

# Analyse textométrique de corpus parallèles français-chinois : « l'Aube » de Jean-Christophe et ses trois traductions chinoises<sup>1</sup>

Jun Miao

Paris III – l'ESIT - 75 016 Paris – France – silaomiao@hotmail.com

## Abstract

This paper aims at highlighting the possibilities, difficulties, limitations, usefulness and even the necessity of textometric methods in French-Chinese literary translation studies. The present work deals with the translation of a French novel, *Jean-Christophe* by Romain Rolland. This book has been translated into Chinese by three different well-known translators. We describe the particular problems which arise in the construction of the quadri-text corpora French-Chinese1-Chinese2-Chinese3 of the first volume "l'Aube". Using the computer programs *Lexico 3* and *mkAlign*, we then present some results in the analysis of these translations that would be difficult to obtain in classical translation studies: The comparison of French and Chinese as well as the comparison of the three translations yield quantitative data that either oppose or strengthen a number of classical hypotheses on French-Chinese translation. The style and the epoch of each translator clearly appear in the textometrical analysis results that we present, thus demonstrating the inescapable contribution of textometric approaches to modern and scientific translation studies.

**Keywords:** textometrical analysis, translation studies, parallel corpora French-Chinese, French literature of the 20<sup>th</sup> century, Chinese translation.

## Résumé

Le but de cette communication est de montrer les possibilités, les difficultés, les limites, l'utilité et même la nécessité des méthodes textométriques pour la recherche traductologique des textes littéraires français-chinois. Le travail porte sur la traduction en chinois d'une oeuvre littéraire française, *Jean-Christophe* de Romain Rolland. Le livre a été traduit en chinois par trois traducteurs célèbres et nous décrivons les particularités et difficultés de la construction du quadri-texte de corpus parallèle français-chinois1-chinois2-chinois3 pour le premier volume « l'Aube ». A l'aide des logiciels *Lexico 3* et *mkAlign*, nous montrons ensuite certains résultats traductologiques qui ne se dégagent que très difficilement dans une approche traductologique classique : aussi bien dans la comparaison du français avec le chinois que dans la comparaison des trois traductions chinoises. On découvre des données quantitatives, qui parfois contredisent, parfois corroborent des hypothèses classiques sur la traduction français-chinois. Le style et l'époque de chaque traducteur apparaissent dans l'analyse textométrique que nous présentons, montrant ainsi le statut incontournable de la textométrie dans une traductologie moderne et scientifique.

**Mots-clés :** analyse textométrique, traductologie, corpus parallèle français-chinois, littérature française du 20<sup>e</sup> siècle, traduction chinoise.

---

<sup>1</sup> L'auteur remercie sincèrement André Salem, Serge Fleury, Kim Gerdes, Paul Plombard de leur aide généreuse et de leur conseil précieux dans ma réalisation de cet article.

## 1. Introduction

Avec le développement d'Internet et l'augmentation de textes disponibles sous forme électronique, les recherches basées sur l'étude du corpus rencontrent de nombreux succès en linguistique, en sociolinguistique et dans l'enseignement des langues étrangères. Les méthodes assistées par ordinateur se diversifient et s'améliorent sans cesse. Quant à la traductologie, Mona Baker<sup>2</sup> y a introduit pour la première fois en 1993 les nouvelles méthodes de la linguistique de corpus. Mais l'adaptation aux besoins de la recherche traductologique requiert un processus long, fastidieux et toujours en cours, avant que ces méthodes ne puissent devenir un standard en traductologie.

La recherche basée sur corpus électronique apporte de nouvelles méthodes en traductologie : on peut analyser statistiquement une grande quantité de matériel et ainsi examiner des hypothèses de « caractéristiques universelles de traduction » (« *universal features of translation* », Baker<sup>3</sup>).

Dans le présent travail, au lieu de nous laisser guider par les idées et impressions survenues lors de la lecture de l'original et des traductions, nous recourons à des méthodes textométriques comme « outillage d'intuition reproductible ». Puis nous interpréterons ces mesures textométriques, parfois en vérifiant certains résultats avec les mêmes outils, parfois en utilisant nos connaissances linguistiques ou littéraires. Nous obtenons ainsi quelques réponses préliminaires sur les spécificités de chaque traduction et sur les choix faits par les traducteurs.

Notre corpus parallèle est constitué du volume « l'Aube » du roman *Jean-Christophe* de Romain Rolland<sup>4</sup> et de ses trois traductions chinoises<sup>5</sup> faites respectivement par Fu Lei (傅雷), Han Hulin (韩沪麟) et Xu Yuanchong (许渊冲). Selon une recherche que nous avons effectuée récemment sur Internet, les corpus électroniques en chinois disponibles sont encore très restreints par rapport aux corpus dans les grandes langues occidentales, bien que l'on constate un accroissement important de l'intérêt qui leur est porté. Pour l'instant, il n'existe pas de corpus parallèle français-chinois et le corpus parallèle utilisé dans cette étude a été construit par nous même.

Le chinois est un système d'écriture idéographique, pour lequel l'informatisation s'est révélée beaucoup plus complexe que celle des systèmes alphabétiques. Ceci a freiné les avancées des méthodes de corpus assistées par ordinateur dans le cadre chinois.

---

<sup>2</sup> Baker. M. (1993). "Corpus Linguistics and Translation Studies: Implications and Applications", *Text and Technology: In Honour of John Sinclair*, Baker, Francis and Tognini-Bonelli (Eds), Amsterdam/Philadelphia, John Benjamins, pp. 233-250.

<sup>3</sup> Ibid.

<sup>4</sup> Romain Rolland (1866-1944) reçut en 1916 le prix Nobel de littérature de 1915 pour son roman fleuve en 10 volumes. Sa version électronique est disponible sur Internet au <http://www.ebooksgratuits.com/ebooks.php>.

