La traduction automatique des textes faciles à lire et à comprendre (FALC) : une étude comparative

Rodriguez Vazquez, Silvia; Kaplan, Abigail; Bouillon, Pierrette; Griebel, Cornelia; Azari, Razieh

How to cite

RODRIGUEZ VAZQUEZ, Silvia et al. La traduction automatique des textes faciles à lire et à comprendre (FALC) : une étude comparative. In: Meta, 2022, vol. 67, n° 1, p. 18–49. doi: 10.7202/1092189ar

This publication URL: https://archive-ouverte.unige.ch/unige:164865
Publication DOI: 10.7202/1092189ar

Meta
Journal des traducteurs
Translators' Journal

META

La traduction automatique des textes faciles à lire et à comprendre (FALC) : une étude comparative

Silvia Rodríguez Vázquez, Abigail Kaplan, Pierrette Bouillon, Cornelia Griebel and Razieh Azari

Volume 67, Number 1, April–May 2022
Pour de nouvelles méthodes en traductologie quantitative
Exploring New Methods in Quantitative Translation Studies

URI: https://id.erudit.org/iderudit/1092189ar
DOI: https://doi.org/10.7202/1092189ar

See table of contents

Publisher(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (print)
1492-1421 (digital)

Explore this journal

Cite this article

Rodríguez Vázquez, S., Kaplan, A., Bouillon, P., Griebel, C. & Azari, R. (2022). La traduction automatique des textes faciles à lire et à comprendre (FALC) : une étude comparative. *Meta*, *67*(1), 18–49. https://doi.org/10.7202/1092189ar

Article abstract

Over the last decade, controlled languages (CL) have received increased attention in machine translation (MT) research. The vast majority of studies have dealt with the impact of CLs on the quality of the final MT output, but very little work has focused on the impact of MT on the accessibility of target texts for people with special needs. This article represents a first attempt to bridge this gap. We present a comparative linguistic study that seeks to explore whether MT systems are a viable option for translating texts that are easy to read and understand (EtR). We tested DeepL, Google Translate, and Yandex with EtR texts from three different domains in four language pairs. Findings show that DeepL is the highest-performing system, and that Spanish and administrative texts in particular seem to present more challenges. The evaluation of the MT output in terms of linguistic accessibility indicates that the highest number of issues are found at a lexical and stylistic level. Although MT systems do not generate EtR texts of acceptable quality yet, our study highlights the potential of this tool, as well as the challenges of creating multilingual content that is accessible for all.

La traduction automatique des textes faciles à lire et à comprendre (FALC) : une étude comparative

SILVIA RODRÍGUEZ VÁZQUEZ
Université de Genève, Genève, Suisse
silvia.rodriguez@unige.ch

ABIGAIL KAPLAN
Spécialiste indépendante en communication multilingue
abigail.kaplan1@gmail.com

PIERRETTE BOUILLON
Université de Genève, Genève, Suisse
pierrette.bouillon@unige.ch

CORNELIA GRIEBEL
Université de Genève, Genève, Suisse
cornelia.griebel@unige.ch

RAZIEH AZARI
Université de Genève, Genève, Suisse
razieh.azari@unige.ch

RÉSUMÉ

Au cours de la dernière décennie, les langages contrôlés (LC) ont toujours été l'objet d'une attention accrue en traduction automatique (TA). La majorité des études ont porté sur l'impact des LC sur la qualité du produit final de la TA, mais très peu d'entre elles se sont intéressées à l'impact de la TA sur l'accessibilité des textes cibles pour les personnes à besoins particuliers. Cet article vise à combler cette lacune. Il cherche à déterminer, par le biais d'une étude linguistique comparative, si les systèmes de TA génériques constituent une option viable pour produire, à partir de textes simplifiés sources, des textes cibles faciles à lire et à comprendre (FALC). Nous avons testé trois outils génériques de TA (DeepL, Google Translate et Yandex) avec des textes FALC de trois domaines différents, dans quatre paires de langues. Les résultats montrent que DeepL est l'outil le plus performant et que l'espagnol et les textes administratifs restent ceux qui occasionnent le plus de problèmes pour la TA. En ce qui concerne l'évaluation de l'accessibilité linguistique, les problèmes aux niveaux lexical et stylistique sont les plus nombreux. Même si la TA ne produit pas encore des textes FALC acceptables, notre étude en souligne le potentiel et met l'accent sur la difficulté de créer du contenu multilingue accessible pour tous.

ABSTRACT

Over the last decade, controlled languages (CL) have received increased attention in machine translation (MT) research. The vast majority of studies have dealt with the impact of CLs on the quality of the final MT output, but very little work has focused on the impact of MT on the accessibility of target texts for people with special needs. This article represents a first attempt to bridge this gap. We present a comparative linguistic study that seeks to explore whether MT systems are a viable option for translating texts that are easy to read and understand (EtR). We tested DeepL, Google Translate, and Yandex with EtR texts from three different domains in four language pairs. Findings show

that DeepL is the highest-performing system, and that Spanish and administrative texts in particular seem to present more challenges. The evaluation of the MT output in terms of linguistic accessibility indicates that the highest number of issues are found at a lexical and stylistic level. Although MT systems do not generate EtR texts of acceptable quality yet, our study highlights the potential of this tool, as well as the challenges of creating multilingual content that is accessible for all.

RESUMEN

A lo largo de la última década, la aplicación de lenguajes controlados (LC) ha sido objeto de estudio en numerosos trabajos de investigación sobre traducción automática (MT). Muchos de ellos se han centrado en analizar el efecto del uso de LC en la calidad del producto final, pero apenas existen estudios que hayan explorado el impacto de la TA en la accesibilidad de los textos meta para las personas con discapacidad. El presente artículo busca contribuir al estado de la cuestión en dicha materia. A través de un estudio lingüístico comparativo, hemos querido comprender si los sistemas de TA son una opción viable para traducir textos en lectura fácil (LF). En el estudio se evaluaron tres sistemas (DeepL, Google Translate y Yandex) con textos en LF de tres campos diferentes, en cuatro combinaciones lingüísticas. Los resultados indican que DeepL es el sistema más eficiente, y que el español y los textos administrativos presentan más dificultades para la TA. En términos de accesibilidad lingüística, se han observado más problemas a nivel léxico y estilístico. Si bien la TA todavía no genera contenido en LF de una calidad aceptable, nuestro estudio destaca el potencial de este tipo de herramientas, así como los retos que supone crear contenido multilingüe que sea accesible para todos.

MOTS CLEFS/KEYWORDS/PALABRAS CLAVE

traduction automatique, accessibilité, facile à lire et à comprendre, DQF-MQM, étude comparative
machine Translation, accessibility, easy-to-read, DQF-MQM, comparative study
traducción automática, accesibilidad, lectura fácil, DQF-MQM, estudio comparativo

The role of genius is not to complicate the simple, but to simplify the complicated.

(Criss Jami 2015 : 94)

1. Introduction

L'accès à l'information est un droit universel, essentiel pour le fonctionnement démocratique des sociétés et pour le bien-être de chaque individu. Les Nations unies promeuvent ce principe dans leur Convention relative aux droits des personnes handicapées[1] et encouragent les États partis à reconnaître et à éliminer les barrières à l'information et à la communication. Au cours des dernières années, plusieurs initiatives ont ainsi été lancées par la communauté internationale, comme la Journée internationale de l'accès universel à l'information[2], le Programme Information pour tous (PIPT)[3] ou encore, sur le plan européen, la publication en 2019 de la nouvelle « Directive du Parlement européen et du Conseil relative aux exigences en matière d'accessibilité applicables aux produits et services »[4]. Rendre les informations accessibles implique d'en assurer l'accès au plus grand nombre, de promouvoir le contenu libre et l'interopérabilité, de favoriser le multilinguisme et de tenir compte des besoins spécifiques des personnes handicapées. Cet article s'inscrit dans ce cadre et propose l'utilisation de la traduction automatique (TA) pour la production de textes facilement compréhensibles.

Dans le contexte de la communication accessible, la langue facile (LF) est souvent considérée comme un véritable instrument d'inclusion (Maaß 2020). Concrètement, la langue facile à lire et à comprendre (FALC) permet la simplification des textes écrits (BFEH 2019) et combine des recommandations linguistiques, comme l'utilisation de mots fréquents et de phrases courtes, avec d'autres règles typographiques ou de mise en page, qui concernent par exemple le contraste de couleurs ou le recours à des images pour illustrer les idées présentes dans le texte. L'objectif de cette étude est de mesurer l'impact de différents systèmes de TA, offerts au grand public, sur la qualité et l'accessibilité linguistiques de textes cibles, rédigés suivant les règles du FALC en langue source[5]. Nos hypothèses de départ sont que, malgré la simplification du texte source et les progrès de la TA, celle-ci ne peut pas encore être utilisée pour produire des textes cibles accessibles en fonction de la paire de langues (de même famille ou de famille différente) et du domaine.

Dans la suite de cet article, nous décrirons d'abord le contexte de cette recherche, ainsi que les raisons qui nous ont conduits à mener une étude sur la performance de la TA, en relation avec le FALC (section 2). Nous définirons ensuite plus en détail les différentes formes de LF et passerons en revue les études déjà publiées sur la TA et l'accessibilité (section 3). Cet état de l'art sera suivi de la méthodologie (section 4) et des principaux résultats de l'étude (section 5). L'article se terminera par les conclusions tirées de notre étude, ses limites et quelques perspectives (section 6).

2. Contexte et motivation

2.1. Création des ressources en FALC : défis et opportunités

Aujourd'hui, malgré le potentiel de la langue FALC pour garantir une communication claire pour tous, très peu de ressources sont encore disponibles. En l'absence de chiffres précis, cette faible prévalence se reflète à différents niveaux. Par exemple, nous ne trouvons aucune indication sur l'utilisation de ressources FALC dans le rapport de l'Agence des droits fondamentaux de l'Union européenne sur le handicap et la situation migratoire dans ses frontières, publié en 2016 (FRANET 2016). Similairement, Luce (2018), dans son rapport sur les réfugiés, mentionne ne pas avoir trouvé de brochures ou de documents FALC relatifs à la demande d'asile. En France, tous les adultes français avec un handicap qui ont un tuteur légal ont finalement obtenu le droit de vote en 2019. Pourtant, lors de l'élection du Parlement européen de 2019, seul un quart des partis politiques (9 sur 36) auraient fourni une version FALC de leur programme, alors qu'ils avaient été encouragés à le faire[6].

Plusieurs facteurs expliquent cette pénurie de ressources FALC. Tout d'abord, leur production implique de nombreuses étapes (analyse du public cible, sélection des informations les plus importantes, simplification, formatage, tests et finalement validation par le public cible), ce qui a un impact négatif sur le *coût de production*. En Suisse par exemple, un traducteur facture au minimum 4 francs suisses par 55 caractères (environ une ligne de 8-11 mots) pour des textes standard, 5,50 francs suisses pour les textes plus complexes et un tarif personnalisé pour les documents techniques[7]. La production de contenu FALC nécessite aussi des ressources humaines importantes. Plus de 60 personnes avec un handicap intellectuel auraient par exemple participé au projet « Orsay facile » pour la publication de deux guides FALC pour le musée d'art d'Orsay (Lamotte et Therwath 2016). Ce facteur est particulièrement

important dans les pays multilingues comme la Suisse où l'accessibilité doit être garantie dans plusieurs langues officielles, tout en assurant le même contenu sémantique du message.

D'autres facteurs plus politiques interviennent aussi, comme la *priorité accordée à l'accessibilité* sur le plan gouvernemental ou le *manque de formation et de sensibilisation* aux normes d'accessibilité existantes. Les stratégies en faveur des personnes handicapées font en général référence aux normes d'accessibilité du Web et à la nécessité de fournir des informations dans des formats faciles à lire, mais ne prescrivent pas de mesures concrètes[8]. Elles contestent même parfois les initiatives en vigueur. Par exemple, un rapport systématique sur le thème de la langue FALC pour les personnes avec un handicap intellectuel, réalisé en 2013-2014, n'inclut pas les règles européennes publiées par Inclusion Europe (2009) et critique même les lignes directrices existantes. Il fait état d'un manque de transparence par rapport à la méthodologie utilisée pour les développer, de l'absence d'une hiérarchisation des règles en fonction de leur impact sur l'accessibilité et d'incohérences des règles (Sutherland et Isherwood 2016). Malgré ces défis, un nombre croissant d'agences de traduction et de groupes de représentation des personnes handicapées, comme l'UNAPEI[9] en France, se spécialisent dans la production et la traduction de publications FALC et travaillent ensemble à la création d'outils informatiques pour aider à rédiger des textes conformes aux règles de la LF. Le « FALC Assistant »[10], par exemple, est un outil en ligne qui facilite la création de textes accessibles, vérifie la conformité des phrases et montre des améliorations possibles. Si ce type d'outil pouvait être utilisé en combinaison avec des technologies de la traduction, comme la traduction automatique, il serait possible de diminuer le coût de la production de contenu FALC et de faire face à la pénurie des ressources.

