

ROSETTA

accès multilingue

RObot de Soutirage Et Toute Traduction Adaptés

Livrable 7.1.1 : Phase préparatoire (1)
Rapport sur les axes de décision des
partenaires





Le projet « ROSETTA »

ROSETTA est un projet collaboratif labellisé par le pôle de compétitivité Cap Digital et subventionné par Bpifrance au titre du Programme d'Investissements d'Avenir (GDN5). Ce projet de recherche et de développement s'inscrit dans la stratégie de Cap Digital, autour des thématiques de l'intelligence artificielle et d'accessibilité des contenus, de développement numérique et de l'inclusion des personnes en situation de handicap par le numérique.

Le projet ROSETTA propose :

1. d'automatiser la chaîne de production de sous-titres multilingues de contenus audiovisuels,
2. d'assister la chaîne de production d'une traduction en Langue des Signes Française (LSF) représentée par l'animation d'un signeur virtuel (ou avatar signant),
3. d'intégrer ces technologies dans un produit finalisé directement utilisables par les utilisateurs.

Durée de projet 36 mois : Octobre 2018 – Novembre 2021

Tous les droits sont réservés

Le document est la propriété des membres du consortium ROSETTA. Aucune copie ou distribution, sous quelque forme ou par tout moyen, n'est autorisée sans l'accord écrit et préalable du (des) propriétaire(s) des droits.

Ce document ne reflète que le point de vue de ses auteurs. Le consortium ROSETTA, les auteurs du document et les financeurs ne peuvent être tenus responsables de l'usage qui pourrait être fait des informations contenues dans ce document.

©2018 ROSETTA

Historique	Date	Modification(s)
V 0.01	12/07/2019	Rédigé par HC (Hadmut Holken)

Résumé

Le projet ROSETTA propose :

1. D'automatiser la chaîne de production de sous-titres multilingues de contenus audiovisuels,
2. De fournir une représentation en Langue des Signes Française (LSF) de ces contenus via l'animation d'avatars signeurs (traducteurs) virtuels,
3. D'intégrer ces technologies dans un produit finalisé directement utilisables par les utilisateurs.

Ce livrable répond à la tâche 7.1.1



Il correspond à la phase préparatoire avec la caractérisation des situations d'accessibilité télévisuelles, intégrant les axes de décision. Il fait l'objet de **plusieurs livrables distincts** et une synthèse :

Phase préparatoire (0) - Rapport du brainstorming avec un public d'experts en accessibilité langue des signes française (LSF) et sous-titres (ST) pour les sourds et malentendants

Phase préparatoire (1) - Rapport sur les axes de décision des partenaires

Phase préparatoire (2) - Rapport du brainstorming avec un public sourd ou malentendant

Phase préparatoire (3) - Rapport du brainstorming avec un public d'étrangers

Phase préparatoire (4) - Rapport du brainstorming avec un public de personnes âgées (à venir)

Phase préparatoire (5) - Synthèse. Phase préparatoire : caractérisation des situations d'accessibilité télévisuelles, intégrant les axes de décision des partenaires

Mots clés

- Secteur(s) d'application : audiovisuel, télévision, plateformes de contenus, formation, communication, information, inclusion, design pour tous.
- Domaine(s) technologiques : intelligence artificielle, apprentissage profond, Big Data, apprentissage automatique, corpus, génération automatique des sous-titrages adaptés multilingues, modélisation, avatar 3D, capture de mouvements.

Sommaire

Sommaire	3
1. Contexte du livrable	4
1.1. Qui ?	5
1.2. Quoi ?	5
1.3. Quand ?	5
1.4. Pourquoi et comment ?	5
2. Résultats issus du questionnaire et du brainstorming	6
2.1. Objectifs des partenaires dans le projet	6
2.1.1. Les objectifs exprimés	6
2.1.2. Les rôles de chacun	8
2.2. Les enjeux du projet	8
2.2.1. Enjeux sur le plan de l'innovation	8
2.2.2. Enjeux sur le plan économique	9
2.2.3. Impact attendu	10
2.3. Les besoins pour atteindre ces enjeux	11
2.4. Les projets R&D antérieurs sur lesquels s'appuyer	12
2.5. Les cibles « clients » ou « utilisateurs » pour les solutions développés	13
2.6. Autres entités extérieures à intégrer dans le développement du projet	14
2.7. Les buts sur le long terme (5 ans, après-projet)	15
2.8. Les points potentiellement bloquants	16



2.9. La planification des tâches	18
2.10. Les attentes face aux autres partenaires	19
2.11. Les ingrédients pour réussir le projet.....	21
3. Conclusion	22
Glossaire	23
Annexes	24
Annexe 1 : Guide d'entretien.....	24
Annexe 2 : Quels traducteurs automatiques ?	26
Annexe 3 : Quelle automatisation envisager ?	29
Annexe 4 : Réflexions autour de la proximité d'une LS vers une autre et simplification	32

Table des tableaux

Tableau 1 : Objectifs des partenaires dans le projet Rosetta	7
Tableau 2 : Rôle des partenaires dans le projet ROSETTA.....	8
Tableau 3 : Les enjeux de l'innovation pour les partenaires	9
Tableau 4 : Les enjeux économiques pour les partenaires.....	10
Tableau 5 : L'impact attendu	10
Tableau 6 : Les besoins pour atteindre le enjeux visés.....	11
Tableau 7 : Projets R&D antérieurs de partenaires	12
Tableau 8 : Les cibles « clients » « et « utilisateurs »	14
Tableau 9 : Les cibles pour interviews potentiels	15
Tableau 10 : Stratégies long terme des partenaires ROSETTA	16
Tableau 11 : Les éléments potentiellement bloquants	17
Tableau 12 : (A) Attentes face à Systran.....	19
Tableau 13 : (B) Attentes face à France Télévisions Accès	20
Tableau 14 : (C) Attentes face à Mocaplab.....	20
Tableau 15 : (D) Attentes face au LIMSI	20
Tableau 16 : (E) Attentes face au LUTIN Userlab.....	20

1. Contexte du livrable

Ce document répond à la tâche 7.1.1 qui correspond à la phase préparatoire avec la caractérisation des situations d'accessibilité télévisuelles, intégrant les axes de décision des partenaires du projet. Il fait l'objet de six livrables distincts, dont une synthèse :

- Phase préparatoire (0) - Rapport du brainstorming avec un public d'experts en accessibilité Langue des Signes Française (LSF) et Sous-Titres (ST) pour les sourds et malentendants
- **Phase préparatoire (1) - Rapport sur les axes de décision des partenaires**
- Phase préparatoire (2) - Rapport du brainstorming avec un public sourd ou malentendant



- Phase préparatoire (3) - Rapport du brainstorming avec un public d'étrangers
- Phase préparatoire (4) - Rapport du brainstorming avec un public de personnes âgées (à venir)
- Phase préparatoire (5) – Synthèse

1.1. Qui ?

Dans un premier temps, nous avons recueillis les besoins **des partenaires** (janvier/février 2019), puis analysé les retours et présenté les résultats (le 27/02/2019). Cette présentation était suivie d'un brainstorming entre partenaires, la discussion autour de la mise en place des services, et l'identification des lieux de test et de l'échantillon.

Dans un deuxième temps, nous avons procédé, en étroite collaboration avec le Lutin Userlab, à la caractérisation des besoins des utilisateurs autour de leur idéal de l'accessibilité pour que les intentions des partenaires y correspondent. Ce cheminement est utile pour la phase de la constitution de la maquette et des nouveaux services, et ces éléments sont structurants pour optimiser les expérimentations par la suite. A ce stade, il s'agit de préfigurer puis de tester des services et contenus *projetés*.

1.2. Quoi ?

L'objet du présent document est la présentation des résultats issus de la collecte de données et du brainstorming avec les partenaires pour recenser leurs besoins dans le cadre du projet.

1.3. Quand ?

Le recueil des besoins des partenaires s'est déroulé en janvier et février 2019. Les résultats ont été présenté et discuté à la séance avec une séance de brainstorming avec l'ensemble des partenaires. Elle a eu lieu le **27/02/2019** au Laboratoire des Usages en Technologies d'Information Numériques de 14h à 18h.

Par la suite, nous avons projeté des situations d'accessibilité télévisuelles auprès de quatre groupes ou échantillons différents au sein du Lutin Userlab et de la Cité des Sciences à Paris :

- Un groupe d'experts opérant dans l'environnement de personnes sourdes et malentendantes (10 avril 2019),
- Un groupe avec des personnes sourdes et malentendantes (10 juillet 2019),
- Un groupe avec des étudiants étrangers (26 juillet 2019),
- Un groupe avec des seniors / personnes âgées (prévu à partir de septembre 2019).

1.4. Pourquoi et comment ?

Concernant le recueil des besoins du côté des partenaires du projet, il s'agissait d'anticiper la pre-maquette et certaines étapes susceptibles de suivre dans le cadre du projet Rosetta.

Dix questions ont été adressées aux partenaires, auxquelles ils ont répondu. Le guide utilisé se trouve en annexe 1.

