

SOMMAIRE

TABLE DES ILLUSTRATIONS	13
INTRODUCTION.....	15
CHAPITRE 1.....	17
TRADUCTOLOGIE, LANGUE DES SIGNES : CONTEXTES ET PROBLEMATIQUES.....	17
1.1 LA LANGUE DES SIGNES FRANÇAISE (LSF)	18
1.1.1 <i>Histoire d'une langue.....</i>	18
Origine et développement	18
100 ans de césure	20
Le Réveil Sourd	21
1.1.2 <i>Législation.....</i>	21
Statut.....	22
Accessibilité.....	22
1.1.3 <i>Description linguistique</i>	23
Simultanéité	24
Iconicité.....	25
Spatialité	28
Ordre propre	29
Pas de forme écrite	30
1.2 TRADUCTOLOGIE	31
1.2.1 <i>Traduire</i>	31
Interprétation ou traduction ?	32
Traduction (processus) et traduction (produit).....	33
Théorie de la traduction	33
1.2.2 <i>Brève rétrospective : du parchemin à l'ordinateur</i>	35
Traducteur, un des plus vieux métiers du monde	35
TAO : qu'est-ce que c'est ?	36
Impacts sur le métier	38
1.2.3 <i>Traductologie des LS.....</i>	39
Contexte	39
Déverbalisation	40
Nécessité des connaissances du monde	41
1.2.4 <i>De la théorie à la pratique: la traduction français-LSF</i>	42
Besoin grandissant pour du contenu accessible	42
Peu de traducteurs fr-LSF.....	43
Des outils mais pas pour les LS.....	43
Comment outiller la traduction des LS ?	44
CHAPITRE 2.....	47
ÉTAT DE L'ART CONCERNANT LA TAO ET L'EVALUATION DE LA TRADUCTION.	47

2.1 TRADUCTION ASSISTEE PAR ORDINATEUR	48
2.1.1 <i>Fondamentaux</i>	48
Forme écrite éditable	49
Linéarité	49
Capitaliser sur les travaux précédents	50
Intégration de modules (pour tout avoir sous la main).....	52
2.1.2 <i>Et pour les LS ?</i>	52
Les outils existants	53
Pourquoi la TAO telle qu'elle existe ne peut s'appliquer aux LS	56
Forme écrite éditable	56
Linéarité	57
Mémoire de traduction	58
2.2 ÉVALUER LA TRADUCTION ET LES OUTILS	59
Entre objectivité et subjectivité	59
2.1.1 <i>Critères pour évaluer une traduction</i>	60
Cohérence de la forme.....	60
Respect du message et de l'intention	60
Évaluation de la traduction automatique.....	62
2.1.2 <i>Comment évaluer les outils d'aide à la traduction ?</i>	64
Évaluation de la TAO	64
Impact de la TAO sur la profession.....	66
Synthèse du chapitre.....	68
CHAPITRE 3.....	71
PROPOSITIONS DE SPECIFICATIONS LOGICIELLES.....	71
3.1 COMPRENDRE LE METIER DE TRADUCTEUR : EXPERIENCES PRELIMINAIRES	72
3.1.1 <i>Traducteurs du texte vers la LSF à l'œuvre</i>	72
Méthodologie.....	73
Résultats.....	76
3.1.2 <i>Brainstorming</i>	83
Méthodologie.....	84
Résultats.....	86
3.1.3 <i>Le processus de traduction</i>	89
Synthèse des résultats	89
Discussion et limites.....	92
3.2 CAHIER DES CHARGES ET PROPOSITIONS.	93
3.2.1 <i>Non-linéarité assistée par des « Blocs »</i>	94
Segmentation libre : la notion de Bloc	94
Contenu des Blocs	95
Organisation temporelle : l'axe vertical	96
Hiérarchie : la structuration des Blocs	96
Flexibilité : la mobilité des Blocs	97
3.2.2 <i>Autres fonctionnalités</i>	98
Gestion du texte source	98
Modules de ressources	98
Prompteur	99
Mémoire de traduction.....	99

CHAPITRE 4.....	101
REALISATION DES PROTOTYPES	101
4.1 CONCORDANCIER BILINGUE FRANÇAIS - LSF.....	102
<i>4.1.1 Base de données d'alignements</i>	<i>102</i>
Corpus utilisé.....	102
Anatomie d'un alignement.....	103
Annotation du texte avec Brat	105
Annotation des vidéos avec Elan.....	106
Création d'une fonction d'alignements.....	107
<i>4.1.2 Implémentation</i>	<i>109</i>
Interface	110
Fonction de recherche : mode d'emploi	111
<i>4.1.3 Exemples d'usage</i>	<i>112</i>
Explorer la langue.....	112
Mot sans signe standard correspondant, mais exemples de traduction similaires entre eux	113
Pas de signe standard correspondant, mais exemples différents	114
Relation de cause à effet.....	115
4.2 L'ENVIRONNEMENT DE TRAVAIL INTEGRE	116
<i>4.2.1 Zone d'édition</i>	<i>117</i>
Création et mouvement des Blocs	117
Charger du contenu dans un Bloc	118
Opération sur les contenus	120
Menu et options.....	121
<i>4.2.2 Modules et outils</i>	<i>121</i>
Carte	123
Historique.....	125
<i>4.2.3 Gestion du texte source</i>	<i>126</i>
Texte et formatage.....	126
Traitement automatique du texte.....	126
<i>4.2.4 Navigation dans l'arbre</i>	<i>127</i>
CHAPITRE 5.....	129
ÉVALUATION PRELIMINAIRE DES OUTILS.....	129
5.1 ÉVALUATIONS	130
<i>5.1.1 Concordancier.....</i>	<i>130</i>
Présentation lors de l'atelier TAO	131
Démonstrations ciblées.....	132
<i>5.1.2 Environnement de travail intégré</i>	<i>133</i>
Méthodologie.....	133
Résultats partiels.....	141
5.2 RETOUR SUR LES RESULTATS PRELIMINAIRES	144
<i>5.2.1 Analyse des réponses.....</i>	<i>144</i>
Concordancier	144
Environnement de travail intégré	144

<i>5.2.2 Modifications et améliorations du protocole</i>	145
Améliorations suggérées pour le prototype par les utilisateurs tests.....	145
Améliorations possibles pour le protocole.....	146
CHAPITRE 6.....	147
CONCLUSION ET PERSPECTIVES	147
6.1 RETOUR SUR LES OBJECTIFS	148
6.2 PRINCIPALES CONTRIBUTIONS ET RESULTATS.....	149
6.3 PERSPECTIVES	151
BIBLIOGRAPHIE.....	155
ANNEXES.....	161
ANNEXE A.....	161
LISTE DES LOGICIELS DE TAO ETUDES AU COURS DES TRAVAUX DE RECHERCHE.....	161
ANNEXE B.....	163
TEXTES D'ELICITATION POUR LA CONSTITUTION DU CORPUS VIDEO DE TRADUCTEURS.	163
<i>Textes session de rodage</i>	163
<i>Textes courts</i>	163
<i>Textes longs</i>	166
<i>(L1)</i>	166
ANNEXE C.....	169
DONNEES EXPORTEES D'ELAN	169
ANNEXE D.....	173
TEXTES PROPOSES LORS D'ATELIERS	173
TEXTES PROPOSES A LA TRADUCTION DANS LE CADRE DU PROTOCOLE D'EVALUATION.....	174
ANNEXE E.....	177
QUESTIONNAIRE D'EVALUATION	177

TABLE DES ILLUSTRATIONS

FIGURE 1 UN OISEAU SURPRIS, CORPUS LS-COLIN	25
FIGURE 2 LE SIGNE PLUIE EN LSF, ET SON ORIGINE MOTIVEE (SOURCE : L. PRONGUE)	26
FIGURE 3 TRANSFERT DE LA TAILLE D'UNE MESURE D'HUILE, CORPUS LS-COLIN	27
FIGURE 4 TRANSFERT DE SITUATION : LE CHEVAL SAUTE UNE BARRIERE, CORPUS LS-COLIN	27
FIGURE 5 TRANSFERTS PERSONNELS : LE SIEUR PREND LE ROLE DU CHEVAL, CORPUS LS-COLIN.....	28
FIGURE 6 L'ESPACE DE SIGNATION (SOURCE : L. VERLAINE)	29
FIGURE 7 EXEMPLES D'ECRITS AVEC SIGNWRITING	30
FIGURE 8 EXEMPLE DE REPRESENTATION AVEC HAMNOSYS.....	30
FIGURE 9 LE TRIANGLE INTERPRETATIF SELON D. SELESKOVITCH (1979).....	34
FIGURE 10 L'INTERFACE D'UN LOGICIEL DE TAO (SDL TRADOS)	37
FIGURE 11 SCHEMA DE DEVERBALISATION.....	40
FIGURE 12 EXEMPLE DE PRECISION DU CONTEXTE (TEXTE SOURCE AU DESSUS ; TRADUCTION EN DESSOUS)..	42
FIGURE 13 LE TEXTE SOURCE DECOUPE (A GAUCHE), ET LES SEGMENTS CIBLES CORRESPONDANTS QUE LE TRADUCTEUR DOIT REMPLIR (A DROITE). TIRE DU LOGICIEL SDL TRADOS STUDIO 2017.....	48
FIGURE 14 LE PRINCIPE DE LINEARITE	50
FIGURE 15 LES AVATARS JADE (EN HAUT), ET PAULA (EN BAS)	54
FIGURE 16 SCHEMA DE LA DIFFUSION EN FONCTION DE LA PRECISION, GIRAUD 1968.....	61
FIGURE 17A UN BINOME A L'OEUVRE.....	74
FIGURE 18 TABLEAU DES TACHES ET DUREES.....	78
FIGURE 19 TABLEAU DES TACHES POUR LES TEXTES LONGS (BINOME 2).....	82
FIGURE 20 LISTE DES QUESTIONS POSEES DURANT LE BRAINSTORMING	85
FIGURE 21 LE TABLEAU DE BRAINSTORMING	86
FIGURE 22 SCHEMA EXPLICATIF DE LA HIERARCHIE DES BLOCS. L'EXEMPLE COMPORTE DEUX HIERARCHIES DE BLOCS (AXE VERTICAL), CHACUN AYANT TROIS NIVEAUX DE HIERARCHIE (AXE HORIZONTAL).	97
FIGURE 23 CAPTURE D'UNE VIDEO DU CORPUS	103
FIGURE 24 CAPTURE DE LA SEGMENTATION DANS BRAT (EN HAUT), ET EXEMPLE DE DONNEES EXPORTEES DEPUIS BRAT (EN BAS).	105
FIGURE 25A ANNOTATION AVEC ELAN (CORPUS 40-BREVES).....	106
FIGURE 26 EXEMPLE D'AFFICHAGE DE RESULTATS POUR LA REQUETE SUR LES TERMES « GLISSEMENT DE TERRAIN »	110
FIGURE 27 TROIS VERSIONS DE « FIN DE NON RECEVOIR »	113
FIGURE 28 TROIS VERSIONS DE « MIS A MAL »	114
FIGURE 29 TROIS VERSIONS DE « UN TSUNAMI PROVOQUE PAR »	115
FIGURE 30 : DEUX VERSIONS DE « UN GLISSEMENT DE TERRAIN PROVOQUE PAR DES PLUIES ABONDANTES »	116
FIGURE 31 INTERFACE PRINCIPALE DE L'ENVIRONNEMENT DE TRAVAIL INTEGRÉ	117
FIGURE 32 : MARQUEURS LORS DU DEPLACEMENT DES BLOCS : LA LIGNE ROUGE PERMET DE VISUALISER L'EMPLACEMENT OU SERA DEPOSE LE BLOC EN COURS DE DEPLACEMENT.	118
FIGURE 33 À PARTIR DU BLOC, IL EST POSSIBLE D'INSERER DIFFERENTS TYPES DE CONTENU : TEXTE, IMAGE, OU VIDEO.....	119
FIGURE 34 MENU DE LA ZONE A BLOCS	121

FIGURE 35 LES RESSOURCES INTEGREGES SONT ACCESSIBLES A DROITE DE L'INTERFACE.....	123
FIGURE 36 LE MODULE DE CARTE ET SES FONCTIONS : EPINGLES, CAPTURE D'ECRAN, GALERIE.....	124
FIGURE 37 INSERTION D'UNE CARTE DANS UN BLOC A PARTIR DE L'OUTIL CARTE	125
FIGURE 38 L'HISTORIQUE DE RECHERCHE INTERACTIF	126
FIGURE 39 LE MENU DE NAVIGATION	127
FIGURE 40 VERSION FRANÇAISE ET VALIDEE DU SUS, PROPOSEE PAR GRONIER, G. & BAUDET, A. (2021)	136
FIGURE 41 SIGNIFICATION DU SCORE AU SUS, PROPOSEE PAR BANGOR, KORTUM ET MILLER (2009)	138
FIGURE 42 GRILLE D'EVALUATION DES TRADUCTIONS	140
FIGURE 43 RESUME DES RESULTATS AUX QUESTIONNAIRES (VERSION ALPHA).....	142
FIGURE 44 RESUME DES RESULTATS PARTIELS QUANT A L'INFLUENCE DE L'UTILISATION DU PROTOTYPE (NOMBRE DE PARTICIPANTS PAR REPONSE)	143

Introduction

Cette thèse est le fruit d'une recherche à la croisée de l'informatique et de la traductologie. Elle explore tout particulièrement la traductologie des langues signées, avec une volonté de chercher à mieux formaliser le métier de traducteur en langues des signes. En effet, avec la publication de plusieurs lois à l'échelle européenne et mondiale concernant le droit des personnes en situations de handicap, la demande pour du contenu accessible en LS, et donc traduit, est croissante.

Cette thèse s'intéresse donc à la description et à l'étude du processus de traduction vers les langues signées, plus particulièrement de la LSF qui est la langue de travail utilisée durant les expérimentations. Elle s'intéresse d'autre part à la traduction assistée par ordinateur des langues vocales, dans le but de l'adapter aux LS, et ainsi de proposer une assistance informatique adaptée spécialement au processus de traduction des LS. Elle aborde également la question de l'évaluation de la traduction, afin de pouvoir jauger et juger de l'efficacité des outils à disposition.

Le présent mémoire s'articule en six chapitres. Le chapitre 1 présente le contexte des langues signées, ainsi que les problématiques liées. Faisant suite à une analyse du contexte initial de nos travaux, il en présente également les questions de recherche. Le chapitre 2 dresse un état de l'art des sujets qui intéressent notre recherche, à savoir la traduction assistée par ordinateur ainsi que l'évaluation de la traduction. Le chapitre 3 aborde les premières spécifications logicielles. Pour ce faire, il présente en premier lieu deux expériences ainsi que leurs résultats, qui permettent de compléter l'état de l'art précédemment dressé. Le chapitre 4 décrit la réalisation de deux prototypes d'outils de traduction assistée par ordinateur : un concordancier

français-LSF, ainsi qu'un environnement de travail intégré. Le chapitre 5 propose une évaluation préliminaire des outils ainsi développés, ainsi que les pistes d'amélioration qui en découlent. Enfin, le chapitre 6 conclue sur l'intégralité des travaux présentés, ainsi que sur les perspectives scientifiques en lien avec ce travail de thèse.

Le manuscrit fait parfois mention d'exemples sous forme vidéo, rassemblés dans une annexe en ligne, triée par chapitre, disponible à cette adresse :

<https://sandbox.zenodo.org/record/1080591>

Chapitre 1

Traductologie, langue des signes : contextes et problématiques

Ce chapitre a pour but d'exposer les notions clés sur lesquelles reposent les travaux rapportés dans cette thèse. La langue des signes française dans un premier temps (section 1.1), puisqu'elle sera notre langue de travail principale avec le français. La LSF est une langue — et non un langage, pas si récente que ce que les préjugés laissent à penser, qui n'est pas non plus universelle, et qui dispose de sa grammaire propre. La traductologie ensuite (section 1.2), qu'est-ce que traduire ? Comment traduire ? Et plus avant encore, comment traduire vers les langues signées ? Ce chapitre conclut sur une synthèse quant à la pratique concrète et à la situation actuelle de la traduction vers les langues des signes en France, pour soulever la question de la valorisation et du développement de ladite pratique.

1.1 La Langue des Signes Française (LSF)

Cette section a pour objectif de présenter la langue des signes française au travers d'éléments historiques, linguistiques et sociolinguistiques. L'enjeu de cette présentation est d'introduire l'objet principal des travaux, en tant que langue propre, et d'en exposer les particularités en tant que langue visuo-gestuelle. Elle se compose d'un résumé historique du développement des LS, suivi d'un point plus détaillé quant à sa situation actuelle, et enfin de plusieurs notions linguistiques essentielles pour mieux se représenter la langue.

1.1.1 Histoire d'une langue

L'évolution des LS, dont la reconnaissance en tant que langue n'est par ailleurs pas identique selon les pays, alterne entre périodes d'interdiction et périodes plus fastes. Nous nous proposons ici de revenir sur les évènements les plus remarquables, et ayant marqué un tournant dans l'histoire des LS.

Origine et développement

S'il est difficile de pointer et dater avec précision les débuts des LS, il est néanmoins possible de retracer ponctuellement des états de la langue au travers la représentation des personnes sourdes au fil des époques.

Pour ce qui est de l'Antiquité, c'est au travers des réflexions d'Aristote quant au «logos» que l'on dégage une première tendance. Selon le philosophe, la parole est le seul moyen d'accéder à la pensée, et par extension à la logique. De ce fait, les personnes ayant recours à la modalité visuo-gestuelle pour s'exprimer ne sauraient y accéder, et se voient de fait incapable de raisonner — et ce bien qu'Aristote observe et reconnaît leur usage de la langue des signes (Aristote, IV siècle av. J-C, cité par Saint-Loup, 1989). Les sourds, selon les réflexions d'Aristote, sont donc relégués dans la catégorie des animaux (à savoir celle des êtres qui ne sollicitent que

la vue, en opposition aux hommes qui eux sollicitent également l'ouïe et la parole).

Le Moyen-Âge marque une première évolution notable dans la considération des personnes sourdes et donc sur les regards portés sur leur usage des LS. L'influence de l'Église catholique sur la société a favorisé l'intégration de plusieurs minorités, y compris celle des personnes sourdes (Saint-Loup, 1989). On trouve mention d'ouvriers ou d'artisans sourds, puisque contrairement aux autres formes de handicaps comme la cécité ou d'autres pathologies physiques, ils peuvent apprendre en observant. C'est également à cette époque que l'on trouve la première trace en ce qui concerne la France, d'une interprétation (par des personnes qui ne sont pas des professionnels rémunérés, à l'époque) entre français et langue des signes (Cantin, 2021). Les sourds semblent à l'époque mener une vie identique à celle des entendants de leur rang. Par exemple, la princesse Jeanne Stuart dite aussi la « Muta Domina » (1428-1493), fille sourde de Jacques Ier d'Écosse, fait un bon mariage et apporte descendance et honneur à son royaume.

Si la situation des sourds semble s'améliorer au fil du temps, ce n'est véritablement qu'à la Renaissance que les LS connaissent leur essor : elles ont maintenant leur place dans l'éducation et l'enseignement. La méthode est alors oraliste : les enfants sourds doivent parler, on les entraîne donc à la langue vocale. Ironiquement, cet entraînement repose sur l'utilisation de signes (code gestuel, des combinaisons de gestes manuels, de la tête et d'expressions du visage). C'est l'engagement de Charles-Michel Lespée (1712-1789), dit Abbé de l'Épée, qui porte la LS au rang de langue d'éducation pour les sourds. Pour ce faire, il complète les signes d'usage en France par une série de signes qu'il appelle « méthodiques », dont le but est de représenter des éléments de la grammaire du français, tels que le genre

ou le nombre, les déterminants... La méthode de l'abbé se démocratise en Europe durant le 18^e siècle.

C'est Roche-Ambroise Auguste Bébian (1789-1839) qui reprend le flambeau, en repensant une méthode sans signes méthodiques cette fois. Il favorise l'éducation bilingue, basée à la fois sur l'usage des signes naturels de la LSF pour l'oral, et sur l'apprentissage du français écrit. On le retrouve notamment à la tête de l'Institution Nationale des Jeunes Sourds de Paris de 1819 à 1821. Il était également le premier à s'atteler à la description de la LSF dans le but d'élaborer un dictionnaire LSF/français.

100 ans de césure

S'il n'était qu'un seul événement à garder en mémoire quant à l'histoire des Sourds et la pratique des LS, il s'agirait sans doute du Congrès de Milan. C'est effectivement à Milan, en 1880, que se déroule le troisième congrès international pour l'amélioration du sort des sourds. Nombre de spécialistes européens (entendants) de la surdité et de l'éducation des jeunes sourds se réunissent dans le but de réviser et uniformiser le système éducatif des enfants sourds. L'assemblée ne compte que trois personnes sourdes, et aucun des débats menés n'a été interprété en LS malgré les interprétations prévues en anglais, français, espagnol et autres langues vocales.

Le congrès est organisé par des défenseurs de la méthode oraliste, qui consiste à éduquer les enfants sourds à la parole et à délaisser les LS. Il débouche sur l'interdiction en Europe de l'usage des langues des signes dans l'éducation des enfants sourds, favorisant de fait le recours à la méthode purement oraliste. Les arguments en faveur de cette décision sont aussi nombreux que variés. On citera "une facilité d'intégration à la société" pour les sourds oralisant, ou encore que, sur l'aspect religieux ou médical, l'oralité permet non seulement « d'élever l'esprit là où les signes ne transmettent que les passions », mais aussi de « donner la vie » aux sourds.

Citons l'Abbé Tarra, qui préside alors le Congrès:

« Le sourd-muet [...] n'a en quelque sorte qu'une respiration incomplète [...] en les habituant à la respiration provoquée par l'émission de la parole, ce n'est donc pas seulement la parole qu'on leur donne, c'est la vie. ».

On ne voit donc la personne sourde que comme une personne malade qu'il faut guérir.

Les motions adoptées pendant ce congrès sont lourdes de conséquences pour le siècle à venir, et plus encore puisque leurs répercussions peuvent encore s'observer aujourd'hui.

Le Réveil Sourd

Faisant suite à l'interdiction de la LSF dans les écoles françaises pendant presque 100 ans, méprisée et marginalisée, c'est dans les années 1970 que se produit en France “le Réveil Sourd”. La rencontre de Jean Grémion, écrivain, journaliste et metteur en scène français, avec un artiste Sourd américain, Alfredo Corrado, donne naissance en 1976 à l'IVT : l'International Visual Theatre, à Paris. La scène permet de requalifier la LSF, et d'en redorer l'image auprès du public si bien qu'en 1977, la LSF est de nouveau « autorisée ». Avec le théâtre comme vecteur principal, l'IVT encourage également les travaux linguistiques sur les LS. Les associations culturelles et militantes Sourdes se multiplient, tout comme les actions pour que les sourds retrouvent une place légitime dans la société. C'est le début des cours de LSF proposés aux entendants, notamment, et les jalons posés d'une culture Sourde renouvelée.

1.1.2 Législation

L'impact du Réveil Sourd conduit à la parution en 1991 d'une première loi quant à la reconnaissance de la LSF et son importance dans l'éducation. Un amendement à la loi du 8 Janvier 1991 permet alors aux parents d'enfants

sourds de choisir entre une communication dite bilingue (LSF et français écrit) et une communication oraliste. S'en suivront d'autres lois concernant l'évolution du statut juridique de la LSF en France, présentées ci-après.

Statut

La loi du 11 février 2005¹, pour l'égalité des droits et des chances, la participation et la citoyenneté des personnes handicapées, reconnaît la LSF comme une langue à part entière:

« Art. L. 312-9-1. -La langue des signes française est reconnue comme une langue à part entière. Tout élève concerné doit pouvoir recevoir un enseignement de la langue des signes française. Le Conseil supérieur de l'éducation veille à favoriser son enseignement. Il est tenu régulièrement informé des conditions de son évaluation. Elle peut être choisie comme épreuve optionnelle aux examens et concours, y compris ceux de la formation professionnelle. Sa diffusion dans l'administration est facilitée. »

Elle n'est cependant pas reconnue comme langue officielle de la République au même titre que le français. L'accessibilité pour les personnes sourdes demeure elle aussi encadrée par d'autres textes de loi.

Accessibilité

Cette même loi du 11 février 2005, renforcée par la Convention relative aux droits des personnes handicapées adoptée par les Nations Unies en 2008, stipule également que chaque citoyen doit pouvoir jouir d'un accès total à l'information du domaine public (administrations et services, lieux publics, transports...).

Elle fait également obligation “aux chaînes de télévision publiques et aux chaînes privées dont l'audience nationale dépasse 2,5 % de l'audience totale

¹ <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000809647/>

des services de télévision de rendre accessibles aux personnes sourdes ou malentendantes la totalité de leurs émissions, en dehors des messages publicitaires et de quelques programmes dérogatoires.”

Elle ne détermine cependant pas les méthodes d’accessibilité à employer, et beaucoup de chaînes concernées optent plus facilement pour du sous-titrage que pour la LSF. D’après le site de la CSA, il existe actuellement un journal d’informations quotidien accessible en LSF et trois journaux sous-titrés sur chacune des trois chaînes d’information de la TNT, du lundi au vendredi. Les week-ends et jours fériés, seuls les journaux sous-titrés demeurent.

Notons cependant que, d’après le rapport Gillot² de 1997, 70% de la population sourde française serait illétrée³ et donc peu à même de tirer profit du sous-titrage proposé. Accessibilité en demi-teinte donc, bien que les services publics et les administrations aient de plus en plus facilement recours à des services d’interprétation pour l’accueil des citoyens signant : ceux qui s’expriment en langue des signes.

Cette faible maîtrise du français observée dans la communauté sourde se justifie sans doute par le fait que les LS sont des systèmes linguistiques qui diffèrent beaucoup du français. La section suivante expose plusieurs caractéristiques linguistiques des langues des signes, notamment celles qui la distinguent des langues audio-vocales.

1.1.3 Description linguistique

Les langues des signes sont un ensemble de langues orales (par opposition à écrites), mais gestuelles, non vocales, et naturelles. C’est-à-dire qu’elles ont naturellement émergé au sein de plusieurs communautés géographiques, par le besoin de communication, et sont donc fortement liées à la culture de ces

² <https://www.vie-publique.fr/sites/default/files/rapport/pdf/984001595.pdf>

³ Peut-on cependant parler d’illettrisme lorsqu’il est question de langue seconde ? (Perini 2008)

mêmes communautés. Si l'on s'imagine souvent “la langue des signes” comme une langue universelle, il n'en est rien : on retrouve certes des universaux grammaticaux dans les langues des signes du monde, mais chacune dispose d'un signaire (le vocabulaire, mais constitué de signes) qui lui est propre.

Les sections ci-après s'intéressent chacune à un point de différence majeur entre les langues vocales et les langues des signes. Les langues de référence utilisées pour fournir des exemples sont les deux langues considérées dans cette thèse, à savoir le français (oral ou écrit en fonction de l'exemple), et la LSF.

Simultanéité

Les langues des signes transmettent du sens grâce à des mouvements des mains, des bras et du buste, mais aussi de la tête, au travers du regard et de l'expression faciale du locuteur. Ces articulateurs pouvant être sollicités simultanément, le locuteur signant est en mesure d'exprimer dans le même temps plusieurs éléments linguistiques là où le locuteur de langue vocale est contraint par la chaîne parlée à ne produire qu'un élément à la fois (si l'on ne tient pas compte de la gestualité coverbale). De plus, les articulateurs de la langue des signes sont visibles, contrairement à ceux des langues des vocales. Les mains, la tête, le visage et le corps sont visibles lorsque le locuteur s'exprime en langue des signes, mais lorsqu'une personne s'exprime par le biais d'une langue vocale, l'interlocuteur ne voit qu'une partie des articulateurs, à savoir les lèvres. Il ne voit ni la langue, ni le voile du palais, ni les cordes vocales. L'apparition simultanée de plusieurs éléments linguistiques est discriminante : l'orientation du regard, ou l'expression du visage du signeur peut modifier le sens du signe. La figure 1 est une capture d'un discours en LSF. La figure 1 présente plusieurs éléments linguistiques : la main gauche d'abord, qui représente le signe

« oiseau » ; le signeur ici conte l'histoire d'un oiseau qui est surpris par les agissements d'un cheval. Son expression faciale seule suffit à signifier que l'oiseau est surpris : les deux éléments sont produits simultanément.

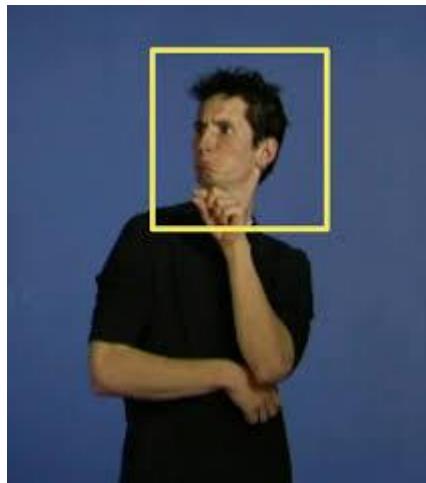


Figure 1 Un oiseau surpris, corpus LS-Colin⁴

Iconicité

La LSF est une langue qui donne à voir, et l'iconicité y a une place centrale. Par iconique, il s'agit de souligner l'aspect motivé du signe, c'est à dire que le signifiant (le signe et sa forme) ressemblent directement au signifié (le concept qu'il représente). Cette ressemblance peut être visuelle, ou reprendre le fonctionnement d'un objet ou d'une action. Par exemple, le signe “pluie”, illustré dans la figure 2.

⁴ Cuxac, Christian, & Boutet, Dominique. (2014). *Corpus LS-Colin sur plusieurs genres discursifs (Laurent Vallo, Nasredine Chab et Anthony Guyon)* (Version 1) <https://doi.org/10.34847/COCOON.6D4B96B9-83C2-38B4-8FD6-ACB62E57DE4D>



Figure 2 Le signe pluie en LSF, et son origine motivée (source : L. Prongué)

On remarque qu'ici, ce sont la forme des doigts et le mouvement qui évoquent les gouttes de pluie qui tombent du ciel.

Cuxac (1996, 2000) oppose la visée illustrative du discours, c'est à dire un discours qui n'a pas ou peu recours aux signes standards (figés dans la langue), aux énoncés non-illustratifs qui eux font usage du signaire standard (peu ou moins iconique), de gestes déictiques tels que les pointages, et de lignes temporelles pour raconter un évènement. La visée illustrative repose sur ce qu'il appelle les structures de grande iconicité : le locuteur ne se contente plus de rapporter des faits, il leur donne vie dans l'espace, voire les incarne. Il existe trois types de structure de grande iconicité, que nous allons détailler ci-après de la moins illustrative à la plus illustrative selon Cuxac:

- Transfert de taille et de forme (TTF) : le locuteur représente un objet du discours en transférant dans l'espace sa taille ou sa forme (ou les deux). Chaque élément constitutif de l'objet est décrit à l'aide de configurations manuelles évoquant sa forme générale, ou alors par le biais de tracés dans l'espace représentant ses contours (Fig. 3).



Figure 3 Transfert de la taille d'une mesure d'huile, corpus LS-Colin

- **Transfert situationnel (TS):** le locuteur représente une situation qui implique le déplacement d'un élément du discours par rapport à un autre, ou dans une scène déjà décrite (Fig. 4). De manière générale, la main dominante joue le rôle de l'élément en mouvement (l'acteur), et l'autre main celui d'un référent stable (le locatif). Chaque main représente son élément avec un proforme : une configuration spécifique selon l'entité représentée. Par exemple, les personnes debout se représentent par un index pointé vers le haut, un bicycle ou un obstacle par une main sur la tranche, une voiture par une main à plat...



Figure 4 Transfert de situation : le cheval saute une barrière, corpus LS-Colin

- **Transfert personnel (TP)** : le locuteur incarne directement l'actant dont il veut parler. Il se met dans la peau de la personne ou de l'élément en question, pour le faire agir ou subir les actions. Le signeur rejoue la scène en y prenant lui-même un rôle (Fig.5), qu'on remarque d'ailleurs par un décrochage du regard sur l'interlocuteur, et une implication du corps plus prononcée que dans tous les autres types de discours. Ce genre de transfert permet de conter, raconter, en somme donner à voir presque tout ce que l'imagination du signeur peut faire naître, du plus concret au plus abstrait.



Figure 5 Transferts personnels : le signeur prend le rôle du cheval, corpus LS-Colin

L'iconicité joue également un rôle important dans la nature spatiale des langues des signes, notamment lorsqu'il est question de transmettre une position géographique.

Spatialité

Les informations exprimées simultanément sont organisées dans l'espace. Le signeur organise son discours dans son espace de signation (Fig.6), c'est à dire l'espace qui l'entoure, verticalement, horizontalement, et dans la profondeur. En ancrant ainsi un élément dans l'espace, il crée une référence partagée avec l'interlocuteur, qu'il peut réactiver à tout moment du discours

par un simple pointage ou un regard, sans avoir à le décrire de nouveau.

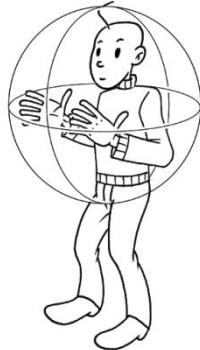


Figure 6 L'espace de signation (source : L. Verlaine)

La spatialisation du discours s'observe également lorsqu'il est question de relation topologique entre des éléments. À gauche, à droite, dessus... La LSF projette la relation dans l'espace. Il en va de même pour tout ce qui relève de l'information géographique. En ce sens, la LSF est d'une part contrainte par la réalité physique, et repose d'autre part sur une prépondérance du raisonnement diagrammatique, en lien avec la pensée visuelle.

Ordre propre

En ce qui concerne la LSF, la nature visuelle et spatiale de la langue induit une tendance à poser le décor dans un premier temps, pour ensuite y faire s'y dérouler l'action. L'ordre de l'énoncé en LSF est généralement le suivant:

Quand - Où - Objet/Acteur - Objet/Patient - Action

La logique systématique consiste à partir du plus général au plus précis. Cette tendance s'observe à l'échelle du discours, qui présente dans un

premier temps des éléments de contexte avant d'énoncer les faits principaux.

Pas de forme écrite

Il n'existe pas d'écriture des langues des signes. En tous cas, pas de forme écrite éditable, comme celles que l'on retrouve pour le français écrit, le russe ou le japonais pour citer des langues ayant recourt à des graphies différentes. Ces systèmes organisés de symboles écrits formalisés permettent de conserver une trace de la langue dans le temps. Ces formes sont éditables, et sont un produit culturel, émergées au sein de diverses sociétés par besoin, et enseignées pour perpétuer la transmission du savoir.