2.2. La traduction automatique : progrès et nouvelles perspectives

La traduction automatique neuronale (TAN) a révolutionné le domaine de la TA ces dernières années. Comme la TA statistique, elle se fonde sur des corpus, mais ceux-ci sont exploités différemment grâce à l'apprentissage profond, qui permet la représentation du sens des mots et des phrases sous la forme d'une représentation numérique (appelée plongements, *word embeddings*) (Forcada 2017). Contrairement aux systèmes statistiques, son fonctionnement repose sur une première étape d'encodage, où le plongement de la phrase est extrait. Ensuite, l'information est décodée à partir de cette représentation sémantique (Koehn 2009/2017). Comme la traduction se fait ainsi à partir de la représentation du sens de la phrase, la TAN aurait tendance à produire des paraphrases (Neubig, Morishita *et al.* 2015). Le résultat brut de la TA contient aussi moins de fautes grammaticales (la traduction est plus fluide), mais peut produire plus d'erreurs sémantiques (Forcada 2017). Des études ont également montré que la TAN produit plus d'omissions et d'ajouts par rapport à la TA statistique (Castilho, Moorkens *et al.* 2017). Il s'avère donc intéressant de voir si la TAN préserve ou pas la qualité du texte FALC et la conformité aux règles FALC. La TAN pourrait en effet augmenter la complexité d'une phrase simple en ajoutant des mots superflus ou en choisissant un lexique ou des structures à éviter.

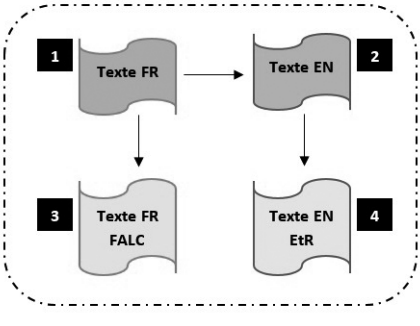
Peu d'études se sont cependant intéressées à la TA en relation avec la langue facile. Le projet « Simplifications des Langues Écrites (SIMPLES) » s'appuie sur un

outil d'apprentissage profond pour la création du contenu FALC. Comme dans notre étude, cette initiative française part de l'hypothèse que, si la production de textes FALC pouvait être facilitée par la technologie, un plus grand nombre de documents FALC serait mis à la disposition des personnes qui en ont besoin (Chehab, Holken *et al.* 2019). Le projet propose de combiner un outil en ligne (« LIREC ») pour la production et l'évaluation des textes accessibles avec un outil de simplification automatique qui tient compte des règles de la langue FALC (*ibid.*). Cette deuxième composante est toujours en cours de développement en collaboration avec une équipe de SYSTRAN, un fournisseur de technologies de TA[11]. Il n'est cependant pas question ici de traduction interlinguistique.

Les articles qui font référence à des technologies de la traduction dans le contexte de la langue FALC s'intéressent plutôt à leur application pour la traduction intralinguistique. Kapler *et al.* (2013) se sont focalisés sur la création de corpus monolingues pour faciliter la TA statistique de l'allemand vers le *Leichte Sprache*. Pour Hansen-Schirra, Nitzke *et al.* (2020), la TAN pourrait traduire d'une langue standard vers la langue facile correspondante, si des corpus parallèles étaient disponibles. Les auteurs notent qu'il faudrait garder des segments vides (sans équivalent) dans le corpus d'entraînement, étant donné que les traductions en langue FALC comportent souvent des informations qui diffèrent du texte source. Le système pourrait ainsi apprendre que certains segments ne sont pas nécessaires dans le texte cible et ne pas les proposer lors de la traduction (*ibid.*). Il n'est cependant pas encore établi que ce type d'approche puisse être mis en place avec des corpus parallèles bilingues de manière à faciliter la TA interlinguistique des textes FALC et simplifier les étapes de traduction (en permettant directement une traduction de la langue source à la langue cible simplifiée, avec un passage direct de l'étape 3 à 4 sur la figure 1). Vu le peu de corpus bilingues existants avec ces caractéristiques et la difficulté d'en créer à cause de la pénurie de textes FALC, il nous semble donc utile d'explorer d'abord le potentiel des systèmes de TAN génériques, disponibles pour le grand public.

Figure 1

Étapes de création de textes FALC dans un contexte multilingue : exemple FR – EN



3. État de l'art : Langue facile, accessibilité et traduction (automatique)

Dans cette section, nous approfondissons les caractéristiques de la langue FALC. Nous parlerons plus en détail des groupes de population cibles et des règles existantes (3.1) et nous expliquerons comment cette étude contribue à l'état de l'art sur la traduction automatique en lien avec le FALC (3.2).

3.1. La langue facile : définition, variantes et règles

Bien que la langue facile (LF, *easy language* en anglais) et la communication claire (CC, *plain language* en anglais) soient parfois utilisées de manière interchangeable dans le contexte de la communication accessible, il s'agit de variétés linguistiques différentes. D'après Maaß (2020, 3 :12-15), la LF serait la variété la plus compréhensible du langage naturel, tandis que la CC, moins stigmatisante, serait plus acceptée par la société (*ibid*.). La LF et la CC ciblent également des groupes différents. La LF est avant tout destinée aux personnes avec un handicap intellectuel, mais d'autres groupes peuvent aussi en bénéficier, comme les apprenants de langues étrangères (avec ou sans handicap), les personnes faiblement alphabétisées, les malentendants ou les personnes âgées (Hansen-Schirra et Maaß 2020). La CC aurait au contraire été pensée pour permettre l'accès au contenu spécialisé par le grand public (*ibid*.), mais certains pays la considèrent aussi comme un instrument de communication inclusive (Maaß 2020).

Plus concrètement, la langue facile à lire et à comprendre (FALC) est une variante de la LF, normalement utilisée dans les cas où l'information est lue, même si cette restriction peut sembler artificielle (du texte écrit disponible en ligne est parfois rendu accessible oralement, par exemple par la synthèse vocale ou des enregistrements audio (*ibid*. 52-56). Le FALC et la CC diffèrent aussi au niveau des consignes de rédaction ; les règles pour rendre les contenus FALC sont plus strictes, surtout au niveau du formatage du texte, qui a aussi une fonction sémantique. Par exemple, la CC ne propose pas de passer à la ligne après chaque idée (Matausch et Nietzio 2012).

Bien que le FALC ne figure généralement pas dans la littérature sur le thème du langage contrôlé (LC), les textes FALC pourraient être considérés comme une forme de LC (Kuhn 2014, 123), même si plusieurs auteurs ont mis en évidence le manque de preuves empiriques en ce qui concerne le processus de définition de ces restrictions linguistiques (Sutherland et Isherwood 2016). Chaque ensemble de règles est en effet propre à une seule langue ; par exemple, l'*Easy to Read* a des règles légèrement différentes de celles du FALC (qui diffère à son tour des autres variétés : *Lectura fácil* [espagnol], *Leichte Sprache* [allemand], *Leitura fácil* [portugais], *Facile da leggere* [italien], etc.). Le FALC impose aussi des restrictions pour ce qui est du lexique (il ne faut pas utiliser de mots difficiles, de mots étrangers, d'abréviations, ni de contractions), ainsi qu'en ce qui concerne la syntaxe (par exemple, il faut utiliser le présent, la voix active et éviter les pronoms ambigus). Aujourd'hui, la communauté académique consacre de plus en plus d'efforts au développement de règles d'accessibilité linguistique fondées sur des preuves scientifiques (cf. Chapitre 3.2 dans Maaß (2020) pour l'allemand). Cependant, les consignes pour la langue FALC proposées à une échelle européenne par Inclusion Europe (2009) dans 16 des 24 langues officielles de l'UE sont toujours considérées comme une référence valide dans plusieurs pays pour

la création de contenus accessibles à tous. Dans le document, les règles sont organisées en fonction du type d'informations (écrites, électroniques, vidéo et audio), et il propose un ensemble de règles générales applicables à ces quatre groupes. Nous présenterons plus en détail les règles que nous avons sélectionnées pour cette étude dans la section 4.3.1.

3.2. Études sur l'accessibilité linguistique en lien avec les langages contrôlés

Différentes études ont été réalisées sur l'accessibilité linguistique et le langage contrôlé. Certains auteurs ont voulu déterminer quels facteurs ont un impact sur la lisibilité et la compréhension des textes par les personnes handicapées et d'autres populations cibles des langages contrôlés, par exemple les personnes avec une déficience intellectuelle légère (DIM) (Feng, Elhadad *et al.* 2009) ou les individus avec le Trouble du spectre de l'Autisme (TSA) (Yaneva et Evans 2015). D'autres études se sont concentrées sur la définition de standards pour les textes simplifiés accessibles (Štajner, Mitkov et Corpas Pastor 2015 ; Yaneva 2015), le développement d'outils d'évaluation automatique de la lisibilité (Feng, Elhadad *et al.* 2009), la création d'outils pour la simplification automatique de textes (SAT) (Štajner, Mitkov et Corpas Pastor 2015 ; Yaneva 2015) ou ont cherché à déterminer si les formules de lisibilité existantes sont capables ou non d'évaluer les résultats de la SAT (Štajner, Mitkov et Corpas Pastor 2015).

Plus concrètement, la lisibilité des informations médicales a fait l'objet de nombreuses recherches et a conduit à l'élaboration de plusieurs corpus simplifiés. Grabar et Cardon (2018) ont construit CLEAR, un corpus comparable de traductions intralinguistiques en français, composé d'entrées d'encyclopédies, de textes de littérature de jeunesse ou tirés de Wikipédia pour le grand public, de notices de médicaments destinées aux professionnels de la santé avec leurs équivalents pour le grand public, ainsi que des revues techniques et simplifiées de Cochrane. Même si ce corpus ne contient pas de textes avec le logo FALC, il témoigne toutefois de la nécessité de fournir des informations de santé simples à tous les non-professionnels, y compris un formulaire facile à lire pour les adultes handicapés. Le travail de Rossetti (2019) se concentre également sur le corpus de résumés de Cochrane en *plain language,* qu'elle a utilisé pour explorer la lisibilité et la compréhensibilité de textes simplifiés traduits automatiquement dans le domaine de la santé.

Felici et Griebel (2019) ont étudié dans quelle mesure les règles de la communication claire sont respectées dans les textes administratifs suisses dans trois des quatre langues officielles de la Suisse (allemand, français et italien), ainsi que le rôle que la traduction humaine joue dans la communication accessible dans les pays multilingues. Elles ont constaté que, bien que l'utilisation d'un langage clair soit considérée comme une priorité dans la communication institutionnelle, dans la pratique, les notices d'assurance suisses sont loin d'avoir une lisibilité optimale. Kaplan, Rodríguez Vázquez *et al.* (2019) se sont aussi centrées sur le contexte suisse et le domaine administratif, mais contrairement à Felici et Griebel (2019), elles ont mis l'accent sur le FALC et la TA. Dans leur étude exploratoire, elles ont analysé la qualité de la TA et la conformité aux règles d'Inclusion Europe (2009) pour la paire français-anglais, avec trois systèmes de TA : DeepL, Google Translate et Yandex. Notre travail fait suite à cette étude exploratoire, en suivant une méthodologie simi-

laire (voir section 4), mais elle en élargit la portée, en ajoutant différentes paires de langues et d'autres domaines.

4. Méthodologie

4.1. Objectifs, questions de recherche et hypothèses

Cette étude vise à déterminer si la TA est un outil adapté pour produire des textes cibles accessibles à partir de textes sources en FALC. Notre hypothèse de départ était que la TA devrait préserver le niveau d'accessibilité linguistique du texte source et sa qualité dans la langue cible, une langue plus compréhensible pour l'homme devant l'être aussi pour la machine.

Pour ce faire, nous avons conçu une étude en deux étapes qui repose d'abord sur une analyse textuelle de la qualité des textes traduits automatiquement (étape I), puis sur une évaluation de la compréhensibilité par des utilisateurs cibles (étape II). Nous décrivons ici les résultats de la première étape, laquelle mesure l'impact de la TA sur deux variables dépendantes (VD) : la qualité de la traduction (VD1) et sa conformité aux règles FALC (VD2). Pour explorer différents scénarios de production textuelle, trois variables indépendantes (VI) ont été manipulées dans l'étude : le système de TA (VI1), le domaine (VI2) et la paire de langues (VI3). Sur la base de ces variables, nous avons défini les questions de recherche et les hypothèses suivantes :

1) Question de recherche QR1. Quelle est la qualité finale de la TA de textes FALC ?
   a. H1.1 Le nombre d'erreurs trouvées dans le texte cible varie selon le système de TA.
   b. H1.2 Le type d'erreurs trouvées dans le texte cible varie selon le système de TA.
   c. H1.3 Le nombre d'erreurs trouvées dans le texte cible varie selon le domaine du texte.
   d. H1.4 Le type d'erreurs trouvées dans le texte cible varie selon le domaine du texte.
   e. H1.5 Le nombre d'erreurs trouvées dans le texte cible dépend de la paire de langues.
   f. H1.6 Le type d'erreurs de traduction dans le texte cible varie en fonction de la paire de langues.

2) Question de recherche QR2. La TA préserve-t-elle l'application des règles FALC ?
   a. H2.1 Le nombre de violations aux règles FALC dans le texte cible varie selon le système de TA.
   b. H2.2 Le type de règles FALC violées dans le texte cible varie selon le système de TA.
   c. H2.3 Le nombre de violations aux règles FALC dans le texte cible varie selon le domaine du texte.
   d. H2.4 Le type de règles FALC violées dans le texte cible varie selon le domaine du texte.
   e. H2.5 Le nombre de violations des règles FALC dans le texte cible dépend de la paire de langues.
   f. H2.6 Le type de règles FALC violées dans le texte cible varie en fonction de la paire de langues.

