

Traduction automatique et marqueurs d'oralité : analyse à partir d'un corpus parallèle allemand-français

Auteur : Minder, Joachim

Promoteur(s) : Letawe, Céline

Faculté : Faculté de Philosophie et Lettres

Diplôme : Master en traduction, à finalité spécialisée

Année académique : 2022-2023

URI/URL : <http://hdl.handle.net/2268.2/17196>

Avertissement à l'attention des usagers :

Tous les documents placés en accès ouvert sur le site le site MatheO sont protégés par le droit d'auteur. Conformément aux principes énoncés par la "Budapest Open Access Initiative"(BOAI, 2002), l'utilisateur du site peut lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces documents, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale (ou prévue par la réglementation relative au droit d'auteur). Toute utilisation du document à des fins commerciales est strictement interdite.

Par ailleurs, l'utilisateur s'engage à respecter les droits moraux de l'auteur, principalement le droit à l'intégrité de l'oeuvre et le droit de paternité et ce dans toute utilisation que l'utilisateur entreprend. Ainsi, à titre d'exemple, lorsqu'il reproduira un document par extrait ou dans son intégralité, l'utilisateur citera de manière complète les sources telles que mentionnées ci-dessus. Toute utilisation non explicitement autorisée ci-avant (telle que par exemple, la modification du document ou son résumé) nécessite l'autorisation préalable et expresse des auteurs ou de leurs ayants droit.

Département de Langues modernes : linguistique, littérature et traduction

Filière Traduction et Interprétation

Traduction automatique et marqueurs d'oralité :
Analyse à partir d'un corpus parallèle allemand-français

Travail de fin d'études présenté par

Joachim Minder

en vue de l'obtention du grade de master en Traduction, à finalité spécialisée

Promotrice : M^{me} Céline Letawe

Co-promoteur : M. Damien Hansen

Lecteur : M. Robert Möller

Année académique 2022-2023

Remerciements

Je voudrais avant tout remercier ma promotrice, Céline Letawe, ainsi que mon co-promoteur, Damien Hansen, qui m'ont guidé dans ma réflexion et qui m'ont tous les deux apporté une aide précieuse, chacun dans son domaine ; leurs connaissances complémentaires m'ont permis de mener ce projet en étant parfaitement entouré. Ils ont su me guider dans ce travail de recherche et me suggérer certaines ressources lorsque cela était nécessaire.

J'aimerais aussi remercier les enseignantes et les enseignants de la filière de m'avoir apporté le bagage nécessaire, et grâce à qui je suis désormais prêt à me lancer dans le monde professionnel.

Enfin, je tiens à remercier mes proches de m'avoir soutenu tout au long de ce travail et de m'avoir remotivé dans les moments plus difficiles. Je les remercie également de m'avoir aidé dans la relecture de ce travail.

Table des matières

1. Introduction.....	1
2. La traduction automatique : développement, fonctionnement et conséquences ...	3
2.1. Historique de la traduction automatique	4
2.1.1. La traduction automatique à base de règles (RBMT)	4
2.1.2. La traduction automatique à base d'exemples (EBMT)	4
2.1.3. La traduction automatique statistique (SMT)	5
2.1.4. La traduction automatique neuronale (TAN).....	6
2.2. Qu'est-ce la TAN et comment fonctionne-t-elle ?	6
2.3. Problématiques liées à l'utilisation de la TAN	8
2.4. Quelle place reste-t-il pour les traducteurs ?	9
3. L'état de l'art des limites de la traduction automatique	11
3.1. Paramètres généraux influençant la TA	11
3.2. Quelques obstacles à la TA	14
3.3. Traduire l'oralité : quelles difficultés pour la TA ?	20
4. Méthodologie	26
5. Analyse du corpus parallèle	29
5.1. Hypothèse.....	29
5.2. Analyse d'extraits.....	30
5.3. Analyse d'exemples isolés	42
5.4. Discussion des résultats.....	64
6. Conclusion	66
Bibliographie	69

Liste des abréviations

- AV *audiovisuel*
- EBMT *example-based machine translation (TA à base d'exemples)*
- LC *langue cible*
- LS *langue source*
- PBSMT *phrase-based statistical machine translation (TA statistique à base de phrases)*
- PE *post-édition*
- RBMT *rule-based machine translation (TA à base de règles)*
- SMT *statistical machine translation (TA statistique)*
- TA *traduction automatique*
- TLA *traduction littéraire automatique*
- TAN *traduction automatique neuronale*
- TC *texte cible*
- TS *texte source*

1. Introduction

Chaque jour, de nouvelles avancées technologiques apparaissent ; le smartphone, l'Internet, l'intelligence artificielle, la réalité virtuelle, l'impression 3D ou encore la robotisation sont autant d'innovations qui changent la vie de nombreuses personnes. Et les métiers de la traduction ne sont pas épargnés par ces progrès.

L'arrivée de la traduction automatique neuronale en 2015 a permis un bond qualitatif indiscutable, au point que la presse généraliste évoque régulièrement ces progrès, qui entraîneraient selon elle la disparition du métier de traducteur. (Loock, 2018, p. 787)

Technologie révolutionnaire pour certains, véritable menace pour d'autres, la traduction automatique ne cesse de faire parler d'elle, surtout depuis ses dernières évolutions. Que ce soit sur les systèmes de messagerie, sur certaines pages Internet, sur les réseaux sociaux, voire dans un cadre professionnel, la traduction automatique est partout autour de nous. Il serait naïf de croire que la machine traduit tout parfaitement. En effet, rares sont ceux qui n'ont pas déjà appuyé sur le bouton « traduire automatiquement » pour ensuite constater qu'ils ne comprenaient pas ce qu'ils étaient en train de lire. C'est de ce constat, entre autres, que naît ce travail.

En effet, le développement rapide des nouvelles technologies de la traduction, que ce soit la traduction automatique à proprement parler ou d'autres outils d'aide à la traduction, constitue un intérêt majeur pour la recherche en traduction. Dans ce contexte, cette étude mêle traduction automatique, traductologie, linguistique contrastive, linguistique de corpus et traductologie de corpus. C'est Mona Baker (1993) qui suggère que la linguistique de corpus, en plein essor depuis le milieu du XX^e siècle, peut servir à la traductologie. Selon elle, un corpus, c'est-à-dire « une collection de textes en format électronique qui sont traités et analysés à l'aide de logiciels spécifiquement créés pour la recherche linguistique » (Zanettin, 2012, p. 7), peut se révéler utile pour l'analyse de traductions. C'est ainsi qu'est née l'expression « traductologie de corpus » (traduction de l'expression anglaise *corpus-based translation studies*). Plus concrètement, ce travail vise à observer comment DeepL, « le traducteur automatique le plus précis et le plus subtil au monde » si l'on en croit son site, traite les éléments propres à l'oralité. En effet, les marqueurs du langage oral n'ayant presque jamais une signification fixe — celle-ci varie selon le contexte dans lequel ces marqueurs sont employés —, il est intéressant d'analyser comment un programme de traduction automatique les traite. Il s'agit d'une étude sur corpus ; celui-ci se compose de

quatre conférences de presse du gouvernement allemand, que l'on peut retrouver sur son site, et de la traduction automatique de ce corpus en français générée par DeepL en octobre 2022. Le choix de conférences de presse garantit la présence d'un certain nombre d'éléments oraux, qui seront exposés et détaillés dans différentes parties du travail.

Avant l'analyse du corpus à proprement parler, ce travail comporte deux parties cruciales. Premièrement, il convient de définir la traduction automatique et d'établir son historique, de sa naissance au milieu du XX^e siècle à son état actuel. Cette partie répondra à de nombreuses questions : qu'est-ce la traduction automatique ? Comment s'est-elle développée ? Comment fonctionne-t-elle désormais ? Quelle place reste-t-il pour les traducteurs humains ou « biotraducteurs » ? En effet, il est crucial de bien avoir conscience du rythme auquel les modèles de traduction automatique se développent et de saisir leur fonctionnement pour comprendre certains aspects de ce travail. Deuxièmement sera présenté un état de l'art sur les limites de la traduction automatique qui ont déjà été étudiées. Celui-ci servira de base à l'analyse à la fois manuelle et automatisée du corpus dit « parallèle » — composé du texte source et de la traduction —, puisque l'analyse du corpus aura comme points de départ certains éléments relevés dans l'état de l'art. L'analyse du corpus sera divisée en deux parties : les observations formulées grâce à l'analyse d'extraits longs seront vérifiées à l'aide de l'analyse d'exemples ou de contre-exemples isolés.

Pour finir, ce travail tentera, par le biais d'un aspect particulier, c'est-à-dire les éléments de l'oralité, d'apporter des éléments de réponse à la question que bon nombre de personnes — traducteurs ou non — se posent : quand la traduction automatique est-elle acceptable, voire bonne, et quand ne l'est-elle pas ? L'analyse du corpus est la partie centrale de ce travail, puisque, bien que l'état de l'art soit assez complet, les technologies de traduction automatique évoluent de manière exponentielle ; dès lors, il est possible que certaines études menées récemment ne soient plus d'actualité. Plus précisément, la finalité de ce travail est d'observer des mécanismes systématiques de DeepL, ou au moins de dégager certaines tendances, surtout en matière de traduction automatique des éléments d'oralité, grâce à l'état de l'art sur les limites de la traduction automatique enrichi par les analyses qualitative et quantitative du corpus parallèle.

2. La traduction automatique : développement, fonctionnement et conséquences

Comme expliqué dans l'introduction, depuis quelques années, et particulièrement depuis les dernières avancées technologiques, l'automatisation est au cœur de nombreuses activités, et les métiers de la traduction n'y échappent pas. De surcroît, la course à la productivité — notamment dans le domaine de la traduction — a eu pour conséquence le développement de la traduction automatique (TA) (Ragni & Nunes Vieira, 2022, p. 137). L'application des programmes de TA ne fait pas consensus parmi les professionnels de la traduction. En effet, alors que certains rejettent complètement leur utilisation, d'autres se montrent plus ouverts à ces dernières technologies et prônent même leur utilisation (Schmidhofer & Mair, 2018, p. 164). Toutefois, une chose est sûre : il n'y aura pas de retour en arrière, nous n'aurons d'autre choix que d'accepter ces avancées et d'utiliser ces technologies à bon escient (*loc. cit.*). En raison de son efficacité et de son caractère abordable, la traduction automatique est déjà omniprésente dans de nombreux domaines. La traduction de pages Internet représente l'une des plus grandes industries pour la TA ; elle permet aux utilisateurs de consulter les sites Internet dans leur langue maternelle ou dans une langue qu'ils comprennent. La TA est également très utilisée dans les domaines de la littérature scientifique et du e-commerce (Wang et al., 2021, p. 10). Si l'on associe la TA à d'autres technologies, il est également possible, par exemple, de traduire des images — grâce à la vision par ordinateur — ainsi que des discours — grâce au traitement de la parole (*ibid.*, pp. 10–11).

Ces révolutions technologiques ont permis de mettre au jour un « nouveau paradigme de traduction automatique » (Forcada, 2017, p. 291) qu'est la traduction automatique neuronale (TAN). Avant la TAN, d'autres paradigmes de traduction automatique étaient utilisés. Tout d'abord sera dressé un bref historique de la traduction automatique. Ensuite seront abordés plus spécifiquement le principe et le fonctionnement de la TAN. Par ailleurs, quelques problématiques liées à l'exploitation des programmes utilisant la TA seront exposées. Au vu de la révolution numérique à laquelle nous assistons, il est également crucial d'évoquer les rôles que le traducteur peut — ou *doit* — encore jouer.

2.1. Historique de la traduction automatique

2.1.1. *La traduction automatique à base de règles (RBMT)*

Les débuts de la traduction automatique remontent à la moitié du XX^e siècle. La RBMT a été la première approche de TA à voir le jour. Elle est restée dominante jusqu'à la fin des années 1980 (Schmidhofer & Mair, 2018, p. 165). C'est dans les années 1950, dans un contexte de rivalité entre les États-Unis et l'URSS et plus particulièrement de la course à l'espace, qu'a été mis au point le premier système de RBMT. À l'Université de Georgetown (Washington D.C.), on confie aux linguistes la tâche de coder les règles linguistiques. Ceux-ci seront ensuite aidés par des ingénieurs. C'est ainsi que l'Université de Georgetown et l'IBM (*International Business Machines Corporation*) créent le premier système de RBMT, qui fonctionne avec la paire de langues russe-anglais (Loffler-Laurian, 1996, pp. 35–36). La RBMT est un modèle de traduction automatique qui repose sur des règles et des entrées de dictionnaires pour une paire de langues donnée. Ce sont ces connaissances qui guident le programme de TA lors de la traduction d'un texte de la langue source vers la langue cible (Shiwen & Xiaojing, 2015, p. 186). Néanmoins, quelques problèmes ont pu être observés avec le modèle de RBMT. Tout d'abord, les règles ne sont pas fixes ; elles évoluent avec le temps. De plus, tous les domaines et toutes les langues n'obéissent pas aux mêmes règles. Par conséquent, les textes multilingues ou les textes qui touchent plusieurs domaines étaient plus complexes à traiter avec la RBMT (Wang et al., 2021, p. 1).

2.1.2. *La traduction automatique à base d'exemples (EBMT)*

À la moitié des années 1980 est créée l'EBMT, le premier des trois modèles de traduction automatique dits « corpus-based » (Hutchins, 1995). Ce nouveau modèle rompt avec les approches à base de règles (RBMT), car l'intervention humaine n'est plus nécessaire pour le codage de règles, et la machine exécute tout le travail. L'EBTM consiste à « extraire des connaissances de traductions existantes (exemples) en vue de faciliter la traduction de nouveaux énoncés » (Wong Tak-ming & Webster, 2015, p. 137). Ce sont donc des corpus bilingues qui alimentent le programme. La traduction automatique à base d'exemples se fait en trois étapes successives. En premier lieu, le système associe un segment du nouveau texte à un segment qu'il trouve dans sa base d'exemples grâce au principe de mesure de similarité entre deux textes. Pendant le processus, les segments peuvent être coupés en sous-segments. Ensuite, la machine identifie la traduction des différents (sous-)segments. En dernier lieu,

elle regroupe les différents (sous-)segments pour produire la traduction (Wong Tak-ming & Webster, 2015, pp. 138–142). La qualité de la traduction dépend du degré de similarité entre le nouvel énoncé et l'énoncé retrouvé dans la base d'exemples. En effet, lorsque le programme identifie des segments similaires, la traduction est plus acceptable. L'EBMT, qui est alimentée par des corpus parallèles, s'utilise aujourd'hui surtout dans les programmes de traduction assistée par ordinateur (Wang et al., 2021, p. 2).

TM [translation memories] can be understood as a 'restricted form of EBMT' (Kit et al. 2002: 57-78) in the sense that both involve storing and retrieving previous translation examples; nevertheless in EBMT the translation output is produced by the system while in TM this is left to human effort. (Wong Tak-ming & Webster, 2015, pp. 137–138)

En effet, dans les programmes d'aide à la traduction, les mémoires de traductions sont des sortes de corpus parallèles alimentés par les traductions déjà réalisées par l'utilisateur.

2.1.3. *La traduction automatique statistique (SMT)*

Dans les années 1990, un nouveau paradigme de traduction automatique est exploité : la traduction automatique statistique (SMT). Ce modèle fonctionne également sur la base de corpus parallèles multilingues, mais ceux-ci sont exploités différemment par la machine que dans l'EBMT (Wang et al., 2021, p. 2). La SMT peut être considérée comme une amélioration du paradigme précédent (l'EBMT). Dans ces corpus parallèles, différentes unités sont mises en relation, du paragraphe au mot, en passant par l'expression et la phrase. Sont ensuite créées des statistiques de fréquence : celles-ci permettent de déterminer à quelle fréquence des mots ou expressions retrouvés dans le corpus se succèdent (Yang & Min, 2015, pp. 201–207). C'est à la toute fin des années 1990 que la SMT est présentée au grand public. En 1999, des chercheurs organisent un atelier à l'Université Johns-Hopkins aux États-Unis et y créent un programme de SMT qu'ils appellent *Egypt*. Dans la foulée sont mis au jour des programmes de *word-based SMT*. Dans ce genre d'approches, les mots sont considérés et traduits seuls. En 2003, la SMT connaît une avancée : les programmes de *phrase-based SMT*. Ceux-ci permettent d'améliorer la qualité de la traduction. En effet, le programme n'observe plus des mots, mais des segments formant des unités sémantiques. Ces *phrase-based SMT* ont été adoptées en 2006 par Google en vue de créer un service de traduction en ligne. Microsoft et Baidu ont ensuite suivi le mouvement (Wang et al., 2021, p. 2). Quelques années plus tard, d'autres progrès technologiques ont vu le jour, et ceux-ci se sont traduits par le développement d'un autre modèle de TA.

2.1.4. La traduction automatique neuronale (TAN)

Ce n'est que depuis quelques années que se développe le dernier paradigme de traduction automatique qu'est la traduction automatique neuronale ; son histoire est « courte, mais météorique » (Ragni & Nunes Vieira, 2022, p. 137). En effet, ce n'est qu'en 2014 que la TAN a vu le jour (Wang et al., 2021, p. 2). Cette dernière est une extension des deux approches précédentes (l'EMBT et la SMT) — par exemple, elle fonctionne elle aussi sur la base d'énormes corpus parallèles —, mais une toute nouvelle technologie la différencie de son prédécesseur : le réseau de neurones (Forcada, 2017, p. 292).

2.2. Qu'est-ce la TAN et comment fonctionne-t-elle ?

L'intelligence artificielle, et plus particulièrement le *deep learning*, est au cœur de la TAN. Tous les « neurones » qui composent ce système sont interconnectés ; et la manière dont ils sont activés et, par conséquent, ce qu'ils produisent dépendent fortement de leur stimulus. Ceux-ci fonctionnent un peu comme les statistiques de fréquence que l'on retrouvait dans le paradigme précédent, c'est-à-dire que tous les neurones ne sont pas connectés entre eux avec le même poids : les exemples que la machine retrouve dans le corpus influencent la relation entre chaque neurone. Le produit final, autrement dit la traduction, est le résultat de toutes les connexions qui se sont effectuées entre les neurones. Plus concrètement, la TAN fonctionne en trois étapes : l'une des étapes se fait en amont du processus de traduction à proprement parler, lui-même divisé en deux grandes phases (*ibid.*, pp. 292–295).

L'étape préliminaire est celle de l'entraînement. Lors de celle-ci, le réseau de neurones doit être entraîné. Autrement dit, « on détermine le poids et la force des connexions entre chaque neurone afin d'obtenir le résultat attendu » (*ibid.*, p. 295). Pour ce faire, la machine a besoin d'un corpus d'entraînement très volumineux. En fonction de la taille du corpus, la phase d'entraînement peut durer de quelques jours à plusieurs mois (*loc. cit.*).

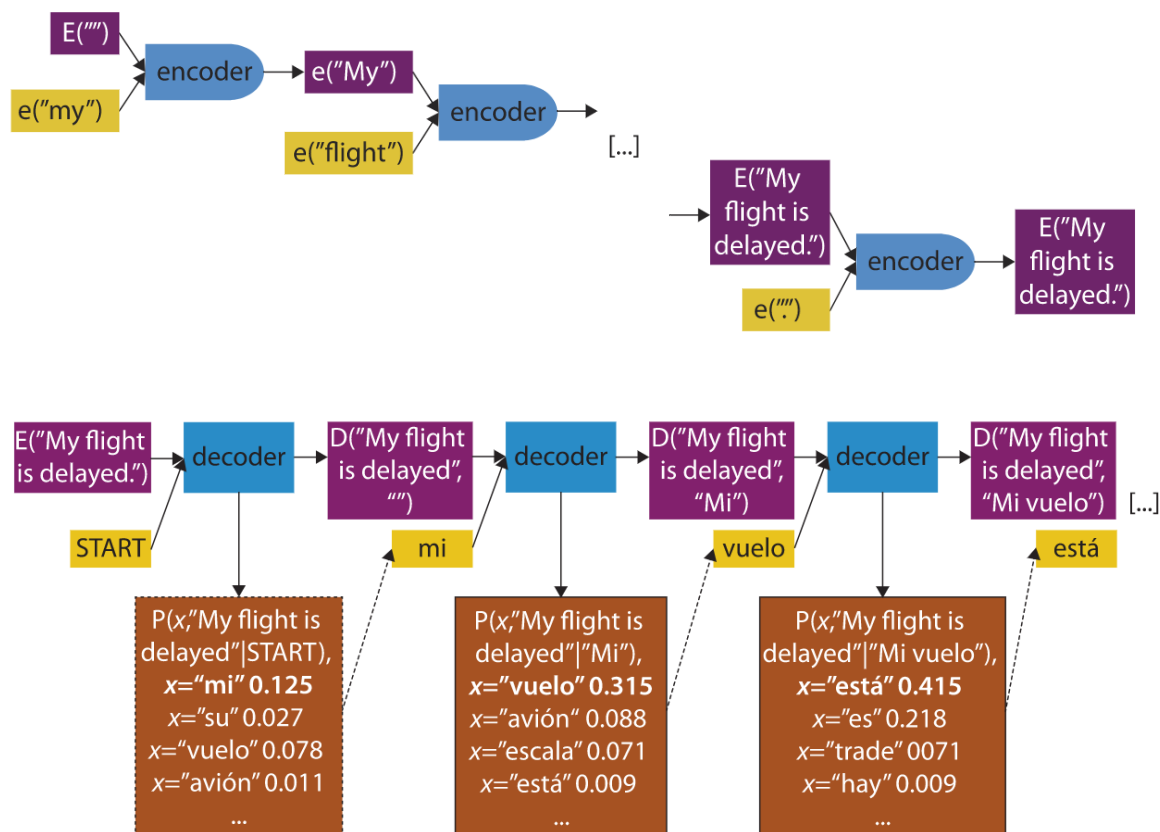
La première étape importante dans le processus de la TAN est alors l'encodage (*encoding*) de la phrase source à traduire. Imaginons que le système informatique veut traduire la phrase « *My father is a teacher* » en français. Dans cette étape d'encodage, le programme recherche dans sa base d'exemples les représentations (E) suivantes : E('My'), qu'il combine à la représentation E('father') pour rechercher la représentation E('My father'), ensuite E('My father is'), pour finalement arriver à la représentation de toute la

phrase $E(\text{'My father is a teacher.'})$. Ce modèle est inspiré de Forcada (2017, p. 297). Cette étape est cruciale, puisque c'est sur la base de cet encodage que la machine peut produire sa traduction.

