



Aix-Marseille Université,
Campus de Luminy,
13009 MARSEILLE



Année universitaire 2014–2015

BLEU:Méthode d'évaluation pour la translation automatique

Xiaoze JI, Yuxing TIAN

Mémoire de TER - M1 Informatique 2015
Master 1 Informatique

Encadré par Carlos Ramisch
Aix Marseille Université

Version du 28/mai/2015

La sommaire

Introduction	1
1. Motivation: traduction automatique d'expressions verbales anglais → français	1
2. Contexte du travail	1
3. Notre travail	1
Description détaillée du sujet	2
1. TED, un corpus pour étudier le problème	2
2. BLEU-méthode d'évaluation la qualité de MT(Machine Translation)	2
La description du travail réalisé	4
1. Travaux préparatoires	4
2. BLEU-méthode d'évaluation la qualité de MT(Machine Translation)	4
2.1) BLEU	4
2.2) Adaptation aux expressions	6
3. Les difficultés, les résolutions et le Test	7
3.1) le problème avec le N-grams	7
3.2) le problème avec le 'focus'	7
3.3) Utilisation était compliquée	8
3.4) L'ordre d'option sur la ligne de commande	8
4. Test fonction sur la ligne de commande	8
4.1 calculer N-grams sur commande Line et interface	8
4.2 calculer N-grams Ignorer majuscules ou en minuscules	8
4.3 calculer ajouter focus et poids=0.2	8
4.4 calculer bleu.....	9
4.5 Résultat N-gram=6.....	9
4.6 calculer bleu.....	9
La conclusion	10
ANNEXE.....	11

Introduction

Aujourd'hui, avec le progrès de technologie d'internet, la distance humaine est diminuée, la communication interculturelle est devenue de plus en plus importante. On peut aller plus loin sans sortir notre domicile. On peut lire le document sans aller à la bibliothèque. On peut savoir la tendance économique sans demander les savants. L'Internet est une grande base de données qui porte plein d'information. Pourtant, il existe un très grand problème qui nous empêche d'apprendre les connaissances inter-culturelles. Ces sont les langues différentes. Il est impossible d'apprendre toutes les langues pour nous, car, pour l'instant, on a plusieurs de langues vivantes dont 141 langues sont les langues officielles[1]. Si l'on laisse le travail de traduction aux traducteurs, il est possible mais ça perd du temps, en plus, la traduction par humaine coûte trop cher et on ne peut l'utiliser qu'une seule fois. Du coup, la traduction automatique est élevée comme un thème très utile pour l'humain de monde entier. On fait juste un clic, et puis la traduction est affichée sur l'écran comme 'Google translation'. Avec le prix bas, la possibilité d'utiliser plusieurs fois, l'économie du temps, et l'augmentation d'efficacité, le sujet 'traduction' automatique est de plus en plus important.

Maintenant, on a de nouveau problème. Comment on peut faire un logiciel qui traduit en haute qualité? Comment peut-on évaluer la qualité de la traduction?

1. Motivation : traduction automatique d'expressions verbales[2] anglais → français

La traduction automatique d'expressions verbales est toujours un problème qui est difficile à résoudre. Il est très difficile d'avoir un bon résultat car il y aurait beaucoup de traductions possibles pour une même phrase. Il est possible d'afficher un mauvais résultat par exemple: 'give up = donner haut', ou 'take off = prendre de', 'se rendre compte = to give account', etc, donc, on a besoin de quelque chose faire une évaluation de traduction automatique.

Pour évaluer la qualité de traduction automatique, on utilise BLEU qui estime la similarité entre les corpus et les références en comparant leurs n-grammes [3].

La traduction d'expressions verbales ne fait pas une partie de notre travail, mais il est très important pour faire comprendre notre sujet, et en plus, ça peut nous aider d'améliorer notre travail.

2. Contexte du travail

De notre côté, le logiciel doit lire le texte source, ensuite il traduit le texte automatique en suivant les règles, puis il met le résultat dans un nouveau fichier et compare avec la ou les références pour vérifier la qualité de la traduction, enfin, la dernière étape est l'évaluation par humaine. Vu que la partie de traduction a été très bien développée dans la suite logicielle Moses, car il est open-source, Notre TER est concentré sur la partie de la vérification de traduction d'anglais vers français en utilisant la méthode 'BLEU' qui ont été développée.