4.2. Variables indépendantes et corpus d'évaluation

4.2.1. Systèmes de TA, domaines et langues sélectionnées pour l'étude

Trois systèmes de traduction automatique (TA) gratuits et en ligne, à base de corpus (VI1), ont été utilisés pour l'étude : DeepL, Google Translate et Yandex. Alors que les deux premiers s'inscrivent dans le paradigme de la traduction automatique neuronale (TAN), Yandex utilise toujours un modèle hybride, qui combine la traduction automatique statistique (TAS) et neuronale (TAN). Au moment de l'étude, les systèmes étaient tous compatibles avec les quatre paires de langues étudiées (VI2), l'anglais, l'allemand, le farsi et l'espagnol, sauf DeepL, qui ne proposait pas le farsi comme langue cible. Les traductions ont été produites le 4 août 2020. Contrairement aux moteurs de TA spécialisés, DeepL, Google Translate et Yandex sont tous considérés comme des systèmes génériques, d'où la décision de les tester avec des textes de trois domaines différents : politique, médical et administratif (VI3).

4.2.2. Corpus d'évaluation

Le corpus d'évaluation est composé de trois textes qui appartiennent chacun à un domaine différent (politique, médical et administratif), rédigés en FALC et classés comme tels par leurs éditeurs (voir annexe 1). Pour minimiser le risque de biais liés à des différences concernant l'application des règles, il a été décidé de sélectionner des textes qui (i) ont été rédigés en Suisse par des organismes publics officiels suisses ou des institutions reconnues sur le plan national et (ii) sont estampillés avec des logos FALC, approuvés par ces institutions. Nous présupposons donc, comme le démontrent certaines études, que les textes FALC produits par des humains et garantis par un logo sont conformes aux lignes directrices suivies (Nietzio, Scheer *et al.* 2012 ; Yaneva 2015).

Il faut souligner que les textes sélectionnés ne sont pas tous des adaptations de documents parallèles non FALC : le texte 1 est un document isolé qui, au moment de l'étude, n'existait que dans sa version FALC. Tous les textes étaient aussi disponibles en ligne au moment de la réalisation de l'étude, soit en format PDF, soit en format Web. Le tableau 1 récapitule les caractéristiques du corpus d'évaluation.

Tableau 1
Caractéristiques des corpus de l'étude

Texte	Nombre de segments	Nombre de mots
1- Politique	283	1958
2- Médical	223	1572
3- Administratif	133	1399
Total	639	4929

Ces différents textes ont ensuite été traduits dans les quatre langues cibles, soit en anglais (EN), en allemand (DE), en farsi (FA) et en espagnol (ES), avec les trois systèmes de TA mentionnés à la section 4.2.1. Nous avons ainsi obtenu un corpus d'évaluation de 7029 segments, avec 1917 segments en allemand, anglais et espagnol, ainsi que 1278 segments en farsi. Comme les textes pouvaient contenir des énumérations ou des titres, les segments ne contenaient pas nécessairement des phrases complètes. La règle de segmentation suivie dans notre étude était l'existence d'un point ou d'un retour à la ligne manuel[12].

4.3. Variables dépendantes

4.3.1. Règles FALC

Pour évaluer les résultats de la TA en matière d'accessibilité linguistique, nous avons pris comme référence les normes européennes visant à rendre l'information facile à lire et à comprendre (Inclusion Europe 2009) (voir section 3). Bien qu'il s'agisse de règles de base formulées pour toutes les langues communautaires et qui ne prennent pas en compte en profondeur les spécificités linguistiques, nous considérons qu'elles sont

Tableau 2

Règles FALC (Inclusion Europe 2009) sélectionnées pour l'étude

Catégorie	N° règle*	Règle
Générale	R01	Utilisez toujours le bon langage qui correspond aux personnes qui utiliseront vos informations.
	R02	N'utilisez pas de mots difficiles. Si vous devez utiliser des mots difficiles, il faut les expliquer clairement.
Mots	R03	Utilisez des mots faciles à comprendre, c'est-à-dire des mots que les gens connaissent bien.
	R04	N'utilisez pas de mots difficiles. Si vous devez utiliser des mots difficiles, expliquez-les clairement.
	R05	Utilisez le même mot pour parler de la même chose dans tout le document.
	R06	N'utilisez pas des idées difficiles comme des métaphores.
	R07	N'utilisez pas de mots d'une langue étrangère.
	R08	Évitez d'utiliser des initiales ou des abréviations. Utilisez le mot en entier lorsque vous le pouvez.
	R09	Essayez de ne pas utiliser de pourcentages ni de grands nombres.
	R10	Faites attention à l'utilisation des pronoms. Vérifiez qu'il est toujours clair de qui ou de quoi parle le pronom. Si ce n'est pas clair, alors utilisez le nom de la personne ou de la chose en entier.
	R11	Essayez d'éviter les caractères spéciaux.
	R12	N'utilisez pas de nombres ordinaux ou avec des suffixes.
	R13	Écrivez les nombres en chiffres et non en toutes lettres. N'utilisez jamais de chiffres romains.
	R14**	N'utilisez pas de contractions (par exemple, « don't » en anglais).
	R15	Lorsque c'est possible, écrivez les dates en entier.
Phrase	R16	Parlez directement aux gens. Utilisez des mots comme « vous ».
	R17	Utilisez des phrases positives plutôt que des phrases négatives quand vous le pouvez.
	R18	Utilisez des phrases actives plutôt que des phrases passives quand vous le pouvez.
	R19	Utilisez une ponctuation simple.
	R20	Utilisez le présent quand vous le pouvez.
Structure	R21	Commencez toujours une nouvelle phrase sur une nouvelle ligne.
	R22	Ne séparez jamais 1 mot sur 2 lignes.
	R23	Faites des phrases courtes. Lorsque c'est possible, 1 phrase devrait tenir sur 1 ligne. Si vous devez écrire 1 phrase sur 2 lignes, coupez la phrase à l'endroit où vous feriez une pause lorsque vous lisez à voix haute.
	R24	Utilisez des titres clairs et faciles à comprendre. Les titres doivent vous expliquer ce qui va suivre juste après.
	R25	Utilisez des points (puces, ou des numéros) pour faire une liste. Une liste de mots séparés par des virgules n'est pas facile à lire.
*** Mots	R26	N'utilisez pas de mots longs. Si vous devez utiliser des mots longs, séparez les mots par un trait d'union.

* Les noms de catégories et les numéros de règles attribués aux règles choisies ont été définis pour cette étude et ne correspondent pas nécessairement à ceux utilisés par Inclusion Europe (2009).

** Cette règle ne s'applique pas dans le cas du français.

*** Règle pour l'allemand uniquement.

adéquates dans le cas de cette étude comparative et interlinguistique pour garantir une validité méthodologique. Comme notre objectif était de tester la performance linguistique de la TA sur des textes FALC, nous avons écarté certaines règles non pertinentes dans notre contexte. Ceci concerne notamment les règles relatives à la conception et au format des documents, par exemple celles liées au type et à la taille de la police, aux couleurs ou à l'utilisation des images. Nous avons finalement choisi 25 règles parmi les règles générales et celles propres aux informations écrites (voir tableau 2), qui sont toutes applicables à toutes les langues étudiées, à l'exception de la règle 14, propre à l'anglais, sur les contractions, ainsi que la règle 26, propre à l'allemand, qui concerne les noms composés. Ces derniers sont en effet bien connus pour être complexes du point de vue sémantique (« Kranken-haus » [hôpital] ou « Stimm-ausweis » [carte de vote]) et doivent être séparés par un trait d'union dans les textes FALC[13].

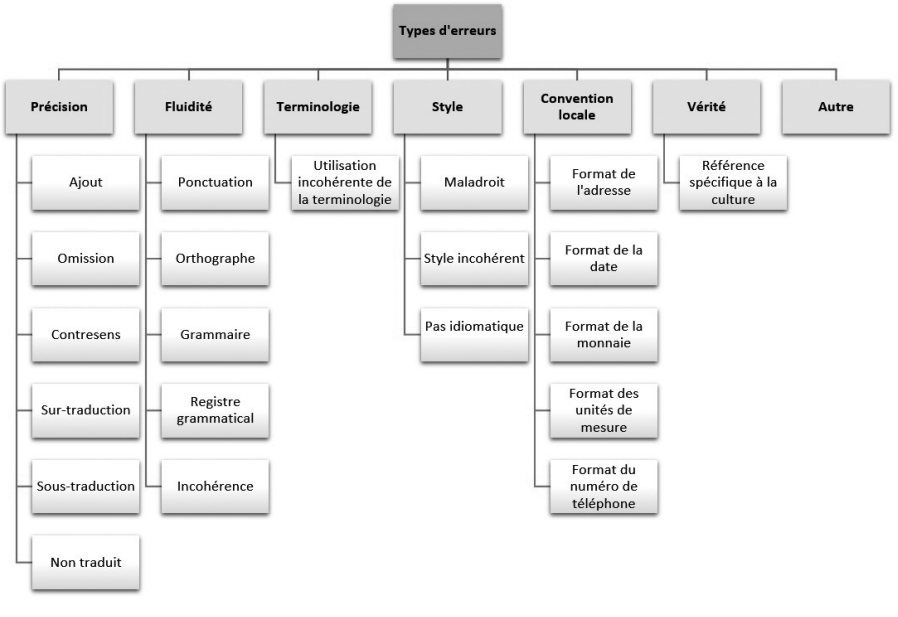
4.3.2. Qualité de la traduction

Afin d'évaluer la qualité de la TA, nous avons utilisé la typologie d'erreur harmonisée « Dynamic Quality Framework/Multidimensional Quality Metrics (DQF-MQM) », adoptée par TAUS[14]. Notre choix s'est porté sur ce modèle, car outre sa popularité auprès des professionnels du secteur privé, il a également été largement utilisé dans le cadre d'études didactiques pour effectuer une analyse manuelle systématique des erreurs de TA (Lommel, Popovic *et al.* 2014 ; Klubička, Toral *et al.* 2017). La typologie, qui peut être adaptée en fonction des besoins de l'utilisateur, propose un vocabulaire de référence pour chaque type d'erreur, organisé autour de 8 catégories principales et 32 sous-catégories. Pour chaque erreur, il est possible d'attribuer quatre niveaux de sévérité différents : critique, majeur, mineur et neutre.

Compte tenu des spécificités du corpus, nous avons adapté la typologie d'erreurs et n'avons conservé que les types d'erreurs pertinents pour nos objectifs de recherche,

Figure 2

Typologie d'erreurs DQF-MQM retenue pour l'étude



notamment nous avons écarté la catégorie d'erreurs « Conception », relative au formatage des documents, à l'affichage du texte, au codage des caractères et à l'utilisation de balises. De même, nous avons supprimé les types d'erreurs qui concernent les mémoires de traduction et les bases de données terminologiques. La figure 2 résume la typologie d'erreurs adoptée, avec 7 catégories et 21 sous-catégories.

4.4. Conception de l'évaluation et plan expérimental

Afin de mesurer l'impact de la TA sur les deux variables dépendantes définies, l'annotation s'est faite à deux niveaux (conformité aux règles FALC et qualité de la traduction). Celle-ci consistait à examiner individuellement chaque segment du corpus d'évaluation et à signaler tous les problèmes constatés par rapport à la typologie d'erreurs DQF-MQM (figure 2) et aux règles FALC (tableau 2). Les deux exercices d'annotation ont été menés à l'aide d'un classeur Microsoft Excel personnalisé.

Chaque annotation a été faite par deux annotateurs différents. Compte tenu de la nature de nos variables indépendantes et dépendantes, nous avons suivi un plan croisé, où *VI1. Système de TA* et *VI3. Domaine* ont été étudiés par une approche intragroupe et *VI2. Langue* par une approche intergroupe. Pour compenser ce que l'on appelle l'effet de séquence, qui pourrait conduire à des biais, l'ordre dans lequel les systèmes de TA choisis ont été évalués a été contrebalancé. En conséquence, chaque annotateur a évalué la sortie de la TA dans un ordre différent. Par exemple, dans le cas du Texte 1, l'annotateur A1 a commencé par Google (A), a continué avec DeepL (B) et a fini par Yandex (C), tandis que l'annotateur A2 a commencé par DeepL (A), a continué avec Yandex (B) et a fini par Google (C), comme l'illustre le tableau 3. De même, il faut noter que les noms des systèmes évalués n'ont pas été divulgués aux annotateurs, afin d'éviter tout biais dû à des préjugés individuels sur la performance générale de ces outils populaires.

Tableau 3
Plan expérimental

	Texte 1 : Politique			Texte 2 : Médical			Texte 3 : Administratif		
Annotateur/ Système TA	Google	DeepL	Yandex	Google	DeepL	Yandex	Google	DeepL	Yandex
A1	A	B	C	B	C	A	C	A	B
A2	C	A	B	A	B	C	B	C	A

Les annotateurs étaient libres de choisir par quelle analyse commencer (soit la conformité aux règles FALC, soit la qualité de la traduction). Ils ne disposaient pas d'un délai particulier pour finaliser l'annotation, mais il leur était conseillé de travailler sur un seul domaine par jour. Tous les annotateurs avaient des compétences en traduction et avaient déjà effectué des annotations d'erreurs de traduction dans le passé, soit dans un contexte universitaire, soit dans un contexte professionnel. Les deux annotateurs pour la langue anglaise avaient déjà de l'expérience avec les règles européennes pour les textes FALC et le modèle DQF-MQM en particulier. Néanmoins, avant l'expérience, ils ont tous été invités à se familiariser avec les règles FALC et les catégories d'erreurs de traduction choisies pour l'étude. En plus des documents complémentaires fournis, des phrases de test ont été incluses dans les

feuilles de travail Microsoft Excel pour permettre aux annotateurs de se former eux-mêmes.