Questions pour les partenaires du projet Rosetta

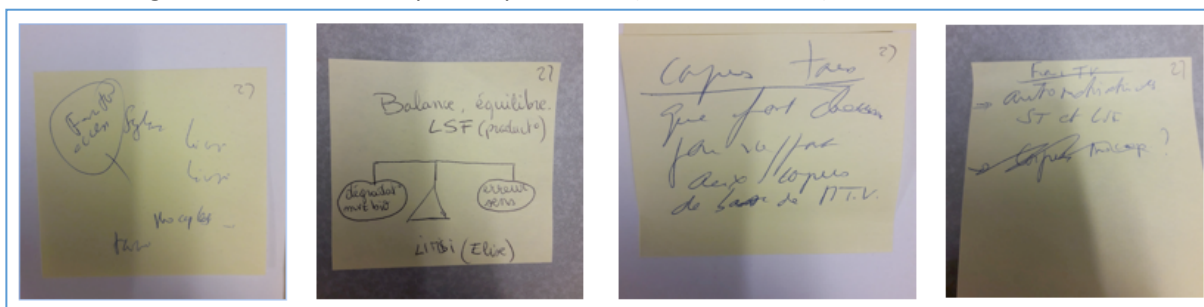
Ce questionnaire prépare

L'Action 7.1 - Phase préparatoire : brainstorming entre partenaires, recueil des besoins, mise en place des services, identification des lieux de test et de l'échantillon

Cibles : Partenaires du projet ROSETTA



Les réponses ont été analysées, restituées lors d'une réunion avec le consortium et discuté lors d'un brainstorming avec l'ensemble des parties prenantes (27 février 2019).



Le présent document résume les éléments-clés issus de l'enquête auprès des partenaires et discutés lors du brainstorming avec eux.

Certains éléments de la retranscription des verbatim par le Lutin (Ammara Bhatti) ont été intégrés dans ce rapport.

Cette approche représente l'un des cinq à six brainstormings prévus dans la phase préparatoire pour le projet ROSETTA (ROBOT de Sous-titrage Et Toute Traduction Adaptés). L'une des séances a eu lieu avec les partenaires du projet.

Toutes les autres séances de brainstorming et les focus groupes qui y étaient associés concernaient les utilisateurs ou experts travaillant avec ces derniers (utilisateurs, bénéficiaires).

Ensemble avec l'équipe du Lutin Userlab, nous avons préparé ces groupes de brainstorming. Durant deux heures chaque groupe d'utilisateurs a abordé, sous l'animation du LUTIN, une discussion autour de l'idéal de l'accessibilité, sous forme d'un brainstorming et d'un focus groupe. Chaque événement fait l'objet d'un rapport dédié. Les contenus de ces rapports sont fusionnés en fin du parcours sous forme d'une synthèse.

2. Résultats issus du questionnaire et du brainstorming

2.1. Objectifs des partenaires dans le projet

2.1.1. Les objectifs exprimés

Les objectifs exprimés par les partenaires sur questionnaire sont comme suit :

Structure	Objectifs dans le projet ROSETTA
FTV	Objectifs centrés sur la politique de service publique <ul style="list-style-type: none"> Gain en productivité par l'automatisation des ST Démultiplier l'offre LSF et ST automatisés pour FTV
Mocaplab	Objectifs stratégiques, technologiques et commerciaux <ul style="list-style-type: none"> Conforter la position de leader technologique pour la motion-capture de la LSF Enregistrer un corpus riche de mocap de LSF



	<ul style="list-style-type: none">• Contribuer au développement et à la commercialisation de nouveaux produits autour des langues des signes pour la valorisation du savoir-faire de Mocaplab
LIMSI	Objectifs de recherche LSF automatisées <ul style="list-style-type: none">• Avancées sur les traitements automatiques, en particulier sur les aspects traduction assistée par ordinateur et génération d'animations de signeur virtuel.
LUTIN	Favoriser la réussite du projet ROSETTA du point de vue des usages des produits et services qui découleront des développements, qu'il s'agisse <ul style="list-style-type: none">• (i) de la validité écologique des corpus et des populations cibles,• (ii) des performances des techniques de transfert d'une modalité à une autre (source sonore vers LSF ou vers sous-titrage),• (iii) d'une langue à une autre (dont FALC),• (iv) de l'évaluation des modules in situ avec les utilisateurs concernés,• (v) de la comparaison avec les solutions d'accessibilité, en France et à l'international.
Systran	Développer son expérience NLP (Natural Language Processing) sur d'autres tâches que la traduction de simplification, comme la génération de sous-titres multilingues automatisée.

Tableau 1 : Objectifs des partenaires dans le projet Rosetta

Par ailleurs, et au-delà de la période du projet, ont été abordés des intérêts concernant la commercialisation des solutions ROSETTA ainsi des objectifs de publication :

(1) Anticipation de la commercialisation des solutions ROSETTA :

Au-delà du projet Rosetta :

- FTV espère avoir gain en productivité par l'automatisation des sous-titres, et de l'offre LSF et sous-titres automatisés. FTV s'attend également à un gain de chiffre d'affaires plus rapide, notamment parce que les sous-titres multilingues permettront une commercialisation plus rapide auprès de leurs cibles/clients à l'étranger.
- Systran pourrait être amené à commercialiser la solution finale auprès de (nouveaux) clients et ainsi potentiellement démultiplier l'offre LSF et ST au-delà des acteurs impliqués dans le projet.
- MOCAPLAB souhaite, par l'enregistrement d'un corpus riche de MOCAP en LSF, contribuer au développement et à la commercialisation de nouveaux produits autour des langues des signes pour la valorisation du savoir-faire de MOCAPLAB.

(2) Publications scientifiques

Les partenaires académiques notamment (LIMSI, CHArt/Lutin Userlab) s'accordent pour souligner leurs objectifs de publication dans des revues scientifiques, tant sur le plan national qu'international.



Durant le brainstorming, il a été souligné qu'à partir du corpus récupéré par France Télévisions Accès, SYSTRAN vise à faire des sous-titres (automatiques), MOCAPLAB s'investit dans l'intégration de la LSF avec LIMSI. Cette intégration se fait par rapport aux niveaux technologiques : SYSTRAN, l'intégrateur, propose le produit qui combine ce qui aura été fait chez MOCAPLAB, ce qu'aura développé le LIMSI.

2.1.2. Les rôles de chacun

La répartition des rôles par partenaire est comme suit :

Structure	Rôle des partenaires dans le projet ROSETTA
FTV	<ul style="list-style-type: none">• Livrer des corpus de l'ensemble des chaînes de FTV avec fichiers sous-titrage.• Mettre en place nos process qualité (ISO 9001) pour les retours des livrables de nos partenaires.
Mocaplab	<ul style="list-style-type: none">• Produire le corpus 3D en mocap (Tâche 2.2 et 2.3)• Faire la synthèse des animations de l'avatar 3D à partir de la représentation de la LSF (Tâche 5.3)
LIMSI	<ul style="list-style-type: none">• Annelies Braffort : LSF, coordination,• Michael Filhol : LSF, représentation AZee, traduction txt→LSF,• Michèle Gouiffès : LSF, traitement d'images, reconnaissance automatique,• Elise Prigent : LSF, perception et compréhension des animations,• François Yvon : sous-titrage, traduction, simplification.
LUTIN	NC
Systran	<ul style="list-style-type: none">• produire les modules de sous-titres adaptés (lot 3)• produire les modules de traduction textuelle (lots 4 et 1)• gérer/coordonner le projet ROSETTA (lot 1). A ce titre, Elise Michon est mobilisée en tant qu'ingénieure de recherche et chef de projet.

Tableau 2 : Rôle des partenaires dans le projet ROSETTA

A partir du rôle de chaque partenaire s'est engagée une discussion autour de la mise en place des traducteurs automatiques et les questionnements liés à cela. Les échanges se trouvent dans l'annexe 2.

2.2. Les enjeux du projet

2.2.1. Enjeux sur le plan de l'innovation

Les partenaires perçoivent leurs enjeux sur le plan de l'innovation de la manière suivante :



Structure	Enjeux sur le plan d'innovation
FTV	<ul style="list-style-type: none">• Intégrer l'IA dans ses logiciels et process de fabrication du sous-titrage SME
Mocaplab	<ul style="list-style-type: none">• Capture d'un important volume de données: indispensable d'optimiser les temps de traitement et d'augmenter la productivité de notre chaîne actuelle• Découverte des applications du machine learning pour la donnée de mouvement• Edition de données de mouvements et identification de paramètres biomécaniques pertinents• Synthèse d'animation par rejeu et combinaison de mouvements enregistrés au préalable
Systran	Développer un modèle pour la tâche inédite de génération automatique de sous-titres à partir de la vidéo et de nouveaux modèles de traduction spécialisés pour le domaine des sous-titres.
LIMSI :	nc
LUTIN	Mettre au point les outils permettant d'évaluer l'innovation sur plusieurs dimensions (nouveau, performance technique, satisfaction des besoins utilisateurs...)

