



Master

2022

Open Access

This version of the publication is provided by the author(s) and made available in accordance with the copyright holder(s).

Traduction automatique dans le domaine juridique : Comparaison de
DeepL et eTranslation de la Commission Européenne

Yildiz, Simge Sirin

How to cite

YILDIZ, Simge Sirin. Traduction automatique dans le domaine juridique : Comparaison de DeepL et eTranslation de la Commission Européenne. Master, 2022.

This publication URL: <https://archive-ouverte.unige.ch/unige:164474>

SIMGE SIRIN YILDIZ

TRADUCTION AUTOMATIQUE DANS LE DOMAINE JURIDIQUE

COMPARAISON DE DEEPL ET ETRANSLATION DE
LA COMMISSION EUROPÉENNE

Directrice : Pierrette Bouillon

Jurée : Marianne Starlander

Mémoire présenté à la **Faculté de traduction et d'interprétation** (Département de traduction et technologies) pour l'obtention de la **Maîtrise universitaire en traduction et technologies**

Université de Genève

Août 2022

J'affirme avoir pris connaissance des documents d'information et de prévention du plagiat émis par l'Université de Genève et la Faculté de traduction et d'interprétation (notamment la *Directive en matière de plagiat des étudiant-e-s*, le *Règlement d'études des Maîtrises universitaires en traduction et du Certificat complémentaire en traduction de la Faculté de traduction et d'interprétation ainsi que l'Aide-mémoire à l'intention des étudiants préparant un mémoire de Ma en traduction*).

J'atteste que ce travail est le fruit d'un travail personnel et a été rédigé de manière autonome.

Je déclare que toutes les sources d'information utilisées sont citées de manière complète et précise, y compris les sources sur Internet.

Je suis consciente que le fait de ne pas citer une source ou de ne pas la citer correctement est constitutif de plagiat et que le plagiat est considéré comme une faute grave au sein de l'Université, passible de sanctions.

Au vu de ce qui précède je déclare sur l'honneur que le présent travail est original.

Simge Sirin Yildiz

Genève, le 17.08.2022



TABLE DE MATIÈRES

LISTE DES FIGURES.....	5
LISTE DES TABLEAUX.....	7
LISTE DES ABRÉVIATIONS.....	9
1. INTRODUCTION.....	10
1.1 CONTEXTE ET OBJECTIF.....	10
1.2 DÉMARCHE.....	11
1.3 PLAN.....	11
2. LA TRADUCTION AUTOMATIQUE.....	13
2.1 DÉFINITION.....	13
2.2 BREF HISTORIQUE.....	15
2.3 LES DIFFÉRENTS APPROCHES ET SYSTÈMES DE LA TA.....	18
2.3.1 INFORMATIONS LINGUISTIQUES.....	19
2.3.2 INFORMATIONS NON LINGUISTIQUES.....	24
2.4 SYSTÈMES UTILISÉS POUR LA COMPARAISON.....	29
2.4.1 DEEPL.....	29
2.4.2 ETRANSLATION.....	30
3. LA TRADUCTION JURIDIQUE.....	34
3.1 DÉFINITION ET TYPOLOGIES.....	34
3.2 TRADUCTION DE TEXTES NORMATIFS EN SUISSE.....	36
3.3 TA DANS LE DOMAINE JURIDIQUE.....	37
4. ÉVALUATION DE LA TA.....	41
4.1 ÉVALUATION HUMAINE.....	42
4.1.1 ÉVALUATION INTRINSÈQUE.....	42

4.1.2 ÉVALUATION EXTRINSÈQUE	45
4.2 ÉVALUATION AUTOMATIQUE	46
4.2.1 PRÉCISION ET RAPPEL	46
4.2.2 WER	47
4.2.3 TER	47
4.2.4 BLEU	47
5. MÉTHODOLOGIE	50
5.1 DONNÉES : PRÉPARATION DU TEXTE SOURCE ET TRADUCTION	50
5.2 ÉVALUATION HUMAINE	51
5.2.1 ANNOTATION D'ERREURS	51
5.2.2 COMPRÉHENSIBILITÉ	53
5.3 ÉVALUATION AUTOMATIQUE	55
5.3.1 BLEU	55
5.3.2 TER	56
5.4 MESURES STATISTIQUES	57
5.4.1 KAPPA	57
5.4.2 COEFFICIENT DE PEARSON	58
6. RÉSULTATS	60
6.1 ÉVALUATION HUMAINE	60
6.1.1 ANNOTATION D'ERREURS	60
6.2 ÉVALUATION AUTOMATIQUE	81
6.3 INTERACCORD	82
6.3.1 KAPPA	82
6.3.2 COEFFICIENT DE PEARSON	83
7. DISCUSSION	93
7.1 SYNTHÈSE ET RÉSULTATS DE L'ÉTUDE	93

7.2 LIMITES DE L'ÉTUDE.....	98
RÉFÉRENCES.....	99

LISTE DES FIGURES

FIGURE 1 ADAPTATION DU SCHEMA DE HUTCHINS ET SOMERS (1992, p. 148) HUMAN AND MACHINE TRANSLATION	13
FIGURE 2 DIRECT MT SYSTEM HUTCHINS&SOMERS 1992, p.72	19
FIGURE 3 INTERLINGUA MODEL WITH TWO LANGUAGE PAIRS (HUTCHINS & SOMERS, 1992, p.74)	21
FIGURE 4 TRANSFER MODEL WITH TWO LANGUAGE PAIRS (HUTCHINS&SOMERS, 1992, p. 75) .	22
FIGURE 5 LE TRIANGLE DE VAUQUOIS (1968) - JURAFSKY, 2014, p. 885.....	23
FIGURE 6 ÉTAPE D'ENTRAÎNEMENT ET DE DÉCODAGE - BOUILLON, 2020	26
FIGURE 7 REPRÉSENTATION DE L'APPROCHE NEURONALE DANS LE TRIANGLE DE VAUQUOIS - SYSTRAN, 2016	27
FIGURE 8 CHRONOLOGY OF MACHINE TRANSLATION DEVELOPMENT QUAH 2006, p. 58	33
FIGURE 9 ADAPTATION DE L'ÉCHELLE DE FIDÉLITÉ ET FLUIDITÉ (KOEHN, 2020, p.46)	43
FIGURE 10 MQM CORE - LOMMEL ET AL., 2015	51
FIGURE 11 GRILLE D'ÉVALUATION POUR L'ANNOTATION D'ERREURS.....	52
FIGURE 12 ÉCHELLE D'ÉVALUATION DE COMPRÉHENSIBILITÉ	53
FIGURE 13 ÉVALUATION DE LA COMPRÉHENSIBILITÉ	54
FIGURE 14 INTERFACE DE INTERACTIVE BLEU SCORE EVALUATOR	56
FIGURE 15 EXTRAIT D'ÉVALUATION PAR 3 JUGES. CHAQUE COLONNE CORRESPOND AUX ÉVALUATIONS D'UN JUGE SUR UNE ÉCHELLE DE 1 À 5	58
FIGURE 16 EXTRAIT D'ÉVALUATION PAR 6 JUGES	58
FIGURE 17 EXTRAIT DU FICHIER EXCEL POUR LE CALCUL DU COEFFICIENT DE PEARSON	59
FIGURE 18 SCORES BLEU POUR ETRANSLATION ET DEEPL.....	81
FIGURE 19 CORRÉLATION ENTRE BLEU ET L'ÉVALUATION DE LA COMPRÉHENSIBILITÉ PAR LES JUGES ALLEMANDS POUR DEEPL	84
FIGURE 20 CORRÉLATION ENTRE BLEU ET L'ÉVALUATION DE LA COMPRÉHENSIBILITÉ PAR LES JUGES ALLEMANDS POUR ETRANSLATION	84
FIGURE 21 CORRÉLATION ENTRE BLEU ET L'ÉVALUATION DE LA COMPRÉHENSIBILITÉ PAR LES JUGES SUISSES POUR DEEPL	85
FIGURE 22 CORRÉLATION ENTRE BLEU ET L'ÉVALUATION DE LA COMPRÉHENSIBILITÉ PAR LES JUGES SUISSES POUR ETRANSLATION.....	85

FIGURE 23 CORRÉLATION ENTRE BLEU ET L'ÉVALUATION DE LA COMPRÉHENSIBILITÉ PAR LES SIX JUGES POUR DEEPL.....	86
FIGURE 24 CORRÉLATION ENTRE BLEU ET L'ÉVALUATION DE LA COMPRÉHENSIBILITÉ PAR LES SIX JUGES POUR ETRANSLATION.....	86
FIGURE 25 CORRÉLATION ENTRE BLEU ET L'ANNOTATION D'ERREURS POUR DEEPL	87
FIGURE 26 CORRÉLATION ENTRE BLEU ET L'ANNOTATION D'ERREURS POUR ETRANSLATION..	87
FIGURE 27 CORRELATION ENTRE L'EVALUATION DE LA COMPREHENSIBILITE ET L'ANNOTATION D'ERREURS POUR DEEPL.....	88
FIGURE 28 CORRELATION ENTRE L'EVALUATION DE LA COMPREHENSIBILITE ET L'ANNOTATION D'ERREURS POUR ETRANSLATION	89
FIGURE 29 CORRELATION ENTRE L'ANNOTATION D'ERREURS ET LA COMPREHENSIBILITE DES JUGES EXPERTS (JUGES ALLEMANDS) POUR DEEPL.....	90
FIGURE 30 CORRELATION ENTRE L'ANNOTATION D'ERREURS ET LA COMPREHENSIBILITE DES JUGES EXPERTS (JUGES ALLEMANDS) POUR ETRANSLATION	90
FIGURE 31 CORRELATION ENTRE L'ANNOTATION D'ERREURS ET LA COMPREHENSIBILITE DES JUGES NON-EXPERTS (JUGES SUISSES) POUR DEEPL.....	91
FIGURE 32 CORRÉLATION ENTRE L'ANNOTATION D'ERREURS ET LA COMPRÉHENSIBILITÉ DES JUGES NON-EXPERTS (JUGES SUISSES) POUR ETRANSLATION.....	91
FIGURE 33 CORRELATION ENTRE LE BLEU ET LE TER POUR DEEPL	92
FIGURE 34 CORRÉLATION ENTRE LE BLEU ET LE TER POUR ETRANSLATION	92

LISTE DES TABLEAUX

TABLEAU 1 DONNEES DU TEXTE SOURCE ET DES TRADUCTIONS PRODUITES PAR DEEPL ET ETRANSLATION	51
TABLEAU 2 NOMBRE D'ERREURS DE DEEPL	60
TABLEAU 3 ERREURS DE DEEPL DANS LA CATÉGORIE FAUSSE TRADUCTION	62
TABLEAU 4 ERREURS DE DEEPL DANS LA CATÉGORIE OMISSION	63
TABLEAU 5 ERREURS DE DEEPL DANS LA CATÉGORIE ABRÉVIATIONS	64
TABLEAU 6 ERREURS DE DEEPL DANS LA CATÉGORIE STYLE.....	66
TABLEAU 7 ERREURS DE DEEPL DANS LA CATÉGORIE TERMINOLOGIE	67
TABLEAU 8 NOMBRE D'ERREURS DE ETRANSLATION	68
TABLEAU 9 ERREURS DE ETRANSLATION DANS LA CATÉGORIE ADDITION	69
TABLEAU 10 ERREURS DE ETRANSLATION DANS LA CATÉGORIE FAUSSE TRADUCTION.....	71
TABLEAU 11 ERREURS DE ETRANSLATION DANS LA CATEGORIE OMISSION.....	73
TABLEAU 12 ERREURS DE ETRANSLATION DANS LA CATÉGORIE NON TRADUIT	74
TABLEAU 13 ERREURS DE ETRANSLATION DANS LA CATÉGORIE ABRÉVIATIONS.....	75
TABLEAU 14 ERREURS DE ETRANSLATION DANS LA CATÉGORIE CONVENTION LOCALE.....	76
TABLEAU 15 ERREURS DE ETRANSLATION DANS LA CATÉGORIE STYLE	78
TABLEAU 16 ERREURS DE ETRANSLATION DANS LA CATÉGORIE TERMINOLOGIE	79
TABLEAU 17 SCORES DE COMPRÉHENSIBILITÉ POUR LES SIX JUGES POUR DEEPL (MAXIMUM = 500).....	80
TABLEAU 18 SCORES DE COMPRÉHENSIBILITÉ POUR LES SIX JUGES POUR ETRANSLATION (MAXIMUM = 500)	80
TABLEAU 19 RÉSULTATS DU CALCUL DE TER POUR DEEPL ET ETRANSLATION	82
TABLEAU 20 SCORES KAPPA POUR DEEPL (ÉVALUATION DE LA COMPRÉHENSIBILITÉ).....	82
TABLEAU 21 SCORES KAPPA POUR ETRANSLATION (ÉVALUATION DE LA COMPRÉHENSIBILITÉ)83	
TABLEAU 22 COEFFICIENT DE PEARSON ENTRE L'ÉVALUATION DE LA COMPRÉHENSIBILITÉ ET LE BLEU	83
TABLEAU 23 COEFFICIENT DE PEARSON POUR L'ANNOTATION D'ERREURS ET LE BLEU	87
TABLEAU 24 COEFFICIENT DE PEARSON POUR L'EVALUATION DE COMPREHENSIBILITE ET L'ANNOTATION D'ERREURS	88
TABLEAU 25 CORRÉLATION ENTRE L'ANNOTATION D'ERREURS ET LA COMPRÉHENSIBILITÉ DES JUGES EXPERTS VS. NON-EXPERTS	89
TABLEAU 26 COEFFICIENT DE PEARSON POUR BLEU ET TER	92

TABLEAU 27 RÉSUMÉ DES SCORES POUR DEEPL ET ETRANSLATION	93
TABLEAU 28 CORRÉLATIONS ENTRE LES DIFFÉRENTES MÉTRIQUES POUR DEEPL ET ETRANSLATION	95

LISTE DES ABRÉVIATIONS

TA : traduction automatique

TEAHQ : traduction entièrement automatique de haute qualité

TAAH : traduction automatique assistée par humain

THAO : traduction humaine assistée par ordinateur

TH : traduction humaine

TAO : traduction assistée par ordinateur

TAN : traduction automatique neuronale

MQM : Multidimensional Quality Metrics

1. INTRODUCTION

Grâce à l'intelligence artificielle, la traduction automatique a fait beaucoup de progrès au cours des dernières années et est devenue un outil indispensable dans notre vie quotidienne. Elle est utilisée dans beaucoup de contextes, pour les bulletins météorologiques, pour les sites web, dans le cadre de grandes organisations et entreprises, par les traducteurs professionnels, etc. Mais même si la traduction automatique a progressé, la qualité des traductions produites par les différents systèmes peut varier. Surtout pour les textes de domaines spécifiques, comme par exemple les textes économiques ou juridiques, une révision et correction des traductions automatiques est presque toujours nécessaire, notamment en raison du vocabulaire spécialisé (Wiesmann, 2019). Dans cette introduction, nous présentons le contexte, l'objectif et la démarche du travail, ainsi que le plan du travail.

1.1 CONTEXTE ET OBJECTIF

Ce mémoire a été réalisé dans le cadre de la formation de Maîtrise du Département de traduction et technologies de la Faculté de Traduction et d'Interprétation de l'Université de Genève.

Il vise à comparer deux systèmes de traduction automatique neuronale, à savoir DeepL et eTranslation de la Commission Européenne pour le domaine juridique.

Vu que les textes juridiques sont en général plus complexes que d'autres types de textes et que la traduction automatique a tendance à être de plus en plus utilisée dans le domaine juridique, il nous a en effet semblé intéressant de vérifier si l'intégration de la traduction automatique neuronale dans ce domaine est efficace ou pas. Notre question de recherche principale est donc la suivante :

La qualité de la traduction automatique neuronale est-elle satisfaisante dans le domaine juridique ?

Cette question comprend deux sous-questions :

- 1. La qualité de la traduction automatique dépend-elle des systèmes ?*
- 2. La qualité de la traduction automatique dépend-elle des évaluateurs ?*

Deux hypothèses sont posées pour ce travail. La première hypothèse suppose que ni eTranslation ni DeepL ne prennent pas en compte le contexte suisse et que eTranslation obtiendrait un meilleur score, parce que ce système est développé par la Commission Européenne et est

spécialisée dans le domaine administratif. La seconde hypothèse suppose que la qualité de la TA dépend des évaluateurs et que dans notre cas, les juges non-experts (les juges suisses) donneront des meilleurs scores en comparaison avec les juges experts (les juges allemands).

1.2 DÉMARCHE

Pour la comparaison des deux systèmes, nous avons choisi comme corpus une partie de la loi COVID-19, entrée en vigueur le 26 septembre 2020. Au total, 100 phrases ont été sélectionnées et ensuite traduites par DeepL et eTranslation. Afin d'évaluer ces traductions, nous avons décidé d'utiliser deux méthodes d'évaluation humaine, deux méthodes d'évaluation automatique. La première méthode d'évaluation humaine a consisté à annoter les erreurs faites par les deux systèmes. Pour ce faire, nous avons élaboré une grille d'évaluation humaine afin d'évaluer les dimensions « fidélité », « terminologie », « style », « abréviations » et « convention locale ». L'annotation d'erreurs a été réalisée par une experte de droit.

Pour la deuxième, nous avons évalué la compréhensibilité des traductions. Pour cette méthode, une autre grille d'évaluation a été créée avec les 100 phrases traduites par DeepL et eTranslation. Les évaluateurs ont donné à chaque traduction un score de 1 (incompréhensible) à 5 (impeccable). L'évaluation de la compréhensibilité a été effectuée par trois locuteurs d'Allemagne et de Suisse.

En ce qui concerne l'évaluation automatique, nous avons retenu les scores BLEU et TER. Finalement, l'interaccord entre les juges et les corrélations entre les différentes métriques ont été calculés.

1.3 PLAN

Dans le chapitre 2, nous définissons la traduction automatique et en donnons un bref historique. Les différentes approches et systèmes de traduction automatique sont expliqués et notamment les deux systèmes de TA utilisés dans cette étude. Le chapitre 3 s'intéresse à la traduction juridique. Il présente les différentes typologies de la traduction juridique, la traduction de textes normatifs en Suisse et l'utilisation de la TA dans le domaine juridique.

Le chapitre 4 donne un aperçu des différentes méthodes d'évaluation de systèmes de traduction automatique, notamment celles utilisées dans ce travail. Le chapitre 5 porte sur la méthodologie du travail.

Dans le chapitre 6, nous présentons les résultats obtenus pour l'évaluation humaine et automatique, ainsi que l'interaccord entre les juges et les corrélations entre les différentes métriques. Dans le chapitre 7, les résultats obtenus sont discutés et les limites de l'étude présentées.

2. LA TRADUCTION AUTOMATIQUE

Ce chapitre vise à présenter la traduction automatique et à expliquer quelques concepts, qui sont essentiels à la compréhension de ce travail. Premièrement, nous expliquons la notion de traduction automatique (section 2.1) et présentons l'évolution du domaine à travers un bref historique (section 2.2). Ensuite, nous donnons un aperçu des différentes approches et des principaux systèmes (section 2.3). Finalement, les deux systèmes utilisés pour la comparaison sont présentés (sections 2.4).

2.1 DÉFINITION

En 2015, Qun et Xiaojun donnent la définition suivante de la traduction automatique :

« Machine Translation (MT) is a sub-field of computational linguistics (CL) or natural language processing (NLP) that investigates the use of software to translate text or speech from one natural language to another. The core of MT itself is the automation of the full translation process, which is different with the related terms such as machine-aided human translation (MAHT), human-aided machine translation (HAMT) and computer-aided translation (CAT). » (Qun & Xiaojun, 2015)

Hutchins et Somers (1992 : 148) distinguent quatre types de traduction, comme illustré dans le schéma suivant (Fig. 1) :

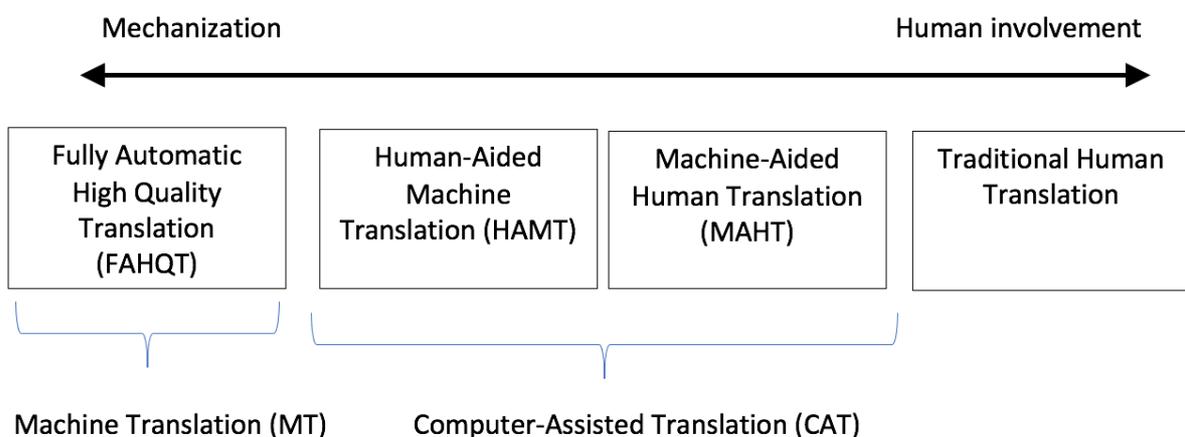


Figure 1 Adaptation du schéma de Hutchins et Somers (1992, p. 148) Human and machine translation

Le spectre montre à droite les méthodes de traduction avec une intervention humaine, et à gauche, les méthodes de traduction avec l'intervention d'une machine.

Tout à droite, nous avons la *traduction humaine traditionnelle* ; au milieu, la *traduction automatique assistée par l'homme* (en français : TAAH, en anglais : HAMT) qui est parfois qualifiée par son hypéronyme *traduction assistée par ordinateur* (en français : TAO, en anglais : CAT), et qui ne peut pas avoir lieu sans intervention humaine, et la *traduction humaine assistée par ordinateur* (en français : THAO, en anglais : MAHT), qui permet aux traducteurs d'avoir accès à des dictionnaires ou à des banques de données terminologiques, etc. À gauche, la traduction entièrement automatique de haute qualité (en français : TEAHQ, en anglais : FAHQT), qui peut se dérouler sans intervention humaine (Bouillon & Clas, 1993).

Jusqu'au début des années 1990, quand l'internet a commencé à être utilisé dans le monde entier, les quatre catégories de traduction présentées par Hutchins et Somers étaient certainement applicables. Plus d'une décennie plus tard, les délimitations de ces quatre catégories sont devenues plus floues, même si beaucoup d'experts dans ce domaine y font toujours référence. La technologie devient en effet de plus en plus multifonctionnelle. Le rythme des changements dans le développement des technologies de traductions est extrêmement rapide ; ce qui est actuel aujourd'hui peut devenir dépassé demain (Quah, 2006).

Jusqu'à maintenant, la traduction entièrement automatique de haute qualité (TEAHQ) reste une tâche difficile à accomplir. Très peu systèmes sont capables de produire des traductions de bonne qualité sans intervention humaine ; la plupart d'entre eux ont besoin d'une intervention humaine soit avant le processus de traduction, dite *pré-édition*, soit après le processus de traduction, dite *post-édition* (Bouillon & Clas, 1993). Si l'intervention est réalisée pendant le processus de traduction, on parle d'un mode interactif (Hutchins & Somers, 1992).

La pré-édition consiste à modifier le texte source en utilisant un langage contrôlé (Ping, 2011). Cette tâche peut être nécessaire, vu qu'il est difficile pour les machines de traiter l'ensemble du langage naturel pour une tâche donnée à cause de l'ambiguïté et de la complexité des langages naturels. Les langage contrôlé vise à résoudre ces types de problèmes en limitant la taille de la grammaire et du vocabulaire afin d'éliminer l'ambiguïté et la complexité (Schwitter, 2015).

La tâche du post-éditeur consiste à corriger les traductions produites par les systèmes de TA conformément à une méthode déterminée. Cela veut dire que la post-édition est réalisée de façon minimale pour les textes destinés uniquement à des fins d'information, et de façon complète pour les textes destinés à être publiés et distribués (Hutchins & Somers, 1992, p. 152).

Le troisième mode d'interaction permet aux systèmes de TA de solliciter l'aide de l'utilisateur pour résoudre les problèmes d'ambiguïté et d'équivalents de traduction. Les systèmes interactifs sont très peu nombreux (Hutchins & Somers, 1992, p. 153). Actuellement, le seul système commercial interactif est *Lilt* (Martikainen, 2022).

La pré-édition est une tâche manuelle qui prend beaucoup de temps et qui peut changer le sens et l'intention du texte source. Par conséquent, la pré-édition a été moins utilisée que la post-édition, mais pour les domaines spécifiques, tels que la traduction des bulletins météorologiques, la pré-édition a prouvé sa viabilité (Kittredge et al. 1976, cité dans Nirenburg, 1987, p. 73).

2.2 BREF HISTORIQUE

Avec l'apparition des premiers ordinateurs après la Seconde Guerre mondiale, apparaît un besoin de traduire automatiquement des textes étrangers (dans le contexte de la guerre froide) et également un besoin de mieux comprendre le fonctionnement du langage. Depuis la fin des années 1940, plusieurs chercheurs développent des algorithmes et des techniques, parfois naïfs, mais qui constituent les premiers pas vers une approche globale de la traduction automatique. Le schéma qui est à la base de tous les processus de cryptographie et de la communication, a été proposé par le mathématicien Warren Weaver en 1949. Ce modèle est très général et peut être appliqué à des contextes multiples : premièrement, un message est encodé par une source humaine ou automatique ; deuxièmement, le message est émis, et finalement, il est décodé par un récepteur. Weaver a écrit également un mémorandum en 1949, qui est considéré comme le point de départ des recherches dans le domaine de la traduction automatique (Poibeau, 2019, p. 39-43).

Dans son mémorandum, Weaver propose des stratégies pour éviter la traduction mot à mot et mentionne que ces stratégies peuvent se baser sur des techniques de la cryptographie en utilisant des ordinateurs modernes (Weaver, 1949). Ce mémorandum, dans lequel Weaver propose l'utilisation de techniques cryptographiques pour l'automatisation de la traduction, a été critiqué par beaucoup de chercheurs qui jugeaient celles-ci trop simplistes et naïves (Quah, 2006, p. 59). Elles seront cependant utilisées dès les années 80 avec les systèmes fondés sur les corpus.

En 1954, les chercheurs américains réussissent pour la première fois à traduire automatiquement du russe vers l'anglais (Trujillo, 1999). Cet événement attire la sympathie du public et contribue à l'augmentation des financements pour plus de recherches dans ce domaine (Poibeau, 2019). Cette démonstration attire aussi l'attention de chercheurs soviétiques, mais aussi de chercheurs du monde entier, qui lancent des recherches similaires à partir de 1955 (Trujillo, 1999).

En 1959, le chercheur israélien, Yehoshua Bar-Hillel publie un rapport, dans lequel il explique qu'aucune méthode ne peut permettre à l'ordinateur de lever les ambiguïtés sémantiques, et de ce fait, la TEAHQ sans intervention humaine n'est pas seulement impossible dans un avenir proche, mais aussi lointain. La machine ne pourrait en effet jamais avoir les connaissances linguistiques nécessaires pour éliminer les ambiguïtés sémantiques.

Les doutes des autorités de financement ont été exprimés en 1966 par le comité ALPAC (Automatic Language Processing Advisory Committee), qui a publié son rapport qui conclut que la post-édition des traductions automatiques n'est pas moins chère ou plus rapide que la traduction humaine. La traduction automatique serait ainsi un échec, parce qu'il n'existe pas de système de traduction automatique de textes scientifiques et qu'un tel système n'est pas faisable dans un proche avenir. Il serait donc préférable de se concentrer sur l'amélioration des stratégies pour la traduction humaine assistée par ordinateur (ALPAC, 1966; Koehn, 2010). Même si ce rapport de ALPAC était biaisé, il a eu une influence sur les recherches menées dans ce domaine. Aux États-Unis, les recherches en TA ont été réduites et elles ont perdu leur statut. Le rapport de ALPAC a aussi eu des effets dans d'autres pays du monde : après 1968, aucun groupe britannique ne s'engage plus dans une recherche active en traduction automatique. De même, les recherches au Japon et en Union soviétique sont réduites. Seulement le groupe français (Centre d'Études pour la Traduction Automatique - CETA) a pu continuer ses recherches. (Hutchins, 1986 : 165-167).

Dans les années 1970, plusieurs pays commencent à reprendre les recherches dans le domaine de la traduction automatique (Quah, 2006, p. 62). Ainsi, le Canada a financé un centre de recherche qui a développé le système TAUM-Météo, un système pour traduire les bulletins météorologiques ; la Commission européenne commence aussi à s'intéresser à la traduction automatique, vu le besoin de traduire entre un nombre grandissant de langues dans l'Union européenne. Ainsi, l'Union européenne décide de travailler avec l'entreprise Systran, née en 1968

aux États-Unis et qui est spécifiquement dédiée à la traduction automatique. En 1976, Systran développe un prototype pour les différentes langues de l'Union européenne. Entre les années 1970 et 1980, d'autres groupes de recherche développent d'autres systèmes commerciaux, comme par exemple l'Université du Texas qui travaille avec l'entreprise Siemens pour développer Metal, un système pour traduire de l'allemand vers l'anglais et vice versa (Poibeau, 2019, p. 64-66). Les recherches dans le domaine de la traduction automatique ont continué pendant les années 1980, afin de trouver de meilleures méthodes et techniques de traduction, et les progrès dans le domaine de la linguistique computationnelle ont également permis à la recherche de développer des approches plus sophistiquées de la traduction (Quah, 2006, p. 62). Jusqu'à la fin des années 1980, deux approches étaient utilisées dans les systèmes de traduction automatique : l'approche directe et l'approche indirecte (Quah, 2006, p. 63). Les différentes approches et systèmes de la TA seront présentés en détail dans la section 2.3.

Dans les années 1990, il y a eu un tournant majeur. Deux nouvelles approches sont apparues : statistique et basée sur l'exemple (en anglais : *example-based*). Contrairement aux premiers systèmes basés sur les règles, ces deux nouvelles approches n'utilisent pas des règles syntaxiques ou sémantiques, mais exploitent de grands corpus de textes. Cependant, les recherches sur les systèmes basés sur les règles se sont poursuivies (Hutchins, 2014). Un autre progrès important a été le développement de la reconnaissance vocale, qui, combinée aux systèmes de TA, permettent la conversion de la parole en texte (Quah, 2006, p. 65).

À la fin des années 1990, les ventes de logiciels de TA ont augmenté, et les premiers systèmes de TA gratuits en ligne, tels que Babelfish et Google Translate, sont apparus. Avec ces innovations, la TA est devenue un produit de masse (Hutchins, 2014).

Dans les années 2000, la traduction statistique est devenue le système dominant de la recherche en TA et l'utilisation de la TA continue à augmenter. De nombreuses entreprises internationales et services de traduction commencent à utiliser la TA, souvent en combinaison avec la pré-édition et post-édition. En ce qui concerne le public général, la TA est presque exclusivement accessible gratuitement en ligne et non plus des logiciels (Hutchins, 2014).