3. Notre travail

Pour ce TER, on a :

Premièrement, on a écrit deux méthodes, une pour lire le fichier de MT, l'autre est pour lire le fichier de référence. Deuxièmement, on calcule 'Bleu' selon son expression. Troisièmement on a ajouté les options pour améliorer le 'BLEU'. Enfin, le résultat de notre programme est assez proche que celui de programme standard en Perl.

Description détaillée du sujet

1. TED, un corpus pour étudier le problème

Le TED est un ensemble de corpus qui contient de nombre de transcription de dialogue[4]. Ces corpus sont donnés en anglais et sont traduits par le monde entier. Ensuite, avant qu'on mette le corpus dans le pipeline de formation MT, nous devons effectuer la tokenization. Vue qu'on veut identifier les expressions verbales en anglais, on doit analyser le corpus anglais.

Après le pré-traitement, la machine va lire le corpus tokenisé automatique et traduire en créant un nouveau fichier dans lequel on aura les textes traduits. Et puis, la machine va comparer les textes traduits avec les références qui sont données par le monde entier. En fin, la machine va afficher le meilleur résultat.

On a utilisé les corpus de TED comme le source à tester pour vérifier si notre BLEU est correct.

2. BLEU-méthode d'évaluation la qualité de MT(Machine Translation) [5]

Vue que la machine peut nous donner un très mauvais résultat à cause de les expressions verbales comme on a dit avant, et la vérification manuel n'est pas possible s'il y a nombre de données, il faudrait un méthode qui estime la qualité de MT, c'est le Bleu.

BLEU(Bilingual Evaluation Understudy) est proposé par IBM, cette méthode pense que si la traduction de la machine est plus proche résultat de la traduction humaine, alors le plus élevé de sa qualité de la traduction.

BLEU exécutant est de calculer N-grams du candidat avec les N-grams de la traduction de référence et compter le nombre de matchs. Ces matchs sont position indépendante. Plus de matchs, la traduction du candidat est meilleure. Autrement dit, la phrase traduite qui correspond la plus à la phrase de référence va être élu. N est un paramètre changeable.

Par exemple :

Candidat MT: the cat stand on the mat .

Référence 1: The cat is on the mat .

$$p_n = \frac{\sum_{C \in (Candidates)} \sum_{n-gram \in C} count_{clip}(n-gram)}{\sum_{C' \in (Candidates)} \sum_{n-gram' \in C'} count_{clip}(n-gram')}$$

Référence 2: There is a cat on the mat .

(Matches n-grams = p_n)

Matches 1-grams = 6/7.

Matches 2-grams = 4/6.

Matches 3-grams = 1/5.

EXPLICATION : $Count_{clip} = \min(Count, Max_Ref_Count)$, c'est-à-dire, s'il y a un 'the' dans le candidat MT et il y en a 2 dans la référence, on dit 'the' apparaît une fois, C et C' sont un ensemble de mot de candidat, n-gramme est une séquence de n mots contigus qui sont corrects dans une phrase de candidat, n-gram' est une séquence de n mots contigus dans une phrase de candidat et p_n est la score de précision pour n-grammes.[6]

Sentence brevity penalty

Avec brevity penalty (BP) une traduction candidat doit maintenant correspondre les traductions de référence de longueur, le choix des mots, et l'ordre des mots. BP est un paramètre très important pour calculer BLEU.

Si l'on a 'c' comme total longueur de corpus de candidat, et 'r' comme le longueur de référence.

$$BP = \begin{cases} 1 & \text{si } c > r \\ e^{(1-r/c)} & \text{si } c \leq r \end{cases}$$

$$\text{Alors } BLEU = BP * \exp \left(\sum_{n=1}^N w_n \log p_n \right) \quad w_n = 1/N$$

Ici, c'est le noyau algorithme pour calculer BLEU.

Cette formule a été implémentée dans un script perl appelé multi-bleu.perl, disponible dans l'outil Moses. Notre travail consiste à ré-implémenter BLEU, puis l'étendre pour prendre en compte les expressions. Pour cela, nous voulons permettre à l'utilisateur de définir certains mots dans la phrase qui sont plus importants que les autres. Ces mots doivent être bien traduits. Pour cela, nous avons modifié BLEU pour concentrer les mots plus importants en complétant une fonction "focus".