Tableau 3 : Les enjeux de l'innovation pour les partenaires

2.2.2. Enjeux sur le plan économique

Sur le plan économique, les partenaires perçoivent leurs enjeux de la manière suivante :

Structure	Enjeux sur le plan économique
FTV	<ul style="list-style-type: none">• Anticiper l'automatisation partielle ou complète de notre chaîne de fabrication du sous-titrage SME. Réorganisation interne et évolution des métiers en conséquence.
Mocaplab	<ul style="list-style-type: none">• Être reconnu comme l'acteur incontournable pour les enregistrements de corpus 3D de LS• Enregistrer un corpus qui pourra ensuite être valorisé et exploité auprès des acteurs de la recherche en linguistique (B2B: laboratoires, entreprises) ou du grand public (B2C, sous forme d'application)• Licencier aux partenaires les résultats obtenus au cours du projet afin de percevoir des retours financiers sur les ventes des produits issus de ROSETTA



Systran	<ul style="list-style-type: none">Développer avec des partenaires privilégiés des solutions techniques intégrables dans un produit.
LIMSI	na
Lutin	na

Tableau 4 : Les enjeux économiques pour les partenaires

Observations :

On peut se poser la question, s'il faut licencier les innovations aux partenaires uniquement, ou aussi à des tiers ? Faut-il travailler sur le B2B, B2C ?

Systran, FTV et Mocaplab devront s'interroger sur les futurs clients potentiels pour l'innovation ROSETTA.

2.2.3. Impact attendu

Chaque partenaire s'attend à un impact de la manière suivante :

Structure	Impact attendu
FTV	Rester leader sur le marché du SME , démultiplier l'offre d'accessibilité sur les antennes de France Télévisions.
Mocaplab	Connaissance par le public (producteurs de contenus, communauté sourde, ...) du savoir-faire de Mocaplab. Communication à l'international pour être l'acteur incontournable pour des projets similaires relatifs aux langues des signes d'autres pays.
Systran	Contribuer à l'accessibilité des programmes télévisuels. Pouvoir communiquer sur cette contribution.
LIMSI	Intégration des traitements automatiques dans des logiciels permettant d'améliorer l'accessibilité.

Tableau 5 : L'impact attendu

Il a été rappelé que l'impact peut être **économique, sociétal, social, écologique, orienté utilisateur / usage**, et les partenaires ont été invités à y réfléchir.

A la suite de la présentation, il s'est posé la question de la nécessité d'échanger avec les partenaires sur ce qui est envisagé, ce qu'il est possible et réaliste de faire, ce que le projet va produire, et quel type d'automatisation est acceptable. L'ensemble des échanges est en annexe 3.



2.3. Les besoins pour atteindre ces enjeux

Pour atteindre les enjeux évoqués (2.1.1), chaque partenaire a listé ce qu'il lui faudrait obtenir d'un ou de plusieurs partenaires du projet.

Structure	Les besoins dans le cadre du projet permettant d'atteindre les enjeux
FTV	Des livrables de la part de Systran pour évaluer la pertinence des fichiers et voir s'ils sont conformes à la charte du sous-titrage.
Mocaplab	Le projet représente un défi technologique important et l'atteinte des objectifs de Mocaplab est conditionnée par le succès d'autres tâches: <ul style="list-style-type: none">• la représentation formelle de la langue des signes,• la traduction automatique du texte vers le mouvement à travers cette représentation• la bonne définition du contenu du corpus: pertinence, représentativité, recouvrement...,• la bonne définition des besoins pour le player 3D (ergonomie, apparence de l'avatar, type de rendu ...)• la bonne définition des cas d'usage du projet pour garantir sa faisabilité et bien délimiter le corpus.
Systran	Mettre en place des conditions de coopération, administratives, pratiques et scientifiques, facilitant la bonne marche du projet et libérant assez de temps pour implémenter des propositions expérimentales innovantes.
LIMSI	Corpus de LSF (vidéo sous-titrée et mocap) pour les études amont et la génération. Méthodes pour le <i>brainstorming</i> et les focus groupe pour les appliquer aux problématiques plus spécifiques de création de logiciel de TAO pour les traducteurs text→LSF et concernant l'animation des signeurs virtuels (leur aspect, l'anonymisation).
LUTIN	Disponibilité des techniques à tester (familiarisation, expertises, utilisateurs

Tableau 6 : Les besoins pour atteindre le enjeux visés

A titre de discussion, SYSTRAN confirme à FTV la venue d'un livrable pour les fichiers générés et une évaluation par rapport à la charte qui existe.

FTV confirme qu'il faudra une évaluation par rapport aux 16 critères de la charte qualité spécifiques à toutes les télévisions.

Du côté de LIMSI on cherche des méthodes pour le brainstorming et les échanges de groupes pour justement réfléchir ensemble aux méthodes, et voir ce qu'on pourrait ajouter pour alimenter les réflexions sur le logiciel TAO et aussi sur l'anonymisation des signeurs virtuels.



2.4. Les projets R&D antérieurs sur lesquels s'appuyer

Structures	Projets R&D&I (antérieurs) et recommandations sur lesquels construisent les partenaires
FTV	SUBTIL (FranceTélévisions Innovations)
Mocaplab	<ul style="list-style-type: none">• Sign3D : La qualité de cette restitution de l'avatar fut jugée très fidèle par le public sourd aux colloques SLTAT (Sign Language Translation and Avatar Technology) à Chicago en 2013 et à Paris en 2015.• SIGN360 (2016). Prototype d'application interactive pour soutenir l'apprentissage de la Langue des Signes Française.• ASTRe (2016-2017: Concours d'Innovation Numérique). Passage de la preuve de concept à une échelle plus industrielle en réduisant notamment les temps et coûts de production.• SignEveil (2018: AAP Langues et numérique de la DGLFLF). Application mobile interactive permettant de découvrir la LSF de manière ludique. Passage de la preuve de concept de SIGN360 à une application grand public diffusée sur les magasins d'applications mobiles.• SubTil (2017 - mai 2019: AAP Accessibilité numérique). Sous-titrage via un avatar 3D en LSF pour des programmes audiovisuels en collaboration avec France Télévisions. Cas d'usage abordés: programmes jeunesse, débat politique.
Systran	Sur les expériences de projets de recherche précédents au sein de l'entreprise, avec la BPI ou d'autres partenaires.
LIMSI	<ul style="list-style-type: none">• LSF : thèse sur la conception de la représentation « Azee », projet sur l'assistance à la traduction, projet pour la constitution d'un corpus de mocap, stages sur la perception et la compréhension des animations de LSF, thèse sur la reconnaissance automatique...• Simplification : nos travaux visent à exploiter les acquis de 10 ans de recherche en traduction automatique / traduction assistée par ordinateur, en particulier nos travaux récents en traduction neuronale. Ils s'appuient aussi sur les travaux en transcription automatique de parole.
LUTIN	Comparaison TVHD – TVSD pour TDF, ROSETTA pour France Télévisions Media4Dplayer

Tableau 7 : Projets R&D antérieurs de partenaires



Une question s'est adressée à Systran : sur quoi construisez-vous concrètement ? La traduction automatique multilingue par le machine / le *deep learning* ? ... des projets européens ?

SYSTRAN a répondu qu'il s'agit d'expériences de projets de recherches précédentes au sein de l'entreprise avec d'autres partenaires qui se situent plutôt du côté de la traduction, de la simplification du côté textuel.

S'intéressant à la génération des sous-titrages de Mocplab, leur représentant explique :

« C'est pas un projet de sous-titrage automatique ou de traduction, c'est un projet d'amélioration de l'accessibilité via les sous-titrages, donc y a une partie sous-titre texte dans laquelle on est pas impliqué ; c'est l'optimisation, de la visualisation des sous-titres, de la spatialisation des sous-titres au niveau des locuteurs des couleurs, etc. y a toute une partie avec une société, qui s'appelle perfectif memory, c'est sur toute la partie base de données, accès à une base de données, tout ce qui est base de donnée sémantique pour accéder, c'est tout ce qui est métadonnée pour faire de la recherche par le contenu et via les sous-titres. »

« Et par contre sur la Langue des signes, c'était dans le cas du développement du nouveau Player de France TV. Dans le cadre du projet c'était faire preuve de concept, et voir si on pouvait intégrer des sous-titres en avatar signeur dans le nouveau Player et donc la langue des signes c'était juste 3-4 usages, un test dans un petit corpus sur 12 minutes de programmes de jeunesse, quatre dessins animés, et un extrait du débat politique Macron-Le Pen. Il n'y avait pas d'apprentissage, pas de grand volume de corpus, pas de notion de traduction, c'était juste de la capture. »

2.5. Les cibles « clients » ou « utilisateurs » pour les solutions développées

Ci-après sont listées clients et utilisateurs que chaque partenaire cible dans le cadre du développement de ROSETTA. Ces cibles seront les clients potentiels et/ou les bénéficiaires de l'automatisation des sous-titres multilingues et/ou de l'avatarisation de l'interprétariat de la langue des signes.

