



ROSETTA

accès multilingue

RObot de SOus-titrage Et TOUTE Traduction ADaptés

Livrable 7.4.1 État de l'art international sur l'accessibilité automatique

Rapport issu des recherches documentaires

Benchmarking des sous-titres automatiques et génération de contenu en langue des signes





Durée de projet 42 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 0.00	30/07/2019	Rédigé par HC
V 0.01	02/08/2019	Relecture, ajouts, modifications par HC
V0.02	16/08/2019	Intégration des observations du LUTIN USERLAB et ajout de la méthode par HC
V0.03	27/09/2019	Rédaction du rapport et intégration des éléments-clés issus des recherches documentaires par HC
V0.04	30/09/2019	Finalisation et ajustements par HC
V0.05	11/10/2019	Intégration des observations du LIMSI finalisation du pré-rapport documentaire par HC
V0.06	03/03/2020	Préparation du rapport final sur la base du pré-rapport. Recherches documentaires complémentaires : les Amériques, monde anglophone, ajout de l'état de l'art, réorganisation et ajouts dans le chapitre « hégémonie américaine », intégration des apports du LIMSI, MOCAPLAB, Lutin par HC
V0.07	07/04/2020	Recherches documentaires complémentaires : Chine, Europe, autres régions du monde. Considération des revues littéraires, de robots signeurs, intégration des apports du LIMSI, MOCAPLAB, Lutin par HC
V0.08	24/04/2020	Recherches documentaires supplémentaires quant aux usages potentiels des outils d'accessibilité audiovisuels automatique par HC
V0.09	29/04/2020	Mise à jour finale avec apport des annexes : Webographie en tableau, références également mises à jour (figures, tableaux, bibliographie) et index des sigles par HC
V.01	07/05/2020	Ajout du graphique heuristique, finalisation, conclusion, <i>Executive Summary</i> et mise à disposition du document aux partenaires par HC
V.02	20/08/2020	Intégration des observations de MOCAPLAB, LIMSI, LUTIN USERLAB par HC
V.03	07/10/2020	Intégration des modifications de LUTIN USERLAB, version finale par HC







Auteurs du livrable 7.4.1

LUTIN USERLAB

- Charles Tijus

HC

- Salima Body
- Marie Chauvière
- Naël Chehab
- Hadmut Holken
- Mathilde Malgrange

LIMSI

- Annelies Braffort
- François Yvon

MOCAPLAB

- Boris Dauriac

Ce livrable répond à la tâche 7.4.1

Afin de relever les autres pratiques ou projets en Europe et à l'international, il convient 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.

Mots clés

- Secteur(s) d'application : benchmark, audiovisuel, télévision, plateformes de contenus, formation, communication, information, inclusion, design pour tous.
- Domaine(s) techniques : intelligence artificielle, apprentissage profond, Big Data, apprentissage automatique, corpus, génération automatique des sous-titrages adaptés multilingues, génération de contenus en langue des signes, avatar signant, signeur virtuel, animation d'avatar, capture de mouvements (motion capture)



Table des matières

EXECUTIVE SUMMARY : ENSEIGNEMENTS-CLE	10
1 CONTEXTE.....	14
1.1 OBJECTIFS : BENCHMARK INTERNATIONAL	14
1.2 LE CONTEXTE GENERAL DE L'ETUDE	14
1.2.1 De l'accessibilité audiovisuelle	14
1.2.2 Une digitalisation globale et croissante au service d'une accessibilité pour tous ?	15
1.3 METHODE GLOBALE DU BENCHMARKING EUROPEEN ET INTERNATIONAL	16
2 CONSTATS PRELIMINAIRES : STREAMING VERSUS TELEVISION	17
2.1 LE STREAMING	17
2.2 LA TELEVISION, UN MODELE ROBUSTE	18
3 L'AVANCEE NORD-AMERICAINE SUR LES LOGICIELS DE RECONNAISSANCE VOCALE.....	20
3.1 GOOGLE – ANDROID.....	20
3.1.1 « Live Caption »	22
3.1.2 « Live Transcribe »	23
3.1.3 La Traduction en direct.....	24
3.1.4 Quelques nuances	24
3.1.5 Netflix : la diffusion de masse impose des normes standardisées	26
3.1.6 Le partenariat Google Chrome : Learning Languages with Netflix	26
3.2 LA RECONNAISSANCE VOCALE AU SERVICE DE LA TRANSCRIPTION AUTOMATISEE	28
3.3 LES AVANCEES ASIATIQUES	30
4 EUROPE : LES SOLUTIONS POUR L'ACCESSIBILITE AUDIOVISUELLE.....	32
4.1 PRODUITS FRANÇAIS EN MATIERE DE SOUS-TITRAGE AUTOMATISE	32
4.1.1 RogerVoice, sous-titres et télécommunication	32
4.1.2 Authôt, la retranscription automatique française.....	33
4.1.3 Fleex, solution française pour apprendre l'anglais.....	33
4.1.4 VoxSigma Suite (Vocapia Research ; un produit Linux)	34
4.2 LES OFFRES POUR UNE ACCESSIBILITE EN LANGUE DES SIGNES	34
4.2.1 SiMAX : la génération semi-automatisée d'un avatar signeur	34
4.2.2 Movistar + 5S – Solution numérique LS et SME (Espagne)	35
4.2.3 BBC iPlayer – une option pour la Langue des Signes	36
5 LES PROJETS SCIENTIFIQUES ET TECHNOLOGIQUES PUBLICS-PRIVES	36
5.1 DE L'ACCESSIBILITE EN GENERAL A LA CONCEPTION UNIVERSELLE	37
5.2 SOUS-TITRAGES : APPORTS DES PROJETS ACADEMIQUES ET COLLABORATIFS	38
5.2.1 La qualité passe-t-elle par la simplification et de la créativité ?.....	39
5.2.2 D'autres modalités techniques pour la génération des sous-titres.....	42



5.2.3	Axes d'amélioration possibles pour les sous-titres générés automatiquement	42
5.3	GENERATION AUTOMATIQUE DE LA LANGUE DES SIGNES	43
5.3.1	Travaux (non exhaustifs) portant sur l'acceptation des avatars.....	44
5.3.2	Représentations des LS basées sur la concaténation/séquence de signes lexicaux.....	45
5.3.3	Autres approches.....	53
5.4	LES ROBOTS AU SERVICE DE LA LANGUE DES SIGNES ?.....	55
5.4.1	Les robots humanoïdes	55
5.4.2	Concentration sur les mains robotisées	57
5.5	PRATIQUES ET NORMES DE L'ACCESSIBILITE DANS LE CADRE TELEVISUEL	59
5.5.1	En Europe	59
5.5.2	En France	60
5.5.3	En Belgique	63
5.5.4	En Espagne.....	64
5.5.5	Au Royaume-Uni.....	67
5.5.6	En Chine.....	68
5.5.7	En Australie.....	71
5.5.8	Dans les Amériques	71
6	OPPORTUNITES DANS D'AUTRES SECTEURS	76
6.1	NOUVEAUX JOURNALISMES SUR TOUS LES SUPPORTS	76
6.2	RECOURS AUX OUTILS D'ACCESSIBILITE DANS LE MONDE DE L'EDUCATION	77
6.2.1	Le projet « MOOC TEAM » : une éducation inclusive à l'environnement	77
6.2.2	Le projet européen TraMOOC	78
6.2.3	StorySign : plateforme d'alphabetisation pour enfants sourds.....	78
6.3	COMMUNICATION INSTITUTIONNELLE : APPEL A DES PRESTATAIRES DE SERVICES EXTERIEURS.....	79
6.3.1	La Commission Européenne fait appel à des interprètes extérieurs.....	79
6.3.2	Entreprises et institutions françaises qui sous-traitent leurs solutions de communication accessibles à des opérateurs-relais privés	79
7	CONCLUSION ET PROCHAINES ETAPES	81
7.1	CONCLUSION	81
7.2	PROCHAINES ETAPES	82
ANNEXES	84
ANNEXE 1	: LEXIQUE DES DIFFERENTS SIGLES.....	84
ANNEXE 2	: RAPPEL DES DIFFERENTS TYPES DE SOUS-TITRES	87
ANNEXE 3	: WEBOGRAPHIE THEMATISEE	88
REFERENCES.....	94







Figures

Figure 1. Typologie des outils accessibilité audiovisuelle	13
Figure 2. YouTube – Kurzgesagt - In a Nutshell (Great Filter).....	18
Figure 3. Nouvelles Fonctionnalités de Android 10	21
Figure 4. Fonctionnalités orientées vers l'accessibilité de Android 10	22
Figure 5. Illustration de l'utilisation de « Live Caption » sur contenu audiovisuel	23
Figure 6. Illustration de l'utilisation de « Live Caption » sur message vocal	23
Figure 7. Android 10 permet de la traduction automatique par image	24
Figure 8. Illustration de "Live Transcribe"	25
Figure 9. Capture écran de l'aide pour utiliser Learning Languages w/ Netflix	27
Figure 10. Capture écran choix de langues (LLN).....	27
Figure 11. Logos des ASR asiatiques, américaines et russe	31
Figure 12. Capture écran de la présentation des deux modalités offertes par RogerVoice (appels sous titrés et interprète LSF)	32
Figure 13. Capture écran des services proposés par Fleex.....	34
Figure 14. Les différents avatars visibles sur le site de SiMAX	35
Figure 15. Capture écran du service Langue des Signes, proposé par Movistar +.....	35
Figure 16. Capture écran du service sous titrage pour malentendants, proposé par Movistar +	36
Figure 17. Capture écran du service Langue des signes/ST par BBC iPlayer.....	36
Figure 18. Capture écran étude des sous-titres augmentés	40
Figure 19. Exemple 1 sous-titrages créatifs	41
Figure 20. Exemple 2 sous-titrage créatif	41
Figure 21. Exemple 3 sous-titrage créatif	41
Figure 23. Capture écran illustrant l'avatar pour les chemins de fer suisses SBB	46
Figure 24. Capture écran de l'illustration de la chaîne de procédés pour générer un avatar	48
Figure 25. Illustration des avatars humains du projet Danois Sand Castle.....	48
Figure 26. Capture de l'émission NPO (projet danois) : dinosaure et personnages de l'émission enfantine	48
Figure 27. HUGO, avatar de Hand Talk	50
Figure 28. Illustration des personnages du système SiSi	51
Figure 29. Capture écran de la vidéo de présentation du projet de Limix	52
Figure 30. Capture écran de la présentation du projet Signall	52
Figure 31. Capture Projet Avatar Signeur de l'équipe EXPRESSION (IRISA)	53
Figure 32. PAULA, de l'université DePaul	54





Figure 33. Capture du projet KEIA	55
Figure 34. Capture écran de la vidéo étude "Momo rencontre les enfants autistes"	56
Figure 35. Capture écran de la vidéo de présentation de ASIMO	56
Figure 36. NAO le robot (photo de présentation).....	57
Figure 37. Capture du document de présentation du projet malaisien SignBot	57
Figure 38. Robot ASLAN épelant la lettre L.....	58
Figure 40. Série – France Ô, exemple de sous-titrage	62
Figure 41. Pourcentage de programmes annuels sous titrés dès 2019.....	63
Figure 42. Quota de sous-titrage et de LSF pour la télévision belge	64
Figure 43. Utilisation des sous-titres en Espagne par genre en 2014-2015	65
Figure 44. Exemple LES (Langue des Signes Espagnols) sur RTVE.....	65
Figure 45. Exemple sous-titrage sur 2 blocs de lignes.	66
Figure 46. Indications musicales en haut à droite de l'écran	66
Figure 47. Représentation des différentes ethnies chinoises.....	69
Figure 48. Capture d'écran d'une vidéo de replay, déclaration officielle du premier ministre et traduction LS par deux interprètes.....	73
Figure 49. Capture d'écran allocution officielle uniquement anglophone et closed captions générées automatiquement.	74
Figure 50. Les normes graphiques du surtitrage des vidéos Brut.....	77
Figure 51. Démonstration de l'avatar signeur STAR de l'application Storysign qui vise à encourager l'alphabétisation des enfants Sourds.....	79
Figure 52. Illustration représentative des services proposés par les opérateurs-relais	80

Tableaux

Tableau 1. Présentation des différents produits issus des technologies ASR.....	29
Tableau 2. Extrait du livre blanc Omniscien montrant la différence de productivité par mots entre le travail humain et la Machine.....	38
Tableau 3. Programmes accessibles en 2018 pour les chaînes dont la part d'audience est supérieure à 2,5 % de l'audience totale des services de télévision (volumes horaires et pourcentages, hors publicité et dérogations)	60
Tableau 4. Programmes accessibles en 2018 pour les chaînes dont la part d'audience est inférieure à 2,5 % de l'audience totale des services de télévision.....	61
Tableau 5. Les engagements spécifiques des chaînes d'information en continu, 2018	61
Tableau 6. Code couleur sous-titrage du CSA.....	62
Tableau 7. Extrait des pourcentages obligatoires de sous-titrages pour les chaînes britanniques.....	68





EXECUTIVE SUMMARY : ENSEIGNEMENTS-CLE

Cette recherche documentaire 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).

Ce rapport est l'un des deux livrables du lot 7.4.1 Benchmark 1. Il a pour objectif d'obtenir une comparaison européenne et internationale des avancées techniques et, d'une part, de permettre aux partenaires du projet ROSETTA d'orienter leurs travaux en connaissance de cause et, d'autre part, de préparer la phase d'une enquête terrain approfondie ; Cette seconde donnant lieu à la réalisation du second livrable.

Sept enseignements-clé sont issus de cette recherche documentaire :

(1) Peu de projets similaires

Un premier constat est que peu de projets similaires existent ou aient aboutis. En effet, bien que la traduction automatique des sous-titres, - dominée par les « *pure players* » américains, voire chinois, soit très avancée, elle ne correspond pas à un cadre de qualité équivalent à celui qui est recommandé par le Conseil Supérieur de l'Audiovisuel (CSA) pour les sous-titres télévisuels en France. Par ailleurs, le développement de la génération de contenu en langue des signes semble nécessiter encore plusieurs années de recherche pour aboutir à une solution qui puisse répondre aux besoins des utilisateurs cibles.

(2) L'automatisation s'accélère par une dynamique forte du streaming

Dans les services de streaming, l'accessibilité par les sous-titres correspond en général davantage à celle de contenus disponibles en temps réel ou à la demande. Ainsi, le sous-titrage peut être complètement automatisé avec l'utilisation de techniques de la reconnaissance automatisée de la parole (ASR pour Automatic Speech Recognition). C'est le cas avec YouTube, plateforme d'hébergement de vidéos, qui permet un sous-titrage automatisé des vidéos, soit en direct (live) soit en différé. Autrement, les contenus sont sous-titrés rapidement, par des laboratoires ou par des amateurs, et tendent ainsi vers une standardisation des formats qui visent à faciliter plutôt la propagation et la diffusion des contenus.

Cette automatisation se développe depuis quelques années grâce aux progrès des machines de transcription et de traduction et de l'intelligence artificielle qui, par sa capacité d'apprentissage rapide et logique, permet dorénavant une plus grande productivité, de meilleures transcriptions et traductions automatisées. Les corrections humaines en postproductions sont néanmoins toujours nécessaires pour une qualité optimale, car les erreurs et approximations sont fréquentes.

(3) Les avancées technologiques nord-américaines et chinoises de la reconnaissance vocale

La technologie nord-américaine s'est affirmée avec des outils de reconnaissance vocale et de transcription automatiques. Ces progrès sont le fait d'une poignée d'entreprises dominantes américaines, les GAFAM (Google, Amazon, Facebook, Apple et Microsoft). Ces géants de l'Internet règnent sans partage sur le web et profitent d'un quasi-monopole technologique nourri par les milliards d'utilisateurs et de leurs données. Celles-ci représentent des corpus énormes favorisant les avancées dans les domaines du *big data* qu'utilise l'intelligence artificielle ; des corpus qui évoluent et progressent à grande vitesse. Le sous-titrage multilingue automatisé, les machines statistiques ou neuronales de traduction (MT et NMT) ont des financements qui proviennent de l'intérêt des GAFAM.



L'avancée des technologies américaines se reflète dans l'offre conséquente de produits commercialisés dans le monde et de nombreux logiciels auteurs sur les marchés audiovisuels.

Les recherches académiques dans les pays occidentaux se nourrissent aussi des apports de ces mêmes technologies, voire sont facilitées par celles-ci.

Ces outils tendent à se diffuser partout ailleurs sauf en Chine qui développe ses propres outils et vise le marché asiatique. Les sociétés asiatiques de l'Internet, les BATX (Baidu, Alibaba, Tencent et Xiaomi) disposent également de bases de données conséquentes d'utilisateurs pour faire progresser leurs technologies. Bien que les marchés chinois, voire asiatiques, soient peu transparents, deux constats s'imposent :

- Leurs outils permettent d'opérer tant la traduction automatique de la parole vers du texte que de faire de la traduction multilingue.
- La multiplicité des dialectes et langues asiatiques (des centaines de langues et dialectes pour les pays tels que la Chine ou l'Inde) pose la question des démarches d'accessibilité, notamment en faveur d'une traduction multilingue automatique. Mais par ailleurs, les politiques et la prise en compte des handicaps n'évoluent pas à la même vitesse que celles en Occident.

Ces éléments font ressortir le positionnement original du projet ROSETTA autour d'une approche de qualité de sous-titrage, encore inexistante aujourd'hui.

(4) La génération de contenus en Langue des Signes, un domaine de recherche très récent

Ce jeune domaine de recherche n'a encore produit que très peu de systèmes opérationnels. Les systèmes de génération automatique de contenus en LS sont généralement basés sur la succession de deux processus :

- La conversion : un système convertit le texte source (éventuellement obtenu après une phase de transcription automatique de la parole) en une représentation intermédiaire plus proche des langues des signes.
- L'animation : à partir de cette représentation intermédiaire, on peut contrôler l'animation d'un avatar, encore appelé signeur virtuel (ou avatar signant), qui va permettre de visualiser le contenu en LS.

Selon la nature de cette représentation intermédiaire (de simples gloses à des représentations symboliques plus détaillées), on dispose d'informations plus ou moins riches du point de vue linguistique. Selon la méthode d'animation employée (synthèse pure ou rejeu de données de capture de mouvement), on dispose d'une animation plus ou moins naturelle. La recherche dans ce domaine est récente, mais les projets de recherches se développent, ainsi que les propositions technologiques (via un avatar, voire un robot). Certains projets arrivent à maturation, parmi lesquels l'autrichien Simax et le brésilien Hand Talk (et VLibras). Ce dernier est disponible gratuitement sur App Store et propose une traduction en Libras (LS brésilienne) depuis l'anglais et le portugais.

(5) Le défi de la génération de contenus face à la spécificité de la langue des signes

A ce jour, la fonction d'interprétariat en langue des signes est quasi exclusivement assurée par des humains à la télévision, mais aussi dans les administrations et entreprises, dans les conférences, dans les services à la clientèle, dans la formation. Les langues des signes sont des langues qui ne peuvent se traduire par une simple séquence de signes (les éléments lexicaux) : il existe d'autres unités gestuelles de nature plus iconique et qui ne peuvent pas être listées dans un dictionnaire. L'ensemble des





éléments corporels à partir du buste sont utilisés pour transmettre l'information de manière simultanée et la grammaire est spatio-temporelle. Le défi de la traduction de contenus en LS sera de tenir compte de l'ensemble de ces aspects ; ce que ne proposent que très peu de projets à l'heure actuelle.

Bien qu'il ne soit pas prévu, dans le projet ROSETTA, de s'intéresser aux robots ou au changement d'apparence de l'avatar, il faut savoir que la recherche sur l'automatisation de la LSF, donc du corps humain et de ses gestes se développe dans les champs de la robotique physique humanoïde et dans celui des agents virtuels conversationnels avec la proposition du changement d'apparence pour représenter plusieurs interlocuteurs. Le constat est que, aucun signeur robotique ou virtuel, à l'heure actuelle, n'est en mesure de représenter la langue des signes qui est à la fois la gestuelle des mains, la dynamique et l'émotion du signeur, les mouvements du buste, de la tête ou les expressions du visage. De ce fait, ils ne peuvent pas (encore) être perçus comme une solution pour assurer la mission d'interprétariat en langue des signes.

Les projets d'avatar signant posent également deux questions : l'acceptabilité par les publics concernés et la représentation que doivent prendre les avatars. Le produit généré automatiquement doit être en mesure de différencier le locuteur du discours signé, c'est-à-dire de changer d'apparence pour signifier le changement de locuteur.

(6) L'aspect social et réglementaire

L'aspect social correspond à l'organisation ou non, au sein des pays, de groupes influents faisant évoluer l'accessibilité audiovisuelle. Cette possibilité d'influer ou non sur la régulation dépend de la capacité de rassembler les intérêts des acteurs concernés pour faire du lobbying, de la taille du pays et pour la génération de contenu en LS, du nombre de locuteurs de la même langue des signes. Par exemple, il y a plus de locuteurs en langue des signes américaine qu'en langue des signes italienne. De ce fait, le poids des doléances ne sera pas le même.

Les pays dans lesquels les associations défendent les intérêts des personnes sourdes portaient souvent plus loin les projets d'accessibilité audiovisuelle, jusqu'à les inscrire dans la législation. L'aspect social et l'aspect législatif sont ainsi liés et peuvent favoriser le développement, voire le déploiement de marchés correspondant. Tout en tenant compte des différents contextes internationaux et/ou variables d'un pays à l'autre, il est possible de réfléchir à comment anticiper de la mise sur le marché des solutions ROSETTA, entre autres par et avec des influenceurs.

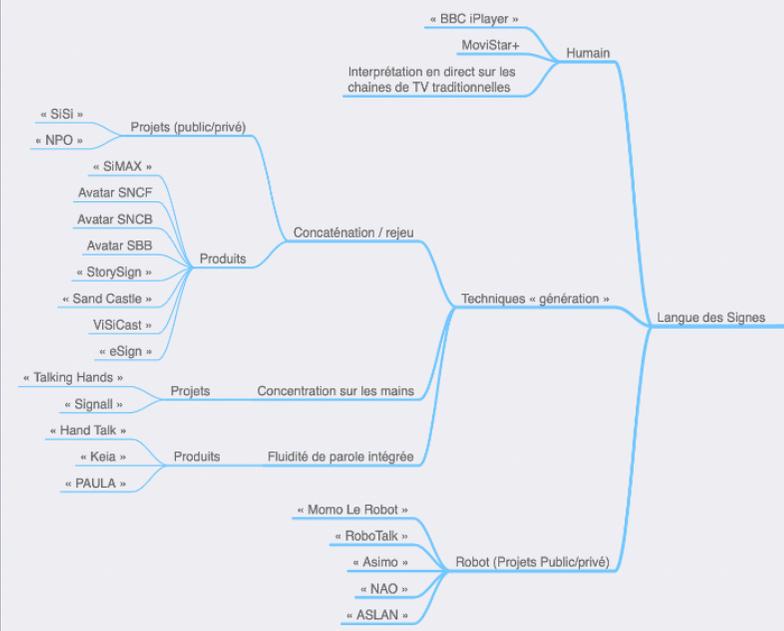
(7) La question de la standardisation pour une diffusion plus large de l'inclusion

La globalisation concerne tous les domaines, y compris l'audiovisuel. Afin de pouvoir diffuser leurs contenus, les producteurs-diffuseurs font face à deux enjeux majeurs : les normes d'accessibilité et les normes de sous-titrages qui sont différentes partout dans le monde. Pour cela, certains groupes, comme Netflix, choisissent de faire sous-titrer leurs films par des prestataires régionaux externes via un guide avec des spécifications très strictes.





Typologie des outils d'accessibilité audiovisuelle : Avatar signeur et sous-titrages automatisés



Outils d'accessibilité recensés pour le projet Rosetta

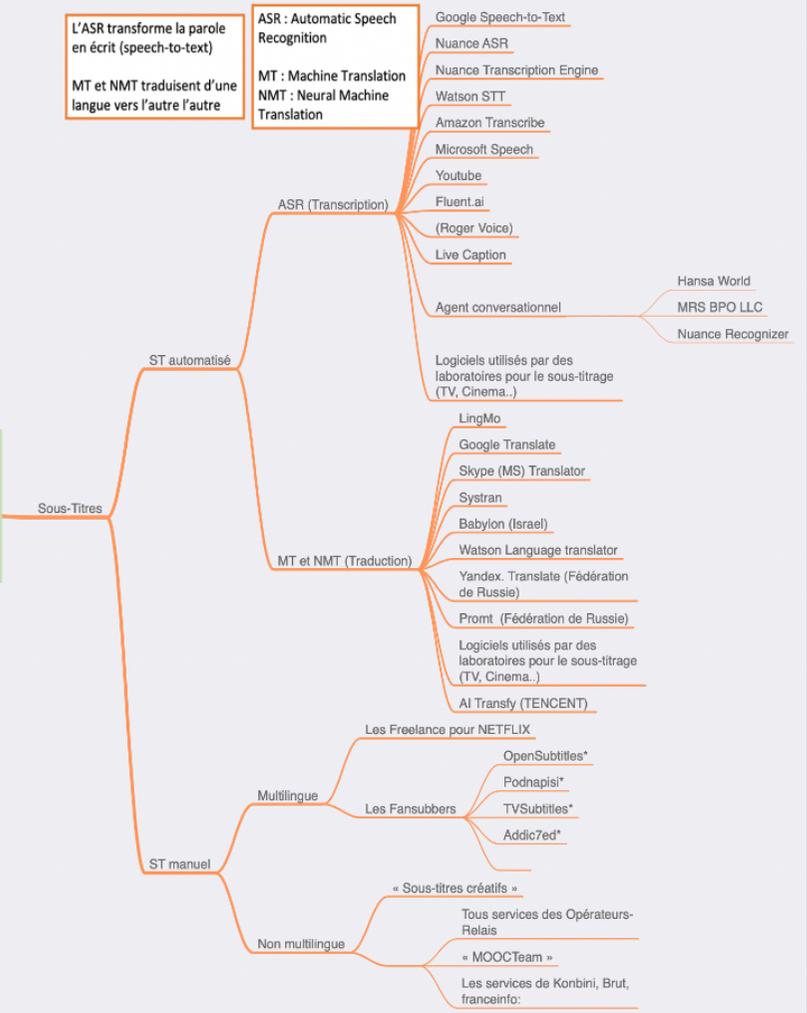


Figure 1. Typologie des outils accessibilité audiovisuelle



1 Contexte

1.1 Objectifs : benchmark international

Le livrable 7.4.1 a pour but la réalisation d'un benchmark à l'échelle européenne et internationale du projet. Il s'agit de réaliser un état de l'art sur ce qui est produit en termes de sous-titrage automatique, de traduction automatique des sous-titres ainsi qu'au niveau de la génération automatique de contenus en LSF (langue des signes française). Ces solutions ont pour cibles les personnes sourdes, malentendantes, les allophones (personnes dont le français n'est pas la langue maternelle), les personnes âgées et de manière générale toutes les personnes ayant des difficultés d'accès à la langue française.

Cette étude s'inscrit dans un contexte d'accessibilité pour développer la traduction multilingue ainsi que la génération de contenus en langue des signes de manière automatique.

Dans les parties qui suivent sont développées les réflexions préliminaires issues de recherches documentaires.

1.2 Le contexte général de l'étude

1.2.1 De l'accessibilité audiovisuelle

Le SME (sous-titrage pour sourds et malentendants) apparaît en 1972 à la télévision américaine¹. Depuis son apparition, le SME n'a cessé d'évoluer et de croître, et nombreux sont les pays qui en proposent pour la télévision. Elle concerne aujourd'hui bon nombre de services et s'étend aux DVD, aux Blu-ray, dans les salles de cinéma, au théâtre ainsi qu'à l'opéra.

Au cours de ces dernières années, afin de permettre et de garantir l'inclusion des publics face à l'accès aux médias audiovisuels, différents gouvernements et pays ont soumis des réglementations définissant les pourcentages minimaux de programmes télévisuels qui devaient bénéficier de sous-titres (ST) à l'attention des personnes malentendantes ou sourdes (ainsi qu'un pourcentage d'audiodescription et d'interprétation en langue des signes, selon les pays).

Par conséquent, il existe d'importants changements - mais aussi des différences dans le domaine de l'accessibilité télévisuelle à travers le monde - en faveur d'une politique d'inclusion pour les droits des personnes présentant un déficit.

Par ailleurs les personnes sourdes de naissance ont pour langue première la langue des signes de leur pays, la langue parlée, même sous sa forme écrite, étant souvent une deuxième langue pour ces personnes.

Les langues des signes ont leur propre syntaxe, grammaire et lexique comme n'importe quelle autre langue naturelle.

Cela met en lumière la question de la surdité pré-linguale et de la surdité post-linguale (connaissance ou non de la langue parlée du pays dans ce cas-là). En raison de ces deux facteurs, il est clair que toutes les personnes sourdes n'ont pas les mêmes préférences de communication.

¹ <https://www.caasem.fr/sous-titrage-adapte/un-peu-d-histoire>



Ainsi préférence pour le sous-titrage ou la langue des signes dépend de la maîtrise de la langue parlée ou de la LS (et donc du niveau acquis dans l'une ou l'autre langue).

Cette nuance permet de mettre en lumière l'intérêt et l'importance de considérer les deux modalités pour garantir l'accessibilité à tous.

Avec l'avènement des nouvelles technologies, l'utilisation d'outils de traitement automatique pour la génération automatique ou l'évaluation de sous-titres pour sourds et malentendants (SME) et de contenu LS sont d'autres options attendues.

