



10.5281/zenodo.11095363

Vol. 07 Issue 04 April - 2024

Manuscript ID: #01338

Validité de la Technique d'alignement automatique en Traduction Assistée par Ordinateur (TAO): Cas de Wordfast Aligner

Wowe Egre Alexis*, Sakwe George Mbotake

Advanced School of Translators and Interpreters (ASTI), University of Buea, Cameroon.

Corresponding author: [wowe81\(at\)gmail.com](mailto:wowe81(at)gmail.com)

RÉSUMÉ

Les outils d'alignement automatique en traduction assistée par ordinateur (TAO) continuent de présenter des limites. Cet article a pour principal but d'évaluer la technique de l'alignement automatique de Wordfast Aligner. Cette étude vise à identifier les lacunes dans les segments du TC au moyen d'une description comparée avec les segments du TS, à évaluer le pourcentage du rendu en nous servant de l'évaluation automatique de Wordfast Aligner, à identifier et évaluer l'algorithme utilisé dans Wordfast Aligner dans la segmentation, et à étudier la qualité de l'alignement produit par Wordfast Aligner et voir si cela est valable dans un cadre général. Quatre hypothèses sont émises dans cette recherche. La première stipule que les lacunes existant dans les segments du TC dans un bi-texte produit automatiquement dans Wordfast Aligner sont identifiables suivant une analogie cognitive. La deuxième postule que l'outil Wordfast Aligner rend tout le sens des segments du TS dans les segments du TC. La troisième stipule que l'algorithme d'alignement automatique utilisé dans Wordfast Aligner est valide. Et la quatrième postule que la technique de l'alignement automatique dans Wordfast Aligner ne bâtit pas une MT optimale. Il ressort de l'analyse de 50 segments pertinents que les lacunes se trouvant dans les segments du TC en rapport avec les segments du TS sont de quatre (4) catégories : des lacunes provenant d'une mauvaise segmentation (2%), des lacunes provenant d'une sur-traduction (16%), des lacunes provenant d'une sous-traduction (64%), des segments avec zéro (0) lacune (18%). Ces résultats montrent également que l'outil Wordfast Aligner, en effectuant l'alignement automatique, ne rend pas toujours tout le sens du segment du texte source. La moyenne générale du rendu par Wordfast Aligner est de 69,68% \approx 70%, ce qui est acceptable, mais pas optimale.

Mots Clés : Ordinateur, Traduction, Traduction Assistée par Ordinateur, Alignement automatique.

ABSTRACT

Machine alignment tools in computer-assisted translation (CAT) continue to have limitations. The main purpose of this paper is to evaluate the Wordfast Aligner machine alignment technique. This study aims to identify gaps in TT segments by means of a comparative description with ST segments, to evaluate the percentage of rendering using Wordfast Aligner's automatic evaluation, to identify and evaluate the algorithm used in Wordfast Aligner in the segmentation, and to study the quality of the alignment produced by Wordfast Aligner and see if this is valid in a general setting. Four hypotheses are put forward in this research. The first states that gaps in the TT segments in a bi-text automatically produced in Wordfast Aligner are identifiable by a cognitive analogy. The second postulates that the Wordfast Aligner tool renders all the meaning of the ST segments into the TT segments. The third hypothesis states that the automatic alignment algorithm used in Wordfast Aligner is valid. And the fourth hypothesis postulates that the automatic alignment technique in Wordfast Aligner does not build an optimal TM. The analysis of 50 relevant segments shows that the gaps found in the TT segments in relation to the ST segments fall into four (4) categories: gaps due to poor segmentation (2%), gaps due to over-translation (16%), gaps due to under-translation (64%), segments with zero (0) gaps (18%). These results also show that the Wordfast Aligner tool, when performing automatic alignment, does not always render the full meaning of the source text segment. The overall average rendering by Wordfast Aligner is 69.68% \approx 70%, which is acceptable, but not optimal.

Keywords: Computer, Translation, Computer-Aided Translation, Automatic Alignment.



This work is licensed under Creative Commons Attribution 4.0 License.

1. Introduction et contexte

Les traducteurs professionnels évoluent actuellement dans un contexte particulier. En effet, la mondialisation a fait naître des besoins croissants en matière de traduction et de rédaction technique. Un grand nombre d'entreprises ont également des besoins croissants en termes de localisation et de terminologie. Elles doivent faire la mise à jour simultanée de leurs produits ou de leurs documentations, et ce, en plusieurs langues. Pour ce faire, il s'avère indispensable de répondre à une exigence bien précise : celle de traduire le plus grand nombre de mots possibles, tout en respectant des délais très courts. Qui plus est, la qualité des travaux rendus doit être irréprochable, ce qui implique notamment de pouvoir s'adapter à n'importe quel domaine de compétences et être capable de rechercher ainsi que d'employer des termes hautement spécialisés.

De nos jours, les logiciels de traduction assistée par ordinateur (TAO) sont florissants dans le secteur de la traduction professionnelle. Le principal avantage de l'utilisation des outils de TAO est évidemment le gain en temps et en efficacité. Des traducteurs professionnels soulignent l'importance de cette technologie car elle garantit une cohérence soutenue de la terminologie et permettent la diminution des risques d'erreurs éventuelles. Ces outils réduisent aussi considérablement le travail de saisie grâce aux suggestions grammaticales et orthographiques pré-traduites et contenues dans des Mémoires de Traduction (MT). Il convient ici de souligner que les MT sont essentiellement basées sur la technique d'alignement. Et il en existe principalement deux types d'alignement à savoir ; l'alignement manuel, c'est-à-dire le traducteur introduit le texte aligné dans le logiciel segment après segment, et l'alignement automatique où le traducteur introduit simplement l'entièreté du TS et du TC dans le logiciel et ce dernier effectue automatiquement l'alignement. Ce dernier type d'alignement permet de gagner en temps. Il existe une pléthore de logiciels d'alignement automatiques tels que WinAlign de Trados (payant), Wordfast Aligner (libre).