En considérant l'habitude différente de compilation, notre programme est disponible pour l'interface et la ligne de commande. Pour les utilisateur d'interface, il est très facile à apprendre. Toute les fonctions sont visible et l'interface est simple. Pour les utilisateurs qui aiment de compiler avec la ligne de commande, il faudrait d'abord saisir le chemin de corpus et le chemin de référence. Ensuite, de saisir les commandes d'option est optionelle. Si l'utilisateur ne chisit rien, le système va afficher le résultat standard. Si l'on veut ajouter option, il est possible de mettre sa commande n'importe où. C'est-à-dire, les commandes comme '-l', '-n', '-f' soient devant le chemin de corpus et le chemin de référence, soient au milieu, soient derrière eux. Attention, derrière de '-n' et de '-d', il faudrait suivre une paramètre. Parce que ces paramètres sont obligé pour que le système sache quel N-gram à utiliser et quelle distance entre le mot élu et le mot non-élu. Sinon, le système ne calcule pas.

Par exemple :

```
j14025412@L-05402300F2-02:~/Traduction/src$ java Blue -focus focus.txt -d 0.2 sy
stem.txt reference frl.txt
Choose option: -l lowercase, -f add focus, -d add weight -n ngram -h help.
Executive order: option (optional), TraductionMachine.txt reference.txt
```

IMG1-0. Les options de BLEU avec la paramètre

Le système Moses est vraiment important pour la traduction automatique. Dans ce sujet on ne fait aucun méthode de traduction automatique, mais le corpus qu'on teste notre BLEU est vient de Moses. Ce système nous a donné un bon corpus et nous a aidé beaucoup pour comprendre et implémenter notre BLEU.

La description du travail réalisé

1.Travaux préparatoires

Au début, il faut qu'on sache la difficulté de la traduction d'anglais vers français. Pour cela, la connaissance linguistique est nécessaire. Les problèmes peuvent être les composés nominaux, les idoms et bien les phrasal verbs. Avec les travaux anciens, on a déjà pas mal de bon algorithme pour faire la traduction, du coup, pour ce TER, on va juste concentrer sur évaluation de la qualité de traduction. Autrement dit, on va programmer un algorithme pour vérifier si cette traduction des expressions est correctes.

Pour commencer notre travail, on a installé le Cygwin sur lequel on a installé Moses. Le Cygwin est un outil pour compiler notre travail avec la ligne de commande.

Le Moses est un système de traduction automatique statistique qui vous permet de former automatique des modèles de traduction pour toute paire de langue. Tout ce que vous avez besoin est un recueil de textes traduits (corpus parallèle). Une fois que vous avez un modèle de formation, un algorithme de recherche efficace trouve rapidement la traduction de probabilité la plus élevée parmi le nombre exponentiel de choix.[7]

Toutes les informations sont disponibles sur GitHub qui est un service web d'hébergement et de gestion de développement de logiciels, utilisant le logiciel de gestion de versions Git.

Tous ces logiciels doivent utiliser dans un environnement Unix, donc on a implémenté le cygwin qui est une collection de logiciels libres à l'origine développés par Cygnus Solutions permettant à différentes versions de Windows de Microsoft d'émuler un système Unix. Il vise principalement l'adaptation à Windows de logiciels qui fonctionnent sur des systèmes POSIX (tels que les systèmes [GNU/Linux](#), [BSD](#), et [Unix](#)). Cygwin simule un environnement Unix sous Windows, rendant possible l'exécution de ces logiciels après une simple compilation.[8]

En fin, on a choisi JAVA comme la langage de programmation pour implémenter le BLEU, car c'est un langage qu'on le sait bien. On a utilisé le MVC comme le patron de conception pour mieux construire notre projet. MVC est un modèle destiné à répondre aux besoins des applications interactives en séparant les problématiques liées aux différents composants au sein de leur architecture respective. Ce paradigme regroupe les fonctions nécessaires en trois catégories: un modèle de données; une vue (présentation, interface utilisateur); un contrôleur (logique de contrôle, gestion des événements, synchronisation).[9]

2.Travail réalisé

2.1) BLEU

Tout d'abord, il faudrait lire le document de Traduction en choisissant le chemin pour que le système sache la longueur de texte, ainsi que pour le document de référence. Pour le document de référence, il faudrait au moins un choix, sinon, le système affichera un erreur. Mais pour le document de Traduction, il n'a qu'un seul choix, parce qu'il est impossible d'évaluer plusieurs traductions en même temps. La traduction et les références sont écrites par humain ou par la machine et sont implémentées dans notre ordinateur. Dans ce travail, les références et les corpus sont donnés par notre professeur. Parce qu'ils sont des fichiers qui viennent de Moses, et on n'utilise pas Moses, donc il faudrait qu'il nous les donne comme l'entrée de notre programme. Ici, si l'on a ajouté trop de référence qui n'a pas besoin, on peut cliquer sur l'interface pour qu'on la supprime. On clique sur le bouton pour que le système calcule lors que la traduction et les références sont bien choisies.

