Les traductions de ChatGPT selon la perspective de la théorie du *skopos* : une évaluation comparative entre prompts

Youssef, Tina

UNIVERSITÉ DE GENÈVE
FACULTÉ DE TRADUCTION ET D'INTERPRÉTATION

Tina Youssef

Les traductions de ChatGPT selon la perspective de la théorie du *skopos* :
une évaluation comparative entre prompts

Directrice : Pierrette Bouillon

Jurée : Marianne Starlander

Mémoire présenté à la **Faculté de Traduction et d'interprétation**
(Département de TIM)
pour l'obtention de la **Maîtrise universitaire en traduction et technologies, mention localisation et traduction automatique**

Université de Genève

Août 2024

Remerciements

Je tiens à remercier toutes les personnes sans lesquelles ce mémoire n'aurait jamais pu voir le jour.

Tout d'abord, je remercie la directrice de ce mémoire, Madame Pierrette Bouillon, pour sa patience, son expertise, ses précieux conseils, sa disponibilité et sa volonté à m'aider surtout quand je me retrouvais devant des difficultés. Ce fut inspirant d'être accompagnée par elle durant la rédaction de ce mémoire.

Je tiens aussi à remercier Madame Marianne Starlander, qui a gentiment accepté d'être ma jurée, et de lire mon mémoire en si peu de temps.

Je remercie également Madame Lucile Davier, qui a eu la volonté et la disponibilité de participer à ce mémoire, et dont les précieux conseils ont débloqué une partie importante de sa conception.

Enfin, j'aimerais remercier la famille Voirol pour leur support durant cette année pleine de hauts et de bas, mais surtout, Ivan, sans qui je n'aurais jamais eu le courage d'affronter mon manque de confiance et de reprendre le travail sur ce mémoire.

Table des matières

Remerciements 2

Table des matières 3

Liste des tableaux 7

Liste des figures 7

1 Introduction 9

1.1 Questions de recherche et objectifs 10

1.2 Méthodologie 11

1.3 Plan 13

2 Les systèmes génératifs et ChatGPT 14

2.1 Les grands modèles de langue 14

2.1.1 Les *transformers* 15

2.1.2 Le pré-entrainement d'un grand modèle de langue 17

2.1.3 L'ajustement d'un grand modèle de langue 17

2.1.4 L'évaluation 18

2.1.5 Les instructions et le « *prompt-engineering* » 19

2.2 ChatGPT 21

2.2.1 Naissance et fonctionnement 21

2.2.2 ChatGPT et traduction 22

2.3 Conclusion 23

3 Traductologie et théorie du *skopos* 24

3.1 Introduction à la traductologie 24
3.1.1 Brève histoire de la traductologie 24
3.1.2 Principes et taxonomie de la traductologie et de ses sujets d'étude 26
3.1.3 Les retombées de l'articles de Holmes 28
3.1.4 Principaux concepts en traductologie 29
3.2 La théorie du *skopos* 30
3.2.1 Origines 30
3.2.2 Principes 31
3.2.3 Règles 31
3.2.4 Le mandat 34
3.2.5 La question de l'intention et de la fonction 34
3.2.6 Évaluation 35
3.2.7 Le concept de loyauté 35
3.2.8 Les retombées 36
3.2.9 État de la littérature : la théorie du *skopos* et la traduction automatique 36
3.3 Conclusion 38
4 Méthodologie 39
4.1 Questions de recherche 39
4.2 Procédure 41
4.2.1 Choix des types de texte 42
4.2.2 Choix des prompts 42

4.2.3 Première évaluation : comparaison des modifications dans les traductions en fonction des prompts 47
4.2.4 Deuxième évaluation : évaluation MQM 47
4.2.5 Troisième évaluation : test-suites 51
4.2.6 Conclusion 53
5 Analyse 54
5.1 Comparaison des modifications dans les traductions en fonction des prompts 54
5.1.1 Analyse quantitative des changements par rapport au prompt 1 54
5.1.2 Analyse qualitative des changements par rapport au prompt 1 57
5.2 Résultats MQM 68
5.3 Test suites 71
5.4 Synthèse et discussion 75
5.5 Conclusion 77
6 Limites et perspectives 78
7 Conclusion 81
Bibliographie 85
Annexes 90
Annexe 1 – Texte journalistique en anglais 90
Annexe 2 – Texte publicitaire en anglais 91
Annexe 3 – Instructions utilisées à l'attention de l'experte en traductologie pour effectuer l'évaluation MQM 92

Annexe 4 – Tableau comparatif des traductions du texte journalistique en fonction des prompts 93

Annexe 5 – Tableau comparatif des traductions du texte publicitaire en fonction des prompts 93

Annexe 6 – Tableaux de l'évaluation MQM 93

Annexe 7 – Tableaux contenant le test suites 93

Liste des tableaux

Tableau 1. Résumé de la structure méthodologique de ce mémoire 13

Tableau 2. Liste des différents prompts utilisés pour effectuer les deux premières évaluations dans ce mémoire 44

Tableau 3. Tableau reprenant les 3 prompts utilisés pour les traductions de l'évaluation MQM 49

Tableau 4. Nombre total des insertions, suppressions, substitutions et déplacement de séquences pour les deux types de texte et pour tous les prompt par rapport au prompt 1 (« traduit ce texte en français ») 55

Tableau 5. Traduction du texte journalistique avec le prompt 1 58

Tableau 6. Traduction du texte publicitaire avec le prompt 1 64

Tableau 7. Résultats MQM 70

Tableau 8. Résultats de la traduction des phrases du test-suites avec le prompt « a » et le prompt « b » 72

Tableau 9. Exemples de phrases du test suites où les acronymes ont été traduits 74

Tableau 10. Exemples de phrases du test suites où les acronymes n'ont pas été traduits 74

Tableau 11. Résumé des résultats trouvés pour chaque question de recherche et hypothèse .. 84

Liste des figures

Figure 1. L'architecture d'un modèle de transformer (Vaswani et al., 2017) 15

Figure 2. Représentation du fonctionnement de la fonction « self-attention » avec un exemple d'une phrase (Jurafsky et Martin, 2024, p. 216 adapté d'Uszkoreit, 2017) 16

Figure 3. Schéma des étapes de pré-entrainement et d'ajustement 18

Figure 4. La carte de Holmes de la traductologie (adaptée de Munday, 2022) 26

Figure 5. Catégorisation des erreurs dans l'évaluation MQM 48

Figure 6. Répartition des résultats MQM 69

Figure 7. Formule utilisée pour calculer le score final MQM, adapté de Lommel et al, page 6 71

1 Introduction

Le 20 novembre 2022, OpenAI, l'entreprise américaine, introduit ChatGPT[1] au monde, l'outil d'intelligence artificielle accessible à tous. Très vite après cela, les médias et réseaux sociaux ont été envahis par des titres d'articles parlant de cet outil, de sa performance, et des différentes tâches qu'il peut accomplir, comme la rédaction de mails, la planification de vacances, ou la recherche de postes vacants dans un domaine particulier. Plusieurs domaines professionnels et académiques ont très vite été impactés, soulevant des problématiques d'éthiques et des questionnements sur les limites de son utilisation. Le domaine de la traduction n'a pas été épargné de ces questionnements, puisque ChatGPT peut aussi traduire des textes dans des combinaisons linguistiques multiples et à une vitesse surhumaine. Cependant, comme tout progrès technologique qui touche le domaine de la traduction, la question de la qualité des traductions produites par les nouveaux systèmes, que ce soit l'IA générative ou les systèmes de traduction neuronaux, constitue toujours un point d'intérêt pour les traducteurs[2] professionnels et les chercheurs dans le domaine.

Cette diversité des tâches linguistiques, dont la traduction, que ChatGPT peut accomplir, et l'accessibilité de cet outil au grand public de domaines très variés, avec des textes eux aussi très variés, nous a donc laissé sceptique quant à la versatilité des traductions que ChatGPT peut produire, et de l'effet du contexte sur ces traductions. Cela nous a entraîné à nous demander si ChatGPT pourrait adapter ses traductions selon leur fonction, comme le prescrirait par exemple la théorie fonctionnaliste du *skopos*. En effet, la règle première de cette théorie de traductologie met en valeur l'importance de la fonction du texte, et prescrit que la traduction doit s'accomplir selon un mandat de traduction particulier qui précise un contexte et une audience cible.

[1] https://chatgpt.com/
[2] Ce mémoire de master a été rédigé au masculin générique en entier.

L'application de théories de traductologie pour évaluer la qualité de traductions humaines a toujours été une pratique fréquente dans le domaine, et après l'essor de la traduction automatique, les recherches se fondant sur des théories comme celles de la théorie du *skopos* afin d'évaluer la qualité des systèmes de traductions automatiques ont commencé à devenir de plus en plus fréquentes (Yamada, 2023).

Nous trouvons ce fondement de recherche intéressant, car nous pensons que les théories de traductologie qui ont fondé le domaine de la traduction, et sur lesquelles les chercheurs se basent pour comprendre l'activité traductive, constituent une base théorique intéressante et un cadre bien-fondé pour juger la qualité de la traduction des systèmes de traduction automatique. Nous pensons aussi que cela est un cadre intéressant afin d'étudier plus en détails la production des prompts, ou instructions, donnés à ChatGPT pour qu'il produise des traductions, puisque nous pensons que ces prompts ont une influence importante sur la qualité du produit final. De même, l'interdisciplinarité nous a toujours intéressé, et nous pensons que le mariage entre un cadre théorique de traduction et des pratiques de développement technologique pourrait peut-être aboutir à des résultats intéressants.

1.1 Questions de recherche et objectifs

En prenant en compte ces intérêts, l'objectif principal de notre mémoire est de savoir si ChatGPT a le potentiel de s'adapter à un mandat de traduction particulier, comme le prescrirait la première règle de la théorie du *skopos* (cf. partie 3.2), qui dit que la traduction d'un texte source donné doit être effectuée selon la fonction voulue du texte cible. Nous nous demandons aussi quelle serait la qualité de telles traductions, dans quelles proportions les traductions produites par ChatGPT sont adaptées aux objectifs précisés dans un prompt.

Par conséquent, nous nous intéressons dans ce mémoire à répondre aux trois questions de recherche principales suivantes :

1) Un mandat de traduction précis, comme le prescrit la première règle de la théorie du *skopos*, influence-t-il la traduction produite par ChatGPT ?
2) La précision d'un mandat dans le prompt conduirait-t-elle à la production d'une traduction de meilleure qualité ?
3) ChatGPT respecte-t-il toujours le mandat précisé dans les prompts ?

Dans la partie suivante, nous détaillerons la méthodologie utilisée pour essayer de répondre à ces questions.

1.2 Méthodologie

Afin de pouvoir répondre aux questions de recherche, nous avons conçu une méthodologie englobant trois méthodes d'évaluation, chacune correspondant à certains éléments dans nos questions.

En premier lieu, afin de répondre à la première question de recherche, nous avons voulu évaluer si le système génératif ChatGPT (partie 2.2) s'adapte ou non à la théorie du *skopos* (partie 3.2), mais surtout à sa première règle qui place le mandat de traduction, ou sa fonction, au premier plan. Nous pensons que cette adaptation pourrait potentiellement se manifester par un certain nombre de changements par rapport à une traduction de référence, et nous aimerions vérifier cette hypothèse. Pour ce faire, nous avons choisi deux textes, un de type journalistique et un de type publicitaire, et nous les avons traduits avec ChatGPT plusieurs fois, en utilisant à chaque fois un prompt (une instruction) différent, et en essayant d'être de plus en plus précise dans nos demandes. En effet, les deux derniers prompts sont les plus précis, car ils contiennent un mandat de traduction avec un contexte clair et spécifique à chaque type de texte. Nous pensons que cette clarté dans le prompt mènera ChatGPT à produire des traductions différentes de celles qu'il pourrait produire avec un prompt non précis. Ensuite, nous avons procédé à une comparaison des changements de chaque traduction en fonction des prompts, par rapport à une

traduction que nous avons considéré comme référence, qui a été traduite avec un prompt simple. Nous avons effectué cette comparaison par le biais d'une analyse quantitative s'inspirant de la méthode « *Translation Edit Rate* » ou *TER* (Snover et al., 2006) pour compter les changements effectués par ChatGPT (partie 5.1.1), et à une analyse qualitative plus descriptive qui montre les types de changements effectués dans chaque traduction (partie 5.1.2).

En second lieu, nous pensons que l'adaptation de ChatGPT à un mandat de traduction précis pourrait l'emmener à produire des traductions de meilleure qualité (Yamada, 2023). Afin de vérifier cette hypothèse, et afin d'évaluer la qualité des traductions données par ChatGPT, nous avons demandé à une experte d'évaluer les traductions données par ChatGPT avec un prompt simple, et aussi les traductions données avec les prompts où nous avons précisé le mandat. La méthode d'évaluation utilisée est la méthode de ***M****ultidimentional* ***Q****uality* ***M****etrics* ou « MQM », qui se base sur une catégorisation d'erreurs, et qui est une des méthodes d'évaluation humaine les plus fiables (partie 4.2.3).

Enfin, nous pensons qu'il est intéressant d'éloigner tout hasard concernant la capacité de ChatGPT de s'adapter ou pas à un mandat donné dans un prompt, et de vérifier d'une manière méthodique la probabilité qu'il s'adapte ou pas à un mandat précis. Nous pensons que plus le mandat est précis dans le prompt, plus la capacité de ChatGPT à s'adapter à ce mandat est élevée. Afin de vérifier cette hypothèse, nous avons effectué une série de tests avec des phrases contenant toutes un phénomène spécifique et facile à repérer : les acronymes. Nous avons ensuite demandé à ChatGPT de les traduire, une fois avec un prompt simple qui lui demande uniquement une traduction de chaque phrase, et une deuxième fois avec un prompt spécifiant explicitement qu'il faut qu'il traduise les acronymes dans les phrases (partie 4.2.5).

Dans le tableau 1 ci-dessous, nous résumons la méthodologie appliquée à notre mémoire, dont les détails se trouvent dans le chapitre 4, et les éléments qu'elle essaie d'évaluer.

Question de recherche	Hypothèse à vérifier	Méthode d'évaluation utilisée
Q1. Un mandat de traduction précis, comme le prescrit la première règle de la théorie du *skopos*, influence-t-il la traduction produite par ChatGPT ?	**H1.** La traduction produite avec un prompt précis est différente par rapport à une traduction produite par un prompt ne précisant pas le mandat ou ayant un mandat peu clair	Comparaison des changements dans les traductions en fonction des prompts et analyse quantitative et qualitative
Q2. La précision d'un mandat dans le prompt conduirait-elle à la production d'une traduction de meilleure qualité ?	**H2.** Si le prompt précise un mandat spécifique, la traduction est de meilleure qualité	Évaluation MQM
Q3. ChatGPT respecte-t-il toujours le mandat précisé dans les prompts ?	**H3.** Plus le prompt est précis, plus il y a une probabilité que ChatGPT produise une traduction adaptée à nos objectifs	Test-suites avec un phénomène spécifique : les acronymes

Tableau 1. Résumé de la structure méthodologique de ce mémoire

1.3 Plan

En vue de tout ce qui précède, et afin de pouvoir répondre à nos questions de recherche et arriver à nos objectifs, nous poserons d'abord le cadre théorique de notre mémoire en parlant, dans le chapitre 2, des systèmes génératifs et ChatGPT, et, dans le chapitre 3, de la traductologie et de la théorie du *skopos*. Ensuite, nous expliquerons en détails la méthodologie utilisée dans le chapitre 4, où nous expliquerons aussi en détails nos questions de recherche. Dans le chapitre 5, nous analyserons les résultats de nos trois évaluations. Dans le chapitre 6, nous parlerons des limites et perspectives de ce mémoire, avant de conclure dans le chapitre 7.

2 Les systèmes génératifs et ChatGPT

Même si la discussion autour des systèmes génératifs et surtout ChatGPT n'est devenue centrale que récemment, la recherche qui l'entoure, elle, date de quelques années. Nous ne pouvons pas nous attarder sur ce sujet sans passer par les éléments de base qui ont permis aux modèles comme celui de ChatGPT d'être développés et de devenir accessibles au public, comme les grands modèles de langue et l'architecture « *transformer* » qui les distingue. Dans la partie 2.1, nous parlerons des grands modèles de langue, en passant par l'architecture des « *transformers* », le pré-entrainement, l'évaluation, l'ajustement et le « *prompt-engineering* ». Dans la partie 2.2, nous allons détailler le modèle « ChatGPT » en parlant de sa naissance et son fonctionnement, avant de nous attarder sur sa relation avec la traduction et de conclure.

2.1 Les grands modèles de langue

Les modèles de langues (*language models*, ou *LMs*) sont des modèles « *transformer* » qui sont capables « d'assigner des probabilités pour le mot qui suivra un autre mot, ou la séquence de mots qui suivra une autre séquence de mots en général » (Jurafsky et Martin, 2024, p. 32, [notre traduction]). En d'autres termes, ce sont des modèles qui peuvent prédire quel texte suivra un texte donné. Toute tâche du traitement des langues (résumé automatique, question-réponse, analyse de sentiments, traduction automatique) pourrait être considérée comme une tâche de prédiction de mots.

Les grands modèles de langue (*large language models*, ou *LLMs*) sont le résultat du pré-entrainement d'un modèle de langue de base, qui est une étape qui permet au modèle d'acquérir des connaissances générales sur le langage, à partir d'un large volume de texte (Jurafsky et Martin, 2024). Les données de pré-entrainement peuvent parfois correspondre à plus de 3 milliards de sites web, comme le corpus « *CommonCrawl* », qui est un des grands corpus de données utilisé pour entrainer ces modèles de langue, comme nous le verrons dans la suite.

2.1.1 Les *transformers*

Un « *transformer* » est un modèle d'architecture de réseau de neurones. Le premier modèle d'architecture de ce type a été conçu par Vaswani et al. en 2017, et est considéré comme la référence dans le domaine. Ce modèle se base exclusivement sur un système d'attention, afin de connecter ces deux éléments principaux : l'encodeur et le décodeur (Vaswani et al., 2017).

L'architecture d'un *transformer* est représentée dans la figure 1 ci-dessous.

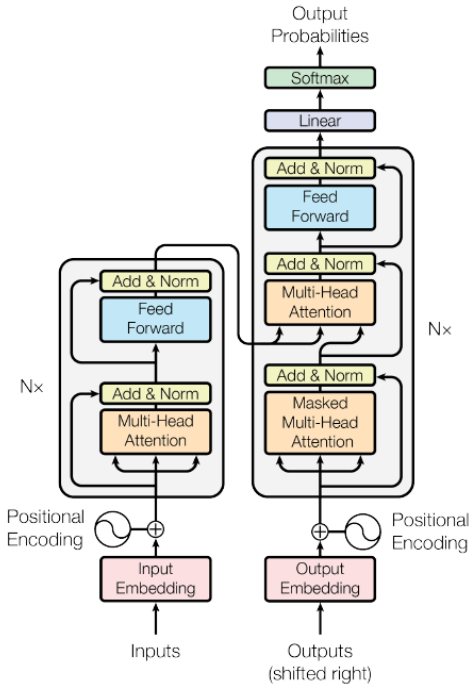

Figure 1. L'architecture d'un modèle de *transformer* (Vaswani et al., 2017)

Tous les *transformers* ne sont pas de type *encoder-decoder*. En effet, il existe des *transformers* qui ne sont pas formés de ses deux composantes, mais uniquement d'encodeurs ou de décodeurs. Le modèle BERT (***B**idirectional **E**ncoder **R**epresentations from **T**ransformers*) est un exemple de *transformer* formé uniquement d'un encodeur, ce qui lui permet de se concentrer sur le contexte bidirectionnel dans un texte, et de le représenter sous forme numérique. Celui-ci permet de comprendre le sens global de la phrase (Devlin et al., 2019). Ce modèle est utilisé pour des tâches comme la classification de mots ou la prédiction de sentiments par exemple.

Les systèmes génératifs, comme ChatGPT notamment, l'outil utilisé dans ce mémoire, sont des systèmes qui n'utilisent que le décodeur, ce qui permet de générer des textes étant donné une prémisse (OpenAI, 2023).

La fonction d'attention, qui est un élément important dans le fonctionnement des *transformers*, est d'aider le modèle de langue à créer des liens entre les mots afin de prendre en compte le contexte et produire de meilleurs résultats.

La fonction de « *self-attention* » permet de créer des liens entre les mots sources afin que le contexte soit clair pour le réseau de neurones. Un exemple de ce fonctionnement est expliqué dans la figure 2, qui contient une représentation d'une phrase sur deux couches (ou *layers*). Les différentes nuances de bleu représentent des valeurs de la fonction « *self-attention* » : plus le bleu est foncé, plus la valeur est élevée. Cela indique que le modèle de langue a correctement représenté le lien entre le mot *« animal »* et le mot *« it »*, qui sont dans ce cas-là interchangeables (Jurafsky et Martin, 2024).