1.2.2 Une digitalisation globale et croissante au service d'une accessibilité pour tous ?

Dans un monde qui se digitalise, de nombreux services innovants se développent. Tous les acteurs, qu'ils soient GAFAM (Google, Amazon, Facebook, Apple, Microsoft), BATX (Baidu, Alibaba, Tencent, XIAOMI) ou non, doivent - afin de proposer des services accessibles au plus grand nombre - tenir en compte les besoins spécifiques de leurs utilisateurs en situation de handicap. Certaines innovations technologiques liées à l'accessibilité, comme la génération automatique de sous-titres, pensée pour les personnes sourdes et malentendantes, peuvent servir à un plus grand nombre. En effet, les sourds et les malentendants ne sont pas les seuls à pouvoir bénéficier de ces technologies de génération automatique de sous-titres, les allophones² eux aussi peuvent en profiter, s'ils sont en cours d'apprentissage de la langue dans le pays d'accueil.

Par ailleurs, les progrès dans le domaine de la traduction automatique peuvent laisser entrevoir, dans un avenir proche, la perspective d'une génération (semi)-automatique de sous-titres multilingues. Dans ce cas, s'ouvriraient également des perspectives de marchés supplémentaires permettant, par exemple, de commercialiser plus rapidement des émissions ou vidéos dans d'autres parties du monde, où les professionnels de l'audiovisuel doivent aujourd'hui encore faire traduire les vidéos (soit en les doublant soit en les sous-titrant) pour comprendre leur contenu avant de les vendre ou les acheter. La traduction (semi)-automatique du français vers d'autres langues faciliterait une commercialisation plus rapide des séries ou films en langue française vers les pays non francophones visés par la traduction automatique.

Toujours dans la lignée de l'accessibilité, l'utilisation d'avatars pour afficher du contenu en LS traduit automatiquement à partir du sous-titrage permettrait une augmentation de l'accessibilité des contenus audiovisuels. Ces contenus en LS s'adresseraient non seulement à des personnes sourdes³ et malentendantes mais aussi à certains individus avec un Trouble du Spectre de l'Autisme (TSA). Les TSA peuvent se manifester par une incapacité de certains à avoir accès au langage parlé ou écrit. Ainsi, il n'est pas rare de voir les langues des signes utilisées pour communiquer avec certaines personnes atteintes de TSA. Par conséquent, développer des solutions d'accessibilité permettraient aux entreprises d'atteindre plus de consommateurs / utilisateurs ayant des besoins spécifiques.

Enfin, si ces solutions d'accessibilité ne sont pas répandues dans tous les domaines, c'est pour une question de rentabilité. La création de sous-titres ou l'utilisation d'un interprète professionnel pour rendre des contenus audiovisuels accessibles ont un coût que certains ne jugent pas justifié pour le

² Personne dont la langue maternelle est une langue étrangère, notamment dans la communauté où elle se trouve.

³ Les locuteurs de LS sont la plupart du temps des personnes sourdes profondes et des entendants qui ont des personnes sourdes dans leur entourage. Les personnes malentendantes pratiquent plus rarement les LS.



gain fourni⁴. C'est pour cela qu'aujourd'hui les GAFAM, et particulièrement Google, semblent s'intéresser aux solutions d'accessibilité produites de manière automatique. En effet, à l'instar de la révolution industrielle, nous pouvons constater à travers l'histoire que la capacité à automatiser la production permet d'augmenter la quantité de production tout en réduisant les coûts.

Au-delà des principes d'égalité d'accès à l'information et à la culture, c'est une course pour un marché d'avenir qui se joue, et qui pour l'instant semble être dominé par les GAFAM.

1.3 Méthode globale du benchmarking européen et international

Sous fond du contexte décrit ci-dessus, nous avons réalisé une recherche documentaire qui a donné lieu à la rédaction de ce rapport présentant les projets et solutions qui se font sur le plan international en matière de traduction automatisée de sous-titres et de génération de contenus en LS. Ces résultats correspondent à la première phase du benchmarking 1 (livrable 7.4.1). Il sera suivi d'une phase terrain avec des entretiens qualitatifs ; cette phase fera l'objet d'un second rapport au sein de ce même livrable. En fin de projet aura lieu une mise à jour appelée benchmarking 2 (livrable 7.4.2).

Le schéma heuristique ci-dessus (figure 1) montre la typologie des outils d'accessibilité télévisuelle à partir duquel nous avons réalisé la présente recherche documentaire.

⁴ Il existe pour certains contenus dont les sous-titres de films ou de contenus scientifiques (allant jusqu'aux tedx, mais aussi les MOOCs) une énorme activité de traduction volontaire faite par des amateurs. Nous avons concentré nos recherches sur les approches ST professionnelles, présupposant un minimum de qualité.



2 Constats préliminaires : streaming versus télévision

Dans un premier temps, nous avons distingué les deux catégories de technologies concernées par l'accessibilité audiovisuelle : la télévision et les services de streaming.

Dans un deuxième temps, nous avons recensé tous les services associés à ces technologies. Nous développons les spécificités et ce qui se fait en termes de génération automatique de sous-titres (multilingues ou non) et en termes de génération automatique de contenus en LS.

Outre le français, nos recherches s'étendent à d'autres langues, principalement celles visées dans le cadre du projet pour la traduction automatique telles que l'anglais, le chinois, l'espagnol et le portugais. Pour correspondre aux objectifs du projet et répondre à la demande des différents partenaires, nous nous appuyons également sur les éléments décrits dans l'annexe technico-économique du projet.

2.1 Le streaming

Streaming est un mot anglais qui signifie diffusion. Contrairement à la télévision classique, c'est une technique permettant de diffuser des flux de vidéos notamment en temps et réel et de manière continue. Un utilisateur peut privilégier les vidéos en streaming pour ne pas avoir à les télécharger.

Les services de streaming produisent davantage de sous-titrages : tout contenu audiovisuel peut recevoir des sous-titres « officiels » ou « amateurs » dans l'heure qui suit la diffusion d'un programme⁵. Cela peut aussi s'expliquer par d'autres facteurs comme d'une part la compatibilité des technologies et d'autre part la standardisation des normes ergonomiques d'interfaces qui permet de diffuser les contenus audiovisuels au plus grand nombre.

Par ailleurs, certains services proposent d'ores et déjà de la génération automatique de sous-titres, comme YouTube, depuis 2009. De plus, YouTube propose parfois des sous-titres multilingues générés de manière automatique ou par les utilisateurs comme dans l'illustration ci-dessous.

⁵ <https://www.europe1.fr/medias-tele/Les-fansubbers-ou-le-sous-titrage-low-cost-656496>

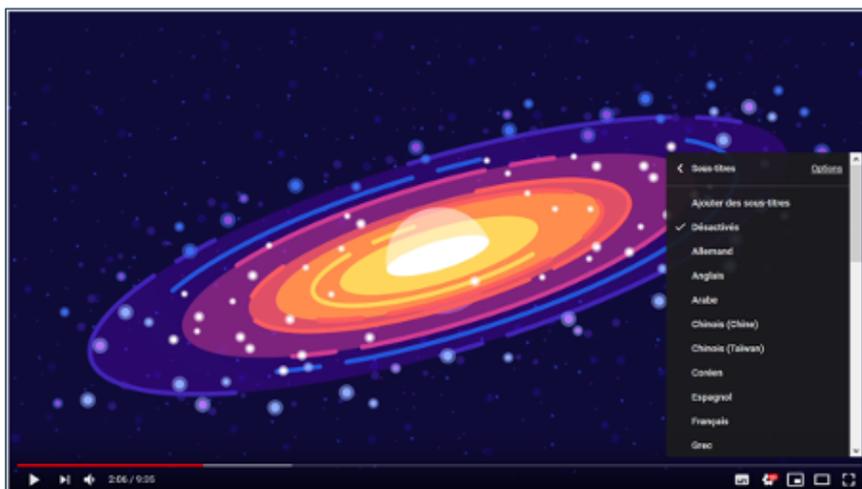


Figure 2. YouTube – Kurzgesagt - In a Nutshell (Great Filter)⁶

L'utilisation de langue des signes semble, pour le moment, cantonnée aux sites de *replay* (rediffusion de contenus télévisuels en ligne). Toutefois, dans le *replay*, on remarque que certains héritages de la télévision, tels que l'utilisation d'un interprète, perdurent. Néanmoins, cette transmission s'accompagne aussi des défauts de désynchronisation présents parfois à la télévision.

Au niveau de la création de systèmes de génération automatique de la LSF ou une autre langue des signes, il n'existe pas de système opérationnel à ce jour. En revanche, il existe des recherches travaillant autour de cette problématique (Marsella, et al., 2013) ; (Dupont de Dinechin & Paljic, 2019) Les travaux de recherches dans ce domaine sont présentés à des conférences dédiées telles que [SLTAT](#)⁷ ou dans certaines revues scientifiques et ne concernent généralement pas le domaine du streaming ou des émissions télévisées.

2.2 La télévision, un modèle robuste

La télévision comparée aux vidéos en streaming propose un modèle davantage accessible. En effet, malgré certains défauts de qualité concernant la fidélité et la synchronisation des sous-titres, les services télévisuels tendent à mettre l'accent sur l'inclusion à travers l'accessibilité, c'est-à-dire l'accès aux contenus audiovisuels pour les personnes en situation d'un handicap. Par ailleurs, bien que les tendances changent, car « en 2016, la consommation moyenne de vidéo mobile a augmenté de 4 heures par rapport à 2012, tandis que la consommation hebdomadaire de télévision a diminué de 2 heures et demie »⁸. La télévision reste encore une technologie répandue, particulièrement auprès des populations vieillissantes^{9, 10}.

La télévision propose un plus large panel de services d'accessibilité dans leurs contenus audiovisuels. En France, les chaînes dont la part d'audience est supérieure à 2,5 % de l'audience totale des services de télévision doivent sous-titrer l'ensemble de leurs programmes¹¹.

⁶ <https://www.youtube.com/watch?v=UjtOGPJ0URM>

⁷ <http://sltat.cs.depaul.edu>

⁸ <https://www.meta-media.fr/2017/02/08/en-2017-lott-nest-plus-une-option-pour-la-tele.html>

⁹ <https://www.zdnet.fr/blogs/digital-home-revolution/le-streaming-toujours-plus-fort-face-a-la-tv-traditionnelle-39862002.htm>

¹⁰ <https://www.zdnet.fr/i/edit/ne/2017/12/PwC003.png>

¹¹ <https://www.csa.fr/Proteger/Garantie-des-droits-et-libertes/Les-droits-des-personnes-handicapees/Le-sous-titrage>



Par ailleurs, les services de télévision semblent être les seuls à proposer en option des sous-titres Sourd et Mal Entendant (SME), de l'audiodescription ou encore l'intégration d'un interprète de langue des signes.



3 L'avancée nord-américaine sur les logiciels de reconnaissance vocale

Concernant la génération automatique de sous-titres interlinguaux (la langue origine n'est pas la langue cible, par exemple de l'anglais vers le français) ou intralinguaux (mêmes langues, origine et cible), l'utilisation d'application ou de logiciel de reconnaissance automatique (ASR, Automatic Speech Recognition) rend depuis quelques années la tâche moins fastidieuse (sept heures de travail à la main pour une heure de vidéo¹², YouTube transcrit en moins d'une heure).

Nous constatons au gré de nos lectures que les technologies américaines pour la génération automatique de sous-titres sont très présentes sur les marchés américains et européens. Dans les parties suivantes, nous présentons les produits les plus prégnants, ici en grande majorité des grandes firmes américaines. En premier lieu, nous présentons l'incontournable Google, proposant des produits pour l'accessibilité comme Live Caption, Live Transcribe ou encore son application de traduction. Nous décrivons également le rapprochement de Google et NETFLIX¹³ (plateforme de vidéos à la demande par abonnement) pour proposer l'apprentissage des langues via le sous-titrage automatique. Ensuite, nous présenterons dans un tableau les logiciels et les produits proposés par d'autres géants américains qui concernent toujours le sous-titrage automatisé. Dans une dernière partie, nous parlerons des quelques produits issus d'une autre puissance mondiale, la Chine.

3.1 Google – Android

En lien certainement avec la pression des lois américaines imposant le sous-titrage à tous les médias audiovisuels aux Etats-Unis (voir partie 5.5), la technologie phare pour accélérer le sous-titrage et son automatisation, l'ASR, est un axe de recherche et d'avancées important pour le marché américain.

A ces fins, Google propose *Google Speech-To-Text*, un outil ASR qui s'appuie sur la technologie de *machine learning* de Google pour traiter des flux en temps réel et des fichiers audio préenregistrés. Il permet une détection automatique de la langue, en temps réel, pour la retranscription. Il y a quatre modèles d'utilisation prédéfinis : commande vocale et recherche, téléphonie, transcription vidéo et modèle par défaut. Fait étonnant, il y a une fonction de filtrage de contenus inappropriés pour les résultats au format texte pour certaines langues.

C'est un outil utilisé ailleurs qu'aux US, par exemple, par Descript, un service de transcription canadien.

Cet outil est évidemment présent sur Android :

« Android [...] est un système d'exploitation mobile fondé sur le noyau Linux et développé actuellement par Google. »¹⁴ Android est présent sur 2,5 milliards d'appareils électroniques¹⁵ : téléphones portables et tablettes (Android¹⁶), montres connectées (Wear OS¹⁷), téléviseurs Android

¹² <https://bfmbusiness.bfmtv.com/le-tete-a-tete-decideurs/les-enjeux-de-la-retranscription-automatique-et-le-sous-titrage-de-vidéos-1446604.html>

¹³ <https://www.netflix.com/fr/>

¹⁴ <https://fr.wikipedia.org/wiki/Android>

¹⁵ <https://www.android.com/what-is-android/>

¹⁶ <https://www.android.com/phones-tablets/>

¹⁷ <https://developer.android.com/wear>





(Android TV ¹⁸), ordinateurs (Chrome OS ¹⁹), voitures (Android Automotive OS ²⁰), (Android Things ²¹ et Android NDK ²²). L'arrivée récente du nouveau système d'exploitation sur smartphone « Android 10 », d'Android - Google (disponible depuis le 03/09/2019) propose **de nouvelles fonctionnalités orientées vers l'accessibilité dont le sous-titrage en direct de n'importe quel contenu média** : musique, vidéo, voix dans une application (GPS par exemple), voix d'un appel ou message audio. Cette solution-là a été conçue avec des sourds et malentendants. Selon les annonces concernant Android, la traduction automatique fonctionnerait dans 25 langues : à partir de la prise d'une image comprenant des mots ou des textes, Android serait capable de traduire en direct, en affichant des sous-titres.

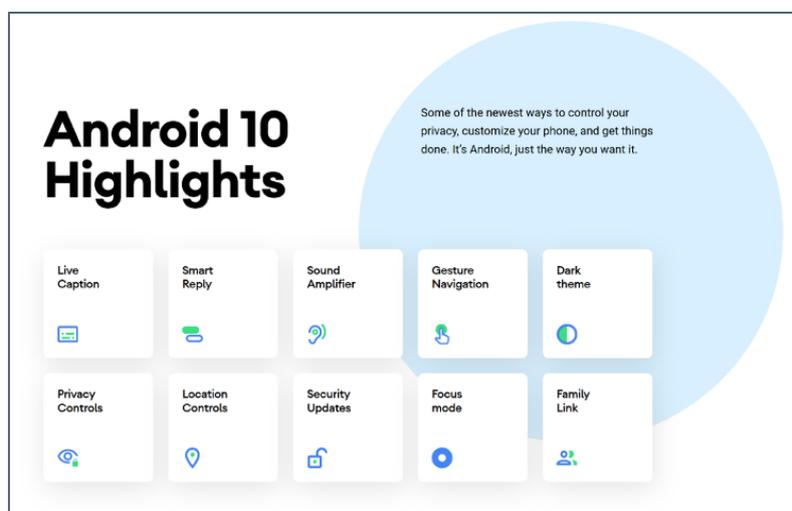


Figure 3. Nouvelles Fonctionnalités de Android 10²³

¹⁸ <https://developer.android.com/tv>

¹⁹ <https://www.google.com/chromebook/chrome-os/>

²⁰ <https://developer.android.com/cars>

²¹ <https://developer.android.com/things>

²² <https://developer.android.com/ndk>

²³ <https://www.android.com/android-10/>



Accessibility	
Live Caption	With a single tap, Live Caption automatically captions videos, podcasts, and audio messages—even stuff you record yourself. Without ever needing wifi or cell phone data.
Hearing aid support	Android 10 now has built-in support for streaming media and calls directly to hearing aids. Using bluetooth low energy so that you can stream all week.
Live Transcribe	Real-time transcriptions for the world around you. Now with sound events that show you when someone is clapping, a dog is barking, someone is whistling, and more.
Sound Amplifier	Filter out background noise and amplify sound according to frequencies personalized to your hearing. Now as a standalone app with a sound visualizer so you can see the sound going on around you.
Accessibility Timeouts	Give yourself more time to interact with your device features. Like the control panel, volume control, and more.

Figure 4. Fonctionnalités orientées vers l'accessibilité de Android 10²⁴

Trois des solutions identifiées sur le site d'Android 10 retiennent notre attention dans le cadre de ce projet : Live Caption, Live Transcribe et Google Traduction pour Android (la traduction en direct). Ces solutions seront exposées dans les parties qui suivent. Nous mettrons ensuite en lumière leur façon de traiter la traduction en direct ainsi que quelques nuances entre ces solutions et finirons par l'intérêt de Netflix pour nos problématiques et son partenariat avec Google Chrome pour « *Learning Languages* » à ce sujet.

3.1.1 « Live Caption »

« Live Caption » (*sous-titrage en direct*) a été conçu avec des sourds et malentendants, ce qui devrait rendre son utilisation ergonomique/adaptée. Dès qu'une voix est détectée, elle est sous-titrée, peu importe le média : voix de film, voix dans applications, voix dans messages, voix dans enregistrements audio, toutes les voix que « Live Caption » arrive à « entendre » ... Cela se fait sans application tierce ni de connexion internet.

²⁴ <https://www.android.com/android-10/>

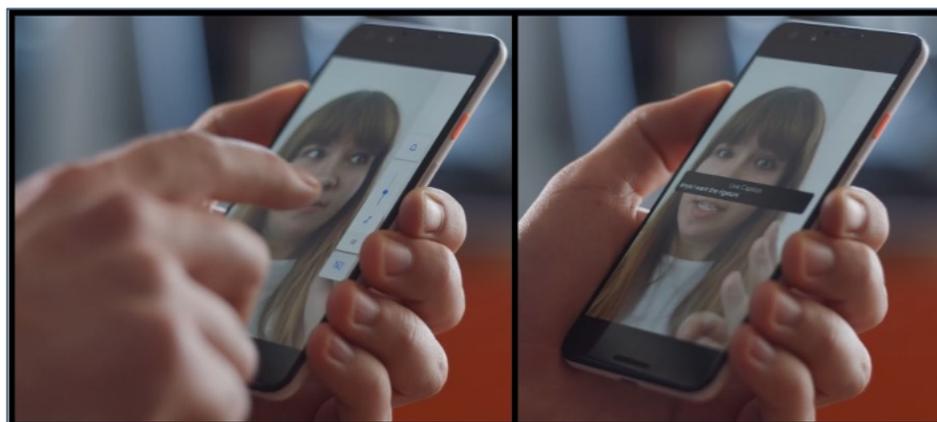


Figure 5. Illustration de l'utilisation de « Live Caption » sur contenu audiovisuel²⁵

Dans les deux illustrations ci-dessus nous pouvons voir l'activation du « Live Caption » (la présentation des options d'accessibilité) et son fonctionnement sur une vidéo / conversation vidéo (avec des sous-titres).

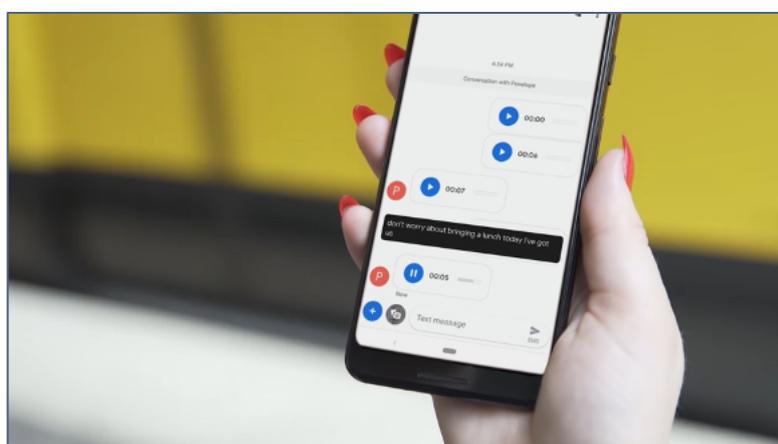


Figure 6. Illustration de l'utilisation de « Live Caption » sur message vocal²⁶

Dans l'illustration ci-dessus nous voyons l'application du « Live Caption » sur des messages audio.

3.1.2 « Live Transcribe »

« Live Transcribe » (*transcription en direct*) est dans la même lignée que « Live Caption ». Comme « Live Caption », la solution a été développée avec des équipes de sourds et malentendants et notamment à travers des collaborations avec l'Université Gallaudet²⁷. L'université de Gallaudet est spécialisée depuis 150 ans dans les apprentissages universitaires pour les sourds et malentendants²⁸.

Contrairement au « Live caption » qui se charge de produire des sous-titres pour les voix qui sortent du téléphone (output), « Live transcribe » transcrit les éléments sonores de l'environnement extérieur au téléphone (input). Il semblerait que cette solution permettrait d'enrichir les sous-titres, en y

²⁵ <https://www.youtube.com/watch?v=YL-8Xfx6S5o>

²⁶ <https://www.youtube.com/watch?v=hPv1Pkj-JO>

²⁷ <https://www.android.com/accessibility/live-transcribe/>

²⁸ <https://www.gallaudet.edu/about>



ajoutant des éléments de transcription de l'audio comme des bruits ambiants tels que des sifflements ou des aboiements.

Une autre différence par rapport à « Live Caption » réside dans le fait que la solution puisse être téléchargée sous la forme d'une application disponible dans le Google Play Store²⁹ et est disponible pour les appareils Android 1.8B à partir du système d'exploitation « Android 5.0 Lollipop » et tous les systèmes d'exploitation jusqu'à « Android 10 ». Par ailleurs l'application nécessite une connexion internet pour fonctionner, contrairement à « Live Caption ».

3.1.3 La Traduction en direct



Figure 7. Android 10 permet de la traduction automatique par image³⁰

Étrangement, « Android 10 » propose un service de traduction par le biais de la caméra du smartphone qui n'est pas mentionnée dans le service de « Live Caption » ou « Live Transcribe ». A partir de l'application Google Traduction, il est en effet possible de prendre une photo d'un texte et de voir sa traduction instantanée s'afficher. Il est possible de télécharger Google Traduction sur Iphone.

Cette option utilise une autre technologie : l'OCR (*optical character recognition*), la reconnaissance automatique de caractères.

3.1.4 Quelques nuances³¹

« Live Caption » ne sera accessible, dans un premier temps, qu'à une petite fraction de modèles de smartphone, notamment le Pixel 3 (téléphone portable de Google) pour des raisons de performances et de coûts.

²⁹<https://play.google.com/store/apps/details?id=com.google.audio.hearing.visualization.accessibility.scribe&hl=fr>

³⁰ <https://www.android.com/what-is-android/>

³¹ <https://venturebeat.com/2019/05/10/how-android-qs-live-caption-works-select-phones/>



Concernant les questions des ressources et de performance : Live Caption utilise (partiellement) la même technologie que celle déployée pour la génération automatique de sous-titres sur YouTube. Puisque Live Caption est sensé fonctionner hors ligne, Android 10 doit contenir tous les algorithmes nécessaires pour la génération automatique de sous-titres, ce qui n'est pas une mince affaire. En effet, la génération automatique de sous-titres sur YouTube passe par une application dématérialisée (en *cloud*) de plusieurs centaines de Giga-octets³². Puisqu'il n'est pas concevable d'occuper une centaine de giga-octets sur un smartphone, Android a dû « alléger » la solution. Ainsi, il semblerait s'agir pour le moment d'une version « *light* » de la version utilisée par YouTube. L'aspect « *light* » de cette version se traduit pour l'instant par la capacité de l'application à ne pouvoir produire uniquement des sous-titres en anglais. Les ressources utilisées (stockage et calcul) ont naturellement un impact sur les performances et les fonctionnalités offertes.

Concernant les questions de coûts : cela se traduit par l'approche de l'exclusivité. Comme nous l'avons dit, « Live Caption » est pour l'instant uniquement disponible sur téléphone « Pixel 3 » qui est un modèle de téléphone produit par Google, incitant ainsi les utilisateurs à investir dans un smartphone Google.

Les solutions « Live Caption », « Live Transcribe » utilisent chacune des technologies différentes. Comme nous l'avons vu, « Live Caption » est basée sur la solution existante pour YouTube et peut fonctionner hors ligne. « Live Transcribe » occupe tout l'écran contrairement à « Live Caption » (cf. illustration ci-dessous). Enfin, « Live Transcribe » s'appuie non pas sur la solution de YouTube mais sur l'API Google Cloud Speech et concerne plus de 70 langues (aucune mention de traduction).

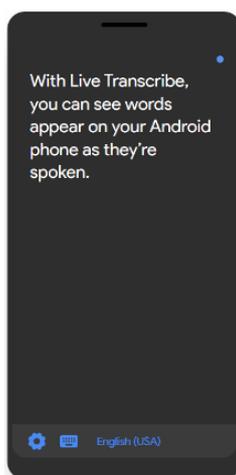


Figure 8. Illustration de "Live Transcribe"³³

Ainsi, comme le dit Kemler, responsable accessibilité des produits Android, dans son interview³⁴ avec le site Venturebeat : la traduction est un autre niveau qui nécessitera des algorithmes différents, une interface qui lui soit propre et adaptée par ailleurs. Toujours lors de cet échange, le responsable a exprimé qu'ils ne souhaitent pas presser les solutions de traductions afin de s'assurer de la qualité de celles-ci. Comme dit dans l'article : « Google doit apprendre à ramper avant de marcher. Et la

³² La taille de l'application n'est probablement pas le facteur principal, mais plutôt les ressources de calcul nécessaires à ces traitements.

³³ <https://www.android.com/accessibility/live-transcribe/>

³⁴ <https://venturebeat.com/2019/05/10/how-android-qs-live-caption-works-select-phones/>



traduction s'apparente plus au fait de courir ». D'autres applications existantes ou produits commercialisés de transcription sont traités dans la partie 4.1.

3.1.5 Netflix : la diffusion de masse impose des normes standardisées

La plateforme américaine de production et de diffusion d'œuvres cinématographiques ou de séries en streaming Netflix se caractérise par sa présence dans le monde entier. Pour s'imposer dans chacun des pays où elle est présente, il lui faut correspondre aux règles et aux normes de diffusion de chaque pays, ces normes n'étant pas, à ce jour, unifiées mondialement (Pedersen, 2018).

Le sous-titrage des vidéos Netflix est assuré par des partenaires extérieurs, dont des micro-entrepreneurs, présents dans les pays de diffusion, qui sont chargés de relire et de corriger les sous-titres. Dans la page des spécifications³⁵ sont recensées et précisées les caractéristiques spécifiques de chacun des pays : cela peut concerner la vitesse de lecture différente selon l'âge, la spécificité des langues (de droite à gauche par exemple), etc. Les formats de sous-titrage sont spécifiés pour être utilisés, dans la mesure du possible, de manière transnationale en dépit des réglementations propres à chaque pays ou région : pour la norme CEA-608 (.scc), la norme européenne EBU (.stl), TTML (.ttml, .dfxp, .xml), ou encore la norme spécifiquement japonaise Lambda Cap (.cap) . Ces derniers sont les seuls autorisés sur la plateforme, lui assurant sa suprématie en matière de diffusions transnationales.

En tenant compte de ces spécificités multiples et des normes singulières de chaque pays, Netflix tend à normaliser la production de sous-titres, s'imposant ainsi par la voie de la standardisation.

3.1.6 Le partenariat Google Chrome : Learning Languages with Netflix

Une nouvelle fois Google se montre en avance face aux problématiques d'accessibilité multilingue. En effet, l'acteur collabore avec Netflix pour mettre en place un système de sous-titrages spécifique à l'apprentissage d'une nouvelle langue. **Learning Languages with Netflix (LLN)** est une extension pour navigateur (Google chrome) créée, en 2019, pour permettre de se perfectionner dans une langue. Cette extension a été créée par David Wilkinson et Ognjen Apic (développeurs Web), deux hommes soucieux d'aider les abonnés Netflix à développer leurs capacités linguistiques. Le format du sous-titrage propose deux langues à afficher. La première langue (en haut) correspond à la langue audio tandis que la seconde langue (en bas) correspond à la langue maternelle de l'utilisateur.

³⁵ <https://partnerhelp.netflixstudios.com/hc/en-us/categories/202282037-SPECIFICATIONS-GUIDES>



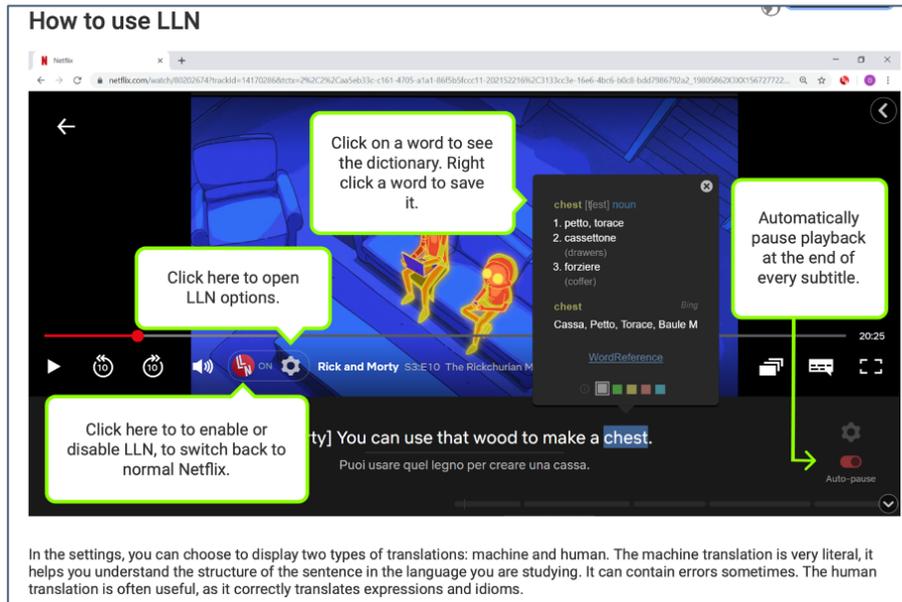


Figure 9. Capture écran de l'aide pour utiliser Learning Languages w/ Netflix

Ces sous-titres sont disponibles dans 11 pays d'Europe et en plusieurs langues : le français, l'anglais, l'allemand, l'italien, le danois, le néerlandais, le norvégien, le portugais, l'espagnol, le suédois ou le turc. Le chinois, le coréen ou encore le japonais ne sont pas encore disponibles.