Une fois que la machine a encodé le segment source, il lui reste à passer à l'étape du décodage (*decoding*), c'est-à-dire à produire la « bonne » traduction. Sur la base de ce que le système a encodé à l'étape précédente, celui-ci génère deux vecteurs : un vecteur à l'état initial $D(\text{'My father is a teacher'}, '')$, dans lequel '' laisse une place pour la traduction, et un vecteur de probabilités $p(x|\text{'My father is a teacher'}, '')$, où x représente toutes les traductions possibles du premier mot. Si le programme a été bien entraîné, il traduit sans doute le mot *my* par *mon*. Ainsi, il lit ensuite $D(\text{'My father is a teacher'}, '')$ ainsi que le mot *mon*, et génère deux autres vecteurs : $D(\text{'My father is a teacher'}, \text{'Mon'})$ et $p(x|\text{'My father is a teacher'}, \text{'Mon'})$. Le système répète cette étape jusqu'à ce qu'il génère la traduction : $D(\text{'My father is a teacher'}, \text{'Mon père est professeur'})$. Ce modèle est également inspiré de Forcada (2017, p. 298).

Les trois étapes décrites ci-dessus appartiennent au modèle de base de la TAN, que l'on appelle l'architecture encodeur-décodeur ou le modèle *seq2seq* (« sequence to sequence ») (Forcada, 2017, p. 299). Toutefois, ce modèle est souvent renforcé par d'autres mécanismes qui assurent une meilleure traduction. Parmi ces mécanismes, il existe le mécanisme d'attention, l'encodage bidirectionnel (Wang et al., 2021, p. 4) ou encore la convolution (Forcada, 2017, pp. 299–300), pour n'en citer que quelques-uns. Le mécanisme d'attention est particulièrement important, car il est au cœur des architectures 'Transformer' qui sont utilisées aujourd'hui. Celui-ci est très utile et performant, et ce, surtout pour la traduction de phrases longues. Il permet aux réseaux de neurones de garder en mémoire chacune des sorties qu'ils ont produites afin d'améliorer la sortie finale (Vaswani et al., 2017).

Voici deux images permettant de visualiser les phases de l'encodage et du décodage dans le cadre du modèle *seq2seq*.



Forcada (2017, pp. 298–299).

2.3. Problématiques liées à l'utilisation de la TAN

Pour pouvoir être efficaces, les programmes qui fonctionnent sur la base de la TAN doivent être entraînés sur un corpus parallèle volumineux.

In the 'real world' (versus the experimental world), DeepL [sic] is reputed to be one of the best automatic translators as far as European languages are concerned, if not the best. [...] The company asserts that one of the explanations for its success is the substantial size of the parallel corpus on which the system is trained. This corpus, Linguee, contains a billion aligned examples. (Blin, 2021, p. 1)

Ce passage laisse sous-entendre l'idée que, si l'on veut augmenter la qualité des productions des programmes de TA, il faut également augmenter le nombre de données d'entraînement. L'utilisation des TAN est-elle par conséquent durable ? En effet, l'exploitation de ces programmes demande énormément d'énergie, et, naturellement, plus le volume de données exploitées est important, plus la consommation d'énergie augmente. Cela va évidemment à l'encontre des objectifs écologiques qui ont été déterminés, notamment dans le cadre de l'Accord de Paris et des Objectifs de développement durable de l'ONU,

d'autant plus que ces programmes ne parviennent toujours pas à générer des traductions aussi qualitatives que celles produites par des professionnels (*ibid.*, p. 2).

Un autre problème s'ajoute à celui de l'impact environnemental des systèmes de TAN. Étant donné qu'il n'est pas aisé de mettre la main sur des données bilingues, il est presque inévitable de devoir intégrer aux données d'entraînement de ces programmes des documents de qualité médiocre. Mais aussi, il se peut que l'on enrichisse les données d'entraînement avec des textes qui ont déjà été traduits par un système de TA (*loc. cit.*). Dans ce cas, il est probable que la machine répète toujours les mêmes erreurs, ce qui renforce les biais. Au vu de ces questions importantes, il est pertinent de se poser les questions suivantes : comment faut-il utiliser ces programmes de traduction automatique et, par conséquent, quelle place reste-t-il pour les traducteurs ?

2.4. Quelle place reste-t-il pour les traducteurs ?

La coopération avec la machine est presque inévitable pour bon nombre de traducteurs professionnels. En effet, bien souvent, le traducteur professionnel améliore et corrige les productions de traduction automatique pour rendre le produit final acceptable (François, 2022, pp. 5–6). Ce phénomène — qui connaît un développement énorme — porte le nom de post-édition. Adopter cette stratégie devrait représenter un gain de temps important, car le traducteur ne doit pas surcorriger la production de la TA, mais « uniquement » corriger les erreurs graves, que François (2022, p. 7) définit comme les « erreurs d'orthographe, de grammaire, de conjugaison, de typographie, de syntaxe et de sens ». Au vu de la demande exponentielle de traductions, l'utilisation de la TA est loin d'être à son apogée. Le développement des technologies et la demande croissante de traductions se traduiront — et se traduisent déjà — par un besoin toujours plus important de post-éditeurs. Une chose est alors sûre : nous aurons toujours besoin de l'expertise de traducteurs qualifiés (Schmidhofer & Mair, 2018, pp. 166–168).

En outre, certains domaines se prêtent moins à la traduction automatique. Par exemple, la TA de textes très techniques comporte souvent des erreurs de traductions. Dans d'autres domaines plus sensibles, tels que les domaines juridiques, la médecine et les sujets liés à la migration, l'« intégrité de traducteurs humains qualifiés » est souvent nécessaire (*ibid.*, p. 167). Enfin, d'un point de vue déontologique, la TA peut également se révéler problématique. En effet, bien que les clients n'en soient pas systématiquement conscients,

la confidentialité n'est pas respectée en cas de TA, puisque les programmes de TA intègrent automatiquement le texte traduit dans leurs corpus (*loc. cit.*). Il existe tout de même certains cas dans lesquels la confidentialité est respectée, notamment lorsqu'il s'agit d'un programme développé et utilisé en interne dans une société ou lorsque l'on utilise une version Pro de DeepL.

Ragni & Vieira (2022, pp. 149–150) évoquent un autre rôle important qui incombe aux traducteurs professionnels :

[Translators] are needed more than just as post-editors – they have the expertise required to understand what clients need and how to achieve it, so they are well placed to contribute to decisions on when and how to draw on NMT.

En effet, les traducteurs humains ont les compétences nécessaires pour prendre des décisions concernant le recours à la TA : dans quels cas faut-il utiliser la TA ? Dans quels cas ne faut-il pas l'utiliser ? Si on l'utilise, comment l'utilise-t-on ? De quoi mon client a-t-il besoin, et quels moyens sont nécessaires pour y parvenir ? Comme expliqué ci-dessus, certains domaines se prêtent mieux à la TA que d'autres. Par conséquent, avant de passer à l'analyse de la TA d'un corpus composé de conférences de presse — qui sont des textes plutôt oraux —, je dresserai un état de l'art des quelques limites de la TA qui ont déjà été observées et étudiées.

3. L'état de l'art des limites de la traduction automatique

La TA n'est pas de mauvaise qualité à tous les niveaux. Si certains professionnels utilisent les programmes de TA, c'est qu'ils sont convaincus de leur potentiel.

[One] inference is very clear: those people using MT [...] must already be satisfied with the level of quality emanating from the MT systems they are deploying, otherwise they would stop using them. That is not the same thing at all as saying that MT quality is perfect, far from it. (Way, 2018, p. 159)

L'idée n'est pas de démontrer que la TAN ne produit que des traductions de mauvaise qualité. L'approche adoptée dans ce travail est nuancée. D'une part, il y a certains domaines ou types de textes pour lesquels une TA est tout à fait recevable. D'autre part, certains documents sont plus complexes à traiter pour la TA. C'est exactement ce dernier point qui sera au cœur de cette partie du travail.

3.1. Paramètres généraux influençant la TA

Avant d'observer les caractéristiques propres à des types de textes bien spécifiques, il est important d'évoquer trois aspects généraux qui peuvent affecter la qualité d'une traduction automatique. Ces trois aspects comprennent la relation qu'entretient la langue du texte source (TS) avec celle du texte cible (TC), les domaines qui sont couverts, et enfin, la longueur des phrases.

L'un des paramètres qui influencent le plus la qualité de la TA est sans aucun doute la relation qu'entretiennent les deux langues en question, c'est-à-dire la langue du TS et la langue de la TA. En effet, plusieurs études ont permis de démontrer que la qualité de la TA dépend du degré de similarité entre la langue source (LS) et la langue cible (LC) :

In general, it is widely accepted that the quality attainable by MT correlates with the level of relatedness between the pair of languages involved. (Toral & Way, 2018, p. 4).

Presque tous les chercheurs et traducteurs s'accordent sur ce point. Burchardt et al. (2016, p. 211), par exemple, expliquent que « la traduction automatique fonctionne particulièrement bien dans les cas où les langues source et cible sont assez similaires en termes de structure, de morphologie, de concepts [...] » (ma traduction). Ils illustrent ceci avec l'exemple de la paire de langues espagnol-portugais, un cas dans lequel la TA devrait être satisfaisante. En ce qui concerne la paire de langues allemand-français, qui fait l'objet

de ce travail, les choses pourraient en revanche se compliquer pour la TA. Tout d'abord, l'allemand et le français n'appartiennent pas à la même famille de langues : la première fait partie des langues germaniques ; la seconde, des langues romanes. Par conséquent, le français et l'allemand étant des langues de familles différentes, de nombreux contrastes linguistiques peuvent être observés dans cette paire. Par exemple, l'allemand avec son rejet du verbe en fin de proposition ou en fin de phrase est assez éloigné du français (*loc. cit.*). D'autres contrastes peuvent être observés entre les deux langues. D'abord, l'allemand a une énorme facilité à créer des mots composés, qui sont déjà difficiles à traduire pour le traducteur professionnel, mais qui peuvent s'avérer encore plus complexes pour la machine. Schreiber (2006, p. 86) note par exemple que, dans les langues romanes, les composés allemands sont souvent traduits – par les biotraducteurs – par des périphrases, c'est-à-dire par une « expression par plusieurs mots d'une notion qu'un seul mot pourrait exprimer » ('Périphrase', *Le Robert*). Wandruszka (1974, p. 320), quant à lui, pense même que, pour les traducteurs humains, la traduction des composés allemands est un problème qui peut se révéler insurmontable, étant donné que même en une phrase entière, il est parfois impossible d'exprimer en français tout le sens que comprend cette structure en allemand. Si cet aspect est déjà complexe pour les biotraducteurs, nul doute qu'il est également problématique pour la TAN. En outre, l'allemand est beaucoup plus favorable à l'utilisation de la voix passive que le français. En effet, dans sa quatrième thèse, Blumenthal (1987, p. 19) explique que le français est une langue plus active que l'allemand.

Cette thèse suggère que la phrase française exprime un degré d'activité plus élevé que la phrase allemande dans de nombreuses constructions typiques des deux langues. Elle fait écho à l'observation de Strohmeyer selon laquelle le français tend à remplacer un énoncé allemand décrivant un processus, parfois même un état, par un énoncé décrivant une action. (*Loc. cit.*, ma traduction)

Par conséquent, alors que l'allemand exprime plutôt un processus (*Vorgangspassiv*) ou un état (*Zustandspassiv*), le français exprimera une action. Toutefois, si le traducteur — ou la machine — décide de rendre une voix passive par une autre voix passive, cela ne constitue pas un problème de sens ; ce « choix » entraînera des conséquences pour l'idiomaticité du texte cible. La dernière différence de système entre les deux langues qui mérite d'être mentionnée, car la liste est longue, est l'adoption d'un style plutôt nominal en allemand, alors que le français tend à verbaliser — et donc à dénominaliser — au maximum. Malblanc (1968, p. 207) observe en allemand « une tendance à remplacer le verbe à sens prédicatif par

un verbe de valeur générale, peu chargé de sens, outil faisant le pont vers un substantif d'action ou d'état, porteur de la signification principale ». Pour illustrer ses propos, prenons l'exemple de *etwas in Betracht ziehen* (qui se traduit littéralement par 'prendre quelque chose en considération' plutôt que 'considérer quelque chose'). Cet aspect n'influencera pas non plus la qualité de la TA en termes de sens, mais plutôt d'un point de vue stylistique.

Un autre paramètre qui influence la qualité de la TA concerne le ou les domaines qui sont couverts. D'une part, comme l'expliquent Burchardt et al. (2016, p. 209), plus le domaine du TS se rapproche du domaine sur lequel a été entraînée la machine, plus la qualité de la TA est bonne. Pour appuyer cet argument, ils prennent l'exemple de la traduction audiovisuelle. Selon eux, étant donné que « le contenu audiovisuel couvre le plus large éventail de thématiques possible », sa traduction représente un défi énorme (*ibid.*, p. 212). Les textes qui se limitent à un domaine se prêtent davantage à la TA. Par conséquent, la TA fonctionne particulièrement bien pour la traduction de textes techniques — plus précis —, alors que la qualité de la TA pour des textes à domaines « ouverts et imprévisibles » est moins bonne (Toral & Way, 2014, p. 175). Cette observation semble contredire celle de Schmidhofer & Mair (2018, p. 167), selon laquelle la TA de textes techniques peut se révéler plutôt mauvaise, comme expliqué au point 2.4. Cette contradiction montre à quel point la traduction automatique est un sujet d'intérêt récent et complexe. Il n'existe en effet pas de réelle explication à cette contradiction ; par exemple, les deux études sont basées sur des corpus similaires : Toral & Way se servent d'un corpus parallèle de deux langues romanes, et Schmidhofer & Mair utilisent un corpus de deux langues germaniques. De plus, l'étude de Toral & Way a été menée en 2014, et celle de Schmidhofer & Mair en 2018. Par conséquent, les résultats de l'étude de Schmidhofer & Mair devraient être meilleurs que ceux obtenus par Toral & Way quatre ans plus tôt.

Enfin, le dernier critère général qui affecte la qualité d'une TA est la longueur des phrases. Toral & Way (2018, p. 1) affirment que la traduction automatique neuronale affiche une qualité supérieure à celle de son prédécesseur, à savoir le dernier paradigme de traduction automatique statistique (*phrase-based SMT*). Bien que cette affirmation vaille à beaucoup de niveaux, ils affirment que plus les phrases sont longues, moins la différence de qualité entre les deux paradigmes (TAN et PBSMT) est visible (*ibid.*, p. 13), ce qui revient à dire que les phrases longues représentent également un défi pour la TAN. Ils ne définissent en revanche pas plus précisément la notion de 'phrases longues'. Cette idée fait tout de même

consensus parmi les chercheurs et les traducteurs. Daems & Ruffo (2022) appuient cet argument. Selon elles, la TA est particulièrement utile pour les textes rédigés avec des phrases courtes ; lorsque la machine est amenée à traduire des phrases longues, un certain nombre d’omissions peuvent être observées. Elles expliquent ce phénomène par le fait que la machine, lorsqu’elle est confrontée à une phrase longue, n’est pas en mesure de tenir compte du co(n)texte relativement éloigné. Matusov (2022), en revanche, vient nuancer cette idée. Alors que la qualité de la TA d’une phrase courte peut se révéler assez bonne dans la plupart des cas, la TA peut s’avérer problématique lorsque la machine est confrontée à une phrase longue, mais aussi lorsqu’elle doit traduire une phrase trop courte. En effet, dans les deux cas, la machine manque de contexte : il est soit trop éloigné — si les phrases sont longues —, soit manquant — si les phrases sont trop courtes.

3.2. Quelques obstacles à la TA

À présent sont traités certains aspects qui représentent des obstacles concrets pour la TA. Ceux-ci sont moins généraux que ceux développés au point 3.1. ; ce sont des phénomènes très précis que l’on ne retrouve pas dans tous les types de textes. L’état de l’art de la TA (neuronale et statistique) est assez riche, mais les recherches se concentrent en majeure partie sur les limites de la traduction littéraire automatique (TLA), avec quelques textes sur le domaine de l’AV également. Dans ce travail, les limites de la TA sont divisées en deux parties. La partie 3.2. s’intéresse à certains phénomènes qui peuvent se retrouver dans les textes oraux, sans pour autant être inhérents à ce type de textes. Les phénomènes présents principalement dans les textes oraux feront l’objet de la partie 3.3.

Tout d’abord, les programmes de TA peuvent être confrontés à un cas très difficile à traiter, même pour les traducteurs professionnels : les éléments culturels, aussi appelés culturèmes ou *realia*. Zotti (2019, p. 227) définit les *realia* comme des « réalités culturo-spécifiques propres à un territoire donné qui contribuent à la création de l’identité de ce territoire ». Originaux et uniques, ces référents n’ont généralement pas d’équivalents dans les autres langues, étant donné qu’ils sont spécifiques à une culture et à un territoire donnés (*loc. cit.*). Déjà problématiques pour les traducteurs professionnels, les *realia* le sont encore davantage pour les programmes de TA.

[Leur] traduction [...] demande la mise en œuvre de compétences variées associées aux différentes stratégies envisageables : la compréhension du référent culturo-spécifique que le mot

désigne et la capacité de reformulation et/ou de recreation dans les cas d'adaptation qui solliciteraient la créativité du traducteur. (*Ibid.*, p. 228)

Les réseaux de neurones qui constituent les modèles de TAN sont-ils à la hauteur en termes de décodage des *realia* et de créativité ? Le traducteur présente l'avantage de pouvoir se demander si c'est la *domestication* — c'est-à-dire l'adaptation de ce référent à la culture cible — ou la *foreignization*, soit le maintien du caractère étranger du texte source, qui est recherchée (Venuti, 2018). La machine, qui ne peut pas se poser cette question, n'est souvent pas apte à décoder ces culturèmes et à produire une traduction correcte. Besacier (2014) illustre notamment un cas dans lequel la TA d'un 'repère culturel' s'est révélée problématique ou même erronée : la machine avait traduit *Charles River* par l'expression *la rivière Charles*. Cette traduction est fautive, car il ne s'agit pas d'une rivière, mais d'un fleuve. Bien que Besacier utilise l'expression 'repère culturel', il s'agit ici plutôt d'un toponyme. Pour résoudre ce problème, Besacier (*loc. cit.*) propose la traduction — ou plutôt le report, comme le suggère Ballard (1998, pp. 201–202) — *la Charles River* ou *la Charles*. Toutefois, les programmes de TA trouvent parfois des traductions intéressantes (Zotti, 2019, p. 238) ; cela dépend évidemment des corpus sur lesquels ils sont entraînés. Mais bien souvent, la TA propose des « équivalents farfelus » (*loc. cit.*) : « [Il est] surprenant de constater que les traductions de ces mots définis ne [sont] pas correctes dans la plupart des cas et reçoivent même des équivalents traductionnels invraisemblables, voire humoristiques. » (*Loc. cit.*)

Un autre aspect qui se révèle souvent problématique est la question du vouvoiement et du tutoiement (Burchardt et al., 2016, p. 215). L'anglais est particulièrement problématique à cet égard. En effet, le pronom *you* peut avoir trois traductions différentes — mais quatre sens différents — en français : *tu*, *vous* (soit à la 2^e personne du pluriel, soit à la forme polie) et *on*. Lors d'une expérience pédagogique, Schumacher (2020, p. 76) observe des irrégularités concernant l'usage du tutoiement et du vouvoiement dans une TA de l'anglais vers le français. La machine passe alors de manière aléatoire du vouvoiement au tutoiement, dans un discours de Donald Trump au peuple américain. Au début du texte, la machine traduit correctement le pronom *you* par *vous*, étant donné que Trump s'adresse à l'ensemble de ses citoyens. Mais à la fin, la machine traduit *You know that* par *Tu le sais bien*, preuve du manque de cohérence de la machine. Le problème, dans le cas du tutoiement et du vouvoiement, n'est pas tellement lié au fait que la machine n'est pas capable d'analyser le contexte : « Un exemple simple est celui du vouvoiement français vu depuis l'anglais : le

choix entre le *tu* et le *vous* n'est pas toujours déductible du contexte proche. » (Segura, 2012, p. 28). En effet, pour traduire correctement le *you* anglais en français ou en allemand, la machine devrait être capable d'observer les relations entre les différents intervenants, les coutumes, les contraintes sociales et les postulats anthropologiques qui existent dans la culture cible. Le recours au vouvoiement en allemand est beaucoup plus fréquent qu'en espagnol ou en italien, par exemple, qui sont des langues dont les locuteurs utilisent plus spontanément le pronom *tu* (Ventura & Bajrić, 2020, pp. 66–67). De l'allemand vers le français, cet aspect ne devrait pas être aussi problématique, étant donné que les deux langues disposent d'une forme polie (les pronoms *Sie* en allemand et *vous* en français). Toutefois, Burchardt et al. (2016, p. 215) ont observé un cas particulier où le pronom allemand *sie* accompagné d'un verbe au singulier (3^e personne du singulier au féminin) avait été traduit en anglais par la machine par le pronom pluriel *they*, soit la 3^e personne du pluriel, alors qu'elle aurait dû utiliser le pronom *she*. Ceci montre que, bien que l'allemand possède des pronoms distincts, leur TA n'est pas toujours correcte et cohérente.

Souvent, le traducteur se retrouve face à un dilemme : doit-il traduire ou non les répétitions ? Pour la TA, le dilemme est le même, bien que le choix de la machine ne soit pas motivé comme il pourrait l'être avec un biotraducteur. Youdale (2020, pp. 23–24) évoque l'un des avantages de la TA par rapport au traducteur humain. Parfois, même le traducteur ou le lecteur le plus attentif ne parvient pas à repérer les répétitions, car elles peuvent être subtiles. La machine, en revanche, se distingue de l'homme précisément à cet égard, étant donné qu'elle peut très facilement détecter les mots, expressions ou structures syntaxiques qui se répètent. Cependant, comment la machine peut-elle « décider » s'il convient de maintenir la répétition ou s'il est préférable de l'effacer, alors que ce choix peut déjà s'avérer complexe pour un traducteur professionnel ? Dans son article « Traduire ou ne pas traduire les répétitions », Prak-Derrington (2011) expose les deux stratégies parmi lesquelles les traducteurs doivent choisir, à savoir la traduction ou non des répétitions. Prak-Derrington constate qu'il y a généralement une suppression, un effacement de la répétition, ce qui signifie par conséquent qu'une variation apparaît à l'endroit où l'auteur original avait fait le choix de la répétition (*ibid.*, p. 1). Il convient toutefois de s'interroger sur ce réflexe que Kundera (1993, p. 130) qualifie de « réflexe de synonymisation » :

Le besoin d'employer un autre mot à la place du plus évident, du plus simple, du plus neutre [...] pourrait s'appeler réflexe de synonymisation, réflexe de presque tous les traducteurs. Avoir une grande réserve de synonymes, cela fait partie de la virtuosité du « beau style » [...]. Ce besoin

de synonymiser s'est incrusté si profondément dans l'âme du traducteur qu'il choisira tout de suite un synonyme [...]. Cette pratique synonymisatrice a l'air innocente, mais son caractère systématique émousse inévitablement la pensée originale.