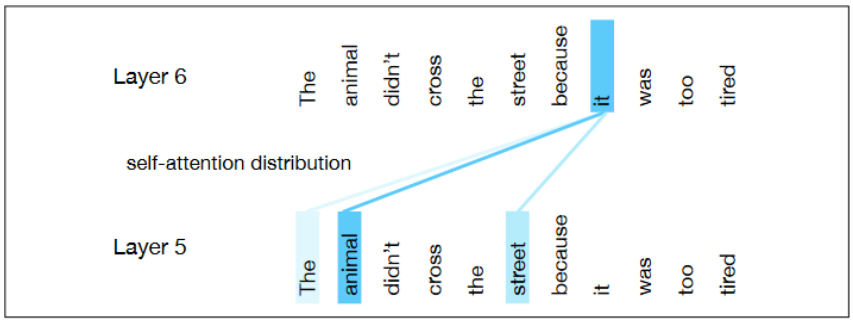

Figure 2. Représentation du fonctionnement de la fonction « *self-attention* » avec un exemple d'une phrase (Jurafsky et Martin, 2024, p. 216 adapté d'Uszkoreit, 2017)

La fonction de « *cross-attention* », utilisée dans les modèles *encoder-decoder*, a un rôle similaire à celui de la fonction de « *self-attention* », mais la différence principale est que la « *self-attention* » relie des mots différents dans une même séquence, alors que la « *cross-attention* » crée des liens entre les mots de deux séquences différentes (Gheini et al., 2021).

Cela est crucial dans les tâches comme la traduction automatique, car cette fonction permet de lier une séquence source à une séquence cible.

2.1.2 Le pré-entrainement d'un grand modèle de langue

L'étape de pré-entrainement est l'étape initiale où le modèle de langue est entraîné, d'une manière non-supervisée, sur un grand corpus de texte afin qu'il apprenne des représentations générales de langue. Durant cette étape, le modèle apprend à prédire des mots masqués dans une phrase, et les prochaines phrases dans un paragraphe (Devlin et al., 2019). L'étape de pré-entrainement consiste ainsi à entrainer le modèle avec un grand nombre de données textuelles diversifiées, afin que le modèle puisse apprendre les logiques et représentations linguistiques (Brown et al., 2020). L'étape de pré-entrainement précède l'étape de l'entraînement supervisé, ou l'ajustement, que nous détaillerons dans la partie 2.1.3.

Les modèles BERT (***B**idirectional **E**ncoder **R**epresentations from **T**ransformers*), par exemple, déjà mentionnés, ont permis de démontrer l'importance de l'étape de pré-entrainement, Celui-ci a été pré-entrainé sur l'entièreté de Wikipédia, et le corpus « *BookCorpus* », et a pu donner des résultats qui ont dépassé ceux des modèles de l'état de l'art (Devlin et al., 2019).

Le pré-entrainement d'un modèle peut être effectué avec des données diversifiées, ne contenant pas uniquement du texte, ce qui permet au modèle d'effectuer différentes tâches en traitement informatique multilingue. Un exemple de corpus communément utilisé pour cela est le corpus « *CommonCrawl* ».

2.1.3 L'ajustement d'un grand modèle de langue

Dans la partie 2.1.2, nous avons expliqué le concept de pré-entrainement. Dans cette partie nous allons détailler l'ajustement, qui est la phase supervisée de l'entraînement des grands modèles de langues pré-entrainés auparavant, et qui nécessite un grand volume de données annotées.

Cette étape vise une meilleure optimisation du réseau de neurone afin de lui permettre d'effectuer une tâche spécifique. Cela est possible grâce à l'apprentissage par transfert, où les représentations du langage apprises par le modèle durant l'étape du pré-entrainement sont transférées à des tâches spécifiques, comme la classification de texte, la réponse à des questions, ou la traduction (Jurafsky et Martin, 2024).

À l'inverse des grands modèles de langue qui résultent du pré-entrainement, qui sont des modèles de fondation généraux, les modèles résultant de l'ajustement sont des modèles affinés, utilisant en général de petits corpus annotés. Dans la figure 3, nous avons adapté une représentation simplifiée des différentes étapes de pré-entrainement et d'ajustement.

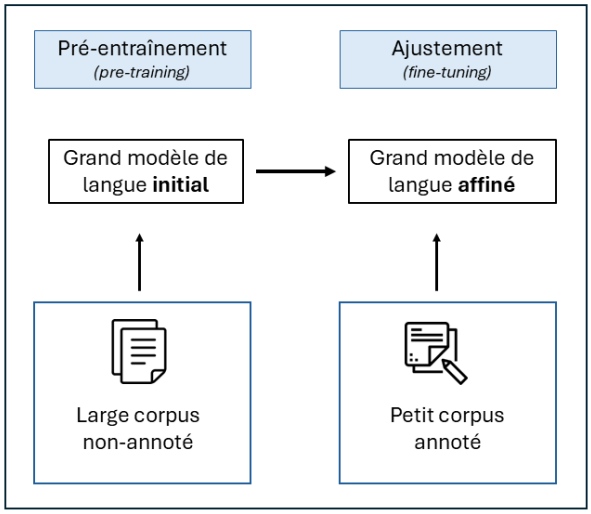

Figure 3. Schéma des étapes de pré-entrainement et d'ajustement[3]

2.1.4 L'évaluation

Plusieurs métriques d'évaluation ont été conçues afin d'évaluer les performances des grands modèles de langues. Nous citons les plus importantes, qui sont la « *GLUE benchmark* » (Wang et al., 2018), qui inclue neuf tâches de compréhension de phrases en anglais, comme la classification de phrases, l'analyse de sentiments, la références à des pronoms dans une phrases,

[3] https://parserdigital.com/2023/09/21/unleashing-the-power-of-generative-ai-understanding-transformers/

entre autres. Les résultats de cette évaluation de modèles de langues pré-entrainés montrent une meilleure performance qu'avec les modèles qui n'ont pas subi de pré-entrainement dans l'évaluation effectuée globalement sur les neuf tâches (Wang et al., 2018). Cependant, les résultats montrent aussi des performances non prometteuses pour les modèles existants, comme l'évaluation grammaticale de la phrase, la coréférence, et la compréhension des liens entre deux phrases (Wang et al., 2018). Ces limites ont montré que même si certaines tâches linguistiques pourraient être effectuées par ces modèles sans commettre des erreurs graves, les liens que les modèles créent entre les mots sont superficiels et manque d'une compréhension plus profonde de la langue.

Afin d'arriver à de meilleurs résultats que ceux trouvés avec GLUE, et pour combler les lacunes démontrées par ses résultats, une nouvelle métrique d'évaluation a été conçue par la même équipe pour améliorer celle qui l'a précédée, appelé SuperGLUE (Wang et al., 20219). Cette nouvelle méthode s'inspire de la première, mais y ajoute de nouvelles tâches linguistiques afin d'évaluer d'une manière plus détaillée les grands modèles de langues. Les résultats montrent que les modèles sont encore derrière les humains en termes de performance d'à peu près 20 points (Wang et al., 2019).

Afin d'obtenir les résultats d'évaluation les plus fiables, il est recommandé d'utiliser à la fois les métriques automatiques susmentionnées, et l'évaluation humaine. Même si les métriques automatiques sont utiles pour les comparaisons standardisées, l'évaluation humaine est cruciale pour évaluer les modèles de langue sur les aspects nuancés de la génération du langage qui dépendent du contexte (Chen et al., 2021).

2.1.5 Les instructions et le « *prompt-engineering* »

Nous appelons « *prompt* » le texte que nous rédigeons dans l'environnement conversationnel des modèles de langues génératifs afin d'obtenir la réponse que nous cherchions. Le prompt est

donc les instructions que nous donnons au modèle pour lui demander de nous donner un résultat, que ce soit une traduction, un texte, une équation, des lignes de code ou des images. C'est avec ces instructions que nous avons pu effectuer nos méthodes d'évaluation dans notre méthodologie de recherche, détaillée dans le chapitre 4.

Le concept de « *prompt-engineering* » est le processus durant lequel l'utilisateur d'un modèle de langue comme ChatGPT affine le texte de son prompt afin d'améliorer les capacités du modèle de générer des réponses pertinentes et précises (Brown et al., 2020). Hormis pour donner des instructions, les prompts peuvent aussi être utiliser pour évaluer les réponses de ces modèles, qui est la méthode que nous avons utilisé dans ce mémoire.

Le processus de « *prompt-engineering* » est aussi utilisé lorsque la tâche que nous souhaitons accomplir avec le modèle de langue n'est pas une tâche pour laquelle il est entraîné. En interrogeant le modèle de langue plusieurs fois, il est possible que les résultats donnés s'améliorent en fonction du nombre et de la forme des prompts donnés au modèle de langue.

Comme nous le montrent Bawden et Yvon (2023) dans leur article s'intitulant « *Investigating the Translation Performance of a Large Multilingual Language Model: the Case of BLOOM* » (évaluation de la performance en traduction d'un grand modèle de langue multilingue : le cas de BLOOM), dont l'objectif était d'évaluer les capacités de traduction automatique du grand modèle de langue « *BLOOM* » en utilisant plusieurs corpus et plusieurs combinaisons linguistiques, les traductions avec des prompts qui donnent des exemples de traduction (par la méthode « *few-shot learning* ») sont de meilleure qualité que les traductions effectuées sans que le modèles ait d'exemples auparavant (« *zero-shot learning* »). Cet article est intéressant non seulement car il évalue les performances en traduction d'un grand modèle de langue, mais aussi car il prouve l'importance de l'entrainement vis-à-vis de la qualité de la traduction.

2.2 ChatGPT

ChatGPT, qui est le sujet principal de notre mémoire, est un *chatbot* utilisant un modèle génératif d'intelligence artificielle, développé par l'entreprise OpenAI aux États-Unis, et devenu accessible au public en 2022. La version gratuite de ce *chatbot*, que nous avons utilisée dans ce mémoire, est disponible et utilisable via le site de l'entreprise. À compter de juin 2023, OpenAI estimait que le logiciel compte plus de 100 millions d'utilisateurs, ce qui le rend un des logiciels dont la croissance est la plus rapide de l'histoire.

2.2.1 Naissance et fonctionnement

ChatGPT est basé sur l'évolution de la technologie GPT-1, ou « *Generative Pre-Training* », développée par une équipe d'OpenAI en juin 2018 (Radford et al., 2018). Le but de ce modèle était de démontrer l'importance du pré-entrainement sur un grand volume de données, suivi de l'ajustement pour entrainer le modèle à effectuer des tâches spécifiques. Le corpus utilisé pour le pré-entrainement de ce modèle est intitulé « *BooksCorpus* », et contient plus de 7000 livres uniques non-publiés, d'une variété de genre comme l'aventure, la fantaisie et la romance (Radford et al., 2018).

En février 2019, OpenAI a développé la deuxième version, GPT-2, qui a été pré-entrainée sur un corpus nommé « *WebText* », composé de données sur Internet. Ce modèle était une amélioration de son prédécesseur, et a pu donner des résultats textuels plus cohérents, en respectant des nuances et un contexte spécifique. Ce modèle a surtout pu montrer de bons résultats avec un pré-entrainement non-supervisé, ce qui l'a rendu plus intéressant que la version précédente, puisque l'entrainement supervisé est coûteux et prend du temps. Cependant, le large volume de données non-annotées a suscité des craintes quant à la mauvaise utilisation du modèle pour des fins non-éthiques (Radford et al. 2019).

En juin 2020, la troisième version du modèle, GPT-3, est devenue accessible, en présentant une amélioration claire de performance, et ayant la capacité d'effectuer des tâches linguistiques avec un niveau minimal d'ajustement. Ces tâches incluent la conversation, la génération de textes de haute qualité, et la traduction de textes, avec des résultats bien plus cohérents et fluides que ceux de GPT-2.

Ensuite, en novembre 2022, le *chatbot* ChatGPT tel que nous le connaissons aujourd'hui devient disponible, et se base sur l'architecture du modèle GPT-3.5, un modèle spécialisé de GPT-3, ajusté par la méthode de l'apprentissage par renforcement, qui consiste à donner une instruction au modèle, pour ensuite donner une réponse (ou réaction) par un évaluateur humain qui contient des informations qui corrigent celles données par le modèle (Ye et al. 2023). Cette étape est importante afin de permettre au public d'utiliser le modèle et d'éviter les utilisations non-éthiques qui pourraient en incomber. C'est cette étape aussi qui a permis au modèle d'effectuer des tâches comme la conversation et la compréhension de contexte, et de donner des réponses pertinentes aux questions posées dans l'environnement de conversation.

2.2.2 ChatGPT et traduction

ChatGPT est capable de générer des textes dans plusieurs langues, comme l'anglais, l'espagnol, le français, l'allemand, le chinois, le japonais, et le coréen. Les données utilisées pour le pré-entrainement sont en plusieurs langues, ce qui lui permet d'effectuer des tâches de traduction (OpenAI, 2023).

Différentes études ont montré que la qualité des traductions varie en fonction de la combinaison linguistique et la complexité des textes. Les traductions peuvent être de moins bonne qualité quand les langues sont peu dotées ou les textes sources utilisent du jargon spécialisé (OpenAI, 2023).

Un article intéressant dans le cadre de notre mémoire est celui s'intitulant « *Optimizing Machine Translation through Prompt Engineering: An Investigation into ChatGPT's Customizability* » (Yamada, 2023) (l'optimisation de la traduction automatique par le biais du « *prompt-engineering* » : une investigation de la customisation de ChatGPT). L'objectif de cette recherche est très semblable au nôtre, car elle cherchait à savoir si le fait d'ajouter des objectifs de traductions clairs et une audience cible dans les prompts permettrait à ChatGPT de produire de meilleures traductions. Cette recherche a utilisé une méthodologie différente à la nôtre : deux prompts ont été utilisés, un spécifiant la fonction de la traduction et l'audience cible, et l'autre demandant à ChatGPT de fournir trois traductions possibles. L'étude a comparé les traductions produites pas ChatGPT à celles produites par Google Translate, DeepL et une traduction simple de ChatGPT. Les traductions ont été analysées par un expert et évaluées selon la similarité cosinus (la distance entre des vecteurs), et les résultats ont montré que les prompts plus précis donnent des traductions de meilleure qualité, et se rapprochent des pratiques professionnelles en traduction (Yamada, 2023).

2.3 Conclusion

Dans ce chapitre, nous avons brièvement exposé le fonctionnement des grands modèles de langue, en expliquant les éléments qui les composent et les étapes de leur développement. Nous avons aussi brièvement mentionné les différentes métriques d'évaluation des grands modèles de langue, et l'importance des prompts dans l'utilisation de ces modèles.

Nous avons de même détaillé le fonctionnement du système génératif ChatGPT qui se base sur l'architecture GPT-3.5 de l'entreprise OpenAI. Les avancements dans l'architecture des modèles génératifs pré-entrainés de 1 à 3.5 montrent un potentiel positif pour l'amélioration de la performance des modèles tels que ChatGPT en général, afin qu'ils puissent effectuer des tâches linguistiques comme la traduction.

3 Traductologie et théorie du *skopos*

Dans ce chapitre, nous avons choisi de présenter une vue d'ensemble de la traductologie en tant que matière d'étude en sciences humaines, et sa place dans le domaine de la traduction, puisque la traductologie représente une partie importante de la finalité de ce mémoire de master. La théorie du *skopos* étant la théorie principale de ce mémoire, nous la présenterons en plus de détails dans la partie 3.2, et nous nous contenterons de parler brièvement de la traductologie dans la partie 3.1.

Ce choix a été fait car nous pensons que le sujet de la traductologie est un sujet d'importance dans le domaine de la traduction en général, et pourrait avoir une importance plus vaste dans le domaine de la traduction automatique en particulier, comme nous essayons de l'étudier dans le cadre de ce mémoire. De plus, puisque le sujet du mémoire porte surtout sur une évaluation de traduction automatique faite par le modèle ChatGPT, nous pensons qu'il est nécessaire d'introduire le domaine de la traductologie afin que la méthodologie que nous avons effectuée dans le chapitre 4 soit claire.

3.1 Introduction à la traductologie

Dans cette partie, nous allons présenter une brève histoire de la traductologie, ainsi que des moments clés dans sa fondation en tant que sujet d'étude. Ensuite, nous aborderons une taxonomie des thèmes abordés en traductologie, de leur évolution, et, enfin, des principaux concepts de traductologie.

3.1.1 Brève histoire de la traductologie

La traductologie en tant que matière en sciences humaines n'est définie qu'en 1972, après que James S. Holmes, poète et traducteur américano-néerlandais, discute dans son article s'intitulant « *The name and nature of translation studies* » durant le troisième congrès international de linguistique appliquée à Copenhague, article qui sera publié ultérieurement en 1988. Il y définit

la traductologie comme étant l'étude des problèmes complexes qui entourent l'activité traductive et la traduction (Holmes, 1972). En parlant de cette discipline, il précise qu'ils existent quatre facteurs principaux qui ont conduit à la croissance de l'importance de la traductologie en tant que discipline d'étude : (1) l'existence de plus en plus de programmes d'études au niveau universitaire en traduction dans le monde, (2) l'organisation de plus en plus de conférences tournant autour de la traduction, ainsi que la publication des livres et des revues sur le sujet, (3) la production de plus en plus de publications, ce qui veut dire un besoin croissant « d'outils généraux et analytiques », comme les anthologies, les bases de données, les encyclopédies et les manuels, et (4) le succès croissant que connaissent les organisations internationales. (Holmes, 1972, p. 11-12).

Même si la traductologie n'a pas vraiment été définie comme discipline avant le 20ème siècle, les pratiques de traduction qu'elle traite, quant à elles, remontent jusqu'au Ier siècle J.C. avec Cicéron et Horace, où les questionnements sur les stratégies de la traduction ont commencé avec la traduction des textes religieux, notamment la traduction de la Bible. Ce domaine de la traduction a relevé plusieurs questions, qui ont notamment inspiré la dichotomie de la traduction « mot pour mot », pour préserver la sainteté des mots de la Bible, et la traduction « sens pour sens », qui est utilisée dans les textes plus modernes. Beaucoup plus tard, au XVIIIème siècle, la traduction était utilisée comme outil dans le domaine de l'apprentissage de langues, suivie de l'essor de la traduction littéraire aux États-Unis dans les années soixante, où une approche de littérature comparée, où la « littérature était étudiée et comparée à un niveau transnational et transculturel, ce qui demandait la lecture de quelques travaux en traduction » (Munday, 2022, p. 13). Dans les années trente et jusqu'aux années soixante, la linguistique contrastive, ou l'étude de deux langues en contraste dans le but d'identifier les différences générales et spécifiques entre elles (Munday, 2022), utilisait des traductions d'une manière assez large afin

d'obtenir des données pour ce genre d'études. En 1960, l'étude de la traduction, ou la traductologie, devient de plus en plus systématique après des publications principales dans le domaine.

3.1.2 Principes et taxonomie de la traductologie et de ses sujets d'étude

Un des éléments clés dans l'article de Holmes est la « carte » de traductologie, qui organise les thèmes et sujet de la traductologie, afin de mieux positionner les études qui ont été faites jusqu'à cette date-là (1972), et pour aider à positionner les études qui seront faites dans l'avenir. Cette taxonomie est détaillée dans la figure 4 ci-dessous :

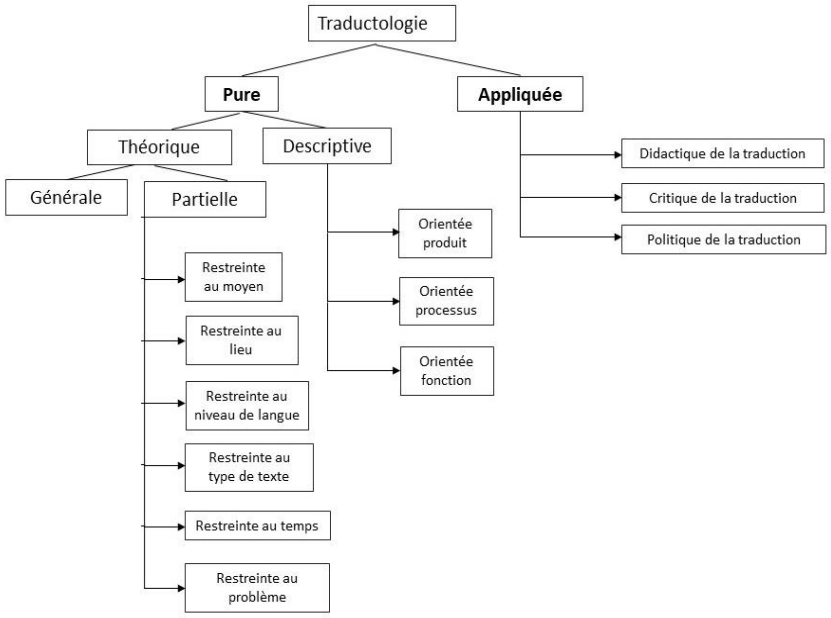

Figure 4. La carte de Holmes de la traductologie (adaptée de Munday, 2022)

Selon Holmes, « la traductologie a […] deux objectifs principaux : (1) décrire le phénomène de l'activité traductive et de la traduction, et (2) établir des principes généraux par lesquelles ces phénomènes peuvent être expliqués et prédits. » (Holmes, 1972, p. 5).