Cet article donc à évaluer l'alignement automatique fait par *Wordfast Aligner* qui est un outil plus accessible.

2. Énoncé du problème de recherche

Dans un contexte de mondialisation, nombre d'entreprises ont besoin de mettre à jour et de localiser leur documentation. Mais, face à cette demande grandissante de la traduction, l'offre s'est également démultipliée : les traducteurs, de plus en plus nombreux, proposent des tarifs de plus en plus compétitifs. D'après les éditeurs de logiciels de traduction assistée par ordinateur (TAO), la solution réside dans l'utilisation d'outils de traduction assistée par ordinateur, tels que les Mémoires de Traduction utilisant la technique d'alignement, afin d'accroître la qualité et les volumes traités et de diminuer les coûts et les délais. Mais il ne pourrait s'agir là que d'un argument commercial. Qu'en est-il vraiment ? L'efficacité de cette technique est-elle avérée, surtout lorsqu'il s'agit d'un texte de type littéraire ?

3. Questions de recherches et objectifs visés

Cette étude est le résultat de quelques questions et des objectifs à atteindre

3.1 Questions de recherches

Cette recherche a fondamentalement été orientée par quatre questions de recherche :

1. Comment peut-on identifier des lacunes dans les segments du texte cible en Alignement automatique dans Wordfast Aligner?
2. L'outil Wordfast Aligner rend-il tout le sens des segments du TS dans les segments du TC ?
3. Quel est la validité de l'algorithme d'alignement utilisé dans Wordfast Aligner ?
4. La qualité de l'alignement produit par Wordfast Aligner est-elle optimale ?

3.2 Objectifs visés

À la suite de la formulation de ces questions de recherche, cet article vise un certain nombre d'objectifs, à savoir :

1. Identifier les lacunes dans les segments du TC au moyen d'une description comparée avec les segments du TS ;

2. Evaluer le pourcentage du rendu en nous servant de l'évaluation automatique de Wordfast Aligner ;
3. Identifier et évaluer l'algorithme utilisé par Wordfast Aligner dans la segmentation.
4. Etudier la qualité de l'alignement produit par Wordfast Aligner, au moyen d'une description comparative du TS avec le TC, et voir si cela est valable dans un cadre général.

4. Méthodologie

Cette recherche étant qualitative, nous nous servons essentiellement de la documentation et d'une observation comparative des bi-textes dans Wordfast Aligner. Nous nous également du module d'évaluation automatique du logiciel Wordfast Aligner.

Les données sont collectées à partir de la documentation existante, de l'étude de la technique d'alignement. La méthode d'alignement est « the process of comparing a source text and its translation, matching the corresponding segments, and binding them together as translation units in a Translation Memory. » (Bowker, 2002 :109). Donc, lesdites données ont été collectées en observant les segments du TC comparativement avec leurs correspondants du TS. C'est dire en même temps qu'en tant que chercheur, nous avons été l'instrument principal de collecte de ces données.

Dans cet article, nous allons procéder à une description et à une analyse des données. Notre étude se veut à la fois quantitative et qualitative. Pour ce qui est des segments de traduction, la description sera la principale méthode d'étude. Ainsi nous ferons une description comparative des segments du TC avec ceux du TS. Nous nous servons également de l'évaluation automatique de chaque segment aligné.

La phase d'analyse de notre travail, précédée de la phase descriptive, consistera en une étude critique de l'alignement produit par Wordfast Aligner, l'objectif étant de voir si cette technique rend de manière absolue tout le sens du segment source dans le segment cible, de déceler l'algorithme utilisé dans le logiciel et d'évaluer l'efficacité de cette méthode dans le processus de l'alignement. À cet effet, nous allons nous servir d'une grille d'analyse comprenant neuf entrées comme le présente le Tableau 1 ci-dessous :

a	Texte Source	Il s'agit de l'extrait qui fait l'objet de l'analyse
b	Texte Cible aligné	Il s'agit du segment traduction aligné par le logiciel
c	Comparaison du TS avec le TC	Il s'agit là des unités du TS ne se trouvant pas dans le TC et vice versa
d	Pourcentage du rendu	Pourcentage du rendu
e	Lacunes constatées	Pourcentage du manquant
f	Justification du pourcentage du rendu	Ce qui justifie le pourcentage du rendu
g	Méthode de Wordfast Aligner	Elle présente la théorie littéraire, la théorie de traduction, la stratégie de traduction et la technique d'alignement utilisées par logiciel
h	Traduction proposée	Lorsque la traduction n'est pas adéquate, nous proposons des pistes d'amélioration
i	Méthode du chercheur	Elle présente la théorie littéraire, la théorie de traduction, la stratégie de traduction et la technique d'alignement utilisées par le chercheur
j	Justification de la méthode du chercheur	Il s'agit de la justification de la méthode du chercheur.

Tableau 4.1 : Grille d'analyse des données

La présente étude s'appuie sur les principes de la Descriptive Translation Studies tels qu'énoncés par Wallmach et Kruger (1997) et sur la critique des traductions de Berman (1995). L'analyse des segments du rendu est faite sur la base de la théorie algorithmique de l'information initiée par Kolmogorov, Solomonov et Chaitin dans les

années 1960, qui vise à quantifier et à qualifier le contenu en information d'un ensemble de données, en utilisant la théorie de la calculabilité et la notion universelle de Machine de Turing.

5. Présentation et analyse des données

Le corpus utilisé dans cette étude est tiré du livre *Things Fall Apart* de Chinua Achebe et de sa traduction *Tout s'effondre* par Pierre Girard. L'outil d'alignement automatique utilisé est *Wordfast Aligner*.

