



Traitement Automatique des Langues

Rapport du projet

« Evaluation du moteur de traduction neuronale
OpenNMT sur un corpus parallèle anglais-français »

Zhongjie LI

Zhiya MA

Perrine QUENNEHEN

Traduction Automatique et Assistée

dispensé par Monsieur SEMMAR Nasredine

Année universitaire 2023-2024

Master 1 / Semestre 2

Sommaire

1. Introduction.....	2
2. Présentation du moteur de traduction neuronale OpenNMT.....	2
Description du fonctionnement d'OpenNMT.....	2
3. Evaluation du moteur de traduction neuronale OpenNMT sur un corpus en formes fléchies.....	3
Décrire le corpus d'apprentissage et d'évaluation (nombre de phrases).....	3
Décrire les métriques d'évaluation : Score BLEU, METEOR, BERT.....	4
Le score BLEU.....	4
Le Score METEOR.....	4
Le Score BERT.....	5
Tableau des résultats.....	6
4. Evaluation du moteur de traduction neuronale OpenNMT sur un corpus en lemmes.....	8
Description des lemmatisateurs.....	8
Tableau des résultats.....	8
5. Points forts, limitations d'OpenNMT et difficultés rencontrées.....	10
Points forts du moteur de traduction.....	10
Limitation du moteur de traduction.....	11
Difficultés rencontrées.....	11
6. Organisation.....	11
Annexes.....	13
Expérimentations en plus.....	13
→ Utilisation du lemmatisateur de Spacy :	13
→ Utilisation des encodeurs/décodeurs transformers d'OpenNMT.....	14
Captures d'écran des résultats.....	18
Bibliographie.....	25

1. Introduction

Dans ce monde diversifié, la multitude de langues rend la traduction essentielle pour la communication humaine, la transmission des cultures et l'accès au savoir. La traduction automatique neuronale a révolutionné le domaine de la traduction de texte en offrant des résultats plus fluides et plus naturels que les méthodes précédentes basées sur des règles ou des statistiques. OpenNMT est un moteur de traduction neuronale open source largement utilisé pour la recherche et les applications pratiques. Notre projet consiste donc à évaluer la performance de ce moteur sur des corpus parallèles anglais et français, permettant ainsi de mesurer sa capacité à gérer des variations linguistiques et à fournir des traductions de qualité.

Comment le moteur de traduction neuronale OpenNMT se comporte-t-il lorsqu'il est appliqué à un corpus parallèle anglais-français ? Quelles sont les forces et les faiblesses de ce système en termes de précision, de fluidité et de gestion des nuances linguistiques ? Pour y répondre, nous examinerons successivement la présentation d'OpenNMT, l'approche employée, les points forts de la traduction neuronale, et enfin, l'évaluation de sa performance sur les corpus donnés.

2. Présentation du moteur de traduction neuronale OpenNMT

Description du fonctionnement d'OpenNMT

OpenNMT est un outil open-source dédié au développement de systèmes de traduction automatique neuronale (NMT), caractérisé par sa simplicité d'utilisation, sa flexibilité et son efficacité, tout en garantissant une précision remarquable dans les traductions. Son architecture repose sur un modèle d'encodeur-décodeur doté d'un mécanisme d'attention, conçu pour estimer la probabilité d'une phrase cible étant donné une phrase source. Le processus de traduction s'articule autour de plusieurs étapes clés.

Pour commencer, lors de l'encodage, chaque mot de la phrase source est converti en un vecteur de mots, puis ces vecteurs sont soumis à un réseau neuronal récurrent (RNN). Les réseaux de neurones récurrents, tels que les LSTM (Long Short-Term Memory) et les GRU (Gated Recurrent Unit), sont utilisés pour capturer les dépendances à long terme dans les séquences de mots.

Ensuite, dans la phase de décodage, le décodeur RNN utilise les vecteurs cachés issus de l'encodage, ainsi qu'une représentation des mots déjà générés, pour prédire les mots suivants de la phrase cible.

Le mécanisme d'attention joue un rôle crucial en pondérant les vecteurs cachés de l'encodeur selon leur pertinence pour la prédiction du prochain mot, permettant ainsi une traduction plus précise et contextuelle. Cela aide à gérer les dépendances à long terme et améliore la précision des traductions en pondérant les informations pertinentes de l'encodeur.

Le modèle est entraîné de bout en bout pour minimiser la log-vraisemblance négative du corpus d'entraînement, ce qui signifie qu'il apprend à prédire les mots de la phrase cible en se basant sur la phrase source de manière optimale.

Approche utilisée : à base de LSTM, Transformers, etc.

OpenNMT utilise une approche diversifiée en intégrant plusieurs techniques avancées pour améliorer ses performances dans le domaine de la traduction automatique neuronale. Initialement, le système se base sur des réseaux de neurones récurrents (RNN), tels que les architectures LSTM (Long Short-Term Memory) et GRU (Gated Recurrent Unit), dotés de mécanismes d'attention, pour encoder et décoder les séquences. Cette approche consiste à encoder la séquence source en utilisant un RNN, qui parcourt la séquence mot par mot en attribuant des poids aux mots en fonction de leur importance dans le contexte global. Les RNN sont particulièrement adaptés pour saisir efficacement les dépendances à long terme et de capturer les informations contextuelles pour la traduction, ce qui les rend robustes pour la traduction de phrases complexes.

Plus récemment, OpenNMT a évolué en adoptant des modèles de type Transformer. Contrairement aux RNN, les Transformers utilisent uniquement des mécanismes d'attention, ce qui leur permet de traiter les relations entre les mots de manière plus globale et parallèle. Les Transformers se composent de plusieurs couches d'attention multi-têtes, chacune étant capable de se concentrer sur différentes parties de la séquence source. De plus, les mécanismes de positionnement sont employés pour fournir des informations sur la position de chaque mot dans la séquence. Cette approche permet une meilleure gestion des dépendances complexes entre les mots, améliorant ainsi la fluidité et la précision des traductions.

En outre, OpenNMT intègre des réseaux neuronaux convolutifs (CNN) pour la traduction. Les CNN, efficaces pour capter les motifs locaux, permettent un traitement plus parallèle des séquences par rapport aux RNN, ce qui accroît la vitesse tant de l'entraînement que de l'inférence. Le système supporte également une diversité de mécanismes d'attention, incluant des variantes telles que l'attention multi-têtes et l'attention hiérarchique, qui renforcent la capacité du modèle à focaliser sur les parties pertinentes de la phrase source, améliorant ainsi la précision et la contextualisation des traductions.

Pour garantir une qualité de traduction optimale, notamment pour les langues à ressources limitées ou pour les termes rares, OpenNMT utilise des stratégies de copie et de couverture. Ces mécanismes permettent au modèle de reproduire directement certains segments de la phrase source dans la phrase cible, tout en suivant les parties déjà traduites de la source. Cette approche facilite la gestion des informations manquantes ou des phrases partiellement traduites, renforçant ainsi la qualité globale des traductions.

Le framework OpenNMT est conçu de manière modulaire, offrant ainsi aux chercheurs la possibilité de remplacer ou d'ajuster facilement les composants du modèle pour répondre à des besoins spécifiques ou expérimenter de nouvelles idées. Cette flexibilité inclut diverses options pour les types d'attention, les mécanismes de copie et les normalisations pour la génération de texte, permettant ainsi une personnalisation optimale du système de traduction. En combinant ces diverses approches, OpenNMT offre une adaptabilité et une puissance exceptionnelles, permettant aux chercheurs et aux praticiens d'explorer et de déployer des systèmes de traduction optimisés pour une variété de tâches et d'applications en traitement du langage naturel.

3. Evaluation du moteur de traduction neuronale OpenNMT sur un corpus en formes fléchies

Décrire le corpus d'apprentissage et d'évaluation (nombre de phrases)

Après avoir séparé les corpus EMEA et Europarl en corpus d'entraînement, chacun des corpus contient ces nombres de phrases :

- **Entraînement** (train) :
 - Europarl (anglais, français) : 100 000 phrases
 - EMEA (anglais, français) : 10 000 phrases
- **Développement** (dev, référence) :

- Europarl (anglais, français) : 3 750 phrases
- **Test** (prédiction) :
 - Europarl (anglais, français) : 500 phrases
 - EMEA (anglais, français) : 500 phrases

Ensuite, nous avons nettoyé les corpus obtenus, ce qui consiste à tokeniser, mettre le texte en minuscule et à supprimer toutes les phrases ayant moins de 80 caractères. Après ce nettoyage, nous avons obtenu des corpus avec moins de phrases :

- **Entraînement** (train) :
 - Europarl (anglais, français) : 98 014 phrases
 - EMEA (anglais, français) : 9 877 phrases
- **Développement** (dev, référence) :
 - Europarl (anglais, français) : 3 673 phrases
- **Test** (prédiction) :
 - Europarl (anglais, français) : 489 phrases
 - EMEA (anglais, français) : 491 phrases

Décrire les métriques d'évaluation : Score BLEU, METEOR, BERT

Le score BLEU

Le Score BLEU (Bilingual Evaluation Understudy) est une métrique standard utilisée pour évaluer la qualité des traductions automatiques en comparant la sortie du modèle de traduction avec une ou plusieurs traductions de référence humaines.

Comment ce calcul le score Bleu :

1. **Comparaison des n-grammes** : Le score BLEU compare les n-grammes (groupes de mots de taille n) de la traduction générée avec ceux de la traduction de référence.
2. **Précision des n-grammes** : La précision mesure la proportion des n-grammes de la traduction générée qui se trouvent également dans la traduction de référence.
3. **Pénalité de longueur** : Pour éviter que les systèmes de traduction ne produisent des phrases trop courtes, une pénalité est appliquée si la traduction générée est plus courte que la traduction de référence.

Comment interpréter ce score :

- Le score BLEU varie de 0 à 100.
- Un score plus élevé indique une meilleure correspondance entre la traduction générée et la traduction de référence.
- Généralement, un score supérieur à 30 est considéré comme acceptable, bien que cela dépende du contexte et de la complexité du texte.

