



Institut National des Langues et Civilisations Orientales  
Département Textes, Informatique, Multilinguisme

## **Système de traduction automatique neuronale français-mongol Historique, technologies, mise en place et évaluations**

Master  
TRAITEMENT AUTOMATIQUE DES LANGUES

Parcours  
Ingénierie Multilingue

par

**Shuai GAO**

Directeur de mémoire :  
**Damien NOUVEL**

Responsable de stage :  
**Qingyu JIANG**

Année universitaire 2020-2021

# DÉCLARATION SUR L'HONNEUR

Je soussigné, Shuai GAO, déclare avoir rédigé ce travail sans aide extérieure ni sources autres que celles qui sont citées. Toutes les utilisations de textes préexistants, publiés ou non, y compris en version électronique, sont signalées comme telles. Ce travail n'a été soumis à aucun autre jury d'examen sous une forme identique ou similaire, que ce soit en France ou à l'étranger, à l'université ou dans une autre institution, par moi-même ou par autrui.

Fait à Alxa, le 22 novembre 2021

Signature :



# PRÉFACE

*机器翻译是自然语言处理领域皇冠上的明珠。*

*Machine translation is the crown pearl in the field of natural language processing.*

*La traduction automatique est la perle de la couronne dans le domaine du traitement des langues.*

Au début de mes études du Traitement automatique des langues (TAL), cette métaphore résonnait à mes oreilles, en particulier dans le cercle des chercheurs TAL chinois. En effet, la traduction automatique a toujours été le rêve de l'humanité depuis longtemps. Elle constitue également un des objectifs importants de l'intelligence artificielle.

Pendant mon stage chez l'UNESCO-ICHEI, Centre international pour l'innovation dans l'enseignement supérieur sous les auspices de l'UNESCO, une grande partie du travail était de produire du contenu multilingue, à partir duquel créer une mémoire de traduction et de gérer une base terminologique. En tant que Centre de catégorie II de l'UNESCO, l'UNESCO-ICHEI publie ses documents officiels ainsi que des bulletins d'information en six langues officielles des Nations unies. Ces tâches m'ont rendu conscient de l'importance capitale de l'exactitude de la traduction et de l'uniformité des termes. C'est ainsi que je me suis penché de plus près sur la traduction automatique.

L'expérience dans cette organisation internationale m'a donné une plus grande vision sur la technologie, l'éducation et l'humanité. La mission de l'UNESCO-ICHEI est étroitement liée à l'Objectif de développement durable 4 (ODD4) des Nations Unies (Éducation 2030) : « Assurer l'accès de tous à une éducation de qualité, sur un pied d'égalité et promouvoir les possibilités d'apprentissage tout au long de la vie ». La langue étant parfois une barrière infranchissable, les ressources d'apprentissage ne sont pas équilibrées entre les pays. Afin de réaliser l'ODD4, il est indispensable de garantir la transmission des connaissances dans de multiples langues. D'où le rôle central que peut jouer la traduction automatique dans notre société, surtout dans cette ère numérique.

Né en Région autonome de Mongolie intérieure de la Chine, je suis témoin du manque de ressources d'apprentissage et de l'arriération du développement numérique en langue mongole par rapport aux autres langues plus parlées telles que l'anglais, le français, le chinois et l'espagnol, etc., sur lesquelles ont été menées de nombreuses recherches. Après avoir vérifié qu'aucune recherche publiée sur la traduction automatique français-mongol n'existe, j'ai eu l'idée de profiter de mon parcours en TAL pour explorer cette problématique et construire un système de traduction français-mongol. Ceci a pour but de promouvoir le développement de cette technologie pour la langue mongole et, surtout, de préserver la diversité de notre ADN culturel en établissant des liens avec d'autres langues.

# REMERCIEMENTS

Mes remerciements les plus vifs vont à toutes les personnes qui ont contribué à la réalisation de ce mémoire et au bon déroulement de mon stage.

Dans un premier temps, je voudrais remercier M. Damien NOUVEL pour son encadrement, sa patience et ses précieux conseils ainsi que des mises en relation avec d'autres professeurs afin de me fournir plus d'aide avec les ressources tant matérielles que linguistiques.

Je voudrais remercier également mon encadrant de mon stage, M. Qingyu JIANG, pour son soutien, sa bienveillance et le partage des connaissances et des expériences.

Je remercie Mme. Ilaine WANG pour ses conseils techniques et l'accès au serveur *Helium* de l'Université Paris Descartes, sans lequel je n'aurais pas pu effectuer l'expérimentation. Tous les modèles de traduction présentés dans ce mémoire ont été entraînés sur ce serveur.

Je remercie Mme. Charlotte MARCHINA pour m'avoir fourni des ressources très utiles sur la langue mongole.

Je remercie Mlle. Yuming ZHAI pour m'avoir recommandé le livre *Machine Translation : Foundations and Models*, qui m'a permis d'acquérir des connaissances de base sur la traduction automatique.

Je tiens particulièrement à remercier tous les participants de l'évaluation humaine : Mme. Urantuya, Mme. Sainjargal Anya, Mme. Surigage, Mme. Sovin, Mme. Hairina et M. Ananda. Sans leurs précieux retours, l'évaluation n'aurait pas pu aboutir.

Merci à M. Burged pour son guide dans l'initiation à la langue mongole et ses commentaires pertinents pour l'analyse basée sur des exemples.

Merci infiniment à mes parents et à ma femme, Mme. Ning DAI, pour leur soutien inconditionnel pendant ces derniers mois.

Finalement, ce mémoire s'appuie sur les ressources open-source. Je tiens à exprimer mes sincères remerciements à tous les contributeurs anonymes.

# RÉSUMÉ

De nos jours, la traduction automatique (MT) est de plus en plus utilisée, surtout la traduction automatique neuronale, qui a porté la qualité des traductions à un nouveau niveau. Malgré son développement rapide, elle ne fonctionne de manière suffisante que pour des langues principales telles que l'anglais, le français, le chinois et l'espagnol, etc. En ce qui concerne les langues moins parlées ou minoritaires, il n'existe pas beaucoup de recherches sur la MT. Notamment, dans le cas de notre mémoire, pour la paire de langues français-mongol, il n'existe aucune recherche publiée dessus.

Ce mémoire consiste à retracer le travail que nous avons effectué pour la mise en place d'un système de traduction automatique neuronale (NMT) pour la paire de langues français-mongol.

Notre intérêt réside en premier lieu dans la traduction automatique en nous intéressant à son développement et aux dernières technologies. Deuxièmement, nous nous pencherons sur la langue mongole en présentant ses caractéristiques ainsi que les travaux précédents liés à la MT. Ensuite, nous nous emploierons à entraîner, à partir d'un corpus parallèle open-source, des modèles de NMT français-mongol avec l'outil *OpenNMT*. Finalement, nous effectuerons deux évaluations automatiques et proposerons une méthode d'évaluation humaine, nommée « *IFF* », ce qui permet de connaître les forces et les faiblesses de notre système par rapport à des moteurs de traduction à usage commercial.

Ce travail entend revitaliser la langue mongole et protéger la diversité de notre ADN culturel en établissant des liens avec d'autres langues principales à l'appui de la NMT.

Mots clés : traduction automatique neuronale, langue mongole, évaluation automatique, évaluation humaine, score BLEU

# TABLE DES MATIERES

<b>DÉCLARATION SUR L'HONNEUR</b> .....	<b>1</b>
<b>PRÉFACE</b> .....	<b>2</b>
<b>REMERCIEMENTS</b> .....	<b>3</b>
<b>RÉSUMÉ</b> .....	<b>4</b>
<b>TABLE DES FIGURES</b> .....	<b>7</b>
<b>TABLE DES TABLEAUX</b> .....	<b>8</b>
<b>1. INTRODUCTION</b> .....	<b>9</b>
1.1. <i>Problématiques</i> .....	10
1.2. <i>Plan</i> .....	11
<b>2. ÉTAT DE L'ART</b> .....	<b>12</b>
2.1. <i>Traduction automatique</i> .....	12
2.1.1. <i>Historique de la traduction automatique</i> .....	12
2.1.2. <i>Méthodologie</i> .....	17
2.1.3. <i>Évaluation</i> .....	21
2.2. <i>Langue mongole</i> .....	28
2.2.1. <i>Locuteurs</i> .....	28
2.2.2. <i>Géographie</i> .....	29
2.2.3. <i>Linguistique</i> .....	30
2.2.4. <i>Écriture</i> .....	31
2.3. <i>Développement de la MT du mongol</i> .....	32
2.3.1. <i>Recherches</i> .....	32
2.3.2. <i>Corpus</i> .....	33
2.3.3. <i>Moteurs, plates-formes et outils</i> .....	34
2.3.4. <i>Défis</i> .....	34
<b>3. EXPÉRIMENTATION</b> .....	<b>35</b>
3.1. <i>Outils</i> .....	35
3.1.1. <i>OpenNMT</i> .....	35
3.1.2. <i>Tilde's Interactive BLEU Score Evaluator</i> .....	36
3.1.3. <i>Serveur Helium</i> .....	38
3.2. <i>Ressources</i> .....	38
3.2.1. <i>OPUS</i> .....	38

3.2.2.	<i>MultiCCAligned v1.1</i> .....	38
3.2.3.	Corpus parallèle français-mongol .....	39
3.3.	<i>Mise en place du prototype de NMT</i> .....	41
3.3.1.	Recueil de données .....	42
3.3.2.	Préparation des données.....	42
3.3.3.	Entraînement.....	43
3.3.4.	Prédiction.....	43
3.3.5.	Évaluation.....	44
<b>4.</b>	<b><i>DISCUSSION</i></b> .....	<b>52</b>
<b>5.</b>	<b><i>CONCLUSION</i></b> .....	<b>57</b>
<b>6.</b>	<b><i>PERSPECTIVES</i></b> .....	<b>59</b>
	<b><i>BIBLIOGRAPHIE</i></b> .....	<b>60</b>
	<b><i>ANNEXE</i></b> .....	<b>67</b>
	<i>Paramètres clés dans « fr_mn.yaml »</i> .....	67
	<i>Participants de l'évaluation humaine</i> .....	68
	<i>Script pour le prétraitement : tmx2txt.py</i> .....	68
	<i>Commandes utilisées</i> .....	69

# TABLE DES FIGURES

Figure 1. Nombre d'utilisateurs Internet par langue .....	9
Figure 2. Langue du contenu des sites web .....	10
Figure 3. Pierre de Rosette.....	12
Figure 4. Développement de la traduction automatique .....	13
Figure 5. Composition de traduction automatique.....	17
Figure 6. Exemple d'un système de traduction automatique statistique.....	18
Figure 7. Exemple d'un système de traduction automatique neuronale .....	20
Figure 8. Plan du groupe mongolique.....	29
Figure 9. « Langue mongole » en écritures cyrillique et classique.....	31
Figure 10. Alphabet mongol (Voyelles) .....	31
Figure 11. Alphabet mongol (Consonants).....	32
Figure 12. Interface de l'Évaluateur de score BLEU interactif .....	36
Figure 13. Exemple d'évaluation.....	36
Figure 14. Exemple des scores BLEU visualisés.....	37
Figure 15. Comparaison phrase par phrase.....	37
Figure 16. WordCloud du corpus français .....	40
Figure 17. WordCloud du corpus mongol .....	40
Figure 18. Schéma général de l'expérimentation .....	41
Figure 19. Téléchargements des corpus fr-mn disponibles sur OPUS .....	42
Figure 20. Fichiers de traductions générées à partir de src-test.....	43
Figure 21. Comparaison des scores BLEU pour la totalité du test set.....	44
Figure 22. Schéma de l'évaluation automatique (2) pour la sélection du test set.....	45
Figure 23. Comparaison des scores BLEU pour la sélection du test set.....	46
Figure 24. Schéma de l'évaluation humaine « IFF » pour la sélection du test set.....	47
Figure 25. Visualisation des scores BLEU de mt3_s (m3) et mt0_s (Google).....	54



# TABLE DES TABLEAUX

Tableau 1. Métriques DEJ.....	22
Tableau 2. Métriques non-DEJ .....	22
Tableau 3. Métriques avec référence .....	25
Tableau 4. Métriques sans référence.....	26
Tableau 5. Système de déclinaison du mongol.....	30
Tableau 6. Configuration du serveur Helium .....	38
Tableau 7. Corpus parallèle ouvert avec Olifant .....	39
Tableau 8. Jeu de données .....	43
Tableau 9. Modèles de traduction sauvegardés .....	43
Tableau 10. Exemple d'un groupe du formulaire d'évaluation « IFF » .....	47
Tableau 11. Stratégie d'évaluation IFF.....	48
Tableau 12. Calculs des scores de l'évaluation humaine.....	50
Tableau 13. Comparaison des scores IFF pour la sélection du test set.....	51
Tableau 14. BLEU-4 pour la totalité du test set.....	52
Tableau 15. BLEU-4 pour la sélection du test set .....	53
Tableau 16. Comparaison des scores BLEU de l'entrée N°1 .....	55
Tableau 17. Scores IFF moyens de l'entrée N°1 .....	55
Tableau 18. Comparaison des scores BLEU de l'entrée N°3 .....	55
Tableau 19. Scores IFF moyens de l'entrée N°3 .....	56
Tableau 20. Comparaison des scores BLEU de l'entrée N°15 .....	56
Tableau 21. Scores IFF moyens de l'entrée N°15 .....	56

# 1. INTRODUCTION

En tant que système symbolique propre à l'être humain, la langue est un véhicule de communication et d'échanges culturels. Selon la dernière édition du *Rapport mondial de l'UNESCO sur les langues 2021*, il existe quelque 7 000 langues parlées dans le monde ([Ibrahimova, 2021](#)). Dans le monde contemporain, nous sommes connectés en permanence à un cyberspace où nous avons l'accès à toutes sortes d'informations en multilingues. Dans le contexte de la mondialisation économique et de l'accélération des échanges, la demande de traductions rapides et peu coûteuses est devenue de plus en plus urgente ([Nakhlé, 2021](#)).

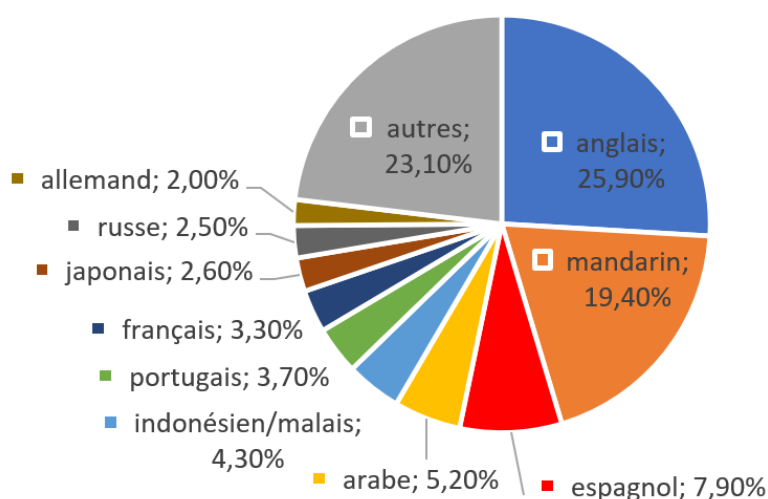


Figure 1. Nombre d'utilisateurs Internet par langue<sup>1</sup>

Internet, principal média de partage d'informations, joue en effet un rôle clé dans la promotion du multilinguisme. Cependant, la répartition des utilisateurs et du contenu des sites web par langue subit un fort déséquilibre. Selon *Internet World Stats* (Figure 1), plus d'une moitié des internautes utilisent l'anglais, le chinois et l'espagnol, ce qui détermine dans une certaine mesure les langues du contenu des sites web : comme montré dans la Figure 2, environ 60 % du contenu disponible sur Internet est en anglais.

D'où l'importance de la traduction pour communiquer avec le monde. De nos jours, la traduction automatique fait des progrès technologiques à grands pas en réponse au besoin des échanges internationaux. Néanmoins, elle ne fonctionne de manière suffisante que pour des langues parlées par un grand nombre de locuteurs et représentant un marché économique pour les grandes entreprises technologiques telles que l'anglais, le chinois et l'espagnol, etc. En ce qui concerne les langues moins parlées ou minoritaires, la langue mongole dans le cas de notre mémoire, son développement fait face à

<sup>1</sup> Source : Internet World Stats

de nombreux défis car il n'existe pas beaucoup de recherches dessus, par exemple, aucune recherche publiée sur la paire de langues français-mongol.

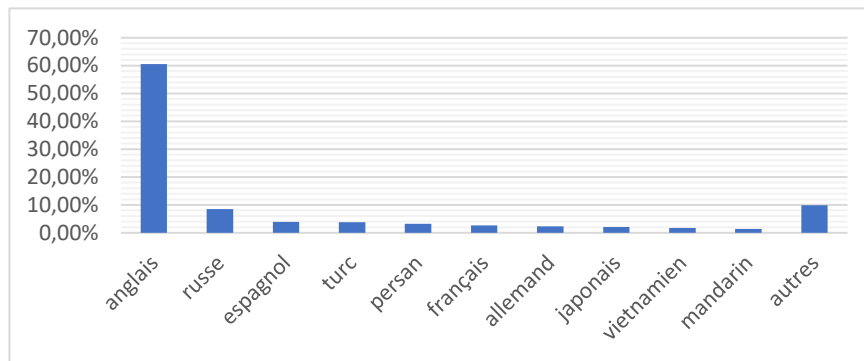


Figure 2. Langue du contenu des sites web<sup>2</sup>

L'objectif de notre mémoire est donc d'entamer une nouvelle étape dans les recherches en traduction automatique pour cette paire de langues peu dotée, tout en utilisant les techniques d'apprentissage automatique, les outils et les données open-source existants afin de construire notre propre système de traduction français-mongol.

## 1.1. Problématiques

À la fin de notre mémoire, nous nous tenterons de répondre à plusieurs questions :

- 1) Quelles sont les raisons du manque de recherches en traduction automatique pour la paire français-mongol ?
- 2) Y a-t-il d'autres recherches sur la traduction automatique de la langue mongole ? Avec quelles langues comme paires ?
- 3) Quelles ressources existent pour mettre en place un système de traduction français-mongol ?
- 4) Comment construire un système de traduction automatique neuronale ? Est-ce difficile ?
- 5) Quelles caractéristiques de la langue mongole mènent à des difficultés supplémentaires ?
- 6) Comment évaluer un système de traduction ?
- 7) Comment pourrait-on mieux faire pour améliorer la performance de notre système ?

<sup>2</sup> Source : W3Techs

## 1.2. Plan

Dans ce mémoire, notre intérêt réside en premier lieu dans la traduction automatique (2.1), son développement (2.1.1), les dernières technologies (2.1.2) et différentes méthodes d'évaluation (2.1.3).

Deuxièmement, nous nous pencherons sur la langue mongole (2.2) en présentant brièvement ses caractéristiques (2.2.3 ; 2.2.4) ainsi que les travaux précédents liés à sa traduction automatique (2.3).

Ensuite, nous effectuerons une expérimentation (3) qui consiste à entraîner des modèles de traduction automatique neuronale pour la paire de langues français-mongol à partir d'un corpus parallèle open-source récupéré sur *OPUS* (3.2.1) avec l'outil *OpenNMT* (3.1.1).

Finalement, nous réaliserons deux évaluations automatiques (3.3.5.1 ; 3.3.5.2) et concevrons une méthode d'évaluation humaine nommée « *IFF* » (3.3.5.3) pour voir les forces et les faiblesses de nos modèles de traduction par rapport à des moteurs à usage commercial.

## 2. ÉTAT DE L'ART

Cette partie explique les notions essentielles concernées et présente les travaux précédents du domaine. Ceci permet de mieux comprendre le travail effectué dans ce mémoire et de contextualiser notre contribution par rapport à l'état de l'art actuel de la traduction automatique pour la paire de langues français-mongol.

### 2.1. Traduction automatique

La traduction automatique, également appelée la traduction machine, ou MT, l'abrégié de l'anglais *Machine Translation*, est une discipline de la linguistique computationnelle et une des tâches importantes du traitement automatique des langues (TAL). Elle consiste à faire passer automatiquement un texte d'une langue « source » dans une autre langue « cible ».

Historiquement, la traduction se faisait manuellement par l'humain. Bien que précise, la traduction humaine manque de productivité et ne peut pas répondre à des besoins urgents et massifs. Avant, la machine capable de réaliser la traduction comme un humain n'apparaissait que dans les films, car la langue naturelle se caractérise par sa diversité et sa complexité. Mais aujourd'hui, grâce au développement de la technologie, la traduction automatique s'est beaucoup améliorée et n'est désormais plus un fantasme de la science-fiction.

#### 2.1.1. Historique de la traduction automatique



Figure 3. Pierre de Rosette

Le concept de « traduction » se forma au fur et à mesure que les êtres humains créèrent la langue et l'écriture. La pierre de Rosette (*Figure 4*), datant de 196 av. J.-C., en est une preuve. C'est une des plus anciennes reliques historiques d'écriture parallèle disponible pour la recherche. Sur la pierre se présente un décret sacerdotal concernant le roi d'Égypte Ptolémée V en trois écritures différentes : les hiéroglyphes égyptiens, l'écriture démotique égyptienne et le grec antique ([Vercher, 2016](#)). Selon certains chercheurs, la pierre de Rosette est un point de départ pour la traduction ([Xiao & Zhu, 2021](#)).

Bien que la notion de traduction existe depuis des milliers d'années dans l'histoire de l'humanité, le développement de la traduction automatique n'a qu'une histoire de moins d'un siècle. Le développement de la traduction automatique a connu des hauts et des bas. Un retour en arrière sur son histoire sera très instructif pour une compréhension approfondie des technologies et des méthodes connexes.

L'histoire de la traduction automatique pourrait être résumée en cinq périodes : naissance, frustration, embellie, développement et prospérité. La *Figure 4*<sup>3</sup> récapitule les travaux importants du domaine dans les périodes correspondantes avec les méthodologies principales.