L'extension inclut également un dictionnaire. Lorsque l'on clique sur un mot, la définition est affichée dans une fenêtre *pop-up*.

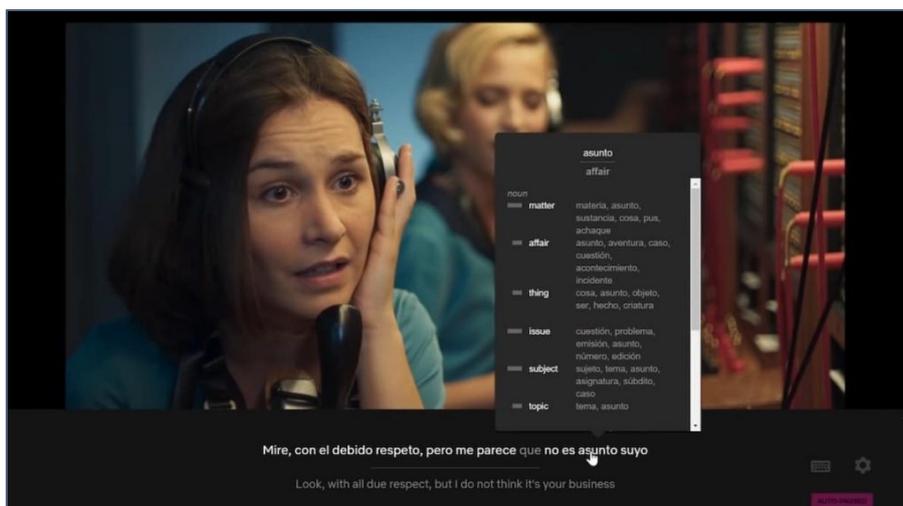


Figure 10. Capture écran choix de langues (LLN)

Des options permettent également de ralentir la vélocité d'une phrase ou de la réécouter. Il existe aussi une liste de recommandation de contenus particulièrement adaptés à l'apprentissage d'une langue. Cette nouvelle application fait écho aux suggestions entendues chez les participants du brainstorming destiné aux étrangers. Par conséquent, nous pouvons supposer que cette technologie répond bien à un besoin existant.



Dans la partie suivante, nous abordons les produits proposés par d'autres firmes et qui concernent la même technologie au service du sous titrage automatique, l'ASR, et les fonctions servies : sous-titrage automatisé.

3.2 La reconnaissance vocale au service de la transcription automatisée

Il existe d'autres produits en matière de reconnaissance vocale au service de la transcription automatisés :

- Pour le sous-titrage synchrone et asynchrone (selon que le programme soit en direct ou non) et pour la traduction automatique, nous constatons que les logiciels et les produits basés sur les avancées technologiques américaines en matière de reconnaissance automatique de la parole sont les plus présents sur le marché occidental.
- Ces logiciels et interfaces (ASR), utilisés dans les produits de transcription automatique tels que les assistants vocaux, sont essentiellement les propriétés de Microsoft (utilisé dans Cortana), Apple (SIRI, créée avec NUANCE, connu pour son logiciel de dictée vocale grand public DRAGON³⁶), Amazon (Alexa), Google (Google Now).

Ces logiciels représentent la suprématie commerciale et technologique des géants du numérique en matière de technologies STT « *speech-to-text* » (transformation de la parole orale en texte écrit).

Ces technologies sont alimentées par les bases de données (corpus) produites par les utilisateurs très nombreux (marchés américains et européens) des différents outils proposés par les GAFAM : par exemple, Facebook compte 2 milliards et demi d'utilisateurs, YouTube, 2 milliards d'utilisateurs, par mois. Notons que les écoutes et les transcriptions de ces géants du numérique étaient « justifiées » et en cours, jusqu'en 2019, dans le cadre « d'amélioration des services »³⁷.

Le tableau ci-dessous synthétise les principales firmes américaines proposant des produits en lien avec la transcription et la traduction, impliquant l'ASR :

<p>IBM (via sa plateforme IA Watson) propose : WATSON STT pour Watson Speech-to-text. Ce logiciel permet la transcription automatique, en temps réel, via la reconnaissance vocale</p>	<p>Basé sur les technologies Machine Learning (Intelligence Artificielle), IBM propose les produits suivants :</p> <ul style="list-style-type: none"> ○ HansaWorld (un chatbot intelligent) ○ LingMo International (pour une traduction multilingue voix et texte) ○ MRS BPO LLC (un Agent conversationnel virtuel, ADAM, sur téléphone), modalité IVR (interactive Voice response)
<p>NUANCE : Nuance est présent dans un produit phare d'APPLE : SIRI</p>	<p>Plusieurs produits basés sur de l'IA (et réseaux neuronaux)</p> <ul style="list-style-type: none"> ○ Nuance Recognizer, un agent conversationnel ; ○ Nuance ASR, pour de la reconnaissance automatique vocale ; ○ Nuance Transcription Engine, qui permet la transcription d'appels/de réunions, pouvant distinguer jusqu'à 6 personnes, et supporte 40 langues environ

³⁶ <https://www.nuance.com/fr-fr/dragon.html>

³⁷ <https://www.france24.com/fr/20190814-facebook-ecoute-enregistrement-sonore-apple-google-amazon-microsoft>





GOOGLE : propose Google Speech-To-Text C'est un outil utilisé par <i>Descript</i> , un service de transcription canadien	Un outil open-source ASR qui s'appuie sur la technologie de <i>machine learning</i> de Google pour traiter des flux en temps réel et des fichiers audio préenregistrés. <ul style="list-style-type: none">○ Permet une détection automatique de la langue, en temps réel, pour la retranscription○ 4 modèles d'utilisation prédéfinis : commande vocale et recherche, téléphonie, transcription vidéo et modèle par défaut○ Fonction de filtrage de contenus inappropriés pour dans les résultats au format texte pour certaines langues
AMAZON Diminutif : AWS pour Amazon Web Services Amazon Transcribe	Utilise également l'ASR pour proposer des services de transcriptions : <ul style="list-style-type: none">○ En flux quasi-réel ou par lots téléchargés,○ Le filtrage de contenu et la reconnaissance des différents locuteurs le cas échéant ;○ L'horodatage des mots permettant leur localisation, cela permet donc un ajout de sous-titrage.
MICROSOFT: remplaçant Bing Speech API and Translator Speech. Kit de développement du service Cognitive Speech (disponible via Azure, le <i>cloud computing</i> de Microsoft)	Kit de développement qui réunit la reconnaissance vocale, la synthèse vocale et la traduction vocale. <ul style="list-style-type: none">○ Un des produits Microsoft utilisant la reconnaissance vocale est Power-Point 365 : il permet de transcrire et afficher la parole en direct, ou sous forme de sous-titres (ne fonctionne pas <i>off line</i>, nécessite le <i>Cloud</i>)○ Skype Translator
YOUTUBE : sous-titrage automatique	Propose de la transcription automatique pour le sous titrage des vidéos. <ul style="list-style-type: none">○ Pour le sous-titrage à la demande (c'est-à-dire à partir d'une vidéo téléchargée), de nombreuses langues sont proposées (allemand, anglais, coréen, espagnol, français italien, japonais, néerlandais, portugais et russe)³⁸○ Pour le sous-titrage d'un flux en direct, seul l'anglais est disponible.
Fluent.ai : Propose une solution basée sur l'IA : Fluent.ai Core	La spécificité de leurs produits réside dans l'apprentissage de <i>patterns</i> (motifs) vocaux, pouvant éviter la traduction du mot-à-mot. Selon eux, cela ouvrirait la voie à la personnalisation des transcriptions, par exemple en prenant en compte les accents, ou de jargon spécifique, et permettrait également le traitement des données hors-ligne (sans accès Internet).

Tableau 1. Présentation des différents produits issus des technologies ASR

Par ailleurs, il existe plusieurs autres services, utilisant l'IA (intelligence artificielle) pour la reconnaissance automatique vocale. Il s'agit de logiciels ou applications de transcription (pour générer automatiquement par exemple des sous-titres), comme Otter.ai,³⁹ Scribie⁴⁰, Speechmatics⁴¹, Trint⁴², Temi⁴³, etc. Beaucoup de sociétés, dont les chaînes de télévision, proposent une hybridation des

³⁸ <https://support.google.com/youtube/answer/6373554?hl=fr>

³⁹ <https://otter.ai/login>

⁴⁰ <https://scribie.com>

⁴¹ <https://www.speechmatics.com>

⁴² <https://trint.com>

⁴³ <https://www.temi.com>



services pour le sous-titrage : il y a d'abord un passage par une transcription automatique (ASR), puis une relecture humaine pour un rendu parfait (Amberscript⁴⁴, Verbit.ai⁴⁵, etc.).

Les services proposés via la reconnaissance automatisée de la parole semblent particulièrement intéresser le monde médical, spécifiquement pour la transcription médicale (pour NUANCE : eScription One, AWS : Amazon Transcribe Medical).

3.3 Les avancées asiatiques

Il se trouve que les entreprises américaines comme Amazon ou Google, à travers leurs propres assistants, ne pénètrent pas le marché chinois avec leurs technologies *Speech-to-Text* de type ASR (*Automatic Speech Recognition*) qui sont à la base des traductions automatisées. Même si seul Siri (Apple) prend en charge le mandarin, son enceinte (Homepod) n'est pas disponible en Chine.

Cela s'explique à la fois par les réglementations très strictes chinoises mais également à cause du traitement de la langue chinoise très complexe et de la multitude des dialectes en l'occurrence.

La Chine souhaite ardemment, depuis des années, mettre fin aux incursions américaines en Asie, et par-là même asseoir de son côté, son hégémonie culturelle (Wang, Zhang, & Kuo, 2020), technologique, sinon économique ⁴⁶. Ceci passe par une politique qui se joue entre stratégie d'influence et affirmation de puissance, en maintenant par exemple les barrières non tarifaires aux sociétés étrangères désireuses de s'implanter ou de vendre en Chine, notamment celles qui refusent de céder leur technologie : c'est l'une des composantes du *soft power* chinois.

Concernant les logiciels existants, l'opacité domine : les recherches et les avancées que les principaux acteurs partagent sur Internet sont peu nombreuses et concernent des logiciels développés en collaboration avec les pays occidentaux. Par exemple l'assistant RAVEN H, produit sous forme de haut-parleur intelligent sans fil, est le fruit d'un partenariat avec le fabricant électronique grand public suédois Teenage Engineering.

Sur sa plateforme DeurOS, BAIDU propose un logiciel conversationnel porté par l'IA, dont les bases de données sont chinoises mais également japonaises (l'éditeur japonais Simeji et son assistant Aladin rachetés par BAIDU en 2011). BAIDU travaille également sur le traitement automatique du langage naturel (TALN), via l'acquisition par exemple d'une start-up dédiée Kitt.ai.

Nous avons également accès aux avancées technologiques proposées par ALIBABA, via son site et son laboratoire DAMO. Le site officiel du Damo Lab présente une majorité de chercheurs ayant obtenu leurs doctorats dans les universités américaines, et ayant travaillé aux États-Unis ou en France avant d'intégrer l'équipe du Language Lab. Pour l'heure, ALIBABA propose ALITRANX, qui pourrait « briser les barrières multilingues entre acheteurs et vendeurs en utilisant les technologies de traduction automatique multilingue et d'analyse lexicale linguistique »⁴⁷.

Par ailleurs, Alibaba propose TMALL Genie, porté par un logiciel IA, AliGenie. Il s'agit encore une fois d'assistant vocal, rival d'Amazon en l'occurrence.

Concernant le 3^{ème} géant chinois, TENCENT, il propose un produit basé à la fois sur l'ASR et sur l'IA : AI Transfy⁴⁸ en lien avec le sous-titrage. AI Transfy (AIT) est une solution couvrant tout le processus de

⁴⁴ <https://www.amberscript.com/fr/>

⁴⁵ <https://verbit.ai>

⁴⁶ <https://www.vie-publique.fr/parole-dexpert/268423-la-chine-puissance-hegemonique-en-asie>

⁴⁷ <https://damo.alibaba.com/labs/language-technology/>

⁴⁸ <https://cloud.tencent.com/product/asr>



sous-titrage depuis la transcription, la traduction, l'édition et la compression audio et vidéo. Cette solution souhaite simplifier l'ensemble du flux de travail de sous-titrage et augmenter l'efficacité globale en améliorant la qualité de sortie tout en réduisant les coûts. Tencent peut se reposer sur une base de données conséquente : le milliard d'utilisateurs de WeChat⁴⁹.

Ainsi, nous pouvons dire que le marché de l'ASR est partagé (non de manière équitable) entre l'Orient et l'Occident entre :

- Alibaba Cloud, Baidu ASR REST, Youdao ASR, Tencent Cloud pour l'Asie
- Google STT, Systran REST⁵⁰, Microsoft STT, IBM Watson, pour les US
- Enfin, dans cette liste apparait également la solution russe Yandex.

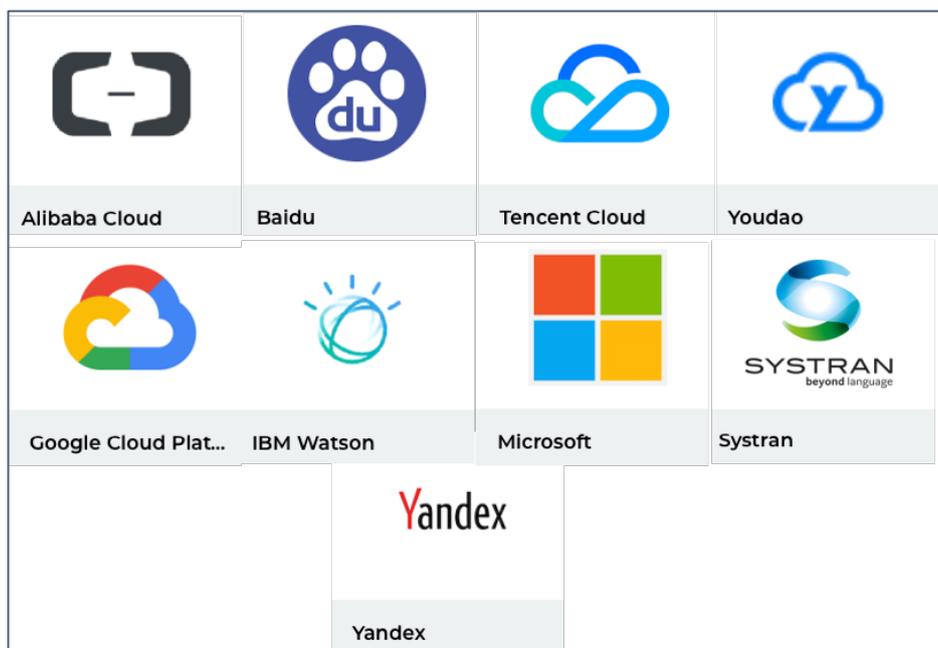


Figure 11. Logos des ASR asiatiques, américaines et russe

En conclusion, même si ces géants ont l'avantage (exclusif) du terrain chinois, voire asiatique, le manque d'accès aux bases de données d'autres langues (et qu'ont les GAFAM), semble aujourd'hui un frein à l'expansion de leurs technologies vers les pays occidentaux.

⁴⁹ <https://www.zdnet.fr/actualites/le-nombre-d-utilisateurs-actifs-quotidiens-de-wechat-depasse-le-milliard-39879067.htm>

⁵⁰ Systran propose des API ASR, voir catalogue sur le lien suivant : <https://inten.to/amp/api-platform/ai/speech/transcribe>



4 Europe : les solutions pour l'accessibilité audiovisuelle

Si en Europe les projets concernant la reconnaissance automatique de la parole ou la génération automatisée de sous-titres ne sont pas légion, monopolisés par les géants du Web, il existe des solutions pour accélérer l'accessibilité audiovisuelle. Nous présentons en premier lieu les produits commercialisés en lien avec le sous-titrage et en second lieu, les propositions d'accessibilité audiovisuelle en langue des signes.

4.1 Produits français en matière de sous-titrage automatisé

4.1.1 RogerVoice, sous-titres et télécommunication

Né en 2015, « RogerVoice »⁵¹ est un service de relais téléphonique français, facilitant la communication et rendant ainsi accessible le smartphone. Application basée sur l'interprétation en LS, la reconnaissance automatique de la parole et la synthèse vocale, RogerVoice est d'abord développée sur Android et distribuée sur la plateforme Google. Proposée au début aux particuliers, elle s'adresse dorénavant également aux entreprises, les centres d'appel (déployée auprès de Oui.Sncf, Allianz Partners), sous forme de SaaS (*Software as a Service*)⁵².

L'application est aujourd'hui disponible dans une douzaine de langues et téléchargeable dans tous les pays, et notamment aux États-Unis.

Sur le site de l'application la page d'aide annonce ainsi que les langues supportées en *Speech-to-text* sont très nombreuses : nous en avons décompté 120.

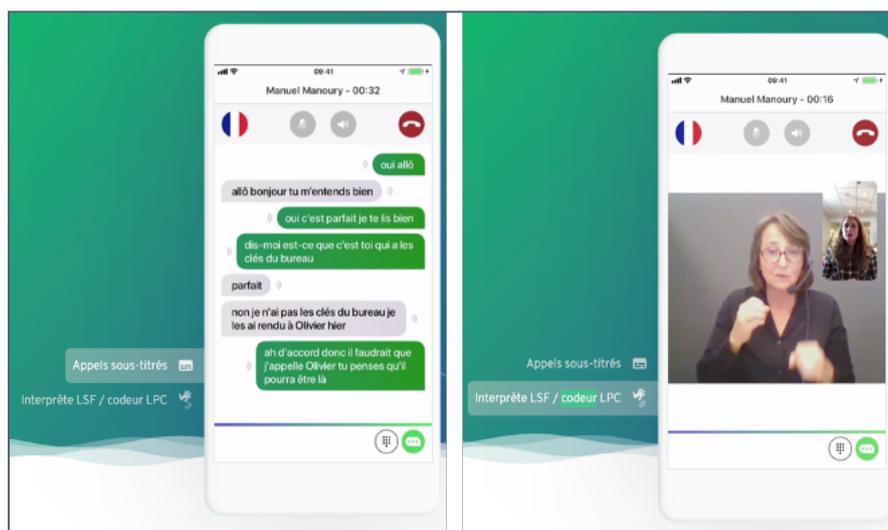


Figure 12. Capture écran de la présentation des deux modalités offertes par RogerVoice (appels sous titrés et interprète LSF)

⁵¹ <https://rogervoice.com/fr/>

⁵² Le SaaS, ou Logiciel en tant que Service, est un modèle de distribution de logiciel à travers le *Cloud*. Les applications sont hébergées par le fournisseur de service.



4.1.2 Authôt, la retranscription automatique française

Authôt⁵³ est une entreprise française récente qui propose une plateforme en ligne de retranscription automatique (40 langues disponibles). Concernant le sous-titrage, si l'utilisateur le souhaite, il peut soit éditer son texte en ligne, soit passer par les équipes de validateurs et de correcteurs, fournissant dans ce cas-là une relecture humaine.

4.1.3 Fleex, solution française pour apprendre l'anglais

Le pendant de la solution proposée par Google Chrome « Learning Language with Netflix » présenté plus haut, est disponible en France via Fleex⁵⁴.

Cette start-up propose d'apprendre l'anglais à travers certaines vidéos/films sélectionnés dans un catalogue, via des « sous-titres intelligents » et des « listes de vocabulaires personnalisés ». Le système serait compatible avec Netflix, Ted, YouTube et parfois même les vidéos en possession des utilisateurs. Trois fonctions sont mises en avant :

- Double sous-titrage : possibilité d'avoir le flux américain et français en même temps ;
- Ralentissement du flux : possibilité de réduire la vitesse de lecture de 1 à 0,9 ou 0,7. A noter que l'intérêt réside surtout dans ces deux vitesses. Le 0.9 ne dénature aucunement la série tandis que le 0,7 la ralentit un peu plus dans le flux vidéo notamment au niveau des mouvements rapides mais permet encore d'avoir une source audio audible et bien ralentie favorisant la compréhension et facilitant la lecture.
- Apprentissage de vocabulaire : lors du passage de la souris sur la ligne de sous-titres la vidéo se mettra spontanément à l'arrêt avec la possibilité de sélectionner un mot et d'en découvrir toutes les traductions possibles. Une option permet également de mémoriser le vocabulaire afin d'accéder via le site Fleex à l'ensemble de ces mots et de réviser.

⁵³ <https://www.authot.com/fr/>

⁵⁴ <https://fleex.tv/fr/>





Claire ne regarde pas *The Walking Dead*...
Claire travaille son anglais avec **fleex**

Apprendre l'anglais avec fleex : comment ça marche

- Vos vidéos préférées**
Fleex fonctionne avec vos vidéos Netflix, les séries et films téléchargés sur votre ordinateur, ainsi qu'une sélection de webséries et de conférences passionnantes.
- Sous-titres intelligents**
Les sous-titres apparaissent sur demande, en anglais et/ou dans votre langue en fonction de vos paramètres. Vous pourrez facilement naviguer entre les dialogues et les faire prononcer plus lentement.
- Listes de vocabulaire personnalisées**
Cliquez sur n'importe quel mot pour afficher ses traductions et les expressions qui le contiennent. Constituez automatiquement votre liste de vocabulaire, au niveau d'une vidéo ou au global.

Figure 13. Capture écran des services proposés par Fleex

4.1.4 VoxSigma Suite⁵⁵ (Vocapia Research ; un produit Linux)

VoxSigma est un outil professionnel de traduction 'speech to text' multilingue. Le processus de transcription s'effectue en 3 étapes : Identification des segments audio avec de la parole, identification de la langue parlée puis conversion du discours en texte. Il est également capable de faire cette conversion avec de la musique. Le résultat est un document XML qui inclut des annotations sur les parties avec du discours et sans discours. Le document peut être directement indexé dans une base de données.

4.2 Les offres pour une accessibilité en langue des signes

Dans cette partie, nous présentons des services existants selon la progression suivante : de la technologie la plus aboutie (avec la représentation d'un signeur virtuel) vers des approches technologiques qui nécessitent encore l'intervention humaine.

4.2.1 SiMAX : la génération semi-automatisée d'un avatar signeur

Le logiciel SiMAX⁵⁶ a été produit par une entreprise autrichienne qui réalise de la traduction de texte vers la LS et affiche le résultat via l'animation d'un avatar. La traduction semble disponible en langues des signes allemande et anglaise. Les traductions sont faites par des professionnels qui sont ensuite enregistrées puis stockées dans une base de données et peuvent donc être rejouées par un avatar.

⁵⁵ <https://www.vocapia.com/>

⁵⁶ <https://simax.media/?lang=en>



Les traductions sont faites, contrôlées et corrigées par des traducteurs sourds professionnels.



Figure 14. Les différents avatars visibles sur le site de SiMAX

4.2.2 Movistar + 5S – Solution numérique LS et SME (Espagne)

Movistar +5S⁵⁷ est une plateforme numérique espagnole accessible sur smartphone, tablette et télévision permettant aux personnes sourdes, malentendantes, non-voyantes et malvoyantes d'accéder à des contenus vidéo (films et séries). Movistar +5S est la version accessible de la chaîne Movistar. La plateforme propose trois systèmes d'accessibilités : des sous-titrages pour les sourds et les malentendants (SME), de l'audiodescription et de la langue des signes espagnole disponible sur l'ensemble du catalogue.



Figure 15. Capture écran du service Langue des Signes, proposé par Movistar +

⁵⁷ <http://www.movistar.es/particulares/movistarplus/guia-rapida/5s-plataforma-contenidos-accesibles/>



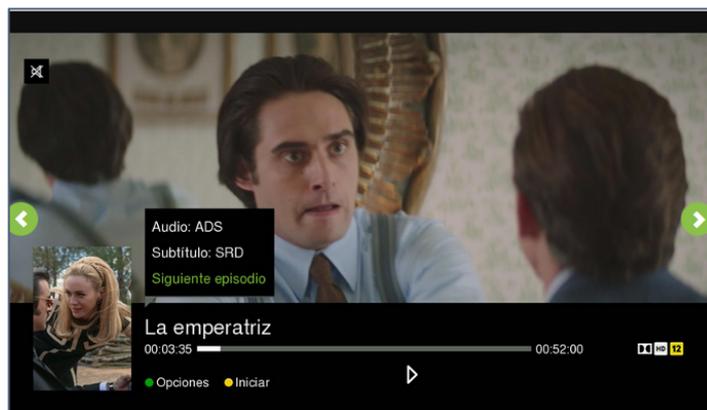


Figure 16. Capture écran du service sous titrage pour malentendants, proposé par Movistar +

4.2.3 BBC iPlayer – une option pour la Langue des Signes

« BBC iPlayer, communément appelé iPlayer, est un service de télévision et de radio par Internet, ainsi qu'une application mobile et informatique développée par la BBC [...] »⁵⁸

Des programmes “signés” sont disponibles sur toutes les versions du BBC iPlayer. Tout comme les programmes en audio description, ils sont parfois disponibles plus tard que la version originale mais sont aussi disponibles durant 30 jours sur BBC iPlayer.

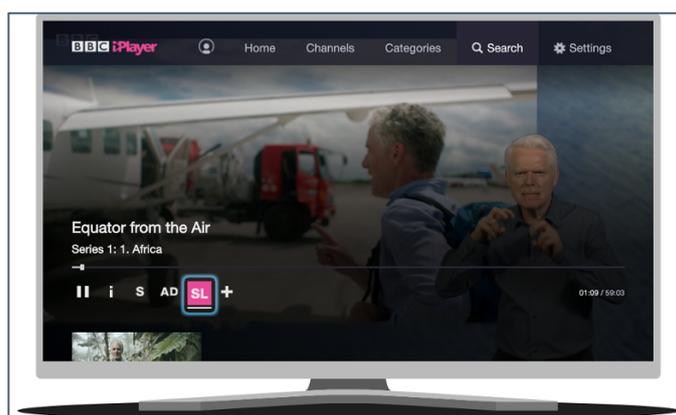


Figure 17. Capture écran du service Langue des signes/ST par BBC iPlayer

5 Les projets scientifiques et technologiques publics-privés

Dans cette partie nous proposons une revue des connaissances scientifiques, portées tant par les recherches académiques universitaires que par les entités de recherche et développement et d'innovation des entreprises. Seront rapportés notamment les apports des projets technologiques issus de partenariats publics-privés.

La ligne de partage entre les régions linguistiques ou les pays s'efface davantage pour les travaux collaboratifs scientifiques publics-privés : les laboratoires gagnent à travailler dans un cadre

⁵⁸ https://fr.wikipedia.org/wiki/BBC_iPlayer



pluridisciplinaire et international et les projets cofinancés par l'Europe sont transfrontaliers par essence. Cela explique pourquoi les résultats de nos recherches ne sont pas présentés selon les pays ou régions linguistiques dans cette partie, mais selon des thèmes d'intérêt. La mention sera explicite lorsqu'il s'agira de travaux nationaux le cas échéant.

Nous commençons par une contextualisation de l'accessibilité, qui est au cœur de ce projet. Dans un deuxième temps nous ferons un état de l'art quant aux études et projets portant sur le sous-titrage automatisé. Un troisième temps portera sur les projets qui traitent de la génération automatique de contenus en LSF. Ensuite, nous nous interrogeons sur d'autres modalités proposées dans le cadre de la langue des signes à travers les robots physiques. Enfin, nous avons jugé utile d'aborder les normes et les pratiques quant à l'accessibilité à travers le monde, en particulier dans les régions linguistiques intéressant le projet ROSETTA, car elles favorisent l'implémentation d'outils d'accessibilité où elles existent.

5.1 De l'accessibilité en général à la conception universelle

Selon « *European Disability Strategy 2010-2020* » de la Commission Européenne, l'accessibilité aux médias est au cœur des priorités de la Commission européenne⁵⁹. C'est pourquoi un projet européen, UMAQ⁶⁰, conduit par l'université autonome de Barcelone, s'intéresse à l'unification, voire la normalisation de l'accessibilité aux médias dans le cadre d'une politique globale européenne en faveur des personnes handicapées. En effet, ces dernières années, les efforts nationaux portaient majoritairement sur l'aspect quantitatif de l'accessibilité, via par exemple des quotas d'accessibilité des médias et de sous-titrages à atteindre. Ce projet permettrait alors de poser les fondements théoriques et appliqués, pour unifier les normes, harmoniser le cadre de l'accessibilité aux médias en Europe (exemple d'une revue spécialisée parue fin 2019⁶¹), mais selon une approche plus qualitative.

L'accessibilité, dans son acception générale, devrait aussi être pensée comme « conception universelle, design pour tous », intégrée en amont et pendant tout le processus de développement selon une méthodologie globale, intuitive et empirique⁶².

C'est le cas par exemple pour le cluster « Accessibilité » réunissant les projets suivants :

- ImAC⁶³ qui s'intéresse à l'accessibilité des médias dans les univers immersifs,
- EasyTV⁶⁴ qui vise à développer des services d'accès améliorés aux médias via de nouvelles fonctionnalités d'accessibilité personnalisées,
- Content4ALL⁶⁵ qui vise la diffusion du contenu en langue des signes, via l'intelligence artificielle et l'animation d'avatars 3D photo-réalistes (ce projet est porté par l'ESN Fincons⁶⁶).