De ce fait, Holmes divise le domaine de la traductologie en deux parties : la traductologie pure, et la traductologie appliquée. La traductologie pure comprend la traductologie théorique et la

traductologie descriptive, qui, selon Holmes, s'alignent le mieux avec les objectifs susmentionnés. Il explique, que, d'un côté, les approches descriptives de la traductologie sont les plus importantes, puisque c'est la catégorie d'études qui reste le plus rapprochée du sujet d'étude. Les approches descriptives de la traductologie comprennent trois sous-catégories : la traductologie descriptive **orientée sur le produit**, qui décrit les traductions déjà existantes, la traductologie descriptive **orientée sur le processus**, qui décrit l'activité de traduction en soi, et la traductologie descriptive **orientée sur la fonction**, qui décrit la fonction de la traduction dans la culture cible, et qui comprend la théorie du *skopos*. Cette théorie est le point sur lequel nous allons le plus nous concentrer dans ce travail de mémoire de master, et qui aura une vue d'ensemble plus détaillée dans la partie 3.2.

D'un autre côté, la catégorie de traductologie théorique « utilise les résultats des approches descriptives de la traductologie afin de faire évoluer des théories et principes en traduction » (Holmes, 1972, p. 6-7), dans le but de développer un ensemble de théories de traduction « complet et inclusif ». La catégorie de traductologie théorique comprend six sous-catégories, que Holmes détaille comme suit :

- La traductologie **restreinte au moyen** : l'étude du moyen utilisé pour traduire, qui pourrait être une traduction humaine, une traduction automatique, et/ou une traduction assistée par ordinateur.
- La traductologie **restreinte au lieu** : elle est restreinte au langage ou à la ou les culture(s) de la paire de langue.
- La traductologie **restreinte au niveau de langue** : elle se concentre sur le niveau linguistique, et peut être exécutée au niveau du mot, de la phrase ou du texte.
- La traductologie **restreinte au type de texte** : l'étude de la traduction de genres spécifiques de textes.

- La traductologie **restreinte au temps** : l'étude de traductions de textes contemporains ou de vieux textes.
- La traductologie **restreinte au problème** : l'étude d'un ou plusieurs problèmes spécifiques dans les théories générales de traduction.

L'autre grande catégorie de traductologie selon Holmes est la traductologie appliquée, qui comprend trois sous-catégories **: la didactique de la traduction, la critique de la traduction, et la politique de la traduction.**

Le domaine de **la didactique de la traduction**, selon Holmes, comprend deux catégories, qui sont l'enseignement de langues étrangères où la traduction joue un rôle important, et l'enseignement de la traduction en tant que domaine en soi au sein de programmes d'études en traduction. Holmes pense que **la critique de la traduction** « reste un domaine peu développé, et n'est pas influencée par les développements dans le domaine de la traduction » (Holmes, 1972). À noter que cet article a été publié en 1972, et depuis, beaucoup de développements ont touché les domaines de la traductologie et de la critique de la traduction.

Enfin, **la politique de la traduction** comprend les sujets de recherche qui relèvent des questions sur le rôle du traducteur ou de la traductrice au sein d'une société, quels textes devraient être traduits, ainsi que le rôle de la traduction dans l'enseignement des langues étrangères.

3.1.3 Les retombées de l'articles de Holmes

L'article publié par James. S. Holmes en 1972 durant le troisième congrès international de linguistique appliquée à Copenhague est considéré comme un élément clé de la naissance et du développement du domaine de la traductologie. Il a eu des retombées assez importantes, notamment dans la création de théories qui questionnent le concept d'équivalence en Allemagne, et l'essor de théories centrées sur les genres de textes (comme la théorie du *skopos*). Suivirent aussi de plus en plus d'études sur l'histoire et la sociologie de la traduction.

L'évolution technologique et le développement de la traduction automatique, ainsi que la traduction audiovisuelle, la localisation, et les études de corpus, ont aussi eu un impact important sur le développement de la recherche en traductologie, et ont engendré de nouvelles questions de recherche dans le domaine.

Des recherches récentes ont aussi pu démontrer l'interdisciplinarité de la traductologie, et la relation de la traduction notamment avec la linguistique, les langues modernes et les études de langues, la littérature comparée, la philosophie, la sociologie, et l'histoire (Munday, 2022). En plus, les domaines de traductions spécialisées ont des relations étroites avec les domaines du droit, la politique, la médecine, la finance, la science, ainsi qu'avec l'informatique lorsqu'il s'agit de la traduction automatique et la traduction assistée par ordinateur.

3.1.4 Principaux concepts en traductologie

Comme mentionné dans la partie 3.1.1 de ce chapitre, l'un des débats les plus vieux dans le domaine de la traductologie est celui de la traduction « mot pour mot » et « sens pour sens ». En effet, cette dichotomie était la base de la traductologie depuis 2 000 ans, jusqu'au XX^{ème} siècle (Munday, 2022). La traduction « mot pour mot » était mal vue et critiquée, et Saint-Jérôme lui-même n'utilisait cette stratégie de traduction que pour les textes religieux.

> "Now I not only admit but freely announce that in translating from the Greek – except of course in the case of the Holy Scripture, where even the syntax contains a mystery – I render not word-for-word, but sense-for-sense." (Saint-Jérôme 395 AC/1997 :25 – cité dans Munday, 2022, p. 27)

À noter que le débat de la traduction « mot pour mot » et « sens pour sens » n'est pas limité à la traduction de la Bible, mais s'étend aussi à d'autres religions, et donc d'autres langues. En effet, nous pouvons constater son envergure dans les textes de religion bouddhiste, en passant par la culture islamique avec l'arabe, jusqu'à la réforme protestante. Ce débat a aussi été reflété dans la traduction par transfert et par interlangue dans le domaine de la traduction automatique.

Le concept de fidélité, qui est un concept qui revient dans beaucoup de théorie de traductologie, a été beaucoup discuté, et on le définissait comme étant l'équivalent de la traduction « mot pour mot ». Ce n'est qu'à partir du XVII^{ème} siècle que la fidélité en traduction faisait référence à la loyauté au sens, et pas aux mots.

3.2 La théorie du *skopos*

La théorie du *skopos* est la théorie qui a guidé l'évaluation effectuée dans ce mémoire. Il est donc nécessaire de la détailler dans les parties suivantes. Nous allons commencer par expliquer les origines et principes de la théorie, les règles, l'importance du mandat, la question de l'intention et de la fonction dans le contexte de la théorie du *skopos*, l'évaluation, le concept de loyauté, les retombées de la théorie et enfin, l'état de la littérature qui lie la théorie du *skopos* avec la traduction automatique et ChatGPT.

3.2.1 Origines

La théorie du *skopos* est une théorie du courant fonctionnaliste en traductologie – un courant qui met en relief l'importance de la fonction d'un texte pour mener à bien une activité traductive. C'est aussi un courant qui met en exergue l'importance du public, texte ou culture cible, à l'inverse d'autres théories de traduction, comme la théorie de l'équivalence, qui garderaient la formalité et l'importance du texte source dans la traduction.

À l'origine de la théorie du *skopos* – qui veut dire « but » ou « objectif » en grec – sont les deux linguistes allemands Katharina Reiß et Hans Vermeer. Ils se sont associés en 1984 pour publier la théorie, répandue en premier lieu en Allemagne, sous le titre : *Towards a General Theory of Translational Action: Skopos Theory Explained* (les fondements d'une théorie générale de la traduction et de l'interprétation). Pour Vermeer, la traduction ne veut pas dire simplement offrir une équivalence d'un texte source vers un texte cible, mais ce serait plutôt de produire un texte dans un contexte cible, pour un but cible, des destinataires cibles, dans des circonstances cibles :

> "To translate means to produce a text in a target setting for a target purpose and target addressees in target circumstances." (Vermeer, 1987, p. 29)

La théorie du *skopos* est basée sur la théorie fondée par Von Wright en 1968 s'appelant la théorie de l'action, qui voulait que la traduction soit un acte de communication. Le sens dépendrait du contexte, et est construit dans le contexte.

3.2.2 Principes

Pour traduire un texte en respectant la théorie du *skopos*, et selon Reiß et Vermeer, il faudrait se poser les quatre questions principales suivantes :

- Pourquoi traduisons-nous le texte source ?
- Quelle est la fonction du texte cible ?
- Qui sont les destinataires du texte cible ?
- Quels sont les besoins de ces destinataires ?

Dans un contexte de mandat de traduction professionnel, ces questionnements devraient trouver leurs réponses dans les informations figurant dans le mandat de traduction lui-même, donné par le client au traducteur.

Selon Reiß et Vermeer, la traduction dans le cadre de la théorie du *skopos* serait un transfert linguistique et culturel ; ces deux aspects ne pouvant être séparés l'un de l'autre.

3.2.3 Règles

Afin de mener à bien une traduction selon la théorie du *skopos*, Nord (2013) et Munday (2016), nous invitent à respecter cinq règles principales, qui nous allons détailler ci-dessous.

La première règle, qui est la plus importante et à qui la théorie du *skopos* doit son nom, est le fait que l'action traductive est déterminée par don *skopos*, donc son objectif. C'est sur cette

règle que nous avons principalement basée notre méthodologie de recherche et notre question de recherche principale (chapitre 4).

La deuxième règle prévoit que la traduction n'est qu'une offre d'information dans une culture source et donne lieu à une offre d'information dans une culture cible, et cela permet au traducteur de prendre de la distance par rapport à la question du sens. Le but de la traduction ne serait pas de transmettre le sens unique d'un texte source dans un texte cible, mais de transmettre une offre d'information qui est disponible dans le texte source et d'en faire une offre d'information dans le texte cible. Une conséquence directe de cette règle de la théorie du *skopos* serait d'avoir une variété d'interprétation de sens et donc de traductions possibles – donc, autant de traductions que d'objectifs, de fonctions du texte source.

La troisième règle de la théorie du *skopos* est que la fonction du texte cible pourrait être différente de celle du texte source. Cette règle vient s'opposer à la théorie de l'équivalence, où la question de l'effet équivalent du texte source et du texte cible est posée. Cette règle impliquerait que la rétro-traduction ne serait plus possible.

La quatrième règle de la théorie du *skopos* met en exergue l'importance de la cohérence intratextuelle, qui signifie une cohérence à l'intérieur de la culture cible, où les destinataires du texte cible pourraient faire sens de ce qui a été traduit.

La cinquième règle de la théorie du *skopos* vient compléter la quatrième, en ajoutant l'importance de la cohérence extratextuelle, qui à son tour, signifie la cohérence entre la traduction et le texte source, et entre la traduction et le mandat, ou le *skopos*. Cela veut dire que le texte cible pourrait refléter le texte source, mais cela pourrait aussi ne pas être le cas, si le mandat de traduction (ou le *skopos*) ne le demande pas.

La théorie du *skopos* préconise une hiérarchie des règles mentionnées ci-dessus, qui est elle-même considérée comme une sixième règle à part entière de la théorie du *skopos* par certains (cf. Du, 2012). Dans cette hiérarchie des règles, Vermeer et Reiß ont établi une importance non égale aux règles de la théorie du *skopos*, qui met en premier la règle de la fonction du texte (ou la règle du *skopos*) qui dicte l'action traductive. C'est d'ailleurs selon cette première règle que les évaluations de traductions pourraient être accomplies. En deuxième lieu, Reiß et Vermeer placent la règle de cohérence intratextuelle, qui implique la cohérence de la traduction avec la culture cible, et la capacité des destinataires de la traduction à faire ses de cette traduction. En troisième et dernier lieu, se place la règle de fidélité, ou la cohérence extratextuelle. Cette règle est la dernière dans cette hiérarchie car, comme mentionné auparavant, il se pourrait que le *skopos*, ou mandat de la traduction voudrait que la fidélité, ou la cohérence extratextuelle, ne soit pas respectée.

Cette hiérarchie des règles émises par Reiß et Vermeer entraînerait plusieurs conséquences sur l'action traductive et le contexte de la traduction.

Selon Schaffner (2009) et Munday (2016), elle conduirait à un détrônement du texte source, qui prend parfois la place la plus importante dans le cadre d'autres stratégies de traduction. Dans une même lignée, la hiérarchie des règles placerait le traducteur dans un rôle d'expert interculturel, car c'est à lui d'évaluer le mandat et de réaliser la fonction de l'action traductive. La dernière conséquence de cette hiérarchie serait, comme mentionné ci-dessus, l'existence de plusieurs possibilités de traductions pour un même texte source, car il est considéré comme une offre d'information qui pourrait prendre la forme d'une autre offre d'information dans le texte cible.

3.2.4 Le mandat

Reiß et Vermeer mettent le mandat de traduction au centre de la théorie du *skopos*, et expliquent l'indispensabilité de son existence. Ce mandat doit contenir le but de la communication (le *skopos*) et les conditions de la traduction (comme le délai, le tarif, etc.). Le *skopos* peut être déterminé durant la négociation entre le traducteur et le client. Le mandat n'est pas considéré comme imposé au traducteur, mais plutôt comme un sujet qui est négocié, et qui doit l'être avant la traduction. Il est aussi important de déterminer le public cible dans le mandat de traduction, car cela est important pour déterminer le *skopos*, puisque la traduction produite est destinée à une culture cible qui doit être en mesure de comprendre et faire sens de cette traduction. Il est donc important de connaître la culture cible dans le mandat de traduction.

3.2.5 La question de l'intention et de la fonction

Les notions de l'intention et de la fonction sont deux notions importantes dans le cadre de la théorie du *skopos*. Elles sont à l'origine d'une critique qu'a faite Christiane Nord (2010) envers Reiß et Vermeer, en disant que l'utilisation des termes « objectif », « intention » et « fonction » comme synonymes poserait un problème et induirait en erreur. Selon elle, chaque terme possède une définition spécifique et différente de l'autre, ce qui brouillerait la compréhension de la théorie du *skopos*. Elle part du principe que l'intention n'est pas égale à la fonction, et que l'intention doit être prise en compte du point de vue de l'auteur. D'un autre côté, la fonction doit être déterminée du point de vue du traducteur ou destinataire, et c'est à eux de déterminer comment la traduction sera utilisée.

Dans la même lignée, Nord critique le fait que l'intention de l'auteur est souvent inaccessible (Nord, 2010), et souvent interprétée par le traducteur. Il n'y a aucune garantie que l'intention de l'auteur soit respectée, ou que son objectif soit atteint, ou même que cela ne soit souhaitable,

parce que tel n'est pas le but de la théorie du *skopos*, qui veut que l'offre d'information cible soir adéquate avec le mandat (Nord, 2013).

3.2.6 Évaluation

Ces éléments que présentent la théorie du *skopos*, notamment les règles et leur hiérarchie, pourraient créer de nouveaux outils et emmener à de nouveaux questionnements sur l'évaluation de la traduction. La nature de la théorie du *skopos* fait qu'il est possible de prendre de la distance par rapport aux concepts de fidélité et d'équivalence. L'accent est mis sur l'adéquation avec le mandat (Nord 2010 et 2013), ce qui donne des opportunités de nouveaux angles d'évaluation par rapport à la révision de traduction et de traduction automatique, surtout avec le concept de « *fit-for-purpose quality* » (une qualité adaptée à l'objectif).

3.2.7 Le concept de loyauté

Le concept de loyauté a été initié par Christiane Nord afin de venir compléter la théorie du *skopos* pour répondre aux critiques qui impliqueraient que la nature de la théorie du *skopos* engendrerait des traductions qui s'éloigneraient du texte source.

Nord met donc en avant l'idée que la loyauté est une catégorie qui implique des gens ou agents et non pas des choses, ou des textes, et que le concept de loyauté va dans les deux sens : il implique et la source, et la cible.

> "Loyalty commits the translator bilaterally to the source and the target sides. Il must not be mixed up with fidelity or faithfulness, concepts that usually refer to a relationship holding between the source and the target texts. Loyalty is an interpersonal category referring to a social relationship between people." (Nord, 1997, p. 125)

D'autres critiques avancent que le concept de loyauté n'était pas nécessaire car la question de fidélité et de responsabilité était déjà posée par Reiß et Vermeer (1984).

Nord ajoute au concept de loyauté que c'est un élément qui doit être inclus durant la négociation du mandat de traduction entre le traducteur et le client, qui est représenté par le respect du mandat et des engagements du traducteur (Nord, 2013).

3.2.8 Les retombées

Les premières retombées de la théorie du *skopos* ont été aperçues surtout en Allemagne, puisque la théorie a été publiée en allemand. Mais au-delà de cela, les plus importantes retombées sont celles sur l'enseignement de la traduction. Nous apercevons souvent des professeurs en traduction appliquée donner des mandats de traduction fictifs aux apprenants durant les cours, afin de leur poser un cadre et de définir d'importants éléments du contexte de la traduction, comme la fonction du texte et les destinataires.

D'autres retombées se représentent par le fait que la théorie du *skopos* introduit de nouveaux outils pour évaluer la traduction, ainsi que pour le développement des outils d'aide à la traduction.

De même, la théorie du *skopos* a eu des retombées positives sur la critique de la traduction (Amann, 1989). Elle a aussi eu une influence sur les analyses des aspects culturels de la traduction (Löwe, 1989) et sur la didactique de et la pratique de la traduction technique (Schmitt, 1989).

3.2.9 État de la littérature : la théorie du *skopos* et la traduction automatique

Durant les recherches entreprises pour la rédaction de ce mémoire, nous avons constaté que les études, articles de recherche, mémoires et thèses tournant autour de la théorie du *skopos*, sont en grand nombre, que ce soit à des fins analytiques, d'évaluation de traductions humaines de différents types de textes, ou pour des objectifs purement théoriques. Cependant, cela n'est pas le cas pour l'évaluation de traduction automatique, et plus particulièrement de la traduction de

systèmes génératifs ou neuronaux. Selon Mona Baker et al., l'engagement théorique avec la traduction automatique au sein de la traductologie était limité pour la majorité du XX^{ème} siècle, puisque les chercheurs en traduction et les chercheurs en traduction automatique étaient réciproquement indifférents par rapport à leurs domaines respectifs (Baker et Chesterman, 1998).

En recherchant des articles tournant autour du sujet, avec des mots-clés comme « ChatGPT », « traduction », « *skopos* » ou même « fonction » (en anglais et en français), sur des sites comme *Google Scholar*, *ResearchGate*, *ACL Anthology* et *arXiv*, nous avons pu trouver quelques études adjacentes. Un premier exemple est l'étude s'intitulant « *Comparative Research on Machine Translation and Human Translation of Examples in Dictionary from the Perspective of Skopos Theory »* (étude comparée sur la traduction automatique et la traduction humaine d'exemples dans le dictionnaire de la perspective de la théorie du *skopos*), dont le but était de trouver des critères afin d'évaluer la qualité de la traduction humaine et la traduction automatique d'exemples, pour pouvoir proposer des moyens d'améliorer la traduction automatique et la traduction d'exemples dans le dictionnaire (Tan et al., 2023). Cette étude a comparé la traduction automatique de *Google Translate* à des traductions humaines de 300 exemples de traduction dans un dictionnaire anglais-chinois. Pour les critères d'évaluation, l'article s'est basé sur trois éléments de la théorie du *skopos* : l'utilité pour les utilisateurs, la fidélité avec le texte source et la cohérence. L'étude a conclu que malgré les avantages que présente la traduction automatique, comme le gain de temps, la traduction humaine reste meilleure, surtout en termes de précision contextuelle et de sensibilité culturelle (Tan et al., 2023).

Un deuxième exemple est la thèse s'intitulant *Country, Nation, and Language: Machine Translation in Iceland* (le pays, la nation et la langue : la traduction automatique en Islande) centrée sur un système de traduction neuronal islandais, et qui a mis en revue les différentes

théories de traductologie, dont la théorie du *skopos*, en utilisant une perspective de traduction automatique (Wagner, 2021). La méthodologie de cette thèse est représentée par une analyse textuelle qualitative de textes de la politique de la langue en Islande, et d'entretiens avec des professionnels de la langue, dont des traducteurs et des développeurs de technologies liées à la traduction automatique. L'objectif était d'évaluer le biais que pourrait poser la traduction automatique en Islande, et son effet sur la langue qui représente une connexion forte au pays. Les résultats ont montré que même si la traduction présente des opportunités d'ouverture pour la langue islandaise, elle crée des défis quant au maintien de l'intégrité de la langue et de sa représentation culturelle (Wagner, 2021).