5.1 Présentation de Wordfast Aligner

Wordfast Aligner est un outil d'alignement automatique en ligne. Il constitue un module du logiciel libre Wordfast très sollicité par bon nombre des professionnels de traduction. Son interface graphique est simple telle que présentée sur la Figure 1 ci-dessous :

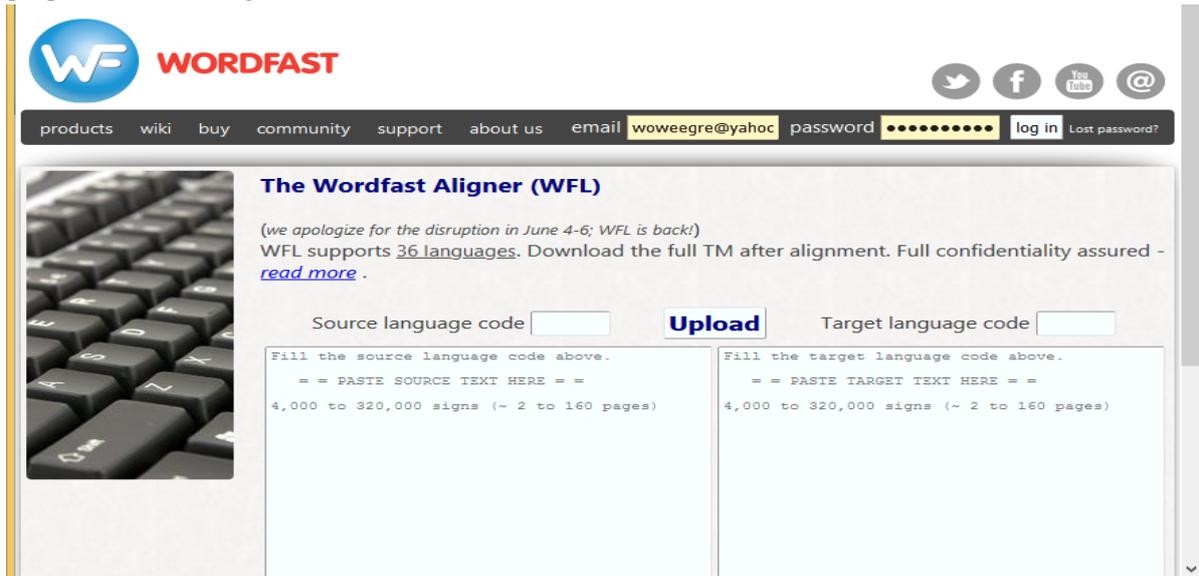


Figure 5.1 : Interface graphique de Wordfast Aligner

Pour pouvoir effectuer un alignement automatique dans Wordfast Aligner, il faut tout d'abord mettre le code des langues du TS et du TC. Ensuite, nous chargeons notre bi-texte dans les espaces correspondants, tel que le présente la Figure 2 ci-dessous :

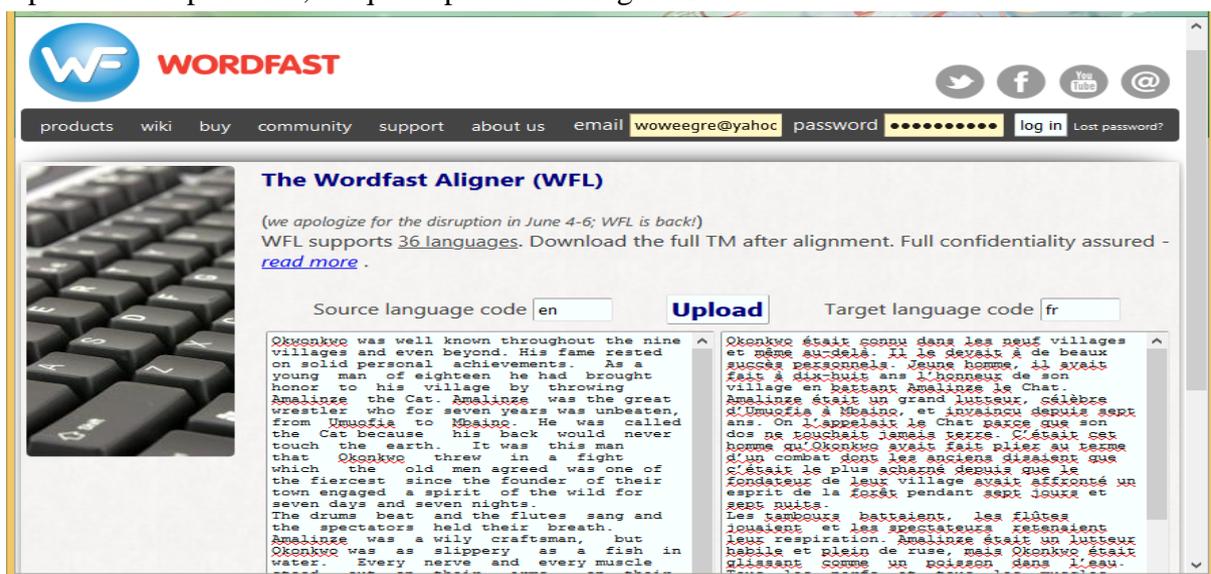


Figure 5.2 : Chargement du bi-texte dans Wordfast Aligner

Après avoir chargé le bi-texte, nous cliquons sur UPLOAD, et le logiciel effectue automatiquement l'alignement, tel que présenté sur la Figure 3 ci-dessous:

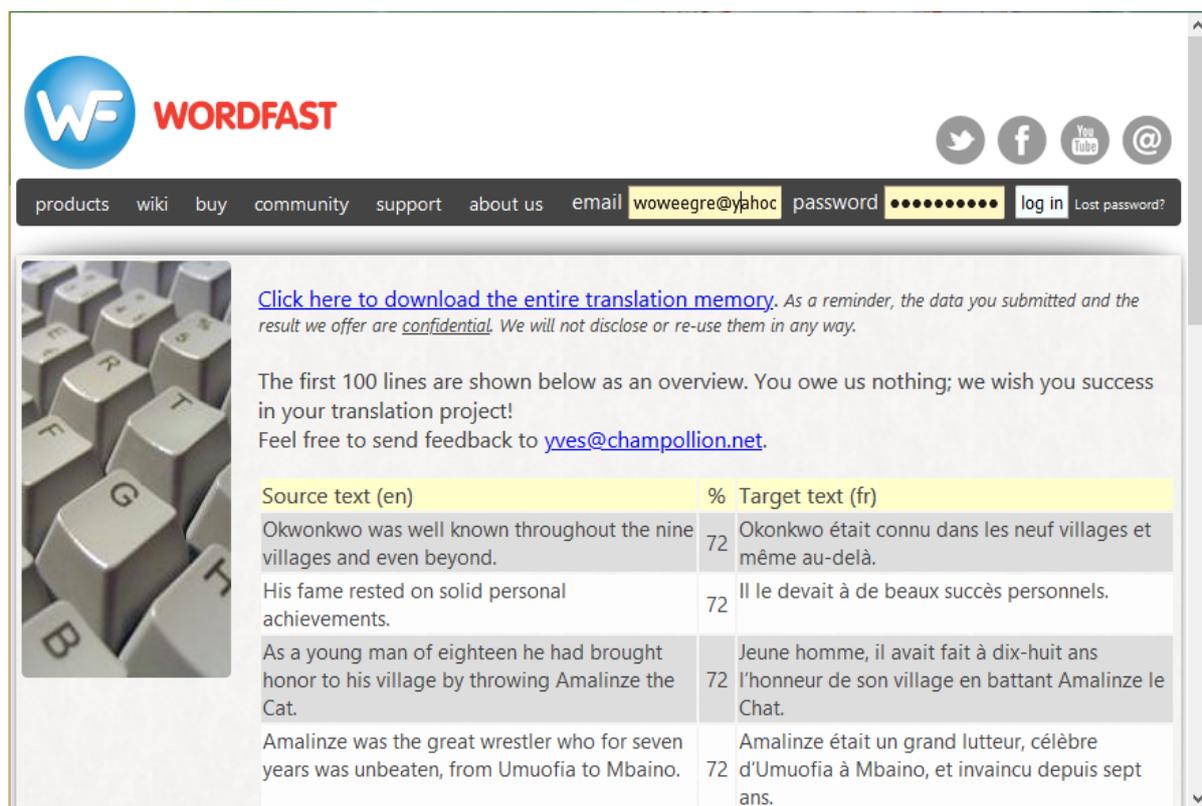


Figure 5.3 : Texte aligné dans Wordfast Aligner

6. Algorithme utilisé

Le mot « algorithme » vient du nom du mathématicien Al-Khwârizmî (latinisé au Moyen Âge en Algoritmi). Un algorithme est une suite ordonnée et finie d'instructions pour la résolution d'un problème précis. Toute technique d'alignement automatique procède tout d'abord à une segmentation.

La phrase peut être définie comme unité linguistique est un problème à part entière. Plusieurs niveaux s'entrecroisent sans permettre d'établir des caractérisations convergentes et consistantes : la prosodie, la ponctuation, la syntaxe, la sémantique et le niveau des propositions logiques.

D'un point de vue syntaxique, on peut difficilement s'en tenir à une définition canonique du type : Phrase = SN + SV. Tout d'abord, parce que cela n'inclut pas les syntagmes non verbaux (du type : « Bien sûr », ou, « Pourquoi pas ? » ou simplement « non »), très courants dans certains genres textuels (notamment les transcriptions de l'oral). Ensuite parce que ce genre de définition a du mal à rendre compte des phrases complexes possédant plusieurs noyaux verbaux, qu'il s'agisse de propositions coordonnées ou subordonnées (énumérations, listes, propositions enchâssées, etc.).

Ainsi, si l'on admet une définition extensive de la phrase basée sur le noyau verbal régissant, on peut aboutir à des phrases atteintes de gigantisme, formant un texte entier s'étalant sur plusieurs pages. Mais si l'on se base seulement sur certains indices typographiques, tels que les sauts à la ligne, n'importe quel syntagme, dûment énuméré, pourra revendiquer son statut de phrase.

Il en résulte que le plan sémantique est insuffisant, lui aussi, à caractériser l'unité de la phrase. Comme l'affirme Rastier (1994 : 115-116), les relations sémantiques ne cessent de transgresser les limites de la phrase. Ainsi « à l'autonomie syntaxique qui refléterait la complétude et l'autosuffisance de la prédication, on doit opposer les relations sémantiques qui rattachent toute phrase à son contexte linguistique et situationnel. Si bien que le découpage d'un texte en phrases n'est pas si simple et la recherche d'un point n'y suffit pas. »

Soient T et T' deux textes parallèles segmentés en phrases. On notera :

$$T = P_1 P_2 \dots P_n$$

$$T' = P'_1 P'_2 \dots P'_m$$

Soit A un alignement de T et T' . A est un sous-ensemble de $T \times T'$, c'est-à-dire un ensemble de k binômes :

$$A = \{(P_{i1}; P'_{j1}), (P_{i2}; P'_{j2}), \dots, (P_{ik}; P'_{jk})\} \text{ avec } 0 \leq i_1 \dots i_k \leq n \text{ et } 0 \leq j_1 \dots j_k \leq m \text{ respectant les conditions de quasi-monotonie et de quasi-bijection.}$$

Pour les phrases n'ayant pas de correspondant, suite à une insertion ou une omission, deux conventions sont possibles : soit on introduit deux phrases vides P_0 et P'_0 , avec lesquelles ces phrases seront alignées (et donc apparaîtront dans A), soit on ne les fait pas figurer dans A . On verra plus loin comment ces deux possibilités peuvent affecter le calcul des indices d'évaluation d'un alignement.

