

Donner accès à l'œuvre de Fu Lei¹

Kim GERDES [1] Jun MIAO [2]

[1] Sorbonne Nouvelle, Université Paris 3 Paris 3 – ILPGA, kim.gerdes@univ-paris3.fr

[2] Sorbonne Nouvelle, Université Paris 3, ESIT, miaojun@miaojun.net

Résumé : Dans le but de donner plus facilement accès à l'œuvre de Fu Lei, nous décrivons la construction d'une base de données de corpus et d'un site incluant des outils lexicométriques adaptés à la recherche traductologique du chinois. Nous exposons les différentes étapes et les outils utilisés dans la construction du corpus ainsi que des exemples d'utilisation du site pour des études en traductologie. Ces outils ouvrent la voie à une autre appréciation, globale et scientifique, de l'œuvre de Fu Lei.

Mots-clés : base de donnée, Fu Lei, traductologie, lexicométrie

中文摘要 : 旨在为进入傅雷译作提供有效捷径, 本文描述如何运用词量学软件来构建其语料库, 而这些软件是专门针对中文翻译学需要而设计的。我们将展示构建此语料库的步骤, 使用工具以及翻译学研究实例。这些技术工具为实现从整体上科学认识傅雷的译作开辟了道路。

关键词 : 语料库, 傅雷, 翻译学, 词量学

Abstract: In order to provide easier access to the work of Fu Lei, we describe the construction of one corpus and one web site which are adapted to Chinese translation studies, by using the lexicometrical tools. We outline the steps and tools used in the construction of this corpus, and we explain, with the concret examples, the use of the web site for the purpose of translation studies. These tools open a new way for the integral and scientific assessment of the Fu Lei's work.

Keywords: corpus, Fu Lei, translation, lexicometrical study

Les traductions de Fu Lei (傅雷, 1908–1966) sont un trésor pour la traductologie et la littérature comparée. Pourtant, fastidieux est l'accès à ses idées dans ses trente-trois traductions des œuvres occidentales, principalement françaises, dont la somme rassemble plus de cinq millions de mots.

Si dans la base de données bibliographiques Fu Lei existante², créée par l'ambassade de France en Chine, on peut consulter les listes des livres traduits du français vers le chinois publiés par des éditeurs de Chine continentale de la fin du XIXe siècle à nos jours, le site ne donne pas un accès direct à ces livres. Pour ceux qui étudient la traduction du français vers le chinois, une vraie base de données de toutes les œuvres traduites par Fu Lei semble nécessaire et urgente, car ses traductions occupent une place importante et très particulière dans le domaine de la traduction chinoise. Nous nous efforcerons dans ce présent article à montrer les problèmes, les procédés et les possibilités de la construction d'une base de données, partielle pour l'instant, et son utilité ainsi que ses fonctions pour une étude lexicométrique (dite aussi « textométrique »).

Choix des corpus

Comme première expérimentation, nous avons choisi le roman de Romain Rolland, *Jean-Christophe*. Cette œuvre a été publiée de 1904 à 1912. Romain Rolland reçut le prix Nobel de littérature de 1915 pour ce roman.

Il y a trois raisons principales à ce premier choix. D'abord, c'est l'œuvre la plus volumineuse que Fu Lei a traduite, contenant 10 tomes, soit environ un demi-million de mots, une taille qui rend l'utilisation de logiciels lexicométriques à la fois nécessaire et puissante. Ensuite, la version électronique de l'original est consultable librement sur le site <http://www.ebooksgratuits.com/ebooks.php>. La traduction de Fu Lei est aussi accessible sur le site <http://www.yifan.net/yihe/novels/foreign/yhkl sdf/kl sdf.html>. Troisièmement, les traductions des œuvres de Romain Rolland représentent le plus le style « Fu Lei ». Comme Fu Lei l'a affirmé lui-même, il a regagné le courage de vivre grâce aux œuvres de Romain Rolland, et parmi toutes ses traductions, ces œuvres sont celles qui, à son avis, ont le mieux conservé « l'esprit » (« 神 » *shen*) de l'original.

1 Les auteurs remercient André Salem et Anne Dister de leur aide chaleureuse pour la réalisation de cet article. Cet article a été présenté dans le Colloque international « Fu Lei et traduction » à l'occasion de son 100e anniversaire, Nanjing, mai 2008.

2 <http://www.fulei.org/>

Fu Lei a mis presque six ans pour traduire et retraduire *Jean-Christophe*. La première publication de *Jean-Christophe* à la maison d'Éditions des Affaires (商务印书馆) date de 1937 à 1941. N'étant pas satisfait de sa propre traduction, Fu Lei a retraduit le texte entier de *Jean-Christophe* et la maison d'Éditions de Pingming (平明出版社) a publié sa retraduction de 1952 à 1953.

Pour obtenir des versions électroniques de bonne qualité, nous avons effectué des vérifications et corrections manuelles sur les textes électroniques originaux³ et traduits, en nous basant respectivement sur l'édition de 1966 parue chez Albin Michel et sur la version de 1998 des Éditions Populaires de l'Anhui (安徽人民出版社) dans le *Recueil des œuvres traduites par Fu Lei* («傅雷译文集»), basée sur la version complète des Éditions Littéraires Populaires (人民文艺出版社) de 1975 (la retraduction de tous les volumes).

Problèmes de la construction des données

Notons deux problèmes de disponibilité des œuvres : 1) Toutes les œuvres et leurs traductions de Fu Lei ne sont pas disponibles sous forme électronique. Faute de versions électroniques existantes, nous devrions établir nous-mêmes les textes électroniques. 2) Il se pose la question de la légalité de l'utilisation des textes. Les droits d'auteur doivent être considérés avant de donner libre accès aux corpus rassemblés.

Il faut signaler que l'objectif de la construction de cette base de données ne vise qu'à faciliter les exploitations scientifiques sur les œuvres traduites par Fu Lei, et n'a pas d'intérêt commercial. La mise en page de la base de données, que nous verrons un peu plus loin, n'est pas du tout convenable pour un public intéressé par une lecture normale. Il est facile d'accéder à des sections particulières du texte, mais une récupération du texte entier ne serait possible qu'après des multiples recherches et des copiés-collés fastidieux. Par conséquent, notre base de données n'empêche pas l'achat normal des livres de traduction de Fu Lei ou des auteurs originaux, tout au contraire, elle pourrait être vue comme une publicité pour ces œuvres.