<sup>5</sup> La première traduction de Fu Lei (1908-1966) était seulement le premier volume de *Jean-Christophe* qui est parue en 1937, mais il a retraduit le texte entier dans les années 50. La version électronique de la traduction de Fu Lei que nous utilisons est au <http://www.yifan.net/yihe/novels/foreign/yhklksdf/klksdf.html>. La traduction de Han Hulin (1939-) a été publiée par les Éditions Yi Lin (译林出版社), 2000. Et la traduction de Xu Yuanchong (1921-) a paru dans la même année, aux Éditions Littéraires du Hunan (湖南文艺出版社). N'ayons accès qu'aux deux dernières traductions de l'édition sur papier, nous avons dû scanner, reconnaître (en utilisant logiciel de reconnaissance de caractères chinois *Presto OCR*) et laborieusement corriger ces deux livres, avant de pouvoir faire cette étude sur les corpus électroniques.

Les espaces utilisés dans les langues occidentales donnent l'impression que la notion de « mot » est relativement facile à établir, bien qu'il existe toujours des ambiguïtés sur cette notion que montrent des exemples comme « parce que », « pomme de terre », « aujourd'hui »... ou encore les différentes notations des pronoms clitiques en espagnol (collés) et en français (séparés par des espaces). Pour le chinois la difficulté d'établir la notion de « mot » est beaucoup apparente, puisque l'écriture chinoise ne marque pas le blanc entre les « mots » chinois (ou plutôt les caractères) : tous sont écrits à la suite, c'est une *scriptura continua*. Ce genre d'écriture rend l'objet « mot » beaucoup moins intuitif et palpable qu'en langues à écriture « découpée ». Les discussions sur la définition du « mot » ainsi que les techniques de segmentation automatique des textes chinois en mots ne se sont développées et améliorées que récemment.

Il n'est donc pas étonnant de voir que les méthodes d'analyses textométriques, qui fleurissent en Occident depuis une vingtaine d'années, ne se sont pas beaucoup développées et appliquées aux textes chinois. On ne trouve que très peu de travaux en traductologie chinoise (Liu et Mu, 2006), surtout à partir de l'exploration concrète, qui se basent sur les méthodes textométriques. Pour le couple français-chinois, cette présente étude est à notre connaissance la première.

En plus des corpus électroniques « bruts », pour le travail textométrique il faut disposer d'outils électroniques adaptés aux langues de l'étude : pour le prétraitement (comme le nettoyage, la correction orthographique et le découpage en mots) et pour l'alignement automatique (ou assisté par ordinateur) des textes avec leurs traductions, et finalement pour l'extraction et l'exploration des données pluri-textuelles (Zimina, 2005) : Comme outillage pour l'exploration textométrique, nous utilisons les logiciels *Lexico3*<sup>6</sup> et *mkAlign*<sup>7</sup> qui sont applicables pour le texte chinois.

## 2. Préparation des textes bruts

Dans le but de dégager plus facilement les informations du texte original français, nous avons changé toutes les majuscules en minuscules. Pour les textes chinois, nous avons choisi le logiciel de segmentation du chinois ICTCLAS<sup>8</sup>, pour découper les phrases chinoises en mots. Après correction et nettoyage, le découpage constitue l'étape la plus importante du traitement des corpus chinois.

Bien qu'il existe des similarités visuelles entre la ponctuation chinoise et la ponctuation occidentale, le système de ponctuation chinoise porte quand même la trace de l'écriture

---

<sup>6</sup> *Lexico3*, un logiciel de lexicométrie, est réalisé par l'équipe universitaire SYLED-CLA2T (Systèmes Linguistiques Énonciation et Discours - Centre de Lexicométrie et d'Analyse Automatique des Textes), fondé en 1997. Ce logiciel a d'abord été développé par André Salem, professeur à l'Université Paris III. Voir le site : <http://www.cavi.univ-paris3.fr/Ilpga/ilpga/tal/lexicoWWW/lexico3.htm>.

<sup>7</sup> *MkAlign* est conçu et développé par Serge Fleury, maître de conférence à Paris III. Ce programme permet de construire, corriger et visualiser un alignement de deux textes via un éditeur à double entrée. Il permet d'afficher simultanément les textes source et cible pour y rajouter ou corriger des segments équivalents : <http://tal.univ-paris3.fr/mkAlign/mkAlignDOC.htm>.

<sup>8</sup> *Institute of Computing Technology, Chinese Lexical Analysis System*, ce logiciel est en développement depuis 2002, disponible au [http://www.nlp.org.cn/project/project.php?proj\\_id=6](http://www.nlp.org.cn/project/project.php?proj_id=6). En effet, il existe plusieurs logiciels de segmentation du chinois, et on peut les trouver sur Internet, comme *Hailanda Segmentation intelligente, version d'essai* (海量智能分词研究版) et *Lucene*. Mais actuellement, aucun logiciel ne donne des résultats parfaits, puisque la technique est en état d'évolution.

chinoise et ses symboles (et les codes informatiques associés) restent donc différents de la ponctuation occidentale. Dans notre essai, nous avons transformé la plupart de ponctuations chinoises en ponctuations occidentales dans le but de lancer une comparaison entre l'original et ses traductions. Cependant, comme certains signes de ponctuation chinoise n'ont pas de correspondants dans le système de ponctuation occidentale, nous les avons gardés tels quels, par exemple «、» (顿号, *dunhao*, la demi-virgule<sup>9</sup>).

Étant donné que notre travail porte sur un corpus parallèle, nous avons besoins d'aligner le texte original et ses traductions chinoises. Pour faciliter ce travail et pour rendre les résultats plus cohérents, nous avons recours à un logiciel d'alignement automatique de textes bilingues. Cette tâche a été réalisée à l'aide de l'*Alignator*<sup>10</sup>. Après avoir bien aligné le texte original et la traduction, on obtient des textes où chaque groupe aligné de paragraphes est signalé par un symbole (par exemple « # ») comme séparateur.

En dernière étape, il faut sauvegarder tous les textes en format *.txt* en utilisant différents encodages, car ces logiciels n'utilisent pas encore tous Unicode (nous utilisons GB2312, Latin-1 et UTF-8).