Il existe bien quelques systèmes graphiques proposés pour décrire la langue (HamNoSys et SignWriting par exemple, Fig. 7). Cependant, ceux-ci ne sont pas utilisés par la communauté Sourde au quotidien à grande échelle, et ne peuvent ainsi pas être qualifié de forme écrite dans la dimension culturelle de l'écriture.

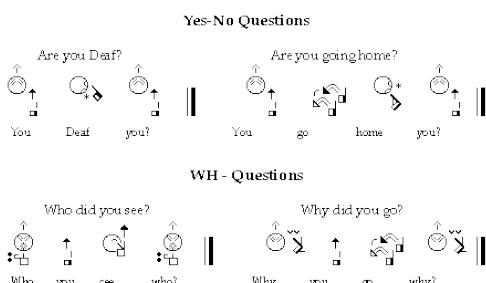


Figure 7 Exemples d'écrits avec Signwriting

Notation du signe	Traduction américaine	Représentation figurative
↓ ↗ ↘ ↑	Where did you go ?	

Figure 8 Exemple de représentation avec HamNoSys

Actuellement, la méthode la plus courante pour garder une trace dans le temps d'un énoncé en LSF demeure la capture vidéo. On ne peut guère la considérer comme l'écriture des LS, puisqu'il ne s'agit pas là d'un système graphique à proprement parlé, mais de l'immortalisation telle quelle d'une production orale. De plus, les vidéos ainsi créées ne sont pas requêtables, et difficilement éditables et indexables pour une consultation ultérieure facilitée au delà du cercle privé.

Comme toutes les langues, les LS peuvent être traduites, c'est-à-dire être la source ou le produit d'une traduction. S'intéresser à la traduction des langues signées revient d'une part à s'intéresser aux langues en question, et d'autre part à s'intéresser à la traduction en elle-même. La section suivante traite ainsi de la traductologie, à savoir la science de la traduction.

1.2 Traductologie

Cette section introduit les principes fondamentaux de l'activité traductive, pour en définir les processus et les enjeux. Émergeant dans le courant du XXe siècle, la traductologie trouve plusieurs définitions à la portée plus ou moins linguistique. Elle est, selon Reiss, « *l'étude scientifique de la traduction en tant que processus et en tant que produit* » (Reiss, 2009, p.1).

Les deux premières parties de cette section s'intéressent à la traduction au sens général, ainsi qu'aux outils qui la concernent. La troisième partie traite quant à elle de la traductologie des langues des signes, c'est-à-dire les spécificités de ce cas de traduction. Enfin, la quatrième et dernière partie marie les deux thèmes précédents, et pose la question de l'outillage informatique de la traduction des langues des signes.

1.2.1 Traduire

Avant de s'intéresser à la traduction des langues des signes, il convient dans un premier temps de statuer sur ce qu'est la traduction en général. Nous

prenons le temps ici de définir ce que nous entendrons par traduction tout au long de ce mémoire. Nous distinguons dans un premier temps la traduction et l'interprétation, puis la traduction en tant que processus ou produit.

Interprétation ou traduction ?

Interprétation et traduction partagent le même objectif, à savoir passer un message d'une langue vers une autre. À cela s'ajoutent plusieurs contraintes d'ordres linguistiques ou culturels, de style ou de format : traduire implique non seulement de transposer le sens du message mais également l'intention du locuteur, et nécessite donc la prise en compte du contexte à l'instant où est produit le discours.

En ce qui concerne la différence entre interprétation et traduction, on cite souvent la différence de modalité en premier lieu. Effectivement, l'interprétation fait référence au passage d'une langue orale à une autre langue orale, là où le terme traduction est utilisé pour désigner le passage d'une langue écrite vers une autre langue écrite. Une autre façon d'appréhender la différence entre les deux concerne la chronologie de la tâche.

L'interprétation est immédiate, dite d'ailleurs *simultanée* : l'interprète traduit les propos au fur et à mesure que le locuteur les produit. Elle est donc contrainte par l'avancée en temps réel du discours. On se focalise sur le sens parfois au détriment de la forme, faute de temps. Une autre pratique de l'interprétation, dite *consécutive*, consiste à prendre note du discours pendant une dizaine de minutes (parfois plus), et à le restituer ensuite en différé sur un laps de temps plus court. Dans les deux cas, le professionnel ne dispose que de très peu, sinon pas du tout de temps pour penser sa traduction : il est contraint par l'immédiateté de la situation, lors de sommets politiques par exemple. De plus, l'interprétation est produite en un seul essai, sans possibilité de modifications futures.

La traduction s'opère en revanche, en différé de l'instant où est produit le discours. Le traducteur dispose d'un délai pour produire, et donc de la possibilité de retravailler sa traduction. Il peut également mener des recherches, et adapter au mieux tant le sens que la portée du message, ou l'intention donnée par le locuteur. S'il existe un recouvrement dans les processus cognitifs impliqués, dans le cadre de la traduction, le professionnel peut travailler la forme de son résultat, mais également de procéder à des corrections si nécessaires (voire d'échanger avec son client pour jauger desdites modifications). Les travaux rapportés dans cette thèse traitent de la traduction, et non de l'interprétation. La participation d'interprètes professionnels dans les études présentées ci-après se justifie du fait de leurs compétences et pratiques traductives en sus de leurs pratiques interprétatives.

Traduction (processus) et traduction (produit)

Nous différencions également la **Traduction** en tant que tâche, c'est-à-dire l'ensemble des processus qui mènent à la traduction, de la traduction en tant que produit, c'est-à-dire le résultat cible. Les travaux rapportés dans cette thèse s'intéressent à l'une comme à l'autre, bien que dans notre cas la traduction-produit ou la source ne soient pas un texte mais une vidéo. La **Traduction** quant à elle, pour éviter toute confusion, sera présentée en tant que « processus de traduction ».

Théorie de la traduction

À compter de la seconde moitié du XXe siècle, la traduction s'est dotée notamment grâce aux travaux linguistiques, de plusieurs théories qui permettent de dépasser la vision bivalente de la question (qui jusque-là était de privilégier soit la langue source, soit la langue cible). Sans entrer dans les détails, il est important de citer l'ouvrage de Lederer et Seleskovitch (1985), *Interpréter pour traduire*, qui pose les jalons des techniques d'interprétation

et de traduction qui sont enseignées. La Théorie Interprétative de la Traduction, également appelée Théorie du Sens, s'oppose à l'idée alors généralisée que la traduction ne se résume qu'au transcodage d'une langue à une autre. Cette théorie présente plutôt la Traduction comme un processus en trois temps, comme représenté Fig. 9 : d'abord comprendre le message depuis la langue source (compréhension), puis en extraire le sens en s'affranchissant de la forme (déverbalisation), et enfin le reformuler et le produire dans la langue cible (réexpression). Bien que cette théorie ait d'abord été développée sur des phénomènes observés durant l'interprétation (ce que les auteurs appellent la traduction orale), elle a plus tard été démontrée comme valable également pour la traduction.

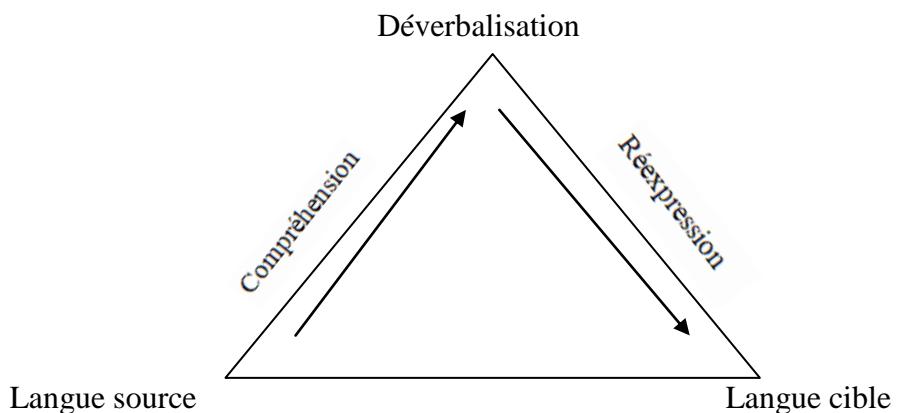


Figure 9 Le triangle interprétatif selon D. Seleskovitch (1979)

La Traduction ne concerne pas seulement le sens, mais également la forme de l'énoncé et l'intention de l'auteur ou du locuteur. En cela, ces données doivent également être considérées dans les phases de compréhension et de réexpression.

Au fil du temps, la traduction s'est modernisée, et l'informatique n'y est pas étrangère. La section suivante présente les bases de la Traduction Assistée par Ordinateur (TAO), mais également l'impact que de pareils outils ont pu

avoir sur le métier de traducteur.

1.2.2 Brève rétrospective : du parchemin à l'ordinateur

Il s'agit ici de considérer le métier de traducteur au fil des siècles. Quels étaient le rôle et la place du traducteur ? Comment travaillait-il ?

Traducteur, un des plus vieux métiers du monde

... ou du moins aussi vieux que les premières sociétés, dès lors qu'elles ont été confrontées les unes aux autres. Si l'approche scientifique de la traduction n'est que très récente, l'approche littéraire, elle, remonte à l'Antiquité. La première trace de traduction remontent à trois siècles avant notre ère : la Septante, une traduction de la bible hébraïque vers le grec koiné. Les textes religieux sont probablement les premières traductions dont nous avons encore trace. La Rome Antique s'intéresse à la traduction d'ouvrages littéraires pour en faciliter la diffusion dans tout l'Empire. Cicéron souligne d'ailleurs dans son *De optimo genere oratorum* (46 av. J.-C.) qu'il ne faut pas traduire *verbum pro verbo*, mot à mot. Au IVème siècle, c'est Saint Jérôme (qui restera d'ailleurs le patron des traducteurs) qui s'attèle à traduire d'anciens textes bibliques de l'hébreu vers le latin cette fois, et qui propose un des premiers guides du traducteur, *De optimo genere interpretandi* (392-395 ap. J-C).

Au VIIIe siècle, avec l'essor de l'Empire arabe, Bagdad sa capitale se dote d'un centre de traduction. On y traduit des ouvrages scientifiques datant de la Grèce antique. Plus tard à Tolède, au XIe siècle, naît la première école de traduction recensée.

Au départ, les traducteurs n'étaient certainement que des copistes spécialisés dans le passage d'une langue à l'autre. Le travail se faisait à la main, ouvrage par ouvrage, une copie à la fois. Avec l'imprimerie d'abord, puis les traitements de textes informatiques, les outils se sont succédés à

mesure du progrès technologique. Aujourd’hui, la traduction se retrouve dotée d’un panel de logiciels dédiés. La section suivante en présente les principes clés.

TAO : qu'est-ce que c'est ?

Dans un premier temps, il faut différencier traduction assistée par ordinateur et traduction automatique, bien que les deux ne soient pas incompatibles. Ici, il est question d’assister le traducteur humain dans sa tâche. Ce dernier demeure expert et décisionnaire dans le processus, alors que dans le cadre de la traduction automatique, le logiciel génère la traduction sans intervention humaine.

On ne commence à parler de TAO qu’à partir des années 1960, à la suite du rapport ALPAC⁵ qui met fin au financement public pour la recherche sur la traduction automatique aux USA. Ce rapport encourage pourtant le développement d’une aide à la traduction humaine, sur la base des glossaires développés quelques années auparavant pour servir de base à la traduction automatique. On cherche cette fois à faciliter le processus de traduction : on transpose les ressources papier en ressources informatiques dont la consultation devient plus rapide, et on pose des questions d’ergonomie, en rassemblant notamment ces outils en une seule et même application.

Les premiers logiciels de TAO consistent en des programmes de gestion terminologique (Hutchins, 2005) pour peu à peu évoluer vers des banques terminologiques multilingues. On développe également les premiers concordanciers bilingues. Les concordanciers sont des logiciels qui

⁵The Automatic Language Processing Advisory Committee, constitué de sept scientifiques avec à leur tête J.R. Pierce, fut créé en 1964 par le gouvernement des USA afin d’évaluer les progrès de la linguistique informatique, et en particulier de la traduction automatique. Son rapport, publié en 1966, est célèbre pour s’être montré très sceptique quant à la recherche sur la traduction automatique, ce qui a conduit le gouvernement à suspendre les fonds de financement pour cette même recherche.

permettent la recherche d'un mot ou d'une séquence de mots dans un corpus afin de consulter son utilisation en contexte. Les concordanciers bilingues permettent de retrouver, pour une paire de langues données, des exemples traductions en contextes.

C'est dans les années 70 que la TAO connaît son véritable essor avec l'apparition des premières mémoires de traduction (Sommers, 2005). Ces logiciels permettent de capitaliser sur le travail passé, en gardant en mémoire des paires de phrases précédemment traduites. Lorsque le traducteur rencontre une phrase déjà traduite (ou très proche) dans la banque de données du logiciel, ce dernier lui suggère la traduction associée. À l'heure actuelle, les logiciels de TAO sont des environnements de travail intégrés (Fig. 10) qui regroupent plusieurs de ces fonctionnalités en sus de prendre en charge des aspects plus classiques de la traduction (gestion du texte source, mise en forme) et de faciliter l'organisation des travaux au sein d'un même service.

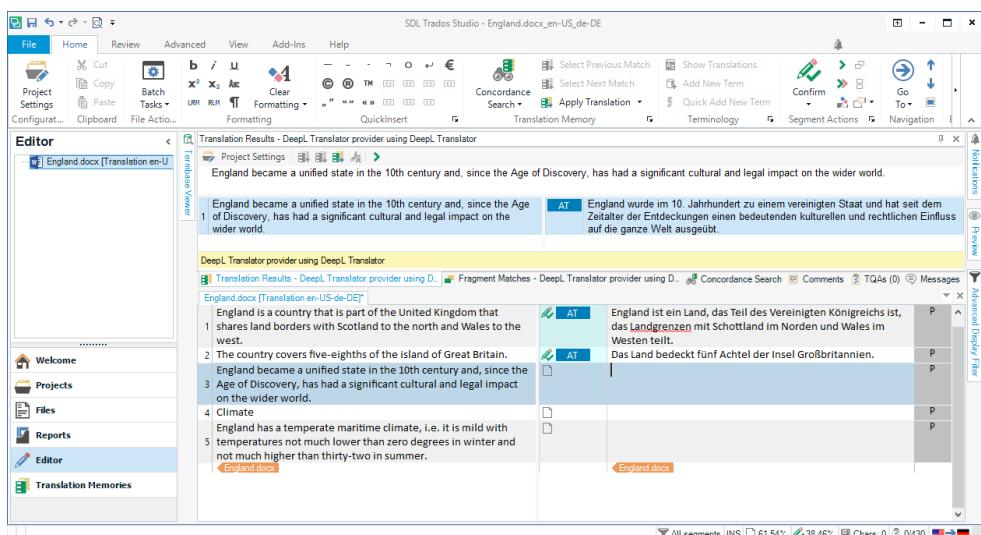


Figure 10 L'interface d'un logiciel de TAO (SDL Trados⁶)

⁶ <https://www.trados.com/fr/>

Impacts sur le métier

L'objectif premier de la TAO est de faciliter le travail du traducteur. Le gain espéré se mesure d'abord en temps : avec tous les outils nécessaires rassemblés au même endroit, mais également via l'automatisation de certaines tâches. Si l'objectif n'est pas de remplacer le traducteur humain, les versions les plus récentes de ces logiciels permettent de réduire considérablement le temps de traduction, notamment de par la redondance des textes traduits (mise à jour de notices ou de catalogues, par exemple). La transition du dictionnaire papier vers le tout TAO ne s'est pas faite du jour au lendemain, et il a d'abord fallu convaincre les professionnels concernés du véritable apport de tels outils — et passer outre cette crainte d'être remplacé par la machine. Bien que l'humain reste aux commandes, l'efficacité des mémoires de traduction d'aujourd'hui est telle que la tâche de traduction s'amoindrit au profit d'une tâche de post-édition : contrôler et éventuellement corriger les suggestions faites par le logiciel. Certains clients optent même pour des versions traduites à moindre coût : les traducteurs reçoivent alors des fichiers dont la majorité du contenu est verrouillé et ne sera ni à traduire, ni à contrôler, mais géré par la traduction automatique. Ces portions de contenu verrouillées sont des correspondances à 100% avec des éléments déjà présents dans la mémoire de traduction, et se retrouvent notamment lorsque le projet de traduction n'est qu'une révision d'un document déjà précédemment traduit : le traducteur ne s'en occupe que lorsque la traduction automatique est en difficulté.

De fait, traductions littéraires et poétiques mises à part, la demande est croissante et les délais de traduction de plus en plus courts. Il a fallu compter une cinquantaine d'années pour que la TAO modifie le métier.

Autre domaine de notre intérêt, la section suivante traite de la traductologie des langues des signes, plus récente encore.

1.2.3 Traductologie des LS

Discipline récente, les premières études académiques concernant l'interprétation et la traduction des langues des signes font suite aux premières études linguistiques. La traductologie des langues des signes se caractérise notamment par son approche plus sociologique que la traductologie générale. On s'intéresse non seulement au métier de la traduction mais également à la place qu'occupent le traducteur et l'interprète LS au sein de la communauté Sourde, ainsi qu'à la déontologie de ces métiers.

Contexte

Nous l'avons vu, la LSF n'est reconnue que depuis peu en France. Cela a un impact sur les métiers qui s'y rapportent. On note notamment l'absence de leur mention dans le code du travail. C'est en ce sens que les premières études s'intéressent à l'aspect sociologique du métier et de la traduction, afin de contribuer à en améliorer la reconnaissance.

L'interprétation demeure plus courante que la traduction en LSF, pour une question d'accessibilité au quotidien. C'est la nécessité d'un travail dit *de liaison* (interprétation de rendez-vous, d'évènements brefs de la vie quotidienne) qui a permis au métier d'interprète de se faire plus vite connaître que le métier de traducteur, et qui justifie peut-être qu'une majorité des études de traductologie des langues des signes concernent l'interprétation plus que la traduction.

Nous avons également souligné précédemment l'absence de forme écrite. Difficile alors de parler de traduction si l'on apprête celle-ci à l'écrit. La traductologie des LS s'intéresse donc à une pratique de la traduction dite orale, où la notion d'écrit est remplacée par la notion d'un discours figé, assimilable à une production orale dont on garderait une trace pérenne.

Déverbalisation

Si la déverbalisation n'est pas propre aux langues des signes puisqu'elle constitue la seconde phase du processus d'interprétation d'après Lederer et Seleskovitch (Fig. 9), elle prend tout son sens ici. Les langues des signes reposent sur la pensée visuelle, aussi enseigne-t-on le recours à la schématisation pour faciliter l'étape de déverbalisation au sein des formations d'interprètes et de traducteurs. Le principe consiste en la reformulation de l'intégralité d'un discours par le biais du dessin ou du schéma, en limitant le recours à la langue écrite. L'objectif est de permettre aux interprètes et aux traducteurs de se décoller du français, de s'émanciper de la forme écrite au profit d'une compréhension plus fine du sens de sorte à en faciliter la reformulation dans la langue cible. (Guitteny, 2007)

La figure 11 propose un exemple de déverbalisation, réalisée par un étudiant interprète en première année de Master, et basée sur l'énoncé « *La Bretagne a connu dimanche sa première grande tempête hivernale, qui a entraîné au large des avaries de bateaux, la disparition d'un marin ainsi que la chute dans la mer de conteneurs transportés par cargos.* »

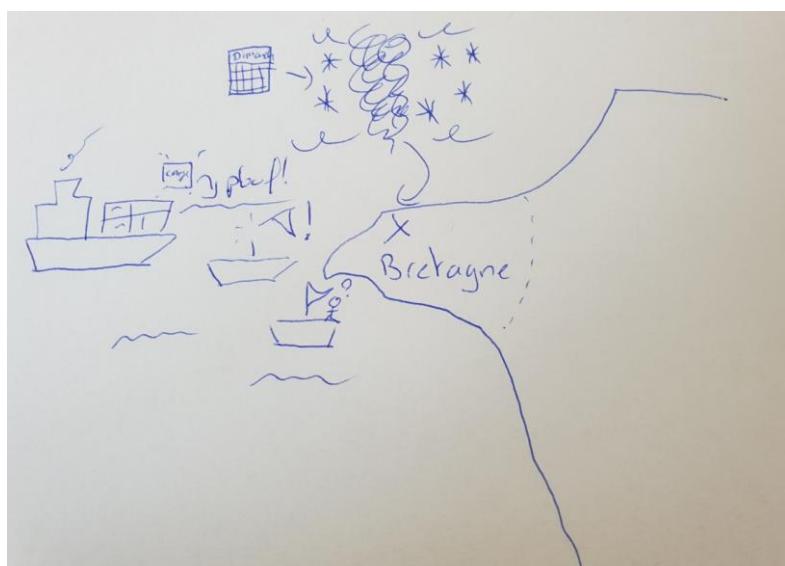


Figure 11 Schéma de déverbalisation

On remarque tous les éléments nécessaires à la mise en place de la scène : la carte de France avec une mise en avant de la Bretagne, ainsi qu'un calendrier qui indique « dimanche ». On note également la représentation de la tempête ainsi que des différents évènements qu'elle a causés. L'utilisation de flèches (entre le calendrier et la tempête, puis entre la tempête et la Bretagne) semble correspondre à une temporalité dans l'énoncé.

Nécessité des connaissances du monde

Un autre aspect est en lien avec la pensée visuelle : l'influence et les contraintes posées par la réalité physique. La langue des signes étant une langue qui donne à voir, faire usage de descriptions est un procédé courant. Qu'il s'agisse de nommer un protagoniste qui n'a pas encore de nom-signé (un signe qui a valeur de prénom), de décrire une pièce mécanique de haute technicité de l'aéronautique ou le fonctionnement des mandibules de la fourmi rouge amazonienne, la nécessité pour des connaissances du monde est d'autant plus marquée avec les LS, d'où l'importance de la recherche encyclopédique durant le processus de traduction.

Les connaissances encyclopédiques sont également nécessaires lorsqu'il est question de restituer du contexte. Comme dit précédemment, les LS disposent d'une structure discursive propre, qui pousse donc à la réorganisation du discours. Se faisant, pour plus d'intelligibilité et de clarté dans l'énoncé, c'est une stratégie du traducteur que de revenir brièvement sur une référence qui aurait pu être sous-entendue dans le texte source, et que la réorganisation vers les LS rend confuse ou indétectable. La figure 12 illustre ce phénomène. Elle présente en italique, le texte source à traduire. En dessous, entre crochets, une translittération de sa traduction en LSF. La première partie de cette traduction, représentée par les gloses en rouge, correspond au rappel du contexte qui permettra la bonne compréhension de l'information principale, représentée ensuite par les gloses bleues. Cette

traduction est proposée par un interprète en LSF. En annexe vidéo, un traducteur commente la nécessité d'ajouter du contexte dans sa traduction.

Le tribunal correctionnel réclame des investigations supplémentaires suite aux évènements de mardi dernier. L'homme a été remis en liberté dans l'attente de son jugement, mais suspendu de ses fonctions d'ici là.

[Rappelez-vous mardi dernier] [à la campagne] [une fête entre jeunes] [paf]
[l'intervention très violente de la police blesse un jeune] [paf] [le policier responsable est arrêté et mis en garde à vue] /

[Aujourd'hui] [le tribunal correctionnel annonce qu'il n'y a pas encore de verdict] [besoin d'une enquête approfondie]/ [cependant, le policier est libre mais suspendu jusqu'au jugement]

Figure 12 Exemple de précision du contexte (texte source au dessus ; traduction en dessous)

À la suite de cette section sur la théorie, nous allons aborder l'étude des pratiques professionnelles. La section suivante synthétise plusieurs observations d'importance à ce sujet.

1.2.4 De la théorie à la pratique: la traduction français-LSF

La langue des signes demeure une langue à faible visibilité, bien qu'en hausse. Cela s'explique d'une part par le peu de locuteurs concernés (283000 signeurs en 2008 d'après une étude de la DREES⁷), d'autre part par sa récente réhabilitation. La question de l'accessibilité se pose néanmoins de plus en plus ces dernières années, et la situation semble en passe d'évoluer encore, comme nous le détaillons ci-après.

Besoin grandissant pour du contenu accessible

Qu'entend-on par contenu accessible ? Il s'agit de rendre l'information quelle qu'elle soit (programmes télévisés, informations dans les transports

⁷ <https://drees.solidarites-sante.gouv.fr/sites/default/files/2020-09/dt131.pdf> p.65

en commun et les lieux publics, sites internet, administrations...) disponible et intelligible pour tous. Pour le public sourd, cela passe par la mise en place de boucles magnétiques pour les personnes appareillées, mais aussi par la création de contenus en LSF, sinon par la traduction de contenus existants. Ce besoin est grandissant suite aux récentes avancées législatives, nous l'avons vu, mais aussi grâce à la dynamique des acteurs culturels et sociaux qui souhaitent s'ouvrir au plus grand nombre. Cinémas, musées, et salles de spectacles rejoignent de plus en plus les institutions françaises et autres grandes entreprises qui font de plus en plus appel au marché de la traduction.

Peu de traducteurs fr-LSF

La demande est présente, mais quid de l'offre ? En France, malgré la création d'un Master de traduction, interprétation et médiation en LSF en 2011, et l'intégration en 2020 de la LSF dans le master LISH de l'Université Paris 8, le nombre de traducteurs professionnels demeure très faible. Nous estimons à environ une quinzaine le nombre de professionnels concernés, pour environ 600 interprètes. Le manque de traducteurs pousse d'ailleurs à une extrapolation du rôle de l'interprète, à qui l'on demande parfois d'entreprendre des tâches de traduction et d'écrivain public faute de moyens humains. Les formations d'interprète et de traducteur en LSF se retrouvent à la fois dans des centres dédiés à la traduction (D-TIM à l'Université de Toulouse, ESIT à l'Université Paris 3) et dans des départements dédiés aux études linguistiques (Université de Lille, de Rouen, ou de Paris 8 – Encrevé, 2014)

Des outils mais pas pour les LS

Si les traducteurs de l'écrit disposent de nombreux outils de TAO pour les assister dans leurs travaux, il semble à première vue qu'aucun ne soit dédié

à la langue des signes. Les traducteurs en LS disposent de dictionnaires en ligne et d'un accès parfois difficile à des vidéos thématiques en LS mal répertoriées. Certaines institutions se sont attelées à la création de glossaires ou de lexiques thématiques, du point de vue de l'enseignement, pour le français ou les mathématiques par exemple, en vue de palier ces difficultés.⁸

La question qui se pose naturellement ensuite est donc la suivante : pourquoi ne pas essayer d'égaliser la situation pour tous les traducteurs, peu importe leurs langues de travail ? Et a fortiori, comment s'y prendre ?

Comment outiller la traduction des LS ?

Les éléments exposés ci-avant nous amènent à réfléchir sur la façon d'outiller les LS. Si les logiciels actuellement populaires dans le milieu de la traduction texte à texte ne sont pas en mesure de les prendre en charge, il convient dans un premier temps de se demander pourquoi. Reposent-ils sur des particularités clés incompatibles avec la LS, et si oui, lesquelles ?

La question peut également être prise dans l'autre sens, à savoir est-ce le processus de traduction du français vers la LSF qui rend cette dernière incompatible avec les logiciels de TAO actuels ? Le changement de modalité implique certainement la prise en compte de données supplémentaires, ainsi que la gestion de formats nouveaux (dans notre cas, la vidéo).

Difficile de répondre à la question « Comment outiller avec la TAO la traduction des LS » sans d'abord s'intéresser à l'un puis à l'autre. C'est ce que se propose de faire ce travail de thèse. Dans un premier temps, nous nous intéresserons à l'anatomie d'un logiciel de TAO type à destination de la traduction texte à texte : quels types d'assistance fournissent-ils, et sur quels grands principes reposent-ils. D'autre part, nous traitons la question de

⁸ <https://ocelles.inshea.fr/fr/accueil>, <https://signmaths.univ-tlse3.fr>

la traduction en LSF : quelles sont les étapes constitutives du processus ? Sont-elles systématiques, ordonnées, automatisables ? Pour répondre à ces questions, les études sont menées directement avec des professionnels des métiers concernés, dans le but de mieux spécifier le métier d'une part. Mieux connaître le métier nous permettra de mieux l'outiller, et de mieux l'enseigner. D'autre part, les connaissances scientifiques formalisées ou du moins théorisées dans le domaine ne sont pas aussi nombreuses que pour les autres langues. Ces travaux contribuent à combler l'écart entre langues vocales et langues signées. Dans un dernier temps, ce travail de thèse propose une mise en application des connaissances rassemblées, par le biais de prototypes logiciels.

Le chapitre suivant dresse un état de l'art en deux parties. La première concerne la TAO, et tente de répondre aux questions soulevées : principes, fonctionnement, adaptabilité et compatibilité avec les LS. La seconde partie traite la question épineuse de l'évaluation de la traduction. Si la TAO améliore la **T**raduction, il nous faut d'abord savoir dans quelle mesure, donc nous traiterons la question de son évaluation.

Chapitre 2

État de l'art concernant la TAO et l'évaluation de la traduction.

Ce second chapitre propose une synthèse des travaux concernant la traduction assistée par ordinateur d'une part (section 2.1), et l'évaluation de la traduction d'autre part (section 2.2). Nous reviendrons tant sur la TAO que sur la traduction automatique, puisqu'aujourd'hui partie intégrante de la TAO, nous le verrons. Le chapitre se termine par une analyse de l'état de l'art, ainsi qu'une synthèse sur les manques qu'il présente.

2.1 Traduction assistée par ordinateur

La traduction est une activité humaine, mais l'on peut confier certaines tâches de son processus à la machine. L'opérateur humain reste expert et maître de sa traduction : c'est là l'objectif de la traduction assistée par ordinateur. La première section fait le point sur ce qu'ont à offrir les outils et les logiciels de TAO : fonctionnalités et fonctionnement, principes de base. La seconde section creuse la question de la TAO à destination des langues des signes. Cette partie se termine sur le constat de la rencontre des deux: pourquoi les logiciels existants ne peuvent-ils prendre en charge les LS

2.1.1 Fondamentaux

Cette section concerne les environnements de travail intégrés. À la lumière de notre problématique, nous avons étudié plusieurs logiciels (liste disponible en annexes) pour en lister les points clés, en cerner le fonctionnement et l'impact dans l'assistance du travail de traduction. Ces points clés ont été identifiés selon un critère simple : leur suppression remet en cause l'intégrité du logiciel, l'objectif étant de rechercher et de mettre en avant des caractéristiques essentielles.

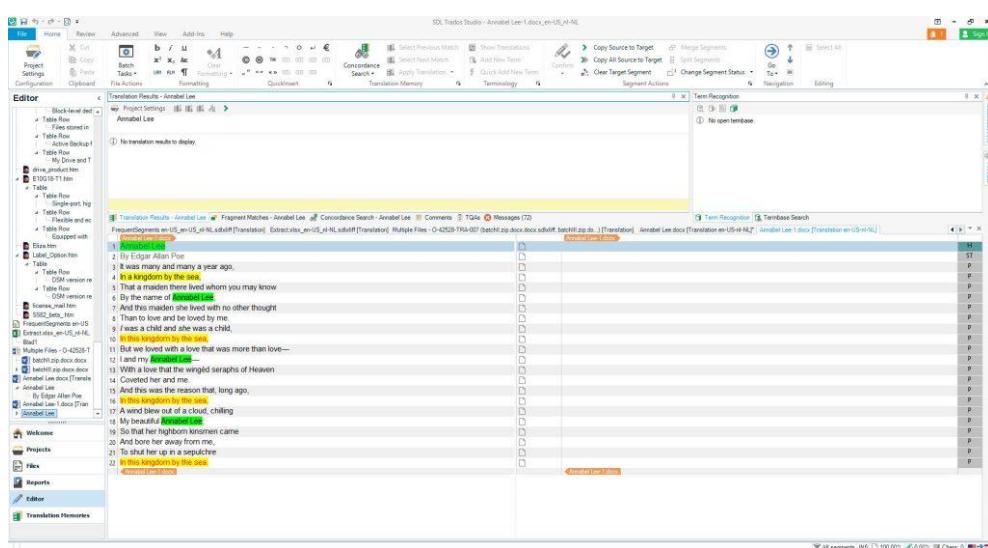


Figure 13 Le texte source découpé (à gauche), et les segments cibles correspondants que le traducteur doit remplir (à droite). Tiré du logiciel SDL Trados Studio 2017

Forme écrite éditable

Tous les logiciels que nous avons étudiés reposent sur l'utilisation d'une forme écrite éditable. La remarque peut sembler triviale, d'autant plus qu'il est question ici de logiciel d'aide à la traduction texte à texte : la place prépondérante qu'y joue l'écrit est une évidence. Pourtant, la modification ou la suppression de cette caractéristique remet en cause le fonctionnement logiciel. En effet, le recours à une forme écrite éditable concerne les textes sources et cibles, mais également le contenu des modules annexes (glossaires, dictionnaires, recherches terminologiques et contenu de la mémoire de traduction). La suppression de cette forme écrite conduit à l'impossibilité pour le traducteur à éditer manuellement sa traduction. Du reste, nous n'avons pas observé la possibilité d'avoir recours à d'autres formats, qu'il s'agisse d'images ou de vidéos.

Linéarité

Le second point clé identifié est ce que nous avons appelé *le principe de linéarité*. Dans un premier temps, le texte source est chargé et découpé automatiquement par le logiciel en segments, c'est-à-dire en unités plus petites. Le grain de ces segments est paramétrable, mais correspond généralement au niveau de la phrase. Chaque segment source est porté en vis-à-vis d'un segment cible, initialement vide, que le traducteur doit remplir (voir figure 13). Pour ce faire, il peut soit traduire ab nihilo, soit avoir recours aux différents modules que le logiciel peut offrir. Le traducteur procède en remplissant chaque segment vide avec la traduction du segment source associé, ligne par ligne. L'ordre des segments n'est pas modifiable: la traduction s'opère obligatoirement de haut en bas dans l'ordre du texte source. Seules certaines exceptions sont possibles, comme la fusion de deux segments consécutifs en un seul, ou l'inversion de deux segments consécutifs. Ces exceptions ne concernent que des segments voisins, et ne permettent pas de modifier plus largement l'ordre dans lequel sont traités ces derniers.