Le système de TA le plus récent apparu au cours des dernières années utilise « l'apprentissage profond ». Il s'agit des « systèmes neuronaux ». Cette nouvelle approche s'est imposée très rapidement (Poibeau, 2019, p. 119-121). Poibeau (2019, p. 121) remarque : « À l'automne 2016, tous les acteurs majeurs du domaine (Google, Bing, Facebook, Systran, etc.) sont passés

d'un coup à l'apprentissage profond, avec dans la plupart des cas une amélioration des performances immédiatement visible. »

Les systèmes neuronaux seront présentés en détail dans la section 2.3.2.3.

2.3 LES DIFFÉRENTS APPROCHES ET SYSTÈMES DE LA TA

Les systèmes de TA peuvent être catégorisés selon différentes caractéristiques, comme par exemple les degrés d'automatisme et de généralité. Mais on peut aussi distinguer les *systèmes bilingues*, qui sont créés pour traiter une seule paire de langues, et les *systèmes multilingues*, qui sont créés pour traiter plusieurs paires de langues. On peut aussi faire la distinction entre les systèmes unidirectionnels, qui traduisent vers une seule langue à la fois, et les systèmes bidirectionnels, qui traduisent vers les deux langues en même temps.

Une autre caractéristique pour distinguer les systèmes de TA est le type d'information, dont le système a besoin pour la traduction. Ces informations peuvent être soit linguistiques (les informations phonologiques, morphologiques, syntaxiques et sémantiques), soit non linguistiques (exemples de traductions). Les informations non linguistiques peuvent être utilisées soit comme alternative aux informations linguistiques, soit comme complément (Bouillon & Clas, 1993, p. 15-16).

Cette section vise à présenter les différents systèmes de la TA, qui sont classés en fonction du type d'information, soit dans la catégorie *informations linguistiques* soit dans la catégorie *informations non linguistiques*.

2.3.1 INFORMATIONS LINGUISTIQUES

Les systèmes qui utilisent les informations linguistiques, aussi appelés *systèmes basés sur les règles* (en anglais : rule-based systems), peuvent effectuer le processus de traduction directement ou indirectement, par transfert ou interlangue (Bouillon & Clas, 1993, p. 16).

2.3.1.1 Systèmes directs

La traduction directe était la première approche utilisée dans le développement de la traduction automatique et elle est considérée comme la première génération de systèmes de TA. La traduction directe étant la première approche développée, son système est plutôt primitif (Quah, 2006, p. 69). Les systèmes directs sont élaborés pour une seule paire de langues et ils sont unidirectionnels (Bouillon & Clas, 1993, p. 16). Ces systèmes fonctionnent principalement à l'aide de dictionnaires (Tucker 1987, cité dans Nirenburg, 1987, p. 23), qui servent à trouver des équivalences au niveau lexical (Poibeau, 2019).

L'approche directe ne prévoit aucune étape intermédiaire dans le processus de traduction. Cela veut dire que le texte en langue source va être traduit directement en langue cible (Hutchins & Somers, 1992, p. 72).

L'approche directe est représentée par Hutchins et Somers (1992) de la manière suivante dans Fig. 2 :



Figure 2 Direct MT system Hutchins&Somers 1992, p.72

La phrase en langue source passe d'abord par une phase d'analyse morphologique, qui sert à l'identification des mots, qui mène directement à la consultation de dictionnaires bilingues pour obtenir des équivalences des mots dans la langue cible, suivi de quelques règles de réarrangement des mots pour produire une traduction acceptable dans la langue cible (Hutchins & Somers, 1992, p. 72).

La traduction directe peut donc être caractérisée comme une traduction mot à mot avec un réarrangement de l'ordre des mots (Hutchins & Somers, 1992, p. 72). Les systèmes directs ne

fonctionnent bien qu'avec des paires de langues très proches dont les structures grammaticales sont similaires (Quah, 2006, p. 70). L'approche de traduction directe ne comprend aucune application de la théorie de la traduction et ne contient qu'une application minimale de la théorie linguistique (Jurafsky & Martin, 2020). En conséquence, les systèmes directs se sont révélés peu fiables et insuffisamment puissants, néanmoins ils ont été adoptés dans presque tous les systèmes de traduction automatique développés avant 1966 (Jurafsky & Martin, 2020). Les systèmes directs comprennent les systèmes Météo, Weidner, CULT et l'ancienne version de Systran (Quah, 2006, p. 70). Dans la pratique, on constate toutefois que la plupart des systèmes de TA ont tendance à combiner plusieurs approches différentes (Jurafsky & Martin, 2020).

2.3.1.2 Systèmes indirects

Les lacunes des systèmes de première génération (systèmes directs) ont conduit au développement de modèles linguistiques plus avancés pour la traduction (Hutchins & Somers, 1992, p. 73). Contrairement aux systèmes directs, les systèmes indirects effectuent une analyse du texte source dans une représentation intermédiaire, afin de rendre explicites les informations implicites du texte source et d'éliminer les éléments non pertinents (Bouillon & Clas, 1993, p. 16). L'approche indirecte comprend deux approches principales : l'approche par transfert (en anglais : transfer) et l'approche par interlangue (en anglais : interlingua). Ces systèmes sont aussi appelés systèmes de deuxième génération (Hutchins & Somers, 1992, p. 73).

Approche par interlangue

Dans l'approche par interlangue, une représentation intermédiaire (appelée interlangue) est produite à partir d'une phrase source (L'Homme, 2008, p. 257). Le processus de traduction est effectué en deux étapes : analyse et génération (Bouillon & Clas, 1993, p. 47). Au moment de l'analyse du texte source, celui-ci est converti dans l'interlangue qui sert à générer la phrase cible indépendamment de la phrase source (L'Homme, 2008, p. 257). L'objectif de l'analyse est la création d'une interlangue. La représentation intermédiaire comprend toutes les informations nécessaires pour la génération du texte cible. La représentation intermédiaire est donc une projection du texte source et sert en même temps de base à la génération du texte cible, c'est donc une représentation abstraite du texte cible ainsi qu'une représentation du texte source (Hutchins & Somers, 1992, p. 73). Cette représentation devrait être valable pour tous les couples de langues traitées, parce qu'elle est définie indépendamment des langues traduites

(L'Homme, 2008, p. 257). C'est la raison pour laquelle l'approche par interlangue est la plus attrayante pour les systèmes multilingues. Dans les systèmes par interlangue, chaque langue dispose d'un module qui est responsable pour l'analyse des phrases sources et la conversion en interlangue, et d'un module qui est responsable pour la génération des phrases cibles à partir de cette interlangue (Trujillo, 1999, p. 167). Chaque module d'analyse est indépendant de tous les autres modules d'analyse et des modules de génération, cela signifie que les langues cibles n'ont aucune influence sur le processus d'analyse (Hutchins & Somers, 1992, p. 73), ce qui permet également l'ajout de nouveaux modules (analyse et génération d'autres langues) sans affecter ou modifier les modules existants du système (Quah, 2006, p. 71).

L'approche par interlangue est représentée par Hutchins et Somers (1992) de la manière suivante dans Fig. 3 :

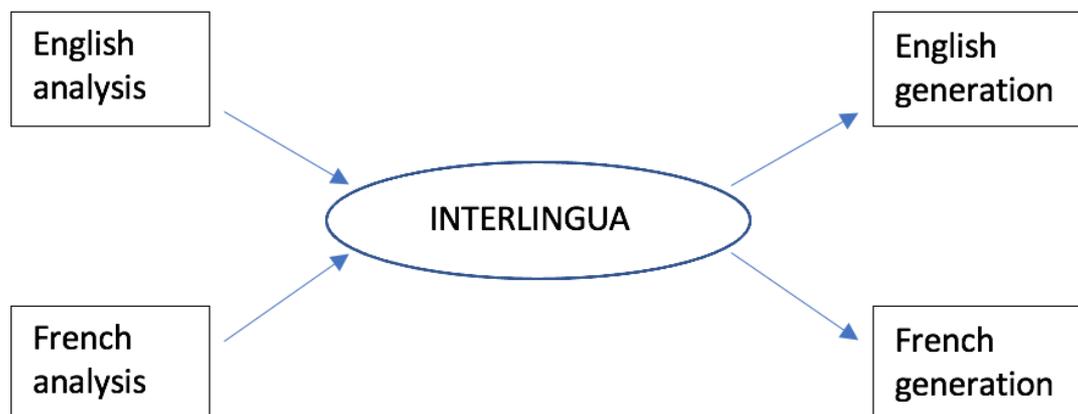


Figure 3 Interlingua model with two language pairs (Hutchins & Somers, 1992, p.74)

Même si l'ajout de nouvelles langues est facile dans les systèmes par interlangue, ces systèmes montrent des inconvénients majeurs : la création d'une interlangue est une tâche très difficile, même pour des langues très proches (Hutchins & Somers, 1992, p. 75), puisqu'elle doit inclure toutes les différences qui existent entre toutes les langues. Marie-Claude L'Homme (2008) donne l'exemple avec le concept désigné par les mots *red* et *rouge*, qui est unique en anglais et français. Mais en espagnol, ce concept est désigné par les mots *rojo* et *tinto*.. C'est la raison pour laquelle l'approche par interlangue n'est pas utilisée comme telle dans les systèmes commerciaux (L'Homme, 2008, p. 258). Le processus sera cependant repris dans les systèmes neuronaux (Poibeau, 2019).

Approche par transfert

Contrairement aux systèmes par interlangue, qui procèdent en deux étapes, les systèmes par transfert effectuent le processus de traduction en trois étapes : analyse, transfert et génération (Bouillon & Clas, 1993, p. 16). L'étape d'analyse consiste à convertir le texte source en une représentation abstraite de la langue source. La deuxième étape, l'étape de transfert, vise à transférer la représentation de la langue source vers sa représentation équivalente dans la langue cible. La dernière étape, l'étape de génération, consiste à générer le texte cible (Quah, 2006, p. 73). Les différentes étapes de l'approche par transfert sont représentées par Hutchins et Somers (1992) de la manière suivante (Fig. 4) :

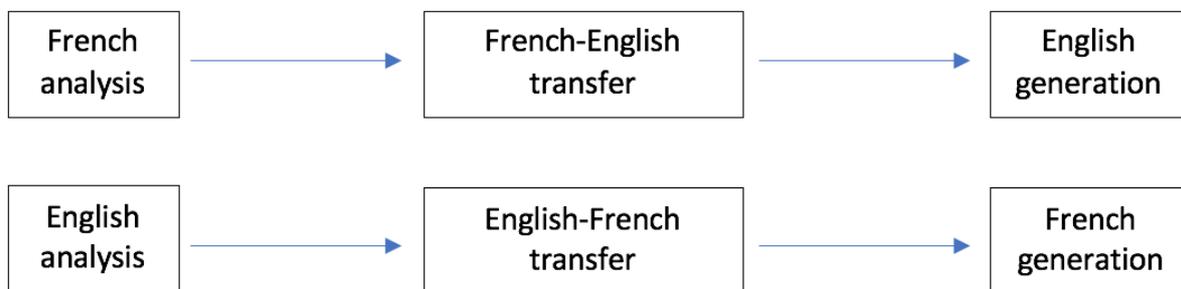


Figure 4 Transfer model with two language pairs (Hutchins&Somers, 1992, p. 75)

Pour l'étape d'analyse et l'étape de génération, l'approche par transfert se base sur des règles qui servent aux systèmes à décrire le fonctionnement de la langue source et de la langue cible. En plus, dans l'approche par transfert, il existe un module de transfert qui crée des correspondances entre les mots et la syntaxe des langues traitées (L'Homme, 2008, p. 256). Dans cette approche, il n'existe donc pas de représentations indépendantes de la langue comme c'est le cas dans l'approche par interlangue. La représentation intermédiaire de la langue source est spécifique à une langue particulière, de même que la représentation intermédiaire de la langue cible (Hutchins & Somers, 1992, p. 75). L'ajout d'une nouvelle langue dans un système par transfert n'est donc pas si simple que dans un système par interlangue. Pour ajouter une nouvelle langue dans un système par transfert, il faudra non seulement ajouter les modules d'analyse et de génération, mais également celui de transfert, qui varie selon le nombre de langues traitées par le système. Par exemple, si on veut ajouter une troisième langue dans un système avec deux langues, on aurait besoin de quatre nouveaux modules de transfert (Hutchins & Somers, 1992, p. 75). L'ajout d'une nouvelle langue pose d'autres problèmes : les règles pour les différentes étapes doivent être redéfinies (L'Homme, 2008, p. 256).

Néanmoins, il y a plusieurs facteurs qui rendent l'approche par transfert attrayante pour la TA. Beaucoup de systèmes sont bilingues et unidirectionnels ; certaines parties des modules de transfert peuvent être partagées s'il s'agit de langues très proches.

Les systèmes de transfert sont généralement considérés comme un compromis pratique entre l'utilisation efficace des ressources des systèmes par interlangue et la facilité de déploiement des systèmes directs. La plupart des systèmes commerciaux appartenait à cette approche (Trujillo, 1999, p. 121).

2.3.1.3 CONCLUSION

Les trois approches présentées peuvent être visualisées avec le triangle de Vauquois de la manière suivante dans Fig. 5 :

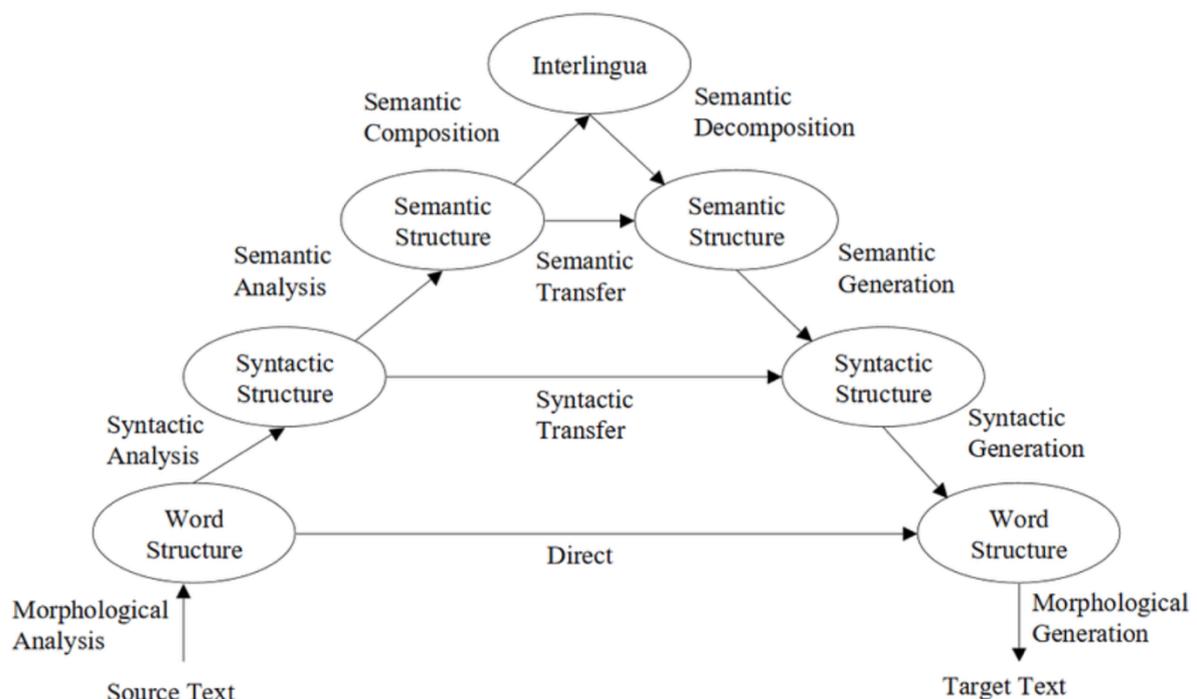


Figure 5 Le triangle de Vauquois (1968) - Jurafsky, 2014, p. 885

Ce triangle est expliqué par Jurafsky (2014, p. 885) :

« The triangle shows the increasing depth of analysis required (on both the analysis and generation end) as we move from the direct approach through transfer approaches to interlingual approaches. In addition, it shows the decreasing amount of transfer knowledge needed as we move up the triangle, from huge amounts of transfer at the direct level (almost all knowledge is transfer knowledge for each word) through transfer

(transfer rules only for parse trees or thematic roles) through interlingua (no specific transfer knowledge). »

Le transfert direct, qui se trouve à la base du triangle, correspond à la traduction mot à mot, pour laquelle un simple dictionnaire bilingue peut suffire. Le transfert syntaxique, qui se trouve au-dessus du transfert direct, permet d'analyser le texte source au niveau syntaxique, pour pouvoir faire appel à des règles de traduction (appelées « règles de transfert »). Le transfert sémantique, qui se trouve au-dessus du transfert syntaxique, permet de choisir le sens des mots selon le contexte de la phrase, et ainsi, d'éliminer l'ambiguïté des mots. Enfin, le transfert par interlangue fait appel à des représentations indépendantes des langues naturelles et permet de traduire directement vers n'importe quel nombre de langues si les modules de génération correspondants sont disponibles (Poibeau, 2019).

2.3.2 INFORMATIONS NON LINGUISTIQUES

Les systèmes qui utilisent des informations non linguistiques se basent sur l'approche analogique. L'approche analogique, aussi appelée approche basée sur les corpus, utilise des exemples de traduction connus pour le processus de traduction. Les systèmes statistiques et les systèmes basés sur l'exemple suivent cette approche. Depuis quelques années, un nouveau type de système statistiques est apparu : les systèmes neuronaux (aussi appelé apprentissage profond, en anglais : deep learning) (Poibeau, 2019, p. 119). Il est également possible de combiner plusieurs approches, ce type de systèmes sont appelés hybrides (Quah, 2006, p. 66).

2.3.2.1 Systèmes basés sur l'exemple

Les systèmes basés sur l'exemple utilisent des exemples de phrases connus pour effectuer le processus de traduction par analogie (Bouillon & Clas, 1993, p. 16). La méthode de base de cette approche est simple : pour traduire une phrase, il faut utiliser des exemples de traductions précédentes qui sont similaires, en partant du principe que la plupart des traductions sont de simples modifications de traductions précédentes. Cette méthode ne permet pas seulement de gagner du temps, mais elle permet également de maintenir la cohérence de la terminologie et du style, ce qui peut être nécessaire dans le cadre d'une traduction juridique (Trujillo, 1999, p. 203).

Les systèmes basés sur l'exemple procèdent typiquement en trois étapes. La première étape consiste à chercher des fragments équivalents à la phrase source à partir d'un corpus bilingue aligné. La deuxième étape consiste à chercher les équivalents en langue cible pour chaque fragment repéré dans la première étape. La dernière étape consiste à combiner les fragments de traduction obtenus, afin de produire une phrase correcte dans la langue cible (Poibeau, 2019, p. 84).

En comparaison avec les systèmes basés sur les règles, les systèmes par analogie montrent certains avantages : le temps de développement est plus limité et la connaissance linguistique est indépendante du système (Bouillon & Clas, 1993, p. 51).

2.3.2.2 Systèmes statistiques

Les systèmes statistiques utilisent des corpus bilingues afin de trouver la phrase cible qui est la plus probable d'être l'équivalent de la phrase source (Bouillon & Clas, 1993, p. 16). Dans l'approche statistique, le texte source est d'abord segmenté en mots ou en segments (Quah, 2006, p. 77). Il existe deux générations de systèmes statistiques : la TA statistique par mots (en anglais : word-based machine translation) qui traduit mot à mot et est similaire dans son fonctionnement aux systèmes directs, et la TA statistique par segments (en anglais : PBMT, phrase-based machine translation) qui traduit par segments (Bouillon, 2020). Les segments identifiés sont ensuite comparés à un grand corpus bilingue aligné qui contient des textes sources avec leurs traductions correspondantes, et finalement, une méthode statistique est appliquée sur le corpus bilingue aligné pour obtenir de nouveaux segments en langue cible. L'hypothèse principale de cette approche est qu'une phrase source peut avoir plusieurs traductions et que chacune d'entre elles a une probabilité variable d'être correcte (Quah, 2006, p. 77). Pour calculer la probabilité, il existe deux formules mathématiques : le modèle *noisy-channel* et le modèle *log-linear*. Ces deux modèles utilisent un modèle de traduction et un modèle de langue. Le modèle de traduction permet de calculer la probabilité d'équivalence entre la phrase source et cible, et le modèle de langue permet de calculer la probabilité de validité de la phrase cible dans la langue cible (Hearne & Way, 2011).

Le fonctionnement des systèmes statistiques est représenté par Bouillon (2020) de la manière suivante dans la Figure 6 :

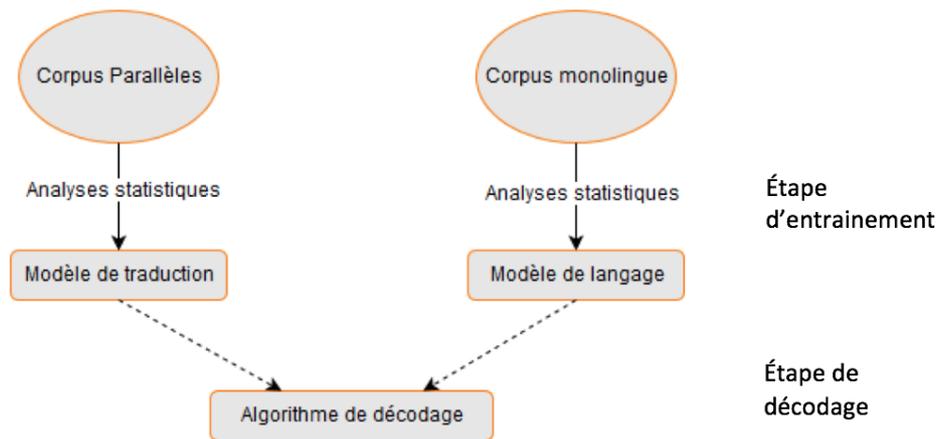


Figure 6 Étape d'entraînement et de décodage - Bouillon, 2020

Pour trouver la traduction la plus probable, les systèmes statistiques procèdent en deux étapes séparées : entraînement et décodage. L'étape d'entraînement consiste à créer le modèle de traduction à partir d'un corpus parallèle et le modèle de langue à partir d'un corpus monolingue. En plus, des modèles supplémentaires peuvent être intégrées à l'étape d'entraînement, comme par exemple le modèle de longueur de phrases, le modèle de réarrangement des mots, etc. L'étape de décodage consiste à produire la traduction en utilisant les modèles créés à l'étape d'entraînement (Hearne & Way, 2011). Chaque traduction possible obtient un score et l'objectif du décodage est de trouver la traduction avec le meilleure score (Koehn, 2010).

Entre les années 1990 et 2016, la plupart des systèmes commerciaux appliquaient l'approche statistiques à base de segments, mais l'assemblage de ces segments de traduction a toujours présenté une limitation majeure de l'approche statistique (Poibeau, 2019, p. 116).

2.3.2.3 Systèmes neuronaux

Avec les systèmes neuronaux, les limites de l'approche statistique ont été résolus à partir du milieu des années 2010. Cette approche, qui est appelée apprentissage profond (en anglais : deep learning), utilise un réseau neuronal avec seulement deux niveaux (input et output) (Melby, 2019, p. 426), et permet de considérer la phrase comme un tout sans la couper en segments avant le processus de traduction. L'approche neuronale utilise les avantages de la traductions statistique, mais effectue le processus de traduction à partir de représentations beaucoup plus riches (Poibeau, 2019, p. 117).

Le concept de l'apprentissage profond a été proposé dans les années 1960, quand les systèmes basés sur les règles étaient dominants, mais ce n'est qu'en 2012 qu'elle a réellement progressé

(Sze et al. 2017, cité dans Melby, 2019). La réintroduction moderne des méthodes neuronales dans la traduction automatique a commencé avec l'intégration de modèles de langage neuronaux dans les systèmes traditionnels de traduction statistique (Koehn, 2020). Les systèmes neuronaux sont donc une nouvelle approche des systèmes basés sur les corpus. Ils sont entraînés en utilisant des corpus bilingues alignés d'énorme taille, qui contiennent des millions d'unités de traduction. En ce sens, les systèmes statistiques et neuronaux fonctionnent de manière similaire, mais les systèmes neuronaux appliquent une approche computationnelle complètement différente : les réseaux neuronaux. Cette dénomination vient du fait que les réseaux neuronaux sont composés de milliers d'unités artificielles qui ressemblent à des neurones (Forcada, 2017).

Les premiers systèmes neuronaux se basaient sur un modèle d'encodage et de décodage (en anglais : encoder-decoder model). La phrase source est lue par l'encodeur du système et représentée comme un vecteur de taille fixe (plongement). À partir de ce vecteur, le décodeur produit la traduction (Bahdanau et al., 2015). En utilisant le triangle de Vauquois, on peut observer des similarités entre l'approche par interlangue et l'approche neuronale (Systran, 2016), la Figure 7 montre cette similarité :

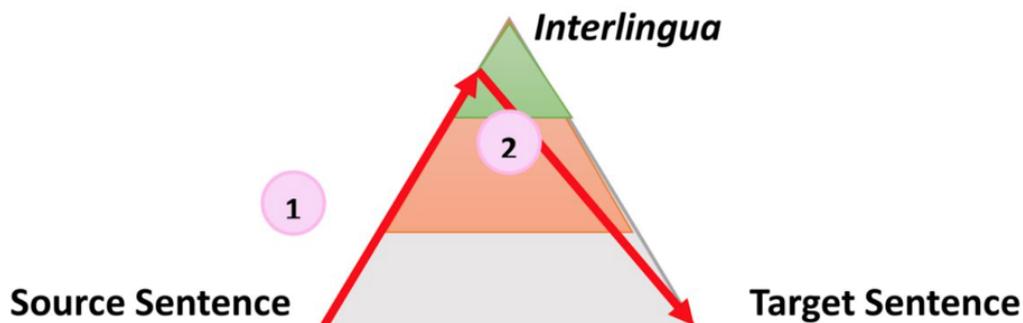


Figure 7 Représentation de l'approche neuronale dans le triangle de Vauquois - Systran, 2016

Comme dans l'approche par interlangue, dans l'approche neuronale, les phrases sources sont converties dans une représentation intermédiaire, mais au lieu de convertir les phrases sources dans une interlangue, les systèmes neuronaux convertissent les phrases sources dans un vecteur. Cette représentation intermédiaire dans l'approche neuronale correspond à un plongement (en anglais : embedding) (Koehn, 2022). Le plongement de la phrase est construit à partir des plongements des mots (en anglais : word embeddings). Les mots qui apparaissent dans des contextes similaires doivent avoir des représentations similaires. Dans les modèles d'espace vectoriel, tels que le plongement, la similarité peut être mesurée par une fonction de distance, c'est-à-dire que

l'angle entre les vecteurs est mesuré. Les plongements permettent de réaliser une inférence sémantique, afin d'obtenir une représentation sémantique : on peut représenter la notion « reine », p.ex. comme « roi + (femme – homme) » (Koehn, 2017).

Les systèmes neuronaux actuels sont beaucoup plus complexes que les premiers systèmes et la plupart de systèmes neuronaux se basent sur des paramètres globaux, appelés *hyperparamètres*, qui sont déterminés empiriquement (Poibeau, 2019, p. 137). L'approche par apprentissage profond semble être l'approche la plus prometteuse jusqu'à aujourd'hui, néanmoins, ils présentent aussi des inconvénients : les systèmes neuronaux requièrent une grande puissance computationnelle. En outre, les systèmes neuronaux ont besoin de données de grande taille ; avec peu de données, ils ne sont pas capables de produire une traduction de qualité acceptable. Les erreurs produites par les systèmes neuronaux sont parfois assez naïves : le système peut produire une traduction qui semble être de bonne qualité dans son contexte, mais qui ne correspond pas au texte source (Koehn, 2022).

2.3.2.1 Systèmes hybrides

Les différentes approches de systèmes peuvent être combinées. Ces systèmes sont appelés *systèmes hybrides*. Il existe des systèmes hybrides avec un système statistique de base. Un tel système utilise le mécanisme des corrélations statistiques des phrases sources et cibles, mais il ajoute une certaine intelligence à ces modèles, en les analysant de manière plus détaillée au niveau linguistique et en utilisant ces annotations lors de la création d'expressions correctes dans la langue cible (Thurmair, 2005).

Il existe aussi des systèmes hybrides avec un système basé sur les règles de base :

« This approach would try to add aspects of language *performance* to the aspects of language competence as documented in the rule base of the system. It would use a standard transfer-based approach, dividing the translation process into the phases of analysis, transfer, and generation; but each of these phases would be re-modelled. » (Thurmair, 2005, p.104).

Jusqu'en 2015, presque tous les systèmes de TA adoptent des approches hybrides dans une certaine mesure (Qun & Xiaojun, 2015).

2.4 SYSTÈMES UTILISÉS POUR LA COMPARAISON

Cette section vise à présenter les deux systèmes de TA (DeepL et eTranslation) qui sont utilisés pour ce travail.

2.4.1 DEEPL

DeepL est un système de traduction automatique en ligne, créé en 2017 par la société allemande DeepL SE. Depuis son lancement en 2017, DeepL Traducteur a atteint une qualité de traduction inégalée, grâce à l'intelligence artificielle. Ce service est destiné non seulement aux entreprises, aux professionnels et aux traducteurs, mais aussi au monde entier pour surmonter les barrières linguistiques (DeepL, 2021).

L'histoire de DeepL commence en 2016 avec une équipe de *Linguee* qui travaille sur la première version de DeepL Traducteur, avec l'objectif de créer un nouveau système de TA ultra-efficace, accessible en ligne et basé sur des réseaux neuronaux. Pour l'élaboration de ce système, l'équipe utilise les données existantes de *Linguee* et parvient à produire une qualité de traduction automatique sans précédent (DeepL, 2021).

En 2017, DeepL Traducteur est accessible en ligne gratuitement et permet de traduire toutes les combinaisons de 7 langues (anglais, allemand, français, espagnol, italien, polonais et néerlandais) (DeepL, 2021).

En 2018, le service offre une nouvelle version payante appelée DeepL Pro. Cette version peut être intégrée aux logiciels de traduction. En plus, DeepL Pro offre une version optimisée du traducteur en ligne et une API (DeepL, 2021). L'API de DeepL permet d'intégrer les traductions directement dans les sites web et les applications (DeepL, 2022).

Deux langues en plus s'ajoutent à DeepL Traducteur en 2018 : le russe et le portugais.

En 2019, DeepL offre une application pour les ordinateurs, qui peut être intégrée aux moteurs d'exploitation Windows et MacOS. Cette nouveauté permet d'utiliser le DeepL Traducteur dans toutes les applications (DeepL, 2021).

En 2019, l'Administration fédérale suisse (en allemand : Bundesverwaltung) a effectué un test sur DeepL avec des traducteurs des services linguistiques de l'Administration fédérale et du parlement suisse, afin d'évaluer la qualité de DeepL Pro pour la traduction des documents dans l'Administration fédérale suisse. Le résultat de ce test montre clairement que DeepL Pro peut

fournir une qualité acceptable pour la compréhension de textes généraux. Cependant, il arrive souvent que la traduction produite par DeepL Pro soit correcte au niveau linguistique et logique au niveau du contenu, mais qu'elle ne corresponde pas au contenu du texte source. En ce qui concerne la traduction de textes spécialisés, DeepL Pro présente clairement des limites. Il a néanmoins été jugé utile de mettre DeepL Pro à la disposition des collaborateurs de l'Administration fédérale (Wymann, 2019).