Étant donné le nombre élevé de segments à examiner par chaque annotateur, nous leur avons demandé de signaler les erreurs au niveau du segment au lieu d'identifier la portée exacte de l'erreur, comme dans Lommel, Popovic *et al.* (2014). Toutefois, à des fins de validité et dans la mesure du possible, il a été demandé aux annotateurs de proposer un texte cible corrigé. Plus d'une erreur pouvait être signalée par segment dans les deux exercices d'annotation. Les annotateurs n'ont pas été payés pour la tâche, mais ont reçu un bon-cadeau à la fin de l'étude.

5. Résultats principaux

Une fois l'évaluation terminée, les annotations des deux exercices ont été extraites pour examiner la performance de la TA. Notre ensemble de données comprenait 14 058 segments (1 917 segments x 2 annotateurs x 3 langues [DE, EN, ES] + 1 278 segments x 2 annotateurs pour le FA), qui ont été révisés selon deux critères différents : la qualité de la traduction et la conformité aux règles FALC.

Lors de l'analyse des données collectées, nous avons calculé le type et le nombre d'erreurs en fonction des différentes variables prises en considération, ainsi que l'accord entre les annotateurs, avec le test Kappa (K) de Cohen (Cohen 1960). Dans les statistiques relatives au nombre d'erreurs, nous avons comptabilisé tous les segments incorrects avec au moins une erreur selon au moins l'un des annotateurs ; dans le cas de l'analyse du type d'erreurs, nous avons comptabilisé les erreurs uniques relevées par les deux annotateurs, en éliminant les doublons. Par exemple, les trois segments du tableau 4 ont été pris en compte dans l'analyse générale, même si l'annotateur A1 n'avait pas signalé d'erreur pour le segment S010. De même, deux erreurs ont été comptabilisées dans le segment S012, tandis que seulement une erreur a été prise en compte dans le cas du segment S043.

Tableau 4

Exemple de segments annotés par rapport à la violation des règles de la langue FALC

ID du segment / Annotateur	A1	A2
S010_EN_D1_TA1		R03
S012_EN_D1_TA1	R03 R04	R04
S043_EN_D1_TA1	R03	R03

Dans les sections suivantes, nous présentons les résultats principaux de l'étude, ce qui nous amènera à confirmer ou à rejeter nos hypothèses.

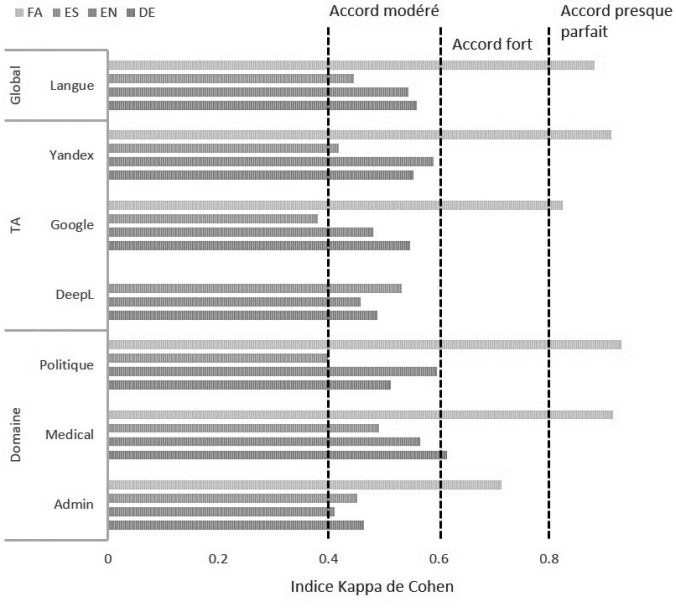
5.1. Qualité de la traduction

Cette section décrit les résultats relatifs à l'impact de la TA sur la qualité de traduction des textes FALC (VD1). La figure 3 montre l'accord entre les annotateurs de manière globale (par langue), par domaine et par système de TA. Pour le farsi, l'accord est fort (> 0,6) ou presque parfait (> 0,8). Pour les autres langues, l'accord est modéré (> 0,4), sauf pour la qualité de la traduction avec Google en espagnol et pour le

domaine politique, avec un indice κ légèrement inférieur (0,398 et 0,380 respectivement). Ces résultats vont de pair avec ceux généralement présentés dans la littérature (Lommel, Popovic *et al.* 2014 ; Klubička, Toral *et al.* 2017).

Figure 3

Accord entre les annotateurs par rapport à la qualité de la traduction selon l'indice Kappa de Cohen



5.1.1. Résultats par système de TA (VI1)

Le tableau 5 montre qu'aucun des systèmes ne traduit correctement les textes FALC. Parmi les trois systèmes évalués, DeepL a été le plus performant avec 50,29 % de segments (N = 964) incorrects et une moyenne de 1,61 erreur par segment avec au moins une erreur (0,81 erreur par rapport au nombre total de segments évalués). DeepL est suivi de Google Translate, avec 53,17 % de segments incorrects avec au moins une erreur. Yandex est le moins performant, avec seulement 28,36 % de segments sans erreurs.

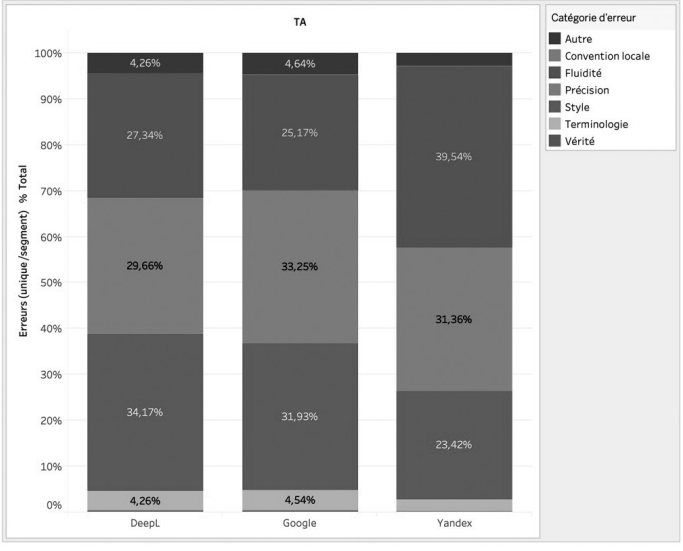
Tableau 5

Segments comportant des erreurs de traduction par système de TA

TA	Segments (total)	Segments incorrects (uniques)	% segments incorrects	Ratio (erreurs/ segment incorrect)	Erreurs (uniques/ segment)	Ratio (erreurs/ segment)	% segments sans erreurs (trad)
Yandex	2556	1831	71,64 %	1,58	2886	1,13	28,36 %
Google	2556	1359	53,17 %	1,56	2114	0,83	46,83 %
DeepL	1917	964	50,29 %	1,61	1551	0,81	49,71 %

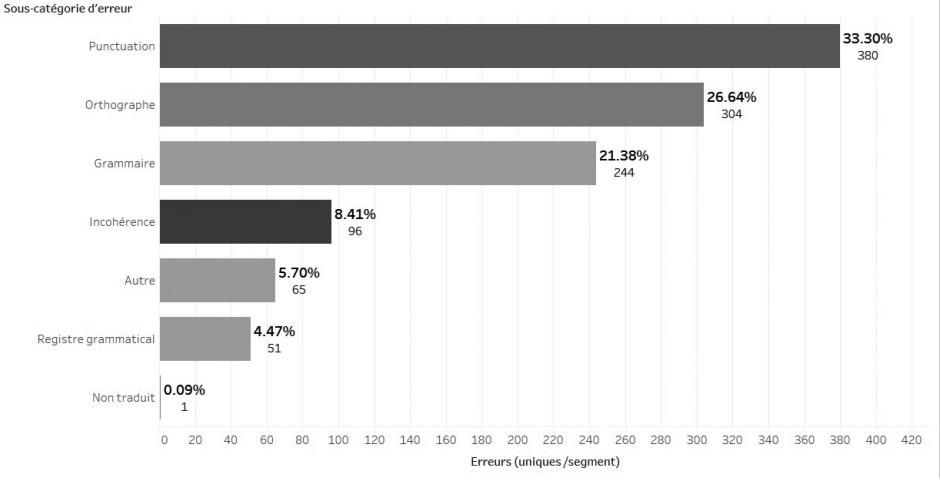
En ce qui concerne le type d'erreurs, la figure 4 montre les différences entre les systèmes neuronaux et le système hybride. Ce dernier totalise 39,54 % d'erreurs de fluidité, dont plus de la moitié correspondent à des fautes d'orthographe ou de ponctuation (voir figure 5). Dans les cas de DeepL et Google Translate, les problèmes de fluidité restent inférieurs à 30 %. La majorité des erreurs restantes appartiennent à deux autres catégories : précision et style.

Figure 4
Erreurs de traduction selon les catégories du modèle DQF-MQM par système de TA



Catégorie d'erreur	DeepL	Google	Yandex
Autre	4,26% (66)	4,64% (98)	2,77% (80)
Convention locale		0,14% (3)	0,07% (2)
Fluidité	27,34% (424)	25,17% (532)	39,54% (1141)
Précision	29,66% (460)	33,25% (703)	31,36% (905)
Style	34,17% (530)	31,93% (675)	23,42% (676)
Terminologie	4,26% (66)	4,54% (96)	2,60% (75)
Vérité	0,32% (5)	0,33% (7)	0,24% (7)
Total	**100,00% (1551)**	**100,00% (2114)**	**100,00% (2886)**

Figure 5
Distribution des erreurs de fluidité de Yandex



5.1.2. Résultats par domaine (VI2)

D'après nos observations, le domaine administratif a causé le plus de problèmes pour la TA, avec 67,19 % de segments incorrects (N = 983) et une moyenne de 1,57 erreur par segment incorrect. Il est intéressant de noter que, même si le texte médical a obtenu le pourcentage le plus bas de segments problématiques, le ratio d'erreurs par segment incorrect est le plus élevé (1,59), comme le montre le tableau 6.

Tableau 6

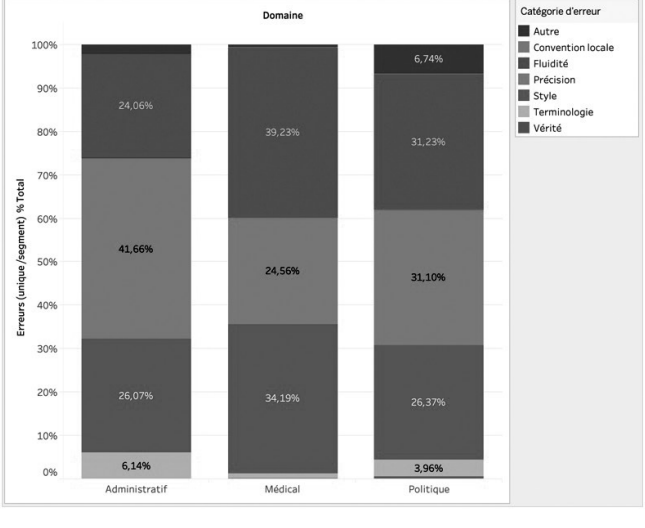
Segments comportant des erreurs de traduction par domaine

Domaine	Segments (total)	Segments incorrects (uniques)	% segments incorrects	Ratio (erreurs/ segment incorrect)	Erreurs (uniques/ segment)	Ratio (erreurs/ segment)	% segments sans erreurs (trad)
Administratif	1463	983	67,19 %	1,57	1576	1,06	32,81 %
Politique	3113	1896	60,91 %	1,57	2981	0,96	39,09 %
Médical	2453	1275	51,98 %	1,59	2024	0,83	48,02 %

Quand nous observons la distribution des erreurs par catégorie dans le domaine administratif, nous constatons qu'une grande partie des problèmes est liée à la précision (41,66 %, N = 644) (voir figure 6). Concrètement, presque 70 % des erreurs sont des contresens (voir figure 7). Pour le domaine médical, les erreurs de fluidité sont les plus nombreuses (39,23 %, N = 794). Enfin, la plupart des erreurs dans le domaine politique se répartissent dans trois catégories principales : fluidité (31,23 %, N = 931), précision (31,10 %, N = 927) et style (26,37 %, N = 786).

Figure 6

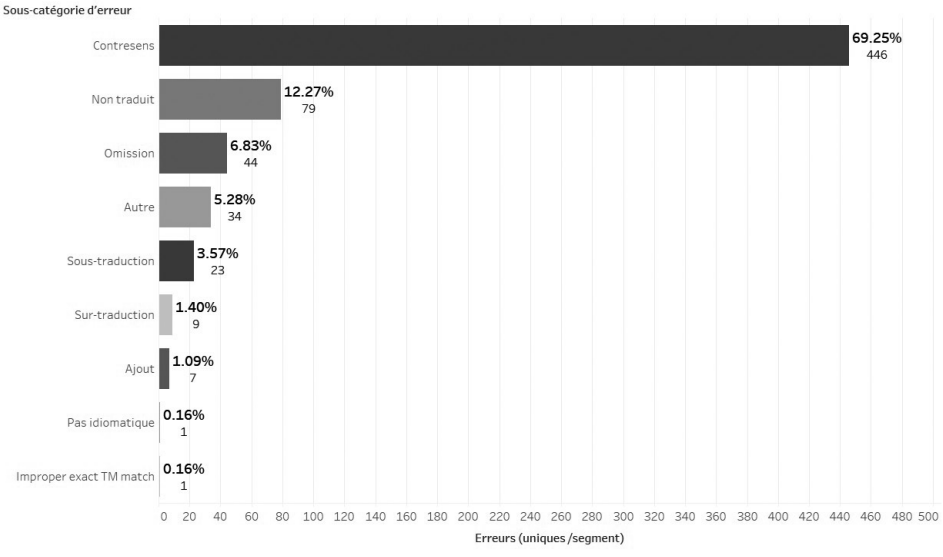
Erreurs de traduction selon les catégories du modèle DQF-MQM par domaine



Catégorie d'erreur	Administratif	Médical	Politique
Autre	2,07% (32)	0,54% (11)	6,74% (201)
Convention locale		0,10% (2)	0,10% (3)
Fluidité	24,06% (372)	39,23% (794)	31,23% (931)
Précision	41,66% (644)	24,56% (497)	31,10% (927)
Style	26,07% (403)	34,19% (692)	26,37% (786)
Terminologie	6,14% (95)	1,19% (24)	3,96% (118)
Vérité		0,20% (4)	0,50% (15)
Total	**100,00% (1546)**	**100,00% (2024)**	**100,00% (2981)**

Figure 7

Distribution des erreurs de précision du domaine administratif



5.1.3. Résultats par langue (VI3)

Finalement, en ce qui concerne la langue, l'étude montre que la qualité de la traduction est plus faible en espagnol que dans les autres langues, avec 79,19 % de segments problématiques, une moyenne de 1,80 erreur par segment incorrect et de 1,43 erreur par segment par rapport au nombre total de segments évalués. L'espagnol est suivi de l'allemand et du farsi (voir tableau 7). Avec 56,49 % de segments sans erreurs, l'anglais est la langue qui obtient les meilleurs résultats.