Le logiciel admet que la concaténation des segments traduits correspond à la traduction des segments source concaténés, comme illustré figure 14. Mathématiquement, le principe de linéarité se formule $f(a+b) = f(a) + f(b)$, où « + » est la mise en séquence des segments, et « f » est la fonction de traduction. Dans cette figure, « Seg. » correspond à segmentation, « Trad. » à traduction, et « Concat. » à la concaténation.

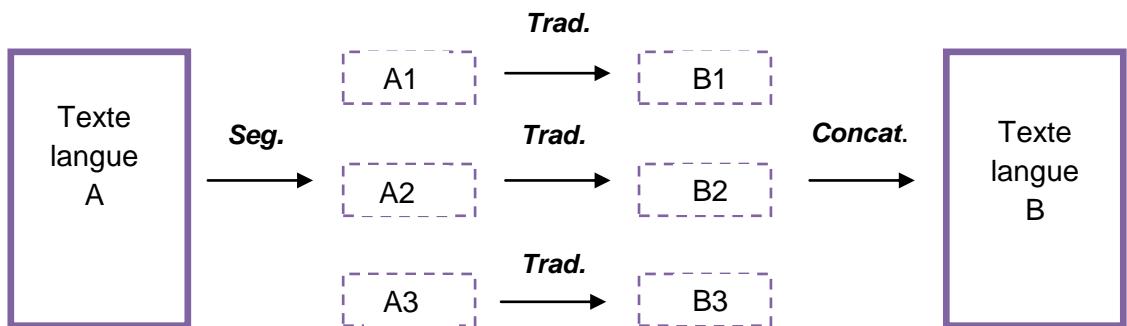


Figure 14 Le principe de linéarité

Capitaliser sur les travaux précédents

Le troisième point clé sur lequel repose le fonctionnement des logiciels de TAO actuels consiste en l'intégration de mémoires de traduction (abrégé MT dans la suite de ce mémoire). Sur le marché depuis presque quarante ans, elles sont aujourd'hui totalement démocratisées à la fois chez les traducteurs freelance, comme au sein de plus gros services de traduction publics ou privés. Certains clients fournissent même la MT associée au projet en cours pour s'assurer des termes utilisés lors de la traduction de leurs produits. Ces outils fonctionnent sur le principe d'une base de données alimentée et consultée en temps réel. À mesure que le traducteur complète les segments cibles, la mémoire de traduction enregistre les paires de segments parallèles : le segment A est la traduction du segment B. Cet enregistrement est automatique. D'autre part, et toujours alors que le traducteur procède à sa traduction, les segments sources cette fois sont analysés par la MT, qui parcourt sa base de données à la recherche d'occurrences semblables au segment en cours d'analyse. La correspondance

peut-être de 100%, appelée alors « perfect match », ou d'un pourcentage de similitude moindre qui ne descend cependant jamais en dessous d'un pourcentage seuil fixé par le traducteur lui-même (généralement supérieur à 75% pour conserver l'efficacité de l'outil). Si la MT trouve une correspondance, qu'elle soit totale ou partielle, elle la suggère alors au traducteur. Ce dernier reste libre d'accepter la suggestion de traduction, de la rejeter, ou de l'accepter avec modifications. Par exemple, supposons que nous avons précédemment traduit “la souris mange le chat” par “the mouse eats the cat”, si nous rencontrons à présent le segment “la souris mange le biscuit”, la MT suggérera automatiquement la traduction “the mouse eats the cat” puisque la structure des deux segments est identique à 80%. Libre alors au traducteur d'accepter la suggestion et de procéder aux modifications nécessaires pour rendre la traduction conforme. Les modifications apportées seront elles aussi enregistrées comme nouvelle entrée de la MT, permettant ainsi de capitaliser sur le travail antérieur.

L'utilisation des MT est associée à des bénéfices en termes de qualité, cohérence et uniformité, rapidité d'exécution, et confort de travail du point de vue du traducteur (O'brien, 1998, p.119; Webb, 1998, p.20; Reinke, 2004). La rapidité s'améliore par le pré-traitement des segments à traduire par la MT, bien que ce dernier tende à transformer la tâche de traduction en tâche de post-édition (c'est-à-dire à la révision d'une traduction proposée par le logiciel). Shreve (2006) souligne que le recours aux MT mène à un processus de traduction qui tend à se micro-contextualiser au niveau du segment, avec des impacts notables en ce qui concerne le style, la cohérence élargie et la structure des documents (Heyn, 1998). La cohérence et l'uniformité concerne aussi bien les différents travaux d'un même traducteur (MT personnelle, vocabulaire lissé au travers des différentes productions dans le temps), que les productions de traducteurs différents sur un même projet commun. Quand les entreprises qui commercialisent ces logiciels affirment qu'ils permettent d'augmenter la *qualité* de la traduction, d'après

Reinke (2004), cela n'est dû qu'à l'augmentation du degré de *cohérence* au sein de la traduction du fait de la répétition des formules et structures contenues dans les outils terminologiques et les mémoires de traduction. Cependant, la notion de *qualité* est complexe à définir, comme nous le verrons dans la section 2.2.

Intégration de modules (pour tout avoir sous la main)

Le terme « environnement de travail intégré » souligne la notion de regroupement des ressources. Au delà des fonctionnalités de traitement des textes sources et cibles, et des mémoires de traduction, ces logiciels regroupent également un ensemble d'outils d'aide à la traduction qui peuvent exister en dehors du logiciel, dans le but de centraliser l'assistance et d'éviter aux traducteurs de devoir naviguer entre plusieurs outils à la fois. On trouve donc directement accessibles depuis la zone de travail : des glossaires, des lexiques terminologiques, des correcteurs ou vérificateurs orthographiques et grammaticaux, des concordanciers et des outils intégrés de traduction automatique. Ces derniers, démocratisés ces dernières années avec les améliorations conséquentes de leurs performances notamment grâce au deep learning, font partie intégrante des logiciels de TAO.

La traduction automatique (TA), qui consiste en la traduction brute d'un texte par une succession ou une combinaison de programmes informatiques, n'est pas incompatible avec la TAO. Dans les environnements de travail intégrés pour la traduction, elle permet notamment de compléter automatiquement les segments cibles. Le traducteur humain n'intervient éventuellement qu'à la fin du programme pour corriger les erreurs potentielles lorsque cela est nécessaire : il s'agit alors d'une tâche de post-édition.

2.1.2 Et pour les LS ?

Les logiciels de TAO aujourd'hui n'incluent pas les langues des signes dans leur catalogue. Dans un premier temps, nous nous intéressons aux outils de

traduction existants pour les LS. En effet, l'absence de logiciels de TAO (qui conserve l'expertise humaine au cœur de la tâche) n'implique pas l'absence de logiciels ou d'outils. Nous investiguons ensuite la question de l'absence de la TAO pour les LS dans la seconde moitié de la section.

Les outils existants

Le domaine informatique n'exclue pas les LS. Au contraire, le mariage des deux est fructueux: le champ des possibles est conséquent, et les pistes à explorer nombreuses. Capture de mouvement, avatars et génération automatique, ressources, traduction automatique, reconnaissance de signes isolés, les travaux de recherche issus du traitement automatique des LS ont pour but commun d'augmenter le degré d'accessibilité du quotidien des personnes sourdes.

Il existe par exemple des sites ressources pour les LS. Ces derniers ne sont pas uniquement destinés aux traducteurs, mais peuvent constituer un ensemble de modules de base exploitables dans le domaine de la traduction. On trouve plusieurs dictionnaires français - LSF en ligne, parmi lesquels les sites Elix et SpreadTheSign⁹. La recherche se fait en français écrit, et le résultat est proposé sous forme de vidéos de LSF (signe isolé). Le site Elix propose également la définition du terme recherché en français, et en LSF si disponible. Il s'agit de sites collaboratifs, dont la base de données est alimentée par les utilisateurs. D'autres sites type dictionnaire ou lexique en ligne, comme The Danish Sign Language Dictionary (Kristoffersen & Troelsgard, 2012) ou le corpus LSFB¹⁰ de l'Université de Namur (Meurant, 2015) proposent en sus une vue concordancier : chaque signe apparaît avec sa définition, accompagnée d'une phrase d'exemple dans laquelle le signe est utilisé en contexte. Le corpus DGS et sa base de données collaborative iLex (qui permet l'annotation de corpus en LS, et de requêter ces

⁹ Elix : <https://dico.elix-lsf.fr/>
SpreadTheSign: <https://www.spreadthesign.com/fr>

¹⁰ <https://www.corpus-lsfb.be/accueil.php>

annotations) propose lui aussi quelques affichages type concordancier, dans l'objectif de travailler sur les sens et nuances de sens (Langer, Müller & Wähl, 2018).

En termes d'accessibilité, on retrouve les travaux concernant la signalisation à l'adresse des personnes sourdes et malentendantes dans les gares françaises. Baptisée Jade, l'avatar retransmet en direct les annonces sonores en gare sur les écrans de la gare¹¹. À l'international, l'Université DePaul de Chicago travaille sur la question depuis plus de quinze ans, et propose aujourd'hui un des avatars les plus aboutis qui soit en termes de rendu naturel des signes (Wolfe et al, 2016), mais qui n'est pas encore intégré dans des logiciels grand public (voir figure 15).

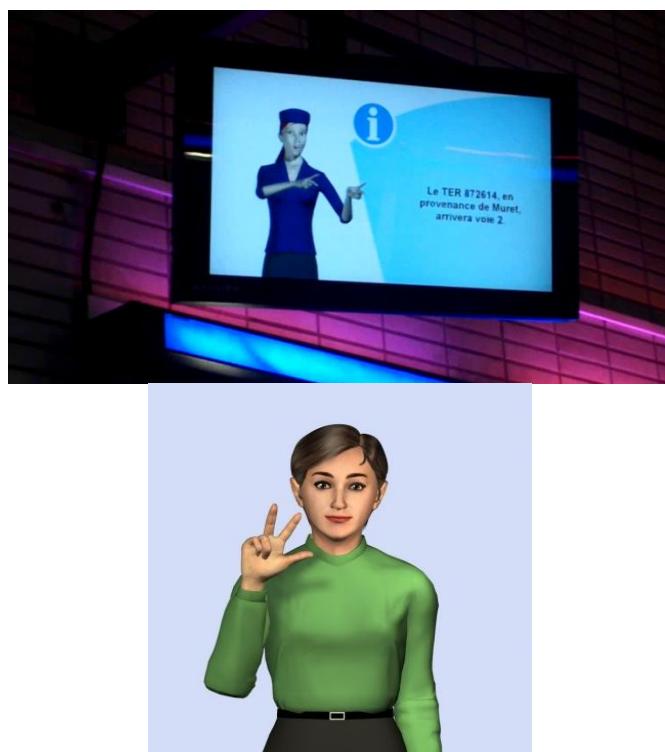


Figure 15 Les avatars Jade (en haut), et Paula (en bas)

Les avatars, ou signeurs virtuels, sont un moyen utilisé pour afficher des

¹¹ <https://www.accessibilite.sncf.com/la-lettre-de-l-accessibilite/lettres/2010/décembre-2010-no2/article/jade-des-mots-pleins-les-mains> : un exemple est disponible dans les annexes vidéos

2.1 Traduction assistée par ordinateur

messages en LS sans que ces derniers ne soient associés à l'identité d'une personne, ou pour visualiser un message en LS en différé (comme le permet l'écrit pour le français par exemple). Les méthodes d'animation pour ces avatars peuvent être très différentes. Il peut s'agir d'un rejet d'animations pré-enregistrées, comme dans les jeux-vidéos par exemple, auquel cas les animations peuvent être le fruit du travail d'infographistes (comme Jade), ou alors être issue de la capture de mouvement. Il peut également s'agir d'une synthèse pure à partir d'une description mathématique des déplacements à appliquer aux segments articulaires de l'avatar, comme c'est le cas pour Paula. D'autres avatars pourraient également être animés par un mélange des deux méthodes précédentes.

Si la TA est répandue dans le domaine de la traduction de texte à texte, elle est de loin beaucoup moins développée et performante pour ce qui est des LS. Les premiers travaux concernaient l'utilisation de gants sensoriels pour capter les signes (Bräffort, 1996). Les capteurs sensoriels ont peu à peu été remplacés par les caméras, et la vision par ordinateur a pris la suite, cette dernière étant plus efficace pour la captation et beaucoup moins invasive et contraignante pour le signeur. Ces travaux concernent la traduction des LS vers les langues vocales. Concernant l'autre sens de traduction, des langues vocales vers les LS, les travaux sont encore plus récents et reposent sur les technologies d'avatar mentionnées plus haut ainsi que sur l'entraînement de modèles qui ne sont pas impactés par la petite taille des corpus d'entraînement (Miyazaki et al., 2020). Ces systèmes de TA n'offrent pas une couverture suffisante du besoin de traduction, contrairement à leurs équivalents texte à texte, ce qui nous oriente d'autant plus vers la nécessité de développer la TAO des LS.

Bien qu'il s'exprime à des degrés variés en fonction des études, l'aspect sociologique n'est jamais bien loin. On peut dissocier deux cas de figure de traduction. Dans le premier cas, dit "d'urgence", c'est la rapidité et la spontanéité de la traduction qui prime. On pense notamment aux annonces

en temps réel, pour le signalement d'incidents (transports) ou d'évènements à risque (inondations), lors de l'intervention des secours, ou de catastrophes d'envergure (attentats, incendies). Dans ces situations, l'important est que le message soit livré et compris rapidement : le sens prime sur la forme, quitte à ce que la traduction soit incomplète ou peu élaborée, pourvue qu'elle soit fidèle au message. Le second cas, dit "commun", met au contraire l'accent sur l'importance de la qualité de la langue. L'objectif est ici d'atteindre une traduction qui soit la meilleure possible, peu importe le temps nécessaire à sa production. Ce sont des traductions que l'on s'attend à trouver au quotidien dans des situations où la transmission de l'information ne se fait pas en temps réel : traduction de contenu d'un site web, d'une œuvre littéraire ou culturelle, de documents administratifs. Dans ces cas, la forme importe autant que le sens, et la traduction se doit d'être complète et fidèle.

La traduction automatique se prête à l'exercice de la traduction dans les cas d'urgence : les diverses applications embarquées permettent une compréhension minimale mais suffisante pour que la communication soit efficace. Pour toutes les autres situations de traduction, la TAO et donc le recours à l'expertise humaine reste préférable, d'autant plus lorsque l'une des langues de travail s'est vue interdite et critiquée comme ont pu l'être les LS au cours de l'histoire.

Pourquoi la TAO telle qu'elle existe ne peut s'appliquer aux LS

La section 2.1.1 a présenté les principes fondamentaux des logiciels de TAO. En regard des informations présentées dans le premier chapitre concernant les langues des signes, ainsi que d'observations personnelles réalisées sur le terrain, nous proposons ci-après d'exposer les limites majeures quant à l'adaptation de la TAO aux LS. En effet, certaines des caractéristiques des LS rendent l'adaptation directe complexe, voire inutile.

Forme écrite éditable

Nous l'avons vu plus haut, les logiciels de TAO concernent la traduction de

texte à texte et reposent de fait sur l'utilisation d'une forme écrite dite éditable. Nous avons également vu que les LS ne disposent pas de forme écrite adoptée par les locuteurs. La première limite identifiée concerne donc le format du contenu à éditer dans les segments cibles, dans le cas où l'on s'intéresse à la traduction du texte vers les LS. La forme écrite devra rester présente puisque le texte source doit pouvoir être pris en charge, mais il faut envisager une solution côté LS, tout comme un recours à la vidéo.

Linéarité

La question du principe de linéarité se pose pour les langues des signes. L'ordre de construction des LS pourrait laisser présager du contraire. Ce principe implique que l'ordre des unités segmentées du texte cible reste le même que celui du texte source. La concaténation des segments cibles traduits équivaut à la traduction des segments sources concaténés, et dans le même ordre. Plus simplement, la méthode de traduction adoptée est celle du phrase à phrase. Doit-il être supposé, peut-il s'appliquer ? Prenons un exemple dans le cadre des LS.

Le texte suivant : « *Quelque 200 personnes pourraient avoir été tuées mercredi matin par un glissement de terrain provoqué par des pluies abondantes dans l'île indonésienne de Java, a annoncé une source policière.* », peut se traduire ainsi (traduction disponible dans les annexes vidéos) :

[Dans l'archipel d'Indonésie, sur l'île de Java] [depuis un certain temps les pluies sont très fortes] / [aujourd'hui mercredi matin] [plusieurs policiers annoncent qu'il se pourrait] [qu'un glissement de terrain ait fait 200 morts]

On observe que le premier élément de la phrase en français (« quelques 200 personnes pourraient avoir été tuées ») se retrouve en dernière position dans l'énoncé en LSF. On notera d'ailleurs que la notion d'incertitude du verbe « pourraient » est dissociée, et se retrouve avant dans l'énoncé signé. De

même, le premier élément signé correspond à un élément de sens qui intervient tard dans l'énoncé en français (la localisation « île de Java »). Ce principe de linéarité permet de respecter strictement la mise en forme du texte et l'ordre dans lequel sont présentées les informations, mais apparaît problématique dans le cas de langues dont la construction du discours ne se résume pas à une succession de phrases bout à bout : il s'agit de la seconde limite.

Mémoire de traduction

La MT stocke des alignements pour une paire de langues donnée (des paires de segments sources et cibles dont l'un est la traduction de l'autre), et permet de capitaliser sur le travail passé en réutilisant ces segments pour accélérer les traductions futures. Dans le cas de la traduction vers les LS, le segment source est d'une forme identique, mais le segment cible lui, devient un segment vidéo. Le recours à la vidéo pose plusieurs questions d'ordre technique. La première concerne leur stockage: la vidéo est plus imposante en termes de stockage, et plus contraignante à manipuler que le format texte. D'autre part, nous pouvons nous interroger sur la pertinence de conserver des fragments de vidéos pour reprendre tel quel le fonctionnement de la MT : comment effectuer une requête dans le contenu-même d'une collection de vidéos ? De même, quel serait l'intérêt d'éditer les uns à la suite des autres plusieurs fragments de traductions vidéos, si la personne qui signe n'est pas la même, ni le fond, ni même au final le contexte ? Le rendu final, en tout cas, n'aurait rien d'une traduction qualitative et professionnelle. De ces simples constats, il apparaît clairement que le principe de MT, s'il est performant pour la traduction texte à texte, nécessite quelques ajustements pour pouvoir être réinvesti dans la traduction texte à signes. Il n'en reste pas moins intéressant à adapter, puisqu'à nouveau, il constitue réellement le cœur de l'innovation de la TAO. Le fait de conserver des exemples de traductions précédentes reste un atout pour l'aide à la traduction, mais le fonctionnement des requêtes ainsi que la réutilisation des résultats

nécessitent une adaptation. Il s'agit de la troisième limite identifiée.

Les outils de TAO ont contribué à faire évoluer le métier de traducteur texte à texte au cours des dernières décennies, mais ne peut s'adapter tel quel pour les LS. Notre objectif est d'améliorer les conditions du métier de traducteur LS, et l'un des critères est la qualité. En effet, évaluer l'efficacité ou l'impact des outils de TAO sur les pratiques professionnelles et sur le processus de traduction en lui-même implique tacitement de s'intéresser à la qualité de la traduction produite. La section suivante s'intéresse à la question de l'évaluation de la traduction et de ses outils d'assistance, traduction automatique comme TAO. Une question vaste, qui peut s'aborder de différents angles.

2.2 Évaluer la traduction et les outils

Qu'est-ce qu'une bonne traduction ? Pour une source donnée, y a-t-il seulement un stéréotype de la bonne traduction, est-elle unique ? Pour deux versions cibles données, y en a-t-il une meilleure qu'une autre, et qu'est-ce qui permet de la qualifier ainsi ? Nous dressons d'abord une liste de critères qui permettent de qualifier (voire quantifier) la qualité d'une traduction produite, à la suite de quoi nous nous intéressons à l'évaluation plus directe des systèmes en eux-mêmes.

Entre objectivité et subjectivité

La traduction demeure une tâche sociale, en ce sens où elle concerne la communication et la culture. Partant de ce constat, on ne peut envisager d'évaluer une traduction sans prendre en compte le facteur humain, de deux sortes : le retour du public concerné pour ce qui est de la qualité du produit en lui-même, et le retour des professionnels qui les ont produits, pour ce qui concerne l'évaluation de leurs pratiques professionnelles et de leurs outils. À l'inverse, on peut se baser uniquement sur des retours subjectifs pour statuer de la qualité et de l'efficacité d'un produit ou d'un outil. Le traitement automatique des langues (TAL) a su développer plusieurs

métriques dédiées à l'évaluation, pour objectiver la qualité de traductions automatiques de langues naturelles. L'objectif et le subjectif sont complémentaires, et permettent de corroborer ou nuancer leurs résultats. Des personnes interrogées sur leurs pratiques auront tendance soit à omettre des tâches qu'elles effectuent pourtant, soit à insister sur des faits qu'une analyse objective permet de déclarer anecdotiques. De même qu'un outil, qui apparaît objectivement peu impactant en termes de productivité, peut être subjectivement considéré comme efficace par les utilisateurs du fait du confort et de l'aisance qu'il apporte.

2.1.1 Critères pour évaluer une traduction

D'après certains experts de la traduction (Bell, 2000, Delisle, 1980, Lederer, 1994, Dancette, 1998), le traducteur doit faire montre d'une palette de compétences : des compétences linguistiques, des compétences extra-linguistiques ou socio-culturelles, ainsi que des compétences discursives et stratégiques. Cette section liste dans un premier temps plusieurs critères quant à l'évaluation de la traduction. Il est question ici de critères d'intérêt, et non pas de méthodes ou de métriques, qui seront abordés dans la section suivante.

Cohérence de la forme

Il s'agit là de l'expression des compétences linguistiques du traducteur. La forme est le reflet de la maîtrise des langues de travail. Il s'agit là de produire une traduction qui soit grammaticalement correcte, orthographiquement irréprochable (quand la langue cible est du texte), et avec une syntaxe de qualité. Ces éléments découlent directement de la capacité de compréhension de la langue source, ainsi que des compétences en langue cible pour un rendu adapté.

Respect du message et de l'intention

Il est question ici de mettre en œuvre les compétences discursives, socio-linguistiques et stratégiques du traducteur. Le propre de la traduction est de

2.2 Évaluer la traduction et les outils

transférer un message d'une langue dans une autre, en en respectant le sens et l'intention, c'est-à-dire la visée du locuteur, l'effet escompté sur l'interlocuteur (ou le lecteur). D'après Guiraud (1968), dont nous reprenons le schéma ci-dessous (Fig. 16) : « Plus la diffusion s'étend, plus le contenu du message se rétrécit ; la base commune, qui n'est jamais complète entre deux individus, s'amenuise à mesure qu'elle s'étend à un nombre plus grand de personnes. »

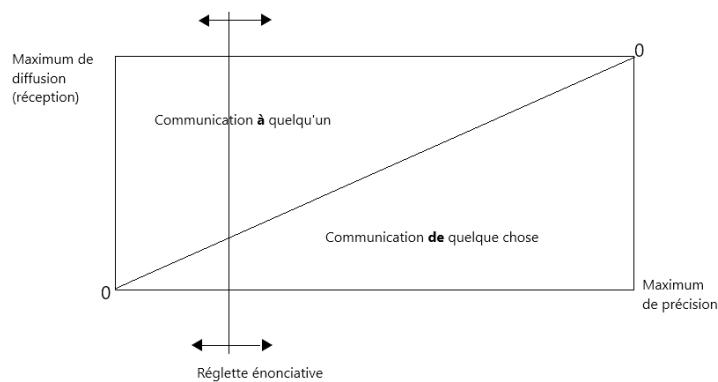


Figure 16 Schéma de la diffusion en fonction de la précision, Giraud 1968

D'après lui, plus nous voulons diffuser un message, c'est-à-dire plus nous le destinons à un grand nombre d'interlocuteurs dans le but de leur faire comprendre, plus le message va perdre en précision : c'est proche du principe de vulgarisation. À l'inverse, si nous voulons conserver toute la précision d'un message, alors il faut accepter que ce dernier ne se destine qu'à un nombre restreint de destinataires. La fidélité du message et le respect de l'intention de l'auteur sont obligatoires. Ceci dit, le traducteur sera toujours amené à faire des choix de traduction, que ce soit pour adapter le message d'une culture à l'autre, ou pour nuancer le propos en contexte. Larose (2002) résume les faits comme suit : « Quel que soit le traducteur et quelle que soit son époque, il est constamment appelé à prendre des décisions le long de ce continuum. Il est toujours placé devant la double nécessité de transmettre une expérience singulière le plus exactement et le plus complètement possible au plus grand nombre d'individus possible.

Mais à quoi bon dire tout à personne ou ne rien dire à tout le monde ? La double postulation entraîne des hésitations d'ordre décisionnel et expose sans cesse le traducteur à des contradictions, puisqu'il est le plus souvent forcé d'aliéner cette singularité de l'original sous prétexte de devoir communiquer un message. » C'est la responsabilité du traducteur de trouver le juste équilibre entre diffusion et précision, et c'est pour garantir le respect de ces critères que le traducteur traduit *vers* sa langue première¹², dont il maîtrise les nuances fines.

Évaluation de la traduction automatique

Il existe plusieurs métriques automatiques pour rendre compte de l'efficacité d'un système de traduction automatique. Là encore, la qualité d'une traduction reste subjective, puisque l'humain reste le destinataire du produit. De fait, ces métriques attribuent certes un score de qualité, mais ce dernier est lié au jugement humain de ce qu'est une bonne traduction. Une bonne métrique doit donc attribuer un score élevé aux traductions qu'un humain notera élevée, et un score faible aux traductions faiblement notées par un humain. La corrélation se fait généralement à deux niveaux : au niveau de la phrase, où les scores sont calculés pour un ensemble de phrases traduites, puis corrélés avec le jugement humain pour ces mêmes phrases, et au niveau du corpus, où les scores sur les phrases sont agrégés pour les jugements humains et les jugements métriques, puis corrélés. Les résultats au niveau de la phrase sont rarement communiqués, bien que Banerjee et al. (2005) donnent des chiffres qui montrent que, au moins pour la métrique automatique METEOR, la corrélation au niveau de la phrase est nettement moins bonne que la corrélation au niveau du corpus. Ils listent également cinq attributs qu'une bonne métrique automatique doit posséder : corrélation, sensibilité, cohérence, fiabilité et généralité. C'est-à-dire qu'une métrique automatique fiable doit présenter une corrélation élevée avec le

¹² Exception faite pour les LS, où les entendants sont amenés à traduire pour les raisons évoquées au chapitre 1.

2.2 Évaluer la traduction et les outils

jugement humain, doit être cohérente et donner des résultats similaires au même système de TA sur un texte similaire. Elle doit être sensible aux différences entre les systèmes de TA et fiable en ce sens que les systèmes de TA qui obtiennent des résultats similaires doivent produire des performances similaires. Enfin, la métrique doit être générale, c'est-à-dire qu'elle doit fonctionner avec différents domaines textuels, dans un large éventail de scénarios et de tâches de TA : champs lexicaux, genres de textes variés...

La métrique BLEU (Bilingual Evaluation Understudy) par exemple, est l'une des premières métriques à pouvoir attester d'une corrélation haute avec le jugement humain (Papineni, 2002). Elle demeure également l'une des métriques automatiques les plus fréquemment utilisées. Les scores sont calculés en comparant le segment à évaluer (généralement une phrase) à plusieurs traductions de référence considérées comme étant de bonne qualité. Pour l'évaluation d'un corpus entier, les scores pour chaque phrase sont lissés de sorte à faire apparaître un score moyen pour l'ensemble. À noter cependant que BLEU ne prend pas en compte le niveau d'intelligibilité ou la correction grammaticale des items évalués. Le score prend la forme d'une valeur entre 0 et 1, 1 représentant une correspondance exacte entre l'item évalué et l'une des traductions de référence. Pour cette raison, même une traduction réalisée par un professionnel humain peut ne pas atteindre le score de 1 : ce n'est pas nécessaire, l'objectif est de s'en approcher le plus possible. Papineni souligne dans ses travaux que plus les traductions de références seront nombreuses, plus le score sera élevé.

Les critères abordés au cours de cette section concernent la qualité de la traduction en elle-même. Concernant les outils d'aide à la traduction, se pose la question de l'efficacité d'une part, et du bénéfice qu'ils peuvent apporter de l'autre. C'est ce à quoi s'intéresse la section suivante.

2.1.2 Comment évaluer les outils d'aide à la traduction ?

La question de l'évaluation de la TAO s'aborde de deux points de vue. Il faut d'une part considérer la performance du logiciel : son utilisation permet-elle de maintenir, ou sinon d'améliorer la qualité du produit final ? D'autre part, comment ces outils sont-ils perçus par les utilisateurs ? Les paragraphes suivants reviennent sur plusieurs pistes possibles pour explorer ces points de vue.

Évaluation de la TAO

Il convient de noter en premier lieu que la majeure partie des études qui s'intéressent à l'évaluation de la TAO le font du point de vue des modules de traduction automatique intégrés dans les logiciels, et par extension, du point de vue de la mémoire de traduction. Elles proposent d'évaluer l'utilité de tel ou tel système de TAO en examinant la qualité de la traduction et/ou la productivité du traducteur, au travers de critères tels que la vitesse de traduction, la quantité de post-édition nécessaire pour parachever une traduction, ou encore le nombre d'erreurs encore présentes dans la traduction finale.

Koehn (2009), quant à l'évaluation d'un outil de TAO qu'il a lui-même participé à développer, place une dizaine de traducteurs en situation. Ces derniers sont amenés à traduire entre français et anglais (l'auteur précise que le choix de langue est en lien avec la performance de la mémoire de traduction) avec l'outil, et leurs interactions avec la machine sont enregistrées (frappe, clic à la souris, durée de traduction). L'outil en question, Caitra¹³, propose trois types d'assistance à la traduction (prédition et complétion de phrases, traduction automatique et post-édition, ainsi qu'une « table de traduction », une liste consultable de traduction possible). Les résultats de cette étude mettent en avant le fait que

¹³ L'auteur mentionne <http://www.caitra.org/>, cependant l'outil ne semble plus maintenu.

2.2 Évaluer la traduction et les outils

l'utilisation de pareille assistance a conduit les traducteurs à traduire plus rapidement, et à produire des traductions de qualité. En moyenne, il observe que les traducteurs sont plus rapides de 16% en utilisant la table de traduction, de 27% en utilisant l'outil de prédition et de complétion de phrases, et de 39% lorsqu'il s'agit seulement de post-éditer une traduction automatique. Il note que certains traducteurs divisent leur temps de traduction par deux. Une analyse plus détaillée des temps de pause (c'est-à-dire des temps morts dans la tâche) a établi que l'assistance a permis d'éliminer les pauses les plus longues pour les traducteurs les plus lents. Du point de vue de la qualité de la traduction, qui était ici évaluée par des experts humains, elle s'avère identique sinon meilleure pour les utilisateurs ayant fait usage de l'assistance. À noter que le critère d'évaluation de la qualité ici concerne le rendu du sens (le sens de la source et de la cible est-il identique ?) ainsi que le niveau de langue, mais sans plus de précisions à ce sujet.

Fiederer et O'brien (2009), dont l'étude compare la qualité d'une traduction humaine à celle d'une traduction automatique post-éditée, montre que des évaluateurs humains jugent ces deux types de traduction équivalents, sinon même la traduction post-éditée meilleure en terme de clarté et de précision. En ce qui concerne le style en revanche, les traductions humaines restent mieux notées. Garcia (2011) compare dans ses travaux également des segments post-édités avec des traductions humaines. Il met en avant que le gain de productivité est minime entre l'un et l'autre. En revanche, selon son étude, la qualité des traductions post-éditées serait meilleure que celle des traductions humaines. Ici encore, les critères d'évaluation concernant la qualité de la traduction sont peu détaillés, ce qui ne nous permet pas de comparer ces différentes études entre elles.

Ce manque de précision quant aux critères d'évaluation de la qualité nous pousse à revenir sur ce point qui, s'il a déjà été rapidement abordé dans la section 2.1.1, mérite d'être développé à la lumière de ces remarques. Larose

(1998) considère que l'évaluation d'une traduction doit s'articuler autour de quatre pôles : l'objet de l'évaluation (qu'est-ce qu'on cherche à évaluer), l'évaluateur (qui évalue ? quels sont ses objectifs, ses intérêts, son parcours ?), les paramètres d'évaluation (quels sont les critères qui permettent de juger le texte), et enfin les méthodes (de quelle manière l'évaluateur va-t-il appliquer les paramètres pour évaluer le texte). Ces quatre pôles mettent en lumière plus de variables possibles pour impacter l'évaluation que les seuls paramètres linguistiques. Selon l'auteur, l'évaluation d'un texte traduit repose historiquement sur une comparaison entre le texte source, et le texte traduit. Cette comparaison concerne la qualité de la langue d'une part, et la fidélité, le respect du sens d'autre part. Il note cependant que cette vision des choses exclut l'idée, pourtant majeure, que la traduction est un outil de communication, et non pas un processus de clonage. Il précise ainsi qu'une telle méthode d'évaluation, pour être la plus impartiale possible, doit également prendre un compte un cahier des charges de la traduction. Ce cahier des charges serait rédigé par le client, et permettrait ainsi pour chaque faute repérée dans le texte traduit, d'en pondérer l'impact et la gravité selon les attentes du client. C'est en ça qu'il propose non plus une vision de la qualité de la traduction comme étant le quotient d'un texte de départ et d'un texte d'arrivée, mais plutôt comme la mesure de l'écart entre résultat et attentes.

Les études précédemment citées statuent sur l'influence de l'utilisation d'un logiciel de TAO sur la traduction (en tant que produit) et sa qualité. Il faut également noter que cette influence se traduit aussi au niveau des pratiques professionnelles, c'est-à-dire sur la Traduction (en tant que processus), et sur les traducteurs eux-mêmes.

Impact de la TAO sur la profession

Le succès et l'efficacité des mémoires de traduction, que ça soit en termes de gain de productivité ou de qualité du produit fini, ont contribué à modifier les pratiques des professionnels. Logiciels largement implantés au

2.2 Évaluer la traduction et les outils

fil de ces quarante dernières années, la maîtrise d'un ou des outils de TAO est aujourd'hui un pré-requis à l'accès au monde de la traduction professionnelle ; si bien que le sujet est désormais enseigné dans les cursus de formation à la traduction.