Structure	Les cibles "clients" ou "utilisateurs" à considérer pour les solutions développées dans le cadre de ROSETTA
FTV	<ul style="list-style-type: none">• FranceTélévisions et ses filiales (FTVEN, FTP, FTD), tous les diffuseurs du paysage audiovisuel français (PAF) ayant des obligations en matière d'accessibilité
Mocaplab	<ul style="list-style-type: none">• Valorisation du corpus: pour les acteurs de la recherche dans le domaine de la linguistique et LSF (laboratoires, entreprises), communauté sourde ou toute personne s'intéressant la LSF.• Valorisation des produits issus de ROSETTA B2B - Sociétés en charge des développements de contenus audiovisuels et de leur sous-titrage.• Autres cas à explorer
Systran	<ul style="list-style-type: none">• Les équipes commerciales qui pourront faire évoluer les nouveaux modèles développés en produit.• Cibles actuelles: services traductions /communication corporate



	<ul style="list-style-type: none"> • Cibles futures : chaînes de télévision, sociétés de production de films, sites de génération de sous-titres
LIMSI	<ul style="list-style-type: none"> • Communauté Sourde, professionnels de production de contenus en LSF, professionnels de la production de sous-titres
LUTIN	<ul style="list-style-type: none"> • Annonces services publics (gares, structures publiques) (note A.B.)

Tableau 8 : Les cibles « clients » « et « utilisateurs »

Selon FTV, côté interprétariat, il n'y a rien à grande échelle à la télévision pour les sourds pour le moment. Donc si le projet ROSETTA arrive à faire quelque chose qui soit compréhensible et fluide, se serait « déjà parfait » pour proposer des *avatars interprètes* comme une solution possible à destination des personnes sourdes et malentendantes. « *Même si ce n'est pas parfait, ce sera déjà super, parce qu'il n'y a rien sur une grande échelle ... il y a seulement 7 minutes par jours, au niveau du coût c'est monstrueux et il n'y a pas assez d'interprètes !* »

2.6. Autres entités extérieures à intégrer dans le développement du projet

Dans le cadre du projet, il serait souhaitable de travailler avec certaines cibles extérieures, notamment pour considérer et rencontrer leurs besoins (clients potentiels, prescripteurs, utilisateurs etc., mais aussi pour recueillir les avis et perceptions des bénéficiaires notamment concernant l'accessibilité, l'acceptabilité, l'utilité et l'utilisabilité des solutions proposées.

A ce stade, les partenaires du projet ROSETTA estiment utiles de considérer les cibles suivantes pour des interviews potentielles (pour la prémaquette, ou dans le cadre de l'enquête prévue ultérieurement).

Structure	Cibles à considérer pour des interviews potentiels (pour la pré-maquette, ou dans le cadre de l'enquête prévue ultérieurement)
FTV	<p>Pas d'avis pour l'instant.</p> <p>Peut-être plus tard le régulateur (CSA) et les associations d'utilisateurs (Unisda).</p>
Mocaplab	<p>Communauté sourde pour définir les besoins relatifs au player (ergonomie, apparence de l'avatar, type de rendu ...).</p> <p>Laboratoires de recherche susceptibles d'être intéressés par le corpus.</p> <p>Société d'édition de supports pédagogiques en langue des signes.</p>
Systran	<p>Dans un premier temps, les utilisateurs finaux afin d'intégrer le plus possible les besoins existants dans la conception de solutions nouvelles.</p> <p>Les personnes générant actuellement les sous-titres manuellement pour les émissions en stock (pas en direct, la chaîne de production en direct ayant déjà été observée lors d'une visite chez France TV Studio) pour connaître leurs contraintes et procédures.</p>
LIMSI	<p>Services de traduction texte vers la LSF (Interpretis, Vice & Versa, Traduquales, etc.),</p>



	Sociétés de production de contenu en LSF (Media’Pi, La manufacture digitale), La FNSF (Fédération Nationale des Sourds de France). Il serait également intéressant de savoir comment travaillent les producteurs de sous-titres (quand ceux-ci ne sont pas produits en « live » par des méthodes de perroquet).
LUTIN	Les <i>end-users</i> en les recensant.

Tableau 9 : Les cibles pour interviews potentiels

Les premières cibles d’utilisateurs ont déjà été entendues sur leur « idéal de l’accessibilité » ; leurs retours ont fait l’objet d’une rédaction dédiée dans le cadre de la tâche 7.1.1 :

- experts travaillant avec des sourds et malentendants,
- des personnes sourdes et malentendantes,
- des étrangers ne maîtrisant pas encore la langue française,
- des seniors (à venir).

Les focus groupes et tests d’utilisateurs seront organisés autour des (pré-)maquettes et /ou prototypes de ROSETTA. Ils auront lieu principalement en France.

Concernant le « benchmark » international, il restera à déterminer les régions géographiques ou linguistiques à traiter en priorité pour l’étude benchmark. HC a mis à disposition des partenaires une proposition de pré-rapport pour avis et discussion (dans le cadre de la tâche 7.4.1 Benchmark 1).

HC sera également force de proposition dans les semaines (et mois) qui viennent, d’une part en fonction des recherches documentaires qui sont en cours, et d’autre part côté utilisateurs et experts audiovisuel, en fonction de l’évolution du projet qui avance au fur et mesure, et en fonction des besoins des partenaires.

2.7. Les buts sur le long terme (5 ans, après-projet)

La vision des partenaires où ils aimeraient être dans 5 ans, au travers ROSETTA, est comme suit :

Structure	But / stratégies long terme (5 ans)
FTV	<ul style="list-style-type: none">• Automatiser tout ou partie de la chaîne de fabrications du sous-titrage SME ;• Démultiplier l’offre d’accessibilité pour FTV (LSF et sous-titrage en multilingues des programmes).
Mocaplab	<ul style="list-style-type: none">• Démocratiser l’accès au grand public des avatars signeurs dans les programmes audiovisuels ;• Etendre l’utilisation de cette technologie à des domaines plus vastes (administration, e-learning, communications...) ;• Etre identifié et choisi comme fournisseur pour la déclinaison de projets similaires à l’international.



Systran	<ul style="list-style-type: none">• Voir les solutions développées dans le projet implémentées à grande échelle sur plusieurs chaînes de télévision,• Avoir contribué à l'évolution des techniques de sous-titrage des programmes télévisuels.
LIMSI	<ul style="list-style-type: none">• Logiciel d'assistance à la traduction texte → LSF déployé dans les services de traduction ;• Logiciel de génération d'animations de signeur virtuel à partir de données de mocap ;• Systèmes de simplification / sous-titrage automatique ou semi-automatique (aide au sous-titrage).
LUTIN	<ul style="list-style-type: none">• Etre spécialiste des technologies numériques pour l'accessibilité

Tableau 10 : Stratégies long terme des partenaires ROSETTA

Pour France Télévisions Accès, la nouveauté consistera à pouvoir proposer sur le long terme:

- l'automatisation des sous-titrages et de la langue de signes pour les programmes,
- un programme en anglais, en français, etc. (plus rapidement) ce qui est intéressant pour la commercialisation des productions de France Télévisions Distribution.

Le LIMSI a souligné l'importance de la proximité d'une langue LS vers une autre, et donc de la « facilité » de pouvoir passer d'une LS vers une autre compte tenu du fait de mettre en avant tout ce qui est commun en langue(s) des signes. Si collectivement les langues vont évoluer pour être toujours simplifiées, puis être productif et économique à la fois, il convient toutefois d'être vigilant dans le cadre de l'avatarisation de la langue des signes pour ne pas trop perdre en expressivité, en productivité, en subtilité, c'est-à-dire autant d'éléments qui constituent le fonctionnement même de la LS¹.

Nous (HC), en concertation avec Le Lutin Userlab et les partenaires, combinerons ces visions stratégiques avec nos approches dans le cadre du benchmarking.

2.8. Les points potentiellement bloquants

Dans le domaine de la recherche, aucun point bloquant n'a été signalé. De manière globale ou sur le plan technologique, voici comment chaque partenaire appréhende certaines difficultés.

Structures	Les points potentiels de blocage	
	Globalement	Sur le plan technologique
FTV	Aucun pour le moment.	na
Mocaplab	Le projet représente un défi technologique important et l'atteinte des objectifs de	na

¹ L'ensemble des verbatim se trouve en annexe 4.



	<p>Mocaplab est conditionnée par le succès d'autres tâches.</p> <p>La qualité de sortie finale de la chaîne "texte → représentation formelle → mouvement" sera déterminante pour mieux identifier les exploitations commerciales possibles.</p>	
Systran	<ul style="list-style-type: none">• Pas de sous-titres en Version Originale à côté des sous-titres en Version Française pour les programmes doublés (par exemple, les sous-titres directement en anglais d'une série originalement en anglais doublée en français).• Peu de programmes proposant de la LSF, et vidéo de l'interprète signeur déjà incrustée dans la vidéo de l'émission.	Le module de génération de sous-titres, grâce à la reconnaissance vocale, pourra générer du texte aligné avec la vidéo, mais ne reconnaîtra pas le type d'information sonore (musique, bruits) et encore moins les locuteurs (personne à l'écran, personne hors champ, voix off), informations normalement codées dans les sous-titres sous forme de couleurs.
LIMSI	<ul style="list-style-type: none">• LSF. Point dur pour le moment : convergence de la représentation linguistique Azee avec les méthodes de génération d'animations par mocap ;• Simplification : très peu de données réalistes disponibles pour la tâche.	na
LUTIN	Tester l'existant et les outils développés dans le cadre du projet	na

Tableau 11 : Les éléments potentiellement bloquants

Idéalement, pour arriver à **débloquer** ces points, les partenaires préconisent des échanges réguliers entre les partenaires :

Ainsi, pour le **LUTIN**, une collaboration inter-partenaires sur les mêmes objectifs est cruciale pour atteindre les objectifs.