3.3 Conclusion

Dans ce chapitre, nous avons donné une brève histoire de la traductologie. Ensuite, nous avons présenté les points essentiels qui expliquent en détails la théorie fonctionnaliste du *skopos*. L'origine du *skopos* est le courant fonctionnaliste en traductologie, et cette théorie présente des règles qu'il faut nécessairement respecter en traduisant, si le but est de traduire selon la théorie du *skopos*. Un point clé de ce chapitre est la partie sur l'évaluation, qui met en relief le fait que la théorie du *skopos* est opportunité de créer un nouveau moyen d'évaluation de traductions, ce qui est surtout pertinent pour le travail que nous voulons effectuer pour ce mémoire. Enfin, nous avons présenté brièvement les retombées de la théorie dans les domaines de la traductologie, de la traduction, et de la didactique de la traduction. Nous avons aussi pu constater le nombre limité de références bibliographiques tournant autour du sujet de la théorie du *skopos* et de la traduction automatique, ce qui pourrait être en même temps un atout et un obstacle pour ce travail de mémoire de master.

4 Méthodologie

En prenant en compte tous les points présentés jusque-là, nous nous intéressons à procéder à une évaluation des traductions produites par ChatGPT de deux textes, un journalistique, et un publicitaire, selon la théorie du *skopos*.

Dans la partie 4.1, nous détaillerons les questions de recherches auxquelles nous aimerions répondre, et dans la partie 4.2 la procédure que nous avons appliquée, et les détails des trois méthodes d'évaluation que nous avons choisies.

4.1 Questions de recherche

Trois questions de recherches principales, avec des sous-questions, ont été la base de la conception de ce mémoire :

Question 1	**Un mandat de traduction précis, comme le prescrit la première règle de la théorie du skopos, influence-t-il la traduction produite par ChatGPT ?**
↳	ChatGPT donne-t-il une traduction différente si le mandat de traduction est précisé dans le prompt ?
Question 2	**La précision d'un mandat dans le prompt conduirait-t-elle à la production d'une traduction de meilleure qualité ?**
↳	Les traductions de ChatGPT sont-elles de meilleure qualité si le mandat est précisé dans le prompt ?
Question 3	**ChatGPT respecte-t-il toujours le mandat précisé dans les prompts ?**
↳	La précision dans notre prompt augmenterait-elle la probabilité que ChatGPT s'adapte à notre mandat ?

Q1. Un mandat de traduction précis, comme le prescrit la première règle de la théorie du *skopos*, influence-t-il la traduction produite par ChatGPT ?

Comme expliqué dans le chapitre 3, la théorie du *skopos* prescrit un ensemble de règles qui amélioreraient la qualité de la traduction si elles sont respectées. La première règle, et celle la plus importante, et que la traduction d'un texte doit être effectuée en suivant un mandat de traduction, ou une fonction particulière. Dans notre mémoire, nous chercherons à évaluer si oui ou non les traductions de ChatGPT qui sont produites avec un prompt qui contient un mandat de traduction explicite, clair et précis, sont différentes par rapport à des traductions produites avec un prompt qui demande simplement à ChatGPT de traduire un texte d'une langue A vers une langue B, sans préciser de mandat. En d'autres termes, nous cherchons à savoir si ChatGPT s'adapte en fonction des prompts donnés, et s'il adapte ses traductions en fonction de la précision du prompt donné. Notre hypothèse est que cela se manifesterait par un certain nombre et par certains types de modifications, comme des changements de style, de lexique ou de ton par exemple. Nous pensons aussi que plus nous serons précis avec notre prompt, plus les changements dans les traductions seront nombreux (Yamada, 2023).

Q2. La précision d'un mandat dans le prompt conduirait-t-elle à la production d'une traduction de meilleure qualité ?

En plus du respect de la théorie du *skopos* et d'un mandat de traduction, il serait important d'évaluer la qualité des traductions produites par ChatGPT. La « qualité » en traduction a toujours été difficile à définir et est un sujet très large (Lommel et al, 2018 ; Fields et al, 2014 ; Hiebl et Gromman, 2023), et qui requiert la considération d'aspects linguistiques, fonctionnels et « axés vers l'utilisateur » (Callison-Burch et al., 2009). Dans le cadre de notre recherche, nous pensons que la précision d'un mandat dans le prompt conduira à la production d'une traduction d'une meilleure qualité, et qu'une adaptation à un prompt précisant un mandat est

manifestée quant à elle par des changements, comme nous l'avons évoqué dans la question de recherche précédente. Afin de pouvoir évaluer la qualité des traductions, nous utiliserons des méthodes d'évaluation humaines existantes, fondées sur la méthode d'évaluation « ***M**ultidimensional **Q**uality **M**etrics* » (MQM) (détaillée dans la partie 4.2.4) pour analyser à quelle point les traductions produites par ChatGPT avec un prompt précisant un mandat sont améliorées par rapport à celles produites avec un prompt ne précisant pas le mandat.

Q3. ChatGPT respecte-t-il toujours le mandat précisé dans les prompts ?

Dans les deux premières questions de recherche, nous voudrions évaluer le produit de ChatGPT, ou sa traduction par rapport à la précision des prompts. Dans cette troisième et dernière question de recherche, nous voudrions évaluer la fréquence avec laquelle ChatGPT arrive à respecter un mandat de traduction donné. En effet, nous pensons que la précision dans un prompt aura plus de chance de donner une traduction adaptée à des objectifs particuliers, et que ChatGPT aura plus tendance à donner une traduction qui remplit les conditions voulues si le prompt précise explicitement le mandat. Afin de vérifier cette hypothèse, nous avons eu recours à un test suites de phrases contenant toutes un élément facilement repérable et traduisible, afin de facilement pouvoir évaluer la fréquence avec laquelle ChatGPT respecte le mandat précisé dans un prompt.

4.2 Procédure

Dans cette partie, nous allons décrire la procédure que nous avons adoptée afin de répondre à nos questions, en commençant par la comparaison entre prompts (partie 4.2.3), puis l'évaluation de la qualité avec la méthode d'évaluation « ***M**ultidimensional **Q**uality **M**etrics* » (partie 4.2.4) et, enfin, la réalisation d'une série de tests afin de mettre en exergue un phénomène de traduction particulier avec ChatGPT (partie 4.2.2).

4.2.1 Choix des types de texte

Pour les textes à traduire, nous avons choisi un texte journalistique et un texte publicitaire, car leurs fonctions et styles sont drastiquement différents. En effet, le texte journalistique a une fonction informative, et est souvent rédigé d'une manière claire et directe, alors que la fonction du texte publicitaire est d'attirer le lecteur et de le convaincre, en utilisant souvent un langage riche en métaphore, en adjectifs et en figure de style.

Pour le texte journalistique, nous avons choisi un texte tournant autour de la visite d'Elon Musk en Chine, issu du journal *The Guardian*[4]. Pour le texte publicitaire, nous avons extrait un texte figurant sur le site de l'hôtel « *Four Seasons* » à Genève[5].

La langue source est l'anglais, et nous avons choisi de faire la traduction vers le français.

4.2.2 Choix des prompts

Afin de répondre à notre première question de recherche, il fallait vérifier si la traduction produite avec un prompt simple, qui demande uniquement une traduction à ChatGPT, est différente des traductions produites avec des prompts précisant des mandats ou des contextes plus clairs. Pour ce faire, nous avons traduit les deux textes (journalistique et publicitaire) avec plusieurs prompts différents, représentés dans le tableau 2 ci-dessus.

Prompt 1	Traduis ce texte en français
Prompt 2	Traduis ce texte en français pour que le texte traduit ait une fonction explicative
Prompt 3	Traduis ce texte en français en respectant les règles de la théorie du *skopos*
Prompt 4	La théorie du *skopos* est une théorie qui dit qu'il faut respecter la fonction du texte et le mandat pour mener à bien l'activité de traduction. Les 6 règles de la théorie du *skopos* sont les suivantes : La première règle, qui est la plus importante et à qui la

[4] https://www.theguardian.com/business/2024/apr/29/tesla-elon-musk-china-visit-li-qiang-mapping-data-baidu
[5] https://www.fourseasons.com/geneva/

	théorie du *skopos* doit son nom, est le fait que l'action traductive est déterminée par don *skopos*, donc son objectif. La deuxième règle prévoit que la traduction n'est qu'une offre d'information dans une culture source et donne lieu à une offre d'information dans une culture cible, et cela permet au traducteur ou à la traductrice de prendre de la distance par rapport à la question du sens. Le but de la traduction ne serait pas de transmettre le sens unique d'un texte source dans un texte cible, mais de transmettre une offre d'information qui est disponible dans le texte source et d'en faire une offre d'information dans le texte cible. Une conséquence directe de cette règle de la théorie du *skopos* serait d'avoir une variété d'interprétation de sens et donc de traductions possibles – donc, autant de traductions que d'objectifs, de fonction du texte source. La troisième règle de la théorie du *skopos* est que la fonction du texte cible pourrait être différente de celle du texte source. Cette règle vient s'opposer à la théorie de l'équivalence, où la question de l'effet équivalent du texte source et du texte cible est posée. Cette règle impliquerait que la rétro-traduction ne serait plus possible. La quatrième règle de la théorie du *skopos* met en exergue l'importance de la cohérence intra-textuelle, qui signifie une cohérence à l'intérieur de la culture cible, où les destinataires du texte cible pourraient faire sens de ce qui a été traduit. La cinquième la règle de théorie du *skopos* vient compléter la quatrième, en ajoutant l'importance de la cohérence extratextuelle, qui à son tour, signifie la cohérence entre la traduction et le texte source, et entre la traduction et le mandat, ou le *skopos*. Cela veut dire que le texte cible pourrait refléter le texte source, mais cela pourrait aussi ne pas être le cas, si le mandat de traduction (ou le *skopos*) ne le demande pas. La théorie du *skopos* préconise une hiérarchie des règles mentionnées ci-dessus, qui est elle-même considérée comme une sixième règle à part entière de la théorie du *skopos* par certains (cf. Du, 2012). Dans cette hiérarchie des règles, Vermeer et Reiß ont établi une importance non égale aux règles de la théorie du *skopos*, qui met en premier la règle de la fonction du texte (ou la règle du *skopos*) qui dicte l'action traductive. C'est d'ailleurs selon cette première règle que les évaluations de traductions pourraient être accomplies. En deuxième lieu, Reiß et Vermeer placent la règle de cohérence intra-textuelle, qui implique la cohérence de la traduction avec la culture cible, et la capacité des destinataires de la traduction à faire ses de cette traduction. En troisième et dernier lieu, se place la règle de fidélité, ou la cohérence extratextuelle. Cette règle est la dernière dans cette hiérarchie car, comme mentionné auparavant, il se pourrait que le *skopos*, ou mandat de la traduction voudrait que la fidélité, ou la cohérence extratextuelle, ne soit pas respectée. En tenant compte de cela, et en respectant les règles du *skopos*, traduis ce texte journalistique en français pour que le texte traduit ait une fonction explicative
Prompt 5	Traduis ce texte journalistique en français en respectant ces règles de la théorie du *skopos* : 1- Le texte source est déterminé par son *skopos*, donc sa fonction. 2- Un texte source est une offre d'information dans une culture cible et un texte cible concernant une offre d'information dans une culture source et un texte source.

	3- Un texte cible ne communique pas des informations qui pourraient être retraduites dans le sens inverse. 4- Un texte cible doit être cohérent en soi. 5- Un texte cible doit être cohérent avec le texte source. 6- Les 5 règles susmentionnées doivent être appliquées par ordre hiérarchique, la règle du *Skopos* étant celle prédominante.
Prompt 6	Traduis ce texte journalistique/publicitaire en français en respectant ces règles de la théorie du *skopos* : 1- Le texte source est déterminé par son *skopos*, donc sa fonction. 4- Un texte cible doit être cohérent en soi. 5- Un texte cible doit être cohérent avec le texte source. 6- Les 5 règles susmentionnées doivent être appliquées par ordre hiérarchique, la règle du *Skopos* étant celle prédominante. La fonction du texte traduit est différente de la fonction du texte source, qui est un texte journalistique/publicitaire. Je veux que le texte cible ait une fonction explicative.
Prompt 7journ	Traduis ce texte en français en respectant le mandat de traduction suivant : Le texte sera publié dans le journal Tribune de Genève L'audience est francophone mais surtout suisse romande.
Prompt 7pub	Un groupe voudrait organiser un voyage touristique à Genève, et cherche des hôtels pour passer leur séjour. Dans le cadre d'un dossier informatif pour les aider à faire leur choix, traduis ce texte en français, pour une audience francophone de France, afin de les informer des détails de cet hôtel, en utilisant un style informatif sans tournures de style et métaphore

Tableau 2. Liste des différents prompts utilisés pour effectuer les deux premières évaluations dans ce mémoire

Ces prompts ont été conçus afin de répondre à notre première question de recherche et de vérifier notre première hypothèse, qui est que la précision du mandat dans le prompt donne une traduction différente que celle produite par un prompt qui ne précise pas le mandat.

Le prompt 1 représente une instruction simple qui demande à ChatGPT de traduire le texte vers le français, et représente la traduction à laquelle nous comparerons les traductions effectuées avec les autres prompts.

Afin d'ajouter de la nuance à notre recherche, et de vérifier toutes les possibilités de mandat, nous avons rédigé des prompts précisant des éléments légèrement différents à chaque fois, pour

couvrir plusieurs options. Nous avons commencé par préciser que nous voulions que la traduction ait une fonction différente de celle du texte source ; c'est pour cela que dans le prompt 2, nous avons demandé que la traduction ait une fonction explicative. Comme expliqué dans la partie 4.2.1 précédente, le texte journalistique a une fonction informative, alors que le texte publicitaire sert à convaincre le lecteur d'un sujet ou produit particulier. Un texte explicatif pourrait être définit comme un texte riche en procédés explicatifs comme la définition et/ou la reformulation, et cherche à répondre à une question comme « comment ? » ou « pourquoi ? ». En demandant à ChatGPT de changer la fonction du texte, nous nous attendions en théorie à ce qu'il adapte les tournures syntaxiques pour qu'elles expliquent l'événement ou le produit présent dans les textes sources.

Dans le prompt 3, nous avons demandé à ChatGPT de traduire en respectant la théorie du *skopos*, qui est un prompt qui certes précise une instruction supplémentaire à suivre à part la traduction, mais qui reste un mandat assez vague, surtout que nous n'avons pas précisé dans le prompt les éléments à prendre compte ou les changements à introduire si la traduction respect la théorie du *skopos*.

Dans le prompt 4, nous avons fourni à ChatGPT une explication détaillée de la théorie du *skopos*, avant de lui préciser aussi que nous aimerions que la traduction ait une fonction explicative, donc différente de celle du texte source. Nous jugeons que ce prompt est plus clair que les précédents, puisque nous avons fournis des détails et un mandat à respecter, même s'il est (intentionnellement) vague.

Dans le prompt 5, nous avons demandé à ChatGPT de respecter les règles de la théorie du *skopos*, cette fois-ci en fournissant explicitement les règles qu'elle prescrit. En théorie, cela voudrait dire que la fonction du texte cible reste la même que le texte source, mais que ChatGPT

devrait prendre en compte des éléments comme la cohérence inter- et intra-textuelle, ce qui veut dire que la traduction produite devrait avoir une meilleure qualité.

Dans le prompt 6, nous avons précisé davantage les règles de la théorie du *skopos* à suivre en ne rédigeant dans le prompt que 4 des 6 règles du skopos, qui tournent autour de la cohérence, et nous avons aussi précisé que nous aimerions que la traduction ait une fonction explicative.

Enfin, nous avons rédigé deux derniers prompts : le prompt 7journ et le prompt 7pub, qui, à leur tour, contiennent des mandats de traduction clairs et qui pourraient être utilisés dans un contexte académique ou professionnel, pour qu'un traducteur humain sache dans quel contexte il faut qu'il traduise son texte. Contrairement, à tous les autres prompts (de 1 à 6), qui ont été utilisés pour traduire les deux textes, nous avons utilisé le prompt 7journ pour traduire uniquement le texte journalistique, et nous avons utilisé le prompt 7pub uniquement pour traduire le texte publicitaire.

Nous avons utilisé ces prompts et les traductions produites dans notre première évaluation (partie 4.2.3) afin d'effectuer une comparaison des changements effectués dans chaque traduction en fonction des prompts et par rapport à la traduction produite avec le prompt 1.

Les traductions produites avec le prompt 1, le prompt 7journ et le prompt 7pub ont aussi été utilisée dans la deuxième évaluation, où nous avons demandé à une experte d'évaluer leur qualité selon la méthode MQM (partie 4.2.4).

À noter que la traduction des textes avec ChatGPT en utilisant ces prompts a été effectuée le 20.05.2024.

4.2.3 Première évaluation : comparaison des modifications dans les traductions en fonction des prompts

Afin de répondre à la première question de recherche, nous avons procédé à la traduction de l'anglais vers le français des deux textes plusieurs fois avec l'outil ChatGPT, en utilisant à chaque fois les prompts dont les détails se trouvent dans la partie 4.2.2 précédente. Comme précisé, afin de pouvoir juger si ChatGPT respecte ou pas une fonction précise du texte, et s'il a la capacité s'adapter à la fonction du texte en fonction des prompts, nous avons traduit les textes avec des prompts différents, dans le but de de comptabiliser le nombre de changements effectués par rapport à cette traduction et d'étudier les différences.

Pour ce faire, nous avons procédé à une analyse quantitative, en se basant sur la méthode de « ***T****ranslation* ***E****dit* ***R****ate* » ou TER (Snover et al., 2006), afin de pouvoir comptabiliser les changements d'une manière concrète et afin d'obtenir des valeurs quantifiables et d'effectuer une comparaison concluante. Pour faire ce compte, nous avons conçu des tableaux Google Sheets pour les deux textes, qui se trouvent en annexe. L'analyse quantitative de cette comparaison, ainsi que ses résultats, se trouvent dans la partie 5.1.1.

Ensuite, et dans les mêmes tableaux, nous avons procédé à une analyse qualitative, en se concentrant non pas sur le nombre de changements mais sur leurs différents types, et en les décrivant en détails afin d'arriver à nos conclusions. Ces différents changements ont été repérés eux aussi dans les même tableaux Google Sheets accessibles en annexe. L'analyse qualitative et ses résultats se trouvent dans la partie 5.1.2.

4.2.4 Deuxième évaluation : évaluation MQM

Cette deuxième évaluation a pour but de répondre à notre deuxième question de recherche et de vérifier notre deuxième hypothèse, où nous pensons qu'un prompt précis qui contient un mandat précis produit une traduction de meilleure qualité. Pour vérifier cette hypothèse, et afin

de répondre à la deuxième question de recherche, nous avons décidé de faire une analyse selon les standards ***M**ultidimensional **Q**uality **M**etrics*, ou MQM, qui est une méthode d'évaluation basée sur une typologie d'erreurs, conçue pour évaluer la qualité de la traduction[6]. Cette méthode a pour but de réduire – voire complètement éviter – la subjectivité des annotateurs lors de l'évaluation d'une traduction, en créant une typologie des erreurs standardisée, sur laquelle il est possible de se baser pour juger la qualité d'une traduction. La catégorisation des erreurs selon la méthode d'évaluation MQM est représentée dans la figure 5 ci-dessous.

Error Category	Error Subcategory	Description
Accuracy	Addition	Translation includes information not present in the source.
	Omission	Translation is missing content from the source.
	Mistanslation	Translation does not accurately represent the source.
	Untranslated text	Source text has been left untranslated.
Fluency	Punctuation	Incorrect punctuation (for locale or style).
	Spelling	Incorrect spelling or capitalization.
	Grammar	Problems with grammar, other than orthography.
	Register	Wrong grammatical register (eg. Inappropriately informal pronouns).
	Inconsistency	Internal insconsistency (not related to terminology).
	Charcater Encoding	Characters are garbled due to incorrect encoding
Terminology	Inappropriate for Context	Terminology is non-standard or does not fit context.
	Inconsistent use	Terminology is used inconsistently.
Style	Unidiomatic	Translation has stylistic problems.
Locale convention	Address format	Wrong format for addresses.
	Currency format	Wrong format for currency.
	Date format	Wrong format for dates.
	Name format	Wrong format for names.
	Telephone format	Wrong format for telephone numbers.
	Time format	Wrong format for time expressions.
Source error		An error in the source.