Dans les cas où une même phrase apparaît dans plusieurs de ces couples, on peut procéder à des regroupements et aboutir à une notation élargie.

Par exemple, pour $T = P_1 P_2 P_3 P_4 P_5 P_6$ et $T' = P'_1 P'_2 P'_3 P'_4 P'_5 P'_6 P'_7$,

l'alignement :

$$A = \{(P_1; P'_1), (P_1; P'_2), (P_2; P'_0), (P_3; P'_3), (P_4; P'_4), (P_5; P'_4), (P_6; P'_5), (P_6; P'_6), (P_6; P'_7)\}$$

s'écrit de manière condensée :

$$A = \{(P_1; P'_1 P'_2) (P_2); (P_3; P'_3) (P_4 P_5; P'_4) (P_6; P'_5 P'_6 P'_7)\} \text{ ou encore,}$$

$$A = \{(P_1; P'_{1..2}) (P_2); (P_3; P'_3) (P_{4..5}; P'_4) (P_6; P'_{5..7})\}$$

Dans le cas général, on notera le $i^{\text{ème}}$ binôme : $B_i = (P_{n_i} \dots m_i; P'_{p_i} \dots q_i)$.

Les binômes de la première notation, sous forme de couple, sont en fait des binômes fragmentaires : la présence du couple $(P_5; P'_4)$ dans cette notation n'indique pas que P_5 est exactement aligné avec P'_4 , mais qu'ils font partie d'un binôme plus large $(P_4 P_5; P'_4)$, à l'intérieur duquel ils sont alignés.

Isabelle & Simard (1996 : 4), se sont intéressés à une catégorie d'alignement vérifiant les conditions suivantes :

– absence de correspondances croisées :

$$\forall (i, i', j, j') [(\{i, j\} \in A) \wedge (\{i', j'\} \in A) \wedge (i' < i) \Rightarrow (j' < j)]$$

– absence de correspondances à chevauchement partiel :

$$\forall (i, j, j') [(\{i, j\} \in A) \wedge (\{i, j'\} \in A) \Rightarrow \forall i' [(\{i', j\} \in A) \Rightarrow (\{i', j'\} \in A)]]$$

– absence de correspondances discontinues :

$$\forall (i, j, j') [(\{i, j\} \in A) \wedge (\{i, j'\} \in A) \Rightarrow \forall j'' [(\{j'', j\} \in A) \wedge (j'' > j) \Rightarrow (\{i, j''\} \in A)]]$$

Ces trois conditions résument et formalisent la propriété de *monotonie* introduite précédemment. Dans le cas d'un alignement monotone et complet, c'est-à-dire attribuant des correspondants (même vides) à toutes les phrases de chaque texte, on peut définir la notion de *chemin*.

Un chemin est défini comme une suite de *transitions* permettant de regrouper, de proche en proche, l'ensemble des phrases des deux textes en binômes d'alignement. Chaque transition n'est autre qu'un couple d'entiers positifs $(p:k)$ exprimant le regroupement de p phrases contiguës avec k phrases contiguës.

Par exemple, pour exprimer l'alignement de deux textes T et T' segmentés en phrases $P_1 P_2 P_3 P_4 P_5 P_6$ et $P'_1 P'_2 P'_3 P'_4 P'_5 P'_6 P'_7$, on peut recourir aux deux systèmes de notations :

– avec la notation ensembliste :

$$A = \{(P_1; P'_1 P'_2)(P_2;) (P_3; P'_3)(P_4 P_5; P'_4)(P_6; P'_5 P'_6 P'_7)\}$$

– avec la notation de type chemin :

$$A = ((1:2), (1:0), (1:1), (2:1), (1:3))$$

Notons qu'un chemin peut aussi être représenté comme suite de *points d'ancrage* si l'on retient les coordonnées de la première phrase de chaque groupement :

$$A = ((1,1), (2,3), (3,3), (4,4), (6,5))$$

C'est dans ce sens que nous parlerons, ultérieurement, des points constituant un chemin. Ce type de notation est similaire au format COAL développé par Isabelle & Simard (1996).

On peut noter l'alignement comme suit :

$$T = P_1 P_2 \qquad T' = P'_1 P'_2 P'_3 P'_4$$

$$A = \{(P_1; P'_1 P'_2), (\emptyset; P'_3), (P_2; P'_4)\}$$

ou bien, si l'on représente ces regroupements par une suite de transitions :

$$A = ((2:1), (0:1), (1:1))$$

7. Algorithme d'évaluation

Ici, nous mettons essentiellement de l'emphase sur les critères d'évaluation et le mode de calcul.

7.1 Critères d'évaluation

Avant de présenter les différentes méthodes, il nous faut examiner rapidement les mesures permettant d'évaluer quantitativement la qualité d'un alignement obtenu automatiquement. Le principe de ce type d'évaluation est de comparer l'alignement produit avec un alignement effectué « manuellement » (c'est-à-dire par l'humain), pris comme alignement de référence. Il est en effet nécessaire de se donner une base de comparaison : on ne peut évaluer un alignement que de manière relative, par rapport à l'alignement considéré comme référence.

Or il est souvent difficile d'établir ce que doit être l'alignement de référence. Les divergences qui apparaissent dans la traduction font qu'il est parfois difficile de choisir entre fusion et insertion.