Le LIMSI compte développer des échanges réguliers notamment avec Mocaplab, en commençant par identifier ensemble les difficultés potentielles avec un extrait du corpus de mocap et du Limsi déjà décrit avec Azee. Pour la simplification/troncation, le laboratoire est en train de constituer un corpus. A propos de la génération de sous-titres avec couleurs, **Systran** estime utile de :

- connaître d'autres projets de recherche impliquant des étapes de reconnaissance dans l'image et dans le son.
- Concernant les sous-titres en VO : contacter les chaînes qui produisent les programmes doublés diffusés sur France Télévisions.
- Programmes en LSF : contacter d'autres chaînes, contacter les services qui produisent les vidéos des interprètes signeurs pour pouvoir les récupérer.



Par ailleurs, une difficulté à contourner est le nombre a priori très faible des corpus. Ainsi FTV dit que très peu de séries sont diffusées en VO sur France TV, et ne pas y avoir accès. Le LIMSI informe de son côté qu'avec les méthodes de génération d'animation par MOCAP et pour la simplification : très peu de données réalistes sont disponibles pour la tâche.

Le Lutin Userlab suggère qu'il aurait des émissions américaines en streaming (Netflix ?) qui pourraient servir pour récupérer des corpus en VO-STO.

Pour SYSTRAN et le LIMSI, il serait utile de contacter des chaînes qui produisent des programmes doublés diffusés sur France Télévisions, notamment pour :

- la solution génération des sous-titres avec couleurs et des sous-titres en version originale,
- pour les programmes en LSF (contacter les services qui produisent les vidéos des interprètes signeurs pour pouvoir les récupérer).

Mais les corpus resteront faibles, car selon FTV, « *si on récupère tous les jours les 7 à 9 min de Télé Matin, et l'après-midi sur France Info, cela peut faire 1 heure à 2 heures par jour.* »

HC propose d'intégrer certaines chaînes (à sélectionner) dans l'échantillon de l'étude benchmark et d'organiser un groupe avec des experts sur ces questions, si cela rencontre l'intérêt des partenaires du projet.

2.9. La planification des tâches

Pour l'ensemble des partenaires, la planification des tâches suit les livrables et deadlines prédéfinis dans l'annexe technique du projet. La priorité pour le LIMSI/Mocpalab et Systran est d'enregistrer le corpus au plus tôt pour fournir les données nécessaires à la suite du projet. Idéalement, une première partie du corpus devrait être tournée dès avril 2019.

La vidéo n'est pas la solution, mais elle permet, selon le LIMSI, d'évaluer les décalages (éventuels) avec les sous-titres. Car si ensuite il y a production des sous-titres de la langue des signes, on ne connaît pas nécessairement l'empan des sous-titres est inconnu, quels décalages il y a typiquement avec ce qu'il y a de l'interprétation versus ce qui vient de la traduction. Pour le LIMSI, l'évaluation de ces vidéos correspond à « *prendre la température dans cette mare de données* ».

Pour SYSTRAN, la traduction simultanée en pratique constitue un bon exemple de synchronisation entre parole et la LSF. S'ils arrivent à bien détecter au niveau de la vidéo les signes qui sont faits, par exemple « *si on peut bien passer d'une personne en train de signer a votre représentation Azee, ce sera un bon exemple de données parallèles entre le texte et la voix et de l'autre côté la représentation* ». Le LIMSI dispose d'exemples de traduction français/langue des signes, et l'ordre dans lequel ils ont le contenu. Parce que en langue des signes on va investir un peu de temps pour positionner les choses dans l'espace comme ça après tout est facile, même si ça ne vient que très tard dans le texte. C'est quelque chose qui serait idéal mais qui souffre de la situation d'interprétation simultanée.

Le LUTIN Userlab est très dépendant des partenaires et souhaite les sensibiliser aux tâches d'évaluation qui nécessite dans la mesure du possible des (pré)maquette et (pré-)prototypes.



Pour la collecte et la mise à disposition de corpus vidéos (avec une préférence pour des vidéos « journalistiques » en raison de leur variété de sujets), les partenaires ont décidé de s'organiser entre eux en fonction des lots².

2.10. Les attentes face aux autres partenaires

De manière générale, les partenaires s'attendaient à une réflexion commune à conduire sur le partage de la propriété intellectuelle et l'exploitation commerciale à la suite du projet (idée avancée par Mocaplab) ainsi qu'une mise à disposition des techniques existantes et des résultats (notamment pour le LUTIN Userlab qui doit les « tester »).

Par ailleurs se sont profilés des attentes spécifiques par rapport à un ou plusieurs partenaires en particulier qui seront clés pour la réussite du projet. Dans ce qui suit, chaque tableau reflète les attentes exprimées face à l'un des partenaires au moment du démarrage du projet. En toute transparence, et pour mieux comprendre ces compétences attendues, nous avons également mis en évidence qui parmi les partenaires a exprimé ses besoins.

Attentes face à SYSTRAN (en provenance de)		
FTV Accès	SYSTRAN	LIMSI
<ul style="list-style-type: none">Automatiser la mise à disposition de ses corpus (vidéo + fichiers ttml et stl) de manière fluide et intelligible pour l'ensemble de ses partenaires.Une fois cette étape réalisée, retour de la part de Systran de fichiers de sous-titrage en français pour évaluation de notre part par rapport à la Charte de qualité du CSA.	Exigence envers soi-même : <ul style="list-style-type: none">L'acquisition d'expérience sur une nouvelle tâche de traitement automatique des langues,La génération de sous-titres automatique.	Echanges réguliers sur les choix technologiques et comparaison des approches.

Tableau 12 : (A) Attentes face à Systran

Attentes face à FTV Accès & la distribution télévisuelle, enjeux des ST automatiques (en provenance de)

² Le LIMSI souhaite définir rapidement la nature des types d'émissions qui seront intégrées dans le démonstrateur final, car cela va déterminer le contenu du corpus de mocap qui va être réalisé (objectif : en avril 2019). Proposition : un magazine ou un documentaire. Par exemple l'équivalent de « Le dessous des cartes » de Arte. Intérêts : il est plus facile d'intégrer un signeur virtuel car il y a soit la journaliste, soit une animation de cartes, intéressant pour les aspects linguistiques en LSF car entités nommées et spatialisation.



MOCAPLAB	SYSTRAN	LIMSI
<ul style="list-style-type: none"> • Aide à la bonne définition des cas d'usage du projet • Même attente envers le Lutin Userlab 	<ul style="list-style-type: none"> • Mise à disposition des données avec un niveau d'information suffisant sur les émissions, • spécifications pratiques des modules attendus 	<ul style="list-style-type: none"> • Liste et contacts des émissions du stock qui contiennent de la LSF (exemple « tout info, tout en signes »). • Contact de franceinfo pour récupérer les JT en LSF.

Tableau 13 : (B) Attentes face à France Télévisions Accès

Attentes face à MOCAPLAB (en provenance de)	
SYSTRAN	LIMSI
Côté LS : collaboration scientifique entre le LIMSI et Mocaplab pour la représentation de la LSF et la création de l'avatar signeur	Côté LS : Un travail de réflexion en commun sur la convergence Azee et génération par mocap

Tableau 14 : (C) Attentes face à Mocaplab

Attentes face au LIMSI (en provenance de)	
MOCAPLAB	SYSTRAN
Pour l'élaboration du contenu du corpus, propositions sur le format des consignes de mouvements à combiner pour animer l'avatar 3D	Collaboration scientifique pour la création, mise en place et coordination des modules de génération de sous-titres textuels et LSF

Tableau 15 : (D) Attentes face au LIMSI

Attentes face au LUTIN Userlab (en provenance de)		
MOCAPLAB	SYSTRAN	LIMSI
Aide à la bonne définition des cas d'usage du projet (par LUTIN + FTV)	Définition précise des besoins utilisateurs et partenaires afin de les répercuter dans la conception de solutions techniques	Se coordonner sur les aspects <i>brainstorming</i> , focus groupes et évaluation afin de pouvoir alimenter les questionnaires sur nos questionnements propres concernant par exemple l'animation et nous aider sur les aspects ergonomiques concernant le logiciel de TAO.

Tableau 16 : (E) Attentes face au LUTIN Userlab



2.11. Les ingrédients pour réussir le projet

Les conseils donnés par les partenaires pour réussir le projet sont de bons sens et se fondent sur l'écoute, l'interaction, la pro-activité et l'intégration de toute la chaîne de valeur, dont les utilisateurs concernés tout au long du projet :

Structure	Les conseils des partenaires pour réussir le projet
FTV	D'y croire à fond et de se donner les moyens d'y parvenir.
Mocaplab	Les possibles exploitations des résultats en fin de projet seront très dépendantes des niveaux de qualité atteints qui sont très difficiles à estimer à ce stade. Il serait peut-être pertinent d'imaginer plusieurs cas d'usage et scénarios avec des niveaux de difficulté graduels.
Systran	Etre à l'écoute des contraintes de chaque partenaire, mais être ferme sur la communication régulière et la production de livrables sur les avancées, afin que le projet avance réellement tout en prenant sa juste place dans l'emploi du temps et les activités de chacun.
LIMSI	Encourager la participation inclusive : participants Sourds le plus tôt possible et tout au long du projet. Non seulement pour la langue des signes mais également sur la question de la lisibilité des écrits produits pour les malentendants.
LUTIN	Une collaboration inter-partenaires sur l'intra-partenaire.