Text Translation: rthnight\Traduction\test\td.phrase.mt.fr

Text Focus :

d : 0.2 ☐ Use Focus

D:\cygwin\home\unearth\night\test\td.ref.fr

Text Reference:

N-gram : 4 ☐ IgnorCase

N-gram	Count_Cor	Count_words
1gram	33653.0	57813.0
2gram	18479.0	55742.0
3gram	11370.0	53671.0
4gram	7076.0	51600.0

MT_length : 57813 RF_length : 57994

bp : 0.9968741120590758

Blue : 0.2727743894751179

Dans la partie de calcul, d'abord, on ajoute les documents de traduction et de référence. Ensuite, il faudrait avoir le lien entre le corpus et la référence. C'est à dire, le premier rang de traduction doit correspondre au premier rang de référence; le deuxième rang de traduction doit correspondre au deuxième rang de référence; pareil pour les suites. Ici, par exemple, on a 3 références pour un seul traduction. Donc, pour chaque phrase de traduction, il va falloir correspondre 3 fois avec toutes les 3 références. Et puis, vue que

$$\text{BLEU} = \text{BP} * \exp \left(\sum_{n=1}^N w_n \log p_n \right) \quad w_n = 1/N,$$

on a besoin de calculer les paramètres comme BP, N, et p_n , pour calculer le paramètre BP (**brevity penalty**), le système va choisir la phrase plus longue dans la référence comme le standard. Si ce standard est plus court que la phrase de traduction, alors le BP égale 1, sinon, le BP égale $e^{(1-r/lc)}$. Le

paramètre N dans $\left(\sum_{n=1}^N w_n \log p_n \right)$ est le n de n-gram, le méthode pour calculer le paramètre p_n , on a écrit dans la partie '**description détaillée du sujet**'. Quand on a tous ces paramètres, le BLEU est facile à calculer. Mais, maintenant, on a une question, pourquoi on doit calculer la somme de $\log p_n$? C'est parce que sinon, le BLEU n'est plus précis. Pour cela, on doit faire une somme de 1 à n de Matches N-grams afin de mettre le résultat plus autorisé. Si un mot de traduction, par exemple 'le', apparaît 3 fois dans la référence1, mais 2 fois dans la référence2, on comptera 3 fois.

Avant qu'on calcule, il faut saisir le n de N-grams, maximum 4. Parce que si l'on saisit supérieur que 4, le résultat de BLEU va être beaucoup diminué.

Quand tout est prêt, clique le bouton 'calculations' pour afficher le BLEU qui est le note de qualité de traduction, la longueur de Traduction, la longueur de référence, et le BP. Le système affiche une table supplémentaire qui contient les mots correspondants et les mots compté dans le N-grams actuel.

```
j14025412@L-05402300F2-02:~/Traduction/src$ java Blue system.txt reference_fr1.txt
Total_length_MT :7, Total_length_RF 1 : 9
1-gram : 5.0 :7.0
2-gram : 3.0 :6.0
3-gram : 2.0 :5.0
4-gram : 1.0 :4.0
BP :0.7514772930752859, Ratio :0.7777777777777778, Blue :0.3266828640925502
```

IMG1-1. Le BLEU original, sans option

2.2) Adaptation aux expressions

Pour améliorer ce système, on a ajouté 5 options :

La Première option ' -l ' comme 'lowercase'. Pour transformer toutes les caractères à lettres minuscules. En considérant pour le système les mots en majuscules sont différents que les mots en minuscules, on ajoute cette option. S'il y a 'conclusion' dans la traduction, et 'CONCLUSION' dans la référence, le système va penser que ces 2 mot ne se correspond pas, le BLEU va diminuer. En fait, cette traduction est vrai. Sur interface, on a un bouton 'IgnoreCase' qui a la même fonction.

```
j14025412@L-05402300F2-02:~/Traduction/src$ java Blue -l system.txt reference_fr1.txt
Total_length_MT :7, Total_length_RF 1 : 9
1-gram : 6.0 :7.0
2-gram : 4.0 :6.0
3-gram : 2.0 :5.0
4-gram : 1.0 :4.0
BP :0.7514772930752859, Ratio :0.7777777777777778, Blue :0.36741454942156665
j14025412@L-05402300F2-02:~/Traduction/src$ ^C
j14025412@L-05402300F2-02:~/Traduction/src$ ^C
j14025412@L-05402300F2-02:~/Traduction/src$ java Blue -focus focus.txt -d 0.2 system.txt reference_fr1.txt
Choose option: -l lowercase, -f add focus, -d add weight -n ngram -h help.
Executive order: option (optional), TraductionMachine.txt reference.txt
```