Le Score METEOR

Description :

Le score METEOR (Metric for Evaluation of Translation with Explicit ORdering) est une métrique d'évaluation de la traduction automatique qui vise à surmonter certaines limitations du score BLEU en prenant en compte les correspondances à un niveau plus fin, comme les synonymes et les variations grammaticales.

Calcul du Score METEOR :

1. Alignement : Trouver les correspondances exactes, les synonymes et les correspondances flexibles (par exemple, les variations grammaticales) entre la traduction générée et la traduction de référence.
2. Précision et Rappel : Calculer la précision (proportion des mots de la traduction générée trouvés dans la traduction de référence) et le rappel (proportion des mots de la traduction de référence trouvés dans la traduction générée).
3. Score F : Combinaison de la précision et du rappel en un score F, pondéré pour favoriser le rappel.
4. Pénalités de Fragmentation : Appliquer des pénalités pour les fragments de correspondance qui sont éloignés les uns des autres, favorisant ainsi des traductions plus fluides et cohérentes.

Interprétation :

- Le score METEOR varie généralement de 0 à 1.
- Un score plus élevé indique une meilleure qualité de traduction, prenant en compte les correspondances lexicales et les variations grammaticales.
- Il est souvent jugé comme étant plus sensible aux variations linguistiques que le score BLEU.

Le Score BERT

Description :

Le score BERT (BERTScore) utilise des représentations contextualisées des mots obtenues à partir de modèles de langage pré-entraînés comme BERT (Bidirectional Encoder Representations from Transformers) pour évaluer la similarité entre la traduction générée et la traduction de référence.

Calcul du Score BERT :

1. Encodage des phrases : Utiliser un modèle pré-entraîné (comme BERT) pour encoder les phrases de la traduction générée et de la traduction de référence en vecteurs de représentation contextuelle.
2. Calcul de la Similarité : Calculer la similarité cosinus entre les vecteurs des mots de la traduction générée et ceux de la traduction de référence.
3. Précision, Rappel, et F1-Score : Calculer la précision et le rappel des similarités des vecteurs, puis combiner ces métriques en un score F1.

Interprétation :

- Le score BERT varie généralement de 0 à 1.
- Un score plus élevé indique une meilleure correspondance entre la traduction générée et la traduction de référence au niveau des représentations sémantiques.
- BERTScore est apprécié pour sa capacité à capturer des nuances sémantiques et contextuelles, souvent ignorées par les autres métriques.

Tableau des résultats

Pour le run 1, nous avons utilisé comme corpus d'apprentissage le corpus Europarl, avec les configurations suivantes :

- **Entraînement :**
 - Corpus anglais : 98 014 phrases
 - Corpus français : 98 014 phrases
- **Développement :**
 - Corpus anglais : 3 750 phrases
 - Corpus français : 3 750 phrases

Concernant le run 2, nous avons élargi le corpus d'apprentissage en incluant à la fois le corpus Europarl et le corpus EMEA, avec les configurations suivantes :

- **Entraînement :**
 - Corpus Europarl anglais : 98 014 phrases
 - Corpus Europarl français : 98 014 phrases
 - Corpus EMEA anglais : 9 877 phrases

- Corpus EMEA français : 9 877 phrases
- Développement :
 - Corpus anglais : 3 750 phrases
 - Corpus français : 3 750 phrases

Voici les résultats obtenus :

Test	Run	BLEU	METEOR	BERT
Europarl_test	run1	29.34	0.4796	Précision moyenne: 0.8175 Rappel moyen: 0.8103 F1 moyen: 0.8134
Europarl_test	run2	27.83	0.4652	Précision moyenne: 0.8117 Rappel moyen: 0.8027 F1 moyen: 0.8067
Emea_test	run1	4.76	0.2019	Précision moyenne: 0.6683 Rappel moyen: 0.6826 F1 moyen: 0.6745
Emea_test	run2	14.63	0.3198	Précision moyenne: 0.7740 Rappel moyen: 0.7755 F1 moyen: 0.7739

1. *Europarl_test_500.tok.true.clean.fr*

Pour le corpus de test Europarl, le run 1 obtient un score BLEU de 29.34, légèrement supérieur au score du run 2 (27.83). Cela indique que le modèle entraîné uniquement sur le corpus Europarl (run 1) produit des traductions légèrement meilleures pour ce type de données par rapport au modèle du run 2, qui a été entraîné sur une combinaison des corpus Europarl et EMEA.

En ce qui concerne METEOR, le run 1 a également un score supérieur (0.4796) comparé au run 2 (0.4652), confirmant que les traductions de run 1 sont de meilleure qualité selon cette métrique.

Pour BERT, les scores de précision, rappel et F1 sont légèrement plus élevés pour le run 1, renforçant l'idée que le modèle est plus performant pour les textes Europarl.

2. *Emea_test_500.tok.true.clean.fr*

Pour le corpus de test EMEA, le run 1 obtient un score BLEU de 4.76, tandis que le run 2 obtient un score de 14.63. Cela montre une amélioration significative dans le run 2. Cette amélioration est due à l'inclusion du corpus EMEA dans l'entraînement du modèle pour le run 2, permettant au modèle de mieux généraliser et de produire des traductions de meilleure qualité pour des textes similaires à ceux du corpus EMEA.

En termes de METEOR, le run 2 surpassé également le run 1 (0.3198 contre 0.2019), indiquant une meilleure qualité de traduction pour EMEA.

Les scores BERT montrent une nette amélioration pour le run 2, avec des scores de précision, rappel et F1 tous supérieurs à ceux du run 1, ce qui corrobore les résultats des autres métriques.

Conclusion

Le run 1, avec un entraînement uniquement sur le corpus Europarl, montre de bonnes performances pour le test Europarl mais très faibles pour le test EMEA. Le run 2, avec un entraînement sur une combinaison des corpus Europarl et EMEA, montre une légère diminution des performances pour le test Europarl mais une amélioration significative pour le test EMEA. Ces résultats suggèrent que l'ajout de données d'entraînement provenant de différents domaines (comme le corpus EMEA) peut améliorer les performances de traduction pour des textes variés, bien que cela puisse légèrement affecter les performances pour des textes très similaires aux données d'entraînement initiales (Europarl).

4. Evaluation du moteur de traduction neuronale OpenNMT sur un corpus en lemmes

Description des lemmatisateurs

La lemmatisation commence par tokeniser le texte avec `nltk.word_tokenize()`. Puis, pour chaque token de l'anglais, sa partie du discours est extraite à l'aide de `nltk.pos_tag()`. En revanche, pour les tokens français, `nltk.pos_tag()` ne permet pas de déterminer leurs parties du discours. La partie du discours joue un rôle essentiel pour perfectionner la lemmatisation des tokens. Le lemmatiseur NLTK pour l'anglais utilisé ici est `WordNetLemmatizer()`, qui prend à la fois le token et sa partie

du discours correspondante en paramètres. Pour la lemmatisation des tokens français, ***FrenchLefffLemmatizer()*** est utilisé, et il prend seulement le token en paramètre. En l'absence des parties du discours, la lemmatisation des tokens français est moins précise.

Tableau des résultats

La distribution des corpus pour l'apprentissage est la même que pour l'apprentissage avec le corpus ayant les formes fléchies, à la seule différence qu'ici nous utilisons les corpus en lemmes.

Voici les résultats obtenus :

Test	Run	BLEU	METEOR	BERT
lemme_Euro parl_test	run1_lemma	28.75	0.4690	Précision moyenne: 0.8147 Rappel moyen: 0.7998 F1 moyen: 0.8068
lemme_Euro parl_test	run2_lemma	24.26	0.4452	Précision moyenne: 0.7733 Rappel moyen: 0.7693 F1 moyen: 0.7707
lemme_Emea _test	run1_lemma	5.24	0.2111	Précision moyenne: 0.6747 Rappel moyen: 0.6841 F1 moyen: 0.6786
lemme_Emea _test	run2_lemma	13.42	0.3165	Précision moyenne: 0.7571 Rappel moyen: 0.7618 F1 moyen: 0.7586

1. *lemme_Europarl_test_500.tok.true.clean.fr*

Pour le corpus de test Europarl en lemmes, le run 1 obtient un score BLEU de 28.75, tandis que le run 2 obtient un score de 24.26. Cela indique que le modèle entraîné uniquement sur le corpus Europarl (run 1) produit des traductions en lemmes de meilleure qualité pour ce type de données par rapport au modèle du run 2, qui a été entraîné sur une combinaison des corpus Europarl et EMEA. La diminution des

performances du run 2 peut être due à la diversité accrue des données d'entraînement, qui peut introduire une variabilité affectant la précision pour des textes très similaires aux données Europarl.

En ce qui concerne METEOR, le run 1 a également un score supérieur (0.4690) comparé au run 2 (0.4452), ce qui confirme la meilleure qualité des traductions pour le run 1.

Pour BERT, les scores de précision, rappel et F1 sont plus élevés pour le run 1, renforçant l'idée que le modèle est plus performant pour les textes Europarl en lemmes.

2. lemme_Emea_test_500.tok.true.clean.fr

Pour le corpus de test EMEA en lemmes, le run 1 obtient un score BLEU de 5.24, tandis que le run 2 obtient un score de 13.42. Cette amélioration significative dans le run 2 montre que l'inclusion du corpus EMEA dans l'entraînement du modèle permet une meilleure généralisation et améliore la qualité des traductions pour des textes similaires à ceux du corpus EMEA.

En termes de METEOR, le run 2 surpassé également le run 1 (0.3165 contre 0.2111), indiquant une meilleure qualité de traduction pour EMEA.

Les scores BERT montrent une nette amélioration pour le run 2, avec des scores de précision, rappel et F1 tous supérieurs à ceux du run 1, ce qui corrobore les résultats des autres métriques.

Conclusion

Les résultats pour les corpus en lemmes montrent une tendance similaire à celle observée avec les formes fléchies, mais avec quelques différences notables. Le run 1_lemma, avec un entraînement uniquement sur le corpus Europarl en lemmes, montre de bonnes performances pour le test Europarl en lemmes mais très faibles pour le test EMEA en lemmes. Le run 2_lemma, avec un entraînement sur une combinaison des corpus Europarl et EMEA en lemmes, montre une légère diminution des performances pour le test Europarl en lemmes mais une amélioration significative pour le test EMEA en lemmes.