L'utilisation de mémoires de traduction, et le degré de complétude qu'elles sont capables d'atteindre aujourd'hui, fait que l'opérateur humain est de moins en moins amené à traduire *from scratch* (proposer une traduction à partir de rien), et de plus en plus rendu à la post-édition. Il s'agit alors de valider ou non les suggestions de la machine, et d'y apporter si nécessaire une correction. Une fois l'intégralité du document traité de la sorte, la phase de révision est maintenue, durant laquelle le traducteur compare document source et produit traduit. Cette modification des pratiques a donné lieu à une révision des standards attendus et nécessaires à une prestation de traduction de qualité (norme ISO 17100¹⁴). À l'instar de Larose, Mellinger (2018) souligne que la notion de qualité d'une traduction peut varier selon le point de vue avec lequel la révision est effectuée, révision qui est pourtant rarement intégrée dans les modèles d'évaluation de la qualité des traductions.

Malgré que l'utilisation d'un logiciel de TAO ajoute un média supplémentaire dans une tâche qui impliquait déjà plusieurs agents (traducteurs, relecteurs, réviseurs), c'est cette technologie qui permet à chacun d'oeuvrer plus facilement de concert en permettant notamment la prise en charge de projets à l'échelle d'un service, et l'édition simultanée par plusieurs utilisateurs d'un même document. Cette collaboration favorisée impacte positivement la productivité, en cela que des projets plus ambitieux peuvent être achevés sur des délais plus courts. La Traduction devient alors l'art de concilier des « *obligations de coût, de temps, et de qualité* » (Bass, 2006).

¹⁴ ISO 17100. 2015. Translation Services – Requirements for Translation Services. Geneva: ISO

Synthèse du chapitre

Ce tour d'horizon des domaines liés à nos questions de recherche nous permet, au-delà de l'inventaire des savoirs pré-existant à nos travaux, de dresser une liste de besoins qui servira de conclusion à ce chapitre.

Nous nous sommes intéressés dans un premier temps à la structure des logiciels de TAO : quels types d'assistance proposent-ils, et sur quels principes majeurs ces assistances reposent-elles. Nous avons identifié trois grands principes, pour lesquels nous avons fait le lien avec les LS :

- une forme écrite éditable, ce dont ne disposent pas les LS ;
- le principe de linéarité, qui n'est pas valable pour les LS ;
- les mémoires de traduction, et aides à la traduction automatique, qui si elles sont performantes et bénéfiques pour la traduction texte à texte, nécessitent des adaptations majeures pour pouvoir être transposées aux LS ne serait-ce qu'en regard du changement de modalité.

Nous avons pu constater qu'aucun des logiciels de TAO que nous avons étudié proposait de prendre en charge la traduction des LS, mais que pour autant, les LS n'étaient pas totalement étrangères au domaine de l'informatique : il existe de nombreuses ressources de langue disponibles en ligne.

Dans un second temps, nous nous sommes penchés sur la question de l'évaluation de ces logiciels, et par extension, sur l'évaluation de la traduction. La notion de qualité de la traduction est plus vaste qu'une simple idée de correction grammaticale, et beaucoup de paramètres sont à prendre en compte (qui évalue, pourquoi, quel est l'objectif de la traduction, comment est-elle évaluée ?).

2.2 Évaluer la traduction et les outils

Il nous manque cependant des informations quant au processus de traduction impliquant les LS. Dans quelle mesure est-il comparable au processus de traduction texte à texte ? Les assistances proposées par la TAO actuelle serait-elle pertinente une fois appliquées aux LS ?

Le chapitre suivant présente deux études menées dans le but de répondre à ces questions. Il s'agit d'identifier clairement les pratiques comme les besoins des professionnels. Nous pourrons par la suite faire le lien avec l'existant, et dresser en détail une liste des modifications à lui apporter d'une part, et des ajouts nécessaires d'autre part. À quoi devrait ressembler un logiciel de TAO pour les LS ? D'ailleurs, pourrons-nous lui appliquer un modèle d'évaluation de la qualité semblable, ou faut-il reconsidérer la question de la qualité de la traduction propre aux LS ?

Chapitre 3

Propositions de spécifications logicielles

L'état de l'art concernant la question de l'assistance logicielle du traducteur humain pour les langues des signes est peu fourni, et vide de références quand il est question d'environnement de travail intégré comme ceux présentés plus tôt et qui s'adressent à la traduction texte à texte. Avant de chercher à réaliser un logiciel adapté, il convient dans un premier temps de s'enquérir des besoins réels des professionnels du métier. Comment passer de besoins exprimés à des fonctions d'assistance, dans un premier temps, puis quelle aide logicielle pour quel besoin ? La première section (3.1) présente l'étude préliminaire menée dans l'objectif de déterminer les besoins des utilisateurs cibles. La seconde section (3.2) présente le cahier des charges établi suite à l'analyse de cette étude.

3.1 Comprendre le métier de traducteur : expériences préliminaires

Dans le but de dresser une première spécification d’interface et d’outils pour un environnement de travail intégré pour les LS, il convient dans un premier temps de déterminer les besoins de la profession ainsi que les obstacles les plus fréquemment rencontrés au cours de la pratique professionnelle. Il s’agit ici de tenter de formaliser le métier pour proposer des assistances informatiques pertinentes. Pour ce faire, nous avons conduit deux études avec l’aide de traducteurs professionnels. Dans un premier temps, nous avons analysé un corpus filmé de professionnels pendant qu’ils travaillaient. En parallèle, nous avons conduit un brainstorming avec un autre groupe ; ces derniers étaient amenés à réfléchir et échanger sur leurs pratiques professionnelles. L’objectif général de cette étude est de spécifier les tâches caractéristiques du processus de traduction, dans ce cas d’étude, du texte vers les LS. Les expériences menées permettent de récolter des données à la fois objectives et subjectives, comme détaillé ci-après. La section 3.1.1 présente la première expérience réalisée, celle des traducteurs au travail, avec sa méthodologie ainsi que les résultats obtenus. La section 3.1.2 présente la seconde expérience menée, le brainstorming, sa méthodologie ainsi que les résultats. Enfin, la section 3.1.3 dresse une synthèse de ces deux expériences préliminaires.

3.1.1 Traducteurs du texte vers la LSF à l’œuvre

Cette première partie de l’étude concerne la collecte de données objectives. Elle repose sur l’analyse d’un corpus pré-existant à cette thèse (voire Filhol, 2021). Des professionnels de la traduction y sont filmés durant différentes sessions de travail, détaillées ci-après, et sans interactions avec l’expérimentateur une fois la consigne donnée. L’analyse de ces vidéos permet non seulement d’observer les tâches constitutives du processus de traduction, mais également de déterminer si ces dernières sont

systématiques, d'une part, et si elles sont réalisées dans un ordre strict ou partiel d'autre part. (Kaczmarek & Filhol, 2020).

Méthodologie

Le groupe de participants à cette première expérience se compose de quatre professionnels employés au sein d'un même service de traduction, répartis en deux binômes. Chaque binôme se compose d'un traducteur Sourd et d'un interprète LSF - français. Cette configuration est celle habituellement utilisée au sein de l'entreprise. L'expérience se déroule dans leurs locaux. Un des deux binômes est plus expérimenté que le second. Le travail en binôme permet aux professionnels de verbaliser leurs questionnements, de discuter des problèmes rencontrés, ce qui les rend observables puisque filmés.

La consigne était la suivante : procéder à la traduction des textes qui leur sont soumis, de la découverte du texte à la production filmée de la traduction, en étant libre d'utiliser le matériel et les ressources de leur choix. L'expérience se déroule en trois sessions, dont une première session de rodage. Les deux binômes ont chacun participé aux trois sessions. Chaque session implique un seul binôme, il y a eu donc en tout six sessions. Les deux membres du binôme reçoivent les mêmes textes à traduire. Les textes proposés à la traduction sont des brèves journalistiques, des textes dits courts de trois à cinq lignes, et des textes dits longs d'environ une demi-page. La session de rodage concernait trois textes courts, distribués un par un, le temps que les participants s'habituent au dispositif. Lors de la seconde session, ils ont été amenés à traduire vingt brèves du même genre, distribuées d'une traite cette fois et les laissant ainsi libres d'organiser leur travail. La troisième et dernière session était dédiée à la traduction de trois textes longs. Voici un exemple de texte court proposé à la traduction, l'intégralité des documents étant disponible dans les annexes :

Le chef des soldats rebelles du Timor oriental, Alfredo Reinado, a opposé une fin de non recevoir mercredi aux mesures présidentielles destinées à rétablir le calme dans le pays, rendant ainsi peu probable un dénouement rapide d'une crise qui perdure depuis plusieurs semaines.

Les enregistrements se sont déroulés sur cinq jours, pour une durée totale de seize heures de séquences, à l'aide de deux caméras. Chacune d'entre elle est positionnée de sorte à ce que son champ couvre un traducteur de face, ainsi que l'écran d'ordinateur de son collègue, comme illustré dans les figures 17a et 17b. Lorsqu'ils étaient prêts à produire la traduction finale, le résultat a été filmé avec les mêmes caméras. Ils pouvaient diviser la prestation en plusieurs prises, mais aucun « lissage » n'a été effectué par la suite au niveau du montage puisque le résultat final n'était pas l'intérêt principal ici.



Figure 17a Un binôme à l'œuvre

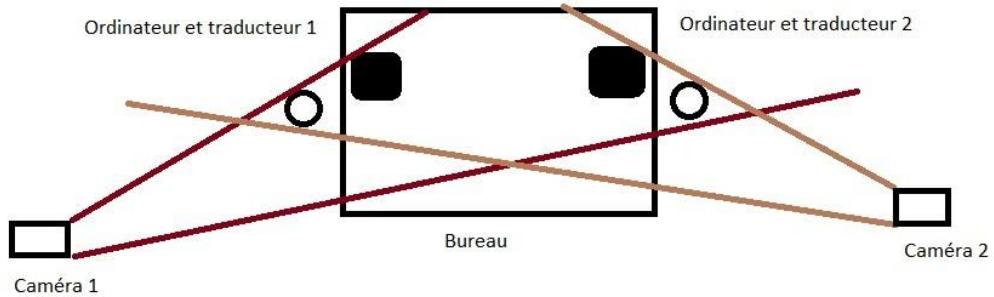


Figure 17b Schéma explicatif du positionnement des caméras

Pour chaque vidéo ainsi obtenue, nous avons annoté les tâches à l'aide du logiciel Elan¹⁵, qui est un logiciel d'annotation de fichiers multimédias. Il permet notamment de créer des pistes d'annotations et des repères temporels pour l'analyse de vidéos. Nous avons utilisé une seule piste. Le processus d'annotation s'est déroulé en deux temps. La première passe sur les vidéos nous a permis d'abord de répertorier toutes les activités observées : que font les traducteurs entre le moment où ils reçoivent le texte à traduire et le moment où ils annoncent la traduction terminée ? Les noms attribués à ces activités l'ont été de façon spontanée. Aussi, entre les deux passes, nous avons pris le temps de contrôler le vocabulaire utilisé, d'une part, et de clairement définir chaque tâche d'autre part, pour que leur identification soit non-ambigüe et facilitée. Lors de la seconde passe, nous avons identifié chaque activité avec le nom de la tâche correspondante, et pour chacune, les balises temporelles correspondantes de début et de fin.

Une fois les pistes combinées et extraites, le résultat est une liste de tâches identifiées avec leurs étiquettes temporelles, pour le processus de traduction de chaque texte. Nous avons ensuite additionné les durées passées pour chaque tâche, puis pour chaque traduction, et enfin, au total. Les annotations peuvent ainsi être comparées entre binômes mais aussi entre textes, afin de

¹⁵ <https://archive.mpi.nl/tla/elan>

déetecter les pratiques répétées et partagées, ainsi qu'un ordre total ou partiel dans le processus de traduction. Nous pouvons également comparer les tâches en fonction de la longueur et de la complexité du texte source.

Résultats

L'étude des traductions de textes courts livre la liste de tâches suivante :

- Recherche lexicale (lexical search) : le traducteur sollicite diverses ressources dans le but de déterminer le signe le plus adéquat pour un concept donné, incluant les noms propres.
- Discussion à propos des signes (discussing signs) : concernant des concepts dont la signification varie légèrement en fonction du contexte, les traducteurs discutent entre eux pour déterminer quel signe correspond le mieux à la situation donnée.
- Recherche de carte (map search) : lorsqu'il est nécessaire d'évoquer des emplacements géographiques relatifs, ou un lieu dont le nom-signe est inconnu des traducteurs même après recherche lexicale, ces traducteurs recherchent des cartes et des plans.
- Recherche de définition (definition look-up) : cela concerne la recherche de définitions de concepts en langue source qui ne sont pas clairs pour le traducteur. Cela peut également s'avérer utile pour trouver une façon de signer si aucun signe standard n'existe.
- Recherche encyclopédique (encycopedic look-up) : lorsque le texte source fait mention d'évènements réels ou de liens entre protagonistes qui sont inconnus du traducteur, ce dernier collecte les informations, les éléments de contexte nécessaires à la bonne compréhension du texte.

- Recherche d'images (picture search) : pour identifier les protagonistes cités dans le texte source, ou pour trouver une périphrase adaptée dans le cas de concept n'ayant pas de signe défini, la recherche d'image permet également de faciliter la description d'items, ou de trouver un nom-signe temporaire pour une personne ou une chose le temps de la traduction. Il s'agit de recherche des illustrations, des images ou photos correspondant à un élément du discours.
- Segmentation et travail sur l'ordre (segmenting and ordering) : cette tâche comprend la division du texte source en plus petites unités, et leur agencement dans un ordre qui convient à la langue des signes. De manière générale, les binômes se mettent d'accord sur les parties à segmenter, puis l'un d'entre eux propose un ordre. Cette tâche inclut également les discussions pour valider ou invalider les propositions, et le cas échéant, les nouvelles tentatives jusqu'à aboutir à un ordre qui satisfait les deux professionnels.
- Mémorisation (memorizing) : le traducteur s'entraîne seul à signer l'intégralité de la traduction, sans intervention de son binôme. Il s'agit de productions personnelles et informelles de ce qui deviendra le résultat final, répétées aussi longtemps que nécessaire jusqu'à mémorisation complète, de sorte à ce que le traducteur puisse procéder en consultant ses notes au minimum.

Un montage vidéo décrivant et illustrant chaque tâche est disponible dans les annexes vidéo.

Les tâches citées ci-avant peuvent être réalisées dans un ordre différent, à l'aide de ressources différentes ou sur une durée plus ou moins longue en fonction du texte à traduire ou du binôme de traducteurs observé. Il est

important de noter qu'elles ne sont pas toutes systématiquement réalisées comme le montre le tableau fig. 18. Il résume les résultats de l'annotation des vidéos en tâches. La colonne de gauche identifie les traductions, où **B** signifie *Beginners* pour le groupe débutant, et **E** signifie *Experts* pour le groupe d'expérience confirmée. La première colonne présente le code correspondant à chaque traduction étudiée : TR signifie translation, le premier chiffre (de 1 à 3) représente le numéro du texte traduit, et le second (de 1 à 2) correspond au binôme l'ayant réalisée. La première ligne du tableau recense toutes les tâches observées durant l'analyse, que nous avons détaillé ci-dessus. Si l'intersection des deux est colorée en vert, cela signifie que cette tâche a été observée au cours de cette traduction, le temps indiqué correspondant au temps passé au total sur cette tâche, ainsi que le nombre d'occurrences. À l'inverse, une case grise signifie que la tâche en question n'a pas été observée pour ce texte. La seconde colonne indique le temps total nécessaire pour achever la traduction, et la dernière ligne le temps total alloué à chaque tâche pour la totalité des six traductions. Ce tableau concerne les trois mêmes textes courts, pour les deux binômes de traducteurs (traductions disponibles en annexes vidéos).

	Transl. time	Lexical search	Discuss signs	Map search	Def. search	Ency whole search	Picture search	Seg & order	Memory
TR 1-1 (B)	1h13	28 min 45 (10)	15 min 52 (13)	1 min 15 (1)		10 min 01 (4)	58 sec (1)	16 min 54 (4)	
TR 2-1 (B)	59 min	17 min 20 (6)	7 min 47 (3)			7 min 44 (1)		26 min 36 (11)	
TR 3-1 (B)	12 min	2 min 56 (2)	2 min 47 (3)	56 sec (1)				5 min 28 (3)	
TR 1-2 (E)	19 min		6 min 58 (6)	3 min 4 (3)			57 sec (1)	7 min 7 (6)	40 sec (1)
TR 2-2 (E)	14 min		4 min 56 (4)		1 min 59 (1)			6 min 29 (3)	44 sec (1)
TR 3-2 (E)	11 min			7 min 2 (4)				4 min 09 (4)	13 sec (1)
Total time	3h08	49 min	37 min	13 min	2 min	18 min	2 min	66 min	2 min

Figure 18 Tableau des tâches et durées

On remarque dans un premier temps que la tâche de segmentation et d'ordre est systématiquement observée durant les six traductions étudiées ici. C'est également celle qui prend le plus de temps au total. Les traducteurs passent donc systématiquement du temps à retravailler l'ordre du texte source pour le traduire. Concernant ces données, avec une durée totale de 66 minutes, la tâche de segmentation et d'ordre représente presque un tiers de la durée totale des traductions analysées. Cet élément tend à démontrer que le principe de linéarité ne peut être supposé lorsqu'il est question de traduire vers les langues signées, ce qui constitue un challenge dans l'adaptation et l'élaboration d'un logiciel de traduction assistée par ordinateur pour ces langues. Autre remarque : la durée de la tâche de recherche lexicale est biaisée par la première traduction du groupe débutant (binôme 1), qui passe près de 40 minutes sur la recherche d'un seul signe. Cette observation nous a amené à repenser la méthode d'analyse des résultats obtenus, et à considérer les durées de chaque tâche ainsi que leurs nombres d'occurrences au lieu d'une simple durée totale, pour plus de pertinence.

Les observations précédentes concernent des textes courts. Nous avons ensuite appliqué le même protocole pour l'analyse de textes longs. Deux textes longs, disponibles en annexe, ont été soumis à chaque binôme. Il s'agit là encore de textes de style journalistique, d'environ une demi-page de longueur. Ces derniers exposent des informations plus denses, impliquant plus de protagonistes, qui entretiennent entre eux des relations complexes. Ils font également référence à des événements passés dont la connaissance est requise à la bonne compréhension du texte. L'analyse des traductions des textes longs a permis d'identifier des tâches déjà repérées pour les textes courts : recherche lexicale, recherche encyclopédique, discussion à propos des signes et segmentation et travail sur l'ordre.

Cependant, le processus de traduction pour des textes plus longs et a fortiori ici plus complexes diffère de la traduction de textes courts sur plusieurs points, que nous décrivons ci-après :

- Repérage dans le texte (text marking) : le traducteur souligne ou surligne des éléments du texte, soit pour ensuite procéder à des recherches plus approfondies à ce sujet, soit pour les garder en mémoire en tant qu'éléments importants du message.
- Compréhension du texte (text comprehension) : puisque le texte source est cette fois plus long et plus complexe, les traducteurs prennent le temps d'une lecture plus posée et plus en détails, en signant parfois certains passages (pour eux-mêmes ou l'un à l'autre) pour s'assurer d'avoir compris le même message, dans les mêmes degrés de nuances. Si ce n'est pas le cas, la tâche évolue en « discuss meaning ».
- Discussion sur la signification (discuss meaning) : lorsque le message ou l'information paraît ambiguë, les traducteurs prennent le temps d'en discuter pour éclaircir les doutes. Cette tâche est souvent associée ou entrecoupée d'épisodes de recherche encyclopédique.
- Discussion sur la traduction (discuss translation) : les traducteurs discutent les choix de traduction (par exemple, est-il préférable d'épeler le nom de telle personne, ou faut-il lui attribuer un nom-signe temporaire), ou discutent à propos de la formulation de la langue cible. L'objectif ici est de trouver la meilleure forme pour transmettre le sens, de trouver la meilleure façon de rendre l'information la plus claire, complète, et précise possible.
- Traduction (translation) : cette tâche s'apparente à une période de

répétition. L'un des traducteurs se filme tandis qu'il signe la traduction pendant que son binôme l'observe. Ces enregistrements intermédiaires et incomplets servent de supports pour les discussions citées ci-dessus, mais également comme exercice de mémorisation. Certaines vidéos produites lors de cette tâche deviendront le résultat final attendu dans le cadre de cette expérience, puisqu'il avait été indiqué aux participants que la post-édition de la vidéo n'était pas nécessaire.

Le tableau figure 19 résume les résultats des analyses pour les textes longs. Le tableau des textes courts présentait les durées totales pour chaque tâche et chaque traduction, mais ne permettait pas de distinguer la prépondérance et la répartition de chaque tâche pour telle ou telle traduction. Cette fois, le tableau présente le nombre d'occurrence pour chaque tâche durant la traduction d'un texte long par le binôme expert (binôme 2, traduction disponible dans les annexes vidéos), de même que le temps total pour cette tâche, ainsi que les durées minimales et maximales observées pour cette tâche.

Tâche	Occurrences	Durée min.	Durée max.	Durée totale
Lexical search	2	11s	30s	41s
Discuss signs	3	13s	19s	51s
Text marking	6	7s	1min10s	4min36s
Discuss meaning	4	17s	3min14s	4min58s
Encyclopedic look-up	9	11s	1min48s	7min48s
Text comprehension	10	5s	3min22s	9min37s

Discuss translation	12	47s	5min34s	26min11s
Translation	22	34s	3min37s	34min14s
Segm. & order	17	24s	4min55s	36min35s

Figure 19 Tableau des tâches pour les textes longs (binôme 2)

On peut directement observer que la tâche de segmentation et d'ordre est celle qui prend le plus de temps, avec un total de 36 minutes et 35 secondes pour 17 occurrences, comme c'était déjà le cas pour les textes plus courts. La deuxième tâche la plus longue est de filmer les traductions (22 occurrences pour un total de 34 minutes 14 secondes). La durée passée sur cette tâche semble dépendre de la longueur et de la complexité du texte source, et du nombre d'essais nécessaires pour parvenir à la version finale de la traduction. Sur une plus grosse quantité de données, le fait d'obtenir une durée normalisée pourrait avoir un sens vis-à-vis du temps total de travail. Ici encore, les tâches ne sont pas exécutées dans un ordre linéaire fixe, mais néanmoins avec une certaine précédence. Dans un premier temps vient la compréhension du texte source, avec lecture et repérage, éventuellement discussion pour clarifier les incertitudes de sens. Ensuite viennent les tâches de recherche, et de segmentation. Une fois la segmentation terminée, les traducteurs travaillent sur l'ordre une première fois, et s'en suivent plusieurs itérations entre travail sur l'ordre, traduction et discussions. Plus le processus de traduction avance, plus les tâches sont centrées autour des essais vidéos et de la discussion concernant la traduction. Si le traducteur se repose sur les notes prises durant toute la durée de l'exercice pour les vidéos intermédiaires, la production finale elle, est intégralement signée de mémoire, pour les textes courts comme pour les textes longs. Le résultat final dure en moyenne entre 20 et 30 secondes pour les textes courts, et entre 2 minutes et 2 minutes 30 pour les textes longs.

Ces résultats apportent des réponses aux questions que nous avions soulevées au chapitre 2. En effet, nous sommes à présent en mesure de mieux formaliser le processus de traduction vers la LSF. Nous avons pu clairement identifier les tâches impliquées, et la quantité de travail qu'elles représentaient au regard de l'intégralité d'une traduction. C'est le cas pour la segmentation par exemple, qui permet par la même de remettre en question le principe de linéarité en ce qui concerne les LS.

Les résultats de cette expérience sont des constats objectifs de traducteurs à l'œuvre. La seconde expérience préliminaire, décrite dans la partie suivante, est dédiée à la récolte d'informations et de retours subjectifs sur la pratique professionnelle que nous pourrons ensuite comparer et mettre en corrélation avec les résultats de la première expérience. (Kaczmarek & Filhol, 2020).

3.1.2 Brainstorming

Le brainstorming avait pour but d'amener les professionnels à réfléchir et verbaliser leurs pratiques, leurs besoins et les obstacles qu'ils rencontrent le plus fréquemment de sorte à pouvoir envisager des solutions. Au travers de la corrélation entre résultats objectifs et résultats subjectifs, nous cherchons à dresser une description plus claire et plus précise de ce qu'implique réellement le processus de traduction. Lorsque l'on interroge un professionnel sur son activité, il se peut qu'il fasse mention de tâches qui ne seront pas observées en réalité. À l'inverse, l'étude vidéo de son activité pourrait révéler des tâches qu'il aurait omises lors de l'entretien. Cette expérience de brainstorming a été réalisée avec la participation d'interprètes en langue des signes française. Il y a deux arguments en cette faveur : dans un premier temps, les traducteurs professionnels demeurent rares en France, et souvent les interprètes sont amenés à assumer la tâche de traduction lorsqu'aucun traducteur n'est disponible. D'autre part, lorsqu'un interprète doit se préparer pour l'interprétation d'un événement majeur planifié, ce

dernier demande en amont la version écrite des discours qu'il devra interpréter, et procède ensuite comme un traducteur travaillerait sur son texte source. Sans dire que le travail est le même, il comporte suffisamment de points communs pour que la participation d'interprètes au brainstorming soit pertinente. Des traducteurs professionnels sourds étaient bien sûr conviés, mais ils n'ont malheureusement pas pu se rendre disponibles.

Méthodologie

La session de brainstorming concernait six interprètes professionnels d'expérience variable, qui travaillent pour le même service d'interprétation. Elle s'est étalée sur une durée de deux heures trente, sous notre supervision, pour la prise de notes ainsi que pour la modération des échanges entre les participants. Les questions étaient proposées par l'intermédiaire d'un diaporama projeté dans la salle, toujours accessible. Les participants se sont vu remettre de quoi écrire, stylos de couleurs et pense-bêtes adhésifs sur lesquels ils étaient libres d'inscrire des idées, à partager et à débattre au fil de la séance. Cette dernière a été structurée en trois parties.

Dans un premier temps, nous avons demandé aux participants de réfléchir quant aux outils numériques et informatiques qui pourraient faciliter le travail d'un traducteur français-LSF (« quels outils numériques pour assister le traducteur français/LSF dans sa tâche ? »). La discussion était orientée sur l'aide logicielle, et ce qu'ils se figuraient que cette dernière puisse ajouter à leur métier. Ces réflexions sont menées d'entrée de jeu, sans tenir compte des connaissances des participants en termes de TAO. Les idées inscrites sur les pense-bêtes étaient ensuite affichées sur un tableau.

Une fois ce premier tour terminé, nous leur avons proposé une courte introduction aux concepts de la TAO. Cette présentation abordait les fonctionnalités logicielles et les outils à disposition des traducteurs, ainsi que la façon dont ils leur permettent de travailler plus rapidement et plus

efficacement. À la fin de cette introduction, nous avons affiché de nouveau le tableau des idées inscrites plus tôt, de sorte à pouvoir les comparer avec les éléments abordés lors de la présentation. La seconde question abordée était : « quels sont les besoins spécifiques à la langue des signes en termes de traduction ? » Les participants étaient alors encouragés à discuter et ajouter librement de nouvelles idées à la liste de fonctionnalités et d'outils établie lors de la première partie du brainstorming. À la suite de cette présentation, ils étaient à même de comparer ce qui existe déjà dans le domaine de la TAO texte à texte et ce qu'ils pensent nécessaire voire indispensable pour outiller la traduction de la LSF, mais ils étaient également à même de réfléchir sur des méthodes et des façons de les matérialiser. Le brainstorming s'est terminé sur un temps d'échange, une discussion libre sur le sujet, durant laquelle chacun était libre de poser plus de questions, d'exprimer son point de vue ou de verbaliser des problèmes qui ne leur étaient pas venus à l'esprit plus tôt. Le tableau figure 20 suivant récapitule les questions posées par nos soins durant la séance.

Questions posées avant l'introduction à la TAO	Questions posées après la présentation à la TAO
Comment pensez-vous qu'un logiciel ou qu'un ordinateur puisse vous aider ? (dans vos pratiques professionnelles)	Quels sont les spécificités et les besoins spécifiques à la langue des signes en termes de traduction ?
Selon vous, quelles fonctionnalités pourraient s'avérer utiles ?	Pouvez-vous donner des exemples précis de fonctionnalités ou d'outils ? Quels outils pour quels besoins ?
Vous imaginez-vous travailler avec pareils outils ? Pourquoi ?	Selon vous, quelle serait la meilleure façon d'adapter les fonctionnalités de la TAO actuelle à la traduction des LS ?

Figure 20 Liste des questions posées durant le brainstorming

Pour toute la durée de la session de brainstorming, nous avons pris des notes des points soulevés par les participants. Nous avons également conservé le tableau d'idées (Fig. 21).

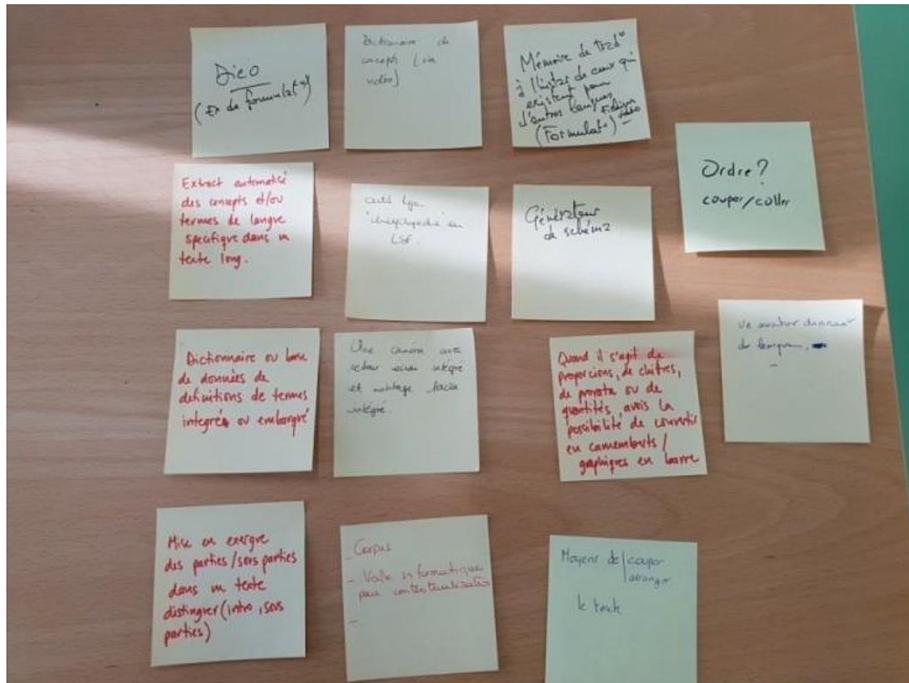


Figure 21 Le tableau de brainstorming

Les résultats exposés ci-après sont issus des pense-bête récoltés en fin de séance, ainsi que des notes prises durant les échanges et les débats. Chaque élément mentionné est accompagné autant que faire se peut d'un élément de contexte ou d'une définition pour la compréhension.

Résultats

Au cours des différents échanges, les participants ont spontanément évoqué les outils et ressources qu'ils utilisent à l'heure actuelle dans leurs pratiques, que nous listons ci-après :

- Les images servent de référence pour l'attribution de signes temporaires, la création de périphrase, ou lorsqu'une description

détaillée est nécessaire. C'est le cas par exemple lors de la traduction de domaines de spécialités, ou de jargon d'entreprise (Harvey, 2002). Elles peuvent également faciliter la compréhension d'un élément à traduire qui est inconnu au traducteur.

- Les dictionnaires (dans la langue source) sont utilisés pour affiner la compréhension du texte source (définitions, nuances).
- Les dictionnaires en LS permettent de retrouver un signe, ou de s'inspirer de sa définition en LSF pour le paraphraser.
- Concernant les idées d'outils ou d'assistances qui pourraient profiter aux traducteurs, nous listons ci-après les remarques faites sur les pense-bête :
 - Des dictionnaires en LS, qui proposeraient en plus des exemples de formulation et d'emploi des éléments recherchés. Il serait question d'une base de données de définitions en LS, avec un lexique plus vaste et plus pointu, ainsi que la possibilité de trier le contenu par thématiques.
 - Certaines fonctions de traitement automatique des langues, comme par exemple la reconnaissance d'entités nommées, à savoir le repérage automatique de noms propres : personnes, lieux, entreprises... Les professionnels font également mention d'automatisation pour le repérage numéral, pour les dates ou les pourcentages, avec une possibilité de générer automatiquement des représentations graphiques des pourcentages rassemblés dans un texte (diagrammes par exemple).
 - Une assistance à la segmentation (mentionnée même avant l'introduction à la TAO). La proposition initiale était celle

d'un outil qui permette d'extraire et de mettre en emphase certaine partie du texte source, ou qui puisse faciliter la découpe des paragraphes pré-existants pour en assembler de nouveaux.

- Une facilitation de la capture et du montage vidéo.
- Un outil qui permette de créer facilement des schémas.
- Un équivalent d'une mémoire de traduction qui puisse fonctionner avec les LS, et qui permettent de retrouver des exemples de formulations.

Une fois la présentation TAO terminée, avec notamment la mention de la segmentation automatique du texte par le logiciel, les professionnels sont tombés d'accord sur l'utilité d'un outil similaire pour les LS, à la condition qu'ici, les unités découpées puissent être librement réorganisées. Les traducteurs ont également fait mention de ce qu'ils ont spontanément appelé des exemples de « comment le dire ». Cette fonction est celle d'un concordancier, à savoir des exemples de traduction en contexte, qui permettent de trouver des formulations ou des exemples propres aux LS. Autre mention semblable, celle d'un recueil d'exemples vidéos de traductions passées : il n'est pas question ici de simples exemples de formulations en langue des signes, mais bien de pouvoir consulter des travaux antérieurs (à la fois le texte source et la façon dont il a été traduit). Il s'agit de l'expression d'un besoin pour un outil semblable à une mémoire de traduction. Il a été fait mention de cette dernière avant la présentation TAO, par l'un des participants qui s'est avéré être un ancien traducteur texte à texte français-anglais. Une fois la présentation terminée, et le concept de mémoire de traduction exploré plus avant, tous se sont accordés sur l'utilité d'un tel outil comme sur le bénéfice qu'il pourrait apporter à la traduction des LS s'il était adapté. Cependant, personne n'a suggéré de piste quant à sa

réalisation.