Dans ces différents projets, la conception universelle est d'autant plus complète car elle inclut, dans les premières phases, jusqu'aux dernières, le public visé mais également tous les publics.

⁵⁹ <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0636:FIN:en:PDF>

⁶⁰ <https://cordis.europa.eu/project/id/752659>

⁶¹ <http://www.jatjournal.org/index.php/jat>

⁶² <https://www.bloghoptoys.fr/cest-quoi-design-pour-tous>

⁶³ IMAC, projet espagnol : <https://www.imac-project.eu>

⁶⁴ <https://easytvproject.eu>

⁶⁵ <http://content4all-project.eu>

⁶⁶ <https://www.finconsgroup.com>





Le projet ROSETTA tente de répondre également à cette attente, à travers la visée systémique proposée par l'approche multidisciplinaire des différents partenaires et l'implication des usagers finaux.

Par ailleurs d'autres projets portés par l'Europe, dans le cadre d'H2020, partagent la vision globale de l'inclusion : Le projet aiD⁶⁷, projet chypriote avec plusieurs objectifs (création de modèles de *Deep Learning* pour une communication interactive entre sourds et entendants ; des réseaux et des modèles intégratifs pour les smartphones ; le partage des connaissances et de l'expertise via la promotion de l'excellence, en direction des entreprises publiques et privées) ou encore, dans la même perspective, le projet ELiTR⁶⁸, qui souhaite approfondir les connaissances sur les MT (Machine Translation), le multilinguisme, la traduction automatique, le NLP (*Natural Language Processing*), et les architectures orientées « Services ».

5.2 Sous-titrages : apports des projets académiques et collaboratifs

Les besoins de sous-titres ont explosé avec la globalisation et l'échange des contenus vidéos à l'international. Les nouveaux usages et les avancées technologiques ont accru la demande en matière de sous-titrages automatisés. Pour les entreprises, ce sont de nouvelles opportunités business permettant de traiter de gros volumes en un temps plus court. Les administrations se saisissent d'une opportunité en matière d'accessibilité de leurs messages au plus grand nombre. Les particuliers quant à eux, peuvent profiter d'une meilleure accessibilité, grâce au sous-titrage automatique, d'un plus grand nombre de contenus sous-titrés ou bien l'apprentissage d'une nouvelle langue.

L'attrait pour l'intelligence artificiel est d'autant plus fort que ses apports sont nombreux. Dès 2010, des travaux démontrent les liens entre la machine et le sous-titrage et les apports pour la traduction de masse (Volk, Senrich, Hardmeier, & Tidström, 2010)

Le premier étant la productivité : en 2017, dans le livre blanc d'Omniscien⁶⁹ (anciennement AsiaOnline) la société démontrait la différence de rendement journalier entre l'humain et la machine.

Editor	Video Title	Words Per Day	
		Human Translation	MT Post-Editing
1	SpongeBob SquarePants Movie	5,400	9,488
2	Pretty Woman	3,816	6,944

Tableau 2. Extrait du livre blanc Omniscien montrant la différence de productivité par mots entre le travail humain et la Machine

Nos lectures permettent d'avancer que l'Intelligence Artificielle (IA) sert énormément à la transcription et à la traduction. Mais elle ne permet toutefois pas une transcription / traduction de niveau fluide. Il y a en effet une distinction entre les niveaux fluides et de veille : la traduction de niveau « veille » permet de comprendre le thème et les relations logiques entre entités, évènements, concepts. Il n'y a pas d'exigence sur la qualité de la formulation (registre de langue, grammaticalité), ce qui explique que

⁶⁷ <https://cordis.europa.eu/project/id/872139>

⁶⁸ <https://elitr.eu/the-project/>

⁶⁹ <https://omniscien.com/more/resources/white-paper/>





la lecture peut être pénible. Une traduction automatique de niveau « fluide » assure un niveau minimum de veille et la bonne qualité de la formulation en plus.

La vraie question concerne donc la qualité du travail effectué : aujourd'hui, l'IA assiste les hommes, car finalement, il y a souvent besoin d'une relecture humaine (les genres et les nuances ne sont pas encore respectés par l'IA, ni les accents, les bruitages parasites, les phrases complexes). C'est une postproduction de qualité permise par les apports de l'automatisation (réduction des temps de travail et du travail répétitif, valorisation des activités humaines (qualité, précision) qui participe, de fait, à une meilleure accessibilité qualitative aux médias (Greco & Jankowska, 2019). La postproduction permet par exemple de simplifier la retranscription pour synthétiser le sous-titrage, de rajouter les éléments de description de sons non visibles à l'écran (un sous-titrage spécifique pour sourds et malentendants).

5.2.1 La qualité passe-t-elle par la simplification et de la créativité ?

Une équipe de chercheurs espagnols (Tamayo & Chaume, 2017) a étudié l'impact de sous-titres en langue simplifiée sur le développement des compétences linguistiques chez des enfants sourds et malentendants. Cette étude fait écho à certaines résultantes du brainstorming du 10/07/2019 au LUTIN avec un public sourd et malentendant autour de leur « l'idéal de l'accessibilité ». Une hypothèse est que la conception d'une option de sous-titres simplifiés pourrait aider les jeunes enfants sourds et malentendants dans l'accès et à la compréhension des contenus audiovisuels sachant que se pose aussi la question de l'effet de la simplification qui fait éviter l'effort cognitif. Les travaux de ces chercheurs espagnols aideront les partenaires à savoir s'il est recommandable de proposer des sous-titres simplifiés.

D'autres études s'intéressent également à la simplification des sous-titres dans le cadre de l'accessibilité. D'après les résultats d'une recherche polonaise, la simplification intralinguale⁷⁰ ne montre pas d'avantage pour la population sourde : même si le verbatim complet est plus long à lire, s'il est affiché plus longtemps, le contenu est davantage compris (Szarkowska, Krejtz, Pilipczuk, Dutka, & Kruger, 2016)

Ce que rejoignent les résultats d'une recherche américaine sur les charges cognitives liées aux sous-titres, démontrant que la compression de texte (dans une visée synthétique) implique une charge cognitive en plus, malgré la suffisance de la compréhension (Angerbauer, Adel, & Vu, 2019). Une autre équipe américaine montre, dès le début des années 2000, les bénéfices et les intérêts du sous-titrage dans le cadre de l'apprentissage de langues étrangères, à travers une compréhension et une production orale améliorées (Williams & Thorne, 2000).

D'une manière générale, un projet européen Cordis, SURE, financé entre 2016 et 2018, porté par l'université de Londres apporte des avancées académiques, via des recherches basées sur la poursuite du regard (*eye-tracking*), quant à l'exploration visuelle des sous-titres, et pour leur amélioration. Leurs différents résultats montrent par exemple que les sous-titres interlinguaux devraient être plus condensés que ceux concernant la même langue (intralinguaux), de plus les sous-titres segmentés de façon naturelle sont plus faciles à traiter (Szarkowska & Gerber-Morón, 2018) (Szarkowska & Gerber-Morón, 2018) (Gerber-Morón, Szarkowska, & Woll, 2018) (Gerber-Morón & Szarkowska, 2018).

D'autre part, parce que les sous-titres automatiquement générés (via l'ASR/*Automatic Speech Recognition*) manquent de fiabilité, une équipe américaine a proposé à des usagers sourds une mention via un balisage du taux de confiance de traduction (mesurant ainsi le taux de fiabilité de l'ASR)

⁷⁰ Le sous titrage intralingual concerne la transcription de la même langue, ce qui n'est pas le cas du sous-titrage interlingual, qui concernera une traduction depuis une autre langue (par exemple une version originale anglaise sous-titrée en français).



apparaissant pendant le sous titrage automatisé. Les résultats d'une étude sur les sous-titres augmentés montrent que la préférence va plutôt vers les sous titrages coutumiers et non transparents (habitudes), sans compter la surcharge cognitive d'informations additionnelles (Berke, Caufield, & Huernerfauth, 2017).

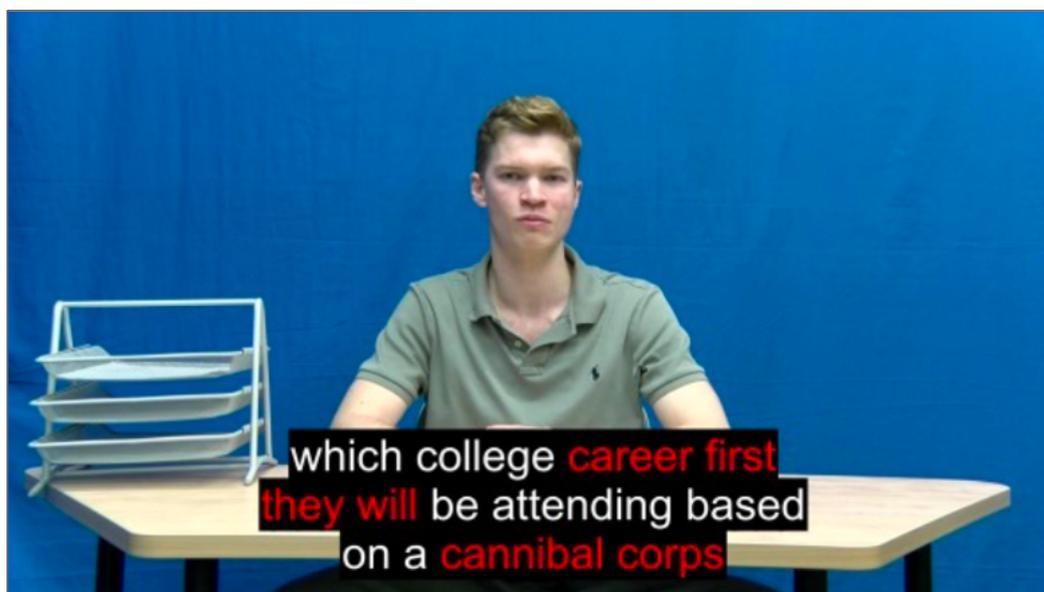


Figure 18. Capture écran étude des sous-titres augmentés

En 2019, deux chercheuses espagnoles ont réalisé une étude autour des sous-titrages à l'attention d'enfants hispanophones présentant des déficiences auditives (Cómitre Narvaez & Sedano Ruiz, 2019) L'enjeu était, ici, de proposer une nouvelle forme de sous-titres dits créatifs.

Leur travail s'inscrivait dans un projet européen qui a réuni un groupe de chercheurs, d'informaticiens, de psychologues et de thérapeutes ayant pour but de développer des méthodes et des outils technologiques adaptés aux personnes présentant des troubles de communication et de langage (autisme, déficience cognitive, handicap sensoriel, auditif, visuel, pluri-handicapés, paralysie cérébrale, etc.) en partenariat avec un centre hospitalier, la Fundació Parc Tauli à Sabadell, des associations et les familles. Ainsi ce projet vise à répondre aux besoins du jeune public hispanophone atteint de surdité de perception ou neurosensorielle et de rendre accessible le matériel audiovisuel (contes en langue française adaptés en films d'animation).

Deux versions du conte animé sont présentées aux enfants : le premier extrait avec sous-titres SME typiques et le second extrait avec les sous-titres créatifs. Toutefois, les sous-titres dits basiques ne respectent pas les normes SME espagnoles. En effet, les couleurs des sous-titres sont pensées en fonction des particularités graphiques du cinéma de Michel Ocelot.

La typographie des sous-titres créatifs sont pensés en fonction du ton, de l'accent ainsi que de l'émotion que l'image est censée transmettre. Il y a également une cohérence entre la typographie du sous-titre et l'esthétique du film d'animation.

L'emplacement est pensé en fonction du mouvement du son pour faciliter l'interprétation et la représentation sonore de la scène.



Figure 19. Exemple 1 sous-titrages créatifs

On remarque une distinction significative entre les deux versions. Dans la version des sous-titres créatifs, nous pouvons voir que l'emplacement des sous-titres a été choisi en fonction des mouvements des personnages. « Ya no somos hormigas » (Maintenant nous ne sommes plus des fourmis) et « Larga vida al héroe que nos salvó » (Longue vie au héros qui nous a sauvé) sont les paroles retranscrites sur les sous-titres.



Figure 20. Exemple 2 sous-titrage créatif

Comme nous pouvons le voir sur l'extrait ci-dessus, la version sous-titre créatif propose une typographie et un format permettant d'accentuer le ton de la scène.

« Tendrás 100 golpes » signifiant « Tu auras 100 coups de bâtons », est un ordre émis par le personnage. Dans l'extrait comportant du SME classique, le ton impératif de la scène est mis en scène par la ponctuation c'est-à-dire le point d'exclamation apposé à la fin de la phrase. Tandis que dans la version créative, l'accent impératif est transmis par la typographie, le texte apparaît en majuscule et inséré dans des flammes. Ce choix de présentation permet aux jeunes spectateurs d'attribuer une émotion (la colère) à la scène qui peut s'avérer plus évidente à comprendre pour des enfants.

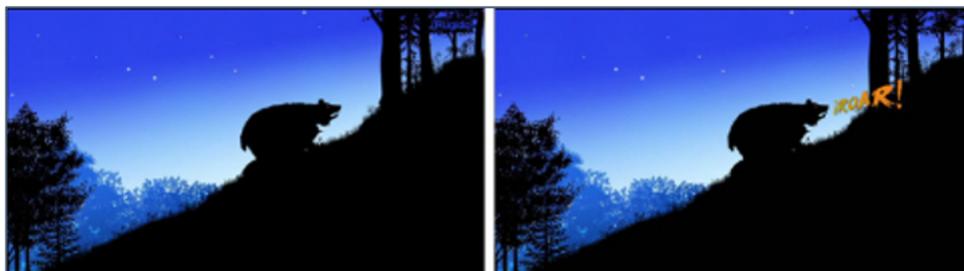


Figure 21. Exemple 3 sous-titrage créatif



Les résultats l'étude dévoilent que le sous-titrage créatif semble parfaitement adapté à la fiction. Avec l'essor des nouvelles technologies, une automatisation des sous-titrages créatifs semble possible. Cette étude nous montre que les alternatives et solutions en termes d'accessibilité audiovisuelle sont nombreuses.

5.2.2 D'autres modalités techniques pour la génération des sous-titres

Concernant la génération de sous-titres automatiques selon d'autres modalités, une équipe chinoise travaille actuellement sur la génération de sous-titres basés sur la lecture des mouvements des lèvres du visage (Lu & Li, 2019). L'alternative est originale et innovante, mais certaines limites sont facilement identifiables. Comment produire / calculer les sous-titres en l'absence de visage, en voix off par exemple ?

Dans la même optique, nous relevons l'intérêt depuis peu pour le *DeepSpeech*, porté par l'IA : ce module open-source de Mozilla, permet grâce via la reconnaissance vocale de convertir un spectrogramme audio en transcription textuelle (Firmansyah, Paul, Bhattacharya, & Urfa, 2020).

D'autres procédés permettent également de convertir la parole en texte : les travaux de (Ghadage & Shelke, 2016) présentent par exemple un système de conversion de la parole en texte multilingue, dans un environnement MATLAB. Il faut noter que comme pour les autres technologies présentées, il faut une base de données conséquente dans ce cas-là. Une alternative est proposée par un laboratoire malaisien grâce à l'implémentation d'un système de reconnaissance vocale via Microsoft Visual Studio : leur résultat montre une précision de l'ordre de 92,69% (Ali & Akmeliawati, 2017) tandis que d'autres laboratoires proposent la traduction multilingue via un outil web auteur (Denisov & Vu, 2019).

D'autres chercheurs proposent également des machines de traduction qui ne nécessitent pas de corpus, donc une intelligence artificielle en passe d'être non supervisée (Chung, Weng, Tong, & Glass, 2019). Les équipes de Facebook (qui réalisait déjà plus de 4 milliards de traductions automatiques) travaillent également sur un projet de traduction de langues rares : dans un article de 2019⁷¹, le directeur du laboratoire FAIR (*Facebook Artificial Intelligence Research*), Antoine BORDES, parlait d'une mise en production imminente de « traduction auto-supervisée » de langues rares, issue elle-même du traitement du langage naturel.

Un autre géant travaille également sur un projet de recherche pour le sous-titrage accessible : IBM, dans le cadre de son IBM Ability-Lab, Media Captioner and Editor⁷². IBM souhaite à travers ses logiciels ASR et d'IA, faciliter la production de sous-titres précis et pertinents pour étendre l'impact du contenu multimédia numérique, le rendant plus accessible aux utilisateurs sourds et malentendants, aux locuteurs de langue étrangère et aux auditeurs affectés par des facteurs situationnels ou environnementaux.

5.2.3 Axes d'amélioration possibles pour les sous-titres générés automatiquement

La génération automatique de sous titres traduits posent des soucis de qualité. Certains chercheurs livrent des réflexions permettant de réfléchir à intégrer des axes d'amélioration.

⁷¹<https://weekend.lesechos.fr/business-story/innovation/0600802801462-la-traduction-automatique-fait-des-pas-de-geant-2248830.php>

⁷² <https://www.ibm.com/able/media-captioner-and-editor.html>



Un cas intéressant porte par exemple sur les traductions interculturelles : en Inde, l'histoire de Laila Majnu est comparable à celle de Romeo & Juliet, d'où l'importance de la « créativité » et des convocations opportunes de références :

- Le contexte de traduction y est important également (traduction de “mère” ou “père” dans le contexte culturel, voire religieux) ou encore les idiomes propres à chaque langue, les “*sensitives languages*” (parole offensante), l'aseptisation des propos offensants, le “slang” ou l'argot, ou encore le jargon spécifique.
- D'autre part, il faudrait prendre en compte les calligraphies et la place ou le sens pris par les caractères des différentes langues : en effet, la qualité de la traduction influe sur la qualité du programme (ou du site web).

Dans leur article (Benaïda, Taleb, & Namoun, 2018) démontrent par ailleurs que Google a du mal à traduire de l'arabe vers l'anglais, les grands textes surtout, à cause de la richesse de langue non atteinte par le moteur de traduction. Au-delà de la richesse des mots, il y a également un souci au niveau des signes diacritiques qui n'apparaissent pas tout le temps, ou qui ne sont pas pris en compte, ayant ainsi pour conséquence une mauvaise traduction de mots.

L'étude indienne de (Gupta, Sharma, Pitale, & Kumar, 2019), financée par Amazon, porte sur les problématiques rencontrées lors de l'automatisation des traductions et démontre en fait la nécessité de viser une performance accrue, de faire appel à la postproduction (étude basée sur la fréquence des erreurs rencontrées, selon 3 catégories : problèmes textuels, de mise en formes et d'adaptabilité de la machine).

En termes de solutions proposées, les sous-titrages de films arabophones proposent par exemple une incrémentation des noms propres pour une contextualisation culturelle via le rajout par exemple d'un micro-commentaire (« restaurant le Fouquet's ») ou d'un nom introducteur (le pain naan, la vague du tsunami), ou même une substitution (« on va voir Chicago : on va voir une comédie musicale »).

5.3 Génération automatique de la langue des signes

La langue des signes est parfois considérée à tort comme « pauvre » (Le Corre, 2007) par les populations non sourdes du fait que l'on pense ne pas pouvoir tout traduire par les signes. L'iconicité de la langue (le fait que les signes ressemblent à ce qu'ils décrivent) fait penser qu'il serait difficile de traduire des sentiments, des pensées, des choses abstraites. En réalité, il s'agit d'une langue naturelle à part entière, dotée des mêmes capacités que toutes les langues naturelles, vocales ou gestuelles, enrichie de spécificités propres, à prendre en compte dans le cadre de sa génération automatisée.

En effet, c'est une langue iconique et tridimensionnelle : elle ressemble à ce qu'elle décrit, dans un espace de signation et dans une temporalité définie (Millet, 2004). Elle inclut non seulement les gestes manuels, mais également les mouvements du buste, des épaules, de la tête, des sourcils, de la bouche, ainsi que la direction du regard et les expressions du visage.

Comme toute autre langue, elle est enrichie de variations régionales, dialectales, voire générationnelles (Delaporte, 2005).

Par ailleurs, la subjectivité à travers la sensibilité (choix de langage, jeux de mots) lors de l'interprétation de la langue des signes (Séro-Guillaume, 2008) est un enjeu à prendre en considération dans le cadre de la génération automatisée de la LSF, ici à nouveau comme dans toute autre langue.



Les recherches concernant la génération, le traitement automatique, ou encore la modélisation des langues signées sont peu nombreuses et assez récentes : (Braffort, 1996), (Losson, 2000) (Filhol, 2008), (Lefebvre-Albaret, 2010), (Segouat, 2010), (Delorme, 2011) ou encore (Kervajan, 2011).

Le champ le plus investigué et que nous aborderons en premier lieu est a priori celui de l'acceptabilité et l'acceptation d'un avatar signant (auprès d'enfants, d'étudiants). Dans une seconde partie, nous recensons les propositions de modalités de cette génération les plus communes. Enfin, nous évoquerons brièvement les modalités plutôt « robotiques » de génération automatique de la langue des signes.

5.3.1 Travaux (non exhaustifs) portant sur l'acceptation des avatars

Au-delà des études portant sur les spécifications techniques de l'avatar (Jenning, Kennaway, & Glauert, 2010), l'intérêt de la recherche se portera par exemple sur l'**acceptabilité** de l'avatar signant, auprès de jeunes enfants sourds (Gilani, et al., 2019). D'autres travaux questionneront par exemple la préférence entre l'avatar réaliste ou l'avatar stylisé : les résultats montrent que la représentation du personnage (stylisé ou réaliste) n'influe pas la lisibilité et la reconnaissance des signes. Pour autant, l'avatar stylisé est perçu comme plus attractif que celui plus réaliste (Adamao-Villani & Anasingaraju, 2016).

Les études portent également sur son **acceptation**⁷³, ou encore sur son **utilité** auprès du public étudiant sourd / malentendant, en prenant en compte l'aspect iconique et visuo-spatial inhérent à la langue des signes : pour comprendre les concepts et les théories dans le cadre universitaire, l'auteure démontre que le support fourni par la réalité virtuelle permet aux étudiants une meilleure compréhension et un meilleur apprentissage des signes en langue des signes (Zirzow, 2015). Ces résultats enrichissent les conclusions de l'étude menée auparavant sur la compréhensibilité perçue (par les Sourds) des avatars (Kipp, Héloir, & Nguyen, 2011). Si la question de l'utilité de l'avatar, et de son intérêt selon les usages possibles, était posée dès les années 2000, elle se discute encore de nos jours (Larboulette & Gibet, 2018).

En 2019, sous l'égide de Microsoft, un groupe de travail propose de réfléchir et de proposer une recherche interdisciplinaire quant à la reconnaissance, la génération et la traduction automatisées de la Langue des Signes. Ce groupe de travail met en avant les aspects déjà évoqués dans la partie précédente, c'est-à-dire, les apports d'un décloisonnement des approches scientifiques, la nécessaire connaissance de la langue des signes, une contextualisation qui manque souvent dans le milieu de la recherche informatique (Bragg, et al., 2019).

L'intérêt pour l'avatar signeur n'est pas, bien entendu, ethno-centré sur le monde occidental.

Par exemple, rendre l'éducation plus inclusive est également un champ de recherche et d'application dans d'autres pays : dans le cadre de l'uniformisation de l'ISL (Indian Sign Language) au sein de l'Éducation via sa génération en 3D, les premiers résultats montrent une nette progression, tant au niveau de la compréhension qu'au niveau de l'acquisition de vocabulaire (Mehta, Pai, & Singh, 2019). Ou encore à travers un avatar virtuel pour réduire le taux d'analphabétisme (30%) au sein de la communauté sourde Brésilienne (De Martino, Silva, & Bolognini, 2017).

⁷³<https://snow.idrc.ocadu.ca/assistive-technology-2/augmentative-processing/animated-signing-characters-signing-avatars/>



5.3.2 Représentations des LS basées sur la concaténation/séquence de signes lexicaux

Dans cette partie nous rapportons différentes modalités de génération automatisée d'animation d'avatar pour la langue des signes : la concaténation d'unité gestuelle et l'animation centrée autour des mains, deux approches qui, comme nous le verrons plus loin, ne répondent que très partiellement aux attentes des cibles visées.

5.3.2.1 Animations par concaténation

Un des premiers systèmes permettant l'animation de signeur virtuel est le JaSigning. Il remplace l'ancien système SiGMLSigning⁷⁴ développé lors des premiers projets ViSiCAST⁷⁵ et eSIGN⁷⁶.

Ce système est le premier issu de la recherche et il est téléchargeable, libre de droit. Il s'agit globalement de l'animation d'un avatar dont les phrases sont la concaténation (suite) de signes lexicaux (dont on peut trouver un mot de sens équivalent dans la LS) : l'entrée du système est une suite de mots (le vocabulaire est issu d'une base de données créée au fur et à mesure), qui ne prend pas compte de la grammaire, de la « prosodie », de la fluidité de la langue des signes, en d'autres termes, des spécificités de la langue des signes (voir introduction 5.3). Les animations sont synthétisées, ce qui leur donne un aspect très robotisé qui nuit à leur compréhensibilité. Néanmoins, cette solution étant la première disponible a été utilisées dans plusieurs projets. Voici quelques exemples de produits issus de partenariats publics-privés :

⁷⁴ <http://vh.cmp.uea.ac.uk/index.php/SiGML>

⁷⁵ http://www.visicast.cmp.uea.ac.uk/Visicast_index.html

⁷⁶ <http://www.visicast.cmp.uea.ac.uk/eSIGN/index.html>



(1) Dans les transports en commun, les chemins de fer

Un système basé sur JASigning a été développé en Suisse, car la compagnie ferroviaire Suisse SBB souhaitait faire appel à la génération automatique de la langue des signes suisse allemande, via ce système⁷⁷. Un prototype avatar était alors capable de signer la station de train, et cela via un système de géolocalisation (Ebling & Glauert, 2016). Ce système pouvait être transposé à d'autres langues des signes. L'équipe travaillant sur le projet comptait à la fois des chercheurs (informaticiens, linguistes) et des experts en langue des signes suisse allemande (*Deutschschweizerische Gebärdensprache*, DSGS).



Figure 22. Capture écran illustrant l'avatar pour les chemins de fer suisses SBB

Ce même projet s'est trouvé répliqué en Belgique, dans le cadre de l'accessibilité de l'information publique dans les gares et trains de la SNCB (Société Nationale des Chemins de fer belges), avec une collaboration entre deux laboratoires universitaires belge et suisse, pour le développement d'un prototype semblable (David & Bouillon, 2018).

Cette approche, basée sur la génération de séquence de signes (et donc portant surtout sur les mains, sans vraiment prendre en considération les spécificités de la langue des signes (énoncées plus haut) est repris aujourd'hui, dans divers projets commercialisés, que nous présenterons plus loin.

D'autres systèmes de type concaténation d'animations ont été développés dans les labos et certains ont donné lieu à des transferts auprès de partenaires privés.

Ainsi, un des prototypes de recherche du Limsi nommé Octopus (Braffort, Bolot & Segouat 2011) a été utilisé dans le cadre d'une innovation destinée à équiper les gare SNCF : Websourd, en partenariat avec le CNRS, souhaitaient en effet, à travers ce projet concevoir une solution d'affichage destinées aux personnes sourdes et malentendantes.⁷⁸ Il s'agissait « d'une concaténation d'animation prédéfinies, pour générer de simples phrases à trous : (...) le train n°#, en provenance de #, arrivera en quai n°#, avec # minutes de retard. (...) toutes les animations avaient été faites à l'avance, et le message était généré en fonction des besoins »⁷⁹. Les animations, de qualité, étaient créées par des

⁷⁷ <https://pub.cl.uzh.ch/demo/signlang/geosign/en/about.html>

⁷⁸ https://www.asso-supelec.org/global/gene/link.php?doc_id=557&fg=1

⁷⁹ Informations issues de propos échangés avec un membre du LIMSI



infographistes sourds, ce qui permettait d'éviter l'effet robotique de JASigning. Ce service est encore en fonction dans les grandes gares en France à l'heure actuelle.



Figure 22. Illustration de l'affichage en gare SNCF

(2) Dans la télévision

Ailleurs en Europe, dès 2012, la R&D de NPO (TV Danoise), se pose la question de la génération de contenu en LS et engage une étude de marché. Un article de l'EBU tech-i⁸⁰ en précise les contours et les résultats.

Une solution avait retenu leur attention : le JASigning. A l'époque, la *motion capture* était également utilisée. Mais même si les résultats sont meilleurs que ceux produits par synthèse pure à partir de descriptions formelles comme SiGML, selon eux, la création de signes ne sert pas vraiment aux programmes tels que les *News* ou les *Talk-Show*. En effet, le niveau de l'automatisation nécessaire n'est pas atteint pour un direct, ni en termes de vocabulaire ni pour la vitesse de parole. Leur intérêt se porte alors pour les programmes jeunes publics : les jeunes enfants ne sont pas en capacité de lire les ST, le choix pour un avatar semble plus évident. Un programme avec peu de paroles et des marionnettes est finalement choisi (Sand Castle).

Un POC (*Proof of Concept*) a été développé par TNO-led, utilisant la technique SiGML : le choix se porte sur la langue danoise signée (c'est-à-dire respectant la syntaxe de la langue écrite et non celle de la LS) plutôt que la langue des signes danoise, pour faciliter les aspects grammaticaux, et faire le travail à partir d'une transcription. Il reste alors à alléger le script, et vérifier/corriger le résultat par une personne sourde signante (afin qu'il soit compris par un enfant sourd) :

- Chaque signe est codé dans un programme d'édition (ESign Editor), en utilisant un système de transcription phonétique, Hamburg Notation system, HamNoSys) avec des symboles (prenant en compte la spatialité et la forme).
- Les phrases sont créées selon les signes voulus et agencées dans le bon ordre.
- Elles sont enrichies d'éléments non manuels (expression faciale, prosodie, mouvement du corps et de la tête).