Comparé aux formes fléchies, les performances globales en lemmes sont légèrement inférieures, ce qui pourrait indiquer que la lemmatisation introduit une certaine complexité supplémentaire dans la modélisation. Cependant, l'ajout de données d'entraînement provenant de différents domaines (comme le corpus EMEA) améliore toujours les performances de traduction pour des textes variés, bien que cela puisse légèrement affecter les performances pour des textes très similaires aux données d'entraînement initiales (Europarl).

En résumé, les résultats montrent l'importance d'utiliser des données d'entraînement diversifiées pour améliorer la robustesse et la généralisation des modèles de

traduction automatique, ce qui est reflété par les scores des différentes métriques d'évaluation utilisées (BLEU, METEOR, BERT).

5. Points forts, limitations d'OpenNMT et difficultés rencontrées

Points forts du moteur de traduction

OpenNMT est un moteur de traduction gratuit, accessible à tout le monde et assez intelligent. Seulement avec un corpus d'entraînement de 100,000 phrases, il est déjà capable de proposer une traduction compréhensible et plutôt fiable sur un texte du même domaine que le corpus d'entraînement. Par ailleurs, le moteur de l'OpenNMT ne requiert aucune expertise linguistique et sa performance penche en grande partie sur la taille, la qualité du corpus d'entraînement. Cela suggère que plus le corpus d'entraînement est pertinent et enrichissant, plus la traduction automatique est raffinée.

En outre, il semble que le moteur de traduction OpenNMT traduit mieux les mots fléchis par rapport aux lemmes. Les résultats des évaluations réalisées évoquent que le moteur obtient un score BLEU moyen d'environ 1.22 de plus sur les mots fléchis. Si les résultats ne sont pas biaisés, cela signifie que l'étape de lemmatisation peut être ignorée lors du prétraitement de corpus.

Limitation du moteur de traduction

Selon les traductions obtenues, il est évident que le moteur de traduction est très faible quand il s'agit d'un texte hors domaine. Dans le cas où le moteur entraîné sur uniquement les textes de l'Europarl traduit un texte de l'EMEA, de nombreux "<unk>" (mot inconnu) sont présents dans la traduction, ce qui la rend totalement incompréhensible. Cette situation est largement améliorée dans le run n°2 avec l'ajout de textes de l'EMEA dans le corpus d'entraînement. Cela reflète indirectement que le moteur de traduction éprouve des difficultés à traiter les textes source d'un domaine inconnu.

L'entraînement du modèle est un processus très long nécessitant des ressources informatiques puissantes. Sur un ordinateur avec 8 Go de RAM sans GPU, l'apprentissage du modèle pour un corpus de 10,000 phrases prend plus d'une heure. En raison de la nature coûteuse de l'apprentissage, cette méthode peut être peu pratique pour les utilisateurs dépourvus de matériel adéquat.

Difficultés rencontrées

Zhiya est bloquée dès l'étape 2 du processus d'entraînement du modèle en raison des capacités limitées de son ordinateur.

L'un de nous ayant un ordinateur équipé d'un GPU, nous avons décidé de l'utiliser. Cependant, nous avons rencontré un problème lors de la construction des modèles pendant l'entraînement, peut-être à cause de Windows, bien que nous n'en soyons pas certains. En effet, après l'entraînement, lorsque nous avons lancé la commande pour la traduction des prédictions, le fichier de sortie ne contenait que des <unk> et aucun mot. Après plusieurs heures de recherche, nous avons trouvé la solution : il fallait modifier l'une des lignes du programme inputters d'OpenNMT.

6. Organisation

- Github du projet
 - Zhongjie LI
 - Perrine QUENNEHEN
 - Zhiya MA
- Code du projet
 - Zhongjie LI
 - Perrine QUENNEHEN
- Rédaction du rapport
 - Zhongjie Li
 - Zhiya MA
 - Perrine QUENNEHEN

Annexes

Expérimentations en plus

→ Utilisation du lemmatisateur de Spacy :

Résultat du score Bleu obtenu :

Test	Run	BLEU	METEOR	BERT
lemme_Europarl_test	run1_lemma_spacy	30.35	0.4919	Précision moyenne: 0.8204 Rappel moyen: 0.8066 F1 moyen: 0.8131
lemme_Europarl_test	run2_lemma_spacy	30.58	0.5015	Précision moyenne: 0.8157 Rappel moyen: 0.8093 F1 moyen: 0.8121
lemme_Emea_test	run1_lemma_spacy	7.86	0.2574	Précision moyenne: 0.7194 Rappel moyen: 0.7138 F1 moyen: 0.7160
lemme_Emea_test	run2_lemma_spacy	16.19	0.3677	Précision moyenne: 0.7960 Rappel moyen: 0.7881 F1 moyen: 0.7913

Comparaison avec les résultats obtenus avec le lemmatiseur de NLTK et FrenchLefffLemmatize :

1. *Europarl test* :

Pour le corpus de test Europarl, les scores BLEU obtenus avec le lemmatiseur de Spacy sont supérieurs à ceux obtenus avec NLTK. Les scores METEOR, qui prennent en compte des correspondances plus flexibles comme les synonymes et les réordonnancements, montrent également une amélioration avec Spacy (0.4919 et 0.5015) par rapport à NLTK (0.4690 et 0.4452). De plus, les mesures de précision, rappel et F1 sont légèrement meilleures avec Spacy, indiquant une meilleure qualité globale des traductions.

2. *Emea test* :

Pour le corpus de test EMEA, les scores BLEU obtenus avec Spacy sont encore une fois supérieurs à ceux obtenus avec NLTK. Les scores METEOR montrent

également une nette amélioration (0.2574 et 0.3677 avec Spacy contre 0.2111 et 0.3165 avec NLTK). Les mesures de précision, rappel et F1 sont également meilleures avec Spacy, indiquant une meilleure performance générale du modèle.

Conclusion

L'utilisation du lemmatiseur de Spacy pour le français a conduit à une amélioration des scores BLEU, METEOR, et des mesures de précision, rappel et F1 par rapport à l'utilisation de NLTK. Ces résultats montrent que le lemmatiseur de Spacy traite mieux les particularités linguistiques du français, ce qui se traduit par des traductions de meilleure qualité. Le choix du lemmatiseur peut donc avoir un impact significatif sur les performances des modèles de traduction automatique.

→ Utilisation des encodeurs/décodeurs transformers d'OpenNMT

Utilisant seulement 5000 train steps et non 10 000, contrairement aux résultats précédent utilisant les décodeurs/encodeurs RNN, car trop long sinon:

Formes Fléchies :

Test	Run	BLEU	METEOR	BERT
Europarl_t est	run1_transform er	32.13	0.5100	Précision moyenne: 0.8292 Rappel moyen: 0.8215 F1 moyen: 0.8250
Europarl_t est	run2_transform er	33.32	0.5226	Précision moyenne: 0.8300 Rappel moyen: 0.8261 F1 moyen: 0.8277
Emea_test	run1_transform er	7.81	0.2434	Précision moyenne: 0.7049 Rappel moyen: 0.7081 F1 moyen: 0.7057
Emea_test	run2_transform er	32.35	0.5072	Précision moyenne: 0.8344 Rappel moyen: 0.8424 F1 moyen: 0.8375

Formes Lemmatisées :

Test	Run	BLEU	METEOR	BERT
lemme_Eur oparl_test	run1_lemma_transformer	33.04	0.5214	Précision moyenne: 0.8102 Rappel moyen: 0.8149 F1 moyen: 0.8121
lemme_Eur oparl_test	run2_lemma_transformer	32.99	0.5239	Précision moyenne: 0.8119 Rappel moyen: 0.8166 F1 moyen: 0.8139
lemme_Ema ea_test	run1_lemma_transformer	6.64	0.2292	Précision moyenne: 0.6806 Rappel moyen: 0.6889 F1 moyen: 0.6837
lemme_Ema ea_test	run2_lemma_transformer	33.51	0.5003	Précision moyenne: 0.8384 Rappel moyen: 0.8374 F1 moyen: 0.8370
lemme_Eur oparl_test	run1_spacy_transformer	35.02	0.5322	Précision moyenne: 0.8205 Rappel moyen: 0.8224 F1 moyen: 0.8211
lemme_Eur oparl_test	run2_spacy_transformer	35.29	0.5347	Précision moyenne: 0.8233 Rappel moyen: 0.8234 F1 moyen: 0.8230
lemme_Ema ea_test	run1_spacy_transformer	8.28	0.2667	Précision moyenne: 0.6977 Rappel moyen: 0.7060 F1 moyen: 0.7010
lemme_Ema ea_test	run2_spacy_transformer	36.87	0.5262	Précision moyenne: 0.8454 Rappel moyen: 0.8432 F1 moyen: 0.8435

Observations Générales

1. Performance Supérieure des Transformers

- Les Transformers obtiennent généralement des scores BLEU plus élevés que les RNNs, indiquant une meilleure performance globale dans les tâches de traduction automatique. Cette tendance est constante pour les corpus en formes fléchies et en formes lemmatisées. Les scores METEOR, Précision, Rappel et F1 des Transformers surpassent également ceux des RNNs, démontrant une qualité supérieure des traductions.

2. Efficacité de l'Entraînement

- Les Transformers atteignent des performances optimales avec seulement 5 000 étapes d'entraînement, comparativement aux 10 000 étapes nécessaires pour les RNNs. Cela suggère une convergence plus rapide des Transformers, permettant un entraînement plus efficace.

Analyse par Type de Corpus

Formes Fléchies

- Europarl_test
 - Les Transformers surpassent les RNNs avec des scores BLEU de 32-33 contre 28-29, ainsi que des scores METEOR et des mesures de Précision, Rappel et F1 plus élevés.
- Emea_test
 - Les Transformers montrent une amélioration spectaculaire, avec des scores BLEU de 7.81 et 32.35 contre 4.76 et 14.63 pour les RNNs, ainsi que des scores METEOR et des mesures de Précision, Rappel et F1 plus élevés.