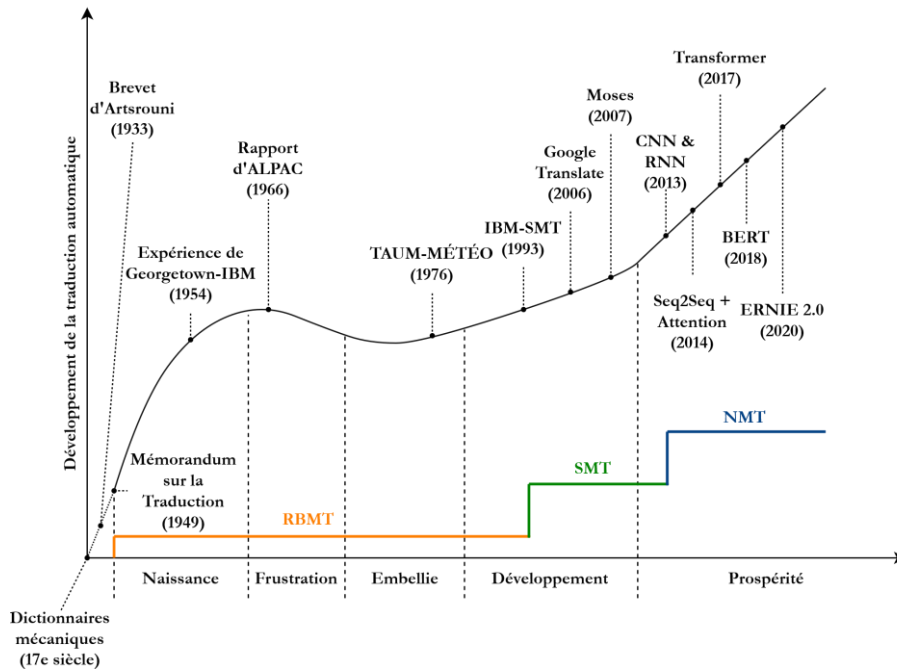


Figure 4. Développement de la traduction automatique

#### 2.1.1.1. Naissance (1933-1964)

Au 17<sup>e</sup> siècle, de nombreux savants tels que René Descartes, Gottfried Wilhelm Leibniz et Cave Beck, ont suggéré pour la première fois l'idée d'utiliser des dictionnaires « mécaniques » pour surmonter les barrières linguistiques ([Knowlson, 1975](#)), ce qui était à l'époque une pensée très avancée. Cependant, ce n'est qu'au 20<sup>ème</sup> siècle que les premières propositions concrètes sont faites ([Hutchins, 1995](#)), à savoir dans les brevets délivrés en 1933 à George Artsrouni et à Petr Smirnov-Troyanskii qui ont proposé l'automatisation de la traduction. Artsrouni a conçu un dispositif de stockage sur bande de papier qui pourrait être utilisé pour trouver l'équivalent de n'importe quel mot dans une autre langue ([Xiao & Zhu, 2021](#)).

En 1949, Warren Weaver et Claude Shannon ont présenté une méthode de traduction automatique basée sur la théorie de l'information ([Shannon & Warren, 1949](#)). Dans

<sup>3</sup> Nota bene : La *Figure 4* est le fruit personnel issu de la rédaction de l'historique en vue d'une visualisation de l'évolution de la traduction automatique.

la même année, Weaver a élaboré un mémorandum intitulé « *Translation* » (le Mémorandum sur la Traduction) qui est considéré comme le véritable début des recherches du domaine de la traduction automatique. Dans ce mémorandum, [Weaver \(1949\[1955\]\)](#) a proposé l'idée d'utiliser la cryptographie pour la traduction automatique. Ces premières explorations ont créé le concept de la « traduction automatique » et fourni certaines pistes de réflexion pour son développement ultérieur.

Après quelques années, la recherche en traduction automatique a commencé sérieusement dans de nombreuses universités américaines. Pour se tenir à ce seul exemple, en 1954 est lancée l'expérience Georgetown-IBM ([Xiao & Zhu, 2021](#)) dans laquelle l'ordinateur IBM 701 a automatiquement traduit 60 phrases russes en anglais pour la toute première fois dans l'histoire. Il s'agissait de la première démonstration publique d'un système de traduction automatique. Cette expérimentation a suscité beaucoup d'intérêt des médias et du public et a grandement favorisé la recherche en traduction automatique.

#### 2.1.1.2. Frustration (1964-1975)

Bien que les chercheurs aient obtenu des succès préliminaires dans ce domaine, de nombreuses personnes sceptiques ont remis en question l'idée de traduire automatiquement par la machine d'autant plus qu'il était très facile de repérer des fautes maladroitement commises par la traduction automatique.

Pour cela, certains chercheurs s'attendaient à pouvoir évaluer de façon objective la faisabilité de la traduction automatique. À cette époque, l'organisme de financement américain a nommé le Comité consultatif sur le traitement automatique des langues (*Automatic Language Processing Advisory Committee*, abrégé ALPAC) pour entreprendre cette tâche. En 1966, l'[ALPAC \(1966\)](#) a publié un rapport intitulé « Language and Machines », qui a nié la faisabilité de la traduction automatique et a déclaré que les machines ne pouvaient pas rivaliser avec la traduction humaine en termes de qualité et que le financement pour cette recherche devrait être arrêté.

Le rapport ALPAC a conduit à une rupture de la recherche dans le monde anglophone. Mais, d'autres pays ont continué à financer leurs équipes de recherche. Les principaux efforts expérimentaux dans cette période étaient axés sur les approches d'*Interlingua*<sup>4</sup>. Toujours est-il que les innovations techniques de cette période étaient très limitées ([Poibeau, 2017](#)).

#### 2.1.1.3. Embellie (1975-1989)

Tous les pays n'ont pas été du même avis que l'ALPAC. Contrairement aux États-Unis, de nombreux pays faisaient face au multilingue, ce qui justifie la poursuite des recherches dans le domaine de la traduction automatique ([Poibeau, 2017](#)).

Au Canada, la politique biculturelle du gouvernement a créé une demande de traduction anglais-français au-delà de la capacité du marché. Pour ce qui est de l'Europe,

---

<sup>4</sup> *Interlingua* : une langue construite ; une représentation linguistiquement neutre de concepts sémantiques et grammaticaux

les problèmes de traduction n'en étaient pas moins aigus, en particulier au sein des Communautés européennes avec des demandes croissantes de traductions de documents vers toutes les langues communautaires. Le centre des activités de traduction automatique est passé des États-Unis à l'Europe et au Canada ([Hutchins, 1995](#)).

Du milieu des années 1970 à la fin des années 1980, sont apparus en Europe de nombreux moteurs de traduction automatique, tels que *Systrans*, *Logos*, *Metal* etc. ([Hutchins, 1996](#)).

Dans les années 1970 au Canada, l'Université de Montréal a développé dans le projet *TAUM* (Traduction Automatique de l'Université de Montréal) le système *MÉTÉO*, qui traduisait les prévisions météorologiques de l'anglais vers le français ([Slocum, 1984](#)). Conçu spécifiquement pour le vocabulaire restreint et la syntaxe limitée des rapports météorologiques, *MÉTÉO* était un programme simple mais suffisamment réussi pour être utilisé jusqu'en 2000 ([Langlais et al., 2000](#)). Le projet *TAUM-MÉTÉO* marque une avancée de la traduction automatique et constitue une étape importante dans son histoire.

#### 2.1.1.4. Développement (1989-2013)

Jusqu'à la fin des années 1980, l'approche dominante de la recherche en traduction automatique était essentiellement basée sur des règles linguistiques. Depuis 1989, cependant, cette domination a été brisée par l'émergence de nouvelles méthodes fondées sur des corpus, plus précisément la « traduction automatique statistique (SMT) » ([Hutchins, 1995](#)).

En 1990, [Brown et al. \(1990\)](#), membres du *Thomas J. Watson Research Center* d'IBM, ont proposé l'idée de traduction automatique statistique, dans laquelle les machines peuvent apprendre automatiquement les connaissances de traduction à partir d'une grande quantité de données au lieu de s'appuyer sur des experts humains pour écrire des règles linguistiques. Ensuite, cette idée a été formellement formulée sous forme de cinq modèles SMT par [Brown et al. \(1993\)](#). En 2003, un nouveau modèle SMT fondé sur des phrases a été proposé par [Koehn et al. \(2003\)](#), qui a encore amélioré la qualité de la traduction. En avril 2006, *Google* a officiellement lancé son service de traduction gratuit, moteur de SMT fondé sur des phrases. Cela a apporté une grande vague de recherches en MT. D'autres entreprises telles que *Microsoft* et *Baidu* ont également lancé leurs services de traduction dans les années suivantes. En 2007, un système open-source important, *Moses*, a été publié ([Koehn et al., 2007](#)), ce qui a grandement favorisé le développement de systèmes de SMT.

Encouragés par le succès de la SMT, de nombreux chercheurs ont proposé de nouveaux modèles basés sur la SMT pour améliorer encore les performances, notamment des modèles SMT basés sur la syntaxe avec des arbres d'analyse ([Liu et al., 2006](#)), des modèles SMT hiérarchiques ([Chiang, 2007](#)), et des modèles SMT factorisés ([Koehn & Hoang, 2007](#)) dans lesquels des informations morphologiques ont été prises en compte.



### 2.1.1.5. Prospérité (2013-aujourd'hui)

Bien que l'utilisation des méthodes statistiques ait beaucoup amélioré la qualité de la traduction, la SMT ne peut pas faire le meilleur emploi des corpus parallèles et la qualité de la traduction demeure loin d'être satisfaisante en raison de l'intégration compliquée de plusieurs composants du système (Poibeau, 2017). La période de prospérité est marquée par l'émergence de la traduction automatique neuronale (NMT).

En 2013, une nouvelle structure d'encodeur-décodeur bout-en-bout (*end-to-end*) pour la traduction automatique a été proposée par Kalchbrenner and Blunsom (2013). Ils ont présenté un genre de modèles de traduction probabilistes appelés « *Recurrent Continuous Translation Models* » (modèles de traduction continus récurrents) qui permettaient d'encoder les textes source en représentations vectorielles continues et de les décoder pour générer les textes en langue cible. La génération est modélisée avec un modèle de langue récurrent (RNN) tandis que le conditionnement sur les phrases source est modélisé avec un modèle de phrase convolutif (CNN). Leurs recherches sur peuvent être considérées comme le début de la traduction automatique neuronale.

Pour la première fois, Bahdanau et al. (2014) a formellement utilisé le terme « *Neural Machine Translation* (traduction automatique neuronale) ». Ils ont introduit le mécanisme d'attention au champ du TAL en utilisant le modèle *Seq2Seq+Attention*. La qualité des traductions a été significativement améliorée. Dans la même année, Sutskever et al. (2014) ont proposé l'utilisation des réseaux à mémoire court et long terme (LSTM, *Long Short-Term Memory*) pour la traduction automatique. Les résultats ont montré que leur système de NMT pouvait surpasser un système de SMT standard.

En 2015, Baidu a publié le premier système de NMT du monde (He, 2015). Un an plus tard, Google a également lancé son système de NMT (Wu et al., 2016), qui a été suivi par d'autres sociétés publiant leurs systèmes. Ainsi, il n'a fallu qu'un an environ pour que la NMT soit déployée en ligne depuis sa première proposition en 2014, tandis qu'il a fallu environ 16 ans pour que les systèmes de SMT soient mis en service (Wang et al., 2021).

En 2017, Vaswani et al. (2017), membres du groupe de recherche de Google, ont proposé une nouvelle architecture *Transformer* qui se compose uniquement du mécanisme de « *self-attention* », sans aucune récurrence ni convolution. *Transformer* a encore amélioré de manière significative la qualité de la traduction par la prise en compte du sens des mots en contexte.

Ce grand succès a attiré de nombreux chercheurs pour développer diverses méthodes, telles que le modèle non autorégressif (Gu et al., 2017), le modèle NMT non supervisé (Artetxe et al., 2019) et les modèles préentraînés *BERT* (Devlin et al., 2018) et *ERNIE 2.0* (Sun et al., 2020), dans le but d'améliorer la qualité et l'efficacité de la traduction automatique neuronale.

Il est à noter que les modèles susmentionnés se basent fondamentalement sur *Transformer* et n'apportent pas de grands changements sur l'architecture. À l'heure actuelle, *Transformer* représente l'état de l'art dans le domaine de NMT en termes de qualité et d'efficacité.

## 2.1.2. Méthodologie

Un moteur de traduction automatique se compose généralement de deux parties : ressources et système (Figure 5).

- Ressources :

Les ressources sont des données « lisibles » par le système de traduction telles que des règles de traduction, des données linguistiques dont des corpus monolingues ou bilingues alignés ou comparables, des bases de connaissances, etc. Sans les ressources, aucun système de MT ne peut fonctionner.

- Système :

Nous appelons l'ensemble d'algorithmes et la mise en œuvre du programme de traduction un « système ». Cela peut être un logiciel ou une architecture développée par des chercheurs. Qu'il s'agisse de règles de traduction, de corpus, de modèles de traduction, ils doivent être lus et utilisés par le système de MT.

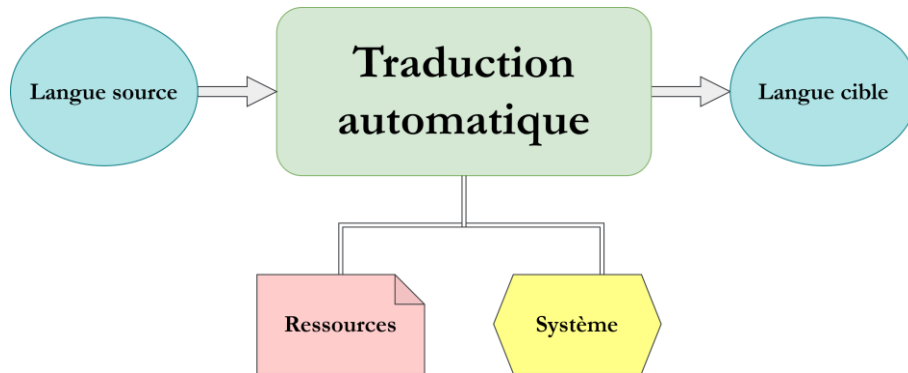


Figure 5. Composition de traduction automatique

Dans l'histoire de la traduction automatique, nous reconnaissons généralement deux méthodologies (Zhang & Zong, 2020):

- 1) Approche orientée connaissances ou basée sur les règles (*knowledge-driven* ou *rule-based approach*) ;
- 2) Approche orientée données ou basée sur les corpus (*data-driven* ou *corpus-based approach*).

L'approche basée sur les règles consiste à élaborer des grammaires et des lexiques pour les paires de langues et à écrire des règles expertes pour traduire une langue en une autre (Wang et al., 2021). Nous n'entrerons pas en détail la traduction automatique basée sur des règles (*Rule-based Machine Translation*, abrégée RBMT) car nous nous intéressons plutôt à la deuxième approche plus moderne et avancée.

L'approche orientée données peut en outre se diviser en traduction automatique statistique (*Statistical Machine Translation*, abrégée SMT) et en traduction automatique neuronale (*Neural Machine Translation*, abrégée NMT).

### 2.1.2.1. Traduction automatique statistique (SMT)

La traduction automatique statistique consiste à utiliser des modèles statistiques pour apprendre automatiquement les connaissances de traduction à partir de corpus

monolingues ou bilingues. Plus précisément, un corpus monolingue permet à la machine d'apprendre le modèle de langue, et avec un corpus parallèle bilingue la machine peut apprendre le modèle de traduction. Le système utilise ensuite ces deux modèles pour modéliser le processus de traduction.

L'ensemble du processus d'apprentissage ne nécessite pas de création manuelle des règles, ni de construction de modèles de traduction à partir d'exemples. La partie du travail humain porte généralement sur la définition des caractéristiques (*features*) requises pour la traduction et d'unités de traduction de base. Les connaissances apprises sont stockées dans les paramètres du modèle.

La *Figure 6* illustre le fonctionnement général d'un système de SMT. Ce processus nécessite deux modèles : le modèle de traduction et le modèle de langue. Le modèle de traduction contient les connaissances en traduction apprises à partir du corpus parallèle bilingue, à savoir une série de textes par paires et leurs probabilités correspondantes, ce qui permet de mesurer la vraisemblance des traductions entre la langue source et la langue cible. Le modèle de langue apprend les règles de génération des séquences à partir du corpus monolingue pour améliorer la fluidité de la traduction en langue cible.

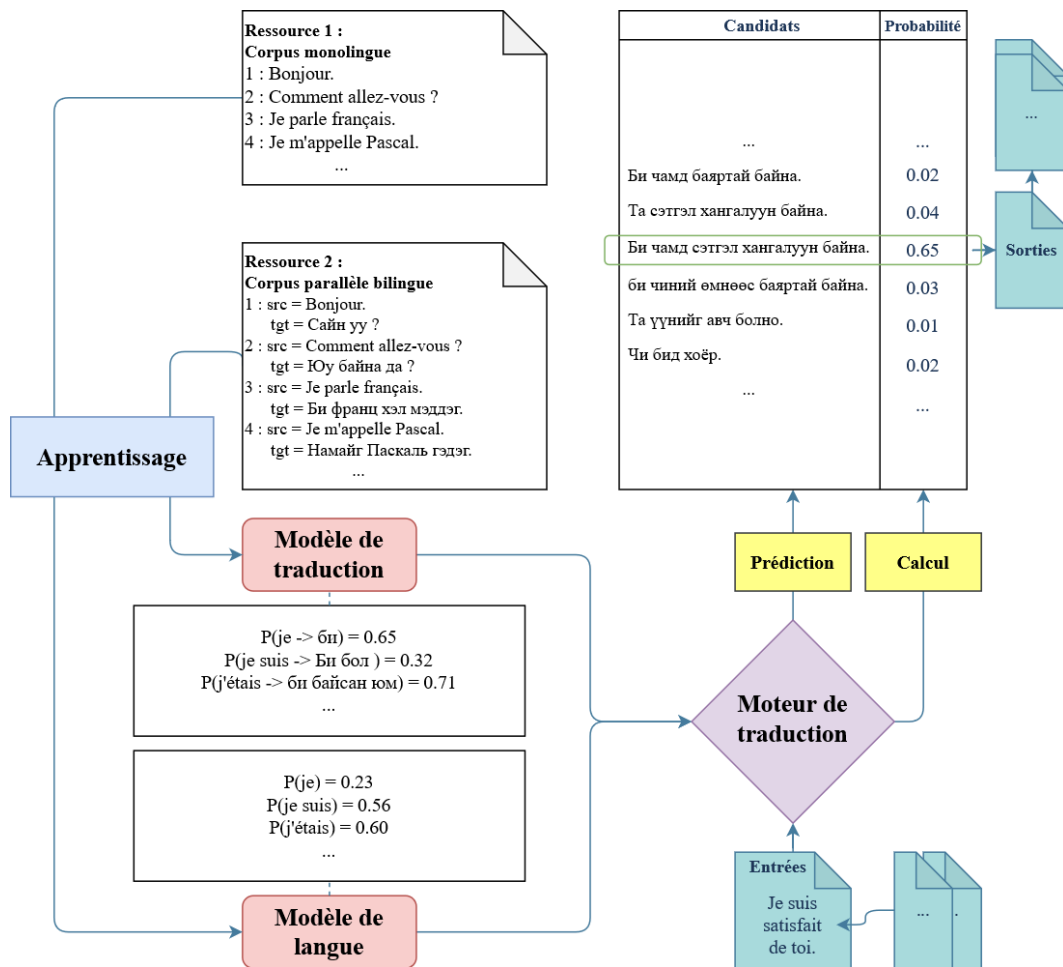


Figure 6. Exemple d'un système de traduction automatique statistique

Ainsi, les deux modèles sont combinés pour former le moteur de traduction qui cherche autant de résultats de traduction que possible et qui calcule ensuite leurs

vraisemblances. Enfin, la traduction avec la vraisemblance la plus élevée est sortie comme résultat final.

La SMT se caractérise par une interprétabilité. Si les sorties ne sont pas satisfaisantes, nous pouvons l'améliorer en revoyant les deux modèles qui constituent le système. Ceci est un grand avantage de la SMT.

Cependant, cette méthode pose également quelques problèmes. Tout d'abord, bien que l'on n'ait plus besoin de définir manuellement des règles ou des modèles de traduction, le système de traduction automatique statistique nécessite toujours des caractéristiques (*features*) manuellement définies. L'amélioration de la qualité des sorties requiert souvent beaucoup d'ingénierie des *features* dont la conception influe de manière décisive sur le système de traduction. Deuxièmement, il existe beaucoup de modules pour la traduction automatique statistique et le développement du système est assez compliqué. Troisièmement, quand la quantité des données augmente pour l'apprentissage, les tailles des deux modèles augmenteront considérablement, ce qui consomme beaucoup de stockage ([Xiao & Zhu, 2021](#)).

#### 2.1.2.2. Traduction automatique neuronale (NMT)

Avec le développement de la technologie d'apprentissage automatique par réseaux de neurones, la traduction automatique neuronale basée sur l'apprentissage profond est en plein essor depuis plusieurs années.

En NMT, la chaîne de mots est représentée comme un vecteur, à savoir une représentation vectorielle distribuée. Ainsi, la traduction n'est plus effectuée sur des mots ou des phrases discrétisés, mais dans un espace vectoriel. Par conséquent, par rapport aux méthodologies précédentes, la NMT présente un changement fondamental dans la manière de représenter les séquences de mots ([Xiao & Zhu, 2021](#)).

En général, la traduction automatique peut être considérée comme la conversion d'une séquence en une autre. Dans la NMT, le processus de conversion séquentielle est mis en œuvre par l'architecture d'encodeur-décodeur. Un réseau d'encodeurs analyse la séquence en langue source et extrait les informations dans une représentation vectorielle, à partir de laquelle un réseau de décodeurs produit la traduction en langue cible. Ce processus est analogue à la traduction humaine. Le modèle NMT « lit » d'abord toute une phrase source ; puis, sur la base de sa compréhension de cette phrase, le modèle génère la phrase cible mot par mot ([Wang et al., 2021](#)).

La *Figure 7* montre un exemple de NMT français-mongol. L'encodeur mappe, par un réseau de neurones, la séquence « Je suis satisfait de toi. » en une représentation vectorielle, (0.6, 0.4, 2, 0.3, -1). Ce vecteur est ensuite envoyé en entrée du décodeur qui le déchiffre et génère une phrase en langue cible, « Би чамд сэтгэл ханхлуун байна. ».

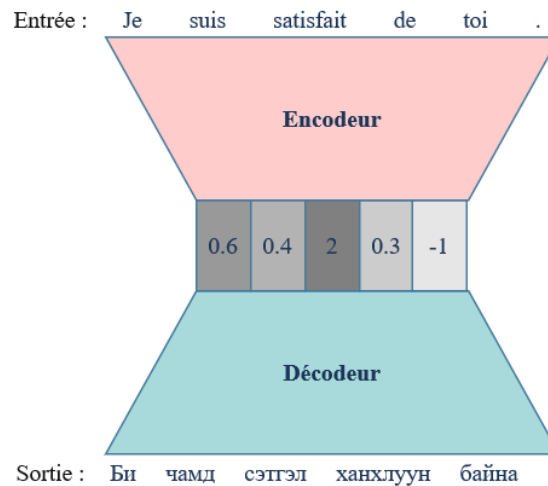


Figure 7. Exemple d'un système de traduction automatique neuronale

Par rapport aux méthodes précédentes telles que RBMT et SMT, la NMT présente un avantage majeur : elle ne nécessite pas d'ingénierie de *features* et toutes les informations nécessaires sont automatiquement extraites de l'entrée par le réseau de neurones. De plus, par rapport à la représentation discrétisée utilisée en SMT, la représentation spatiale continue distribuée des mots et des phrases en NMT peut fournir des informations plus riches pour la modélisation. En outre, une méthode basée sur le gradient peut être utilisée pour optimiser le modèle. Enfin, la NMT consomme généralement moins de stockage que la SMT (Xiao & Zhu, 2021).