⁸⁰ https://tech.ebu.ch/docs/tech-i/0790%20Tech-i%20issue%2026_for%20web.pdf



- Le texte est ensuite converti dans un format XML, telle que SiGML (université d'Anglia Est), basée sur HamNoSys.
- Un logiciel génère alors une animation pour chaque signe



Figure 23. Capture écran de l'illustration de la chaîne de procédés pour générer un avatar

Le logiciel peut insérer alors des transitions pour améliorer la fluidité entre les signes.

Plusieurs avatars ont été créés mais l'effet "robot" des animations créées par synthèse pure pouvait rebuter, le choix s'est alors porté sur un dinosaure pour qui une manière de signer non naturelle serait plus acceptée.

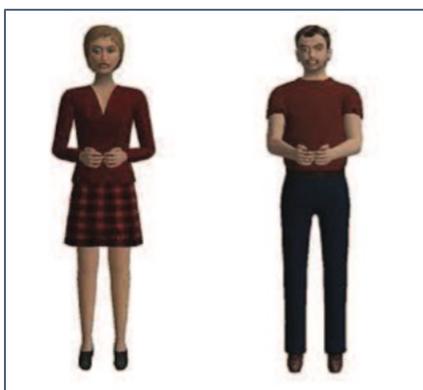


Figure 24. Illustration des avatars humains du projet Danois Sand Castle



Figure 25. Capture de l'émission NPO (projet danois) : dinosaure et personnages de l'émission enfantine





Le dinosaure créé proposait néanmoins des expressions faciales limitées. L'autre difficulté était de signer selon les personnes qui parlent à l'écran : il a été décidé de changer graphiquement l'apparence de l'avatar (couleur du vêtement) selon la personne qui parlait.

L'évaluation du POC (16 sourds et malentendants, dont des enfants de 6 à 8 ans, et leurs profs) montrait un enthousiasme, la reconnaissance de la plus-value pour le rajout de l'avatar mais aussi quelques points négatifs ou à améliorer concernant le respect nécessaire des expressions faciales et les mouvements naturels. Un second épisode mis en ligne, a reçu différentes réactions : même si certains adultes n'ont pas aimé les mouvements altérés et les insuffisances au niveau des expressions faciales, l'accueil était généralement bon pour les parents et les enfants, qui souhaitaient voir d'autres épisodes. Par ailleurs, certains adultes connaissant des difficultés de lecture de sous-titres étaient vivement intéressés par cet avatar.

(3) Pour le produit HAND TALK⁸¹

Commercialisé par une firme brésilienne, les informations sur les technologies / systèmes utilisés ne filtrent pas, mais nous pouvons avancer quelques informations le concernant : Il s'agit encore ici de concaténation et non de traduction, les animations sont prédéfinies, c'est une succession de signes de la Libras (la LS brésilienne). Sur la vidéo de présentation, nous voyons que le visage est figé, il n'y a aucune animation du visage ou du corps. Il y a une application smartphone et un logiciel « VLibras » intégrables sur les navigateurs web qui permettent une traduction à partir d'une base de données.

Par ailleurs, cette même application contient aussi une fonctionnalité « lecture d'écran » (*text-to-speech*) à destination des aveugles et malvoyants, assurant la lecture de la page et traduisant l'écrit en audio au moyen d'un logiciel open source (NVDA). Les vidéos #HugoEnsina (sur YouTube entre autres) permettent un apprentissage de la langue des signes brésilienne : sur ces vidéos, le visage est moins figé, il ne s'agit plus de génération automatisée mais bien d'animation d'un personnage en 3D.

Le 26 septembre 2016, à l'occasion de la Journée Nationale des Sourds, le gouvernement brésilien a officiellement mis à disposition VLibras, l'extension de Hand Talk, sur tous les sites officiels du gouvernement. Sur ces sites, l'utilisateur n'est pas obligé de télécharger l'extension, elle est directement intégrée au site Internet. La start-up a bénéficié d'une subvention de Google.org, à hauteur de 750 000\$⁸² dans le cadre du *Google AI Impact Challenge*. Ce prix vise à récompenser les organisations à but non lucratif, les entreprises sociales et les instituts de recherche œuvrant à relever les défis sociétaux.

⁸¹ <https://www.handtalk.me/br>

⁸² <https://lavca.org/2019/05/07/google-ai-impact-challenge-invests-us750k-grant-in-brazilian-app-hand-talk-as-one-of-the-20-global-winners/>



Figure 26. HUGO, avatar de Hand Talk



(4) Projet étudiants

Dans un souci d'exhaustivité du Benchmark, nous signalons qu'il existe de nombreuses autres initiatives estudiantines. A titre d'exemple, nous n'en citons qu'un seul :

Sisi (pour *Say It Sign It*)⁸³ est un projet étudiant de l'Université East Anglia (Royaume-Uni) et de SYS Consulting (IBM) datant de 2012. Il s'agissait d'un outil de reconnaissance de parole et de conversion utilisant la même technologie JaSigning. Pour cet outil, nous retrouvons les mêmes limites que précédemment : il n'offre pas de traduction et il ne s'agit que d'une simple concaténation de signes. L'animation est gérée automatiquement, sans fluidité et le résultat semble alors incompréhensible.



Figure 27. Illustration des personnages du système Sisi

5.3.2.2 Avatars portant uniquement sur l'expression des mains

Outre les avatars générés automatiquement et qui sont basés sur le système JASigning, il y a également des prototypes universitaires d'avatars dont seules les mains bougent. Il n'y a donc aucun mouvement de buste, de tête, ni d'expressions faciales. Ci-après trois types de projets et problématiques :

- i. **Talking hands** est un projet italien, proposé par une start-up, LIMIX (créée en 2015), en lien avec l'université de Camerino (UNICAM). Le système propose une reconnaissance automatique de signes isolés : Il y aurait d'abord un enregistrement des mouvements, puis une traduction de ces mouvements, un transfert de données vers un smartphone et enfin la prononciation de la phrase traduite via une voix synthétique. Il s'agirait a priori d'un gant (une seule main) pour traduire la langue des signes, proposé ainsi comme une technologie AAC (Augmentative and Alternative Communication), créée pour faciliter la communication. Connaissant les spécificités de la langue des signes, un seul gant, pour une seule main qui signe, pose la question de l'efficacité de ce système. A ce jour, nous ne savons pas si le projet est en cours de réalisation, le site du produit est un appel à financement auprès du fonds Chivas Venture ou une proposition pour le fonds européen MARCHE, dans le cadre H2020.

⁸³ <http://mqtt.org/projects/sisi>

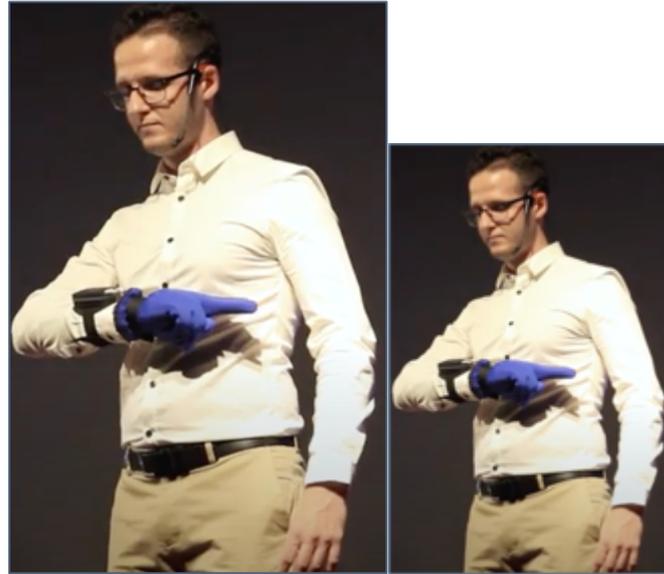


Figure 28. Capture écran de la vidéo de présentation du projet de Limix

- ii. **Signall** est un projet hongrois (en collaboration avec l'université de Gallaudet) qui propose à travers des gants colorés, une reconnaissance des signes via une caméra Kinect. La personne sourde/malentendante signe devant une caméra. La caméra reconnaît les signes grâce aux gants et traduit alors les signes. L'entendant parle au micro et peut ainsi se faire comprendre par la personne malentendante. Il s'agira également de reconnaissance de signes isolés, pour des applications d'apprentissage. Le système de reconnaissance via les caméras Kinect semble être compliqué et limité.

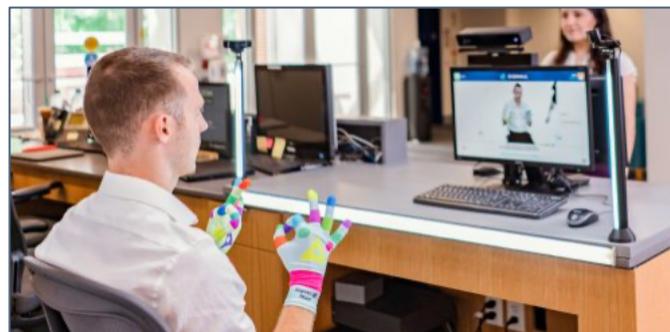


Figure 29. Capture écran de la présentation du projet Signall

- iii. Ailleurs dans le monde, on porte également un intérêt pour un avatar généré automatiquement. Par exemple, le signeur virtuel pour la langue des Signes Sinhala, au Sri Lanka, cependant également centré sur les mains (Punchimudiyanse & Meegama, 2015); ou encore un autre, pour la langue arabe, (Lakhfif & Laskri, 2010).





5.3.3 Autres approches

D'autres approches plus récentes tentent d'atteindre des résultats plus qualitatifs. Ci-dessous seront montrés d'abord le projet universitaire Avatar Signeur⁸⁴, puis le projet PAULA⁸⁵. Tous deux présentent des similitudes : les mouvements sont plus « fluides », mais pour le projet Avatar signeur, il s'agit encore de concaténation (donc d'une séquence) de signes, pour une compréhension globale imparfaite. Puis sera fait mention de la solution KEIA⁸⁶ en France qui offre une traduction automatique du français vers la LS par le *Deep Learning*.

- i. **Les projets Avatar Signeur de l'IRISA « SignCom » et « Sign3D »**, ont été coordonnés par l'équipe de recherche EXPRESSION, au sein du laboratoire IRISA (Institut de Recherche en Informatique et Systèmes Aléatoires) rattaché aux Universités de Bretagne Sud et Rennes 1, dans le cadre d'un projet ANR et d'un projet Investissement d'Avenir. Dans les deux cas, il s'agissait d'un partenariat avec l'association WEBSOURD⁸⁷. Le premier projet porte sur le développement d'une chaîne concaténative de données capturées et annotées, puis animée via un avatar signeur, et pour le second projet, le développement d'une plateforme numérique pour « *enregistrer, accéder et exploiter des données mouvements en haute définition en LSF* ». Ce deuxième projet était en partenariat avec le MOCAPLAB⁸⁸. Les travaux ont pris fin en 2015.



Figure 30. Capture Projet Avatar Signeur de l'équipe EXPRESSION (IRISA)

- ii. **L'avatar PAULA**, issu des travaux d'une équipe pluridisciplinaire de l'université DePaul (informaticiens, linguistes, animateurs 3D, mathématiciens, spécialistes de la culture sourde ou encore des interactions Homme-machine) est encore à l'heure actuelle, en cours de perfectionnement (les premiers travaux datent du SIGGRAPH 99).

⁸⁴ <http://lsf.irisa.fr>

⁸⁵ http://asl.cs.depaul.edu/project_info.html

⁸⁶ <https://www.keia.io/solutions/>

⁸⁷ Par ailleurs, WEBSOURD, association coopérative d'intérêt collectif, née de la rencontre des mondes associatifs sourds et du réseau d'entreprises coopératives, n'existe plus depuis 2015. Aujourd'hui, il semblerait que l'association Media'Pi ait pris le relais, en termes d'action socioculturelle en faveur des sourds.

⁸⁸ <http://lsf.irisa.fr/archives.html>



Cet avatar intègre les paramètres linguistiques de la langue des signes américaine, à travers sa clarté et son débit presque naturel. De plus, PAULA a participé à plusieurs projets connexes, de l'éducation des sourds et de la formation d'interprètes, au tutorat des soignants de personnes à la fois sourdes et handicapées mentales.

Un partenariat avec le Limsi a été initié récemment, et PAULA permet maintenant de générer des contenus en LSF avec une très bonne qualité d'animation et une couverture linguistique incluant certains aspects iconiques des LS, basée sur une représentation intermédiaire qui prend en compte toutes les spécificités linguistiques de la LSF (Hadjadj, Filhol & Braffoer 2018), ce qui en fait une des approches les plus prometteuses à l'heure actuelle (Filho McDonald 2020).

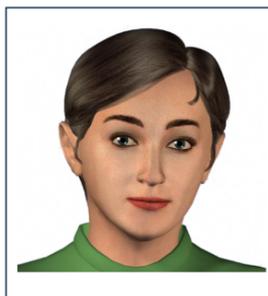


Figure 31. PAULA, de l'université DePaul

- iii. Pour la solution française KEIA⁸⁹, il s'agit « d'Interprétation Artificielle en Langue des Signes » : c'est-à-dire que la traduction du français vers la LS est générée par l'intelligence artificielle et plus précisément grâce au *Deep Learning*. La start-up du même nom (KEIA) a pu préciser les contours de la technologie au service de cet avatar en collaboration avec l'Académie de la LSF (ALSF)⁹⁰ par les moyens suivants :
- Un corpus parallèle (langue française / langue des signes française) de 100 000 petites phrases simples a été constitué.
 - L'animation de l'avatar s'est construite par l'utilisation de dispositifs de type mocap (*motion capture* ou capture de mouvements).
 - L'IA et l'apprentissage profond permettent de créer à partir de phrases en français (input) l'animation virtuelle en LSF (output).
 - Enfin, KEIA développe aussi l'intégration de la spatialisation (des emplacements prédéfinis dans l'espace de signation)
 - D'une part, cette solution est proposée par des personnes bilingues et d'autre part, la traduction automatique laisse néanmoins une place à la post-édition au niveau de l'animation. Par ailleurs, aujourd'hui, le vocabulaire est limité à 500 signes et les structures de phrases sont simples.
 -

⁸⁹ <https://www.keia.io>

⁹⁰ <http://www.languedessignes.fr>

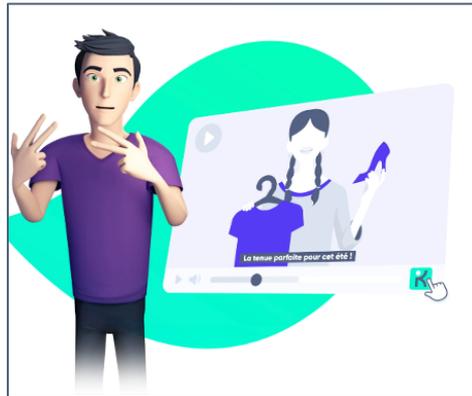


Figure 32. Capture du projet KEIA

5.4 Les robots au service de la langue des signes ?

Au-delà de l'animation d'un avatar virtuel, il existe d'autres travaux pour traiter la langue des signes de manière automatique. Différents laboratoires de recherche à travers le monde s'intéressent au « robot », soit en se centrant uniquement sur les mains robotisées, soit en animant un robot humanoïde communiquant en langue des signes.

L'outil développé dans ROSETTA (**RO**bot de **S**ous-titrage **E**t **T**oute **T**raduction **A**daptés) sert à la traduction automatique pour différentes applications. Le projet est de créer un signeur virtuel, non pas un robot dans le sens d'un « *dispositif mécatronique (alliant mécanique, électronique et informatique) conçu pour accomplir automatiquement des tâches imitant ou reproduisant, dans un domaine précis, des actions humaines. La conception de ces systèmes est l'objet d'une discipline scientifique, branche de l'automatisme nommé robotique.* »⁹¹

ROSETTA se basera uniquement sur l'utilisation de signeurs virtuels (ou avatars signeurs) : un module s'occupe de la conversion / traduction des sous-titres vers la traduction intermédiaire (SYSTRAN, LIMSI), puis l'affichage de la langue des signes vers l'animation d'un signeur virtuel (LIMSI – MOCAPLAB). Il est probable que la génération des sous-titres sera automatique en fin de projet, si tout va bien, et que le signeur virtuel nécessitera encore plusieurs années de développement par la suite. Bien qu'il n'y ait pas de « robots » dans ce projet, le benchmark des outils existants rapporte ci-après dans un souci de complémentarité les exemples de « robots » issus de l'Europe, de l'Asie (Japon), de la Turquie, du monde anglophone (américain essentiellement), de Malaisie.

5.4.1 Les robots humanoïdes

- i. En Europe, des robots physiques couplés à des avatars (Scassellati, et al., 2018), ou encore une récente étude RoboTalk (en lien avec le projet français InMoov⁹²) proposent une interaction homme-machine via deux bras d'un robot humanoïde (impression 3D), signant en langue des signes allemande, et équipée d'un système de reconnaissance vocale (Homburg, Thieme, Völker, & Stock, 2019). Ce même robot issu du projet InMoov a également servi à interroger

⁹¹ <https://fr.wikipedia.org/wiki/Robot>

⁹² <http://inmoov.fr/project/>



l'intérêt de la langue des signes pour les enfants atteints de troubles autistiques (Axelsson, 2018), sous le nom de Momo le Robot ^{93, 94}.

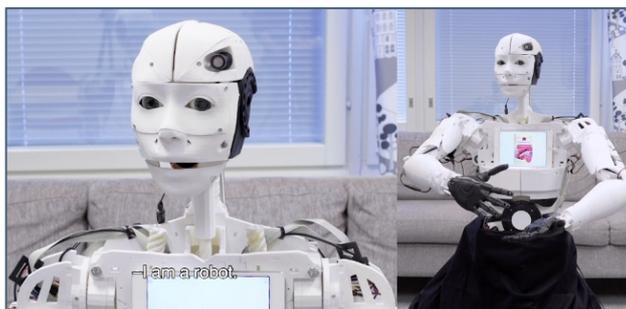


Figure 33. Capture écran de la vidéo étude "Momo rencontre les enfants autistes"

- ii. En Asie, d'autres robots sont proposés : dans une perspective « humanoïde » et de robot qui se déplace, dès le début des années 2000, le constructeur japonais Honda Motor, réfléchissait à un robot assistant, ASIMO. La dernière version (Asimo 5) était en capacité de signer lors de son dernier show⁹⁵. Des éléments concernant les degrés de liberté mains/buste ont même été améliorés entre 2000 et 2011 dans cette perspective. Nous ne savons pas à l'heure actuelle si et comment le robot a évolué depuis.



Figure 34. Capture écran de la vidéo de présentation de ASIMO

- iii. Il y a également un intérêt pour le robot humanoïde en Turquie, notamment avec l'université ITU, qui travaille avec le robot H-25 NAO, pour l'apprentissage la langue des signes auprès de jeunes enfants. Même si l'étude peut être intéressante car elle interroge surtout l'acceptation d'un robot, elle pose question quant à l'efficacité du robot : en effet, il n'est doté que de 3 doigts dans chaque main...

⁹³ <https://awards.ixda.org/entry/2019/momo-the-robot-meets-children-with-autism/>

⁹⁴ <https://vimeo.com/302898833>

⁹⁵ <https://www.slashgear.com/honda-asimo-learns-sign-language-and-masters-stairs-18325765/>



Figure 35. NAO le robot (photo de présentation)

5.4.2 Concentration sur les mains robotisées

- i. D'autres recherches et prototypes imaginés par des chercheurs aux États-Unis et dans le monde anglophone, recourent à un robot illustré par une ou des mains. Dès 1977, une main robotique qui pouvait épeler des mots en utilisant l'alphabet manuel avait été construite pour aider les personnes malentendantes. La version améliorée de ce modèle, à savoir « Dexter II », intégrait un mouvement du poignet (Jaffe, 1994). En 1992, une autre main robotique, capable de produire des lettres en mouvements fluides, a été développée. Enfin, en 1994, la version la plus avancée, RALPH (Robotic ALPHabet) était contrôlé par une interface utilisateur pilotée par un ordinateur distant (Hersh & Johson, 2003). Depuis, même si les techniques de fabrication, d'automatisation des bras ont évolué, il s'agit toujours de « *finger-spelling* », épellation aux doigts.

Par exemple, le prototype malaisien, SignBot (Ali, Bakr, Fauzi, & Akmeliawati, 2019), est un robot issu d'impression 3D, se centrant sur les doigts et les mains, où il s'agira surtout d'épeler des mots via les doigts, à travers des séquences :

Letter	Features	Servo used	Sequences of servos
A		2, 3, 4, 5	Simultaneously 2, 3, 4, 5
B		1	1
H		1, 4, 5, 6, 7	Simultaneously 1, 4, 5 then 6 and 7
I		1, 2, 3, 4	Simultaneously 1, 2, 3, 4

Figure 36. Capture du document de présentation du projet malaisien SignBot



- ii. Un autre robot, également imprimé en 3D et aux moteurs servo, ASLAN (Antwerp's Sign Language Actuating Node), imaginé en 2014 par 3 étudiants (Fierens, Huys et Slaets), était un projet parrainé par l'Institut Européen d'ORL, Sint-Augustinus Antwerpen), pour « traduire » du texte en langue des signes. Les quelques vidéos circulant sur ce projet montrent un bras, avec des doigts articulés, épelant des mots. Il était question à l'époque de rajouter un visage imitant les expressions faciales (Hersh & Johnson, 2003). Le projet semble terminé (mais non abouti), les pages du projet n'étant plus accessibles sur le site de l'Université d'Anvers.

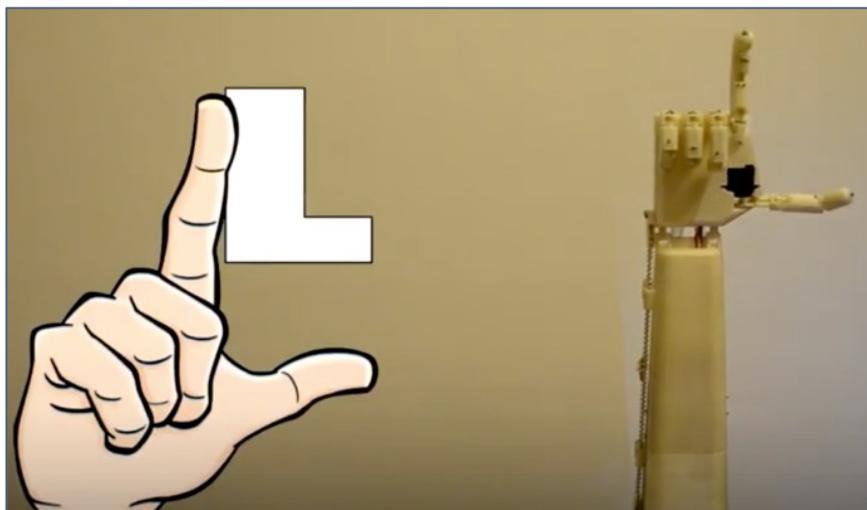


Figure 37. Robot ASLAN épelant la lettre L

38 Cette revue de littérature montre qu'une majorité de projets portés par les laboratoires de recherche s'organisent autour de solutions techniques et logiciels qui se concentrent sur les mains et les bras. la plupart des anciens projets s'appuyaient sur le système JaSigning (voir la partie « 5.3.2 Avatars basés sur la concaténation/séquence et sur les mains ») ou s'en rapprochaient. Ils ne prenaient pas compte les autres indicateurs de la langue des signes, c'est-à-dire les expressions faciales, les mouvements du buste et de la tête, la vitesse des mouvements. De plus, il est important de prendre en compte les autres éléments nécessaires à une bonne compréhension du discours signé :

- Le style du langage : la conversation est continue, les mots ne sont pas isolés, il y a un effet de coarticulation, qui rend le discours fluide, une prosodie à prendre en considération
- Les caractéristiques propres : les accents, la vitesse de la parole, l'âge (débit et langage "cognitif").
- Les spécificités linguistiques : la grammaire spatiale, les structures linguistiques iconiques.

Des travaux plus récents tentent de s'affranchir de ces limites. L'initiative publique Paula et son pendant dans le privé (Keia) en sont deux exemples. Par ailleurs, ces recherches dévoilent un intérêt commun pour l'inclusion des personnes en situation de handicap, particulièrement les populations sourdes, au-delà de la diffusion de l'information audiovisuelle. En effet, cela montre que le développement d'une traduction automatique en LSF s'adresserait à plusieurs secteurs d'activités et non seulement au secteur de l'audiovisuel.



5.5 Pratiques et normes de l'accessibilité dans le cadre télévisuel

Depuis quelques années, même si des disparités subsistent à travers le monde, des changements aux niveaux nationaux ainsi que dans les unions internationales (UE, ONU) montrent l'intérêt et la prise en compte de l'accessibilité, en tant que droit universel : ces droits ont été déclarés droits, entre autres, dans la Convention relative aux droits des personnes handicapées par l'ONU (2006).

L'un des facteurs majeurs favorisant la création et l'introduction de l'usage d'un produit ou service d'accessibilité audiovisuelle est la réglementation. Les pays dans lesquels les technologies de sous-titrages pour sourds et malentendants se développent sont également les pays dans lesquels la législation en matière d'accessibilité est la plus forte, la plus complète.

Dans les parties qui suivent seront traitées successivement et synthétiquement les normes et pratiques en Europe, en Chine, et sur le continent américain. Cette contextualisation de la réglementation d'accessibilité audiovisuelle permettra aux acteurs technologiques du projet ROSETTA de se positionner par rapport à l'environnement social dans lequel leurs technologies pourraient être utilisées.

5.5.1 En Europe

L'Union européenne a défini un cadre pour les réglementations d'abord par la Directive 2006/65/CE « Services de médias audiovisuels » du Parlement européen et du Conseil du 11 décembre 2007⁹⁶. Ces directives ont favorisé considérablement l'augmentation du volume des programmes sous-titrés pour les sourds allant parfois jusqu'à rendre le SME obligatoire⁹⁷. Depuis ont suivi d'autres directives la dernière étant la Directive **2019/882**⁹⁸ qui vise une meilleure coordination, en matière d'accessibilité, entre les pays de l'espace européen.

Par ailleurs, ces directives se font en consultation avec des organisations non gouvernementales européennes soutenant les personnes sourdes et malentendantes, comme par exemple la *European Federation of Hard of Hearing* (EFHOH)⁹⁹, composée d'experts européens, malentendants ou non, dont la fonction est d'améliorer l'inclusion et la participation des malentendants à la société.

Ainsi, afin de permettre et de garantir l'inclusion des publics face à l'accès aux médias audiovisuels, différents gouvernements et pays ont instauré des réglementations nationales définissant les pourcentages de programmes télévisuels qui devaient bénéficier de sous-titres à l'attention des déficients auditifs (ainsi qu'un pourcentage d'audiodescription et d'interprétation en langue des signes, selon les pays).

Ces évolutions législatives sont très variables d'un pays à l'autre, comme montré dans les parties suivantes, mais représentent le 13 mars 2019, pour Sophie Cluzel, Secrétaire d'État chargée des Personnes handicapées, « une opportunité pour les entreprises européennes, un **appel à l'innovation** pour proposer des produits et des services mieux conçus pour l'usage d'un plus grand nombre¹⁰⁰ ».

⁹⁶ <https://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=URISERV:l24101a&from=FR>

⁹⁷ http://media.wix.com/ugd/c2e099_098aaba4b05ee8f7fed8b0add0b8c332.pdf
http://media.wix.com/ugd/c2e099_0921564404524507bed2ff3648781a3c.pdf

⁹⁸ <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A32019L0882>

⁹⁹ <https://efhoh.org>

¹⁰⁰ <https://handicap.gouv.fr/presse/communiqués-de-presse/article/la-directive-europeenne-d-accessibilite-des-biens-et-des-services>



5.5.2 En France

En France, le sous-titrage pour les sourds et les malentendants est une obligation inscrite aux cahiers des charges des chaînes publiques depuis 1984. Mais c'est seulement avec la loi du 1er août 2000 que le sous-titrage pour sourds et malentendants est devenu une obligation pour les chaînes hertziennes.

La loi du 11 février 2005 (n°2005-102 du 11 Février 2005) fait obligation aux chaînes de télévisions publiques et aux chaînes privées dont l'audience dépasse 2,5 % de l'audience totale des services de télévision de sous-titrer la totalité de leurs programmes, hors publicités et programmes dérogatoires.

La loi confie au CSA (Conseil Supérieur de l'Audiovisuel) le soin de veiller à l'accessibilité des programmes télévisés aux personnes handicapées (ainsi qu'à la représentation du handicap à l'antenne).

Pour les chaînes hertziennes dont l'audience est inférieure à 2,5 % de l'audience totale des services de télévision, une convention conclue avec le CSA fixe les proportions des programmes accessibles (entre 20 et 40%).

Depuis 2010, les chaînes télévisées ont l'obligation légale (« Loi pour l'égalité des droits et chances ») de proposer aux téléspectateurs des programmes sous-titrés pour les personnes malentendantes ou audiodécrits pour les non-voyants.

Ci-dessous les tableaux présentant les objectifs de sous-titrage pour les chaînes représentant 2,5% d'audience dans un premier temps, et moins de 2,5%, dans un second temps ¹⁰¹.

Chaîne	Volume annuel accessible (en heures)	Réalisé en % du volume total
France 2	7802	100 %
France 3 national	6831	100 %
France 4	8415	100 %
France 5	8140	100 %
France Ô	6834	100 %
TF1	6762	100 %
Canal+	7980	100 %
M6	6881	100 %
C8	5494	100 %
W9	7216	100 %
TMC	6 551	100 %

Source : Estimations fournies par les chaînes début 2019.