En 2020, la qualité de DeepL Traducteur montre des améliorations grâce à des innovations dans le domaine de l'architecture des réseaux neuronaux. Le chinois et le japonais s'ajoutent à DeepL Traducteur et une nouvelle fonctionnalité appelée « glossaire » permet aux utilisateurs de choisir comment certains mots ou expressions doivent être traduits (DeepL, 2021).

En 2021, DeepL lance une nouvelle application mobile pour iPhone et iPad, ce qui permet d'accéder encore plus facilement à DeepL et ses traductions (DeepL, 2021).

2.4.2 ETRANSLATION

Le service de traduction automatique de la Commission européenne, appelée eTranslation, a été élaboré par la direction générale de la traduction (DGT) de la Commission européenne et a été lancé officiellement en 2017. « Ce service est destiné aux administrations publiques européennes, aux petites et moyennes entreprises et aux départements de langues des universités, ainsi qu'aux projets relevant du mécanisme pour l'interconnexion en Europe » (Commission européenne, 2022b). Le service de eTranslation permet de traduire à partir et vers toutes les langues officielles de l'Union européenne (il y a 24 langues officielles dans l'UE, donc 552 combinaisons sont possibles). Sur son site web, la Commission européenne (2022) explique : « eTranslation's goal is to provide quality machine translation for all EU languages in a secure system that protects privacy and passes on Intellectual Property rights to the requestor of the machine translation ».

L'histoire de la traduction automatique à la Commission européenne commence en 1975 avec le système appelé « Systran ». Ce système était l'un des premiers systèmes de traduction automatique de l'anglais vers le français, fondé sur une architecture linguistique (section 2.3.1). En 1981, le service appelé « ECMT » a été créé à partir de ce prototype. Il était capable de traduire 28 paires de langues après presque trois décennies de développement et d'utilisation. Le service « ECMT » a finalement été utilisé pour traduire environ 2 millions de pages par an et mis à

disposition des utilisateurs dans d'autres institutions européennes et dans les États membres (Eisele, 2017).

ECMT était un système qui se basait sur une interaction complexe de règles linguistiques pour l'analyse, le transfert et la génération. Lorsque 10 pays supplémentaires ont rejoint l'UE en 2004, augmentant ainsi le nombre de langues officielles de 12 à 21, il était clair que le développement de systèmes similaires pour 9 langues complexes appartenant à 4 nouvelles familles de langues ne serait pas faisable. (Eisele, 2017).

En 2006, motivée par les progrès significatifs dû à la nouvelle stratégie statistique, la Commission européenne commence à financer une nouvelle série de projets de recherche avec l'objectif d'appliquer systématiquement les technologies de TA les plus avancées à toutes les paires de langues de l'UE, avec une attention particulière pour les langues des nouveaux États membres (Eisele, 2017).

En 2009, la direction générale de la traduction (DGT), en coopération avec la direction générale de la société de l'information et des médias (DG INFSO), a commencé à élaborer une nouvelle stratégie pour la TA et le nouveau projet, appelé MT@EC, a été lancé. Après plusieurs mois de travail sur les systèmes statistiques, la première génération de systèmes de TA a été mise à la disposition des utilisateurs dans la DGT. En 2013, la première version du service MT@EC a été officiellement inaugurée, couvrant les 24 langues de l'UE (Eisele, 2017).

Depuis son inauguration, l'utilisation de MT@EC n'a cessé d'augmenter. En 2016, plus de 24 millions de pages ont été traduites par ce système. Mais les observations des utilisateurs de la DGT ont rapidement révélé que le système comportait certaines erreurs difficiles à détecter et à corriger par les traducteurs (Eisele, 2017).

Le multilinguisme qui caractérise l'UE a donné un nouvel élan aux activités de TA au sein de la Commission européenne, et ainsi, le nouvel service appelé « eTranslation » a été créé par la DGT, afin de rendre tous les « Digital Service Infrastructures » (DSIs) multilingues (Eisele, 2017).

Étant donné que les résultats de la recherche sur les systèmes neuronaux en 2016 étaient très prometteurs et que cette nouvelle méthode avait un grand potentiel, l'équipe du service MT@EC a commencé à travailler sur l'application de systèmes neuronaux sur les langues cibles pour lesquelles les performances des systèmes statistiques étaient les plus faibles (Eisele, 2017). Au début, en 2017, eTranslation utilisait des réseaux neuronaux pour la traduction de l'anglais vers l'allemand, l'hongrois, le finnois et l'estonien, et vice versa. À partir de 2018, a été d'ajouter une nouvelle paire de langues au système neuronal chaque mois :

« The general strategy is to upgrade to neural technology first for languages which are poorly served by the current statistical methods. Statistical engines remain available as well, initially through the MT@EC web page, and in future by choosing ‘MT@EC’ as a domain from the eTranslation web page. » (Commission européenne, 2022a).

Selon le site web de la Commission européenne, eTranslation présente les caractéristiques suivantes :

« Niveau élevé de sécurité : toutes les données traitées par le système sont protégées par les pare-feux de la Commission. Elles ne sont pas accessibles aux personnes extérieures ; Traduction à partir de et vers toutes les langues officielles de l’UE ; Meilleurs résultats obtenus avec les textes portant sur l’UE ; Gratuit jusqu’à nouvel avis [...] » (Commission européenne, 2022b)

Markus Foti (2022) explique dans son article que la traduction automatique n’est pas aussi fiable que la traduction humaine, mais qu’elle est moins onéreuse et plus rapide :

« C’est ainsi qu’eTranslation a permis à la Commission de transmettre rapidement des informations sur les voyages pendant la pandémie de Covid, par le biais de la page internet Re-Open EU (réouverture des frontières de l’UE). [...] La traduction professionnelle humaine reste la référence suprême, mais la technologie de la traduction automatique nous permet d’atteindre l’ensemble du public européen mieux que jamais auparavant. À une époque où les traducteurs humains produisent également davantage et subissent plus de pression que jamais, les 205 millions de pages traduites en 2021 prouvent que la traduction, qu’elle soit humaine ou automatique, a un rôle crucial à jouer pour permettre une conversation véritablement européenne. » (Foti, 2022).

CONCLUSION

Dans ce chapitre, la notion de la traduction automatique a été définie et son évolution a été décrite. Après la présentation des différentes approches et systèmes de traduction automatique, l'aperçu de l'évolution élaboré par Quah (2006) est repris en résumé dans la Figure 8 :

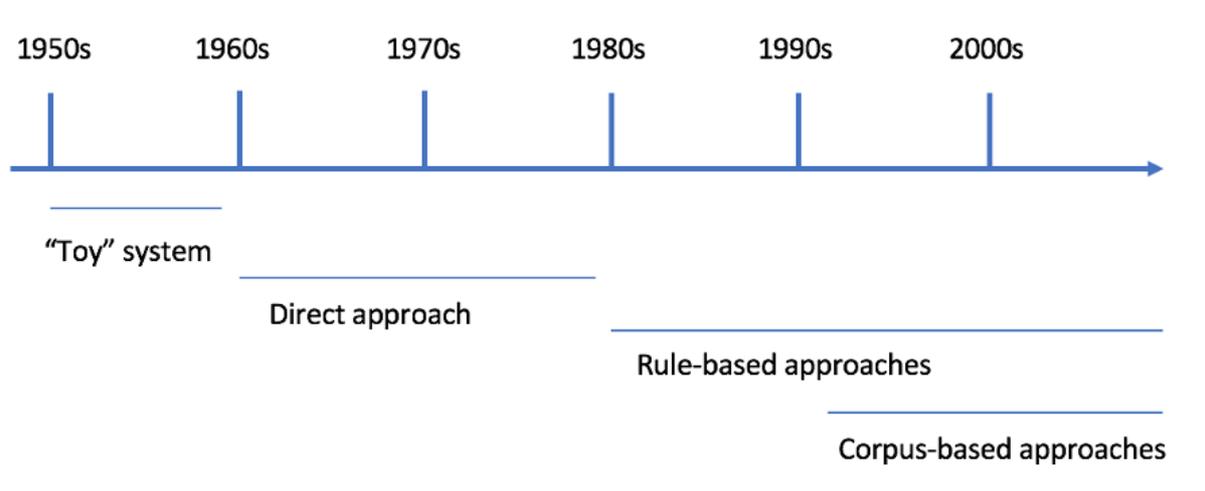


Figure 8 Chronology of machine translation development Quah 2006, p. 58

Avec l'apparition des systèmes neuronaux à partir du milieu des années 2010, les limites des autres approches ont été résolues. Cette nouvelle approche neuronale semble être la plus prometteuse jusqu'à maintenant, mais ils présentent aussi des inconvénients : les systèmes neuronaux requièrent une grande puissance computationnelle et les erreurs produites par les systèmes neuronaux peuvent parfois être assez naïves, parce que leurs traductions peuvent sembler être de bonne qualité, même si elles ne correspondent pas au texte source.

3. LA TRADUCTION JURIDIQUE

Ce chapitre vise à donner une définition de la traduction juridique et à présenter les différentes typologies de la traduction juridique (section 3.1). Ensuite, nous expliquons la traduction de textes normatifs en Suisse (section 3.2). Finalement, l'utilisation de la TA dans le domaine juridique est présentée (section 3.3).

3.1 DÉFINITION ET TYPOLOGIES

La traduction peut être classée en plusieurs catégories (Cao, 2007). Une typologie très répandue distingue la traduction générale (p.ex. la traduction des discours écrits), la traduction littéraire (la traduction d'œuvres littéraires) et la traduction spécialisée (la traduction des textes scientifiques, économiques, etc.) (Pelage, 2001, p. 62-65). La traduction juridique fait partie de la catégorie de la traduction spécialisée (Cao, 2007), qui implique une utilisation spécialisée du langage, c'est-à-dire le langage dans un contexte juridique (Cao, 2013, p. 415). Ce langage spécialisé, ou langue de spécialité (Koutsivitis, 1988), présente les caractéristiques suivantes : « une nomenclature lexicale différente de la langue courante, une sémantique particulière où les termes jouent un rôle principal, une syntaxe et une stylistique particulière. » (Pelage, 2001, p. 30). Ce langage diffère d'une culture et d'une langue juridique à l'autre (Gémar, 2011, p. 51). Gémar (2011) décrit trois caractéristiques qui distinguent les textes juridiques des autres textes :

« Il s'agit d'un texte le plus souvent normatif – la loi, le jugement ou le contrat ; il présente un vocabulaire généralement peu courant ; il est rédigé dans un style propre à un genre : législatif, judiciaire, contractuel, notarial, doctrinal, etc. » (Gémar, 1995, cité dans Gémar, 2011, p. 51)

Le langage juridique peut être utilisé dans différents buts de communication : informatif, argumentatif ou normatif (Ruano, 2020, p. 281). Selon Monjean-Decaudin (2014), « la traduction juridique répond à une nécessité spécifique résultant d'une obligation normative : un texte juridique est traduit afin de se conformer aux dispositions d'une norme (internationale, régionale ou nationale) ». Les textes normatifs, tels que les lois, les traités internationaux ou les contrats de droit privé sont juridiquement valides et donc contraignants. Cet aspect distingue les textes juridiques des autres textes spécialisés (Sarcevic, 1999, p. 103). Legault (1977, cité dans

Sarcevic, 1999, p. 103) explique : « on ne peut pas se permettre de traduire une loi comme on traduit un autre texte, car la traduction est aussi loi ».

La traduction juridique elle-même peut également être classée en fonction de différents critères. Les classifications traditionnelles se basent sur la fonction ou l'utilisation des textes juridiques originaux dans la langue source, sans tenir compte de différents facteurs de la langue cible, tels que les fonctions ou le statut des textes traduits. Mais il est nécessaire de considérer les facteurs de la langue cible en complément de celles de la langue source (Cao, 2013, p. 415-416). Cao (2013) propose trois classifications de la traduction juridique sur la base des buts des textes de la langue cible :

« (i) normative purpose, i.e. the production of equally authentic legal texts in bilingual and multilingual jurisdictions of domestic laws and international legal instruments and other laws; (ii) informative purpose, e.g. the translation of statutes, court decisions, scholarly works and other types of legal documents if the purpose of the translation is to provide information to the target readers; and (iii) general legal or judicial purpose. »

Monjean-Decaudin (2014) fait la distinction entre la traduction-instrument et la traduction-document en se basant sur la typologie proposée par Christiane Nord dans le cadre de la théorie du skopos, qui a été adaptée à la traduction des textes législatifs par Valérie Dullion : « Pour les textes auxquels sont attachés des effets juridiques, la traduction est susceptible de remplir deux catégories de fonctions : informer sur le droit ; produire un texte qui aura une valeur en droit » (Dullion, 2000, cité dans Monjean-Decaudin, 2014). Selon Monjean-Decaudin (2014), la traduction-instrument vise à produire un texte qui aura une valeur directement utilisable dans le pays, et la traduction-document, par contre, vise à informer sur le droit.

Chaque pays crée son propre système juridique et le droit ne tient pas la même place dans tous les pays, ce qui pose différents problèmes aux traducteurs juridiques (Pelage, 2001). Selon Sandrini (1999), la traduisibilité des textes juridiques est directement liée à la proximité des systèmes juridiques :

« Gehören beide Rechtsordnungen demselben Rechtskreis an, kann eine relative Nähe der Rechtsinhalte angenommen werden; bei unterschiedlichen Rechtskreisen führen Tradition und Rechtsauffassung zu völlig verschiedenen Rechtslösungen mit entsprechenden Auswirkungen auf die Übersetzbarkeit. » (De Groot, 1991, cité dans Sandrini, 1999, p. 17)

Il est donc difficile de traduire des notions et concepts qui existent dans la langue source, mais pas dans la langue cible, ce qui demande l'application des différentes stratégies de traduction.

Gémar (2011, p. 59) présente quatre catégories différentes de contexte culturel, linguistique et juridique, qui ont été proposées par Jacques Vanderlinden en 2002 : Dans la première catégorie se retrouvent les pays unilingues qui possèdent un seul système juridique (p.ex. la France, le Mexique, le Brésil) ; la deuxième catégorie est composée de pays bilingues (p.ex. le Canada, La Belgique) ou multilingues (p.ex. la Suisse ou l'Espagne) ; dans la troisième groupe se trouvent des pays où règnent à la fois un bi- ou multilinguisme et un bijuridisme – c'est-à-dire une coexistence de deux traditions juridiques (Gouvernement du Canada, 2021) – comme c'est le cas dans certains pays à régime fédéral (Belgique, Canada, Inde) ; la quatrième catégorie est composée de pays à la fois bi- ou multilingues et bi- ou multisystémiques (Canada, Sri Lanka, Afrique du Sud).

La section suivante vise à présenter plus en détail la traduction de textes normatifs en Suisse.

3.2 TRADUCTION DE TEXTES NORMATIFS EN SUISSE

La Suisse, qui est un État qui reconnaît trois langues officielles nationales (l'allemand, le français et l'italien), mais avec un seul système juridique, peut être classée dans la deuxième catégorie proposée par Vanderlinden (dans la section 3.1). La Chancellerie fédérale suisse (2022) explique sur leur site web :

« Les linguistes de la Chancellerie fédérale veillent à ce que, dans chaque langue, les lois et les autres textes officiels soient compréhensibles tout en étant corrects sur le plan juridique. Dans le cas des textes législatifs, c'est un travail subtil parce que les textes sont équivalents sur le plan juridique dans les trois langues officielles que sont l'allemand, le français et l'italien. »

La Commission interne de rédaction (CIR), une commission interdisciplinaire et interdépartementale, responsable de cette tâche, est composée de linguistes des Services linguistiques centraux de la Chancellerie fédérale et de juristes de l'Offices fédérale de la justice.

« La CIR soumet les projets de textes législatifs importants (Constitution fédérale, arrêtés fédéraux, lois, ordonnances importantes) à une procédure de corédaction : elle

examine simultanément les versions allemande et française des projets – parfois aussi la version italienne-, les compare et les remanie. » (Chancellerie fédérale, 2022).

Selon Schweizer et al. (2011), il semble problématique que l'allemand tienne une place dominante, parce que ce fait influence la qualité du droit multilingue. Pour éviter ce problème, les auteurs recommandent de développer au moins une terminologie dans la phase d'élaboration, qui soit utile et applicable pour toutes les langues officielles.

La traduction juridique passe généralement par le droit comparé, parce que les notions juridiques, définies dans le contexte de chaque société, sont différentes d'une culture à l'autre (Dullion, 2014). « [...] le droit comparé a pour objet de rechercher et d'expliquer les différences entre les divers ordres juridiques à une époque déterminée » (Muir Watt, 2022). Mais dans les systèmes juridiques multilingues, « c'est un seul et même droit qui se dit dans plusieurs langues, les termes juridiques font théoriquement référence à des notions uniformes, définies par les sources du système en question, si bien que le droit comparé devient superflu » (Dullion, 2014). En Suisse, le droit comparé est toutefois utile dans le cadre du processus législatif, quand p.ex. une institution étrangère est mentionnée dans un rapport, dont le législateur suisse pourrait s'inspirer, ou quand il s'agit de légiférer dans un domaine qui évolue rapidement, ou encore, lorsqu'il a affaire à des nom de tribunaux (Dullion, 2014).

La traduction de textes normatifs ne consiste pas à simplement reproduire le contenu à des fins d'information, mais à produire un texte normatif qui sera interprété et appliqué par les tribunaux et autres institutions. La tâche du traducteur consiste toujours à créer un texte qui permet d'interpréter et d'appliquer de manière uniforme tous les textes sources et cibles de la même loi, du même traité ou de la même convention. Le traducteur devient ainsi un participant responsable au processus de communication entre le législateur et celui qui applique le droit (Sarcevic, 1999, p. 105).

3.3 TA DANS LE DOMAINE JURIDIQUE

Les textes juridiques sont considérés comme les textes spécialisés les plus complexes (Killman, 2014). Ils présentent toute une série de caractéristiques (comme les différences entre les systèmes juridiques, entre les branches du droit et entre les types de texte) qui posent de grands défis aux systèmes de TA. Ces défis incluent toutes les caractéristiques qui posent un problème aux systèmes de TA en général (Wiesmann, 2019). Matthiesen (2017) liste les défis

suivants : la complexité des phrases, les divergences et les ambiguïtés lexicales et syntaxiques, les phraséologies, , la terminologie, les abréviations, les collocations, et les erreurs dans le texte source. Tous ces défis conduisent à la question de savoir si la TA est déjà capable de traduire des textes juridiques ou au moins certains types ou parties de textes juridiques, sans un grand effort de post-édition (Wiesmann, 2019).

Yates (2006) évalue la performance de Babel Fish (un système de TA direct (section 2.3.1.1) élaboré par Systran) dans le domaine de la traduction juridique en comparant les traductions de textes juridiques en espagnol et en allemand, et elle constate :

« Babel Fish's performance in this study was poor, but it was not uniformly poor. Babel Fish translated the German texts less poorly than the Spanish ones. Furthermore, the civil code excerpts were translated more accurately than the press releases in Spanish, but less accurately in German. These discrepancies suggest that further research – with different languages and possibly different types of texts – is needed to more broadly determine Babel Fish's relative strengths and weaknesses. »

Kit et Wong (2008) évaluent et comparent les systèmes de TA, accessibles en ligne, pour l'utilisation dans le domaine juridique, de différentes langues vers l'anglais. Les conclusions de cette étude montrent :

« It is a fact that the MT quality is still far from satisfactory, let alone perfect, and the 'gistable' translation by MT is far from publishable. However, it should also be recognized that the currently available MT technologies, if properly utilized, have been good enough to serve most of the translation demands for the purposes of information access, interchange, and assimilation. Even for the purpose of information dissemination, professional use of MT can minimize one's workload towards publishable translation. In this regard, MT is more specifically aimed at enhancing human translators' productivity and creativity by not only releasing them from routine work, which is best for machines, but also by providing them with more translation possibilities and, more importantly, the best choices for translating each word, phrase, and clause. »

Farzindar et Lapalme (2009) développent un système statistique basé sur des phrases, appelé TransLI (Translation of Legal Information), pour la traduction automatique des jugements de

la Cour du Canada, de l'anglais vers le français et du français vers l'anglais. L'évaluation de ce système montre :

« Although still not of publishable quality, the translations of the TransLI system that we developed in this project can be readily used for human revision, with promising productivity gains. [...] Legal organizations need solutions that enable them to handle quickly a high volume of translations. Our goal was to study the ability to train translation systems on a specific domain or subject area like the legal field so as to radically increase translation accuracy. This process recycles existing translated content to train the machine on the terminology and style of the requested domain. To our knowledge this is one of the first times that an SMT engine has been developed specifically for judicial texts and evaluated in a pilot study. We managed to establish that an SMT engine trained on an appropriate corpus can produce a cost-effective revisable text. »

Les études menées par Matthiesen (2017) montrent que les systèmes neuronaux sont de meilleure qualité que les systèmes linguistiques et statistiques, vu qu'ils produisent beaucoup plus de traductions grammaticalement correctes. Matthiesen (2017) constate un progrès dans la traduction automatique neuronale, mais il remarque également que ce progrès est dû en grande partie à l'amélioration remarquable de la syntaxe, c'est-à-dire aux traductions plus fluides et grammaticalement plus correctes ; néanmoins, les phrases cibles contiennent encore souvent des erreurs de sens ou le contenu de la phrase source est mal traduit.

Afin de répondre à la question si la TA est capable de traduire des textes juridiques, ou au moins certains types de textes juridiques, vers une autre langue juridique à un niveau suffisant pour que l'effort de post-édition soit limité, Wiesmann (2019) évalue et compare deux systèmes : DeepL Traducteur (système de TA) et MateCat (système de TAO qui intègre la TA). Les résultats de cette étude montrent :

« Overall, the results are considered insufficient to give post-editing of machine-translated legal texts a bigger place in translation pedagogy. As the evaluation of the correspondence between source and target text was fundamentally worse than with regard to the meaningfulness of the target text, translation pedagogy should respond by raising awareness about differences between machine translation output and human translation in this field, and by improving translation approach and strengthening legal expertise. »

Vieira et al. (2020) publient un article qui présente une méta-analyse qualitative de documents officiels et de recherches publiées sur l'utilisation de la TA dans le domaine médical et juridique. Les résultats de l'étude montrent :

« First, the review shows that research on MT use in healthcare and law can often disregard the complexities of language and language translation. The article calls for cross-disciplinary research conducted within the medical and legal sectors. Second, the review highlights a broad societal need for higher levels of awareness of the specific strengths and, crucially of the limitations of MT. Finally, the article concludes that MT technology can in its current state exacerbate social inequalities and put certain communities of users at greater risk. We highlight this as a persistent issue that merits further attention from researchers and policymakers. »

En conclusion, on peut dire que tous les auteurs sont d'accords que la TA n'est pas encore capable d'être utilisée dans le domaine juridique sans une intervention humaine. Une post-édition et révision sont toujours nécessaires pour éviter les conséquences risquées d'une fausse traduction dans ce domaine.

Ce mémoire s'intéresse à l'état actuel de la performance de deux systèmes de TA (DeepL et eTranslation) dans le domaine juridique.

CONCLUSION

Dans ce chapitre, nous avons donné une définition de la traduction juridique et présentés les différentes typologies de la traduction juridique. Ensuite, nous avons expliqué la traduction de textes normatifs en Suisse et l'utilisation de la TA dans le domaine juridique.

4. ÉVALUATION DE LA TA

Ce chapitre vise à présenter les méthodes d'évaluation de la traduction automatique. Celles-ci sont catégorisées en évaluation humaine (section 4.1) et en évaluation automatique (4.2).

La première question soulevée lorsqu'on analyse la qualité d'une traduction est de savoir quels critères peuvent être utilisés pour mesurer la qualité (Prieto Ramos, 2015). Selon Poibeau (2019), une bonne traduction doit être le plus fidèle possible au texte source, c'est-à-dire qu'elle doit respecter les caractéristiques, le ton, le style, le détail des idées et de l'argumentation du texte source ; elle doit être intelligible et linguistiquement correcte ; idéalement, le lecteur ne devrait pas se rendre compte qu'il s'agit d'une traduction, donc une bonne traduction doit être idiomatique.

D'autres questions se posent lorsqu'on analyse la qualité d'une traduction réalisée automatiquement : comment mesurer la qualité des systèmes de TA ? Comment comparer deux systèmes de TA entre eux ? Comment mesurer le progrès d'un même système de TA au cours du temps ? (Poibeau, 2019, p. 153). Il n'existe pas de méthode d'évaluation standard. L'une des raisons de l'absence d'une méthode d'évaluation standard est l'existence d'une grande variété de systèmes ; une autre raison est la diversité des groupes intéressés à l'évaluation (Quah, 2006, p. 129). Afin de répondre à la diversité et aux défis de l'évaluation de la TA, différents types de mesures qualitatives et quantitatives ont été développés (Kit & Wong, 2015). Une méthode possible est de demander des juges humains d'évaluer la qualité des traductions, ou de comparer la similarité entre l'output d'un système de TA avec une traduction réalisée par un traducteur humain (Koehn, 2010, p. 217). L'élaboration de mesures d'évaluation automatique au cours de la dernière décennie a été un tournant important dans le domaine de l'évaluation de la TA (Kit & Wong, 2015).

Les deux types d'évaluation (humaine et automatique) seront présentées dans les sections 4.1 et 4.2.

4.1 ÉVALUATION HUMAINE

La méthode la plus reconnue pour évaluer l’output d’un système de TA est de regarder la traduction et de juger si elle est correcte ou non. Les juges bilingues qui comprennent à la fois la langue source et la langue cible sont les plus qualifiés pour effectuer cette évaluation (Koehn, 2010). Selon Kit et Wong, l’évaluation humaine comporte deux aspects : l’évaluation intrinsèque et extrinsèque. L’évaluation intrinsèque vise à juger l’output de systèmes de TA, et l’évaluation extrinsèque vise à juger l’utilisabilité de l’output dans le contexte dans lequel il est utilisé (Kit & Wong, 2015).

4.1.1 ÉVALUATION INTRINSÈQUE

Kit et Wong (2015) distinguent trois stratégies d’évaluation intrinsèque, qui seront présentées dans les sections suivantes.

JUGEMENT INTUITIF

Avec le jugement intuitif, les juges doivent juger intuitivement les traductions phrase par phrase (Kit & Wong, 2015). Le plus souvent, les traductions sont évaluées selon deux critères : fidélité (en anglais : adequacy) et fluidité (en anglais : fluency). La fidélité juge la précision avec laquelle la traduction reproduit le sens exact de la phrase source et la fluidité montre si la traduction est grammaticalement correcte, claire, lisible et idiomatique (Jurafsky, 2020). Selon Poibeau (2019, p. 156), ces critères « semblent plus faciles à évaluer qu’affecter un score global aux traductions produites, mais ils restent subjectifs [...] ». Les juges doivent être bilingues afin d’évaluer la fidélité de la traduction, mais ils peuvent être monolingues s’il existe une traduction humaine comme référence. L’évaluation de la fluidité peut être effectuée par des juges monolingues (Kit & Wong, 2015). Afin d’effectuer l’évaluation de la fidélité et la fluidité, les juges donnent des scores aux phrases sur une échelle, qui varie suivant les études. Un exemple d’échelle d’évaluation de la fidélité et de la fluidité est illustré dans la figure 9 (Koehn, 2020).

	Adequacy	Fluency
5	all meaning	flawless English
4	most meaning	good English
3	much meaning	non-native English
2	little meaning	disfluent English
1	none	incomprehensible

Figure 9 Adaptation de l'échelle de fidélité et fluidité (Koehn, 2020, p.46)

En principe, la fidélité et la fluidité sont indépendantes l'une de l'autre. Des études ont cependant montré qu'elles sont en fait fortement liées (White, 2001, cité dans Kit & Wong, 2015). Il est donc possible de n'évaluer qu'une d'entre elles et d'inférer ensuite l'autre (Kit & Wong, 2015).

ÉVALUATION COMPARATIVE

Au lieu d'évaluer la fidélité et la fluidité selon une échelle, il est aussi possible de comparer l'output de deux ou plusieurs systèmes de TA, phrase par phrase. Selon Koehn (2020), cette méthode d'évaluation est plus facile que l'évaluation de la fidélité et fluidité. On constate en effet que les juges évaluent de manière plus cohérente si l'output d'un système est meilleur, pire ou identique que la fidélité ou la fluidité.

Une autre méthode est de demander aux juges de classer (en anglais : rank) les traductions de différents systèmes de TA du meilleur au pire. La performance des systèmes de TA est alors représentée par le nombre moyen de fois où leur output sont classés plus haut que les autres. Le « ranking » de traductions était la méthode officielle d'évaluation humaine dans les workshops de TA statistique depuis 2008 et a remplacé l'évaluation de la fidélité et fluidité selon un jugement intuitif (Callison-Burch et al., 2009, cité dans Kit & Wong, 2015). Ce changement est dû au fait qu'on obtient un meilleur score pour l'accord entre les juges (score Kappa) avec cette méthode qu'avec l'évaluation de la fidélité et fluidité (Koehn, 2020). Le score Kappa peut être calculé avec la formule suivante :

$$\kappa = \frac{P(a) - P(e)}{1 - P(e)}$$

P(a) représente la probabilité de l'accord observé, qui est le nombre de fois où les juges ont choisi la même catégorie, divisé par le nombre total de jugements. P(e) représente la probabilité d'un accord aléatoire (dû au hasard), c'est-à-dire la « probabilité que si un juge dit une

annotation au hasard, l'autre dise la même chose au hasard » (Bouillon, 2021). L'accord entre les juges est parfait si le résultat du calcul est égal à 1, et il y a un désaccord si le résultat du calcul est égal à 0. Il existe différents scores Kappa selon les méthodes d'annotation : le Cohen's Kappa est utilisé lorsqu'il y a deux juges et que les deux juges ont annoté tous les exemples ; le Light's Kappa est utilisé s'il y a plus que deux juges et si tous les juges ont annoté tous les exemples ; et le Fleiss' Kappa est utilisé lorsqu'il y a plus de deux juges et si les juges sont sélectionnés au hasard pour chaque exemple (Bouillon, 2021).

ANALYSE D'ERREURS

L'analyse (ou l'annotation) d'erreurs est une autre méthode pour évaluer la qualité de systèmes de TA et pour analyser si l'output d'un système de TA est acceptable pour la post-édition (Alvarez-Vidal et al., 2021). Cette méthode est un indice de la quantité de travail nécessaire pour corriger l'output « brut » d'un système de TA, afin de le rendre acceptable (Hutchins & Somers, 1992, p. 164). Pour effectuer l'analyse d'erreurs, il faut d'abord créer une liste des différents types d'erreurs, et pendant l'évaluation, toutes les erreurs sont comptées (Arnold et al., 1994, p. 174). Les erreurs peuvent être classées de différentes manières en fonction de la variété et de la complexité des systèmes de TA, en fonction des caractéristiques grammaticales des textes dans divers domaines ou langues, et en fonction des demandes des utilisateurs (Kit & Wong, 2015).