### 3. Explorations et résultats

Dans cette partie, nous allons démarrer notre expérience textométrique proprement dite. En notant ce que nous observons et commentant certains résultats thématiques, nous mettons en parallèle la théorie de la traduction et les faits obtenus par l'analyse quantitative.

#### 3.1. Informations textométriques générales du corpus parallèle français-chinois

Dans le but d'avoir une vue d'ensemble sur le corpus parallèle français-chinois, nous avons mis l'original et les trois traductions chinoises dans un même fichier. *Lexico3* nous affiche les caractéristiques textométriques de nos corpus selon plusieurs aspects.

##### 3.1.1. Principales caractéristiques du corpus

Num	Partie	Occurenc	Formes	Hapax	Fmax	Forme
✓ 1	chinois1	38458	5567	2759	2847	的
✓ 2	chinois2	41117	6824	3627	2297	他
✓ 3	chinois3	39812	5721	2907	2100	的
✓ 4	français	41650	5938	3405	1886	il

Tableau 1 : Principales caractéristiques de l'*Aube de Jean-Christophe* et ses trois traductions chinoises

Il est facile de voir, à partir du tableau 1, que le nombre total de mots (occurrences) est plus important pour le texte français (41 650 contre 38 458, 41 117 et

39 812). On note aussi que les nombres de formes des mots dans les traductions chinoises 1 et 3 (Fu Lei, 傅雷 et Xu Yuanhong, 许渊冲) sont moins élevés que dans le texte français, cependant c'est la traduction chinoise 2 (Han Hulin, 韩沪麟) qui possède les formes les plus variées (6 824). Le nombre d'hapax (les formes qui n'apparaissent qu'une seule fois dans un texte) est le moins élevé dans la traduction chinoise 1 (2 759). Il est intéressant de constater que le nombre de hapax de la traduction chinoise 2 (3 627) dépasse celui du texte français.

<sup>9</sup> Cette ponctuation est utilisée pour exprimer une petite pause entre deux ou plusieurs éléments lors d'une énumération dans une phrase.

<sup>10</sup> Ce programme, conçu par Kim Gerdes, maître de conférence à Paris III, permet l'alignement automatique des paragraphes, basé sur leur longueur, et la correction à la main des paragraphes mis en correspondance.

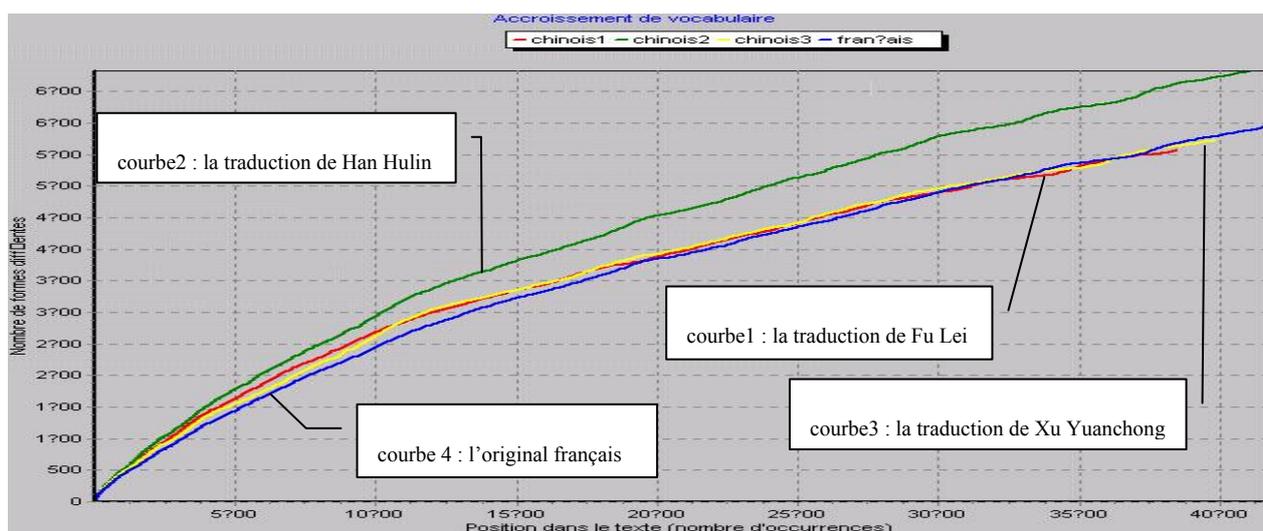
« 的 » (en pinyin *de*), l'équivalent approximatif du français « de » (ou plutôt de l'anglais « 's » – la marque du génitif), est le mot le plus fréquent dans les traductions chinoises 1 et 3, mais la traduction chinoise 1 possède plus de ce mot (Fmax 2 847) que la traduction 3 (Fmax 2 100). Le mot chinois « 他 » (*ta*), correspondant plus ou moins au français « il », est la forme la plus fréquente dans la traduction 2. On constate également que « il » est le mot le plus fréquent dans le texte français. Cependant, le mot « 他 » (*ta*, *il*) en chinois dans la traduction chinoises 3 est plus nombreux que le « il » en français dans l'original.

Le tableau 1 nous offre une vue d'ensemble des informations sur les mots de l'original et de ses trois traductions chinoises. Nous pouvons en tirer une idée sur nos corpus : l'original est plus long que ses traductions chinoises en nombre de mots. Fu Lei a utilisé le moins de mots (38 458 mots) et le moins de hapax (2759,  $38\,458/2\,759 = 13,93$  mots par hapax), par contre, l'emploi d'unités de mot différentes pour sa traduction (38 458 mots divisés par 5 567 formes = 6,91 mots en moyenne par nouveau mot) occupe une position moyenne parmi les trois traductions.