Prak-Derrington (2011, p. 2) assimile même ce réflexe de synonymisation au mouvement des Belles infidèles, car supprimer la répétition ne reflète pas la pensée et la logique de l'auteur original. Parfois, la répétition peut aussi avoir une fonction particulière, ce qui est notamment le cas de l'anaphore. L'anaphore est un « procédé visant à un effet de symétrie, d'insistance, etc., par répétition d'un même mot ou groupe de mots au début de plusieurs phrases ou propositions successives » ('Anaphore', *CNRTL*). Un exemple très célèbre d'anaphore est celui qui provient d'un discours de Charles de Gaulle, lors duquel il prononce les mots suivants : « Paris ! Paris outragé ! Paris brisé ! Paris martyrisé ! Mais Paris libéré ! » (Charles de Gaulle, 1944). Dans un tel cas, la reproduction de la répétition est en effet nécessaire, étant donné que la répétition entraîne un effet d'insistance, qui peut avoir différentes vocations en fonction du contexte. Bien que la machine puisse détecter très facilement les répétitions, comment peut-elle « décider » s'il convient de les traduire ou non ?

Ensuite, comment la machine traite-t-elle les marqueurs de genre ? D'une langue à l'autre, l'utilisation de ces marqueurs peut considérablement changer. Comment traduire d'une langue comme l'anglais, par exemple, qui possède peu de marqueurs de genre, vers l'allemand, qui est une langue très inclusive ? Ces questions, déjà cruciales pour le biotraducteur, le sont davantage lorsque la traduction repose sur des réseaux de neurones. Même s'il existe souvent des stratégies pour contourner cette difficulté, il arrive qu'il n'y en ait pas. Matusov (2022) évoque l'exemple du russe, une langue dans laquelle la forme du passé dépend du genre, ce qui peut engendrer de nombreux soucis lors d'une traduction, surtout automatique, car la machine pourrait manquer de co(n)texte et, par conséquent, attribuer le mauvais genre à une personne. Kuzman (2022) a étudié la traduction automatique de romans de l'anglais vers le slovène. Lors de l'analyse, elle a également relevé un problème en ce qui concerne la question des genres : le système de TAN de Google avait modifié le genre d'un personnage. Elle note, toutefois, que cette étude date d'avant 2020 ; l'on pourrait, par conséquent, s'attendre à ce que la machine ait progressé à cet égard. En développant un modèle de traduction automatique anglais-français entraîné sur des données de jeux vidéo, Hansen & Houlmont (2022, pp. 263–264) ont, quant à eux, réussi à « apprendre » à la machine comment traiter les marqueurs de genre. Alors que ce n'est pas

le cas en anglais, les mots utilisés pour décrire les personnages de jeux vidéo ont souvent un genre en français, et la machine, grâce à un entraînement sur des données pertinentes, a appris les bonnes stratégies pour pallier ce problème.

[...] The common translation strategy is to neutralize the potential gender pitfall by deleting every direct mention of gender [...], not using tenses that require a feminine or masculine form [...], omitting gendered words [...] or using generic terms/paraphrases in their stead [...]. Our custom engine has applied this strategy systematically, interestingly showing that NMT can learn not only specific vocabulary but also translation strategies that can anticipate common sources of error [...]. (*Loc. cit.*)

Néanmoins, il s'agit d'un système adapté à la traduction de jeux vidéo, ce cas ne peut donc pas être généralisé. En effet, Mohar et al. (2020, p. 130) ont, de leur côté, observé des erreurs dans la traduction des genres lors d'une analyse de la TA d'un texte de Margaret Atwood. Parmi les nombreuses failles qu'ils ont relevées, l'on peut retrouver des erreurs dans la traduction des genres des pronoms personnels, notamment le pronom anglais *I*, que la machine a considéré comme masculin. Cette traduction du pronom personnel sujet, censé être féminin, a donc engendré des problèmes de sens, étant donné que, comme en russe, la forme passée des verbes dépend du genre du pronom. Ces observations montrent à quel point la question du genre peut se révéler épineuse pour la TAN. D'une manière similaire, la machine renforce généralement les biais, notamment les biais de genre (Kayser-Bril, n. d.). En effet, Vanmassenhove et al. (2021, p. 2204) observent :

Especially in a field such as MT, the implicit gender in a language such as English and its consecutive translations into morphologically richer languages with gender agreement, makes it relatively easy to expose and study biases related to gender in a contrastive linguistic setting.

Kayser-Bril (n. d.) donne l'exemple de Google Translate, qui, au début des années 2010, avait traduit la phrase « Men should clean the kitchen » par « Frauen sollten die Küche sauber [machen] » en allemand. Ce phénomène semble être dû à une représentation inégale, dans les données d'entraînement, de femmes et d'hommes dans ce contexte. Kayser-Bril explique que la TA s'est certes améliorée à cet égard, mais que « de nombreux problèmes persistent » (ma traduction).

Une autre question, très intimement liée à la notion de créativité, concerne la traduction automatique des néologismes et, dans la même logique, des jeux de mots. Lors d'une étude sur la créativité en traduction, Guerberof-Arenas & Toral (2022) ont mesuré les

scores de créativité de trois différentes traductions : la première étant une traduction humaine ; la deuxième, une PE ; la dernière, une TA. Le constat est clair : la traduction qui enregistre les meilleurs scores de créativité est la traduction humaine, suivie de la PE et, enfin, de la TA. Ceci implique que la machine est loin d’être au niveau du biotraducteur en termes de créativité. Le néologisme étant un « mot de création récente ou emprunté depuis peu à une autre langue ou toute acception nouvelle donnée à un mot ou à une expression qui existaient déjà dans la langue » (‘Néologisme’, *Larousse*), l’on peut s’attendre à ce qu’il soit plutôt mal traité par la machine. En effet, entraîné sur des corpus (énormes, certes), le système de TA a du mal à créer de nouveaux mots, de nouvelles acceptions ou de nouvelles expressions. Mohar et al. (2020, p. 138) le confirment :

Both tools used in this research [Google Translate and Translator.eu] failed in translating words coined by Atwood or used in her text in an unusual way, for instance, the adjective ‘well-thumbed tears.’

Une étude de Shafiull & Awadh (2020) démontre également à quel point la traduction de néologismes est une tâche complexe, à la fois pour les traducteurs humains (dans ce cas, des étudiants) et pour la machine. Toutefois, la machine se montre plus mauvaise que les étudiants à cet égard (*ibid.*, p. 7). Alors que certains étudiants parviennent à trouver des néologismes équivalents acceptables, la machine n’y arrive pas. Shafiull & Awadh (*ibid.*, pp. 7–8) expliquent que cette difficulté est probablement due à l’inutilité des dictionnaires, mais aussi des corpus sur lesquels est entraînée la machine, en ce qui concerne la traduction de telles expressions. Dans la catégorie des néologismes, l’on retrouve aussi les acronymes non attestés. Alors que certains étudiants parviennent à trouver la bonne traduction, ce type de néologismes représentent un obstacle majeur pour la machine qui, bien souvent, ne les traduit pas (*loc. cit.*). D’une manière assez similaire, les représentations abstraites, imagées, et les métaphores représentent elles aussi un défi majeur pour la TA. Mohar et al. (2020, pp. 138–139) illustrent cette hypothèse à l’aide d’une phrase : « The lives of the lives in question had their chances, most of which they blew. » (Atwood, 2006, p. 3). Entendu dans son acception courante, le verbe anglais *blow* signifie ‘souffler’ ou ‘exploser’. En revanche, dans cette phrase, il a une signification imagée : la collocation *blow their chances* veut dire ‘gâcher leurs chances’. Les deux traducteurs automatiques que Mohar et al. ont utilisés lors de cette expérience, à savoir Translator.eu et Google Translate, ont traduit le verbe anglais *blow* par ‘exploser’ (en slovène) (Mohar et al., 2020, p. 139). Ce phénomène peut s’expliquer assez simplement. Les performances de la machine dépendent des corpus

sur lesquels elle est entraînée. Étant donné que les mots sont employés de manière imagée, abstraite, et qu'ils sont détournés de leur acception originelle, la traduction que la machine retrouve dans sa base d'exemples est celle qui correspond à la signification première du terme.

3.3. Traduire l'oralité : quelles difficultés pour la TA ?

Dans cette partie du travail sont abordés certains aspects susceptibles de se retrouver dans un texte oral (comme les transcriptions de discours, par exemple). Il s'agit d'exposer de quelle manière ceux-ci peuvent se révéler problématiques au moment de traduire un tel texte avec un programme de TA. Pourquoi, en règle générale, les marqueurs d'oralité sont-ils difficiles à traiter par les systèmes de TA ? Tout d'abord, ces systèmes sont en majeure partie entraînés sur des corpus de textes écrits relevant d'un registre courant, voire soutenu. Par conséquent, les aspects oraux des transcriptions de discours ne sont pas retrouvés en assez grand nombre dans les données d'entraînement pour que la machine produise une traduction adéquate (Burchardt et al., 2016, p. 212). En outre, obtenir des données d'entraînement 'orales', surtout si les langues concernées sont peu représentées, n'est pas simple (*loc. cit.*). Cette partie a été plus difficile à élaborer, car la TA de marqueurs d'oralité semble peu étudiée jusqu'ici. Dès lors, la plupart des références concernent principalement la traduction littéraire. Toutefois, certains des aspects abordés dans les travaux que j'ai lus peuvent être mis en lien avec des éléments du langage oral. En cela, l'état de l'art de la traduction automatique des particularités du langage oral s'établit sur la base de recherches en TLA, avec quelques éléments provenant de recherches sur la TA audiovisuelle et la TA de jeux vidéo.

Hansen & Houlmont (2022, p. 263) expliquent que le registre représente un certain obstacle pour la TA. Lorsqu'ils ont traduit deux jeux vidéo, l'un relevant d'un registre assez soutenu et l'autre familier, voire grossier, la TA a « neutralisé » les caractéristiques des deux registres. Burchardt et al. (2016, p. 212) ont fait une observation similaire et ont évoqué le cas particulier du langage oral, de la transcription de discours et de dialogues.

[...] AVT [audiovisual transcription] of subtitles and dubbing scripts, by its very nature, deals with written representation of spoken dialogue and has characteristics that can make it difficult for MT. [...] This situation creates a whole set of new challenges for MT. (*Loc. cit.*)

Il est alors clair que le caractère oral des transcriptions de discours et de dialogues peut fortement affecter la qualité de la TA. Outre les aspects qui ont déjà été exposés aux points 3.1. et 3.2., il existe en effet toute une série de paramètres propres à la retranscription de discours qui peuvent s'avérer particulièrement difficiles à traduire pour la machine.

D'abord, le phénomène des phrases incomplètes est problématique pour la TA, et les textes oraux sont plus susceptibles de comporter un nombre assez important de phrases incomplètes (Burchardt et al., 2016, p. 213). Hansen et Houlmont (2022, p. 264) expliquent, en effet, que le langage oral, riche en pauses et en phrases incomplètes, pose un réel problème à la machine. Ces phrases sont presque automatiquement complétées, ce qui entraîne naturellement des erreurs de sens pouvant se révéler problématiques. Comme l'explique Bouamor (2010, p. 52), il est possible que les corpus soient nettoyés lorsqu'ils sont constitués, c'est-à-dire que les éventuelles fautes d'orthographe, erreurs de traduction et phrases incomplètes peuvent être corrigées ou complétées. Ceci implique que les corpus ne contiendraient plus de phrases incomplètes, d'où leur traitement problématique par la TA.

Le problème que posent les phrases longues a été évoqué au point 3.1., mais il est important de rappeler que les phrases (trop) courtes peuvent elles aussi s'avérer complexes à traiter pour la machine (Mohar et al., 2020, p. 130). Il existe une raison qui pourrait expliquer ce phénomène. En effet, il se peut que, lorsque confrontée à une phrase très courte, la machine manque de co(n)texte et ne puisse, par conséquent, pas produire une traduction correcte, que ce soit d'un point de vue grammatical ou sémantique.

Se pose également la question de la traduction des marqueurs pragmatiques (aussi appelés marqueurs discursifs, mots du discours ou particules discursives), qui peuvent apparaître dans les textes oraux. Il s'agit de « petits mots de la langue parlée » qui ont une valeur et une fonction particulière (Dostie & Pusch, 2007, pp. 3–4).

[Les marqueurs pragmatiques] ne contribuent pas au contenu propositionnel des énoncés et c'est pourquoi leur présence ou leur absence ne modifie pas la valeur de vérité des énoncés auxquels ils sont joints. Ils ont tendance à constituer des unités prosodiques indépendantes [...]. [Ils] n'entrent pas dans une structure argumentale et ils peuvent occuper différentes positions par rapport à un énoncé, s'ils ne sont pas utilisés comme mots-phrases. (*Ibid.*, p. 4)

En français, l'on peut trouver parmi les marqueurs pragmatiques *ben*, *bon*, *là*, *heu*, *hein*, *tu vois*, *bien*, *donc*, *enfin*, *si vous voulez*, *toujours* (dans le sens de « tu peux *toujours* demander ») (*ibid.*, pp. 3–10). Quant à l'anglais, il y a notamment les marqueurs

pragmatiques *I mean, oh, so, you see, well, anyway, because, you know* et *uh* (B. Fraser, 1988, pp. 19–32). En allemand, les marqueurs pragmatiques les plus répandus sont les particules modales (Dostie & Pusch, 2007, p. 6), qui seront traitées plus tard dans ce travail. Liefländer-Koistinen (2004, p. 550) explique que les particules modales allemandes sont liées à la situation et rendent compte de la réaction du locuteur à ce qui a été précédemment dit ; elles se rapportent à tout un énoncé, et non à un mot en particulier, et donnent une valeur supplémentaire à ce qui est dit. En bref, la particule modale allemande vient nuancer un énoncé. Schoonjans (2014, p. 402) identifie 18 particules en allemand : *bloß, aber, auch, doch, denn, eben, eh, einfach, eigentlich, etwa, halt, mal, ja, nur, schon, ruhig, wohl* et *vielleicht*. En se concentrant uniquement sur les particules modales *ja* et *doch*, il explique qu'en fonction du co(n)texte, ces particules peuvent apporter de nombreuses nuances : la justification, l'explication, le reproche, la concession, la surprise, et plus encore (*ibid.*, p. 403). Si, rien que pour ces deux particules, les possibilités de traduction sont immenses, l'on peut effectivement s'attendre à ce que les systèmes de TA rencontrent des difficultés lorsqu'ils sont confrontés à des marqueurs pragmatiques quels qu'ils soient. Burchardt et al. (2016, p. 213) affirment, en effet, que les textes oraux peuvent être constitués de nombreux marqueurs discursifs, qui ne sont pas souvent retrouvés dans les données d'entraînement. Par conséquent, les systèmes de TA peuvent avoir du mal à rendre la nuance de ces marqueurs.

Par ailleurs, les problèmes de syntaxe peuvent eux aussi considérablement affecter la qualité d'une TA. C'est le cas, par exemple, de la syntaxe elliptique ou non conventionnelle, aspect propre à l'oralité. Dans leur étude concernant la TA de *Life Stories* de Margaret Atwood, Mohar et al. (2020, pp. 131–133) observent comment plusieurs phrases à la syntaxe non conventionnelle ont été traduites par deux programmes de TA (Google Translate et Translator.eu). La première phrase dont ils ont analysé la traduction des deux programmes est celle-ci : « *I was born, I would have begun, once* » (Atwood, 2006, p. 3). Cette phrase est problématique à plusieurs égards. Premièrement, Atwood a volontairement omis les guillemets autour de 'I was born', ce qui rend la lecture difficile. Deuxièmement, déterminer comment ces deux phrases sont liées est une tâche compliquée. Troisièmement, il est difficile de savoir sur laquelle des deux propositions l'adverbe 'once' porte. Cette phrase à la syntaxe peu conventionnelle est par conséquent difficile à décoder. En réalité, la première proposition est un discours direct ; la deuxième, l'incise déclarative ; et l'adverbe 'once' est à relier à la deuxième proposition. Les deux systèmes de TA ont rencontré des difficultés à

traduire cette phrase. En effet, ils ont considéré les deux phrases comme des propositions indépendantes ; or, elles sont intrinsèquement liées. Les phrases à la syntaxe ambiguë ou non conventionnelle représentent par conséquent un défi pour la TA (Mohar et al., 2020, p. 131). Il est assez simple de le confirmer si l'on observe comment DeepL traduit cette phrase (en octobre 2022) : « Je suis né, j'aurais commencé, une fois. » Burchardt et al. (2016, p. 211) ont observé un phénomène similaire dans le cadre de leur recherche. Une phrase elliptique — même averbale, dans ce cas — comme « Wieder ein Wochenende vorbei » influencera elle aussi la qualité de la TA. La machine a traduit cette proposition comme suit en anglais : « Again a weekend pass. » En revanche, le système de TA traduit correctement la phrase complète lorsque le verbe est réintroduit. Dans la même logique, les propositions indépendantes juxtaposées par des virgules — ce qui forme des phrases à rallonge — peuvent également se révéler complexes à traduire avec un système de TA (Mohar et al., 2020, p. 132). Dans l'exemple qui suit, la syntaxe non conventionnelle présente dans une phrase très longue a été difficile à traiter par la machine également.

They should have spotted the photographer in the bushes, they shouldn't have chewed with their mouths open, they shouldn't have worn the strapless top, they shouldn't have yawned, they shouldn't have laughed: so unattractive, the candid denture. (Atwood, 2006, p. 3)

La première erreur des deux programmes de TA a été de ne pas considérer toutes les propositions comme une unité. Le système Translator.eu produit une traduction qui n'a aucun sens ; Google Translate traduit cette phrase un peu mieux, mais sa traduction reste médiocre. Mais ce qui a surtout posé problème est la dernière partie de la phrase, qui est une proposition elliptique. En effet, à cause de la post-modification de l'adjectif, les systèmes de TA ne sont pas en mesure d'identifier la relation entre l'adjectif 'unattractive' et le nom 'denture'. Cette mauvaise compréhension engendre une erreur grammaticale, puisque l'adjectif n'est pas accordé avec le nom (Mohar et al., 2020, p. 132). Toutes ces analyses corroborent ainsi ce que Tezcan et al. (2019) ont observé : plus la phrase est longue et complexe, plus la qualité de la TA sera médiocre. Le calque syntaxique, ou le fait de reproduire la syntaxe du TS dans la traduction, reste un défaut aussi bien pour le biotraducteur que pour la machine (Besacier, 2014, p. 5), comme le montrent les exemples cités précédemment.

Pour finir, les transcriptions de discours ou d'autres textes oraux peuvent causer des problèmes de références pronominales, que l'on appelle également anaphores pronominales.

En effet, dans de longues phrases avec beaucoup d'informations, les pronoms peuvent rapidement devenir ambigus, et le traducteur, mais aussi la machine, peuvent facilement perdre le fil. Comme l'expliquent Burchardt et al. (2016, p. 211), les phénomènes qui requièrent une connaissance extralinguistique sont susceptibles de pousser la TA à ses limites ; c'est notamment le cas des pronoms ambigus et des anaphores pronominales. Hardmeier & Federico (2010, p. 283) définissent les anaphores pronominales comme « l'utilisation de pronoms en vue de faire référence à quelque chose de mentionné précédemment dans le discours ». À ce niveau, le biotraducteur est avantagé par rapport à la machine. En effet, il a la capacité de retourner quelques lignes plus haut pour identifier le référent du pronom, ce que la machine n'est pas toujours capable de faire, surtout lorsque les phrases sont longues et que la distance entre le pronom et son référent est importante (Hardmeier & Federico, 2010, p. 283 ; Mohar et al., 2020, p. 134). Par exemple, Mohar et al. ont tenté de faire traduire ce passage par la TAN : « I'm working on my own life story. I don't mean I'm putting it together; no, I'm taking it apart » (Atwood, 2006, p. 3). Lors de leur analyse, ils ont remarqué que la machine avait rencontré un problème avec la traduction du pronom *it*, se référant à *my own life story* ; en effet, elle l'avait traduit par un pronom masculin, alors que le substantif est féminin (Mohar et al., 2020, p. 134). Ils en ont alors conclu que la machine n'était, en général, pas capable de prendre en compte le contexte au-delà de la phrase, ce qui, par conséquent, pose des problèmes lors de la traduction des pronoms lorsque leur référent est éloigné. Lors de leur étude sur la TA des anaphores pronominales de l'allemand vers l'anglais, Hardmeier & Federico (2010, p. 284) ont formulé les observations suivantes : la TA traduit correctement les pronoms dans à peine plus de 60 % des cas ; pour les pronoms démonstratifs, elle produit une traduction correcte dans 90 % des cas ; en revanche, en ce qui concerne les pronoms personnels de politesse ainsi que les pronoms réfléchis, la TA est particulièrement mauvaise. Il convient toutefois de préciser que, bien que ces études soient relativement récentes, le nouveau paradigme qu'est la TAN évolue très rapidement. Par conséquent, ces problèmes pourraient être bien mieux traités aujourd'hui.

La traduction de transcriptions de discours ou d'autres documents relevant d'un registre oral est, en conclusion, un défi de taille pour la machine. Le fond du problème est commun à (presque) tous les aspects détaillés ci-dessus.

If the MT engine translating the text has been trained on traditional written text, the features used in spoken text may not be represented accurately in the training data and the engine will therefore

have no relevant examples from which it can produce an accurate translation. [...] One issue that arises here is the relative difficulty of obtaining such a corpus, particularly in lesser-resourced language pairs. (Burchardt et al., 2016, p. 212)

L'idéal serait, en effet, d'enrichir les systèmes de TAN avec des données similaires. Cependant, comme expliqué précédemment, ce type de données n'est pas facile à trouver. De plus, cela pose certains problèmes, comme évoqués au point 2.3., notamment en ce qui concerne la durabilité et la qualité de ces données.

Par ailleurs, la TAN est un paradigme en constante évolution. Bien que les recherches sur la base desquelles est construite cette partie n'appartiennent pas à un passé lointain, le développement exponentiel de cette technologie pourrait avoir pour conséquence que toutes ces observations ne soient plus d'actualité. Ce travail permettra notamment de vérifier si, malgré son évolution rapide, la machine rencontre toujours autant de difficultés lors de la traduction de certains aspects détaillés ci-dessus.

4. Méthodologie

L'objectif de ce travail étant d'observer comment la TAN, en l'occurrence le système de DeepL, traduit certains aspects propres à l'oralité, il était nécessaire de constituer un corpus pertinent. Pour la compilation de celui-ci, une série de critères étaient déterminants.

Tout d'abord, il était nécessaire d'opter pour une paire de langues. D'une part, pour ne pas totalement dépayser le programme de TA, j'ai choisi deux langues dont la machine devrait posséder suffisamment de ressources dans ses corpus d'entraînement, à savoir l'allemand pour la langue source et le français pour la langue cible. D'autre part, une autre raison pour laquelle j'ai opté pour cette paire de langues est la suivante. Comme expliqué au point 3.1., il y a une certaine distance entre ces deux systèmes, l'une des langues étant germanique et l'autre romane. De cette manière, la machine ne part ni avec un trop gros handicap ni avec un avantage notoire.