Vu que certaines erreurs sont plus graves que d'autres, chaque type d'erreur sera multiplié par poids (Arnold et al., 1994, p. 174), qui peut dépendre de l'impact de l'erreur ou de la difficulté de la correction (Bouillon, 2021). Mossop (2020) fait la distinction entre les erreurs majeures, mineurs et critiques (en anglais : major, minor and critical errors) : l'erreur est considérée comme majeure quand l'aspect central du message est mal traduit, elle est considérée comme mineure si elle concerne un autre aspect, p.ex. s'il y a une erreur dans une note de bas de page, vu que les notes de bas de page contiennent rarement des aspects centraux du message. L'erreur est considérée comme critique lorsqu'elle a un impact négatif non seulement sur la compréhension du message, mais aussi sur la santé, la sécurité, les finances, la responsabilité juridique du client qui publie la traduction, etc. (Mossop, 2020).

Selon Trujillo (1999, p. 261), l'analyse d'erreurs pose au moins trois problèmes : premièrement, déterminer ce qu'est une traduction acceptable est une question subjective ; deuxièmement, la détermination du poids de certains types d'erreurs peut impliquer des motivations conflictuelles ; et troisièmement, la classification des erreurs peut être problématique.

4.1.2 ÉVALUATION EXTRINSÈQUE

Kit et Wong (2015) distinguent trois méthodes d'évaluation extrinsèque : l'extraction d'informations, l'évaluation de compréhensibilité et la post-édition. Deux de ces méthodes seront présentées dans la section suivante.

ÉVALUATION DE LA COMPRÉHENSIBILITÉ

Une méthode pour évaluer la compréhensibilité est de donner aux juges des segments de traductions réalisées automatiquement et des segments de traductions réalisées par des traducteurs humains, et ensuite, de poser des questions par rapport aux segments. Cette méthode permet d'évaluer si les traductions réalisées automatiquement par un système de TA sont précises et compréhensibles (Kit & Wong, 2015).

Une autre stratégie présentée par Reeder est de demander aux juges de déterminer si les segments de traductions sont produits automatiquement ou par un traducteur humain (Reeder, 2001, cité dans Kit & Wong, 2015).

Il est aussi possible d'emprunter des méthodes qui viennent des sciences cognitives et par exemple de masquer certains mots dans la traduction produite automatiquement et de demander les juges de compléter ces trous dans le texte (« gap filling ») La qualité de l'output de la TA par rapport à la lisibilité ou à la compréhensibilité est indiquée par la mesure dans laquelle la TA permet aux juges de deviner correctement les mots manquants (Kit & Wong, 2015).

POST-ÉDITION

La post-édition et l'annotation des erreurs sont des tâches fortement liées. En fait, la post-édition peut être considérée comme une annotation d'erreurs implicite, puisque chaque opération d'édition est en fait une correction d'une erreur de traduction. Néanmoins, ces deux tâches ont presque toujours été séparées et effectuées indépendamment (Popovic & Arcan, 2016). Les tâches d'un post-éditeur consistent à éditer, modifier et/ou corriger un texte traduit par un système de TA d'une langue source vers une langue cible (Alvarez-Vidal et al., 2021). La qualité d'un système de TA dépend de l'effort de post-édition. Cette méthode permet de mesurer la rentabilité de la post-édition. Dans le pire des cas, la post-édition peut prendre encore plus de temps que la traduction humaine du texte source (Kit & Wong, 2015).

Ramlow (2009, p. 298) distingue trois catégories d'effort de post-édition : effort maximal (qui concerne les erreurs qui nécessitent un effort maximal pour la correction et pour lesquelles une comparaison entre le texte source et le texte cible est nécessaire), effort moyen (qui reprend les erreurs qui nécessitent une reformulation de mots et de phrases, sans avoir besoin de comparer le texte source avec le texte cible), et effort minimal (pour les erreurs qui nécessitent une petite correction, sans avoir besoin de comparer le texte source et le texte cible).

Krings (2001) distingue également trois catégories d'effort de post-édition : effort de temps (en anglais : time effort), effort technique (en anglais : technical effort) et effort cognitif (en anglais : cognitive effort). Une méthode pour mesurer l'effort de temps de post-édition est de mesurer le temps nécessaire pour la post-édition. Cette méthode donne une indication claire de la qualité de l'output du système de TA, puisque le coût de la post-édition se reflète directement dans le temps de post-édition (Kit & Wong, 2015). Une méthode pour mesurer l'effort technique de post-édition est de compter le nombre de modifications nécessaires pour post-éditer un output d'un système de TA (Kit & Wong, 2015). L'effort cognitif est l'effort physique et mental nécessaire pour corriger les erreurs d'un système de TA (Krings, 2001).

4.2 ÉVALUATION AUTOMATIQUE

Pour l'évaluation automatique de l'output de systèmes de TA, aucune intervention humaine est nécessaire pendant l'exécution (Kit & Wong, 2015) ; par contre, une traduction produite par un traducteur humain est nécessaire comme référence pour que le système puisse faire la comparaison entre l'output du système de TA et la traduction humaine (Koehn, 2020). Contrairement à l'évaluation humaine qui peut prendre beaucoup de temps et être onéreuse, l'évaluation automatique est beaucoup plus rapide et moins chère, mais elle aussi est moins précise que l'évaluation humaine. L'évaluation automatique peut quand-même être utile pour avoir une idée de la qualité d'une traduction (Jurafsky, 2020).

Cette section vise à présenter quelques méthodes d'évaluation automatique.

4.2.1 PRÉCISION ET RAPPEL

La précision d'une traduction est mesurée en divisant le nombre de mots corrects dans la traduction produite par un système de TA par le nombre total de mots de cette traduction. La traduction du système de TA est comparée avec une ou plusieurs traductions humaines, qui sont appelées la traduction de référence (en anglais : reference translations). Le rappel est mesuré en

divisant le nombre de mots corrects dans la traduction produite par un système de TA par le nombre total de mots de la traduction de référence. Ces deux métriques peuvent être biaisées, parce qu'il est possible d'avoir une précision élevée (l'output contient seulement des mots corrects), et en même temps d'avoir un rappel très bas (beaucoup de mots de la traduction de référence ne sont pas contenus dans l'output du système de TA) ; ou il est possible d'avoir une précision basse et un rappel élevé (Koehn, 2010). Une méthode commune de combiner les deux métriques (précision et rappel) est la **f-measure**. La f-measure est mesurée en divisant le nombre de mots corrects par la moitié de la somme de mots total de l'output et de la traduction de référence (Koehn, 2010).

4.2.2 WER

Le « Word error rate » (WER) est une métrique beaucoup utilisée en reconnaissance vocale. Contrairement à la précision, le rappel et la f-measure, le WER prend en compte l'ordre des mots (Bouillon, 2021). Cette métrique se base sur la distance de Levenshtein, qui est définie comme le nombre minimal de modifications nécessaire pour arriver à la traduction de référence (Koehn, 2010). Le WER peut être mesuré en divisant le nombre d'étapes d'édition minimales nécessaires pour arriver à la référence (insertions, suppressions, substitutions) par le nombre de mots de la référence (Bouillon, 2021).

4.2.3 TER

Le « translation error rate » ou « translation edit rate » (TER) mesure l'effort d'édition de l'output d'un système de TA pour arriver à la traduction de référence (Shterionov et al., 2017). Pour mesurer le TER, on compte les insertions, les suppressions et les substitutions des mots, mais aussi les déplacements (en anglais : shifts) des mots (Koehn, 2010).

Le TER peut être calculé en divisant le nombre de modifications par le nombre moyen de mots dans la référence (Snover et al., 2006).

4.2.4 BLEU

BLEU (BiLingual Evaluation Understudy) est la métrique la plus utilisée pour l'évaluation automatique de la TA (Kit & Wong, 2015), notamment parce qu'elle est indépendante du langage (Shterionov et al., 2017). La tâche du BLEU consiste à comparer les n-grammes de la traduction automatique avec les n-grammes de la traduction de référence et à compter le nombre

de correspondances (Papineni et al., 2002). Pour effectuer cette comparaison, la traduction automatique et la traduction de référence sont coupées en segments de longueur 1 à n, afin de compter le nombre de segments communs. Généralement, il est considéré que le résultat le plus fiable est obtenu avec des segments de longueur $n = 4$ (Poibeau, 2019, p. 158). Les segments communs sont indépendants de leur position, et plus il y a de segments communs, plus la traduction produite automatiquement est de bonne qualité (Papineni et al., 2002). BLEU permet donc de mesurer la précision de l'output d'un système de TA (Shterionov et al., 2017). Si tous les segments de la traduction automatique et de la traduction de référence sont identiques, alors le score BLEU est égal à 1 ; si aucun segment n'est commun, le score est égal à 0. Donc, plus il y a des segments communs, plus le score BLEU sera proche de 1 (Poibeau, 2019, p. 158).

Même si le score BLEU est beaucoup utilisé, il présente quelques limites : le BLEU ne prend pas en compte l'importance des mots, et tous les mots sont traités de la même manière et avec le même poids. Une autre limite du score BLEU est qu'il ne prend pas en compte la cohérence grammaticale de la phrase dans son ensemble (Koehn, 2010).

Il existe différentes métriques similaires au BLEU, par exemple les scores NIST et METEOR. Le score NIST repose sur les mêmes principes que le score BLEU : « les deux textes à comparer, 'Tref' et 'Tauto' sont découpés en segments, et la mesure est fondée sur une comparaison du nombre de segments de 'Tauto' contenus dans 'Tref' » (Poibeau, 2019, p. 159). La différence entre le score BLEU et le score NIST est que le score NIST donne des poids différents aux mots. Le score NIST prendrait ainsi mieux compte la diversité informationnelle des textes à traduire (Poibeau, 2019, p. 159). Le score METEOR a été conçu pour répondre explicitement à plusieurs faiblesses observées dans le score BLEU. Contrairement au score BLEU, le score METEOR prend en compte non seulement les mots identiques, mais aussi les variantes morphologiques et les synonymes (Banerjee & Lavie, 2005). Le score METEOR prendrait donc mieux en compte les particularités linguistiques des textes analysés (Poibeau, 2019, p. 159).

Des études menées par Doddington en 2002 ont montré que les systèmes avec des scores automatiques bas ont également obtenu un score d'évaluation humaine bas ; et à l'inverse, les systèmes avec des scores automatiques élevés ont obtenu des scores d'évaluation humaine élevés. Cette constatation montre la forte corrélation entre les scores humains et automatiques. La méthode la plus souvent utilisée pour calculer la corrélation entre deux métriques est le coefficient de Pearson (en anglais : Pearson's correlation coefficient). Cette méthode permet de mesurer la corrélation entre les métriques d'évaluation automatique humaine (Koehn, 2010).

CONCLUSION

Nous avons vu que l'évaluation de la TA est une tâche complexe et qu'il y a beaucoup de méthodologies différentes pour l'évaluer. Dans ce chapitre, nous avons présenté quelques méthodes d'évaluation de la TA. Tandis que l'évaluation humaine peut être plus précise et fiable que l'évaluation automatique, l'évaluation automatique est beaucoup plus rapide, objective et moins chère que l'évaluation humaine. Pour choisir l'évaluation adéquate, il faut toujours tenir compte du contexte et de l'objectif de l'évaluation qu'on veut effectuer.

5. MÉTHODOLOGIE

Ce chapitre vise à expliquer la méthodologie utilisée pour l'évaluation des deux systèmes de TA pour le domaine juridique. Dans la suite, nous expliquons la démarche de préparation du texte source et de la traduction (section 5.1). Ensuite, la méthodologie de l'évaluation humaine (section 5.2) et de l'évaluation automatique (section 5.3) est expliquée. Finalement, la méthodologie pour mesurer l'interaccord entre les juges et la corrélation entre les différentes métriques utilisées est présentée (section 5.4).

5.1 DONNÉES : PRÉPARATION DU TEXTE SOURCE ET TRADUCTION

Pour répondre à la question de recherche : « *La qualité de la traduction automatique neuronale est-elle satisfaisante dans le domaine juridique ?* », avec les deux sous-questions : « *La qualité de la traduction automatique dépend-elle des systèmes ?* » et « *La qualité de la traduction automatique dépend-elle des évaluateurs ?* », un texte juridique de 100 segments avec 2106 mots en français a été choisi. Le corpus choisi est la loi COVID-19 qui est entrée en vigueur le 26 septembre 2020. Ce corpus comprend l'article 1 (Objet et principes), l'article 2 (Mesures dans le domaine des droits politiques), l'article 3 (Mesures dans le domaine des capacités sanitaires), l'article 4 (Mesures dans le domaine de la protection des travailleurs), l'article 5 (Mesures dans le domaine des étrangers et de l'asile), l'article 6 (Mesures en cas de fermeture des frontières), et quelques notes de bas de page de la loi COVID-19. Ce corpus a été choisi, parce que le sujet « COVID » est très actuel, qui contient en partie des nouveaux termes. L'idée de voir comment les deux systèmes de TA traduisent les nouveaux termes nous a semblé très intéressante.

Les 100 segments sélectionnés ont été nettoyés, en particulier les chiffres de référence pour les renvois aux notes de bas de page ont été enlevés du texte original, puis ont été traduits par les deux systèmes de TA vers l'allemand. Les traductions de ces 100 segments avec DeepL et avec eTranslation ont été utilisées pour l'évaluation humaine et pour l'évaluation automatique.

Le tableau 1 résume les données du texte source et des traductions produites par les deux systèmes de TA :

Nombre de mots du texte source	Nombre de mots du texte cible produit par DeepL	Nombre de mots du texte cible produit par eTranslation
2106	1725	1721

Tableau 1 Données du texte source et des traductions produites par DeepL et eTranslation

5.2 ÉVALUATION HUMAINE

Pour l'évaluation humaine, deux méthodes ont été choisies : l'annotation d'erreurs et l'évaluation de la compréhensibilité, qui sont expliquées dans les sections suivantes.

5.2.1 ANNOTATION D'ERREURS

Pour l'annotation d'erreurs (section 4.1.1), une grille d'évaluation a été créée, qui se base sur les « Multidimensional Quality Metrics » (MQM), publié par le Centre allemand de recherche sur l'intelligence artificielle (DFKI) en 2015. Le DFKI propose une liste de catégories (en anglais : Core) qui représentent les problèmes les plus courants dans l'évaluation de la qualité des textes traduits. « The Core represents a relatively high level of granularity suitable for many tasks. Where possible, users of MQM are encouraged to use issues from the Core to promote greater interoperability between systems. » (Lommel et al., 2015).

Les différentes catégories proposées par le DFKI sont résumées dans la Fig. 10 :

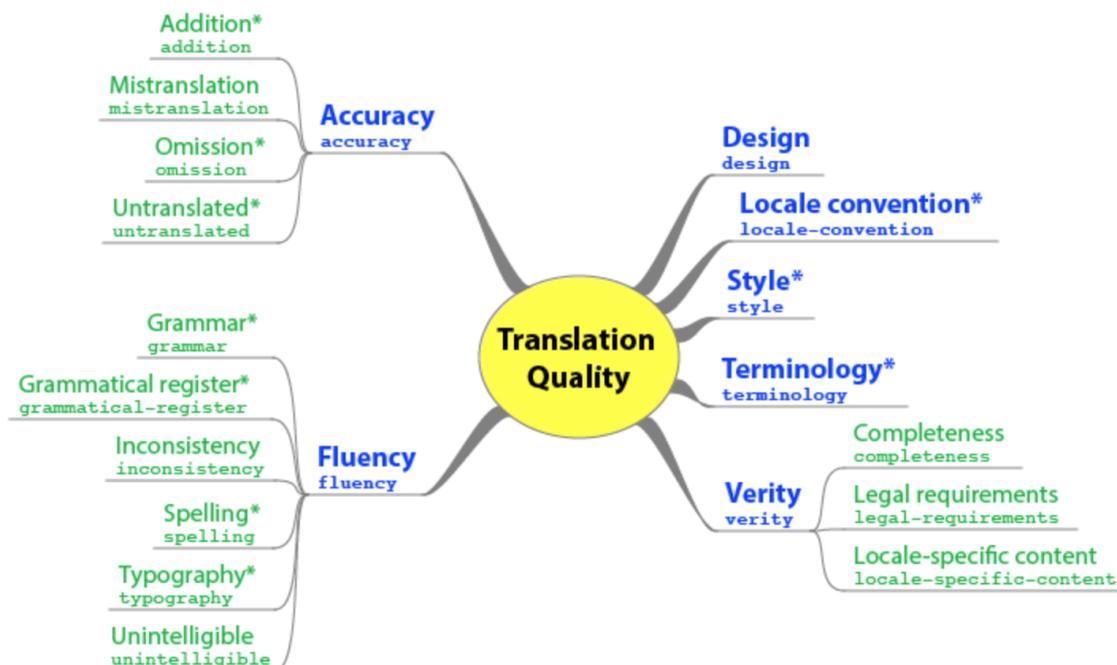


Figure 10 MQM Core - Lommel et al., 2015

Pour ce travail, 8 catégories d’erreurs ont été retenues. Dans la catégorie « accuracy » (fidélité) se trouvent 4 sous-catégories : addition, omission, fausse traduction (en anglais : mistranslation) et non traduit (en anglais : untranslated). Vu que les systèmes neuronaux sont capables de produire des traductions très fluides, la catégorie « fluency » (fluidité) a été omise. La cinquième catégorie de cette grille d’évaluation est le style : « Style issues relate to what is commonly known as “Style”, defined both formally (in style guides) and informally (e.g., a “light style” or an “engaging style”). These issues are closely related to fluency, but are often treated separately by tools and quality processes and so are grouped as a separate dimension in MQM. » (Lommel et al., 2015). Vu que le texte source est une loi suisse, la sixième catégorie est la convention locale : « Issues in Locale convention relate to the formal compliance of content with locale-specific conventions [...] » (Lommel et al., 2015). La septième catégorie est la terminologie : « Terminology issues relate to the use of domain- or organization-specific terminology (i.e., the use of words to relate to specific concepts not considered part of general language). » (Lommel et al., 2015). La dernière catégorie « abrégations » pourrait être classée dans la catégorie de la terminologie, mais vu que les abrégations apparaissent très fréquemment dans les textes juridiques, il a été décidé de séparer ces deux catégories.

Trois poids d’erreurs différents ont été attribués à chaque catégorie : mineur, majeur et critique. Le score d’une erreur mineure est de 1, celui d’une erreur majeure de 10 et d’une erreur critique de 100. Le score de pénalité pour chaque segment est obtenu en divisant la somme de tous les scores d’erreur par le nombre de mots contenus dans le segment.

La grille d’évaluation pour l’annotation des erreurs est illustrée dans la Fig. 11 :

	Source	Translation	Accuracy									Abbreviations	Locale convention			Style			Terminology			words	Penalties							
			Addition			Mistranslation			Omission				Untranslated																	
			minor	major	critical	minor	major	critical	minor	major	critical		minor	major	critical	minor	major	critical	minor	major	critical			minor	major	critical	minor	major	critical	
	Loi fédérale sur les bases légales des ordonnances du Conseil fédéral visant à surmonter l’épidémie de COVID-19	Bundesgesetz über die Rechtsgrundlagen für Verordnungen des Bundesrates zur Überwindung der COVID-19-Epidemie																										12	0	
	(Loi COVID-19)	(COVID-19-Gesetz)																											1	0
	du 25 septembre 2020 (Etat le 1 ^{er} janvier 2022)	vom 25. September 2020 (Stand am 1. Januar 2022)																											9	0
	L’Assemblée fédérale de la Confédération suisse,	Die Bundesversammlung der Schweizerischen Eidgenossenschaft,																											5	0

Figure 11 Grille d’évaluation pour l’annotation d’erreurs

L’annotation d’erreurs a été effectuée par une experte de droit bilingue (français et allemand).

Les résultats de cette annotation d'erreurs seront présentés dans le chapitre 6 et discutés dans le chapitre 7.

5.2.2 COMPRÉHENSIBILITÉ

Pour l'évaluation de la compréhensibilité, une autre grille d'évaluation a été créée. Pour cette évaluation, la méthode du jugement intuitif (section 4.1.1) a été utilisé. Une échelle d'évaluation de la compréhensibilité a été créée sur la base de l'échelle d'évaluation de Koehn (2020, p. 46), qui est reprise dans la Fig. 12 :

Score	Comprehensibility
5	Flawless
4	Good
3	Unclear
2	Difficult to understand
1	Incomprehensible

Figure 12 Échelle d'évaluation de compréhensibilité

Pour cette tâche, 6 juges, dont 3 locuteurs d'Allemagne et 3 de la Suisse, ont évalué la compréhensibilité des traductions produites par DeepL et eTranslation. Les 3 juges suisses ne sont pas des professionnels ni dans le domaine du droit, ni dans le domaine de la traduction et ils sont monolingues. Les 3 juges allemands, par contre, sont tous des étudiants en traduction de l'Université de Genève et ils sont multilingues, mais dans notre cas, ce facteur ne joue aucun rôle, vu que le texte source n'était pas accessible aux 6 juges, mais seulement les traductions produites par DeepL et eTranslation et la traduction de référence. La version allemande de la loi COVID-19 était à leur disposition pour vérifier le sens original de la phrase dans le cas où la traduction produite par les systèmes de TA n'était pas du tout compréhensible.

La Figure 13 illustre le formulaire d'évaluation de la compréhensibilité :

ID	Translation 1	Score	Translation 2	Score
1	Bundesgesetz über die Rechtsgrundlagen für Verordnungen des Bundesrates zur Überwindung der COVID-19-Epidemie	4	Bundesgesetz über die Rechtsgrundlagen der Bundesratsverordnungen zur Überwindung des COVID-19-Ausbruchs	4
2	(COVID-19-Gesetz)	5	(Covid-19-Gesetz)	5
3	vom 25. September 2020 (Stand am 1. Januar 2022)	5	vom 25. September 2020 (Staat am 1. Januar 2022)	3
4	Die Bundesversammlung der Schweizerischen Eidgenossenschaft,	5	Die Bundesversammlung der Schweizerischen Eidgenossenschaft,	5
5	gestützt auf die Artikel 68 Absatz 1, 69 Absatz 2, 92, 93, 101 Absatz 2, 102, 113, 114 Absatz 1, 117 Absatz 1, 118 Absatz 2 Buchstabe b, 121 Absatz 1, 122, 123 und 133 der Bundesverfassung (BV),	4	gestützt auf die Art. 68 Abs. 1, 69 Abs. 2, 92, 93, 101 Abs. 2, 102, 113, 114 Abs. 1, 117 Abs. 1, 118 Abs. 2, Buchst. b, 121 Abs. 1, 122, 123 und 133 der Verfassung (Verfassung),	5
6	nach Einsicht in die Botschaft des Bundesrates vom 12. August 2020,	3	unter Hinweis auf die Botschaft des Bundesrates vom 12. August 2020,	4
7	beschliesst:	5	beschließt:	5
8	SR 101	5	RS 101	3
9	BBL 2020 6363	3	FF 2020 6363	3
10	Art. 1 Zweck und Grundsätze	5	Art. 1 Gegenstand und Grundsätze	5
11	Dieses Gesetz regelt besondere Kompetenzen des Bundesrates zur Bekämpfung der COVID-19-Epidemie und zur Bewältigung der Auswirkungen der Bekämpfungsmaßnahmen auf Gesellschaft, Wirtschaft und Behörden.	5	Dieses Gesetz regelt die besonderen Zuständigkeiten des Bundesrates zur Bekämpfung des COVID-19-Ausbruchs und zur Bewältigung der Folgen von Bekämpfungsmaßnahmen für Gesellschaft, Wirtschaft und Behörden.	5
12	Der Bundesrat macht von diesen Befugnissen nur in dem Maße Gebrauch, wie es für die Überwindung der COVID-19-Epidemie erforderlich ist.	4	Der Bundesrat nutzt diese Befugnisse nur in dem zur Überwindung des COVID-19-Ausbruchs erforderlichen Umfang.	5
13	Insbesondere macht er von diesen Befugnissen keinen Gebrauch, wenn das angestrebte Ziel auch im Rahmen des ordentlichen oder dringlichen Gesetzgebungsverfahrens rechtzeitig erreicht werden kann.	4	Insbesondere nimmt er diese Befugnisse nicht in Anspruch, wenn das angestrebte Ziel auch im Rahmen des ordentlichen oder dringenden Gesetzgebungsverfahrens rechtzeitig erreicht werden kann.	4

Figure 13 Évaluation de la compréhension

La traduction produite par DeepL a été nommée « Translation 1 » et la traduction produite par eTranslation a été nommée « Translation 2 », pour que les juges puissent évaluer les deux traductions le plus objectivement possible. La traduction de référence, soit la loi COVID-19 en allemand, était à la disposition des juges dans un document séparé. Les 6 juges ont évalué les 100 segments traduits par DeepL et les 100 segments traduits par eTranslation, donc ils ont évalué 200 segments en total.

Le score maximum possible était de 500, vu qu'il y a 100 segments et que l'échelon le plus élevé a le score 5. Les résultats de l'évaluation de la compréhension seront présentés dans le chapitre 6 et discutés dans le chapitre 7.

5.3 ÉVALUATION AUTOMATIQUE

L'évaluation automatique a été effectuée avec les métriques BLEU et TER, afin d'évaluer la précision des deux systèmes de TA.

5.3.1 BLEU

Pour calculer le score BLEU (section 4.2.4), le « Interactive BLEU score evaluator » (<https://www.letsmt.eu/Bleu.aspx>, 19.06.2022) de « Tilde Custom Machine Translation » a été utilisé. Cet outil permet d'effectuer des évaluations comparatives de la qualité des traductions avec un ou plusieurs systèmes de TA et de comparer l'output de ces systèmes avec une traduction humaine :

« With the Interactive BLEU tool, users can see their system evaluation results in a dynamic, sentence-by-sentence graph, or compare two systems to see which performs better. This enables users to evaluate MT quality for each sentence individually, providing deeper insight into the performance of MT engines. » (Tilde, 2022)

Pour évaluer le score BLEU, l'Interactive BLEU score évaluateur demande d'effectuer quatre étapes, dont deux étapes sont optionnelles. La première étape consiste à télécharger le texte source (dans notre cas, le corpus avec les 100 segments de la loi COVID-19 en français), mais cette étape est optionnelle. La deuxième étape consiste à télécharger la traduction humaine de référence. Vu que les lois en Suisse existent dans toutes les langues officielles de la Suisse, nous avons utilisé la version allemande de la loi COVID-19. La troisième étape consiste à télécharger la traduction produite par un système de TA (dans notre cas, p.ex. la traduction produite par DeepL). La dernière étape est également optionnelle et consiste à télécharger la traduction du deuxième système de TA (dans notre cas, p.ex. la traduction produite par eTranslation). L'outil demande de télécharger les fichiers en format « .txt », et « permet aussi à l'utilisateur de choisir s'il souhaite prendre en compte la casse dans le calcul de BLEU (Lowercase) et s'il souhaite que le texte soit tokenisé. » (Bouillon, 2022). La tokenisation consiste à décomposer le texte brut en mots séparés. Pour les langues latines, ça signifie que les mots sont séparés des ponctuations (Koehn, 2010). Dans notre cas, la tokenisation a été choisie, ainsi que la casse (Lowercase).

Après le téléchargement de tous les fichiers et la sélection de toutes les options souhaitées, l'outil est prêt pour calculer le score BLEU.

L'interface de l'outil « Interactive BLEU score evaluator » est illustrée dans la Fig. 14 :

Interactive BLEU score evaluator

Perform comparative quality evaluations of files translated with one or more MT systems. This allows you to compare MT output with human translations and compare the BLEU scores of various MT systems. [Click here to learn more.](#)

Step 0: Pick source file (Optional) Aucun fichier choisi .txt

Step 1: Pick human translated file Aucun fichier choisi .txt

Step 2: Pick machine translated file Aucun fichier choisi .txt

Step 3: Pick second machine translated file (Optional) Aucun fichier choisi .txt

Lowercase Calculate BLEU Display

Tokenized

Difference highlighting

Figure 14 Interface de Interactive BLEU score evaluator

Les résultats du score BLEU seront présentés dans le chapitre 6 et discutés dans le chapitre 7.

5.3.2 TER

Pour calculer le score TER (section 4.2.3), le plateforme MT3 a été utilisé. MT3 (Machine Translation Training Tool) est une plateforme open-source, qui a été élaborée par l'Université de Cordoba en collaboration avec l'Université de Genève. « The main goal of this tool is to make abstraction of the technical details by letting users focus on the important processing steps and helping them understand the internals by visualizing intermediate processing results. » (Cuenca et al., 2020).

Le plateforme MT3 permet l'évaluation automatique de trois métriques : BLEU, WER et TER. Pour calculer le TER, il faut importer deux fichiers en format .txt : premièrement, la traduction de référence (dans notre cas la loi COVID-19 en allemand), et deuxièmement, la traduction automatique (appelée « Hypothesis » sur la plateforme). Le score TER est ensuite calculé automatiquement.

5.4 MESURES STATISTIQUES

Pour mesurer l'accord entre les juges, nous avons utilisé le score Kappa (section 4.1.1) ; le coefficient de Pearson a été choisi pour mesurer la corrélation entre la métrique BLEU et l'évaluation de la compréhensibilité, ainsi que la corrélation entre la métrique BLEU et l'annotation d'erreurs, et la corrélation entre les deux métriques automatiques BLEU et TER. Les scores obtenus sont présentés dans les sections suivantes.

5.4.1 KAPPA

Le score Kappa est calculé pour mesurer l'accord entre les 6 juges qui ont évalué la compréhensibilité des traductions produites par DeepL et eTranslation. Plus les juges sont d'accord, plus on a de chance que l'évaluation soit valide (Bouillon, 2021). Le score Kappa a été calculé en utilisant l'outil R, qui est un langage et un environnement gratuitement disponible pour le calcul statistique et les graphiques (CRAN, 2022). Pour ceci, le Notebook dans la plateforme Colab de Google a été utilisé, qui a été créé par Pierrette Bouillon et Johanna Gerlach dans le cadre du cours de traduction automatique 2 à l'Université de Genève. Cet outil permet d'importer des données et de calculer l'accord observé, le Kappa de Cohen (pour deux juges) et le Kappa de Light (pour plus de deux juges). Pour ce travail, le Kappa de Light est calculé, parce que 6 juges ont évalué les traductions. L'objectif est de calculer 6 scores Kappa : premièrement, le score Kappa pour DeepL évalué par les 3 juges allemands ; deuxièmement, le score Kappa pour eTranslation évalué par les 3 juges allemands ; troisièmement, le score Kappa pour DeepL évalué par les 3 juges suisses ; quatrièmement, le score Kappa pour eTranslation évalué par les 3 juges suisses ; cinquièmement, le score Kappa pour DeepL évalué par les 6 juges ; et finalement, le score Kappa pour eTranslation évalué par les 6 juges.