La traduction chinoise 2 est la plus longue (41 117 mots) au niveau des occurrences et aussi la plus riche au niveau des formes de mot (6 824 formes différentes, ce qui fait  $41\,117/6\,824 = 6,03$  mots en moyenne par nouveau mot), et au niveau de nombre d'hapax (3627,  $41\,117/3\,627 = 11,34$  mots par hapax). Mais il faut insister simplement sur le fait que la différence de longueur de la traduction 2 par rapport à l'original est considérable : elle utilise 2 659 mots de plus pour rendre le texte, ce qui fait une augmentation de 7%.

La traduction 3 de Xu Yuanchong se trouve en position intermédiaire par rapport aux deux traductions précédentes en ce qui concerne la longueur du texte (39 812 mots et 5 721 formes différentes) et des hapax ( $39\,812/2\,907 = 13,7$  mots par hapax). Mais c'est elle qui utilise le moins de formes différentes par rapport à sa longueur :  $39\,812/5\,721 = 6,96$  mots en moyenne par nouveau mot. Elle est donc le texte le plus « homogène » du côté de son vocabulaire.

### 3.1.2. Accroissement de vocabulaire de chaque partition du corpus



**Figure 1** : Accroissement de vocabulaire de *l'Aube de Jean-Christophe* et ses trois traductions chinoises

La figure 1, qui rend compte de l'apparition de nouveaux mots au fur et à mesure que l'on parcourt le texte, nous permet d'observer les mots que le traducteur utilise au cours du processus de traduction. D'emblée, il est évident que les courbes 1 (correspondant à la

traduction de Fu Lei), 3 (la traduction de Xu Yuanchong), et 4 (l'original français) sont très similaires, en se superposant aux nombreux endroits. La seule courbe particulière est la courbe 2 (la traduction de Han Hulin), qui se situe nettement au-dessus des autres courbes, et qui est donc aussi celle qui est la plus éloignée de la courbe 4 de l'original.

On s'attendrait normalement à ce que le français, une langue flexionnelle dans laquelle les lemmes (« mots ») changent de forme selon leur rapport grammatical aux autres lemmes, possède plus de formes dans son texte que dans le même texte écrit en chinois, qui est une langue typiquement isolante. Mais d'après les résultats de la figure 1, il est étonnant de voir que la courbe 4 partage beaucoup de similarités avec les courbes 1 et 3, en plus, elle est au dessous de ces deux courbes dès le début jusqu'au moment où il y a environ 20 000 mots dans le texte. Ceci indique que pendant cette période, la vitesse d'accroissement du vocabulaire du texte français est moins rapide que celle des traductions chinoises.

Nous avons également étudié la forme des courbes. Bien que les courbes 1 et 3 soient plus proches de la courbe 4, la forme de la courbe 2 est par contre plus similaire à celle de la courbe 4. Remarquons les bosses de la courbe 4 qui apparaissent en 11 000, 20 000, 30 000, 350 000 et 380 000 sur l'axe X, la courbe 2 suivit presque la même trace.

A partir de nos analyses, certaines hypothèses sont émises à propos de nos trois traductions chinoises : la traduction 1 garde l'ensemble du contenu de l'original, mais elle a tendance à simplifier certaines choses ; par contre, le traducteur de la traduction 2, Han Hulin, adopte une manière explicite et un style assez littéral pour sa traduction ; quant à la traduction 3 de Xu Yuanchong, elle partage de nombreuses similarités avec la traduction 1. Xu Yuanchong aurait-il utilisé de la traduction de Fu Lei comme référence quand il a lui-même traduit cette oeuvre ? Mais ce ne sont justement que nos hypothèses, qu'il faudra vérifier ultérieurement en fournissant des indices concrets.

### 3.1.3. Spécificités des mots dans la traduction

*Lexico3* nous permet également d'extraire les mots spécifiques de chaque texte. Comme nous désirons respecter certaines limites éditoriales, nous n'entrons pas ici dans le détail des spécificités des mots de chaque traduction chinoise, nous soulignerons simplement quelques points intéressants relevés à partir des listes de mots les plus spécifiques et les plus non-spécifiques de la traduction de Fu Lei par rapport à deux autres traductions.

Formes/SR	chinois1	chinois2	chinois3
象	23	*** 2	-21 1
的	2847	+43 2233	-11 2100
曼希沃	82	+41 0	0
菁	678	+41 440	-2 253
鲁意莎	50	+26 0	0
跟	78	+22 8	-10 13
可是	104	+20 46	4
克利斯朵夫	266	+19 260	+13 0
时候	107	+18 15	-14 42
约翰·米希尔	31	+16 0	0
那些	53	+14 10	-5 7
望	49	+13 7	-6 10
屋子	28	+12 0	3
高脱弗烈特	23	+12 0	0
地下	24	+12 0	1
人家	70	+11 6	-14 36
一忽儿	21	+11 0	0

**Tableau 2 :** Les mots les plus spécifiques de la traduction de Fu Lei

Formes/SR	chinois1	chinois2	chinois3
地	9	*** 365	+36 246
横	3	-36 110	+5 124
、	35	-24 191	+14 130
时	15	-22 167	+31 51
却	25	-12 47	-5 133
感到	3	-11 48	+5 35
就	146	-10 207	-3 320
房间	1	-7 25	+4 17
更加	2	-7 28	+4 22
一直	1	-6 12	26
已	4	-6 45	+13 7
这	139	-6 231	+4 201
不过	3	-6 23	27
看来	1	-6 12	24
曲子	6	-5 25	26
向	33	-5 98	+11 37
他	1856	-5 2297	+6 2046

**Tableau 3 :** Les mots les plus non-spécifiques de la traduction de Fu Lei

En fournissant des informations inaccessibles par la lecture ordinaire, les tableaux 2 et 3 montrent leur grande utilité pour les recherches traductologiques. Prenons le premier mot affiché dans le tableau 2 comme exemple, nous voyons que le mot « 象 » (*xiang*) est le mot le plus spécifique de la traduction Fu Lei, avec 123 apparitions. En revanche, on ne le trouve que 2 fois dans la traduction 2 et 1 fois dans la traduction 3. Bizarrement, dans le tableau 3, « 像 » (*xiang*), au deuxième rang du tableau 3, apparaît très souvent dans les deux dernières traductions (avec 110 fois, spécificité + 5, et 124 fois, spécificité +10), mais très peu dans la traduction de Fu Lei (3 fois d'apparition, spécificité -36).