Figure 5. Catégorisation des erreurs dans l'évaluation MQM

Ce type d'évaluation humaine compte sur un score de poids des erreurs, catégorisées entre des erreurs de petite sévérité, qui ne sont pas graves aux yeux de l'évaluateur (« *minor errors* ») et des erreurs de grande sévérité, qui sont plus graves et nécessitent une plus grande attention (« *major errors* »), pour représenter la sévérité des erreurs repérées. La pénalité numérique des « *minor errors* » est donc inférieure à celle des « *major errors* », et cela se reflètera lors du calcul des scores.

[6] https://themqm.org/

Pour les fins de notre recherche, nous avons demandé à une experte d'évaluer quatre traductions produites par ChatGPT selon la méthode MQM :

- La traduction du texte journalistique avec le prompt 1, qui ne précise aucun mandat ;
- La traduction du texte publicitaire avec le prompt 1, qui ne précise aucun mandat ;
- La traduction du texte journalistique avec le prompt 7journ, qui donne un mandat précis spécifique au texte journalistique à ChatGPT ;
- La traduction du texte publicitaire avec le prompt 7pub, qui donne un mandat précis spécifique au texte publicitaire à ChatGPT ;

Pour faciliter la lisibilité, nous reprenons les prompts en question dans le tableau 3 ci-dessous :

Prompt 1	Traduis ce texte en français
Prompt 7journ	Traduis ce texte en français en respectant le mandat de traduction suivant : Le texte sera publié dans le journal Tribune de Genève L'audience est francophone mais surtout suisse romande.
Prompt 7pub	Un groupe voudrait organiser un voyage touristique à Genève, et cherche des hôtels pour passer leur séjour. Dans le cadre d'un dossier informatif pour les aider à faire leur choix, traduis ce texte en français, pour une audience francophone de France, afin de les informer des détails de cet hôtel, en utilisant un style informatif sans tournures de style et métaphore

Tableau 3. Tableau reprenant les 3 prompts utilisés pour les traductions de l'évaluation MQM

Pour clarifier, ces prompts sont les mêmes présentés dans la partie 4.2.2 ; aucun changement n'y a été introduit.

Comme pour la première évaluation, le prompt 1 est considéré comme un prompt simple qui ne précise aucun mandat ou contexte. Le prompt 7journ a été utilisé pour traduire le texte journalistique et pour lui donner un contexte plus clair et une audience cible, afin qu'il adapte possiblement sont vocabulaire. Le prompt 7pub a été utilisé pour traduire le texte publicitaire,

afin que ChatGPT puisse possiblement adapter le texte publicitaire en un texte plus simplifié et sans figures de style.

Après la première évaluation, qui nous avait permis de préciser des fonctions différentes et de voir comment ChatGPT changeait ses traductions en fonction de prompts plus ou moins précis, cette deuxième évaluation nous permettra de vérifier si les traductions produites par un mandat précis sont de meilleure qualité, grâce à l'évaluation MQM.

Après la traduction des deux textes sur ChatGPT en utilisant les prompts 1, 7journ et 7pub susmentionnés, nous avons envoyer à l'évaluatrice un fichier contenant les textes sources et les traductions, ainsi qu'un fichier Microsoft Excel, contenant l'environnement d'évaluation MQM avec les explications de chaque catégorie, et un calculateur de score automatique. Le dossier contenait de même un document Excel par traduction (donc 4 documents Excel, un pour chaque traduction), et un fichier Microsoft Word avec des instructions expliquant à l'évaluatrice comment procéder. Tous ces documents sont accessibles dans les annexes suivantes : annexe 1 pour le texte journalistique en anglais, annexe 2 pour le texte publicitaire en anglais, annexe 3 pour les instructions, et annexe 6 pour les tableaux de l'évaluation MQM.

Dans les instructions, nous avons expliqué que l'évaluatrice devait d'abord se familiariser avec la catégorisation des erreurs MQM, puis lire les traductions et leurs texte sources. Après avoir identifié des potentielles erreurs dans les traductions, l'évaluatrice devait copier et coller dans le tableau Excel correspondant les phrases contenant des erreurs (selon la catégorisation MQM). Ensuite, elle devait choisir la catégorie d'erreur correspondante dans le tableau Excel, et ajouter la sévérité de l'erreur. L'évaluatrice pouvait aussi ajouter un commentaire si cela était nécessaire. Cette méthode d'évaluation a été conçue de telle sorte à ce que l'évaluatrice puisse faire l'évaluation en contexte, et pas segment par segment, car il était important d'évaluer le texte dans son ensemble plutôt que phrase par phrase.

Nous n'avons pas précisé dans les instructions quel texte était traduis avec quel prompt, pour éviter tous biais possible.

4.2.5 Troisième évaluation : test-suites

Cette troisième évaluation sert à répondre à notre troisième question de recherche et à vérifier notre troisième hypothèse. Nous pensons que plus le prompt est précis, plus il y a une probabilité que ChatGPT produise une traduction adaptée. En d'autres termes, nous pensons que la fréquence de son adaptation dépendra de la précision du prompt qui lui est donné. Dans cette troisième évaluation, et afin d'arriver à vérifier cette hypothèse, nous avons choisi de procéder à un test-suite, qui est une méthode d'études de la traduction de phénomènes spécifiques (King et Falkedal, 1990). Cette méthode consiste en un choix d'un corpus, composé de phrases hors contexte n'appartenant pas à un même texte, mais contenant toutes le même phénomène linguistique que le chercheur désire évaluer et selon ses besoins (King et Falkedal, 1990). En traduction, et selon les combinaisons linguistiques, ces phénomènes peuvent être les anaphores, l'accord des articles avec leurs noms, ou, dans notre cas, la traduction d'acronymes.

Par exemple, Avramidis et al. (2019) ont voulu évaluer 16 systèmes de traduction automatique, avec la combinaison linguistique anglais allemand, afin d'évaluer leur performance dans la traduction de divers phénomènes grammaticaux, comme l'accord du verbe avec le sujet, l'ambiguïté, la subordination, la ponctuation, parmi d'autres : l'étude a été faite sur 107 phénomènes grammaticaux divisés en 14 catégories. De même, Burchardt et al. (2017) ont comparé la performance de trois systèmes de traduction automatique (*rule-based, phrase-based* et les systèmes neuronaux) en utilisant des test suites pour évaluer leurs traductions de phénomènes linguistiques.

Dans le contexte de ce mémoire, et afin d'arriver à ce choix, nous nous sommes basées sur les résultats de l'analyse qualitative (expliqués dans la partie 5.1.2) pour effectuer un test suite sur

des phrases contenant des acronymes, afin d'évaluer à quelle fréquence ChatGPT les traduisait, et si la précision dans notre prompt augmenterait cette fréquence. Nous considérons que cette question est importante, puisque nous pensons que la réponse aurait une influence sur comment ChatGPT est utilisé pour traduire des textes.

Afin d'effectuer notre test-suite, nous avons sélectionné un corpus de **50** phrases, sourcées de textes de domaines différents. Le corpus est donc réparti comme suit :

- **10** phrases de rapports d'organisations internationales ;
- **10** phrases d'articles journalistiques ;
- **10** phrases de rapport annuels et de communication d'entreprise ;
- **10** phrases d'articles scientifiques ;
- **10** phrases de publications de réseaux sociaux.

Dans ces **50** phrases, nous avons compté **78** d'occurrences d'acronymes.

Toutes ces phrases contiennent donc des acronymes spécifiques à ces domaines. La variété de ce corpus est intentionnelle, puisque cela favorise des résultats englobant plusieurs domaines, le but de notre mémoire n'étant pas d'analyser la traduction des acronymes en soi, qui est un domaine assez large (Lymar, 2021 ; Abdul-Razzaq, 2008). Pour ce test suites, avons effectué la traduction sur ChatGPT phrase par phrase en utilisant les prompts suivants :

- Une première fois en utilisant le prompt « a » suivant : « traduis cette phrase en français » ;
- Une deuxième fois en utilisant le prompt « b » suivant : « traduis cette phrase en français en traduisant les acronymes ».

4.2.6 Conclusion

Dans ce chapitre, nous avons détaillé les questions de recherche auxquelles nous aimerions répondre dans ce mémoire, et la méthodologie utilisé pour répondre à chacune de ces questions. Pour notre première question de recherche, où nous nous demandons si un prompt contenant un mandat précis influencerait la qualité des traductions données par ChatGPT, nous effectuerons une comparaison de différentes traductions en fonction de prompts par rapport à un prompt de référence, et nous analyserons les changements d'une manière quantitative et qualitative. Pour la deuxième question de recherche, nous nous demandons si une traduction qui s'adapte à la première règle de la théorie du *skopos*, donc adaptée à un mandat, est de meilleure qualité, et pour cela, nous utiliserons la méthode d'évaluation MQM et analyserons les résultats. Enfin, pour la troisième question de recherche, nous cherchons à jauger la fréquence avec laquelle ChatGPT respecte un mandat précis, et nous pensons que plus le mandat est précis, plus il est capable de produire une traduction adaptée. Pour cela, nous effectuerons un test suites avec des phrases contenant toutes des acronymes, et nous les traduirons à deux reprises, une fois avec un prompt simple, et une fois avec un prompt qui précise explicitement que ChatGPT doit traduire les acronymes dans les phrases.

Dans le chapitre 5 suivant, nous exposerons et analyserons les résultats de ces trois méthodes d'évaluation.

5 Analyse

Comme expliqué dans le chapitre 4, notre recherche repose sur trois méthodes principales afin d'arriver à répondre à nos questions de recherches, à savoir, la comparaison des modifications apportées grâce à des prompts différents, l'évaluation MQM avec des traductions de prompts précis, et la traduction de phrases contenant un phénomène spécifique sous forme de test suites.

Dans ce chapitre, nous effectuerons une analyse des résultats de ces trois méthodes, et nous en tirerons des conclusions après avoir fait une synthèse de tous les résultats trouvés.

5.1 Comparaison des modifications dans les traductions en fonction des prompts

Dans les parties suivantes, nous présenterons les résultats de l'analyse quantitative en premier lieu (partie 5.1.1), où nous avons compté le nombre de changements dans les traductions par rapport au prompt 1, et l'analyse qualitative en second lieu (partie 5.1.2), où nous présenterons les types de changements effectués par ChatGPT dans les différentes traductions.

5.1.1 Analyse quantitative des changements par rapport au prompt 1

Afin de mieux représenter tous les changements dans les traductions dans les prompts de 2 à 7journ et 7pub par rapport au prompt 1, nous avons compté le nombre de modifications, en nous inspirant du score « ***T****ranslation* ***E****dit* ***R****ate* » (TER) (Snover et al., 2006), qui calcule les insertions, les suppressions, les substitutions et les changements de place dans la phrase par rapport à une traduction de référence, qui est, dans notre cas, la traduction donnée avec le prompt 1. Le nombre de toutes ces modifications, ainsi que le nombre total des modifications par texte et par prompt, est représenté dans le tableau 4 ci-dessous.

Type de texte	Prompts	Insertions	Suppressions	Substitutions	Changement de place	Total de changements
Texte journalistique	Prompt 2	4	0	19	0	23
	Prompt 3	2	2	27	7	38
	Prompt 4	2	2	26	7	37
	Prompt 5	1	4	16	0	21
	Prompt 6	15	2	32	7	56
	Prompt 7journ	14	15	47	0	76
Texte publicitaire	Prompt 2	5	0	12	0	17
	Prompt 3	5	0	12	0	17
	Prompt 4	3	0	11	0	14
	Prompt 5	2	0	12	0	14
	Prompt 6	3	1	15	2	21
	Prompt 7pub	4	37	67	0	108

Tableau 4. Nombre total des insertions, suppressions, substitutions et déplacement de séquences pour les deux types de texte et pour tous les prompt par rapport au prompt 1 (« traduit ce texte en français »)

Comme nous pouvons le constater, la majorité des changements effectués par ChatGPT pour les prompts différents consiste en des substitutions, le remplacement d'un mot par un autre.

À partir des résultats présentés dans le tableau 4 ci-dessus, nous pouvons déjà constater qu'il y a des changements dans les traductions de tous les prompts par rapport à la traduction du prompt 1 pour les deux textes, journalistique et publicitaire.

En effet, le nombre total des changements pour la traduction du texte journalistique avec le prompt 2 est de **23**, le nombre total des changements pour la traduction avec le prompt 3 est de **38**, le nombre total des changements pour la traduction avec le prompt 4 est de **37**, le nombre total des changements pour la traduction avec le prompt 5 est de **21**, le nombre total des

changements pour la traduction avec le prompt 6 est de **56**, et enfin, le nombre total des changements pour la traduction avec le prompt 7journ est de **76**.

Pour ce qui est de la traduction du texte publicitaire, nous pouvons constater que le nombre total des changements pour la traduction avec le prompt 2 est de **17**, le nombre total des changements pour la traduction avec le prompt 3 est lui aussi de **17**, le nombre total des changements pour la traduction avec le prompt 4 est de **14**, le nombre total des changements pour la traduction avec le prompt 5 est luis aussi de **14**, le nombre total des changements pour la traduction avec le prompt 6 est de **21**, et enfin, le nombre total des changements pour la traduction avec le prompt 7pub est de **108**.

Nous pouvons aussi constater que le nombre total des changements s'élève avec chaque prompt d'une manière plus ou moins constante, et ce pour les deux types de texte. Nous remarquons que pour les deux types de texte, le plus grand nombre de changements est celui remarqué dans la traduction avec les derniers prompts, à savoir **76** changements avec le prompt 7journ pour le texte journalistique, et **108** changements avec le prompt 7pub pour le texte publicitaire.

Si nous analysons les détails de ces changements, nous pouvons constater que toujours pour les deux types de texte, le type de modification qui a le plus grand nombre est les substitutions. En effet, pour le texte journalistique, le total de substitutions pour les traductions de tous les prompts est de **47**, alors que le total des insertions est de **14**, celui des suppressions est de **15**, et celui des changements de place est de **0**. Pour le texte publicitaire, nous pouvons faire le même constat, puisque le total des substitutions est de **67**, alors que celui des insertions est de **4**, celui des suppressions est de **37** et celui des changements de place est de **0**.

Enfin, nous pouvons constater que le nombre de changement le plus bas, quant à lui, figure dans les traductions produites par le prompt 5, avec **21** changements pour le texte journalistique et **14** changements pour le texte publicitaire (en égalité avec le prompt 4).

En se basant sur ces résultats, nous pouvons conclure que les prompts contenant des mandats plus précis (ou les prompts 7journ et 7pub) entraînent un nombre de changement assez élevé par rapport à un prompt simple, et même par rapport à des prompts contenant des mandats qui ne sont pas précis.

5.1.2 Analyse qualitative des changements par rapport au prompt 1

Afin de mener à bien cette comparaison des traductions faites par les prompts de 2 à 6 avec la traduction faite avec le prompt 1 (notre traduction de référence pour cette évaluation), nous avons placé les traductions dans un tableau Excel à côté de la traduction références pour faciliter la lecture, et nous avons marqué en rouge les différences dans chaque phrase en décrivant les types de changements effectués dans une colonne « commentaires ». Nous avons décrit les changements que nous avons considéré être les plus flagrants dans les parties 5.1.2.1 (analyse qualitative du texte journalistique) et 5.1.2.2 (analyse qualitative du texte publicitaire), et ces tableaux peuvent être consultés dans l'annexe 4 pour le tableau contenant les comparaisons du texte journalistique et dans l'annexe 5 pour le tableau contenant les comparaisons du texte publicitaire.

5.1.2.1 Texte journalistique

Comme mentionné, un des deux textes utilisés pour cette évaluation est de type journalistique. Nous allons procéder dans cette partie à une analyse qualitative des changements effectués par ChatGPT dans les traductions des prompts de 2 à 7journ par rapport à la traduction effectuée avec le prompt.

Pour faciliter la compréhension et la lisibilité de notre analyse, nous avons présenté dans le tableau 5 ci-dessous la traduction du texte journalistique avec le prompt 1. Nous rappelons que cette traduction a été utilisée comme traduction « référence » pour cette première évaluation, et que la comparaison effectuée dans l'analyse qualitative et quantitative est par rapport à cette

traduction produite à l'aide du prompt 1. Les numérotations des paragraphes sont utilisées dans notre analyse qualitative, et nous les avons inclues dans le tableau 5 pour faciliter la lisibilité.

# paragraphe	Traduction du texte journalistique avec le prompt 1
1	Moment décisif pour Tesla alors que la visite d'Elon Musk en Chine récolte rapidement des récompenses.
2	Un accord pour utiliser les données de cartographie du géant chinois de la recherche sur le web, Baidu, constitue une avancée majeure vers le lancement de la technologie d'assistance à la conduite sur le plus grand marché automobile du monde.
3	La visite d'Elon Musk en Chine aurait rapporté des récompenses immédiates avec un accord permettant à Tesla d'utiliser les données de cartographie fournies par la société de recherche sur le web Baidu, une avancée majeure dans l'introduction de la technologie d'assistance à la conduite sur le plus grand marché automobile du monde.
4	Musk a effectué une visite non annoncée en Chine ce week-end. Le milliardaire a publié une photo de sa rencontre avec le Premier ministre chinois, Li Qiang, sur X, le réseau social qu'il a pris en charge en 2022.
5	Baidu, qui domine la recherche sur le web en Chine, fournira des fonctions de cartographie et de navigation pour aider Tesla à exploiter sa technologie d'assistance à la conduite, qu'il appelle "conduite entièrement autonome" ou FSD, selon des sources citées par Bloomberg News. Les services de cartographie - cruciaux pour les technologies d'assistance à la conduite - sont strictement contrôlés par le gouvernement chinois.
6	Malgré son nom, la FSD ne fournit pas de capacités de conduite autonome : elle nécessite un conducteur qui a "les mains sur le volant et est prêt à reprendre le contrôle à tout moment". Cependant, son lancement en Chine pourrait aider Tesla dans la lutte acharnée pour la part de marché dans le pays et fournir plus de revenus. Son coût est de 8 000 dollars ou 99 livres sterling par mois, bien qu'elle ne soit pas disponible dans de nombreux pays.
7	Musk adopte souvent une attitude combative dans ses relations avec les politiciens, comme des critiques virulentes à l'encontre du président américain Joe Biden, ou une impasse au Brésil contre un gouvernement qu'il accuse de censurer X, anciennement connu sous le nom de Twitter. Cependant, il a adopté un ton plus conciliant envers le deuxième politicien le plus puissant de Chine, déclarant qu'il était "honoré" de rencontrer Li.

Tableau 5. Traduction du texte journalistique avec le prompt 1

En comparant les résultats des 6 autres prompts au prompt 1 (« **traduis ce texte en français** »), nous pouvons constater les résultats suivants :

5.1.2.1.1 Comparaison prompt 1 avec prompt 2

Lien vers la feuille Google Sheets contenant la comparaison prompt 1 avec prompt 2

Dans cette traduction, nous avons pu constater que les différences principales sont restreintes au début du texte (le titre et les deux premiers paragraphes). La différence la plus flagrante est

le changement dans la syntaxe du titre : en effet, ChatGPT a transformé le titre en un phrase complète en ajoutant un verbe, des articles, un connecteur logique et de la ponctuation. Le titre est donc transformé de « *Moment décisif pour Tesla alors que la visite d'Elon Musk en Chine récolte rapidement des récompenses* » en « *La visite d'Elon Musk en Chine est un moment décisif pour Tesla, car elle a rapidement donné des résultats positifs* ». Les autres différences sont surtout des remplacements de mots par leurs synonymes, et plus précisément le remplacement de l'expression *« récolte rapidement des récompenses »* par *« rapidement donné des résultats positifs »* dans le titre, et un remplacement d'article qui ne change pas le sens de phrase dans ce contexte. ChatGPT a aussi remplacé le mot « permettant » par le mot « *pour que* » avant « *Tesla utilise les données …* ». Il est important aussi de noter que, comme dans la traduction effectuée avec le prompt 1, a commis une erreur dans la traduction de la valeur monétaire présente dans le paragraphe 6 : « *Son coût est de 8 000 dollars ou 99 livres sterling par mois* ». Le montant mentionné dans le texte source (qu'il est possible de consulter dans l'annexe 1) est de 99 dollars américains et non livres sterling.