Une fois l'alignement de référence établi, il est aisé de définir des mesures quantitatives permettant d'évaluer un alignement quelconque. Depuis les campagnes d'évaluation menées au sein du projet ARCADE (Véronis, 1997), destinées à fournir une base de comparaison solide pour évaluer des systèmes d'alignement de conceptions différentes, un certain consensus s'est établi : les mesures quantitatives généralement employées sont le *rappel* et la *précision*.

$$\text{Rappel} = \text{Nombre d'alignements corrects} / \text{nombre d'alignements de référence}$$

$$\text{Précision} = \text{Nombre d'alignements corrects} / \text{Nombre d'alignements proposés}$$

Le rappel indique la proportion de binômes de l'alignement de référence présents dans l'alignement évalué, et la précision indique la proportion de binômes corrects dans l'alignement évalué. Ces deux indicateurs sont les complémentaires du silence et du bruit :

$$\text{Silence} = 1 - \text{Rappel}$$

$$\text{Bruit} = 1 - \text{Précision}$$

Traduit dans notre système de notation, on obtient, dans l'évaluation de A par rapport à $A_{réf}$ (la notation d'un ensemble entre barres représente le cardinal):

$$\text{Précision}(A/Aréf) = \frac{|A \cap Aréf|}{|A|}$$

$$\text{Rappel}(A/Aréf) = \frac{|A \cap Aréf|}{|Aréf|}$$

En recherche d'information, rappel et précision évoluent souvent de manière antagoniste : plus une méthode obtient un bon rappel, plus la précision est faible. En ce qui concerne l'alignement, un bon rappel permet parfois d'améliorer la précision car il y a consolidation réciproque des binômes, du fait des hypothèses de quasi-bijection et monotonie : la donnée d'un binôme correct permet d'aligner plus sûrement les phrases qui suivent et qui précédent.

Enfin, pour tenir compte en même temps des deux mesures, nous reprendrons la *Fmesure* utilisée dans le projet ARCADE (reprise de van Rijsbergen, 1979), comme indice d'évaluation globale synthétisant rappel et précision :

$$F = \frac{2 \cdot (P \cdot R)}{(P + R)}$$

Cette mesure, qui représente la moyenne harmonique de *P* et *R*, n'est autre que le coefficient Dice appliqué à la comparaison de *A* et *A_{réf}* :

$$F = \text{Dice} = \frac{2 \cdot |A \cap Aréf|}{|A| + |Aréf|}$$

De par son caractère multiplicatif, elle défavorise, à moyenne arithmétique égale, les configurations où *P* et *R* sont éloignés :

si $P = 0,5$ $R = 0,5$ alors *Moyenne* = 0,5 et $F = 0,5$

si $P = 0,3$ $R = 0,7$ alors *Moyenne* = 0,5 et $F = 0,42$

7.2 Mode de calcul

Il existe plusieurs façons de quantifier les cardinaux de *A* et *A_{réf}*. Nous en distinguerons trois, suivant que l'on calcule 1/ le nombre de binômes, 2/ les longueurs cumulées des binômes et 3/ les surfaces des binômes.

– Nombre de binômes

La méthode la plus simple consiste à calculer le nombre de couples contenus dans *A*, *A_{réf}*, et $A \cap A_{réf}$. Mais cette mesure ne prend pas en compte l'importance relative des phrases : l'alignement incorrect d'une phrase de deux cents mots sera traité de la même manière que l'alignement incorrect d'une phrase de 2 mots.

– Longueurs cumulées des binômes

On peut bien sûr pallier ce défaut par la prise en compte des longueurs de chaque binôme (exprimées en nombre de caractères ou en nombre de mots).

La longueur d'un couple (*P*, *P'*) pouvant être exprimée par $l(P, P') = l(P) + l(P')$, on a dans l'exemple suivant :

$$A_{réf} = \{(P_1; P'_1 P'_2)(P_2;)(P_3; P'_3)(P_4 P_5; P'_4)(P_6; P'_5 P'_6 P'_7)\}$$

$$A = \{(P_1; P'_1)(P_3; P'_3)(P_4; P'_4)(P_5; P'_5 P'_6)(P_6; P'_7)\}$$

$$L(A_{réf})=l(P_1)+l(P'_1)+l(P'_2)+l(P_2)+l(P_3)+l(P'_3)+l(P_4)+l(P_5)+l(P'_4)+l(P_6)+l(P'_5)+l(P'_6)+l(P'_7)$$

$$L(A)=l(P_1)+l(P'_1)+l(P_3)+l(P'_3)+l(P_4)+l(P'_4)+l(P_5)+l(P'_5)+l(P'_6)+l(P_6)+l(P'_7)$$

$$L(A \cap A_{réf})=l(P_1)+l(P'_1)+l(P_3)+l(P'_3)+l(P_4)+l(P'_4)+l(P_6)+l(P'_7)$$

Le problème d'une telle mesure est qu'elle n'est pas homogène avec la distributivité des couples d'alignement :

$$l(P_i;P'_iP'_2)=l(P_i)+l(P'_i)+l(P'_2) \neq l(P_i;P'_i)+l(P_i;P'_2)=2 \cdot l(P_i)+l(P'_i)+l(P'_2)$$

– *Surface des binômes*

Pour corriger cette anomalie, les consignes du projet ARCADE, s'inspirant des travaux de Simard & Isabelle (1996), suggèrent de pondérer chaque binôme (P_i, P'_j) par l'aire représentée par le produit des longueurs de P_i et P'_j . Par exemple :

$$\begin{aligned} \text{card}(A) &= \text{card}(\{(P_1;P'_1), (P_1;P'_2), (P_2;P'_0), (P_3;P'_3), (P_4;P'_4), (P_5;P'_4)\}) \\ &= \text{Aire}(P_1;P'_1) + \text{Aire}(P_1;P'_2) + \text{Aire}(P_2;P'_0) + \text{Aire}(P_3;P'_3) + \text{Aire}(P_4;P'_4) + \text{Aire}(P_5;P'_4) \\ &= l(P_1) \cdot l(P'_1) + l(P_1) \cdot l(P'_2) + l(P_2) \cdot l(P'_0) + l(P_3) \cdot l(P'_3) + l(P_4) \cdot l(P'_4) + l(P_5) \cdot l(P'_4) \end{aligned}$$

où $l(P)$ représente le nombre de caractères de P .