Avec une lecture attentive de trois traductions, nous nous rendons compte qu'il s'agit de l'emploi du vocabulaire dans le cadre de l'évolution du chinois. En effet, « 象 » (*xiang*), avec les sens « éléphant », « l'apparence des objets », « imitation » et « 像 » (*xiang*), qui véhicule « l'image faite par l'homme », « ressembler », « comme », étaient deux mots différents bien qu'ils se prononcent identiquement. Mais à l'époque de Fu Lei, dans les années 50-60, on cherchait à simplifier les caractères chinois, et « 象 » (*xiang*) a été utilisé comme le caractère simplifié de « 像 » (*xiang*)<sup>11</sup>. Mais vingt ans après, sur les conseils des linguistes du chinois, le Conseil des Ministres de la Chine a publié en 1986 le tableau des « éléments de caractères » modifié, dans lequel on a annulé l'emploi de « 象 » en remplacement de « 像 ». De même, nous relevons aussi que le deuxième mot le plus spécifique « 的 » (*de*, équivalent à *de*, qui est utilisé pour : déterminant + 的 + nom) de Fu Lei et son mot le plus non-spécifique « 地 » (*de*, équivalent à *de*, qui est utilisé pour : adverbe + 地 + verbe) ont le même problème. Ainsi, de cette manière, le logiciel informatique nous aide à repérer plus facilement la diachronie des graphismes dans les traductions de différentes époques.

Il suffit d'exploiter attentivement les tables ci-dessus pour en tirer des informations utiles à l'approfondissement de notre recherche traductologique. Par exemple, au 7<sup>e</sup> rang du tableau 2, « 可是 » (*keshi*, mais) est un mot très spécifique de la traduction de Fu Lei avec 104 apparitions, spécificité +20, mais ce mot apparaît seulement 46 fois dans la traduction de Han Hulin et 4 fois dans la traduction de Xu Yuanchong. Comment les deux derniers traducteurs ont-ils traduit cette conjonction ? Cette interrogation va forcément devenir un point intéressant pour notre prochaine étude. Un autre exemple, « 一忽儿 » (*yihui'er*, un peu plus tard) est un mot du dialecte du nord, avec 20 apparitions dans la traduction de Fu Lei, mais aucune apparition chez les deux autres traducteurs. Nous ne pouvons nous empêcher de supposer que Fu Lei avait une préférence pour certaines formes régionales du chinois dans sa traduction. Nous avons bien conscience de ce que nos analyses dans les paragraphes précédents sur les tableaux 2 et 3 restent encore très grossières et simples, mais ce qui est important, c'est que les tableaux obtenus grâce à ce logiciel peuvent nous fournir quantité d'indices utiles dans notre future recherche traductologique.

### 3.2. Analyses textométriques par terme

La fonction « groupe des formes de mots » du logiciel (en anglais *Generalized Types or Tgen(s)*) nous a rendu un grand service (Lamalle et Salem, 2002<sup>12</sup>) : les groupes de formes

<sup>11</sup> Dans le tableau des « éléments de caractères » pour la Réforme de l'Écriture publié par le Conseil des Ministres de la République Populaire de Chine en 1964, « 象 » a été défini comme le caractère simplifié pour « 像 ». Mais il y avait une note disant qu'en cas de confusion possible, on pourrait utiliser quand même le caractère « 像 ». Voir : <http://zhidao.baidu.com/question/27456439.html>

<sup>12</sup> Cf. S. Fleury. et M. Zimina. *Exploring Translation Corpora with mkAlign*, disponible au <http://www.translationdirectory.com/articles/article1263.htm> consulté le 17 octobre 2007.

sont des unités textuelles définies par l'utilisateur à l'aide d'outils automatiques, qui permettent de regrouper certains événements dans le texte. Dans la partie suivante, nous comptons introduire une étude thématique sur notre corpus en nous appuyant sur la fonction « groupe des formes » dans *Lexico3* et sur une confirmation par l'affichage simultané du texte source et cible dans *mkAlign*.

### 3.2.1. Pronoms personnels

Chaque langue possède un système de pronoms personnels, mais leur utilisation est différente selon la grammaire et les usages d'une langue à l'autre. Après avoir groupé respectivement les pronoms du texte français et ceux de trois traductions chinoises, nous introduisons les résultats de groupes de formes dans les listes des spécificités des mots de *Lexico3*, et nous obtenons les tableaux suivants.

Formes/SR	français
il	1886
lui	453
ils	156
eux	30
elle	225
elles	30
je	80
nous	8
tu	106
vous	27

**Tableau 4** : Spécificité des pronoms personnels français

Formes/SR	chinois1	chinois2	chinois3
他	1856	-5 2297	+6 2046
他们	165	151	164
她	159	185	178
她们	1	-3 3	15 +5
我	127	128	154 +2
我们	10	10	13
你	141	132	157
你们	1	2	1
您	28	+6 13	2 -6

**Tableau 5** : Spécificités des pronoms personnels chinois dans les trois traductions chinoises

Avant d'interpréter les informations des tableaux 4 et 5, nous voudrions parler brièvement des différences entre les systèmes de pronoms du français et du chinois. Nous savons que le français utilise différentes formes du pronom personnel de la 3<sup>e</sup> personne pour le sujet (« il » et « elle »), l'objet direct (« le » et « la ») et l'objet indirect (« lui »).