Néanmoins, la NMT n'échappe pas à certains défis. Bien qu'elle n'ait pas besoin de l'ingénierie des *features*, l'architecture du réseau de neurones requiert toujours une conception humaine. Même si l'architecture est prête, le réglage du système et le paramétrage des hyperparamètres s'appuie sur un grand nombre d'expériences (Xiao & Zhu, 2021). De plus, la traduction automatique neuronale repose fortement sur les données. La performance d'un système NMT dépend directement de la quantité et de la qualité des données. Surtout lorsque les données sont rares pour certaines langues, il est difficile d'entraîner un bon système de traduction. En outre, il se peut que la NMT « oublie » de traduire certains éléments dont l'entropie est importante. En effet, plus la langue cible est informativement dense, et plus l'entropie d'un mot est grande, plus il aura des chances d'être omis par le système. Finalement, le plus grand inconvénient de la NMT, c'est le manque d'interprétabilité. Les systèmes de traduction réalisés par apprentissage de réseaux de neurones ont toujours été marqués par un effet « boîte noire ». Seul le système peut lire et « comprendre » les représentations vectorielles. Quand un problème se pose, il est difficile de savoir quels phénomènes linguistiques vont poser un problème au système et pourquoi (Nakhlé, 2021). Donc, le degré d'intervention humaine reste très limité.

À en juger par la recherche et l'application actuelles de la traduction automatique, la NMT s'avère aujourd'hui être la méthode dominante du domaine. Cela est principalement dû à son court cycle de développement : on peut réaliser rapidement un prototype après avoir collecté une quantité minimale de données (Xiao & Zhu, 2021).

Puisque le coût de l'acquisition des données devient de plus en plus faible et que son développement se caractérise par une simplicité, la NMT est privilégiée par de plus en plus de chercheurs.

### 2.1.3. Évaluation

Afin de mesurer la performance d'un système de MT, il faut une évaluation pour déterminer la qualité de ses productions. Depuis le début de son histoire, l'évaluation a été considérée comme nécessaire, plus que dans d'autres domaines du traitement automatique des langues, probablement parce que la traduction automatique était considérée comme un domaine applicatif. Et donc, des résultats très concrets étaient attendus (Poibeau, 2017). Cette étape a un rôle crucial à jouer, que ce soit dans les itérations du développement d'un système de MT ou dans de nombreux scénarios d'application, pour décider quelles techniques donnent de meilleurs résultats et quelles techniques ont été dépassées. La traduction automatique n'aurait pas pu devenir ce qu'elle est aujourd'hui sans l'évaluation (Xiao & Zhu, 2021).

Il existe un éventail de méthodes d'évaluation qui se divisent principalement en deux catégories : l'évaluation humaine et l'évaluation automatique.

#### 2.1.3.1. Évaluation humaine

L'évaluation humaine est considérée comme celle dans laquelle l'humain intervient lors de l'évaluation elle-même, à la différence de l'évaluation automatique que nous présenterons dans la partie 2.1.3.2.

Selon Chatzikoumi (2020), l'évaluation humaine peut en outre être divisée en deux sous-catégories selon le critère suivant : si un jugement directement exprimé (*directly expressed judgment* en anglais, abrégé DEJ) est utilisé ou non. Ces deux sous-catégories sont ainsi appelées les « métriques DEJ » et les « métriques non-DEJ ».

Les Tableaux 1 et 2 récapitulent succinctement les métriques DEJ et non-DEJ avec leur idée principale.

Métriques DEJ	
<b><i>Adequacy and fluency annotation tasks</i></b> (Tâches d'annotation d'adéquation et de fluidité)	L'évaluateur utilise des outils, tels que <i>DQF</i> <sup>5</sup> et <i>UCCA</i> <sup>6</sup> , pour évaluer la traduction en termes d' <u>adéquation et de fluidité souvent avec une échelle de 1 à 4.</u>
<b><i>Ranking</i></b> (Classement)	L'évaluateur utilise des outils, tels que <i>DQF</i> et <i>Appraise</i> <sup>7</sup> , pour <u>effectuer une comparaison en choisissant</u>

<sup>5</sup> *DQF* : *Dynamic Quality Framework*, plate-forme consistant à standardiser l'évaluation des traductions humaine et automatique, développée par TAUS en 2011, disponible sur : <https://www.taus.net>

<sup>6</sup> *UCCA* : *Universal Conceptual Cognitive Annotation*, projet interlinguistique pour l'annotation sémantique, développé par Abend et Rappoport

<sup>7</sup> *Appraise* : boîte à outils open-source pour l'évaluation de la traduction, développée par Federmann

	la <u>meilleure traduction</u> parmi plusieurs productions de différents systèmes.
<b>Direct Assessment</b> (Évaluation directe)	L'évaluateur exprime <u>un jugement sur la qualité de la traduction produite dans une échelle de notation continue</u> , qui représente le degré auquel une traduction est meilleure qu'une autre en termes d'adéquation et de fluidité.
<b>Quality-checking annotation tasks</b> (Tâches d'annotation de contrôle qualité)	L'évaluateur <u>annote la traduction avec trois étiquettes possibles</u> : « <i>acceptable</i> », « <i>can easily be fixed</i> » et « <i>none of both</i> ».

Tableau 1. Métriques DEJ

<b>Métriques non-DEJ</b>	
<b>Semiautomated metrics</b> (Métriques semi-automatisées)	Les métriques semi-automatisées sont <u>des variantes de métriques automatisées avec l'intervention humaine</u> , comme dans les cas de <i>HTER</i> <sup>8</sup> , <i>HBLEU</i> <sup>9</sup> et <i>HMETEOR</i> <sup>10</sup> .
<b>Task-based evaluation</b> (Évaluation basée sur les tâches)	L'évaluateur <u>évalue indirectement le degré</u> auquel les concepts du texte source sont exprimés dans la traduction <u>via certaines tâches concrètes, sans porter de jugement direct</u> sur la qualité de la traduction.
<b>Error classification and analysis</b> (Classification et analyse des erreurs)	L'évaluateur effectue <u>une classification des erreurs</u> de la traduction en utilisant <i>MQM</i> <sup>11</sup> , ce qui est de préférence suivi par <u>une analyse</u> avec <i>DQF</i> .
<b>Postediting</b> (Post-édition)	La post-édition est définie comme la tâche par laquelle la sortie MT est transformée en une traduction livrable. Elle peut également être utilisée comme méthode d' <u>évaluation indirecte, en calculant les efforts temporel et cognitif avec des outils</u> , telles que <i>CASMACAT</i> <sup>12</sup> et <i>PET</i> <sup>13</sup> , etc., <u>qui fournissent des informations statistiques pertinentes</u> .

Tableau 2. Métriques non-DEJ

<sup>8</sup> *HTER* : Human-targeted Translation Edit Rate<sup>9</sup> *HBLEU* : Human-targeted Bilingual Evaluation Understudy<sup>10</sup> *HMETEOR* : Human-targeted Metric for Evaluation of Translation with Explicit Ordering<sup>11</sup> *MQM* : Multidimensional Quality Metrics, métriques qui définissent la qualité des traductions humaine et automatique et qui décrit le processus d'évaluation avec des critères spécifiques, développés par QTLaunchPad<sup>12</sup> *CASMACAT* : Cognitive Analysis and Statistical Methods for Advanced Computer Aided Translation, disponible sur : <https://github.com/casmacat><sup>13</sup> *PET* : Post-Editing Tool, disponible sur : <http://www.clg.wlv.ac.uk/projects/PET/>

Généralement, les métriques DEJ font preuve de plus de subjectivité que les métriques non-DEJ. Parce que le traitement cognitif impliqué dans les deux types d'évaluation est assez différent : dans les métriques DEJ, l'évaluateur effectue directement une évaluation sur la qualité de la traduction, telle que sa fidélité, sa fluidité, etc., tandis que, dans les métriques non-DEJ, le processus est beaucoup plus axé sur les tâches concrètes sous la forme de classification, post-édition, questions à choix multiples et questions à trous, etc. ([Chatzikoumi, 2020](#)).

Dans l'évaluation humaine, plusieurs évaluateurs évaluent, de manière directe ou indirecte, la qualité de la traduction, puis les résultats de tous les évaluateurs sont combinés pour donner un résultat final. Les évaluateurs, également appelés annotateurs, constituent donc un facteur important dans le processus d'évaluation humaine. Pour assurer la fiabilité, le choix d'évaluateurs doit répondre à certains critères.

Les évaluateurs peuvent être monolingues ou bilingues, c'est-à-dire des locuteurs natifs ou quasi natifs de la langue cible ou des langues source et cible. Il est préférable que les évaluateurs soient des traducteurs professionnels et des experts du domaine concerné, mais le coût de l'évaluation augmentera considérablement. Dans la plupart des cas, on recrute des évaluateurs qui ne sont pas des traducteurs professionnels et parfois qui ne sont pas bilingues ([Nakhlé, 2021](#)). Par ailleurs, la formation des évaluateurs, le guide d'évaluation avec des exemples, ainsi que la familiarité de l'évaluateur avec le domaine auquel appartiennent les textes, constituent des conditions préalables au projet d'évaluation humaine ([Chatzikoumi, 2020](#)).

L'évaluation humaine peut refléter avec précision la qualité de la traduction et est considérée comme la méthode la plus adéquate et la plus crédible. Selon [Graham et al. \(2013\)](#), des annotations humaines s'avèrent être nécessaires dans le traitement automatique des langues naturelles « afin d'estimer dans quelle mesure un système donné imite les activités traditionnellement effectuées par l'humain ». Donc, les métriques humaines sont toujours un élément important de l'évaluation dans les ateliers annuels de MT comme le WMT.

Cependant, l'évaluation humaine présente évidemment plusieurs inconvénients tels que le coût élevé, le manque de réutilisabilité, son caractère chronophage et son retour différé ainsi qu'un accord inter-annotateur difficile à assurer eu égard à la subjectivité ([Chatzikoumi, 2020](#)). En conséquence, dans le développement réel, l'évaluation purement humaine ne sera pas utilisée trop fréquemment. Il faut donc considérer des moyens pour baisser les coûts tout en gardant la valeur informationnelle de l'évaluation ([Nakhlé, 2021](#)), d'où l'évaluation automatique.

### 2.1.3.2. Évaluation automatique

Dans le développement de la MT, l'évaluation automatique revêt une importance capitale pour une estimation rapide de la performance d'un système. Lors de l'entraînement du modèle, les développeurs se servent des scores automatiques pour savoir si l'entraînement s'est bien passé ou pour mesurer le progrès entre des versions différentes ([Nakhlé, 2021](#)). En outre, la plupart des décisions de sélection d'architecture, de réglage d'hyperparamètres et de filtrage de données reposent sur l'évaluation automatique à un



rythme et à une échelle qui ne seraient pas soutenables avec l'évaluation humaine ([Freitag et al., 2020](#)).

De nos jours, il existe de nombreuses métriques automatiques, lesquelles pourraient être classées en deux<sup>14</sup> types ([Xiao & Zhu, 2021](#)) :

- 1) Métrique avec référence
- 2) Métrique sans référence

Les métriques avec référence consistent à donner pour les sorties d'un système de MT un score basé sur le degré de similarité par rapport aux traductions de référence.

Les métriques sans référence, également appelées « estimation de la qualité (*Quality Estimation*, abrégé QE) », visent à évaluer la performance d'un système de MT en classant les sorties d'un système de MT en fonction de leurs niveaux de qualité sans se référer aux traductions de référence (qui n'existent pas dans ce cas-là). En fait, elles ne comptent pas comme métriques d'évaluation proprement dites, mais sont considérées comme intermédiaires ([Specia et al., 2009](#)) ou moyens alternatifs d'évaluation ([Bojar et al., 2016](#)).

Les *Tableaux 3* et *4* récapitulent succinctement les métriques avec et sans référence.

---

<sup>14</sup> *Nota bene* : Il existe un autre type d'évaluation automatique nommée « évaluation diagnostique basée sur les points de contrôle linguistiques (*Diagnostic Evaluation based on Linguistic Checkpoints*) » qui est peu utilisé. Nous ne l'aborderons pas dans ce mémoire.

<b>Métriques avec référence</b>	
<p style="text-align: center;"><b><i>Edit Distance</i></b> (Distance d'édition)</p>	<p>Cette méthode d'évaluation automatique est basée sur <u>la distance d'édition</u>, plus précisément, <u>la distance de Levenshtein</u> (<a href="#">Levenshtein, 1966</a>), qui consiste à <u>calculer le nombre minimum d'opérations d'édition nécessaires</u>, telles que l'insertion, l'élimination et la substitution, <u>pour convertir la sortie MT à la traduction de référence.</u></p> <p>En voici quelques métriques connues :</p> <ul style="list-style-type: none"> <li>• <i>Word Error Rate</i> (WER) (<a href="#">Nießen et al., 2000</a>)</li> <li>• <i>Translation Edit Rate</i> (TER) (<a href="#">Snover et al., 2006</a>)</li> <li>• <i>Sentence Error Rate</i> (SER) (<a href="#">Tomás et al., 2003</a>)</li> <li>• <i>Cover Disjoint Error Rate</i> (CDER) (<a href="#">Leusch et al., 2006</a>)</li> </ul>
<p style="text-align: center;"><b><i>Precision and Recall</i></b> (Précision et rappel)</p>	<p>Une autre méthode très utilisée se base sur la précision et le rappel.</p> <p><u>La précision est le rapport entre les N-grams acceptables dans la sortie MT</u>, c'est-à-dire les N-grams également trouvés dans au moins une des traductions de référence, <u>au nombre de N-grams dans la même sortie MT.</u></p> <p><u>Le rappel est le rapport des N-grams acceptables dans la sortie MT</u>, c'est-à-dire les N-grams également trouvés dans au moins une des traductions de référence, <u>au nombre de N-grams de la traduction de référence</u> (<a href="#">Chatzikoumi, 2020</a>).</p> <p>En voici quelques métriques connues :</p> <ul style="list-style-type: none"> <li>• <i>Bilingual Evaluation Understudy</i> (BLEU) (<a href="#">Papineni et al., 2002</a>)</li> <li>• <i>National Institute of Standards and Technology</i> (NIST) (<a href="#">Doddington, 2002</a>)</li> <li>• <i>Recall-Oriented Understudy for Gisting Evaluation</i> (ROUGE) (<a href="#">Lin &amp; Och, 2004</a>)</li> <li>• <i>General Text Matcher</i> (GTM) (<a href="#">Melamed et al., 2003</a>)</li> <li>• <i>Metric for Evaluation of Translation with Explicit ORdering</i> (METEOR) (<a href="#">Banerjee &amp; Lavie, 2005</a>)</li> </ul>

Tableau 3. Métriques avec référence

<b>Métriques sans référence</b>	
<p><b>Quality Estimation</b> (Estimation de la qualité)</p>	<p>L'estimation de la qualité (QE) est utilisée pour prédire « <u>la qualité de la sortie d'un système pour une entrée donnée, sans aucune information sur la sortie attendue</u> » (<a href="#">Specia et al., 2009</a>).</p> <p>La QE n'a pas le même objectif que les métriques avec référence, et n'entend pas non plus les remplacer. Elle en sert d'un complément pour, d'une part, combler le vide créé dans les cas où aucune traduction de référence n'existe et, d'autre part, répondre aux besoins de QE au niveau du segment où d'autres métriques ont de mauvais résultats (<a href="#">Specia et al., 2010</a>).</p> <p>En voici quelques plates-formes de QE :</p> <ul style="list-style-type: none"> <li>• <i>QuEst</i><sup>15</sup> (<a href="#">Specia et al., 2013</a>)</li> <li>• <i>Asiya</i><sup>16</sup> (<a href="#">Giménez &amp; Gonzalez, 2013</a>)</li> </ul>

Tableau 4. Métriques sans référence

Manifestement, l'évaluation automatique présente par rapport à l'évaluation humaine plusieurs avantages tels que la rapidité, le faible coût, et la réutilisabilité. Il est à noter que la réutilisabilité constitue l'avantage le plus important lors du développement des systèmes de MT, ce qui permet des modifications, des améliorations et des réévaluations ([Banerjee & Lavie, 2005](#); [Lavie, 2011](#)). De plus, l'évaluation automatique fait acte de plus de cohérence car une métrique automatique donne toujours le même score étant donné que tous les paramètres restent inchangés, contrairement à l'évaluation humaine où l'on ne peut pas garantir que l'évaluateur donne toujours le même score au même texte en fonction de sa vigueur et de son attention, sans parler de l'accord inter-annotateur ([Chatzikoumi, 2020](#)).

L'évaluation automatique est donc privilégiée par les développeurs de MT. Avec le développement continu de la technologie d'évaluation, les résultats de l'évaluation automatique sont devenus de plus en plus instructifs et peuvent aider les développeurs à comprendre rapidement la performance du système actuel.

Néanmoins, elle n'échappe non plus à quelques inconvénients. Par exemple, les scores automatiques s'avèrent souvent difficiles à interpréter ([Koehn, 2010](#)) et ils ne sont pas capables de fournir des informations sur les forces et les inconvénients exacts d'un système de MT ([Zhou et al., 2008](#)). De plus, il s'agit de faibles performances dans les textes courts ([Koehn, 2007](#)), du manque de *stemming* (racinisation), de

<sup>15</sup> *QuEst*, disponible sur : <https://github.com/lspcia/quest>


<sup>16</sup> *Asiya*, disponible sur : <https://asiya.lsi.upc.edu/>

lemmatisation, d'utilisation des synonymes et des paraphrases ([Callison-Burch et al., 2006](#); [Lavie, 2011](#)), et du fait qu'il n'y a pas de distinction entre des scores très faibles dans des traductions de très mauvaise qualité ou des traductions très libres ([Coughlin, 2003](#)), etc.

Même si l'évaluation automatique est essentielle pour la phase d'entraînement et de développement ([Nakhlé, 2021](#)), pour obtenir une appréciation finale sur la performance d'un modèle, l'évaluation humaine est toujours considérée comme déterminante dans la communauté scientifique ([Bojar et al., 2016](#)).

En conséquence, l'évaluation humaine et l'évaluation automatique sont souvent utilisées ensemble dans le développement réel pour avoir une meilleure compréhension du système de MT ([Xiao & Zhu, 2021](#)).

## 2.2. Langue mongole

La langue mongole (cyrillique : Монгол хэл ; traditionnel : ) appartient au groupe mongolique, à la famille altaïque, l'une des principales familles linguistiques d'Asie centrale et du nord-est ([Janhunén, 2005](#)).

La langue mongole, parlée par plus de 7 millions de locuteurs et langue officielle de la République populaire de Mongolie, peut *de facto* faire référence à plusieurs langues et dialectes similaires. La plupart des variétés de la langue mongole partagent quelques caractéristiques importantes, telles que l'ordre des phrases SOV (sujet-objet-verbe), l'agglutination (un processus linguistique où des mots complexes sont formés en ajoutant des morphèmes significatifs aux racines des mots) et l'harmonie des voyelles ([Gaunt, 2004](#)).

### 2.2.1. Locuteurs

L'Empire mongol (1206-1368), fondé par Gengis Khan (1160-1227), était le plus grand empire terrestre de l'histoire de tous les temps. Cet empire et ses États successeurs ont exercé une énorme influence sur l'histoire du monde ultérieure ([Janhunén, 2005](#)). Aujourd'hui, les Mongols vivent principalement en Mongolie et dans la région autonome de Mongolie intérieure de la République populaire de Chine, ainsi qu'au Xinjiang, sur le plateau tibétain et dans certaines parties de la Russie.

Historiquement, les Mongols (locuteurs ethniques du mongol) sont connus comme les créateurs de l'Empire mongol et les descendants du peuple unifié par Gengis Khan. À l'origine, le mongol était le nom d'une unité sociale limitée, ou d'une tribu, mais comme il s'agissait de l'unité dont descendait Gengis Khan, le terme a finalement été étendu pour englober l'ensemble de la population qui parlait ou venait à parler la même langue ([Janhunén, 2005](#)).

Avec la diversification du mongol au fil du temps est né un groupe des langues et dialectes apparentés, nommé « mongolique ». Certaines populations conservent aujourd'hui l'ethnonyme commun, ou ses variantes, et continuent d'être désignées par des noms tels que « Mongol », « Mongghul », « Mangguer » ou « Moghol » ([Janhunén, 2005](#)). La plupart des populations mongoliques, cependant, portent des ethnonymes différents, principalement basés sur d'autres noms ethniques ou tribaux anciens. Indépendamment de leurs ethnonymes, tous les locuteurs des langues mongoles modernes peuvent être considérés comme des descendants plus ou moins directs des Mongols historiques ([Janhunén, 2005](#)).

Il faut cependant souligner que la taxonomie ethnique officielle n'est pas toujours cohérente avec les distinctions linguistiques et culturelles réelles. Au total, le nombre de locuteurs du mongol peut être estimé à 7 millions, ce qui est étonnamment peu compte tenu du grand impact historique et de la dispersion géographique moderne des Mongols. La population totale des groupes ethniques concernés est un peu plus importante, car, surtout en Chine, l'identité ethnique officielle peut également être héritée

sans maintien de la langue maternelle. Ainsi, il y a environ 5 millions de Mongols ethniques en Chine, mais probablement pas plus de 3 millions de locuteurs de la langue mongole. Même ainsi, il y a plus de locuteurs de mongol en Chine qu'en Mongolie ([Janhunen, 2005](#)).

## 2.2.2. Géographie

Les langues mongoliques sont parlées sur une large étendue du continent eurasiatique, de la région caspienne à l'ouest à la Mandchourie méridionale et centrale à l'est. Au sud, le territoire mongolique s'étend au nord de l'Afghanistan ainsi qu'à la région du Gansu-Qinghai (Amdo) au nord du Tibet, tandis qu'au nord il comprend la région du Baïkal et une partie du bassin de l'Amour. Le centre de gravité politique et historique de ce territoire est situé sur le plateau mongol, une région dominée par les steppes et les montagnes au cœur de l'Asie ([Janhunen, 2005](#)).