Tableau 7

Segments comportant des erreurs de traduction par langue

Langue	Segments (total)	Segments incorrects (uniques)	% segments incorrects	Ratio (erreurs/ segment incorrect)	Erreurs (uniques/ segment)	Ratio (erreurs/ segment)	% segments sans erreurs (trad)
ES	1917	1518	79,19 %	1,80	2732	1,43	20,81 %
DE	1917	1106	57,69 %	1,64	1810	0,94	42,31 %
FA	1278	696	54,46 %	1,22	851	0,67	45,54 %
EN	1917	834	43,51 %	1,39	1158	0,60	56,49 %

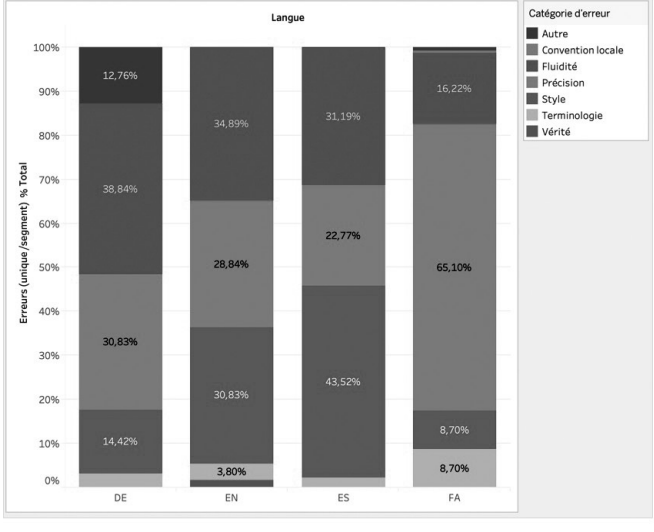
Quand les erreurs de traduction sont triées par catégorie, les différences entre les langues sont également notables, comme le montre la figure 8. Les problèmes de précision sont nombreux en farsi (65,10 %, N = 554), tandis qu'en espagnol, les erreurs de style sont les plus élevées (43,52 %, N = 1189). Les figures 9 et 10 illustrent comment ces erreurs sont distribuées par sous-catégorie. Nous pouvons observer que les problèmes de contresens sont les plus élevés en farsi (72,92 %) et que, en espagnol, les tournures non idiomatiques constituent 62,99 % des erreurs de style. En anglais, la plupart des erreurs sont distribuées en trois catégories principales : fluidité (34,89 %, N = 404), précision (28,84 %, N = 334) et style (30,83 %, N =357). En allemand, la plus

grande partie des erreurs concerne la fluidité. Des exemples pour les paires de langues FR-ES et FR-EN sont montrés dans le tableau 8. L'une des erreurs fréquentes est causée par des différences d'ordre des mots entre le français et l'allemand au niveau de la position du verbe. Si une phrase est coupée à des fins d'accessibilité (en raison des sauts de ligne par exemple), le système de TA place en effet le verbe dans le segment où il se trouve dans le texte source, comme dans cet exemple :

> *« Nous choisissons également les personnes qui vont aller au Conseil des États. »*
>
> „Wir wählen auch die Menschen, die gehen zu Ständerat."[15]

FIGURE 8

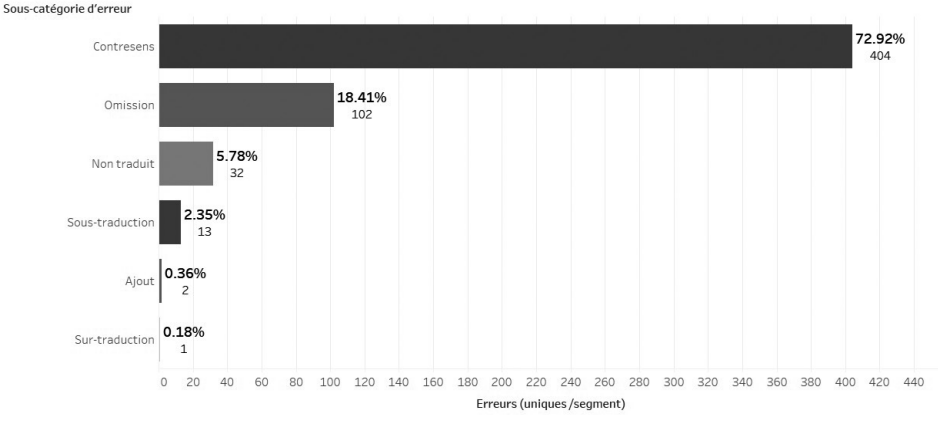
Erreurs de traduction selon les catégories du modèle DQF-MQM par langue



Catégorie d'erreur	DE	EN	ES	FA
Autre	12,76% (231)		0,26% (7)	0,71% (6)
Convention locale				0,59% (5)
Fluidité	38,84% (703)	34,89% (404)	31,19% (852)	16,22% (138)
Précision	30,83% (558)	28,84% (334)	22,77% (622)	65,10% (554)
Style	14,42% (261)	30,83% (357)	43,52% (1189)	8,70% (74)
Terminologie	3,15% (57)	3,80% (44)	2,27% (62)	8,70% (74)
Vérité		1,64% (19)		
Total	100,00% (1810)	100,00% (1158)	100,00% (2732)	100,00% (851)

FIGURE 9

Erreurs de précision en farsi



Il est intéressant de noter que 12,76 % (N = 231) des erreurs dans cette langue appartiennent à la catégorie « Autre ». En regardant les données plus en détail, nous constatons qu'il s'agit surtout de problèmes de terminologie, liés au *locale* (DE-DE par rapport à DE-CH) et à la traduction des abréviations, pour lesquels il n'y avait pas de sous-catégorie d'erreurs spécifique sous « Terminologie ».

Figure 10
Erreurs de style en espagnol

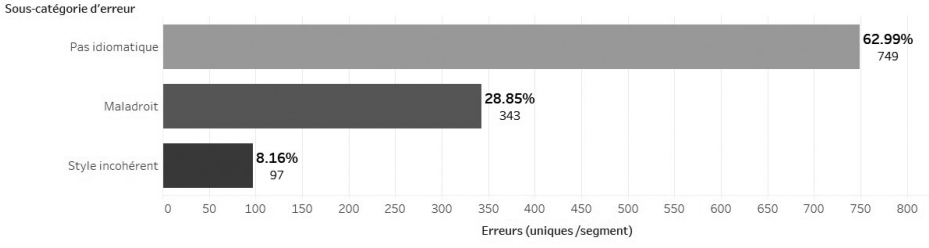


Tableau 8
Sélection d'exemples d'erreurs de traduction (paires de langues FR-EN et FR-ES)

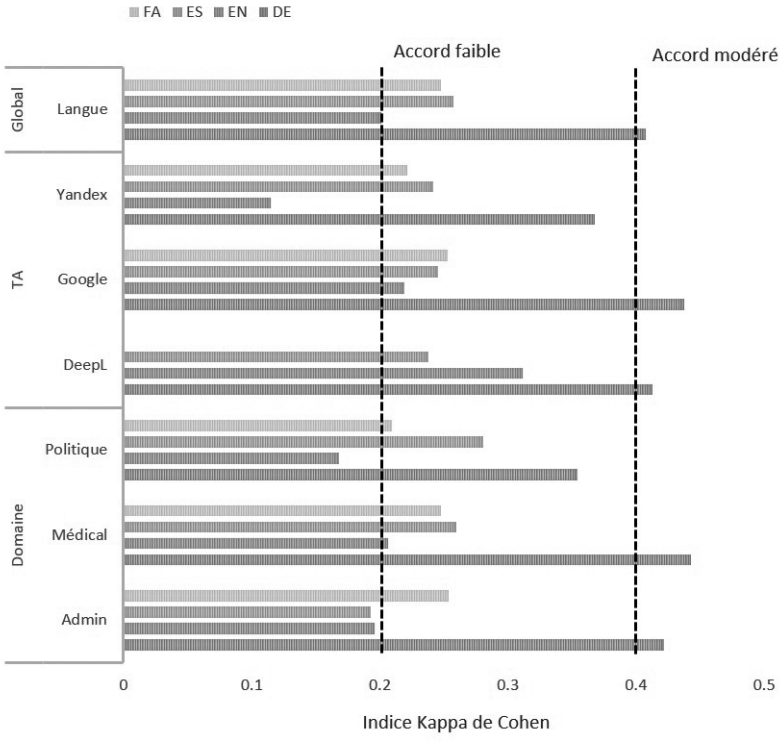
Paire de langues	Catégorie d'erreurs	Source	Traduction automatique
FR-ES	Précision – Contresens	C'est quoi le **bien** de l'enfant ?	¿Qué es lo **bueno** del niño ?
	Précision – Contresens	L'**APEA** peut par exemple décider que la famille doit aller dans un centre de conseil.	Por ejemplo, la **Asociación de Padres y Maestros** puede decidir que la familia vaya a un centro de asesoramiento.
	Style – Pas idiomatique	**Ils font ce qu'il faut** pour que l'enfant se porte bien.	**Hacen lo que tienen que hacer** para asegurarse de que el niño esté bien.
FR-EN	Précision – Contresens	L'APEA :	EAPC :
	Fluidité – Grammaire	• **évalue** la situation	• **assess** the situation
	Fluidité – Grammaire	• **décide** s'il faut nommer un curateur	• **decide** whether to appoint a curator

5.2. Niveau d'accessibilité linguistique : violation des règles de la langue FALC

Cette section décrit les résultats relatifs à l'impact de la TA sur l'accessibilité linguistique du texte cible, c'est-à-dire sur sa conformité par rapport aux règles FALC (VD2). La figure 11 montre l'accord entre les annotateurs de manière globale (par langue), par domaine et par système de TA. Dans la plupart des cas, l'accord est faible (> 0,2), sauf pour l'allemand, où l'accord est souvent modéré (> 0,4). L'indice κ est très faible (entre 0,1 et 0,2) en anglais pour les traductions faites par Yandex et le domaine politique.

Figure 11

Accord entre les annotateurs par rapport à l'accessibilité linguistique selon l'indice Kappa de Cohen



5.2.1. Résultats par système de TA (VI1)

D'un point de vue statistique, la performance de la TA par rapport à la proportion de segments incorrects en ce qui concerne la conformité aux règles d'accessibilité linguistique est meilleure que dans le cas de l'analyse de la qualité de la traduction. Comme l'indique le tableau 9, Yandex est toujours le système le plus faible, avec 52,78 % de segments problématiques et une moyenne de 1,35 erreur par segment incorrect (0,71 erreur par rapport au nombre total de segments évalués). Il est intéressant de noter que, même si DeepL montre les meilleurs résultats, avec 56,13 % de segments *a priori* accessibles, le ratio d'erreurs par segment incorrect est légèrement supérieur à celui de Yandex (1,36).

Tableau 9

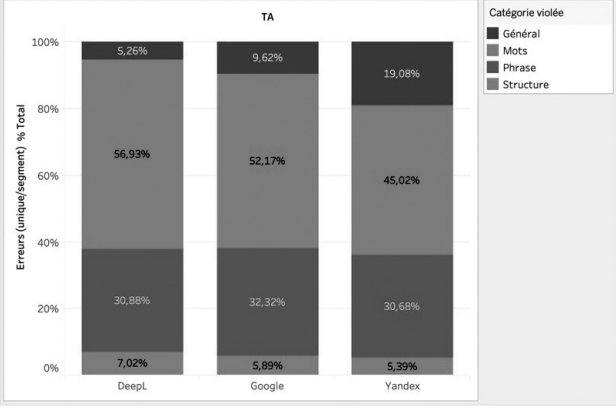
Segments comportant des erreurs d'accessibilité par système de TA

TA	Segments (total)	Segments incorrects (uniques)	% segments incorrects	Ratio (erreurs/ segment incorrect)	Erreurs (uniques/ segment)	Ratio (erreurs/ segment)	% segments sans erreurs (acc)
Yandex	2556	1349	52,78 %	1,35	1819	0,71	47,22 %
Google	2556	1154	45,15 %	1,28	1476	0,58	54,85 %
DeepL	1917	841	43,87 %	1,36	1140	0,59	56,13 %

En ce qui concerne le type de règles violées, les données collectées montrent à nouveau une différence notoire entre les systèmes de TAN et le système hybride Yandex (voir figure 12). En effet, les traductions produites par DeepL et Google Translate présentent plus de problèmes d'accessibilité sur le plan lexical (catégorie « Mots »), avec respectivement 56,93 % (N = 649) et 52,17 % (N = 770) de segments signalés. La figure 13 montre que, de manière générale, les règles les plus violées sont la règle R03, relative à l'utilisation de mots faciles à comprendre, et la R05, qui porte sur la cohérence terminologique. La règle R26 (propre à l'allemand) a également été fortement transgressée. Il est intéressant de noter que DeepL obtient de moins bons résultats avec cette règle-là, ainsi qu'avec la règle R14 (« N'utilisez pas de contractions »). Sur le plan de la phrase et de la structure, les résultats sont similaires dans les trois systèmes (voir figure 12).