En chinois, par contre, il existe une distinction de pronom par genre, mais pas par fonction grammaticale dans la phrase. C'est-à-dire que pour le pronom personnel de la 3<sup>e</sup> personne on a « 他 » (*ta*) pour il, et « 她 » (*ta*) pour « elle », mais « 他 » (*il*) et « 她 » (*elle*) ne changent pas de forme selon que leur fonction est celle de l'objet direct ou de l'objet indirect. De plus, le mot « vous » en français est utilisé à la fois comme forme de politesse du pronom de la 2<sup>e</sup> personne au singulier et comme pronom de la 2<sup>e</sup> personne au pluriel. Mais en chinois il s'agit de deux mots distincts : « 您 » (*nin*) qui est la forme de politesse, et « 你们 » (*nimen*) qui est le pronom personnel de la 2<sup>e</sup> personne du pluriel.

Une rapide comparaison entre les tableaux 4 et 5 montre que la fréquence de « 他 » (*ta*) dans les traductions chinoises 2 (2 297 fois, spécificité +6) et 3 (2 046 fois) est beaucoup plus élevée que dans l'original français (1 886 fois). La raison, comme nous l'avons expliqué dans le paragraphe précédent, est que « 他 » (*ta*) correspond à la fois à « il » et à « lui » (partiel). Le résultat de « 他 » dans la traduction 2 avec 2 297 apparitions confirme le nombre de formes maximal (2 297) pour « 他 » obtenu dans le tableau 1 cité plus haut. Il est nécessaire de signaler que le mot chinois « 他 » (*ta*) est beaucoup plus utilisé dans la traduction 2 que dans la traduction 3 (2 297 – 2 046 = 251). Mais ce qui est intéressant, c'est de noter que dans la traduction 1, « 他 » (1 858 fois, spécificité -5) apparaît moins souvent que « il » en français. Nous pensons que « 他 » (*ta*) est un des indicateurs de la caractéristique de la traduction, car dans la composition chinoise, de manière générale, on préfère privilégier la concision en évitant les répétitions des pronoms personnels.

Examinons une citation originale et ses trois traductions chinoises glosées, en regard de notre traduction directe pour que les francophones puissent comparer leurs styles de traduction.

Dans l'original de cet exemple, « il » comme pronom personnel sujet apparaît trois fois, et comme pronom indéfini une fois. Dans les traductions de Fu Lei et de Xu Yuanchong « 他 » (*il*) n'est traduit qu'une fois pour « il ne dînait pas », et pour les verbes suivants comme « aller » et « croire », ils ont tous les deux omis le sujet, et les traductions de ces actions partagent donc le même sujet de « dîner ». Regardons la traduction 2 de Han Hulin ; il a traduit tous les pronom sujet français « il » en chinois « 他 » (*il*). Ce genre de traduction n'est pas faux, mais simplement, lorsqu'on lit le texte, on a l'impression que la mélodie n'est pas jolie et le style n'est pas concis, car il y a trop de répétition du même ton. Ce petit exemple  
Ex : *le jour venu, il ne dînait pas, il s'agitait comme une âme en peine, il allait regarder cinquante fois l'horloge, il croyait que le soir n'arriverait jamais ;* (Suf. = Suffixe, nég. = négation, Cl. = Classificateur)

**Chinois 1 :** 到了那天, 他吃不下饭, 好象担着重大的心事, 骚乱不堪, 跑去  
arriver-Suf. le jour, il manger-nég. descendre repas, comme s'inquiéter-Suf. grand de souci, trembler infini, courir aller  
对时钟看了几十次, 以为天不会黑的了。  
vérifier horloge regarder-Suf. plusieurs-dix fois, croire ciel nég. possible noir -Suf.

*(Le jour venu, il ne mange plus son repas, comme s'il était accablé d'énorme de soucis, tremblant sans cesse, en courant aller vérifier des dizaines fois l'horloge, et en croyant qu'il ne fasse pas nuit.)*

**Chinois 2 :** 到演出那天, 他连晚饭也不吃, 像灵魂受到煎熬似的躁动不安。他一次  
arriver présentation le jour, il même dîner aussi nég. manger, comme âme subi souffrance comme trembler. il une fois  
又一次跑去看钟。他以为天永远不会黑下来；  
encore une fois courir aller regarder horloge. il croire ciel jamais nég. possible noir détendre ;

*(Le jour de présentation venu, il n'arrive même pas à manger le dîner, comme si son âme en peine, tout en tremblant, il court regarder une fois encore une fois l'horloge, il croit qu'il ne fasse jamais nuit ;)*

**Chinois 3 :** 演出的日子一到, 他连晚餐都吃不下, 激动得像受苦受难的  
présentation de jour une arriver, il même dîner aussi manger nég. descendre, excité jusqu'à comme subi souffrance de  
鬼魂一样；一个小时要看五十次钟, 以为天永远不会黑了；  
mânes pareil ; une-Cl. heure vouloir regarder cinquante fois horloge, croire ciel jamais nég. possible noir Suf.;

*(Le jour de présentation venu, il n'arrive même pas à manger son dîner, excité comme des mânes de souffrance ; en regardant l'horloge cinquante fois pendant une heure, et croyant qu'il ne fasse jamais nuit ;)*

illustre aussi pourquoi Han Hulin utilise plus de mots chinois pour rendre le même texte en chinois (voir le tableau 1 et la figure 1).

La citation concrète nous donne aussi l'occasion d'amener un examen « qualitatif » des traductions. Au niveau de contenu, les trois traducteurs ont tous traduit l'ensemble des informations contenues dans l'original. Au niveau de style, tous les trois ont adopté style plus ou moins littéraire. Le style de la traduction de Fu Lei et celui de la traduction de Han Hulin sont assez similaires, bien que la dernière semble un peu lourde dans l'emploi du pronom personnel. D'un point de vue général, ces deux traductions sont plutôt proches du style de l'original. Il nous semble que Xu Yuanchong a un style de traduction un peu exagéré, en ajoutant « excité », « pendant une heure », et en traduisant une « âme » en « des mânes ».