D'une manière générale, les participants ont souligné l'importance des ressources en LS, qu'elles devraient être plus nombreuses, plus accessibles également, et mieux référencées. Selon nos participants, ces dernières sont rares, et souvent mal organisées, ce qui rend leur consultation laborieuse et parfois infructueuse. D'autre part, la capture vidéo ainsi que le dessin ou la schématisation (représenter les informations du texte au moyen de schémas ou de dessins) sont considérés comme des méthodes de prise de notes, ou des façons de réfléchir sur la traduction en elle-même. Le schéma permet de poser visuellement l'organisation du discours, et les captures vidéo intermédiaires servent de brouillons.

La majorité des réflexions et des idées suggérées tendent à se rapprocher d'outils déjà existant en TAO, tels que des glossaires, des concordanciers, ou une mémoire de traduction. Les deux expériences que nous venons de présenter se sont déroulées avec le concours de professionnels différents, pourtant les méthodes énoncées et observées semblent cohérentes. La section suivante met en relation les résultats de ces deux protocoles, et s'attèle à brosser un portrait plus précis du processus de traduction en LS.

3.1.3 Le processus de traduction

Synthèse des résultats

Les résultats recueillis lors des deux expériences sont cohérents entre eux, mais les deux séries de résultats présentent également des contrastes. La première nous a permis d'identifier des pratiques dont les participants au brainstorming n'ont pas fait mention. C'est le cas par exemple de la recherche de cartes, de la recherche encyclopédique pour du contexte ou encore les tâches de mémorisation ou de répétition de la traduction. De même, le brainstorming nous a permis de prendre connaissance de

problèmes qui ne sont pas apparus lors de la première expérience : difficultés d'accès aux ressources en LS, besoin de consultation de travaux antérieurs (produit par d'autres ou soi-même).

L'étude des traducteurs à l'œuvre ne nous a pas permis d'identifier un ordre strict des tâches dans le processus de traduction. Celui-ci semble dépendre du texte source lui-même, mais également de la façon personnelle à chaque traducteur d'aborder le travail. Cependant, nous avons pu identifier une forme de précédence dans l'exécution des tâches. Dans un premier temps, le traducteur prend connaissance du texte qu'il doit traduire. Cette première étape inclut les tâches de lecture et de repérage dans le texte. Il s'agit ici de comprendre le sens au niveau global d'abord, puis d'affiner la compréhension jusqu'aux plus fines nuances pour rester le plus fidèle possible au sens du document source. Les tâches de recherches (lexicale, encyclopédique, pour des cartes ou illustrations, ainsi que pour des définitions) apparaissent dans un second temps, pour pallier les éventuels manques d'information nécessaire à cette compréhension fine.

Une fois que le texte est pleinement compris et intégré par le traducteur, ce dernier va travailler sur l'ordre de présentation des informations, via la tâche de segmentation et d'ordre. Cette tâche sera répétée autant de fois que nécessaire jusqu'à obtention d'un résultat jugé satisfaisant. L'étape suivante consiste en la production de la traduction : essais vidéos, mémorisation, et production finale. Les tâches dites de discussion (sur le sens, les signes, les choix de traduction) peuvent s'intercaler sporadiquement au cours des autres tâches.

Il apparaît également que toutes les tâches ne sont pas systématiques. On observe cependant les mêmes tâches pour un même texte, ce qui montre une influence du texte source sur le processus lui-même. Par exemple, durant la traduction du premier texte (fig. 18), les deux binômes ont recherché des

cartes et des images. Idem pour le troisième texte, où les deux groupes ont recherché une carte. La longueur et la complexité du sujet jouent également un rôle : nous avons identifié de nouvelles tâches lorsqu'il était question de traduire des textes plus longs. La traduction de textes brefs peut se résumer en une succession de tâches utilitaires : recherches, travail sur l'ordre, production. La traduction de textes plus longs intègre des tâches de réflexion et de recherches plus poussées, ainsi qu'une tâche de traduction à part entière, qui est la seconde tâche la plus longue derrière celle de segmentation et d'ordre. Les deux expériences soulignent l'importance de la tâche de segmentation et d'ordre : c'est la seule tâche systématique, qui apparaît aussi bien dans la traduction de textes courts que dans la traduction de textes longs, mais également la tâche sur laquelle les traducteurs passent le plus de temps au total.

Les preuves produites ici concernant la non-application aux LS du principe de linéarité (travail systématique et de longue durée sur la réordonnance du texte source) souligne la non-pertinence de la segmentation automatique du texte source lorsqu'il est question de LS (plusieurs essais d'ordre souvent réalisés). Si une assistance à la segmentation est sollicitée par les professionnels, et apparait comme nécessaire en regard des résultats ci-avant, il est important que cette segmentation puisse s'effectuer librement par le traducteur.

Compte tenu de la différence d'expérience professionnelle entre les deux binômes lors de la première expérience, nous ne ferons que quelques observations à ce sujet. Dans un premier temps, l'expérience semble avoir une influence sur les tâches mises en place. Seul le groupe 1 (débutant) procède à des recherches lexicales et encyclopédiques, là où seul le groupe 2 (expert) ajoute une tâche de mémorisation (fig. 18). Le temps passé sur chaque traduction représente une autre différence notable entre les deux

binômes (1h13 pour le premier texte chez les débutants contre 19 minutes pour les experts).

Discussion et limites

Ce jeu de données est trop maigre pour poser une description définitive et formelle du processus de traduction des LS, et se base sur trop peu de professionnels différents pour pouvoir généraliser. En outre, les textes proposés à la traduction sont tous de type journalistique. Les expériences présentées ci-avant méritent d'être réitérées avec plus de participants et des types de textes différents pour affiner les observations et en conséquence, les spécifications. Les traductions étudiées ici sont le produit de collaboration entre deux ou plusieurs professionnels. Les observations faites sont basées sur des discussions, aussi serait-il intéressant de comparer ces résultats avec d'autres obtenus lorsque le traducteur travaille seul et qu'il ne peut pas se baser sur son collègue pour obtenir un retour.

Cependant, ils constituent une base pour définir de premières spécifications pour un logiciel de TAO à destination des LS, en se référant aux pratiques observées et relatées durant les expériences. Ils nous guident également quant à une première liste de fonctionnalités qu'un tel logiciel pourrait proposer :

- Un outil d'aide à la segmentation apparaît nécessaire.
- Le fait que la tâche de traduction elle-même soit la plus courante, ajouté aux différents essais repérés pour chaque traduction, implique un besoin d'interface flexible qui permettrait aux utilisateurs de modifier indéfiniment le projet sans avoir à repartir de zéro.
- L'intérêt pour une mémoire de traduction étant mentionné à plusieurs reprises, il conviendrait d'envisager un outil similaire.

L'intégration de pareille fonction dans un logiciel soulève un nouveau défi technique, puisque le principe de linéarité n'étant pas respecté, les segments sources et cibles qui composent normalement les alignements d'une mémoire de traduction ne pourraient ici correspondre, de même que les alignements ne pourraient être enregistrés automatiquement du fait de la segmentation non-automatique. L'alignement nécessiterait l'intervention du traducteur lui-même.

Faisant suite à ces deux expériences, la prochaine étape est de dresser une liste de critères que notre logiciel se devra de respecter. Pour ce faire, nous reprenons les besoins et les problèmes soulevés lors des études, pour associer chacun d'entre eux à une ou plusieurs solutions envisagées.

Enfin, seule la traduction du texte vers la LSF a été étudiée. Ces données pourraient être complétées par une étude de l'autre sens de traduction.

3.2 Cahier des charges et propositions.

Le cahier des charges que nous présentons ici s'articule autour des trois points clés de la TAO identifiés au chapitre 2, et qui représentent les trois défis majeurs dans la création d'un logiciel d'assistance à la traduction des LS. À chacun est associé une liste de spécifications, ainsi que les fonctionnalités envisagées pour y répondre, l'objectif étant de permettre aux utilisateurs de gagner en temps et en confort.

Le logiciel devra avoir pour interface une présentation classique, à savoir un affichage côté à côté avec le texte source à gauche, et la zone de travail du traducteur à droite. La traduction en LS étant produite d'une traite, et non au fur et à mesure comme la traduction texte à texte, l'interface doit également être pensée en ce sens et permettre d'assister le travail préparatoire sur le

texte. La section 3.2.1 présente la notion de « Blocs », que nous introduisons ici comme fonctionnalité majeur de notre prototype, et la section 3.2.2 détaille les autres fonctionnalités envisagées.

3.2.1 Non-linéarité assistée par des « Blocs »

Le principe de linéarité admet que la concaténation des segments cibles traduits équivaut à la traduction des segments sources concaténés. Il est assisté par la segmentation automatique intégrée dans les logiciels de TAO existants. Ce principe n'étant, nous l'avons vu, pas admis en LSF, il apparaît inutile d'envisager une fonction de segmentation automatique pour le texte source. En revanche, comme la tâche de segmentation et d'ordre est non seulement la seule à être systématique, mais également celle qui prend le plus de temps au traducteur, le travail sur l'ordre (à savoir découper le texte et replacer les informations dans l'ordre propre de la LSF) doit être assisté d'une autre façon (Kaczmarek & Larroque, 2021).

Segmentation libre : la notion de Bloc

Le recours à la segmentation dans le cadre de la traduction texte à texte, et dans le cadre de la traduction des LS, n'a pas le même but. Dans le premier cas, il s'agit de diviser le texte en unités plus courtes, dans l'objectif peut-être ergonomique d'une part, de faciliter la lisibilité de l'affichage pour le traducteur, mais surtout pour permettre d'alimenter la mémoire de traduction. Cette segmentation affranchit d'une certaine manière le traducteur du contexte : il traduit des phrases désolidarisées du texte dans lequel elles sont initialement produites. En revanche, pour ce qui est de la segmentation dans le cadre des LS, l'objectif est de pouvoir réagencer les informations en faveur de l'organisation du discours propre à la langue, et du contexte. Le choix de la découpe nécessite une compréhension fine du discours, et ne peut se voir automatisée ici. C'est pourquoi la segmentation doit être libre : c'est au traducteur qu'il revient de créer autant de parties

qu'il le souhaite, et de les délimiter à sa convenance.

Pour ce faire, nous proposons la notion de **Bloc**, qui va pouvoir être manipulé de manière libre. Ces Blocs représentent chacun une portion plus ou moins longue du discours cible.

Contenu des Blocs

Au vu des divers types d'informations manipulés par les participants durant les expérimentations, une autre proposition consiste à pouvoir utiliser ces Blocs pour d'autres types d'éléments utiles au processus de traduction (des illustrations, des vidéos). Une fois créés, les Blocs doivent être remplis avec de l'information, et peuvent donc accueillir différents types de contenu de sources diverses :

- Du texte, dans un premier temps, qu'il s'agisse d'un copié/collé ou d'une entrée au clavier. Il peut s'agir d'informations prélevées dans un site extérieur, d'extraits du texte source ou de notes prises par le traducteur.
- De vidéos ensuite, qu'elles proviennent elles aussi d'un site internet, ou qu'elles soient importées depuis l'ordinateur de l'utilisateur. Elles peuvent également être filmées directement depuis l'application, voire depuis le Bloc.
- Des éléments d'illustration, tels que des cartes et images, des photographies ou des schémas, qui peuvent être soit prélevés sur internet, soit importés également depuis l'ordinateur de l'utilisateur.
- Des schémas pourraient également être dessinés directement dans l'application par le biais d'un outil dédié.

Organisation temporelle : l'axe vertical

Le traducteur pourra créer autant de Blocs que nécessaire et les agencer librement, mais l'axe vertical dans l'interface aura un rôle particulier : il sera utilisé pour organiser les Blocs temporellement, le temps s'écoulant du haut vers le bas. Cet axe est un repère quant au déroulé temporel de la production en LS.

Hiérarchie : la structuration des Blocs

Afin de rendre lisible la manière dont les Blocs sont utilisés dans le processus de traduction, nous proposons un mécanisme de hiérarchisation des Blocs (Fig. 22). Chaque Bloc peut être subdivisé en autant de Blocs enfants que désiré. Les Blocs enfants peuvent être utilisés comme une façon de compléter le contenu de son parent, en ajoutant un niveau de détails supplémentaire, par exemple une illustration en lien avec le texte source contenu dans le Bloc initial. Ainsi, un Bloc enfant dépend toujours d'un autre Bloc. La relation entre un Bloc enfant et un Bloc parent est une relation de partition : l'intersection des Blocs enfants d'un même parent est vide, et leur union représente l'intégralité du Bloc parent. Un Bloc enfant peut lui-même générer un Bloc enfant, créant ainsi une structure arborescente, où chaque enfant dépend de son parent et représente une sous-portion de son parent. Ces Blocs sont ordonnés temporellement. Une illustration du principe est donnée figure 22, où chaque colonne correspond à un niveau de détail du moins fin (à gauche) au plus fin (à droite).

Le but de cette représentation est de privilégier le visuel et donc la lisibilité de l'interface en évitant la création de Blocs non structurés.

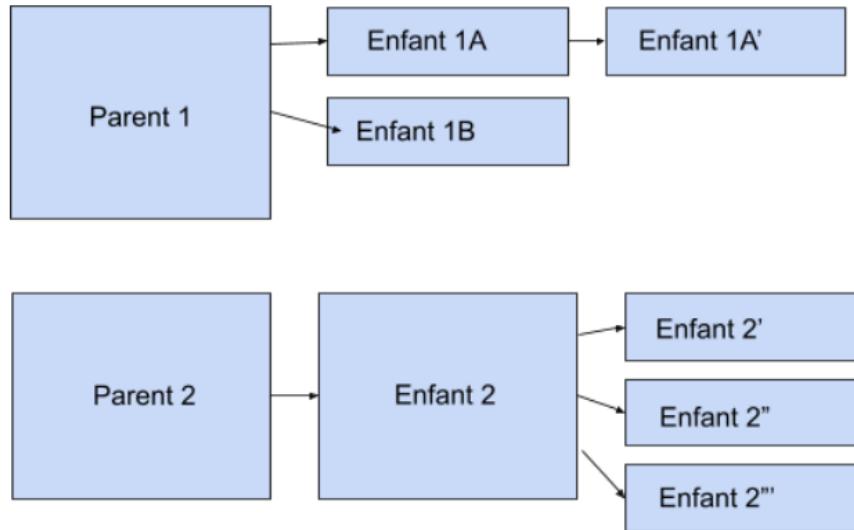


Figure 22 Schéma explicatif de la hiérarchie des Blocs. L'exemple comporte deux hiérarchies de Blocs (axe vertical), chacun ayant trois niveaux de hiérarchie (axe horizontal).

Flexibilité : la mobilité des Blocs

Nous avons remarqué que l'ordre des informations dans la traduction finale était le fruit de plusieurs essais, et qu'il était donc sujet à modifications. Pour permettre aux traducteurs de gagner du temps, et de ne pas avoir à repartir de zéro à chaque nouvelle tentative, notre proposition est que les Blocs constitutifs de notre interface soient mobiles. Ils doivent pouvoir être déplacés, sans altération de leur contenu. Lors du déplacement, un Bloc parent entraîne logiquement avec lui ses enfants. Par ailleurs, si un Bloc parent possède plusieurs enfants, alors ces derniers peuvent-être réorganisés entre eux, entraînant alors eux-mêmes leurs enfants.

Reprendons pour exemple la configuration de la figure 22. La mobilité des Blocs peut opérer à la verticale, ou à l'horizontale. Si la première tentative d'ordre réalisée par le traducteur ne lui convient pas, il peut :

- Échanger les Parents 1 et 2, pour placer temporellement la lignée 2 avant la lignée 1, auquel cas les deux lignes entières s'échangent.



- Échanger les Enfants 2', 2'' et 2''' entre eux sans rien impacter d'autre.
- Échanger les Enfants 1A et 1B, le Bloc Enfant 1A' descend alors d'une ligne en suivant Enfant 1A.
- Déplacer le Bloc Enfant 2 à la ligne d'en dessous, créant alors un Bloc Parent 3 et où Enfant 2', 2'' et 2''' deviennent respectivement Enfant 3 A, B et C.
- Déplacer le Bloc Enfant 2 comme enfant du Bloc Enfant 1B. Il devient alors Enfant 1B', et les Blocs Enfants 2', 2'' et 2''' deviennent des Blocs Enfants 1B'1, 2, et 3.

3.2.2 Autres fonctionnalités

D'autres fonctionnalités peuvent être intégrées pour permettre de gérer les autres composantes de la traduction, comme le texte source, ou l'accès à différentes ressources.

Gestion du texte source

Le logiciel doit offrir la possibilité de charger et d'afficher le texte sur lequel travaille le traducteur. Il doit pouvoir prendre en charge les formats les plus communs, ainsi qu'en conserver la mise en page. En réponse aux suggestions émises lors du brainstorming quant à la possibilité de pré-traiter le texte, l'application doit proposer des fonctionnalités de traitement automatique, telles que la reconnaissance de dates ainsi que la reconnaissance d'entités nommées : à l'activation d'une fonction, ces données pourraient apparaître en évidence dans le corps du texte, comme les traducteurs surlignent un passage dans le texte.

Modules de ressources

Pour permettre aux traducteurs de gagner du temps lors de leurs recherches,

et pour faciliter le transfert de contenu entre un site extérieur et un Bloc de traduction, l'application doit accueillir un volet d'accès à différentes ressources. Chaque onglet correspond un type de recherche : lexicale, encyclopédique, ainsi qu'une carte interactive. Chacun des onglets agrège les résultats de recherche au sein de plusieurs ressources différentes. Un historique de recherche global permet de rapidement réitérer une recherche antérieure.

Prompteur

Pour assister l'étape de production filmée, nous envisageons une fonction de type prompteur à intégrer à l'interface. Celle-ci permettrait au traducteur, une fois ses Blocs remplis et organisés à sa guise, d'en sélectionner plusieurs de sorte à les faire défiler selon un rythme et un mode de présentation sélectionné (défilement vertical ou diaporama...).

Mémoire de traduction

La mémoire de traduction est l'outil central de la TAO, aussi son adaptation aux LS est primordiale. L'absence de forme écrite pour les LS ne permet pas de reprendre l'outil tel quel, mais la fonctionnalité doit être conservée. La mémoire de traduction permet de réinvestir les travaux passés en suggérant à l'utilisateur des paires précédemment traduites. Il faut envisager une fonctionnalité qui permette à la fois de conserver le travail en cours et de le garder accessible en contexte pour des usages futurs. Un fonctionnement identique aux mémoires de traduction existantes semble complexe à atteindre de par le recours à la vidéo. En effet, bien que les requêtes concernent le texte (dans le cas de la traduction du français vers la LSF), les éléments vidéos suggérés par la MT ne peuvent être réutilisés directement dans la traduction en cours comme c'est le cas avec le texte.

Nous avons choisi d'adapter cette fonctionnalité sous la forme d'un concordancier bilingue français - LSF. Tout comme la MT, le concordancier

repose sur une base de données d'alignements. La différence est que ces alignements ne sont pas générés automatiquement au cours de la traduction, et que le traducteur doit faire la démarche de les consulter (le logiciel ne suggère pas automatiquement). Il permet cependant d'explorer la langue, et d'en consulter des exemples de traduction, comme suggéré par les participants du brainstorming.

La plupart des propositions décrites dans ce chapitre a été mises en œuvre au sein de prototypes qui sont présentés dans le chapitre suivant.

Chapitre 4

Réalisation des prototypes

La réalisation du cahier des charges présenté en chapitre 3 s'est déroulée en deux parties. Dans un premier temps, nous avons développé un concordancier bilingue français – LSF (section 4.1). C'est un outil autonome, élaboré en parallèle de l'environnement de travail intégré en lui-même, et qui est directement accessible mais aussi intégrable à l'environnement de travail. Dans un deuxième temps, l'environnement de travail a été développé (section 4.2). Les deux sections de ce chapitre détaillent l'implémentation de ces deux prototypes.

4.1 Concordancier bilingue français - LSF

Un concordancier est, quelles que soient les langues de travail, un moteur de recherche qui peut parcourir des corpus et lister toutes les occurrences d'une requête en contexte. Dans le domaine de la traduction, il s'agit de concordanciers bilingues, dont les alignements rassemblent deux segments qui sont la traduction l'un de l'autre pour une paire de langues données. La requête est effectuée dans une langue source et les résultats sont fournis avec une traduction alignée dans la langue cible. Ce type d'outil permet aux traducteurs de rechercher des mots ou des expressions dans leur contexte, afin de déterminer leur fréquence d'utilisation, ou le style de discours auquel ils sont le plus souvent associés. La section 4.1.1 décrit le processus de création de la base de données d'alignements du concordancier. La section 4.1.2 détaille son implémentation, et la section 4.1.3 fournit plusieurs exemples d'utilisation.

4.1.1 Base de données d'alignements

Nos deux langues de travail sont le français écrit et la LSF. Les alignements se composent donc d'une partie de texte, qui pourra être requêtée, et d'une portion de vidéo en LSF identifiée par des bornes temporelles. Aucune base de données préexistante ne correspondait aux besoins identifiés pour ce projet (corpus bilingues et alignés), nous avons donc créé notre propre échantillon pour équiper notre prototype. Cet échantillon repose cependant sur un corpus déjà existant.

Corpus utilisé

Nous avons utilisé un corpus parallèle bilingue existant, nommé « 40-brèves » (Filhol et Tannier, 2014), accessible en ligne sur la plateforme Ortolang¹⁶. Ce dernier se compose de quarante textes journalistiques courts d'environ trois à cinq lignes, et chacun est traduit par trois traducteurs

¹⁶ <https://www.ortolang.fr/market/corpora/40-breves/v1>

professionnels différents, pour un total de 120 vidéos. La figure 23 montre un exemple de brève, ainsi qu'une capture de la mise en scène vidéo. Le texte en français est le suivant :

Quelque 200 personnes pourraient avoir été tuées mercredi matin par un glissement de terrain provoqué par des pluies abondantes dans l'île indonésienne de Java, a annoncé une source policière.



Figure 23 Capture d'une vidéo du corpus

Anatomie d'un alignement

Par définition, un alignement implique que les deux segments qui le composent soient équivalents en sens (l'un étant la traduction de l'autre). Il peut s'agir de correspondance lexicale (un signe isolé aligné avec le mot ou l'expression auquel il correspond), ou d'éléments plus complexes et plus longs. Ces alignements ont dû être réalisés manuellement, faute de moyen d'automatisation existant.

Chaque vidéo dure environ 30 secondes, et chacune d'entre elles fournit déjà un alignement en soi puisque aligner des textes de trois lignes avec trente secondes de vidéo ne pose pas de problème technique. Les professionnels consultés ayant fait remonter que bon nombre des ressources

en LSF étaient difficilement consultables, des requêtes qui renverraient comme résultats des quarts d'heure entiers de vidéos n'auraient pas d'intérêt. Ainsi, pour permettre des recherches plus fines dans le concordancier, nous avons ensuite segmenté les textes sources en plus petites unités que nous avons appairés avec leurs traductions respectives. Le grain est varié, pour tirer profit des trois traductions existantes pour chaque texte, mais se veut tout de même relativement fin pour que le concordancier ne renvoie pas en résultats des vidéos qui durent plusieurs minutes, mais un extrait précis.

Les segments ont été sélectionnés pour l'absence de signes lexicaux correspondant spécifiquement au même contenu sémantique, ou par la variété de leur construction : figure de style, expressions idiomatiques ou construction propre à l'une ou l'autre des deux langues. Des segments plus aléatoires sont également inclus, pourvu qu'ils disposent d'un segment traduit identique. Seuls les segments convenables pour l'alignement ont été conservés, c'est-à-dire ceux pour lesquels le contenu était l'exact identique de celui de la vidéo, afin de créer des données parallèles convenablement alignées. Par exemple, on ne peut pas aligner un adverbe du texte avec seulement l'expression faciale du signeur, puisque la vidéo comporterait d'autres éléments linguistiques superposés, comme le mouvement des mains qui lui, ne serait pas couvert par le texte.

Pour contourner les difficultés liées à l'utilisation de la vidéo, nous avons rassemblé les caractéristiques de chaque segment en données textuelles, selon le format suivant :

<TxtID, start_pos, length, VidID, start_time, duration>

où *TxtId* correspond au code d'identification du texte, ce qui permet de les

retrouver dans la base de données. *start_pos* correspond à la position du premier caractère du segment dans le texte source (en ne considérant que du texte brut) et *Length* représente la longueur en nombre de caractères du segment. De la même façon, *VidID* correspond à l'identifiant de la vidéo, *start_time* représente le marqueur temporel qui correspond au début du segment vidéo, et *duration* représente la durée du segment vidéo en secondes. Pour ce faire, les textes sources ont été annotés avec le logiciel Brat, et les vidéos avec le logiciel Elan.

Annotation du texte avec Brat

Le logiciel Brat¹⁷ est un logiciel d'annotation de textes. Chaque segment sélectionné est associé à une clé d'identification unique, qui lorsque l'on exporte les données, est également caractérisée par les numéros du premier et dernier caractère du segment dans le texte source. La figure 24 montre un exemple d'annotation des segments de textes. La partie haute montre la découpe en segment d'une phrase du texte. La partie basse montre le format des données exportées, avec chaque segment identifié de T1 à T7, leurs positions de début et de fin dans le texte brut, ainsi qu'un rappel du contenu du segment.

Alignement		Alignement
un glissement de terrain provoqué par des pluies abondantes dans l'île indonésienne de Java,		
T1	Alignement 15 63	Quelque 200 personnes pourraient avoir été tuées
T2	Alignement 64 72	mercredi
T3	Alignement 83 142	un glissement de terrain provoqué par des pluies abondantes
T4	Alignement 148 153	l'île
T5	Alignement 170 174	Java
T6	Alignement 176 208	a annoncé à une source policière
T7	Alignement 121 142	des pluies abondantes

Figure 24 Capture de la segmentation dans Brat (en haut), et exemple de données exportées depuis Brat (en bas).

Nous avons segmenté et aligné au total 40 textes (les 40 brèves).

¹⁷ <https://brat.nlplab.org/>

Annotation des vidéos avec Elan

Pour l'annotation des vidéos, nous avons utilisé Elan. Pour chaque segment de texte identifié, sa traduction en LSF est localisée dans la vidéo. Ici, seule une piste est nécessaire, dans laquelle est créée une étiquette où l'on annote la clé d'identification du segment texte. La figure 25 montre un exemple d'annotation sous Elan.



Figure 25a Annotation avec ELAN (corpus 40-brèves)

Les données exportées depuis Elan rassemblent un ensemble de lignes qui associe à chaque clé d'identification des repères temporels de début et de fin dans une vidéo elle aussi identifiée.

0.05	5.11	T5
5.14	7.56	T6
7.66	9.69	T3
9.74	12.58	T1
12.61	14.93	T2
15.04	20.0	T4

Figure 25b Annotation avec ELAN (corpus 40-brèves)

La figure 25b représente les données exportées depuis Elan pour une vidéo (et donc un texte). La troisième colonne, qui comporte les mentions T1 à

T6, est celle des clés d'identification. La première correspond aux marqueurs temporels du début du segment, et la deuxième, ceux de fin. Notons que cela confirme encore le non respect du principe de linéarité dans le cas des LS: les clés d'identification sont attribuées au fur et à mesure, ce qui signifie que dans le texte, le segment T1 apparaît en premier, puis le segment T2 et ainsi de suite. Dans l'affichage des données Elan qui elles, apparaissent dans l'ordre de la vidéo, on remarque que les segments textes apparaissent dans un ordre nouveau.

Par la suite, nous avons rassemblés les données issues de ces deux passes d'annotation pour créer les alignements. Cette méthodologie a été appliquée pour la création des 600 premiers alignements, basés sur un seul corpus parallèle, les 40-brèves.

En sus de la liste des alignements, des vidéos et des textes, la base de données comporte plusieurs tables de métadonnées qui conservent des informations complémentaires concernant les textes et les vidéos qui constituent le corpus. Le but est de garder une trace accessible d'informations dont la mention pourrait être pertinente dans l'interface du concordancier, ou pour créer des filtres de recherche. Les tables de métadonnées (un pour les textes, un pour les vidéos) sont destinées à lister des informations telles que les auteurs, les informateurs (les personnes signant dans les vidéos), les sources, les titres originaux ou encore les thèmes abordés.

Création d'une fonction d'alignements

L'objectif, pour rendre l'outil plus performant, est bien entendu d'agrandir cette base de données, en y ajoutant des alignements supplémentaires mais également en variant le contenu aligné puisque pour l'heure, elle ne contient que des brèves journalistiques. La tâche est fastidieuse, aussi nous avons pris le temps de réfléchir à la création d'une fonction d'alignement à

intégrer au concordancier, qui puisse d'une part assister la création d'alignements, et d'autre part, permettre à des utilisateurs externes de participer à l'agrandissement de la base de données du concordancier. Une interface permet de générer un alignement et de l'ajouter à la base de données en quelques clics. Dans un premier temps, l'utilisateur doit sélectionner une paire texte-vidéo sur laquelle il va travailler. Il choisit donc un texte parmi la liste des fichiers disponibles, puis la vidéo qu'il souhaite aligner, sachant que les vidéos proposées sont déjà identifiées comme étant des traductions du texte sélectionné. À noter qu'il est également possible de sélectionner d'abord la vidéo puis le texte. Une fois la paire de contenus validée, la fonction d'alignement est affichée comme illustré figure 26.

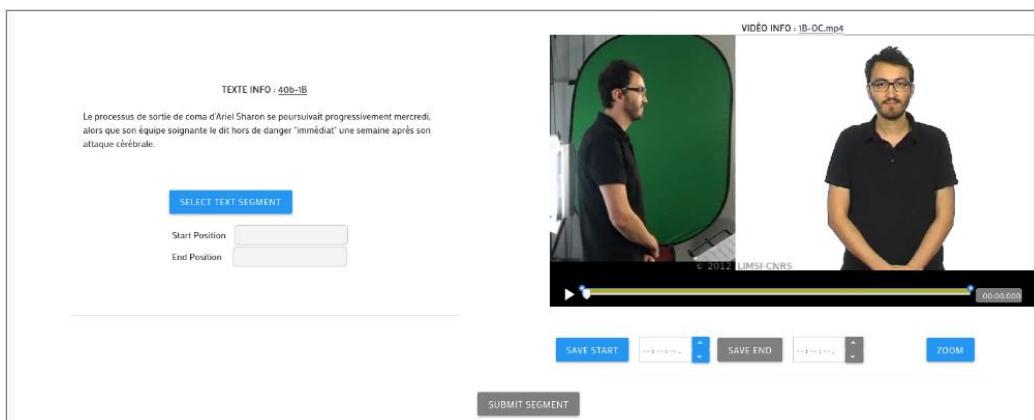


Figure 26 Interface de création d'alignements

La création d'un alignement comporte deux étapes :

- Dans le texte, on sélectionne le segment que l'on désire aligner, en le surlignant avec la souris. Les indices des premier et dernier caractères de la chaîne apparaissent dans les fenêtres dédiées à titre indicatif (non éditables manuellement). On clique ensuite sur *select text segment* pour valider le choix.

- Dans la vidéo, on repère la traduction du segment précédemment sélectionné, et on identifie son début et sa fin avec les boutons *save start* et *save end*. Ces deux boutons permettent de positionner des

bornes dans la barre de défilement de la vidéo, et d'extraire le temps correspondant, affiché dans la fenêtre adjacente. Le bouton *Zoom* permet, si la vidéo est trop longue, de zoomer sur la barre de défilement pour faciliter le positionnement des bornes.

Une fois ces deux étapes réalisées, l'utilisateur peut enregistrer son alignement en cliquant sur le bouton *Submit segment*. Ce dernier est alors ajouté dans la base de données. À noter que chaque utilisateur dispose de son propre fichier d'alignements, ce qui nous permet de garder plus facilement la main sur ce qui est ajouté dans le concordancier, pour modification ou suppression en cas d'erreur. À terme, il sera possible, lors du lancement d'une requête dans le concordancier, de filtrer par contributeurs les fichiers dans lequel la recherche est effectuée. Cette fonction constitue un gain de temps considérable dans la création des alignements, faute d'automatisation possible. Au total, l'échantillon d'alignements pour ce prototype comporte actuellement 800 entrées, soit 200 entrées supplémentaires réalisées à l'aide de cette fonction directement intégrée dans le concordancier (réalisées par nos soins et l'aide de quelques bénévoles curieux).

4.1.2 Implémentation

Nous avons cherché à fournir aux traducteurs, pas toujours familiers des environnements informatiques, un logiciel facilement accessible et ne nécessitant pas d'installation ou mise à jour avant l'utilisation. Ainsi, le concordancier a été implémenté sous la forme d'un site web accessible librement. Son développement a été réalisé dans le cadre d'une collaboration avec PostLab¹⁸. L'accès nécessite de créer un compte utilisateur, mais il existe un compte de test public dont les informations de connexion peuvent être communiquées sur demande. L'outil repose sur un

¹⁸ <https://www.postlab.fr/>

programme réalisé sous Python 3.

La section ci-après en détaille les différentes fonctionnalités.

Interface

Le concordancier propose deux fonctions : la fonction d'alignement, présentée ci-avant, ainsi que l'outil principal, la fonction de recherche. L'une et l'autre sont accessibles depuis une page d'accueil. Lors de la première utilisation, la fonction de recherche se présente simplement sous la forme d'un formulaire de recherche, à l'intérieur duquel l'utilisateur renseigne sa requête en français écrit. Le programme parcourt alors la base de données à la recherche d'une correspondance. S'il n'en trouve aucune, il renvoie un message d'information pour le signaler. Si le segment est contenu dans plusieurs alignements, le résultat proposé sera toujours celui de l'alignement le plus restreint autour du segment recherché. Les résultats sont listés verticalement sur la page et chaque occurrence est affichée comme l'illustre la figure 27. Dans l'exemple ci-dessous, la requête de l'utilisateur est « glissement de terrain ».