Tableau 3. Programmes accessibles en 2018 pour les chaînes dont la part d'audience est supérieure à 2,5 % de l'audience totale des services de télévision (volumes horaires et pourcentages, hors publicité et dérogations)

¹⁰¹<https://www.csa.fr/Informer/Informations-publiques-et-ressources-humaines/Les-rapports-annuels-et-bilans-du-CSA/Le-rapport-annuel-2018-du-CSA>



Chaîne	Obligation de sous-titrage en 2018	Volume annuel accessible (en heures)	Réalisé en % du volume total
Chaînes de la TNT gratuite			
TFX	60 %	5 095	75 %
NRJ 12	40 %	3081	49,98 %
Cstar	30 %	2300	32 %
Gulli	20 %	3993	51,26 %
TF1 Séries Films	40 %	6450	89 %
L'Équipe	40 %	3242	41 %
6ter	60 %	4583	64 %
RMC Story	40 %	4271	57,9 %
RMC Découverte	40 %	4939	66 %
Chérie 25	50 %	3659	58,13 %
Chaînes de la TNT payante			
Canal+ Cinéma	40 %	6230	83 %
Canal+ Sport	40 %	2572	42 %
Paris Première	40 %	3853	55 %
Planète+	40 %	3499	44 %

Source : Estimations fournies par les chaînes début 2019.

Tableau 4. Programmes accessibles en 2018 pour les chaînes dont la part d'audience est inférieure à 2,5 % de l'audience totale des services de télévision

Ainsi que le précise le Rapport Annuel relatif à l'accessibilité, « il n'existe pas d'obligation de traduire des émissions en LSF, hormis les engagements spécifiques des chaînes d'information en continu (...) leurs conventions prévoient qu'elles doivent mettre à l'antenne, en plus des 3 journaux télévisés sous titrés, un journal télévisé, traduit en LSF, du lundi au vendredi » (CSA, 2019). Pour sa part, franceinfo propose deux journaux télévisés en LSF par jour.

Chaîne	Tranche horaire de la traduction en LSF	Nombre de journaux télévisés	Volume annuel accessible (en heures)
franceinfo:	Du lundi au vendredi à 12h et 17h et le weekend à 11h et 19h	840	182 heures
BFMTV	Du lundi au vendredi à 13h	294	70 heures
Cnews	Du lundi au vendredi à 16h	274	67 heures
LCI	Du lundi au dimanche à 20h	341	36 heures

Tableau 5. Les engagements spécifiques des chaînes d'information en continu, 2018

En 2018, le CSA compte 47 heures 34 de plus aux 355 heures prévues. Seul LCI, la Chaîne Info du groupe tf1, a proposé 36 heures de programme signé en moins.

Selon les coûts d'interprétariat en LSF, cela représente un budget compris entre environ 300 000 et 2,6 millions d'euros.



Figure 39. Série – France Ô, exemple de sous-titrage

Le code couleur a son importance dans le sous-titrage pour malentendants. Selon le code couleur du Conseil Supérieur de l’Audiovisuel il signifie :

⇒ Blanc	Le locuteur est à l'image
⇒ Jaune	Le locuteur est hors-champ ou il y a une voix off
⇒ Cyan	Pensée ou flash-back du locuteur
⇒ Magenta	Indications musicales
⇒ Rouge	Indications sonores et descriptions de bruits
⇒ Vert	Langue étrangère (sans traduction)

Tableau 6. Code couleur sous-titrage du CSA

Le bandeau sombre vient renforcer le code couleur. Afin de faciliter la lisibilité des sous-titres DVB_Subtitling, notamment lorsque des sous-titres avec des caractères de couleur blanche sont diffusés sur une vidéo comportant un fond blanc, le Conseil a invité dès mars 2011, les éditeurs de la TNT (Télévision Terrestre Numérique) à diffuser des sous-titres DVB_Subtilting sur un bandeau sombre à l'image des sous-titres télétexte appelés à disparaître avec l'arrêt de la télévision analogique. Des dysfonctionnements quant à la réception des sous-titrés diffusés avec ce bandeau sombre ont été observés durant l'été 2011. À la suite, des travaux d'analyse et d'adaptation technique ont été menés par le Conseil afin que les équipements de génération des sous-titres des chaînes respectent bien la norme de sous-titrage EN 300 743. Les sous-titres diffusés sur certaines chaînes avec un bandeau sombre sont actuellement bien reçus par le parc de récepteurs TNT.

Dans ce même rapport, est présenté une « charte de qualité pour l'usage de la Langue des Signes Française dans les programmes télévisés (Janvier 2015) ». Issue d'une collaboration entre des



associations des communautés sourdes et malentendantes et les chaînes de TV concernées, rappelle les éléments qualitatifs de l'accessibilité.

Par ailleurs, cette charte propose également d'explorer de nouvelles solutions offertes par la télévision connectée.

5.5.3 En Belgique

Depuis Janvier 2019, le CSA (Conseil Supérieur de l'Audiovisuel) belge a mis en place de nouveaux objectifs d'accessibilité. Ils sont depuis fixés en fonction de l'audience et non plus comme anciennement par rapport aux chiffres d'affaires. Ci-dessous, un tableau récapitulatif des objectifs d'accessibilités 2019/2024 pour les chaînes belges, selon leurs audiences, à travers le prisme de la chaîne francophone RTBF ¹⁰².

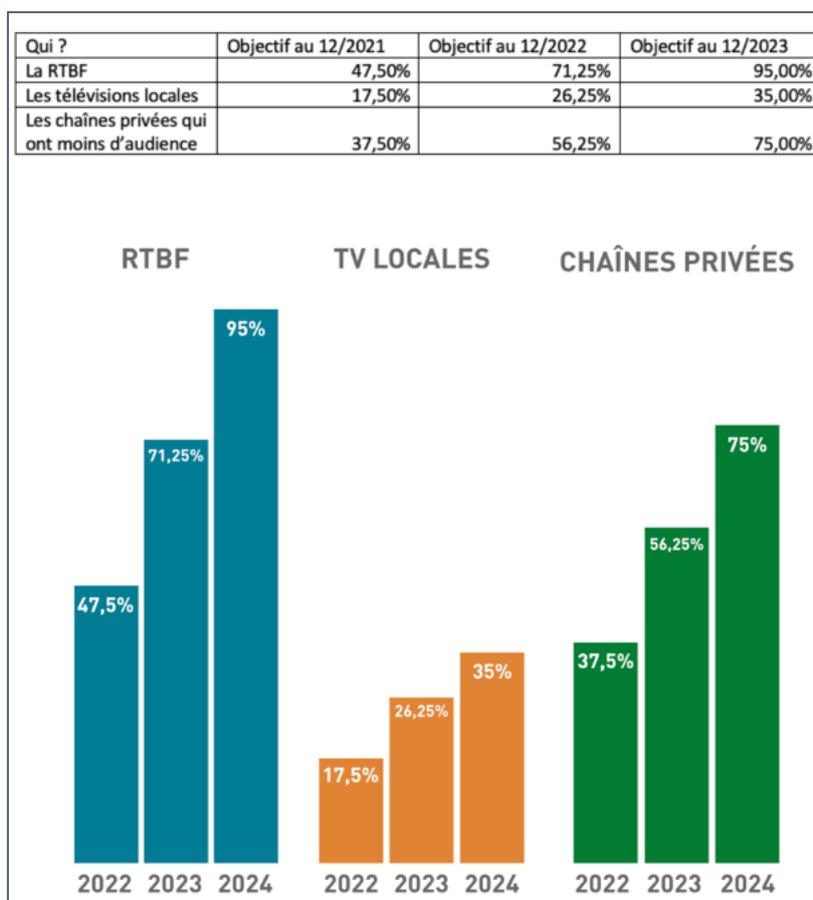


Figure 39. Pourcentage de programmes annuels sous titrés dès 2019

Concernant les VOD (Vidéo à la demande), elles doivent être accessibles pour 25% minimum des programmes.

Pour la langue des signes et malgré la pénurie d'interprètes, il s'agit de tout mettre en œuvre pour que les messages **d'intérêt général de sécurité à caractère urgent** soient, **si possible**, interprétés en langue des signes. Aussi, les messages **d'intérêt général de santé publique**, doivent être interprétés en langue des

¹⁰² <http://www.ffsb.be/audiovisuel/nv-reglement-access/>



signes. Il n'y a pas de quotas spécifiques pour la langue des signes : celle-ci est incluse dans l'obligation de sous-titrage¹⁰³ : (voir figure ci-dessous).

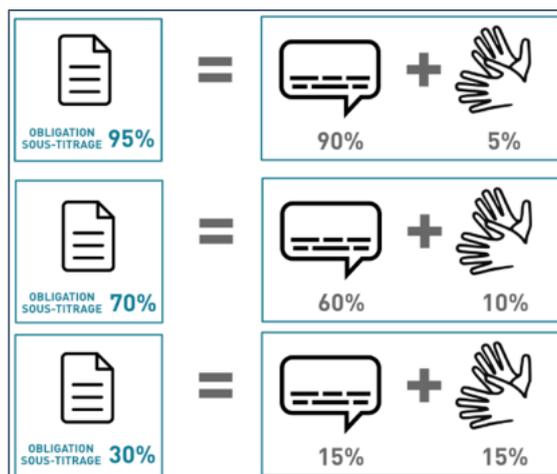


Figure 40. Quota de sous-titrage et de LSF pour la télévision belge

Concernant la communauté flamande, le principe de non-discrimination dans les réclames et les messages d'intérêt public a été introduit dans le décret du 27 Mars 2009. Il prévoit de plus l'accessibilité d'une proportion importante des programmes (sans préciser ce que signifie « proportion importante »). 100% des programmes d'information doivent être sous-titrés (public et privé) depuis la modification du décret (13 juillet 2012). En pratique, nombreux sont les programmes sous-titrés, en raison du bilinguisme de la Belgique.

La communauté germanophone n'a pas de législation en termes d'accessibilité.

5.5.4 En Espagne

La loi Générale de Droits des personnes handicapées et son inclusion sociale (loi 1/2013)¹⁰⁴ expose, dans un premier temps, les conditions d'exclusion dans lesquelles les personnes en situation de handicap se retrouvent quotidiennement à cause de la structure actuelle et du fonctionnement de la société et mentionne en particulier la restriction des droits et des libertés fondamentales dont ils subissent les conséquences. Malgré les importants progrès sociaux en la matière, la loi souligne que les droits des personnes handicapées à ce jour sont encore limités en termes d'accès aux environnements et aux processus, ce qui les empêche de mener une vie pleinement autonome et indépendante.

Selon la dernière enquête en 2008, de l'Institut national de statistique sur le handicap, l'autonomie personnelle et les situations de dépendance¹⁰⁵, le nombre de personnes handicapées en Espagne atteint 3,85 millions de personnes, soit 8,5 % de la population totale. Sur ce nombre, près d'un million (967 445) sont des personnes malentendantes. Un chiffre important puisqu'il s'agit d'un secteur où la population montre des difficultés à accéder aux technologies de l'information et de la communication.

¹⁰³ https://www.rtf.be/entreprise/contact-et-question/faq/detail_accessibilite?id=9309319

¹⁰⁴ <https://www.boe.es/buscar/doc.php?id=BOE-A-2013-12632>.

¹⁰⁵ https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176782&menu=resultados&secc=1254736194716&idp=1254735573175#!tabs-1254736194716



Toutefois, une étude de 2015¹⁰⁶ montre qu'il y a d'important progrès en termes d'accessibilité des programmes de la télévision espagnole, entre 2014 et 2015, impulsés par la loi générale de la communication audiovisuelle (7/2010). Ainsi cette étude comparative révèle que la production de sous-titres s'étend significativement aux chaînes télévisuelles espagnoles, en 2015.

La moyenne des programmes diffusés avec sous-titres est de 71,2% en 2015 tandis que la moyenne des programmes diffusés, en 2014, est de 67,2%. Il y a donc une augmentation significative.

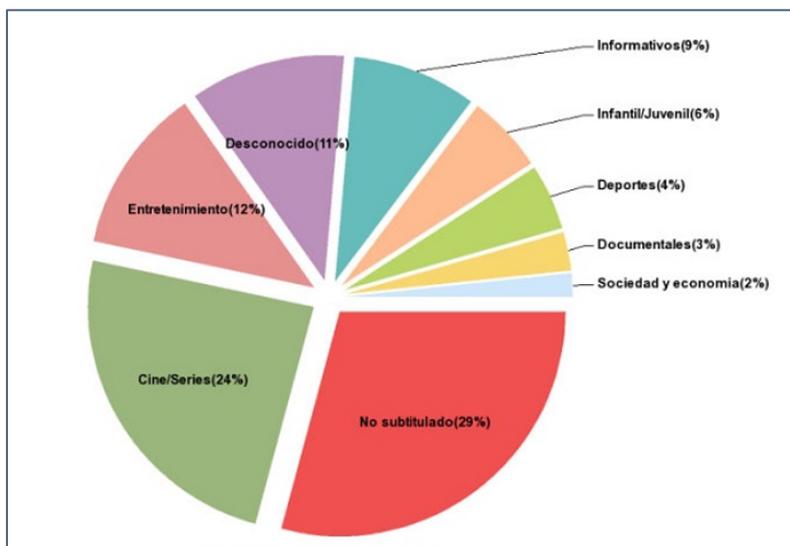


Figure 41. Utilisation des sous-titres en Espagne par genre en 2014-2015

De plus RTVE, chaîne publique espagnole, est déclarée, en 2011, la chaîne la plus accessible¹⁰⁷. Depuis 2014, la chaîne espagnole étend son service à d'avantage de programmes. Une traduction en langue des signes est également proposée¹⁰⁸. On note ainsi des progrès importants en matière d'accessibilité audiovisuelle pour tous en Espagne.



Figure 42. Exemple LES (Langue des Signes Espagnols) sur RTVE

¹⁰⁶ <http://www.cesya.es/articulos/201601211256>

¹⁰⁷ <http://www.rtve.es/rtve/20120713/tve-cadena-mas-programas-accesibles-para-personas-discapacidad-dio-2011-mas-30000-horas-subtituladas/546047.shtml>

¹⁰⁸ <http://www.rtve.es/alcarta/videos/en-lengua-de-signos/lengua-signos-07-07-19/5323440/>



Il y a une importante mobilisation des secteurs télévisuels et internet en Espagne autour de l'accessibilité de l'information audiovisuelle. Depuis 2010, grâce à La loi Générale de l'Audiovisuel¹⁰⁹ qui établit plusieurs paramètres au niveau de l'accessibilité des chaînes de télévisions : celles-ci doivent se conformer aux règles et respecter les quotas horaires de diffusion de sous-titrages SME, d'audiodescription ainsi que de langue des signes. Bien qu'il n'existe pas encore, en Espagne, de génération automatique de sous-titres, il y a un engouement notable autour des problématiques de l'accessibilité audiovisuelle.



Figure 43. Exemple sous-titrage sur 2 blocs de lignes.

En Espagne, les sous-titres sourds et malentendants sont répartis sur 2 blocs de lignes maximum. Chaque ligne peut compter jusqu'à 37 caractères pour 15 caractères par seconde. Les sous-titres sont généralement centrés en bas de l'image, sauf les indications musicales qui apparaissent en haut à droite de l'écran, en bleu entre parenthèses.



Figure 44. Indications musicales en haut à droite de l'écran

Afin de distinguer les personnages, les Espagnols ont recours à différentes couleurs et/ou typologies d'écriture. Par exemple lors de la manifestation d'une voix off à l'écran ou d'incorrection du langage le sous-titre apparaît en italique. Le traitement des couleurs en Espagne est moins strict qu'en France, c'est-à-dire que les couleurs sont généralement assignés à un personnage (selon le genre de film) et la

¹⁰⁹ <https://www.boe.es/buscar/pdf/2010/BOE-A-2010-5292-consolidado.pdf> (art.8)



palette de couleurs étant limitée, ils ont davantage recours aux étiquettes, aux guillemets ainsi qu'aux tirets. L'utilisation d'étiquettes consiste à attribuer une étiquette à chaque sous-titre, de manière à identifier explicitement le personnage. Les étiquettes sont utilisées lorsqu'il n'y a pas assez de couleurs pour identifier les personnages. Les guillemets et les tirets sont utilisés pour indiquer un dialogue.

Le blanc est généralement utilisé lorsqu'un personnage non central à l'histoire s'exprime tandis que le bleu est assigné aux effets sonores.

Un autre point lié à la multitude des versions des langues espagnoles attire notre attention. L'exemple le plus probant fut il y a quelque temps le sous-titrage mis en place par NETFLIX en espagnol (d'Espagne) du film mexicain d'Alfonso Cuarón : ROMA. Netflix a dû retirer les sous-titrages en espagnol castillan, jugé offensant par les espagnols ibériques (et par le réalisateur lui-même). Il est important en effet de se saisir et de prendre en compte cette différence, entre l'espagnol ibérique ou castillan, propre à l'Europe et l'espagnol, voire les différents espagnols américano-américain¹¹⁰ et de ne pas imposer le castillan par défaut.

5.5.5 Au Royaume-Uni

L'Ofcom¹¹¹ est l'autorité régulatrice des communications (télévisions, radios, VOD, lignes téléphoniques fixes, mobiles, services postaux, etc.) au Royaume Uni. Il œuvre en faveur de l'accessibilité et s'assure que les diffuseurs fournissent des proportions minimales de programmes avec sous-titrages, langue des signes et description audio.

Pour le sous-titrage : actuellement, 70 chaînes doivent fournir un niveau minimal de sous-titrage, la chaîne publique BBC s'engageant à sous-titrer 100% de sa programmation.

Pour la langue des signes : les chaînes à faible audience peuvent, au lieu de transmettre un quota d'émissions signées, verser une somme à la *British Sign Language Broadcasting Trust* : 60 chaînes ont adhéré à ce programme et des programmes sont ainsi visibles sur BSL Zone (Zone du *British Sign Language*) et sur le site internet de la BSLBT¹¹².

¹¹⁰ <https://www.bbc.com/mundo/noticias-46814069>

¹¹¹ <https://www.ofcom.org.uk/home>

¹¹² http://www.ofcom.org.uk/static/archive/itc/itc_publications/codes_guidance/sign_language_dtt/index.asp.html



Channel	Subtitling		Audio Description		Signing	
	Annual Quota	Achieved (2015)	Annual Quota	Achieved (2015)	Annual Quota	Achieved (2015)
Level One						
BBC1	100%	99.9%	10%	22.1%	5%	5.3%
BBC2	100%	99.9%	10%	24.8%	5%	5.2%
BBC3	100%	100%	10%	32.4%	5%	5.6%
BBC4	100%	100%	10%	36.3%	5%	6.0%
BBC News	100%	100%	Exempt		5%	5.7%
CBBC	100%	99.9%	10%	32.2%	5%	5.7%
CBeebies	100%	100%	10%	30.0%	5%	5.3%
ITV	90%	98.4%	10%	20.3%	5%	5.7%
ITV2	80%	94.6%	10%	30.0%	5%	5.7%
ITV3	80%	97.1%	10%	47.4%	5%	5.2%
ITV4	72%	86.0%	10%	29.8%	4%	5.0%
CITV	70%	97.3%	10%	29.8%	30 mins sign-presented a month	110 mins sign-presented a month
Channel 4	90%	100%	10%	25.7%	5%	5.2%
E4	80%	100%	10%	48.9%	5%	5.2%
Film 4	70%	100%	10%	25.6%	Alt	
More 4	70%	100%	10%	27.8%	Alt	
4Seven	30.8%	100%	5.7%	22.6%	Alt	
Channel 5	80%	88.2%	10%	14.7%	5%	10.0%

Tableau 7. Extrait des pourcentages obligatoires de sous-titrages pour les chaînes britanniques.

5.5.6 En Chine

La République Populaire de Chine (RPC) est composée de 56 ethnies¹¹³, toutes égales en droits et en devoirs selon la Constitution. Les linguistes parlent de 129 langues utilisées en Chine. Le Han est parlé ou compris par 92 % des ethnies en Chine¹¹⁴.

¹¹³ <http://www.stats.gov.cn/english/Statisticaldata/CensusData/rkpc2010/html/A0106a.htm>

¹¹⁴ https://fr.wikipedia.org/wiki/GROUPES_ethniques_de_Chine

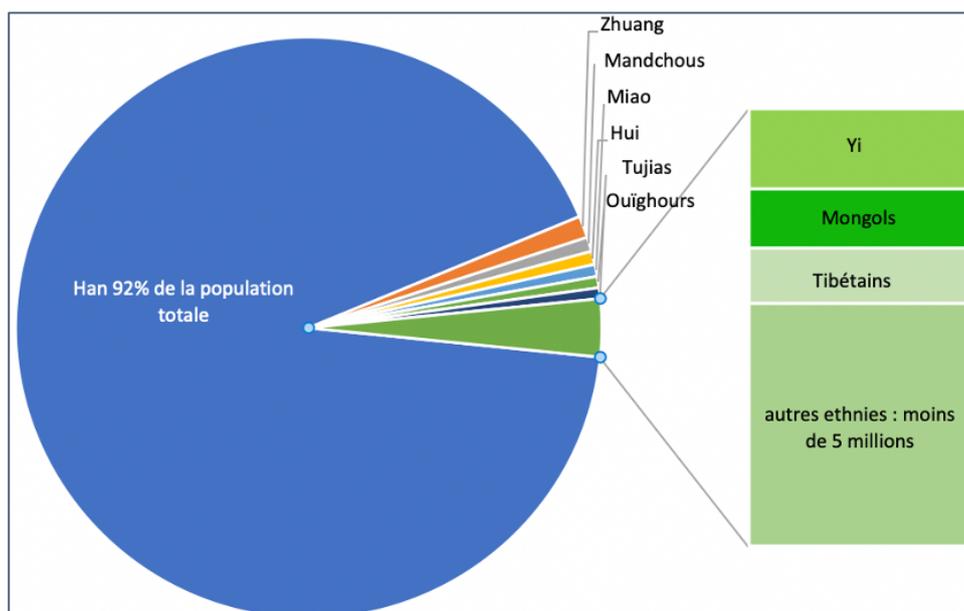


Figure 45. Représentation des différentes ethnies chinoises

En 2013, la chercheuse chinoise Che évoquait plus 2000 chaînes de télévisions en Chine (Che, 2013). Parmi celles-ci, la CCTV (Groupe public d'État), les chaînes de télévisions provinciales et les stations municipales.

Le groupe CCTV diffuse 22 chaînes dont des chaînes pour l'international : en anglais, en français, en russe, en arabe et une dernière pour l'Afrique.

Il n'y a pas de standards nationaux unifiés (Rovira-Esteva & Casas-Tost, 2017). On peut voir un film sous-titré en une seule langue, parfois deux, et quelque fois, pas sous-titré du tout. Les sous-titres peuvent être sur une ligne, ou bien sur deux. La couleur, la taille, et la police de caractères varient selon les différentes compagnies. D'un autre côté, les centres de traduction de langues minoritaires peuvent avoir une procédure uniformisée (exemple de Huace, compagnie privée). Le doublage demeure une technique dominante pour les films étrangers, voire les films sur les chaînes de cinéma... le sous-titrage sera préféré dans le cadre des plateformes de *streaming*.

La télévision s'adresse à un large public hétérogène, tandis que les plateformes de *streaming* s'adressent à un public ayant un niveau scolaire en anglais suffisant pour suivre le programme.

De plus concernant les traductions en Chine : l'anglais n'est pas la langue cible la plus importante pour les films chinois, par contre, cela concernera plus les langues cibles de la BRI (Belt and Road Initiative, nouvelle route de la soie chinoise¹¹⁵) : 20 langues dont le Swahili, l'Hausa, le russe etc.

Peu de films étrangers sont officiellement admis en Chine (33 en 2014, 28 en 2015, 51 en 2016, entre 60 et 70 ces dernières années¹¹⁶) et leurs traductions sont exclusivement faites par quatre des centres étatisés : Shanghai Film Translation and Production Studio, August First Film Studio, Changchun Translation and Production Studio, and China Film Group Corporation et où le travail est partagé équitablement.

¹¹⁵ Il s'agit ici de liens commerciaux, politiques, culturels privilégiés par la RPC

¹¹⁶ <https://larevuedesmedias.ina.fr/infographie-le-marche-du-cinema-chinois>



Notons que les centres de traduction de langues minoritaires sont importants pour la préservation des langues ethniques en République Populaire de Chine : Il y a aujourd'hui 11 centres étatiques qui traduisent et doublent les films vers 38 dialectes et 17 langues ethniques mineures (Jin & Gambier, 2018).

Les films diffusés en antenne, par la CCTV sont de l'ordre de 400, tandis que ceux diffusés sur les chaînes Internet (Youku, Sohu, Tencent, Iqiyi), sont de l'ordre de 1000, par an. Les traductions sont faites pour ces dernières soit par des sociétés privées, soit via le fansubbing¹¹⁷. Bien qu'il s'écarte (ou au contraire, parce qu'il s'en écarte) des conventions de traduction, le fansubbing, et les fansubbers sont présentés comme des « héros de l'interculturalité », et ce malgré les fermetures des sites (YYeTs, Shooter¹¹⁸) par les autorités, pour violations des droits d'auteur.

Selon le département de la réhabilitation psychologique du *Chinese Rehabilitation Research Center*, les chiffres du recensement (1987-1996), la Chine comptait 32,175 millions de sourds et malentendants, soit 2,3 % de la population totale et 36% de l'ensemble des personnes handicapées, à savoir la déficience la plus présente parmi les populations handicapées¹¹⁹.

Depuis 2008, la loi de Protection des personnes handicapées (art 43 et suivants¹²⁰) incite fortement à la prise en compte des handicaps en favorisant l'accès à la culture sous toutes ses formes (télévision, cinéma, littératures, etc.) Mais c'est une accessibilité très relative, il n'y pas d'obligation en tant que telle, c'est plutôt une incitation : il y a encore aujourd'hui un déséquilibre en termes de développement en faveur des personnes handicapées¹²¹.

L'État est tenu de normaliser et de promouvoir la langue des signes et le braille¹²² à travers les objectifs suivants : Les « Esquisses du plan de réforme et de développement à long et moyen terme des langues parlées et écrites de la Chine (2012-2020) » et le « Plan national de réforme des langues parlées et écrites pendant la 13e période quinquennale (2016-2020) ».

La langue des signes et le braille étaient inscrits dans ce plan d'ensemble. Le « Lexique des expressions communes en langue des signes nationale chinoise » et le « Schéma braille commun chinois » ont été publiés et sont entrés en vigueur en 2018. Des réunions importantes telles que les sessions plénières de l'Assemblée Nationale Populaire (organe législatif suprême) ont une interprétation en langue des signes en direct, et China Media Group et certaines stations de télévision locales proposent une interprétation en langue des signes pour certains programmes importants. En 2018, les stations de télévision provinciales et préfectorales / urbaines ont diffusé 295 programmes en langue des signes et les stations de radio ont diffusé 230 programmes pour les personnes handicapées¹²³.

Il est constaté par certains (et notamment dans les rares associations) que la société civile aurait été jusqu'à très récemment mise à l'écart dans les processus ou mises en œuvre d'initiatives politiques formelles impactant les personnes handicapées (et cela en vertu de la Convention des Nations Unies, ratifiée par la Chine en 2008). En effet, le handicap est perçu « médicalement » comme une maladie

¹¹⁷ Sous-titrage réalisé par des amateurs de série

¹¹⁸ YYeTs (人人影视) est la plus grande communauté de fansubbers, Shooter(射手网) est un portail de fansubbing populaire

¹¹⁹ <https://www.dinf.ne.jp/doc/english/asia/resource/z00ap/003/z00ap00308.html>

¹²⁰ http://www.cdpf.org.cn/english/Resources/lawsregulations/201707/t20170730_601596.shtml

¹²¹ http://french.xinhuanet.com/2019-07/25/c_138256906.htm

¹²² http://www.chinadaily.com.cn/regional/2019-07/25/content_37497820_8.htm

¹²³ http://www.cdpf.org.cn/english/Resources/documents/201907/t20190726_658048.shtml



que l'on traite, des handicaps dont il faut atténuer les impacts pratiques¹²⁴). En conséquence, il s'agit pour les associations de défendre les droits des personnes handicapées, en sensibilisant le public, les entreprises. Ainsi, une alliance, fondée en 2013 et coparrainée par l'Information Accessibility Research Society, Alibaba, Tencent, Baidu, Microsoft-Chine, vise à promouvoir le développement de l'accessibilité à l'information¹²⁵.

A l'heure actuelle, la plupart des produits ou services chinois tels que QQ, QQ Space, WeChat, (messageries instantanées) Tencent, Tencent News, App Bao, Penguin FM (assistant vocal), etc. ont été optimisés pour être « *barrier-free* », c'est-à-dire, accessibles. En 2018, Tencent remportait le prix UNESCO « pour sa contribution exceptionnelle à l'autonomisation des personnes handicapées »¹²⁶.

5.5.7 En Australie

Selon l'ACMA (Autorité australienne des communications et des médias¹²⁷), les sous-titrages dits « Captions », sont obligatoires pour les diffuseurs publics (ABC, SBS) sur la totalité de leurs programmes de 6h à minuit (y compris les rediffusions), sauf si les programmes sont :

- En langue étrangère (pas en anglais),
- Musicaux,
- Des vidéos à la demande,
- Des programmes communautaires,
- Des publicités.

Pour les chaînes privées, cela dépend du caractère du programme. Les chaînes de cinéma ont une obligation (pour 2024) de sous-titrer à hauteur de 100%, tandis que pour les chaînes d'infos, de sports ou de musique, c'est respectivement de l'ordre de 65, 65 et 55% à atteindre en 2024.

Concernant les messages d'alerte, tous les diffuseurs ont obligation de sous-titrer, mais pas obligatoirement de traduire en langue des signes (AUSLAN pour l'Australie).

5.5.8 Dans les Amériques

5.5.8.1 Au Brésil : l'adoption d'un avatar signeur sur les sites du gouvernement

Le 6 juillet 2015, la loi n°13.146 oblige les sociétés brésiliennes ou ayant une antenne au Brésil à assurer l'accessibilité de leur site Web aux personnes handicapées. Cette loi impose aux entreprises publiques ou privées de garantir aux personnes handicapées l'accès aux informations disponibles « conformément aux meilleures pratiques, directives, normes d'accessibilité adoptées à l'échelle internationale »¹²⁸.