Pour pouvoir calculer le score Kappa, il faut d'abord importer les évaluations des juges en format .txt, comme illustré dans les Fig. 15 et Fig. 16 :

5	5	5
5	5	5
5	5	5
4	4	5
3	4	5
5	5	5
5	4	5
3	3	2
5	5	4
5	4	4
4	5	5
4	4	4
5	5	5
3	2	2
4	3	4
4	5	3
5	4	5
5	5	5
5	5	5
4	4	5
5	5	5
4	4	5
4	3	5
5	5	4
5	5	5
5	5	5
4	4	4
5	5	4

Figure 15 Extrait d'évaluation par 3 juges.
Chaque colonne correspond aux évaluations
d'un juge sur une échelle de 1 à 5

5	5	5	5	5	5	5	5	5	5
5	5	5	5	5	5	5	5	4	5
5	5	5	5	5	5	5	5	5	5
4	4	5	5	5	5	5	5	4	5
3	4	5	5	5	5	5	5	5	4
5	5	5	5	5	5	5	5	5	5
5	4	5	5	5	5	5	3	5	5
3	3	2	2	1	3	5	5	5	5
5	5	4	4	5	5	5	5	5	5
5	4	4	4	5	5	5	5	5	5
4	5	5	5	5	5	5	5	5	5
4	4	4	4	5	5	5	5	4	5
5	5	5	5	5	5	5	5	4	5
3	2	2	2	4	3	5	5	5	5
4	3	4	4	5	5	5	5	4	5
4	5	3	4	3	4	5	5	5	5
5	4	5	5	5	5	5	5	4	5
5	5	5	5	5	5	5	5	5	5
5	5	5	5	5	5	5	5	5	5
4	4	5	5	5	5	5	5	5	5
5	5	5	5	5	5	5	5	5	5
4	4	5	5	5	5	5	5	5	5
4	3	5	4	5	5	5	5	5	5
5	5	4	5	5	5	5	5	5	5
5	5	5	5	5	5	5	5	5	5
4	4	4	4	5	5	5	5	5	5
5	5	4	5	5	5	5	5	5	5
5	5	5	5	5	5	5	5	5	5
4	4	4	4	5	5	5	5	5	5
5	5	4	5	4	4	4	5	5	5
5	5	4	5	4	4	4	4	5	5
5	5	4	5	4	4	4	4	5	5

Figure 16 Extrait d'évaluation par 6 juges

Après l'importation des 6 fichiers en format .txt, il faut effectuer 4 étapes pour obtenir le score Kappa : La première étape consiste à exécuter la commande `sample 3 ← read.table ('nom du fichier')`, afin d'écrire les données des fichiers dans une variable qui est appelée ici « sample 3 ». La deuxième étape consiste à exécuter une commande pour vérifier l'importation des données dans la variable « sample 3 ». La troisième étape consiste à exécuter une commande pour obtenir un résumé des données dans la variable « sample3 ». La dernière étape consiste à calculer le score Kappa de Light pour mesurer l'accord.

Les résultats des scores Kappa seront présentés dans le chapitre 6 et discutés dans le chapitre 7.

5.4.2 COEFFICIENT DE PEARSON

Afin de mesurer la corrélation entre le score BLEU et l'évaluation de la compréhension, ainsi que la corrélation entre le score BLEU et l'annotation d'erreurs et la corrélation entre le BLEU et le TER, le coefficient de Pearson a été calculé en utilisant la formule PEARSON dans Excel, qui permet de calculer le coefficient de Pearson entre les valeurs de deux colonnes. Les valeurs de ces deux colonnes correspondent d'un part aux scores BLEU pour

chaque segment, et d'autre part aux scores de compréhensibilité (moyennes pour les juges) et de MQM obtenus pour chaque segment, comme illustré dans la Fig. 17 :

	A	B	C
1	variable x (DeepLDE)	variable y (DeepLDE)	
2	p.ex. score BLEU	p.ex. moyenne des scores de compréhensibilité	
3	57.57		4
4	100		5
5	100		5
6	100		5
7	100		4.33
8	91.31		4
9	100		5
10	100		4.66
11	37.99		2.66
12	37.99		4.66
13	90.62		4.33
14	44.38		4.66
15	40.42		4
16	17.2		5
17	4.34		2.33
18	51.87		3.66
19	58.28		4
20	37.07		4.66
21	32.87		5
22	15.62		5
23	33.66		4.33
24	87.02		5

Figure 17 Extrait du fichier Excel pour le calcul du coefficient de Pearson

CONCLUSION

Dans ce chapitre nous avons expliqué la méthodologie utilisée pour l'évaluation des deux systèmes dans le domaine juridique. Pour ce faire, un extrait de la loi COVID-19 en français a été choisi comme texte source, afin d'évaluer les deux systèmes de TA, notamment DeepL et eTranslation, avec des différentes métriques. Pour l'évaluation humaine, l'annotation d'erreurs et l'évaluation de la compréhensibilité a été effectué et pour l'évaluation automatique, les métriques BLEU et TER ont été calculées. Finalement, la corrélation entre les différentes métriques a été calculée avec le coefficient de Pearson.

6. RÉSULTATS

Ce chapitre vise à présenter les résultats de l'évaluation humaine (section 6.1) et de l'évaluation automatique (section 6.2), ainsi que les résultats de l'accord entre les juges et entre les métriques (section 6.3).

6.1 ÉVALUATION HUMAINE

6.1.1 ANNOTATION D'ERREURS

DEEPL

Dans l'annotation d'erreurs, DeepL a obtenu un score de pénalité de 257.03. Ce score est la somme de tous les scores de pénalité pour chaque segment. Sur les 100 segments au total, DeepL a réussi à traduire 29 segments sans erreurs. Les 71 autres segments contiennent 138 erreurs en total. Les nombres d'erreurs dans chaque catégorie ont été résumés dans le tableau 2 :

	Mineur	Majeur	Critique	TOTAL
Addition	0	0	2	2
Fausse traduction	1	5	4	10
Omission	1	0	0	1
Non traduit	0	0	0	0
Abréviations	1	0	7	8
Convention locale	24	3	0	27
Style	29	2	0	31
Terminologie	34	13	12	59
TOTAL	90	23	25	138

Tableau 2 Nombre d'erreurs de DeepL

Dans la catégorie addition, DeepL a fait 2 erreurs qui ont été classées comme critique selon l'experte de droit : Pour la traduction des phrases sources « Il ne prend les mesures visées à l'al. 2, [...] » et « S'il prend des mesures au sens de l'al. 1, [...] », DeepL ajout le chiffre 1 : « Er

trufft die Massnamen nach Absatz 1. 2 [...] » et « Trifft er Massnamen im Sinne von Absatz 1. 1 [...] ». À cause de ces deux additions, les articles mentionnés dans la traduction ne correspondent pas aux articles mentionnés dans les phrases sources.

Dans la catégorie fausse traduction, DeepL a fait 10 erreurs en total, dont 1 mineure, 5 majeures et 4 critiques, qui sont illustrées dans le tableau 3 :

Fausse traduction	Phrase source	Traduction de DeepL	Loi COVID-19 en allemand
Mineur	Si la population adulte souhaitant être vaccinée a reçu une dose suffisante de vaccin, les restrictions de capacité applicables aux établissements et aux entreprises accessibles au public ainsi qu'aux manifestations et aux rassemblements privés doivent être levées.	Wenn die erwachsene Bevölkerung, die geimpft werden möchte, eine ausreichende Dosis des Impfstoffs erhalten hat, müssen die Kapazitätsbeschränkungen für öffentlich zugängliche Einrichtungen und Unternehmen sowie für private Veranstaltungen und Versammlungen aufgehoben werden.	Ist der impfwillige erwachsene Teil der Bevölkerung ausreichend geimpft, so sind die Kapazitätsbeschränkungen für öffentlich zugängliche Einrichtungen und Betriebe sowie Veranstaltungen und private Zusammenkünfte aufzuheben.
Majeur	Loi fédérale sur les bases légales des ordonnances du Conseil fédéral visant à surmonter l'épidémie de COVID-19	Bundesgesetz über die Rechtsgrundlagen für Verordnungen des Bundesrates zur Überwindung der COVID-19-Epidemie	Bundesgesetz über die gesetzlichen Grundlagen für Verordnungen des Bundesrates zur Bewältigung der Covid-19-Epidemie
Majeur	Art. 1 Objet et principes	Art. 1 Zweck und Grundsätze	Art. 1 Gegenstand und Grundsätze
Majeur	toute autre analyse, si ceci est nécessaire en vue de garantir les capacités de test et de laboratoire requises pour lutter contre l'épidémie de COVID-19.	jede andere Analyse, wenn dies notwendig ist, um die zur Bekämpfung des COVID-19- Ausbruchs erforderlichen Test- und Laborkapazitäten zu gewährleisten.	anderen Analysen, wenn dies zur Sicherstellung der zur Bekämpfung der Covid-19- Epidemie erforderlichen Test- und Laborkapazitäten notwendig ist.
Majeur	prendre d'autres mesures nécessaires au maintien des capacités .	andere Maßnahmen ergreifen, die zur Aufrechterhaltung der Fähigkeiten erforderlich sind.	weitere zur Sicherstellung der Kapazitäten erforderliche Massnahmen zu treffen.

Majeur et Critique	Le Conseil fédéral peut obliger les fabricants, les distributeurs , les laboratoires, les établissements de santé et d'autres établissements des cantons à communiquer leurs stocks de produits thérapeutiques, d'équipements de protection et d'autres biens médicaux importants pour le maintien des capacités sanitaires (biens médicaux importants).	Der Bundesrat kann Hersteller, Händler , Laboratorien, Gesundheitseinrichtungen und andere Einrichtungen der Kantone verpflichten, ihre Bestände an Heilmitteln, Schutzausrüstungen und anderen medizinischen Gütern, die für die Aufrechterhaltung der gesundheitlichen Leistungsfähigkeit wichtig sind (wichtige medizinische Güter), zu melden.	Der Bundesrat kann Hersteller, Vertreiber , Laboratorien sowie Gesundheitseinrichtungen und weitere Einrichtungen der Kantone verpflichten, ihren Bestand an Heilmitteln, Schutzausrüstungen und weiteren für die Gesundheitsversorgung wichtigen medizinischen Gütern (wichtige medizinische Güter) zu melden.
Critique	Il informe régulièrement le Parlement, en temps utile et de manière exhaustive, de la mise en oeuvre de la présente loi.	Es informe le Parlement régulièrement, rechtzeitig und umfassend über die Umsetzung dieses Gesetzes.	Er informiert das Parlament regelmässig, frühzeitig und umfassend über die Umsetzung dieses Gesetzes.
Critique	Dans le cadre de sa stratégie, il veille à ce que la vie économique et sociale soit restreinte le moins possible et le moins longtemps possible ; pour ce faire, la Confédération et les cantons devront tout d'abord exploiter toutes les possibilités offertes par les plans de protection, par les stratégies de dépistage et de vaccination et par le traçage des contacts.	Im Rahmen seiner Strategie sorgt er dafür, dass das wirtschaftliche und soziale Leben so wenig und so lange wie möglich eingeschränkt wird; dazu müssen der Bund und die Kantone zunächst alle Möglichkeiten ausschöpfen, die sich aus den Schutzplänen, den Screening- und Impfstrategien und der Rückverfolgung von Kontakten ergeben.	Er richtet seine Strategie auf die mildest- und kürzestmögliche Einschränkung des wirtschaftlichen und gesellschaftlichen Lebens aus, indem Bund und Kantone zuvor sämtliche Möglichkeiten von Schutzkonzepten, von Test- und Impfstrategien sowie des Contact-Tracing ausschöpfen.
Critique	Il garantit que, malgré la fermeture des établissements de restauration ordonnée par les autorités, suffisamment d'installations sanitaires sont à la disposition des conducteurs de camion et que ceux-ci peuvent se restaurer dans des établissements de restauration.	Es stellt sicher, dass trotz behördlich angeordneter Schließung von Gastronomiebetrieben genügend sanitäre Einrichtungen zur Verfügung stehen und Lkw-Fahrer in Gastronomiebetrieben essen können.	Der Bundesrat stellt sicher, dass den LKW-Fahrerinnen und LKW-Fahrern trotz der behördlichen Schliessung von Gastrobetrieben genügend sanitäre Einrichtungen zur Verfügung stehen und dass die LKW-Fahrerinnen und LKW-Fahrer sich in Gastrobetrieben verpflegen können.

Tableau 3 Erreurs de DeepL dans la catégorie fausse traduction

L'erreur mineure concerne la traduction de « manifestations et aux rassemblements privés », qui a été traduit par DeepL comme « private Veranstaltungen und Versammlungen ». Dans la phrase source, l'adjectif « privés » modifie seulement le mot « rassemblements », mais dans la traduction de DeepL, l'adjectif fait référence aux deux mots « manifestations » et « rassemblements ».

Les 5 erreurs majeures concernent la traduction des mots « bases légales » (traduction : Rechtsgrundlagen), « Objet » (traduction : Zweck), « distributeurs » (traduction : Händler), « capacités » (traduction : Fähigkeiten), et « l'épidémie de COVID-19 » (traduction : COVID-19-Ausbruch). La traduction de ces mots par DeepL ne correspond pas à cent pour cent aux idées des phrases sources.

Les 4 erreurs critiques concernent la traduction des mots « capacités sanitaires » (traduction : gesundheitliche Leistungsfähigkeit), deux fois « Il » (traduction : Es), « FF » (traduction : BBL), et « le moins longtemps possible » (traduction : so lange wie möglich). DeepL n'a pas réussi de traduire correctement ces exemples : « gesundheitliche Leistungsfähigkeiten » fait référence aux humains et non pas aux institutions ; « Il » fait référence à la Confédération et devrait être traduit comme « Er » ou « Der Bundesrat » ; et « so lange wie möglich » est le contraire de « le moins longtemps possible », donc « le plus longtemps possible » ;

Dans la catégorie omission, DeepL a fait une erreur mineure. Cette erreur concerne l'omission de deux points, comme illustré dans le tableau 4 :

Omission	Phrase source	Traduction de DeepL	Loi COVID-19 en allemand
Mineur	La Confédération assure, en collaboration avec les cantons, l'existence d'un système de traçage des contacts (système TTIQ) qui fonctionne dans toute la Suisse. À cette fin, elle peut notamment :	Der Bund sorgt in Zusammenarbeit mit den Kantonen für ein schweizweit funktionierendes System zur Rückverfolgung von Kontakten (TTIQ-System). Zu diesem Zweck kann er insbesondere	Der Bund stellt in Zusammenarbeit mit den Kantonen ein schweizweit funktionierendes Test- und Contact-Tracing-System (TTIQ-System ²¹) sicher. Er kann zu diesem Zweck insbesondere:

Tableau 4 Erreurs de DeepL dans la catégorie omission

Dans la catégorie « non traduit », DeepL n'a pas fait des erreurs.

Dans la catégorie abrégations, DeepL a fait 8 erreurs en total, dont 1 mineure et 7 critiques, qui sont présentées dans le tableau 5 :

Abréviation	Phrase source	Traduction de DeepL	Loi COVID-19 en allemand
Mineur	CNA	SUVA	Suva
Critique	FF 2020 6363	BBL 2020 6363	BBI 2020 6563
Critique	LEI	AuIG	AIG
Critique	LEI	AuG	AIG
Critique	LEI	LEI	AIG

Tableau 5 Erreurs de DeepL dans la catégorie abrégations

L'erreur mineure concerne la traduction de « CNA ». Bien que DeepL ait utilisé l'abréviation correcte en allemand, il a mis des majuscules au lieu de minuscules.

Les erreurs critiques concernent la traduction de « FF » et « LEI ». L'abréviation « FF 2020 6363 » fait référence au « Message concernant la loi fédérale sur les bases légales des ordonnances du Conseil fédéral visant à surmonter l'épidémie de COVID-19 (loi COVID-19) ». Cette abréviation française correspond à l'abréviation allemande « BBI 2020 6363 » qui fait référence à « Botschaft zum Bundesgesetz über die gesetzlichen Grundlagen für Verordnungen des Bundesrates zur Bewältigung der Covid-19-Epidemie (Covid-19-Gesetz) ». DeepL n'a pas réussi à traduire cette abréviation correctement : il l'a traduit comme « BBL » qui fait référence à l'Office fédéral des constructions et de la logistique (OFCL), en allemand : « Bundesamt für Bauten und Logistik (BBL) ».

L'abréviation française « LEI », qui correspond à l'abréviation allemande « AIG », a été traduite par DeepL de trois manières différentes : une fois comme « AuIG », une fois comme « AuG » et quatre fois comme « LEI ».

Dans la catégorie « convention locale », DeepL a fait 27 erreurs en total, dont 24 mineures et 3 majeures. Les 24 erreurs mineures concernent l'utilisation de « ß » (qui est utilisé en Allemagne) au lieu de « ss » (qui est utilisé en Suisse).

Les 3 erreurs majeures concernent le mot « Ausschüsse », qui est plutôt utilisé en Allemagne (« Kommissionen » en Suisse) et qui de ce fait a été jugé comme une erreur majeure.

Dans la catégorie « style », DeepL a fait 31 erreurs en total, dont 29 mineures et 2 majeures. Le tableau 6 montre quelques exemples de ce type d'erreurs :

Style	Phrase source	Traduction de DeepL	Loi COVID-19 en allemand
Mineur	Loi fédérale sur les bases légales des ordonnances du Conseil fédéral visant à surmonter l'épidémie de COVID-19	Bundesgesetz über die Rechtsgrundlagen für Verordnungen des Bundesrates zur Überwindung der COVID-19-Epidemie	Bundesgesetz über die gesetzlichen Grundlagen für Verordnungen des Bundesrates zur Bewältigung der Covid-19-Epidemie
Mineur	Dans le cadre de sa stratégie, il veille à ce que la vie économique et sociale soit restreinte le moins possible et le moins longtemps possible; pour ce faire, la Confédération et les cantons devront tout d'abord exploiter toutes les possibilités offertes par les plans de protection , par les stratégies de dépistage et de vaccination et par le traçage des contacts.	Im Rahmen seiner Strategie sorgt er dafür, dass das wirtschaftliche und soziale Leben so wenig und so lange wie möglich eingeschränkt wird; dazu müssen der Bund und die Kantone zunächst alle Möglichkeiten ausschöpfen, die sich aus den Schutzplänen , den Screening- und Impfstrategien und der Rückverfolgung von Kontakten ergeben.	Er richtet seine Strategie auf die mildest- und kürzestmögliche Einschränkung des wirtschaftlichen und gesellschaftlichen Lebens aus, indem Bund und Kantone zuvor sämtliche Möglichkeiten von Schutzkonzepten , von Test- und Impfstrategien sowie des Contact-Tracing ausschöpfen.
Mineur	Il associe les gouvernements cantonaux et les associations faitières des partenaires sociaux à l'élaboration des mesures qui touchent leurs compétences.	Er bezieht die Kantonsregierungen und die Dachverbände der Sozialpartner in die Ausarbeitung von Maßnahmen ein, die ihre Zuständigkeiten betreffen .	Er bezieht die Kantonsregierungen und die Dachverbände der Sozialpartner bei der Erarbeitung von Massnahmen ein, die ihre Zuständigkeit betreffen .
Mineur	Il peut, pour garantir un approvisionnement suffisant de la population en biens médicaux importants:	Er kann, um eine ausreichende Versorgung der Bevölkerung mit wichtigen medizinischen Gütern zu gewährleisten:	Er kann zur Gewährleistung einer ausreichenden Versorgung der Bevölkerung mit wichtigen medizinischen Gütern:
Mineures et Majeur	Afin de renforcer les services de santé sollicités par la crise COVID-19, les cantons financent les réserves de capacités nécessaires pour affronter les pics d'activité .	Um die durch die COVID-19-Krise beanspruchten Gesundheitsdienste zu stärken, finanzieren die Kantone die Kapazitätsreserven , die zur Bewältigung von	Zur Stärkung der durch die Covid-19-Krise beanspruchten Gesundheitsversorgung finanzieren die Kantone die zur Abdeckung von Auslastungsspitzen nötigen Vorhaltsleistungen .

		Spitzenbelastungen erforderlich sind.	
Mineur et Majeur	Afin de promouvoir l'exercice des droits politiques, le Conseil fédéral peut prévoir que les demandes de référendum ou d'initiative populaire munies du nombre de signatures requis doivent être déposées auprès de la Chancellerie fédérale avant l'expiration du délai applicable aux référendums et aux initiatives populaires, qu'elles soient munies ou non des attestations de la qualité d'électeur.	Um die Ausübung der politischen Rechte zu fördern, kann der Bundesrat vorsehen, dass Referendums- und Volksinitiativbegehren, die mit der erforderlichen Anzahl Unterschriften versehen sind , vor Ablauf der für Referenden und Volksinitiativen geltenden Frist bei der Bundeskanzlei eingereicht werden müssen, unabhängig davon, ob sie mit Stimmrechtsbescheinigungen versehen sind oder nicht .	Der Bundesrat kann zur Unterstützung der Ausübung der politischen Rechte vorsehen, dass Referendums- und Initiativbegehren mit der nötigen Anzahl Unterschriften, jedoch auch ohne Stimmrechtsbescheinigung innerhalb der Referendums- und Initiativfrist bei der Bundeskanzlei einzureichen sind.

Tableau 6 Erreurs de DeepL dans la catégorie style

Les erreurs de cette catégorie sont souvent dû à la traduction mot-à-mot ou à la reprise de la syntaxe de la phrase source, ce qui rend la traduction en allemand moins naturelle. Même si les traductions de DeepL sont compréhensibles, parfois elles ne semblent pas être très naturelles ou appropriées dans le contexte juridique.

Dans la catégorie « terminologie », DeepL a fait 59 erreurs en total, dont 34 mineures, 13 majeures et 12 critiques. Le tableau 7 montre quelques exemples de ce type d'erreurs :

Terminologie	Phrase source	Traduction de DeepL	Loi COVID-19 en allemand
Mineur	(Loi COVID-19)	(COVID-19 -Gesetz)	(Covid-19 -Gesetz)
Mineur	Art. 1a Critères et valeurs de référence	Art. 1a Kriterien und Referenzwerte	Art. 1a Kriterien und Richtwerte
Mineur	mettre en place un traçage électronique des contacts qui soit complet et efficace; les données du traçage devront être anonymisées ou supprimées après avoir été analysées, mais au plus tard	eine umfassende und wirksame elektronische Rückverfolgung der Kontakte einrichten; die Rückverfolgungsdaten müssen anonymisiert oder nach der Analyse gelöscht werden, spätestens	umfassendes, wirksames und digitales Contact-Tracing ; die Daten des Contact-Tracing sind nach Abschluss der Datenauswertung, spätestens aber zwei Jahre nach

	deux ans après avoir été prélevées;	jedoch zwei Jahre nach der Erhebung.	ihrer Erhebung zu anonymisieren oder zu löschen;
Mineur	TTIQ = Tests, traçage, isolement et quarantaine	TTIQ = Tests, Rückverfolgung, Isolierung und Quarantäne	TTIQ = Testen, Tracing, Isolation, Quarantäne
Majeur	Art. 6a Certificat sanitaire	Art. 6a Gesundheitszeugnis	Art. 6a Impf-, Test- und Genesungsnachweise
Mineur, Majeur et Critique	obliger les cantons à améliorer, dans le cadre du traçage des contacts , la situation relative aux données concernant les foyers épidémiques et les sources d'infection présumés et les dédommager pour les dépenses en découlant;	die Kantone verpflichten, im Rahmen der Rückverfolgung von Kontakten die Datenlage zu mutmaßlichen Ausbrüchen und Infektionsquellen zu verbessern und sie für die damit verbundenen Ausgaben zu entschädigen;	die Kantone verpflichten, im Contact-Tracing die Datenlage bezüglich vermuteter Cluster und Infektionsquellen zu verbessern (Rückverfolgung) und die Kantone für die entsprechenden Aufwände entschädigen;
Critique	FF 2020 6363	BBL 2020 6363	BBI 2020 6363
Critique	le regroupement familial (art. 47 LEI),	die Familienzusammenführung (Art. 47 LEI),	den Familiennachzug (Art. 47 AIG),

Tableau 7 Erreurs de DeepL dans la catégorie terminologie

12 de 34 erreurs mineures concernent la traduction du mot « COVID-19 », qui a été reprise par DeepL dans la traduction allemande, mais dans la version officielle allemande de la loi COVID-19, on peut observer que les majuscules manquent : « Covid-19 ». La traduction de termes comme « valeurs de référence » et « traçage des contacts » a également été jugée comme des erreurs mineures, vu que les termes traduits par DeepL ne correspondent pas aux termes corrects utilisés en allemand, mais on peut quand-même comprendre de quoi il s'agit. De même, la traduction de « TTIQ = Tests, traçage, isolement et quarantaine » a également été jugée comme une erreur mineure, parce que la traduction de DeepL « TTIQ = Tests, Rückverfolgung, Isolierung und Quarantäne » montre que l'abréviation ne correspond pas aux lettres initiales des mots traduits, qui donneraient l'abréviation « TRIQ ».

Les erreurs majeures concernent la traduction de termes, comme p.ex. « certificat sanitaire » ou « dépenses », qui ont été traduits par DeepL sans prendre en compte le contexte du texte.

Les erreurs critiques, par contre, concernent la traduction des abréviations, qui font partie de la terminologie juridique. Si les abréviations ne sont pas traduites correctement, il est impossible de p.ex. trouver la même loi ou la même institution, etc. De même, les termes comme « foyers

épidémiques » et « regroupement familial », ont été mal traduits par DeepL, vu que ces traductions font référence à des autres concepts.

ETRANSLATION

Dans l’annotation d’erreurs, eTranslation a obtenu un score de pénalité de 770.25. Ce score est la somme de tous les scores de pénalité pour chaque segment. Sur 100 segments en total, eTranslation a réussi à traduire 14 segments sans erreurs. Les 86 autres segments contiennent 247 erreurs en total. Le nombre d’erreurs dans chaque catégorie a été résumé dans le tableau 8 :

	Mineur	Majeur	Critique	TOTAL
Addition	4	0	0	4
Fausse traduction	5	7	19	31
Omission	4	4	4	12
Non traduit	1	0	1	2
Abréviations	1	0	17	18
Convention locale	33	3	3	39
Style	34	9	1	44
Terminologie	30	28	39	97
TOTAL	112	51	84	247

Tableau 8 Nombre d’erreurs de eTranslation

Dans la catégorie addition, eTranslation a fait 4 erreurs mineures, qui sont présentées dans le tableau 9 :

Addition	Phrase source	Traduction de DeepL	Loi COVID-19 en allemand
Mineur	prévoir des dérogations aux dispositions sur l’évaluation de la conformité des dispositifs médicaux et aux dispositions sur la procédure d’évaluation et la mise sur le marché d’équipements de protection;	Ausnahmen von den Bestimmungen über die Konformitätsbewertung von Medizinprodukten und den Bestimmungen über das Evakuierungsverfahren und das Inverkehrbringen von Schutzausrüstungen vorzusehen;	Ausnahmen von den Bestimmungen über die Konformitätsbewertung von Medizinprodukten sowie von den Bestimmungen über das Konformitätsbewertungsverfahren und das Inverkehrbringen von Schutzausrüstungen vorsehen ;

Mineur	ordonner la confiscation de biens médicaux importants, contre indemnisation;	die Einziehung wichtiger medizinischer Vermögenswerte gegen Entschädigung anzuordnen ;	die Einziehung von wichtigen medizinischen Gütern gegen Entschädigung anordnen ;
Mineur et mineur	obliger les cantons à améliorer, dans le cadre du traçage des contacts, la situation relative aux données concernant les foyers épidémiques et les sources d'infection présumés et les dédommager pour les dépenses en décaissant;	die Kantone zu verpflichten, im Rahmen der Kontaktnachverfolgung die Situation in Bezug auf die Daten zu mutmaßlichen Seuchenausbrüchen und Infektionsquellen zu verbessern und sie für die damit verbundenen Ausgaben zu entschädigen;	die Kantone verpflichten , im Contact-Tracing die Datenlage bezüglich vermuteter Cluster und Infektionsquellen zu verbessern (Rückverfolgung) und die Kantone für die entsprechenden Aufwände entschädigen ;

Tableau 9 Erreurs de eTranslation dans la catégorie addition

Les quatre erreurs mineures concernent l'addition du mot « zu » à la forme infinitive du verbe en allemand. Si les verbes dans les phrases sources étaient placés après le mot « pour », l'addition du mot « zu » en allemand serait correcte, mais vu que dans les phrases sources seulement la forme infinitive des verbes est utilisée, cette addition n'est pas correcte.

Dans la catégorie fausse traduction, eTranslation a fait 31 erreurs, dont 5 mineures, 7 majeures et 19 critiques. Quelques exemples de ce type d'erreurs sont présentés dans le tableau 10 :

Fausse traduction	Phrase source	Traduction de DeepL	Loi COVID-19 en allemand
Majeur et Mineur	Loi fédérale sur les bases légales des ordonnances du Conseil fédéral visant à surmonter l'épidémie de COVID-19	Bundesgesetz über die Rechtsgrundlagen der Bundesratsverordnungen zur Überwindung des COVID-19- Ausbruchs	Bundesgesetz über die gesetzlichen Grundlagen für Verordnungen des Bundesrates zur Bewältigung der Covid-19-Epidemie
Majeur et Majeur	Le Conseil fédéral peut obliger les fabricants, les distributeurs , les laboratoires, les établissements de santé et d'autres établissements des cantons à communiquer leurs stocks de produits thérapeutiques, d'équipements de protection et d'autres biens médicaux importants pour le maintien des capacités sanitaires	Der Bundesrat kann Hersteller, Händler , Labors, Gesundheitseinrichtungen und andere kantonale Einrichtungen verpflichten, ihre Bestände an Heilmitteln, Schutzausrüstungen und anderen medizinischen Gütern, die für die Erhaltung der Gesundheitskapazitäten von Bedeutung sind	Der Bundesrat kann Hersteller, Vertreiber , Laboratorien sowie Gesundheitseinrichtungen und weitere Einrichtungen der Kantone verpflichten, ihren Bestand an Heilmitteln, Schutzausrüstungen und weiteren für die Gesundheitsversorgung wichtigen medizinischen Gütern (wichtige medizinische Güter) zu melden.