Figure 12

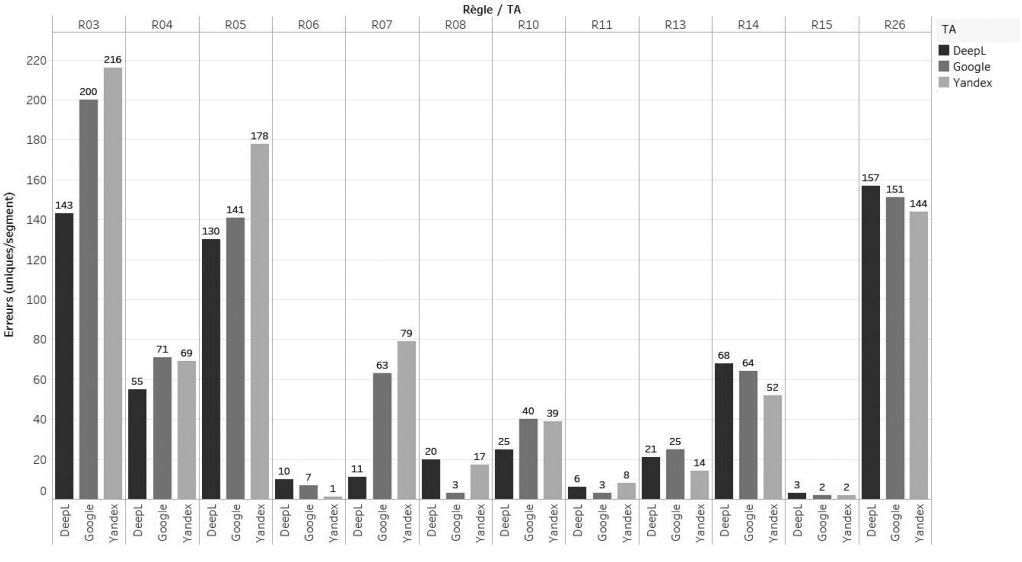
Erreurs d'accessibilité selon la catégorie de la règle violée par le système de TA



Catégorie violée	DeepL	Google	Yandex
Général	5,26% 60	9,62% 142	19,08% 347
Mots	56,93% 649	52,17% 770	45,02% 819
Phrase	30,88% 352	32,32% 477	30,68% 558
Structure	7,02% 80	5,89% 87	5,39% 98
Total	**100,00% 1140**	**100,00% 1476**	**100,00% 1819**

Figure 13

Erreurs d'accessibilité linguistique de la TA par règle (catégorie « Mots »)



5.2.2. Résultats par domaine (VI2)

Le texte administratif contient le plus grand nombre d'erreurs d'accessibilité (60,01 % des segments signalés, avec une moyenne de 1,53 erreur par segment). La performance de la TA par rapport à la violation des règles de la langue FALC dans les deux autres domaines (médical et politique) est assez similaire, comme le montre le tableau 10.

Tableau 10

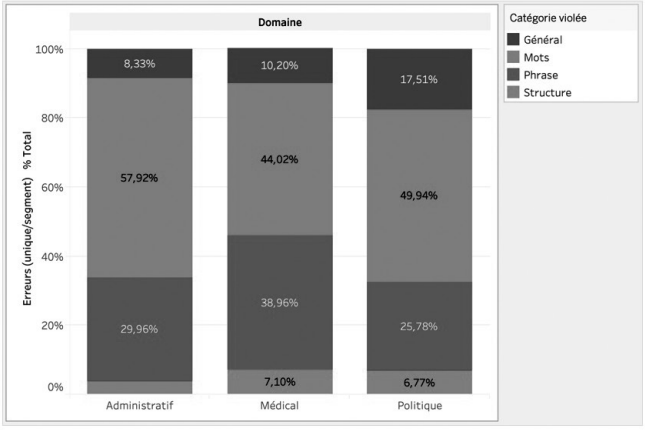
Segments comportant des erreurs d'accessibilité par domaine

Domaine	Segments (total)	Segments incorrects (uniques)	% segments incorrects	Ratio (erreurs/ segment incorrect)	Erreurs (uniques/ segment)	Ratio (erreurs/ segment)	% segments sans erreurs (acc)
Administratif	1463	878	60,01 %	1,53	1345	0,92	39,99 %
Médical	2453	1120	45,66 %	1,27	1422	0,58	54,34 %
Politique	3113	1346	43,24 %	1,24	1668	0,54	56,76 %

Les règles sous la section « Mots » sont les plus violées dans tous les domaines (voir figure 14). Le texte médical semble poser plus de problèmes sur le plan de la phrase (38,96 % des erreurs, N = 554) que les textes administratif et politique. Le texte politique contient plus de violations de règles sous la section « Général » (17,51 %, N = 292) que les autres domaines.

Figure 14

Erreurs d'accessibilité selon la catégorie de la règle violée par domaine



Catégorie violée	Administratif	Médical	Politique
Général	8,33% 112	10,20% 145	17,51% 292
Mots	57,92% 779	44,02% 626	49,94% 833
Phrase	29,96% 403	38,96% 554	25,78% 430
Structure	3,79% 51	7,10% 101	6,77% 113
Total	**100,00% 1345**	**100,00% 1422**	**100,00% 1668**

5.2.3. Résultats par langue (VI3)

Environ la moitié des segments dans le texte cible n'était pas accessible. L'espagnol comptabilise le plus grand nombre de violations aux règles FALC (1312 erreurs, avec une moyenne de 1,33 erreur par segment signalé), suivi du farsi, de l'allemand et de l'anglais (voir tableau 11). L'allemand obtient le ratio d'erreurs par segment problématique le plus élevé (1,48).

Tableau 11

Segments comportant des erreurs d'accessibilité

Langue	Segments (total)	Segments incorrects (uniques)	% segments incorrects	Ratio (erreurs/ segment incorrect)	Erreurs (uniques/ segment)	Ratio (erreurs/ segment)	% segments sans erreurs (acc)
ES	1917	988	51,54 %	1,33	1312	0,68	48,46 %
FA	1278	637	49,84 %	1,27	810	0,63	50,16 %
DE	1917	882	46,01 %	1,48	1302	0,68	53,99 %
EN	1917	837	43,66 %	1,21	1011	0,53	56,34 %

Si on compare les résultats par langue, domaine et système, la plus grande variation est observée dans la distribution des erreurs d'accessibilité par langue. Comme le montre la figure 15, la catégorie de règles la plus transgressée en allemand et en anglais est celle de type « Mots », avec respectivement 78,80 % et 59,45 % des erreurs. Le tableau 12 inclut quelques exemples dans le cas de la paire FR-EN (les mots problématiques et les erreurs sont soulignés en gras). En allemand, c'est la règle R26 qui est la moins respectée (voir figure 16). Cette règle, qui prescrit de séparer les éléments lexicaux dans les substantifs composés, n'est en effet pas appliquée par les systèmes TA qui ne sont pas entraînés avec des corpus de textes en LF et qui proposent donc la composition « standard » (par exemple « Wahlcouvert » au lieu de « Wahl-couvert »).

Dans le cas de l'espagnol, la plupart des problèmes ont été signalés sur le plan de la phrase (65,55 %), notamment la règle R16 (« Parlez directement aux gens ») n'a pas été respectée 555 fois (44,05 % du total d'erreurs de « Phrase », voir figure 17). Des exemples sont montrés dans le tableau 12. Les problèmes au niveau de la phrase sont moins fréquents en farsi, où les erreurs identifiées sont de caractère général (43,33 %) ou liées aux mots (49,88 %).

Figure 15

Erreurs d'accessibilité selon la catégorie de la règle violée par langue

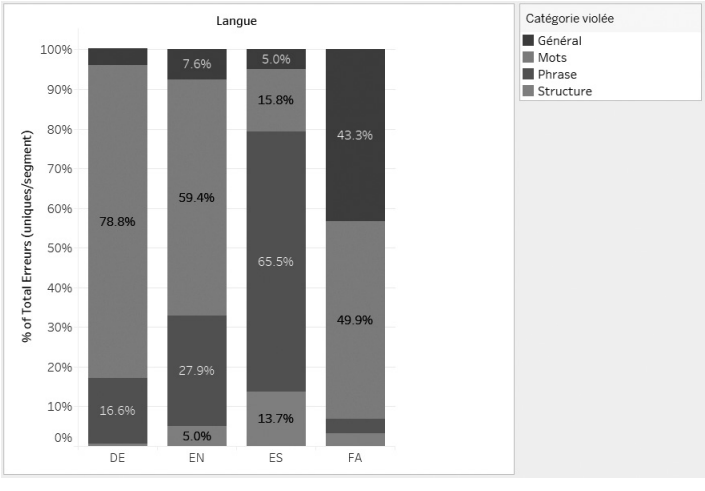


FIGURE 16

Erreurs en allemand par règle (catégorie « Mots »)

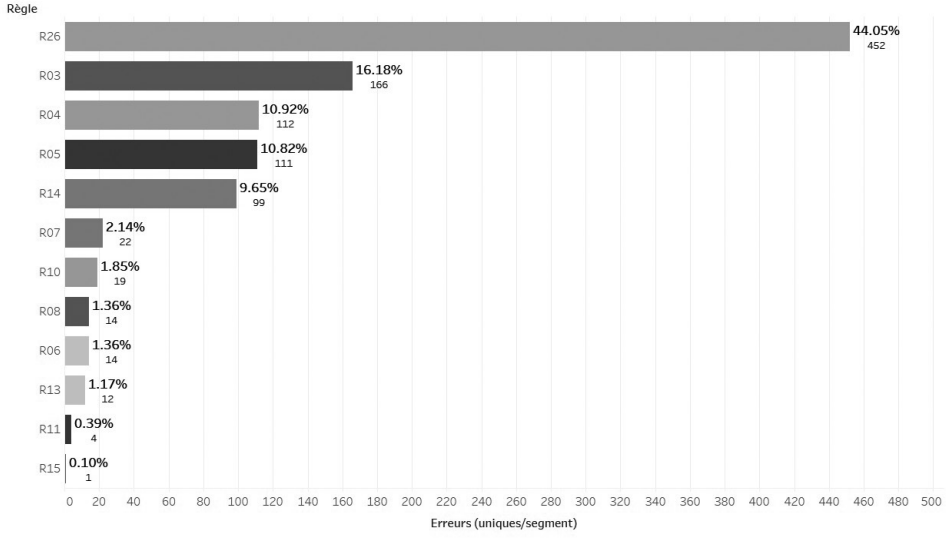


FIGURE 17

Erreurs en espagnol par règle (catégorie « Phrase »)

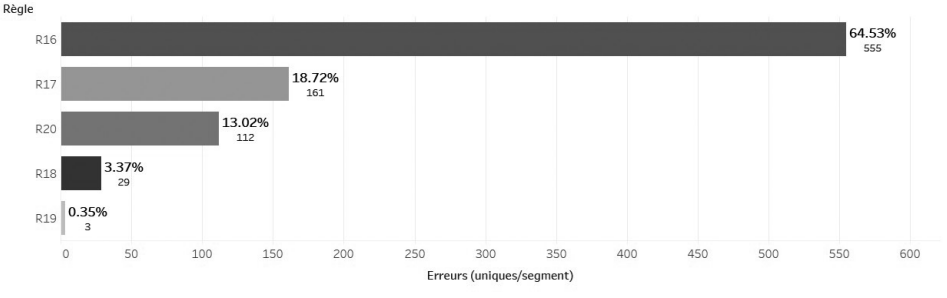


Tableau 12

Sélection d'exemples d'erreurs d'accessibilité linguistique (paires de langues FR-ES et FR-EN)

Paire de langues	Règle	Source	Traduction automatique
FR-ES	R16 - Parlez directement aux gens. Utilisez des mots comme « vous ».	**On s'occupe** de lui et on lui donne de l'affection.	**Se ha ocupado** de él y le ha dado afecto.
	R16 - Parlez directement aux gens. Utilisez des mots comme « vous ».	**On écoute** son avis.	**Escuchamos** su opinión.
	R18 - Utilisez des phrases actives plutôt que des phrases passives quand vous le pouvez.	**Il faut l'envoyer** dans les délais. Le délai est normalement de 30 jours.	**Tiene que ser enviado** a tiempo. El plazo es normalmente de 30 días.
FR-EN	R05 - Utilisez le même mot pour parler de la même chose dans tout le document.	Les parents ne peuvent pas ou ne veulent pas toujours protéger le **bien** de leur enfant.	Parents are not always able or willing to protect the **good** of their child.
	R05 - Utilisez le même mot pour parler de la même chose dans tout le document.	Le **bien** de l'enfant est en danger quand un danger physique ou un danger psychologique le menacent.	A child's **property** is in danger when there is a physical or psychological threat to the child.
	R03 - Utilisez des mots faciles à comprendre, c'est-à-dire des mots que les gens connaissent bien. R05 - Utilisez le même mot pour parler de la même chose dans tout le document.	L'APEA **intervient** quand le **bien** de l'enfant est en danger.	The APEA **intervenes** when the child's **welfare** is in danger.