Par ailleurs, il a fallu choisir un type de textes ni trop écrit ni trop oral, pour ne pas tomber dans le stéréotype de l'oralité, c'est-à-dire dans une représentation démesurée des caractéristiques de l'oralité. Une piste sur laquelle il semblait judicieux de se pencher était, par conséquent, les discours du monde politique, qui devraient *a priori* rester assez soutenus. Toutefois, afin d'assurer un certain caractère oral, j'ai fait le choix de me pencher sur des conférences de presse du gouvernement fédéral allemand. En effet, l'on peut retrouver des retranscriptions fidèles de différentes conférences de presse directement sur la page Internet du gouvernement allemand¹. La raison pour laquelle j'ai décidé de sélectionner des conférences de presse (plutôt que des discours, par exemple) est que celles-ci fonctionnent de manière spontanée : un intervenant pose une question, et un membre du *Bundesregierung* répond sur-le-champ, sans avoir le temps de préparer une réponse déjà toute faite. Ceci permet d'assurer une certaine spontanéité, et, par conséquent, une représentation de certains éléments d'oralité, contrairement aux discours préparés.

Afin de délimiter le corpus, j'ai sélectionné des conférences de presse traitant toutes d'un seul et même sujet, soit l'invasion de l'Ukraine par la Russie, thème le plus actuel au moment de créer le corpus. Par souci de temps, je n'ai choisi que quatre conférences : celles du 28 février, des 2, 7 et 14 mars 2022. Au total, le corpus allemand comprend 32 899 mots.

¹ <https://www.bundesregierung.de>

Après avoir créé un document reprenant les quatre conférences de presse, je l'ai fait traduire par DeepL. La TA date d'octobre 2022. La traduction du corpus compte 38 705 mots.

Pour la suite des opérations, j'ai créé un bitexte aligné sur la page Internet de *YouAlign* (voir Image 1, page 28), programme appartenant à *Terminotix*. J'ai ensuite exporté ce bitexte au format TMX (*Translation Memory Exchange*). C'est ainsi qu'a été constitué mon corpus parallèle, contenant d'une part le texte allemand et d'autre part sa TA en français.

Avant de passer à l'exploitation de ce corpus parallèle, j'ai analysé manuellement le TS afin de repérer les éléments d'oralité sur lesquels j'allais me concentrer dans le cadre de ce travail. Ces aspects seront définis et détaillés au point suivant.

Enfin, après avoir identifié les éléments oraux, j'ai exploité mon corpus parallèle sur le logiciel *Xbench* (voir Image 2, page 28). Celui-ci est souvent utilisé comme outil d'aide à la traduction. Grâce à une fonctionnalité de recherche terminologique, on peut trouver la traduction d'un mot, d'une expression ou d'une phrase se trouvant dans le corpus parallèle avec lequel on a alimenté le programme. Cet usage n'est pas celui que je souhaitais en faire. Néanmoins, j'ai jugé utile d'utiliser ce logiciel, étant donné que, comme j'avais déjà pré-identifié les segments sur lesquels concentrer mon analyse, il me suffisait de les rechercher grâce à cette fonctionnalité.

Image 1

Télécharger l'alignement complet : [format TMX](#) [format LogiTerm \(HTML\)](#)

Allemand	Français
Conférence0203_de.docx	Conférence0203_fr.docx
Regierungs (...)	Conférence (...)
Sprecher: (...)	Intervenant (...)
Vorsitzende Welty eröffnet die Pressekonferenz und begrüßt SRS'in Hoffmann sowie die Sprecherinnen und Sprecher der Ministerien.	La présidente Welty ouvre la conférence de presse et souhaite la bienvenue à SRS'in Hoffmann ainsi qu'aux porte-parole des ministères.
Vorsitzend (...)	Présidente (...)
Kall: Herr (...)	Kall : Mon (...)
Aber wie Sie wissen, gibt es keine Binnengrenzkontrollen, jedenfalls keine regelmäßigen Binnengrenzkontrollen, sondern nur stichprobenartige.	Mais comme vous le savez, il n'y a pas de contrôles aux frontières intérieures, en tout cas pas de contrôles réguliers aux frontières intérieures, mais seulement des contrôles aléatoires.
Deswegen i (...)	C'est pour (...)
Frage: Her (...)	Question : (...)
Das Bundes (...)	L'Office f (...)
Von welchen Szenarien, welchen Zahlen gehen Sie zurzeit aus?	De quels scénarios, de quels chiffres vous basez-vous actuellement ?
Kall: Aus Sicht des Bundesinnenministeriums ist es nach wie vor nicht	Kall : Du point de vue du ministère fédéral de l'Intérieur, il n'est

LOGITERM

ALIGNFACTORY

BRIDGETERM

SYNCHROTERM

TRANSEARCH

Copyright © Terminotix Inc. 2009-2022

Capture d'écran du fichier aligné obtenu sur YouAlign², à exporter au format TMX

Image 2

Source	#	Target
Wie sieht das die Bundesregierung?	1	Comment le gouvernement fédéral voit-il les choses ?
Was hält die Bundesregierung davon?	1	Qu'en pense le gouvernement allemand ?
Unterstützt die Bundesregierung dies?	1	Le gouvernement fédéral soutient-il cette démarche ?
Wie reagiert die Bundesregierung darauf?	1	Comment le gouvernement fédéral réagit-il à cela ?
Schließt sich dem die Bundesregierung an?	1	Le gouvernement fédéral s'y associe-t-il ?
Wer die Zustimmung der Bundesregierung nötig?	1	L'accord du gouvernement fédéral était-il nécessaire ?
Davon wird die Bundesregierung nicht abrücken.	1	Le gouvernement fédéral ne s'en écartera pas.
Die Bundesregierung ist dafür nicht zuständig.	1	Le gouvernement fédéral n'est pas compétent en la matière.
Wie sehen Sie das seitens der Bundesregierung?	1	Comment voyez-vous cela du côté du gouvernement fédéral ?
Wird sich die Bundesregierung daran beteiligen?	1	Le gouvernement fédéral y participera-t-il ?
Die Bundesregierung kommentiert das jetzt nicht.	1	Le gouvernement fédéral ne commente pas cette décision.
Zusatz: Sie richtet sich an die Bundesregierung	1	Complément : Elle s'adresse au gouvernement fédéral.
Was sagt die Bundesregierung über diesen Schritt?	1	Que pense le gouvernement fédéral de cette démarche ?
Das kann die Bundesregierung natürlich nur begrüßen.	1	Le gouvernement fédéral ne peut évidemment que s'en réjouir.
Das Gas wird im Auftrag der Bundesregierung eingekauft.	1	Le gaz est acheté pour le compte du gouvernement fédéral.
Wie steht die Bundesregierung zu einem solchen Embargo?	1	Que pense le gouvernement fédéral d'un tel embargo ?
Wie geht die Bundesregierung mit dieser realen Gefahr um?	1	Comment le gouvernement fédéral gère-t-il ce risque réel ?
Säverin: Die Bundesregierung arbeitet mit Hochdruck daran.	1	Säverin : Le gouvernement fédéral y travaille d'arrache-pied.
Wie will es die Bundesregierung schaffen, dass das aufhört?	1	Comment le gouvernement fédéral va-t-il faire pour que cela cesse ?

116 entries found. 19 entries displayed. 62 ms elapsed.

Capture d'écran du logiciel Xbench : recherche d'une expression et affichage des entrées trouvées (TS et TC alignés)

² <https://youalign.com/>

5. Analyse du corpus parallèle

Cette partie est consacrée à l'analyse de la TA de DeepL d'octobre 2022. Celle-ci porte sur certains aspects mentionnés dans l'état de l'art. En effet, il convient d'observer comment DeepL traite en particulier les éléments propres à l'oralité, c'est-à-dire entre autres :

- les phrases longues (de plus de 40 mots³),
- les phrases à la syntaxe non conventionnelle,
- les phrases incomplètes ou coupées,
- les phrases averbales,
- les anaphores et les répétitions,
- les expressions et mots étrangers,
- certains marqueurs pragmatiques tels que les particules modales et les déictiques.

Dans un premier temps, j'analyserai deux extraits relativement longs du corpus parallèle contenant une discussion entière et non uniquement des interventions isolées, ce qui me permet d'avoir une vue d'ensemble sur deux passages du corpus et ainsi de pouvoir observer le texte en tant qu'unité, c'est-à-dire d'observer la cohésion et la cohérence textuelles. Ensuite, j'illustrerai mes observations au moyen d'exemples (ou de contre-exemples) isolés, afin de voir s'il est possible de tirer des conclusions générales sur le fonctionnement de DeepL. L'objectif est effectivement d'analyser s'il existe des phénomènes récurrents dans une TA de DeepL et d'évaluer leur(s) effet(s) dans la TA.

5.1. Hypothèse

L'état de l'art établi au point 3 de ce travail a permis de démontrer que, bien que la TAN soit en constante évolution, elle produit encore souvent des traductions bancales, voire mauvaises, lorsqu'elle fait face aux aspects mentionnés ci-dessus. Par conséquent, l'on peut s'attendre à ce que quelques problèmes de sens, de syntaxe ou de style surviennent à ces endroits-là. En effet, étant donné que DeepL, ainsi que d'autres programmes de TA, sont alimentés principalement par des textes écrits, les textes oraux ne seraient *a priori* pas appropriés à la TA. Toutefois, au vu du développement rapide de la TAN, il est possible que DeepL produise à certains moments une traduction acceptable. Dès lors, l'un des objectifs sera également d'expliquer pourquoi ces éléments sont tantôt bien, tantôt mal traités, ou, *a contrario*, de conclure qu'il s'agit dans certains cas d'une question de hasard.

³ Au vu du manque de définition scientifique de la notion de « phrase longue », j'ai décidé de fixer cette limite.

5.2. Analyse d'extraits

	TS (Conférence de presse du 28/02/2022)	TA (DeepL, octobre 2022)
1	<i>Ist also gewährleistet, dass trotz diesen partiellen SWIFT-Aussetzungen Zahlungen bei den russischen Lieferanten ankommen können, jetzt?</i>	<i>Est-il donc garanti que, malgré ces suspensions partielles de SWIFT, les paiements peuvent parvenir aux fournisseurs russes, maintenant ?</i>
5	<i>Baron: Ich wiederhole es noch einmal: Es ist genau so ausgestaltet, dass die Banken, die gelistet sind, von SWIFT abgekoppelt werden, und zwar diese Banken. Damit ist sichergestellt, dass</i>	<i>Baron : Je le répète encore une fois : il est exactement conçu de telle sorte que les banques qui sont listées sont découplées de SWIFT, et ce sont ces banques. Cela garantit que les</i>
10	<i>Zahlungsverpflichtungen für Energie- oder Rohstofflieferungen möglich sind und auch ankommen.</i>	<i>obligations de paiement pour les livraisons d'énergie ou de matières premières sont possibles et qu'elles arrivent à destination.</i>
	<i>STS Hebestreit: Vielleicht darf ich ergänzen. Im Sanktionspaket wurden</i>	<i>STS Hebestreit : Je peux peut-être compléter. Dans le paquet de sanctions,</i>
15	<i>Energielieferungen dezidiert ausgeschlossen. Diese werden also nicht sanktioniert. Insofern sind auch Zahlungsströme, die diese</i>	<i>les livraisons d'énergie ont été résolument exclues. Elles ne sont donc pas sanctionnées. Par conséquent, les flux de paiement qui concernent ces</i>
	<i>Energielieferungen betreffen, nicht Teil</i>	<i>livraisons d'énergie ne font pas partie du</i>
20	<i>des Sanktionspaketes.</i>	<i>paquet de sanctions.</i>
	<i>Zusatzfrage: Ist es aber nicht ein Widerspruch in sich, auf der einen Seite den SWIFT-Zahlungsverkehr weitgehend abzukoppeln, dann aber auf dem großen</i>	<i>Question supplémentaire : N'est-ce pas contradictoire de découpler en grande partie le trafic des paiements SWIFT d'une part, et de permettre à nouveau une</i>
25	<i>Sektor der Energielieferung im Grunde doch wieder Geschäftstätigkeit wie in der Vergangenheit - ich will nicht sagen</i>	<i>activité commerciale comme par le passé - je ne veux pas dire "business as usual" - dans le grand secteur de la livraison</i>

	<i>„business as usual“ - zu ermöglichen? Ist das nicht ein Widerspruch in sich?</i>	<i>d'énergie d'autre part ? N'est-ce pas une contradiction en soi ?</i>
30	<i>STS Hebestreit: Ich würde Ihrem Eindruck, dass das „business as usual“ sei -</i>	<i>STS Hebestreit : Je ne voudrais pas donner l'impression que c'est "business as usual" -</i>
	<i>Zusatz: Eben nicht!</i>	<i>Complément : justement pas !</i>
35	<i>STS Hebestreit: - stark widersprechen. Es sind wahrscheinlich die härtesten Wirtschaftssanktionen, die jemals verhängt wurden, und zwar in sehr kurzer Zeit. Überlegen Sie: Etwa vor einer Woche kündigte der russische Präsident</i>	<i>STS Hebestreit : - contredire fortement. Ce sont probablement les sanctions économiques les plus dures jamais imposées, et ce en très peu de temps. Réfléchissez : il y a environ une semaine,</i>
40	<i>an, Luhansk und Donezk anzuerkennen. Wir haben den Überfall Russlands auf die Ukraine von Donnerstag. Heute ist Montag. Wir haben ein am Freitag</i>	<i>le président russe a annoncé qu'il reconnaissait Louhansk et Donetsk. Nous avons l'incursion de la Russie en Ukraine de jeudi. Nous sommes aujourd'hui lundi. Nous avons adopté un deuxième train de</i>
45	<i>zweites Sanktionspaket beschlossen, ein drittes jetzt über das Wochenende. Ich glaube also, „business as usual“ ist dafür nicht das richtige Wort.</i>	<i>sanctions vendredi, et un troisième ce week-end. Je pense donc que "business as usual" n'est pas le mot approprié.</i>
50	<i>Zu den Energielieferungen: Es wurde eben genau - ich sprach das an anderer Stelle hier auch schon an - über die Sanktionen, die man trifft - - Da braucht es eine gewisse Durchhaltefähigkeit. Es bringt nichts, eine Sanktion zu erlassen, die man nach wenigen Wochen wieder</i>	<i>Concernant les livraisons d'énergie : On vient de parler - j'en ai déjà parlé ailleurs - des sanctions que l'on prend - - Il faut une certaine capacité à tenir le coup. Il ne sert à rien de prendre une sanction et de devoir l'annuler au bout de quelques semaines parce qu'on n'est pas</i>
55	<i>kassieren muss, weil man es selbst nicht durchhalten kann.</i>	<i>capable de la supporter soi-même.</i>

Cet extrait de la conférence de presse du 28 février 2022 est intéressant, car il offre un large éventail d'éléments d'oralité. Les phrases sont analysées dans l'ordre ; ainsi, il est possible de tenir compte de la logique interne de cet extrait.

Dans la première question, l'on peut observer un premier déictique : *jetzt* (ligne 4), qui signifie littéralement *maintenant*. Il convient tout de même de s'interroger sur le sens de cet adverbe dans cette question. Il serait plus logique de l'interpréter d'une manière moins précise sur le plan temporel. Ainsi, plutôt que le traduire par *maintenant* (moment précis), il serait plus approprié de le comprendre comme *pour l'instant* ou *pour le moment*, indépendamment du reste de la phrase. Par conséquent, la traduction de *jetzt* par *maintenant* n'est pas incorrecte, mais elle n'est pas tout à fait adéquate dans ce contexte.

Ensuite, dans la réponse de Baron, il y a un problème d'anaphore pronominale. En effet, à la ligne 5 de la TA, il est impossible de dire à quoi fait référence le pronom *il*. DeepL semble alors n'avoir pas pu déchiffrer correctement le pronom *es*, et sa traduction est dès lors problématique. Comme le pronom *es* n'a pas d'antécédent précis et qu'il a une valeur générale, le *il* ne convient pas. Aux lignes 8 et 9 de la TA, la proposition *et ce sont ces banques* paraît maladroite. Ce problème est probablement causé par *und zwar* dans le TS, qui est censé venir apporter une précision sur ce qui vient d'être dit. Ce lien n'est pas clair en français.

La « question supplémentaire » est particulièrement riche en éléments d'oralité. Le premier aspect qui peut être observé dans le TS est l'anaphore, grâce à *Ist das (aber) nicht ein Widerspruch in sich*, que l'on retrouve aux lignes 21-22 et 28-29. Cette anaphore semble avoir partiellement disparu dans la TA de DeepL. D'abord, la première des deux propositions utilise l'adjectif *contradictoire*, alors que la seconde utilise le substantif *contradiction*. Il ne semble toutefois pas possible d'expliquer pourquoi DeepL supprime (partiellement) l'anaphore et pourquoi le substantif *Widerspruch* est traduit tantôt par un substantif et tantôt par un adjectif. Ensuite, dans la première proposition, DeepL semble avoir omis *in sich*, que l'on retrouve pourtant dans la seconde (*en soi*). Cette anaphore a un certain effet en allemand : elle permet de donner à l'énoncé un caractère accusateur, ce qui manque alors en français. Une autre omission peut également être observée dans cette question. La proposition *im Grunde* ne se retrouve pas dans la TA. Enfin, l'expression anglaise *business as usual* a été reportée en français. Cette expression s'utilise souvent en français dans un

contexte politique, mais il serait intéressant de voir si d'autres expressions étrangères ont été traduites ou reportées dans la TA.

Un problème majeur peut être observé entre les lignes 30 et 34. L'intervenant Hebestreit a été interrompu par un autre locuteur, ce qui semble avoir posé problème à DeepL. En effet, la phrase de la TA n'a plus aucune logique ni aucun sens : *je ne voudrais pas donner l'impression que c'est « business as usual » contredire fortement*. La machine n'est pas en mesure de détecter que la deuxième proposition est en réalité la suite de la première. À la ligne 37, l'on rencontre à nouveau *und zwar*, qui semble avoir été mieux traité par DeepL dans ce cas-ci. Cet effet de précision a été rendu dans la TA par *et ce*. Aux lignes 41 et 42, il y a une phrase très orale : *Wir haben den Überfall Russlands auf die Ukraine von Donnerstag*. En effet, dans ce contexte, le verbe allemand *haben* n'est pas utilisé dans sa première acception. Dans cette phrase-ci, *wir haben* veut plutôt dire *il y a eu*. DeepL a traduit littéralement cette phrase : *nous avons l'incursion de la Russie en Ukraine de jeudi*. La traduction littérale avec le verbe *avoir* ne fonctionne pas en français. Cela corrobore l'une des observations reprises dans l'état de l'art, selon laquelle la TA est problématique lorsqu'un mot n'est pas utilisé dans son acception d'origine.

Le paragraphe suivant s'est également révélé problématique pour DeepL, surtout à cause de nombreux éléments oraux, à commencer par les déictiques. La définition de cet élément permet de le considérer dans ce travail comme un élément d'oralité. En effet, l'on appelle déictique « tout élément linguistique dont la fonction consiste à articuler l'énoncé sur la situation particulière dans laquelle il est produit ou à l'inscrire dans un discours » ('Déictique', *Larousse*). Dans le TS, aux lignes 49 et 50, il y a une combinaison de plusieurs déictiques : *an anderer Stelle hier auch schon*. DeepL semble avoir omis l'un de ces éléments dans la TA : *hier*. Or, cet élément est important, car il indique que le locuteur en a déjà parlé *ici* (lors de cette conférence de presse). S'il manque cet élément, le lecteur peut avoir l'impression que le locuteur en a déjà parlé à un autre endroit. Cette impression est d'ailleurs renforcée à cause de la traduction de *an anderer Stelle* par *ailleurs*, qui n'est pas correcte. En effet, le locuteur veut dire qu'il en a parlé plus tôt dans la même conférence, ce qui ne transparaît pas dans la TA. Ainsi, *an anderer Stelle* n'est pas une locution adverbiale de lieu, mais plutôt une locution adverbiale temporelle. Comme expliqué dans l'état de l'art, lorsqu'une expression n'est pas utilisée dans son acception première, ce qui est le cas ici avec *an anderer Stelle*, la TA peut se révéler problématique. Cette combinaison de déictiques a dès lors engendré un contresens dans la TA. À la ligne 51, on retrouve le mot *da* dans le

TS. Stukenbrock (2018) explique que l’adverbe de lieu (« Lokaladverb ») *da* peut être un déictique de lieu, c’est-à-dire qu’il peut servir à désigner un endroit plus ou moins éloigné du locuteur ; cependant, toujours selon Stukenbrock (2018), l’adverbe de lieu *da* peut également servir de référence anaphorique. Ici, à la ligne 51, l’adverbe de lieu *da* est effectivement une référence anaphorique, il se rapporte aux « sanctions que l’on prend ». Cependant, ce lien anaphorique n’est pas présent dans la TA, ce qui affecte la compréhension. En effet, à la lecture de la TA, il n’est pas facile de comprendre pourquoi et pour quoi il faut une certaine capacité à tenir le coup. En revanche, DeepL fait un ajout intéressant à la ligne 49 : *on vient de parler*. Dans le TS, il manque le verbe principal de la phrase (*es wurde eben X*). Il s’agit effectivement du bon verbe. Toutefois, cela reste un ajout, et un point abordé dans l’état de l’art peut dès lors être confirmé : DeepL tend à compléter les phrases incomplètes. Il est possible que DeepL ait fait une connexion avec le verbe utilisé dans l’incise. Comme aux lignes 27 et 28, DeepL conserve l’incise avec les tirets aux lignes 49 et 50. Le fait que DeepL conserve les incises telles quelles dans la TA peut affecter la lisibilité et le style des phrases, surtout si elles sont déjà longues. Il serait dès lors intéressant de voir si DeepL procède parfois à des restructurations.