IMG1-2. Le BLEU avec option lowercase '-l'

La Deuxième option ' -n ' comme 'N-grams'. cette option est de laisser l'utilisateur saisir le **n** de N-grams, cette paramètre doit supérieur que 0. Mais on vous conseille de saisir les chiffres inférieure que 5, parce que sinon, on a risque de trouver une valeur de BLEU = 0 . Dans la ligne de commande, on tape "-n 3" pour dire 3-gram. Sur l'interface, on juste saisi la chiffre dans la colonne. Pour notre projet, la paramètre doit entre 1 et 4, sinon le système affichera un warning.

```
j14025412@L-05402300F2-02:~/Traduction/src$ java Blue -n 6 system.txt reference_fr1.txt
Total_length_MT :7, Total_length_RF 1 : 9
1-gram : 5.0 :7.0
2-gram : 3.0 :6.0
3-gram : 2.0 :5.0
4-gram : 1.0 :4.0
5-gram : 0.0 :3.0
6-gram : 0.0 :2.0
WARNING : BLUE is zero, because no correct X-gram was found. Try -n with a lower valuer.
BP :0.7514772930752859, Ratio :0.7777777777777778, Blue :0.0
```

IMG1-3. Le BLEU avec option de saisir 'N' '-n'

La Troisième option '-f' comme 'focus'. C'est pour concentrer le mot qu'on s'intéresse le plus. On choisit un ou plusieurs mots le plus important, si le système l'a bien traduit, on donne le BLEU plus haut parce que les autres mot sont pas très importants. On a fait un nouveau fichier qui indique la localisation de ce mot par la chiffre, donc il faut qu'on ajoute le chemin de ce fichier avant qu'on utilise la fonction '-f'.

La Quatrième option '-d ' comme 'distance'. Comme on a le mots concentré, il faudrait un méthode qui peut montrer que le mot élu est le plus important. Du coup, on va pondérer chaque mot par les chiffre entre 0 et 1. L'écart entre les 2 mots est la 'distance'. Le mot concentré pondère 1, les chiffres des mots de ses côtés diminuent en soustrayant 'distance'. Autrement dit, le mot plus loin de

mot élu, la pondération est plus petite. Si l'on a 2 mots élus, il y aura 2 pondérations pour un mot qui se trouve entre ces mots, on prendra la pondération la plus grande.

```
j14025412@L-05402300F2-02:~/Traduction/src$ java Blue -f focus.txt -d 0.2 system
.txt reference_fr1.txt
WARNING : Sentence 1 's focus, position 3 is out of sentence border. Ignoring it
Total_length_MT :7, Total_length_RF 1 : 9
1-gram : 3.9999999999999996 :5.4
2-gram : 3.0 :5.3999999999999995
3-gram : 2.0 :4.8
4-gram : 1.0 :4.0
BP :0.7514772930752859, Ratio :0.7777777777777778, Blue :0.3419373059461623
```

IMG1-4. Le BLEU avec option focus et distance'-f' '-d'

La Cinquième option '-h'. Cette option est pour afficher le manuel de ce système.

3. Les difficultés, les résolution et le Test

3.1) le problème avec le N-grams

Quand on vient de commencer notre travail, on s'est trompé la signification de n-grammes. On pensait que chaque fois quand le mot de corpus apparaît dans la référence doit être compté.

Par exemple, MT: le le

Référence: le le le
on calcule 6.

Pour la deuxième fois, on a bien révisé le document, donc cette fois-ci, le méthode de calculer le BLEU a été corrigé, mais il nous restait encore de problème--- notre BLEU donnait le mauvais résultat. IL était très différent que le résultat standard. Après relire le programme, le problème a été trouvé. C'était parce que chaque fois quand on veut calcule le BLEU, on a compté 2 fois la référence, donc le résultat était loin de standard.

Par exemple, MT: le le

Référence: le le le

il y a deux "le" dans MT, donc on supprime une , il calcule 3.

Pour la troisième fois, après avoir corrigé, le résultat était devenu bon. Le problème de bleu est résolu.