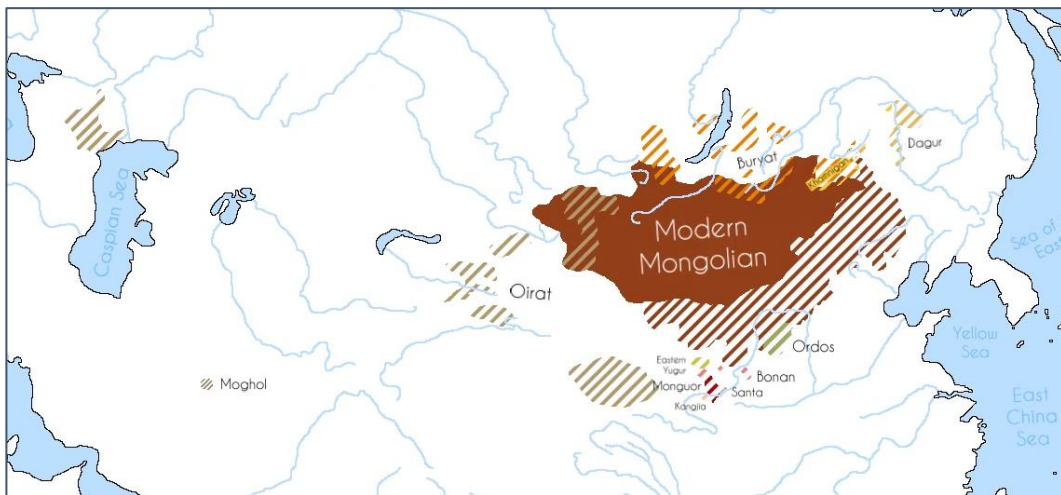


Figure 8. Plan du groupe mongolique

La *Figure 8* illustre la répartition géographique des langues mongoliques ([Kim, 2020](#)). La langue officielle de Mongolie, mongol moderne dans la figure, s'appelle à juste titre le mongol Khalkha, qui forme le noyau absolu du groupe mongolique. Autour de ce noyau, il y a plusieurs langues étroitement apparentées : « ordos » au sud, « oirat » à l'ouest, « bouriate » au nord et le « khamnigan » au nord-est. Plus loin du noyau, il y a encore d'autres langues : « moghol » à l'ouest, « dagur » au nord-est et le complexe « gansu-qinghai » ou le « monguor », comprenant « shira yughur », « mongghul », « mangghuer », « bonan » et « santa », au sud ([Janhunen, 2005](#)).

Le désert de Gobi est traditionnellement considéré comme la frontière physique entre la Mongolie et la Chine. La Grande Muraille de Chine est un symbole permanent de la frontière ethnique sino-mongole. Politiquement, la langue mongolique la plus importante est le mongol proprement dit (mongol propre, ou *Mongolian Proper* en anglais), parlé dans le territoire central des Mongols historiques. Son dialecte Khalkha sert aujourd'hui de langue officielle de la République populaire de Mongolie, l'ancienne région chinoise de la Mongolie extérieure. Et des formes dialectales très

similaires sont utilisées par les Mongols vivant dans la région autonome chinoise de la Mongolie intérieure avec un statut officiel de deuxième langue de la région après le mandarin ([Janhunen, 2005](#)).

### 2.2.3. Linguistique

La langue mongole est une langue agglutinante, dans laquelle les changements grammaticaux sont effectués par l'ajout des suffixes à la racine du mot et chaque suffixe exprime un seul sens grammatical.

Le mongol appartient à la famille des langues altaïques et partage certaines similitudes grammaticales avec le turc, le mandchourien-toungouse, le japonais et le coréen ([Tserenpil & Kullmann, 2008](#)). Le vocabulaire mongol est riche en éléments turcs anciens, sanskrits, tibétains et chinois qui reflètent l'héritage nomade, bouddhiste et asiatique du pays. Le mongol continue d'adopter des mots du russe et de l'anglais, mais conserve en grande partie son propre caractère distinctif ([Tserenpil & Kullmann, 2008](#)).

Il existe dans la langue mongole cinq modes grammaticaux : indicatif, dubitatif, impératif, conditionnel, subjonctif, optatif.

Cas	Suffixe	Usage	Exemple : nom (livre)
nominatif	-	sujet ; objet indéfini direct	nom
accusatif	-ig	objet défini direct ; sujet d'une proposition subordonnée	nomig
génitif	-in	sujet d'une construction participiale ; nom attributif ; nom possesseur	nomin
datif	-d	objet indirect ; emplacement ; direction (vers)	nomd
ablatif	-As	direction (depuis)	nomoos
instrumental	-Ar	instrument de l'action ; direction (à travers ; le long de)	nomoor
comitatif	-tAi	possession	nomtoi

Tableau 5. Système de déclinaison du mongol

Le mongol a son système de déclinaison avec sept cas, lesquels sont résumés avec leur usage dans le *Tableau 5*. Les lettres sont écrites ici en translittération latine. Le « A » représente une voyelle qui peut être composée de « a », « e » et « o » ([Janhunen, 2005](#)).

## 2.2.4. Écriture

Tout au long de leur histoire, les Mongols ont utilisé plusieurs alphabets. Au début du 13<sup>e</sup> siècle, le mongol était écrit en alphabet ouïghour ancien (avec 5 voyelles et 18 consonants). Au fil du temps, cela a été adapté en une écriture verticale, appelée « Mongol Bichig », ou écriture mongole classique/traditionnelle.

Ensuite, le mongol s’est écrit successivement en Phags-pa (13<sup>e</sup> siècle), Galik (16<sup>e</sup> siècle), Soyombo (17<sup>e</sup> siècle) pour faciliter l’écriture multilingue en chinois, sanskrit ou tibétain (Simon, 2021).

Pendant une courte période dans les années 1930, le mongol a été écrit en écriture latine. Cependant, peu de temps après, l’alphabet cyrillique a été adoptée en février 1941 par le gouvernement de la République populaire de Mongolie. Cette écriture est en usage pour écrire le mongol Khalkha jusqu’aujourd’hui. Néanmoins, les Mongols de la région autonome de Mongolie intérieure en Chine utilisent toujours l’écriture mongole classique (Tserenpil & Kullmann).

Depuis 1994, des efforts ont été déployés en Mongolie pour réintroduire l’écriture classique et elle est maintenant enseignée dans une certaine mesure à l’école, bien qu’elle soit principalement utilisée à des fins décoratives par des artistes, des calligraphes et des poètes. La personne moyenne en Mongolie connaît peu ou rien de l’écriture mongole traditionnelle (Simon, 2021).

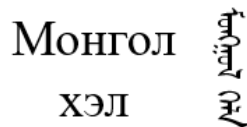


Figure 9. « Langue mongole » en écritures cyrillique et classique

La Figure 9 est un exemple de « langue mongole » écrit respectivement en écritures cyrillique et classique. Les Figures 10 et 11 (Simon, 2021) compare les lettres cyrillique et classique du mongol avec la translittération latine.

L’écriture mongole cyrillique est essentiellement le même que l’alphabet cyrillique utilisé en russe avec l’ajout de deux nouvelles lettres, « Өө » et « Үү », bien que la prononciation ne soit pas exactement la même.

Pour ce qui est de l’écriture mongole classique, il est à noter que chaque lettre a normalement trois formes en fonction de sa position : initiale, médiane et finale.

Initial	ᠠ	ᠡ	ᠢ	ᠣ	ᠤ	ᠥ	ᠦ	ᠨ
Medial	ᠠ᠊	ᠡ᠊	ᠢ᠊	ᠣ᠊	ᠤ᠊	ᠥ᠊	ᠦ᠊	ᠨ᠊
Final	ᠠ᠎	ᠡ᠎	ᠢ᠎	ᠣ᠎	ᠤ᠎	ᠥ᠎	ᠦ᠎	ᠨ᠎
Cyrillic	А	Э	И	О / У	Ө / Ү	Е	Ё / Ю	Я
Latin	A	E	I	O / U	Ö / Ü	Ye	Yo/Yu	Ya

Figure 10. Alphabet mongol (Voyelles)



Initial	ᠠ	ᠡ	ᠢ	ᠣ	ᠤ	ᠥ	ᠦ	ᠨ	ᠬ
Medial	ᠠᠢ	ᠡᠢ	ᠢᠢ	ᠣᠢ	ᠤᠢ	ᠥᠢ	ᠦᠢ	ᠨᠢ	ᠬᠢ
Final	ᠠᠨ	ᠡᠨ	ᠢᠨ	ᠣᠨ	ᠤᠨ	ᠥᠨ	ᠦᠨ	ᠨᠨ	ᠬᠨ
Cyrillic	Н	Н	Б	П	Х	Г	М	Л	Х
Latin	N	Ng	B	P	Kh, Q	Gh, γ	M	L	H
Initial	ᠭ	ᠭᠢ	ᠭᠡ	ᠭᠢ	ᠭᠣ	ᠭᠤ	ᠭᠥ	ᠭᠦ	ᠭᠨ
Medial	ᠭᠠ	ᠭᠡ	ᠭᠢ	ᠭᠣ	ᠭᠤ	ᠭᠥ	ᠭᠦ	ᠭᠨ	ᠭᠯ
Final	ᠭᠠᠨ	ᠭᠡᠨ	ᠭᠢᠨ	ᠭᠣᠨ	ᠭᠤᠨ	ᠭᠥᠨ	ᠭᠦᠨ	ᠭᠨᠨ	ᠭᠨᠬ
Cyrillic	Г	С	Ш	Т	Д	Ч	Ж	Й	Р
Latin	G	S	Sh	T	D	Ch	J	Y	R
Initial	ᠪ	ᠪᠢ	ᠪᠡ	ᠪᠢ	ᠪᠣ	ᠪᠤ	ᠪᠥ	ᠪᠦ	ᠪᠨ
Medial	ᠪᠠ	ᠪᠡ	ᠪᠢ	ᠪᠣ	ᠪᠤ	ᠪᠥ	ᠪᠦ	ᠪᠨ	ᠪᠯ
Final	ᠪᠠᠨ	ᠪᠡᠨ	ᠪᠢᠨ	ᠪᠣᠨ	ᠪᠤᠨ	ᠪᠥᠨ	ᠪᠦᠨ	ᠪᠨᠨ	ᠪᠨᠬ
Cyrillic	В	Ф	Ч / Ц	Г	К	Ц	Э	Х	Лх
Latin	V	F	Ch / Ts	G	K	Ts	Z	H	Lkh

Figure 11. Alphabet mongol (Consonants)

## 2.3. Développement de la MT du mongol

Nous évoquons dans l'introduction le manque de recherches sur la MT du mongol, surtout pour la paire de langues français-mongol. En effet, nous constatons que jusqu'à présent aucune recherche sur la MT n'a été publiée pour français-mongol (cyrillique ou classique). En revanche, il existe d'autres recherches pour mongol-anglais et mongol-chinois qui peuvent servir de références.

### 2.3.1. Recherches

[Ochir and Serguleng \(2003\)](#) ont développé un système de RBMT anglais-mongol (classique) et l'ont amélioré avec une méthode basée sur des corpus ; [Hou and Liu \(2007\)](#) ont proposé une méthode de traduction mongol<sup>17</sup>-chinois basée sur des exemples ; [Su \(2014\)](#) a étudié les caractéristiques de la paire mongol-chinois et a proposé un modèle de SMT mongol-chinois basé sur l'approche hiérarchique (*Hierarchical Phrase-based System*).

Ensuite, une série de recherches sur la traduction automatique neuronale mongol-chinois ont été menées avec le développement de l'apprentissage profond. [Wu \(2017\)](#) a réalisé un modèle de NMT mongol-chinois basé sur les réseaux de neurones récurrents avec le mécanisme d'attention. [Shen \(2017\)](#) a utilisé différents modèles *Word2Vec* et a optimisé le modèle NMT mongol-chinois. [Wang \(2018\)](#) a entraîné un modèle de NMT mongol-chinois basé sur des réseaux de neurones convolutifs et un autre modèle basé sur *Transformer*. [FAN et al. \(2018\)](#) ont proposé un modèle de NMT mongol-chinois avec de l'information préalable qui permet d'enrichir des *features* du modèle pour

<sup>17</sup> *Nota bene* : Pour la paire mongol-chinois, le mongol ici est toujours en écriture classique.

améliorer la performance de traduction. [Liu \(2018\)](#) a proposé un modèle de NMT LSTM basé sur le codage de morphèmes, qui s'avère être plus performant, surtout en termes de dépendance à long terme. [Cao \(2020\)](#) a exploré l'application du corpus monolingue en tant que complément du corpus parallèle dans la NMT mongol-chinois et a proposé trois modèles qui ont grandement amélioré la performance de traduction. [Wang et al. \(2020\)](#) ont transféré des paramètres d'un modèle de NMT anglais-chinois préentraîné à un autre système de NMT mongol-chinois.

### 2.3.2. Corpus

La langue mongole étant une langue minoritaire, son développement de la MT est étroitement lié à la construction des corpus.

En 1990, l'Institut de recherche sur la langue mongole de l'Université de Mongolie intérieure a construit la *Base de données sur la langue mongole moderne* qui comprend 1,2 millions de mots en mongol classique ([Hua, 1997](#)). [Jaimai and Chimeddorj \(2008\)](#) ont construit de façon semi-automatique un corpus à 5 millions de mots. [Zhang \(2009\)](#) a mis en place une plate-forme expérimentale qui permet de maintenir un corpus bilingue chinois-mongol et d'y rechercher des informations selon les étiquettes sur les caractéristiques linguistiques des phrases. [Bao \(2016\)](#) a créé à partir des émissions télévisées et des actualités un corpus mongol classique qui compte 1,75 millions de mots. La *China Conference on Machine Translation (CWMT)* possède également des corpus parallèles mongol-chinois pour l'évaluation annuelle de la MT à l'échelle nationale, par exemple *IMU-CWMT2013*, *IMU-CWMT2015* et *IMU-corpus-CWMT2017*, etc. ([Yang et al., 2019](#)). [Крылов \(2017\)](#) a développé *Mongolian Corpus*<sup>18</sup>, corpus mongol cyrillique en ligne permettant d'étudier le lexème, le morphème et la distance entre les mots.

À force d'efforts des chercheurs, la langue mongole a connu des progrès effectifs concernant la construction des corpus. Selon [Fei et al. \(2019\)](#), il existe aujourd'hui des corpus du mongol classique ou cyrillique de toutes sortes, tels que le *corpus des livres anciens*, le *corpus historique*, le *corpus des proverbes mongols*, le *corpus des expressions idiomatiques* et les *corpus parallèles mongol-chinois, mongol-russe, mongol-anglais, mongol-japonais*, etc. Actuellement, un corpus mongol classique de 200 millions de mots est en cours de construction.

Il est pourtant à noter qu'aucun des corpus susmentionnés n'est open-source. Ils sont, soit les propriétés intellectuelles de certaines organisations protégées par le droit d'auteur, soit les données exclusives accessibles uniquement sur l'inscription pour l'évaluation organisée par CWMT (tout abus hors de l'évaluation sera sanctionné) ([CWMT, 2017](#)). En conséquence, il ne serait possible de trouver des corpus mongols

---

<sup>18</sup> *Mongolian Corpus*, développé par Сергей Александрович Крылов, disponible sur : <http://web-corpora.net/MongolianCorpus/search/>

accessibles que sur *OPUS*<sup>19</sup>, collection des corpus parallèles open-source ([Tiedemann, 2012](#)).

### 2.3.3. Moteurs, plates-formes et outils

Il est de notoriété que *Google Traduction*<sup>20</sup> est aujourd'hui le moteur de traduction plus utilisé. Cela est dû à sa gratuité, sa précision et au fait qu'il prend en charge le plus de langues, y compris, dans notre cas, la paire de langues peu dotée français-mongol (cyrillique). Il y a de plus deux moteurs de traduction commerciaux payants qui prennent en charge cette paire, *NiuTrans*<sup>21</sup> et *PoliLingua*<sup>22</sup>. Toujours est-il qu'aucune publication ne vienne décrire le fonctionnement ou l'évaluation du système pour cette paire de langues en particulier.

Certaines entreprises en Mongolie intérieure ont développé des plates-formes et outils pour promouvoir le développement de la MT du mongol : *Oyun*<sup>23</sup> est une plate-forme multifonctionnelle pour chinois-mongol (classique et cyrillique), qui fournit la MT textuelle et vocale, l'OCR, la synthèse de la parole, etc. ; *Yijinyun*<sup>24</sup> est une autre plate-forme similaire à *Oyun* offrant presque les mêmes fonctionnalités au niveau logiciel, mais il se concentre davantage sur le développement des matériels tels que l'encreinte intelligente, le robot et le traducteur portable, etc. En juillet 2021, *Yijinyun* a publié un stylo traducteur (mongol-chinois-japonais-russe-coréen-tibétain-anglais) qui fait fureur dans le cercle de chercheurs en MT du mongol ; *Menksoft*<sup>25</sup> est une entreprise clé pour la numérisation du mongol, qui se concentre sur la méthode de saisie, les polices, l'encodage et la bureautique en écriture classique ([Ao et al., 2011](#)).

### 2.3.4. Défis

Les défis auxquels est confrontée la MT du mongol peuvent être attribués aux facteurs suivants : manque de ressources, encodage incohérent (du mongol classique), création difficile des corpus parallèles à grande échelle, recherche insuffisante sur la langue mongole ([Wu, 2017](#)).

---

<sup>19</sup> *OPUS : The Open Parallel Corpus*, disponible sur : <https://opus.nlpl.eu/>

<sup>20</sup> *Google Traduction*, disponible sur : <https://translate.google.fr/>

<sup>21</sup> *NiuTrans*, disponible sur : <https://niutrans.com/>

<sup>22</sup> *PoliLingua*, disponible sur : <https://www.polilingua.com/>

<sup>23</sup> *Oyun*, disponible sur : [www.nmgoyun.com](http://www.nmgoyun.com)

<sup>24</sup> *Yijinyun*, disponible sur : <https://www.mengguyu.cn/>

<sup>25</sup> *Menksoft*, disponible sur : <http://www.menksoft.com/>

## 3. EXPÉRIMENTATION

Une fois l'état de l'art mis en exergue, nous passons à l'expérimentation qui consiste à mettre en place un système de NMT pour la paire de langues français-mongol avec des ressources open-source existantes.

Pour donner une vue d'ensemble, nous récapitulons ici rapidement notre travail effectué : nous avons choisi l'outil *OpenNMT*<sup>26</sup> comme notre cadre de développement (*framework*) et avons récupéré un corpus parallèle français-mongol en format .tmx depuis *OPUS* dans le projet *MultiCCAligned v1.1* ; ensuite, nous avons préparé le corpus aux exigences d'*OpenNMT* ; après de longues périodes d'entraînement, nous avons obtenu quatre modèles de traduction, lesquels ont été finalement menés à deux évaluations automatiques et une évaluation humaine originale nommée « *IFF* » pour en savoir les forces et les faiblesses par rapport à certains moteurs de traduction commerciaux existants.

### 3.1. Outils

La mise en place d'un système de NMT implique l'apprentissage automatique. Les outils nécessaires sont tant logiciels que matériels.

- Logiciel : une boîte à outils (*toolkit*) ou un cadre de développement (*framework*) qui structure l'expérimentation et fournit la base du projet ; un outil d'évaluation automatique
- Matériel : un serveur qui héberge les données ainsi que le cadre de développement prêt pour l'apprentissage automatique

Les outils les plus importants dans notre expérimentation sont : *OpenNMT*, *Tilde's Interactive BLEU Score Evaluator* et le serveur *Helium*.

#### 3.1.1. *OpenNMT*

*OpenNMT*, lancé en 2016, est le premier *toolkit* open-source à large public pour concevoir, entraîner et déployer des modèles de traduction automatique neuronale. Il prend en charge plusieurs architectures de modèles telles que *Transformer* et RNN etc., et également des tâches connexes telles que la génération automatique de textes (GAT) et la modélisation des langues naturelles. Les objectifs de l'initiative d'*OpenNMT* sont la facilité d'utilisation, l'efficacité, la modularité, l'extensibilité et la disponibilité de production (Klein et al., 2020).

Il existe deux versions d'implémentation :

---

<sup>26</sup> *OpenNMT*, disponible sur : <https://opennmt.net/>

- *OpenNMT-py* : implémentation simple et multimodale bénéficiant de la facilité d'utilisation et de la polyvalence de *PyTorch*.
- *OpenNMT-tf* : implémentation modulaire et stable basée sur *TensorFlow 2.0*.

Toutes les deux versions fournissent une bibliothèque *Python* permettant aux utilisateurs de configurer, entraîner et exécuter des modèles simplement par la ligne de commande.

Pour notre expérimentation, nous choisissons *OpenNMT-py* comme notre cadre de développement.

### 3.1.2. Tilde's Interactive BLEU Score Evaluator

*Tilde*<sup>27</sup> est une entreprise de technologie linguistique européenne qui offre des outils pour la traduction, la localisation et des activités de R&D en langue, dont *Interactive BLEU Score Evaluator*<sup>28</sup> (Évaluateur de score BLEU interactif).

Figure 12. Interface de l'Évaluateur de score BLEU interactif

*Interactive BLEU Score Evaluator* permet de calculer les scores BLEU (individuel ou cumulatif à 1-4 gram) et de visualiser les résultats dans un graphique dynamique phrase par phrase, offrant ainsi une comparaison des performances de différents moteurs de traduction automatique.

<b>BLEU:</b>	<b>8.60</b>	<b>8.46</b>
<b>Precision x brevity:</b>	9.82 x 87.58	9.78 x 86.56
<b>Type</b>	<b>1-gram 2-gram 3-gram 4-gram</b>	<b>1-gram 2-gram 3-gram 4-gram</b>
<b>Individual</b>	25.71 11.65 6.78 4.59	25.90 11.65 6.72 4.51
<b>Cumulative</b>	22.51 15.15 11.09 8.60	22.42 15.04 10.96 8.46
<b>Export data</b>	<b>CSV</b>	

Figure 13. Exemple d'évaluation

<sup>27</sup> *Tilde*, disponible sur : <https://www.tilde.com/>

<sup>28</sup> *Interactive BLEU Score Evaluator*, disponible sur : <https://www.letsmt.eu/Bleu.aspx>

La *Figure 12* montre son interface où l'on peut charger les fichiers de traduction par l'humaine et la machine. Un clique sur « Score » permet d'obtenir le score BLEU de la traduction machine chargée. La *Figure 13* est un exemple d'évaluation. L'ensemble des scores pour chaque phrase peuvent être exportées en cliquant sur le bouton « CSV ».

Nous avons choisi d'utiliser cet outil pour notre expérimentation parce qu'au lieu d'obtenir un simple score, nous voudrions nous pencher sur des exemples concrets en vue d'une analyse détaillée.