Ce qui est curieux c'est que les hypothèses que nous avons formulées d'après les courbes d'accroissement du vocabulaire ne sont pas tout à fait justes, à savoir que Fu Lei et Xu Yuanchong partagent beaucoup de similarités de traduction, et que Han Hulin a adopté la façon plus littérale pour sa traduction. Comme le montre l'exemple précédent, nous pouvons simplement en tirer la conclusion que Han Hulin colle plutôt à la syntaxe française dans sa traduction. Évidemment, nous avons conscience qu'une conclusion définitive est prématurée et qu'il reste bien des points à approfondir dans nos recherches à venir. De tout ce qui précède, nous voyons qu'une pure analyse statistique (textométrique) est loin d'être suffisante

et qu'il faut vérifier abondamment les résultats de l'analyse textométrique et les hypothèses dans le contexte.

### 3.2.2. Expressions figées (idiomatiques) chinoises

Les expressions figées chinoises, dites *chengyu* (成语, « *set phrase* » en anglais), sont une catégorie de tournures idiomatiques. Composées généralement de quatre caractères (il y en a aussi de trois caractères ou de cinq caractères...), ces expressions tirées des classiques, d'anecdotes ou de légendes folkloriques, sont très liées à la culture chinoise. L'emploi des expressions figées chinoises dans les traductions des oeuvres étrangères, qui reflète fortement les choix et la technique de traduction du traducteur, constitue toujours un point important dans la traductologie.

*Lexico3* nous permet de scruter sans peine l'utilisation des expressions figées chinoises dans les trois traductions chinoises. Après avoir entré «.....» (chaque caractère chinois occupant 2 octets) dans les groupes des formes en accrochant « [c'est] une expression rationnelle », nous obtenons une liste des expressions figées chinoises, dans laquelle la fréquence de chaque expression est aussi indiquée à sa droite. Cependant, nous remarquons que certaines chaînes de quatre caractères collés ensemble qui ne sont pas des expressions figées entrent aussi dans la liste, telles que « 小提琴手 » (*violoniste*), et « 这就是说 » (*c'est-à-dire*), parce que l'ordinateur ne prend en compte que la forme graphique des mots. Nous obtenons semi-automatiquement la liste des expressions figées chinoises utilisées dans chaque traduction chinoise en enlevant les mots à quatre caractères qui ne sont pas des *chengyu*. Ceci montre que bien que logiciel soit conçu pour faciliter la recherche, la vérification et la correction manuelles sont quand même nécessaires (Olohan, 2004 : 71).

La localisation de ces termes dans la carte du texte rend possible de sonder la distribution de l'emploi des expressions figées chinoises dans chaque traduction.

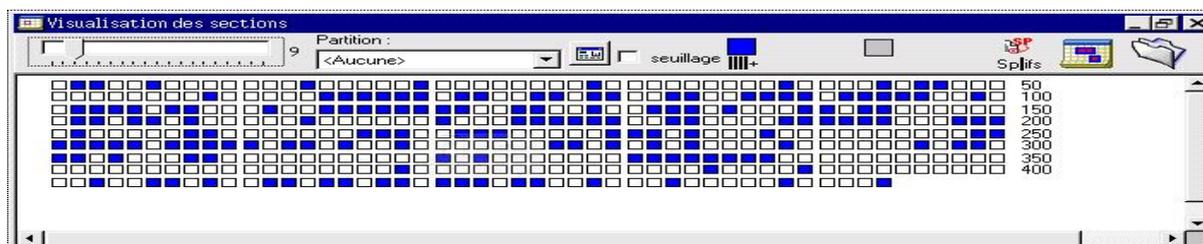


Figure 2 : Localisation des expressions figées chinoises dans la traduction de Fu Lei

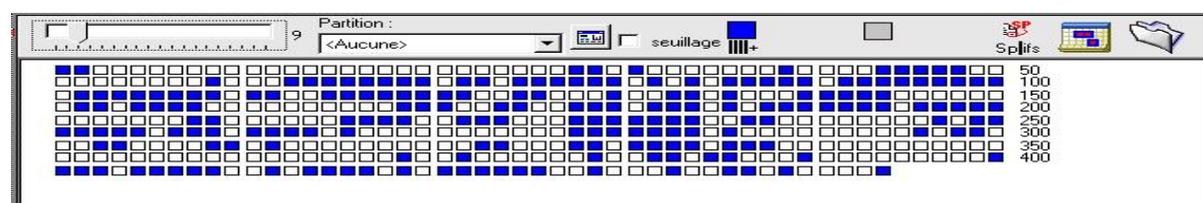


Figure 3 : Localisation des expressions figées chinoises dans la traduction de Han Hulin

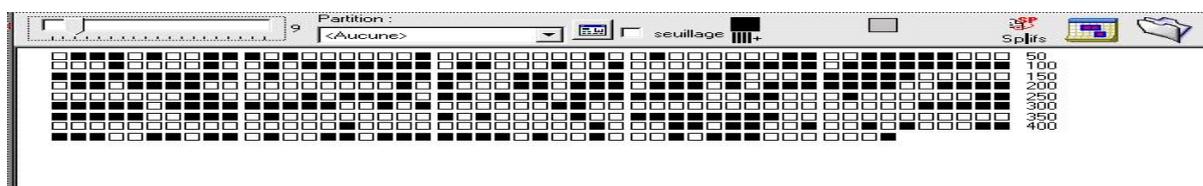


Figure 4 : Localisation des expressions figées chinoises dans la traduction de Xu Yuanchong

Selon les figures au-dessus 2, 3 et 4, nous remarquons que la fréquence d'emploi des expressions chinoises dans la traduction de Fu Lei est beaucoup plus faible que dans les deux autres traductions, mais la dispersion en est plus uniforme. Dans les traductions de Han Hulin et de Xu Yuanchong, l'utilisation des expressions figées chinoises se concentre au milieu du texte.

Nous pouvons d'ailleurs nous interroger *sur* les spécificités des expressions chinoises dans chaque traduction chinoise à l'aide de *Lexico3*. Pour mieux savoir quelles phrases ou mots dans l'original ont été traduits en expressions figées chinoises, nous avons recours à *mkAlign* qui affiche simultanément les expressions que nous cherchons dans les volets source et cible.