Formes Lemmatisées

- Avec NLTK
 - Europarl_test
 - Les scores BLEU des Transformers sont de 33.04 et 32.99, comparés à 28.75 et 24.25 pour les RNNs. Les scores METEOR et les mesures de Précision, Rappel et F1 sont également supérieurs.
 - Emea_test
 - Les scores BLEU des Transformers (6.64 et 33.51) sont nettement supérieurs à ceux des RNNs (5.24 et 13.42), ainsi que les scores METEOR et les mesures de Précision, Rappel et F1.
- Avec Spacy
 - Europarl_test
 - Les Transformers affichent des scores BLEU de 35.01 et 35.29 contre 30.35 et 30.58 pour les RNNs, ainsi que des scores METEOR et des mesures de Précision, Rappel et F1 plus élevés.

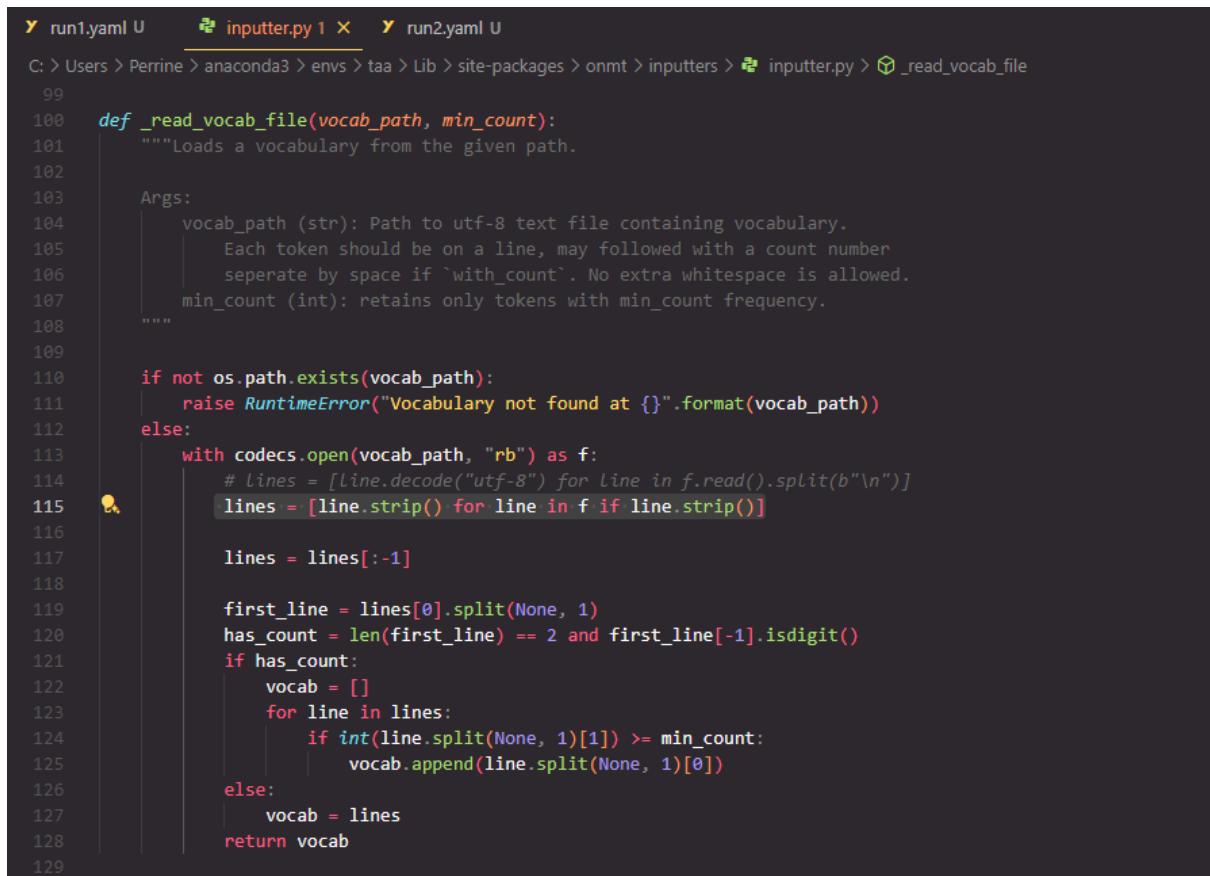
- Emea_test
 - Les Transformers obtiennent des scores BLEU de 8.28 et 36.87, tandis que les RNNs obtiennent 7.86 et 16.18, avec des scores METEOR et des mesures de Précision, Rappel et F1 plus élevés.

Conclusion

Les résultats montrent que les encodeurs/décodeurs basés sur les Transformers surpassent largement ceux basés sur les RNNs en termes de performance (score BLEU, METEOR, Précision, Rappel et F1) et d'efficacité d'entraînement (nombre d'étapes). Les Transformers sont donc recommandés pour des tâches similaires de traduction automatique ou de traitement du langage naturel, offrant des traductions de meilleure qualité et une convergence plus rapide. Cependant, ils nécessitent impérativement une machine équipée d'un GPU. En effet, même avec deux fois moins d'étapes que les encodeurs/décodeurs RNNs, l'entraînement avec les Transformers a pris au moins 4 heures, contre seulement 15 à 20 minutes pour les RNNs.

Captures d'écran des résultats

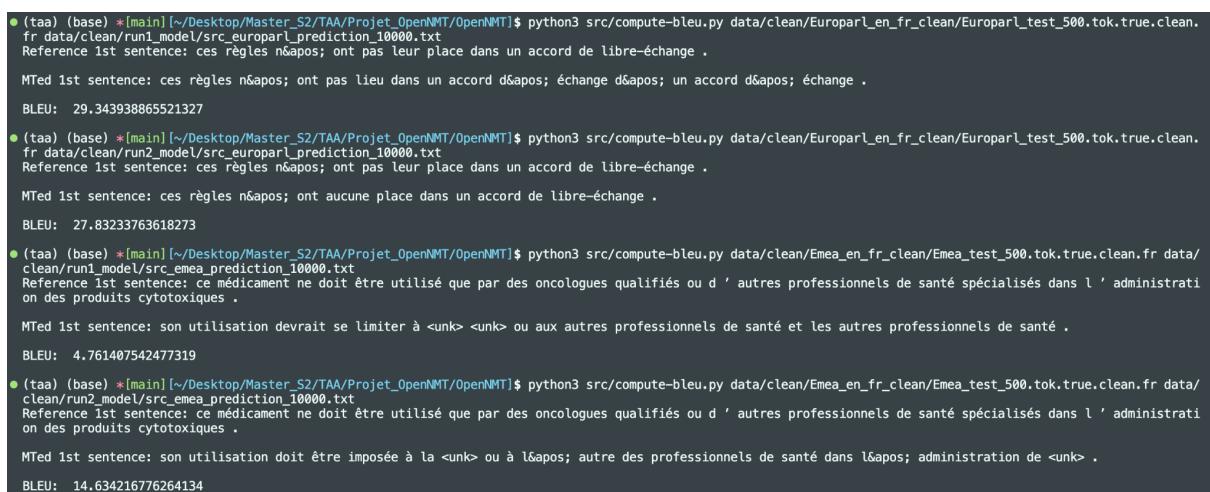
Correction ligne de code inputters.py:



```
run1.yaml U      inputters.py 1 X    run2.yaml U
C: > Users > Perrine > anaconda3 > envs > taa > site-packages > onmt > inputters > inputters.py > _read_vocab_file
99
100 def _read_vocab_file(vocab_path, min_count):
101     """Loads a vocabulary from the given path.
102
103     Args:
104         vocab_path (str): Path to utf-8 text file containing vocabulary.
105         Each token should be on a line, may be followed with a count number
106         separated by space if `with_count`. No extra whitespace is allowed.
107         min_count (int): retains only tokens with min_count frequency.
108
109     if not os.path.exists(vocab_path):
110         raise RuntimeError("Vocabulary not found at {}".format(vocab_path))
111     else:
112         with codecs.open(vocab_path, "rb") as f:
113             # Lines = [line.decode("utf-8") for line in f.read().split(b"\n")]
114             lines = [line.strip() for line in f if line.strip()]
115
116             lines = lines[:-1]
117
118             first_line = lines[0].split(None, 1)
119             has_count = len(first_line) == 2 and first_line[-1].isdigit()
120             if has_count:
121                 vocab = []
122                 for line in lines:
123                     if int(line.split(None, 1)[1]) >= min_count:
124                         vocab.append(line.split(None, 1)[0])
125
126             else:
127                 vocab = lines
128
129     return vocab
```

Exercice 3 :

Score Bleu



```
(taa) (base) *[main] [/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT]$ python3 src/compute-bleu.py data/clean/Europarl_en_fr_clean/Europarl_test_500.tok.true.clean.fr data/clean/run1_model/src_europarl_prediction_10000.txt
Reference 1st sentence: ces règles n&apos;ont pas leur place dans un accord de libre-échange .
MTed 1st sentence: ces règles n&apos;ont pas lieu dans un accord d&apos;échange d&apos;un accord d&apos;échange .
BLEU: 29.343938865521327

(taa) (base) *[main] [/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT]$ python3 src/compute-bleu.py data/clean/Europarl_en_fr_clean/Europarl_test_500.tok.true.clean.fr data/clean/run2_model/src_europarl_prediction_10000.txt
Reference 1st sentence: ces règles n&apos;ont pas leur place dans un accord de libre-échange .
MTed 1st sentence: ces règles n&apos;ont aucune place dans un accord de libre-échange .
BLEU: 27.832337636182739

(taa) (base) *[main] [/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT]$ python3 src/compute-bleu.py data/clean/Emea_en_fr_clean/Emea_test_500.tok.true.clean.fr data/clean/run1_model/src_emea_prediction_10000.txt
Reference 1st sentence: ce médicament ne doit être utilisé que par des oncologues qualifiés ou d'autres professionnels de santé spécialisés dans l'administration de produits cytotoxiques .
MTed 1st sentence: son utilisation devrait se limiter à <unk> <unk> ou aux autres professionnels de santé et les autres professionnels de santé .
BLEU: 4.761407542477319

(taa) (base) *[main] [/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT]$ python3 src/compute-bleu.py data/clean/Emea_en_fr_clean/Emea_test_500.tok.true.clean.fr data/clean/run2_model/src_emea_prediction_10000.txt
Reference 1st sentence: ce médicament ne doit être utilisé que par des oncologues qualifiés ou d'autres professionnels de santé spécialisés dans l'administration de produits cytotoxiques .
MTed 1st sentence: son utilisation doit être imposée à la <unk> ou à l'administration de <unk> .
BLEU: 14.634216776264134
```