5.1.2.1.2 Comparaison prompt 1 avec prompt 3

Lien vers la feuille Google Sheets contenant la comparaison prompt 1 avec prompt 3

Dans le résultat de ce prompt, nous pouvons remarquer des changements qui s'étendent plus au reste du texte, même s'ils ne sont pas différents des changements constatés dans la partie précédente (5.1.2.1.1), à savoir, la transformation du titre en une phrase complète et le remplacement par des synonymes. En effet, le titre dans la traduction du prompt 1 était « *Moment décisif pour Tesla alors que la visite d'Elon Musk en Chine récolte rapidement des récompenses* » et est devenu dans la traduction du prompt 3 « *"Un moment décisif" pour Tesla lors de la visite d'Elon Musk en Chine qui récolte rapidement des récompenses* ». ChatGPT a aussi remplacé « *un accord pour utiliser* » par « *l'accord visant à* » au début du paragraphe 2,

et il a aussi remplacé le complément de nom « *de cartographie* » par l'adjectif « *cartographiques* », et ce deux fois dans le texte. Il a aussi remplacé l'adjectif « *majeure* » par l'adjectif « *significative* » en référence au mot « avancée » dans le paragraphe 2. Nous pouvons aussi constater que dans le paragraphe 5 du texte, ChatGPT a changé l'ordre des mots : il a déplacé le syntagme prépositionnel *« Selon des sources citées par Bloomberg News »,* au début de la phrase. Ce changement n'a aucune incidence sur le sens de la phrase ou du paragraphe, mais pourrait être considéré comme important si un traducteur voulait mettre l'emphase sur les sources de l'information donnée par cette phrase.

5.1.2.1.3 Comparaison prompt 1 avec prompt 4

Lien vers la feuille Google Sheets contenant la comparaison prompt 1 avec prompt 4

Pour ce prompt, qui est le plus détaillé en termes de la signification de la théorie du *skopos* (d'où sa longueur), les changements sont similaires à ceux de tous les autres prompts (remplacement par des synonymes, léger changement de syntaxe). Comme pour la traduction du prompt 3, ChatGPT a changé l'emplacement de la proposition *« Selon des sources citées par Bloomberg News »*, pour la placer au début de la phrase au lieu de sa fin. Le titre a aussi été changé de « *Moment décisif pour Tesla alors que la visite d'Elon Musk en Chine récolte rapidement des récompenses* » à « *Un moment décisif pour Tesla lors de la visite d'Elon Musk en Chine qui récolte rapidement des récompenses* ». Nous avons aussi repéré d'autres changements comme le remplacement du complément de nom « *de cartographie* » par l'adjectif « *cartographiques* », et le remplacement de l'expression « *avancée majeure* » par l'expression « *étape importante* ».

À noter que dans le résultat généré par ChatGPT après lui avoir donné ce prompt, le logiciel a donné l'explication suivante : *« Selon la théorie du skopos, l'action traductive doit être déterminée par l'objectif du texte, ou le skopos. La traduction n'est pas simplement la*

transmission d'un sens unique du texte source au texte cible, mais plutôt une offre d'information dans une culture source qui est transformée en une offre d'information dans une culture cible. Ainsi, la fonction du texte cible peut différer de celle du texte source, et la cohérence intratextuelle et extratextuelle est essentielle. Dans cette optique, voici la traduction du texte journalistique, en respectant ces principes : ».

5.1.2.1.4 Comparaison prompt 1 avec prompt 5

Lien vers la feuille Google Sheets contenant la comparaison prompt 1 avec prompt 5

Les changements dans les résultats de ce prompt sont toujours similaires à ceux des traductions précédentes, comme les légers changements de syntaxe et les remplacements de mots par des synonymes. Le titre dans cette traduction, qui est « *Moment décisif pour Tesla lors de la visite d'Elon Musk en Chine qui récolte rapidement des récompenses* » est resté presque le même que dans la traduction du prompt 1 qui est « *Moment décisif pour Tesla alors que la visite d'Elon Musk en Chine récolte rapidement des récompenses* ». ChatGPT a remplacé « *alors que* » par « *lors de* », ce qui change la nuance sémantique, et a ajouté « *qui* » avant le verbe « *récolte* ». Comme dans les traductions des prompts précédents, ChatGPT a remplacé le complément de nom « *de cartographie* » par l'adjectif « *cartographiques* ». Un autre changement qu'il est important de noter est le fait que dans cette traduction, et contrairement à la traduction avec le prompt 1, ChatGPT a considéré que la fonctionnalité « *FSD* » est au masculin, en utilisant « *le* » et « *il* » pour y faire référence.

Comme pour le prompt 4 précédent, ChatGPT a aussi donné une explication avant de donner la traduction : *« Dans le respect des règles de la théorie du skopos, voici la traduction du texte journalistique en français : ».*

5.1.2.1.5 Comparaison prompt 1 avec prompt 6

Lien vers la feuille Google Sheets contenant la comparaison prompt 1 avec prompt 6

Dans ce texte, même si les changements sont plus ou moins similaires en type (synonymes, changement de syntaxe), ils sont très dispersés dans le texte. Un seul changement mérite d'être noté pour les objectifs de ce mémoire : en traduisant des valeurs monétaires en dollars américains et en sterling, ChatGPT a remplacé la valeur mentionnée en sterling par celle en euro, ce qui pourrait être considéré comme un essai (même s'il est erroné) d'adaptation à l'audience cible française, que nous n'avions toujours pas précisé à ce stade dans le prompt. En effet, la traduction de la phrase « *It costs $8,000, or $99 (£80) a month* » est « *Il coûte 8 000 $, ou 99 £ (80 €) par mois* ».

À noter que pour ce prompt aussi, ChatGPT a donné une explication avant la traduction : *« Dans le respect des règles de la théorie du skopos, voici la traduction du texte journalistique en français avec une fonction explicative : ».*

5.1.2.1.6 Comparaison prompt 1 avec prompt 7journ

Lien vers la feuille Google Sheets contenant la comparaison prompt 1 avec prompt 7journ

Ce prompt est celui que nous avons utilisé pour traduire le texte journalistique, mais cette fois-ci avec un mandat de traduction proche d'un mandat de traduction réel, qui précise des éléments pratiques à respecter qui pourraient avoir une influence sur le style de rédaction dans la traduction. Pour ce qui est des changements, ChatGPT a commencé par ajouter l'article « *un* » au début du titre et à mettre deux points avant d'expliquer quel est le moment décisif pour Tesla. Il a aussi remplacé l'expression « *récolte rapidement des récompenses* » par l'expression « *porte rapidement ses fruits* ». Dans la première phrase, il a remplacé le complément « *de cartographie* » par l'adjectif « *cartographiques* » pour désigner les données. Comme dans le titre, ChatGPT a aussi remplacé l'expression « *aurait rapporté des récompenses immédiates* » par l'expression « *aurait déjà porté ses fruits* ». Dans la même phrase, il a remplacé l'expression « *une avancée majeure* » par l'expression « *une étape importante* ». ChatGPT a aussi ajouté

dans cette traduction le nom « *Elon* » dans le quatrième paragraphe, alors qu'il n'avait gardé que le nom « *Musk* » dans la traduction du prompt 1. Dans le paragraphe 5, nous remarquons que ChatGPT n'a pas traduit l'acronyme « *FSD* » (*full self-driving*), mais l'a gardé en anglais entre des guillemets. Il est important aussi de noter que dans le paragraphe suivant, il fait référence à la fonctionnalité « *FSD* » au masculin en utilisant « *le* » et « *il* », alors que dans la traduction avec le prompt 1, il y faisait référence au féminin, en utilisant « *elle* » et « *la* ». Toujours dans le même paragraphe, et contrairement à la traduction avec le prompt 1, ChatGPT a bien traduit les devises monétaires, en gardant la devise de dollar américains pour les deux valeurs. Il a aussi fait des remplacements de mots par leurs synonymes dans tout le texte, comme le remplacement de « *à l'encontre* » par « *envers* » dans la proposition « *ses critiques virulentes envers le président américain* », et le remplacement du mot « *politicien* » par le mot « *homme* » dans la phrase « *le deuxième homme le plus puissant de Chine* ».

5.1.2.2 Texte publicitaire

Comme pour le texte journalistique, et pour faciliter la compréhension et la lisibilité de notre analyse, nous avons présenté dans le tableau 6 ci-dessous la traduction du texte publicitaire avec le prompt 1. Nous rappelons que cette traduction a été utilisée comme traduction « référence » pour cette première évaluation, et que la comparaison effectuée dans l'analyse qualitative et quantitative est par rapport à cette traduction effectuée avec le prompt 1.

# paragraphe	Traduction du texte publicitaire avec le prompt 1
1	Un hôtel emblématique du lac Léman depuis 1834
2	Avec des vues enchanteresses sur le lac et les Alpes enneigées au loin, ainsi que sur la vieille ville, notre hôtel de luxe du lac Léman reste le premier choix des voyageurs sophistiqués et des hommes d'État du monde entier.

3	Venez vous détendre après votre journée avec un traitement dans notre spa sur le toit et créez des liens avec vos collègues autour de la gastronomie italienne au Il Lago étoilé au Michelin avant de vous retirer dans votre chambre conçue par Pierre-Yves Rochon pour une bonne nuit de sommeil. Le Four Seasons Hotel des Bergues Genève mêle un sens revitalisé de l'histoire à un service personnel chaleureux et authentique en plein cœur de la ville.
4	Suite Royale
5	Évoquant le style riche et élégant de Versailles, notre Suite Royale impressionne à chaque tournant, avec de superbes reproductions françaises et de grandes fenêtres donnant sur le lac Léman.
6	Suite Mont Blanc
7	Ouvrez vos fenêtres sur les vues du célèbre Jet d'Eau du lac Léman depuis ces grandes suites, dotées d'un espace de vie richement conçu et d'une chambre isolée.
8	Suite Présidentielle – Genève
9	Notre suite récemment rénovée, nommée d'après les vues qu'elle offre, est un espace chic et contemporain où s'installer, avec des finitions contrastées du Vieux Monde.

Tableau 6. Traduction du texte publicitaire avec le prompt 1

Pour le texte de type publicitaire, et en comparant les résultats des 6 autres prompts au prompt 1 (**« traduis ce texte en français »**), nous pouvons constater les résultats suivants :

5.1.2.2.1 Comparaison prompt 1 avec prompt 2

Lien vers la feuille Google Sheets contenant la comparaison prompt 1 avec prompt 2

Pour cette traduction du texte publicitaire avec le prompt 2, nous pouvons noter que les changements sont restreints au début du texte (titre et deux premiers paragraphes). Dans ce prompt, il y a eu une transformation du titre en une phrase complète, avec l'ajout d'un verbe, de connecteur de temps, ainsi que d'une explication qui ne paraissait pas dans la traduction du premier prompt. En effet, le titre de la traduction du prompt 1 était « *Un hôtel emblématique du lac Léman depuis 1834* » et est devenu dans la traduction du prompt 2 « *Depuis 1834, cet hôtel emblématique du lac Léman offre une expérience exceptionnelle* ». Le reste des changements sont simplement des remplacements de mots par leurs synonymes, comme le remplacement du verbe « *reste* » par le verbe « *demeure* » dans le paragraphe 2. ChatGPT a aussi remplacé l'expression « *en plein cœur de la ville* » par l'expression « *au cœur même de la ville* » dans le

paragraphe 3. ChatGPT a aussi ajouté l'adjectif « *majestueuses* » en référence aux Alpes enneigées dans le paragraphe 2 ; adjectif qui n'était pas présent dans la traduction du prompt 1.

5.1.2.2.2 Comparaison prompt 1 avec prompt 3

Lien vers la feuille Google Sheets contenant la comparaison prompt 1 avec prompt 3

Les changements dans cette traduction sont identiques à ceux constatés dans la traduction du prompt 2 ans la partie 5.1.2.2.1 précédente, avec la même transformation du titre de « *Un hôtel emblématique du lac Léman depuis 1834* » pour la traduction du prompt 1 à « *Depuis 1834, cet hôtel emblématique du lac Léman offre une expérience exceptionnelle »* pour la traduction du prompt 3, l'ajout de l'adjectif « *majestueuses* » et le remplacement du verbe « *reste* » par le verbe « *demeure* » dans le paragraphe 2, et le remplacement de l'expression « *en plein cœur de la ville* » par l'expression « *au cœur même de la ville* » dans le paragraphe 3.

5.1.2.2.3 Comparaison prompt 1 avec prompt 4

Lien vers la feuille Google Sheets contenant la comparaison prompt 1 avec prompt 4

Toujours restreints aux deux premiers paragraphes, les changements sont similaires à ceux constatés dans les traductions des prompts 2 et 3. Pour cette traduction, ChatGPT n'a pas changé le titre, il est donc resté comme suit : « *Un hôtel emblématique du lac Léman depuis 1834* » pour la traduction avec le prompt 1 et celle avec le prompt 4. Au début du paragraphe 2, il a remplacé l'expression « *avec des vues enchanteresses* » par l'expression « *offrant une vue enchanteuse* ». Il est important de noter que ChatGPT n'a pas traduit l'expression « *Michelin-starred* » figurant dans le paragraphe 3, en la laissant simplement en anglais ; une faute qu'il n'avait pas commise jusque-là.

5.1.2.2.4 Comparaison prompt 1 à prompt 5

Lien vers la feuille Google Sheets contenant la comparaison prompt 1 avec prompt 5

Le résultat de ce prompt est identique à celui du prompt 4 analysé dans la partie 5.1.2.2.3 précédente, avec les mêmes changements et la même erreur. Le titre reste le même « *Un hôtel emblématique du lac Léman depuis 1834* » pour les deux traductions. ChatGPT a fait les mêmes remplacements : au début du paragraphe 2, il a remplacé l'expression « *avec des vues enchanteresses* » par l'expression « *offrant une vue enchanteuse* ». Il a aussi laissé l'expression « *Michelin-starred* » sans qu'elle ne soit traduite.

À noter qu'avant de donner la traduction, ChatGPT a donné l'explication suivante : « *Dans le respect des règles de la théorie du skopos, voici la traduction du texte publicitaire en français :* ».

5.1.2.2.5 Comparaison prompt 1 à prompt 6

Lien vers la feuille Google Sheets contenant la comparaison prompt 1 avec prompt 6

Dans la traduction du prompt 6, le type des changements est à peu près le même que les changements constatés dans les traductions des prompts précédents, à savoir, les remplacements des mots ou expressions par leurs synonymes.

Nous pouvons remarquer, dans le paragraphe 3, quelques choix de traduction qui seraient encouragés dans le cas d'une traduction humaine : l'utilisation de l'expression idiomatique *« tissez des liens »* au lieu de l'utilisation d'un verbe comme *« créer »* ou *« établir »* ; l'ajout de l'expression *« pour profiter »* dans le paragraphe 7 afin de clarifier la phrase ; et le remplacement de la phrase *« finitions contrastées du Vieux Monde »* par la phrase *« finitions contrastées à l'ancienne »* dans le paragraphe 9, qui semble être stylistiquement plus correcte. Il est aussi important de noter que ChatGPT a cette fois-ci traduit l'expression « *Michelin-starred* » par « *étoilé au Michelin* » dans le paragraphe 3.

5.1.2.2.6 Comparaison prompt 1 et prompt 7pub

Lien vers la feuille Google Sheets contenant la comparaison prompt 1 avec prompt 7pub

Le prompt 7pub est le prompt où nous avons rédigé un contexte précis pour le faire respecter à ChatGPT dans sa traduction du texte publicitaire. En effectuant une lecture générale de la traduction produite avec ce prompt, nous constatons qu'il y a beaucoup de changements qui touchent des phrases entières, ce qui n'était pas le cas pour les traductions précédentes. Nous commençons par le titre, où ChatGPT a supprimé l'article « *un* » du début du titre, et a jouté « *au bord du* » avant le mot « *lac* », pour transformer le titre de « *Un hôtel emblématique du lac Léman depuis 1834* » dans la traduction du prompt 1 à « *Hôtel emblématique au bord du lac Léman depuis 1834* ». Ensuite, dans le premier paragraphe, ChatGPT a supprimé l'adjectif « *enchanteresses* » après le mot « *vues* », et a remplacé l'expression « *hommes d'état* » par « *dirigeants* ». Dans le paragraphe suivant, nous notons le premier changement de phrase en entier susmentionné, avec la phrase « *Venez vous détendre après votre journée avec un traitement dans notre spa sur le toit* » dans la traduction du prompt 1, qui est devenue « *Profitez de soins dans notre spa sur le toit* » dans la traduction du prompt 7pub. Dans le même paragraphe, ChatGPT a remplacé l'expression « *créez des liens* » par « rencontrez » avant « *collègues* », et le mot « *gastronomie* » par le mot « *cuisine* ». Toujours dans le même paragraphe, il a remplacé la proposition « *sens revitalisé de l'histoire à un service personnel chaleureux et authentique* » par la proposition « *offre un mélange d'histoire et de service personnalisé* ». À noter que tout comme dans la traduction du prompt 1, ChatGPT n'a pas correctement traduit l'expression « *Michelin-starred* » en parlant du restaurant, et produit l'expression « *étoilé au Michelin* ». Dans le cinquième paragraphe, ChatGPT a presque complètement changé la phrase, en supprimant pas mal de mots. En effet, cette phrase dans la traduction du prompt 1 est comme suit : « *Évoquant le style riche et élégant de Versailles, notre*

Suite Royale impressionne à chaque tournant, avec de superbes reproductions françaises et de grandes fenêtres donnant sur le lac Léman », alors que dans la traduction du prompt 7pub, elle est comme suit : « *Notre Suite Royale, inspirée par le style de Versailles, présente des reproductions françaises et de grandes fenêtres donnant sur le lac Léman* ». Cela est aussi le cas pour la phrase du septième paragraphe, qui est passé de « *Ouvrez vos fenêtres sur les vues du célèbre Jet d'Eau du lac Léman depuis ces grandes suites, dotées d'un espace de vie richement conçu et d'une chambre isolée* » à « *Ces grandes suites offrent des vues sur le Jet d'Eau du lac Léman, avec un salon décoré et une chambre séparée* ». Enfin, dans le dernier paragraphe, ChatGPT a gardé le début de la phrase tel qu'il était dans la traduction du prompt 1, mais a changé la fin qui est passée de « *est un espace chic et contemporain où s'installer, avec des finitions contrastées du Vieux Monde* » à « *est un espace moderne avec des finitions classiques* ».

5.2 Résultats MQM

Dans la partie 4.2.4, nous avons détaillé la procédure méthodologique utilisée dans cette deuxième évaluation, en y décrivant les détails et les prompts utilisés. Dans cette partie, nous exposerons les résultats trouvés de cette évaluation.

Pour rappel, pour cette évaluation, nous avons traduis les deux textes en utilisant le prompt 1, qui est un prompt simple ne précisant pas de mandat, et le prompt 7journ pour le texte journalistique et 7pub pour le texte publicitaire, qui sont les deux prompts qui précisent des mandats spécifiques pour chaque texte.

Afin de comprendre au mieux les résultats MQM, il est important de les analyser sur deux volets : une analyse du nombre d'erreurs par catégories, par texte et par prompt, et ensuite, une analyse des résultats MQM finaux pour chaque texte traduit avec les deux prompts.

La répartition des erreurs par catégorie est représentée dans la figure 6 ci-dessous :

		Texte journalistique		Texte publicitaire	
Error Category	**Error Subcategory**	Prompt 1	Prompt 7journ	Prompt 1	Prompt 7pub
Accuracy	Addition	0	0	0	0
	Omission	0	1	0	2
	Mistanslation	0	0	3	0
	Untranslated text	1	0	0	0
Fluency	Punctuation	0	0	0	0
	Spelling	0	0	0	0
	Grammar	0	0	0	0
	Inconsistency	0	0	0	0
	Link/Cross-reference	2	1	3	1
Terminology	Inappropriate for Context	0	0	0	0
	Inconsistent use	0	0	0	0
Style	Unidiomatic	18	7	14	12
Locale convention	Address format	0	0	0	0
	Currency format	0	0	0	0
	Date format	0	0	0	0
	Name format	0	1	3	2
	Telephone format	0	0	0	0
	Time format	0	0	0	0
Source error		0	0	2	0
Other		1	0	0	0

Figure 6. Répartition des résultats MQM

À partir de la figure 6, nous pouvons constater que les erreurs de styles sont les plus fréquentes pour les deux textes et pour les deux prompts. En effet, le texte journalistique traduit avec le prompt 1, celui ne précisant pas le mandat, contient **18** erreurs dans cette catégorie, et le même texte traduit avec le prompt 7journ contient **7** erreurs de ce type. Le texte publicitaire, quant à lui, contient **14** erreurs avec le prompt 1 et **12** erreurs avec le prompt 7pub.

Comme précisé dans l'explication de l'analyse par la méthode MQM dans la partie 4.2.4, cette méthode d'évaluation permet d'assigner des sévérités aux erreurs repérées. Ces sévérités auront une influence importante sur les résultats finaux, puisqu'elles donnent un poids plus important aux erreurs jugées très sévères (« *major errors* ») par l'évaluatrice. À noter que ce type d'erreur figure très rarement dans les traductions analysées : sur les quatre textes, il n'y a que le texte publicitaire traduit avec le prompt 1 qui contient **3** erreurs considérées comme graves par l'évaluatrice. Pour notre évaluation, nous avons choisi d'assigner la sévérité de 1 pour les

erreurs de petite sévérité (« *minor errors* ») et de 2 pour les erreurs de grande sévérité (« *major errors* »). En d'autres termes, les erreurs graves valent deux erreurs moins graves.