	(biens médicaux importants).	(wichtige medizinische Güter), zu melden.	
Majeur et Critique	prévoir des dérogations au régime de l'autorisation pour des activités en relation avec des biens médicaux importants ou adapter les conditions d'autorisation ;	Ausnahmen von der Genehmigungsregelung für Tätigkeiten im Zusammenhang mit wichtigen medizinischen Gütern vorsehen oder die Bedingungen für die Selbstbestimmung anpassen;	Ausnahmen von der Bewilligungspflicht für Tätigkeiten im Zusammenhang mit wichtigen medizinischen Gütern vorsehen oder die Bewilligungsvoraussetzungen anpassen;
Critique	du 25 septembre 2020 (Etat le 1 ^{er} janvier 2022)	vom 25. September 2020 (Staat am 1. Januar 2022)	vom 25. September 2020 (Stand am 1. Januar 2022)
Critique	Dans le cadre de sa stratégie, il veille à ce que la vie économique et sociale soit restreinte le moins possible et le moins longtemps possible ; pour ce faire, la Confédération et les cantons devront tout d'abord exploiter toutes les possibilités offertes par les plans de protection, par les stratégies de dépistage et de vaccination et par le traçage des contacts.	Im Rahmen seiner Strategie dafür Sorge zu tragen, dass das wirtschaftliche und soziale Leben so wenig wie möglich und so lange wie möglich bleibt; dazu müssen Bund und Kantone zunächst alle Möglichkeiten nutzen, die sich aus den Schutzplänen, den Screening- und Impfstrategien und der Kontaktnachverfolgung ergeben.	Er richtet seine Strategie auf die mildest- und kürzestmögliche Einschränkung des wirtschaftlichen und gesellschaftlichen Lebens aus, indem Bund und Kantone zuvor sämtliche Möglichkeiten von Schutzkonzepten, von Test- und Impfstrategien sowie des Contact-Tracing ausschöpfen.
Critique	Si la population adulte souhaitant être vaccinée a reçu une dose suffisante de vaccin, les restrictions de capacité applicables aux établissements et aux entreprises accessibles au public ainsi qu'aux manifestations et aux rassemblements privés doivent être levées.	Wenn die erwachsene Bevölkerung, die geimpft werden möchte, eine ausreichende Impfstoffdosis erhalten hat, sollten die Kapazitätsbeschränkungen für öffentlich zugängliche Einrichtungen und Unternehmen sowie für Demonstrationen und private Versammlungen aufgehoben werden.	Ist der impfwillige erwachsene Teil der Bevölkerung ausreichend geimpft, so sind die Kapazitätsbeschränkungen für öffentlich zugängliche Einrichtungen und Betriebe sowie Veranstaltungen und private Zusammenkünfte aufzuheben.
Critique	prévoir des dérogations aux dispositions sur l'évaluation de la conformité des dispositifs médicaux et aux dispositions sur la procédure d'éva-	Ausnahmen von den Bestimmungen über die Konformitätsbewertung von Medizinprodukten und den Bestimmungen über das	Ausnahmen von den Bestimmungen über die Konformitätsbewertung von Medizinprodukten sowie von den Bestimmungen über das

	luation et la mise sur le marché d'équipements de protection;	Evakuierungsverfahren und das Inverkehrbringen von Schutzausrüstungen vorzusehen;	Konformitätsbewertungsverfahren und das Inverkehrbringen von Schutzausrüstungen vorzusehen;
Critique	ordonner la confiscation de biens médicaux importants, contre indemnisation;	die Einziehung wichtiger medizinischer Vermögenswerte gegen Entschädigung anzuordnen;	die Einziehung von wichtigen medizinischen Gütern gegen Entschädigung anordnen;
Critique	permettre d'assouplir, de raccourcir ou d'abolir progressivement l'obligation de quarantaine si des mesures de rechange telles que la vaccination ou les tests réguliers peuvent garantir une réduction comparable de la propagation du virus.	Ermöglichung der Lockerung, Verschärfung oder Abschaffung der Vierkantenblüte , wenn durch alternative Maßnahmen wie die Impfung oder regelmäßige Tests eine vergleichbare Verringerung der Ausbreitung des Virus gewährleistet werden kann.	Möglichkeiten, die Quarantäne schrittweise zu lockern, zu verkürzen oder aufzuheben, wenn durch Alternativen wie Impfung, regelmässige Tests oder andere Massnahmen eine vergleichbare Reduktion der Verbreitung gesichert werden kann.
Critique	La Confédération peut mettre un système pour la délivrance du document à la disposition des cantons et de tiers.	Der Bund kann Entanten und Dritten ein System für die Ausstellung des Dokuments zur Verfügung stellen.	Der Bund kann den Kantonen sowie Dritten ein System für die Erteilung von Nachweisen zur Verfügung stellen.

Tableau 10 Erreurs de eTranslation dans la catégorie fausse traduction

Les erreurs mineures concernent la traduction de mots, comme p.ex. « l'épidémie » qui a été traduit par eTranslation comme « Ausbruch ». Même si ce mot a été mal traduit par eTranslation, on peut quand-même le comprendre avec le contexte.

Les erreurs majeures concernent les mots, dont la traduction n'est pas appropriée pour ce contexte juridique, comme p.ex. la traduction de « distributeurs » qui correspond au mot allemand « Vertreiber », mais qui a été traduit par eTranslation comme « Händler ».

Les erreurs critiques, par contre, concernent toutes les traductions qui sont totalement mal traduites et qui, par conséquence, posent des problèmes de sens, c'est-à-dire que soit l'idée contraire de la phrase source est exprimée, soit des mots non existants sont utilisés : p.ex. « le moins longtemps possible » a été traduit par eTranslation comme « so lange wie möglich » ; « État » a été traduit comme « Staat » au lieu de « Stand » ; « cantons » a été traduit comme « Entanten », ce qui n'existe pas en allemand, ainsi que « Vierkantenblüte », la traduction de eTranslation pour le mot « quarantaine ».

Dans la catégorie omission, eTranslation a fait 12 erreurs, dont 4 mineures, 4 majeures et 4 critiques, qui sont présentées dans le tableau 11 :

Omission	Phrase source	Traduction de DeepL	Loi COVID-19 en allemand
Mineur	Il informe régulièrement le Parlement, en temps utile et de manière exhaustive, de la mise en oeuvre de la présente loi.	Er unterrichtet das Parlament regelmäßig rechtzeitig und umfassend über die Durchführung dieses Gesetzes.	Er informiert das Parlament regelmässig, frühzeitig und umfassend über die Umsetzung dieses Gesetzes.
Mineur	Il peut, pour garantir un approvisionnement suffisant de la population en biens médicaux importants:	Um eine ausreichende Versorgung der Bevölkerung mit wichtigen medizinischen Gütern zu gewährleisten, kann er	Er kann zur Gewährleistung einer ausreichenden Versorgung der Bevölkerung mit wichtigen medizinischen Gütern:
Majeur	S'il prend des mesures au sens de l'al. 1, il prévoit que leur exécution relève des organes d'exécution de la loi du 13 mars 1964 sur le travail ²³ et de la Caisse nationale suisse d'assurance en cas d'accidents (CNA), et que les frais résultant de cette exécution sont financés par le supplément de prime destiné aux frais liés à la prévention des accidents et maladies professionnels prévu à l'art. 87 de la loi fédérale du 20 mars 1981 sur l'assurance-accidents ²⁴ .	Wenn er Maßnahmen im Sinne von Absatz 1 trifft, sieht er vor, dass deren Durchführung in den Zuständigkeitsbereich der Vollzugsbehörden des Arbeitsgesetzes vom 13. März 1964 und der Schweizerischen Unfallversicherungskasse (CNA) fällt und dass die Kosten, die sich aus dieser Ausführung ergeben, aus dem Zuschlag für die Kosten der Verhütung von Berufsunfällen und Berufskrankheiten gemäß Artikel 87 des Bundesgesetzes vom 20. März 1981 über die Unfallversicherung finanziert werden.	Ergreift er Massnahmen nach Absatz 1, so sieht er vor, dass der Vollzug den Durchführungsorganen des Arbeitsgesetzes vom 13. März 1964 sowie der Schweizerischen Unfallversicherungsanstalt (Suva) obliegt und dass die dafür anfallenden Vollzugskosten aus dem Prämienzuschlag für die Verhütung von Berufsunfällen und Berufskrankheiten nach Artikel 87 des Bundesgesetzes vom 20. März 1981 über die Unfallversicherung finanziert werden.
Majeur	Le Conseil fédéral peut édicter des dispositions dérogeant à la loi fédérale du 16 décembre 2005 sur les étrangers et l'intégration (LEI) ²⁹ et à la loi du 26 juin 1998 sur l'asile (LAsi):	Der Bundesrat kann vom Bundesgesetz vom 16. Dezember 2005 über Ausländer und Integration (LEI) und vom Asylgesetz vom 26. Juni 1998 (LAsi) abweichen :	Der Bundesrat kann vom Ausländer- und Integrationsgesetz vom 16. Dezember 2005 ²⁹ (AIG) und vom Asylgesetz vom 26. Juni 1998 (AsylG) abweichende Bestimmungen erlassen über:

Critique	Dans le cadre de sa stratégie, il veille à ce que la vie économique et sociale soit restreinte le moins possible et le moins longtemps possible; pour ce faire, la Confédération et les cantons devront tout d'abord exploiter toutes les possibilités offertes par les plans de protection, par les stratégies de dépistage et de vaccination et par le traçage des contacts.	Im Rahmen seiner Strategie dafür Sorge zu tragen, dass das wirtschaftliche und soziale Leben so wenig wie möglich und so lange wie möglich bleibt ; dazu müssen Bund und Kantone zunächst alle Möglichkeiten nutzen, die sich aus den Schutzplänen, den Screening- und Impfstrategien und der Kontaktnachverfolgung ergeben.	Er richtet seine Strategie auf die mildest- und kürzestmögliche Einschränkung des wirtschaftlichen und gesellschaftlichen Lebens aus, indem Bund und Kantone zuvor sämtliche Möglichkeiten von Schutzkonzepten, von Test- und Impfstrategien sowie des Contact-Tracing ausschöpfen.
Critique	Art. 5 ²⁸ Mesures dans le domaine des étrangers et de l'asile	Art. 5 Ausländer- und Asylmaßnahmen	Art. 5 ²⁸ Massnahmen im Ausländer- und Asyl bereich
Critique	Il peut soutenir des mesures prises par les cantons afin de simplifier l'entrée dans la vie professionnelle, rendue difficile par la crise du coronavirus, des jeunes qui terminent leur formation scolaire.	Er kann Maßnahmen der Kantone unterstützen, um den Eintritt in das Berufsleben zu vereinfachen, der durch die Coronavirus-Krise erschwert wurde und die ihre Schulbildung beendet .	Der Bundesrat kann Massnahmen der Kantone fördern, die darauf abzielen, Schulabgängerinnen und Schulabgängern den Berufseinstieg, der durch die Covid-19-Epidemie erschwert ist, zu erleichtern.

Tableau 11 Erreurs de eTranslation dans la catégorie omission

Les 4 erreurs mineures concernent l'omission des signes de ponctuation. Les 4 erreurs majeures concernent l'omission de la traduction de quelques mots : « le supplément de prime » a été traduit par eTranslation comme « Zuschlag », donc « de prime » a été omis. De la même manière, dans l'exemple « édicter des dispositions dérogeant à la loi » a été traduit par eTranslation comme « vom Gesetz abweichen », donc « déroger à la loi ».

Les 4 erreurs critiques concernent l'omission des mots qui sont indispensables pour la compréhension des phrases.

Dans la catégorie non traduit, eTranslation a fait 2 erreurs, dont une mineure et une critique, qui sont présentées dans le tableau 12 :

Non traduit	Phrase source	Traduction de DeepL	Loi COVID-19 en allemand
-------------	---------------	---------------------	--------------------------

Mineur	vu les art. 68, al.1, 69, al. 2, 92, 93, 101, al. 2, 102, 113, 114, al. 1, 117, al. 1, 118, al. 2, let. b, 121, al. 1, 122, 123 et 133 de la Constitution (Cst.) ,	gestützt auf die Art. 68 Abs. 1, 69 Abs. 2, 92, 93, 101 Abs. 2, 102, 113, 114 Abs. 1, 117 Abs. 1, 118 Abs. 2, Buchst. b, 121 Abs. 1, 122, 123 und 133 der Verfassung (Verfassung) ,	gestützt auf die Artikel 68 Absatz 1, 69 Absatz 2, 92, 93, 101 Absatz 2, 102, 113, 114 Absatz 1, 117 Absatz 1, 118 Absatz 2 Buchstabe b, 121 Absatz 1, 122, 123 und 133 der Bundesverfassung (BV) ,
Critique	Introduit par le ch. I de la LF du 18 déc. 2020 (Culture, cas de rigueur, sport, assurance-chômage, amendes d'ordre) (RO 2020 5821); FF 2020 8505).	Eingeführt durch die ch. I der BG vom 18. Dezember 2020 (Kultur, Härtefälle, Sport, Arbeitslosenversicherung, Ordnungsgeld) (AS 2020 5821); FF 2020 8505).	Eingeführt durch Ziff. I des BG vom 18. Dez. 2020 (Kultur, Härtefälle, Sport, Arbeitslosenversicherung, Ordnungsbussen) (AS 2020 5821; BBI 2020 8819).

Tableau 12 Erreurs de eTranslation dans la catégorie non traduit

L'erreur mineure concerne l'abréviation de la Constitution « Cst. », qui n'a pas été traduite par eTranslation comme une abréviation, comme dans la version allemande de la loi COVID-19 (« BV »), mais comme « Verfassung ». Cette erreur a été jugée comme mineure, parce que l'abréviation dans ce cas n'est pas très importante, vu qu'elle suit le mot « la Constitution » ou « Verfassung » en allemand.

L'erreur critique concerne l'abréviation « ch. », qui doit être traduite en allemand comme « Ziff. », mais qui n'a pas été traduite par eTranslation. Cette erreur est critique, parce que le public suisse-allemand n'associe pas cette abréviation « ch. » avec l'abréviation allemand « Ziff. », mais plutôt avec l'abréviation « CH » qui représente la Suisse.

Dans la catégorie abréviations, eTranslation a fait 18 erreurs, dont une mineure et 17 critiques.

Quelques exemples de ce type d'erreurs sont présentés dans le tableau 13 :

Abréviations	Phrase source	Traduction de DeepL	Loi COVID-19 en allemand
Mineur	vu les art. 68, al.1, 69, al. 2, 92, 93, 101, al. 2, 102, 113, 114, al. 1, 117, al. 1, 118, al. 2, let. b, 121, al. 1, 122, 123 et 133 de la Constitution (Cst.) ,	gestützt auf die Art. 68 Abs. 1, 69 Abs. 2, 92, 93, 101 Abs. 2, 102, 113, 114 Abs. 1, 117 Abs. 1, 118 Abs. 2, Buchst. b, 121 Abs. 1, 122, 123 und 133 der Verfassung (Verfassung) ,	gestützt auf die Artikel 68 Absatz 1, 69 Absatz 2, 92, 93, 101 Absatz 2, 102, 113, 114 Absatz 1, 117 Absatz 1, 118 Absatz 2 Buchstabe b, 121 Absatz 1, 122, 123 und 133 der Bundesverfassung (BV) ,

Critique	RS 101	RS 101	SR 101
Critique	FF 2020 6363	FF 2020 6363	BBI 2020 6563
Critique	Nouvelle teneur selon le ch. I de la LF du 19 mars 2021 (Cas de rigueur, assurance-chômage, accueil extra-familial pour enfants, acteurs culturels, manifestations), en vigueur du 20 mars 2021 au 31 déc. 2031 (RO 2021 153, 878 ch. II al. 3; FF 2021 285, 2515).	Neuer Inhalt gemäß Ziff. I der BG vom 19. März 2021 (Stressfall, Arbeitslosenversicherung, außerfamiliäre Aufnahme für Kinder, kulturelle Akteure, Veranstaltungen), gültig vom 20. März 2021 bis zum 31. Dezember 2031 (AS 2021 153, 878 PS. II Absatz 3; FF 2021 285, 2515).	Fassung gemäss Ziff. I des BG vom 19. März 2021 (Härtefälle, Arbeitslosenversicherung, familienergänzende Kinderbetreuung, Kulturschaffende, Veranstaltungen), in Kraft vom 20. März 2021 bis zum 31. Dez. 2031 (AS 2021 153, 878 Ziff. II Abs. 3; BBI 2021 285, 2515).
Critique	CNA	CNA	Suva
Critique	LEI	LEI	AIG
Critique	LAsi	LAsi	AsylG

Tableau 13 Erreurs de eTranslation dans la catégorie abréviations

L'erreur mineure concerne la traduction de l'abréviation « Cst. », qui doit être traduite en allemand comme « BV », mais eTranslation n'a pas réussi à le traduire.

Les 17 erreurs critiques concernent les abréviations françaises qui ont été reprises par eTranslation dans sa traduction allemande, comme p.ex. « CNA », « LEI » ou « LAsi », respectivement « Suva », AIG » et « AsylG » en allemand.

Dans la catégorie convention locale, eTranslation a fait 39 erreurs, dont 33 mineures, 3 majeures et 3 critiques, qui sont présentées dans le tableau 14 :

Convention locale	Phrase source	Traduction de DeepL	Loi COVID-19 en allemand
Mineur	arrête:	beschließt:	beschliesst:
Majeur	Ceux-ci informent immédiatement leurs commissions respectives.	Diese setzen ihre jeweiligen Ausschüsse unverzüglich in Kenntnis.	Diese informieren umgehend ihre Kommissionen .
Majeur	Introduit par le ch. I de la LF du 18 déc. 2020 (Culture, cas de rigueur, sport, assurance-chômage, amendes d'ordre) (RO	Eingeführt durch die ch. I der BG vom 18. Dezember 2020 (Kultur, Härtefälle, Sport, Arbeitslosenversicherung,	Eingeführt durch Ziff. I des BG vom 18. Dez. 2020 (Kultur, Härtefälle, Sport, Arbeitslosenversicherung,

	2020 5821); FF 2020 8505).	Ordnungsgeld) (AS 2020 5821); FF 2020 8505).	Ordnungsbussen) (AS 2020 5821; BBI 2020 8819).
Critique	Au besoin, la Chancellerie fédérale transmet les listes de signatures au service compétent selon le droit cantonal pour attester la qualité d'électeur.	Erforderlichenfalls übermittelt das Bundeskanzleramt die Unterschriftenlisten an die nach kantonales Recht zuständige Stelle, um die Wählereigenschaft zu bescheinigen.	Die Bundeskanzlei stellt nötigenfalls die Unterschriftenlisten der Amtsstelle zu, die nach kantonalem Recht für die Stimmrechtsbescheinigung zuständig ist.
Critique et Critique	Afin de promouvoir l'exercice des droits politiques, le Conseil fédéral peut prévoir que les demandes de référendum ou d'initiative populaire munies du nombre de signatures requis doivent être déposées auprès de la Chancellerie fédérale avant l'expiration du délai applicable aux référendums et aux initiatives populaires , qu'elles soient munies ou non des attestations de la qualité d'électeur.	Um die Ausübung politischer Rechte zu fördern, kann der Bundesrat vorsehen, dass Anträge auf Volksabstimmung oder Volksinitiative , die mit der erforderlichen Anzahl von Unterschriften versehen sind, vor Ablauf der für Volksabstimmungen und Volksinitiativen geltenden Frist beim Bundeskanzleramt eingereicht werden müssen, unabhängig davon, ob sie mit Wahlbescheinigungen versehen sind oder nicht.	Der Bundesrat kann zur Unterstützung der Ausübung der politischen Rechte vorsehen, dass Referendums- und Initiativbegehren mit der nötigen Anzahl Unterschriften, jedoch auch ohne Stimmrechtsbescheinigung innerhalb der Referendums- und Initiativfrist bei der Bundeskanzlei einzureichen sind.

Tableau 14 Erreurs de eTranslation dans la catégorie convention locale

31 de 33 erreurs mineures concernent l'utilisation de « ß » (qui est utilisé en Allemagne), au lieu de « ss » (qui est utilisé en Suisse). Les 2 autres erreurs mineures concernent la traduction des mots « Constitution » (eTranslation : « Verfassung » ; Covid-19-Gesetz : « Bundesverfassung »), et « présidents des commissions » (eTranslation : « Vorsitzende der zuständigen Kommissionen » ; Covid-19-Gesetz : « Präsidenten der zuständigen Kommissionen »).

Les 3 erreurs majeures concernent la traduction du mot « commissions » qui a été traduit par eTranslation comme « Ausschüsse », ce qui est plutôt utilisé en Allemagne ; en Suisse le mot « Kommissionen » est utilisé. De la même manière, le mot « amendes d'ordre » a été traduit par eTranslation comme « Ordnungsgeld », mais ce mot est également plutôt utilisé en Allemagne ; en Suisse, le mot « Ordnungsbussen » est utilisé.

Les 3 erreurs critiques concernent la traduction de « Chancellerie fédérale » (eTranslation : « Bundeskanzleramt » ; Covid-19-Gesetz : « Bundeskanzlei »), et de « référendum ou initiative

populaire » (eTranslation : « Volksabstimmung oder Volksinitiative » ; Covid-19-Gesetz : « Referendum und Initiative »).

Dans la catégorie style, eTranslation a fait 44 erreurs, dont 34 mineures, 9 majeures et 1 critique.

Quelques exemples de ce type d'erreurs sont présentés dans le tableau 15 :

Style	Phrase source	Traduction de DeepL	Loi COVID-19 en allemand
Mineur	Le Conseil fédéral n'utilise de ces compétences que dans la mesure nécessaire pour surmonter l'épidémie de COVID-19.	Der Bundesrat nutzt diese Befugnisse nur in dem zur Überwindung des COVID-19-Ausbruchs erforderlichen Umfang .	Der Bundesrat macht von diesen Befugnissen nur so weit Gebrauch, als dies zur Bewältigung der Covid-19-Epidemie notwendig ist.
Mineur	vu le message du Conseil fédéral du 12 août 2020,	unter Hinweis auf die Botschaft des Bundesrates vom 12. August 2020,	nach Einsicht in die Botschaft des Bundesrates vom 12. August 2020,
Mineur, Mineur et Mineur	Il associe les gouvernements cantonaux et les associations faitières des partenaires sociaux à l'élaboration des mesures qui touchent leurs compétences.	Er bezieht die kantonalen Regierungen und die Dachverbände der Gesellschaftspartner in die Ausarbeitung von Maßnahmen ein, die ihre Zuständigkeiten berühren .	Er bezieht die Kantonsregierungen und die Dachverbände der Sozialpartner bei der Erarbeitung von Massnahmen ein, die ihre Zuständigkeit betreffen.
Majeur	interdire ou restreindre des examens et traitements médicaux non-urgents;	Verbot oder Einschränkung nicht dringlicher ärztlicher Untersuchungen und Behandlungen;	medizinisch nicht dringend angezeigte Untersuchungen und Behandlungen zu verbieten oder einzuschränken;
Majeur	Le Conseil fédéral garantit que les professionnels du secteur agricole et de la construction ainsi que les artisans et les ouvriers en déplacement professionnel ont la possibilité de se restaurer dans des établissements de restauration malgré la fermeture ordonnée par les autorités.	Der Bundesrat stellt sicher, dass Fachleute aus dem Agrar- und Bausektor sowie Handwerker und Wanderarbeiter die Möglichkeit haben, sich trotz der von den Behörden angeordneten Schließung in Restaurationsbetrieben zu ernähren .	Der Bundesrat stellt sicher, dass trotz der behördlichen Schliessung von Gastrobetrieben Berufsleute aus dem Landwirtschaftssektor und dem Bausektor sowie Handwerkerinnen und Handwerker und Berufstätige auf Montage die Möglichkeit haben, sich in Gastrobetrieben zu verpflegen.

Critique	Il doit être personnel, infalsifiable et, dans le respect de la protection des données, vérifiable; il doit être conçu de manière que seule une vérification décentralisée ou locale de son authenticité et de sa validité soit possible et qu'il puisse, dans la mesure du possible, être utilisé par son détenteur pour entrer dans d'autres pays et en sortir.	Er muss persönlich, fälschungssicher und unter Wahrung des Datenschutzes überprüfbar sein; er muss so gestaltet sein, dass nur eine dezentrale Prüfung seiner Echtheit und Gültigkeit möglich ist und dass er, soweit möglich, von seinem Besitzer verwendet werden kann, um in andere Länder einzureisen und aus diesem zu gelangen .	Der Nachweis muss persönlich, fälschungssicher, unter Einhaltung des Datenschutzes überprüfbar und so ausgestaltet sein, dass nur eine dezentrale oder lokale Überprüfung der Authentizität und Gültigkeit von Nachweisen möglich ist und er möglichst für die Ein- und Ausreise in andere Länder verwendet werden kann.
----------	---	---	--

Tableau 15 Erreurs de eTranslation dans la catégorie style

Comme les erreurs de DeepL dans la catégorie style, les erreurs de eTranslation dans cette catégorie sont également dû à la traduction mot-à-mot ou à la reprise de la syntaxe de la phrase source, ce qui rend la traduction en allemand moins naturelle. Même si les traductions de eTranslation sont compréhensibles, parfois elles ne semblent pas être très naturelles ou appropriées dans le contexte juridique.

Dans la catégorie terminologie, eTranslation a fait 97 erreurs en total, dont 30 mineures, 28 majeures et 39 critiques. Quelques exemples de ce type d'erreur sont présentés dans le tableau 16 :

Terminologie	Phrase source	Traduction de DeepL	Loi COVID-19 en allemand
Mineur	COVID-19	COVID-19	Covid-19
Mineur	Art. 1a ⁶ Critères et valeurs de référence	Art. 1a ⁶ Kriterien und Referenzwerte	Art. 1a ⁶ Kriterien und Richtwerte
Mineur, Majeur, Critique, Majeur, Critique et Critique	Nouvelle teneur selon le ch. I de la LF du 19 mars 2021 (Cas de rigueur, assurance-chômage, accueil extra-familial pour enfants, acteurs culturels, manifestations), en vigueur du 20 mars 2021 au 31 déc. 2031 (RO 2021 153, 878 ch. II al. 3; FF 2021 285, 2515).	Neuer Inhalt gemäß Ziff. I der BG vom 19. März 2021 (Stressfall , Arbeitslosenversicherung, außerfamiliäre Aufnahme für Kinder, kulturelle Akteure , Veranstaltungen), gültig vom 20. März 2021 bis zum 31. Dezember 2031 (AS 2021 153, 878 PS . II Absatz 3; FF 2021 285, 2515).	Fassung gemäss Ziff. I des BG vom 19. März 2021 (Härtefälle, Arbeitslosenversicherung, familienergänzende Kinderbetreuung, Kulturschaffende, Veranstaltungen), in Kraft vom 20. März 2021 bis zum 31. Dez. 2031 (AS 2021 153, 878 Ziff. II Abs. 3; BBl 2021 285, 2515).

Majeur	Art. 3 Mesures dans le domaine des capacités sanitaires	Art. 3 Maßnahmen im Bereich der Gesundheitskapazitäten	Art. 3 Massnahmen im Bereich der Gesundheitsversorgung
Majeur	prévoir des dérogations au régime de l'autorisation pour des activités en relation avec des biens médicaux importants ou adapter les conditions d'autorisation;	Ausnahmen von der Genehmigungsregelung für Tätigkeiten im Zusammenhang mit wichtigen medizinischen Gütern vorsehen oder die Bedingungen für die Selbstbestimmung anpassen;	Ausnahmen von der Bewilligungspflicht für Tätigkeiten im Zusammenhang mit wichtigen medizinischen Gütern vorsehen oder die Bewilligungsvoraussetzungen anpassen;
Critique	prévoir des dérogations aux dispositions sur l'évaluation de la conformité des dispositifs médicaux et aux dispositions sur la procédure d'évaluation et la mise sur le marché d'équipements de protection;	Ausnahmen von den Bestimmungen über die Konformitätsbewertung von Medizinprodukten und den Bestimmungen über das Evakuierungsverfahren und das Inverkehrbringen von Schutzausrüstungen vorzusehen;	Ausnahmen von den Bestimmungen über die Konformitätsbewertung von Medizinprodukten sowie von den Bestimmungen über das Konformitätsbewertungsverfahren und das Inverkehrbringen von Schutzausrüstungen vorsehen;
Critique	S'il prend des mesures au sens de l'al. 1, il prévoit que leur exécution relève des organes d'exécution de la loi du 13 mars 1964 sur le travail ²³ et de la Caisse nationale suisse d'assurance en cas d'accidents (CNA), et que les frais résultants de cette exécution sont financés par le supplément de prime destiné aux frais liés à la prévention des accidents et maladies professionnels prévu à l'art. 87 de la loi fédérale du 20 mars 1981 sur l'assurance-accidents ²⁴ .	Wenn er Maßnahmen im Sinne von Absatz 1 trifft, sieht er vor, dass deren Durchführung in den Zuständigkeitsbereich der Vollzugsbehörden des Arbeitsgesetzes vom 13. März 1964 und der Schweizerischen Unfallversicherungskasse (CNA) fällt und dass die Kosten, die sich aus dieser Ausführung ergeben, aus dem Zuschlag für die Kosten der Verhütung von Berufsunfällen und Berufskrankheiten gemäß Artikel 87 des Bundesgesetzes vom 20. März 1981 über die Unfallversicherung finanziert werden.	Ergreift er Massnahmen nach Absatz 1, so sieht er vor, dass der Vollzug den Durchführungsorganen des Arbeitsgesetzes vom 13. März 1964 sowie der Schweizerischen Unfallversicherungsanstalt (Suva) obliegt und dass die dafür anfallenden Vollzugskosten aus dem Prämienzuschlag für die Verhütung von Berufsunfällen und Berufskrankheiten nach Artikel 87 des Bundesgesetzes vom 20. März 1981 über die Unfallversicherung finanziert werden.

Tableau 16 Erreurs de eTranslation dans la catégorie terminologie

Les erreurs mineures concernent les termes qui ont été traduits mot à mot ou qui ont été repris de la phrase source, comme p.ex. le mot « COVID-19 » qui ne s’écrit pas avec des majuscules en allemand.

Les erreurs majeures de cette catégorie concernent la traduction des termes qui ne sont pas vraiment appropriés dans le contexte juridique, comme p.ex. « Stressfall » au lieu de « Härtefall » (en français : cas de rigueur).

Les erreurs critiques concernent la traduction des termes qui ne sont pas du tout appropriés dans le contexte juridique et qui heurtent la compréhensibilité de la traduction.