Si l'on prend : $A_{réf} = \{(P_1;P'_1P'_2)(P_2;)(P_3;P'_3)(P_4P_5;P'_4)(P_6;P'_5P'_6P'_7)\}$

$$A = \{(P_1;P'_1)(P_2;P'_2)(P_3;P'_3)(P_4;P'_4)(P_5;P'_5P'_6)(P_6;P'_7)\}$$

L'évaluation de $A / A_{réf}$ peut être exprimée par les rapports des différentes aires.

Comme le remarquent Simard et Isabelle (1996 : 4), le recours à cette représentation bidimensionnelle permet la comparaison de toute forme d'alignement, quelle qu'en soit la granularité, du moment que l'on calcule les aires au niveau des caractères : « Comme la segmentation en caractères est une donnée qui ne laisse aucune place à l'interprétation, les alignements mettront tous en rapport les mêmes objets, et seront de ce fait directement comparables.

Et en nous servant des calculs des indices basés sur les longueurs, les transfuges et les cognats nous obtenons le rapport de chaque alignement.

8. Analyse quantitative

Nous présentons dans cette partie une statistique de l'analyse de cinquante (50) segments.

➤ L'analyse révèle que les lacunes se trouvant dans les segments du TC en rapport avec les segments du TS sont de quatre (4) catégories : les lacunes provenant d'une mauvaise segmentation, nous avons une (1) occurrence ; des lacunes provenant d'une sur-traduction, nous avons huit (8) occurrences ; des lacunes provenant d'une sous-traduction, nous avons trente-deux (32) occurrences ; des segments avec zéro (0) lacune, nous avons neuf (9) occurrences.

Nous dressons un tableau récapitulatif de ces résultats comme suit :

Catégorie des lacunes	Nombre d'occurrences	Pourcentage
Mauvaise segmentation	01	2%
Segments sans lacunes	09	18%
Sur-traduction	08	16%
Sous-traduction	32	64%
TOTAL	50	100%

Tableau 8.1 : Lacunes dans les segments du TC en rapport avec les segments du TS

➤ Il ressort de l'analyse que l'outil Wordfast Aligner, en effectuant l'alignement automatique, ne rend pas toujours tout le sens du segment du texte source. Ici nous avons également remarqué deux (2) cas de figures : tout le sens non-rendu avec quarante-et-une (41) occurrences, tout le sens rendu avec neuf (9) occurrences. Nous récapitulons ces résultats dans le tableau ci-dessous :

Type de sens rendu	Nombre d'occurrences	Pourcentage
Tout le sens rendu	9	18%
Tout le sens non rendu	41	82%
TOTAL	50	100%

Taleau 8.2 : Récapitulatif des segments du TC ayant le sens des segments du TS

➤ Nous ressortons la moyenne générale des pourcentages des cinquante segments alignés.

Extrait	Pourcentage du rendu	Extrait	Pourcentage du rendu	Extrait	Pourcentage du rendu
1	72%	18	100%	35	67%
2	72%	19	66%	36	01%
3	72%	20	100%	37	83%
4	72%	21	66%	38	83%
5	74%	22	66%	39	83%
6	83%	23	06%	40	83%
7	73%	24	67%	41	100%
8	53%	25	67%	42	100%
9	73%	26	67%	43	83%
10	68%	27	87%	44	83%
11	61%	28	77%	45	85%
12	61%	29	47%	46	31%
13	61%	30	67%	47	100%
14	100%	31	48%	48	53%
15	05%	32	100%	49	38%
16	61%	33	100%	50	58%
17	61%	34	100%		
Moyenne générale du pourcentage du rendu			69,68%		

Tableau 8.3: Evaluation du rendu

➤ l'analyse des segments montre que sur les cinquantes, nous avons neuf (9) cas où le TC est rendu de manière optimale, tandis que les quarante-un (41) autres ne sont pas optimal, comme le montre le tableau ci-dessous :

Alignement	Nombre d'occurrences	Pourcentage
Alignement optimal	9	18%
Alignement non optimal	41	82%
TOTAL	50	100%

Tableau 8.4 : évaluation de l'optimalité du rendu

9. Synthèse des résultats

Les résultats que nous présentons ici ont été obtenus de l'analyse faite au point précédent :

- Au niveau de la première question, l'analyse révèle que l'alignement automatique dans Wordfast Aligner présente un nombre considérable de lacunes.
- Au niveau de la deuxième question, il ressort de l'analyse que l'outil Wordfast Aligner, en effectuant l'alignement automatique, ne rend pas toujours tout le sens du segment du texte source.
- Pour ce qui est de la validité de l'algorithme d'alignement automatique utilisé par Wordfast Aligner, il ressort que le sens du TS dans le TC est à environ 70%, ce qui est acceptable, mais pas optimal.
- Pour ce qui est de la qualité de l'alignement produit par Wordfast Aligner, l'analyse nous montre tout simplement que la technique d'alignement automatique sous Wordfast Aligner produit 18% d'alignement optimal contre 82% d'alignement non optimal.