- 1 *STS Hebestreit: Wir werben international natürlich für unsere Position, die in breiten Teilen Europas geteilt wird, aber auch in weiten Teilen der Vereinigten Staaten, Kanadas, Japans und anderen.*
- 5 *Aber natürlich muss jeder Staat für sich selbst entscheiden, wie er sich in dieser Phase verhält. Wir bleiben bei unserer Position. Wir halten sie auch für richtig.*
- 10 *Lassen Sie mich diese Gelegenheit vielleicht auch noch einmal zum Anlass nehmen, zu sagen, dass wir auch von dieser Stelle aus Russland noch einmal eindringlich auffordern, die Kampfhandlungen einzustellen - es gibt massive Opfer auch unter der Zivilbevölkerung; es wird viel zerstört -, seine Truppen zurückzuziehen und mit diesem durch nichts gerechtfertigten Angriff auf die Ukraine unverzüglich aufzuhören. Das gehört auch dazu. Das haben verschiedene Vertreter der Bundesregierung, auch der Bundeskanzler gestern in der Regierungserklärung deutlich gemacht.*
- 20 *Bei dieser Haltung bleibt es.*
- Sie haben den Sicherheitsrat angesprochen. Auch das hat der*
- STS Hebestreit : Nous faisons bien sûr la promotion de notre position au niveau international, qui est partagée par une grande partie de l'Europe, mais aussi par une grande partie des États-Unis, du Canada, du Japon et d'autres pays. Mais il va de soi que chaque État doit décider pour lui-même de la manière dont il se comporte dans cette phase. Nous maintenons notre position. Nous pensons également qu'elle est juste.*
- Permettez-moi de profiter de cette occasion pour dire que nous demandons une nouvelle fois à la Russie de mettre un terme aux combats - il y a de nombreuses victimes, y compris parmi la population civile, et beaucoup de destructions -, de retirer ses troupes et de mettre fin immédiatement à cette attaque injustifiée contre l'Ukraine. Cela fait aussi partie de la démarche. Différents représentants du gouvernement fédéral, y compris le chancelier fédéral, l'ont clairement fait savoir hier dans la déclaration gouvernementale. Cette position est maintenue.*
- Vous avez évoqué le Conseil de sécurité. Le chancelier fédéral l'a également*

- 30 *Bundeskanzler gestern in seiner Regierungserklärung dargelegt. Im Sicherheitsrat kann man so weit gehen - die Kollegin Sasse möge mir das nachsehen -: Eine Enthaltung ist in dem Sinne eine Zustimmung, weil es eine Sanktion, eine Verurteilung nicht*
- 35 *verhindert hätte. Russland musste sein eigenes Vetorecht benutzen, um einer Verurteilung zu entgehen. Der Bundeskanzler nannte das eine Schande.*
- 40 *Frage: Herr Hebestreit, unter der großen Überschrift „Wie sich Deutschland schützt“, will ich fragen, welche Vorkehrungen womöglich getroffen werden, wenn man auf den Aspekt schaut, dass Putin mit Atomwaffen gedroht hat.*
- 45 *Gibt es dafür Vorkehrungen? Welche?*
- In dem Zusammenhang vielleicht eine Frage an Herrn Collatz: Können Sie etwas dazu sagen, wo diese Atomwaffen stationiert sind und wie weit sie reichen?*
- 50 *STS Hebestreit: Wir nehmen die Äußerung des russischen Präsidenten natürlich sehr, sehr ernst. Es ist auch klar, wie unverantwortlich allein schon eine solche Drohung ist.*
- expliqué hier dans sa déclaration gouvernementale. Au Conseil de sécurité, on peut aller jusqu'à dire - que Mme Sasse me pardonne - qu'une abstention est une approbation dans le sens où elle n'aurait pas empêché une sanction, une condamnation. La Russie a dû utiliser son propre droit de veto pour éviter une condamnation. Le chancelier allemand a qualifié cela de honte.*
- Question : Monsieur Hebestreit, sous le grand titre "Comment l'Allemagne se protège", je voudrais demander quelles dispositions sont éventuellement prises si l'on considère l'aspect selon lequel Poutine a menacé d'utiliser des armes nucléaires. Y a-t-il des dispositions à cet effet ? Lesquelles ?*
- Dans ce contexte, peut-être une question à M. Collatz : pouvez-vous nous dire où sont stationnées ces armes nucléaires et quelle est leur portée ?*
- STS Hebestreit : Nous prenons bien sûr la déclaration du président russe très, très au sérieux. Il est également clair qu'une telle menace est à elle seule irresponsable.*

55 *Gleichzeitig haben im Parallèlement, les armes nucléaires ont*
Bedrohungsszenario auch Nuklearwaffen toujours joué un rôle dans le scénario de
immer eine Rolle gespielt und ist insoweit menace et la défense contre les armes
andererseits auch die Abwehr von nucléaires a toujours été prise en compte
Nuklearwaffen immer mitgedacht. Viel dans ce contexte. Je n'en dirai pas plus à
60 *mehr will ich von dieser Stelle aus dazu ce stade.*
nicht sagen.

Collatz: Ich kann dazu natürlich auch Collatz : Je ne peux bien sûr qu'ajouter à
nur ergänzen - das ist ja allgemein ce sujet - tout le monde le sait - que la
bekannt -, dass die Bundeswehr nicht Bundeswehr ne dispose pas d'armes
65 *über Atomwaffen verfügt, sondern die nucléaires, mais qu'elle met à disposition*
Flugzeuge bereitstellt, um im Rahmen der les avions pour fournir, le cas échéant,
nuklearen Teilhabe gegebenenfalls bei ses propres prestations dans le cadre de
der Ausbringung der Mittel, wenn es l'Alliance, dans le cadre de la
soweit sein sollte, was hoffentlich nie der participation nucléaire, lorsque les
70 *Fall sein wird, im Rahmen des moyens seront déployés, ce qui, espérons-*
Bündnisses eigene Leistungen zu le, ne sera jamais le cas.
erbringen.

Zusatzfrage: Sie meinen jetzt bei der Question supplémentaire : vous voulez
Abwehr, oder? Ich hatte gefragt, ob Sie dire maintenant à la défense, n'est-ce pas
75 *sagen können, wo die russischen ? J'ai demandé si vous pouviez dire où*
Atomwaffen stationiert sind und wie weit sont stationnées les armes nucléaires
sie reichen. russes et quelle est leur portée.

J'ai également choisi d'analyser cet extrait de la conférence du 28 février 2022, car celui-ci permet de corroborer, d'infirmer ou de nuancer les observations de l'extrait précédent.

Dans le premier paragraphe, on remarque une progression presque totalement à thème constant, c'est-à-dire que, mis à part dans la deuxième phrase, toutes les phrases commencent par le thème, à savoir le pronom personnel sujet *wir*. Dans ce contexte, cette progression a une importance particulière, puisque cela signifie que le porte-parole du gouvernement fédéral, M. Hebestreit, souhaite insister sur la position du gouvernement. Cette progression peut être considérée comme une figure de rhétorique, et celle-ci doit dès lors apparaître dans la TA. Les trois phrases dont il est question ici commencent également par le pronom personnel sujet *nous* dans la TA.

Le paragraphe suivant commence par une phrase longue allant de la ligne 9 à la ligne 20. Comme mentionné dans l'état de l'art au point 3, Daems & Ruffo (2022) affirment que plus la phrase du TS est longue, plus le risque d'omissions est important. À la ligne 10, la suite d'adverbes *auch noch einmal* ne se retrouve pas dans la TA, ce qui semble confirmer l'observation de Daems & Ruffo. En revanche, le groupe adverbial *noch einmal* à la ligne 12 a été traduit par DeepL. Si l'omission du premier *noch einmal* avait été effectuée par un biotraducteur, il aurait été possible de la justifier comme une stratégie pour éviter la répétition de la même expression. Or, il est impossible de déterminer avec certitude pourquoi DeepL a procédé à cette omission. Toutefois, il peut être supposé que le premier groupe (ligne 10) a été omis, car il s'agit d'une suite d'adverbes plus longue. Ensuite, entre la ligne 14 et la ligne 16, il y a une longue incise, qui constitue une phrase à elle seule. Ce genre de « digressions » est très fréquent à l'oral, mais il convient tout de même d'observer si les incises ne perturbent pas le rythme, le style et la syntaxe de la TA. Ici, DeepL conserve l'incise et ne modifie ni sa forme ni son articulation avec le reste de la phrase. Dans la TA, l'incise est mal articulée avec le cotexte. DeepL a en revanche légèrement modifié le contenu de cette digression. Alors qu'en allemand, il y a deux phrases dans cette digression, on n'en voit plus qu'une en français, étant donné que DeepL a utilisé un seul verbe pour les deux parties de l'incise. Toutefois, l'incise, malgré sa longueur, ne semble pas avoir perturbé la syntaxe de la fin de la phrase. La progression des idées à partir de la ligne 20 est également intéressante : les deux phrases commencent par *das*, pronom qui fait référence aux trois demandes qui viennent d'être énumérées. Cependant, cette progression des idées n'est pas respectée dans la TA de DeepL, ce qui affaiblit l'effet percutant de ces deux phrases. Enfin, le dernier point qui mérite d'être évoqué dans ce paragraphe est l'adverbe pronominal *dazu*

(ligne 20), qui sert de référence anaphorique à la *Position* (lignes 2 et 8) qu'évoque le locuteur. Cependant, comme la référence est bien plus haut, DeepL a commis un glissement de sens. Le *dazu* fait référence à la position du gouvernement fédéral. Il est tout de même intéressant de voir que DeepL semble avoir été entraîné à traduire ce genre de mots par des termes généraux, ici *démarche*, qui ne permet pas de faire le lien avec *position*.

Dans le paragraphe suivant, l'on peut à nouveau observer une incise, plus courte cette fois-ci (lignes 30 et 31). À la lecture, cette incise ne semble pas poser de problème. En revanche, si cette phrase était destinée à être prononcée oralement, la syntaxe pourrait engendrer une ambiguïté, car l'incise ainsi que la proposition complétive du verbe *dire* commencent toutes les deux par la conjonction *que*. Ceci peut donner l'impression que l'incise (*que Mme Sasse me pardonne*) constitue la proposition complétive du verbe *dire*, ce qui n'est pas le cas. Il est d'ailleurs aussi intéressant d'observer que le substantif allemand *Kollegin* est devenu *Mme* dans la TA. L'utilisation de ce substantif dans le texte source est ici un exemple d'emploi informel. Il serait intéressant de vérifier l'hypothèse suivante : les quelques cas de niveau de langue informel seraient-ils supprimés dans la TA et transposés dans un niveau plus soutenu ? C'est en tout cas, comme mentionné dans l'état de l'art, ce que suggèrent Hansen & Houlmont (2022) : la TA tend à neutraliser les particularités du registre informel.

À la ligne 46, il y a une phrase averbale (*In dem Zusammenhang vielleicht eine Frage an Herrn Collatz*, traduite par *Dans ce contexte, peut-être une question à M. Collatz*), ce qui revient assez souvent dans le corpus, surtout lorsque le locuteur annonce sa question. Ce type de phrases étant probablement peu représentées dans les données d'entraînement de DeepL, il est intéressant d'observer comment ce programme les traite. Bien que la TA soit acceptable, ce n'est sans doute pas ainsi que l'on annoncerait une question spontanément en français. Au lieu de *dans ce contexte*, un locuteur francophone aurait plutôt tendance à dire *à ce sujet*, voire *tant que nous y sommes* ou *puisque'on en parle*.

La première partie de la réponse donnée par M. Hebestreit (lignes 50 à 54) est globalement correcte. Il n'y a pas vraiment d'éléments relatifs au caractère oral, si ce n'est l'insistance créée par la répétition de l'adverbe *sehr* (ligne 52). DeepL conserve cette répétition, ce qui est approprié ici. Toutefois, par curiosité, j'ai fait retraduire cette phrase à DeepL (en novembre 2022), mais cette fois-ci sans la virgule séparant les deux *sehr*. Dans ce cas, DeepL ne le répète pas : *nous prenons bien sûr très au sérieux la déclaration du président russe*.

En revanche, dans le paragraphe suivant (lignes 55 à 61), les choses se compliquent. En effet, certaines omissions peuvent être observées, alors que la phrase n'est pas particulièrement longue. Il convient dès lors de s'interroger sur la cause de ces omissions. Premièrement, deux *auch* ne se retrouvent pas dans la TA. Ce mot invariable peut avoir de nombreuses significations différentes. Dans ces deux cas, il pourrait s'agir de particules modales. Dans une phrase déclarative, « [l']énoncé en *auch* est un commentaire qui souligne une concordance générale, une absence de hiatus entre le contenu relaté et un contenu antérieur » (Métrich et al., 1998, p. 253). Ici, la question posée plus haut portait sur la menace nouvelle des armes nucléaires. En réponse à cela, Hebestreit affirme que les armes nucléaires ont toujours été une menace et sous-entend alors qu'il ne s'agit pas d'un phénomène récent. Cet emploi du *auch* est confirmé par Métrich et al. (1998, p. 253) :

Le locuteur commente les propos de l'interlocuteur ; la réplique peut prendre des allures de reproche dans la mesure où la concordance soulignée par *auch* suggère que l'énoncé antérieur n'avait pas lieu d'être [...].

Cette nuance n'est pas présente dans la TA, ce qui engendre un léger glissement de sens, ou du moins un « glissement d'effet ». Une solution simple pour rendre cette nuance serait par exemple d'utiliser la conjonction de coordination *mais*. D'autres omissions ou problèmes de traduction peuvent être relevés dans cette phrase. La mauvaise traduction de l'adverbe *insoweit* (qui semble avoir été traduit par *dans ce contexte*) entraîne une répercussion importante sur la logique du texte. En effet, *insoweit* apporte en quelque sorte un lien de cause-conséquence : comme les armes nucléaires ont toujours joué un rôle dans le scénario de menace, *alors puisque c'est le cas*, la défense contre ces armes a toujours été prise en compte. Ce lien n'est pas clair dans la TA. L'adverbe *andererseits* n'a pas été traduit non plus. Or, il vient ajouter un élément à la phrase : il prend d'une certaine manière la signification première du *auch* en tant que particule de mise en relief ; dans ce cas, l'on peut dire qu'il « porte sur un élément ou membre de phrase qu'il met en relief en présentant l'information qu'il exprime comme s'ajoutant à une ou plusieurs autres de même type » (*ibid.*, p. 238). Enfin, à la ligne 59, l'adverbe *viel* n'a pas été reproduit dans la TA, ce qui n'entraîne qu'une répercussion mineure. On constate dès lors dans ce paragraphe que les cascades d'adverbes sont difficiles à traiter pour la TA, d'autant plus qu'ici, ces adverbes ont des nuances très précises. Les cascades d'adverbes et de circonstants sont très révélatrices du caractère oral des conférences de presse.

La phrase suivante est également intéressante, puisqu'il s'agit d'une phrase longue (57 mots). On peut par conséquent s'attendre à ce que cela affecte la lisibilité de la TA et à ce qu'il y ait certaines omissions. Deux omissions peuvent en effet être relevées. La première est celle de la particule modale *ja* (ligne 63). Selon Kefer (n.d., p. 1), la particule *ja* peut servir à marquer l'étonnement du locuteur par rapport ce qu'il exprime. Blumenthal (1997, pp. 96–98) confirme cette idée et ajoute même une deuxième nuance importante de la particule modale *ja* :

Ja unterstellt das gemeinsame Vorwissen implizit, während eine explizite Formulierung wie *wir beide wissen bekanntlich seit langem, daß...* wohl adäquat in jede Sprache übersetzt werden kann. (*Ibid.*, p. 98)

Si l'on en croit le contexte, dans cette phrase, le locuteur utilise la particule *ja* pour créer une connivence. Toutefois, étant donné que cette particule se retrouve dans l'incise *tout le monde le sait*, ajoute-t-elle vraiment quelque chose, tant en allemand que dans la TA ? Cela aurait été redondant de traduire cette proposition par *vous savez que tout le monde le sait*. Une autre omission, plus importante, peut être observée : *wenn es soweit sein sollte* (lignes 68 et 69), qui veut dire littéralement *si ça devait aller si loin*. On pourrait argumenter que cette proposition a, à peu de choses près, la même valeur sémantique que l'adverbe *gegebenenfalls* (ligne 67), d'autant plus qu'ils portent tous les deux sur le même syntagme. En revanche, on remarque que cet adverbe ne porte pas sur le même groupe dans la TA (ligne 66). Par conséquent, l'omission de *wenn es soweit sein sollte* est importante. Par ailleurs, l'incise (ligne 63) a été conservée telle quelle dans la TA, ce qui corrobore l'observation faite dans l'analyse de l'extrait précédent.

Pour conclure l'analyse qualitative et exploratoire de ces deux extraits, l'on peut souligner quelques problèmes récurrents. Les problèmes principaux sont les omissions, surtout des déictiques et des adverbes lorsque ceux-ci sont en cascade, ainsi que les références et anaphores pronominales. Ces aspects posent *a priori* davantage de problèmes lorsque les phrases sont plus longues. D'autres aspects moins bien traités ont également pu être observés, notamment les phrases interrompues ou coupées, les incises, les anaphores et les particules modales. Par ailleurs, une phrase averbale a pu être observée dans l'un de ces extraits, et sa traduction n'était pas convaincante. Enfin, les expressions et les mots étrangers, qui apparaissent davantage à l'oral qu'à l'écrit, méritent également une brève analyse. Ce sont ces observations qu'il convient particulièrement de confirmer ou d'infirmer grâce à une analyse quantitative, c'est-à-dire à l'aide de plusieurs exemples isolés.

5.3. Analyse d'exemples isolés

- **Les déictiques**

Le premier point abordé dans l'analyse des extraits est celui des déictiques. Pour limiter l'échantillon, je me penche exclusivement sur les deux déictiques *jetzt* (littéralement *maintenant*) et *hier* (littéralement *ici*). Il est important de noter que les adverbes *jetzt* et *hier* peuvent également servir de références anaphoriques ou même cataphoriques. Une anaphore fait référence à un élément mentionné précédemment dans le discours (Kleiber, n.d., p. 3) ; une référence cataphorique porte sur un élément mentionné plus tard (Perdicoyanni-Paléologou, 2009, p. 61). En revanche, les déictiques sont « des morphèmes qui expriment une représentation de l'espace et une représentation du temps en termes d'espace » (T. Fraser & Joly, 1980, p. 22). Il est tout à fait possible, dans certains cas, de traduire ces déictiques de manière littérale, mais dans ce contexte de conférence de presse, ceux-ci peuvent avoir une signification plus nuancée, c'est pourquoi il convient d'observer la manière dont DeepL les traite. Voici quelques exemples dans lesquels *jetzt* et *hier* (déictiques uniquement) sont omis ou mal traités dans la TA.

EXEMPLE 1

<i>Insofern sind Drittstaatsangehörigen ganz überwiegend mit erfasst, weil es bei den Menschen, die es jetzt aus diesem Kriegsgebiet heraus schaffen, natürlich darum geht - das hat die Bundesinnenministerin auch am Wochenende immer wieder gesagt -, tatsächlich Leben zu retten und das nicht vom Pass abhängig zu machen, sondern Drittstaatsangehörigen dort natürlich genauso zu helfen. (Conférence du 07/03/2022)</i>	<i>Dans cette mesure, les ressortissants de pays tiers sont principalement concernés, car pour les personnes qui parviennent à sortir de cette zone de guerre, il s'agit bien entendu - comme l'a répété la ministre fédérale de l'Intérieur ce week-end - de sauver des vies et de ne pas faire dépendre cela du passeport, mais d'aider les ressortissants de pays tiers de la même manière.</i>
--	--

Comme expliqué dans l'état de l'art et dans l'analyse des extraits, c'est surtout dans les phrases longues que le risque d'omissions est présent. Dans cet exemple, le déictique *jetzt* ne semble pas avoir été traduit par DeepL. Or, ce déictique a ici une signification importante.

Renforcé grâce au déterminant démonstratif *diesem*, le déictique *jetzt* souligne l'idée qu'il s'agit des personnes qui parviennent à sortir de cette zone de guerre *actuellement*, ce qui influence l'ancrage spatio-temporel de cette réalité.

EXEMPLE 2

<i>Ich habe die Lieferländer jetzt nicht im Kopf, aber dort ist eine Abkehr von russischen Importen sicherlich am einfachsten möglich, obwohl das natürlich im Detail auch immer schwierig ist.</i>	<i>Je n'ai pas les pays fournisseurs en tête, mais c'est certainement là qu'il est le plus facile de se détourner des importations russes, même si c'est bien sûr toujours difficile dans le détail.</i>
--	--

(Conférence du 14/03/2022)

Dans l'exemple 2, on remarque également que l'adverbe *jetzt* n'apparaît pas dans la TA. Il aurait été envisageable de dire : *je n'ai pas les pays fournisseurs en tête là tout de suite*, ou quelque chose de semblable. Toutefois, l'omission de DeepL n'affecte pas vraiment le sens de la phrase. En effet, lorsqu'il lit dans la TA *je n'ai pas les pays fournisseurs en tête*, le lecteur comprend facilement qu'il ne les a pas en tête à l'instant où il parle. Cette idée étant implicite en français, le déictique n'est pas indispensable. Toutes les omissions ne sont donc pas problématiques ; celle-ci est même opportune.

EXEMPLE 3

<i>Sie können sich vorstellen, dass ich angesichts eines solchen Umfangs jetzt nicht einzeln darauf eingehen kann. Ich kann Ihnen für die Maßnahmen, die beim BMI laufen, jetzt sagen, dass davon zum Teil schon etwas abgeschlossen ist.</i>	<i>Vous pouvez imaginer qu'étant donné l'ampleur de la question, je ne peux pas l'aborder individuellement. Je peux vous dire que certaines des mesures en cours au BMI sont déjà achevées.</i>
---	---

(Conférence du 14/03/2022)

Dans cet extrait, le déictique *jetzt* est présent à deux reprises dans le TS. Cependant, il n'apparaît jamais dans la TA, ce qui signifie que DeepL a omis ce déictique deux fois de suite. Ces omissions ont un effet sur le sens de l'énoncé. Le premier *jetzt* a presque une valeur spatiale ; le locuteur veut dire que l'ampleur de la question ne lui permet pas d'y répondre seul *maintenant, ici*, c'est-à-dire à la conférence de presse, ce qui ne veut pas pour

autant dire qu'il ne peut pas l'aborder seul dans un autre contexte. La valeur du second *jetzt* est légèrement différente ; le locuteur explique que *pour le moment, à l'heure à laquelle il parle*, certaines mesures sont achevées. Par conséquent, l'omission de ces deux déictiques influence sensiblement la valeur et l'ancrage spatio-temporel du message du locuteur.

EXEMPLE 4

*Kall: Ich möchte mich **jetzt** nicht zu einzelnen Vereinen äußern. (Conférence du 28/02/2022)* *Kall : Je ne souhaite pas m'exprimer sur les clubs en particulier.*

Toutefois, comme l'illustre ce passage, il n'y a pas que dans les phrases longues que DeepL omet les déictiques. En effet, ici aussi, l'omission de l'adverbe *jetzt* dans la TA entraîne un glissement de sens. Au lieu de dire qu'il ne préfère pas s'exprimer sur ces clubs au moment où il parle, le locuteur semble vouloir dire qu'il n'a pas envie de s'exprimer sur ce sujet, quels que soient l'endroit et le moment.