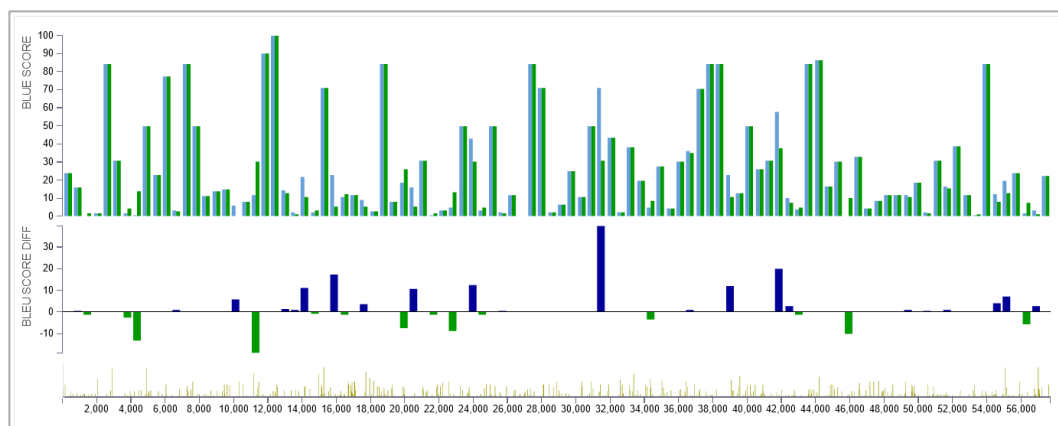


Figure 14. Exemple des scores BLEU visualisés

Quand nous mettons deux traductions machine, l'évaluateur génère un graphe comme montré dans la *Figure 14* qui visualise dynamiquement les scores de deux traductions produites par deux systèmes permettant de comparer leurs performances de manière visible avec la différence (BLEU SCORE DIFF). Nous pouvons glisser horizontalement sur ce graphe dynamique pour voir d'autres phrases qui se trouvent hors de la vue actuelle.

<< < 21 > >> From 1 to 100 of all 57725 sentences.			
Sentence	BLEU	Length ratio	Text
Human	100.00	1.00	Абacre Бараа материалын удирдлага , хяналт нь Windows нь бараа материалын программ хангамж нь шинэ үе юм . Энэ нь хөдөлмөрийн менежмент захиалга , нэхэмжлэл , худалдан авах , бараа материал авах бүх үйл ажиллагааг хамарна . хэрэглэгчийн интерфэйс нь анхааралтай өндөр хурдны үйлчлүүлэгчийн тулд оролт , нийтлэг <b>алдаанаас</b> урьдчилан сэргийлэх оновчтой байна . Энэ нь хэд хэдэн компьютер дээр ашиглаж <b>зориулалттай</b> , найдвартай , аюулгүй зөвшөөрөл хэмжээгээр агуулдаг юм .
Machine	89.83	0.97	Бараа материалын удирдлага , хяналт нь Windows нь бараа материалын программ хангамж нь шинэ үе юм . Энэ нь хөдөлмөрийн менежмент захиалга , нэхэмжлэл , худалдан авах , бараа материал авах бүх үйл ажиллагааг хамарна . хэрэглэгчийн интерфэйс нь анхааралтай өндөр хурдны үйлчлүүлэгчийн тулд оролт , нийтлэг <unk> урьдчилан сэргийлэх оновчтой байна . Энэ нь хэд хэдэн компьютер дээр ашиглаж <unk> найдвартай , аюулгүй зөвшөөрөл хэмжээгээр агуулдаг юм .
Machine	89.83	0.97	Бараа материалын удирдлага , хяналт нь Windows нь бараа материалын программ хангамж нь шинэ үе юм . Энэ нь хөдөлмөрийн менежмент захиалга , нэхэмжлэл , худалдан авах , бараа материал авах бүх үйл ажиллагааг хамарна . хэрэглэгчийн интерфэйс нь анхааралтай өндөр хурдны үйлчлүүлэгчийн тулд оролт , нийтлэг <unk> урьдчилан сэргийлэх оновчтой байна . Энэ нь хэд хэдэн компьютер дээр ашиглаж <unk> найдвартай , аюулгүй зөвшөөрөл хэмжээгээр агуулдаг юм .

Figure 15. Comparaison phrase par phrase

En cliquant sur un score affiché dans la *Figure 14*, on voit en-dessous une comparaison phrase par phrase (*Figure 15*), ce qui facilite une étude basée sur des exemples concrets.

### 3.1.3. Serveur *Helium*

Toute notre expérimentation se déroule sur « *Helium* », serveur distant possédé par l'Université Paris Descartes dont l'accès nous est fourni par l'*ERTIM*<sup>29</sup>.

*Helium* est un serveur sans GPU, mais possède plusieurs CPU suffisamment puissants pour y construire un environnement d'apprentissage automatique par réseaux de neurones. Mais il se peut que l'entraînement soit assez long. Voici quelques informations clés sur la configuration du serveur dans le *Tableau 6*.

<b>Architecture</b>	X86_64
<b>CPU</b>	Intel Xeon E5-2690 v3 @ 2.60GHz
<b>RAM</b>	32 GB
<b>HDD</b>	80 GB

*Tableau 6. Configuration du serveur Helium*

## 3.2. Ressources

Pour construire notre propre modèle de traduction par l'apprentissage automatique, un corpus parallèle est théoriquement obligatoire. Au tout début de l'expérimentation, nous prévoyions la difficulté d'en trouver eu égard au manque de recherches connexes. Heureusement, après une fouille chronophage, nous avons pu récupérer un corpus parallèle français-mongol sur le site *OPUS*.

### 3.2.1. *OPUS*

*OPUS* est une collection croissante de ressources linguistiques. Dans le projet *OPUS*, les auteurs essaient de convertir et d'aligner des données en ligne gratuites, d'ajouter des annotations linguistiques et de fournir un corpus parallèle accessible au public. L'objectif global du projet *OPUS* est de rendre disponibles gratuitement des ressources parallèles, en mettant particulièrement l'accent sur des langues peu dotées ([Tiedemann, 2012](#)).

Les données originaires du corpus récupéré depuis *OPUS* sont collectées par le projet *MultiCCAligned v1.1*.

### 3.2.2. *MultiCCAligned v1.1*

La collection *MultiCCAligned v1.1* a été créée à partir de 68 *Common Crawl Snapshots*<sup>30</sup> (jusqu'en mars 2020). Les documents sont découpés en phrases en fonction des ponctuations et une déduplication est effectuée. Les paires de phrases ont été

<sup>29</sup> *ERTIM* : Équipe d'Accueil au carrefour de l'ingénierie multilingue et du Traitement Automatique des Langues, son site web disponible sur : <http://www.inalco.fr/recherche/ertim>

<sup>30</sup> *Common Crawl* : disponible sur : <https://commoncrawl.org/>

extraites à l'aide des scores de similarité des intégrations *LASER*<sup>31</sup> des paires de documents (similarité minimale 1,04, triée en fonction du score de similarité décroissant). Il manque certaines langues non couvertes par *LASER* (El-Kishky et al., 2020). Tout le prétraitement est effectué automatiquement et aucune correction manuelle n'a été effectuée.

### 3.2.3. Corpus parallèle français-mongol

Le corpus parallèle français-mongol est en format .tmx (*Translation Memory eXchange*). Il contient 288 623 paires de phrases. La *Tableau 7* montre un aperçu du corpus ouvert avec *Olifant*<sup>32</sup>.

Key	Flag	Text_FR	Text_MN
288608	<input type="checkbox"/>	Les visions passées	Нууц үгээ мартсан уу?
288609	<input type="checkbox"/>	Rouleaux de la materail	хамгийн сүүлийн үеийн мэдээ
288610	<input type="checkbox"/>	Enceintes (12 Live)	Бондиц (158 Live)
288611	<input type="checkbox"/>	Fétichisme des pieds	POV өсвөр Насныхан
288612	<input type="checkbox"/>	Méthode d'installation correcte Inccrect méthode d'installation	Арга Inccrect арга суулгах засах
288613	<input type="checkbox"/>	À 317.3 km / 197.2 milles	90 - 100 км/ч
288614	<input type="checkbox"/>	SIMARGL QUANT FUND	Forex Trading дохионууд
288615	<input type="checkbox"/>	Ascenseur télescopique de boom	Дулаан болон дуу тусгаарлалт
288616	<input type="checkbox"/>	&lt; start="928.259" dur="4.339"> LE PARC MOI MAIS C'E...	&lt; start="928.259" dur="4.339"> ПАРК ХҮНИЙГ ӨӨРИЙ...
288617	<input type="checkbox"/>	13 Fente Pages Casino - Casino Téléphone £, €, à partir de \$, ...	16 Хуудсууд казино Slot - Гар утас Казино £, €, \$ авсан, £ ...
288618	<input type="checkbox"/>	CE Y ISO9001	200 кг/ч
288619	<input type="checkbox"/>	Gimåfors Tennis Club: Les dirigeants Ingemar Holmvall 070-33...	Холбоо барих: 060-910 09, Ассамблей / Indalens вэб pasto...
288620	<input type="checkbox"/>	&lt; start="893.725" dur="2.936"> POUVEZ TROUVER L'É...	&lt; start="893.725" dur="2.936"> ТОНОГ ТӨХӨӨРӨМЖ...
288621	<input type="checkbox"/>	65880 Parti	762143 Тийн
288622	<input type="checkbox"/>	voiture 10 Porte - Fusible	Car 10 ТХГН-ийн гол эзэмшигч
288623	<input type="checkbox"/>	G9 18-40W Ampoules halogènes Four Lampe ROHS Four Am...	G9 18-40W галоген чийдэн Зуух гэрэл ROHS CE Зуух Stea...

Tableau 7. Corpus parallèle ouvert avec *Olifant*

Après avoir survolé le contenu du corpus, nous remarquons que certaines entrées sont des phrases complètes normales tandis que d'autres sont des bouts de phrase, des mots propres, des locutions, du code ou des textes très techniques. En outre, il existe des entrées avec des caractères spéciaux pas encore normalisées et certaines entrées non alignées. La qualité du corpus est donc médiocre. Une correction humaine serait bien trop longue à réaliser, nous conservons donc le corpus en l'état pour notre travail.

Nous constatons également que son thème est très probablement « les publicités sur internet ». Pour autant que nous avons exploré, les domaines couverts sont : médecine, informatique, enseignement supérieur, technologie, chimie, jeux vidéo, commerce, tourisme, santé, météorologie, musique, publicité, peinture, électroménager, stylisme, automobile, etc.

<sup>31</sup> *LASER* : *Language-Agnostic SEntence Representations*, disponible sur : <https://github.com/facebookresearch/LASER>

<sup>32</sup> *Olifant* : un éditeur .NET de mémoire de traduction, disponible sur : <http://okapi.sourceforge.net/Release/Olifant/ReadMe.htm>



En vue d’une compréhension plus intuitive du corpus, nous avons recours au nuage de mots (*Word Cloud*). Pour ce faire, il faut d’abord éliminer les mots vides. Les listes des mots vides sont trouvées sur *GitHub* pour le français<sup>33</sup> et le mongol<sup>34</sup>, auxquelles nous avons apporté des améliorations en y ajoutant respectivement 1 et 46 entrées.

Voici deux nuages de mots français (*Figure 16*) et mongols (*Figure 17*) générées par la librairie *Python* « *WordCloud 1.8.1*<sup>35</sup> ».

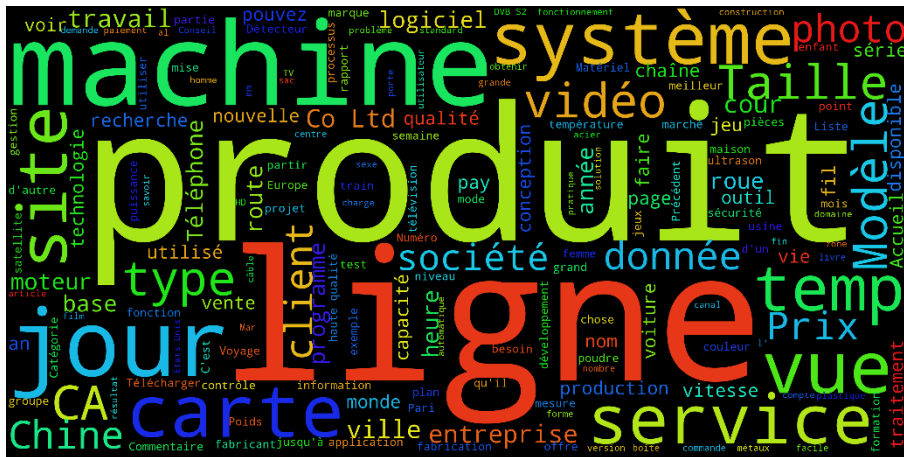


Figure 16. WordCloud du corpus français

Les mots les plus importants sont : « produit », « ligne », « machine ». Nous voyons aussi « site », « service », « entreprise » et « prix », etc. Dans les grandes lignes, cette représentation visuelle semble cohérente avec notre premier constat sur le thème du corpus.



Figure 17. WordCloud du corpus mongol

<sup>33</sup> Stopwords\_fr, disponible sur : <https://github.com/stopwords-iso/stopwords-fr>

<sup>34</sup> Stopwords\_mn, disponible sur : [https://github.com/tugstugi/mongolian-nlp/blob/master/datasets/most\\_frequent\\_words.csv](https://github.com/tugstugi/mongolian-nlp/blob/master/datasets/most_frequent_words.csv)

<sup>35</sup> *WordCloud*, disponible sur : <https://pypi.org/project/wordcloud/>

Il est *a priori* de même pour le corpus mongol. Malgré de légères différences, nous trouvons également des mots clés similaires tels que « татаж » (télécharger), « хуудас » (page), « машин » (machine), « компани » (entreprise), « холбоо » (communication).

### 3.3. Mise en place du prototype de NMT

Les outils et les ressources présentés, nous commencerons à mettre en place un prototype de NMT afin de comparer sa performance avec les moteurs de traduction à usage commercial. Le processus de l'expérimentation peut être décrit par le schéma ci-dessous (Figure 18) :

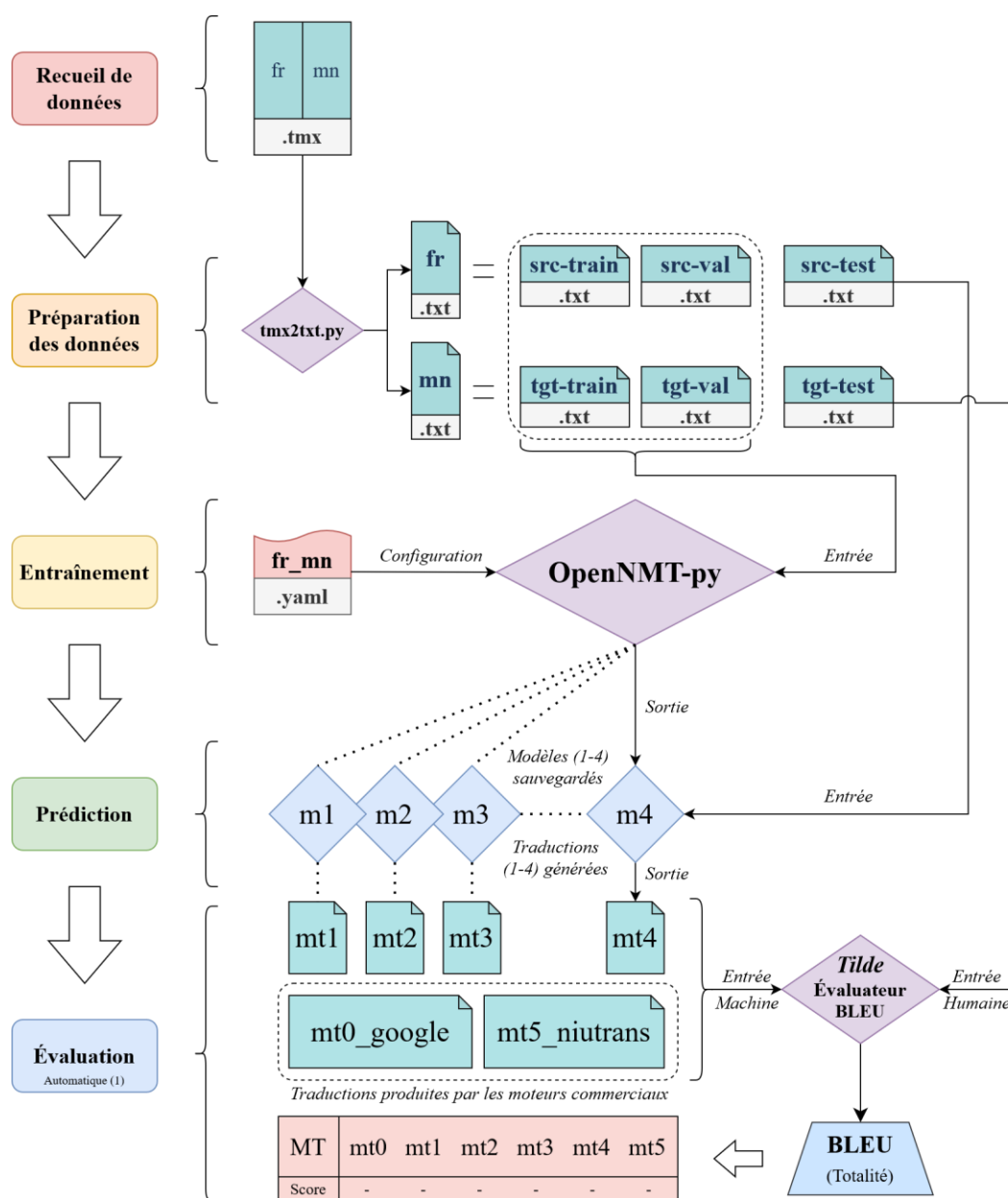


Figure 18. Schéma général de l'expérimentation

### 3.3.1. Recueil de données

Nous sommes allés sur le site *OPUS* pour télécharger le corpus français-mongol. Comme montré dans la *Figure 19*, il existe plusieurs corpus disponibles sous différents formats.

Search & download resources:     show all versions

Language resources: click on [ [tmx](#) | [moses](#) | [xces](#) | [lang-id](#) ] to download the data! (raw = untokenized, ud = parsed with universal dependencies, alg = word alignments and phrase tables)

<http://localhost/opusapi/?source=fr&target=mn&version=latest>

corpus	doc's	sent's	fr tokens	mn tokens	XCES/XML	raw	TMX	Moses	mono	raw	ud	alg	dic	freq	other files
JW300 v1c	1678	0.2M	3.7M	2.8M	xces fr mn	fr mn			fr mn	fr mn				fr mn	sample
MultiCCAligned v1.1	1	0.3M	3.1M	2.6M	xces mn	fr mn	tmx	moses	fr mn	fr mn				fr mn	sample
GNOME v1	649	0.2M	1.4M	1.2M	xces fr mn	fr mn	tmx	moses	fr mn	fr mn		alg smt		fr mn	sample
TED2020 v1	266	24.5k	0.5M	0.4M	xces fr mn	fr mn	tmx	moses	fr mn	fr mn				fr mn	sample
Mozilla-I10n v1	1	1.0k	0.7M	7.5k	xces fr mn	fr mn			fr mn	fr mn				fr mn	sample
QED v2.0a	174	23.3k	0.3M	0.2M	xces fr mn	fr mn	tmx	moses	fr mn	fr mn		alg smt	dic	fr mn	sample
XLEnt v1.1	1	40.0k	93.1k	94.5k	xces fr mn	fr mn	tmx	moses	fr mn	fr mn				fr mn	sample
Tatoeba v2021-07-22	1	18	0.1k	98	xces fr mn	fr mn	tmx	moses	fr mn	fr mn				fr mn	sample
Ubuntu v14.10					xces fr mn	fr mn	tmx	moses	fr mn	fr mn			dic	fr mn	sample
<b>total</b>	<b>2771</b>	<b>0.8M</b>	<b>9.8M</b>	<b>7.4M</b>	<b>0.8M</b>	<b>0.4M</b>	<b>0.6M</b>								

Figure 19. Téléchargements des corpus fr-mn disponibles sur OPUS

Ce dont nous avons besoin est un corpus parallèle, nous avons donc téléchargé la version en .tmx du corpus *MultiCCAligned v1.1* qui compte 288 623 entrées avec une taille de 76,4 Mo.

Pour l'entraînement, il faut généralement diviser la mémoire de traduction en deux fichiers monolingues. Nous remarquons que sur la page de téléchargements il existe également des fichiers monolingues dans la colonne « mono », susceptibles d'avoir une plus grande quantité et de nous éviter la division. Cependant, après une vérification, ces fichiers ne sont pas alignés et n'ont pas le même nombre d'entrées (fr : 102415642 lignes ; mn : 506785 lignes). Nous revenons donc vers la mémoire de traduction.

Un prétraitement du corpus en .tmx est donc nécessaire.

### 3.3.2. Préparation des données

Pour que le corpus soit lisible par *OpenNMT*, nous avons fait un prétraitement conformément à ses exigences de format.

Nous avons d'abord écrit un script *Python* « [tmx3txt.py](#) » pour transformer le fichier .tmx parallèle en deux fichiers .txt monolingues. Ensuite, nous avons employé deux rapports de division (abrégés « rd » ci-après) pour diviser respectivement ces deux corpus monolingues en « *train* » (ensemble d'entraînement), « *val* » (ensemble de validation) et « *test* » (ensemble de test) :

- rd1 = *train* : *val* : *test* = 6 : 2 : :2
- rd2 = *train* : *val* : *test* = 7 : 1 : :2

Ceci a pour but de vérifier l'influence de l'augmentation du *train* sur la performance. Nous en discuterons dans la partie [Discussion](#).

Voici un récapitulatif sur la répartition du jeu de données (*Tableau 8*) :

rd	train	val	test
6 : 2 : :2	172 892	57 725	57 725
7 : 1 : :2	201 754	28 861	57 725

Tableau 8. Jeu de données

### 3.3.3. Entraînement

Nous nous sommes connectés au serveur *Helium* via le protocole *Secure Shell* (*SSH*) pour y déposer les données avec « *rsync* » et créer un environnement de développement virtuel.

L'entraînement sur *OpenNMT* se fait simplement avec un fichier de [configuration](#) et des [commandes](#) : « *src-train* », « *tgt-train* », « *src-val* » et « *tgt-val* » passent par *OpenNMT* en entrée et le fichier de configuration « *fr\_mn.yaml* » donne au système des hyperparamètres pour configurer l'apprentissage. Nous utilisons *Transformer* pour l'encodeur et le décodeur avec respectivement six couches identiques.

Nous avons fait quatre entraînements avec la configuration de base<sup>36</sup> de *Transformer* fournie dans la documentation d'*OpenNMT*. La présentation de certains hyperparamètres clés se trouve dans l'annexe.