Score METEOR

```
(taa) *[main] [/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT]$ python3 src/meteor-score.py data/clean/Europarl_en_fr_clean/Europarl_test_500.tok.true.clean.fr  
r data/clean/run1_model/src_europarl_prediction_10000.txt  
Executing command: java -jar src/meteor-1.5/meteor-1.5.jar data/clean/run1_model/src_europarl_prediction_10000.txt data/clean/Europarl_en_fr_clean/Europarl_t  
est_500.tok.true.clean.fr -l fr -norm  
METEOR Score: Final score: 0.47964957150977366  
(taa) *[main] [/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT]$ python3 src/meteor-score.py data/clean/Europarl_en_fr_clean/Europarl_test_500.tok.true.clean.fr  
r data/clean/run2_model/src_europarl_prediction_10000.txt  
Executing command: java -jar src/meteor-1.5/meteor-1.5.jar data/clean/run2_model/src_europarl_prediction_10000.txt data/clean/Europarl_en_fr_clean/Europarl_t  
est_500.tok.true.clean.fr -l fr -norm  
METEOR Score: Final score: 0.46651679812067612  
(taa) *[main] [/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT]$ python3 src/meteor-score.py data/clean/Emea_en_fr_clean/Emea_test_500.tok.true.clean.fr data/c  
lean/run1_model/src_emea_prediction_10000.txt  
  
Executing command: java -jar src/meteor-1.5/meteor-1.5.jar data/clean/run1_model/src_emea_prediction_10000.txt data/clean/Emea_en_fr_clean/Emea_test_500.tok.  
true.clean.fr -l fr -norm  
METEOR Score: Final score: 0.281921335538312  
(taa) *[main] [/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT]$ python3 src/meteor-score.py data/clean/Emea_en_fr_clean/Emea_test_500.tok.true.clean.fr data/c  
lean/run2_model/src_emea_prediction_10000.txt  
  
Executing command: java -jar src/meteor-1.5/meteor-1.5.jar data/clean/run2_model/src_emea_prediction_10000.txt data/clean/Emea_en_fr_clean/Emea_test_500.tok.  
true.clean.fr -l fr -norm  
METEOR Score: Final score: 0.3197504521397011
```

Score BERT

```
(taa) *[main] [/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT]$ python3 src/bert-score.py data/clean/lematised/lemme_Europarl_test_500.tok.true.clean.fr data/  
clean/run1_model_lemma/src_europarl_prediction_10000.txt  
calculating scores...  
computing bert embedding.  
100%|██████████| 16/16 [00:12<00:00,  1.33it/s]  
computing greedy matching.  
100%|██████████| 8/8 [00:00<00:00, 105.42it/s]  
done in 12.12 seconds, 40.34 sentences/sec  
Précision moyenne: 0.8147  
Rappel moyen: 0.7998  
F1 moyen: 0.8068  
(taa) *[main] [/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT]$ python3 src/bert-score.py data/clean/lematised/lemme_Europarl_test_500.tok.true.clean.fr data/  
clean/run2_model_lemma/src_europarl_prediction_10000.txt  
calculating scores...  
computing bert embedding.  
100%|██████████| 16/16 [00:11<00:00,  1.33it/s]  
computing greedy matching.  
100%|██████████| 8/8 [00:00<00:00, 120.42it/s]  
done in 12.09 seconds, 40.44 sentences/sec  
Précision moyenne: 0.7733  
Rappel moyen: 0.7693  
F1 moyen: 0.7707  
(taa) *[main] [/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT]$ python3 src/bert-score.py data/clean/lematised/lemme_Emea_test_500.tok.true.clean.fr data/c  
lean/run2_model_lemma/src_emea_prediction_10000.txt  
calculating scores...  
computing bert embedding.  
100%|██████████| 14/14 [00:09<00:00,  1.43it/s]  
computing greedy matching.  
100%|██████████| 8/8 [00:00<00:00, 120.33it/s]  
done in 9.85 seconds, 49.83 sentences/sec  
Précision moyenne: 0.7571  
Rappel moyen: 0.7618  
F1 moyen: 0.7586  
(taa) *[main] [/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT]$ python3 src/bert-score.py data/clean/lematised/lemme_Emea_test_500.tok.true.clean.fr data/c  
lean/run1_model_lemma/src_emea_prediction_10000.txt  
calculating scores...  
computing bert embedding.  
100%|██████████| 14/14 [00:13<00:00,  1.05it/s]  
computing greedy matching.  
100%|██████████| 8/8 [00:00<00:00, 95.82it/s]  
done in 13.46 seconds, 36.48 sentences/sec  
Précision moyenne: 0.6747  
Rappel moyen: 0.6841  
F1 moyen: 0.6786
```

Exercice 4 :

BLEU

```
● (taa) *(main)[~/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT]$ python3 src/compute-bleu.py data/clean/lematised/lemme_Europarl_test_500.tok.true.cl  
ean.fr data/clean/run1_model_lemma/src_europarl_prediction_10000.txt  
Reference 1st sentence: ces règle n &apos ; ont pas leur place dans un accord de libre-échange .  
MTed 1st sentence: ces règle n ' ont aucune place dans un accord de libre-échange libre .  
BLEU: 28.754770083442274  
● (taa) *(main)[~/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT]$ python3 src/compute-bleu.py data/clean/lematised/lemme_Europarl_test_500.tok.true.cl  
ean.fr data/clean/run2_model_lemma/src_europarl_prediction_10000.txt  
Reference 1st sentence: ces règle n &apos ; ont pas leur place dans un accord de libre-échange .  
MTed 1st sentence: ces règle n &apos ; ont pas de place dans un accord de libre-échange .  
BLEU: 24.25517752387918  
● (taa) *(main)[~/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT]$ python3 src/compute-bleu.py data/clean/lematised/lemme_Emea_test_500.tok.true.clean.  
fr data/clean/run1_model_lemma/src_emea_prediction_10000.txt  
Reference 1st sentence: ce médicament ne doit être utilisé que par des oncologues qualifiés ou d ' autres professionnel de santé spécialisés dans l ' admin  
istration des produit cytotoxiques .  
MTed 1st sentence: elle devrait être limitée de se limiter à des <unk> ou autres professionnel de la santé dans les administration .  
BLEU: 5.243602298627247  
● (taa) *(main)[~/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT]$ python3 src/compute-bleu.py data/clean/lematised/lemme_Emea_test_500.tok.true.clean.  
fr data/clean/run2_model_lemma/src_emea_prediction_10000.txt  
Reference 1st sentence: ce médicament ne doit être utilisé que par des oncologues qualifiés ou d ' autres professionnel de santé spécialisés dans l ' admin  
istration des produit cytotoxiques .  
MTed 1st sentence: il s &apos ; agit de l &apos ; utilisation de l &apos ; utilisation de la santé ou d &apos ; autres professionnel de la santé  
BLEU: 13.422551347664125
```

METEOR

```
(taa) *[main][~/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT]$ python3 src/meteor-score.py data/clean/lematised/lemme_Europarl_test_500.tok.true.clean.fr da  
ta/clean/run1_model_lemma/src_europarl_prediction_10000.txt  
Executing command: java -jar src/meteor-1.5/meteor-1.5.jar data/clean/run1_model_lemma/src_europarl_prediction_10000.txt data/clean/lematised/lemme_Europarl  
_test_500.tok.true.clean.fr -l fr -norm  
METEOR Score: Final score: 0.46896623505185897  
(taa) *[main][~/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT]$ python3 src/meteor-score.py data/clean/lematised/lemme_Europarl_test_500.tok.true.clean.fr da  
ta/clean/run2_model_lemma/src_europarl_prediction_10000.txt  
Executing command: java -jar src/meteor-1.5/meteor-1.5.jar data/clean/run2_model_lemma/src_europarl_prediction_10000.txt data/clean/lematised/lemme_Europarl  
_test_500.tok.true.clean.fr -l fr -norm  
METEOR Score: Final score: 0.44517516830472304  
(taa) *[main][~/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT]$ python3 src/meteor-score.py data/clean/lematised/lemme_Emea_test_500.tok.true.clean.fr data/c  
lean/run2_model_lemma/src_emea_prediction_10000.txt  
Executing command: java -jar src/meteor-1.5/meteor-1.5.jar data/clean/run2_model_lemma/src_emea_prediction_10000.txt data/clean/lematised/lemme_Emea_test_50  
0.tok.true.clean.fr -l fr -norm  
METEOR Score: Final score: 0.3164610622072863  
(taa) *[main][~/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT]$ python3 src/meteor-score.py data/clean/lematised/lemme_Emea_test_500.tok.true.clean.fr data/c  
lean/run1_model_lemma/src_emea_prediction_10000.txt  
Executing command: java -jar src/meteor-1.5/meteor-1.5.jar data/clean/run1_model_lemma/src_emea_prediction_10000.txt data/clean/lematised/lemme_Emea_test_50  
0.tok.true.clean.fr -l fr -norm  
METEOR Score: Final score: 0.21108904049135394
```