Par exemple, MT: le le

Référence1: le le le

Reference2 : le le le le

il y a deux "le" dans MT(Tvalue), quatre dans Référence(RTvalue), donc on calcule
le code :

```
if (RTotalNgram.containsKey(key)) {
    int RTvalue = Integer.parseInt(RTotalNgram.get(key) .toString());
    int Tvalue = Integer.parseInt(Tngram[i].get(key) .toString());
    if (Tvalue < RTvalue) {
        correct[i] += Tvalue;
    } else {
        correct[i] += RTvalue;
    }
}
```

3.2) le problème avec le 'focus'

Le 'focus' n'est pas facile à réaliser car il a le lien avec la fonction '-d'. Il faut ajouter les 2 fonction en même temps. Mais heureusement, c'était juste la compréhension de 'focus', la pondération et la distance doivent entre 0 et 1. Si l pondération est inférieur que 0, alors on la met 0.

On utilise la méthode matrice pour sauvegarder sa poids, double[] Focus, chaque N-gram de mots correspond sa poids

le code :

```
for (int j = 0; j < focus.length; j++) {
    if (i == 0) {
        double m = 1 - d * Math.abs(f - j);
        if (m < 0) {
            focus[j] = 0.0;
        } else {
            BigDecimal bg = new BigDecimal(m);
            focus[j] = bg.setScale(2, BigDecimal.ROUND_HALF_UP).doubleValue();
        }
    } else {
        double m = 1 - d * Math.abs(f - j);
        if (focus[j] < m) {
            BigDecimal bg = new BigDecimal(m);
            focus[j] = bg.setScale(2, BigDecimal.ROUND_HALF_UP).doubleValue();
        }
    }
}
```

3.3) Utilisation était compliqué

Au début, on a un bouton en bas de l'interface pour vider tous les champs de l'interface. Mais pour chaque nouvelle calcul, il faudrait vider les records. Sinon, les mots de référence et les mots correspondent allaient augmenter en base de l'ancienne donnée. Pour le corriger, on a supprimé le bouton 'clear' et rajoutait le méthode de 'clear' sous le méthode de 'Calculations'. C'est-à-dire, chaque fois qu'on calcule, les anciennes données sont supprimées, le système va calculer avec les nouvelles données.

3.4) L'ordre d'option sur la ligne de commande

Pour améliorer le système, l'ordre d'option est important aussi. Parce que ce système est pour tout le monde mais pas juste pour nous qui savons très bien le programme, du coup, **l'ordre d'option doit être libre s'il n'y a pas contrainte de l'un à l'autre. Ce n'est pas un problème très compliqué, ça nous fait 2 heures pour le résoudre.**

4. Test fonction sur la ligne de commande

4.1 calculer N-grams sur commande Line et interface:

MT: le pomme est sur le Table .

Référence: le pomme est sur un table de table .

Résultat:

```
Total_length_MT :7, Total_length_RF 1 : 9
1-gram : 5.0 :7.0
2-gram : 3.0 :6.0
3-gram : 2.0 :5.0
4-gram : 1.0 :4.0
c'est correct.
```

4.2 calculer N-grams Ignorer majuscules ou en minuscules

Résultat:

```
Total_length_MT :7, Total_length_RF 1 : 9
1-gram : 6.0 :7.0
2-gram : 4.0 :6.0
3-gram : 2.0 :5.0
4-gram : 1.0 :4.0
c'est correct.
```

4.3 calculer ajouter focus et poids=0.2:

focus: 2 4 15

Résultat:

WARNING : Sentence 1 's focus, position 3 is out of sentence border. Ignoring it
Total_length_MT :7, Total_length_RF 1 : 9
1-gram : 3.9999999999999996 :5.4
2-gram : 3.0 :5.3999999999999995
3-gram : 2.0 :4.8
4-gram : 1.0 :4.0
il y a un warning : 15 > total length MT, donc warning afficher le place d'erreur.

4.4 calculer bleu:

Résultat:

Total_length_MT :7, Total_length_RF 1 : 9
1-gram : 5.0 :7.0
2-gram : 3.0 :6.0
3-gram : 2.0 :5.0
4-gram : 1.0 :4.0
BP :0.7514772930752859, Ratio :0.7777777777777778, Blue :0.3266828640925502
c'est correct.