Lors de la réunion de présentation s'en est suivie une discussion sur l'importance de déterminer la nature des types d'émissions qui seront impliqués dans le démonstrateur final, déterminer les contenus du corpus mocap.

Il faudrait aussi déterminer ce qu'il faut tester tant sur ces émissions que sur les maquettes à venir.

Un débat s'est engagé sur la nature des traductions vs. simplification. La simplification serait a priori à éviter, car en général, une tâche de traduction c'est une tâche qui consiste à dire dans la langue cible la même chose que dans la langue source. Si simplification il y a, celle-ci ne doit rien perdre par rapport à la langue source.

Il a été souligné l'importance de comprendre ce que l'interlocuteur de SYSTRAN attend de chaque partenaire dans ce domaine (entre autres). Selon le LIMSI et Mocaplab, la traduction du français vers la langue des signes (et vice versa) pose problème, donc il leur faut l'avis de SYSTRAN avant tout démarrage.

Les actions à envisager étaient de mettre en place des rendez-vous entre certains partenaires concernés pour avancer ensemble sur les tâches qui les concernent.



3. Conclusion et perspectives

En fonction des objectifs des rôles exprimés, les enjeux du projet, que ce soit sur le plan de l'innovation ou économique, de l'impact attendu ou les besoins pour atteindre ces enjeux, il a été clairement exprimé le souhait de favoriser les échanges réguliers et inter-partenaires.

Cela permet de mieux cerner les besoins sur court et long terme de chacun et notamment d'anticiper et de résoudre des points de blocage prévus ou non, des échanges d'ordre technologique, organisationnel ou autre. Des échanges réguliers sur les choix technologiques et une comparaison des approches liées à la méthode tout comme s'entraider à prendre les bonnes décisions est une demande récurrente, ou encore des aller-retours pour les solutions développés (à développer).

La transparence, la mise à disposition de données ou encore le montage de plusieurs cas d'usage et de scénarios avec des difficultés graduels font partie des souhaits d'interaction entre les partenaires. Nous avons déjà commencé à intégrer certains des souhaits ou besoins exprimés dans le cadre des brainstorming et focus groupes avec les utilisateurs et experts dans le cadre de ce lot 7. Par la suite, nous préconiserons des échanges plus rapprochés entre les partenaires pour les focus groupes et tests à venir.

Par ailleurs, les cibles « clients » ou « utilisateurs » des partenaires seront considérées dans le cadre de l'étude benchmarking à venir,

Au delà, nous préconisons au moins une réunion par trimestre avec l'ensemble du consortium, notamment pour comprendre l'état d'avancement du projet et afin de pouvoir adapter les étapes suivantes et organiser les focus groupes tests in situ en fonction des besoins et attentes des partenaires.

L'identification des lieux de test et de l'échantillon n'a pas encore été abordée, car le projet n'était pas suffisamment avancé pour cette approche au moment de la collecte des données et de la présentation des données recueillies.



Glossaire

Allophones	Personnes dont la langue première est une autre langue que la ou les langues officielles d'un pays, et qui réside habituellement sur ce territoire
Rosetta	RObot de Sous-titrage Et Toute Traduction Adaptés
TAO	Traduction Assistée par Ordinateur
TSA	Troubles du Spectre de l'Autisme



Annexes

Annexe 1 : Guide d'entretien

Questions pour les partenaires du projet Rosetta

Ce questionnaire prépare

L'Action 7.1 - Phase préparatoire : brainstorming entre partenaires, recueil des besoins, mise en place des services, identification des lieux de test et de l'échantillon

Cibles : Partenaires du projet ROSETTA

Nous procédons à une phase préparatoire de brainstorming ou de recueil des besoins et intentions des partenaires par rapport à la constitution de la maquette et des services à constituer, éléments structurants et nécessaires pour développer la prémaquette avant l'expérimentation. Il s'agit de pouvoir tester des services et contenus projetés.

Le projet ROSETTA lui-même propose :

4. d'automatiser la chaîne de production de sous-titres multilingues de contenus audiovisuels,
5. de fournir une représentation en Langue des Signes Française (LSF) de ces contenus via l'animation d'avatars signeurs virtuels,
6. d'intégrer ces technologies dans un produit finalisé directement utilisables par les utilisateurs.

Objectif du questionnaire

Préparation de la pre-maquette et préparation de certaines étapes susceptibles de suivre dans le cadre du projet Rosetta.

Méthode :

- (1) Recueillir des avis, besoins et desiderata des partenaires.
- (2) Recueillir éventuellement l'un ou l'autre avis extérieur (autre chaîne TV, client, fournisseur, groupement d'utilisateur, etc.).
- (3) Discuter des éléments obtenus lors d'une prochaine réunion projet pour confronter les idées et fixer les suites de la pré-maquette.
- (4) Rédiger le livrable en fonction des éléments obtenus.

Nous vous remercions d'avance pour votre participation et la réponse aux questions ci-dessous !

1. Quels sont vos objectifs (celle de votre organisation) dans ce projet ?

Pouvez-vous aussi préciser votre rôle (ou vos rôles, si plusieurs interlocuteurs) dans le cadre du projet ?

2. Pour votre entreprise ou organisation, quels sont les enjeux du projet ?

- Sur le plan de l'innovation ?
- Sur le plan économique ?
- Concernant l'impact attendu ?
- Autres ?

3. Quels sont vos besoins dans le cadre du projet pour atteindre les enjeux développés ci-dessus ?

4. Sur quel(s) projet(s) de R&D&I antérieur(s) comptez-vous vous appuyer afin de construire sur les résultats et recommandations obtenues (déjà existants) ?



5. Pour votre entreprise/organisation, à qui s'adresse la ou les solutions qui vont être développées dans le cadre du projet ?

6. Qui voudriez-vous interroger dans le cadre du projet (pour la pré-maquette, ou dans le cadre de l'enquête prévue ultérieurement) ? Autre chaîne TV, client, fournisseur, confrère, groupement d'utilisateur, régulateur, autres ? Qui le cas échéant ?

7. Quels buts voulez-vous atteindre sur le long terme (5 ans) après le projet ?

8. Voyez-vous des points bloquants pour le moment ? Et durant le projet ?

- Sur le plan technologiques. Lesquels ?
- Dans le domaine de la recherche. Lesquels ?
- Autres ?

Idéalement, que faudrait-il faire pour arriver à débloquer ces points ?

9. Comment comptez-vous planifier vos tâches ?

10. Qu'attendez-vous de chaque partenaire pour vous interfacier avec chacun ?

(ex. : mise à disposition des base de données de signes, mise à disposition d'un fichier d'utilisateurs, techniques utilisées pour la motion capture, etc

Quelles sont vos attentes face à :

- Systran :
- MFP/FTV/la distribution télévisuelle et les enjeux du ST automatique (et LS) :
- Mocalab (côté LS) :
- LIMSI :
- LUTIN :

11. Quels conseils donneriez-vous pour réussir le projet ?

Souhaitez-vous ajouter quelque chose ?

Merci pour votre coopération.



Annexe 2 : Quels traducteurs automatiques ?

Note de retranscription (par A. Bhatti, Lutin Userlab) :

Lors de la discussion autour des rôles de chaque partenaire s'est poursuivi un questionnement autour de la mise en place des traducteurs automatiques :

Systran souhaitait savoir :

- combien de phrases représentent un corpus de 5 minutes par exemple ?
- S'il y a l'idée de faire procéder en deux étapes d'abord à partir des sous-titres générés en français et passer à la LSF, et aussi de passer directement de la vidéo à la LSF ?

En guise de réponse :

- Selon MOCAPLAB, SYSTRAN intervient aussi dans la tâche 5, pour la traduction française vers la LSF, pensant que SYSTRAN pourrait passer du français à la représentation formelle-intermédiaire qui sera proposée par le LIMSI.
- Selon le LIMSI, « la traduction automatique fondamentalement suppose un ordre des éléments de la langue source et aussi de la langue des signes et la représentation ici est plutôt arborescente, ce n'est pas quelque chose qui peut être réduit à un ordre par exemple. Pour ça il y a un changement à sortir, même si on parle de quelques choses de textuelle, et si on parle de vidéo, double peine, on ne part ni d'un ordre d'entrée et ni d'un ordre de sortie, là, à mon avis il n'y a rien qui existe, qui marche. Je ne suis pas sûre qu'on peut promettre quelques choses sur ce terrain, mais par contre et regarder et vouloir aller se serait une partie exploratoire qui serait passionnant. On peut regarder à partir d'une vidéo avoir la représentation linguistique en LS paraît difficile, il n'y a rien qui existe. »

SYSTRAN suggère de passer d'abord par exemple la détection de l'entité nommée, pour que soit un point d'ancrage qui puisse «simplifier» la synchronisation.