BERT

```
(taa) *(main)[~/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT]$ python3 src/bert-score.py data/clean/lematised/lemme_Europarl_test_500.tok.true.clean.fr data  
/clean/run1_model_lemma/src_europarl_prediction_10000.txt  
calculating scores...  
computing bert embedding.  
100%|██████████| 16/16 [00:12<00:00,  1.33it/s]  
computing greedy matching.  
100%|██████████| 8/8 [00:00<00:00, 105.42it/s]  
done in 12.12 seconds, 40.34 sentences/sec  
Précision moyenne: 0.8147  
Rappel moyen: 0.7998  
F1 moyen: 0.8068  
(taa) *(main)[~/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT]$ python3 src/bert-score.py data/clean/lematised/lemme_Europarl_test_500.tok.true.clean.fr data  
/clean/run2_model_lemma/src_europarl_prediction_10000.txt  
calculating scores...  
computing bert embedding.  
100%|██████████| 16/16 [00:11<00:00,  1.33it/s]  
computing greedy matching.  
100%|██████████| 8/8 [00:00<00:00, 120.42it/s]  
done in 12.09 seconds, 40.44 sentences/sec  
Précision moyenne: 0.7733  
Rappel moyen: 0.7693  
F1 moyen: 0.7707  
(taa) *(main)[~/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT]$ python3 src/bert-score.py data/clean/lematised/lemme_Emea_test_500.tok.true.clean.fr data/c  
lean/run1_model_lemma/src_emea_prediction_10000.txt  
calculating scores...  
computing bert embedding.  
100%|██████████| 14/14 [00:09<00:00,  1.43it/s]  
computing greedy matching.  
100%|██████████| 8/8 [00:00<00:00, 120.33it/s]  
done in 9.85 seconds, 49.83 sentences/sec  
Précision moyenne: 0.7571  
Rappel moyen: 0.7618  
F1 moyen: 0.7586  
(taa) *(main)[~/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT]$ python3 src/bert-score.py data/clean/lematised/lemme_Emea_test_500.tok.true.clean.fr data/c  
lean/run2_model_lemma/src_emea_prediction_10000.txt  
calculating scores...  
computing bert embedding.  
100%|██████████| 14/14 [00:13<00:00,  1.05it/s]  
computing greedy matching.  
100%|██████████| 8/8 [00:00<00:00, 95.82it/s]  
done in 13.46 seconds, 36.48 sentences/sec  
Précision moyenne: 0.6747  
Rappel moyen: 0.6841  
F1 moyen: 0.6786
```

Spacy

BLEU

```
(taa) *[main] [/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT]$ python3 src/compute-bleu.py data/clean/lematised_spacy/lemme_Europarl_test_500.tok.true.clean.fr data/clean/run1_model_lemma_spacy/src_europarl_prediction_10000.txt
Référence 1st sentence: ce règle n&apo ; avoir pas son place dans un accord de libre - échange .
MTed 1st sentence: ce règle n&apo ; avoir pas lieu dans un accord de libre - échange .

BLEU: 30.353478663617846

(taa) *[main] [/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT]$ python3 src/compute-bleu.py data/clean/lematised_spacy/lemme_Emea_test_500.tok.true.clean.fr data/clean/run1_model_lemma_spacy/src_emea_prediction_10000.txt
Référence 1st sentence: ce médicament ne devoir être utiliser que par un oncologue qualifier ou d ' autre professionnel de santé spécialiser dans l ' administration de produit cytotoxique .

MTed 1st sentence: lui devoir être limiter à l&apo ; appel qualifié ou à autre professionnel de santé de l'apos ; administration dans l'apos ; administration de SPG .

BLEU: 7.863586295088196

(taa) *[main] [/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT]$ python3 src/compute-bleu.py data/clean/lematised_spacy/lemme_Europarl_test_500.tok.true.clean.fr data/clean/run2_model_lemma_spacy/src_europarl_prediction_10000.txt
Référence 1st sentence: ce règle n&apo ; avoir pas son place dans un accord de libre - échange .

MTed 1st sentence: ce règle n&apos ; avoir aucun place dans un accord de libre - échange .

BLEU: 30.58154882349212

(taa) *[main] [/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT]$ python3 src/compute-bleu.py data/clean/lematised_spacy/lemme_Emea_test_500.tok.true.clean.fr data/clean/run2_model_lemma_spacy/src_emea_prediction_10000.txt
Référence 1st sentence: ce médicament ne devoir être utiliser que par un oncologue qualifier ou d ' autre professionnel de santé spécialiser dans l ' administration de produit cytotoxique .

MTed 1st sentence: il falloir se limiter à le base de l'apos ; emploi ou à autre professionnel de santé dans l'apos ; administration de l'apos ; agent a .

BLEU: 16.189072549808355
```

BERT

```
(taa) *[main] [/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT]$ python3 src/bert-score.py data/clean/lematised_spacy/lemme_Europarl_test_500.tok.true.clean.fr data/clean/run1_model_lemma_spacy/src_europarl_prediction_10000.txt
calculating scores...
computing bert embedding.
100%|██████████| 16/16 [00:11<00:00, 1.36it/s]
computing greedy matching.
100%|██████████| 8/8 [00:00<00:00, 101.24it/s]
done in 11.83 seconds, 41.33 sentences/sec
Précision moyenne: 0.8204
Rappel moyen: 0.8866
F1 moyen: 0.8131
(taa) *[main] [/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT]$ python3 src/bert-score.py data/clean/lematised_spacy/lemme_Europarl_test_500.tok.true.clean.fr data/clean/run2_model_lemma_spacy/src_europarl_prediction_10000.txt
calculating scores...
computing bert embedding.
100%|██████████| 16/16 [00:12<00:00, 1.32it/s]
computing greedy matching.
100%|██████████| 8/8 [00:00<00:00, 95.82it/s]
done in 12.21 seconds, 40.04 sentences/sec
Précision moyenne: 0.8157
Rappel moyen: 0.8893
F1 moyen: 0.8121
(taa) *[main] [/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT]$ python3 src/bert-score.py data/clean/lematised_spacy/lemme_Emea_test_500.tok.true.clean.fr data/clean/run1_model_lemma_spacy/src_emea_prediction_10000.txt
calculating scores...
computing bert embedding.
100%|██████████| 14/14 [00:08<00:00, 1.65it/s]
computing greedy matching.
100%|██████████| 8/8 [00:00<00:00, 88.88it/s]
done in 8.62 seconds, 56.98 sentences/sec
Précision moyenne: 0.7194
Rappel moyen: 0.7338
F1 moyen: 0.7168
(taa) *[main] [/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT]$ python3 src/bert-score.py data/clean/lematised_spacy/lemme_Emea_test_500.tok.true.clean.fr data/clean/run2_model_lemma_spacy/src_emea_prediction_10000.txt
calculating scores...
computing bert embedding.
100%|██████████| 14/14 [00:08<00:00, 1.58it/s]
computing greedy matching.
100%|██████████| 8/8 [00:00<00:00, 126.70it/s]
done in 8.94 seconds, 54.93 sentences/sec
Précision moyenne: 0.7968
Rappel moyen: 0.7881
F1 moyen: 0.7913
```

METEOR

```
(taa) *[main] [/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT]$ python3 src/meteor-score.py data/clean/lematised_spacy/lemme_Europarl_test_500.tok.true.clean.fr data/clean/run1_model_lemma_spacy/src_europarl_prediction_10000.txt
Executing command: java -jar src/meteor-1.5/meteor-1.5.jar data/clean/run1_model_lemma_spacy/src_europarl_prediction_10000.txt data/clean/lematised_spacy/lemme_Europarl_test_500.tok.true.clean.fr -l fr -norm
METEOR Score: Final score: 0.4919332113961734
(taa) *[main] [/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT]$ python3 src/meteor-score.py data/clean/lematised_spacy/lemme_Europarl_test_500.tok.true.clean.fr data/clean/run2_model_lemma_spacy/src_europarl_prediction_10000.txt
Executing command: java -jar src/meteor-1.5/meteor-1.5.jar data/clean/run2_model_lemma_spacy/src_europarl_prediction_10000.txt data/clean/lematised_spacy/lemme_Europarl_test_500.tok.true.clean.fr -l fr -norm
METEOR Score: Final score: 0.5015041858867812
(taa) *[main] [/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT]$ python3 src/meteor-score.py data/clean/lematised_spacy/lemme_Emea_test_500.tok.true.clean.fr data/clean/run1_model_lemma_spacy/src_emea_prediction_10000.txt
Executing command: java -jar src/meteor-1.5/meteor-1.5.jar data/clean/run1_model_lemma_spacy/src_emea_prediction_10000.txt data/clean/lematised_spacy/lemme_Emea_test_500.tok.true.clean.fr -l fr -norm
METEOR Score: Final score: 0.2574120187051249
(taa) *[main] [/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT]$ python3 src/meteor-score.py data/clean/lematised_spacy/lemme_Emea_test_500.tok.true.clean.fr data/clean/run2_model_lemma_spacy/src_emea_prediction_10000.txt
Executing command: java -jar src/meteor-1.5/meteor-1.5.jar data/clean/run2_model_lemma_spacy/src_emea_prediction_10000.txt data/clean/lematised_spacy/lemme_Emea_test_500.tok.true.clean.fr -l fr -norm
METEOR Score: Final score: 0.36772656464785966
```

Transformers :

fléchies :

BLEU

```
(taa) *[main]~/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT$ python3 src/compute-bleu.py data/clean_transformers/Europarl_en_fr_clean/Europarl_test_500.tok
.true.clean.fr data/clean_transformers/run1_model/src_europarl_pred_5000.txt
Reference 1st sentence: ces règles n&apos; ont pas leur place dans un accord de libre-échange .
MTed 1st sentence: ces règles n&apos; ont pas sa place dans un accord de libre échange .
BLEU: 32.13390576751825

(taa) *[main]~/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT$ python3 src/compute-bleu.py data/clean_transformers/Europarl_en_fr_clean/Europarl_test_500.tok
.true.clean.fr data/clean_transformers/run2_model/src_europarl_pred_5000.txt
Reference 1st sentence: ces règles n&apos; ont pas leur place dans un accord de libre-échange .
MTed 1st sentence: ces règles n&apos; ont aucune place dans un accord de libre-échange .
BLEU: 33.32174586488234

(taa) *[main]~/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT$ python3 src/compute-bleu.py data/clean_transformers/Emea_en_fr_clean/Emea_test_500.tok.true.cl
ean.fr data/clean_transformers/run1_model/src_emea_pred_5000.txt
Reference 1st sentence: ce médicament ne doit être utilisé que par des oncologues qualifiés ou d ' autres professionnels de santé spécialisés dans l ' administration des produits cytotoxiques .
MTed 1st sentence: son utilisation devrait être réservée aux <unk> de <unk> ou à d&apos; autres professionnels de la santé spécialisées dans l&apos; administration des agents <unk> .
BLEU: 7.814724566891177

(taa) *[main]~/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT$ python3 src/compute-bleu.py data/clean_transformers/Emea_en_fr_clean/Emea_test_500.tok.true.cl
ean.fr data/clean_transformers/run2_model/src_emea_pred_5000.txt
Reference 1st sentence: ce médicament ne doit être utilisé que par des oncologues qualifiés ou d ' autres professionnels de santé spécialisés dans l ' administration des produits cytotoxiques .
MTed 1st sentence: son utilisation devrait être confiné à la <unk> qualifié ou à d ' autres professionnels de la santé dans l ' administration .
BLEU: 32.35190208731724
```