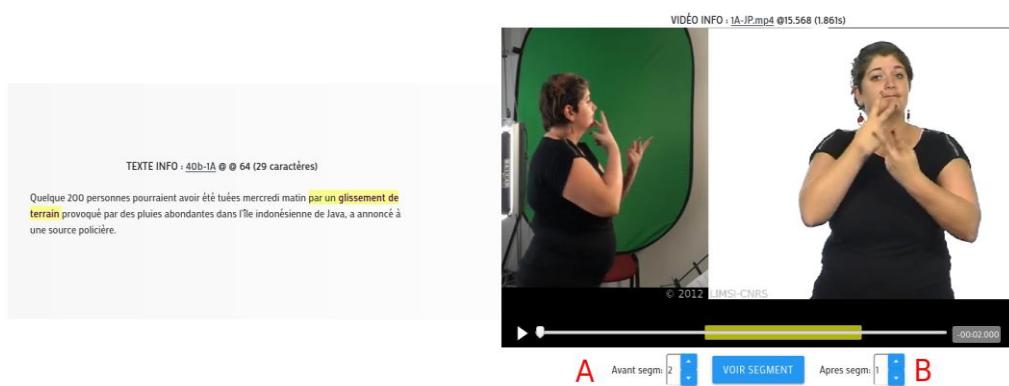


Figure 26 Exemple d'affichage de résultats pour la requête sur les termes « glissement de terrain »

Une occurrence donnée se présente en deux parties : le texte sur la gauche et la vidéo sur la droite. Le segment aligné qui contient la requête est surligné en jaune dans le texte, qui apparaît d'ailleurs en entier pour fournir du

contexte à l'utilisateur. La requête exacte est quant à elle mise en gras. Le titre situé au dessus du texte fournit les informations le concernant : l'identifiant du texte (40b-1A ici), ainsi que la position de départ (caractère 64) et la longueur (29 caractères) du segment. Cliquer sur l'identifiant du texte ouvre un onglet supplémentaire qui contient l'intégralité du texte (pour le cas où celui-ci serait trop long à afficher entièrement sur la page de résultats du concordancier). De façon similaire, le segment vidéo est surligné en jaune dans la barre de défilement de la vidéo. Le titre affiche l'identifiant de la vidéo (1A-JP) qui permet d'ouvrir la vidéo dans un nouvel onglet, le temps de départ (15.568s) et la durée (1.861s) pour le segment affiché. L'utilisateur peut choisir d'afficher plus ou moins de contexte à droite et à gauche du segment vidéo via les boutons dédiés (A et B sur la figure 27). Le bouton *voir segment* permet de recentrer le lecteur sur ce dernier. La vidéo est jouée en boucle.

Fonction de recherche : mode d'emploi

Bien que l'utilisateur attende généralement des correspondances exactes lors des requêtes effectuées, on peut envisager des expressions composées qui comportent un paramètre. C'est le cas par exemple de la formulation «quelque XX personnes», où XX est une valeur numérique. La recherche par correspondance exacte pourrait renvoyer des résultats corrects, mais qui seraient noyés au milieu d'autres résultats non-pertinents puisque ne comportant que l'un ou l'autre des composés (« quelques », « personnes », ou « XX » uniquement). C'est pourquoi la fonction de recherche inclut également la possibilité d'une requête partielle complétée par un suffixe ou un affixe :

- Le suffixe « # » ajouté à la fin d'un radical permet de faire correspondre toutes les occurrences d'un radical peu importe sa

terminaison, avec ou sans limite de longueur. Par exemple, la requête Europ# renverrait comme résultats : Europe¹⁹, européen, européens, européenne, européennes... Pour déterminer une éventuelle limite, il suffit d'ajouter la longueur maximum désirée à la suite. Ainsi la requête Europ#3 ne va renvoyer que des résultats contenant 3 caractères maximum après le radical, soit ici Europe, et européen.

- L'affixe « ## » permet de rechercher une séquence de mots contenant un terme intermédiaire. Par exemple, la requête « quelque ## personnes » pourrait renvoyer comme résultats « quelque 200 personnes », « quelque deux cents personnes », « quelque mille et une personnes. ». On peut également poser une condition sur le nombre d'éléments intermédiaires tolérés, en ajoutant le nombre maximum désiré à la suite de l'affixe. La requête quelque ##1 personnes ne reverrait dans notre exemple que le résultat « quelque 200 personnes ».

4.1.3 Exemples d'usage

La sous-section suivante présente différents usages qui peuvent être fait de ce concordancier, du point de vue de la traduction mais également de celui de la langue en général. Les traductions complètes desquelles sont issues les captures en illustration sont disponibles dans les annexes vidéos.

Explorer la langue

Comme l'ont mentionné les professionnels interrogés lors du brainstorming, les ressources en langue des signes sont rares. De même, les ressources existantes sont pour la plupart mal documentées voire difficiles d'accès. Au delà de l'utilité du concordancier dans la traduction, c'est également une

¹⁹ La recherche n'est pas sensible à la casse.

façon unique d'explorer la langue. Il pourrait servir ainsi à illustrer l'enseignement de la LSF au travers d'exemples, mais également comme outil pédagogique dans la formation des interprètes et des traducteurs professionnels. Il permet en effet de trouver et d'afficher non seulement du lexique et des phénomènes grammaticaux en contexte, mais également de mettre en lumière certains choix de traduction.

Afin d'évaluer l'utilisabilité de notre concordancier, nous avons organisé un atelier de présentation de l'outil qui est décrit chapitre 5. Nous avons proposé aux professionnels de la traduction présents de consulter plusieurs textes, et d'isoler les expressions qu'ils seraient le plus amené à rechercher dans notre concordancier. Voici trois des exemples les plus souvent mentionnés. (Kaczmarek & Filhol, 2020).

Mot sans signe standard correspondant, mais exemples de traduction similaires entre eux

Certaines expressions figées du français n'ont pas d'équivalent direct en LSF. C'est le cas par exemple de l'expression « fin de non recevoir ». Le fait qu'il n'y ait pas de signe lexical spécifique pour une expression ne signifie pas qu'elle est intraduisible pour autant. Quand on recherche « fin de non recevoir » dans notre concordancier, on obtient trois résultats. Notons que les trois traductions pour chaque brève dans le corpus sont indépendantes. Elles utilisent pourtant ici toutes les trois une structure similaire, à savoir le signe “refuser, rejeter”. Les expressions du visage sont également identiques, tout comme la structuration spatiale des traductions.



Figure 27 Trois versions de « fin de non recevoir »

Le fait que la forme soit trois fois identique peut inspirer confiance malgré l'absence de signe lexical spécifique. Ce genre de résultats peut conforter le traducteur dans ses choix de traduction, ou l'encourager à faire usage de la même structure dans sa propre traduction.

Pas de signe standard correspondant, mais exemples différents

Dans d'autres cas, on peut obtenir des traductions différentes. Il s'agit par exemple de l'expression « mis à mal », qui ne présente pas d'équivalent hors contexte en LSF. On trouve trois exemples de traduction dans notre base de données, et ces trois occurrences sont différentes les unes des autres. La phrase qui contient le segment est la suivante : « Les espoirs de paix au Sri Lanka sont de nouveau mis à mal après le lancement mercredi d'une grande offensive militaire contre la rébellion tamoule dans le nord de l'île ».



Figure 28 Trois versions de « mis à mal »

Pour rendre le sens de l'expression « mis à mal », le signeur de gauche utilise le signe « casser », et celui de droite opte pour « difficile ». Le signeur du milieu utilise une structure iconique basée sur le signe « espoir », qui ici s'écroule. Dans ce dernier cas, la traduction est notablement plus influencée par le contexte que précédemment : le signeur reprend un élément du texte en français (espoirs) qu'il adapte de façon iconique en LSF. Si l'expression avait été employée dans une autre phrase, son choix de traduction aurait été différent.

Trois traductions différentes sont proposées pour une même expression dans un même contexte. Cet exemple met en lumière tout l'intérêt du concordancier, à savoir la variété des occurrences que l'on peut obtenir pour un même contexte. Les professionnels peuvent capitaliser sur le travail passé, et les apprenants de la langue peuvent mieux appréhender la finesse et la richesse de la langue.

Relation de cause à effet

Le concordancier est également une façon intéressante d'observer certains phénomènes sémantiques. Par exemple, la relation de cause à effet peut être traduite de nombreuses façons en LSF, en fonction du contexte et des choix du traducteur.

Dans l'exemple « un tsunami provoqué par » illustré figure 30, cette relation a été traduite par l'utilisation du signe «conséquence», par le signeur de droite. Les deux autres signeurs ont fait le choix d'utiliser des structures iconiques qui décrivent l'événement, à savoir : un séisme sous-marin provoque un tsunami. Dans ces cas, la relation de cause à effet réside dans la succession des événements, accompagnée par les expressions des visages des traducteurs, et dans la dynamique du discours. On remarque typiquement un temps de transition spécifique très bref entre les deux événements mentionnés, ainsi qu'un haussement des sourcils pour marquer la conséquence.



Figure 29 Trois versions de « un tsunami provoqué par »



Figure 30 : Deux versions de « un glissement de terrain provoqué par des pluies abondantes »

Sur la figure 31, la phrase à traduire était : «un glissement de terrain causé par des pluies abondantes ». Les deux traducteurs décrivent la pluie, le sol gorgé d'eau, puis le glissement de terrain. Les captures d'écran sont prises juste avant le signe du glissement de terrain. Ici encore la relation de cause à effet réside dans la succession des événements décrits, dans les expressions du visage, mais également dans le temps de transition très bref entre les deux, avec les sourcils haussés et le menton relevé.

Ce concordancier est fonctionnel en tant qu'outil à part entière. Cependant, notre objectif à long terme est de proposer un logiciel complet de TAO qui l'intégrera en tant que module. Cet environnement de travail complet correspond à notre second prototype, détaillé dans la partie suivante.

4.2 L'environnement de travail intégré

Le prototype d'environnement de travail intégré a été réalisé en collaboration avec une stagiaire ingénierie de Télécom Paris. Sa conception a été guidée par le cahier des charges préalablement établi, auquel se sont ajoutées quelques nécessités d'ordre pratique : pouvoir sauvegarder son projet en cours, d'une part, et d'autre part, tenir compte du fait que le métier d'interprète LSF n'induit pas forcément d'aisance avec les pratiques informatiques. Tout comme pour le concordancier, nous avons choisi

4.2 L'environnement de travail intégré

d'implémenter le logiciel sous la forme d'une application web permet aux utilisateurs d'y accéder en suivant un simple lien URL. Nous avons décidé de réaliser le développement avec l'environnement de développement d'applications web Angular²⁰ en raison de sa puissance et de sa modularité.

L'interface se découpe en plusieurs zones dont chacune assure un rôle déterminé, présentées dans les sections suivantes, et illustrées figure 32.

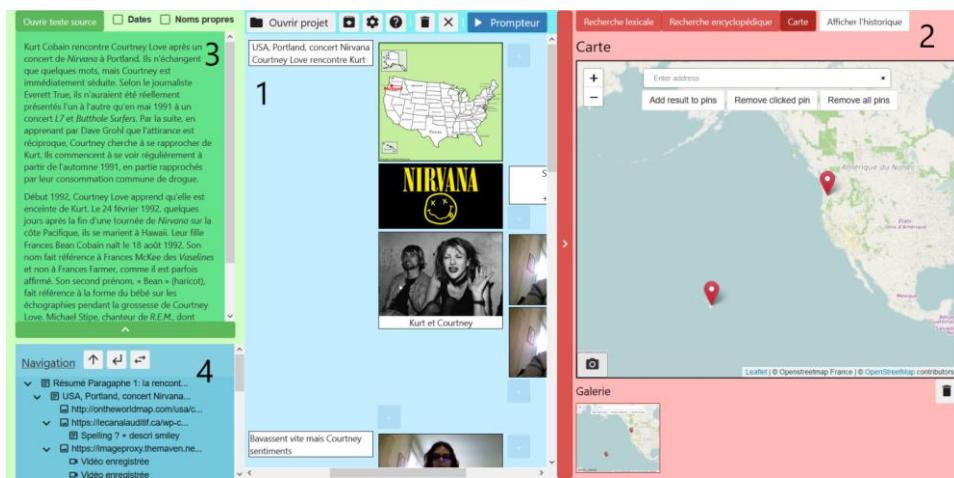


Figure 31 Interface principale de l'environnement de travail intégré

4.2.1 Zone d'édition

La zone centrale (bleue ciel, 1) est la zone de travail principale du traducteur. C'est là que l'utilisateur pourra créer de nouveaux Blocs, les remplir avec du contenu de son choix, et les déplacer pour établir une hiérarchie à sa convenance. Cette zone permettant de prendre des notes, ainsi que d'insérer le résultat de différentes recherches réalisées au préalable par l'utilisateur, remplit l'exigence d'avoir une structure modulaire, facilement réorganisable afin de faciliter la phase de réorganisation de la traduction.

Création et mouvement des Blocs

Un bouton + permet la création d'un nouveau Bloc, qu'il s'agisse du

²⁰ <https://angular.io/>

premier Bloc du projet, ou d'un Bloc enfant. Chaque parent dispose également d'un bouton +. Comme stipulé dans le cahier des charges, ces Blocs sont mobiles et hiérarchisés. Ils sont déplaçables par cliqué-glissé, et l'utilisation de phénomènes de transparence et de grisé en cours de déplacement et un jeu de marqueurs permettent de faciliter la visualisation des modifications en cours (fig. 33).

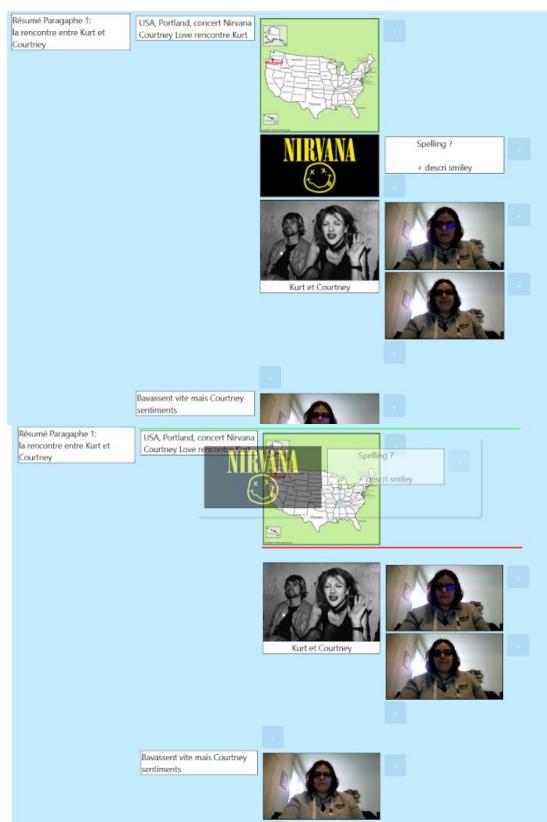


Figure 32 : Marqueurs lors du déplacement des Blocs : la ligne rouge permet de visualiser l'emplacement où sera déposé le Bloc en cours de déplacement.

charger du contenu dans un Bloc

Une fois les Blocs créés, il est possible de les remplir avec plusieurs types de contenu. Lorsqu'un Bloc n'est pas encore rempli, des icônes permettant à l'utilisateur d'ajouter facilement du contenu sont affichées à l'intérieur, en

grisé dans l'arrière-plan du Bloc afin de ne pas les confondre avec du contenu réel que l'utilisateur aurait inséré par la suite (figure 34). Si le Bloc n'est pas vide, il est possible d'ajouter du contenu supplémentaire en ouvrant le menu contextuel à l'aide d'un clic droit. L'utilisateur peut ensuite choisir le type de contenu à rajouter, ce qui ouvre le dialogue d'ajout de contenu approprié.

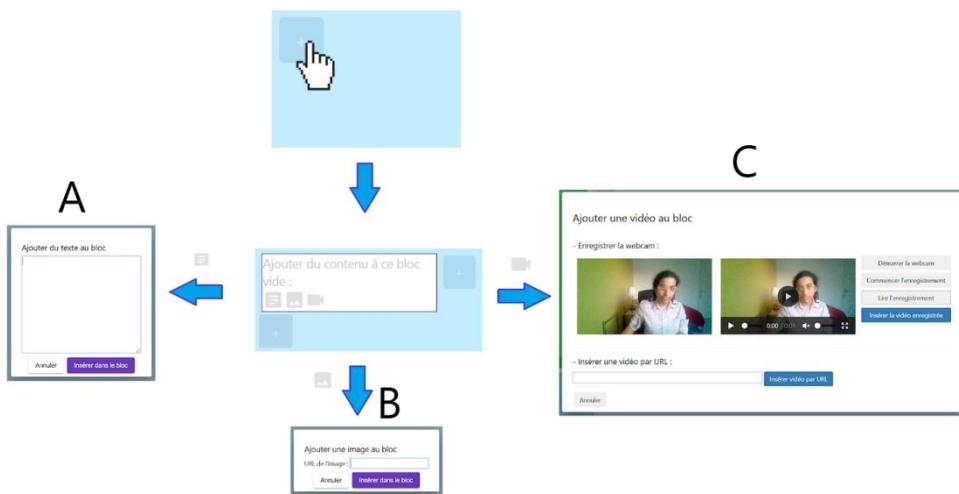


Figure 33 À partir du Bloc, il est possible d'insérer différents types de contenu : texte, image, ou vidéo.

Il est possible d'ajouter du contenu au Bloc sous plusieurs formes :

- Du texte, entré par l'utilisateur au clavier (fenêtre A figure 34). Le traducteur peut ainsi saisir des notes dans des Blocs, à l'image de posts-its. Couplé avec les autres types de contenu, ce texte peut également servir de légende.
- Des images, insérées via leur adresse URL ou à l'aide d'un cliqué/glissé depuis l'image jusqu'à un Bloc vide (fenêtre B). Le traducteur peut ainsi insérer des images qu'il juge pertinentes lors de la préparation de sa traduction,

- Des vidéos, insérées depuis d’autres sites web via leur adresse URL, ou enregistrées par l’utilisateur lui-même par sa webcam, via un popup d’enregistrement (fenêtre C), afin de prendre des notes directement en LSF. Cette fonctionnalité avait été sollicitée par les traducteurs interviewés lors des premières études.

Opération sur les contenus

D’autres actions sur les Blocs sont possibles, comme la fission du contenu d’un Bloc parent vers des Blocs enfants, ou l’inverse, laissant à l’utilisateur la possibilité de récupérer les contenus de tous les enfants et de les insérer dans le Bloc parent. Ces opérations permettent, à l’instar de la mobilité des Blocs, de faciliter la manipulation des contenus : il n’est pas nécessaire de les supprimer pour les ajouter de nouveau dans un autre Bloc.

La fission concerne les Blocs qui possèdent au moins un élément de contenu. L’opération, pour un Bloc sélectionné, crée autant de Blocs enfants que le Bloc parent possède de contenus différents. Elle vide ensuite le Bloc parent pour distribuer un contenu par Bloc enfant. L’opération inverse, dite de fusion, rassemble tous les contenus de Blocs enfants dans leur Bloc parent.

Un seul Bloc peut regrouper plusieurs contenus. Dans ces cas-là, les contenus sont réorganisables à l’intérieur d’un même Bloc à l’aide d’un menu contextuel, si l’utilisateur veut par exemple afficher le texte au-dessus ou en-dessous de la vidéo. De même, il est possible pour l’utilisateur de supprimer ou de modifier individuellement chaque contenu du Bloc, de supprimer tous les contenus afin de le vider tout en gardant la structure de l’arbre intacte, ou de supprimer le Bloc-même ainsi que tous ses enfants.

Menu et options



Figure 34 Menu de la zone à Blocs

Cette zone contient également la barre de menu (figure 35), dont les fonctions sont réparties comme suit :

1. Charger un projet déjà existant (sous forme d'archive).
2. Sauvegarder et télécharger le projet en l'état actuel (sous forme d'archive également).
3. Accéder et modifier certains paramètres, comme le mode d'affichage des Blocs (vertical ou horizontal).
4. Afficher le tutoriel utilisateur.
5. Supprimer le Bloc sélectionné (et son contenu).
6. Vider le contenu du Bloc sélectionné (mais sans supprimer le Bloc).
7. Épingler le Bloc sélectionné
8. Lancer le défilement sous forme de prompteur (fonctionnalité prévue mais non implantée dans cette version du prototype).

L'export de projet au format ZIP permet à l'utilisateur de sauvegarder l'avancement de son projet pour conservation, ou pour le partager avec des collaborateurs. Les images et vidéos générées par l'utilisateur, telles que les captures de cartes ou les enregistrements à la webcam, sont enregistrés respectivement au format PNG et MP4. Une simple extraction du projet zippé permet alors à l'utilisateur de récupérer ces fichiers images et vidéos si jamais il désire y accéder en dehors du logiciel. Les ressources puisées en ligne par l'utilisateur (images tirées d'encyclopédie, vidéos tirées de dictionnaires...) ne sont pas sauvegardées dans le fichier du projet afin de ne pas en augmenter inutilement la taille, ainsi que pour des problématiques de licences ou de droits, seuls les liens sont conservés.

4.2.2 Modules et outils

La partie la plus à droite de l'interface (2) accueille les modules annexes de

recherche.

Elle est rétractable pour laisser plus de place à la zone d'édition lorsqu'elle n'est pas utilisée. Elle regroupe, à l'heure actuelle, un module de recherche lexicale, un module de recherche encyclopédique, le concordancier, et un module de carte interactive. Il suffit à l'utilisateur de surligner un élément dans le texte pour l'envoyer directement dans la barre de recherche du module de recherche.

Recherche lexicale et recherche encyclopédique

Cet onglet se compose de la sorte :

1. la fonction de recherche lexicale
2. la fonction de recherche encyclopédique
3. la carte interactive
4. le bouton pour réduire l'historique
5. l'historique interactif
6. la barre de recherche

L'onglet de recherche lexicale permet de lancer une même recherche sur plusieurs sites à la fois, et agrège les résultats organisés en onglets par site. Actuellement, les recherches peuvent s'effectuer sur les sites *Elix*²¹ et *Spread The Sign*²², deux dictionnaires français - LSF collaboratifs en ligne. De la même façon, l'onglet de recherche encyclopédique permet de procéder à une recherche sur Wikipédia directement depuis l'interface du logiciel. Depuis ces pages, l'utilisateur peut par exemple sélectionner une image sur la page de résultat, et la faire glisser directement à l'intérieur d'un Bloc pour la conserver.

²¹ <https://dico.elix-lsf.fr/>

²² <https://www.spreadthesign.com/fr/fr/search/>

4.2 L'environnement de travail intégré

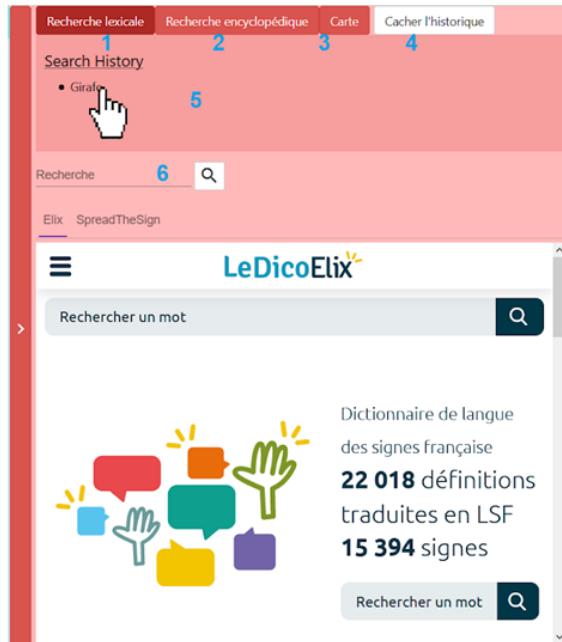


Figure 35 Les ressources intégrées sont accessibles à droite de l'interface

Carte

Le module de recherche cartographique contient une carte interactive sur laquelle l'utilisateur peut placer des épingle afin de sauvegarder certaines localisations. Il est également possible de prendre des captures pour ensuite les insérer dans des Blocs de la zone d'édition (figure 37).

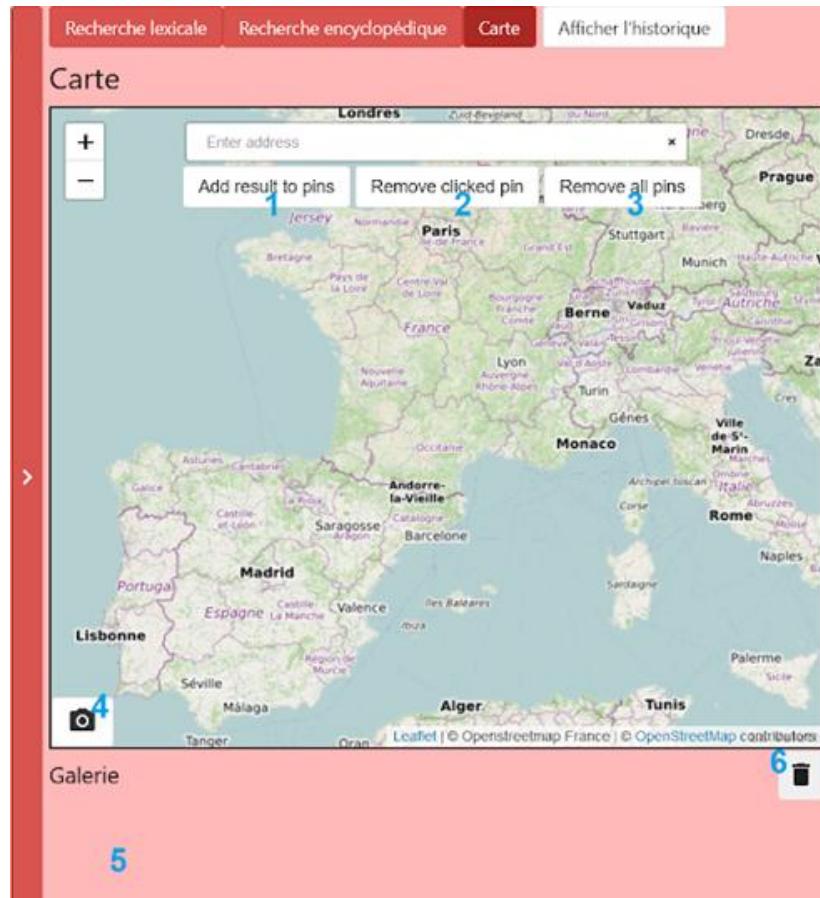


Figure 36 Le module de carte et ses fonctions : épingle, capture d'écran, galerie

L'outil carte propose plusieurs actions à l'utilisateur. Une barre de recherche permet de lancer une requête, ce qui centre la carte sur l'endroit désiré, mais permet également les actions suivantes :

1. ajouter une épingle à l'endroit recherché, pour le conserver
2. retirer la dernière épingle placée
3. retirer toutes les épingles sur la carte
4. prendre une capture de la carte
5. afficher toutes les captures faites durant la session. La galerie se vide automatiquement à la sauvegarde du projet
6. vider manuellement la galerie

Les captures réalisées sont stockées dans la galerie et peuvent être déplacées directement dans un Bloc (fig. 38). L'affichage simultané de plusieurs

4.2 L'environnement de travail intégré

épingles permet une visualisation des relations topographiques entre plusieurs lieux.

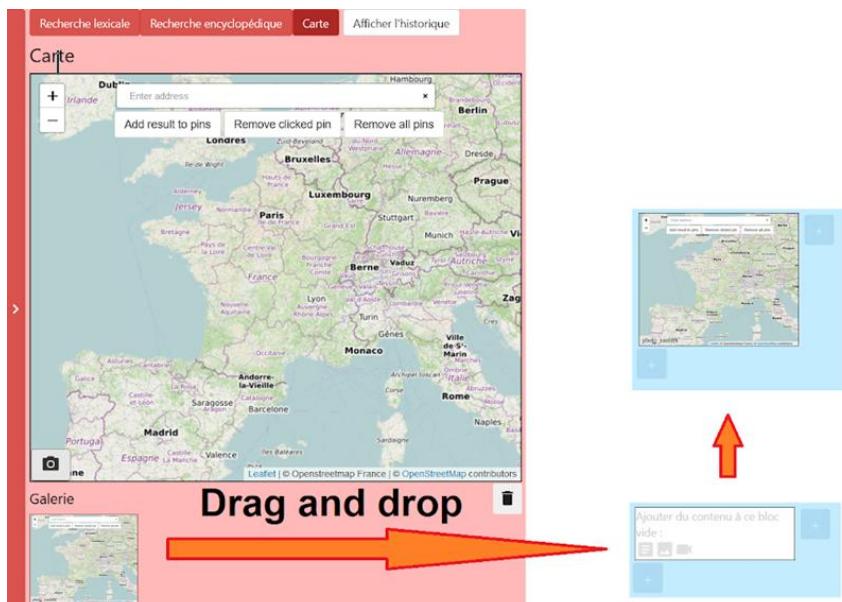


Figure 37 Insertion d'une carte dans un Bloc à partir de l'outil carte

Historique

L'outil propose un historique de recherche commun et interactif. À chaque recherche lancée, c'est-à-dire à chaque terme renseigné et validé dans la barre de recherche, ou surligné dans le texte source, ce dernier est enregistré et listé dans l'historique de recherche (fig. 39). Cliquer sur un élément de la liste de l'historique permet de relancer instantanément cette même recherche. Cette option en plus d'offrir un gain de temps, permet de garder une vue sur les recherches déjà effectuées. À noter cependant que l'historique n'est actuellement pas maintenu lors de la sauvegarde, et apparaîtra vide lorsque le projet sera chargé.

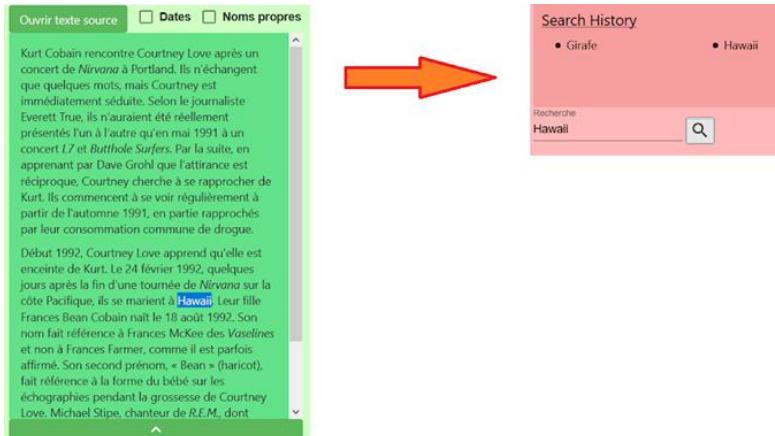


Figure 38 L'historique de recherche interactif

Pour illustrer ces modules, et résumer la section 4.2.2, une vidéo de démonstration de l'utilisation de l'application est disponible dans les annexes vidéo.

4.2.3 Gestion du texte source

Le texte source est chargé par l'utilisateur, et apparaît dans la partie en haut à gauche (3, partie verte figure 32). Pour ce faire, il suffit de cliquer sur le bouton dédié, et de sélectionner dans l'explorateur le fichier désiré. Les formats texte sans mise en page (extensions .txt) et avec mise en page (extension .docx) sont actuellement pris en charge.

Texte et formatage

Le format du texte source est conservé lors du chargement dans le logiciel. Cela concerne notamment les mentions en gras, en italique, la mise en page (saut de lignes, titres, etc). Le texte défile manuellement via une barre de navigation. Ce dernier n'est pas modifiable à l'intérieur du logiciel. L'utilisateur peut néanmoins surligner et copier le texte s'il le désire.

Traitements automatiques du texte

Une fois chargé, des fonctions simples de détection d'entités (noms propres et dates) sont disponibles. En effet, lors de l'observation des traducteurs

professionnels, ces derniers avaient souvent le réflexe de surligner toutes les dates et/ou noms propres, et ce de manière systématique. Ces fonctions permettent d'assister la tâche. Actuellement, elles sont gérées par des expressions régulières, mais dans le futur, nous aimerions ajouter des fonctions de TAL plus poussées afin de détecter des expressions idiomatiques ou des expressions temporelles plus subtiles (eg. « à la fin des années 80 », « il y a 3 jours » ...).

4.2.4 Navigation dans l'arbre

Comme le nombre de Blocs n'est pas limité, et que l'espace de la zone d'édition demeure malgré tout restreint par l'écran, les structures de Blocs peuvent rapidement devenir très grandes. Le défilement vertical et horizontal nuit à l'aspect « pensée visuelle » que nous cherchions à mettre en avant. Pour pallier ces inconvénients, la structure de l'arbre est affichée dans un encart de navigation (4, en bas à gauche, fig. 32). La navigation permet à l'utilisateur de conserver une vision globale de l'arborescence en cours (voir figure 40).

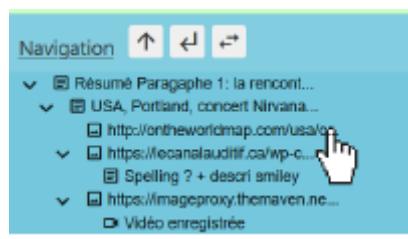


Figure 39 Le menu de navigation

L'icône en début de ligne rappelle la nature du contenu du Bloc : image, texte, ou vidéo. Depuis ce menu, l'utilisateur peut choisir de centrer l'affichage de la zone d'édition sur une portion de l'arbre en particulier. En cliquant sur un élément de la liste, on sélectionne une nouvelle racine pour l'affichage, sachant que sélectionner un Bloc dans cette liste n'affichera l'arborescence qu'à partir du sous-arbre de ses enfants. Le Bloc en question,

devenu racine, ne sera pas affiché. La racine actuelle est toujours indiquée en rouge dans la navigation. Les différents boutons fléchés permettent de naviguer entre les racines ou de revenir à l'affichage total dans la zone d'édition

Ce chapitre a détaillé la réalisation de deux outils prototypes, qu'il convient à présent d'évaluer : c'est le sujet du chapitre suivant. Il s'agit une fois encore d'impliquer les professionnels à qui s'adressent ces propositions logicielles, et de collecter des retours subjectifs, qui seront mis en regard avec d'autres données plus objectives, à l'image des méthodologies précédemment mobilisées.

Chapitre 5

Évaluation préliminaire des outils

Le développement et l'implémentation terminés, il reste deux étapes importantes concernant nos prototypes logiciels. La première consiste à les soumettre à une utilisation en situation. Cette étape d'évaluation sollicite la participation de professionnels du métier, et permet de repérer des failles de conception d'une part, et d'autre part de mesurer l'impact que peuvent avoir nos outils sur les pratiques professionnelles. C'est ce que présente la section 5.1 : l'évaluation du concordancier, de l'environnement de travail intégré, ainsi que les pistes d'amélioration qui en découlent. Cette dernière partie permet de faire le lien avec la section 5.2 de ce chapitre, à savoir les pistes d'améliorations. Les expériences préliminaires menées avant le développement des prototypes ainsi que des réflexions spontanées menées au cours de celui-ci ont permis de dresser des listes de fonctionnalités que nous n'avons pas pu toutes implémenter encore, mais qui pourraient s'avérer riches d'intérêt pour améliorer encore nos outils. Les résultats obtenus lors des évaluations fournissent eux aussi des perspectives à explorer. Cette partie traite également du devenir général des prototypes.