4.5 Résultat N-gram=6:

Total_length_MT :7, Total_length_RF 1 : 9
1-gram : 5.0 :7.0
2-gram : 3.0 :6.0
3-gram : 2.0 :5.0
4-gram : 1.0 :4.0
5-gram : 0.0 :3.0
6-gram : 0.0 :2.0
WARNING : BLUE is zero, because no correct X-gram was found. Try -n with a lower valuer.
BP :0.7514772930752859, Ratio :0.7777777777777778, Blue :0.0
il y a un warning: Bleu = 0, on doit réduire la valeur de N.

4.6 calculer bleu:

Source: ted.phrase.mt.fr

Référence: ted.ref.fr

Résultat:

Standard Total_length_MT :57921, Total_length_RF: 58293 BP :0.994, Ratio :0.994

Test Total_length_MT :57921, Total_length_RF 1 : 58293 BP :0.9935, Ratio :0.9936

	1-gram	2-gram	3-gram	4-gram	Bleu
Standard	0,5850	0,3330	0,3230	0,1380	0,2731
Test	0.5848	0.3333	0.2130	0.1378	0.2732
Strandard ignore maj et min	0,6140	0,3580	0,2320	0,1520	0,2948
Test ignore maj et min	0.6142	0.3582	0.2320	0.1524	0.2950

Table 1-1. comparer notre donnée avec la donnée standard

Les données 'standard' et 'Strandard ignore maj et min' sont sorties de fichier 'multi-bleu.perl', les données 'Test' et 'Test ignore maj et min' sont sorties de notre BLEU.

Test fonction sur i'interface (voir sur l'annexe)

La conclusion

Avant qu'on fasse ce TER, on pense que la traduction automatique est un sujet très abstrait et notre compétence est loin de le résoudre. Autrement dit, c'est un sujet pour les chefs de programmation. Pourtant, quand on se met à faire ce sujet, on trouve que c'est un sujet intéressant qui a besoin de données et de trouver un meilleur algorithme afin d'augmenter la qualité de traduction. Pour cela, il faudrait une bonne construction de programme de sorte que le système lise plus de référence et trouve un résultat qui corresponde le plus de la réponse standard, un grand nombre de données et un bon système qui peut bien valoir la qualité comme BLEU. En sachant les parties plus importantes, on va être bien orienté pour aller plus loin sur le sujet MT. Du coup, la traduction automatique n'est plus un sujet pour les chefs mais pour tous les informaticiens.

Dans ce TER, on a fait le BLEU qui est une simple et pratique de logiciels pour évaluer la qualité de MT (traduction automatique). Ensuite, nous avons ajouté des options, pour que notre BLEU soit mieux avec plus de fonction. Cette fois-ci, on fait tous les travaux ensemble parce que sinon on va prendre plus de temps pour relire et rattraper l'ancien travail, en plus, il est facile d'être confus et le conflit apparaîtra.

Pour le BLEU, il y a encore beaucoup d'insuffisances, on peut l'améliorer dans le futur:

1 On peut ajouter une fonction "Tokenisateur" qui peut séparer la phrase plus précise, puis notre calcul soit plus précis.

2 On peut ajouter d'autres types de documents pour la MT et pour la Référence, par exemple document xml, html.

Cette fois-ci, avec l'aide de M. **Carlos Ramisch**, nous avons beaucoup appris, nous avons bien profité de JavaEclipse et comprenons mieux le langage JAVA. En plus, on a su plus comment fonctionne-t-il le MT, on a vu son noyau, son problème et sa future. Si l'on avait plus de temps, on l'approfondirait sans aucun doute. Nous espérons que ce BLEU peut aider les autres qui s'intéressent à la traduction automatique.

Références:

[1] http://fr.wikipedia.org/wiki/Liste_des_langues_officielles

[2] The MWE types we are interested in Phrasal verbs, <http://aim-west.imag.fr/what-are-mwes/> .

[3] Carlos Ramisch and Laurent Besacier and Alexander Kobzar, 2013, How hard is it to automatically translate phrasal verbs from English to French ?, http://pageperso.lif.univ-mrs.fr/~carlos.ramisch/download_files/publications/2013/p01.pdf , p.01 -p.02

[4] Carlos Ramisch and Laurent Besacier and Alexander Kobzar, 2013, How hard is it to automatically translate phrasal verbs from English to French ? , http://pageperso.lif.univ-mrs.fr/~carlos.ramisch/download_files/publications/2013/p01.pdf , p.03 – p.04

[5] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu, July 2002, BLEU: a Method for Automatic Evaluation of Machine Translation , <http://www.aclweb.org/anthology/P02-1040.pdf> , p.03- p.05