6.1.2 ÉVALUATION DE LA COMPRÉHENSIBILITÉ

Dans l’évaluation de la compréhensibilité, DeepL a obtenu un score de 458.6 sur 500. Ce score est le score moyen de tous les six juges : Les annotateurs suisses ont donné respectivement les scores 472, 462 et 493, ce qui donne un score moyen de 475.6 ; et les annotateurs allemands ont donné les scores 452, 450 et 423, ce qui correspond au score moyen de 441.6. Les scores des juges pour DeepL sont présentés dans le tableau 17 :

DeepL	Juge 1	Juge 2	Juge 3	Moyenne
CH	472	462	493	475.6
DE	452	450	423	441.6
Moyenne Total				458.6

Tableau 17 Scores de compréhensibilité pour les six juges pour DeepL (maximum = 500)

eTranslation, par contre, a obtenu un score de 424.15 sur 500, qui correspond à la moyenne des six annotateurs : les suisses ont donné les scores 429, 438 et 459 ce qui donne un score moyen de 442 ; les allemands ont donné les scores 404, 413 et 402, ce qui donne un score moyen de 406.3. Les scores des juges pour eTranslation sont présentés dans le tableau 18 :

eTranslation	Juge 1	Juge 2	Juge 3	Moyenne
CH	429	438	459	442
DE	404	413	402	406.3
Moyenne Total				424.15

Tableau 18 Scores de compréhensibilité pour les six juges pour eTranslation (maximum = 500)

6.2 ÉVALUATION AUTOMATIQUE

6.2.1 BLEU

Les résultats du score BLEU sont illustrés dans la Figure 18, où le score BLEU pour eTranslation est représenté en bleu et pour DeepL en vert :

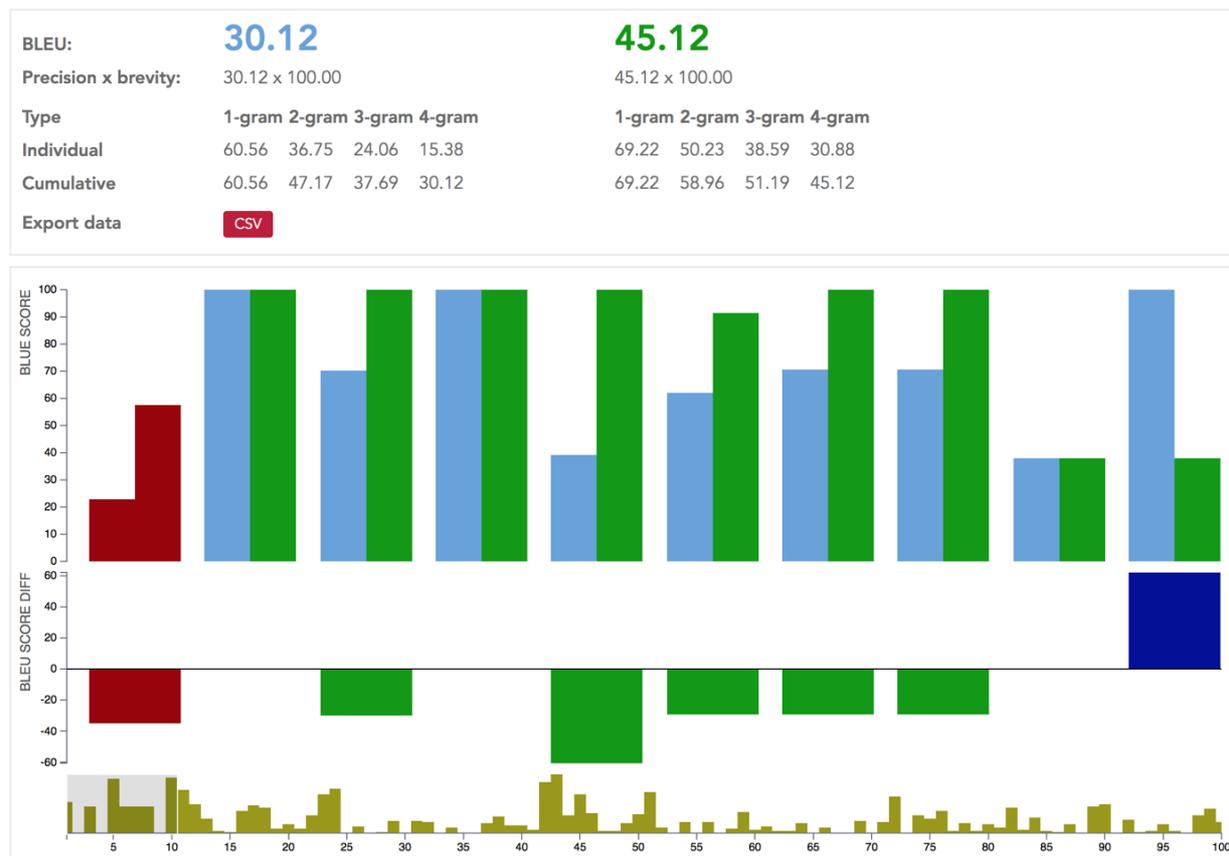


Figure 18 Scores BLEU pour eTranslation et DeepL

On peut observer que le score BLEU pour DeepL (45.12) est plus haut que le score BLEU pour eTranslation (30.12). Ces deux scores correspondent au score cumulé de tous les N-grammes (de 1 à 4). On peut également observer que les scores pour les unigrammes, bigrammes, trigrammes et quadragrammes, cumulés et non cumulés, pour DeepL sont plus élevés que les scores pour eTranslation. Dans les deux systèmes, les unigrammes ont obtenu le score le plus élevé, suivi des bigrammes, des trigrammes ; les quadragrammes ont obtenu le score le moins élevé. Une autre observation est que les scores cumulés dans les deux systèmes sont plus élevés que les scores non cumulés.

6.2.2 TER

Les résultats du calcul de TER pour DeepL et eTranslation sont présentés dans le tableau 19 :

	DeepL	eTranslation
shifts-tried	295	356
segments-scored	100	100
beam-search-calls	395	456
n-edits	877	1058
ref-words	1597	1597
total-ter	0.549	0.662

Tableau 19 Résultats du calcul de TER pour DeepL et eTranslation

« Shifts-tried » représente le nombre de déplacements de mots (shifts) pour chaque système est qui est plus haut pour eTranslation que pour DeepL. « Segments-scored » montre le nombre de segments en total, dans notre cas 100 segments. « N-edits » représente le nombre d'édicions au total, qui est dans ce cas également plus haut pour eTranslation que pour DeepL. « Ref-word » montre le nombre total de mots, qui est égal à 1597 mots dans notre cas.

Le calcul de TER pour DeepL donne un résultat de 0.549 et pour eTranslation un résultat de 0.6622. Moins le score TER est élevé, mieux est la traduction produite par le système de TA, dans notre cas, c'est donc DeepL qui a obtenu un meilleur score TER.

6.3 INTERACCORD

6.3.1 KAPPA

Les résultats de scores Kappa pour l'évaluation de la compréhensibilité de DeepL par les six juges sont représentés dans le tableau 20 :

	DeepL DE	DeepL CH	DeepL DE & CH
Subjects	100	100	100
Raters	3	3	6
z	1.83	-0.124	0.00531
p-value	0.0678	0.901	0.996
Kappa	0.16	-0.0472	0.0573

Tableau 20 Scores Kappa pour DeepL (évaluation de la compréhensibilité)

L'accord entre les juges pour DeepL est très faible. Le score Kappa pour DeepL évalué par les 3 juges allemands est de 0.16 ; le score Kappa pour DeepL évalué par les 3 juges suisses est de -0.0472 ; et le score Kappa pour DeepL, évalué par tous les 6 juges, est de 0.0573.

Les résultats de scores Kappa pour l'évaluation de la compréhensibilité de eTranslation par les six juges sont représentés dans le tableau 21 :

	eTranslation DE	eTranslation CH	eTranslation DE & CH
Subjects	100	100	100
Raters	3	3	6
z	25.3	2.36	0.32
p-value	0	0.0185	0.749
Kappa	0.558	0.229	0.276

Tableau 21 Scores Kappa pour eTranslation (évaluation de la compréhensibilité)

On peut observer que l'accord entre les juges pour l'évaluation de la compréhensibilité de eTranslation est plus élevé que l'accord entre les juges pour DeepL. Le score Kappa pour eTranslation évalué par les 3 juges allemands est de 0.558 ; le score Kappa pour eTranslation évalué par les 3 juges suisses est de 0.229 ; et le score Kappa pour eTranslation évalué par les 6 juges est de 0.276. Ces résultats ne semblent pas être très surprenants vu qu'il est souvent plus facile d'évaluer de mauvaises traductions que des bonnes traductions.

6.3.2 COEFFICIENT DE PEARSON

La corrélation entre les différentes métriques d'évaluation a été calculée avec le coefficient de Pearson. Les résultats pour la corrélation entre le BLEU et l'évaluation de la compréhensibilité sont présentés dans le tableau 22 :

	DeepL	eTranslation
DE	-0.025824976	0.221336866
CH	0.049218199	0.083354152
DE & CH	0.168465709	0.192677827

Tableau 22 Coefficient de Pearson entre l'évaluation de la compréhensibilité et le BLEU

Toutes les corrélations sont très faibles et l'une est négative, ce qui montre qu'il n'y a presque pas de corrélation entre ces deux métriques. Toutefois, on peut observer que la corrélation entre le BLEU et l'évaluation de la compréhensibilité par les juges allemands est plus forte que la corrélation entre le BLEU et l'évaluation de la compréhensibilité par les juges suisses. La corrélation faible négative entre le score BLEU de DeepL et l'évaluation de la compréhensibilité par les juges allemands pour DeepL est illustrée dans la Figure 19 :

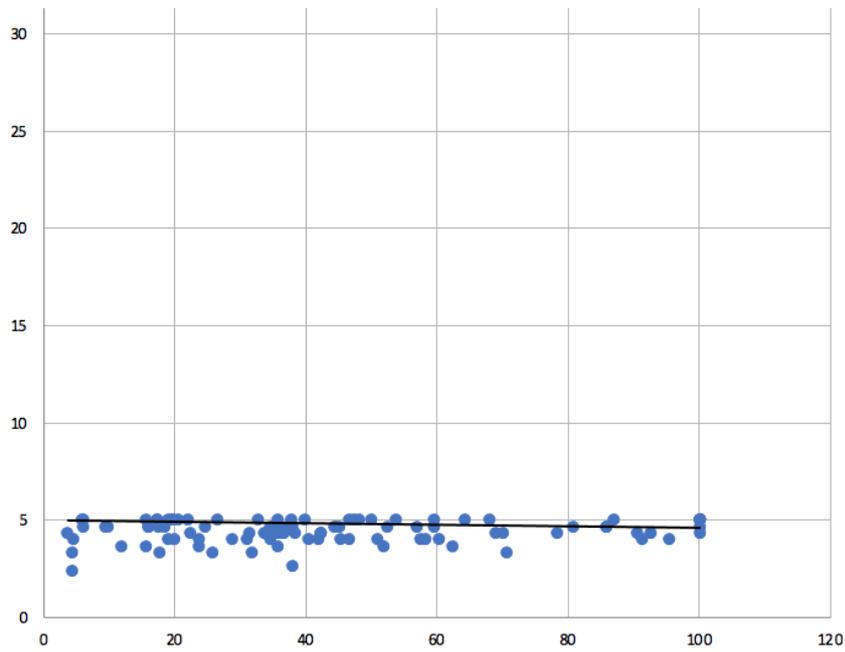


Figure 19 Corrélation entre BLEU et l'évaluation de la compréhension par les juges allemands pour DeepL

La figure 20 illustre la corrélation faible positive entre le BLEU et l'évaluation de la compréhension par les juges allemands pour eTranslation :

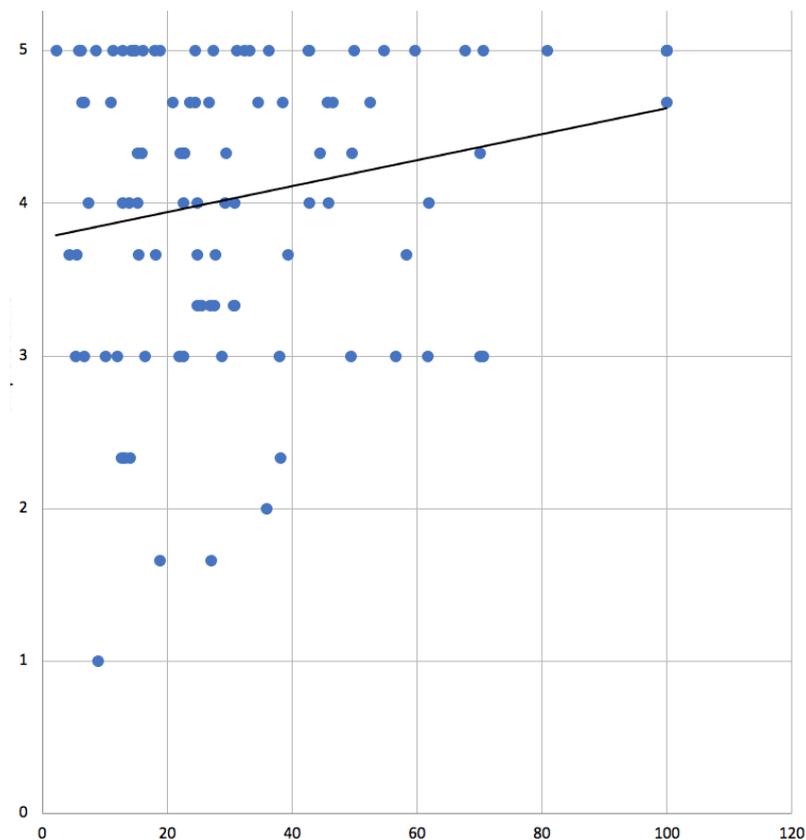


Figure 20 Corrélation entre BLEU et l'évaluation de la compréhension par les juges allemands pour eTranslation

La figure 21 illustre la corrélation faible positive entre le BLEU et l'évaluation de la compréhension par les juges suisses pour DeepL :

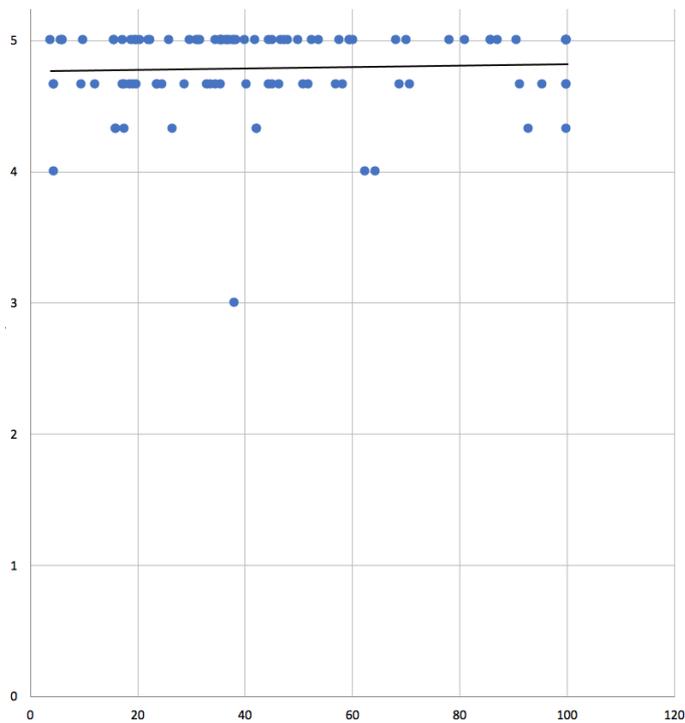


Figure 21 Corrélation entre BLEU et l'évaluation de la compréhension par les juges suisses pour DeepL

La figure 22 illustre la corrélation faible positive entre le BLEU et l'évaluation de la compréhension par les juges suisses pour eTranslation

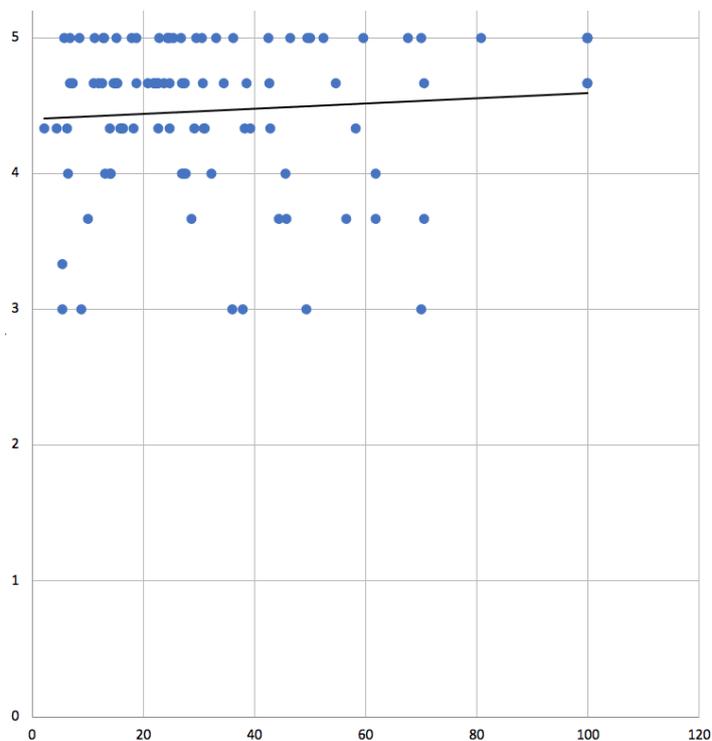


Figure 22 Corrélation entre BLEU et l'évaluation de la compréhension par les juges suisses pour eTranslation

La figure 23 illustre la corrélation faible positive entre le BLEU et l'évaluation de la compréhension par les six juges pour DeepL :

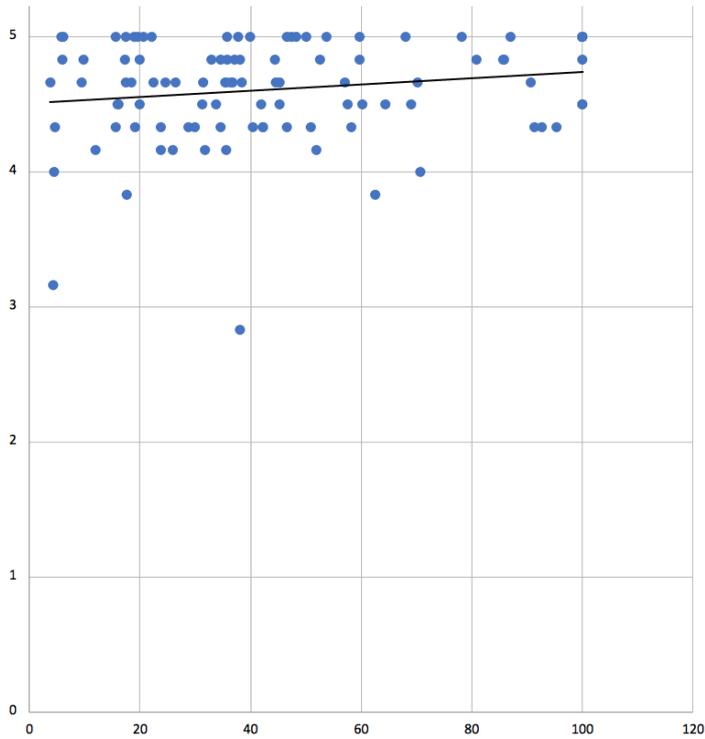


Figure 23 Corrélation entre BLEU et l'évaluation de la compréhension par les six juges pour DeepL

La figure 24 illustre la corrélation faible positive entre le BLEU et l'évaluation de la compréhension par les six juges pour eTranslation :

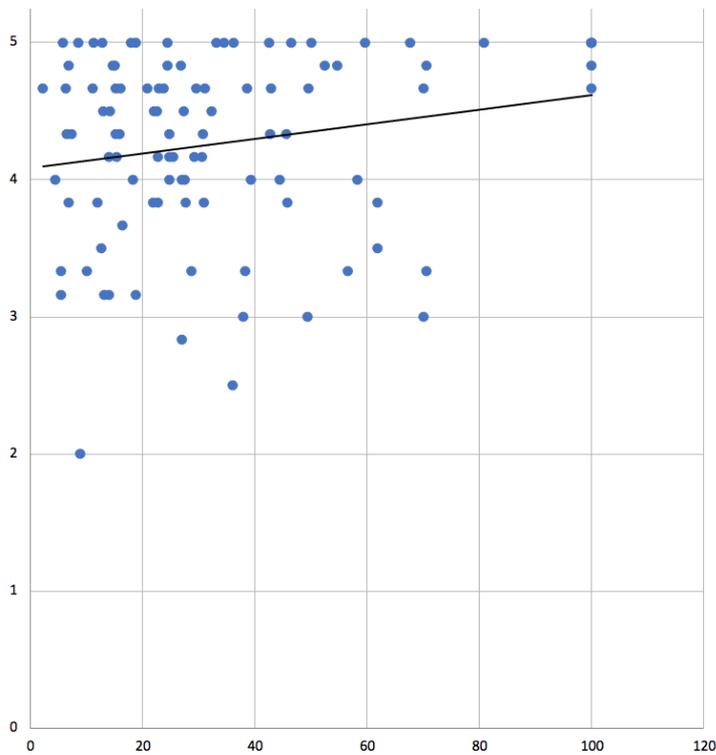


Figure 24 Corrélation entre BLEU et l'évaluation de la compréhension par les six juges pour eTranslation

La corrélation entre les résultats du BLEU et de l'annotation d'erreurs a également été calculée avec le coefficient de Pearson. Cette corrélation pour DeepL est de 0.085804434 et pour eTranslation de -0.1052777392. Le tableau 23 résume les scores pour les deux systèmes de TA :

	DeepL	eTranslation
Coefficient de Pearson	0.085804434	-0.1052777392

Tableau 23 Coefficient de Pearson pour l'annotation d'erreurs et le BLEU

La figure 25 illustre la corrélation faible positive entre le BLEU et l'annotation d'erreurs pour DeepL :

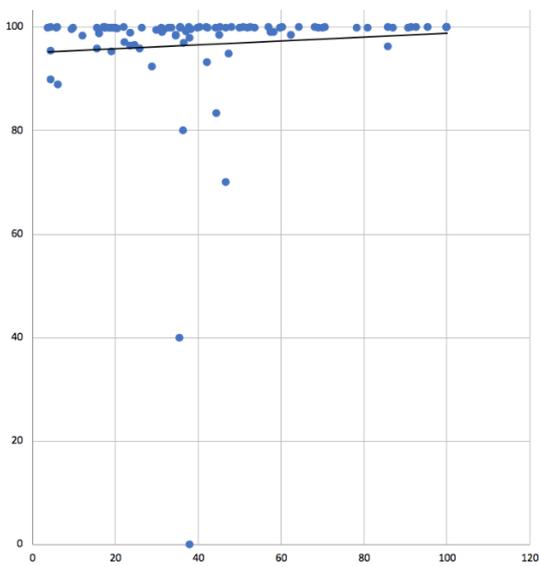


Figure 25 Corrélation entre BLEU et l'annotation d'erreurs pour DeepL

La figure 26 illustre la corrélation faible négative entre le BLEU et l'annotation d'erreurs pour eTranslation :

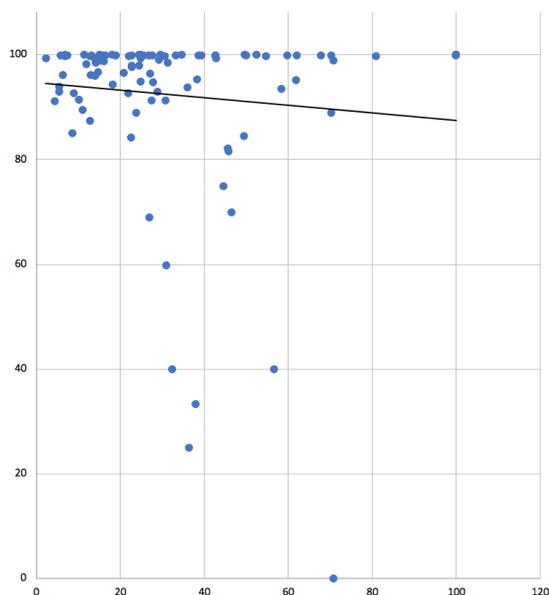


Figure 26 Corrélation entre BLEU et l'annotation d'erreurs pour eTranslation

La corrélation entre l'évaluation de la compréhensibilité et l'annotation d'erreurs a aussi été calculé avec le coefficient de Pearson, pour voir si le nombre d'erreurs affecte ou pas la compréhensibilité. Cette corrélation pour DeepL (évalué par les 6 juges) est de 0.42819815, donc une corrélation moyenne, et pour eTranslation (évalué par les 6 juges) de 0.24470299, donc une corrélation faible. Le tableau 24 résume les scores pour les deux systèmes de TA :

	DeepL	eTranslation
Coefficient de Pearson	0.42819815	0.24470299

Tableau 24 Coefficient de Pearson pour l'évaluation de compréhensibilité et l'annotation d'erreurs

La figure 27 illustre la corrélation moyenne entre l'évaluation de la compréhensibilité et l'annotation d'erreurs pour DeepL :

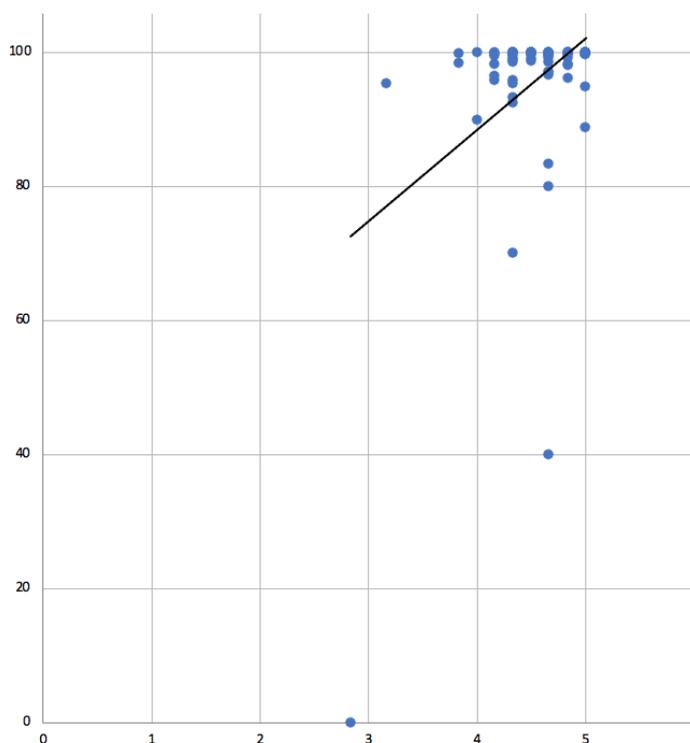


Figure 27 Corrélation entre l'évaluation de la compréhensibilité et l'annotation d'erreurs pour DeepL

La figure 28 illustre la corrélation faible entre l'évaluation de la compréhensibilité et l'annotation d'erreurs pour eTranslation :

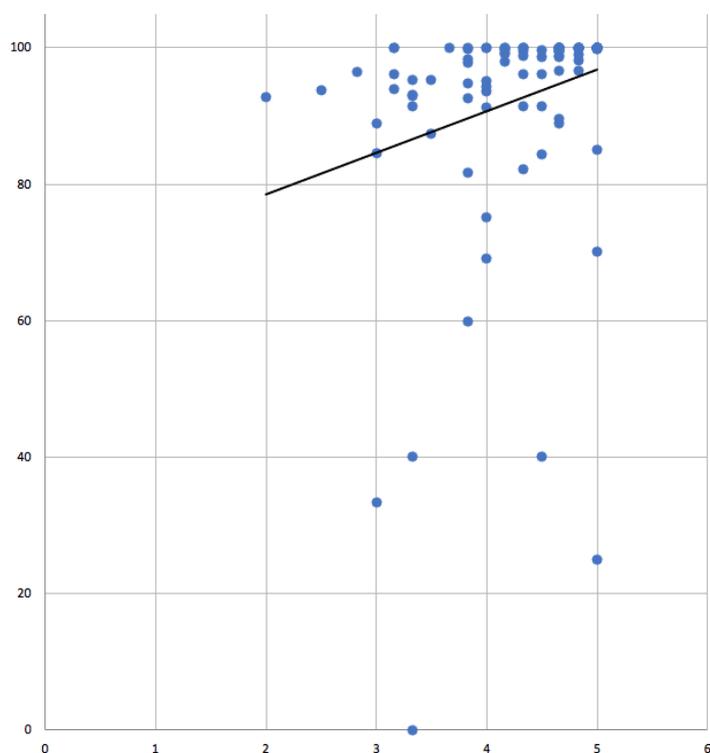


Figure 28 Corrélation entre l'évaluation de la compréhensibilité et l'annotation d'erreurs pour eTranslation

Afin de voir s'il y a une plus forte corrélation entre les erreurs et la compréhensibilité des juges experts et les juges non-experts, le coefficient de Pearson a également été calculé. Le tableau 25 résume les scores obtenus pour les deux systèmes et pour les deux groupes de juges :

	DeepL	eTranslation
Coefficient de Pearson pour les juges experts (juges allemands)	0.06806352	0.14998658
Coefficient de Pearson pour les juges non-experts (juges suisses)	0.45635712	0.33829313

Tableau 25 Corrélation entre l'annotation d'erreurs et la compréhensibilité des juges experts vs. non-experts

On peut observer, que la corrélation entre l'annotation d'erreurs et la compréhensibilité des juges non-experts (juges suisses) est plus forte que la corrélation entre l'annotation d'erreurs et la compréhensibilité des juges experts (juges allemands) est faible pour

les deux systèmes, ce qui montre que les juges experts font plus attention aux fautes que les juges non-experts.

