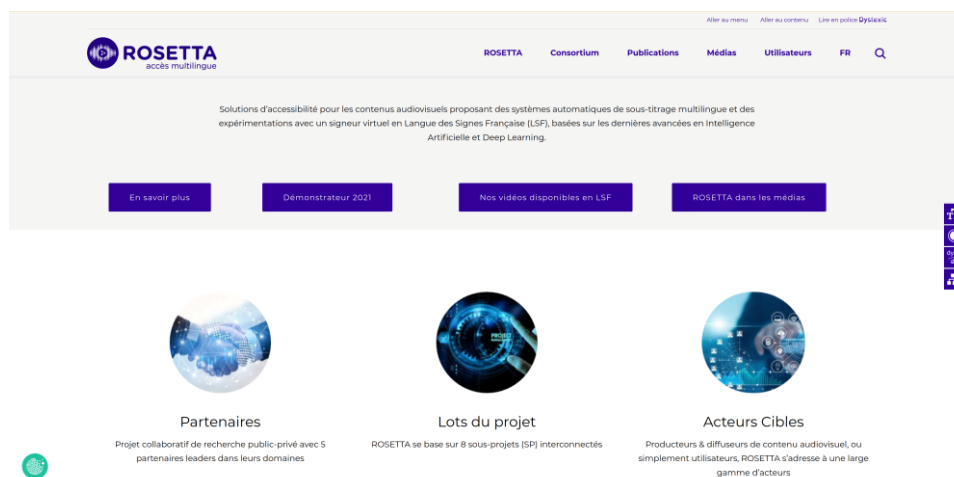
Bien que cette tâche soit appelée 8.2, il s'agit des premiers éléments de communication mis en place permettant de créer l'identité propre de ROSETTA et permettant de créer les canaux de communication correspondant à la valorisation.



A partir d'avril 2020, HC ont travaillé en tant que sous-traitants du LISN (anciennement LIMSI) sur la création d'une image de marque propre et bien distincte pour le projet ROSETTA. Cette identité visuelle, suivie et validée par l'ensemble des partenaires, a été la base pour créer l'ensemble du matériel de communication du projet. Celle-ci s'exprime à travers un logo avec une *baseline* (**ROSETTA accès multilingue**) et des supports graphiques de communication (bannières, poster, kakémono, *leaflets*, etc.). Ces supports ont pu être utilisés tant pour la communication en présentiel qu'en distanciel.



Le site www.rosettaccess.fr constitue la principale source d'information en ligne sur le projet. A l'image de l'accès qu'il propose pour les contenus multimédias, il est multilingue : français, anglais et LSF sous forme de vidéos. Il fournit des informations détaillées sur le projet ROSETTA telles que sa nature, ses objectifs, sa méthodologie, sans oublier les renseignements techniques comme la présentation des sous-projets (lots) et les résultats (sous forme de rapports et publications). Il recense les apparitions du projet dans les médias et a permis de relayer l'appel à contributions auprès de groupes d'utilisateurs.



En complément à la communication sur le site web, les membres du consortium ont opté pour l'utilisation des réseaux sociaux adaptés à ROSETTA : twitter (diffusion fréquente d'information) et LinkedIn (création de réseau professionnel autour du projet et partage d'information).

Un calendrier éditorial était régulièrement mis à disposition des partenaires sur la Dropbox commune au projet, ce calendrier organisait et suivait l'évolution et l'avancement des publications ROSETTA sur les réseaux sociaux.

Tâche 8.3 : Démonstrateur

Développement d'un démonstrateur en ligne valorisant les différentes avancées du projet. Les tâches de développement seront effectuées dans les autres lots, cette sous-tâche en sera juste l'intégration.

#	Échéance	Intitulé livrable	Type	Responsable	État
8.3.1	T0+9	Spécification du démonstrateur	Document	SYSTRAN	Description intégrée dans le rapport 8.1.6
8.3.2	T0+27	Prototype	Document	SYSTRAN	Description intégrée dans le rapport 8.1.6
8.3.3	T0+39	Version Finale du démonstrateur	Logiciel	SYSTRAN	Livré P3



Avancées

Le démonstrateur mis en ligne (<https://rosettaccess.fr/index.php/demonstrateur-final-rosetta/>) montre sur un court extrait de titres de journaux le résultat des recherches menées dans les modules techniques de ROSETTA, à la fin du projet (novembre 2021) :

- Lot 3 : sous-titres français générés automatiquement par le LISN
- Lot 4 : sous-titres multilingues générés automatiquement par SYSTRAN
- Lot 5 : avatar signant en LSF :
 - corpus enregistré par MOCAPLAB
 - algorithme de traduction du français vers la LSF en passant par AZee par SYSTRAN et le LISN
 - annotation AZee, coordination et vérifications LSF par le LISN
 - génération de l'avatar 3D par MOCAPLAB



Ce démonstrateur n'est **pas un "produit", mais un aperçu du meilleur des technologies développées** dans chacun de ces lots techniques pour la production de contenu annexe permettant l'accessibilité d'une vidéo initiale : sous-titres et vidéo LSF. Les utilisateurs peuvent alors apprécier l'état de bon nombre de résultats du projet sur un même support, en utilisant les différentes fonctionnalités du FTV *player* (<https://rosettaccess.fr/index.php/access-player/>) notamment pour l'affichage des sous-titres et la vitesse de défilement.

Ce démonstrateur est le résultat d'un travail conjoint des différents partenaires, depuis la conception et le choix du matériel jusqu'à l'intégration dans le FTV *player*. La véritable valeur constituée à la fin du projet réside dans toute la recherche qui a été menée pendant les 3 ans et demi du projet, ce démonstrateur commun vise à en montrer un aperçu pour :

- **visualiser de façon immédiate** l'intégration de fichiers générés par des modèles automatiques pour améliorer l'accessibilité d'une vidéo
- **garder une trace** du point atteint par la recherche dans ce domaine dans le cadre de ce projet
- **donner envie** d'en savoir plus par la lecture des rapports et publications scientifiques associées