[6] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu, July 2002, BLEU: a Method for Automatic Evaluation of Machine Translation , <http://www.aclweb.org/anthology/P02-1040.pdf> , p.02

[7] Système de traduction probabiliste Moses, <http://statmt.org/amos/>, ligne 2-3

[8] Cygwin <http://fr.wikipedia.org/wiki/Cygwin> ligne 1-3

[9] Modèle-vue-contrôleur, <http://fr.wikipedia.org/wiki/Mod%C3%A8le-vue-contr%C3%B4leur>, ligne 01-07

ANNEXE

Test fonction sur i'interface:

Blue

Text Translation:

025412/Traduction/src/system.txt

Upload

Text Focus :

j14025412/Traduction/src/focus.txt

Upload

d :

0.2

☐ Use Focus

Text Reference:

Upload

/filer/etudiants/g5/j14025412/Traduction/src/reference_fr1.txt

N-gram :

4

☐ IgnorCase

Calculation

N-gram	Count_Cor	Count_words
1gram	5.0	7.0
2gram	3.0	6.0
3gram	2.0	5.0
4gram	1.0	4.0

MT_length : 7

RF_length : 9

bp : 0.7514772930752859

Blue : 0.3266828640925502

IMG1-5. Le BLEU original sur interface,

Blue

Text Translation:

025412/Traduction/src/system.txt

Upload

Text Focus :

j14025412/Traduction/src/focus.txt

Upload

d :

0.2

☒ Use Focus

Text Reference:

Upload

/filer/etudiants/g5/j14025412/Traduction/src/reference_fr1.txt

N-gram :

6

☒ IgnorCase

Calculation

N-gram	Count_Cor	Count_words
1gram	4.6	5.4
2gram	3.6	5.3999999999999995
3gram	2.0	4.8
4gram	1.0	4.0

MT_length : 7

RF_length : 9

bp : 0.7514772930752859

Blue : 0.0

WARNING : Sentence 1 's focus, position 3 is out of sentence border. Ignoring it

WARNING : BLUE is zero, because no correct X-gram was found. Try -n with a lower valuer.

IMG1-6. Le BLEU paramètre N est supérieur que 4

Blue

Text Translation:

025412/Traduction/src/system.txt

Upload

Text Focus :

j14025412/Traduction/src/focus.txt

Upload

d :

0.2

☒ Use Focus

Text Reference:

Upload

/filer/etudiants/g5/j14025412/Traduction/src/reference_fr1.txt

N-gram :

4

☐ IngnorCase

Calculation

N-gram	Count_Cor	Count_words
1gram	3.9999999999999996	5.4
2gram	3.0	5.3999999999999995
3gram	2.0	4.8
4gram	1.0	4.0

MT_length : 7

RF_length : 9

bp : 0.7514772930752859

Blue : 0.3419373059461623

WARNING : Sentence 1 's focus, position 3 is out of sentence border. Ignoring it

IMG1-7. Le BLEU avec la paramètre faute de focus

Blue

Text Translation:

2/Traduction/src/ted.phrase.mt.fr

Upload

Text Focus :

j14025412/Traduction/src/focus.txt

Upload

d :

0.2

☐ Use Focus

Text Reference:

Upload

/filer/etudiants/g5/j14025412/Traduction/src/ted.ref.fr

N-gram :

4

☒ IgnorCase

Calculation

N-gram	Count_Cor	Count_words
1gram	35573.0	57921.0
2gram	20006.0	55850.0
3gram	12477.0	53779.0
4gram	7878.0	51708.0

MT_length : 57921

RF_length : 58293

bp : 0.9935980393863326

Blue : 0.29505639379939463

IMG1-8. Le BLEU en choisissant 'ingorcase' équivalent de '-l'

14

Blue

Text Translation:

2/Traduction/src/ted.phrase.mt.fr

Upload

Text Focus :

j14025412/Traduction/src/focus.txt

Upload

d :

0.2

☐ Use Focus

Text Reference:

Upload

/filer/etudiants/g5/j14025412/Traduction/src/ted.ref.fr

N-gram :

4

☐ IgnorCase

Calculation

N-gram	Count_Cor	Count_words
1gram	33873.0	57921.0
2gram	18619.0	55850.0
3gram	11455.0	53779.0
4gram	7125.0	51708.0

MT_length : 57921

RF_length : 58293

bp : 0.9935980393863326

Blue : 0.27327556261961383

IMG1-ç. Le BLEU sans choisir 'ingorcase'