5.1 Évaluations

Nous l'avons vu, ces deux prototypes sont le fruit d'interactions et d'itérations avec les professionnels du milieu. Les fonctionnalités intégrées dans ces deux outils reposent sur leurs suggestions et leurs remarques, que nous nous sommes efforcés de restituer au mieux lors du développement. Cette section présente l'évaluation des prototypes, réalisée en deux parties. La section 5.1.1 détaille les tests ouverts et l'évaluation du concordancier en tant qu'outil indépendant. La section 5.1.2 détaille la méthodologie et les protocoles mis en place afin d'évaluer l'environnement de travail intégré, ainsi que les résultats obtenus et leur analyse. Enfin, la section 5.1.3 dresse une liste des fonctionnalités à améliorer en regard des résultats obtenus, et des pistes pour ce faire.

5.1.1 Concordancier

Le concordancier en tant qu'outil indépendant a fait l'objet de plusieurs présentations, notamment dans le cadre d'un atelier organisé par nos soins à l'Université de Toulouse en décembre 2019 sur le thème de la TAO, et d'une conférence à l'Université de Lille en novembre 2021, sur le thème de la TAO, proposée aux étudiants du Master ILSF. Il a également fait l'objet de démonstrations plus ciblées, qui ont permis de récolter des retours quant à son utilisation. Rappelons que le concordancier comporte deux fonctions différentes, recherche et alignement, qui sont également présentées de façon distincte lors des différentes sessions. En effet, pour cerner l'intérêt de la fonction d'alignement, il est préférable que le public ait pu au préalable prendre connaissance du fonctionnement du concordancier via la fonction de recherche, et par extension de l'importance de la création d'alignements pour le rendre de plus en plus performant.

Présentation lors de l'atelier TAO

Cet atelier concernait aussi bien le public étudiant que le public professionnel (traducteurs, interprètes, enseignants), pour un total d'environ soixante personnes. L'objectif est d'une part de présenter la TAO à un public peu familier, et d'autre part de mettre en avant nos travaux et les outils que nous avons à proposer. Le public est impliqué dans la présentation, qui comporte notamment deux mises en situation de traduction.

Dans un premier temps, les participants sont mis dans une situation de traduction : ils sont libres de se répartir en groupes de trois ou quatre, et doivent traduire un court texte élaboré spécialement pour l'occasion. Nous en avons mis quatre à disposition : la description anatomique de la lamproie, le mode d'emploi pour changer les cordes d'une harpe, deux haikus et une brève journalistique relatant des inondations à Londres. Chaque groupe choisit un texte, et doit le traduire en ayant recours au matériel et ressources de leur choix, pourvu qu'il prenne note de la méthode utilisée pour y parvenir.

- Un premier temps d'échange permet ensuite de mettre en commun les méthodes et les difficultés, ainsi que les traductions réalisées.
- Ce temps est suivi d'une présentation de la TAO d'un point de vue général, puis plus spécifiquement de nos travaux. Cette présentation s'achève sur la présentation du concordancier, et d'un essai libre de ce dernier.
- La deuxième mise en situation de traduction propose de traduire une des brèves journalistiques ayant servi à alimenter le corpus, avec cette fois la possibilité d'avoir recours au concordancier comme ressource.

- Pour conclure chaque session, la fonction d’alignement est présentée. Cette présentation est ponctuée d’un appel à contribution pour la base de données d’alignements sur laquelle repose le concordancier.
- La journée s’achève alors sur des échanges concernant cette deuxième partie de la session, qui constitue pour nous une évaluation qualitative du prototype.

Les textes proposés lors de l’activité sont disponibles en annexe.

L’accueil réservé à l’outil s’est avéré favorable, tant lors de la présentation que pendant la mise en situation qui a suivi. Dans la première mise en situation de traduction (étape 1), les participants avaient eu recours à des ressources en ligne : dictionnaires de LSF, recherche d’images et de cartes... Dans la deuxième mise en situation de traduction (étape 4), le concordancier leur a permis de lever des ambiguïtés de sens, tout comme de vérifier ou d’inspirer leurs traductions. En moyenne, et en considérant uniquement les textes de natures similaires (hors haïkus, donc), le second exercice ne prend qu’une vingtaine de minutes aux participants, là où le premier exercice nécessitait au moins quarante-cinq minutes. Nous avons d’ailleurs dû interrompre l’exercice pour respecter le planning de la journée. Les utilisateurs comprennent rapidement le fonctionnement de l’outil, et y ont recours à plusieurs reprises sans accroc. L’accès à du contenu précédemment traduit et à de la langue en contexte permet de réduire le temps nécessaire à la traduction.

Démonstrations ciblées

Le concordancier a également été présenté dans un cadre plus restreint, à une douzaine d’autres professionnels du métier. Il s’agit de démonstrations ciblées de l’outil, de ses usages et de ses fonctionnalités. Par respect des

contraintes sanitaires particulières en vigueur à cette période, ces démonstrations ont été réalisées sous forme de visio-conférence. Le concordancier étant déployé en ligne, les participants pouvaient suivre et manipuler sans problèmes. Les rencontres virtuelles s'articulent autour d'une brève mise en contexte de nos travaux, puis d'une visite guidée commentée du logiciel agrémentée d'une démonstration pour chaque fonction.

La fonction d'annotation, également présentée, a été utilisée par quatre personnes sur la douzaine ayant assisté aux démonstrations.

Ces démonstrations en groupes restreints de collègues ont été plus prolifiques quant aux retours qu'ils nous ont permis d'obtenir. Par le biais de retour sur l'instant, d'une part, mais également de bilan faisant suite à plusieurs utilisations.

5.1.2 Environnement de travail intégré

L'évaluation de l'environnement de travail intégré que nous proposons repose sur deux volets, l'un qualitatif et l'autre quantitatif. Nous nous intéressons aux retours des utilisateurs d'une part, et aux performances du prototype d'autre part. Ces deux types d'évaluation relèvent d'un seul protocole, impliquant la participation de plusieurs groupes de professionnels. Ce protocole se déroule en deux temps : un premier passage qui a pour but de récolter des premières impressions, et d'évaluer la prise en main de l'outil (protocole à court terme), et une seconde passe au terme d'une période d'usage libre pour cette fois récolter des avis basés sur une utilisation prolongée et moins guidée pour évaluer l'impact à long terme de l'outil (protocole à long terme).

Méthodologie

Dans un premier temps, le protocole d'évaluation décrit ci-dessous s'adresse à des traducteurs LSF professionnels. Il nécessite la participation d'un

groupe test, c'est-à-dire des participants qui manipulent le prototype durant l'évaluation, d'un groupe témoin, qui lui se voit proposer les mêmes tâches à réaliser sans l'aide du prototype, ainsi qu'un groupe d'experts, qui sont amenés à évaluer les traductions réalisées par les deux autres groupes. Il s'agit pour les groupes test et témoin de réaliser une tâche de traduction. Seul le protocole à court terme nécessite la participation du groupe témoin. Chaque groupe doit se composer d'un nombre proche de professionnels, indépendamment de leur expérience dans le métier (bien que ce paramètre est pris en compte lors de l'évaluation par les experts, voir ci-après).

Le protocole à court terme propose aux groupes test et témoin de procéder à la traduction de trois textes, du français vers la LSF. Ces trois textes, de type journalistique, sont identiques pour tous les groupes de participants : deux textes courts d'une dizaine de lignes, et un texte plus long de plus d'une page au format A4 orientée portrait. Ces textes ont été choisis pour leur complexité, et les besoins de recherches qu'ils pourraient susciter. Les critères appliqués pour cette sélection se basent sur les tâches constitutives du processus de traduction que nous avons identifiées durant nos travaux préliminaires (recherche de cartes, recherches encyclopédiques), mais également du fait qu'ils contiennent de nombreuses dates ou de nombreux noms propres. Les textes proposés sont consultables dans les annexes.

La tâche de traduction va de la découverte du texte à traduire à l'étape de production filmée. Une session d'évaluation s'étend sur 4 heures. Nous recevons deux participants à la fois, durant autant de sessions que nécessaire. Chaque session est intégralement filmée.

Les participants, lors de leur arrivée, sont installés à un bureau équipé en fonction du groupe auquel ils appartiennent. Les participants test disposent d'un ordinateur équipé avec le prototype à évaluer. Les participants témoins disposent d'un ordinateur sans le prototype, connecté à internet. Ces

derniers sont libres de procéder comme à leur habitude pour traduire. Les textes sont remis un par un, c'est à dire qu'un participant doit avoir terminé de traduire le premier texte pour passer au second, et le second pour passer au troisième. Chaque tâche de traduction est filmée, et chronométrée. Une fois les trois textes traduits, les participants tests sont invités à remplir un questionnaire de type System Usability Scale (SUS, Brooke, 1996). Ce dernier comporte dix questions, qui reposent sur des concepts issus de la norme ISO 9241-11 (efficacité, fiabilité, satisfaction), et dont les réponses sont des degrés attribués sur une échelle de Likert (pas du tout d'accord, pas d'accord, neutre, d'accord, tout à fait d'accord) comme illustré Fig. 41. Le prototype propose un service dont nous cherchons à évaluer l'utilisabilité. Pour ce faire, il faut pouvoir mesurer l'expérience faite par les utilisateurs. Nous avons opté pour ce type de questionnaire car il est à la fois simple de fonctionnement, et rapide à remplir par ces derniers. Les items à évaluer ciblent le retour subjectif de l'utilisateur, et permettent de le transposer d'une valeur qualitative vers une valeur quantitative par le biais d'une échelle de score.

Nous apportons cependant des modifications à trois des items à évaluer présentés dans la version française de Gronier et Baudet (2021).

Item à évaluer	1: Pas du tout d'accord	2: Pas d'accord	3: Neutre	4: D'accord	5: Tout à fait d'accord
1. Je voudrais utiliser ce système fréquemment					
2. Ce système est inutilement complexe					
3. Ce système est facile à utiliser					
4. J'aurais besoin du soutien d'un technicien pour être capable d'utiliser ce système					
5. Les différentes fonctionnalités de ce système sont bien intégrées					
6. Il y a trop d'incohérences dans ce système					
7. La plupart des gens apprendront à utiliser ce système très rapidement					
8. Ce système est très lourd à utiliser					
9. Je me suis senti·e très en confiance en utilisant ce système					
10. J'ai eu besoin d'apprendre beaucoup de choses avant de pouvoir utiliser ce système					

Figure 40 Version française et validée du SUS, proposée par Gronier, G. & Baudet, A. (2021)

L'item numéro 1 dans un premier temps, ne permet pas de rejeter complètement l'hypothèse d'une réutilisation future. Si l'utilisateur répond « Pas du tout d'accord » à l'affirmation « je voudrais utiliser ce système fréquemment », le désaccord porte sur la fréquence, et non pas sur l'intention de l'utilisation elle-même. Nous supprimons donc l'adverbe dans

la formulation de l'item pour éviter ce biais. D'autre part, les items numéro 6 et 8 comportent des termes génériques qui nécessitent d'être précisés. La notion d'incohérence citée dans l'item 6 sera définie, dans le cadre de ce protocole, comme toute manipulation faite dans le logiciel qui peut entrer en contradiction avec une autre, ou toute manipulation donnant lieu à un résultat qui n'est pas celui attendu. Par exemple, si un clic sur le terme A de l'historique de recherche renvoie des résultats sur un terme B, il y a incohérence. La notion de « lourdeur d'utilisation » citée dans l'item 8 sera définie dans le cadre de ce protocole comme correspondant au degré d'effort que l'utilisateur doit déployer pour parvenir à ses fins lors de l'utilisation du logiciel. Par exemple, le nombre de clics nécessaires pour réussir une action, ou le nombre d'étapes intermédiaires pour en réaliser une autre (ajouter une image, se filmer, etc.).

Ce questionnaire donne lieu à un score sur 100, en tenant compte des items pairs et impairs, qui sont des items à polarité inversée. Les items impairs correspondent à des assertions subjectivement positives sur le système. Plus la réponse de l'utilisateur sera favorable, plus le score augmente. À l'inverse, les items pairs correspondent à des assertions subjectivement négatives sur le système. Cette fois, plus la réponse de l'utilisateur est favorable à l'assertion, plus le score décroît. Par exemple, un utilisateur répond « D'accord » à l'item 1, et « Pas d'accord » à l'item 2.

Le calcul se réalise comme suit:

1. Soustraire un point au score coché par l'utilisateur pour les items 1, 3, 5, 7 et 9 (items impairs, non inversés).
2. Calculer 5 moins le score coché par l'utilisateur pour les items 2, 4, 6, 8 et 10 (items pairs, inversés).
3. Additionner les 10 nouveaux scores ainsi obtenus, et multiplier le

tout par 2.5

La multiplication par 2.5 a pour but de faciliter l'interprétation des résultats, plus lisibles sur une échelle de 0 à 100 que de 10 à 50, ce qui correspond à la dispersion des scores avant la multiplication.

Ce score global peut être rapporté à des adjectifs pour faciliter son interprétation, comme illustré figure 42.

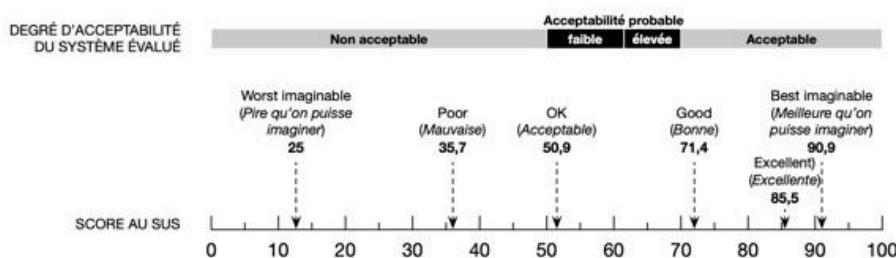


Figure 41 Signification du score au SUS, proposée par Bangor, Kortum et Miller (2009)

La seconde partie du questionnaire comporte des questions à choix multiples ou à réponses ouvertes pour justifier les réponses apportées au SUS, et ainsi mieux identifier les points à améliorer. Ces questions constituent la partie qualitative de l'évaluation. Elles interrogent l'aisance de la prise en main, le confort d'utilisation, l'utilité des solutions proposées par l'outil, le degré de confiance porté par l'utilisateur envers l'assistance, ainsi que le sentiment général quant à l'utilisation (rapidité, bénéfice). Le questionnaire est disponible en annexe.

L'évaluation objective du prototype repose elle sur des mesures indépendantes du retour des participants. La première concerne le temps. Chaque traduction pour chaque utilisateur est chronométrée, dans le but d'effectuer plusieurs tests statistiques par texte et par groupe, et de pouvoir évaluer si oui ou non l'application offre un gain de temps significatif lors du processus de traduction. Nous nous intéressons ensuite à l'usage que font les

utilisateurs des différentes fonctionnalités proposées par le prototype. Pour ce faire, un compteur de clics s'incrémente de 1 à chaque appel de fonctionnalité. L'analyse de ces données permet d'évaluer la pertinence des fonctionnalités, ainsi que la fréquence de leur utilisation. Le calcul de l'écart-type permet de lisser l'influence du paramètre personnel pour les cas où une minorité d'utilisateurs ait généré la majorité des utilisations comptabilisées.

Une fois toutes les tâches de traduction terminées, les résultats récoltés sont transmis au groupe d'experts pour qu'ils en évaluent la qualité. Le groupe d'experts est constitué de traducteurs professionnels avec une expérience plus élevée que la moyenne de celle des participants. Ces derniers ne sont pas informés du protocole afin de ne pas biaiser leurs retours. Leur tâche est de procéder à l'évaluation de chacun des textes traduits par les participants.

Cette évaluation est guidée par une grille de critères, pour rendre le jugement le plus objectif possible (voir fig. 43). Cette dernière est inspirée des travaux de Toudic et al. (2014) quant à la proposition d'une grille multicritères pour l'évaluation de la qualité en traduction spécialisée (texte à texte). Nous avons choisi cette grille en particulier pour servir de base à nos propres critères car elle est le fruit d'une réflexion qui intègre l'influence de la TAO dans le résultat de la traduction. Nous utilisons dans ce protocole une version simplifiée, et adaptée à l'évaluation des LS, c'est-à-dire qui ne comporte plus de critères non-pertinents pour les LS (correction orthographique, pas exemple), et qui ne traite pas de la typologie des erreurs plus précisément que par thématiques. Nous avons également renommé plusieurs critères pour les rendre plus explicites.

Critère évalué \ score attribué	0	1	2	3	4
Correction grammaticale					
Précision					
Fluidité du discours					
Fidélité de la traduction					
Score total					

Figure 42 Grille d'évaluation des traductions

La grille comporte cinq critères thématiques pour lesquels les experts doivent attribuer un score allant de 0 (valeur la plus faible) à 4 (valeur la plus forte).

- La correction grammaticale des énoncés correspond à la justesse de leur construction. Un score de 4 correspond à un énoncé grammaticalement correct, dans une langue des signes travaillée et la plus standard possible. Chaque erreur de langue doit conduire l'expert à abaisser le score attribué.
- La précision correspond à la justesse des informations traduites, i.e. vocabulaire dédié et précis. Un vocabulaire trop générique impacte négativement le score.
- La fluidité du discours correspond à la facilité avec laquelle on peut le suivre (exposition claire des informations, construction logique, peu d'efforts à fournir pour la compréhension).
- Enfin, la fidélité de la traduction prend en compte le respect de

l'intention du message source, la capacité de l'énoncé traduit à remplir sa fonction (informer, mettre en garde, mode d'emploi...), mais également le nombre et la nature d'éventuels ajouts et omissions.

Dans le calcul final des scores, chaque thématique correspond à un coefficient 1, exception faite de la fidélité qui est coefficient 2, puisque cette thématique regroupe 4 critères à elle seule (intention, fonction, ajouts et omissions). Chaque traduction se voit donc attribuée un score total sur 20 points, que nous pouvons ensuite comparer les uns aux autres pour faire émerger ou non des tendances, et d'évaluer ainsi l'impact du prototype sur la qualité de la traduction. Pour neutraliser le biais lié à la variabilité inter-évaluateurs, chaque traduction doit être évaluée par chaque expert, et est alors caractérisée par la moyenne de chaque note obtenue. En sus de ces scores attribués, le groupe d'évaluateurs est encouragé à commenter : impression globale sur la traduction, pourquoi tel score, etc, pour ajouter au pan qualitatif de la traduction.

Résultats partiels

Cette étude a pour l'instant été menée à petite échelle, mais nous sommes néanmoins en mesure de présenter les résultats issus des tests du protocole concernant la phase à court terme. L'intégralité du protocole n'a pas pu être mise en place à grande échelle dans le contexte des restrictions sanitaires, mais certains participants aux démonstrations ciblées ont pu manipuler librement le logiciel et traduire quelques textes. Nous avons ensuite demandé à ces cinq participants de remplir les questionnaires d'évaluations, à titre indicatif, et pour tester l'efficacité du questionnaire à proprement parler (compréhension des questions, échelles de réponses, redondances...), pour attester au mieux de la validité de nos métriques. A noter que parmi ces 5 participants, 2 sont des connaissances de l'évaluateur, mais que ces deux

participants s'avèrent être les plus critiques (résultats les plus faibles au score SUS). Il ne s'agit donc pas d'une évaluation complète, mais plutôt d'une mise à l'épreuve des questionnaires adressés aux testeurs. La figure 44 présente les scores SUS obtenus, ainsi qu'un résumé des réponses aux autres questions.

Participant	Score SUS	Interprétation	Réutilisation	Bénéfice ?	Top 3 fonctionnalités
1	67.5	Acceptable	Oui	Qualité = Temps -	Ressources, Blocs, carte
2	82.5	Bon	Oui	Temps et qualité +	Ressources, Blocs, carte
3	90	Excellent	Oui	Temps ++ Qualité ?	Blocs, carte, vidéos
4	82.5	Bon	Oui	Qualité et temps +	Blocs, vidéos, TAL
5	67.5	Acceptable	Oui	Qualité et temps +	Blocs, carte, TAL

Figure 43 Résumé des résultats aux questionnaires (version alpha)

Les scores SUS obtenus sont supérieurs à 50 et correspondent à des degrés positifs d'acceptabilité de l'outil. Les Blocs mobiles apparaissent pour ces cinq questionnaires dans le top trois des fonctionnalités favorites proposées par le prototype.

Critère / Réponse	Très en deçà	En deçà	Égale	Supérieure	Très supérieure
Rapidité	0	1	0	3	1

Qualité	0	0	1	3	0
---------	---	---	---	---	---

Figure 44 Résumé des résultats partiels quant à l'influence de l'utilisation du prototype (nombre de participants par réponse)

Deux participants sur les cinq indiquent que plus de trente minutes leur a été nécessaire pour s'accoutumer à l'outil, deux autres indiquent entre quinze et trente minutes, et le cinquième a répondu entre cinq et quinze minutes. Tous répondent souhaiter l'utiliser de nouveau ultérieurement, à un niveau de fréquence élevé (de 3 à 5 où 5 est le maximum). Trois participants estiment leurs traductions un peu meilleures, un de qualité égale, et le cinquième ne sait pas se prononcer. Trois estiment avoir gagné un peu de temps avec l'aide du prototype (fig. 45), un estime avoir gagné beaucoup de temps, et un participant estime en avoir perdu un peu. Dans le protocole complet, cette impression sera à confronter aux mesures objectives. Pour les trois participants ayant répondu qu'ils n'ont pas été en mesure d'exploiter toutes les fonctionnalités du logiciel, ou qu'ils n'ont pas réussi à accomplir tout ce qu'ils auraient souhaité, les raisons invoquées sont le manque de temps (test trop bref) pour les trois, la complexité des manipulations pour un, et pour un autre le fait qu'il ne connaissait pas de prime abord toutes les fonctionnalités intégrées (malgré la présentation rapide du logiciel en début d'expérience). Deux remarques libres émergent de ces questionnaires. La première concerne l'efficacité de l'outil, qui d'après ce participant, serait plus destiné à assister la traduction de textes plus longs. La seconde concerne une éventuelle possibilité d'imprimer le résultat assemblé dans la zone de travail une fois la préparation terminée. On peut interroger le but final de cette volonté d'impression, et le relier à un besoin identifié d'une forme de prompteur.

Ces résultats constituent une première piste pour proposer des améliorations, que cela soit pour le protocole d'évaluation en lui-même ou pour le prototype qu'il concerne. La section suivante s'intéresse aux ajouts

et aux modifications à apporter au prototype pour l'améliorer, ainsi qu'à son devenir à plus long terme.

5.2 Retour sur les résultats préliminaires

La section 5.2.1 s'intéresse d'une part à l'analyse des résultats obtenus lors de l'évaluation partielle, et la section 5.2.2 aux améliorations à apporter au protocole qui en découlent.

5.2.1 Analyse des réponses

Concordancier

En ce qui concerne le concordancier, les utilisateurs potentiels ont exprimé leur intérêt pour l'outil, et se sont accordés sur son utilité, bien qu'en l'état actuel, la base de données des alignements reste trop limitée pour qu'il puisse être aussi performant que ses équivalents texte à texte. La possibilité d'accéder à du contenu préalablement traduit permet de nourrir une réflexion sur sa propre traduction, et de guider certains choix de traduction. L'amélioration la plus importante consiste à enrichir la base de données des alignements, en nombre d'alignements d'une part, et en variété de contenus d'autre part. Pour l'heure, elle ne permet d'accéder qu'à des alignements issus de discours journalistiques. Le type de discours et l'acte de langage associé influent sur la forme que prend ce dernier, et conséutivement sur la forme de sa traduction (vocabulaire, emphase, prise de rôle).

Environnement de travail intégré

Le score SUS moyen obtenu est de 78 sur 100, ce qui correspond au degré d'acceptabilité **bon** sur l'échelle de Bangor, Kortum et Miller. Les items ayant fait perdre le plus de points au système concernent la nécessité du soutien d'un technicien pour utiliser correctement le système, ainsi que

l'aisance avec laquelle la majorité des personnes pourront apprendre à l'utiliser. Ces deux points nous ramènent au postulat de départ que le métier étant jusqu'à lors peu outillé, les professionnels pourraient être frileux quant à l'utilisation d'un prototype complet. Il apparaît d'après ces réponses que, si le système ne leur apparaît pas comme étant pénible à utiliser, il demeure néanmoins dense pour des utilisateurs peu familiers avec l'outil informatique. Pour répondre à cette demande d'assistance de la part d'un technicien, ou pour alléger le temps de prise en main général, une proposition pourrait être d'élaborer un tutoriel plus complet pour détailler le mode d'emploi de chaque fonctionnalité, ainsi qu'un guide de l'utilisateur plus global pour traiter des interactions qu'entretiennent les fonctions entre elles.

5.2.2 Modifications et améliorations du protocole

Les réponses à la seconde liste de questions (hors SUS donc) marquent un intérêt pour l'outil en cela que chacun répond vouloir l'utiliser de nouveau dans le futur (sur une échelle de 1 à 5 où 1 correspond à jamais et 5 à très souvent, la réponse la plus faible obtenue est de 3). Ces réponses soulignent plusieurs points important dans l'optique d'améliorer le prototype, mais nous permettent également d'ajuster le protocole en conséquence. Nous les présentons ci-après par catégories.

Améliorations suggérées pour le prototype par les utilisateurs tests

Les participants évoquent la possibilité d'imprimer les notes une fois le travail d'analyse terminé, de sorte à pouvoir l'emporter indépendamment de l'ordinateur. Ils suggèrent également d'améliorer le tutoriel intégré : le rendre plus explicite, et couvrir toutes les fonctionnalités disponibles au moment du test.

Améliorations possibles pour le protocole

Deux remarques émergent des premiers tests. La première serait soit d'allonger le temps alloué au test pour permettre aux utilisateurs les moins familiers de ne pas être désavantagés par le temps de prise en main, soit de détailler plus avant la présentation du logiciel et de ses fonctionnalités proposées pour faciliter l'utilisation du logiciel. La seconde remarque est de proposer à la traduction des textes plus longs dès la première étape de l'évaluation, pour inciter les participants à faire usage de toutes les fonctionnalités.

Les participants les plus à l'aise avec l'informatique se familiarisent plus rapidement au logiciel, tandis que ceux qui le sont moins rapportent des difficultés à identifier les fonctionnalités. Ils parviennent cependant à les manipuler correctement. Parmi les commentaires recueillis, un participant indique que le logiciel lui a fait perdre un peu de temps pour la traduction d'un texte court, mais que ce dernier trouverait plus d'utilité dans le cas de textes longs. Ce dernier point pourra être exploré plus avant lors de l'évaluation à long terme, qui laisse aux participants volontaires beaucoup plus de temps pour manipuler l'outil, et le choix libre des textes qu'ils traduisent en l'utilisant. L'échantillon étant réduit, il convient de modérer les résultats.

Chapitre 6

Conclusion et perspectives

Ce dernier chapitre s'ouvre sur un rappel de nos problématiques et de nos objectifs (section 6.1). Il résume ensuite l'intégralité de ce manuscrit et de ses contributions (section 6.2), avant de proposer quelques perspectives (6.3).

6.1 Retour sur les objectifs

Dans le chapitre 1, nous avons présenté les caractéristiques ainsi que la complexité des langues des signes. Ces langues, à part entière, disposent d'une grammaire et d'un lexique propre, qui reposent sur une modalité visuo-gestuelle, l'utilisation de l'espace ainsi que l'iconicité. Ces langues ne disposent pas de formes écrites. Nous avons exposé un besoin croissant pour du contenu accessible en langue des signes, et donc traduit. Ce chapitre présente également le travail de traduction, plus particulièrement la traduction des langues signées ainsi que la traduction assistée par ordinateur. Nous avons constaté que ces outils ne prenaient pas en charge l'assistance à la traduction des langues des signes. Nous avons alors posé l'objectif de proposer un outil qui soit en mesure de le faire, pour assister les encore rares traducteurs professionnels des LS dont la charge de travail est toujours croissante.

L'état de l'art présenté chapitre 2 s'intéresse aux caractéristiques des logiciels de TAO. Nous en dégageons trois grands principes d'organisation, à savoir l'utilisation prépondérante d'une forme écrite éditable, la disponibilité d'une mémoire de traduction ainsi que ce que nous avons appelé le principe de linéarité selon lequel la concaténation de segments cibles traduits correspond à la traduction des segments sources concaténés, principe qui n'est pas respecté pour les langues signées. Ce chapitre répertorie également plusieurs points clés de l'évaluation de la traduction, mais également de la qualité et la pertinence de l'assistance apportées par les outils de TAO.

Toutes ces observations donnent naissance à plusieurs questions quant à la création d'un outil capable de gérer les LS tout en apportant une assistance notable aux traducteurs. Dans un premier temps, il était nécessaire de chercher à formaliser la tâche de traduction pour pouvoir l'assister. Pour

cela, nous avons étudié le processus de traduction des LS : quelles sont ses tâches constitutives ? Sont-elles systématiques et ordonnées ?

Ces questions ont eu pour but de spécifier les questions de recherche de ces travaux, à savoir : quelles spécifications pour un logiciel de TAO à destination des LS ? Quels principes respecter, comment les constituer ? Quelles aides peuvent-ils apporter au métier de traducteur des LS ?

La recherche s'est donc articulée autour d'objectifs découlant de ces questions. Dans un premier temps, il s'agit d'analyser les points clés des logiciels de TAO ainsi que les étapes constitutives du processus de traduction. En parallèle, nous avons également identifié les besoins des professionnels tout comme les obstacles les plus fréquemment rencontrés, de sorte à pouvoir déterminer quelles aides logicielles pourraient leur bénéficier et ainsi dresser une première spécification logicielle.

6.2 Principales contributions et résultats

La première contribution de nos travaux concerne l'identification des tâches constitutives du processus de traduction des LS. Le chapitre 3 détaille les expériences mises en œuvre pour ce faire, à savoir d'une part l'analyse de vidéos de traducteurs en situation de traduction, issues d'un corpus pré-existant, et d'autre part l'organisation d'un brainstorming avec des professionnels du métier. Ces deux études nous ont permis de repérer et de caractériser les étapes de la traduction, mais également les difficultés les plus courantes lors de cette activité, ainsi que les outils actuellement sollicités pour y parvenir.

Ces observations et conclusions initiales ont permis une première formalisation du métier, qui gagnerait à être enrichie par exemple en réitérant les analyses sur un corpus plus conséquent, tant en terme de participants que de textes traduits.

La seconde contribution correspond à l'élaboration d'un cahier des charges quant à la création d'un logiciel de TAO à destination des LS. C'est l'objet des chapitres 3 et 4. La seconde partie du chapitre 3 reprend une à une les notions clés de l'aide informatisée à la traduction des LS, ainsi que les choix techniques mis en place pour les réaliser au travers de nos deux prototypes d'outils. Le chapitre 4 détaille cette réalisation technique, et ainsi la conception de deux logiciels, qui sont le fruit de nos travaux. Le premier est un concordancier bilingue français / Langue des Signes Française, librement accessible en ligne, que nous avons mis en place dans le but de servir de mémoire de traduction (principe essentiel de la TAO mais difficilement transposable aux LS en l'état). Le second est un prototype d'environnement de travail intégré pour le traducteur en LS, qui regroupe divers modules d'aide à la traduction issus de nos observations précédentes. Il comporte également une zone de travail dédiée, dont l'ergonomie a été pensée pour répondre au mieux aux besoins de la traduction des LS, faite de Blocs déplaçables à même d'accueillir plusieurs types de contenu (texte, vidéo, image...).

C'est une première tentative de rapprochement entre TAO et LS. Basés sur des observations et des réflexions issues directement des professionnels concernés, ces outils ont vocation à s'améliorer au fil du temps et des retours que leur utilisation pourra susciter.

La troisième contribution principale de ces travaux concerne les méthodes d'évaluation de ces outils. Les pré-évaluations et les présentations publiques menées jusqu'à lors, comme décrites dans le chapitre 5, permettent d'affirmer un intérêt pour les outils et les solutions que nous proposons. Les résultats obtenus lors des évaluations finales déterminent l'évolution des outils, et contribuent à identifier des modifications ou des ajouts nécessaires pour améliorer la performance et la pertinence de l'assistance, d'une part, et

le confort de l'utilisateur d'autre part. Afin de proposer une méthode d'évaluation pour nos prototypes, nous avons mené une réflexion particulière quant à la question de l'évaluation des traductions.

6.3 Perspectives

Cette première tentative de logiciel de TAO à destination des LS a abouti à des prototypes fonctionnels. Ces prototypes nécessitent des évolutions et certains de nos choix sont à confronter à la pratique pour les discuter et les améliorer, ainsi qu'améliorer l'ergonomie et l'utilisabilité du système. De nouvelles perspectives issues des retours une fois le logiciel testé, ainsi que des données collectées au travers de nouvelles études nous permettront de poursuivre le développement d'une application qui se veut issue de la collaboration entre chercheurs et professionnels de terrain.

Dans un premier temps, il s'agira notamment d'inclure de nouveaux modules, tels qu'un prompteur pour assister la livraison finale, par exemple en faisant défiler les Blocs sélectionnés lorsque le traducteur se filme en fin de processus. Le concept du prompteur télévisuel est de faire défiler à vitesse de lecture le discours qui doit être tenu devant la caméra. Ici l'idée est tout à fait semblable : faire défiler devant le traducteur les éléments constitutifs de sa traduction, à savoir les Blocs qu'il a préalablement préparés. Les arbres pouvant rapidement devenir conséquent, il serait libre de sélectionner parmi tous les Blocs créés seulement ceux qu'il désire afficher, ainsi que la vitesse de défilement. La forme que pourrait prendre cette fonctionnalité n'a pas encore été définie, et mérite d'autres consultations avec les professionnels concernés.

Nous envisageons également de faire une plus grande place aux modules de traitement automatique des langues, afin de générer du contenu directement à partir du texte source. Un module d'extraction et de résumé automatique

pourrait être utile, tout comme une fonction qui puisse permettre de repérer ou sélectionner dans un texte à traduire des éléments numériques tels que les pourcentages et d'en réaliser automatiquement une représentation graphique qui sera utile pour élaborer sa traduction, visuelle et donc basée sur cette représentation, en LSF.