LIMSI : Cela serait plutôt pour assister à la traduction assistée, c'est un peu l'approche que l'on a nous, c'est de dire dans le projet, mettre une partie, une thèse en l'occurrence sur la traduction assistée c'est-à-dire que quelque **part il y a un humain qui va traduire mais avec un logiciel qui l'aide, par exemple en détectant les entités humaines, en détectant les liens entre les choses, en permettant d'ordonner des choses, etc. comme il existe des logiciels de traductions assistés de texte en texte.**

Si c'est des informations qui sont sur un compteur et qui seront lues, soit c'est le fil de sous titres qui



a été reconnues. Sa dépend si c'est pour traduire ensuite, il faut avoir quelques choses, je ne sais pas ce que ça vaut, est ce que le traducteur voudrait avoir des informations simplifiées. Pour un sous-titrage c'est très bien mais comme source pour un traducteur humain qui voudra revenir dans un équivalent en LSF, je ne suis pas sûre que de dégradé d'emblée ce que l'on nous propose comme ressources soit la meilleure solution, **un traducteur au contraire voudra a priori avoir la version pure.**

LIMSI : On n'a pas vraiment envisagé qu'ils partent de l'audio, parce que les pratiques dans les métiers de la traduction c'est de partir des textes, parce que quand on fait audio vers langue des signes, ce qui ne veut pas dire qu'il ne faut pas le faire mais **l'idée était plutôt de proposer un équivalent de ce qui existe dans le monde du texte, dans la traduction textuelle.**

La traduction depuis l'audio en général, c'est une situation d'interprétation simultanée, il n'y a pas de logiciels où on travaille les traductions y a un truc ou on signe à côté, c'est de l'interprétation en fait.

LIMSI : Sa dépend complètement de l'émission, si c'est style Télé matin, c'est du texte qui est préparé peu avant, et c'est aussi lu par l'interprète peu avant, puis l'interprétation se fait simultanément, mais du coup en connaissant un peu du contenu, c'est un mixe. Mais il y a des émissions, qui sont complètement préparées à l'avance, se sont de vraies traductions. Il y a des cas, où on a filmé l'interprète qui improvise sur place sans savoir de quoi il va parler. Donc là, nous, si on fait un logiciel pour assister quelqu'un, c'est évidemment en situation de traduction, qui nous prépare ensuite en vue d'être monté dans un truc. L'ambition n'est pas d'assister l'interprétation simultanée.

MOCAPLAB : SYSTRAN a un rôle, tout un travail, sur un module qui permet de passer d'un écrit à une représentation de la LSF, c'est un truc auquel, nous, MOCAPLAB on s'intéressait ; on est un peu loin du projet mais si on ne comprend pas ce qu'il se passe à ce niveau-là, ce qui se passe à ce niveau-là nous intéresse aussi pour d'autres idées, ou pour favoriser certaines traductions, pour moi c'est un gros morceau du projet. **Le rôle de SYSTRAN pourrait passer d'un écrit vers un morceau de traduction.**

SYSTRAN : en ce moment c'est ce que l'on est en train de mettre en place, avec l'équipe de LIMSI, pour avoir justement à disposition de tous les partenaires à la fois le fichier de sous-titres que vous fournissez, et une transcription verbatim, et donc forcément automatique et potentiellement avec certaines, mais un document de travail de départ, c'est apprendre tout ce qui est du type d'adaptation qui attendaient des sous-titres, en faites pour permettre un modèle d'attente d'opération. **On préfère partir d'un format « texte », de texte à texte,** sachant que justement dans les côtés technologiques



on va de plus en plus vers des modèles qui prennent le format idéal en entrée et le format idéal en sortie, et qui justement essaye de faire plusieurs tâches en même temps donc du coup je sais qu'il y a une étude de François Hillon qui s'est intéressé aussi de passer de la vidéo directement aux sous titres et de travailler ses aspects-là. Et c'est dans ce sens-là que j'avais compris du côté de SYSTRAN, on **fournit les sous-titres qui seraient générés en textuel du coup**, qui pourraient **à la fois servir du côté langue des signes française et comme un point d'entrée plus facile à la vidéo**, et qu'il y avait aussi une idée de passer de la vidéo à la représentation LSF directement, il y avait les deux voies qui étaient envisagées, et que forcément **partir d'un format texte est plus simple pour réduire la complexité du départ**.

LIMSI : Les gens qui font la reconnaissance de traitement de vidéo, ce n'est pas moi, mais même d'un petit peu plus loin ça paraît difficile d'imaginer que ce soit réaliste dans le temps du projet, partir directement d'une vidéo et d'avoir sa représentation linguistique de langue des signes, ça me paraît difficile et même depuis le texte, mais ça ne veut pas dire qu'il ne faut pas essayer.

Dans ce cadre-là, j'ai l'impression **qu'il y a deux questions, en faites à partir des sous-titres**, par exemple si on part du texte vers une représentation de la LSF, là il y a un rôle, une représentation à trouver du coup qui vous intéresse du côté de la LSF et donc moi ; j'avais eu l'impression qu'on nous pas proposé de représentation.

LIMSI : mais ça on l'a, partir d'un verbatim (la reconnaissance vocale) sans traitement, et arriver sur de la forme ça je pense sans forcément être l'expert de l'état d'art en la matière qu'il n'y a pas.

MOCAPLAB : **Nous on va le concevoir, et il faut qu'on sache comment il va servir**, est ce qu'on fait tous les mots du dictionnaire, est ce qu'on fait 50 phrases sur des situations différentes, est-ce ce que l'on fait. Qu'est-ce qu'on veut valider, et qu'est-ce qu'on veut évaluer, et les objectifs du projet, en tout cas de notre point de vue, sur la partie LSF, on sait qu'on s'attaque à quelque chose d'extrêmement difficile, si on dit on va connecter les deux.

Ce que l'on veut évaluer dépendra de ce que l'on peut viser, et ce que l'on va être capable de faire ... on ne le sait pas trop, donc on est plutôt coincé, comment on peut évaluer quelque chose dont on ne connaît pas la nature et en même temps c'est la problématique et c'est là qu'il faut que l'on discute maintenant, et je me suis intéressé du retour de SYSTRAN et qu'ils nous disent un peu ce qui est envisageable ou pas, ce qui est accessible ou pas, et on sait comment on se situera par rapport à ça.

FTV : au niveau des corpus, je vais pouvoir avoir l'émission régionale, ainsi que les émissions de France



infos, tous les jours il va y avoir des corpus LSF, en plus de télé matin.

Annexe 3 : Quelle automatisation envisager ?

Note de retranscription (par A. Bhatti, Lutin Userlab) :

SYSTRAN : on est orienté du côté recherche, en tout cas sur la tâche textuelle pour la transformation des textes et on a peu investigué l'autre partie du texte vers un autre format de sortie, on n'a jamais vu (l'équipe de recherche), entre nos mains cet autre format de sortie, donc on n'a pas spécialement investigué sur ce que l'on pourrait faire. Du côté de l'entreprise, on essaye de se débarrasser des règles, et d'aller plutôt vers des algorithmes de types Learning et justement d'apprentissages, le format d'entrée ou de sortie n'est pas forcément un problème à condition qu'on ait beaucoup de données. Mais c'est vrai que vous de votre côté si vous attendez un retour avec plus de règles, il faut donc qu'on en parle et qu'on voit ce que l'on peut faire, ce qui est réaliste et ce qu'on va pouvoir concevoir.

MOCAPLAB : il faut du texte versus de la représentation formelle intérieure, donc il n'y a pas besoin de MOCAP.

On l'a dit dès le début que c'était d'une part très compliqué, deux niveaux de Deep Learning combinés, et il faudrait en plus des milliards de données.

LIMSI : C'est un peu illusoire, il y a quelque chose de fondamental, de sémantique, dans ce que l'on fait en langue des signes. Vouloir traduire sans comprendre avec simplement des exemples, je pense qu'il y a une productivité à la langue des signes.