A part de cela, il existe encore deux problèmes techniques concrets. L'un concerne la langue chinoise, l'autre l'alignement des corpus parallèles entre le français et le chinois.

1. La langue chinoise

L'informatisation du chinois s'est révélée beaucoup plus complexe que celle des systèmes basés sur l'utilisation d'un alphabet réduit (Miao et Salem : 2008).

L'utilisation des approches lexicométriques est facilitée par la segmentation des textes en mots. Ainsi, sachant que la langue chinoise est une *scriptura continua*, et que ses caractères s'écrivent les uns après les autres sans espace entre eux, si l'on veut donner accès à des corpus chinois par le biais de la recherche par mots clés, il faut dans un premier temps segmenter les phrases chinoises en mots. Mais la segmentation du chinois reste toujours un grand obstacle, malgré des progrès considérables pendant les dernières années. Si ce problème a un lien avec l'absence de consensus sur la définition du « mot » en chinois, la morphologie compositionnelle du chinois ajoute de la difficulté à la segmentation des phrases chinoises (voir par exemple Xia : 2000).

Pour l'instant, dans notre corpus parallèle, nous utilisons la segmentation du texte chinois faite par le logiciel ICTCLAS⁴ (*Institute of Computing Technology, Chinese Lexical Analysis System*). Il existe en effet plusieurs logiciels de segmentation du chinois, libres ou en vente sur Internet, comme par exemple *Hailanda Segmentation intelligente, version d'essai* (海量智能分词研究版) et *Lucene*⁵. Pour nos besoins en lexicométrie, le logiciel ICTCLAS a donné des résultats les plus cohérents, mais actuellement, aucun logiciel ne donne des segmentations tout à fait satisfaisantes, et

3 Dans le but de dégager plus facilement les informations du texte original français, nous avons changé toutes les majuscules en minuscules.

4 Ce logiciel est en développement depuis 2002, disponible sur http://www.nlp.org.cn/project/project.php?proj_id=6

5 Voir respectivement les sites : <http://www.hylanda.com/server/> et <http://lucene.apache.org/java/docs/index.html>

les techniques sont en état d'évolution constante.

Dans notre corpus de la traduction de Fu Lei de *Jean-Christophe*, nous avons surtout dû corriger la segmentation des traductions des noms propres des personnages. Par ailleurs, dans le but de rendre plus facilement comparable le texte chinois avec le texte français, nous avons changé la ponctuation chinoise en ponctuation française où l'établissement d'une équivalence nous semblait naturelle, mais pour certaines ponctuations qui n'ont pas de correspondances en français, nous les avons gardées telles quelles (par exemple, « 、 », la demi-virgule). Voici un paragraphe de la traduction originale et segmentée :

Tableau 1 : Extrait de la traduction chinoise de *Jean-Christophe* originale et segmentée

original	他们不再说话了。约翰·米希尔坐在壁炉旁边，鲁意莎坐在床上，都在那里黯然神往。老人嘴里是这么说，心里还想着儿子的婚事非常懊丧。鲁意莎也想着这件事，埋怨自己，虽然她没有什么可埋怨的。
segmenté	他们不再说话了。约翰·米希尔坐在壁炉旁边，鲁意莎坐在床上，都在那里黯然神往。老人嘴里是这么说，心里还想着儿子的婚事非常懊丧。鲁意莎也想着这件事，埋怨自己，虽然她没有什么可埋怨的。

Par ailleurs, nous avons rencontré des problèmes d'encodage du corpus chinois. Bien que la norme *Unicode* fournisse une bonne solution pour représenter les caractères chinois, quelques problèmes existent cependant à cause de l'unification des jeux idéographiques utilisés dans différentes langues, avec une calligraphie légèrement différente et parfois signifiante⁶ – une norme satisfaisante et généralement acceptée n'est pas encore en vue.

2. L'alignement du texte français et chinois

Le français et le chinois sont des langues lointaines, et il n'y a pas de cognats⁷ entre elles. Ainsi, l'alignement automatique de ces deux textes est beaucoup plus difficile qu'entre deux textes dans des langues voisines. En plus, nous relevons, dans notre corpus, que Fu Lei a modifié de manière considérable la forme du texte original : parfois, il démantèle un paragraphe original en deux paragraphes indépendants dans sa traduction; parfois, il regroupe deux paragraphes en une grande unité de paragraphe. Par contre, il s'agit d'une traduction littéraire et donc assez précise de l'original sans trop d'ajouts ou d'émissions du traducteur.

Faute de logiciels adaptés à l'alignement franco-chinois⁸, nous avons développé nous-mêmes un logiciel d'alignement semi-automatique : *Alignator*, programmé en Python (pour le calcul) et en Javascript (pour l'interface graphique). En l'absence de cognats, *Alignator* utilise un algorithme dynamique de dilatation temporelle pour deviner certaines paires de traduction. L'idée est que les paires de traduction ont une distribution similaire dans le texte entier – leurs positions peuvent être légèrement décalées par des insertions et des émissions dans la traduction, mais au total leurs signaux d'apparition se ressemblent plus qu'entre des mots qui ne sont pas des traductions mutuelles. Notons que ce procédé ne dépend pas d'une segmentation préalable en mots. Pour une *scriptura continua*, *Alignator* utilise les caractères simples.