EXEMPLE 5

***Jetzt** gibt es **heute** den Vorschlag der Kommission. (Conférence du 02/03/2022)* *Nous avons **maintenant** la proposition de la Commission.*

L'adverbe *jetzt* se retrouve ici dans la TA de DeepL, mais il convient tout de même de s'interroger sur ce déictique. En effet, la phrase allemande comporte deux déictiques : *jetzt* et *heute*, le dernier ayant été omis. Même si la phrase en elle-même n'est pas nécessairement parfaite, notamment à cause de l'utilisation du verbe *avoir*, l'omission d'un des deux adverbes n'est pas problématique, car — selon moi — ils ont le même sens sur le plan temporel dans cette phrase.

EXEMPLE 6

*Zusatzfrage: Da bin ich **jetzt** aber sehr überrascht. (Conférence du 02/03/2022)* *Question supplémentaire : Cela me surprend beaucoup.*

Dans cet exemple, le déictique *jetzt* a à nouveau été omis. Bien que cette omission ne soit pas majeure d'un point de vue purement sémantique, elle atténue l'effet de surprise en réaction à ce qui est dit juste avant cet énoncé. La traduction littérale de *jetzt* ne résoudrait pas le problème ; ici, pour bien faire, une reformulation serait nécessaire. Ma proposition

serait par exemple : *voilà qui me surprend beaucoup*. Cette traduction renforcerait l’ancrage de l’énoncé dans la situation. Le déictique *jetzt* serait alors repris dans le *voilà*, qui apporte la nuance temporelle que renferment le *jetzt* et le *da* allemand.

EXEMPLE 7

<i>Aber darüber hat das THW gerade</i>	<i>Mais le THW vient de donner des</i>
<i>informiert; das kann ich jetzt hier nicht</i>	<i>informations à ce sujet ; je ne peux pas en</i>
<i>komplett vortragen. (Conférence du</i>	<i>parler ici en entier.</i>
<i>02/03/2022)</i>	

On retrouve dans cet exemple une succession de deux adverbes, à savoir *jetzt* et *hier*. Comme cela a été discuté dans l’analyse des extraits, DeepL semble omettre un adverbe lorsque celui-ci se retrouve dans une « cascade » d’adverbes, ce qui est le cas ici. En effet, le déictique *jetzt* n’est pas explicitement reporté dans la TA. Toutefois, dans le contexte de la conférence de presse, l’adverbe *ici*, qui est la traduction littérale de *hier*, sous-entend également l’idée de *maintenant*, puisque l’adverbe *ici* pourrait être remplacé par *à cette conférence* (d’un point de vue tant temporel que spatial).

EXEMPLE 8

<i>Dazu habe ich jetzt hier nichts Neues</i>	<i>Je n’ai rien de nouveau à apporter à ce</i>
<i>beizutragen. (Conférence du 14/03/2022)</i>	<i>sujet.</i>

Dans cette phrase de huit mots, les deux adverbes *jetzt* et *hier* sont omis dans la TA. Ceci montre que, bien que les recherches semblent pointer du doigt les omissions dans les phrases longues principalement, ce phénomène n’est pas inhérent aux phrases longues.

Parmi les 183 *jetzt* (déictiques) observés dans le corpus allemand, 123 sont traduits (67 %) et 60 ne le sont pas (33 %). La différence entre les phrases courtes (moins de 20 mots) et les phrases longues n’est pas frappante. Dans les phrases courtes, 33 *jetzt* sur 103 sont omis (32 %), alors que l’omission atteint un taux de 34 % dans les phrases longues. En revanche, lorsque *jetzt* et *hier* sont combinés, il y a une omission d’au moins un des deux adverbes dans 5 cas sur 6. Quant aux *hier*, dans les phrases de moins de 20 mots, 29 % ne sont pas traduits. En revanche, dans les phrases longues, 45 % ne sont pas traduits ; au total, 36 % des *hier* déictiques ne sont pas traduits. Les quatre *hier* cataphoriques (phrases courtes et longues) sont tous traduits et 18 % des *hier* anaphoriques ne sont pas traduits. Un autre

problème peut être soulevé quant à ces deux déictiques : leur traduction est (presque) systématiquement littérale, et elle ne correspond, par conséquent, pas toujours à la valeur des déictiques dans le TS.

- **Les phrases incomplètes ou interrompues**

Dans l'analyse du premier extrait long ont été rencontrées à deux reprises des phrases incomplètes ou interrompues. L'un de ces deux cas de figure (extrait 1, lignes 30-34) a permis de remarquer que lorsqu'une personne coupait la parole au locuteur, il n'y avait plus aucune logique entre les deux parties de phrase. Dans l'autre exemple (extrait 1, lignes 48-51), il a pu être observé que DeepL complétait la phrase à laquelle il manquait une partie, comme indiqué dans l'état de l'art (Hansen & Houlmont, 2022, p. 264). Trois autres exemples de phrases incomplètes apparaissent dans le corpus.

EXEMPLE 1

Zusatzfrage: Können Sie etwas zum zeitlichen Verlauf sagen? Ist das am Freitag - - -

STS Hebestreit: Ich glaube, dazu habe ich mich schon geäußert. Ich habe gesagt: über das Wochenende, seit Freitag. (Conférence du 28/02/2022)

Question supplémentaire : Pouvez-vous dire quelque chose sur le déroulement dans le temps ? Est-ce que c'est le

vendredi - - -

STS Hebestreit : Je crois que je me suis déjà exprimé à ce sujet. J'ai dit : pendant le week-end, depuis vendredi.

Dans cet exemple, la proposition *Ist das am Freitag - - -* n'est pas complète. Contrairement à ce à quoi l'on pourrait s'attendre au vu de l'état de l'art, DeepL se tient à ce qui est dit explicitement dans le TS et ne complète pas l'énoncé. Ceci semble pertinent, puisque la suite de la phrase est difficilement déductible du co(n)texte, à l'inverse de l'exemple que l'on retrouve aux lignes 48-51 du premier extrait long.

EXEMPLE 2

Nein. Das war die Frage, ob die Bundesregierung, der Kanzler vorweg, der ja in der Lage ist, wie mehrfach - - - Ich

Non. C'était la question de savoir si le gouvernement fédéral, le chancelier en premier lieu, qui est en effet en mesure,

meine, der Begriff „Bazooka“ verbietet sich in diesem Zusammenhang mit dem Sondervermögen für die Bundeswehr.
(Conférence du 28/02/2022)

comme à plusieurs reprises - - - je veux dire que le terme de "bazooka" s'interdit dans ce contexte de fonds spéciaux pour la Bundeswehr.

Cet exemple permet d'affirmer que DeepL ne complète pas systématiquement les phrases incomplètes. L'exemple 3 ci-dessous permet également de confirmer cette observation.

EXAMPLE 3

*Zusatzfrage: Zu näheren Hintergründen will man wahrscheinlich aus - - -
Alexandrin: Man will aus sicherheitspolitischen Gründen nichts sagen, genau richtig.* (Conférence du 14/03/2022)

*Question supplémentaire : Pour en savoir plus, il est probable que l'on veuille - - -
Alexandrine : On ne veut rien dire pour des raisons de politique de sécurité, c'est tout à fait exact.*

Il reste tout de même difficile d'affirmer pourquoi DeepL ne complète pas du tout les phrases incomplètes ou interrompues, alors que l'état de l'art confirme l'inverse. Est-ce grâce à l'évolution des systèmes de TAN ? Il est également possible que DeepL ne complète pas les phrases à cause des trois tirets qui marquent la « fin » de la phrase. En effet, si l'on observe comment DeepL traduit ces phrases lorsque ces trois tirets disparaissent, les phrases sont complétées de manière aléatoire. L'exemple 2 devient : *non. C'était la question de savoir si le gouvernement fédéral, le chancelier en premier lieu, qui est en mesure, comme nous l'avons dit à plusieurs reprises, d'assurer la sécurité de l'approvisionnement en eau potable, était en mesure de le faire* (TA de décembre 2022). L'exemple 3 est quant à lui complété, mais en moindre mesure : *Il est probable que l'on veuille en savoir plus sur le contexte* (TA de décembre 2022). L'hypothèse selon laquelle ce sont les trois tirets qui empêchent DeepL de compléter les phrases pourrait dès lors être vérifiée si le corpus contenait davantage d'exemples similaires.

- **Les incises ou digressions**

Il a pu être observé dans l'analyse des extraits longs que les incises ou digressions, entourées par des tirets, gardaient la même structure et la même position dans la TA de DeepL. Il convient dès lors d'analyser quelques exemples du corpus parallèle afin de voir si l'on est en mesure — ou non — de tirer des conclusions sur le fonctionnement de DeepL à cet égard.

EXEMPLE 1

<i>Juristisch - das will ich auch noch einmal</i>	<i>D'un point de vue juridique - je tiens à le</i>
<i>klarstellen - ziehen EU-Sanktionen keine</i>	<i>préciser - les sanctions de l'UE</i>
<i>Entschädigungspflicht nach sich.</i>	<i>n'entraînent aucune obligation</i>
(Conférence du 28/02/2022)	<i>d'indemnisation.</i>

Cet exemple permet de confirmer les observations faites précédemment. En effet, l'incise ne fait pas l'objet d'une restructuration et reste à la même place. En revanche, DeepL omet un groupe adverbial qui a une importance majeure : *noch einmal*. Celui-ci ajoute l'idée que le locuteur a déjà évoqué ce point en amont et vient nuancer la valeur de l'énoncé, étant donné que *noch einmal* exprime en quelque sorte l'agacement du locuteur, qui doit se répéter. Cette omission avère l'hypothèse formulée plus haut, selon laquelle les omissions se produisent tant dans les phrases longues que dans les phrases courtes.

EXEMPLE 2

<i>Ich glaube, die Erkenntnis, was das, was der</i>	<i>STS Hebestreit : Je pense que la prise de</i>
<i>Bundeskanzler als Zeitenwende bezeichnet</i>	<i>conscience de ce que le chancelier fédéral a</i>
<i>hat - das hat er, glaube ich, bereits am</i>	<i>qualifié de changement d'époque - je crois</i>
<i>Donnerstag in einer ersten Stellungnahme</i>	<i>qu'il l'a déjà fait jeudi dans une première</i>
<i>getan -, am Ende für uns alle bedeutet, war</i>	<i>prise de position - signifie en fin de compte</i>
<i>ein Prozess, der auch europäisch und dann</i>	<i>pour nous tous, a été un processus qui s'est</i>
<i>international abgelaufen ist. (Conférence du</i>	<i>également déroulé au niveau européen et</i>
28/02/2022)	<i>ensuite international.</i>

Cet exemple corrobore également les observations faites plus haut au sujet des incises. Néanmoins, dans ce cas, la digression semble perturber la syntaxe de la phrase. En effet, l'on se rend rapidement compte que, sans l'incise, la phrase n'a pas de logique. Cependant, il est possible que cette perturbation ne soit pas uniquement due à l'incise : la phrase du TS est tout aussi mal formulée. Dans cette phrase, il y a également un problème de référence

anaphorique. Le pronom *das* dans l’incise ne semble pas avoir été déchiffré correctement par DeepL : celui-ci est censé faire référence aux propos du chancelier fédéral, ce qui n’est pas clair dans la TA.

EXEMPLE 3

<i>Ich bin jetzt etwas im Zwiespalt, weil sich immer mehr Hände zum Thema der Ukraine heben - dafür wollen wir uns natürlich die Zeit nehmen -, gleichzeitig aber der Vertreter des Arbeitsministeriums in fünf Minuten weg muss. (Conférence du 07/03/2022)</i>	<i>Je suis un peu partagée, parce que de plus en plus de mains se lèvent sur le sujet de l'Ukraine - nous voulons bien sûr prendre le temps de le faire - mais en même temps, le représentant du ministère du Travail doit partir dans cinq minutes.</i>
--	--

Une fois de plus, l’exemple ci-dessus confirme les observations précédentes. Il est aussi intéressant de souligner que DeepL supprime la virgule que l’on retrouve dans le TS après l’incise. Il aurait été plus judicieux de la conserver, étant donné que la conjonction *mais* est généralement précédée d’une virgule. À la première ligne, le déictique *jetzt* a à nouveau été omis. Par ailleurs, un problème de référence anaphorique, présente dans le TS grâce à l’adverbe pronominal *dafür*, peut également être relevé. Bien que cela ne soit pas visible directement, cet adverbe pronominal fait référence au sujet de l’Ukraine (*Thema Ukraine*). Or, dans la TA de DeepL, le pronom *le* ainsi que le verbe *faire* dans l’incise n’ont pas de référent, et le lecteur ne peut dès lors pas comprendre de quoi le locuteur parle.

EXEMPLE 4

<i>Dasselbe - das kann ich sagen - gilt im Grunde auch für die Türkei. (Conférence du 02/03/2022)</i>	<i>Je peux dire qu'il en va de même pour la Turquie.</i>
---	--

Cet exemple fait exception aux comportements systématiques de DeepL observés jusqu’ici. En effet, l’incise a disparu dans la TA, et la phrase a par conséquent été totalement restructurée. L’incise est devenue la proposition principale qui introduit la subordonnée complétive. Dès lors, il convient d’observer que DeepL ne conserve pas automatiquement toutes les incises telles qu’elles apparaissent dans le TS, bien que cela soit une tendance générale. À titre informatif, dans mon échantillon de phrases comportant une incise, cette dernière ne disparaît que dans une phrase sur 34, c’est-à-dire celle de l’exemple 4.

- **Les anaphores et la progression des idées**

Il ressort de l'analyse du premier des deux extraits qu'une anaphore n'a pas été totalement reproduite dans la TA, ce qui atténue l'effet produit par celle-ci dans le TS. Il convient par conséquent d'observer davantage d'extraits afin de voir si la suppression de l'anaphore est fréquente et si cette suppression a un effet ou non.

EXEMPLE 1

<p><i>Das hat auch der Wirtschafts- und Klimaschutzminister Robert Habeck deutlich gemacht und das haben auch andere gesagt, und auch der Bundeskanzler sieht sich in dieser Reihe.</i> (Conférence du 28/02/2022)</p>	<p><i>Le ministre de l'Économie et de la Protection du climat, Robert Habeck, l'a dit clairement, tout comme d'autres, et le chancelier fédéral se considère lui aussi dans cette lignée.</i></p>
--	---

Dans cette phrase, l'anaphore permet au locuteur d'insister sur l'objet plutôt que sur le sujet. Au niveau de la progression des idées, cela a un effet aussi : l'information connue (le thème) est placée en début de phrase, et le locuteur finit avec le rhème, c'est-à-dire la nouvelle information. En français, cette progression n'est pas présente, puisque l'information nouvelle est placée en début de phrase. Il aurait été possible de coller plus à la syntaxe allemande et de commencer la phrase par *c'est ce que*.

EXEMPLE 2

<p><i>Ist China mit im Augenmerk? Ist mit im Augenmerk, dass man China aus Europa bitten könnte, sich darum zu kümmern?</i> (Conférence du 07/03/2022)</p>	<p><i>La Chine est-elle concernée ? Est-ce que la Chine pourrait être invitée à s'en occuper depuis l'Europe ?</i></p>
---	--

Dans cet exemple, l'anaphore créée grâce à la répétition de la question *ist [...] mit im Augenmerk* n'a pas été reproduite par DeepL, alors que les deux expressions répétées en allemand sont proches l'une de l'autre. Cette suppression entraîne l'atténuation de l'insistance. Il est curieux de voir comment DeepL traite cette phrase quelques mois plus tard : *La Chine fait-elle l'objet d'une attention particulière ? Une attention particulière est-elle accordée à la possibilité de demander à la Chine d'Europe de s'en occuper ?* (TA de DeepL de janvier 2023). Cette TA, bien qu'elle ne soit pas parfaite, notamment à l'égard de l'expression *la Chine d'Europe*, conserve l'effet de l'anaphore.

EXEMPLE 3

Haben Sie Hinweise darauf, dass Belarus auf die Sanktionen reagiert, zum Beispiel, indem man die Menschen, die im Grenzgebiet sind, möglicherweise doch in die Ukraine lässt, wenn sie so waghalsig wären, sich durch das Kriegsgebiet durchzuschlagen? **Haben Sie darauf Hinweise?** [...] Es gab in der Grenzregion zwischen Belarus und der Ukraine Flüchtlinge, die dort ausgeharrt haben, um nach Europa zu kommen, und die sind lange im Grenzgebiet gewesen und nicht weitergekommen. **Haben Sie Hinweise darauf**, dass sich an dieser Situation etwas geändert hat? (Conférence du 14/03/2022)

Avez-vous des indications selon lesquelles la Biélorussie réagirait aux sanctions, par exemple en laissant éventuellement entrer en Ukraine les personnes qui se trouvent dans la zone frontalière, si elles étaient assez téméraires pour traverser la zone de guerre ? Avez-vous des indications à ce sujet ? [...] Il y a eu des réfugiés dans la région frontalière entre la Biélorussie et l'Ukraine qui ont attendu là pour venir en Europe, et ils sont restés longtemps dans la région frontalière sans pouvoir avancer. Avez-vous des indications selon lesquelles cette situation a changé ?

Les répétitions de la question *haben Sie Hinweise darauf* sont éloignées, contrairement à l'anaphore de l'exemple 2. Or, dans ce passage, l'anaphore est conservée dans la TA. En effet, DeepL a traduit ces expressions par *avez-vous des indications selon lesquelles* ou *à ce sujet*. Cela permet dès lors d'affirmer que la distance entre les éléments répétés n'influencerait *a priori* pas la manière dont sont traitées les anaphores.

EXEMPLE 4

Es geht nicht ausschließlich **nur um** die Energielieferungen. **Es geht** zum Beispiel **auch darum**, dass der Zahlungsverkehr mit Russland in verschiedenen anderen Bereichen in irgendeiner Form weiterfunktionieren muss. **Das betrifft** unter anderem humanitäre Hilfsleistungen; **das betrifft** legitime wirtschaftliche Geschäfte; **das betrifft** die Zivilgesellschaft. Es geht also weit über

Il ne s'agit pas uniquement des livraisons d'énergie. Il s'agit par exemple aussi du fait que les paiements avec la Russie doivent continuer à fonctionner sous une forme ou une autre dans différents autres domaines. Cela concerne entre autres l'aide humanitaire ; cela concerne les transactions économiques légitimes ; cela concerne la société civile. Cela va donc bien au-delà du domaine de l'énergie.

den Energiebereich hinaus. (Conférence du 28/02/2022)

Dans l'exemple 4, deux anaphores différentes peuvent être observées : la première avec l'expression *es geht um/darum* et la seconde avec le verbe *betreffen*. Aucune des répétitions n'a été supprimée dans la TA, ce qui permet de conserver l'effet d'insistance créé par ces répétitions.

EXEMPLE 5

<i>Wir sehen furchtbare Bilder und Informationen aus der Ukraine, was die ukrainischen Städte und die Lage dort angeht, was Artilleriebeschuss, was das Aushungern von Großstädten angeht, was die Versorgung insgesamt angeht.</i>	<i>Nous voyons des images et des informations terribles en provenance d'Ukraine, en ce qui concerne les villes ukrainiennes et la situation qui y règne, en ce qui concerne les tirs d'artillerie, en ce qui concerne l'affamement des grandes villes, en ce qui concerne l'approvisionnement en général.</i>
--	---

(Conférence du 07/03/2022)

Ici, dans le TS, le verbe *angehen* est utilisé à trois reprises, alors que dans la TA, le verbe correspondant, c'est-à-dire *concerner*, est présent à quatre reprises. Par conséquent, l'on peut affirmer que la traduction de DeepL renforce l'effet produit par la répétition.

EXEMPLE 6

<i>Es geht auch darum, dass die jeweiligen Konfliktpartner - so will ich sie jetzt nennen - auch weiterhin im Gespräch miteinander sind und sich auch die internationale Staatengemeinschaft verantwortlich fühlt. [...] Darum geht es.</i>	<i>Il s'agit également de faire en sorte que les différents partenaires du conflit - c'est ainsi que je veux les appeler maintenant - continuent à dialoguer entre eux et que la communauté internationale se sente également responsable. [...] C'est de cela qu'il s'agit.</i>
---	--

(Conférence du 07/03/2022)

Ici aussi, dans le texte source, le même verbe est utilisé à deux reprises, comme c'est le cas dans la TA. La dernière phrase a toute son importance, tant sa répétition que sa progression des idées. En effet, le locuteur entend d'une part résumer ce qu'il vient de dire au moyen de

la répétition, et d'autre part insister sur ce dont il s'agit en plaçant *darum* en tête de phrase. À cet égard, la TA respecte tout à fait le texte source.

Par conséquent, en ce qui concerne les anaphores et la progression des idées, il semble difficile de tirer des conclusions définitives sur la tendance de DeepL. En effet, indépendamment de la distance entre les éléments répétés, DeepL reproduit ou non les anaphores dans la TA. Quant à la progression des idées, aucune observation ne permet de conclure à une tendance ou à une autre. En ce qui concerne cet aspect, l'on peut dès lors conclure à l'hypothèse du hasard.

- **Les expressions et mots étrangers**

L'analyse d'un extrait long, a permis de remarquer que l'expression anglaise *business as usual* présente dans le TS allemand avait été reportée dans la TA. Ce report ne posait pas de problème, puisqu'on entend souvent parler de *business as usual* en français dans les discussions politiques et économiques. Aux trois occurrences de cette expression, celle-ci a été conservée dans la TA. Toutefois, le report des expressions étrangères – qu'il soit opportun ou problématique – est-il un comportement systématique de DeepL ?

EXEMPLE 1

<i>Insofern gibt es auch gewisse Abwehrmaßnahmen dagegen. Aber wir werden sie hier nicht en détail ausführen.</i>	<i>Il existe donc aussi certaines mesures de défense. Mais nous ne les détaillerons pas ici.</i>
--	--

(Conférence du 28/02/2022)

Cet exemple est particulier, puisque le locuteur allemand utilise une expression française. On pourrait dès lors s'attendre à un simple report de l'expression dans la TA. Toutefois, DeepL a ici reformulé la phrase ; au lieu de traduire le verbe *ausführen* de manière littérale et de conserver le groupe prépositionnel *en détail*, la machine semble avoir regroupé les deux idées en un verbe : *détailler*. Cela est probablement dû au fait que DeepL est programmé sur la paire de langues allemand-français, le français étant dès lors compris dans cette paire de langues.

EXEMPLE 2

<i>Wir reden jetzt, wenn ich das richtig im Kopf habe, von einer Größe más o menos von 48 Milliarden Euro, vielleicht von etwas weniger. (Conférence du 28/02/2022)</i>	<i>Nous parlons maintenant, si je me souviens bien, d'une taille más o menos de 48 milliards d'euros, peut-être un peu moins.</i>
--	--

Dans cet exemple, on retrouve l'expression espagnole *más o menos*, qui se traduit littéralement par *plus ou moins*. D'abord, sans explication apparente, celle-ci semble avoir été conservée telle quelle dans la TA. Mais surtout, celle-ci engendre un problème de syntaxe. En effet, si l'on utilisait l'expression équivalente en français, l'on dirait *d'une taille de plus ou moins 48 milliards d'euros*, et non *d'une taille plus ou moins de 48 milliards d'euros*. L'expression espagnole semble dès lors perturber la machine. Il semble ainsi adéquat de conclure que, étant paramétré sur la paire de langues allemand-français, DeepL n'a pas pu déchiffrer l'expression espagnole, puisque, s'il avait pu la déchiffrer, il n'y aurait probablement pas eu d'erreur de syntaxe.