5.3. Discussion

Les données recueillies nous ont permis de valider nos hypothèses en ce qui concerne la qualité de la TA pour le FALC (QR1, voir section 4.1). Malgré les progrès récents, les systèmes génériques ne sont pas encore capables de traduire des textes FALC de manière entièrement automatique, indépendamment du système, de la langue et du domaine. En ce qui concerne les systèmes, DeepL s'est révélé l'outil le plus performant pour traduire les textes FALC, même si le type de fautes de traduction était très similaire à celui de Google Translate. De son côté, Yandex obtient le plus grand nombre d'erreurs de fluidité, ce qui rejoint les conclusions tirées par Kaplan, Rodríguez Vázquez *et al.* (2019) dans leur étude exploratoire, centrée sur le domaine administratif. Ces résultats confirment donc la supériorité souvent notée de DeepL (Macketanz, Ai *et al.* 2018), également pour la traduction de textes contrôlés (Marzouk et Hansen-Schirra 2019). Pour les langues, il est surprenant de constater que l'espagnol a causé le plus de problèmes, étant donné qu'il appartient à la même famille de langue que la langue source. Il est important de noter, cependant, que la plupart des erreurs signalées par les annotateurs concernaient le style. De même, notre étude montre que les systèmes de TA doivent encore faire des progrès sur le plan de la précision pour le farsi.

Dans cette étude, nous avons également validé nos hypothèses par rapport à l'accessibilité des textes produits et aux règles FALC violées (QR2, voir section 4.1).

De manière générale, le niveau lexical reste celui qui présente le plus de problèmes, indépendamment du système, du domaine et des langues. Ceci a été également démontré par Kaplan, Rodríguez Vázquez *et al.* (2019) et peut s'expliquer par le fonctionnement des systèmes neuronaux, qui ont tendance à faire plus de fautes sémantiques (voir section 2.1). En allemand, le fait que les systèmes de TA contiennent en majorité des corpus en allemand d'Allemagne a aussi eu un impact significatif en ce qui a trait à l'accessibilité linguistique. Par exemple, la traduction de « enveloppe de vote » par « Abstimmungsumschlag » au lieu de « Wahlcouvert », utilisé dans le contexte suisse, représente une incohérence lexicale qui pourrait entraîner des problèmes de compréhension pour le public cible, et plus particulièrement pour ceux ayant besoin d'un lexique connu et utilisé dans la langue parlée. L'origine du texte pourrait aussi expliquer le fait que le domaine administratif a obtenu les moins bons résultats par rapport aux autres domaines.

Pour les deux évaluations, il est intéressant de noter que l'accord entre les annotateurs a été beaucoup plus élevé lors de la détection de problèmes de traduction que dans l'évaluation de conformité aux règles de la langue FALC, ce qui montre la subjectivité des règles, ainsi que la difficulté de l'évaluation de l'accessibilité linguistique. Dans le futur, il serait souhaitable d'étudier plus en détail l'accord par catégorie d'erreurs ou par règle violée, afin de définir des mécanismes pour clarifier les règles quand elles sont trop subjectives. Il serait par ailleurs intéressant de comparer les résultats de l'évaluation humaine avec les métriques automatiques couramment utilisées pour mesurer la simplification (indices de lisibilité, fréquences des mots, analyse syntaxique des textes, par exemple [Grabar 2019]), en tenant compte des limitations liées à ces formules quand elles sont utilisées dans le contexte de l'accessibilité linguistique (cf. Kaplan 2021 : 39-42).

De manière générale, ces deux évaluations montrent que la TA brute ne peut pas encore être utilisée pour produire des textes FALC. Les résultats obtenus par type de segments (sans erreurs, avec des erreurs d'accessibilité linguistique ou de qualité de traduction uniquement, ou avec les deux types d'erreurs ; voir figure 18) nous permettent de tirer les conclusions suivantes :

Segments sans erreurs : si l'on considère que seuls les segments sans erreurs sont utiles pour produire du contenu FALC, les meilleurs résultats sont obtenus par DeepL et pour le domaine médical et la langue anglaise, comme résumé dans le tableau 13 (voir pourcentages marqués en gras).

Segments avec des erreurs de qualité de traduction : il n'est sans doute pas utile de produire un texte cible conforme aux règles de la langue FALC s'il contient des problèmes de traduction, notamment de fidélité[16]. Si c'est le cas, il serait donc souhaitable de faire postéditer le texte en bilingue et le faire valider ensuite à nouveau par rapport aux bonnes pratiques d'accessibilité linguistique. Dans un scénario idéal, les deux tâches seraient accomplies dans une phase unique par des postéditeurs formés en FALC, afin d'optimiser le processus de production. Dans des cas exceptionnels, où la traduction automatique ne contient pas d'erreurs de fidélité, on pourrait envisager une postédition monolingue. Dans notre étude, ce type de segments (avec des erreurs de qualité de traduction uniquement) sont plus fréquents en anglais, avec DeepL et dans le domaine médical (voir tableau 13).

Segments avec des erreurs d'accessibilité linguistique : pour les segments de ce type, la TA pourrait aussi être utile. Il serait en effet possible de consulter un expert

en FALC qui se focaliserait sur la dernière étape du processus (passage de l'étape 2 à 4, voir figure 1 dans la section 2.1). Selon nos observations, DeepL serait toujours l'outil le plus recommandable. Même si Yandex obtient le pourcentage le plus bas en ce qui concerne uniquement le nombre de segments avec des violations des règles du FALC, il faut tenir compte du pourcentage total (52,68 %). Les meilleurs résultats pour ce scénario sont obtenus pour l'anglais et le domaine politique, qui ont le pourcentage le plus faible de segments inaccessibles (voir tableau 13).

Ces pistes de réflexion, fondées sur la littérature et les observations de notre corpus annoté, seront complétées et affinées lors de la deuxième étape de notre étude (évaluation par des utilisateurs cibles, voir section 4.1). Dans le futur, il serait aussi intéressant de voir comment ajouter la postédition dans ce processus de création de contenu accessible.

Figure 18

Type de segments selon les erreurs trouvées

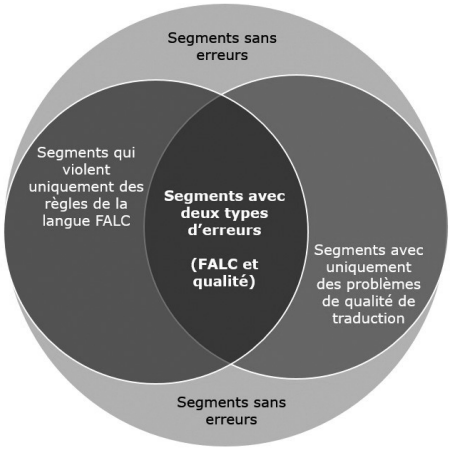


Tableau 13

Type de segments selon les erreurs trouvées (nombre et %)

Segments/ variables	Système			Domaine			Langue			
Sans erreurs	689	902	553	300	843	1001	836	483	556	269
	35,94 %	35,29 %	21,64 %	20,51 %	**34,37 %**	32,16 %	**43,61 %**	37,79 %	29,00 %	14,03 %
2 types d'erreurs	577	859	1177	698	785	1130	590	538	627	858
	30,10 %	33,61 %	46,05 %	47,71 %	**32,00 %**	36,30 %	**30,78 %**	42,10 %	32,71 %	44,76 %
Erreurs FALC	841	1154	1349	878	1120	1346	837	637	882	988
	43,87 %	45,15 %	52,78 %	60,01 %	45,66 %	**43,24 %**	**43,66 %**	49,84 %	46,01 %	51,54 %
Erreurs FALC (isolées)	313	359	209	228	392	261	273	109	272	227
	37,22 %	31,11 %	**15,49 %**	25,97 %	35,00 %	**19,39 %**	32,62 %	**17,11 %**	30,84 %	22,98 %
Erreurs traduction	964	1359	1831	983	1275	1896	834	696	1106	1518
	50,29 %	53,17 %	71,64 %	67,19 %	**51,98 %**	60,91 %	**43,51 %**	54,46 %	57,69 %	79,19 %
Erreurs traduction (isolées)	436	564	691	333	547	811	270	168	496	757
	45,23 %	41,50 %	**37,74 %**	**33,88 %**	42,90 %	42,77 %	32,37 %	**24,14 %**	44,85 %	49,87 %
Segments (total)	1917	2.556	2556	1463	2.453	3113	1917	1278	1917	1917

6. Conclusions

Dans cet article, nous avons présenté la première étude à notre connaissance sur la traduction automatique de textes faciles à lire, avec différents systèmes, domaines et paires de langues. Nous avons également créé un corpus multilingue, annoté selon les critères de qualité de la traduction et d'accessibilité linguistique, et qui pourrait être exploité par la communauté à de nouvelles fins de recherche[17].

Globalement, nous avons observé la même tendance dans les deux évaluations : comme dans d'autres études avec des textes en langue standard, DeepL offre les meilleurs résultats. Cependant, la qualité reste insuffisante pour pouvoir suggérer l'utilisation de la TA brute pour la production de textes FALC. Pour ce qui est des langues et des domaines, l'espagnol et les textes administratifs semblent poser le plus de défis du point de vue de la qualité de la traduction finale et de l'accessibilité linguistique.

Malgré les limites de notre étude, notamment en ce qui concerne le nombre d'annotateurs, la subjectivité de l'évaluation humaine de la qualité (Lommel, Popovic *et al.* 2014) et la taille de notre corpus, nous pensons qu'elle a apporté un éclairage intéressant sur les faiblesses de la TA à différents niveaux, qui pourrait être pris en compte lors de la création de corpus d'entraînement pour la TAN avec des textes accessibles. De même, nos résultats pourraient aider à définir des règles de pré- ou de postédition ciblant les problèmes d'accessibilité linguistique les plus récurrents. Enfin, notre étude peut constituer un point de départ pour des réflexions sur le processus de production de textes FALC, ainsi que les ressources nécessaires, notamment par rapport à la formation des postéditeurs.

Dans la suite, nous prévoyons de procéder à une analyse plus fine des résultats, en nous concentrant sur les règles de la langue FALC les plus problématiques, ainsi que sur la gravité des erreurs de traduction. Cela nous permettra de mieux corréler les données et d'examiner si certains problèmes de traduction entraînent des problèmes d'accessibilité particuliers. En outre, nous procéderons à une évaluation humaine auprès de groupes de population cibles afin d'évaluer la compréhensibilité de la TA brute par rapport à la TA postéditée.

REMERCIEMENTS

Nous tenons à remercier les huit annotateurs pour leur temps et leur intérêt pour la communication accessible. Cette recherche a été menée dans le cadre du projet « P-16 : Proposition et mise en œuvre d'un centre de recherche suisse pour une communication sans obstacle (2017-2020) », codirigé par la Haute école spécialisée zurichoise (ZHAW) et l'Université de Genève (UNIGE). Le projet a été approuvé par swissuniversities et financé par des contributions fédérales.

NOTES

1. Nations Unies (2006) : Convention relative aux droits des personnes handicapées. Genève : Nations unies. Consulté le 2 mai 2022, <https://www.un.org/disabilities/documents/convention/convoptprot-f.pdf>.
2. *Nations unies* (Dernière mise à jour : 28 septembre 2021) : Consulté le 2 mai 2022, <https://www.un.org/fr/observances/information-access-day>.
3. Unesco (Dernière mise à jour : 21 juin 2021) : Consulté le 2 mai 2022, <https://fr.unesco.org/programme/ifap>.
4. *Journal officiel de l'Union européenne* (Dernière mise à jour : 17 avril 2019) : Consulté le 2 mai 2022, <https://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:32019L0882&qid=1603884427271&from=EN>.

5. Bien que les textes FALC soient de nature multimodale, nous nous sommes uniquement concentré sur leurs caractéristiques linguistiques principales.
6. Ministère de l'Intérieur, France (2018) : *Élections des représentants au Parlement européen du 26 mai 2019 – Mémento à l'usage des candidats*. France : République française. Consulté le 2 mai 2022, <https://www.interieur.gouv.fr/Elections/Etre-candidat/Elections-des-representants-au-Parlement-europeen-du-26-mai-2019>.
7. *Textoh !* (Dernière mise à jour : 15 avril 2022) : Consulté le 2 mai 2022, <https://www.textoh.ch/traduction/>.
8. Commission Européenne (2017) : *Progress Report on the implementation of the European Disability Strategy (2010-2020)*. Bruxelles : Commission européenne (Dernière mise à jour : 2 février 2017) : Consulté le 2 mai 2022, <https://ec.europa.eu/social/BlobServlet?docId=16995&langId=en>.
9. *Unapei* (Dernière mise à jour : 15 avril 2022) : Consulté le 2 mai 2022, <https:// www.unapei.org/>.
10. *FALC-Assistant* (Dernière mise à jour : XX mois ANNÉE) : Consulté le 2 mai 2022, <https://frh-fondation.ch/en/falc-assistant/>.
11. *Simples* (Dernière mise à jour : 12 novembre 2019) : Consulté le 2 mai 2022, <http://51.91.138.70/simples/>.
12. La règle 19 des Règles spécifiques aux informations écrites des consignes européennes pour une information facile à lire et à comprendre dit : « Faites des phrases courtes. Lorsque c'est possible, 1 phrase devrait tenir sur 1 ligne. Si vous devez écrire 1 phrase sur 2 lignes, coupez la phrase à l'endroit où vous feriez une pause lorsque vous lisez à voix haute » (Inclusion Europe 2009). Dans notre étude, nous avons également voulu examiner comment la TA traite ce type de sauts de ligne forcés dans les textes FALC, introduits à des fins d'accessibilité.
13. Une proposition plus récente est le « Mediopunkt », c'est-à-dire un point au lieu du trait d'union qui est susceptible de faciliter ultérieurement la compréhension de mots composés (Bredel et Maaß 2017).
14. *Harmonized DQF-MQM Error Typology,* TAUS. Consulté le 2 mai 2022, <https://www.taus.net/qt21-project#harmonized-error-typology/>.
15. La traduction littérale aussi erronée que celle en allemand serait « Nous votons aussi les Hommes qui au Conseil des États vont » (la syntaxe est invertie). La variante correcte, parmi d'autres serait : Wir wählen auch die Menschen/Personen, die in den Ständerat gehen.
16. Il existe des situations où l'utilisation du résultat brut de la TA est acceptable, même s'il contient des erreurs de traduction ; par exemple, pour le *gisting* (Forcada, Scarton *et al.* 2018). Pourtant, dans le cas de la langue FALC, il y aurait des considérations socioéthiques à prendre en compte, étant donné les caractéristiques des groupes de population cibles.
17. Le corpus annoté est disponible sur demande.