Les paires de « mots » ainsi dégagées sont ensuite utilisées dans un procédé plus standard provenant de l'alignement par cognats : l'algorithme cherche un chemin diagonal minimal à travers une

6 Des questions équivalentes dans un contexte européen seraient par exemple « Est-ce que la lettre « A » en gothique est le même symbole que le « A » dans une police standard ? » ou encore, « Est-ce que la lettre alpha en majuscule du grec est le même symbole que le « A », pourtant identique du point de vue de sa forme ? ». Cf par exemple <http://fr.wikipedia.org/wiki/Unicode>

7 Le *cognat* signifie une paire de formes qui se ressemblent et qui sont des traductions mutuelles.

8 La plupart des logiciels d'alignement automatique se basent sur la présence de cognat et sont donc adaptés aux paires de langues proches (voir Simard et al. 1992). De plus, les logiciels applicables au chinois sont difficilement disponibles car ils utilisent de grands lexiques.

matrice de distances entre les paragraphes. Les résultats sont meilleurs pour des bitextes de grande taille (ou les signaux apparaissent plus clairement) et d'un parallélisme assez précis, sans trop d'insertions ou torsions, conditions que vérifie notre corpus de *Jean-Christophe*.

L'idée de base de l'alignement de *Alignator* est qu'on ne veut jamais découper un paragraphe en deux, mais on veut aligner les frontières de paragraphes le plus précisément possible. De plus, l'algorithme part du principe, souvent faux et à corriger manuellement, que tout paragraphe dispose d'un (ou plusieurs) correspondant(s), autrement dit, il n'existe pas de paragraphe sans équivalent de traduction, comme par exemple des introductions à l'œuvre écrites par le traducteur. Les résultats obtenus automatiquement ne sont que rarement parfaits et doivent normalement être corrigés manuellement, mais cette deuxième étape manuelle est considérablement facilitée par l'alignement automatique.

La mise en correspondance peut directement être corrigée manuellement dans l'interface de l'application en ligne. Le résultat peut être exporté dans différents formats, entre autres, dans un format adapté pour *Alignoscope*. Dans le paragraphe suivant, nous expliquons brièvement la fonctionnalité de l'interface de correction. On peut utiliser des boutons ou bien le glisser-déposer pour travailler le bitexte :

I. les boutons au bord du bloc :

Vers le haut : permettre de faire une manipulation d'un bloc vers le cadre plus haut.

^ : monter dans un bloc vide (cas rare où le paragraphe n'a pas de correspondant);

^^ : joindre au bloc précédent;

^^^ : joindre au bloc précédent et recalculer les blocs suivants, en conservant les alignements déjà effectués.

Vers le bas : permettre de faire une manipulation d'un bloc vers le cadre plus bas.

Les règles sont de même que les boutons vers le haut.

II. glisser-déposer (Drag and Drop):

Un clic de la souris sur le paragraphe choisi et un tirage vers le paragraphe correspondant permettent leur mise en correspondance intuitive et dynamique. Un popup contenant les premiers mots des deux paragraphes choisis va s'afficher pendant la manipulation. Le symbole de chaîne indique ensuite cette mise en correspondance qui ne serait pas modifiée par un recalcul automatique.

La figure 1 suivante illustre la correction manuelle de l'alignement entre le texte original *Jean-Christophe* et sa traduction de Fu Lei. Le résultat de cet alignement peut être obtenu par la *manipulation 1* en utilisant le bouton vers le bas et également par la *manipulation 2* à l'aide du glisser-déposer.

Dès que la correction manuelle est faite, on peut exporter l'alignement résultant en XML selon les besoins de la recherche : choix libres des marqueurs de séparation, choix des places des marqueurs (devant ou après les paragraphes alignés), choix entre la sortie d'un seul bitexte aligné ou de deux textes avec balisage parallèle. Voici un extrait du bitexte aligné par un marqueur # après chaque paragraphe.

Tableau 2: Extrait de l'alignement du corpus parallèle français-chinois

```
- bon dieu ! qu'il est laid ! fit le vieux, d'un ton convaincu. il alla reposer la lampe sur la table.
#
" 天 哪 ! 他 多 丑 ! " 老 人 语 气 很 肯 定 的 说 .
  他 把 灯 放 在 了 桌 上 .
#
louis fit une moue de petite fille grondée. jean-michel la regarda du coin de l'œil, et rit.
- tu ne voudrais pas que je te dise qu'il est beau ? tu ne me croirais pas. allons, ce n'est pas de ta
faute. ils sont tous comme cela.
#
  鲁 意 莎 撅 着 嘴 , 好 似 挨 了 骂 的 小 姑 娘 , 约 翰 ^ 米 希 尔 觑 着 她 笑 道 : " 你 总 不 成 要
我 说 他 好 看 吧 ? 说 了 你 也 不 会 信 . 得 了 罢 , 这 又 不 是 你 的 错 , 小 娃 娃 都 是 这
样 的 . "
#
```

Manipulation 1

l'enfant s'éveille et pleure. son regard trouble s'agite. quelle épouvante ! les ténèbres, l'éclat brutal de la lampe, les hallucinations d'un cerveau à peine dégagé du chaos, la nuit étouffante et grouillante qui l'entoure, l'ombre sans fond d'où se détachent, comme des jets aveuglants de lumière, des sensations aiguës, des douleurs, des fantômes : ces figures énormes qui se penchent sur lui, ces yeux qui le pénètrent, qui s'enfoncent en lui, et qu'il ne comprend pas !... il n'a pas la force de crier ; la terreur le cloue immobile, les yeux, la bouche ouverts, soufflant du fond de la gorge. sa grosse tête boursouflée se plisse de grimaces lamentables et grotesques ; la peau de sa figure et de ses mains est brune, violacée, avec des taches jaunâtres...