Dans un second temps, nous envisageons d'intégrer des fonctionnalités supplémentaires qui seraient cette fois issues de la recherche d'autres collègues. Un outil permettant de créer un schéma ou de dessiner directement dans le logiciel serait un apport majeur, considérant la nature visuelle des LS et le recours à cette pratique lors de la prise de notes de la part de personnes sourdes ou de la déverbalisation chez les traducteurs et interprètes. Le traitement automatique des langues *signées* peut également être mis à contribution. En effet, bien que récentes, ces études s'intéressent en particulier à la formalisation des LS et à l'élaboration de modèles descriptifs permettant la génération de contenu en LS, pour en faciliter le traitement automatique. Ce type d'outil pourrait servir pour générer un prompteur anonymisé. Elles s'intéressent aussi à la reconnaissance automatique de signes isolés dans une vidéo (Belissen, 2018 ; Huang, 2019), l'aide à la segmentation du flux et le sous-titrage automatique (Bull, 2020), ou l'anonymisation des signeurs (Bigand, 2019). Ces fonctionnalités pourraient être envisagées dans le cadre de notre environnement intégré.

Ces ajouts successifs conduisent vers un objectif à plus long terme, à savoir rendre le logiciel fonctionnel également dans l'autre sens de traduction, des LS vers les langues écrites.

Cet outil a vocation à être le fruit de collaborations variées sur le thème des LS et de leur traduction. On peut ainsi, au même titre que la TAO est enseignée dans les cursus de formations des traducteurs de texte à texte, envisager une place pour ces outils dans les formations des traducteurs et

interprètes en LS. La collaboration facilite la diffusion, et la diffusion participe à la sensibilisation du public concerné.

L'évaluation de la qualité de la traduction, tout comme des outils qui ont permis de les produire, constituent deux notions importantes dans ce mémoire de thèse. À long terme, nous aimerais explorer la question de l'évaluation de la traduction des LS. Le projet européen QT21²³ s'intéresse à l'évaluation de la traduction automatique, et propose un cadre permettant de décrire et de définir des métriques d'évaluation de qualité de la traduction (MQM, Multidimensional Quality Metrics). On retrouve notamment dans les travaux associés à ce projet, une typologie détaillée des erreurs de traduction ainsi qu'un guide pour les annotateurs d'erreurs assorti d'un arbre de décision. QT21 cherche à améliorer les modèles de traduction automatique existants de sorte à les rendre plus performants lorsqu'il s'agit de traduire des langues considérées comme « challenging ». Il s'agit de langues européennes qui sont soumises à une grammaire particulière, ou à un ordre des mots qui diffère de la majorité des langues. En cela, puisque c'est également le cas pour les LS, ces travaux constituent une base solide pour étendre la problématique non plus seulement vers les langues vocales mais également vers les LS, et entamer une réflexion sur un modèle d'évaluation commun pour leur traduction.

Cette recherche, ces recherches, et leurs fruits, de par les points de vue nouveaux qu'elles apportent sur les langues et les pratiques, peuvent trouver leur place au sein d'un cursus universitaire. En tant qu'assistance possible à la traduction ou à l'exploration de la langue (dans le cas du concordancier), elles concernent directement la formation des professionnels du milieu. Elles offrent également de nouvelles perspectives linguistiques et informatiques, pour que la technologie puisse se porter au service des LS et

²³ <https://www.qt21.eu/>

de leurs utilisateurs, au quotidien, et en tant qu'égales des langues vocales.

Bibliographie

- Banerjee, S. and Lavie, A. (2005) "METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments" in *Proceedings of Workshop on Intrinsic and Extrinsic Evaluation Measures for MT and/or Summarization at the 43rd Annual Meeting of the Association of Computational Linguistics (ACL-2005), Ann Arbor, Michigan, June 2005*
- Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies*, 4(3), 114-123.
- Belissen, V., Braffort, A., & Gouiffès, M. (2018, April). Analyse de vidéos de Langue des Signes Française à des fins de reconnaissance automatique. In *Handiversité 2018*.
- Bigand, F., Prigent, E., & Braffort, A. (2019, October). Retrieving Human Traits from Gesture in Sign Language: The Example of Gestural Identity. In *Proceedings of the 6th International Conference on Movement and Computing* (pp. 1-4)
- Braffort, A. (1996). Reconnaissance et comprehension de gestes, application à la langue des signes,
- Brook, J. (1996). SUS: a "quick and dirty" usability scale. *Usability evaluation in industry*.
- Bull, H., Gouiffès, M., & Braffort, A. (2020, August). Automatic segmentation of sign language into subtitle-units. In *European Conference on Computer Vision* (pp. 186-198). Springer, Cham.

Cantin, Y. (2021). “Interprètes en langue des signes de la Révolution jusqu’au Moyen Âge: une analyse historique,” in *Traductologie et langues des signes*, ed. F. Encrevé (Paris: Classiques Garnier), 17–30

Cuxac, C., & Boutet, D. (2014). *Corpus LS-Colin sur plusieurs genres discursifs (Laurent Vallo, Nasredine Chab et Anthony Guyon)* (Version 1) <https://doi.org/10.34847/COCOON.6D4B96B9-83C2-38B4-8FD6-ACB62E57DE4D>

Cuxac, C. (1996). *Fonctions et structures de l'iconicité des langues des signes: analyse descriptive d'un idiolecte parisien de la langue des signes française* (Doctoral dissertation, Paris 5).

Cuxac, C. (2000) La langue des signes française, les voies de l'iconicité, Paris, Ophrys.

Dancette, J., (1998) Parcours de la traduction, Lille, Presses Universitaires de Lille.

Delisle, J. (1980). L’analyse du discours comme méthode de traduction, Ottawa. *Éditions de l’Université d’Ottawa*.

Encrevé, F. (2014) Les spécificités historiques des formations d’interprètes LSF/français en France. *Double Sens, revue de l’association française des interprètes et traducteurs en langue des signes*. 2, 7–18.

Fiederer, R., and O’Brien S. (2009). Quality and Machine Translation: A realistic objective? in *The Journal of Specialised Translation* 11, 52–74
Filhol, M. (2021). *Modélisation, traitement automatique et outillage logiciel des langues des signes* (Doctoral dissertation, Université Paris-Saclay).

Garcia, I. (2011). Translating by post-editing: is it the way forward? in *Machine Translation* 25, 217–237.

Gronier, G. & Baudet, A. (2021) Psychometric Evaluation of the F-SUS: Creation and Validation of the French Version of the System Usability Scale, *International Journal of Human–Computer Interaction*, 37:16, 1571–1582.

Guiraud, P. (1968). Langage et théorie de la communication. In A. Martinet (éd.), *Le Langage*, 145-68.

Guitteny, P. (2007) Langue des signes et schémas, *Trait. Autom. des Langues*, 48(3), 201-229.

Hanke, T. (2004). HamNoSys-representing sign language data in language resources and language processing contexts. In *LREC*, volume 4, 1-6

Harvey, M. (2002). What's so special about legal translation?. *Meta: journal des traducteurs/Meta: Translators' Journal*, 47(2), 177-185.

Heyn, M. (1998) Translation memories: Insights and prospects. In L. Bowker, M. Cronin, D. Kenny & J. Pearson (Eds.), *Unity in diversity? Current trends in translation studies* (123-136), Manchester: St. Jerome Publishing.

Huang, J., Zhou, W., Zhang, Q., Li, H., & Li, W. (2018, April). Video-based sign language recognition without temporal segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 32, No. 1).

Hutchins, J. (2005). Example-based machine translation: a review and commentary. *Machine Translation*, 19(3), 197-211.

Kaczmarek, M., Filhol, M . (2020) Use cases for Sign Language Concordancer , in *proceedings of 9th Workshop on the Representation and Processing of Sign Languages (RPSL2020)*, Marseille.

Kaczmarek, M., Filhol, M.,(2020) Alignments Data Base for a Sign Language concordancer, in *proceedings of 12th International Conference on Language Resources and Evaluation (LREC 2020)*, Marseille.

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

Kaczmarek ,M., Larroque, A. (2021) Traduction Assistée par Ordinateur des Langues des Signes : élaboration d'un premier prototype, in *proceedings of Traitement Automatique des Langues Naturelles (TALN 2021)*, Lille.

Koehn, P. (2009). A process study of computer-aided translation. *Machine Translation*, 23(4), 241-263.

Kristoffersen, J. H., & Troelsgard, T. (2012). The electronic lexicographical treatment of sign languages: The Danish Sign Language Dictionary. In *Electronic Lexicography*, Oxford University Press.

Langer, G., Müller, A., & Wähl, S. (2018, May). Queries and Views in iLex to Support Corpus-based Lexicographic Work on German Sign Language (DGS). In *sign-lang@ LREC 2018* , 107-114.

Larose, R. (1998) Méthodologie de l'évaluation des traductions, in *Meta*, 43(2), 163-186.

Mellinger, C. D. (2018) Re-thinking translation quality: Revision in the digital age. *Target. International Journal of Translation Studies*, 30(2), 310-331.

Meurant, L. (2015) Corpus LSFB. Un corpus informatisé en libre accès de vidéos et d'annotations de la langue des signes de Belgique francophone (LSFB). Laboratoire de Langue des signes de Belgique francophone (LSFB-Lab). FRS-F.N.R.S et Université de Namur.

Miyazaki, T., Morita, Y., & Sano, M. (2020, May) Machine translation from spoken language to Sign language using pre-trained language model as encoder. In *Proceedings of the LREC2020 9th Workshop RPSL* (pp. 139-144).

O'Brien, S. (1998) Practical experience of computer-aided translation tools in the software localization industry. *Bowker, L.; Cronin, M.; Kenny, D*, 115-122.

Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002) Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th*

annual meeting of the Association for Computational Linguistics (pp. 311-318).

Reinke, U. (2004) Translation memories: Systeme Konzepte linguistische Optimierung. Frankfurt am Main: Lang.

Reiss, K. (2009) Problématiques de la traduction, Paris, Ed. ECONOMICA.

Seleskovitch, D. (1979) *Traduction et mécanismes du langage*.

Seleskovitch, D., & Lederer, M. (1984) *Interpréter pour traduire*. Didier érudition.

Shreve, G. M. (2006) The deliberate practice: translation and expertise. *Journal of translation studies*, 9(1), 27-42.

Somers, H. (2005) Machine translation: latest developments. In *R. Mitkov, éditeur: The Oxford Handbook of Computational Linguistics*, Oxford University Press, 512-528.

Toudic, D., Hernandez Morin, K., Moreau, F., Barbin, F., & Phuez, G. (2014) Du contexte didactique aux pratiques professionnelles: proposition d'une grille multicritères pour l'évaluation de la qualité en traduction spécialisée. *ILCEA. Revue de l'Institut des langues et cultures d'Europe, Amérique, Afrique, Asie et Australie*, (19).

Webb, L. E. (1998) Advantages and disadvantages of translation memory. *A Cost-Benefit Analysis*.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., Sloetjes, H. (2006) ELAN: a Professional Framework for Multimodality Research. In: Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation.

Wolfe, R., Efthimiou, E., Glauert, J., Hanke, T., McDonald, J., & Schnepf, J. (2016) eds. Special issue: recent advances in sign language translation and avatar technology, Springer International Publishing.

ANNEXES

ANNEXE A

Liste des logiciels de TAO étudiés au cours des travaux de recherche.

- SDL Trados Studio : <https://www.trados.com/fr/products/trados-studio/>
- MemoQ : <https://www.memoq.com/fr>
- OmegaT : <https://omegat.org/fr/>
- Similis : <http://www.lingua-et-machina.com/>
- Wordfast : <http://www.wordfast.net/>
- Déjà Vu : <http://www.atril.com/>
- Anaphraseus : <http://sourceforge.net/projects/anaphraseus/files/>
- Ecco : <http://primatrans.tripod.com/ecco/>
- Olifant (MT seulement) :
<http://www.translate.com/technology/tools/olifant/ReadMe.htm>

ANNEXE B

Textes d'élicitation pour la constitution du corpus vidéo de traducteurs.

Textes session de rodage

- (1) L'aide humanitaire commençait à arriver sur la côte caraïbe du Nicaragua, après le passage dévastateur du cyclone Felix qui a fait au moins 38 morts et 50.000 sinistrés dans cette région pauvre, majoritairement peuplée d'indiens Miskitos.
- (2) Le président George W. Bush a catégoriquement exclu jeudi que les troupes américaines se retirent battues d'Irak, 24 heures après avoir admis pour la première fois une analogie possible avec la guerre du Vietnam.
- (3) Une partie de la ville d'Oxford (centre) était engloutie sous plus d'un mètre d'eau mercredi, alors que le pic de la Tamise se déplaçait vers l'est, en direction de Londres, menaçant de déborder sur plusieurs communes comme Reading ou Henley.

Textes courts

- (1) Une forte tempête qui sévissait dans la nuit de jeudi à vendredi en mer du Nord a contraint plusieurs groupes pétroliers à fermer des plateformes au large de la Norvège, tandis que Royaume-Uni, Pays-Bas et Allemagne étaient en alerte pour parer à d'éventuelles déferlantes annoncées comme les pires depuis 20 ans.
- (2) Des Casques bleus de l'ONU ont été déployés jeudi sur la frontière entre le Liban et Israël, où des tirs ont été échangés mercredi soir

entre soldats libanais et israéliens, premier incident armé depuis la fin de la guerre menée cet été par Israël contre le Hezbollah chiite libanais.

- (3) L'état d'urgence a été décrété dimanche en Californie (ouest) où 11 incendies ont fait un mort, provoqué des dizaines de milliers d'évacuations et ravagé plus de 12.000 hectares de broussailles et des bâtiments, menaçant aussi des maisons de célébrités dans la ville de Malibu.
- (4) La cour d'assises spéciale jugeant Yvan Colonna pour l'assassinat du préfet Erignac s'est offert dimanche une audience à ciel ouvert le long du golfe d'Ajaccio, un déplacement qui a ébranlé les thèses de l'accusation, a claironné la défense, tandis que les parties civiles les voyaient au contraire "confortées".
- (5) La guérilla des FARC a fait mardi un geste spectaculaire avec l'annonce de la libération de trois otages dont Clara Rojas, l'assistante d'Ingrid Betancourt, tout en opposant une fin de non-recevoir aux dernières propositions du président colombien Alvaro Uribe.
- (6) Le président américain George W. Bush se rendra pour la première fois en janvier à Jérusalem et en Cisjordanie pour favoriser la conclusion d'un accord de paix entre Israéliens et Palestiniens dans les derniers mois de son mandat.
- (7) La police égyptienne a arrêté mardi dix personnes soupçonnées d'implication dans le triple attentat qui a frappé la veille la ville touristique de Dahab sur la mer Rouge, tuant 18 personnes dont des étrangers.

- (8) L'Égypte a refusé jeudi le passage par le canal de Suez du Clemenceau, en route vers l'Inde pour son désamiantage, quelques heures après une spectaculaire opération coup de poing de Greenpeace, dont deux militants sont montés à bord du porte-avions français.
- (9) Un jeune adjoint au shérif local d'une petite ville du Wisconsin (nord des États-Unis) a tué dimanche, au petit matin au cours d'une fusillade, six personnes, dont sa petite amie, participant à une soirée, ont annoncé les autorités locales.
- (10) Le Premier ministre irakien Nouri al-Maliki s'est rendu mercredi dans la ville sainte chiite de Kerbala, au sud de Bagdad, au lendemain de violents affrontements pendant un pèlerinage qui ont fait au moins 52 morts et 300 blessés, a indiqué son bureau.
- (11) La Russie a recommencé jeudi matin à alimenter l'Europe en pétrole via l'oléoduc Droujba, qui avait été fermé lundi en raison d'un différend avec le Bélarus, a annoncé le vice-président du monopole russe des oléoducs Transneft, Sergueï Grigorev.
- (12) Bretagne, Basse-Normandie, Loire-Atlantique et Vendée ont subi dimanche leur première grande tempête hivernale qui a entraîné au large du Finistère la disparition d'un marin, des avaries de bateaux et la chute en mer de conteneurs transportés par des cargos.

Textes longs

(L1)

L'un des avocats de Jean-Louis Gergorin, Me Paul-Albert Iweins, a expliqué mercredi à la cour d'appel de Paris qu'il comprenait que Dominique de Villepin n'ait pas dit toute la vérité dans l'affaire Clearstream car il devait protéger le président Chirac. Jean-Louis Gergorin et le général Philippe Rondot affirment qu'à plusieurs reprises en 2004, Dominique de Villepin s'est recommandé d'instructions du président de la République pour enquêter sur les listings Clearstream qui mettaient notamment en cause Nicolas Sarkozy. "C'est tout à son honneur: quand on est ministre, on n'implique pas le chef de l'Etat", a plaidé Me Iweins. Quant aux rendez-vous que Dominique de Villepin nie avoir eus avec Jean-Louis Gergorin au premier semestre 2004, "je sais pourquoi ces rendez-vous ne sont pas reconnus", a-t-il dit, "car Jean-Louis Gergorin n'en a pas parlé au départ", lors de ses premières dépositions. Aux yeux de l'avocat, il est donc normal que Dominique de Villepin, lorsqu'il est entendu, "confirme les déclarations de Jean-Louis Gergorin, mais il ne va pas aller plus loin". "Mettez-vous à sa place", a-t-il supplié, en rappelant la frénésie médiatique de l'époque. Ainsi, à la sortie de la garde à vue de Jean-Louis Gergorin, en juin 2006, l'avocat se souvient avoir été assailli par une nuée de caméras, avec une seule question à la bouche: "Est-ce que votre client met en cause Dominique de Villepin? Si j'avais dit +oui+, le gouvernement sautait."

Publié le 12/09/2009 à 12:15 UTC par l'AFP

(L2)

Kurt Cobain rencontre Courtney Love après un concert de *Nirvana* à Portland. Ils n'échangent que quelques mots, mais Courtney est immédiatement séduite. Selon le journaliste Everett True, ils n'auraient été réellement présentés l'un à l'autre qu'en mai 1991 à un concert *L7* et *Butthole Surfers*. Par la suite, en apprenant par Dave Grohl que l'attirance est réciproque, Courtney cherche à se rapprocher de Kurt. Ils commencent à se voir régulièrement à partir de l'automne 1991, en partie rapprochés par leur consommation commune de drogue.

Début 1992, Courtney Love apprend qu'elle est enceinte de Kurt. Le 24 février 1992, quelques jours après la fin d'une tournée de *Nirvana* sur la côte Pacifique, ils se marient à Hawaii. Leur fille Frances Bean Cobain naît le 18 août 1992. Son nom fait référence à Frances McKee des *Vaselines* et non à Frances Farmer, comme il est parfois affirmé. Son second prénom, « Bean » (haricot), fait référence à la forme du bébé sur les échographies pendant la grossesse de Courtney Love. Michael Stipe, chanteur de *R.E.M.*, dont Cobain admire la carrière et le travail, est choisi comme parrain de leur fille. Stipe deviendra par ailleurs un ami proche de Courtney Love par la suite.

Extrait de Wikipédia (fr) du 2/10/2017

ANNEXE C

Données exportées d'ELAN

T1-1

00:00:09.207 Discussing signs x2
00:02:29.261 Lexical search x3
00:01:08.469 Discussing signs x2
00:12:19.630 Lexical search
00:00:37.342 Picture search
00:03:06.953 Lexical search
00:02:58.155 Lexical search
00:01:20.067 Discussing signs
00:00:55.733 Lexical search
00:01:56.695 Encyclopedic look-up
00:01:58.858 Encyclopedic look-up
00:00:21.252 Picture search
00:01:58.047 Discussing signs
00:02:01.422 Discussing signs
00:00:18.401 Lexical search
00:03:03.589 Discussing signs
00:02:20.026 Discussing signs
00:04:26.083 Lexical search
00:00:17.207 Discussing signs
00:03:46.160 Encyclopedic look-up
00:02:21.231 Encyclopedic look-up
00:01:38.865 Discussing signs
00:02:14.326 Lexical search
00:01:10.837 Discussing signs
00:00:48.336 Discussing signs
00:01:15.028 Map search
00:06:39.521 Ordering
00:07:07.287 Ordering
00:01:17.136 Ordering
00:01:51.292 Ordering

T2-1

00:02:33.270 Discussing signs
00:02:22.062 Ordering
00:07:44.371 Encyclopedic look-up
00:01:30.627 Ordering
00:03:06.813 Ordering

00:01:13.043 Lexical search
00:03:28.982 Lexical search
00:01:14.076 Lexical search
00:00:34.500 Ordering
00:02:05.624 Discussing signs
00:03:28.085 Lexical search
00:03:52.664 Ordering
00:06:20.174 Ordering
00:04:07.471 Lexical search
00:03:50.000 Lexical search + discussing signs
00:03:17.307 Ordering
00:02:19.335 Ordering
00:03:19.556 Ordering
00:01:47.139 Ordering
00:01:22.945 Ordering

T3-1

00:00:42.960 Discussing signs
00:01:41.859 Discussing signs
00:01:00.525 Ordering
00:02:11.517 Lexical search
00:00:24.468 Discussing signs
00:00:42.468 Lexical search
00:01:45.732 Ordering
00:00:56.572 Map search
00:02:43.114 Ordering

T1-2

00:00:57.558 Picture search
00:00:56.210 Discussing signs
00:01:14.066 Map search
00:00:51.001 Map search
00:00:59.611 Map search
00:04:14.074 Ordering
00:00:36.854 Discussing signs
00:00:14.783 Discussing signs
00:00:51.750 Discussing signs
00:00:52.250 Ordering
00:03:43.233 Discussing signs
00:00:41.431 Ordering
00:00:38.379 Discussing signs
00:00:37.850 Ordering
00:00:43.173 Ordering
00:00:22.587 Ordering

00:00:40.549 Memorising

T2-2

00:00:23.709 Discussing signs
00:00:44.734 Memorising
00:01:40.261 Ordering
00:01:59.666 Definition look up
00:01:09.261 Discussing signs
00:01:03.476 Ordering
00:03:46.359 Ordering
00:01:57.819 Discussing signs
00:01:27.117 Discussing signs

T3-2

00:01:33.189 Map search
00:02:23.444 Map search
00:00:42.316 Ordering
00:02:09.639 Map search
00:01:34.666 Ordering
00:00:57.742 Map search
00:01:01.432 Ordering
00:00:52.910 Ordering
00:00:13.861 Memorising

ANNEXE D

Textes proposés lors d'ateliers

Texte 1 : Une partie de la ville d'Oxford (centre) était engloutie sous plus d'un mètre d'eau mercredi, alors que le pic de la Tamise se déplaçait vers l'est, en direction de Londres, menaçant de déborder sur plusieurs communes comme Reading ou Henley.

Texte 2 : Les lampreies sont caractérisées par une large ventouse entourant la bouche, et d'appareil lingual complexe dont l'axe est constitué par une grosse tige cartilagineuse : le cartilage piston. Des plaques antérieures viennent constituer une armature à la région buccale. Ces plaques appartiennent au splanchnocrâne qui se prolonge caudalement par une série de travées cartilagineuses.

Texte 3 : Notre méthode pour changer une corde cassée:

- Faites un nœud autour du tourillon. Pour les cordes en nylon, un nœud supplémentaire est conseillé.
- Glissez la corde dans l'œillet à travers la table, puis dans la cheville. Serrez la corde, en vous assurant que la corde fasse au moins deux tours de cheville. Ajustez la hauteur de la note.
- Coupez ensuite les bouts de cordes qui dépassent (à l'intérieur de la caisse et au dessus de la cheville, pour éviter les vibrations.

Texte 4 :

Aiguilles chagrines
fils invisibles du temps
ourlez mon linceul

Retombé au sol

Le cerf-volant

A égaré son âme.

Textes proposés à la traduction dans le cadre du protocole d'évaluation

Texte court 1 :

Kurt Cobain rencontre Courtney Love après un concert de *Nirvana* à Portland. Ils n'échangent que quelques mots, mais Courtney est immédiatement séduite. Selon le journaliste Everett True, ils n'auraient été réellement présentés l'un à l'autre qu'en mai 1991 à un concert *L7* et *Butthole Surfers*. Par la suite, en apprenant par Dave Grohl que l'attraction est réciproque, Courtney cherche à se rapprocher de Kurt. Ils commencent à se voir régulièrement à partir de l'automne 1991, en partie rapprochés par leur consommation commune de drogue.

Début 1992, Courtney Love apprend qu'elle est enceinte de Kurt. Le 24 février 1992, quelques jours après la fin d'une tournée de *Nirvana* sur la côte Pacifique, ils se marient à Hawaii. Leur fille Frances Bean Cobain naît le 18 août 1992. Son nom fait référence à Frances McKee des *Vaselines* et non à Frances Farmer, comme il est parfois affirmé. Son second prénom, « Bean » (haricot), fait référence à la forme du bébé sur les échographies pendant la grossesse de Courtney Love. Michael Stipe, chanteur de *R.E.M.*, dont Cobain admire la carrière et le travail, est choisi comme parrain de leur fille. Stipe deviendra par ailleurs un ami proche de Courtney Love par la suite.

Source : Extrait Wikipédia

Texte court 2 :

Le swahili, langue de travail de l'Union Africaine, est également langue officielle de la Communauté des pays d'Afrique de l'Est qui regroupe le Kenya, la Tanzanie, l'Ouganda, le Rwanda, le Burundi, le Soudan du sud et

depuis le printemps 2021, La République démocratique du Congo. Il est également une des langues officielles de la SADC ou Communauté de Développement Economique de l'Afrique australe. Il est présent dans les publications des grandes institutions internationales (organismes onusiens, diverses ONG...). Diffusé sur les antennes des chaînes mondiales d'information (RFI, BBC, Voix de l'Amérique, Deutsche Welle, Pars Today, Radio Chine Internationale...) il est, en dehors de sa région d'origine, enseigné dans de nombreuses universités dans le monde et dans le secondaire en Afrique du Sud.

La tradition littéraire écrite en swahili remonte à plusieurs siècles (à l'origine en caractères arabes, depuis la fin du XIXème siècle en caractères latins). La place du swahili dans l'enseignement en Tanzanie et au Kenya a conduit à l'émergence d'un grand nombre d'auteurs contemporains. L'apprentissage du swahili permet d'aborder aisément celui des autres langues de la même famille (bantu) parlées dans une aire qui va du Cameroun à l'Afrique du Sud et de la République Démocratique du Congo aux Comores.

Source : INALCO (<http://www.inalco.fr/langue/swahili>)

Texte long :

Halfdan Ragnarson, à ne pas confondre avec Halfdan III de Vestfold, dit « Halfdan le Noir », est un chef viking actif dans les îles Britanniques dans les années 870. Halfdan fait partie des chefs de la Grande Armée qui débarque en Angleterre en 865. En l'espace de quelques années, cette armée parvient à conquérir les royaumes anglo-saxons de Northumbrie et d'Est-Anglie. La première mention de Halfdan dans les sources date de l'année 870, lorsqu'il conduit ses troupes contre le Wessex à l'automne ou au début de l'hiver. Une série de batailles entre Vikings et Anglo-Saxons s'ensuit, notamment marquée par la défaite de Halfdan à la bataille d'Ashdown, le 8 janvier 871. À la fin de l'année, l'armée viking quitte sa base de Reading et se retire à Londres pour y passer l'hiver. Halfdan y fait frapper des pièces à son nom, signe de son autorité.

À l'automne 872, Halfdan part vers le nord et hiverne à Torksey, dans le Lindsey. Certaines versions de la *Chronique anglo-saxonne* indiquent qu'il fait campagne en Northumbrie cette année-là, peut-être pour mater la rébellion d'Ecgberht. Lors d'une campagne ultérieure, en 873 ou 874, les Vikings marchent sur Repton et chassent le roi de Mercie Burgred. Après la soumission de la Mercie, la Grande Armée se sépare en deux et Halfdan s'installe sur la Tyne pour combattre les Pictes et les Bretons du royaume de Strathclyde. Les *Annales d'Ulster* mentionnent une grande bataille entre les

Pictes et les Vikings en 875. La *Chronique anglo-saxonne* indique pour l'année 876 que les Vikings s'établissent pour de bon en Northumbrie. Halfdan s'intéresse quant à lui à l'Irlande. En 875, le roi viking de Dublin Oistin mac Amlaíb est tué par un *Albann* qui semble être Halfdan. Au terme d'une dernière campagne, il trouve la mort en affrontant les Vikings d'Irlande lors de la bataille de Strangford Lough en 877.

D'après des textes tardifs, Halfdan est l'un des fils du héros légendaire Ragnar Lodbrok, et le frère d'Ivar le Désossé, Sigurd Œil de Serpent, Björn Côte-de-Fer. Il pourrait être identifié à Hvitserk, un fils de Ragnar qui n'est jamais mentionné en même temps que lui. Cependant, pour la plupart des historiens modernes, les légendes entourant Ragnar et ses fils sont dénuées de fondement historique. Le seul lien de parenté certain de Halfdan est qu'il est le frère d'Ivar, un autre chef de la Grande Armée viking qui correspond au *Ivar le Désossé* des sagas. Un autre frère est cité comme ayant mené les Vikings au combat lors de la bataille de Cynuit en 878. Ce troisième frère serait Ubbe, mais cette identification n'est pas certaine.

Halfdan III de Vestfold lui, est le père d'Harald dit « à la belle chevelure ». Harald fut le premier roi à unifier la Norvège, durant son règne de 872 à 931. Succédant à la mort de son père alors qu'il n'a que dix ans, il monte sur le trône de plusieurs petits royaumes assez dispersés. Halfdan les avait soit conquis, soit en avait hérité, et ces royaumes se situaient surtout dans le Sud-Est de la Norvège.

Selon le récit des Sagas, c'est en raison du dédain de la fille d'un roi voisin, Eric de Hordaland, qu'Harald fait le vœu de ne pas couper ou peigner ses cheveux jusqu'à ce qu'il soit le seul roi de Norvège. Cela lui prend dix ans. En 866, il commence une série de conquêtes sur les nombreux petits royaumes qui composent alors la Norvège. Entre 872 et 885 (les avis divergent), après une grande victoire à Hafrsfjord près de Stavanger, il se trouve roi de la totalité du pays. Le monument *Les Épées sur les rochers* commémore sa victoire.

ANNEXE E

Questionnaire d'évaluation

Questionnaire de fin de test

Item à évaluer	1: Pas du tout d'accord	2: Pas d'accord	3: Neutre	4: D'accord	5: Tout à fait d'accord
1. Je voudrais utiliser ce système fréquemment					
2. Ce système est inutilement complexe					
3. Ce système est facile à utiliser					
4. J'aurais besoin du soutien d'un technicien pour être capable d'utiliser ce système					
5. Les différentes fonctionnalités de ce système sont bien intégrées					
6. Il y a trop d'incohérences dans ce système					
7. La plupart des gens apprendront à utiliser ce système très rapidement					
8. Ce système est très lourd à utiliser					
9. Je me suis senti·e très en confiance en utilisant ce système					
10. J'ai eu besoin d'apprendre beaucoup de choses avant de pouvoir utiliser ce système					

1: Si vous aviez la possibilité, réutiliseriez-vous cet outil dans le futur ?	<input type="checkbox"/> oui <input type="checkbox"/> non
2: Si oui, à quelle fréquence ?	Rarement  Très souvent 1 2 3 4 5
3: Si non, pour quelles raisons ? (plusieurs choix possibles)	<input type="checkbox"/> Prise en main complexe <input type="checkbox"/> Manque de fonctionnalités <input type="checkbox"/> Je suis satisfait de mes outils actuels <input type="checkbox"/> autre:
4: Au bout de combien de temps vous êtes-vous senti à l'aise avec l'outil ?	<input type="checkbox"/> Moins de cinq minutes <input type="checkbox"/> Entre 5 et 15 minutes <input type="checkbox"/> Entre 15 et 30 minutes <input type="checkbox"/> Pour de 30 minutes
5: Pensez-vous avoir été en mesure d'exploiter toutes les fonctionnalités de l'outil lors du test ?	<input type="checkbox"/> oui <input type="checkbox"/> non
6: Avez-vous pu réaliser toutes les opérations que vous souhaitiez ?	<input type="checkbox"/> oui <input type="checkbox"/> non

7: Vis-à-vis des questions 5 et 6, si non, pourquoi ?	<input type="checkbox"/> Manque de fonctionnalités <input type="checkbox"/> Manque de temps <input type="checkbox"/> Manipulation trop complexe <input type="checkbox"/> Interface trop peu intuitive <input type="checkbox"/> autre:
8: Pour vous, quel est l'avantage majeur de cet outil par rapport à vos outils actuels ? (Les classer de 1 à X selon l'importance que vous attribuez)	<input type="checkbox"/> Le regroupement des ressources (Encyclopédiques, lexicales, cartes...) <input type="checkbox"/> La possibilité d'insérer des vidéos <input type="checkbox"/> La possibilité d'insérer des images <input type="checkbox"/> La structure de notes-blocs réorganisable <input type="checkbox"/> La détection automatique de dates / noms propres dans le texte source <input type="checkbox"/> L'historique des recherches <input type="checkbox"/> La carte interactive <input type="checkbox"/> Autre : <input type="checkbox"/> Cet outil n'a aucun avantage par rapport à mes outils actuels

<p>9: Pour vous, quel est l'inconvénient majeur de cet outil par rapport à vos outils actuels ? (Les classer de 1 à X selon l'importance que vous attribuez)</p>	<p><input type="checkbox"/> Impossible de travailler hors-ligne</p> <p><input type="checkbox"/> Temps d'apprentissage du logiciel</p> <p><input type="checkbox"/> Impossible de travailler à plusieurs dessus</p> <p><input type="checkbox"/> Autre :</p> <p><input type="checkbox"/> Cet outil n'a aucun inconvénient par rapport à mes outils actuels</p>
<p>10: Comment jugez-vous la qualité de la traduction finale que vous avez faite en vous aidant de cet outil par rapport à votre travail habituel ?</p>	<p><input type="checkbox"/> Bien moins bonne</p> <p><input type="checkbox"/> Un peu moins bonne</p> <p><input type="checkbox"/> De la même qualité</p> <p><input type="checkbox"/> Un peu meilleure</p> <p><input type="checkbox"/> Bien meilleure</p> <p><input type="checkbox"/> Je ne sais pas</p>
<p>11: L'utilisation de cet outil vous a fait:</p>	<p><input type="checkbox"/> Gagner beaucoup de temps</p> <p><input type="checkbox"/> Gagner un peu de temps</p> <p><input type="checkbox"/> Même durée de travail</p> <p><input type="checkbox"/> Perdre un peu de temps</p> <p><input type="checkbox"/> Perdre beaucoup de temps</p> <p><input type="checkbox"/> Je ne sais pas</p>

12: Avez-vous des remarques ou des suggestions à faire sur l'application ?