Les autres catégories d'erreurs présentent des nombres négligeables par rapport aux erreurs de style, et ne dépassent jamais **3** erreurs par type de texte et par prompt. Par exemple, en ne prenant en compte que la catégorie de style, nous constatons donc une amélioration entre le prompt 1 et les prompt 7journ et 7pub pour les deux types de texte, avec **11** erreurs de style de moins pour le texte journalistique traduit avec le prompt 7journ, et **4** erreurs de style de moins pour le texte publicitaire traduit avec le prompt 7pub.

Comme mentionné, la deuxième étape de notre analyse se repose sur les scores MQM finaux, représentés dans le tableau 7 ci-dessous :

Texte	Score MQM
Texte journalistique – prompt 1	**0,06044**
Texte journalistique – prompt 7journ	**0,02747**
Texte publicitaire – prompt 1	**0,08242**
Texte publicitaire – prompt 7pub	**0,04670**

Tableau 7. Résultats MQM

Ces scores finaux représentent la qualité globale de chaque traduction, et peuvent être répartis par type d'erreurs et par sévérité d'erreur, comme expliqué dans le chapitre 4.

Ces scores sont calculés de telle sorte à arriver à un nombre représentant le taux d'erreurs par mot, selon la formule suivante :

$$P = (Issues_{minor} \times 1 + Issues_{major} \times 2) \div Word\ count$$

Où :

- *P* (pour « *penalty* » en anglais) représente le nombre d'erreurs par mot ;

- $Issues_{minor}$ représente le nombre total d'erreurs de petite sévérité dans toutes les catégories ;
- *1* représente la pénalité attribuée aux erreurs de petite sévérité ;
- $Issues_{minor}$ représente le nombre total d'erreurs de grande sévérité dans toutes les catégories ;
- *2* représente la pénalité attribuée aux erreurs de grande sévérité ;
- *Word count* représente le nombre total de mots dans le texte évalué.

Cette formule a été adapté de la formule représentée dans la figure 7 ci-dessous afin de mieux satisfaire les buts de notre recherche :

$$P = (Issues_{minor} + Issues_{major} \times 5 + Issues_{critical} \times 10) \div Word\ count$$

Figure 7. Formule utilisée pour calculer le score final MQM, adapté de Lommel et al, page 6

À noter que la formule dans la figure 7 a été donnée à titre d'exemple dans l'article référencé, et que les pénalités peuvent être changées – comme cela a été fait pour calculer les scores MQM dans notre mémoire.

Pour ce qui est des scores finaux, nous pouvons constater qu'il y a une amélioration entre la traduction du texte journalistique traduit avec le prompt 1 – où P = **0,06044**, et le même texte traduit avec le prompt 7journ – où P = **0,02747**, et idem pour le texte publicitaire traduit avec le prompt 1 – où P = **0,08242**, et avec le prompt 7pub, où P = **0,04670**. Cela signifie que ChatGPT a fait moins d'erreurs avec les prompts 7journ et 7pub, ceux qui contiennent un mandat plus précis, qu'avec le prompt 1, le prompt simple qui ne contient pas de mandat précis, et ce pour la traduction des deux types de texte.

5.3 Test suites

La troisième méthode d'évaluation choisie est celle des test suites. Pour rappel, nous avons traduit les phrases du test suites (dont les détails se trouvent dans la partie 4.2.5) avec ChatGPT à deux reprises :

- Une première fois en utilisant le prompt « a » suivant : « traduis cette phrase en français » ;
- Une deuxième fois en utilisant le prompt « b » suivant : « traduis cette phrase en français en traduisant ou adaptant les acronymes ».

Cette méthode – la comparaison de la traduction de ChatGPT entre deux prompts différents – nous a semblé intéressante car dans la première étape de notre méthodologie, où nous avons fait un exercice similaire, non pas avec un test suite mais avec deux textes différents, nous avons constaté que ChatGPT n'a traduit l'acronyme qui figurait dans le texte journalistique, à savoir, la « *FSD* » (ou *full self-driving* en anglais), que pour le prompt 1, celui le plus simple et ne précisant pas le mandat de traduction. Nous avons donc voulu approfondir la question davantage en faisant ce test-suite, afin d'éloigner tout biais ou hasard.

Nous avons ensuite procédé à une évaluation simple de chaque phrase et de sa traduction, une par une, dans un tableau Microsoft Excel, et nous avons choisi pour chaque phrase une des deux valeurs suivantes : « oui » pour les fois où ChatGPT a traduit les acronymes, et « non » pour les fois où cela n'a pas été fait.

Les résultats de cet exercice sont présentés dans le tableau 8 ci-dessous :

	Traduction avec prompt « a »	Traduction avec prompt « b »
Nombre de fois où les acronymes ont été traduits	**34**/78**,** ou **43%**	**37**/78, ou **47%**
Nombre de fois où les acronymes n'ont pas été traduits	**44**/78, ou **56%**	**41**/78, ou **53%**

Tableau 8. Résultats de la traduction des phrases du test-suites avec le prompt « a » et le prompt « b »

Comme nous pouvons le constater, le nombre d'acronymes traduits par ChatGPT avec le prompt « a » est de **34,** ou **43%**, et celui avec le prompt « b » est de **37,** ou **47%**, alors que le

nombre d'acronymes non traduits par ChatGPT avec le prompt « a » est de **44,** ou **56%,** et celui avec le prompt « b » est de **41,** ou **53%**.

Ces résultats montrent une très légère amélioration lorsque le prompt « b », un prompt plus précis et qui demande explicitement à ChatGPT de traduire les acronymes, est utilisé. Cependant, les résultats étant très proches de la moitié des occurrences des acronymes (**78 ÷ 2 = 39**), nous pouvons considérer que ChatGPT traduit les acronymes presque une fois sur deux, et ce même quand cela est explicitement demandé dans le prompt.

Pour montrer un extrait du test suites, nous avons représenté dans les deux tableaux 9 et 10 ci-dessous quelques exemples de phrases où les acronymes ont été traduits et de phrases où les acronymes n'ont pas été traduits.

Phrase source	This report – the first programme implementation report prepared under my leadership – summarizes two years of **ILO** performance
Résultat prompt « a »	Ce rapport – le premier rapport d'exécution du programme préparé sous ma direction – résume deux années de performance de **l'OIT**.
Résultat prompt « b »	Ce rapport – le premier rapport de mise en œuvre du programme préparé sous ma direction – résume deux ans de performance de **l'OIT**.
Phrase source	Elements of Friday's global **IT** outage, which grounded planes and hit services from banking to healthcare, have occurred before and until more contingencies are built into networks, and organisations put better back-up plans in place, it will happen again.
Résultat prompt « a »	Des éléments de la panne mondiale des **systèmes informatiques** de vendredi, qui a cloué les avions au sol et impacté des services allant de la banque aux soins de santé, se sont déjà produits auparavant et, tant que davantage de mesures de contingence ne seront pas intégrées dans les réseaux et que les organisations ne mettront pas en place de meilleurs plans de sauvegarde, cela se reproduira.
Résultat prompt « b »	Des éléments de la panne **informatique** mondiale de vendredi, qui a cloué les avions au sol et perturbé des services allant de la banque aux soins de santé, se sont déjà produits auparavant et, tant que plus de mesures de contingence ne seront pas intégrées dans les réseaux et que les organisations ne mettront pas en place de meilleurs plans de secours, cela se reproduira.
Phrase source	As in many neurodevelopmental disorders, comorbidities are more the rule than the exception. Therefore, **ASD** often comes with comorbidities.
Résultat prompt « a »	Comme dans de nombreux troubles neurodéveloppementaux, les comorbidités sont plus la norme que l'exception. Par conséquent, les **troubles du spectre autistique** sont souvent associés à des comorbidités.
Résultat prompt « b »	Comme dans de nombreux troubles neurodéveloppementaux, les comorbidités sont plus la règle que l'exception. Par conséquent, le **TSA** est souvent accompagné de comorbidités.

Phrase source	#**DYK** that loggerheads are the most common nesting turtle that we document in Broward County?
Résultat prompt « a »	**Savez-vous que** les tortues caouannes sont les tortues de nidification les plus courantes que nous documentons dans le comté de Broward ?
Résultat prompt « b »	#**SAVIEZVOUS** que les tortues caouannes sont les tortues nidificatrices les plus courantes que nous documentons dans le comté de Broward ?

Tableau 9. Exemples de phrases du test suites où les acronymes ont été traduits

Phrase source	The **SDS-WAS** integrates research and user communities (e.g. medical, aeronautical, agricultural users).
Résultat prompt « a »	Le **SDS-WAS** intègre les communautés de recherche et d'utilisateurs (par exemple, les utilisateurs médicaux, aéronautiques, agricoles).
Résultat prompt « b »	Le **SDS-WAS** intègre les communautés de recherche et d'utilisateurs (par exemple, les utilisateurs médicaux, aéronautiques et agricoles).
Phrase source	To take a country of nearly 170 million people off the Internet is a drastic step, one we haven't seen since the likes of the Egyptian revolution of 2011," said John Heidemann, chief scientist of the networking and cybersecurity division at **USC** Viterbi's Information Sciences Institute.
Résultat prompt « a »	Déconnecter un pays de près de 170 millions de personnes d'Internet est une mesure drastique, que nous n'avons pas vue depuis des événements comme la révolution égyptienne de 2011 », a déclaré John Heidemann, scientifique en chef de la division des réseaux et de la cybersécurité à l'Institut des sciences de l'information de **USC** Viterbi.
Résultat prompt « b »	Couper l'Internet à un pays de près de 170 millions d'habitants est une mesure radicale, une chose que nous n'avons pas vue depuis la révolution égyptienne de 2011 », a déclaré John Heidemann, scientifique en chef de la division des réseaux et de la cybersécurité à l'Institut des sciences de l'information de l'**USC** Viterbi.
Phrase source	At the **AMEC** Awards 2022 – accolades that recognise excellence in communications measurement – Access Intelligence won five awards and was nominated for a further six.
Résultat prompt « a »	Lors des **AMEC** Awards 2022 – des distinctions qui reconnaissent l'excellence dans la mesure des communications – Access Intelligence a remporté cinq prix et a été nominée pour six autres.
Résultat prompt « b »	Lors des **AMEC** Awards 2022 – des récompenses qui reconnaissent l'excellence dans la mesure des communications – Access Intelligence a remporté cinq prix et a été nominé pour six autres.
Phrase source	**TENS** delivers a low-frequency, low-voltage repetitive stimulus of approximately 8 to 12 mA for 500 milliseconds (ms).
Résultat prompt « a »	Le **TENS** délivre un stimulus répétitif à basse fréquence et basse tension d'environ 8 à 12 mA pendant 500 millisecondes (ms).
Résultat prompt « b »	La **TENS** délivre un stimulus répétitif de basse fréquence et basse tension d'approximativement 8 à 12 mA pendant 500 millisecondes (ms).

Tableau 10. Exemples de phrases du test suites où les acronymes n'ont pas été traduits

5.4 Synthèse et discussion

Dans cette partie, nous ferons une synthèse des résultats que nous avons eu avec nos trois méthodes d'évaluation, (analyse du nombre et types de modifications, évaluation MQM et test-suites), et nous essaierons d'en tirer des conclusions.

La première évaluation, qui la comparaison des changements dans les traductions en fonction de plusieurs prompts et par rapport à un prompt 1 simple nous a montré que lorsque le prompt contient un mandat clair et précisant un contexte, cela a une influence sur la traduction produite, qui se manifeste par des changements par rapport à une traduction produite avec un prompt simple. Nous pouvons remarquer cela avec le nombre total de changements introduits dans les traductions des prompts 7journ et 7pub, qui sont plus élevés par rapport aux autres totaux. Nous avons aussi remarqué que pour le texte publicitaire, le nombre total de suppressions pour la traduction produite avec le prompt 7pub est de **37,** et nous pensons que cela pourrait être lié au faut que dans notre mandat, nous avons demandé à ChatGPT de « simplifier » le texte en utilisant moins de figures de style. Nous pensons donc que ces suppressions pourraient refléter cette instruction, et que ChaGPT a voulu rendre les phrases moins longues.

Il est aussi important de noter que les prompts que nous avons utilisés qui sont plutôt vagues, qui donnent un mandat non précis (donc les prompts de 2 à 6) ont entraîné des modifications dans leurs traductions, mais en effectuant l'analyse qualitative, nous avons constaté qu'elles n'étaient pas conséquentes, et n'influençaient pas énormément la fonction globale du texte. En effet, le type de changement qui a été introduit le plus dans les traductions des deux types de texte pour ces prompts sont le remplacement de mots ou d'expressions par leurs synonymes, ce qui a été reflété aussi par l'analyse quantitative et par le nombre élevé de substitutions.

La deuxième évaluation, qui était l'évaluation MQM a été conçue dans le cadre de notre mémoire afin d'évaluer si les traductions résultant de prompts contenant un mandat de

traduction précis étaient des traductions de meilleure qualité. Les résultats de cette évaluation ont montré que les traductions faites avec des prompts clairs, précisant le mandat de traduction, contiennent moins d'erreurs par mot que celles qui sont produites avec un prompt simple, qui demande uniquement à ChatGPT une traduction sans préciser le contexte ni le mandat. Nous pensons important de noter que non seulement les traductions résultant d'un prompt avec un mandat précis contiennent moins d'erreurs, mais elles contiennent surtout moins d'erreurs de style. La catégorie « style » une catégorie qui comprend les parties de la traduction considérées non-idiomatiques, qu'un traducteur professionnel dont la langue maternelle est la même que la langue cible n'utiliserait pas dans sa traduction, puisqu'elle ne respecte pas les expressions idiomatiques couramment utilisées dans la langue. Les standards MQM la définisse comme suit :

> « Appropriateness of a text with respect to general language and any special language requirements with respect to lexical and usage conventions, as well as compliance with the organizational style guide component of the translation specifications »[7].

Il est aussi intéressant de noter que l'amélioration constatée avec le texte publicitaire est plus importante que celle avec le texte journalistique. En effet, comme mentionné dans la partie 5.2, l'évaluatrice a repéré 3 erreurs graves dans la traduction du texte publicitaire faite avec le prompt 1, mais cela n'était pas le cas dans la traduction du texte publicitaire faite avec le prompt 7pub. Nous pensons que cela est dû au fait que le mandat que nous avons précisé dans le prompt 7pub est plus clair, car il contient des instructions plus concrètes concernant les changements à introduire dans le texte, comme l'élimination des figures de styles et la simplification des phrases.

[7] https://themqm.org/mqm-terminology/

Enfin, la troisième évaluation est représentée par le test suites, effectué avec **50** phrases contenant **78** occurrences d'acronymes. Nous avons traduit ces phrases une par une à deux reprises, une première fois avec un prompt simple, et une deuxième fois avec un prompt qui demande explicitement qu'il faut traduire les acronymes. Les résultats de cette analyse confirment ceux des deux méthodes d'évaluation précédentes, car elle a montré que ChatGPT a traduit les acronymes à une fréquence plus élevée avec le prompt explicitant cette instruction qu'avec le prompt ne le précisant pas. Cependant, il est important de noter que malgré cela, les résultats montrent des occurrences où même quand on le lui demande explicitement, ChatGPT ne traduit pas les acronymes.

Les résultats des trois évaluations nous montreraient l'importance de la rédaction d'un prompt aussi détaillé et explicite que possible, afin d'arriver à la traduction la plus proche de nos besoins (Yamada, 2023).

Nous précisons aussi que même si un prompt contenant un mandat précis entraîne une amélioration dans la traduction, cela ne veut pas dire que la traduction est parfaite, ni qu'elle respecte le mandat à 100% : en effet, les résultats obtenus surtout avec l'évaluation MQM et le test suites montrent que ChatGPT commet des erreurs de traductions et ne respectent pas toujours le mandat demandé.

5.5 Conclusion

Dans ce chapitre, nous avons exposé et analysé les résultats des trois méthodes d'évaluation utilisées dans la méthodologie. La première évaluation consistait à comparer des modifications dans des traductions en fonction de prompts différents par rapport à un prompt simple. Au terme de cette évaluation, nous avons voulu vérifier si un mandat précis, ressemblant à un mandat réel de traduction produirait des traductions différentes que celles produites avec un prompt simple, ne précisant pas le mandat. Nous avons constaté, grâce aux deux analyses quantitative et

qualitative, que les traductions effectuées à l'aide d'un prompt précis présentaient des modifications par rapport à celle produites avec un prompt simple. Les prompts les plus précis, à savoir le prompt 7journ et le prompt 7pub, ont engendré le plus grand nombre de modifications dans les traductions. La deuxième évaluation consistait à l'évaluation par une experte de traductions produites à l'aide d'un prompt simple et de prompts contenant un mandat spécifique pour chaque type de texte, selon les standards de la méthode MQM. Les résultats ont montré que ChatGPT commet moins d'erreurs – donc, sa traduction est de meilleure qualité – avec les prompts précisant le mandat. Enfin, la troisième évaluation consistait en la traduction d'une série de phrases contenant toutes des acronymes, une fois avec un prompt simple et une fois avec un prompt demandant explicitement la traduction des acronymes. Au terme de cette analyse, nous avons constaté qu'il est légèrement plus probable que ChatGPT traduise les acronymes quand cela est explicitement demandé que quand cela n'est pas demandé.

6 Limites et perspectives

Bien que ces conclusions soient intéressantes, nous reconnaissons les limites de ce mémoire, et nous nous demandons quelles potentielles perspectives notre travail pourrait avoir.

Durant la conception de la méthodologie de ce mémoire, nous avons fait face à quelques obstacles conceptuels, surtout en rapport avec quelles méthodes d'évaluation choisir en fonction des objectifs que nous voulions atteindre et des questions de recherche auxquelles nous aimions répondre. Nous avons choisi d'utiliser la méthode d'évaluation MQM dans ce mémoire, mais nous pensons que de meilleures méthodes d'évaluation pourraient être utilisées – ou conçues – pour mener à bien un travail dans la même lignée que le nôtre.

Notre analyse qualitative (partie 5.1.2) a montré que ChatGPT introduisait des changements aux traductions quand les prompts étaient plus précis, mais la nature de ces changements n'était

pas toujours concluante : la majorité du temps, ces modifications étaient des remplacements de mots par leurs synonymes, ce qui ne rend pas nécessairement la traduction plus adaptée à un mandat. Nous avons surtout pu constater cela grâce à l'analyse quantitative, où le nombre de substituions était toujours le plus élevé. Cela n'a pas pu nous dire si la traduction était adaptée au mandat ou pas, et nous pensons qu'il serait intéressant d'étudier cela en utilisant des méthodes d'évaluation différentes de celles que nous avons utilisé dans ce mémoire.

De plus, notre étude s'est limitée à deux types de texte – journalistique et publicitaire – ce qui en soi limite la versatilité des résultats. Cela est lié à la taille du mémoire et aux délais de temps, mais il serait intéressant d'étudier la traduction de ChatGPT avec d'autres types de textes, comme des textes juridiques, institutionnels ou scientifiques, qui pourraient relever des problématiques différentes que celles de notre mémoire. Nous notons surtout que le texte publicitaire a été plus influencé que le texte journalistique par le prompt contenant un mandat précis, et nous pensons que cela est une variable qui mériterait d'être étudiée plus en détail.

Dans ce mémoire, nous nous sommes aussi limités à une seule combinaison linguistique, à savoir de l'anglais vers du français, et nous pensons qu'il serait intéressant d'évaluer la performance de ChatGPT avec d'autres langues, et surtout avec une combinaison linguistique qui n'inclue pas l'anglais, puisque c'est la langue sur laquelle ChatGPT a été le plus entraîné.

Nous pensons aussi qu'il serait intéressant de comparer les traductions obtenues avec un prompt contenant un mandat précis avec les traductions que pourraient produire des systèmes de traduction neuronaux, comme *Google Translate* ou *DeepL*, surtout par rapport à la qualité de ces traductions.

Nous pensons que les résultats trouvés et les questions de recherche sont intéressants, et pourraient aboutir à des ouvertures pertinentes dans le domaine. Par exemple, il serait intéressant de pouvoir se concentrer sur le côté technique de l'architecture des *transformers*

(expliqué dans la partie 2.1.1), et d'étudier quel type d'architecture serait le mieux adapté à des tâches de traduction, et encore plus précisément, à une la traduction avec des combinaisons linguistiques différentes.