– bon dieu ! qu'il est laid ! fit le vieux, d'un ton convaincu. il alla reposer la lampe sur la table.

block 7
⇓

孩子醒过来,哭了,惊慌的眼睛在那儿乱转,多可怕啊!无边的黑暗,剧烈的灯光,浑沌初萌的头脑里的幻觉,包围着他的那个闷人的、蠕动不已的黑夜,还有那深不可测的阴影中,好似耀眼的光线一般透出来的尖锐的刺激,痛苦,和幽灵,一使他真名片妙的那些巨大的脸正对着他,眼睛瞪着他,直透到他心里去...他没有气力叫喊,吓得不能动弹,睁着眼睛,张着嘴,只在喉咙里喘气,带点虚肿的大胖脸扭做一堆,变成可笑而又可怜的怪样子;脸上与手上的皮肤是棕色的,暗红的,还有些黄黄的斑点。

"天哪!他多丑!"老人语气很肯定的说。

他把灯放在了桌上。

Manipulation 2

l'enfant s'éveille et pleure. son regard trouble s'agite. quelle épouvante ! les ténèbres, l'éclat brutal de la lampe, les hallucinations d'un cerveau à peine dégagé du chaos, la nuit étouffante et grouillante qui l'entoure, l'ombre sans fond d'où se détachent, comme des jets aveuglants de lumière, des sensations aiguës, des douleurs, des fantômes : ces figures énormes qui se penchent sur lui, ces yeux qui le pénètrent, qui s'enfoncent en lui, et qu'il ne comprend pas !... il n'a pas la force de crier ; la terreur le cloue immobile, les yeux, la bouche ouverts, soufflant du fond de la gorge. sa grosse tête boursouflée se plisse de grimaces lamentables et grotesques ; la peau de sa figure et de ses mains est brune, violacée, avec des taches jaunâtres...

– bon dieu ! qu'il est laid ! fit le vieux, d'un ton convaincu. il alla reposer la lampe sur la table.

block 6
⇓

孩子醒过来,哭了,惊慌的眼睛在那儿乱转,多可怕啊!无边的黑暗,剧烈的灯光,浑沌初萌的头脑里的幻觉,包围着他的那个闷人的、蠕动不已的黑夜,还有那深不可测的阴影中,好似耀眼的光线一般透出来的尖锐的刺激,痛苦,和幽灵,一使他真名片妙的那些巨大的脸正对着他,眼睛瞪着他,直透到他心里去...他没有气力叫喊,吓得不能动弹,睁着眼睛,张着嘴,只在喉咙里喘气,带点虚肿的大胖脸扭做一堆,变成可笑而又可怜的怪样子;脸上与手上的皮肤是棕色的,暗红的,还有些黄黄的斑点。

"天哪!他多丑!"老人语气很肯定的说。

– bon dieu ! qu'il est l...

to "天哪!他多丑!..."

他把灯放在了桌上。

block 6
⇓

l'enfant s'éveille et pleure. son regard trouble s'agite. quelle épouvante ! les ténèbres, l'éclat brutal de la lampe, les hallucinations d'un cerveau à peine dégagé du chaos, la nuit étouffante et grouillante qui l'entoure, l'ombre sans fond d'où se détachent, comme des jets aveuglants de lumière, des sensations aiguës, des douleurs, des fantômes : ces figures énormes qui se penchent sur lui, ces yeux qui le pénètrent, qui s'enfoncent en lui, et qu'il ne comprend pas !... il n'a pas la force de crier ; la terreur le cloue immobile, les yeux, la bouche ouverts, soufflant du fond de la gorge. sa grosse tête boursouflée se plisse de grimaces lamentables et grotesques ; la peau de sa figure et de ses mains est brune, violacée, avec des taches jaunâtres...

block 7
⇓

"天哪!他多丑!"老人语气很肯定的说。

他把灯放在了桌上。

Résultat

– bon dieu ! qu'il est laid ! fit le vieux, d'un ton convaincu. il alla reposer la lampe sur la table.

louisa fit une moue de petite fille grondée. jean-michel la regarda du coin de l'œil, et rit.

– tu ne voudrais pas que je te dise qu'il est beau ? tu ne me croirais pas. allons, ce n'est pas de ta faute, ils sont tous comme cela.

block 7
⇓

"天哪!他多丑!"老人语气很肯定的说。

他把灯放在了桌上。

block 8
⇓

鲁意莎撅着嘴,好似挨了骂的小姑娘,约翰·米希尔觑着她笑道:"你总不成要我说他好看吧?说了你也不会信,得了罢,这又不是你的错,小娃娃都是这样的。"

Figure 1 : Alignement des textes français-chinois dans l'Alignator (extrait)

Analyses du corpus

Le texte aligné de cette manière peut directement être importé dans notre logiciel lexicométrique *Alignoscope*, qui est une sorte de concordancier, permettant de voir graphiquement la distribution des occurrences de certaines formes dans le texte entier. Les fonctions de recherches donnent directement accès aux paires de paragraphes alignés contenant ou excluant certaines formes dans la langue source ou cible. Par ailleurs, le logiciel calcule également les spécificités du sous-corpus constitué des paragraphes correspondant à la recherche. La liste des mots les plus spécifiques est intéressante pour l'étude des variations de la traduction et des collocations dans le corpus. Sur le site actuellement en ligne, on trouve déjà le bitexte de *Jean-Christophe* (<http://miaojun.net/alignoscope>). D'autres textes sont en cours de préparation à la mise en ligne.

En effet, la concordance est largement utilisée dans les études littéraires assistées par ordinateur : on peut examiner les phrases contenant les mots recherchés dans la plupart des logiciels existants, cependant, il est difficile de discerner le contexte plus large des mots recherchés ainsi que leur distribution globale. Dans le logiciel *Alignoscope*, la fonction de l'affichage simultané rend possible l'examen du contexte des unités de forme, et cela sans l'installation locale d'un logiciel – un simple navigateur conforme aux normes internationales suffit.

En choisissant les termes qui qualifient l'homme dans le texte original, nous allons fournir des exemples thématiques pour concrétiser l'utilité de notre base de données aux traductologues ainsi qu'aux apprenants avancés du français.