METEOR

```
((base) *[main]~/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT$ conda activate taa
(taa) *[main]~/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT$ python3 src/meteor-score.py data/clean_transformers/Europarl_en_fr_clean/Europarl_test_500.tok
.true.clean.fr data/clean_transformers/run1_model/src_europarl_pred_5000.txt
Executing command: java -jar src/meteor-1.5/meteor-1.5.jar data/clean_transformers/run1_model/src_europarl_pred_5000.txt data/clean_transformers/Europarl_en_
fr_clean/Europarl_test_500.tok.true.clean.fr -l fr -norm
METEOR Score: Final score: 0.5099817366083398
(taa) *[main]~/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT$ python3 src/meteor-score.py data/clean_transformers/Europarl_en_fr_clean/Europarl_test_500.tok
.true.clean.fr data/clean_transformers/run2_model/src_europarl_pred_5000.txt
Executing command: java -jar src/meteor-1.5/meteor-1.5.jar data/clean_transformers/run2_model/src_europarl_pred_5000.txt data/clean_transformers/Europarl_en_
fr_clean/Europarl_test_500.tok.true.clean.fr -l fr -norm
METEOR Score: Final score: 0.5225667218351999
(taa) *[main]~/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT$ python3 src/meteor-score.py data/clean_transformers/Emea_en_fr_clean/Emea_test_500.tok.true.cl
ean.fr data/clean_transformers/run1_model/src_emea_pred_5000.txt
Executing command: java -jar src/meteor-1.5/meteor-1.5.jar data/clean_transformers/run1_model/src_emea_pred_5000.txt data/clean_transformers/Emea_en_fr_clean
/Emea_test_500.tok.true.clean.fr -l fr -norm
METEOR Score: Final score: 0.24337418975586318
(taa) *[main]~/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT$ python3 src/meteor-score.py data/clean_transformers/Emea_en_fr_clean/Emea_test_500.tok.true.cl
ean.fr data/clean_transformers/run2_model/src_emea_pred_5000.txt
Executing command: java -jar src/meteor-1.5/meteor-1.5.jar data/clean_transformers/run2_model/src_emea_pred_5000.txt data/clean_transformers/Emea_en_fr_clean
/Emea_test_500.tok.true.clean.fr -l fr -norm
METEOR Score: Final score: 0.5072293911308532
```

BERT

```
((taa) *[main]~/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT$ python3 src/bert-score.py data/clean_transformers/Europarl_en_fr_clean/Europarl_test_500.tok.t
rue.clean.fr data/clean_transformers/run1_model/src_europarl_pred_5000.txt
calculating scores...
computing bert embedding.
100%|██████████| 16/16 [00:12<00:00, 1.29it/s]
computing greedy matching.
100%|██████████| 8/8 [00:00<00:00, 115.11it/s]
done in 12.46 seconds, 39.24 sentences/sec
Précision moyenne: 0.8292
Rappel moyen: 0.8215
F1 moyen: 0.8258
(taa) *[main]~/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT$ python3 src/bert-score.py data/clean_transformers/Europarl_en_fr_clean/Europarl_test_500.tok.t
rue.clean.fr data/clean_transformers/run2_model/src_europarl_pred_5000.txt
calculating scores...
computing bert embedding.
100%|██████████| 16/16 [00:12<00:00, 1.30it/s]
computing greedy matching.
100%|██████████| 8/8 [00:00<00:00, 113.45it/s]
done in 12.40 seconds, 39.43 sentences/sec
Précision moyenne: 0.8300
Rappel moyen: 0.8261
F1 moyen: 0.8277
(taa) *[main]~/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT$ python3 src/bert-score.py data/clean_transformers/Emea_en_fr_clean/Emea_test_500.tok.true.cle
an.fr data/clean_transformers/run1_model/src_emea_pred_5000.txt
calculating scores...
computing bert embedding.
100%|██████████| 14/14 [00:10<00:00, 1.38it/s]
computing greedy matching.
100%|██████████| 8/8 [00:00<00:00, 113.66it/s]
done in 10.24 seconds, 47.93 sentences/sec
Précision moyenne: 0.7049
Rappel moyen: 0.7081
F1 moyen: 0.7057
(taa) *[main]~/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT$ python3 src/bert-score.py data/clean_transformers/Emea_en_fr_clean/Emea_test_500.tok.true.cle
an.fr data/clean_transformers/run2_model/src_emea_pred_5000.txt
calculating scores...
computing bert embedding.
100%|██████████| 13/13 [00:09<00:00, 1.34it/s]
computing greedy matching.
100%|██████████| 8/8 [00:00<00:00, 111.39it/s]
done in 9.82 seconds, 50.00 sentences/sec
Précision moyenne: 0.8344
Rappel moyen: 0.8424
F1 moyen: 0.8375
```

Lemmes

BLEU

```
((taa) *[main]~/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT$ python3 src/compute-bleu.py data/clean_transformers/lematised/lemme_Europarl_test_500.tok.true.clean.fr data/clean_transformers/run1_model_lemma/src_europarl_pred_5000.txt
Référence 1st sentence: ces règle n & apes ; ont pas leur place dans un accord de libre-échange .
MTed 1st sentence: ces règle n & apes ; ont pas sa place dans un accord de libre-échange .
BLEU: 33.045689040090366

((taa) *[main]~/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT$ python3 src/compute-bleu.py data/clean_transformers/lematised/lemme_Europarl_test_500.tok.true.clean.fr data/clean_transformers/run2_model_lemma/src_europarl_pred_5000.txt
Référence 1st sentence: ces règle n & apes ; ont pas leur place dans un accord de libre-échange .
MTed 1st sentence: ces règle n & apes ; ont aucune place dans un accord de libre-échange .
BLEU: 32.98884234212623

((taa) *[main]~/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT$ python3 src/compute-bleu.py data/clean_transformers/lematised/lemme_Emea_test_500.tok.true.clean.fr data/clean_transformers/run1_model_lemma/src_emea_pred_5000.txt
Référence 1st sentence: ce médicament ne doit être utilisé que par des oncologues qualifiés ou d' autres professionnel de santé spécialisés dans l ' administration des produit cytotoxiques .
MTed 1st sentence: cette utilisation devrait se limiter aux <unk> professionnel de la santé ou d & apes ; autres secteur sanitaire dans l & apes ; administration des <unk> .
BLEU: 6.644983491643204

((taa) *[main]~/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT$ python3 src/compute-bleu.py data/clean_transformers/lematised/lemme_Emea_test_500.tok.true.clean.fr data/clean_transformers/run2_model_lemma/src_emea_pred_5000.txt
Référence 1st sentence: ce médicament ne doit être utilisé que par des oncologues qualifiés ou d' autres professionnel de santé spécialisés dans l ' administration des produit cytotoxiques .
MTed 1st sentence: son utilisation doit être limitée en matière de <unk> qualifiée ou d & apes ; autres professionnel de la santé dans le corps des agent <unk> .
BLEU: 33.51008207072152
```

```
((taa) *[main]~/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT$ python3 src/compute-bleu.py data/clean_transformers/lematised_spacy/lemme_Emea_test_500.tok.true.clean.fr data/clean_transformers/run1_model_lemma_spacy/src_emea_pred_5000.txt
Référence 1st sentence: ce médicament ne devoir être utiliser que par un oncologue qualifier ou d ' autre professionnel de santé spécialiser dans l ' administration de produit cytotoxique .
MTed 1st sentence: son utilisation devoir se limiter à <unk> et à autre professionnel de santé <unk> dans l'apres ; administration de agent <unk> .
BLEU: 8.284032285796304

((taa) *[main]~/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT$ python3 src/compute-bleu.py data/clean_transformers/lematised_spacy/lemme_Europarl_test_500.tok.true.clean.fr data/clean_transformers/run1_model_lemma_spacy/src_europarl_pred_5000.txt
Référence 1st sentence: ce règle n&apo ; avoir pas son place dans un accord de libre - échange .
MTed 1st sentence: ce règle n&apo ; avoir aucun place dans un accord de libre - échange .
BLEU: 35.017086841573494

((taa) *[main]~/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT$ python3 src/compute-bleu.py data/clean_transformers/lematised_spacy/lemme_Europarl_test_500.tok.true.clean.fr data/clean_transformers/run2_model_lemma_spacy/src_europarl_pred_5000.txt
Référence 1st sentence: ce règle n&apo ; avoir pas son place dans un accord de libre - échange .
MTed 1st sentence: ce règle n&apos ; avoir pas son place dans un accord de libre - échange .
BLEU: 35.29501197098429

((taa) *[main]~/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT$ python3 src/compute-bleu.py data/clean_transformers/lematised_spacy/lemme_Emea_test_500.tok.true.clean.fr data/clean_transformers/run1_model_lemma_spacy/src_emea_pred_5000.txt
Référence 1st sentence: ce médicament ne devoir être utiliser que par un oncologue qualifier ou d ' autre professionnel de santé spécialiser dans l ' administration de produit cytotoxique .
MTed 1st sentence: il falloir en tenir compte de <unk> qualifier ou d&apo ; autre professionnel de santé dans l'apos ; administration de agent <unk> .
BLEU: 36.871549136946136
```