Dénomination	Rapport de division	Nombre d'étapes
m1_rd1_5000	6 : 2 : 2	5 000
m2_rd1_50000	6 : 2 : 2	50 000
m3_rd2_50000	7 : 1 : 2	50 000
m4_rd2_100000	7 : 1 : 2	100 000

Tableau 9. Modèles de traduction sauvegardés

Les entraînements terminés, nous avons obtenu quatre modèles de traduction, nommé « *m1* », « *m2* », « *m3* » et « *m4* », décrits dans le *Tableau 9*.

### 3.3.4. Prédiction

Une fois modèles sauvegardés, il est temps de faire des prédictions pour les tester : « *src-test* » passe respectivement par les quatre modèles en entrée. Les prédictions/traductions sont générées (*Figure 20*), nommées « *mt1* », « *mt2* », « *mt3* » et « *mt4* ».





Nom	Taille
 mt1_rd1_5000.txt	6 230 Ko
 mt2_rd1_50000.txt	3 991 Ko
 mt3_rd2_50000.txt	4 047 Ko
 mt4_rd2_100000.txt	4 017 Ko

Figure 20. Fichiers de traductions générées à partir de *src-test*

<sup>36</sup> Configuration de base, disponible sur : <https://opennmt.net/OpenNMT-py/examples/Translation.html>

### 3.3.5. Évaluation

Nous procédons ensuite à l'évaluation pour ces quatre traductions de manière à comparer les performances des quatre modèles entraînés. Pour ce faire, nous avons d'abord effectué deux évaluations automatiques, l'une pour la totalité du *test set*, l'autre pour une sélection du *test set* composée de 50 phrases complètes. Nous utilisons le score BLEU car c'est la métrique automatique la plus utilisée. Ensuite, nous avons proposé une méthode d'évaluation humaine nommée « *IFF* » pour cette sélection.

#### 3.3.5.1. Évaluation automatique (1)

Le principe de notre première évaluation automatique est décrit dans la [Figure 18](#). En vue d'une comparaison qualitative, nous avons préparé « mt0 » et « mt5<sup>37</sup> », traductions de « src-test » produites par les moteurs commerciaux, *Google* et *NiuTrans*<sup>38</sup>. Les six traductions passent respectivement par l'Évaluateur de score BLEU interactif de *Tilde* comme « entrée machine » tandis que « tgt-test » y passe comme « entrée humaine ».

Nous avons ainsi obtenu les scores BLEU de ces six traductions pour la totalité du *test set* ([Figure 21](#)). Il faut normalement faire référence aux scores BLEU cumulatifs à 1-4 gram lors de la description de la performance d'un système de génération de texte ([Brownlee, 2017](#)).

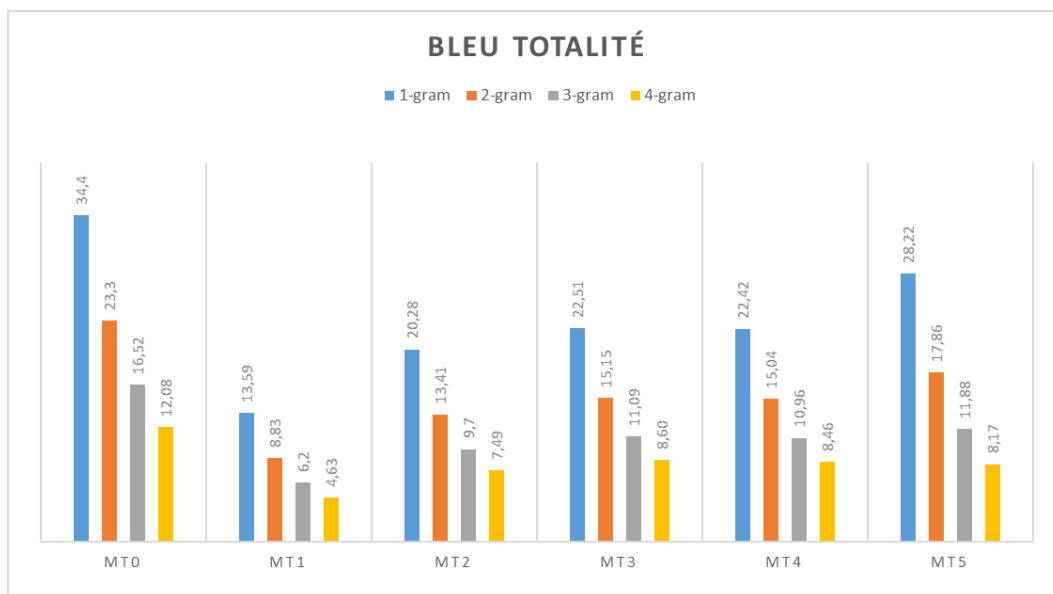


Figure 21. Comparaison des scores BLEU pour la totalité du *test set*

<sup>37</sup> *Nota bene* : L'ajout de « mt5 » a été effectué à la fin de notre expérimentation et après que l'évaluation humaine avait été terminée. Avant, nous n'avions pas connaissance de l'existence de *NiuTrans* donc ne savions pas que ce moteur de traduction prenait en compte la rare paire de langues français-mongol. En conséquence, nous n'avons pas eu le temps de l'ajouter pour l'évaluation humaine mais que pour l'évaluation automatique.

<sup>38</sup> *NiuTrans* : moteur de traduction développé par l'Université du Nord-Est de la Chine, disponible sur : <https://niutrans.com/trans?type=text>

À la première vue des résultats de l'évaluation automatique, nous constatons que les traductions de certains de nos modèles, à savoir « mt3 » et « mt4 », sont déjà comparables à « mt5 » produite par *NiuTrans*, voire supérieures si l'on regarde le BLEU-4. Et « mt2 », différente de « mt3 » seulement en rapport de division, s'avère moins bonne. Pour ce qui est de « mt1 », son modèle « m1 » n'est entraîné qu'à 5 000 étapes et sert en fait de *toy model* (modèle réduit) pour assurer le bon fonctionnement et comprendre les mécanismes du cadre de développement *OpenNMT-py*. Et on y constate la confirmation que le nombre d'itérations impacte fortement la performance. En ce qui concerne « mt0 », rien d'étonnant à ce que *Google Traduction* soit toujours le meilleur moteur de traduction.

Les résultats de l'évaluation automatique (1) sur la totalité du *test set* sont plutôt satisfaisants.

### 3.3.5.2. Évaluation automatique (2)

Étant donné que le corpus n'est pas d'une bonne qualité, nous craignons que l'évaluation automatique (1) sur la totalité du *test set* soit unilatérale. Aussi avons-nous procédé à une autre évaluation automatique axée sur une sélection de 50 phrases complètes, dont la procédure est décrite dans le schéma ci-dessous (*Figure 22*).

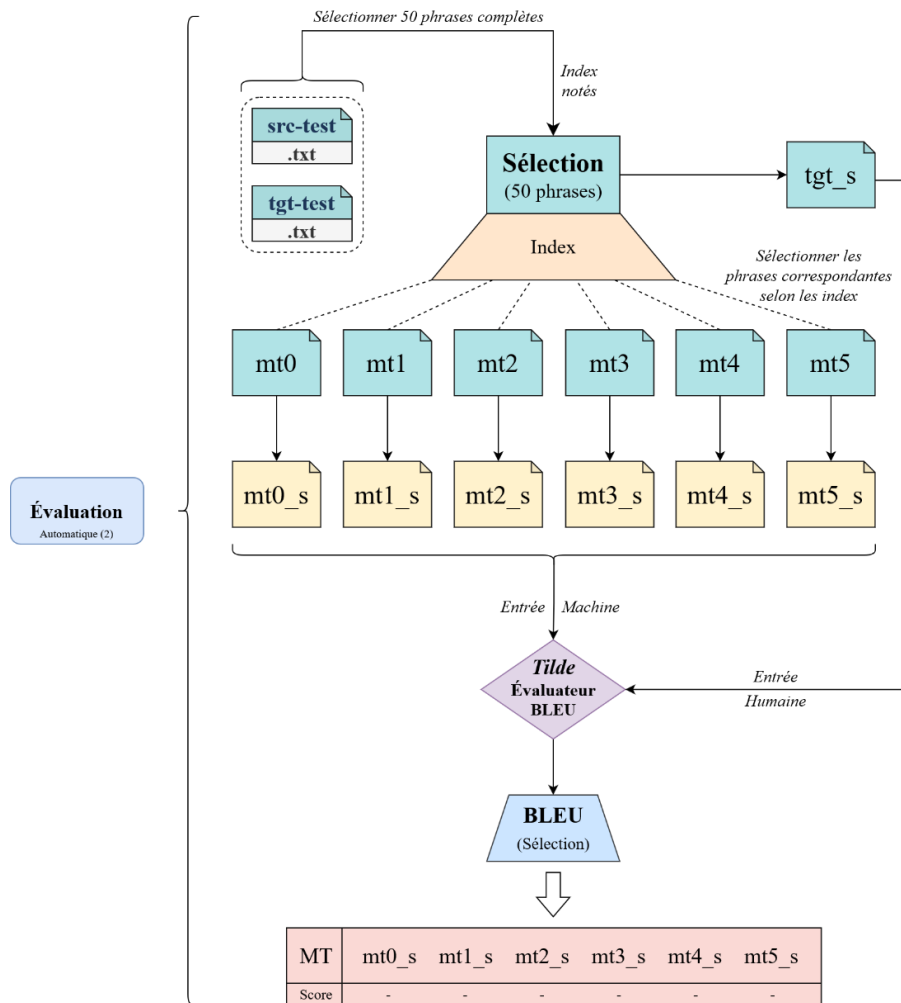


Figure 22. Schéma de l'évaluation automatique (2) pour la sélection du test set

La sélection a été faite manuellement à partir du *test set*, c'est-à-dire 50 parmi 57 725 lignes. Leurs index ont été notés afin de pouvoir facilement récupérer les phrases correspondantes dans « tgt-test » et dans les productions des modèles/moteurs de traduction. D'où « tgt\_s » et les « mtx\_s<sup>39</sup> », qui passent ensuite par l'Évaluateur de score BLEU interactif de *Tilde* comme « entrée humain » et « entrée machine ». Nous avons ainsi obtenu les scores BLEU pour la sélection (*Figure 23*).

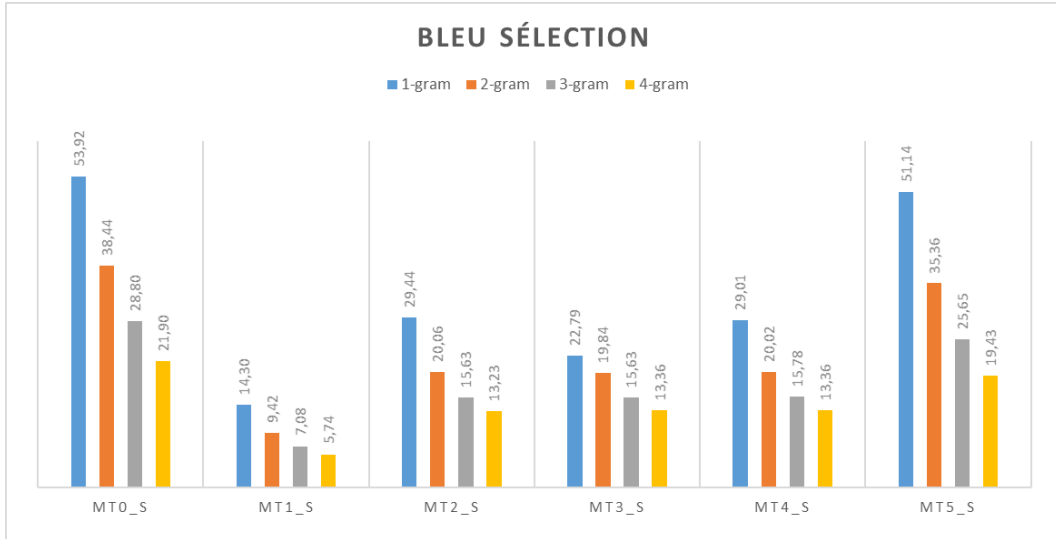


Figure 23. Comparaison des scores BLEU pour la sélection du test set

Nous constatons tout de suite que tous les scores ont largement augmenté, surtout ceux de *Google* et *NiuTrans*, et que nos modèles, même s'ils ont également obtenu des performances bien meilleures, restent à bonne distance des deux moteurs commerciaux. Il n'y a pas trop de différences entre nos trois derniers modèles.

Les résultats de l'évaluation automatique (2) sur la sélection du *test set* sont moins satisfaisants.

### 3.3.5.3. Évaluation humaine

Afin d'assurer la fiabilité de l'évaluation automatique (2) et de mieux comprendre les résultats, nous avons conçu une méthode d'évaluation humaine nommée « *IFF* » pour la sélection du *test set*. La méthodologie est illustrée dans la *Figure 24*.

Il fallait d'abord recruter des évaluateurs. Pour ce faire, nous avons publié sur plusieurs réseaux sociaux un avis dans le but de recruter des locuteurs natifs du mongol en cyrillique qui sachent de préférence parler également le français. Finalement, seulement six locuteurs ont accepté de participer à l'évaluation humaine, mais ils ne sont pas tous bilingues français-mongol comme escompté.

Parmi eux, deux évaluateurs viennent de la République populaire de Mongolie, qui travaillent en Europe et qui sont bilingues français-mongol, tandis que les quatre autres évaluateurs viennent de la Région autonome de Mongolie intérieure de la Chine,

<sup>39</sup> mtx\_s : traductions sélectionnées, « s » pour « sélection » ; « x » = 1, 2, 3, 4, 5

dont l’une est trilingue chinois-français-mongol, les trois autres sont bilingues chinois-mongol et ne savent pas parler le français. Il est à noter que la langue maternelle des quatre derniers n’est pas le mongol cyrillique mais traditionnel et qu’ils ont appris l’écriture cyrillique dans leurs vies scolaires ou professionnelles, ce qui sera pris en considération quand nous mettrons en commun leurs retours. Afin d’aider les non-françophones à comprendre le contenu du texte de source, nous ajoutons la traduction automatique en chinois comme une version intermédiaire.

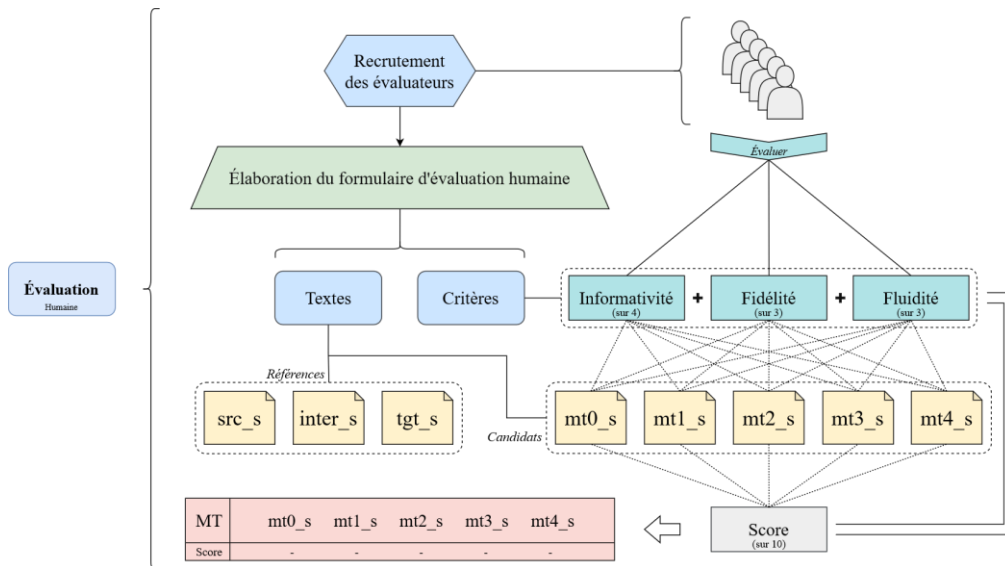


Figure 24. Schéma de l'évaluation humaine « IFF » pour la sélection du test set

Ensuite, nous avons élaboré le formulaire d'évaluation (Tableau 10) à partir de la sélection du *test set*. Le formulaire se compose de deux parties : textes et critères. Notre méthode est appelée « IFF », qui est un sigle de « Informativité », « Fidélité » et « Fluidité », trois attributs que nous définissons pour une traduction.

评测 Évaluation 文本代号 Nom des textes	文本 5 Texte 5	信息度 Informativité	忠实度 Fidélité	流畅度 Fluidité	总分 Score
src_s	Tous les transferts et téléchargements sont cryptés à l'aide du cryptage SSL 256 bits. Ainsi, les données de vos documents Excel et PDF ne seront pas susceptibles d'être soumises à un accès non autorisé.	X			
inter_s	所有传输和下载均使用 256 位 SSL 加密进行加密。这将防止您的 Excel 和 PDF 文档中的数据受到未经授权访问。				
tgt_s	Бүх байршуулалт болон таталтыг 256 битийн SSL кодчиллол ашиглан шифрлэдэг. Үүнийг хийснээр Excel болон PDF баримтаас өгөгдөл нь зөвшөөрөлгүй хандалтанд өртөмтгий биш болно.				
mt0_s	Бүх дамжуулалт, татан авалтыг 256 битийн SSL шифрлэлт ашиглан шифрлэдэг. Энэ нь таны Excel болон PDF баримт бичигт байгаа өгөгдлийг зөвшөөрөлгүй хандахаас урьдчилан сэргийлэх болно.	4	2	3	9
mt1_s	Бүх байршуулалт болон таталтыг 256 битийн SSL кодчиллол ашиглан шифрлэдэг. Үүнийг хийснээр PDF болон <unk> өртөмтгий биш болно.	2	1	1	4
mt2_s	Бүх байршуулалт болон таталтыг 256 битийн SSL кодчиллол ашиглан шифрлэдэг. Үүнийг хийснээр PDF болон PPT <unk> өгөгдөл нь зөвшөөрөлгүй <unk> өртөмтгий биш болно.	3	2	2	7
mt3_s	Бүх байршуулалт болон таталтыг 256 битийн SSL кодчиллол ашиглан шифрлэдэг. Үүнийг хийснээр PDF болон <unk> өгөгдөл нь зөвшөөрөлгүй <unk> өртөмтгий биш болно.	3	2	2	7
mt4_s	Бүх байршуулалт болон таталтыг 256 битийн SSL кодчиллол ашиглан шифрлэдэг. Үүнийг хийснээр PDF болон <unk> өгөгдөл нь зөвшөөрөлгүй <unk> өртөмтгий биш болно.	3	2	2	7

Tableau 10. Exemple d'un groupe du formulaire d'évaluation « IFF »



Il y a au total 50 groupes de textes dans le formulaire. Chaque groupe contient huit éléments :

- 1) **src\_s** (texte source en français)
- 2) **inter\_s** (texte intermédiaire en chinois)
- 3) **tgt\_s** (texte cible en mongol cyrillique)
- 4) **mt0\_s** (traduction de *Google* en mongol cyrillique)
- 5) **mt1\_s** (traduction N°1 en mongol cyrillique)
- 6) **mt2\_s** (traduction N°2 en mongol cyrillique)
- 7) **mt3\_s** (traduction N°3 en mongol cyrillique)
- 8) **mt4\_s** (traduction N°4 en mongol cyrillique)

Les trois premiers éléments constituent la référence et ne participent pas à la notation tandis que les cinq derniers éléments sont les candidats et seront notés selon leurs attributs : « Informativité », « Fidélité » et « Fluidité ». Ces trois derniers et la grille de notation constituent notre stratégie d'évaluation (*Tableau 11*).

Score	Critère		
	Informativité —— Quantité d'informations de la référence que contient la traduction	Attributs Fidélité —— Degré de similarité du sens de la traduction à celui de la référence	Fluidité —— Fluidité de la traduction
0	La traduction ne contient <b>aucune information</b> contenue dans la référence.	Le sens de la traduction <b>n'a rien à voir</b> avec la référence.	La traduction <b>n'a pas de syntaxe</b> et elle <b>n'est pas du tout fluide</b> .
1	La traduction contient <b>peu d'informations</b> dans la référence.	Le sens de la traduction est <b>quelque peu lié</b> à la référence.	La traduction <b>a la syntaxe</b> , mais elle <b>n'est pas très fluide</b> .
2	La traduction contient <b>certaines informations</b> contenues dans la référence.	Le sens de la traduction est <b>à peu près le même</b> que la référence.	La traduction <b>est conforme à la syntaxe</b> et elle est <b>en grande partie fluide</b> .
3	La traduction contient <b>la plupart des informations</b> contenues dans la référence.	Le sens de la traduction est <b>exactement le même</b> que la référence.	La traduction est <b>conforme à la syntaxe</b> et elle est <b>parfaitement fluide</b> .
4	La traduction contient <b>toutes les informations</b> contenues dans la référence.		

Tableau 11. Stratégie d'évaluation IFF

Notre stratégie d'évaluation est comme suit : le score complet est de 10. À noter que la totalité des points n'est pas la même pour ces trois attributs : 4 pour « Informativité », 3 pour « Fidélité », 3 pour « Fluidité ». Étant donné la qualité du corpus, nous avons l'intuition des performances médiocres de nos modèles entraînés, et nous considérons donc que pour nos modèles, c'est la quantité des informations contenues dans la traduction qui importe plus. D'où le privilège de l'informativité avec un point de plus.

Une fois le formulaire prêt, nous l'avons envoyé aux évaluateurs. Tous leurs retours (*Tableau 12*) ne sont pourtant pas utilisés :

- On élimine les scores donnés par Hairina car ses scores font preuve d'un optimisme extrême et d'une fausseté évidente peut-être faute d'une bonne maîtrise de l'écriture cyrillique.

- On élimine également les scores donnés par Ananda car il n'a pas pu terminer toute l'évaluation en raison de santé et il n'a noté que 18 phrases sur 50 qui sont selon lui plus claires. Donc, ses scores manquent d'exhaustivité et d'objectivité.

Cette décision a été faite selon les conseils des autres locutrices natives participantes après maintes vérifications.

Ce sont donc les scores moyens « sans Hairina ni Ananda » donnés par les quatre premières évaluatrices qui ont été employés pour calculer les scores finaux de nos modèles de traduction.