La figure 29 illustre la corrélation faible entre l'annotation d'erreurs et la compréhensibilité des juges experts (juges allemands) pour DeepL :

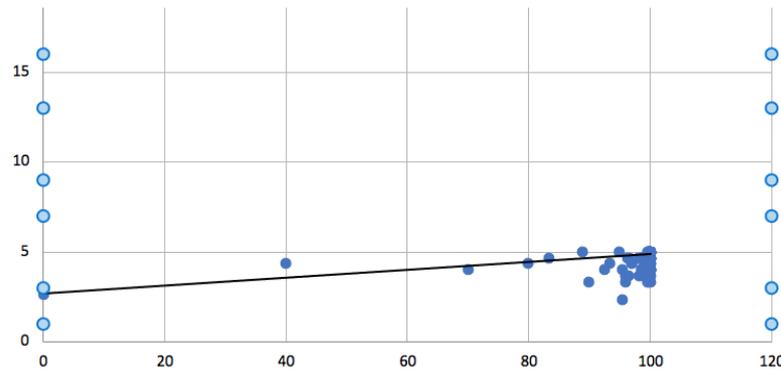


Figure 29 Corrélation entre l'annotation d'erreurs et la compréhensibilité des juges experts (juges allemands) pour DeepL

La figure 30 illustre la corrélation faible entre l'annotation d'erreurs et la compréhensibilité des juges experts (juges allemands) pour eTranslation :

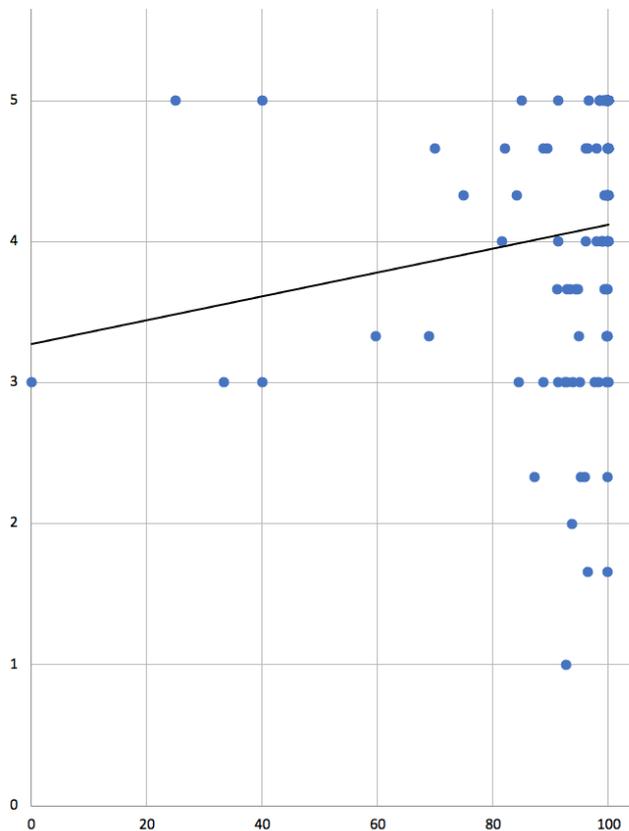


Figure 30 Corrélation entre l'annotation d'erreurs et la compréhensibilité des juges experts (juges allemands) pour eTranslation

La figure 31 illustre la corrélation moyenne entre l'annotation d'erreurs et la compréhensibilité des juges non-experts (juges suisses) pour DeepL :

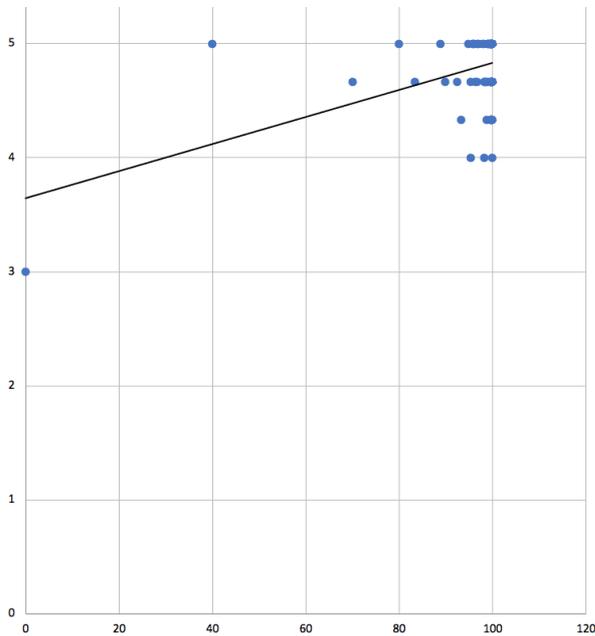


Figure 31 Corrélation entre l'annotation d'erreurs et la compréhensibilité des juges non-experts (juges suisses) pour DeepL

La figure 32 illustre la corrélation moyenne entre l'annotation d'erreurs et la compréhensibilité des juges non-experts (juges suisses) pour eTranslation :

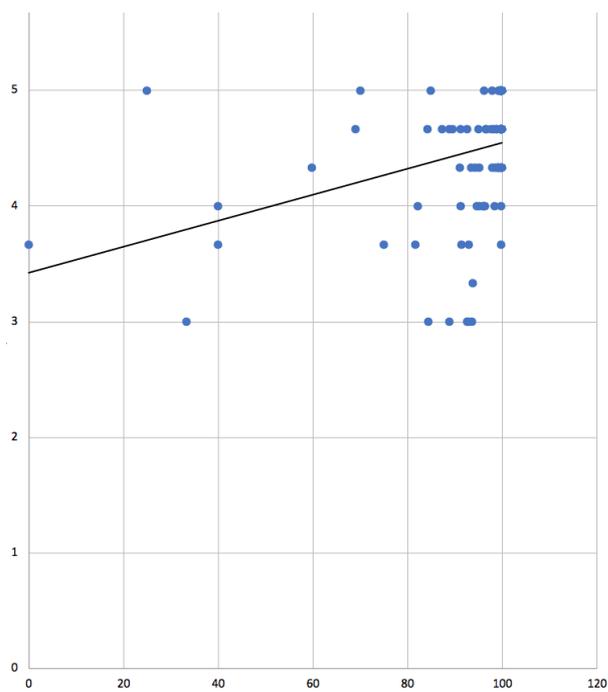


Figure 32 Corrélation entre l'annotation d'erreurs et la compréhensibilité des juges non-experts (juges suisses) pour eTranslation

La corrélation entre les scores BLEU et les scores TER a également été calculée avec le coefficient de Pearson. Cette corrélation pour DeepL est de -0.790202014 et pour eTranslation de -0.730531436, ce qui montre que les deux métriques arrivent aux mêmes conclusions. Le tableau 26 résume les scores pour les deux systèmes de TA :

	DeepL	eTranslation
Coefficient de Pearson	-0.790202014	-0.730531436

Tableau 26 Coefficient de Pearson pour BLEU et TER

La figure 33 illustre la corrélation forte négative entre le BLEU et le TER pour DeepL :

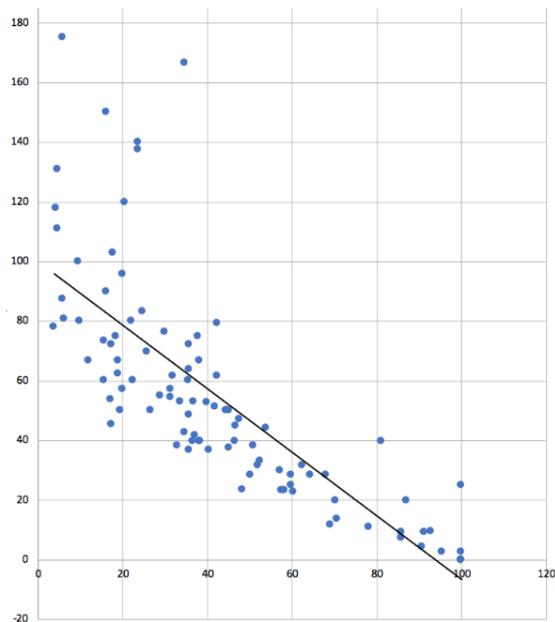


Figure 33 Corrélation entre le BLEU et le TER pour DeepL

La figure 34 illustre la corrélation forte négative entre le BLEU et le TER pour eTranslation :

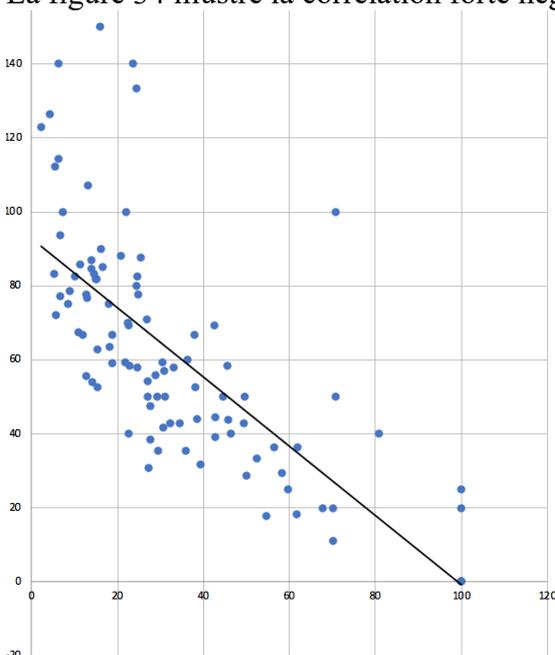


Figure 34 Corrélation entre le BLEU et le TER pour eTranslation

7. DISCUSSION

Ce chapitre vise à résumer et discuter les résultats d'évaluations des deux systèmes de TA, notamment DeepL et eTranslation (section 7.1). Après cette courte discussion, les limites de l'étude seront présentées (section 7.2).

7.1 SYNTHÈSE ET RÉSULTATS DE L'ÉTUDE

Les résultats de l'évaluation humaine, ainsi que de l'évaluation automatique ont montré que DeepL est capable de produire de meilleures traductions dans le domaine juridique que eTranslation, mais une post-édition est nécessaire pour les deux traductions produites par DeepL et eTranslation. Le tableau 27 résume tous les scores obtenus pour les deux systèmes de TA :

	DeepL	eTranslation
Nombre d'erreurs	138	247
Nombre d'erreurs mineures	90	112
Nombre d'erreurs majeures	23	51
Nombre d'erreurs critiques	25	84
Score de pénalité	257.03	770.25
Nombre de segments sans erreurs	29	14
Moyenne du score de compréhensibilité	458.6 (de 500)	424.15 (de 500)
BLEU	45.12	30.12
TER	0.549	0.662
KAPPA	0.0573	0.276

Tableau 27 Résumé des scores pour DeepL et eTranslation

On peut observer que les différents scores pour les deux systèmes de TA sont encore loin d'être parfaits, mais en les comparant, on remarque que DeepL a obtenu des meilleurs scores. Dans l'annotation d'erreurs, DeepL a obtenu un score de pénalité de 257.03 (138 erreurs en total, avec 90 erreurs mineures, 23 erreurs majeures et 25 erreurs critiques) et eTranslation un score de pénalité de 770.25 (247 erreurs en total, avec 112 erreurs mineures, 51 erreurs majeures et 84 erreurs critiques). Ce score de eTranslation est plus que le double du score de pénalité de DeepL ; il faudra donc plus de temps pour post-éditer la traduction produite par eTranslation

que la traduction produite par DeepL, mais il faut toutefois mentionner que les deux systèmes de TA contiennent plus d'erreurs mineures que d'erreurs majeure et critiques. On peut également observer que les catégories qui contiennent le plus d'erreurs dans les deux systèmes de TA sont les catégories « terminologie », « convention locale », et « style ». En ce qui concerne la convention locale, on remarque que eTranslation ne prend pas du tout en compte le contexte suisse, et que DeepL le prend très peu en compte, comme p.ex. l'utilisation de « ss » (qui est utilisé en Suisse) au lieu de « ß » (qui est utilisé en Allemagne) : DeepL traduit en effet une seule fois avec le « ss » et eTranslation ne l'utilise jamais. En outre, la terminologie n'est pas toujours adaptée au contexte suisse, les deux systèmes de TA utilisent souvent des termes de l'Allemagne au lieu des termes de la Suisse, comme par exemple le terme « Ordnungsgeld », qui est utilisé en Allemagne, au lieu du terme « Ordnungsbussen », qui est utilisé en Suisse. Il faut toutefois mentionner que DeepL a fait 27 erreurs dans la catégorie « convention locale » (dont 24 mineures et 3 majeures), donc 12 erreurs moins que eTranslation qui a fait 39 erreurs (dont 33 mineures, 3 majeures et 3 critiques) dans cette catégorie.

En ce qui concerne les abréviations qui font partie de la terminologie, on peut observer que eTranslation a seulement été capable de traduire deux abréviations (de sept en total) correctement : « LF » (en allemand : « BG »), et « RO » (en allemand : « AS »). DeepL, par contre, a été capable de traduire 5 abréviations correctement : « RS » (en allemand : « SR »), « LF » (en allemand : « BG »), « RO » (en allemand : « AS »), « FF » (en allemand : « BBI »), « LAsi » (en allemand : « AsylG »), et une abréviation presque correctement : « CNA » (en allemand : « Suva »), DeepL a traduit cette abréviation avec des majuscules, ce qui est considéré comme une erreur mineure, vu que le public cible peut le comprendre quand-même.

En ce qui concerne les segments sans erreurs, DeepL a réussi à traduire 29 segments sans erreurs et eTranslation a réussi à traduire 14 segments sans erreurs.

Les scores de la compréhensibilité montrent également que DeepL est le meilleur système de TA entre les deux (score de DeepL : 458.6 de 500 possibles ; et score de eTranslation : 424.15 de 500 possibles).

En ce qui concerne l'évaluation automatique, DeepL a également obtenu de meilleurs scores pour les métriques BLEU et TER : le score BLEU de DeepL est de 45.12 et de eTranslation de 30.12 ; le score TER de DeepL est de 0.549 et de eTranslation de 0.662. En observant les résultats pour le score KAPPA, on peut voir que l'accord entre les juges pour eTranslation est plus grand que pour DeepL, ce qui n'est pas surprenant, vu qu'il est souvent plus facile d'évaluer de mauvaises traductions que des bonnes traductions. Ces résultats obtenus concordent,

même si le coefficient de Pearson ne montre pas de corrélation forte entre les mesures automatiques et les mesures humaines, mais il y a une forte corrélation entre les deux métriques automatique BLEU et TER.

Les résultats du coefficient de Pearson pour les différentes métriques sont résumés dans le tableau 28 :

	DeepL	eTranslation
Coefficient de Pearson pour l'évaluation de la compréhensibilité et BLEU	0.168465709	0.192677827
Coefficient de Pearson pour l'annotation d'erreurs et le BLEU	0.085804434	-0.1052777392
Coefficient de Pearson pour l'évaluation de la compréhensibilité (par les 6 juges) et l'annotation d'erreurs	0.42819815	0.24470299
Coefficient de Pearson pour l'annotation d'erreurs et la compréhensibilité des juges experts (juges allemands)	0.06806352	0.14998658
Coefficient de Pearson pour l'annotation d'erreurs et la compréhensibilité des juges non-experts (juges suisses)	0.45635712	0.33829313
Coefficient de Pearson pour BLEU et TER	-0.790202014	-0.730531436

Tableau 28 Corrélations entre les différentes métriques pour DeepL et eTranslation

La corrélation entre le BLEU et l'évaluation de la compréhensibilité pour les deux systèmes de TA est faible positive, c'est-à-dire qu'il n'y a presque pas de corrélation entre ces deux métriques. Ainsi, la corrélation entre le BLEU et l'annotation d'erreurs pour les deux systèmes sont très faibles, mais la corrélation pour eTranslation est plus forte que pour DeepL, même si cette valeur est négative, c'est-à-dire si la valeur du BLEU augmente, la valeur de l'annotation d'erreurs diminue. La corrélation entre le BLEU et l'annotation d'erreurs pour DeepL devrait également être négative, mais elle est très faible positive, ce qui est contre-intuitif. La corrélation entre les deux métriques humaines, l'évaluation de la compréhensibilité et l'annotation des erreurs, est moyenne pour DeepL et faible pour eTranslation, ce qui montre qu'il y a une corrélation entre le nombre d'erreurs faites par les deux systèmes de TA et la compréhensibilité des traductions. Les corrélations entre l'annotation d'erreurs et la compréhensibilité des juges non-experts (juges suisses) pour les deux systèmes sont plus fortes que les corrélations entre

l'annotation d'erreurs et la compréhensibilité des juges experts (juges allemands), ce qui montre que les juges experts font plus attention aux fautes que les juges non-experts. Les corrélations entre les deux métriques automatiques, le BLEU et le TER, par contre, sont pour les deux systèmes négatives, très fortes selon l'interprétation de Koehn (2020), c'est-à-dire si la valeur du BLEU augmente, la valeur du TER diminue et vice versa.

Pour répondre à la question de recherche qui était : « *La qualité de la traduction automatique neuronale est-elle satisfaisante dans le domaine juridique ?* », on peut dire que les deux systèmes de TA utilisés pour ce travail sont encore loin de pouvoir être utilisés sans aucune intervention humaine, mais les bons scores suggèrent qu'ils pourraient être utilisés pour la traduction juridique. L'effort de post-édition ne serait en effet pas très élevé, vu que les deux systèmes de TA contiennent moins d'erreurs critiques et majeures que d'erreurs mineures. Afin de savoir s'il serait plus rapide de post-éditer une traduction produite par un système de TA ou s'il serait plus rapide de produire une traduction humaine à partir de zéro, il faudra mesurer le temps de post-édition et le temps nécessité par le traducteur humain et les comparer, c'est-à-dire qu'il faudra mesurer l'effort de post-édition dans une étude de productivité. En observant les résultats obtenus pour l'annotation d'erreurs, on peut assumer que l'effort de post-édition pour eTranslation serait plus grand que l'effort de post-édition pour DeepL, parce que eTranslation a fait 247 erreurs en total, ce qui donne un score de pénalité de 770.25, tandis que DeepL a fait 138 erreurs, ce qui donne un score de pénalité de 257.03

Pour répondre aux deux sous-questions : « *Est-ce que la qualité de la traduction automatique dépend des systèmes ?* » et « *Est-ce que la qualité de la traduction automatique dépend des évaluateurs ?* », on peut dire que les résultats des évaluations humaines, ainsi que les résultats des évaluations automatiques ont prouvé que DeepL est capable de produire une meilleure traduction que eTranslation ; et que la qualité de la traduction automatique peut dépendre des évaluateurs, parce que dans notre cas, l'annotation d'erreurs et l'évaluation de la compréhensibilité sont des tâches très subjectives, ce qui a été montré par le score KAPPA. Cette subjectivité est surtout visible dans l'évaluation de la compréhensibilité : les trois évaluateurs allemands, qui sont tous les trois des étudiants de traduction, ont été plus critiques que les trois évaluateurs suisses, qui ne sont pas des professionnels ni dans le domaine juridique ni dans le domaine linguistique.

En ce qui concerne les deux hypothèses, on peut dire que la première, qui supposait que ni eTranslation ni DeepL n'allaient prendre en compte le contexte suisse et que eTranslation aurait un meilleur score, a été partiellement confirmée. La seconde hypothèse, qui supposait que la qualité de la TA dépend des évaluateurs et que les juges non-experts (juges suisses) donneraient des meilleurs scores que les juges experts (juges allemands), a été confirmée.

Afin d'améliorer la qualité des deux systèmes de TA, il faut surtout éviter ou minimiser les erreurs dans les catégories « terminologie », « convention locale », et « style », vu que ce sont les catégories avec les plus d'erreurs. En outre, les deux systèmes devraient prendre en compte le contexte suisse en utilisant le « ss » au lieu de « ß » et en n'utilisant que des termes utilisés en Suisse. Une solution possible pour résoudre ce problème pourrait être la spécification de la région pour la langue cible, donc une possibilité pour les utilisateurs de choisir p.ex. l'allemand utilisé en Suisse, comme c'est déjà le cas pour l'anglais (sélection entre anglais américain et anglais britannique). En incluant des règles, qui p.ex. interdisent l'utilisation de « ß » dans le contexte suisse, les deux systèmes de TA pourraient traduire avec le « ss ». De la même manière, l'inclusion d'un glossaire terminologique avec des termes suisses pourrait être la solution pour éviter ou minimiser les erreurs dans les catégories « terminologie » et « convention locale ».

7.2 LIMITES DE L'ÉTUDE

L'étude qui a été conduite dans le cadre de ce travail est une étude assez courte avec de nombreuses limites.

Tout d'abord, les deux systèmes de TA utilisés pour ce travail ont été comparés sur un sous-domaine de la traduction juridique, il est donc difficile de généraliser les résultats obtenus dans ce travail pour tous les sous-domaines de la traduction juridique. Ensuite, les évaluations effectuées dans ce travail portaient sur des corpus relativement courts et les évaluations humaines ont été effectuées avec un petit nombre de juges qui portaient seulement sur la paire de langues français vers allemand. Afin de vérifier la fiabilité des résultats obtenus dans ce travail, il faudra mener des études de plus grande ampleur.

Ce travail compare la qualité des deux systèmes de TA choisis pour ce travail, mais sans prendre en compte la traduction humaine. Les résultats obtenus montrent que DeepL est capable de produire de meilleures traductions que eTranslation, mais ne montrent pas dans quelle mesure l'utilisation de ce système de TA permettrait un gain de temps et de productivité par rapport à la traduction humaine. Afin de savoir, quelle stratégie entre les deux (post-édition de la TA et traduction humaine) est la plus efficace, il faudra mettre en place une étude de productivité.

En outre, les résultats obtenus pour la corrélation entre les différentes métriques remettent en question la fiabilité d'évaluations qui reposent uniquement sur une traduction de référence. Comme expliqué par Koehn (2010, p. 227) : « Given the variability in translation, it is harsh to require matches of the system output against a single human reference translation. If multiple human reference translations are used, it is more likely that all acceptable translations of ambiguous parts of sentences show up ». Donc, on pourrait obtenir des résultats plus fiables si plusieurs traductions de référence (traductions humaines) étaient utilisées.

Bien que les résultats obtenus soient concluants, cette étude mériterait d'être approfondie pour les raisons des limites mentionnées ci-dessus. Il serait judicieux de poursuivre ces recherches sur des corpus de plus grande taille et dans d'autres sous-domaines juridiques, et avec un plus grand nombre d'évaluateurs humains, afin d'obtenir de meilleurs scores KAPPA. Il faut toutefois mentionner que l'annotation d'erreurs a été faite par une seule personne, mais cette personne est une experte. Il serait également intéressant d'évaluer d'autres paires de langues et d'autres systèmes de TA.

RÉFÉRENCES

Alcaraz, E., & Hughes, B. (2014). *Legal Translation Explained*. Routledge.

ALPAC. (1966). *Language and Machines. Computers in Translation and Linguistics*. Automatic Language Processing Advisory Committee, Division of Behavioral Sciences, National Academy of Sciences, National Research Council.

Alvarez-Vidal, S., Oliver, A., & Badia, T. (2021). What Do Post-Editors Correct? A Fine-Grained Analysis of SMT and NMT Errors. *Revista Traudumatica, Numero 19*, 131-147.

Arnold, D., Balkan, L., Lee Humphreys, R., Meijer, S., & Sadler, L. (1994). *Machine Translation. An Introductory Guide*. Arnold, Balkan, Lee Humphreys, Meijer, Sadler.

Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural Machine Translation by Jointly Learning to Align and Translate. *International Conference on Learning Representations, ICLR 2015, San Diego, California*.

Banerjee, S., & Lavie, A. (2005). METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgements. *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, 65-72.

Bouillon, P. (2020). *Cours de traduction automatique 1—Systèmes et architecture*.

Bouillon, P. (2021). *Cours de traduction automatique 2*.

Bouillon, P. (2022). *Exercice évaluation automatique*. Cours de traduction automatique 2. Université de Genève.

Bouillon, P., & Clas, A. (1993). *La Traductique*. Les Presses de l'Université de Montréal.

Cao, D. (2007). Law, Language and Translation. In *Translating Law. Multilingual Matters*.

Cao, D. (2013). Legal translation studies. In *The Routledge Handbook of Translation Studies* (p. 415-424). Routledge.

Chancellerie fédérale. (2022a). *Commission interne de rédaction (CIR)*. Confédération suisse. [https://www.bk.admin.ch/bk/fr/home/soutien-gouvernement/accompagnement-legislatif/R%C3%A9daction_1%C3%A9gislative/Commission_interne%20de%20r%C3%A9daction%20\(CIR\).html](https://www.bk.admin.ch/bk/fr/home/soutien-gouvernement/accompagnement-legislatif/R%C3%A9daction_1%C3%A9gislative/Commission_interne%20de%20r%C3%A9daction%20(CIR).html)

Chancellerie fédérale. (2022b). *Langues*. Confédération suisse. <https://www.bk.admin.ch/bk/fr/home/soutien-gouvernement/langues.html>

Commission européenne. (2022a). *Machine Translation Help*. <https://webgate.ec.europa.eu/etranslation/help.html>

Commission européenne. (2022b). *Traduction automatique pour les administrations publiques—ETranslation*. https://ec.europa.eu/info/resources-partners/machine-translation-public-administrations-etranslation_fr#aide

CRAN. (2022). *The Comprehensive R Archive Network*. <https://stat.ethz.ch/CRAN/>

Cuenca, E., Estrella, P., Bruno, L., Mutal, J. D., Girletti, S., Volkart, L., & Bouillon, P. (2020). Re-design of the Machine Translation Training Tool MT3. *22nd Annual Conference of the European Association for Machine Translation (EAMT)*.

DeepL. (2021). *À propos de DeepL*. DeepL Traducteur. <https://www.deepl.com/fr/publisher/>

DeepL. (2022). *L'API de DeepL*. DeepL Traducteur. <https://www.deepl.com/fr/docs-api/>

Dr. Eisele, A. (2017). Machine Translation at the European Commission. In *Maschinelle Übersetzung. Grundlagen für den professionellen Einsatz*. BDÜ Fachverlag.

Dullion, V. (2014). *Droit comparé pour traducteurs : De la théorie à la didactique de la traduction juridique*. Springer Science+Business Media.

Forcada, M. L. (2017). Making sense of neural machine translation. *Translation Spaces*, 6(2), 291-309. <https://doi.org/10.1075/ts.6.2.06for>

Foti, M. (2022). *eTranslation. Le système de traduction automatique de la Commission européenne en appui d'une Europe numérique*. Traduire [En ligne], 246. Mis en ligne le 15 juin 2022, consulté le 31 juillet 2022. <http://journals.openedition.org/traduire/2793>

Gémar, J. C. (2011). Langue, droit et traduction : L'apport de la jurilinguistique. In *Les frontières avancées du savoir du juriste : L'anthropologie juridique et la traductologie juridique*. Etablissements Emile Bruylant.

Gouvernement du Canada. (2021). *À propos du bijuridisme*. <https://www.justice.gc.ca/fra/sjc-csj/harmonization/bijurilex/aproposb-aboutb.html>

Hearne, M., & Way, A. (2011). Statistical Machine Translation : A Guide for Linguists and Translators. *Language and Linguistics Compass*, 5 (5), 205-226.

Hutchins, W. J. (1986). *Machine Translation Past, Present, Future*. Ellis Horwood Limited.

Hutchins, W. J. (2014). The history of machine translation in a nutshell. *Retrieved December, 20, 2009*.

- Hutchins, W. J., & Somers, L. H. (1992). *An Introduction to Machine Translation*. ACADEMIC PRESS.
- Jurafsky, D. (2014). Machine Translation. In *Speech and Language Processing*.
- Jurafsky, D. (2020a). Machine Translation and Encoder-Decoder Models. In *Speech and Language Processing*.
- Jurafsky, D. (2020b). *Speech and Language Processing*.
- Jurafsky, D., & Martin, J. H. (2020). Regular Expressions, Text Normalization, Edit Distance. In *Speech and Language Processing*.
- Killman, J. (2014). Vocabulary accuracy of statistical machine translation in the legal context. *Proceedings of the 11th Conference of the Association for Machine Translation in the Americas.*, 85-98.
- Kit, C., & Wong, T. (2015). Evaluation in machine translation and computer-aided translation. In *The Routledge encyclopedia of translation technology*. Routledge.
- Koehn, P. (2010). *Statistical Machine Translation*. Cambridge University Press.
- Koehn, P. (2017). Neural Machine Translation. In *Statistical Machine Translation*. arXiv:1709.07809. <https://arxiv.org/abs/1709.07809>
- Koehn, P. (2020). *Neural Machine Translation*. Cambridge University Press.
- Koehn, P. (2022). The State of Neural Machine Translation (NMT) by Philipp Koehn. *Omniscien Technologies*. <https://omniscien.com/blog/state-neural-machine-translation-nmt/>
- L'Homme, M.-C. (2008). *Initiation à la traductique*. Linguattech éditeur.
- Lommel, A., Burchardt, A., & Uszkoreit, H. (2015). *Multidimensional Quality Metrics (MQM) Definition*. Deutsches Forschungszentrum für Künstliche Intelligenz GmbH. <http://www.qt21.eu/mqm-definition/definition-2015-12-30.html>
- Martikainen, H. (2022). Ghosts in the machine : Can adaptive MT help reclaim a place for the human in the loop? *hal-03548696*.
- Matthiesen, A. J. (2017). *Maschinelle Übersetzung im Wandel. Die Auswirkungen von künstlicher Intelligenz auf maschinelle Übersetzungssysteme. Mit einer vergleichenden Untersuchung von Google Translate und Microsoft Translator*. Epubli.
- Melby, A. K. (2019). Future of machine translation. In *The Routledge Handbook of Translation and Technology*. Routledge.

- Monjean-Decaudin, S. (2014). Théorie et pratique de la traduction juridique ou sur les lieux d'une rencontre. *De Gruyter Mouton. Semiotica*, 201, 81-101.
- Mossop, B. (2020). Quality assessment. In *Revising and Editing for Translators* (p. 228-234). Routledge.
- Muir Watt, H. (2022). *DROIT - Droit comparé*. Encyclopaedia Universalis. <https://www.universalis.fr/encyclopedie/droit-droit-compare/>
- Nirenburg, S. (1987). *Machine Translation. Theoretical and methodological issues*. Cambridge University Press.
- Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002). BLEU: a Method for Automatic Evaluation of Machine Translation. *Association for Computational Linguistics (ACL)*.
- Pelage, J. (2001). *Éléments de traductologie juridique. Application aux langue romanes*. Jacques Pelage.
- Ping, K. (2011). Machine Translation. In *Routledge Encyclopedia of Translation Studies*. Routledge.
- Poibeau, T. (2019). *Babel 2.0 Où va la traduction automatique?* Odile Jacob.
- Popovic, M., & Arcan, M. (2016). PE2rr Corpus : Manuel Error Annotation of Automatically Pre-annotated MT Post-edits. *LREC*. <https://aclanthology.org/L16-1005.pdf>
- Prieto Ramos, F. (2015). *Quality Assurance in Legal Translation : Evaluating Process, Competence and Product in the Pursuit of Adequacy*. <https://doi.org/10.1007/s11196-014-9390-9>
- Quah, C. K. (2006). *Translation and Technology*. Palgrave Macmillan.
- Qun, L., & Xiaojun, Z. (2015). Machine Translation General. In *The Routledge encyclopedia of translation technology*. Routledge.
- Ramlow, M. (2009). *Die maschinelle Simulierbarkeit des Humanübersetzens. Evaluation von Mensch-Maschine-Interaktion und der Translatqualität der Technik*. Frank&Timme.
- Ruano, M. R. M. (2020). Legal translation. In *Routledge Encyclopedias of Translation Studies* (p. 280-285). Routledge.
- Sandrini, P. (1999). *Übersetzen von Rechtstexten. Fachkommunikation im Spannungsfeld zwischen Rechtsordnung und Sprache*. Gunter Narr Verlag Tübingen.
- Sarcevic, S. (1999). Das Übersetzen normativer Rechtstexte. In *Übersetzten von Rechtstexten. Fachkommunikation im Spannungsfeld zwischen Rechtsordnung und Sprache* (Vol. 54). Gunter

Narr Verlag Tübingen.

Schweizer, R. J., Borghi, M., Baumann, J., Berther, A., Janczak, P., Lötscher, A., Raveglia, G., De Rossa Gisimundo, F., & Scheffler, J. (2011). *Mehrsprachige Gesetzgebung in der Schweiz : Thesen und Empfehlungen*. Confédération suisse.

Schwitler, R. (2015). Controlled Language. In *Routledge Encyclopedia of Translation Technology*. Routledge.

Shterionov, D., Superbo, R., Nagle, P., Casanellas, L., O'Dowd, T., & Way, A. (2017). *Human versus automatic quality evaluation of NMT and PBSMT*. Springer Science+Business Media.

Snover, M., Dorr, B., Schwartz, R., Micciulla, L., & Makhoul, J. (2006). A Study of Translation Edit Rate with Targeted Human Annotation. *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers*, 223-231.

Systran. (2016). *How does Neural Machine Translation work?* Systran Blog. <https://www.systransoft.com/blog/how-does-neural-machine-translation-work/>

Thurmain, G. (2005). Hybrid Architectures for Machine Translation Systems. *Language Resources and Evaluation, Vol. 39, No. 1*, 91-108.

Tilde. (2022). *Interactive BLEU*. Tilde Machine Translation. <https://www.tilde.com/products-and-services/machine-translation/features/interactive-bleu>

Trujillo, A. (1999). *Translation Engines : Techniques for Machine Translation*. Springer.

Weaver, W. (1949). *Translation*. The Rockefeller Foundation.

Wiesmann, E. (2019). MACHINE TRANSLATION IN THE FIELD OF LAW: A STUDY OF THE TRANSLATION OF ITALIAN LEGAL TEXTS INTO GERMAN. *Comparative Linguistics*, 37.

Wymann, A. (2019). *Bericht DeepL-Test. Auswertung der Testergebnisse und Empfehlungen der Arbeitsgruppe « Maschinelle Übersetzung »*. Confédération suisse. <https://www.newsadmin.ch/newsd/message/attachments/59735.pdf>

Yates, S. (2006). Scaling the Tower of Babel Fish : An Analysis of the Machine Translation of Legal Information. *Law Library Journal*, 98 (No.3), 481-502.