Enfin, nous pensons qu'il est important de prendre en compte l'influence des prompts (ou instructions) dans lesquels nous demandons à ChatGPT de traduire un texte, et la manière avec laquelle ils sont rédigés. Contrairement à d'autres systèmes automatiques de traduction neuronaux (comme *Google Translate* ou *DeepL*), ChatGPT suit des instructions données par l'utilisateur pour effectuer une tâche ; ces tâches ne sont pas limitées à la traduction, contrairement aux autres systèmes cités. La problématique de la rédaction des prompts mériterait d'être étudié pleinement, car c'est un nouveau facteur qui vient s'insérer dans les variables à étudier, et est l'objet de plus en plus d'études (Bawden et al., 2023). En effet, nous pensons que cette nouvelle variable exerce une influence importante sur le texte produit par ChatGPT.

7 Conclusion

Pour conclure ce mémoire, nous reviendrons sur les hypothèses et les questions de recherche que nous avons expliquées au début de ce mémoire, et nous essaierons de juger si les résultats auxquels nous sommes arrivés dans notre méthodologie ont réussi à y répondre.

Q1. Un mandat de traduction précis, comme le prescrit la première règle de la théorie du *skopos*, influence-t-il la traduction produite par ChatGPT ?

Afin de répondre à cette question, nous avons effectué une comparaison des changements dans les traductions de deux textes en fonction de prompts différents : un prompt simple qui ne précise pas le mandat (prompt 1), des prompts qui sont plus détaillés mais qui précisent des mandats vagues (prompt 2 à 6) et des prompts qui contiennent un mandat clair et qui précise le contexte (prompts 7journ et 7pub) pour chaque type de texte. Cette étape était la première dans notre méthodologie, et notre hypothèse était qu'un mandat précis donnerait une traduction différente d'une traduction produite avec un prompt simple. Après l'analyse des résultats, nous avons pu relever des changements que ChatGPT a introduit dans les traductions en fonction des prompts différents. Nous avons aussi pu constater que plus le mandat était précis, plus les changements étaient nombreux : en effet, selon les résultats présentés dans l'analyse quantitative (partie 5.1.1), les prompts 7journ et 7pub ont engendré le plus grand nombre de changements à leurs traductions, avec **76** changements pour le prompt 7journ et **108** changements pour le prompt 7pub. Cela répond donc à notre première question de recherche, et valide notre première hypothèse.

Q2. La précision d'un mandat dans le prompt conduirait-t-elle à la production d'une traduction de meilleure qualité ?

Comme mentionné dans la partie 4.1, la qualité en traduction est un sujet large, et la qualité des traductions de ChatGPT spécifiquement mériterait d'être étudiée plus en détails, en prenant en considération des variables qui n'ont pas été étudiées dans le cadre de ce mémoire. Cependant, nous pensons que cette question est importante, puisque la qualité de la traduction d'un outil en dit long sur la capacité de cet outil à effectuer des tâches de traduction, et, dans notre cas, de s'adapter à un mandat précis, et poserait un point d'entrée sur les attentes que pourrait avoir les utilisateurs de cet outil. Afin de répondre à cette question, nous avons utilisé la méthode d'évaluation MQM pour évaluer les traductions produites par ChatGPT des deux textes à l'aide du prompt 1, qui ne précise pas de mandat, et les traductions produites avec les prompts 7journ et 7pub, qui précisent des mandats. Les résultats de l'évaluation MQM ont montré que les traductions produites avec les prompts précis contiennent moins d'erreurs, et sont donc de meilleure qualité, que celles produites avec le prompt ne précisant pas le mandat, ce qui valide notre deuxième hypothèse.

Q3. ChatGPT respecte-t-il toujours le mandat précisé dans les prompts ?

Pour cette troisième et dernière question de recherche, notre hypothèse était que plus le prompt est précis, plus il y a une probabilité que ChatGPT produise une traduction adaptée à l'objectif voulu. En effet, durant la première évaluation, nous avons remarqué des manques de cohérence liés aux changements introduits par ChatGPT dans les traductions, sans qu'il y ait d'explications claires à ce phénomène. Un exemple de cela est la traduction ou pas d'un acronyme qui figurait dans le texte journalistique (la *FSD*) qui a été traduit quelques fois mais pas d'autres. Afin d'éloigner tout hasard dans nos résultats et d'aiguiser notre recherche, nous avons recouru à un test-suites contenant des acronymes, car nous pensions que cela pourrait nous donner une

réponse claire sur notre questionnement concernant la capacité de ChatGPT à suivre une instruction précise (un mandat) lorsqu'il s'agit de traduction, et la probabilité que la traduction corresponde à l'objectif voulu. Une fois effectué, les résultats du test-suites ont montré que sur **78** occurrences d'acronymes, et en lui demandant explicitement, ChatGPT traduit plus de fois les acronymes que si la demande ne figurait pas dans le prompt. Avec ces résultats, nous considérons que notre troisième hypothèse a été elle aussi validée qu'en partie, car même si des améliorations entre les deux prompts ont été remarquées, la différence entre les deux n'était pas très grande, et surtout, que la fréquence à laquelle ChatGPT a traduit les acronymes était toujours moins élevée que celle à laquelle il ne les a pas traduits, malgré les améliorations constatées.

Nous avons repris les résultats de ce mémoire dans le tableau 11 ci-dessous afin de faciliter la lecture.

Question de recherche	**Hypothèse à vérifier**	**Méthode d'évaluation utilisée**	**Résultats trouvés**	**Validation ou non de l'hypothèse**
Q1. Un mandat de traduction précis, comme le prescrit la première règle de la théorie du *skopos*, influence-t-il la traduction produite par ChatGPT ?	**H1.** La traduction produite avec un prompt précis est différente par rapport à une traduction produite par un prompt ne précisant pas le mandat ou ayant un mandat peu clair	Comparaison des changements dans les traductions en fonction des prompts et analyse quantitative et qualitative	Les deux analyses ont montré qu'un prompt plus précis a donné des traductions avec plus de changements que celles avec un prompt simple, et que les prompts contenant un mandat spécifique ont conduit au plus grand nombre de changements	Hypothèse validée
Q2. La précision d'un mandat dans le prompt conduirait-t-elle à la production d'une traduction de meilleure qualité ?	**H2.** Si le prompt précise un mandat spécifique, la traduction et de meilleure qualité	Évaluation MQM	L'évaluation MQM a montré que les traductions avec un prompt précisant un mandat clair contiennent moins d'erreurs que celles avec un prompt simple	Hypothèse validée

Q3. ChatGPT respecte-t-il toujours le mandat précisé dans les prompts ?	**H3.** Plus le prompt est précis, plus il y a une probabilité que ChatGPT produise une traduction adaptée à nos objectifs	Test-suites avec un phénomène spécifique	Les résultats du test-suites ont montré que ChatGPT a plus tendance à donner une traduction adaptée aux objectifs voulus lorsque le mandat le précise	Hypothèse validée en partie

Tableau 11. Résumé des résultats trouvés pour chaque question de recherche et hypothèse

Pour finir, nous pensons qu'il est important de revenir sur un concept important prescrit par la théorie du *skopos* et mentionné dans la partie 3.2.6 de ce mémoire, et qui est le concept des traductions d'une qualité adaptée à l'objectif (« *fit-for-purpose quality* »). Nous pensons que les évaluations effectuées dans ce mémoire et les résultats qu'elles ont apportés montrent l'importance d'une traduction qui est adéquate par rapport à un objectif. Nous pensons de même que ce concept est un élément important à garder en tête durant l'utilisation d'un outil comme ChatGPT pour des tâches de traduction, que ce soit par un traducteur professionnel, ou un utilisateur désirant traduire un texte, car comme certains de nos résultats le montrent, cela a des probabilités plus élevées de donner un rendu plus qualitatif.

Nous pensons que les résultats trouvés dans notre mémoire liés à la rédaction des prompts et à leur influence sur la qualité des traductions pourraient aider les utilisateurs et les traducteurs qui utilisent ChaGPT à obtenir de meilleurs résultats et d'une manière plus efficace.

Bibliographie

Abdul-Razzaq, H. H. (2008). Translating Acronyms of Media & U.N. Organizations. ALBAHITH ALALAMI, 1(4), Article 4. https://doi.org/10.33282/abaa.v1i4.466

Avramidis, E., Macketanz, V., Strohriegel, U., & Uszkoreit, H. (2019). Linguistic Evaluation of German-English Machine Translation Using a Test Suite. Proceedings of the Fourth Conference on Machine Translation (Volume 2: Shared Task Papers, Day 1), 445-454. https://doi.org/10.18653/v1/W19-5351

Baker, M., & Chesterman, A. (1998). "Machine Translation." In M. Baker (Ed.), Routledge Encyclopedia of Translation Studies (pp. 156-160). London: Routledge.

Bawden, R., & Yvon, F. (2023). *Investigating the Translation Performance of a Large Multilingual Language Model : The Case of BLOOM* (arXiv:2303.01911). arXiv. http://arxiv.org/abs/2303.01911

Baziotis, C., Mathur, P., & Hasler, E. (2023). Automatic Evaluation and Analysis of Idioms in Neural Machine Translation. Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics, 3682-3700. https://doi.org/10.18653/v1/2023.eacl-main.267

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020a). *Language Models are Few-Shot Learners* (arXiv:2005.14165). arXiv. https://doi.org/10.48550/arXiv.2005.14165

Burchardt, A., Macketanz, V., Dehdari, J., Heigold, G., Peter, J.-T., & Williams, P. (2017). A Linguistic Evaluation of Rule-Based, Phrase-Based, and Neural MT Engines. The Prague

Bulletin of Mathematical Linguistics, 108(1), 159-170. https://doi.org/10.1515/pralin-2017-0017

Callison-Burch, C. (2009). Fast, cheap, and creative: Evaluating translation quality using Amazon's Mechanical Turk. Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing Volume 1 - EMNLP '09, 1, 286. https://doi.org/10.3115/1699510.1699548

Chen, M., Tworek, J., Jun, H., Yuan, Q., Pinto, H. P. de O., Kaplan, J., Edwards, H., Burda, Y., Joseph, N., Brockman, G., Ray, A., Puri, R., Krueger, G., Petrov, M., Khlaaf, H., Sastry, G., Mishkin, P., Chan, B., Gray, S., … Zaremba, W. (2021). *Evaluating Large Language Models Trained on Code* (arXiv:2107.03374). arXiv. https://doi.org/10.48550/arXiv.2107.03374

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT : Pre-training of Deep Bidirectional Transformers for Language Understanding* (arXiv:1810.04805). arXiv. https://doi.org/10.48550/arXiv.1810.04805

Du, X. (2012). A Brief Introduction of *Skopos* Theory. *Theory and Practice in Language Studies*, *2*(10). https://www.academia.edu/6154098/A_Brief_Introduction_of_*Skopos*_Theory

Fields, P., Hague, D. R., Koby, G. S., Lommel, A., & Melby, A. (2014). What Is Quality? A Management Discipline and the Translation Industry Get Acquainted. Tradumàtica Tecnologies de La Traducció, 12, 404-412. https://doi.org/10.5565/rev/tradumatica.75

Gheini, M., Ren, X., & May, J. (2021). Cross-Attention is All You Need : Adapting Pretrained Transformers for Machine Translation. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 1754-1765. https://doi.org/10.18653/v1/2021.emnlp-main.132

Hiebl, B., & Gromann, D. (s. d.). Quality in Human and Machine Translation : An Interdisciplinary Survey.

Holmes, J. S. (1972, août 21). The Name and Nature of Translation Studies. *Third International Congress of Applied Linguistics*, 14.

Improving language understanding with unsupervised learning. (s. d.). Consulté 3 juillet 2024, à l'adresse https://openai.com/index/language-unsupervised/

Jurafsky, D., & Martin, J. H. (2024). *Speech and Language Processing* (3rd éd.).

King, M., & Falkedal, K. (1990). Using test suites in evaluation of machine translation systems. Proceedings of the 13th Conference on Computational Linguistics -, 2, 211-216. https://doi.org/10.3115/997939.997976

Liu, E., Chaudhary, A., & Neubig, G. (2023). Crossing the Threshold : Idiomatic Machine Translation through Retrieval Augmentation and Loss Weighting (arXiv:2310.07081). arXiv. http://arxiv.org/abs/2310.07081

Lommel, A. R., Burchardt, A., & Uszkoreit, H. (s. d.). Multidimensional quality metrics : A flexible system for assessing translation quality.

Lymar, M. Yu. (2021). ENGLISH ABBREVIATIONS AND ACRONYMS IN THE SOCIO-POLITICAL DISCOURSE : TYPES AND APPROACHES IN TRANSLATION. « Scientific Notes of V. I. Vernadsky Taurida National University », Series: « Philology. Journalism », 1(5), 256-260. https://doi.org/10.32838/2710-4656/2021.5-1/44

Munday, J. (2022). *Introducing Translation Studies: Theories and Applications* (5e éd.). Routledge. https://doi.org/10.4324/9780429352461

Nord, C. (2014). *Translating as a Purposeful Activity: Functionalist Approaches Explained*. Routledge. https://doi.org/10.4324/9781315760506

OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao, H., Bavarian, M., Belgum, J., … Zoph, B. (2024). *GPT-4 Technical Report* (arXiv:2303.08774). arXiv. https://doi.org/10.48550/arXiv.2303.08774

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving Language Understanding by Generative Pre-Training. *OpenAI*.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). *Language Models are Unsupervised Multitask Learners.*

Reiss, K., & Vermeer, H. J. (2014). *Towards a General Theory of Translational Action: Skopos Theory Explained*. Routledge.

Snover, M., Dorr, B., Schwartz, R., Micciulla, L., & Makhoul, J. (s. d.). *A Study of Translation Edit Rate with Targeted Human Annotation*.

Tan, R., Long, X., & Bamigbade, O. O. (2023). *Comparative Research on Machine Translation and Human Translation of Examples in Dictionary from the Perspective of Skopos Theory*. 342-353. https://doi.org/10.2991/978-94-6463-230-9_41

Transformer : A Novel Neural Network Architecture for Language Understanding. (s. d.). Consulté 12 août 2024, à l'adresse http://research.google/blog/transformer-a-novel-neural-network-architecture-for-language-understanding/

Unleashing the Power of Generative AI : Understanding Transformers – Parser Digital. (s. d.). Consulté 12 août 2024, à l'adresse https://parserdigital.com/2023/09/21/unleashing-the-power-of-generative-ai-understanding-transformers/

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2023). *Attention Is All You Need* (arXiv:1706.03762). arXiv. https://doi.org/10.48550/arXiv.1706.03762

Wagner, A. C. (2021). *Country, Nation, and Language. Machine Translation in Iceland* [Thesis]. https://skemman.is/handle/1946/38074

Wang, A., Pruksachatkun, Y., Nangia, N., Singh, A., Michael, J., Hill, F., Levy, O., & Bowman, S. R. (2020). *SuperGLUE : A Stickier Benchmark for General-Purpose Language Understanding Systems* (arXiv:1905.00537). arXiv. https://doi.org/10.48550/arXiv.1905.00537

Wang, A., Singh, A., Michael, J., Hill, F., Levy, O., & Bowman, S. R. (2019). *GLUE : A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding* (arXiv:1804.07461). arXiv. https://doi.org/10.48550/arXiv.1804.07461

Wang, S., Zhang, G., Wu, H., Loakman, T., Huang, W., & Lin, C. (2024). MMTE : Corpus and Metrics for Evaluating Machine Translation Quality of Metaphorical Language (arXiv:2406.13698). arXiv. http://arxiv.org/abs/2406.13698

Yamada, M. (s. d.). *Optimizing Machine Translation through Prompt Engineering : An Investigation into ChatGPT's Customizability*.

Ye, J., Chen, X., Xu, N., Zu, C., Shao, Z., Liu, S., Cui, Y., Zhou, Z., Gong, C., Shen, Y., Zhou, J., Chen, S., Gui, T., Zhang, Q., & Huang, X. (2023). *A Comprehensive Capability Analysis of GPT-3 and GPT-3.5 Series Models* (arXiv:2303.10420; Version 1). arXiv. https://doi.org/10.48550/arXiv.2303.10420

Annexes

Annexe 1 – Texte journalistique en anglais

‘Watershed moment’ for Tesla as Elon Musk’s visit to China reaps quick reward

Deal to use mapping data from web search giant Baidu is a big step towards launching driver assistance tech in world’s biggest car market.

Elon Musk’s visit to China has reportedly reaped immediate rewards with a deal for Tesla to use mapping data provided by web search company Baidu, a big step in introducing driver assistance technology in the world’s largest car market.

Musk made an unannounced visit to China over the weekend. The billionaire posted a picture of his meeting with the Chinese premier, Li Qiang, on X, the social network he took over in 2022.

Baidu, which dominates web search in China, will provide mapping and navigation functions to help Tesla operate its driver assistance technology, which it calls “full self-driving”, or FSD, according to sources cited by Bloomberg News. Mapping services – crucial to driver assistance technologies – are strictly controlled by China’s government.

Despite its name, FSD does not provide autonomous driving abilities: it requires a driver who has “hands on the wheel and is prepared to take over at any moment”. However, launching it in China could help Tesla in the fierce competition for market share in the country, and provide more income. It costs $8,000, or $99 (£80) a month, although it is not available in many countries.

Musk is often combative in his dealings with politicians, such as strident criticism of the US president, Joe Biden, or a standoff in Brazil against a government he claims is censoring X, formerly known as Twitter. However, he adopted a more emollient tone towards China’s second most powerful politician, saying he was “honoured” to meet Li.

Annexe 2 – Texte publicitaire en anglais

A landmark Lake Geneva Hotel since 1834

With enchanting views of the lake and the snow-capped Alps in the distance, as well as the Old Town, our Lake Geneva luxury Hotel remains the first choice of sophisticated travellers and world statesmen. Come and unwind from your day with a treatment at our rooftop Spa and build connections with colleagues over Italian gastronomy at Michelin-starred Il Lago before retreating to your Pierre-Yves Rochon–designed room for a good night's sleep. Four Seasons Hotel des Bergues Geneva blends a revitalized sense of history with warm and genuine personal service in the very heart of the city.

Royal Suite

Echoing the rich and elegant style of Versailles, our Royal Suite impresses as every turn, with stunning French reproductions and tall windows overlooking Lake Geneva.

Suite Mont Blanc

Open your windows to views of Lake Geneva's celebrated Jet d'Eau from these large suites, featuring a richly designed living area and secluded bedroom.

Presidential Suite – Genève

Our newly renovated suite, named after the views in which it provides, is a chic and contemporary space to settle into, with contrasting Old-World finishes.

Annexe 3 – Instructions utilisées à l'attention de l'experte en traductologie pour effectuer l'évaluation MQM

Dans le cadre de ce mémoire qui vise à évaluer la traduction du système d'IA générative ChatGPT dans la perspective de la théorie du *skopos*, votre travail consistera à évaluer des textes traduits par ChatGPT, afin de nous permettre d'analyser la qualité de ses traductions.

Plusieurs documents sont à votre disposition à cet effet :

- Le présent document Word, contenant les instructions ;
- Quatre documents Excel, contenant les critères d'évaluation sur lesquels vous devrez vous baser pour effectuer l'évaluation, et que vous devrez remplir ;
- Deux documents Word contenant les deux textes sources en anglais, un journalistique et un publicitaire ;
- Quatre documents Word contenant différentes traductions en français, à raison de deux traductions pour chaque type de texte, obtenues avec des « prompts » différents sur ChatGPT.

Afin d'effectuer l'évaluation, voici les étapes à suivre :

1. Vous familiariser avec la classification des erreurs selon le tableau MQM (Multidimensional Quality Metrics) ;
2. Lire les traductions avec leurs textes sources respectifs ;
3. Identifier les erreurs dans chaque texte ;
4. Copier-coller les segments contenant ces erreurs, ainsi que les segments sources correspondants, dans le tableau associé (par exemple les erreurs du document *Txt_journ_promptA* dans le tableau *Tableau_MQM_journ_promptB*, etc.) ;
5. Spécifier la catégorie de chaque erreur identifiée, sa sous-catégorie (si cela s'applique), ainsi que la sévérité de l'erreur ;
6. Si besoin, ajouter un commentaire pour clarifier vos choix.

Le but de ce mémoire étant une évaluation du texte dans son ensemble, nous vous encourageons, dans la mesure du possible, à privilégier la lecture des traductions de telle sorte à prendre en compte le contexte, et non en lisant chaque segment de phrase comme s'il était isolé.

Annexe 4 – Tableau comparatif des traductions du texte journalistique en fonction des prompts

Lien vers le tableau Google Sheets

Annexe 5 – Tableau comparatif des traductions du texte publicitaire en fonction des prompts

Lien vers le tableau Google Sheets

Annexe 6 – Tableaux de l'évaluation MQM

Lien vers le fichier contenant les 4 tableaux d'évaluation MQM

Annexe 7 – Tableaux contenant le test suites

Lien vers le tableau Google Sheets