évaluateur/trice/modèle	mt0			mt1			mt2			mt3			mt4			Score
	Informativité	Fidélité	Fluidité	Informativité	Fidélité	Fluidité	Informativité	Fidélité	Fluidité	Informativité	Fidélité	Fluidité	Informativité	Fidélité	Fluidité	
Uruntuwa	3,76	2,37	2,19	0,66	0,08	0,20	0,90	0,72	0,16	0,30	1,18	0,64	0,16	0,28	1,24	
Sainjagal Anya	3,94	2,66	2,64	0,58	0,12	0,10	0,80	0,78	0,10	0,12	1,00	0,84	0,14	0,12	1,10	
Surrigage	3,88	2,54	2,54	0,64	0,10	0,10	0,82	0,80	0,12	0,16	1,06	0,90	0,12	0,16	1,14	
Sovin	3,66	2,44	2,88	0,54	0,14	0,34	1,02	0,52	0,16	0,42	1,08	0,58	0,18	0,46	1,20	
Hairina	3,78	2,82	2,82	1,76	1,16	1,24	4,16	2,20	1,50	1,60	5,30	1,98	1,48	1,72	5,18	
Ananda	3,67	3,00	3,00	0,28	0,28	0,28	0,83	1,33	1,17	1,22	3,72	1,56	1,56	1,61	4,72	
Moyen	3,78	2,64	2,68	0,74	0,31	0,38	1,42	1,06	0,53	0,64	2,22	1,08	0,61	0,73	2,43	
Moyen sans Hairina	3,78	2,60	2,65	0,54	0,14	0,20	0,87	0,83	0,34	0,44	1,61	0,90	0,43	0,53	1,88	
Moyen sans Hairina ni Ananda	3,81	2,50	2,56	0,61	0,11	0,19	0,89	0,71	0,14	0,25	1,08	0,74	0,15	0,26	1,17	

Tableau 12. Calculs des scores de l'évaluation humaine

Les scores BLEU pour l'évaluation automatique se présentent sous forme de pourcentage. Afin que les scores de l'évaluation humaine soient aussi clairs que les scores BLEU, nous les avons transformés pour chaque attribut (4-informativité, 3-fidélité, 3-fluidité) également en pourcentage. Par exemple, « informativité » de « mt0\_s » donne  $3,81/4 * 100\% = 95,25\%$ , « fidélité » de « mt0\_s » donne  $2,50/3 * 100\% = 83,43\%$ .

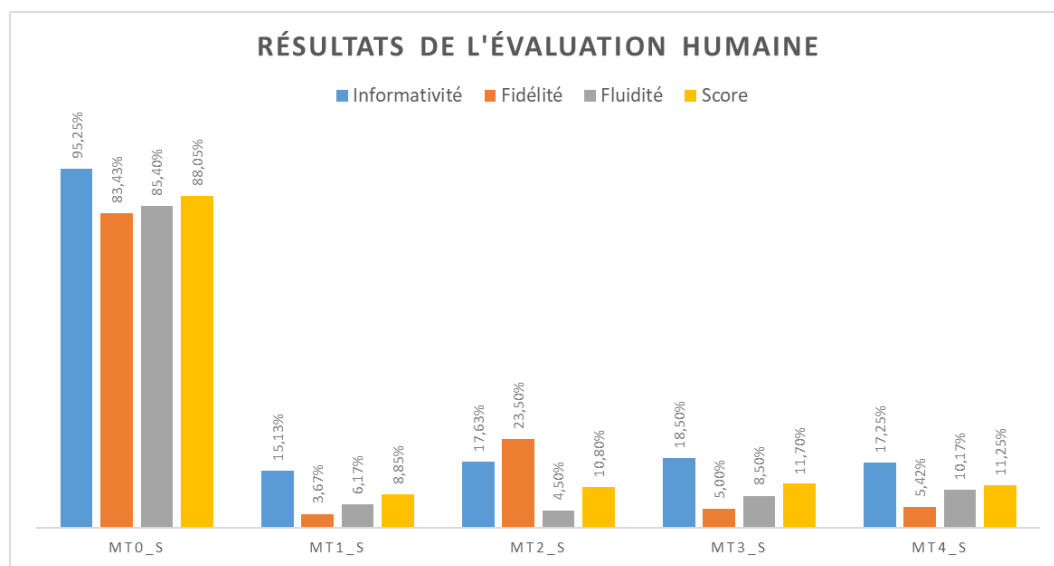


Tableau 13. Comparaison des scores IFF pour la sélection du test set

Les scores *IFF* des traductions sont comparés dans le *Tableau 13*. Nous constatons que *Google Traduction* remporte sur tous nos modèles entraînés avec une supériorité écrasante. Les jugements humains sont sans appel, nos modèles ne rivalisent pas avec celui de *Google*, ce qui n'est pas surprenant mais est encore plus visible lors de cette évaluation. Cela confirme également l'intérêt de l'évaluation humaine, qui est bien plus sévère que les scores BLEU.

Somme toute, les résultats de l'évaluation humaine sur la sélection du *test set* sont frustrants, mais ne sont pas surprenants vu que nous nous comparons à un système mis au point par une grande entreprise technologique disposant de moyens considérables.

## 4. DISCUSSION

Dans cette partie, nous expliquerons notre méthodologie et conception pour l'évaluation, dresserons un bilan des résultats obtenus et mènerons une analyse basée sur des exemples.

Afin d'obtenir une appréciation la plus exhaustive possible sur les performances de nos modèles entraînés, nous avons effectué trois évaluations, deux automatiques et une humaine. Ceci suit en fait un ordre de priorités : nous effectuons d'abord une évaluation automatique avec la métrique de notre choix, le score BLEU, qui peut donner des retours tout de suite sans investissement humain ; si les résultats sont satisfaisants, nous concevons ensuite une méthode d'évaluation humaine pour savoir comment les locuteurs natifs trouvent les traductions produites par nos modèles par rapport à celles produites par les moteurs de traduction à usage commercial, ce qui prend beaucoup de temps et requiert un grand investissement humain.

L'évaluation automatique (1) a donné les résultats, dans une certaine mesure, satisfaisants car nous constatons une petite progression prometteuse entre « m1 » et « m4 », et que les scores de « m3 » et de « m4 » ne sont pas très loin de *Google*, même supérieurs à un autre moteur de MT payant, *NiuTrans*. À noter que « m1 » sert de *toy model* (modèle réduit) pour tester le bon fonctionnement du cadre de développement *OpenNMT-py*.

Ensuite, nous sommes passés à l'évaluation humaine, dans laquelle nous ne pouvions inclure qu'une petite partie du corpus étant donné son caractère chronophage. C'est pourquoi nous avons ensuite préparé une sélection. Lors de la création manuelle de la sélection, nous avons survolé le contenu du corpus et nous nous sommes rendu compte de sa qualité médiocre : il existe beaucoup de phrases incomplètes ; certaines entrées ne sont pas normalisées et même pas alignées. Aussi avons-nous pensé à effectuer de nouveau l'évaluation automatique sur la sélection composée de 50 phrases complètes, dont la qualité est, plus ou moins, assurée après une vérification humaine.

m0	m1	m2	m3	m4	m5
12,08	4,63	7,49	8,60	8,46	8,17

Tableau 14. BLEU-4 pour la totalité du test set

Dans l'évaluation automatique (1), bien que les scores de nos certains modèles soient proches de *Google*, tous les scores (BLEU-4) se situent autour de 10. Selon l'interprétation du score BLEU proposée par *Google*<sup>40</sup>, les traductions produites par nos modèles sont toutes les « traductions presque inutiles », ce qui est très possiblement dû à la mauvaise qualité du corpus utilisé lors de l'entraînement.

<sup>40</sup> Interprétation du score BLEU de *Google*, disponible sur : <https://cloud.google.com/translate/automl/docs/evaluate#bleu>

m0	m1	m2	m3	m4	m5
21,90	5,74	13,23	13,36	13,36	19,43

Tableau 15. BLEU-4 pour la sélection du test set

Dans l'évaluation automatique (2), l'ensemble des scores a beaucoup augmenté tandis que les écarts des scores se sont élargis : les scores de deux moteurs à usage commercial sont devenus assez proches, autour de 20, qui indique que « l'idée générale apparaît clairement, mais le texte comporte de nombreuses erreurs grammaticales » ; les scores de nos modèles utiles (« mt1 » excepté) sont autour de 13 qui signifie « l'idée générale est difficilement compréhensible ».

À partir des résultats de ces deux évaluations automatiques, nous pouvons donc affirmer que l'amélioration de la qualité du *test set* (de la totalité à la sélection) fait ressortir l'écart des performances entre les moteurs de traduction à usage commercial et nos modèles. De plus, si l'on regarde les différences entre nos modèles entraînés, nous pourrions dire que l'augmentation du nombre d'itérations peut à l'origine améliorer considérablement la performance (« m1 et m2 »), mais cela n'a pas changé grand-chose après 50 000 étapes (« m3 et m4 ») ; et que l'augmentation du *train set* (rapports de division, 6 : 2 : 2 contre 7 : 1 : 2) peut améliorer légèrement la performance.

Pour ce qui est de l'évaluation humaine, nous avons comparé nos modèles avec le moteur de *Google*, mais n'avons pas eu le temps d'inclure « mt5\_s » de *NiuTrans* comme expliqué dans la note de bas de page 37. Mais étant donné les résultats de l'évaluation automatique (2) (« m0 » est proche de « m5 »), nous pourrions supposer que pour la sélection d'une bonne qualité, les résultats de l'évaluation humaine de *NiuTrans* soient presque les mêmes que ceux de *Google*.

Selon les résultats *IFF* montrés dans le *Tableau 13*, nos modèles ne peuvent absolument pas rivaliser avec *Google Traduction*, ce qui est normal et prévisible eu égard à la quantité et surtout à la qualité des données que l'on a utilisées. Mais nous pourrions tout de même constater que :

- 1) plus le nombre d'itérations est grand pour l'entraînement, plus la fluidité est bonne (« m1 » excepté en tant que modèle réduit) ;
- 2) l'augmentation du *train set* (différence entre « m2 » et « m3 ») n'a amélioré qu'un peu le score final et la fluidité, mais a grandement fait dégrader la fidélité, peut-être à cause de la réduction de la taille du jeu de validation ;
- 3) « m4 » à 100 000 étapes n'a pas obtenu un score très différent par rapport à « m3 » à 50 000 étapes avec le même rapport de division, même a un peu dégradé en termes d'informativité et de score final ;
- 4) le nombre d'itérations n'influe pas tellement l'informativité en regardant tous les 4 modèles.

Il est par ailleurs à noter que dans le formulaire d'évaluation humaine, il arrive que nos modèles (surtout « m3 » et « m4 ») omettent de traduire certaines entrées, ce qui a eu des effets négatifs sur les scores *IFF*.

Comme évoqué à la fin de la partie [2.1.4](#), « l'évaluation humaine est toujours considérée comme déterminante dans la communauté scientifique ». Malgré les résultats

passables dans deux évaluations automatiques, nous reconnaissons, d'après l'évaluation humaine, que nos quatre modèles entraînés laissent beaucoup à désirer.

Toujours est-il que nous voudrions nous pencher sur quelques exemples concrets pour avoir une idée sur les faiblesses et, éventuellement, les forces de notre meilleur modèle par rapport à *Google Traduction*. Compte tenu des résultats des trois évaluations, nous estimons que « m3 » est le meilleur modèle qui a été entraîné à 50 000 étapes avec l'augmentation du *train set* (rapport de division : 7 : 1 : 2).

Nous avons recours à l'Évaluateur de score BLEU interactif de *Tilde* afin d'obtenir une visualisation des scores BLEU pour « m3 » et *Google Traduction* (Figure 25). Le premier histogramme en haut montre les scores BLEU pour les deux traductions de chaque phrase (« m3 » en bleu ; *Google* en vert). Le deuxième histogramme au milieu présente les différences de chaque paire des scores BLEU pour ces deux traductions. Si une colonne bleue s'affiche, cela veut dire que « m3 » est meilleur que *Google* ; et vice versa, si une colonne verte s'affiche, « m3 » est pire que *Google*. Le dernier histogramme permet d'ajuster le champ visuel pour afficher plus ou moins d'entrées sur le nombre d'entrée total (ici la sélection contient 50 entrées).

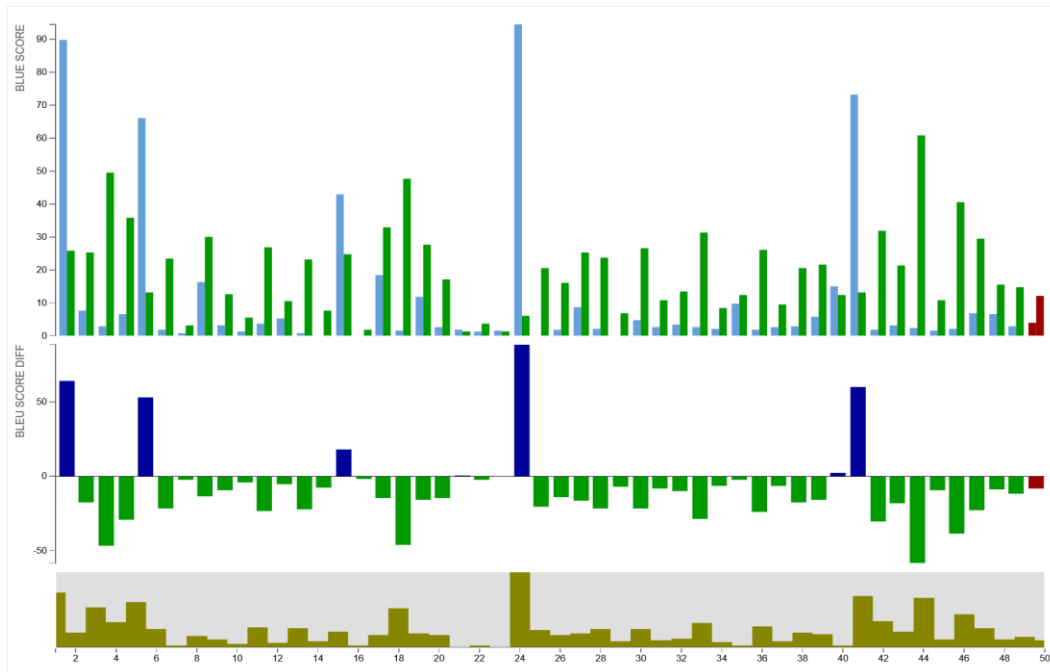


Figure 25. Visualisation des scores BLEU de mt3\_s (m3) et mt0\_s (Google)

Nous remarquons que parmi 50 entrées, nous avons cinq entrées dont les traductions faites par « m3 » sont meilleures que celles de *Google*. En combinant également les retours de l'évaluation humaine, nous décidons d'analyser les trois entrées suivantes de la sélection : N°1 (Tableau 16), N°3 (Tableau 18) et N°15 (Tableau 20).

Ces choix ont été faits avec un principe de couvrir tous les trois cas pertinents :

- Entrée N°1 : « m3 » est bien meilleur que *Google* ;
- Entrée N°3 : *Google* est bien meilleur que « m3 » ;
- Entrée N°15 : la différence entre « m3 » et *Google* n'est pas très grande.

Les analyses ci-dessous se basent sur les scores BLEU, les scores *IFF* moyens convertis en pourcentage, ainsi que les commentaires de M. Burged, locuteur natif de la langue mongole.

Sentence 1	BLEU	Length ratio	Text
Source	-	-	Gestion et contrôle des stocks est une nouvelle génération de logiciel d'inventaire pour Windows. Il couvre toutes les opérations de prise de commandes, facturation, achats, gestion des stocks et de travail à . L'interface utilisateur est entièrement optimisé soins pour l'entrée à grande vitesse de l'ordre d'un client et la prévention des erreurs les plus courantes. Il est conçu pour une utilisation sur plusieurs ordinateurs, et contient des niveaux d'autorisation fiables et sécurisées.
Human	100.00	1.00	Абаце Бараа материалын удирдлага , хяналт нь Windows нь бараа материалын программ хангамж нь шинэ үе юм . Энэ нь хөдөлмөрийн менежмент захиалга , нэхэмжлэл , худалдан авах , бараа материал авах бүх үйл ажиллагааг хамарна . хэрэглэгчийн интерфэйс нь анхааралтай өндөр хурдны үйлчлүүлэгчийн тулд оролт , нийтлэг алдаанаас урьдчилан сэргийлэх оновчтой байна . Энэ нь хэд хэдэн компьютер дээр ашиглаж зориулалттай , найдвартай , аюулгүй зөвшөөрөл хэмжээгээр агуулдаг юм .
Machine	89.83	0.97	Бараа материалын удирдлага , хяналт нь Windows нь бараа материалын программ хангамж нь шинэ үе юм . Энэ нь хөдөлмөрийн менежмент захиалга , нэхэмжлэл , худалдан авах , бараа материал авах бүх үйл ажиллагааг хамарна . хэрэглэгчийн интерфэйс нь анхааралтай өндөр хурдны үйлчлүүлэгчийн тулд оролт , нийтлэг <unk> урьдчилан сэргийлэх оновчтой байна . Энэ нь хэд хэдэн компьютер дээр ашиглаж <unk> найдвартай , аюулгүй зөвшөөрөл хэмжээгээр агуулдаг юм .
Machine	25.92	1.00	Бараа материалын менежмент ба хяналт бол Windows - ийн шинэ үеийн бараа материалын програм хангамж юм . Энэ нь захиалга авах , нэхэмжлэх хийх , худалдан авах , бараа материалын менежмент , ажил хийх хүртэлх бүх үйл ажиллагааг хамарна . Хэрэглэгчийн захиалгыг хурдан оруулах , нийтлэг алдаанаас урьдчилан сэргийлэх зорилгоор хэрэглэгчийн интерфэйсийг бүрэн оновчтой болгосон . Энэ нь олон компьютер дээр ашиглаж зориулалттай бөгөөд найдвартай , найдвартай зөвшөөрлийн түвшинг агуулдаг .

Tableau 16. Comparaison des scores BLEU de l'entrée N°1

Pour l'entrée N°1, selon les scores BLEU, « m3 » est bien meilleur que *Google*, 89,83 contre 25,92. Mais nous constatons que dans la traduction de « m3 », il existe des <unk> qui signifie « mot inconnu ». Cela veut dire que dans l'entraînement, « m3 » n'a pas lu le mot qui devrait s'y trouver.

Modèle/Moteur	Informativité	Fidélité	Fluidité	Score
m3	68,75 %	58,33 %	50,00 %	60,00 %
Google	87,50 %	91,67 %	66,67 %	82,50 %

Tableau 17. Scores IFF moyens de l'entrée N°1

Cependant, selon les scores *IFF*, *Google* remporte sur « m3 ».

D'après Burged, dans l'entrée N°1, la logique de la traduction de « m3 » est correcte. Les phrases simples avec une grammaire de base sont bien traduites, mais les phrases longues sont peu naturelles. La traduction de *Google* semble relativement fluide car *Google* aurait probablement un plus grand vocabulaire.

Nous remarquons ici donc l'opposition entre les scores automatique et humain.

Sentence 3	BLEU	Length ratio	Text
Source	-	-	Portable Document Format est un format de fichier développé dans les années 90 pour présenter les documents, y compris le formatage de texte et les images, de manière indépendante des logiciels d'application, du matériel et des systèmes d'exploitation.
Human	100.00	1.00	Зөврийн баримт бичгийн формат нь 1990 - ээд онд боловсруулсан файлын хэлбэр бөгөөд текст формат болон дүрсийг багтаасан баримт бичиг , програм хангамж , техник хангамж , үйлдлийн системээс
Machine	3.01	1.41	Хэвлэл мэдээллийн формат нь бие даасан болон бусад <unk> агуулгын системийн нийцсэн болон гаргаж авах : <unk> баримт бичиг нь илгээсэн тоног төхөөрөмж , машин гэх мэт хүчтэй алдааны хэлбэр нь үүнээс <unk> ба гаргаж авсан бөгөөд үүнийг идэвхжүүлсэн байна .
Machine	49.61	1.07	Зөврийн баримт бичгийн формат нь хэрэглээний програм хангамж , техник хангамж , үйлдлийн системээс үл хамааран текст форматлах , зураг оруулах зэрэг 1990 - ээд онд боловсруулсан файлын формат юм .

Tableau 18. Comparaison des scores BLEU de l'entrée N°3

Pour l'entrée N°3, selon les scores BLEU, *Google* est bien meilleur que « m3 » où il y a plus de mots inconnus (<unk>).



Modèle/Moteur	Informativité	Fidélité	Fluidité	Score
m3	12,50 %	8,33 %	8,33 %	7,50 %
Google	87,50 %	75,00 %	75,00 %	80,00 %

Tableau 19. Scores IFF moyens de l'entrée N°3

Les scores *IFF* donnent le même résultat que les scores BLEU (« m3 » < Google).

D'après Burged, il semble que « m3 » n'a pas appris la bonne syntaxe car sa traduction est un peu en désordre. Son vocabulaire ne semble pas non plus pertinent. La traduction de *Google* ressemble beaucoup à la traduction de référence.

Ici, nous constatons la cohérence des scores automatique et humain.

Sentence 15	BLEU	Length ratio	Text
Source	-	-	Télécharger la version complète de fonctionnalités essai de 30 jours ( 6.15 MB)
Human	100.00	1.00	( Б ? рэн боломжуудтай 30 хоногийн туршилтын хувилбарыг татаж авах 6.15 MB )
Machine	42.83	0.71	<unk> боломжуудтай 30 хоногийн туршилтын хувилбарыг татаж авах <unk> <unk>
Machine	24.86	0.79	30 хоногийн туршилтын бүрэн хувилбарыг татаж авах ( 6.15 MB )

Tableau 20. Comparaison des scores BLEU de l'entrée N°15

Pour l'entrée N°15, selon les scores BLEU, la traduction de « m3 » est meilleure que celle de *Google*. Il semble que « m3 » a appris toute la phrase sauf le début et la fin.

Modèle/Moteur	Informativité	Fidélité	Fluidité	Score
m3	25,00 %	16,67 %	25,00 %	27,50 %
Google	100,00 %	91,67 %	75,00 %	90,00 %

Tableau 21. Scores IFF moyens de l'entrée N°15

Néanmoins, selon les scores *IFF*, « m3 » n'égale toujours pas *Google*. C'est encore une fois l'opposition entre les scores automatique et humain.

Burged pense que la traduction de référence est même pire que celle de *Google*. En outre, nous remarquons que la phrase source en français est grammaticalement incorrecte. Nous en inférons que la traduction de référence est très probablement un produit de traduction automatique de cette fausse phrase source en français, ce qui en dit long sur la mauvaise qualité du corpus utilisé dans notre expérimentation.

Toute réflexion faite, un corpus de qualité est crucial pour construire un bon système de NMT.