RÉFÉRENCES

BFEH (2019) : *Langue facile à lire. Fiche d'information à l'intention de l'administration fédérale*. Confédération suisse : Bureau fédéral de l'égalité pour les personnes handicapées (BFEH).

Bredel, Ursula et Maass, Christiane (2017) : Wortverstehen durch Wortgliederung – Bindestrich und Mediopunkt in Leichter Sprache. *In* : Bettina M. Bock, Ulla Fix et Daisy Lange, dir. *"Leichte Sprache" im Spiegel theoretischer und angewandter Forschung*. Kommunikation – Partizipation – Inklusion. Berlin : Frank & Timme, 211-228.

Castilho, Sheila, Moorkens, Joss, Gaspari, Federico *et al.* (2017) : Is Neural Machine Translation the New State of the Art ? *The Prague Bulletin of Mathematical Linguistics*. 108:109-120.

Chehab, Nael, Holken Hadmut et Malgrange, Mathilde (2019) : *Étude Recueil des besoins FALC*. Issy-Les-Moulineaux : Holken Consultants & Partners. Consulté le 2 mai 2022, <http://51.91.138.70/simples/docs/SIMPLES_Etude_Recueil_desBesoins_FALC_HC.pdf>.

Cohen, Jacob (1960) : A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*. 20(1):37-46.

Felici, Annarita et Griebel, Cornelia (2019) : The Challenge of Multilingual 'Plain Language' in Translation-Mediated Swiss Administrative Communication : A Preliminary Comparative Analysis of Insurance Leaflets. *Translation Spaces*. 8(1):167-191.

Feng, Lijun, Elhadad, Noémie et Huenerfauth, Matt (2009) : Cognitively Motivated Features for Readability Assessment. *In* : Alex Lascarides, Claire Gardent, Joakim Nivre, dir.

Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics on - EACL '09, 229-237. Athènes : Association for Computational Linguistics. Consulté le 2 mai 2022, <https://doi.org/10.3115/1609067.1609092>.

Forcada, Mikel L. (2017) : Making sense of neural machine translation. *Translation Spaces*. 6(2):291-309.

Forcada, Mikel L., Scarton, Carolina, Specia, Lucia, *et al.* (2018) : Exploring gap filling as a cheaper alternative to reading comprehension questionnaires when evaluating machine translation for gisting. *In* : Ondřej Bojar, Rajen Chatterjee, Christian Federmann *et al.*, dir. *Proceedings of the Third Conference on Machine Translation : Research Papers*, 192–203. Bruxelles : Association for Computational Linguistics. Consulté le 2 mai 2022, <https://doi.org/10.18653/v1/W18-6320>.

FRANET (2016) : *Monthly data collection on the current migration situation in the EU - August 2016 monthly report*. Bruxelles : European Union Agency for Fundamental Rights (FRA). Consulté le 2 mai 2022, <https://fra.europa.eu/sites/default/files/fra_uploads/fra-august-2016-monthly-migration-disability-focus_en.pdf>.

Grabar, Natalia (2019) : *Adaptation de documents techniques pour les locuteurs non spécialisés*. Thèse pour l'obtention d'une habilitation à diriger la recherche de l'Université de Paris-Sud. Paris : Université Paris-Sud.

Grabar, Natalia et Cardon, Rémi (2018) : CLEAR – Simple Corpus for Medical French. *In* : Arne Jönsson, Evelina Rennes, Horacio Saggion *et al.*, dir. *Proceedings of the 1st Workshop on Automatic Text Adaptation (ATA)*, 3-9. Tilbourg : Association for Computational Linguistics. Consulté le 2 mai 2022, <https://doi.org/10.18653/v1/W18-7002>.

Hansen-Schirra, Silvia et Maass, Christiane, dir. (2020) : *Easy Language Research : Text and User Perspectives. Vol. 2. Easy - Plain - Accessible*. Berlin, Heidelberg : Frank & Timme.

Hansen-Schirra, Silvia, Nitzke, Jean, Gutermuth, Silke, *et al.* (2020) : Technologies for the Translation of Specialised Texts into Easy Language. *In* : Silvia Hansen-Schirra et Christiane Maass, dir. *Easy Language Research : Text and User Perspectives*. Berlin : Frank & Timme, 99-127.

Inclusion Europe (2009) : *Information for all. European standards for making information easy to read and understand*. Bruxelles : Commission européenne. Consulté le 2 mai 2022, <https://www.inclusion-europe.eu/easy-to-read-standards-guidelines/>.

Kaplan, Abigail (2021) : Suitability of Neural Machine Translation for Producing Linguistically Accessible Text : Exploring *the Effects of Pre-Editing on Easy-to-Read Administrative Documents*. Mémoire de maîtrise non publié. Genève : Université de Genève (UNIGE). Consulté le 2 mai 2022, <https://archive-ouverte.unige.ch/unige:151207>.

Kaplan, Abigail, Rodríguez Vázquez, Silvia et Bouillon, Pierrette (2019) : Measuring the Impact of Neural Machine Translation on Easy-to-Read Texts : An Exploratory Study. Présenté à : Conference on Easy-to-Read Language Research (Klaara 2019), Helsinki. Consulté le 2 mai 2022, <https://archive-ouverte.unige.ch/unige:123648>.

Klubička, Filip, Toral, Antonio et M. Sánchez-Cartagena, Víctor (2017) : Fine-Grained Human Evaluation of Neural Versus Phrase-Based Machine Translation. *The Prague Bulletin of Mathematical Linguistics (PBML)*. 108:121-132.

Koehn, Philipp (2009/2017) : Draft of Chapter 13 : Neural Machine Translation. *Statistical Machine Translation*, 2e éd. Cambridge : Cambridge University Press. Consulté le 2 mai 2022, <https://arxiv.org/pdf/1709.07809.pdf>.

Kuhn, Tobias (2014) : A Survey and Classification of Controlled Natural Languages. *Computational Linguistics*. 40(1):121–170.

Lamotte, Helen et Therwath, Alexandre (2016). Orsay facile : Inclure les personnes déficientes intellectuelles dans l'élaboration de documents adaptés. *I :* Emma Nardi et Cinzia Angelini, dir. *Best Practice 5 : A tool to improve museum education internationally*. Rome : Edizioni Nuova Cultura, 61-70. Consulté le 2 mai 2022, <https://ceca.mini.icom.museum/fr/publications/best-practice/>.

Lommel, Arle, Popovic, Maja et Burchardt, Aljoscha (2014) : Assessing Inter-Annotator Agreement for Translation Error Annotation. *In* : *Proceedings of MTE : Workshop on Automatic and Manual Metrics for Operational Translation Evaluation - LREC*. Reykjavik, Islande. Consulté le 2 mai 2002, <https://www.dfki.de/fileadmin/user_upload/import/7445_LREC-Lommel-Burchardt-Popovic.pdf>.

Luce, Amy (2018) : *Asylum Seekers and Refugees with Intellectual Disabilities in Europe*. Oakville : Samuel Centre for Social Connectedness. Consulté le 2 mai 2022, <https://www.socialconnectedness.org/wp-content/uploads/2019/10/Asylum-Seekers-and-Refugees-with-Intellectual-Disabilities-in-Europe-1-1.pdf>.

Maass, Christiane (2020) : *Easy Language – Plain Language – Easy Language Plus : Balancing comprehensibility and acceptability*. Vol. 3. Easy - Plain - Accessible. Berlin : Frank & Timme. Consulté le 2 mai 2022, <https://library.oapen.org/handle/20.500.12657/42089>.

Macketanz, Vivien, Ai, Renlong, Burchardt, Aljoscha *et al.* (2018) : TQ-AutoTest – A Semi-Automatic Test Suite for (Machine) Translation Quality. *In* : Nicoletta Calzolari, Khalid Choukri, Christopher Cieri *et al.*, dir. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation*. Miyazaki : European Language Resources Association (ELRA), 886-892.

Marzouk, Shaimaa et Hansen-Schirra, Silvia (2019) : Evaluation of the Impact of Controlled Language on Neural Machine Translation Compared to Other MT Architectures. *Machine Translation*. 33(1-2):179-203.

Matausch, Kerstin et Nietzio, Annika (2012) : Easy-to-read and plain language : defining criteria and refining rules. *In* : Klaus Miesenberger, Andrea Petz, Kerstin Matausch, *et al.*, dir. *Proceedings of the W3C Easy-to-Read on the Web Symposium*, Paper 11. Consulté le 2 mai 2022, <http://www.w3.org/WAI/RD/2012/easy-to-read/paper11/>.

Neubig, Graham, Morishita, Makoto et Nakamura, Satoshi (2015) : Neural Reranking Improves Subjective Quality of Machine Translation. *In* : Toshiaki Nakazawa, Hideya Mino, Isao Goto *et al.*, dir. *Proceedings of the 2nd Workshop on Asian Translation (WAT)*. Kyoto : Workshop on Asian Translation, 35-41. Consulté le 2 mai 2022, <https://aclweb.org/anthology/papers/W/W15/W15-5003/>.

Nietzio, Annika, Scheer, Birgit et Bühler, Christian (2012) : How Long Is a Short Sentence ? – A Linguistic Approach to Definition and Validation of Rules for Easy-to-Read Material. *In* : Klaus Miesenberger, Arthur Karshmer, Petr Penaz *et al.*, dir. *Computers Helping People with Special Needs*. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 7383:369-376. Consulté le 2 mai 2022, <http://dx.doi.org/10.1007/978-3-642-31534-3_55>.

Rossetti, Alessandra (2019) : *Simplifying, Reading, and Machine Translating Health Content : An Empirical Investigation of Usability*. Thèse de doctorat non publiée. Dublin : Dublin City University (DCU).

Štajner, Sanja, Mitkov, Ruslan et Corpas Pastor, Gloria (2015) : Simple or Not Simple ? A Readability Question. *In* : Núria Gala, Reinhard Rapp et Gemma Bel-Enguix, dir. *Language Production, Cognition, and the Lexicon*. Cham : Springer International Publishing, 48:379-398. Consulté le 2 mai 2022, <https://doi.org/10.1007/978-3-319-08043-7_22>.

Sutherland, Rebekah Joy et Isherwood, Tom (2016) : The Evidence for Easy-Read for People With Intellectual Disabilities : A Systematic Literature Review. *Journal of Policy and Practice in Intellectual Disabilities*. 13(4):297-310. Consulté le 2 mai 2022, <https://doi.org/10.1111/jppi.12201>.

Yaneva, Victoria (2015) : Easy-read Documents as a Gold Standard for Evaluation of Text Simplification Output. *In* : Irina Temnikova, Ivelina Nikolova et Alexander Popov, dir. *Proceedings of the Student Research Workshop associated with RANLP 2015*. Hissar : INCOMA Ltd, 30-36.

Yaneva, Victoria et Evans, Richard (2015) : Six Good Predictors of Autistic Text Comprehension. *In* : Ruslan Mitkov, Galia Angelova et Kalina Bontcheva, dir. *Proceedings of the International Conference Recent Advances in Natural Language Processing*. Hissar : INCOMA Ltd, 697-706. Consulté le 2 mai 2022, <https://www.aclweb.org/anthology/R15-1089>.

ANNEXES

Annexe 1 – Corpus d'évaluation

Texte 1 – Domaine politique

Titre : Un guide pour voter – 20 octobre 2019
Auteur : Pro Infirmis / Insieme.ch [Avec le soutien du Bureau fédéral de l'égalité pour les personnes handicapées, BFEH]

Source : https://insieme.ch/fr/produit/un-guide-pour-voter/
Consulté le 10 octobre 2020

Texte 2 – Domaine médical

Titre : Informations sur le coronavirus
Auteur : Office fédéral de la santé publique (OFSP), Confédération suisse

Source : https://www.bag.admin.ch/bag/fr/home/krankheiten/ausbrueche-epidemien-pandemien/aktuelle-ausbrueche-epidemien/novel-cov/barrierefreie-inhalte/leichte-sprache/leichte-sprache-informationen-zum-coronavirus.html
Consulté le 10 octobre 2020

Texte 3 – Domaine administratif

Titre : Informations sur la protection de l'enfant
Auteur : Conférence en matière de protection des mineurs et des adultes (COPMA)

Source : https://www.copma.ch/application/files/2414/9390/8817/Aide_memoire_protection_enfant_langage_simplifie.pdf Consulté le 10 octobre 2020