1. Des unités de forme

En tant que premier exemple, nous examinons comment Fu Lei a traduit le mot français « honnête » en chinois. En entrant ce mot dans le champ de recherche « contient » du côté original, nous obtenons l'affichage suivant :

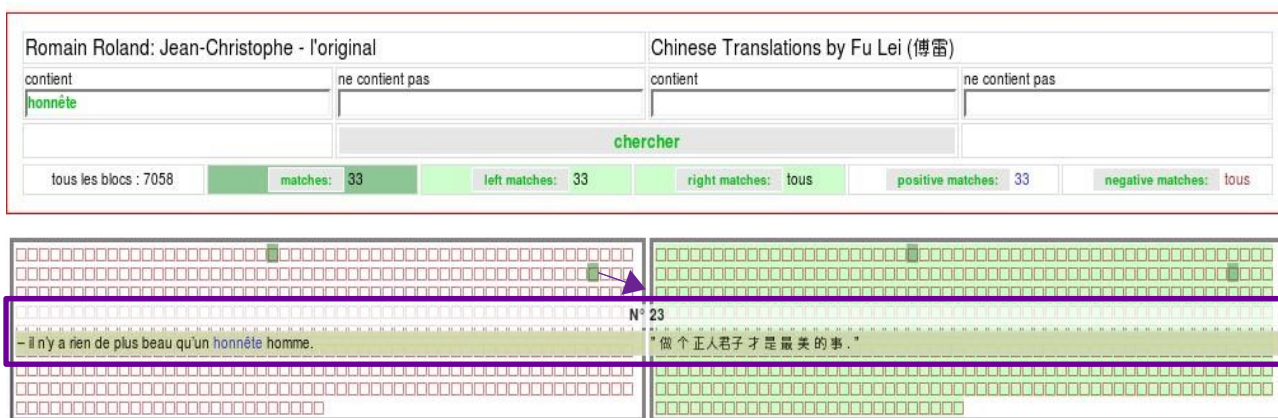


Figure 2 : L'affichage simultané de « honnête » dans la base de données (extrait)

Pour l'ensemble du corpus, il y a au total 7058 blocs, chaque bloc comprend une unité de paragraphes alignés. La page des résultats de la recherche (Figure 2) montre que le bitexte contient 33 blocs contenant « honnête » dans le côté français. Le carré signalé par la couleur vert foncé signifie un bloc correspondant complètement à notre recherche. En passant avec la souris au-dessus d'un carré, le bloc bilingue est chargé et affiché, comme dans la Figure 2. Un clic sur le bloc l'ouvre de manière permanente. Ce chargement « sur demande » permet une grande rapidité de démarrage du logiciel et une charge de mémoire très faible pour l'ordinateur personnel. L'alternative, le chargement dès le lancement du logiciel des dix volumes de *Jean-Christophe*, serait très lente ainsi que lourde en analyse. Dans la fenêtre popup, on distingue les mots correspondant à la recherche sous des couleurs différentes. Dans la phrase originale à gauche du popup, le mot « honnête » dans l'exemple « – il n'y a rien de plus beau qu'un honnête homme. » est signalé par la couleur bleue. Dans le popup droit, on voit la traduction de Fu Lei : “做个正人君子才是最美的事.” Il est donc facile de relever que Fu Lei utilise l'expression figurée chinoise « 正人君子 »

(zhengrenjunzi, « l'homme d'honneur ») pour rendre « un honnête homme » en chinois. Le numéro « 23 » au milieu du popup marque sa place dans l'ensemble des blocs. Un second clic dans le vide ou dans la ligne du titre du popup (avec le numéro) va fermer le popup. De cette manière, nous avons le moyen d'examiner systématiquement et précisément les mots auxquels Fu Lei a eu recours pour traduire « honnête ». L'Alignoscope peut également sortir tous les blocs résultant de la recherche pour faciliter une recherche détaillée de toutes les traductions.

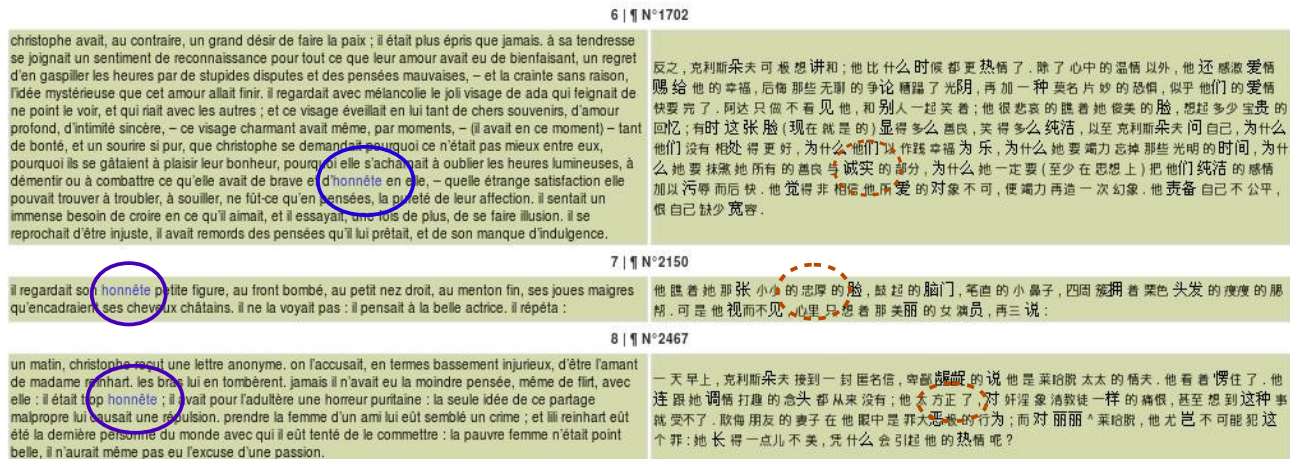


Figure 3 : Résultats de la recherche de « honnête » dans la base de donnée (extrait)

Un simple examen de la figure 3 montre que Fu Lei s'est servi de divers mots chinois tels que, « 诚实的 » (chengshi, « honnête »), « 忠厚的 » (zhonghou, « fidèle et doux »), « 方正 » (fangzheng, « carré ») pour traduire le mot « honnête » dans les contextes variés.

De cette manière nous n'avons vu que la forme exacte « honnête » mais pas les autres formes voisines. Pour saisir toute la notion de « honnête » dans l'original, nous devons prendre en considération toutes les formes et dérivations du mot « honnête ». Ceci est possible en cherchant l'expression « honnête|honnêtes|honnêteté|honnêtement » (la barre droite signifie la disjonction ou, donc cette expression correspond à un de ces mots séparés par les barres).