## 5. CONCLUSION

Dans ce mémoire, nous avons décrit l’historique de la traduction automatique et les technologies du domaine. Nous avons présenté la langue mongole et défini un état de l’art concernant la traduction automatique pour la paire de langues peu dotée, français-mongol. Faute de recherches connexes, nous avons entamé une nouvelle étape dans le domaine de MT en développant notre propre système (quatre modèles) de traduction automatique neuronale français-mongol. Le développement s’avère être coûteux en temps et en moyens. Pour ce faire, nous avons trouvé des données sur la paire français-mongol qu’il a fallu préparer. Après l’entraînement, nous avons ensuite effectué deux évaluations avec la métrique automatique populaire, BLEU. Mais nous ne nous en sommes pas contentés et avons en outre proposé « *IFF* », métrique originale pour l’évaluation humaine en vue de mieux comprendre les forces et les faiblesses de notre système, ce qui a requis un grand investissement du temps et des ressources humaines. Finalement, nous avons mené une analyse comparative basée sur des exemples à partir des résultats obtenus des évaluations et des commentaires d’un locuteur natif.

Au cours de la réalisation du mémoire, nous avons pu répondre aux questions posées dans l’introduction :

**1) Quelles sont les raisons du manque de recherches en traduction automatique pour la paire français-mongol ?**

Les locuteurs de la langue mongole ne sont pas très nombreux. Il n’y aurait *a priori* pas beaucoup de chercheurs en traduction automatique pour cette langue. Par ailleurs, il existe peu de ressources open-source pour le mongol, sans lesquelles il est difficile de mener des recherches en traduction automatique.

**2) Y a-t-il d’autres recherches sur la traduction automatique de la langue mongole ? Avec quelles langues comme paires ?**

Oui. Il existe de plus en plus de recherches sur la traduction automatique de la langue mongole. Le mongol cyrillique va souvent avec l’anglais ; et le mongol classique, quasiment toujours avec le chinois.

**3) Quelles ressources existent pour mettre en place un système de traduction français-mongol ?**

Il existe *de facto* pas mal de ressources pour la langue mongole (classique) en Chine, dont la majorité sont les propriétés de certaines organisations. Cependant, ces ressources sont généralement pour la paire de langues chinois-mongol (classique). Pour la paire français-mongol, pour autant que nous avons pu trouver après une longue recherche, seul le corpus parallèle français-mongol de *MultiCCAligned v1.1* utilisé dans notre expérimentation permet de mettre en place un système de traduction automatique. Mais sa qualité n’est pas du tout satisfaisante.

**4) Comment construire un système de traduction automatique neuronale ?  
Est-ce difficile ?**

Voici les étapes que nous avons suivies pour construire notre système de traduction automatique : choisir d'abord un cadre de développement (*framework*) ; recueillir les ressources linguistiques, à savoir un corpus parallèle avec une qualité aussi bonne que possible ; adapter le corpus à des exigences du cadre de développement pour qu'il soit lisible ; configurer l'apprentissage ; démarrer l'entraînement ; sauvegarder les modèles entraînés qui constituent le système de NMT.

La construction est plutôt coûteuse que difficile, en temps et en moyens. Les étapes nécessaires peuvent être résumées en quelques lignes comme ci-dessus, mais pour chaque étape, il y a de nombreux paramètres à gérer. Par exemple, le corpus de qualité n'est pas facile à trouver pour certaines paires de langues ; l'entraînement est chronophage et nécessite souvent un matériel puissant ; le paramétrage est délicat et demande de compétences techniques pointues, etc.

**5) Quelles caractéristiques de la langue mongole mènent à des difficultés supplémentaires ?**

Le mongol cyrillique ne mène pas en soi à des difficultés supplémentaires tandis que le mongol classique, oui. Selon certains travaux précédents cités pour la MT chinois-mongol, l'encodage, la direction de l'écriture et la police pose plus de problèmes que d'autres langues.

**6) Comment évaluer un système de traduction ?**

Pour évaluer un système de traduction automatique, il faut utiliser les métriques automatiques pour connaître rapidement la performance d'un système ; une fois que les résultats sont satisfaisants pour l'évaluation automatique, il faut procéder à une évaluation humaine pour avoir une appréciation finale sur la performance du système.

**7) Comment pourrait-on mieux faire pour améliorer la performance de notre système ?**

Nous laissons cette dernière question pour la partie [Perspectives](#) où nous allons regarder en arrière notre travail et en ouvrir des pistes de réflexion.

## 6. PERSPECTIVES

Dans cette dernière partie, nous présenterons les limites de notre travail et envisagerons des plans pour les travaux futurs.

Notre travail aurait plusieurs limites :

Dans un premier temps, pour ce qui est des ressources, la qualité de notre corpus utilisé pour l'entraînement est médiocre. Cela a eu un effet négatif critique sur la performance de notre système.

Dans un deuxième temps, en ce qui concerne l'entraînement, nous n'avons pas assez étudié le paramétrage et n'avons utilisé que la configuration de base pour tous les entraînements sans chercher à tester plusieurs configurations. Nous ne savons donc pas si la configuration de base est problématique.

Dans un troisième temps, pour l'évaluation humaine, nous n'avons pas pu recruter suffisamment d'évaluateurs. De plus, nous n'avons pas pu former les évaluateurs. Ils n'avaient pas tous des connaissances nécessaires pour comprendre parfaitement le contenu des textes dans le formulaire. Certains de leurs retours font preuve d'un grand désaccord inter-annotateur, c'est pourquoi nous avons finalement adopté uniquement quatre retours sur six.

Pour la suite de notre travail, il faudrait essayer différentes configurations, même plusieurs cadres de développement pour trouver le meilleur paramétrage possible. Ensuite, pour l'évaluation humaine, nous pourrions recruter plus tôt les évaluateurs pour qu'ils viennent plus nombreux, qu'ils puissent être formés sur les techniques d'évaluation et qu'ils se familiarisent avec le contenu du corps. Finalement, c'est particulièrement dans les données qu'il faut mettre l'effort pour aller plus loin dans le domaine de NMT, puisqu'elle est orientée données. Nous envisageons donc, pour nos travaux futurs, de développer un corpus parallèle français-mongol (cyrillique et classique) d'une meilleure qualité.

Inévitablement imparfait, notre travail entend revitaliser la langue mongole, protéger la diversité de notre ADN culturel en établissant des liens avec d'autres langues principales à l'appui de la NMT, et revaloriser la paire de langues peu dotée, français-mongol. Ce mémoire, au lieu d'une fin d'études, marque en réalité le début de notre carrière de recherche pour la construction d'un corpus français-mongol open-source. La construction, n'échappant pas à son amplitude intrinsèque, serait un défi herculéen. Au contraire, sa difficulté justifie une exploration continue.

# BIBLIOGRAPHIE

Abend, O., & Rappoport, A. (2013). Universal conceptual cognitive annotation (UCCA). Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers),

ALPAC. (1966). *Language and Machines*.

Ao, M., Xiong, Z., & Hu, H. (2011). 基于蒙科立输入法的蒙古语同形异码词研究. 第十一届全国人机语音通讯学术会议 (NCMMSC2011) 论文集,

Artetxe, M., Labaka, G., & Agirre, E. (2019). An effective approach to unsupervised machine translation. *arXiv preprint arXiv:1902.01313*.

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Banerjee, S., & Lavie, A. (2005). METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization,

Bao, M. (2016). Construction of Mongolian Language Media Texts Corpus. *内蒙古师范大学学报: 哲学社会科学版*, 45(4), 70-74.

Bojar, O., Federmann, C., Haddow, B., Koehn, P., Post, M., & Specia, L. (2016). Ten years of WMT evaluation campaigns: Lessons learnt. Proceedings of the LREC 2016 Workshop “Translation Evaluation—From Fragmented Tools and Data Sets to an Integrated Ecosystem,

Brown, P. F., Cocke, J., Della Pietra, S. A., Della Pietra, V. J., Jelinek, F., Lafferty, J. D., Mercer, R. L., & Roossin, P. S. (1990). A Statistical Approach to Machine Translation. *Computational Linguistics*, 16(2), 79-85.

Brown, P. F., Pietra, V. J. D., Pietra, S. A. D., & Mercer, R. L. (1993). The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics*, 19(2), 263-311.

Brownlee, J. (2017, le 19 décembre 2019). *A Gentle Introduction to Calculating the BLEU Score for Text in Python*. <https://machinelearningmastery.com/calculate-bleu-score-for-text-python/>

Callison-Burch, C., Osborne, M., & Koehn, P. (2006). Re-evaluating the role of BLEU in machine translation research. 11th Conference of the European Chapter of the Association for Computational Linguistics,

Cao, Y. (2020). *Research on Chinese-Mongolian Neural Machine Translation Based on Monolingual Corpora* University of Science and Technology of China].

Chatzikoumi, E. (2020). How to evaluate machine translation: A review of automated and human metrics. *Natural Language Engineering*, 26(2), 137-161.

Chiang, D. (2007). Hierarchical phrase-based translation. *Computational Linguistics*, 33(2), 201-228.

Coughlin, D. (2003). Correlating automated and human assessments of machine translation quality. Proceedings of MT summit IX,

CWMT. (2017). CWMT 2017 MACHINE TRANSLATION EVALUATION PARTICIPATING SITE AGREEMENT. In.

Debret, J. (2020). *Les normes APA françaises : Guide officiel de Scribbr basé sur la septième édition (2019) des normes APA*.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Doddington, G. (2002). Automatic evaluation of machine translation quality using n-gram co-occurrence statistics. Proceedings of the second international conference on Human Language Technology Research,

El-Kishky, A., Chaudhary, V., Guzmán, F., & Koehn, P. (2020). CCAIined: A Massive Collection of Cross-Lingual Web-Document Pairs. *Empirical Methods in Natural Language Processing*,

FAN, W., HOU, H., WANG, H., WU, J., & LI, J. (2018). Mongolian-Chinese Neural Machine Translation with Priori Information. *Journal of Chinese Information Processing*, 06.

Federmann, C. (2010). Appraise: An Open-Source Toolkit for Manual Phrase-Based Evaluation of Translations. LREC,

Fei, D., Yuan, L., & Quan, C. (2019). Analysis of the construction of minority corpus oriented to information processing. *无线互联科技*, 19.

Freitag, M., Grangier, D., & Caswell, I. (2020). BLEU might be guilty but references are not innocent. *arXiv preprint arXiv:2004.06063*.

Gaunt, J. (2004). *Modern Mongolian: A course-book*. Routledge.

Giménez, J., & Gonzalez, M. (2013). An Open Toolkit for Automatic Machine Translation (Meta-) Evaluation.

Graham, Y., Baldwin, T., Moffat, A., & Zobel, J. (2013). Continuous measurement scales in human evaluation of machine translation. Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse,

Gu, J., Bradbury, J., Xiong, C., Li, V. O., & Socher, R. (2017). Non-autoregressive neural machine translation. *arXiv preprint arXiv:1711.02281*.

He, Z. (2015). Baidu translate: Research and products. Proceedings of the Fourth Workshop on Hybrid Approaches to Translation (HyTra),

Hou, H., & Liu, Q. (2007). 基于实例的汉蒙机器翻译. *中文信息学报*, 21(4), 65-72.

Hua, S. (1997). 实现 500 万词级《现代蒙古语文数据库》的主要措施.

Hutchins, J. (1996). The state of machine translation in Europe. AMTA,

Hutchins, W. J. (1995). Machine Translation: A Brief History.

Ibrahimova, M. (2021). *Décryptage : Quelles langues parle le cyberspace ?* <https://en.unesco.org/courier/2021-2/languages-cyberspace>

Indiana University, I. *Mongolian*. <https://celcar.indiana.edu/materials/language-portal/mongolian/index.html>

Jaimai, P., & Chimeddorj, O. (2008). Corpus building for Mongolian language. Proceedings of the 6th Workshop on Asian Language Resources,

Janhunnen, J. (2005). *The Mongolic Languages*. Taylor & Francis e-Library.

Kalchbrenner, N., & Blunsom, P. (2013). Recurrent continuous translation models. Proceedings of the 2013 conference on empirical methods in natural language processing,

Kim, S. (2020). *The History of the Mongolic Languages*. <https://www.youtube.com/watch?v=hkrHAzFathQ>

Klein, G., Kim, Y., Deng, Y., Nguyen, V., Senellart, J., & Rush, A. M. (2020). OpenNMT: Neural Machine Translation Toolkit: 2020 Edition.

Knowlson, J. (1975). Universal Language Schemes in England and France 1600-1800. *University of Toronto Press*.

Koehn, P. (2007). Euromatrix-machine translation for all european languages. Proceedings of Machine Translation Summit XI: Invited papers,

Koehn, P. (2010). Statistical Machine Translation. In: Cambridge University Press.

Koehn, P., & Hoang, H. (2007). Factored translation models. Proceedings of the 2007 joint conference on empirical methods in natural language processing and computational natural language learning (EMNLP-CoNLL),

Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., & Zens, R. (2007). Moses: Open source toolkit for statistical machine translation. Proceedings of the 45th annual meeting of the association for computational linguistics companion volume proceedings of the demo and poster sessions,

Koehn, P., Och, F. J., & Marcu, D. (2003). *Statistical phrase-based translation*.

Lacruz, I., Denkowski, M., & Lavie, A. (2014). Cognitive demand and cognitive effort in post-editing.

Langlais, P., Leplus, T., Gandrabur, S., & Lapalme, G. (2000). From the Real World to Real Words: The METEO case.

Lavie, A. (2011). Evaluating the Output of Machine Translation Systems.

Leusch, G., Ueffing, N., & Ney, H. (2006). CDER: Efficient MT evaluation using block movements. 11th Conference of the European Chapter of the Association for Computational Linguistics,

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. Soviet physics doklady,

Lin, C.-Y., & Och, F. J. (2004). Automatic evaluation of machine translation quality using longest common subsequence and skip-bigram statistics. Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-04),

Liu, W. (2018). *Research on Mongolian-Chinese Machine Translation Based on LSTM Neural Network* Inner Mongolia University of Technology].

Liu, Y., Liu, Q., & Lin, S. (2006). Tree-to-string alignment template for statistical machine translation. Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics,

Melamed, I. D., Green, R., & Turian, J. (2003). Precision and recall of machine translation. Companion Volume of the Proceedings of HLT-NAACL 2003-Short Papers,

Nakhlé, M. (2021). *Évaluation globale d'un système de traduction automatique de documents structurés dans le domaine financier* Université de Paris Nanterre].

Nießen, S., Och, F. J., Leusch, G., & Ney, H. (2000). An Evaluation Tool for Machine Translation: Fast Evaluation for MT Research. LREC,

Ochir, & Serguleng, W. (2003). The Design of an English-Mongolian Machine Translation System. *内蒙古大学学报: 自然科学版*, 34(5), 582-587.



Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002). Bleu: a method for automatic evaluation of machine translation. Proceedings of the 40th annual meeting of the Association for Computational Linguistics,

Poibeau, T. (2017). *Machine Translation*.

Shannon, C. E., & Warren, W. (1949). *The mathematical theory of communication*. The mathematical theory of communication.

Shen, Z. (2017). *基于注意力神经网络的蒙汉机器翻译系统的研究* [内蒙古大学].

Simon, A. (2021). *Omniglot - writing systems and languages of the world*. [www.omniglot.com](http://www.omniglot.com)

Slocum, J. (1984). *Machine Translation: its history, current status, and future prospects*. MIT Press,

Snover, M., Dorr, B., Schwartz, R., Micciulla, L., & Makhoul, J. (2006). A study of translation edit rate with targeted human annotation. Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers,

Specia, L., Raj, D., & Turchi, M. (2010). Machine translation evaluation versus quality estimation. *Machine translation*, 24(1), 39-50.

Specia, L., Shah, K., De Souza, J. G., & Cohn, T. (2013). QuEst-A translation quality estimation framework. Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations,

Specia, L., Turchi, M., Cancedda, N., Cristianini, N., & Dymetman, M. (2009). Estimating the Sentence-Level Quality of Machine Translation Systems. EAMT,

Su, C. (2014). *基于层次短语模型的蒙—汉统计机器翻译研究* [内蒙古大学].

Sun, Y., Wang, S., Li, Y., Feng, S., Tian, H., Wu, H., & Wang, H. (2020). Ernie 2.0: A continual pre-training framework for language understanding. Proceedings of the AAAI Conference on Artificial Intelligence,

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. *Advances in neural information processing systems*,

Tiedemann, J. (2012, may). Parallel Data, Tools and Interfaces in OPUS. *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)* Istanbul, Turkey.

Tomás, J., Mas, J. À., & Casacuberta, F. (2003). A quantitative method for machine translation evaluation. Proceedings of the EACL 2003 Workshop on Evaluation Initiatives in Natural Language Processing: are evaluation methods, metrics and resources reusable?,

Tserenpil, D., & Kullmann, R. (2008). *Mongolian grammar*. ADMON Co.,Ltd.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, L., & Polosukhin, I. (2017). *Attention Is All You Need* International Conference on Neural Information Processing,

Vercher, E. (2016). *La célèbre pierre de Rosette. Comment a-t-elle été traduite ?* <https://www.agestrad.com/blog-traduccion/la-celebre-pierre-rosette>

Wang, H. (2018). *多粒度蒙古文汉文神经网络机器翻译研究* [内蒙古大学].

Wang, H., Wu, H., He, Z., Huang, L., & Church, K. W. (2021). Progress in Machine Translation. *Engineering*.

Wang, Y., Su, Y., Zhao, Y., Sun, X., & Ren, Q. (2020). Mongolian-Chinese Neural Machine Translation Model Based on Parameter Transfer. *Computer Applications and Software*, 37(9), 81-87. <https://doi.org/10.3969/j.issn.1000-386x.2020.09.014>

Weaver, W. (1949[1955]). Translation. *Cambridge: Technology Press*, 14, 15-23.

Wu, J. (2017). *多方法融合蒙汉机器翻译与译文重排序研究*. [内蒙古大学, 1-z.

Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., & Macherey, K. (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*.

Xiao, T., & Zhu, J. (2021). *Machine Translation: Foundations and Models*. <https://opensource.niutrans.com/mtbook/homepage.html>

Yang, M., Hu, X., Xiong, H., Wang, J., Jiaermuhamaiti, Y., He, Z., Luo, W., & Huang, S. (2019). Ccmt 2019 machine translation evaluation report. China Conference on Machine Translation,

Zhang, G. (2009). *The Experimental Study and Realization of Mongolian-Chinese Alignment corpora* [Inner Mongolia Normal University].

Zhang, J., & Zong, C. (2020). Neural machine translation: Challenges, progress and future. *Science China Technological Sciences*, 1-23.

Zhou, M., Wang, B., Liu, S., Li, M., Zhang, D., & Zhao, T. (2008). Diagnostic evaluation of machine translation systems using automatically constructed linguistic

check-points. Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008),

Крылов, С. А. (2017). Монгольские аналитические конструкции в количественном аспекте. *Oriental Studies*(5 (33)).

## ANNEXE

### Paramètres clés dans « fr\_mn.yamll »

Catégorie	Nom	Valeur	Signification
<b>Batching</b>	valid_batch_size	16	Taille maximale du lot (batch) pour la validation
<b>Optimization</b>	model_dtype	fp32	Type de données du modèle
	optim	adam	Méthode d'optimisation
	learning_rate	2	Taux d'apprentissage de départ
	decay_method	noam	Taux de décroissance
<b>Model</b>	encoder_type	transformer	Type de couche d'encodeur à utiliser
	decoder_type	transformer	Type de couche de décodeur à utiliser
	enc_layer	6	Nombre de couches dans l'encodeur
	dec_layer	6	Nombre de couches dans le décodeur
	rnn_size	512	Taille des états cachés de RNN
	word_vec_size	512	Taille de <i>word-embeddings</i> pour src et tgt
	transformer_ff	2048	Taille du <i>feed-forward</i> caché de <i>Transformer</i>
	dropout_steps	0	Étapes auxquelles le <i>dropout</i> change
	dropout	0.1	Probabilité de <i>dropout</i>
	share_decoder_embeddings	true	Utilisez une matrice de pondération partagée pour les <i>word-embeddings</i> d'entrée et de sortie dans le décodeur
	share_embeddings	true	Partager <i>word-embeddings</i> entre l'encodeur et le décodeur. Cela nécessite d'utiliser un dictionnaire partagé pour cette option.

## Participants de l'évaluation humaine

Évaluateurs/trices	Titre
Urantuya	Interprète au Service d'Aide à l'Intégration en Suisse
Sainjargal Anya	Ancienne étudiante à la Sorbonne Université
Surigage	Étudiante à l'Université libre de Bruxelles
Sovin	Animatrice à la Station de radiodiffusion et de télévision de la Ligue Alxa
Hairina	Étudiante à l'Université nationale de Mongolie
Ananda	Comptable

## Script pour le prétraitement : tmx2txt.py

```

20 # Importer les bibliothèques nécessaires
21 from bs4 import BeautifulSoup
22 import sys
23
24 def tmx2txt(tmx, source, target):
25     # ouvrir le fichier .tmx
26     with open(tmx) as input:
27         # lire le fichier .tmx
28         file = input.read()
29         print("-----Reading file-----")
30
31         # parser le fichier .tmx
32         soup = BeautifulSoup(file, "html.parser")
33         print("-----Analyzing the file structure-----")
34
35         # trouver les éléments contenant les phrases et les transformer en texte
36         list_fr = soup.find_all("tuv", attrs={"xml:lang":"fr"})
37         list_mn = soup.find_all("tuv", attrs={"xml:lang":"mn"})
38
39         # créer le fichier de sortie
40         with open(source, "w", encoding="utf-8") as output_src:
41             print("-----Generating fr.txt-----")
42             for item_fr in list_fr:
43                 text_fr = item_fr.get_text()+"\n"
44                 output_src.write(text_fr)
45
46
47         with open(target, "w", encoding="utf-8") as output_tgt:
48             print("-----Generating mn.txt-----")
49             for item_mn in list_mn:
50                 text_mn = item_mn.get_text()+"\n"
51                 output_tgt.write(text_mn)
52         print("-----Finished-----")

```

## Commandes utilisées

### Serveur Helium

S'y connecter

```
ssh user@helium.lab.parisdescartes.fr -p xxxx
```

(Saisir ensuite le mot de passe de l'utilisateur)

### WordCloud

Installer WordCloud

```
pip install wordcloud
```

Générer un wordcloud

```
wordcloud_cli --text corpus/mn_288624.txt --stopwords stop-words_mn.txt --imagefile wordcloud_mn.png --scale 10
```

### OpenNMT

Installer OpenNMT-py

```
pip install --upgrade pip  
pip install OpenNMT-py
```

Créer un environnement virtuel

```
python -m venv venv_nmt
```

Activer l'environnement virtuel

```
source venv_nmt/bin/activate
```

Construire le vocabulaire du corpus

```
onmt_build_vocab -config fr_mn.yaml -n_sample -1
```

Lancer l'entraînement

```
onmt_train -config fr_mn.yaml -share_vocab
```

Traduction

```
onmt_translate -model data/models/model_frmn_step_x.pt -src data/corpus/src-test.txt -output data/corpus/tgt-pred_x.txt -cpu 0 -verbose
```