10. Importance de l'étude

À l'issue de cette étude, il ressort clairement que l'alignement automatique s'avère nécessaire dans la constitution des MT. Mais cette technique nécessite encore beaucoup d'attention de la part du traducteur professionnel qui devra passer en revue tous les segments afin de les polir avant de constituer son MT

La constitution de Mémoires de Traduction est la création de tables d'équivalences entre texte source et texte cible. Pour ce faire, le logiciel divise le texte à traduire en segments. Au moment où le traducteur valide le texte cible correspondant, le logiciel mémorise le segment source et le segment cible comme étant des équivalents linguistiques. Si le segment source apparaît une nouvelle fois dans le texte (les répétitions peuvent être fréquentes dans les textes techniques) le logiciel propose automatiquement la traduction mémorisée. Lors de la mise à jour de la version source d'un texte déjà traduit, le logiciel reprend automatiquement les parties déjà traduites et signale au traducteur les éléments nouveaux ou modifiés.

Les logiciels les plus sophistiqués reconnaissent les segments approximativement identiques, et les signalent au traducteur en marquant les éléments qui diffèrent du segment mémorisé. Donc, l'alignement automatique permet au professionnel de la traduction de gagner en temps matériel.

11. Conclusion et recommandations

L'utilité de la TAO étant manifeste au niveau du principe, il ne faut toutefois pas perdre de vue que son efficacité varie en fonction du type de textes à traduire, que sa mise en œuvre peut exiger d'importants investissements, et que sa rentabilité n'est pas immédiate, les dictionnaires et mémoires de traduction ne s'enrichissant que progressivement. En revanche, le recours aux mémoires de traduction offre une solution au problème des délais lorsque le texte à traduire doit subir des modifications de dernière minute. Dans un tel cas la TAO permet de traduire une première version puis, lorsque le texte source est définitif, de reprendre automatiquement toutes les parties non modifiées et de ne traiter manuellement que les modifications, mises en évidence par le logiciel de TAO. Notre étude traitait

essentiellement de l'évaluation de l'alignement automatique sous Wordfast Aligner en ligne, ceci ouvre d'avantage le vaste champ de recherches sur d'autres outils d'alignement existants. Aux firmes informatiques, nous recommandons de faire d'améliorations sur les outils d'alignement déjà créés, et pourquoi pas d'en inventer d'avantage.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Chinua, A., (1959). Thingsfallapart. First AnchorBooksEdition,1994, Pp. 3-7.
- Chinua, A., (1959). *Tout s'effondre, traduit de l'Anglais* (Nigéria) par PierreGirard. Rales : Actes Sud.
- Mounin, G.,(1963).*LesProblèmesthéoriquesde latraduction*.Paris:Editions Seuil
- Nida, E. A., &Taber, Ch., (1969). *The theory and practice of translation*.Leiden: E.J. Brill.
- Berman, A., (1984). *L'Épreuve de l'étranger : Culture et traduction dans l'Allemagne romantique*.Paris : Edition Gallimard.
- Newmark, P. (1988). *A Textbook of Translation*.New York: Prentice Hall.
- Amparo, H. A., (1990). *La Notion de fidélité en traduction, dans Collection de traductologie*. N°5. Paris : Didier Erudition.
- Lederer, M., &Fortunato, I. (1991). *La liberté en traduction: les acte du colloque international tenu à l'ESIT les 7,8 et 9 juin 1990*. Paris: Didier Érudition.
- Isabelle, P., (1992). Bi-textual aids for translators. In Proceedings of the Eight annual Conference of the UW Centre for the New OED and Text Research, University of Waterloo, Waterloo, pp. 76-89.
- Isabelle, P., (1992). La bi-textualité : vers une nouvelle génération d'aides à la traduction et la terminologie. *META Journal des Traducteurs*, VI. 35. No. 4, pp. 721-731.
- Boualem, F. (1993). ML-TASC: système de traduction automatique multilingue SS'93 Symposium, High Performance Computing.Calgary. Pp. 1-13.
- Melby, A., (1995). Why can't a computer translate more like a person? Consulté le 10/12/2019 sur, <http://www.ttt.org/theory/barker.html>.
- Reiss, K., (2000). *Translation criticism – the potentials & limitations: Categories and Criteria for Translation Quality Assessment*, Erroll F. Rhodes (trad.), London: St. Jerome Publishing.
- Munday, J., (2001). *Introducing Translation Studies*.New York, Routledge
- Coordonnier, J.-L.,(2002).Aspectsculturelsdela tradition:quelques notionsclés,in*Meta: journaldestraducteurs*,VI47.N°1, pp.38-50.
- Mandoki, K., M., (2005). Pour optimisation de l'outil informatique dans l'environnement traductionnel.. Mémoire de Master Inédit. ASTI, Université de Buea.
- Bauer, L.,Holmes, J., &Warren, P.(2006).*Language matters*.NewYork: Palgrave Macmillan.
- Kelese, E. K., (2006). The impact of computer assisted translation (cat) tools on the translation activity in Cameroon's public and Private Sectors. Mémoire de Master Inédit. ASTI, Université de Buea.
- Sauron, V. (2007). Les outils d'aide à la traduction : ennemis ou alliés du traducteur ? », (Ed.) Franco Angeli, Milan, pp199-226.
- Peraldi, S. (2010). *Traduction assistée par ordinateur : entre théorie et Pratique*. Les Cahiers du GEPE, Corpus et mémoires de traduction,Strasbourg : Presses universitaires.
- Dima, El H. (2010). *La traduction assistée par ordinateur dans le contexte égyptien*, Tralogy. Session 1 - Terminology and Translation / Terminologie et Traduction. Paris : Tralogy.
- Delpech. E., (2013). *Traduction assistée par ordinateur et corpus comparables: contributions à la traduction compositionnelle Informatique et langage*. Nantes :PresseUniversitaire.