On peut ainsi obtenir un survol de la traduction du mot « honnête », mais il reste au traductologue d'interpréter les traductions de Fu Lei, et de faire des hypothèses sur sa vision de l'honnêteté.

2. La spécificité

Avec la recherche des occurrences des blocs, l'Alignator effectue un calcul statistique sur l'ensemble des blocs correspondant aux critères de recherche : le test exact de Fisher, i.e. que pour chaque mot on calcule la somme des probabilités que le mot apparaisse dans le sous-corpus des blocs correspondant aux critères de recherche autant de fois qu'observé. Si un mot apparaît beaucoup plus qu'une distribution équilibrée laisserait prédire, on dit que le mot est spécifique pour ce sous-corpus. Évidemment, les mots recherchés apparaissent toujours parmi les mots spécifiques parce que le sous-corpus est justement constitué en cherchant ces mots. Mais d'autres mots aussi font leur apparition dans la liste, car ils apparaissent plus souvent qu'attendu, c'est-à-dire que ces mots cooccurrent avec les mots recherchés pour des raisons syntaxiques, sémantiques ou textuelles. Même si nous avons sélectionné les blocs sur un critère unilingue, les listes de spécificité mettent en évidence les mots spécifiques des deux sous-corpus. La liste de la 2e langue inclut différentes traductions des mots recherchés.

Inversement, on peut aussi obtenir la liste des mots apparaissant moins souvent que le hasard le prédirait. Ces mots ont souvent une relation antonymique, dans un sens large qui inclut les constructions syntaxiques et sémantiques, avec les mots recherchés (cf. Lebart et Salem 1994).

Dans notre exemple, en entrant « honnête|honnêtes|honnêteté|honnêtement »⁹ dans le volet du texte original, les spécificités positives et négatives pour les deux langues s'affichent. Pour respecter la limite de pages d'édition, nous comparons ici seulement les premières spécificités positives du texte original et de la traduction.

Tableau 3: Les spécificités des mots « honnête » dans le corpus parallèle français-chinois (extrait)

spec	freq sub	freq tot	token (texte original)	spec	freq sub	freq tot	token (texte traduit)
∞	35	35	honnête	∞	5	5	维廉哀
∞	24	24	honnêtes	∞	4	4	坦华塞
∞	16	16	honnêteté	∞	3	3	阿迦玛衣,
∞	3	3	villiers	∞	3	3	巡礼
∞	3	3	blondine	∞	3	3	守本分
∞	2	2	écœurée	∞	3	3	君子
∞	2	2	traduits	∞	3	3	鲍东
∞	2	2	tannhäuser	∞	2	2	阿丹丽
∞	2	2	lucie	∞	2	2	贪心
∞	2	2	imitateurs	∞	2	2	苏阿勃
∞	2	2	fromage	∞	2	2	眉开眼笑
∞	2	2	effrontés	∞	2	2	特维廉哀
∞	2	2	difformités	∞	2	2	正派
∞	2	2	affectent	∞	2	2	戏谑
∞	1780	53289	;	∞	2	2	前夫
∞	1218	38746	,	∞	2	2	信奉
∞	848	24673	de	∞	2	2	作料
∞	696	28498	.	∞	2	2	何谓
∞	474	19078	il	∞	2	2	任意
∞	458	13658	et	∞	1673	48512	的
∞	419	14671	la	∞	1640	49890	,
∞	325	12092	à	∞	854	34292	.
∞	302	11489	le	∞	452	17791	他
∞	300	10436	l	∞	376	10714	是
∞	298	8632	les	∞	281	9442	不
∞	282	7523	d	∞	275	13680	了
∞	254	8535	elle	∞	176	8133	在
∞	243	7492	un	∞	169	7607	着
∞	243	6289	qui	∞	94	8727	"
∞	238	7302	qu	39	19	21	诚实
∞	234	6194	des	13	26	87	'
∞	211	6584	en	10	11	27	规矩
∞	206	7008	ne	9	6	11	依斐日尼
∞	198	6806	;	9	4	6	饼
∞	190	6355	était	8	26	114	老实
∞	173	6131	pas	8	3	4	伯伯
∞	168	6513	se	6	4	9	帝王
∞	166	6081	-	6	3	6	神幻
∞	154	5538	lui				
∞	152	5725	que				
∞	134	5133	s				
∞	124	5172	-				
∞	88	4543	christophe				
∞	28	12092	a				
∞	18	14671	là				
∞	2	7008	né				
∞	2	6194	dès				
∞	2	3142	dû				
7	3	4	aristocrates				
6	5	10	moralité				

Côté gauche du *tableau 3*, nous voyons que les mots « honnête », « honnêtes », « honnêteté » apparaissent respectivement 35, 24, 16 fois. Nous voyons que les « mots de l'honnêteté » apparaissent dans un contexte de commentaires sur les personnages du roman : on rencontre des mots comme « écoeurée », « affectent », « moralité » et « vertueuse », etc., (manuellement entourés ici en bleu). Dans ces paragraphes, Romain Rolland compare les qualités et les vices de l'homme, ce qui explique l'apparition des noms des personnes dans la liste. De plus, on trouve de la ponctuation et certains mots fonctions qui co-occurrent avec les mots recherchés ; on tombe aussi sur du bruit comme « fromage » – des mots qui se retrouvent dans la liste parce que le sous-corpus contient par

9 On peut aussi utiliser l'expression régulière équivalente « honnêt.* », signifiant tout mot commençant par « honnêt ». L'Alignoscope propose aussi d'autres types d'expressions régulières décrits sur le site du logiciel.

hasard toutes leurs apparitions.

Côté droit, les spécificités remarquables des mots comme « 守本分 » (*shoubenfeng*, « celui qui tient ses devoirs »), « 君子 » (*junzi*, « l'homme d'honneur »), « 正派 » (*zhengpai*, « probe ») trahissent, automatiquement et sans recours manuel au texte, que ces mots sont très probablement utilisés par Fu Lei en tant que traduction indirecte du concept de l'honnêteté. On trouve aussi des mots tels que « 诚实 » (*chengshi*, « honnête », spécificité +39), « 规矩 » (*guiju*, « correct », +18), « 老实 » (*laoshi*, « droit », +10) qui sont employés par Fu Lei à cause de leur proximité sémantique avec « honnêteté ».