EXEMPLE 3

<i>Ihr Schlachtruf „Nein zum Krieg!“, „Njet woinoi!“ ist ja mittlerweile weltbekannt. (Conférence du 02/03/2022)</i>	<i>Leur cri de guerre "Non à la guerre !", "Njet woïnoï !" est désormais connu dans le monde entier.</i>
---	--

Cet exemple est lui aussi curieux. Dans le TS, on retrouve l'expression russe „*Njet woinoi*“. Celle-ci n'a pas été reproduite telle quelle dans la TA. En effet, l'orthographe a été modifiée. Cela semble être une erreur, puisque je ne trouve aucune occurrence de *njet woïnoï* sur Google. Une explication est toutefois possible : il pourrait s'agir d'une simple règle de transcription phonétique ; ainsi, le lecteur francophone prononcerait *woinoi* avec le son /v/ et non avec le son /w/. De la même manière, les <i> pourraient être devenus des <ï> afin que le lecteur francophone prononce la diphtongue /ɔi/ plutôt que le son /wa/. Il n'est cependant pas possible d'expliquer pourquoi la machine reproduit ce « choix » visiblement conscient. Il se peut, par exemple, que cette transcription soit présente dans les données d'entraînement de DeepL, mais il n'est pas possible de le vérifier.

EXEMPLE 4

<i>Im Sinne des „burden sharing“ und der damit verbundenen Aufgaben, die man übernommen hat, steht man nie alleine bei der Landesverteidigung. (Conférence du 28/02/2022)</i>	<i>Dans l'esprit du "burden sharing" et des tâches qui y sont liées et que l'on a acceptées, on n'est jamais seul en matière de défense nationale.</i>
---	--

Dans cet exemple, l'expression anglaise *burden sharing* a été reportée en français. Ce cas est similaire à celui de *business as usual*, puisqu'il s'agit également d'une expression courante en politique. Son report est dès lors tout à fait justifié.

En conclusion, les expressions étrangères restent souvent non traduites, sauf lorsqu'il s'agit d'une expression française dans le TS (voir exemple 1). Il arrive également qu'une expression étrangère perturbe la syntaxe de la TA (voir exemple 2). Deux causes peuvent potentiellement expliquer cela : soit, l'expression étrangère pose un problème lorsqu'elle n'est pas entre guillemets ; soit, elle est problématique lorsqu'elle n'est pas en anglais. La dernière hypothèse semble plus pertinente, puisque, dans l'exemple 3, c'est une expression russe que DeepL a mal traitée.

- **Les particules modales *ja* et *doch***

Les deux particules modales que j'analyse sont *ja* et *doch*, puisque ce sont les deux particules que Kefer (n.d.) identifie comme des particules de connivence. Dès lors, dans le contexte de la conférence de presse, on peut s'attendre à ce que ces deux particules soient souvent employées. Kefer explique en outre que *ja* peut servir à marquer l'étonnement du locuteur. Selon Métrich et al. (2002, pp. 121–135), il faut faire une distinction entre la particule modale *ja* accentuée et la particule modale *ja* non accentuée. Lorsqu'elle est non accentuée, *ja* en tant que particule modale « présente le contenu de l'énoncé (fait, argument, opinion etc.) comme relevant du connu ou de l'évident, comme acquis [...] » (*ibid.*, p. 121). En revanche, lorsque *ja* est une particule modale accentuée, elle « souligne l'importance que le locuteur ou la personne dont il parle attache à la réalisation du contenu de l'énoncé » (*ibid.*, p. 132).

Au total, sur 128 occurrences de la particule modale *ja*, seules 13 ont été traduites, c'est-à-dire que presque 90 % des particules *ja* n'ont pas été traduites. Trois traductions différentes de *ja* peuvent être observées.

EXEMPLE 1

Frage: Eine Frage an das Auswärtige Amt und gegebenenfalls an BMI und Justizministerium.

Question : Une question au ministère des Affaires étrangères et, le cas échéant, au BMI et au ministère de la Justice.

Nach dem EU-Außenministertreffen hat Kommissionspräsidentin von der Leyen ein europaweites Verbot der russischen Staatssender RT, Sputnik und ihrer Tochtergesellschaften angekündigt. Können Sie mir sagen, wie dieses Verbot aussehen soll, wie die Bundesregierung dazu steht und wie das europaweit geschehen soll? Medienregulierung ist ja Sache der Nationalstaaten. (Conférence du 28/02/2022)

Après la réunion des ministres des Affaires étrangères de l'UE, la présidente de la Commission von der Leyen a annoncé une interdiction à l'échelle européenne des chaînes publiques russes RT, Sputnik et de leurs filiales. Pouvez-vous me dire comment cette interdiction doit se présenter, quelle est la position du gouvernement fédéral à ce sujet et comment cela doit se faire à l'échelle européenne ? La régulation des médias est en effet l'affaire des Etats nationaux.

Dans cet exemple, la particule modale *ja* a été traduite par *en effet*. Le *ja*, utilisé ici comme particule modale, est non accentué, c'est-à-dire qu'il présente – comme l'expliquent Métrich et al. (2002) – le fait décrit comme connu et évident. On peut dès lors affirmer que cette particule modale est une particule de connivence, comme l'expliquent Kefer (n.d.) et Blumenthal (1997). En effet, cette explication est logique, puisque le locuteur demande comment la régulation des médias peut se faire à l'échelle européenne alors qu'il s'agit d'un pouvoir des États nationaux. Le locuteur suppose donc que son interlocuteur est conscient de cette information. Cette nuance n'apparaît pas dans la TA avec l'utilisation de la locution *en effet*, qui ne sert qu'à indiquer une explication ou à souligner une affirmation ('En effet', Larousse). Dès lors, une piste de traduction plus correcte serait : *vous savez très bien que la régulation des médias est une affaire des États nationaux.*

EXEMPLE 2

Frage: An Herrn Hebestreit und gegebenenfalls Frau Sasse: Frau von der

Question : A M. Hebestreit et éventuellement à Mme Sasse : Mme von

Leyen hat gestern im Hinblick auf die EU-Perspektive der Ukraine wörtlich gesagt: Sie gehören zu uns und wir wollen sie drin haben. Umfassen dieses „uns“ und das „wir“ auch die Position der Bundesregierung?

der Leyen a dit hier textuellement, en ce qui concerne la perspective de l'Ukraine dans l'UE : ils font partie de nous et nous voulons qu'ils y soient. Est-ce que ce "nous" et ce "nous" englobent également la position du gouvernement fédéral ?

STS Hebestreit: Dazu kann ich auf die Ratsschlussfolgerung vom 24. Februar, also vom Donnerstag - das war der Abend nach dem Überfall auf die Ukraine - verweisen. Darin heißt es: Wir erkennen - Zitat - „die europäischen Bestrebungen der Ukraine und ihre Entscheidung für Europa, wie sie im Assoziierungsabkommen zum Ausdruck kommen, an.“ Genau dazu steht auch die Bundesregierung.

STS Hebestreit : Je peux vous renvoyer à la conclusion du Conseil du 24 février, c'est-à-dire du jeudi - c'était le soir après l'attaque contre l'Ukraine. On y lit ceci : Nous reconnaissons - je cite - "les aspirations européennes de l'Ukraine et son choix pour l'Europe, tels qu'ils sont exprimés dans l'accord d'association". C'est exactement ce que soutient le gouvernement allemand.

Zusatzfrage: Bedeutet das eine positive Stellungnahme zu dem, was auch Selensky jetzt noch einmal aktuell gefordert hat, nämlich die Aufnahme der Ukraine in die EU möglicherweise nach einem beschleunigten Verfahren? Denn das ist ja eine Konkretisierung des Prozesses. Unterstützt die Bundesregierung dies? (Conférence du 28/02/2022)

*Question supplémentaire : Est-ce que cela signifie une prise de position positive sur ce que Selensky vient de réclamer à nouveau, à savoir l'admission de l'Ukraine dans l'UE, éventuellement selon une procédure accélérée ? Car il s'agit **bien** d'une concrétisation du processus. Le gouvernement fédéral soutient-il cette démarche ?*

Métrich et al. (2002, p. 122) évoquent la traduction de *ja* par *bien*. Selon eux, lorsque le locuteur souhaite présenter ses propos comme évidents, la traduction de *ja* par *bien* est envisageable. Cela implique aussi une certaine forme de connivence. Dans cet exemple, c'est en effet de cela qu'il s'agit.

EXEMPLE 3

Vorsitzende Welty: Kann uns das Finanzministerium zu SWIFT weiterhelfen?

La présidente Welty : Est-ce que le ministère des Finances peut nous aider sur SWIFT ?

*Kuhn: Ich kann nur unterstreichen, was die Kollegin gerade ausgeführt hat. Ziel dieser Maßnahmen ist es - darauf hat die Bundesregierung mit den europäischen, den amerikanischen und den übrigen internationalen Partnern hingewirkt; Japan ist **ja** noch dazugestoßen -, dass diese Maßnahmen so ausgerichtet werden, dass sie zielgerichtet sind. (Conférence du 28/02/2022)*

*Kuhn : Je ne peux que souligner ce que vient de dire ma collègue. L'objectif de ces mesures - le gouvernement fédéral a œuvré dans ce sens avec les partenaires européens, américains et les autres partenaires internationaux ; le Japon s'y est **d'ailleurs** ajouté - est que ces mesures soient orientées de telle sorte qu'elles soient ciblées.*

Métrich et al. (2002, p. 122) mentionnent une nuance subtile que le *ja* peut ajouter lorsqu'il est utilisé dans une phrase déclarative. Il arrive que la particule modale *ja* soit insérée dans une phrase pour que cette dernière se présente comme un rappel, un commentaire ou un argument. Ici, la proposition *Japan ist ja noch dazugestoßen* peut tout à fait être comprise comme un commentaire secondaire, presque comme une parenthèse, ce qui peut s'expliquer par l'utilisation d'une incise. Dès lors, la traduction de *ja* par *d'ailleurs* semble dans ce cas pertinente.

Il y a aussi, comme évoqué plus haut, de nombreux cas dans lesquels la particule modale *ja* n'a pas été traduite. En voici quelques exemples.

EXEMPLE 4

Frage: Zum Thema Flüchtlinge: Wie viele Menschen aus der Ukraine sind hier schon angekommen?

Question : A propos des réfugiés : Combien de personnes d'Ukraine sont-elles déjà arrivées ici ?

*Es gab **ja** schon vor dem Krieg sehr viele Flüchtlinge aus Belarus, die in die Ukraine gelangt sind. Haben Sie Erkenntnisse, ob*

Il y avait déjà avant la guerre de très nombreux réfugiés de Biélorussie qui se sont rendus en Ukraine. Avez-vous des informations sur le nombre de ces

und, wenn ja, wie viele von denen sich auf den Weg in den Westen gemacht haben? *personnes qui ont pris le chemin de l'Ouest ?*

(Conférence du 28/02/2022)

Ici, la particule modale *ja* est une particule non accentuée, ce qui signifie qu'elle a pour but de présenter les propos du locuteur comme connus et évidents. Ainsi, la phrase agit plutôt comme un rappel du locuteur à son interlocuteur. Cette nuance n'est pas présente dans la TA. Ici, une traduction explicitante serait envisageable : *comme nous le savons*. Toutefois, cette traduction est discutable, car elle entraîne une lourdeur dans le TC.

EXEMPLE 5

<i>Frage: Meine Frage geht ebenfalls an das BMG. Wie wichtig ist gerade jetzt die Anwendung auch der sektoralen Impfpflicht - sie wird ja diese Woche in Kraft treten -, speziell auch vor dem Hintergrund der derzeit steigenden Zahlen, die Sie gerade erwähnt haben?</i>	<i>Question : Ma question s'adresse également au BMG. Quelle est l'importance de l'application de l'obligation vaccinale sectorielle - elle entrera en vigueur cette semaine -, en particulier dans le contexte de l'augmentation actuelle des chiffres que vous venez de mentionner ?</i>
--	--

(Conférence du 14/03/2022)

Dans cet exemple, la particule modale *ja*, non accentuée, est utilisée comme un commentaire, une parenthèse. Elle n'a pas été traduite dans la TA. En revanche, comme elle se trouve dans une proposition située en incise, elle constitue déjà une sorte de parenthèse. Toutefois, à l'oral, cela ne transparaîtrait pas ; l'utilisation de la locution *d'ailleurs* peut être une piste d'amélioration.

Pour conclure cette partie sur la particule modale *ja*, on peut voir que, dans environ 90 % des cas, elle n'est pas traduite, ce qui engendre parfois une perte d'effet et de nuance. Parmi les cas de non-traduction, celle-ci s'avère problématique – c'est-à-dire qu'elle engendre une perte dans la TA – dans environ 80 % des cas. Par ailleurs, lorsque la particule modale *ja* est traduite, c'est en grande partie par la locution *en effet*, qui ne rend que très rarement l'effet de la particule allemande. En cas de particule modale *ja* non accentuée, la traduction par *en effet* est occasionnelle (Métrich et al., 2002, p. 121). Lorsque *ja* est une particule modale accentuée, la traduction par *en effet* n'est pas envisageable (*ibid.*, p. 132).

La particule modale *doch* est également riche en nuances. Dans le corpus, cette particule est restée non traduite dans 8 cas sur 13. Les deux traductions qui se retrouvent dans la TA sont *bien* et *tout de même*, qui sont des traductions que Métrich et al. (1995, pp. 123–146) évoquent pour la particule modale *doch*.

EXEMPLE 1

<i>Frage: [...] Reden wir davon, dass diese Mittel in den nächsten - ich erfinde jetzt eine Zahl - zwei, drei Jahren abfließen können oder muss man sich eher 20 Jahre vorstellen? [...]</i>	<i>Question : [...] Est-ce que nous parlons de la possibilité d'écouler ces fonds dans les deux ou trois prochaines années - je vais inventer un chiffre - ou faut-il plutôt envisager 20 ans ? [...]</i>
<i>STS Hebestreit: Vielleicht darf ich noch ergänzen: Heute ist Montag. Dieses Sondervermögen ist am Sonntag verkündet worden, und es ist in den Tagen davor entwickelt worden. Es wäre doch etwas kühn, jetzt schon konkrete Projekte und Aufteilungen zu erwarten.</i>	<i>STS Hebestreit : Je peux peut-être encore compléter : Nous sommes lundi. Ce fonds spécial a été annoncé dimanche, et il a été développé dans les jours précédents. Il serait tout de même un peu audacieux de s'attendre à des projets concrets et à des répartitions dès maintenant.</i>

(Conférence du 28/02/2022)

Dans la réponse de Hebestreit, la particule *doch* présente « le contenu de l'affirmation et sa valeur argumentative [...] comme relevant de l'évidence » (Métrich et al., 1995, p. 128). En effet, la question est de savoir si les fonds vont être écoulés dès maintenant ou dans plusieurs années. Après avoir donné quelques preuves, Hebestreit affirme que les projets concrets vont commencer dès maintenant ; cet argument est renforcé par la particule modale *doch*. Dans la TA, *doch* est devenu *tout de même*, qui permet également de rendre cette nuance. La traduction *quand même* aurait également été possible.

EXEMPLE 2

<i>[...] der Finanzminister [hat heute] deutlich gemacht, dass man im Rahmen dessen bei einer Neuverschuldung von 99,7 Milliarden Euro bleiben möchte, um eben</i>	<i>[...] le ministre des Finances a clairement indiqué aujourd'hui qu'il souhaitait rester dans le cadre d'un nouvel endettement de 99,7 milliards d'euros, afin de pouvoir</i>
--	---

<p><i>auch in den nächsten Jahren weiterhin das stemmen zu können, was man sich jetzt vorgenommen hat. Insofern ist jede einzelne Maßnahme sicherlich diskussionswürdig und Diskussion berechtigt, aber das muss am Ende, und das ist die vornehmste Aufgabe des Finanzministers, alles zusammenpassen und sich rechnen. Insofern kann man jetzt zu jeder einzelnen Maßnahme sagen: Wenn es für die Verteidigung ein Sondervermögen in Höhe von 100 Milliarden Euro gibt, dann muss doch für dieses oder jenes auch noch Geld da sein.</i></p>	<p><i>continuer à assumer ce que nous avons décidé de faire dans les années à venir. Dans ce contexte, chaque mesure individuelle mérite certainement d'être discutée et justifiée, mais au final, et c'est la tâche la plus importante du ministre des Finances, tout doit s'accorder et se calculer. Dans cette mesure, on peut maintenant dire pour chaque mesure individuelle : s'il existe un fonds spécial de 100 milliards d'euros pour la défense, il doit bien y avoir de l'argent pour ceci ou cela.</i></p>
---	---

(Conférence du 28/02/2022)

Dans cet exemple, *doch* a également une valeur argumentative et présente un propos censé être connu et évident ; dans ce sens-ci, la particule modale *doch* peut souvent être utilisée avec des verbes de modalité tels que *müssen*, ce qui est ici le cas (Métrich et al., 1995, p. 127). Cette explication est logique, comme l'explique le locuteur : s'il y a de l'argent disponible pour telle ou telle chose, alors il y a bien de l'argent pour d'autres choses. Dans ce contexte, le *bien* utilisé dans la TA est tout à fait correct et contient la même nuance que le *doch*.

Cependant, comme expliqué plus haut, dans la plupart des cas, la particule modale *doch* n'est pas traduite dans la TA.

EXEMPLE 3

<p><i>Frage: Herr Wede, aus jedem Bericht über die Situation zum Beispiel an den Bahnhöfen wird klar, dass es ohne die Ehrenamtlichen innerhalb von wenigen Stunden zum reinen Chaos kommen kann. Wie reagiert die Bundesregierung darauf?</i></p>	<p><i>Question : Monsieur Wede, il ressort clairement de chaque rapport sur la situation, par exemple dans les gares, que sans les bénévoles, le chaos pur et simple peut s'installer en quelques heures. Comment le gouvernement fédéral réagit-il</i></p>
--	---

<i>Im Grunde ist die Versorgung von Flüchtlingen doch eine staatliche Aufgabe. Der Eindruck ist, dass Sie dieser nicht hinreichend nachkommen.</i>	<i>à cela ? En principe, la prise en charge des réfugiés est une tâche étatique. On a l'impression que vous ne vous en acquittez pas suffisamment.</i>
---	--

(Conférence du 14/03/2022)

Dans ce cas, *doch* est utilisé dans son emploi le plus courant, comme l'expliquent Métrich et al. (1995, p. 126) :

Doch marque d'une façon générale que le contenu de la déclarative est présenté comme indubitablement établi (acquis, connu, évident) face à un interlocuteur ou un tiers (qui peut être le locuteur lui-même [...]) qui donne le sentiment, par ses propos ou son comportement, de penser le contraire.

Dès lors, *doch* est ici utilisé comme un argument et marque un reproche, voire l'agacement du locuteur (*loc. cit.*). La TA n'a pas la même force argumentative que le TS et ne donne pas l'impression qu'il s'agit d'un reproche. La non-traduction de *doch* est alors quelque peu problématique et engendre une perte dans la TA.

EXEMPLE 4

<i>Frage: An das Auswärtige Amt zu den Verhandlungen im Normandie-Format: Ein Argument dafür, dass man keine Waffen geliefert hat, war ja auch gewesen, dass man neutral ist, also dass Deutschland und Frankreich nur zwischen der Ukraine und Russland vermitteln.</i>	<i>Question : Au ministère des Affaires étrangères concernant les négociations en format Normandie : l'un des arguments pour ne pas livrer d'armes était que l'on était neutre, c'est-à-dire que l'Allemagne et la France ne faisaient que de la médiation entre l'Ukraine et la Russie.</i>
--	--

[...]

[...]

<i>Zusatzfrage: Hat sich durch die Entscheidung der Bundesregierung am Samstag, dass man der Ukraine jetzt Waffen liefern möchte, nicht etwas verändert? Damit hat man doch Partei bezogen - ob zu Recht oder zu Unrecht - und kann jetzt nicht mehr damit</i>	<i>Question supplémentaire : la décision prise samedi par le gouvernement allemand de livrer des armes à l'Ukraine n'a-t-elle pas changé quelque chose ? On a pris parti - à tort ou à raison - et on ne peut plus argumenter en disant que l'on est un médiateur neutre.</i>
---	---

*argumentieren, dass man ein neutraler
Vermittler ist.*

(Conférence du 28/02/2022)

Dans cet exemple, la particule *doch* sert à présenter les propos du locuteur comme confirmés et maintenus face à un fait ou à une idée contraire à ces propos (Métrich et al., 1995, p. 123). En effet, l'Allemagne a pris parti en décidant de livrer des armes à l'Ukraine, ce qui ne correspond pas à la neutralité que l'Allemagne envisageait au départ. Cette nuance n'apparaît pas clairement dans la TA. Il aurait été possible d'explicitier cette nuance en utilisant *car*, une traduction que proposent Métrich et al. (*loc. cit.*) : *car on a pris parti [...]*.

Quant à la particule modale *doch*, lorsqu'elle est traduite dans la TA, sa traduction est correcte et plus ou moins équivalente à la valeur du *doch* dans le TS. En revanche, lorsqu'elle n'est pas traduite, on constate une perte d'effet dans la TA.

5.4. Discussion des résultats

Les six aspects qui ont fait l'objet d'une étude quantitative sont les déictiques *jetzt* et *hier*, les phrases interrompues ou incomplètes, les incises et digressions, les anaphores, les expressions et mots étrangers, ainsi que les deux particules modales allemandes de connivence *ja* et *doch*.

Tout d'abord, alors que l'état de l'art suggère que les omissions ont lieu principalement dans les phrases « longues », l'analyse de la TA des déictiques *jetzt* et *hier* n'a pas permis de corroborer ce phénomène. Même si ces deux adverbes sont souvent omis — surtout *hier*, qui atteint un taux d'omission de plus de 35 % —, ce n'est en réalité pas la longueur des phrases qui permet d'expliquer ce phénomène. En effet, d'après cette étude, c'est plutôt la quantité d'adverbes et de circonstanciels dans une phrase qui est un facteur décisif ; par exemple, en cas de cascades d'adverbes, au moins l'un d'eux est omis dans la TA. D'ailleurs, lorsque *jetzt* et *hier* sont utilisés dans une même phrase, il est rare que les deux adverbes soient traduits. En outre, dans la plupart des cas, la TA de ces déictiques ne comprend pas leur valeur et leur ancrage spatio-temporel présents dans le TS. Il apparaît par conséquent que DeepL n'est pas en mesure de déchiffrer ces déictiques et de les transposer à la situation et dans le contexte de l'énoncé.