LIMSI : Nous on avait prévu des choses, et nous on devait réanimer des choses à partir de MOCAP enregistré et en partant des représentations c'est-à-dire en gros réutiliser des choses enregistrées et vous, vous aviez prévu de rentrer dans justement cette représentation, en apprendre les aboutissants et voir comment elles marchent. Du coup, là, il y aurait de quoi voir et apprendre et de voir parce que oui il y a ce problème, pour éviter de discuter trop longtemps, je ne peux pas m'empêcher de dire l'expérience de pensée suivante ? Et si on les avait ? Et si on avait des millions d'heure représentée dans le formalisme, et traduit ? Quand même je me pose la question d'un point de vue théorique, est ce que le paradigme est le même ? Est ce qu'on pourrait faire du Deep Learning là-dessus ? Je ne suis pas sûre, parce que justement il y a tellement de sémantique dans la manière de produire la langue des signes, c'est gouverner par, pas juste les mots d'entrés, c'est gouverner par ce que l'on veut dire, et comment on veut le montrer et c'est ça qui fait la richesse, et le point de vigilance de toute la



communauté sourde, et le point de vigilance de tous les traducteurs du métier vers la langue des signes, ils font attention à tout ce qui est illustratif et propre à la langue des signes, que on va être juger, ce n'est pas quelque qu'on peut faire sans, et ça sa demande d'avoir un contexte, je donne un exemple la lave coule le long du volcan, se sera différent si cela coule vers la mer, ou le long d'un autre truc, ou bien le bateau accoster par exemple de Somalie ça va être d'Est vers l'Ouest et non pas du Nord vers le Sud comme si il arrivait vers Alger, et ça c'est connaissance du monde qu'il faut avoir et qui ne seront pas donné par des millions d'exemples sauf théoriquement à dire qu'il peut y avoir des millions d'exemples avec des bateau, avec Alger, etc. .. Du point de vue théorique on aime dire que la langue est une productivité infinie et que les entrées sont des entrées uniques à chaque fois, et donc on est en train de montrer qu'on ne peut pas avoir un nombre suffisant d'exemples et que donc la sémantique est incontournable, donc je ne sais pas quoi faire à partir de cela, est ce qu'on prend des parties de phrases et on les applique à un humain, etc. j'aimerais votre avis ?

MOCAPLAB : Comment on sait que même si on a des millions, et qu'on sait qu'on va saturer, la question est quelle est la mise en jeu, soit c'est une aide et y a un humain qui intervient, et on se dit que c'est juste une aide à un humain pour la traduction, soit on se dit que dans certains cas y a un niveau de qualité.

LIMSI : soit on peut le faire, mais a un choix prêt, et après il n'y aura qu'à choisir, cela se pourrait être une manière de vraiment accélérer le truc en disant vers ou la lave coule et je le fais.

SYSTRAN : pour la lave qui coule vous avez en face plusieurs représentations, de « elle coule par là ou par la », dans ce cas un choix est envisageable, mais si vous avez qu'un seul exemple de « une fois lave qui coule » et si elle coule par la, nous on va apprendre quelques choses de différent si vous avez un mot et une traduction.

MOCAPLAB : **la question est, est ce que c'est une aide à un traducteur humain, ou est ce qu'il y a une version qui fait la moitié du boulot, qui s'arrête fait des conclusions, et est automatique.** Cela peut dans certains contextes être intéressant parce que quand on ne comprend rien à une langue, Google traduction est mieux que rien.

SYSTRAN : **quel type d'automatisation est acceptable ?** Je me souviens d'analyses qui disaient qu'un mouvement trop mécanique est détectée par une zone du cerveau différente, et ça par exemple ce n'est pas acceptable. Mais par exemple dans la reconnaissance *Texte to speech*, une voix mécanique



n'est pas agréable à l'oreille, on arrive à sentir que ce n'est pas une voix naturellement humaine mais c'est acceptable à défaut du contexte.

LIMSI : il y a un peu deux niveaux. Moi je ne serais pas pour vous dire que n'importe quel caprice doit pouvoir nous bloquer, mais il y a aussi une réalité, là c'est un peu plus politique, si on part du principe qu'une traduction pas forcément propre et jolie ; il y a un niveau d'erreur acceptable parce que c'est une traduction automatique, c'est déjà vrai pour le texte, le problème c'est que là on est dans une langue qui a été carrément discriminée, il y a une situation politique qui n'est pas vraiment la même, et si l'on se dit que tout ce qui est la richesse de la langue des signes, je parlais de point de vigilance quasi obsessionnelle parfois, on peut pas faire les mêmes arbitrages simples, y a des choses dans lesquels on peut arriver parfois. Bon de toute façon tu vois de quoi ça parle, mais je ne suis pas convaincue que la communauté des bénéficiaires supposés acceptent la même chose.

MOCAPLAB : il y a deux niveaux de problématique sur la qualité finale, la qualité avec acoustique de la parole qui sort et une voix synthétique ... c'était désagréable, cela n'avait rien avoir avec le sens où les mots qui enchaînaient c'était la qualité acoustique du signal qui était désagréable, et dans la langue des signes cela peut être pareil dans la qualité de langue au niveau gestuel ; les sourds sont à mon avis très exigeants dans le respect de fluidité, il y a des gestuelles désagréables. Ils seraient comme nous face à une parole désagréable, et si par contre la traduction fait des erreurs, des faux sens, des approximations, mais que la qualité du signal fait que c'est traduit par un enfant de 5ans, mais que la qualité acoustique/gestuelle n'est pas insupportable, sont deux paramètres très différents, c'est pas à nous de faire l'analogie avec le son et dire on sait ce qui l'en est, et en même je pense qu'il faut être extrêmement vigilant pour plein de raisons **il ne faut pas faire n'importe quoi mais les sourds eux-mêmes ne sont pas forcément conscients** de ce à quoi ils se sont .. Je parle au niveau perceptif un mauvais mouvement est désagréable.

MOCAPLAB : Un point très délicat du problème, parce que si dès le début on se trompe et de l'usage possible, et des méthodes, il faut qu'on trouve très tôt le réglage. Il y a dans la langue des signes un coin fragile.



Annexe 4 : Réflexions autour de la proximité d'une LS vers une autre et simplification

Verbatim retranscrits (par A. Bhatti, Lutin Userlab) :

LIMSI : Est-ce que la question c'est, que tout ce qu'on aura à faire c'est de passer vers la langue des signes, est ce que **c'est simple de passer vers une autre langue des signes ? La réponse est oui, et beaucoup plus qu'une autre langue parlée. Toutes les LS** ont un dialecte d'une même langue, à prendre avec des pincettes, mais elles **ont une proximité**. C'est des langues qui émergent régulièrement d'ailleurs dans des groupes assez spontanément, et qui ont quelques choses de tellement humains, enfin font appel à une projection en 3D, il y a beaucoup moins de linéarisation, donc on se retrouve proche à chaque fois de quelques choses, y a pleins de choses qui diffèrent mais ces différences ne sont pas la structurantes que ça a la lumière **des modèles que l'on développe pour les représenter, ce qui est structurant sont partagés**.

Aucune langue des signes dit par exemple « à droite de » sans mettre à droite de parce que c'est juste normal, c'est immédiat, c'est humain et dans tous les **processus d'évolution des langues le principe d'économie rentre évidemment en jeu, et ça répond aux deux critères qui sont tensions**. qui est d'être productif et économique à la fois, **collectivement les langues vont évoluer pour être toujours simplifiées, mais on perd à ce moment-là en expressivité, en productivité, en subtilité**. En précisant, c'est donc une autre force qui tire, qui dit qu'il faut pas trop simplifier le vocabulaire, qu'il faut garder les nuances, et en langue des signes « à droite de », tu mets « à droite » c'est ce qui a de plus simple, de plus productif, d'immédiat, et tu peux même dire « à droite de mais un peu moins », c'est hyper précis tout en étant hyper expressif, tout en étant hyper économique. C'est normal que sa se retrouve dans toutes les langues de signes, ce n'est possible que grâce au canal gestuel, enfin grâce à son nombre de dimensions disponibles.

LIMSI : ça dépend ce qu'on veut internationaliser, parce que si c'est un système qui passe vers une représentation qui suffirait de modifier pour coller à une autre langue, en faites je parlais tout à l'heure de politiquement correcte, mais chaque linguiste de communauté de signes va avoir son petit terrain et dire je ne sais pas pour la langue de signes de machin, mais chez nous c'est comme ça.

Encore un truc politiquement incorrecte c'est parler de **langue de signe internationale** parce que les



codes ne sont pas forcément partagés mais on dit *International Sign Language*, ça fait des années qu'il y a des conférences internationale en langues des signes, qui mettent tout le monde d'accord **grâce au *International Sign***, et ça ne peut se faire que grâce à ce qu'on disait que c'est quelques choses de plus ou moins proche, et il y a quelques empreintes qu'ils ont avec la langue des signes américaines parce qu'ils dominent le monde, etc. mais c'est assez mineure, et **on va faire usage justement de tout ce qui est commun en langue des signes**, et c'est tellement large, que en faites on arrive à parler avec presque que ça, et donc on ne parle pas de langue des signes internationales.

LIMSI : il faut bien comprendre que c'est une langue qui émerge entre gens qui ne parlent pas la même langue, mais qui vont se comprendre, le truc c'est que ça commence à inclure plein de monde dans le monde, le statut de ce truc n'est pas encore clair.

Mais comme les langues de signes ne sont pas très différentes, ça vient vite, avec un effort minime.

La nécessité de linéariser et donc de trouver un code pour le faire, et donc de rendre la chose non universelle est plus forte en langues des signes.



Consortium

france•tv
access

SYSTRAN
beyond language

MOCAPLAB

ÉCOLE PRATIQUE
des HAUTES ÉTUDES | PSL

Lutin Userlab
Cité des sciences et de l'industrie

LISN
LABORATOIRE INTERDISCIPLINAIRE
DES SCIENCES DU NUMÉRIQUE

technologies DaIA
data - intelligence artificielle
Partenaire valideur non financé

bpifrance
Co-financeur

Holken Consultants & Partners
Sous-traitant valorisation