Cependant, la liste des spécificités des mots ne remplace pas la recherche manuelle dans le bitexte, malgré le fait que ces spécificités peuvent fournir des idées et des tendances dans les traductions de Fu Lei. Dans le but de dégager précisément les correspondants chinois dont Fu Lei s'est servi pour la notion « honnête », nous revenons à l'affichage simultané de l'original et de sa traduction. Un examen comparatif précis donne le recensement suivant. Nous mettons une glose simple, pas toujours tout à fait convenable après chaque mot chinois, en espérant au moins montrer une partie de la délicatesse dans le choix des mots effectués par Fu Lei.

Tableau 4 : Les traductions pour le groupe de mot « honnête » dans la traductions de Fu Lei

Mots originaux	Traductions	Pinyin	Glose	F	Traductions	Pinyin	Glose	F
honnête	正人君子	zhengrenjunzi	l'homme d'honneur	1	笃厚的	duhoude	doux	1
	善	shan	gentil	1	规矩诚实	guijuchengshi	correct et honnête	2
	老实的	laoshide	droit	1	良家妇女的	liangjiafunvde	bonne femme	2
	老实	laoshi	droit	6	规矩	guiju	correct	4
	真正的	zhengzhengde	véritable	1	清清白白的	qingqingbaibaide	transparent	1
	守本分的	shoubenfende	celui qui tient ses devoirs	1	胆小	qingbaide	transparent	1
	诚实的	chengshide	honnête	6	道德	zhongshi	fidèle	1
	忠厚的	zhonghoude	fidèle et doux	2	安分守己的	anfenshoujide	celui tient ses devoirs et sans histoire	1
	方正	fangzhengde	carré	1	认真的	renzhengde	sérieux	1
	本色的	benshede	de couleur naturelle	1	老实人的	laoshirende	homme candide	1
	真诚	zhengcheng	sincère	1	完全的	wanquande	complet	1
honnêtes	老实的	laoshide	droit	4	诚实的	chengshide	honnête	4
	泰然的	tairande	tranquille	1	正派的	zhengpaide	probe	1
	老实话的	laoshihuade	discours francs	1	简单的	jiandande	simple	1
	老实人	laoshiren	homme candide	4	懦弱的好人	nuoruodehaoren	bon homme lâche	1
	胆小	danxiao	timide	1	好好先生	haohaoxiansheng	bon monsieur	1
honnêteté	道德	daode	morale	2	正人君子	zhengrenjunzi	l'homme d'honneur	1
	大贤大德	daxiandade	grande bonté et grande vertu	1	诚实	chengshi	honnêteté	5
	老成	laocheng	sérieux	1	奉公守法	fenggongshoufa	celui qui tient ses devoirs et respecte les lois	1
	老实	laoshi	droit	3	守本分	shoubenfen	celui qui tient ses devoirs	1
honnête-ment	老老实实	laolaoshishi	honnêtement	1				

Trois choses se dégagent à partir du tableau 4 : 1) L'étape précédente des spécificités nous fournit des idées des recherches à effectuer, car tous les mots dans la liste des spécificités se retrouvent dans l'examen du contexte. 2) Un examen manuel du contexte est nécessaire, permettant de discerner les nuances des mots dans leurs contextes. 3) Certains mots chinois ne peuvent se détecter que sous l'examen du contexte, à cause de la segmentation automatique de la phrase chinoise, par

exemple, l'expression « 懦弱的好人 » (*nuoruodehaoren*, « bon homme lâche ») est segmentée en trois unités « 懦弱 », « 的 », et « 好人 ». La spécificité ne reconnaît pas facilement ce genre de traductions, car le calcul est basé sur les unités de mots.

2. D'autres moyens de traduire

Les traductologues s'intéressent aux variations de traduction. Dans cette étape, nous essayons de montrer comment dégager les divergences et la variété des traductions de Fu Lei dans *L'Alignoscope*. Ainsi sont mises en évidence la créativité du traducteur et sa technique de traduction.

L'Alignoscope propose des champs de recherche positifs (« contient ») et des champs de recherche négatifs (« ne contient pas »). Si on met une expression dans le champ de recherche négatif, on exclut des résultats tout bloc qui contient ces unités. En effet, avec les champs « contient », et « ne contient pas » on peut déterminer précisément les critères de la recherche dans les deux volets (original et traduction). Les résultats d'une telle recherche complexe s'affichent dans *L'Alignoscope* à l'aide des combinaisons de couleurs, véhiculant deux dimensions de signification.

I. la couleur de l'arrière plan du carré :

reflète le remplissage complet des conditions de la recherche ou le remplissage partiel dans un des deux volets de langue.

- vert foncé : correspondance parfaite dans les deux langues
- vert clair : correspondance parfaite dans le volet concerné mais pas de correspondance parfaite dans les deux langues
- vide : ne répond pas aux conditions de recherche du volet concerné

II. la couleur du carré-même : reflète les exigences transversales.

- bleu : le bloc correspond aux deux expressions positives des deux langues
- rouge : le bloc contient tous les mots exclus dans les deux langues
- noir : pas de correspondance ni positive ni négative dans les deux langues

Par exemple, un carré bleu avec un arrière-plan vert foncé signifie que le contenu du bloc correspond parfaitement aux exigences de la recherche.

Les mêmes couleurs sont utilisées pour marquer les mots recherchés dans le popup :

- bleu : les mots recherchés (« contient »);
- rouge : les mots exclus (« ne contient pas »).

A titre d'exemple, nous cherchons les traductions alternatives pour un groupe de mots au sujet de l'« amour ». Entrons « amour|aim.* » dans le champ de recherche « contient » du côté de texte original. De cette manière, toutes les formes de conjugaison du verbe « aimer » seront incluses dans notre recherche. Étant donné que cette recherche inclut également toutes les formes du mot « aimable » que nous ne souhaitons pas analyser, nous mettons « aimable.* » dans le champ de recherche « ne contient pas », autrement dit, les mots « aimable », « aimables » et « aimablement » n'ont pas d'intérêt pour cette recherche.