En ce qui concerne les phrases incomplètes, l'état de l'art semble indiquer que DeepL tend à les compléter. Or, cela ne semble pas être le cas, sauf lorsqu'il manque un verbe dans le TS et que celui-ci est facile à déduire du co(n)texte. Dans ces cas-ci, DeepL tend à compléter les phrases. Cela dit, il est difficile de déterminer avec certitude pourquoi DeepL ne complète pas les phrases auxquelles il manque toute une partie. Deux explications sont possibles : soit, cela est dû au développement de la TAN et à l'entraînement de celle-ci ; soit, cela est dû à un signal inséré dans le TS pour signaler que la phrase n'est pas complète. La deuxième explication semble plus plausible. En effet, la plupart des phrases incomplètes se terminent par « - - - », ce qui — *a priori* — empêche DeepL de compléter la phrase. En effet, lorsque l'on supprime ces signaux dans le TS, DeepL tend à compléter ces phrases aléatoirement, ce qui constitue un ajout significatif, et même un problème de sens. Quant aux phrases interrompues, elles sont souvent difficiles à traiter pour DeepL ; par exemple, lorsque le locuteur A parle, qu'un locuteur B lui coupe la parole, et que le locuteur A reprend ensuite sa phrase, la syntaxe est totalement perturbée, et la phrase n'a plus de logique.

Ensuite, les incises et digressions, entourées par des tirets, présentes dans le TS sont, à une exception près, toujours conservées telles quelles dans la TA. Ce phénomène n'est pas un problème en soi, mais il arrive que ces incises affectent la lisibilité de la TA, voire qu'elles engendrent des problèmes de syntaxe en français. Dans le corpus parallèle, l'on peut tout de même observer un exemple intéressant de restructuration : une incise a été transformée en proposition principale.

En ce qui concerne la traduction ou non des anaphores, il ne semble toutefois pas possible de dégager une tendance. En effet, indépendamment de la longueur de la phrase et de la distance qui sépare les éléments constituant l'anaphore, DeepL peut — ou non — reproduire l'anaphore. Toutefois, la non-traduction des anaphores engendre souvent une perte d'effet, par exemple celui de l'insistance, et peut parfois affaiblir la force argumentative de l'énoncé.

DeepL semble, par ailleurs, faire face à certaines difficultés lorsqu'il traduit des expressions et mots étrangers. Lorsqu'il est paramétré sur une paire de langues définies, ici allemand-français, il semble effectivement ne pas toujours être en mesure de déchiffrer les mots autres qu'allemands et français. Toutefois, le corpus ne contient pas suffisamment d'expressions étrangères pour pouvoir corroborer cette observation.

Enfin, les particules modales *ja* et *doch* se sont révélées problématiques elles aussi, surtout la première. À peine plus de 10 % des *ja* ont été traduits ; bien que leur traduction ne soit pas systématiquement inappropriée, elle reste problématique dans la majorité des exemples. Par exemple, *ja* a été traduit principalement par *en effet*, qui ne correspond que très rarement à la particule allemande. Par ailleurs, dans 80 % des cas de non-traduction de *ja*, il y a une perte importante. Quant à la particule *doch*, lorsqu'elle est traduite, sa traduction est correcte. Néanmoins, comme pour *ja*, quand elle n'est pas traduite, il manque une nuance importante dans la TA. Ici aussi, rien ne semble pouvoir expliquer pourquoi ces particules sont traduites dans certains cas et ne le sont pas dans d'autres.

6. Conclusion

L'objectif de cette étude était d'analyser la manière dont DeepL, un programme de traduction automatique neuronale, traite les éléments propres à l'oralité. La question principale de ce travail était de savoir s'il était possible de dégager certaines tendances, voire d'observer des phénomènes récurrents dans une TA de DeepL. Pour ce faire, j'ai compilé un corpus parallèle composé de conférences de presse du gouvernement allemand portant sur l'invasion de l'Ukraine par la Russie et de leur traduction automatique. Ce corpus a été exploité à la fois manuellement et automatiquement sur le logiciel XBenCh. L'état de l'art des limites de la traduction automatique neuronale ainsi que l'analyse qualitative et exploratoire des extraits longs ont permis d'établir une liste de marqueurs d'oralité à analyser de manière quantitative. L'hypothèse de départ était que, au vu des récentes améliorations de la TAN, il est possible que la TA soit globalement acceptable ; cependant, étant donné que les données d'entraînement de la machine consistent principalement en des textes écrits, certains marqueurs d'oralité devraient se révéler problématiques pour DeepL. Grâce à cette analyse quantitative de la traduction automatique produite en octobre 2022 par DeepL, il a été possible, dans certains cas, de dégager des phénomènes récurrents, ou du moins certaines tendances.

L'un des points principaux de l'état de l'art concernait les phrases longues : selon les recherches, plus les phrases sont longues, plus la TA est problématique, notamment en termes d'omissions. Or, les observations formulées dans ce travail ne permettent pas de corroborer ce phénomène. En effet, de nombreuses omissions peuvent être observées, mais celles-ci ne sont pas dues à la longueur des phrases. Par exemple, l'analyse de la TA des déictiques a permis de démontrer que c'est plutôt la quantité d'adverbes et de circonstanciels qui influence le taux d'omissions ; ainsi, en cas de cascades d'adverbes, les omissions sont souvent importantes. En ce qui concerne les déictiques *jetzt* et *hier*, qui ont été analysés dans ce travail, leur non-traduction ou leur traduction inappropriée a un effet important au niveau de l'ancrage spatio-temporel de l'énoncé. Qu'ils soient traduits littéralement ou qu'ils ne soient simplement pas traduits, la valeur de l'énoncé est effectivement différente ; DeepL ne semble dès lors pas être en mesure de replacer les énoncés dans leur contexte. En revanche, l'état de l'art indiquait également que les phrases longues pouvaient affecter la lisibilité — et parfois même la syntaxe — de la TA. Ce phénomène a, quant à lui, pu être confirmé dans

l'analyse du corpus parallèle. À plusieurs reprises, la syntaxe de la TA de DeepL n'est pas correcte.

Un autre point qui ne permet *a priori* pas de confirmer l'état de l'art concerne les phrases incomplètes. Plusieurs chercheurs affirment que DeepL tend à compléter ces phrases, ce qui ne se produit qu'à de très rares occasions dans le corpus. Cela semble toutefois être dû à l'ajout de trois tirets à la fin des phrases incomplètes dans le texte source, qui empêchent DeepL de compléter la phrase de manière aléatoire. Le corpus ne comprend toutefois pas suffisamment d'exemples pour que cette hypothèse puisse être vérifiée.

Enfin, il est fait mention dans l'état de l'art des « marqueurs pragmatiques », qui ont été analysés dans ce travail grâce aux particules modales *ja* et *doch*. L'état de l'art est clair à cet égard : ces marqueurs d'oralité sont problématiques pour la TAN. Ce travail a permis de corroborer cette observation. D'un côté, lorsque ces particules sont traduites, leur traduction ne reflète que rarement la valeur des particules modales allemandes, qui peuvent apporter de nombreuses nuances différentes en fonction du co(n)texte. De l'autre, lorsqu'elles ne sont pas traduites, l'on peut constater une perte dans la TA. À l'instar des déictiques *jetzt* et *hier*, les particules modales *ja* et *doch* se révèlent particulièrement problématiques pour les systèmes de TAN.

Pour conclure, la traduction des éléments propres à l'oralité n'est pas toujours mauvaise ; au contraire, dans certains rares cas, DeepL procède par exemple à des restructurations intéressantes et évite certaines lourdeurs et répétitions. En dépit des avancées spectaculaires que connaissent la TAN et l'intelligence artificielle en général, comme mentionné dans la première partie de ce travail, il apparaît que certains aspects — notamment certains marqueurs d'oralité — ne peuvent (toujours) pas être traités de manière satisfaisante par l'autoproclamé « meilleur traducteur au monde ». Ces différents marqueurs, dont avant tout les déictiques et les particules modales, ont un point commun qui peut expliquer les difficultés auxquelles DeepL fait face. En effet, ces éléments relèvent de la pragmatique (dans son acception linguistique). Comme le suggère la définition même de la pragmatique, ces marqueurs d'oralité n'ont une signification que lorsque le contexte dans lequel ils sont utilisés et leur vocation discursive sont connus et pris en compte. Par conséquent, DeepL n'étant qu'un réseau de neurones artificiels, il ne peut pas aller au-delà des informations qui lui sont explicitement données, c'est-à-dire le texte à proprement parler. Ainsi, même avec la révolution née de la TAN, notamment l'introduction du mécanisme

d'attention, ces éléments pragmatiques ne sont pas à la portée des programmes de TAN. Dès lors, c'est entre autres pour des aspects comme celui-ci que l'intervention du biotraducteur — qui peut se servir de son intelligence *humaine* — sera toujours nécessaire.

Depuis plusieurs années, l'on observe en traductologie un intérêt croissant pour la recherche en traduction automatique. L'entraînement de modèles de TAN occupe une place de plus en plus prépondérante dans le domaine. L'une des hypothèses principales de ce travail concernait les données d'entraînement des programmes de TA. En effet, les corpus sur lesquels sont entraînés ces programmes sont constitués en majeure partie de textes écrits. Cette étude pourrait être approfondie de plusieurs manières. Il serait notamment intéressant d'observer si l'entraînement d'un système de TAN sur des données appropriées pourrait améliorer les performances de DeepL en matière de traduction des marqueurs d'oralité et, plus largement, des marqueurs pragmatiques. Cependant, comme expliqué dans ce travail, les textes oraux traduits sont beaucoup plus rares que les textes écrits, et les modèles de TAN ont besoin d'un volume important de données d'entraînement pour être performants ; néanmoins, comme expliqué au point 2.3., l'entraînement de programmes de TAN soulève certains problèmes, notamment en ce qui concerne leur impact environnemental. Il serait également intéressant de vérifier certaines hypothèses observées dans l'analyse du corpus parallèle, mais qui n'ont pas pu être vérifiées de manière quantitative. Par exemple, les quelques cas de niveau de langue informel sont-ils automatiquement supprimés dans la TA et transposés dans un niveau plus soutenu ? Une autre hypothèse qu'il serait intéressant de vérifier concerne les anaphores pronominales ; l'état de l'art mentionnait leur caractère problématique pour la TA. Même si l'analyse quantitative ne portait pas sur les anaphores pronominales, il a pu être observé à quelques reprises qu'elles étaient mal traitées par DeepL, qui, souvent, ne parvient pas à repérer leur référent. Par conséquent, les anaphores pronominales sont-elles un réel obstacle pour DeepL ? Il serait possible de vérifier ces hypothèses au moyen d'une étude quantitative sur corpus, comme cela a été fait dans ce travail.

Bibliographie

Corpus

Bundesregierung, Pressekonferenz. (28/02/2022). <https://www.bundesregierung.de/breg-de/suche/regierungspressekonferenz-vom-28-februar-2022-2008756>

Bundesregierung, Pressekonferenz. (02/03/2022). <https://www.bundesregierung.de/breg-de/aktuelles/regierungspressekonferenz-vom-2-maerz-2022-2009448>

Bundesregierung, Pressekonferenz. (07/03/2022). <https://www.bundesregierung.de/breg-de/suche/regierungspressekonferenz-vom-7-maerz-2022-2010748>

Bundesregierung, Pressekonferenz. (14/03/2022). <https://www.bundesregierung.de/breg-de/suche/regierungspressekonferenz-vom-14-maerz-2022-2015546>

Articles scientifiques

Awadh, A. N. M., & Shafiull, K. A. (2020). Challenges of translating neologisms comparative study: Human and machine translation. In *Dil ve Dilbilimi Çalışmaları Dergisi*, 16(4), pp. 1987–2002.

Baker, M. (1993). Corpus Linguistics and Translation Studies – Implications and Applications. In *Text and Technology : In Honour of John Sinclair*, pp. 233–249.

Ballard, M. (1998). La traduction du nom propre comme négociation. In *Palimpsestes*, 11, pp. 199–223.

Blin, R. (2021). Neural machine translation, corpus and frugality. <https://doi.org/10.48550/ARXIV.2101.10650>

Bouamor, H. (2010). Construction d'un corpus de paraphrases d'énoncés par traduction multiple multilingue. In *Actes de la 17e conférence sur le Traitement Automatique des Langues Naturelles. Rencontres jeunes Chercheurs en Informatique pour le Traitement Automatique des Langues*, pp. 44–53.

Burchardt, A., Lommel, A., Bywood, L., Harris, K., & Popović, M. (2016). Machine translation quality in an audiovisual context. In *Target. International Journal of Translation Studies*, 28(2), pp. 206–221.

Dostie, G., & Pusch, C. D. (2007). Présentation. Les marqueurs discursifs. Sens et variation. In *Langue Française*, n° 154(2), pp. 3–12.

- Forcada, M. L. (2017). Making sense of neural machine translation. In *Translation Spaces*, 6(2), pp. 291–309.
- François, F. (2022). Post-édition, mode d'emploi. Retour sur une expérience personnelle. In *Traduire*, 246, pp. 5–9.
- Fraser, B. (1988). Types of English Discourse Markers. In *Acta Linguistica Hungarica*, 38(1/4), pp. 19–33.
- Fraser, T., & Joly, A. (1980). Le système de la déixis (2): Endophore et cohésion discursive en anglais. In *Modèles Linguistiques*, 2(2), pp. 22–51.
- Guerberof-Arenas, A., & Toral, A. (2022). Creativity in translation: Machine translation as a constraint for literary texts. In *Translation Spaces*. <https://doi.org/10.1075/ts.21025.gue>
- Hansen, D., & Houlmont, P.-Y. (2022). A Snapshot into the Possibility of Video Game Machine Translation. In *Proceedings of the 15th Biennial Conference of the Association for Machine Translation in the Americas (Volume 2: Users and Providers Track and Government Track)*, 2, pp. 257–269.
- Hardmeier, C., & Federico, M. (2010). Modelling Pronominal Anaphora in Statistical Machine Translation. In *Proceedings of the 7th International Workshop on Spoken Language Translation*, pp. 283–289.
- Hutchins, W. J. (1995). Machine Translation: A Brief History. In *Concise History of the Language Sciences*, pp. 431–445. Elsevier.
- Kayser-Bril, N. (n. d.). Automated translation is hopelessly sexist, but don't blame the algorithm or the training data. *Algorithm Watch*.
- Kleiber, G. (n.d.). Peut-on définir une catégorie générale de l'anaphore? In *Vox Romanica*, 48, pp. 1–14.
- Liefländer-Koistinen, L. (2004). Modalpartikeln als Übersetzungsproblem. In *Übersetzung – Translation – Traduction. Ein internationales Handbuch zur Übersetzungsforschung. 1. Teilband*. (Harald Kittel, pp. 550–554).
- Loock, Rudy. (2018). Traduction automatique et usage linguistique : une analyse de traductions anglais-français réunies en corpus. In *Meta*, 63(3), pp. 786-806.

- Mohar, T., Orthaber, S., & Onič, T. (2020). Machine Translated Atwood: Utopia or Dystopia? In *ELOPE: English Language Overseas Perspectives and Enquiries*, 17(1), pp. 125–141.
- Perdicoyanni-Paléologou, H. (2009). Le concept d'anaphore, de cataphore et de déixis en linguistique française. In *Revue Québécoise de Linguistique*, 29(2), pp. 55–77. <https://doi.org/10.7202/039441ar>
- Prak-Derrington, E. (2011). Traduire ou ne pas traduire les répétitions. In *Nouveaux Cahiers d'Allemand : Revue de Linguistique et de Didactique*, 29(3), pp. 293–305.
- Ragni, V., & Nunes Vieira, L. (2022). What has changed with neural machine translation? A critical review of human factors. In *Perspectives*, 30(1), pp. 137–158. <https://doi.org/10.1080/0907676X.2021.1889005>
- Schmidhofer, A., & Mair, N. (2018). Machine Translation in Translator Education. In *Salamanca: Ediciones Universidad de Salamanca*, 4(2), pp. 163–180.
- Schoonjans, S. (2014). Zu den französischen Entsprechungen der deutschen Modalpartikeln ja und doch in literarischen Texten. In *Neuphilologische Mitteilungen*, 114(4), p. 401.
- Schumacher, P. (2020). La traduction automatique neuronale: Technologie révolutionnaire ou poudre de perlimpinpin? Compte rendu d'une expérience pédagogique. In *Al-Kīmiyā*, 18, pp. 67–89.
- Shiwen, Y., & Xiaojing, B. (2015). Rule-based Machine Translation. In S. Chan (Ed.), *The Routledge Encyclopedia of Translation Technology*, pp. 186–200. Routledge, Taylor & Francis Group.
- Stukenbrock, A. (2018). Blickpraktiken von SprecherInnen und AdressatInnen bei der Lokaldeixis: Mobile Eye Tracking-Analysen zur Herstellung von *joint attention*. In *Gesprächsforschung - Online-Zeitschrift Zur Verbalen Interaktion*, 19, pp. 132–168.
- Tezcan, A., Daems, J., & Macken, L. (2019). When a 'sport' is a person and other issues for NMT of novels. In *Proceedings of the Qualities of Literary Machine Translation*, pp. 40–49.
- Toral, A., & Way, A. (2014). Is machine translation ready for literature? In *Proceedings of Translating and the Computer*, 36.

- Toral, A., & Way, A. (2018). What level of quality can neural machine translation attain on literary text? In *Translation Quality Assessment*, pp. 263–287.
- Vanmassenhove, E., Shterionov, D., & Gwilliam, M. (2021). Machine Translationese: Effects of Algorithmic Bias on Linguistic Complexity in Machine Translation. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pp. 2203–2213.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need.
<https://doi.org/10.48550/ARXIV.1706.03762>
- Ventura, D., & Bajrić, S. (2020). L’alternance des termes d’adresse entre tutoiement et vouvoiement: Français, espagnol, italien. In *Studia Romanica et Anglicae Zagradiensia*, 65, pp. 65–70.
- Wandruszka, M. (1974). Deutsch im Übersetzungsvergleich. In U. Engel, P. Grebe, & H. Moser (Eds.), *Sprachsystem und Sprachgebrauch: Festschrift für Hugo Moser zum 65. Geburtstag, Teil I*. Pädagogischer Verlag Schwann.
- Wang, H., Wu, H., He, Z., Huang, L., & Ward Church, K. (2021). Progress in Machine Translation. In *Engineering*.
- Way, A. (2018). Quality Expectations of Machine Translation. In *Translation Quality Assessment. Machine Translation: Technologies and Applications* (Vol. 1, pp. 159–178).
- Wong Tak-ming, B., & Webster, J. J. (2015). Example-based Machine Translation. In S. Chan (Ed.), *The Routledge Encyclopedia of Translation Technology* (pp. 137–151). Routledge, Taylor & Francis Group.
- Yang, L., & Min, Z. (2015). Statistical Machine Translation. In S. Chan (Ed.), *The Routledge Encyclopedia of Translation Technology* (pp. 201–212). Routledge, Taylor & Francis Group.
- Zotti, V. (2019). Ressources numériques pour la traduction des mots désignant des realia. In *Dictionnaires et culture numérique dans l’espace francophone. 1: Portrait actuel de la lexicographie en ligne* (Vol. 149). Didier Érudition, Klincksieck.

Chapitres de livre

Atwood, M. (2006). Life Stories. In *The Tent* (Anchor, p. 3).

Conférences

Besacier, L. (2014). Traduction automatisée d'une œuvre littéraire: Une étude pilote. *Traitement Automatique du Langage Naturel (TALN)*, Marseille, France.

Daems, J., & Ruffo, P. (2022). From raw MT to user-centricity: What we've learned so far about quality, textual features, and user expectations in literary MT. *Literary Machine Translation as a Human-Machine Dialectic*, Liège.

Kuzman, T. (2022). Evaluation of Neural Machine Translation Models Tailored to Novels for Literary Translation from English to Slovene. *Literary Machine Translation as a Human-Machine Dialectic*, Liège.

Matusov, E. (2022). Giving translator full control: Customizable MT for literary translation. *Literary Machine Translation as a Human-Machine Dialectic*, Liège.

Ouvrages de référence

Centre national de ressources textuelles et lexicales (en ligne). <https://www.cnrtl.fr/>

Larousse (en ligne). <https://www.larousse.fr/dictionnaires/francais>

Le Robert (en ligne). <https://dictionnaire.lerobert.com/>

Ouvrages scientifiques

Chan, Sin-wai, éd. (2002). *Translation and information technology*. Hong Kong: Chinese University Press.

Blumenthal, P. (1987). *Sprachvergleich Deutsch-Französisch*. Max Niemeyer Verlag.

Blumenthal, P. (1997). *Sprachvergleich Deutsch-Französisch, 2: Vol. Neubearbeitete und ergänzte Auflage*. Max Niemeyer.

Loffler-Laurian, A.-M. (1996). *La traduction automatique*. Presses universitaires du Septentrion.

Malblanc, A. (1968). *Stylistique comparée du français et de l'allemand* (4e éd.). Didier.

- Métrich, R., Faucher, E., & Courdier, G. (1995). *Les invariables difficiles: Dictionnaire allemand-français des particules, connecteurs, interjections et autres 'mots de la communication'*.
- Métrich, R., Faucher, E., Courdier, G., Dalmas, M., Fernandez Bravo, N., & Rubenach, S. (1998). *Les invariables difficiles: Dictionnaire allemand-français des particules, connecteurs, interjections et autres mots de la communication*. Association des Nouveaux cahiers d'allemand.
- Métrich, R., Fernandez-Bravo, N., & Rubenach, S. (2002). *Les invariables difficiles: Dictionnaire allemand-français des particules, connecteurs, interjections et autres 'mots de la communication'*.
- Schreiber, M. (2006). *Grundlagen der Übersetzungswissenschaft: Französisch, italienisch, spanisch*. Max Niemeyer.
- Venuti, L. (2018). *The translator's invisibility: A history of translation* (Third edition). Routledge.
- Youdale, R. (2020). *Using computers in the translation of literary style: Challenges and opportunities*. Routledge/Taylor & Francis Group.
- Zanettin, F. (2012). *Translation-driven Corpora*. St Jerome Publishing.

Autres

- Charles de Gaulle. (1944). *Discours du 25 août 1944*.
- Kefer-Malotau, M. (n.d.). *Grammaire allemande*.
- Kundera, M. (1993). *Les testaments trahis: Essai*. Gallimard.
- Segura, J. (2012). *Mémoires partagées d'alignements sous-phrastiques bilingues*. Université Montpellier II - Sciences et Techniques du Languedoc.

Logiciels et programmes utilisés

- ApSIC Xbench. Version 2.9.0 Build (474).
- DeepL Pro. (en ligne). <https://www.deepl.com/translator>. Traduction automatique générée au mois d'octobre 2022.
- YouAlign. (en ligne). <https://youalign.com>