METEOR

```
((taa) *[main]~/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT$ python3 src/meteor-score.py data/clean_transformers/lematised/lemme_Europarl_test_500.tok.true.clean.fr data/clean_transformers/run1_model_lemma/src_europarl_pred_5000.txt
Executing command: java -jar src/meteor-1.5/meteor-1.5.jar data/clean_transformers/run1_model_lemma/src_europarl_pred_5000.txt data/clean_transformers/lematised/lemme_Europarl_test_500.tok.true.clean.fr -l fr -norm
METEOR Score: Final score: 0.521444621972628
((taa) *[main]~/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT$ python3 src/meteor-score.py data/clean_transformers/lematised/lemme_Europarl_test_500.tok.true.clean.fr data/clean_transformers/run2_model_lemma/src_europarl_pred_5000.txt
Executing command: java -jar src/meteor-1.5/meteor-1.5.jar data/clean_transformers/run2_model_lemma/src_europarl_pred_5000.txt data/clean_transformers/lematised/lemme_Europarl_test_500.tok.true.clean.fr -l fr -norm
METEOR Score: Final score: 0.5239335603922811
((taa) *[main]~/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT$ python3 src/meteor-score.py data/clean_transformers/lematised/lemme_Emea_test_500.tok.true.clean.fr data/clean_transformers/run1_model_lemma/src_emea_pred_5000.txt
Executing command: java -jar src/meteor-1.5/meteor-1.5.jar data/clean_transformers/run1_model_lemma/src_emea_pred_5000.txt data/clean_transformers/lematised/lemme_Emea_test_500.tok.true.clean.fr -l fr -norm
METEOR Score: Final score: 0.5002576505589411
((taa) *[main]~/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT$ python3 src/meteor-score.py data/clean_transformers/lematised/lemme_Emea_test_500.tok.true.clean.fr data/clean_transformers/run2_model_lemma/src_emea_pred_5000.txt
Executing command: java -jar src/meteor-1.5/meteor-1.5.jar data/clean_transformers/run2_model_lemma/src_emea_pred_5000.txt data/clean_transformers/lematised/lemme_Emea_test_500.tok.true.clean.fr -l fr -norm
METEOR Score: Final score: 0.22925324838179342
((taa) *[main]~/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT$ python3 src/meteor-score.py data/clean_transformers/lematised_spacy/lemme_Europarl_test_500.tok.true.clean.fr data/clean_transformers/run1_model_lemma_spacy/src_europarl_pred_5000.txt
Executing command: java -jar src/meteor-1.5/meteor-1.5.jar data/clean_transformers/run1_model_lemma_spacy/src_europarl_pred_5000.txt data/clean_transformers/lematised_spacy/lemme_Europarl_test_500.tok.true.clean.fr -l fr -norm
METEOR Score: Final score: 0.5321837449746265
((taa) *[main]~/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT$ python3 src/meteor-score.py data/clean_transformers/lematised_spacy/lemme_Europarl_test_500.tok.true.clean.fr data/clean_transformers/run2_model_lemma_spacy/src_europarl_pred_5000.txt
Executing command: java -jar src/meteor-1.5/meteor-1.5.jar data/clean_transformers/run2_model_lemma_spacy/src_europarl_pred_5000.txt data/clean_transformers/lematised_spacy/lemme_Europarl_test_500.tok.true.clean.fr -l fr -norm
METEOR Score: Final score: 0.534726728667625
((taa) *[main]~/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT$ python3 src/meteor-score.py data/clean_transformers/lematised_spacy/lemme_Emea_test_500.tok.true.clean.fr data/clean_transformers/run1_model_lemma_spacy/src_emea_pred_5000.txt
Executing command: java -jar src/meteor-1.5/meteor-1.5.jar data/clean_transformers/run1_model_lemma_spacy/src_emea_pred_5000.txt data/clean_transformers/lematised_spacy/lemme_Emea_test_500.tok.true.clean.fr -l fr -norm
METEOR Score: Final score: 0.2666839443769637
((taa) *[main]~/Desktop/Master_S2/TA/Projet_OpenNMT/OpenNMT$ python3 src/meteor-score.py data/clean_transformers/lematised_spacy/lemme_Emea_test_500.tok.true.clean.fr data/clean_transformers/run2_model_lemma_spacy/src_emea_pred_5000.txt
Executing command: java -jar src/meteor-1.5/meteor-1.5.jar data/clean_transformers/run2_model_lemma_spacy/src_emea_pred_5000.txt data/clean_transformers/lematised_spacy/lemme_Emea_test_500.tok.true.clean.fr -l fr -norm
METEOR Score: Final score: 0.5262489641454393
```

BERT

```
(taa) *[main]~/Desktop/Master_S2/Projet_OpenNMT/OpenNMT$ python3 src/bert-score.py data/clean_transformers/lematised/lemme_Europarl_test_500.tok.true.clean.fr data/clean_transformers/run1_model_lemma/src_europarl_pred_5000.txt
calculating scores...
computing bert embedding.
100%|██████████| 16/16 [00:11<00:00, 1.34it/s]
computing greedy matching.
100%|██████████| 8/8 [00:00<00:00, 111.89it/s]
done in 12.08 seconds, 40.49 sentences/sec
Précision moyenne: 0.8102
Rappel moyen: 0.8149
F1 moyen: 0.8121
(taa) *[main]~/Desktop/Master_S2/Projet_OpenNMT/OpenNMT$ python3 src/bert-score.py data/clean_transformers/lematised/lemme_Europarl_test_500.tok.true.clean.fr data/clean_transformers/run2_model_lemma/src_europarl_pred_5000.txt
calculating scores...
computing bert embedding.
100%|██████████| 16/16 [00:11<00:00, 1.37it/s]
computing greedy matching.
100%|██████████| 8/8 [00:00<00:00, 126.52it/s]
done in 11.77 seconds, 41.54 sentences/sec
Précision moyenne: 0.8119
Rappel moyen: 0.8166
F1 moyen: 0.8139
(taa) *[main]~/Desktop/Master_S2/Projet_OpenNMT/OpenNMT$ python3 src/bert-score.py data/clean_transformers/lematised/lemme_Emea_test_500.tok.true.clean.fr data/clean_transformers/run1_model_lemma/src_emea_pred_5000.txt
calculating scores...
computing bert embedding.
100%|██████████| 13/13 [00:10<00:00, 1.20it/s]
computing greedy matching.
0%|          | 0/8 [00:00<?, ?it/s]
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
100%|██████████| 8/8 [00:00<00:00, 102.09it/s]
done in 10.95 seconds, 44.85 sentences/sec
Précision moyenne: 0.6806
Rappel moyen: 0.6889
F1 moyen: 0.6837
(taa) *[main]~/Desktop/Master_S2/Projet_OpenNMT/OpenNMT$ python3 src/bert-score.py data/clean_transformers/lematised/lemme_Emea_test_500.tok.true.clean.fr data/clean_transformers/run2_model_lemma/src_emea_pred_5000.txt
calculating scores...
computing bert embedding.
100%|██████████| 13/13 [00:08<00:00, 1.51it/s]
computing greedy matching.
100%|██████████| 8/8 [00:00<00:00, 133.89it/s]
done in 8.69 seconds, 56.52 sentences/sec
Précision moyenne: 0.8384
Rappel moyen: 0.8374
F1 moyen: 0.8370

(taa) *[main]~/Desktop/Master_S2/Projet_OpenNMT/OpenNMT$ python3 src/bert-score.py data/clean_transformers/lematised/lemme_Europarl_test_500.tok.true.clean.fr data/clean_transformers/run1_model_lemma/src_europarl_pred_5000.txt
calculating scores...
computing bert embedding.
100%|██████████| 16/16 [00:11<00:00, 1.34it/s]
computing greedy matching.
100%|██████████| 8/8 [00:00<00:00, 111.89it/s]
done in 12.08 seconds, 40.49 sentences/sec
Précision moyenne: 0.8102
Rappel moyen: 0.8149
F1 moyen: 0.8121
(taa) *[main]~/Desktop/Master_S2/Projet_OpenNMT/OpenNMT$ python3 src/bert-score.py data/clean_transformers/lematised/lemme_Europarl_test_500.tok.true.clean.fr data/clean_transformers/run2_model_lemma/src_europarl_pred_5000.txt
calculating scores...
computing bert embedding.
100%|██████████| 16/16 [00:11<00:00, 1.37it/s]
computing greedy matching.
100%|██████████| 8/8 [00:00<00:00, 126.52it/s]
done in 11.77 seconds, 41.54 sentences/sec
Précision moyenne: 0.8119
Rappel moyen: 0.8166
F1 moyen: 0.8139
(taa) *[main]~/Desktop/Master_S2/Projet_OpenNMT/OpenNMT$ python3 src/bert-score.py data/clean_transformers/lematised/lemme_Emea_test_500.tok.true.clean.fr data/clean_transformers/run1_model_lemma/src_emea_pred_5000.txt
calculating scores...
computing bert embedding.
100%|██████████| 13/13 [00:10<00:00, 1.20it/s]
computing greedy matching.
0%|          | 0/8 [00:00<?, ?it/s]
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
100%|██████████| 8/8 [00:00<00:00, 102.09it/s]
done in 10.95 seconds, 44.85 sentences/sec
Précision moyenne: 0.6806
Rappel moyen: 0.6889
F1 moyen: 0.6837
(taa) *[main]~/Desktop/Master_S2/Projet_OpenNMT/OpenNMT$ python3 src/bert-score.py data/clean_transformers/lematised/lemme_Emea_test_500.tok.true.clean.fr data/clean_transformers/run2_model_lemma/src_emea_pred_5000.txt
calculating scores...
computing bert embedding.
100%|██████████| 13/13 [00:08<00:00, 1.51it/s]
computing greedy matching.
100%|██████████| 8/8 [00:00<00:00, 133.89it/s]
done in 8.69 seconds, 56.52 sentences/sec
Précision moyenne: 0.8384
Rappel moyen: 0.8374
F1 moyen: 0.8370
```

Bibliographie

Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002). *BLEU: a method for automatic evaluation of machine translation*. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics* (pp. 311-318).

Banerjee, S., & Lavie, A. (2005). *METEOR: An automatic metric for MT evaluation with improved correlation with human judgments*. In *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization* (pp. 65-72).

Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., & Artzi, Y. (2020). *BERTScore: Evaluating Text Generation with BERT*. In *International Conference on Learning Representations (ICLR)*.

Guillaume Klein, Yoon Kim, Yuntian Deng, Jean Senellart, and Alexander Rush. 2017. OpenNMT: Open-Source Toolkit for Neural Machine Translation. In *Proceedings of ACL 2017, System Demonstrations*, pages 67–72, Vancouver, Canada. Association for Computational Linguistics.