¹²⁴ <https://www.newamerica.org/cybersecurity-initiative/digichina/blog/how-ai-can-better-serve-people-with-disabilities-in-china/>

¹²⁵ <http://www.capa.ac/ai>

¹²⁶ <https://fr.unesco.org/news/prix-unescoemir-jaber-al-ahmad-al-ahmad-al-jaber-al-sabah-lautonomisation-personnes-handicapees>

¹²⁷ <https://www.acma.gov.au/captions-tv-broadcasters-must-provide>

¹²⁸ https://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2015/Lei/L13146.htm



Cela signifie que le Brésil tend à faire adopter systématiquement les normes internationales d'accessibilité telles que proposées par le W3C. Le paysage audiovisuel étant très fragmenté, cette loi n'impose pas de normes d'accessibilité pour les émissions de télévision.

L'article 63 de la loi 13.146 du 6 juillet 2015¹²⁹ oblige les sites Web brésiliens à adopter les normes d'accessibilité garantissant l'accès aux informations aux personnes en situation de handicap. Depuis, les sites publics et notamment le site du gouvernement ont leur version accessible en Libras, bien que certains linguistes critiquent cet outil développé par Hand Talk. Un plug-in installé sur le moteur de recherche permet de traduire les contenus écrits en Libras, la langue des signes portugaise.

5.5.8.2 Au Canada : le défi du bilinguisme

Le Conseil de la Radiodiffusion et des Télécommunications Canadiennes (CRTC) a créé, en 2007, des normes de sous-titrages dans une démarche collaborative regroupant des producteurs/diffuseurs de médias, des spécialistes du captioning ainsi que des représentants de sourds et malentendants. Ces normes sont tant au niveau quantitatif que qualitatif.

Dès 2012, les travaux d'un groupe de travail francophone ont permis l'élaboration de normes universelles du sous-titrage codé à l'intention des télédiffuseurs canadiens de langue française¹³⁰.

Aujourd'hui, le CRTC impose à la plupart des télédiffuseurs¹³¹ le sous-titrage codé intégral des émissions diffusées durant la journée (de 6h à minuit). En ce qui concerne les émissions préenregistrées, les télédiffuseurs doivent viser un taux de précision du sous-titrage codé de 100 %. Pour la programmation en direct, le sous-titrage codé de langue française doit atteindre un taux de précision de 85 %. Pour le sous-titrage codé en direct de la langue anglaise, ce taux de précision est attendu au moins 98 %¹³².

A ce stade de l'étude, nous n'avons pas d'information concernant l'écart d'exigence à la précision entre les sous-titres en langue anglaise et les sous-titres en langue française, pour les contenus en direct. En effet, depuis septembre 2019, le CRTC a fait adopter de nouvelles normes de qualité obligatoires pour le sous-titrage codé de langue anglaise relatives au taux de précision de la programmation en direct. Depuis ce jour, le taux de précision du sous-titrage en langue anglaise est évalué par le NER (Named Entity Recognition¹³³), un outil de comparaison de l'expérience du spectateur utilisant les sous-titres à celle du spectateur auditif. Pour ce faire, le NER compare, de manière automatisée, une transcription exacte de l'audio et le contenu des sous-titres sur le même morceau de contenu. Cet outil de mesure du taux de précision n'est pour l'instant appliqué que pour les sous-titrages codés de langue anglaise relatifs aux programmes en direct.

¹²⁹ https://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2015/Lei/L13146.htm

¹³⁰ Normes universelles du sous-titrage codé à l'intention des télédiffuseurs canadiens de langue française, Document révisé en fonction de la politique réglementaire de radiodiffusion CRTC 2011-741 et CRTC 2011-741-1 Mai 2012. In :

https://www.cab-acr.ca/french/societal/captioning/normes_universelles.pdf

¹³¹ Par exemple les télédiffuseurs publics sont obligés de sous-titrer à 100% leurs contenus. Les télédiffuseurs "indépendants" ont une obligation moindre, comme en France.

¹³² https://crtc.gc.ca/fra/info_sht/b321.htm

¹³³ Il s'agit d'un outil d'extraction de données. Sur une énorme base de données, cet outil est capable de reconnaître les mots et leur type. De dire que telle donnée est une date, telle donnée est un nom, telle donnée un lieu etc. A partir de cet outil, les extraits sont comparés automatiquement. Ainsi est calculé le taux d'exactitude du sous-titrage tel qu'il a été diffusé, par rapport à une transcription à posteriori. Voir : <https://www.smalsresearch.be/named-entity-recognition-une-application-du-nlp-utile/>



Les normes qualitatives imposées par le CRTC concernant, quant à elles, l'ensemble des contenus (en langue anglaise comme en langue française). Le Conseil a mis en place un protocole de surveillance sur la qualité du sous-titrage (marge d'erreur de 5%, temps de latence minimum entre l'audio et le texte, correction des erreurs avant rediffusion) et la possibilité pour les individus de déposer des plaintes concernant le sous-titrage.

Les interventions officielles du gouvernement sont diffusées dans les deux langues, le locuteur énonce le discours en anglais et en français, successivement. Sur les sites internet correspondants aux chaînes de télévision canadienne accessibles depuis la France (en l'occurrence la chaîne CPAC¹³⁴), trois possibilités de visionnage d'une même vidéo sont proposées : en français, en anglais ou dans la langue originelle de l'audio. Choisir l'option « français » traduit toutes les parties anglaises en langue française par le biais d'une voix off qui n'intervient que lorsque le locuteur parle anglais. La voix off est assurée en direct par un traducteur. Choisir de conserver la langue originale de l'audio permet de regarder le contenu dans les deux langues, selon la langue parlée par le locuteur, dans le respect du bilinguisme total de la région.

Sur le site de replay de la chaîne d'information CPA « les vidéos sont équipées de deux médaillons représentant deux interprètes langue des signes : l'un en langue des signes américaine, l'autre en langue des signes française (ou québécoise) ».



Figure 46. Capture d'écran d'une vidéo de replay, déclaration officielle du premier ministre et traduction LS par deux interprètes

Dans le cas où l'allocution est faite par le premier ministre de l'Ontario par exemple, une province uniquement anglophone, le discours est traduit par un seul interprète en langue des signes américaines. Sur les vidéos en direct, les *closed captions* sont disponibles et générées de manière automatique¹³⁵ par une société française : Dalet¹³⁶.

¹³⁴ <https://www.cpac.ca/fr/programs/covid-19-la-reponse-du-canada/episodes/66189568/>

¹³⁵ <https://www.tvtechnology.com/equipment/the-case-for-automating-closed-captioning>

¹³⁶ <https://www.dalet.com/fr/platforms/dalet-amberfin>



Figure 47. Capture d'écran allocution officielle uniquement anglophone et closed captions générées automatiquement.

5.5.8.3 États-Unis¹³⁷ : la loi fédérale américaine en matière d'accessibilité à l'épreuve d'un système audiovisuel typiquement privé

En 1998, les États-Unis promulguent une loi en faveur de l'accessibilité du contenu vidéo diffusé en ligne¹³⁸. L'article 508 exige l'accessibilité numérique dans les programmes ou services fédéraux. Ceux qui ne respectent pas cette loi se verront refuser les aides de l'État.

Adoptée dès 1990, l'ADA (*Americans with Disabilities Act*) a fixé des exigences d'accessibilité historiques qui affectent à la fois les entités privées et publiques. La portée de l'ADA est très large, elle peut à la fois exiger l'accès aux bâtiments par une rampe d'accès pour les personnes handicapées mais aussi exiger des médias de diffusion de fournir un sous-titrage pour toutes vidéos promotionnelles, présentations ou publicités diffusées dans un lieu public.

Le FCC (*Federal Communications Commission*) a œuvré en faveur de l'accessibilité de la télévision en permettant aux utilisateurs de personnaliser le contenu, de grossir la taille de police ou changer la couleur. La commission a permis d'inclure la notion de qualité dans la législation, notion que l'ADA n'imposait pas.

En 2010, le CVAA (*Communications and Video Accessibility Act*) a élargi le champ de l'ADA à tous les médias en ligne qui ont préalablement été diffusés à la télévision. C'est-à-dire que les contenus diffusés sur Internet, après avoir été diffusés à la TV, doivent eux-aussi être dotés de sous-titrages.

A ce jour, aux États-Unis les sous-titres apparaissent par le déroulement (« roll-up »), le sous-titre défile sur trois lignes, de bas en haut et se situe en bas de l'écran. Lorsque cela n'est pas possible (si le sous-titres n'est pas visible à cause de la couleur de fond) les sous-titres apparaîtront sur deux lignes en haut de l'écran

Bien souvent, ces législations que l'on retrouve au Canada et aux États-Unis, sont portées par des collectifs et associations influents, représentatifs des populations sourdes et malentendantes.

Or, spécifions qu'aux États-Unis, il existe trois systèmes de diffusion de la télévision : la télévision hertzienne (disponible gratuitement) et la télévision par câble ou par satellite qui exigent quant à elle un abonnement payant. Sur le réseau hertzien, les Américains ont une diffusion décentralisée, il n'existe pas de bouquet de chaînes diffusées au niveau national. Aujourd'hui, les chaînes locales diffusent leurs propres émissions d'information, le reste est acheté à des sociétés de productions

¹³⁷ <https://www.ada.gov/cguide.htm>

¹³⁸ Modification des articles 504 et 508 du Rehabilitation Act de 1973



indépendantes. Cette indication permet de penser différemment le système de diffusion audiovisuel américain et de mettre en perspective la question d'outils d'accessibilité généralisés.



6 Opportunités dans d'autres secteurs

Dans cette partie nous questionnons les usages potentiels des outils d'accessibilité audiovisuels automatiques susmentionnés, à savoir le sous-titrage et la transcription automatiques, la traduction automatique et la génération de contenus en langue des signes.

Durant la phase de recherche documentaire, nous avons recensé des domaines susceptibles d'être intéressés par ces solutions. Seront présentés d'abord les opportunités qu'offrent les nouveaux journalismes.

Puis nous nous intéresserons au domaine de l'éducation en faisant un état des lieux des méthodes éducatives innovantes et accessibles. Nous interrogerons également les pratiques de communications des entreprises multinationales ou encore des administrations qui tendent à plus d'accessibilité pour une meilleure diffusion de leurs messages.

6.1 Nouveaux journalismes sur tous les supports

L'émergence de vidéos de *news* sur les réseaux sociaux s'explique par deux facteurs : l'usage du smartphone et des réseaux sociaux comme premier canal de consultation des *news*¹³⁹ et la surpondération de l'image animée par rapport à l'écrit sur les flux de contenus des réseaux sociaux. L'actualité est d'autant plus consultée par les 18-34 ans sur les réseaux sociaux et la visibilité est meilleure si cette actualité est sous-forme de vidéo.

Les acteurs principaux de ces magazines digitaux en France sont Brut et Konbini ou encore Franceinfo depuis quelques temps. La formule de ces vidéos est très normée : des vidéos courtes et surtitrées diffusées sur des canaux tels que Twitter et Facebook. On parle de surtitrage et non de sous-titrage puisqu'il s'agit là d'une transcription audio du discours qui prend une grande place sur la vidéo : le surtitrage devient un élément graphique de la vidéo, symbole de l'identité du média¹⁴⁰.

Le surtitrage des vidéos de Brut par exemple est très normé : ils ont fait le choix du jaune ou du vert pour se distinguer du blanc, pour mettre l'accent sur un terme. Pour s'adapter au mieux aux pratiques existantes sur leurs canaux de diffusion, ils ont choisi de conserver des éléments spécifiques des réseaux sociaux comme les flèches faites à la main, les gifs¹⁴¹ ou les mèmes¹⁴² intégrés.

¹³⁹ <https://www.ifop.com/wp-content/uploads/2019/02/115960-Présentation-version-publiée.pdf>

¹⁴⁰ Définition : Le mot surtitre (qui s'écrit sans trait d'union) désigne à l'origine la traduction simultanée des paroles d'un opéra (ou d'un texte chanté ou encore d'une pièce de théâtre) projetée sur un écran placé habituellement au-dessus du cadre de scène (angl. proscenium arch). Le mot sous-titre (qui s'écrit avec un trait d'union) désigne cette même traduction, mais placée au bas d'un écran. Dans le monde de la presse écrite, un surtitre désigne une ligne placée au-dessus du titre proprement dit et qui fournit un aspect complémentaire d'information. In : <https://roberge.mus.ulaval.ca/gdrm/09-surti.html>

¹⁴¹ Format d'image animé

¹⁴² Élément culturel transmis par les réseaux sociaux, décliné en masse



Figure 48. Les normes graphiques du surtitrage des vidéos Brut.

A ce jour, ces termes « surtitrage » et « sous-titrage » peuvent toujours porter à confusion. Le surtitrage tels que chez Konbini, en gros caractères, est devenu leur identité graphique. Lors des recherches nous avons pris connaissance d'offres de stage proposées par Konbini pour réaliser les sous-titrages (qui en réalité sont des surtitrages) de leurs vidéos, l'annonce indiquait qu'il s'agissait d'un sous-titrage manuel ; le recours à une transcription automatisée peut être une solution souhaitable pour ces médias.

6.2 Recours aux outils d'accessibilité dans le monde de l'éducation

La diffusion des savoirs au plus grand nombre est au cœur des objectifs des acteurs de l'éducation, en ce sens, l'accessibilité des apprentissages, quels qu'ils soient, est un enjeu fondamental.

Nous avons répertorié certaines initiatives allant en ce sens comme l'existence de MOOC (Massive Open Online Course) accessibles ou bien le projet européen TraMOOC visant à créer un outil de traduction des MOOC. Nous avons également relevé une application d'apprentissage de la lecture à destination d'enfants Sourds développé par Huawei. Ces initiatives ont pour objectif d'accroître l'accessibilité des apprentissages.

6.2.1 Le projet « MOOC TEAM » : une éducation inclusive à l'environnement

En 2019, la plateforme « FUN MOOC »¹⁴³ a hébergé un module intitulé « Agir ensemble pour réduire la présence de métaux toxiques dans notre assiette. Transferts environnementaux des contaminants métalliques »¹⁴⁴. Il s'agit d'un cours accessible à tous selon les critères européens d'accessibilité. Deux versions sont proposées de ce MOOC : une version en voix off sous-titrée et une version en LSF sous-titrée. L'apprenant choisit l'une ou l'autre, sachant que ces deux versions sont proposées pour tous les cours du MOOC. Sur la même page WEB s'affichent les deux vidéos et l'individu clique où il souhaite. Des scripts et des diapositives sont à disposition.

¹⁴³ France Université Numérique, plateforme française de MOOC lancée en 2013 : <https://www.fun-mooc.fr>

¹⁴⁴https://www.fun-mooc.fr/courses/course-v1:univ-toulouse+101020+session01/about?fbclid=IwAR2j7lapGMkxumGd8kdIj6PRZ8eWIE_aG90FxoVP6SSiYnfM8-KksTIHM9M



Les cours ne sont pas signés par un interprète mais traduits par des traducteurs sourds qui ont reçu préalablement le script et ont préparé leur traduction vers la langue des signes.

La correction a été effectuée par une personne scientifique sourde dont la spécificité est que la LSF soit sa seconde langue. Ce projet novateur, appelé « MOOC TEAM », œuvre pour une éducation inclusive à l'environnement (Laffont, et al., 2019).

6.2.2 Le projet européen TraMOOC

Le projet TraMOOC¹⁴⁵, financé dans le cadre de H2020 a été lancé en février 2015. Il avait pour but de fournir une traduction automatique fiable et adaptée aux MOOC¹⁴⁶ : une traduction qui portait de l'anglais vers onze langues européennes ainsi que les langues des BRIC (Brésil, Russie, Inde, Chine).

Habituellement, les solutions de traduction automatique MT (Machine Translation) sont utilisées pour des usages commerciaux et sont performantes dans leur domaine spécifique. La traduction automatique neuronale apporte une amélioration significative de la qualité de traduction par rapport aux systèmes de Machine Translation. Cette dernière permettait à l'outil TraMOOC d'être adapté au domaine éducatif et ainsi d'être capable de gérer la nature multiple des informations contenues dans les MOOC.

La qualité des traductions devait être évaluée par des traducteurs professionnels. L'évaluation de la qualité de la technologie portait sur le calcul de l'effort post-édition requis pour que les traductions soient publiables. Ce projet dont les partenaires sont uniquement des universitaires a pris fin en 2018 et ne semble pas avoir été commercialisé.

6.2.3 StorySign : plateforme d'alphabétisation pour enfants sourds

En 2018, Huawei a dévoilé une application pour smartphone « STORYSIGN » dont la fonction est de traduire en langue des signes des livres pour enfants. Version en LSF est montrée via un avatar appelé « STAR » qui signe en temps réels les histoires captées par l'appareil photo du smartphone grâce à « l'intelligence artificielle et la réalité augmentée »¹⁴⁷. Initialement, l'application propose toute une bibliothèque de livres en langue des signes américaine, en effet, l'ASL serait la langue des signes la plus répandue au monde¹⁴⁸. L'application se diversifie et offre désormais 13 langues.

¹⁴⁵ <https://cordis.europa.eu/project/id/644333/reporting>

¹⁴⁶ <https://www.k4all.org/project/tramooc/>

¹⁴⁷ <https://consumer.huawei.com/fr/campaign/storysign/>

¹⁴⁸ <https://www.handirect.fr/storysign-apprentissage-de-la-lecture-aux-enfants-sourds/>

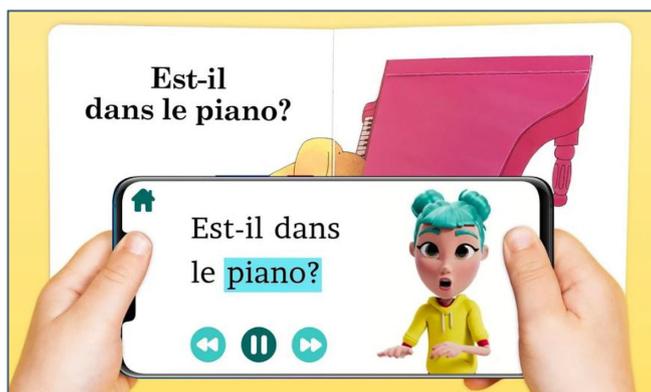


Figure 49. Démonstration de l'avatar signeur STAR de l'application Storysign qui vise à encourager l'alphabétisation des enfants Sourds

L'application a été développée avec le studio Aardman (Wallace et Gromit) et a tenu compte des conseils d'experts des communautés de personnes sourdes et malentendantes.

6.3 Communication institutionnelle : appel à des prestataires de services extérieurs

Les administrations et institutions comptent parmi leurs fonctions celle de communiquer avec ses administrés ou ses bénéficiaires par exemple. Les entreprises communiquent elles aussi à l'extérieur dans un cadre professionnel et /ou auprès du grand public. Leurs modèles de communication, notamment sur Internet à destination du grand public se doivent d'être accessibles. Nous avons interrogé les modèles de communication de la commission européenne, mais aussi de certains services publics français ou entreprises.

6.3.1 La Commission Européenne fait appel à des interprètes extérieurs

La Commission Européenne comptait, en 2019, deux eurodéputés sourds qui ont chacun été assistés de deux interprètes (l'un pour la langue parlée vers la langue des signes, l'un pour la langue des signes vers la langue parlée). Selon le responsable de l'unité d'interprétation de la Commission Européenne, les équipes ne disposent pas d'interprète salarié en langue des signes. Quand les organisateurs de conférences ou de réunions le demandent, ils font appel à des interprètes extérieurs (freelance). Dans ces cas précis, ils utilisent majoritairement la langue des signes internationale (et non pas la langue des signes américaine) ou bien la langue des signes de l'un des pays européens concernés.

Dernièrement, la Commission Européenne a publié une vidéo à propos du Covid-19 et à destination de tous les Etats membres. Cette vidéo était disponible en anglais, français, allemand, italien, espagnol et langue des signes britannique. Des sous-titres étaient disponibles pour chacune des 28 langues de l'UE. La Commission Européenne a donc recours à des services de traduction (d'une langue parlée vers une autre ainsi que d'une langue parlée vers une langue des signes, et inversement) humains et non automatiques.

6.3.2 Entreprises et institutions françaises qui sous-traitent leurs solutions de communication accessibles à des opérateurs-relais privés

Nous avons également questionné les modes de communications des services publics français tels que la Caf (Caisse d'allocation familiales) ou Pôle Emploi. Ainsi que des grandes entreprises telles qu'EDF



ou Air France. Tous ces organismes ont en commun de sous-traiter leurs systèmes de communication accessibles. Chacun de ces organismes fait appel à des plateformes extérieures d'opérateurs-relais.

Lorsqu'un bénéficiaire de la Caf, ayant une déficience auditive par exemple, tente de joindre le service public par téléphone, on lui propose un RDV téléphonique par visio-conférence dans lequel il y aura trois participants : le conseiller Caf, le bénéficiaire ainsi que l'interprète en langue des signes. Une autre option est proposée au bénéficiaire (ou au client, selon la structure contactée) : il s'agit de la transcription écrite. Le transcripteur retranscrit manuellement le discours du conseiller et transforme ainsi l'échange en une sorte de « chat » sans que l'entreprise source n'en soit équipée.

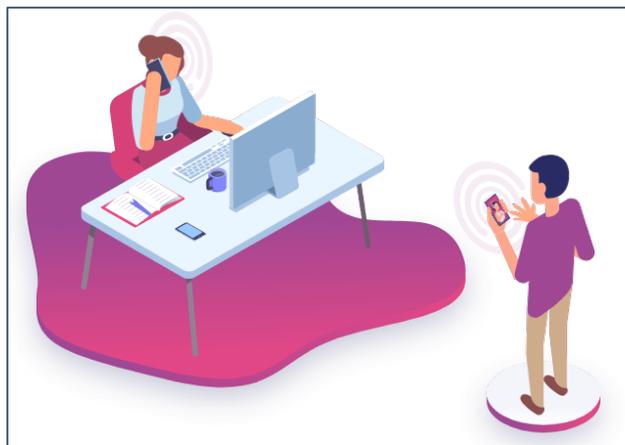


Figure 50. Illustration représentative des services proposés par les opérateurs-relais

Ces prestataires privés de transcription et d'interprétation en langue des signes sont peu nombreux et assurent aujourd'hui leurs fonctions grâce à des interprètes salariés. Un responsable de l'une de ces entreprises a jugé utile, dans un récent échange, d'accéder à l'automatisation de la transcription dans le cas d'appels nocturnes par exemple.

La création d'outils de transcription, de traduction, de sous-titrages automatiques ainsi que la génération de contenu en langue des signes, peut occasionner des opportunités dans la communication institutionnelle : d'une part autour des vidéos proposées (événementielles, publicitaires, informationnelles) et d'autre part autour de nombreux domaines hors audiovisuel.

Ces outils peuvent faciliter la production des vidéos d'information mais aussi la diffusion des savoirs à travers des MOOC 100% accessibles ou des logiciels à destination de populations aux besoins spécifiques. Ils peuvent aussi s'avérer très utiles pour les entreprises, institutions tant dans leurs modes de communication internes qu'externes.

Les recherches documentaires effectuées montrent, de manière générale, que les entreprises et institutions rendent accessibles leurs services en sous-traitant leurs plateformes de service à la clientèle. A ce jour, nous n'avons pas trouvé d'informations révélant que certaines entreprises ou institutions, à l'international, utilisent des outils d'accessibilité automatiques pour communiquer de manière accessible avec leurs clients ou leurs bénéficiaires.

Cependant, il semble évident que de plus en plus d'entreprises, marques, institutions rendent accessibles leur site Internet en permettant de modifier l'interface en fonction des besoins du visiteur. L'Oréal ou encore d'Unilever par exemple, se sont engagés en faveur de l'inclusion à travers leurs sites internet.

Le site www.loreal.com affiche être en conformité avec le WCAG 2.0 (*Web Content Accessibility Guide Online* ou guide d'accessibilité des contenus en ligne). Unilever affiche sa politique d'inclusion en



précisant que chacune des vidéos de leur site internet est accompagné d'une vidéo textuelle détaillée par exemple. Ces initiatives en faveur d'actions inclusives permettent de penser que certaines institutions ou entreprises puissent s'intéresser au champ d'application du projet ROSETTA.

7 Conclusion et prochaines étapes

7.1 Conclusion

Le but de cette étude comparative était de recenser ce qui est produit (projets, solutions, littérature) en termes de **sous-titrage automatique, de traduction automatique des sous-titres ainsi qu'au niveau de la génération automatique de contenu en LSF (langue des signes) et l'animation d'un avatar signant**. *In fine*, ces solutions ont pour cibles les personnes sourdes, malentendantes, les personnes étrangères en voie d'apprentissage de la langue nationale, les personnes âgées et de manière générale toutes les personnes ayant des difficultés d'accès à la langue parlée (française en France).

Les résultats recensent les avancées techniques et portent sur les régions linguistiques visées par le projet ROSETTA (hispanophones, anglophones, francophones ainsi que la Chine, Italie et le Brésil), mais pas seulement. Ce document pose un état des lieux précis et factuels et recense la littérature (académique essentiellement), autour de ces problématiques pour confronter les solutions pensées par le milieu de la recherche et les alternatives élaborées. Les références sont indiquées au fur et à mesure des descriptifs des projets. Les travaux cités se trouvent dans l'annexe.

L'état de l'art témoigne qu'il n'existe pas de **projets aussi ambitieux que ROSETTA**. Mais il existe une importante dynamique et mobilisation des pays occidentaux (Espagne, Angleterre, France et États-Unis, Brésil) au niveau de l'inclusion des personnes en situation de handicap dans les secteurs de l'audiovisuel. Ces changements sont impulsés par les réformes et les lois pour l'inclusion. Cependant, il est important de souligner les différences de services proposées par les voies *streaming* et télévisuelles. Bien que la télévision semble être un modèle plus robuste que celui du streaming, comme nous l'avons vu, la popularité du streaming présage un détournement de la télévision au profit de ces services, ce que semble aussi indiquer le CSA¹⁴⁹.

L'automatisation du sous-titrage s'accélère. La principale raison est la dynamique forte du streaming provoqué par l'hégémonie américaine dans le monde occidental sur le plan des technologies de reconnaissance vocale et de transcription automatique. Les GAFAM disposent d'énormes quantités de données exploitables et desservent de manière dominante l'ensemble des pays anglophones et ceux utilisant les langues européennes. Deux observations s'imposent :

- Leur dominance s'arrête aux frontières avec la Chine qui maîtrise les langues chinoises et ses différents dialectes ainsi que ses rapprochements avec les autres cultures asiatiques.

¹⁴⁹ <https://www.csa.fr/Informer/Collections-du-CSA/Thema-Toutes-les-etudes-realisees-ou-co-realisees-par-le-CSA-sur-des-themes-specifiques/Les-etudes-du-CSA/Videos-en-ligne-ou-television-chez-les-jeunes-publics-etude-econometrique>



- Le sous-titrage dans le monde du streaming ne s'apparente pas aux qualités de sous-titrage télévisuel en France ou dans d'autres pays européens qui imposent également des règles.

Par ailleurs, les techniques de traduction et de sous-titrages automatiques dans le monde télévisuel devront affronter l'hétérogénéité des normes et des règles des différents pays dès qu'une internationalisation est à l'ordre du jour. Seul Netflix semble capable, à l'heure actuelle, de maîtriser un déploiement télévisuel à l'international en respectant les réglementations nationales pour les sous-titres de ses offres. Leur solution consiste à faire travailler des prestataires locaux (humains) pour le sous-titrage.

La production de contenu en LS, et a fortiori la traduction automatique, doit répondre aux défis suivants :

- ⇒ Représenter les LS dans leur intégrité linguistique et pas uniquement comme une séquence de gestes manuels,
- ⇒ Assurer une bonne qualité des animations des avatars qui vont permettre de visualiser le contenu en LS.

Malgré les avancées technologiques et la digitalisation des pratiques, les solutions d'accessibilité audiovisuelles automatiques de qualité ne sont pas encore déployées à l'échelle internationale. Les objectifs de Rosetta aideront les communautés concernées à accéder à des services de qualité, et cela, lorsque les solutions auront dépassé de nombreux obstacles. Au-delà de l'apport de solutions pour les communautés concernées, il convient de réfléchir à la plus-value que représente ROSETTA pour les professionnels du sous-titrage et de la traduction en langue des signes et comment cela modifie et aide leurs métiers respectifs.

Une réflexion similaire s'impose pour la distribution/commercialisation télévisuelle (film, série, documentaire) à l'étranger, car la traduction sous-titrée multilingue peut favoriser des ventes plus rapides. L'aspect social et réglementaire, tout comme des standards adaptés pourraient favoriser l'introduction de ce type de projets d'innovation sociale et leur mise sur le marché.

7.2 Prochaines étapes

Cette photographie des projets, solutions et autres pratiques à travers le monde est très utile pour avoir une vue d'ensemble et pour positionner le projet ROSETTA par rapport aux autres projets et acteurs. Cependant il est nécessaire d'étudier et investiguer de nouveaux marchés pour ce projet, liant l'inclusion, l'innovation sociale et sociétale et la haute performance technologique de l'intelligence artificielle.

C'est donc dans la continuité des recherches menées au niveau des techniques et de l'innovation audiovisuelle que la prochaine étape vise une phase d'entretiens qualitatifs.

L'objectif de cette phase sera de recenser avec précision où en sont les producteurs-diffuseurs en matière d'automatisation du sous-titrage multilingue et de la génération de contenu en langue des signes : quelles sont leurs avancées et leurs besoins ?

L'étude comparative (*benchmarking*) qualitative se poursuivra auprès de producteurs-diffuseurs de l'audiovisuel (TV, Web, festivals), laboratoires (auteurs intégrés), certains auteurs *free-lance* (sous-titres/adaptateurs/traducteurs), entreprises et administrations dont des universités produisant des vidéos (dont des MOOC), des associations d'utilisateurs et d'experts.