Du côté du texte traduit, sachant que « 爱 » (*ai*, « adorer ») et « 喜欢 » (*xihuan*, « aimer ») et « 爱情 » (*aiqing*, « amour ») sont couramment utilisés pour « aimer » et « amour », nous les mettons tous dans le champ de recherche « ne contient pas », en laissant vide le champ « contient » de la traduction dans le but d'analyser la variation dans la traduction de Fu Lei. Voici le résultat :



Figure 4 : L'affiche de la recherche des variations de traduction au sujet de l' « amour » dans la traduction de Fu Lei (extrait)

La Figure 4 montre qu'il y a 170 blocs dans le corpus parallèle entier qui correspondent exactement à notre exigence de recherche, c'est-à-dire que dans 170 endroits Fu Lei n'a pas utilisé « 爱 » ni « 爱情 » ni « 喜欢 » pour rendre « amour|aim.* » en chinois. Ces 170 blocs sont signalés par la couleur vert foncé et le carré bleu. L'affichage du bloc N° 19, par exemple, montre que Fu Lei a choisi « 我多疼你 » (*woduotengni*, « comme je te chéris ») pour traduire « comme je t'aime ».

Le carré bleu clair avec l'affichage de N° 409 signifie que côté texte original le mot « aimerai » répond au critère de recherche établi, mais côté texte traduit le mot chinois « 喜欢 » (*xihuan*, « aimer ») souligné en rouge est le mot dans le champ de recherche « ne contient pas ». C'est la raison pour laquelle cette paire de paragraphes s'affiche avec le cadre bleu sans arrière-plan.

Étant donné que nous avons laissé vide le champ de recherche positive (« contient ») du côté de la traduction, un paragraphe ne contenant pas les mots chinois exclus apparaît en vert clair, comme par exemple le bloc N° 767. Mais côté gauche, nous avons un carré noir sans arrière-plan, puisque dans l'original, il n'y a aucun des mots des champs de recherche positive ou négative.

Dans le bloc N°687, par exemple, les mots « aimable » et « 喜欢 » (*xihuan*, « aimer ») apparaissent dans le volet de leur langue respective, mais ce sont des mots exclus (« ne contient pas »), et ils sont donc marqués en rouge.

Cette coloration permet un survol facile de la situation et permet un accès direct aux différentes traductions de Fu Lei. On voit aussi d'un coup la distribution globale des 170 blocs contenant une traduction alternative des mots «amour|aim.* ».

Conclusion

Il nous semble important que les méthodes de travail sur corpus assisté par ordinateur se répandent aussi en Chine dans le domaine de la traductologie ou plus généralement en études littéraires. Bien qu'il y ait des chercheurs¹⁰ qui ont commencé à prêter attention à la construction de corpus depuis les années 80, leur objectif consiste premièrement à promouvoir la réforme de l'enseignement de l'anglais en Chine, en fournissant des données pour le «Programme pédagogique des cours de l'anglais à l'université» et le «Programme des examens d'anglais au niveau universitaire, CET (College English Test)» (Wei : 2005). En bref, ce genre de corpus sert généralement à la pédagogie et à la théorie des langues étrangères et est spécialisé à l'anglais. Tout récemment, on peut constater un essor des constructions de corpus en Chine¹¹ mais il n'existe pas encore de corpus parallèle français-chinois. D'ailleurs, peu nombreux sont les recherches et les travaux qui approfondissent un domaine de recherche en particulier à cause du manque d'outils adéquats et applicables au texte chinois.

La recherche traductologique se situe souvent dans une voie de recherche empirique. La construction de bases de données et les méthodes lexicométriques vont amener la recherche vers une voie plus scientifique et efficace. La proposition de la construction de la base de données a pour objectif de rendre plus facilement accessibles toutes les œuvres traduites par Fu Lei, et également de faciliter une recherche scientifique traductologique sur ces trésors, précieux non seulement pour le domaine de la traduction mais d'un grand intérêt littéraire universel.

Bibliographie

Lebart L. et Salem A. 1994. *Statistique textuelle*. Paris, Dunod.

Liu Kanglong & Mu Lei (刘康龙, 穆雷). 2006. Corpus Linguistics and Translation Studies (语料库语言学与翻译研究). in *Chinese Translators Journal* (中国翻译). Beijing, Vol.27 N° 1, p59-64.

Miao J. et Salem A. 2008. Comparaison textométrique de traductions franco-chinoises, in *Lexicometria*, <http://www.cavi.univ-paris3.fr/ilpga/ilpga/tal/lexicowww/navigations/Mult2.pdf>

Simard M., Foster G. and Isabelle P. 1992. *Using Cognates to Align Sentences in Bilingual Corpora*. Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation TMI-92 (pp. 67-81), Montréal.

Wei Naixing, Li Wenzhong & Pu Jianzhong, (卫乃兴, 李文中, 濮建忠). 2005. *Corpora in use* (语料库应用研究), Shanghai, 上海外语教育出版社.

Xia F. 2000. *The Segmentation Guidelines for the Penn Chinese Treebank (3.0)* Consultable sur le site www.cis.upenn.edu/~chinese/segguide.3rd.ch.pdf (Lu le 14 février 2008).

10 Yang Huizhong (杨惠中) a établi le premier corpus des documents anglais JDEST en Chine (Jiaoda Corpus of English for Science and Technology: <http://corpus.sjtu.edu.cn/>), puis on peut voir des corpus des CLEC (Chinese Learner English Corpus : <http://www.clal.org.cn/corpus/ChiSearchEngine.aspx>) construit par Gui Shichun et Yang Huizhong (桂诗春, 杨惠中), CEEC (The Corpora of English Education in China) par He Anping (何安平), etc.

11 Cf. Liu Kanglong & Mu Lei (刘康龙, 穆雷). 2006. Corpus Linguistics and Translation Studies (语料库语言学与翻译研究). in *Chinese Translators Journal* (中国翻译). Beijing. Vol.27 N° 1, p59-64.