Elle permettra de :



- Comprendre les pratiques et besoins pour produire des contenus d'accessibilité audiovisuelle, la nature des contenus concernés, les difficultés et solutions envisagés,
- Se positionner de manière concrète, en fonction des besoins, pour anticiper la mise sur le marché des innovations.



Annexes

Annexe 1 : Lexique des différents sigles

A	
ADA	
American with Disabilities Act	65
ASR	
Automatic Speech Recognition.....	15
B	
BRI	
Belt and Road Initiative	60
C	
CCTV	
China Central Television (le réseau national de télévision chinois).....	60
CRTC	
Conseil de la Radiodiffusion et des Télécommunications Canadiennes	63
CSA	
Conseil Supérieur de l'Audiovisuel	52
CVAA	
Communications and Video Accessibility Act (Loi sur l'accessibilité des communications et des vidéos)	65
D	
DSGS	
Deutschschweizerische Gebärdensprache, langue des signes allemande	39
E	
ESN	
Entreprise de Services Numériques.....	32
F	
FAIR	
Facebook Artificial Intelligence Research	36
FCC	
Federal Communications Commission (Commission fédérale américaine).....	65
I	
IA	
Intelligence Artificielle	33
IRISA	
Institut de Recherche en Informatique et Systèmes Aléatoires	45
M	
mocap	
motion capture ou capture de mouvements.....	46
MT	
Machine Translation	32





N	
NLP	
Natural Language Processing.....	32
O	
OCR	
Optical Character Recognition (reconnaissance optique des caractères)	19
OS	
Operating System ou système d'exploitation	16
P	
POC	
Proof Of Concept (prototype qui fonctionne)	40
R	
RPC	
République Populaire de Chine	60
RTBF	
Radio Télévision Belge Francophone.....	55
RTVE	
RadioTelevisión Española (Radiotélévision espagnole).....	57
S	
SaaS	
Software As a Service (logiciel en tant que service)	27
SiGML	
Signing Gesture Markup Language (permet la transcription des gestes).....	39
SLTAT	
Sign Language and Avatar Technology	14
SME	
Sous-titrage pour sourds et malentendants.....	11
SNCB	
Société Nationale des Chemins de fer belges.....	40
ST	
Sous-titre.s.....	11
STT	
Speech-to-Text.....	23
T	
TALN	
Traitement automatique du langage naturel	25
TNT	
Télévision Terrestre Numérique	54
V	
VOD	
Video On Demand (Vidéo à la demande)	55
W	
WCAG	





Web Content Accessibility Guide (guide d'accessibilité des contenus sur le Web)..... 70



Annexe 2 : Rappel des différents types de sous-titres

Dans un souci de clarté, nous proposons ci-dessous un lexique des expressions utilisées dans le champ général du sous-titrage :

- Il existe deux formats principaux de sous titrages : externe (Closed) et interne (Open). Le premier est indépendant de la vidéo (souvent sous format '.srt' ou '.ass', avec un lecteur associé) tandis que le second est directement intégré à la vidéo.
- De plus, il y a également deux types de traductions au sein même des sous-titres, en lien avec les usages finaux : dans la même langue (intralinguale) et dans une autre langue (extralinguale) : ils peuvent en effet servir de support à l'apprentissage d'une nouvelle langue ou au renforcement de l'acquisition d'une langue (maternelle ou autre).
- Les closed captions (CC) : Texte au bas de l'écran présentant une transcription du dialogue, agrémenté de la description des signaux audio dans une vidéo. Les closed captions sont utiles aux téléspectateurs sourds ou malentendants car ils incluent des bruits de fond, la différenciation des locuteurs et d'autres informations importantes. Ces légendes peuvent être désactivées ou activées.
- Les sous-titres (ST) sont une version traduite, synthétique, de la transcription d'une vidéo, qui s'affichent également sous forme de texte en bas de l'écran. Les sous-titres, support écrit de l'oral, sont généralement destinés aux téléspectateurs, entendants, qui ne parlent pas la langue utilisée dans l'audio de la vidéo. De plus, ces sous titres n'incluent pas d'autres éléments audio, tels que la musique, les effets sonores ou la différenciation des haut-parleurs.
- Le sous-titrage pour les sourds et les malentendants (SDH, ou SME en France) est une forme de combinaison entre CC et les ST classiques ou traditionnels. Plus récents, Le SDH est apparu à l'origine en Amérique dans l'industrie du DVD. Le terme SDH fait généralement référence à des sous-titres réguliers, dans la langue d'origine de la vidéo, qui incluent également des informations importantes sans dialogue, telles que l'identification du locuteur.
- SDH est le type de sous-titre ou de légende le plus complet, puisqu'il comprend aussi les sons non verbaux qui contribuent à l'expérience de visionnement. Par exemple, les sous-titres SDH peuvent noter les bruits ambiants qui aident à créer les paramètres de la vidéo. Les haut-parleurs de la vidéo peuvent réagir aux bruits hors écran ; lorsque les sous-titres SDH le soulignent, les téléspectateurs sourds ou malentendants peuvent mieux comprendre ce qui se passe dans la vidéo. SDH combine à la fois des informations audio et linguistiques pour créer un fichier de sous-titres. De cette façon, le fichier peut être accessible à un public de langue étrangère qui peut également être sourd ou malentendant.





Annexe 3 : Webographie thématisée

— ARTICLES SUR LE NET (HORS CITATIONS BIBLIOGRAPHIQUES)

Sources	Liens
Article sur le net : BBC	https://www.bbc.com/mundo/noticias-46814069
Article sur le net : BFMBusiness	https://bfmbusiness.bfmtv.com/le-tete-a-tete-decideurs/les-enjeux-de-la-retranscription-automatique-et-le-sous-titrage-de-videos-1446604.html
Article sur le net : Caasem, histoire du sous-titrage	https://www.caasem.fr/sous-titrage-adapte/un-peu-d-histoire
Article sur le net : centre national du sous-titrage et de l'audiodescription (en espagnol)	http://www.cesya.es/articulos/201601211256
Article sur le net : China Daily	http://www.chinadaily.com.cn/regional/2019-07/25/content_37497820_8.htm
Article sur le net : China Global Television Network	https://news.cgtn.com/news/3d3d414e3551444f34457a6333566d54/index.html
Article sur le net : Europe 1	https://www.europe1.fr/medias-tele/Les-fansubbers-ou-le-sous-titrage-low-cost-656496
Article sur le net : France Info	https://www.meta-media.fr/2017/02/08/en-2017-lott-nest-plus-une-option-pour-la-tele.html
Article sur le net : France24	https://www.france24.com/fr/20190814-facebook-ecoute-enregistrement-sonore-apple-google-amazon-microsoft
Article sur le net : Handirect	https://www.handirect.fr/storysign-apprentissage-de-la-lecture-aux-enfants-sourds/
Article sur le net : La Revue des médias / INA	https://larevuedesmedias.ina.fr/infographie-le-marche-du-cinema-chinois
Article sur le net : Les Echos	https://weekend.lesechos.fr/business-story/innovation/0600802801462-la-traduction-automatique-fait-des-pas-de-geant-2248830.php
Article sur le net : Newamerica	https://www.newamerica.org/cybersecurity-initiative/digichina/blog/how-ai-can-better-serve-people-with-disabilities-in-china/
Article sur le net : News Guandong	http://www.newsgd.com/news/2019-05/24/content_187604771.htm
Article sur le net : RTVE (en espagnol)	http://www.rtve.es/rtve/20120713/tve-cadena-mas-programas-accesibles-para-personas-discapacidad-dio-2011-mas-30000-horas-subtituladas/546047.shtml
Article sur le net : TVTechnology	https://www.tvtechnology.com/equipment/the-case-for-automating-closed-captioning
Article sur le net : Venturebeat	https://venturebeat.com/2019/05/10/how-android-qs-live-caption-works-select-phones/





Article sur le net : Venturebeat	https://venturebeat.com/2019/05/10/how-android-qs-live-caption-works-select-phones/
Article sur le net : Vie Publique	https://www.vie-publique.fr/parole-dexpert/268423-la-chine-puissance-hegemonique-en-asie
Article sur le net : Xinhuanet	http://french.xinhuanet.com/2019-07/25/c_138256906.htm
Article sur le net : Zdnet	https://www.zdnet.fr/blogs/digital-home-revolution/le-streaming-toujours-plus-fort-face-a-la-tv-traditionnelle-39862002.htm
Article sur le net : Zdnet	https://www.zdnet.fr/i/edit/ne/2017/12/PwC003.png
Communiqué de presse Handicap.gov	https://handicap.gouv.fr/presse/communiques-de-presse/article/la-directive-europeenne-d-accessibilite-des-biens-et-des-services

— DIFFÉRENTES VIDÉOS CITÉES ET ACCESSIBLES SUR LE NET

Sources	Liens
Vidéo de présentation du projet Kinect	https://msutoday.msu.edu/news/2019/new-technology-breaks-through-sign-language-barriers/
Vidéo de présentation du robot ASIMO	https://www.slashgear.com/honda-asimo-learns-sign-language-and-masters-stairs-18325765/
Vidéo sur Vimeo	https://vimeo.com/302898833
Vidéo sur YouTube	https://www.youtube.com/watch?v=UjtOGPJOURM
Vidéo sur YouTube	https://www.youtube.com/watch?v=hPv1PkjJ-J0

— DIVERS

Sources	Liens
Information sur le design pour tous	https://www.bloghoptoys.fr/cest-quoi-design-pour-tous
Informations sur les différentes technologies pour l'éducation (site US)	https://snow.idrc.ocadu.ca/assistive-technology-2/augmentative-processing/animated-signing-characters-signing-avatars/
Livre blanc association des sourds chinoise	http://www.cdpc.org.cn/english/Resources/documents/201907/t20190726_658048.shtml
Livre blanc Omniscien	https://omniscien.com/more/resources/white-paper/
Page du cours MOOC signé	https://www.fun-mooc.fr/courses/course-v1:univ-toulouse+101020+session01/about?fbclid=IwAR2j7lapGMkxumGd8kdIj6PRZ8eWIE_aG90FxoVP6SSiYnfM8-KksTIHM9M
Page Google Store	https://play.google.com/store/apps/details?id=com.google.audio.hearing.visualization.accessibility.scribe&hl=fr
Page présentation Alliance Capa	http://www.capa.ac/ai





Prix Unesco Tencent	https://fr.unesco.org/news/prix-unescoemir-jaber-al-ahmad-al-ahmad-al-jaber-al-sabah-lautonomisation-personnes-handicapees
---------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

— LES GUIDES DE SOUS-TITRAGES (PAYS)

Sources	Liens
Guide sous titrages anglais britannique	http://www.ofcom.org.uk/static/archive/itc/itc_publications/codes_guidance/sign_language_dtt/index.asp.html
Guide sous-titrage américain	https://www.ada.gov/cguide.htm
Guide sous-titrage belge francophone	https://www.rtf.be/entreprise/contact-et-question/faq/detail_accessibilite?id=9309319
Guide sous-titrage canadien (anglophone)	https://www.cab-acr.ca/french/societal/captioning/normes_universelles.pdf
Guide sous-titrage espagnol	http://www.rtve.es/alacarta/videos/en-lengua-de-signos/lengua-signos-07-07-19/5323440/

— PAGES ACCUEIL SITES DIVERS

Sources	Liens
Page d'accueil Amberscript	https://www.amberscript.com/fr/
Page d'accueil Association EFHOH	https://efhoh.org
Page d'accueil Authot	https://www.authot.com/fr/
Page d'accueil CPAC (Cable Public Affairs Channels)	https://www.cpac.ca/fr/programs/covid-19-la-reponse-du-canada/episodes/66189568/
Page d'accueil DALET	https://www.dalet.com/fr/platforms/dalet-amberfin
Page d'accueil Flincons Group	https://www.finconsgroup.com
Page d'accueil Fleex	https://fleex.tv/fr/
Page d'accueil FUN MOOC	https://www.fun-mooc.fr
Page d'accueil Huawei (projet Storysign)	https://consumer.huawei.com/fr/campaign/storysign/
Page d'accueil JATJOURNAL	http://www.jatjournal.org/index.php/jat
Page d'accueil KEIA	https://www.keia.io/solutions/
Page d'accueil L'Oréal	https://www.loreal.com
Page d'accueil Labo DAMO d'Alibaba	https://damo.alibaba.com/labs/language-technology/
Page d'accueil Académie LSF	http://www.languedessignes.fr
Page d'accueil MIPTV	https://www.miptv.com





Page d'accueil NetFlix	https://www.netflix.com/fr/
Page d'accueil OFCOM	https://www.ofcom.org.uk/home
Page d'accueil Otter.ai	https://otter.ai/login
Page d'accueil RogerVoice	https://rogervoice.com/fr/
Page d'accueil Scribie	https://scribie.com
Page d'accueil Simax	https://simax.media/?lang=en
Page d'accueil SLTAT	http://sltat.cs.depaul.edu
Page d'accueil Speechmatics	https://www.speechmatics.com
Page d'accueil Temi	https://www.temi.com
Page d'accueil Tencent ASR	https://cloud.tencent.com/product/asr
Page d'accueil Trint	https://trint.com
Page d'accueil Université Gallaudet	https://www.gallaudet.edu/about
Page d'accueil Verbit.ai	https://verbit.ai
Page d'accueil Vocapia	https://www.vocapia.com/

— PAGES EN LIEN AVEC LES LÉGISLATIONS SELON LES DIFFÉRENTS PAYS

Sources	Liens
Page législation Australie	https://www.acma.gov.au/captions-tv-broadcasters-must-provide
Page législation Belgique	http://www.ffsb.be/audiovisuel/nv-reglement-access/
Page législation Brésil	https://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2015/Lei/L13146.htm
Page législation Chine (associatif)	http://www.cdpc.org.cn/english/Resources/lawsregulations/201707/t20170730_601596.shtml
Page législation Espagne	https://www.boe.es/buscar/doc.php?id=BOE-A-2013-12632 .
Page législation Espagne	(art.8)
Page législation Europe	https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0636:FIN:en:PDF
Page législation Europe	https://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=URISERV:l24101a&from=FR
Page législation Europe	https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A32019L0882

— PAGES INFORMATION WIKIPEDIA

Sources	Liens
Page Wikipédia groupes ethniques en Chine	https://fr.wikipedia.org/wiki/GROUPES_ETHNIQUES_DE_CHINE





Page Wikipédia ANDROID	https://fr.wikipedia.org/wiki/Android
Page Wikipédia BBC Player	https://fr.wikipedia.org/wiki/BBC_iPlayer

— PAGE DE PRÉSENTATION DES DIFFÉRENTS PROJETS PUBLICS PRIVÉS

Sources	Liens
Page présentation projet : Avatar Signeur Irisa	http://lsf.irisa.fr/archives.html
Page présentation projet : Handtalk	https://www.handtalk.me/br
Page présentation projet : InMoov	http://inmoov.fr/project/
Page présentation projet : Limix	http://www.limix.it/en/
Page présentation projet : Momo the Robot	https://awards.ixda.org/entry/2019/momo-the-robot-meets-children-with-autism/
Page présentation projet : Signall	https://www.signall.us/product_info/
Page présentation projet : Signlang	https://pub.cl.uzh.ch/demo/signlang/geosign/en/about.html
Page présentation projet TraMOOC	https://www.k4all.org/project/tramooc/
Page présentation projet Visicast	http://www.visicast.cmp.uea.ac.uk/Visicast_index.html
Page présentation projet Visicast	http://www.visicast.cmp.uea.ac.uk/eSIGN/index.html
Page présentation recherche : IBM	https://www.ibm.com/able/media-captioner-and-editor.html
Page présentation SiSi	http://mqtt.org/projects/sisi
Page projet Cordis AID	https://cordis.europa.eu/project/id/872139
Page projet Cordis CONTENT4ALL	http://content4all-project.eu
Page projet Cordis EASYTV	https://easytvproject.eu
Page projet Cordis ELITR	https://elitr.eu/the-project/
Page projet Cordis IMAC	https://www.imac-project.eu
Page projet Cordis TRAMOOC	https://cordis.europa.eu/project/id/644333/reporting
Page projet Cordis UMAQ	https://cordis.europa.eu/project/id/752659
Page projet UKRI ExTOL	https://gtr.ukri.org/projects?ref=EP%2FR03298X%2F1

— PAGES DE SUPPORTS TECHNIQUES

Sources	Liens
Page support technique : Movistarplus	http://www.movistar.es/particulares/movistarplus/guia-rapida/5s-plataforma-contenidos-accesibles/





Page support technique YouTube	https://support.google.com/youtube/answer/6373554?hl=fr
Page technique Netflix	https://partnerhelp.netflixstudios.com/hc/en-us/categories/202282037-SPECIFICATIONS-GUIDES
Page technique SIGML	http://vh.cmp.uea.ac.uk/index.php/SiGML
Page technique ANDROID	https://www.android.com/what-is-android/
Page technique ANDROID	https://www.android.com/phones-tablets/
Page technique ANDROID	https://developer.android.com/wear
Page technique ANDROID	https://developer.android.com/tv
Page technique ANDROID	https://www.google.com/chromebook/chrome-os/
Page technique ANDROID	https://developer.android.com/cars
Page technique ANDROID	https://developer.android.com/things
Page technique ANDROID	https://developer.android.com/ndk
Page technique ANDROID	https://www.android.com/android-10/
Page technique ANDROID	https://www.android.com/android-10/
Page technique ANDROID	https://www.youtube.com/watch?v=YL-8Xfx6S5o
Page technique ANDROID	https://www.android.com/accessibility/live-transcribe/
Page technique ANDROID	https://www.android.com/what-is-android/
Page technique ANDROID	https://www.android.com/accessibility/live-transcribe/

— PRÉSENTATION DE DIFFÉRENTS RAPPORTS (FRANCE ET ÉTRANGER)

Sources	Liens
Présentation de résultats IFOP	https://www.ifop.com/wp-content/uploads/2019/02/115960-Présentation-version-publiée.pdf
Rapport CSA : études sur la consommation audiovisuelle des jeunes	https://www.csa.fr/Informer/Collections-du-CSA/Thema-Toutes-les-etudes-realisees-ou-co-realisees-par-le-CSA-sur-des-themes-specifiques/Les-etudes-du-CSA/Videos-en-ligne-ou-television-chez-les-jeunes-publics-etude-econometrique
Rapport CSA : normes sous-titrage	https://www.csa.fr/Proteger/Garantie-des-droits-et-libertes/Les-droits-des-personnes-handicapees/Le-sous-titrage
Rapport CSA : Rapport Annuel 2018	https://www.csa.fr/Informer/Informations-publiques-et-ressources-humaines/Les-rapports-annuels-et-bilans-du-CSA/Le-rapport-annuel-2018-du-CSA
Rapport de l'EFHOH (2011)	http://media.wix.com/ugd/c2e099_098aaba4b05ee8f7fed8b0add0b8c332.pdf
Rapport de l'EFHOH (2015)	http://media.wix.com/ugd/c2e099_0921564404524507bed2ff3648781a3c.pdf



Rapport statistique Espagne	https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176782&menu=resultados&secc=1254736194716&idp=1254735573175#ltabs-1254736194716
Rapport Statistique gouvernement Chinois	http://www.stats.gov.cn/english/Statisticaldata/CensusData/rkpc2010/html/A0106a.htm
Rapport Statistique handicaps en Chine	https://www.dinf.ne.jp/doc/english/asia/resource/z00ap/003/z00ap00308.html

Références

Adamao-Villani, N., & Anasingaraju, S. (2016). Toward the Ideal Signing Avatar. *EAI Endorsed TRansactions on e-Learning(ICST.ORG)*.

Ali, R., & Akmeliawati, R. (2017). Speech to Text Translation for Malay Language. *Conference Series: Materials Science and Engineering*.

Ali, R., Bakr, N., Fauzi, N., & Akmeliawati, R. (2019). SignBot, Sign-Language Performing Robot, Based on Sequential Motion of Servo Motor Arrays. Dans J.-H. e. Kim, *Robot Intelligence Technology and Application 5*. Springer.

Angerbauer, K., Adel, H., & Vu, T. (2019). Automatic Compression of Subtitles with neural Networks and its effect on user experience. *Interspeech 2019*, pp. 594-598.

Axelsson, M. (2018, 8). Designing an InMoov Robot to Teach assistive Sign Language to Children with Autism. *Thesis*. Helsinki.

Baillet, C. (2013). Culture sourde et culture de consommation. La nécessité d'une approche pluridisciplinaire. *Management & Avenir*, pp. 135-152.

Benaida, M., Taleb, A., & Namoun, A. (2018). The link between automated translation of the arabic language and quality of websites. *International Journal of Computer Science and Network Security*, 18 (9).

Berke, L., Caufield, C., & Huernerfauth, M. (2017). Deaf and Hard-Of-Hearing Perspectives on Imperfect Automatic Speech Recognition for Captioning One-on-One Meetings. *ASSETS'17*. Baltimore, MD.

Bragg, D., et al. (2019). Sign Language Recognition, Generation, and Translation : an interdisciplinary Perspective. *ASSETS 2019*. ACM.

Che, D. (2013). *La dimension publique et la télévision en Chine : les exemples de CCTV et de Phénix TV*. Strasbourg: Université de Strasbourg.

Chung, Y., Weng, W., Tong, S., & Glass, J. (2019). Towards Unsupervised Speech-to-text Translation. *ICASSP*.

CSA . (2019). *L'accessibilité des programmes de télévision aux personnes handicapées et la représentations du handicap à l'antenne*.

Cómitre Narvaez, I., & Sedano Ruiz, E. (2019). Sous-titrage créatif pour enfants sourds et malentendants : les contes au cinéma. *Palimpseste*, pp. 32.

David, B., & Bouillon, P. (2018). Traduction automatique de la parole vers la langue des signes de Belgique francophone. Evaluation d'un avatar destiné aux transports en commun par la communauté





sourde. Dans G. Bourghis, & Slimane, M. (Éd.), *10ème conférence de l'IFRATH sur les technologies d'assistance. Recherches pluridisciplinaires pour l'autonomie des personnes en situation de handicap*. Paris: sn.

De Martino, J., Silva, I., & Bolognini, C. (2017). Signing avatars : making education more inclusive. *Universal Access in The Information Society*, 16, pp. 793-808.

Delaporte, Y. (2005, novembre). La variation régionale en langue des signes française. *Marges Linguistiques : Langues régionales*(10), pp. 118-132.

Delorme, M. (2011). Modélisation du squelette pour la génération réaliste de postures de la LSF.

Denisov, P., & Vu, T. (2019). IMS-Speech: A Speech to Text Tool. *PrePrint*.

Dupont de Dinechin, G., & Paljic, A. (2019). Automatic Generation of Interactive 3D Characters and Scenes for Virtual Reality from a Single-Viewpoint 360-Degree Video. *IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. Osaka, Japan.

Ebling, S., & Glauert, J. (2016). Building a swiss German Sign Language Avatar with JASigning and evaluating it among the Deaf Community. *Universal access informatic society*, 15, pp. 577-587.

Filhol, M. (2008). Modèle descriptif des signes pour un traitement automatique des LS.

Firmansyah, M., Paul, A., Bhattacharya, D., & Urfa, G. (2020). A.I. based embedded Speech to Text Using DeepSpeech. *preprint*.

Gerber-Morón, O., & Szarkowska, A. (2018). Line Breaks in subtitling : an Eye tracking study on viewer preferences. *Journal of the eye Movement Research*, 11(3).

Gerber-Morón, O., Szarkowska, A., & Woll, B. (2018). The Impact of text segmentation on subtitle reading: eye tracking study on cognitive load and reading performance. *Journal of Eye Movement research*, 11(4).

Ghadage, Y., & Shelke, S. (2016). Speech to Text conversion for multilingual languages.

Gilani, S., Traum, D., Sortino, R., Gallagher, G., Aaron-Lozano, K., Padilla, C., . . . Petitto, L. (2019). Can a Signing Virtual Human engage a Baby's Attention ? *IVA'19*.

Greco, G., & Jankowska, A. (2019). Framing media accessibility quality. *Journal of Audiovisual Translation*, 2 (2) , pp. 1-10.

Gupta, P., Sharma, M., Pitale, K., & Kumar, K. (2019). Problems with automating translation of movie/TV show subtitles. (*preprint*).

Hersh, M., & Johnson, M. (2003). *Assistive technology for the hearing impaired, deaf and deaf-blind*. Berlin: Springer.

Hersh, M., & Johson, M. (2003). *Assistive technology for the hearing impaired, deaf and deaf-blind*. Berlin: Springer.

Homburg, D., Thieme, M., Völker, J., & Stock, R. (2019). RoboTalk - Prototyping a Humanoid Robot as Speech-to-Sign Language Translator. *HICSS*.

Jaffe, D. (1994). Evolution of mechanical fingerspelling hands, for people who are deaf-blind. *Journal of rehabilitation*, 31, 236-244.

Jenning, E., Kennaway, R., & Glauert, J. (2010). Requirements For A Signing Avatar.

Jin, H., & Gambier, Y. (2018). Audiovisual Translation in China : A dialogue between Yves Gambier and Jin Haina. *Journal of Audiovisual Translation*, pp. 26-39.



Kervajan, L. (2011). *Contribution à la traduction automatique Français/LSF au moyen de personnages virtuels*.

Kipp, M., Héloir, A., & Nguyen, Q. (2011). *Sign Language Avatars : Animation and Comprehensibility. IVA*.

Laffont, L. et al. (2019). *Education inclusive à la santé environnementale pour la transition écologique : MOOC "TEAM", Transferts Environnementaux des contaminants Métalliques*. Toulouse: université Paul Sabatier Toulouse 3, Toulouse INP, Toulouse Jean Jaurès.

Lakhfif, A., & Laskri, M. (2010). *Un Signeur Virtuel 3D pour la Traduction Automatique de Textes Arabes vers la Langue des Signes Algérienne*. Université d'Annaba.

Larboulette, C., & Gibet, S. (2018). *Avatars signeurs: que peuvent-ils nous apprendre. J.Enaction 2018 : Journée sur l'énaction en animation, simulation et réalité virtuelle*. Poitiers.

Le Corre, G. (2007). *La langue des signes française (LSF)*. *Enfance*, pp. 228-236.

Lefebvre-Albaret. (2010). *Traitement automatique de vidéos en LSF : modélisation et exploitation des contraintes phonologiques du mouvement*.

Losson, O. (2000). *Modélisation du geste communicatif et réalisation d'un signeur virtuel de phrases en LSF*.

Lu, Y., & Li, H. (2019). *Automatic Lip-Reading System Based on Deep Convolutional Neural Network and Attention-Based Long Short-Term Memory*. *Applied Sciences*, 9 (8), p. 1599.

Marsella, S. et al. (2013). *Virtual character performance from speech*. *Proceedings of the 12th ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (pp. 25-35). ACM.

Mehta, N., Pai, S., & Singh, S. (2019, Juillet). *Automated 3D sign Language caption generation for video*. *Universal Access in the Information Society*.

Millet, A. (2004). *La langue des signes française (LSF) : une langue iconique et spatiale méconnue*. *Cahiers de l'APLIUT*, XXIII (2).

Pedersen, J. (2018). *From Old Tricks to Netflix : How local are interlingual subtitling norms for streamed television*. *Journal of Translation*, 1(1), pp. 81-100.

Prado, D. (2010). *La traduction automatisée : le cas des langues romanes*. *Hermès, La Revue*, pp. 95-102.

Punchimudiyanse, M., & Meegama, R. (2015). *3D Signing Avatar for Sinhala Sign Language*. *IEEE International Conference on Industrial and Information System*.

Rovira-Esteva, S., & Casas-Tost, H. (2017). *In search for standards for subtitling in Chinese*. *Sino-foreign Audiovisual Translation & dubbing cooperation Workshop*. Shanghai: Xina.

Scassellati, B. et al. (2018). *Teaching Language to Deaf Infants with a Robot and a Virtual Human*. *CHI'18*. Montréal.

Segouat. (2010). *Modélisation de la coarticulation en LSF pour la diffusion automatique d'informations en gare ferroviaire à l'aide d'un signeur virtuel*.

Séro-Guillaume, P. (2008). *Le métier d'interprète en Langue des Signes*. *Connaissances surdités*.

Szarkowska, A., & Gerber-Morón, O. (2018). *Two or Three Lines : a mixed-methods study on subtitle processing and preferences*. *Perspectives. Studies in Translation Theory and Practice*.





Szarkowska, A., & Gerber-Morón, O. (2018). Viewers can keep up with fast subtitles : evidence from eye movements. *PLoS One*.

Szarkowska, A., Krejtz, I., Pilipczuk, O., Dutka, L., & Kruger, J. (2016). The effects of text editing and subtitle presentation rate on the comprehension and reading patterns of interlingual and intralingual subtitles among deaf, hard of hearing and hearing viewers. *Across Languages and Cultures*, 17 (2), pp. 183-204.

Tamayo, A., & Chaume, F. (2017). Subtitling for d/deaf and Hard-of-Hearing Children : Current practices and New Possibilities ti Enhance Language development. *Brain Sciences*, 7 (7), p. 75.

Volk, M., Senrich, R., Hardmeier, C., & Tidström, F. (2010). Machine Translation of TV Subtitles for large Scale Production. Dans V. Zhechev (Éd.), *Bringing MT to the user : Research on Integrating MT in the Translation Industry (JEC'10)*, (pp. 53-62). Denver, CO.

Wang, D., Zhang, X., & Kuo, A.-Y. (2020). Researchin inter-asian audiovisual translation. *Perspectives*.

Williams, H., & Thorne, D. (2000). The value of teletext subtitling as a medium for language learning. *System*, 28 (2), pp. 217-228.

Wolton, D. (2010). Avant-propos : La traduction, passeport pour accéder à l'autre. Hermès, La Revue, p. 56.

Zirzow, N. (2015). Signing avatars : using virtual reality to support students with hearing loss. *Rural Special Education*, 34(3), pp. 33-36.