Après une comparaison attentive entre l'original les trois traductions, dont nous ne présentons pas ici tous les détails, nous tirons la conclusion suivante : Fu Lei utilise moins d'expressions figées chinoises dans sa traduction, mais il a réussi à garder le plus possible les expressions de l'original. Sa traduction semble plus fidèle à l'original. Han Hulin est celui qui emploie le plus d'expressions chinoises dans sa traduction, et il semble ainsi que sa traduction contient beaucoup d'explications et que son style s'enracine dans la tradition culturelle chinoise. Quant à la traduction de Xu Yuanchong, elle présente une fréquence d'emploi des expressions figées chinoises intermédiaire entre celles des deux autres traductions, mais parfois ces emplois sont un peu décalés, ce qui donne un effet un peu étrange.

### 3.2.3. Traductions de mot-clé de l'original : « Rhin »

Le « Rhin » étant un des objets les plus importants dans l'original, il constitue d'un côté l'endroit où vivait Jean-Christophe, d'un autre le symbole de vie du héros sans cesse en mouvement. Il nous paraît indispensable d'étudier comment les traducteurs chinois rendent le mot « Rhin » en chinois.

Nous utilisons ici *mkAlign* pour lancer notre expérience (dans cette étape, nous utilisons la traduction de Fu Lei). D'abord, une comparaison de l'apparition de « Rhin » entre les listes « spécificités du vocabulaire » des volets source et cible relève que « Rhin » apparaît au total 6 fois dans l'original. « 莱茵河 » (*laiyinhe*, *le Rhin*), la traduction chinoise la plus spécifique pour le « Rhin », se présente 8 fois dans le texte entier, dont 5 fois dans les parties où « Rhin » apparaît dans l'original. Fu Lei a donc traduit à trois endroits d'autres mots que « Rhin » avec le mot « 莱茵河 » (*le Rhin*). L'affichage simultané de l'original et la traduction dans *mkAlign* confirment notre remarque (nous ne la présentons pas ici en raison de la limite éditoriale des pages).

Outre les endroits contenant le « Rhin » dans l'original, Fu Lei a traduit les mots « Vatterrhein » (para.116), « (les pays) rhénans » (para. 117) et le « fleuve » dans la phrase « il s'interrompait pour montrer le poing au fleuve » (para. 98) en « 莱茵河 » (*le Rhin*). Cependant, nous remarquons qu'il y a un carré rouge vide dans le volet cible, qui signifie que Fu Lei n'a pas utilisé le mot « 莱茵河 » (*le Rhin*) pour le « Rhin » dans l'original. En y regardant de plus près, nous constatons que « vin du Rhin » (para. 253) est traduit en « 莱茵美酒 » (*laiyinmeijiu*, vin supérieur du Rhin) au lieu de « 莱茵河酒 » (*laiyinhejiu*, vin du Rhin).

En adoptant la même méthode, nous examinons les traductions de Han Hulin et de Xu Yuanchong au sujet de « Rhin ». Outre sept endroits identiques à ceux de la traduction de Fu Lei, Han Hulin a traduit aussi « (la petite ville) rhénane » (para. 45) en chinois « 莱茵河 » (*le Rhin*). Quant à la traduction de Xu Yuanchong, elle est très similaire à la traduction de

Hanhulin, sauf à un endroit où « Vatterrhein » (para.116) a été traduit en « 莱茵河水 » (*laiyinheshui*, l'eau du Rhin) au lieu de « 莱茵河 » (*laiyinhe*, le Rhin).

L'examen du « Rhin » dans les traductions de Han Hulin et Xu Yuanchong nous amène à nous interroger en retour sur la raison pour laquelle Fu Lei n'a pas traduit « (la petite ville) rhénane » (para. 45) en chinois « 莱茵河 » (*le Rhin*) ? Retournant à la traduction, nous trouvons que Fu Lei a précisé en chinois que cette petite ville se trouve dans la vallée du Rhin « 莱茵流域 » (*laiyinliuyu*). En revanche, quand nous revenons aux traductions de Han Hulin et de Xu Yuanchong pour constater comment ils ont traduit « fleuve » dans « il s'interrompait pour montrer le poing au fleuve », nous constatons qu'ils ont tous les deux traduit le « fleuve » en chinois « 河 » (prononcé *he*, *fleuve*).

De tout ce qui précède, nous sommes d'avis que bien que les trois traductions donnent une première impression de similarité pour traduire le « Rhin », la nuance de traduction existe quand même. Il nous semble que Fu Lei a porté plus d'attention au contexte par rapport à deux autres traducteurs, en traduisant le « fleuve » dans la phrase « il s'interrompait pour montrer le poing au fleuve » (para. 98) en « 莱茵河 » (*le Rhin*). Car selon le contexte, le vieux Krafft se promenait au bord du Rhin, bien que le nom de fleuve ne soit pas mentionné.

#### 4. Conclusions

En 2005, M. Zimina a mis l'accent sur la perspective de la méthode textométrique, en écrivant que la textométrie permet d'accéder à de nouvelles dimensions d'analyse des données de traduction entre des phénomènes langagiers relevant de différents niveaux de l'analyse linguistique.

A travers les résultats et analyses que nous avons exposés, nous avons la conviction que les méthodes textométriques ont un rôle à jouer dans la recherche traductologique. Nous espérons qu'à partir de cet essai, nous pourrions améliorer les méthodes textométriques pour servir à nos recherches traductologiques, à la fois quantitative et qualitative, dans un domaine très littéraire et psychologique où les données « dures » se font habituellement rares.

#### Références

- Granger S., Lerot J. and Peth-Tyson S. (eds.) (2003). *Corpus-based Approaches to Contrastive Linguistics and Translation Studies*. Amsterdam – New York, Editions Rodopi.
- Habert B., Nazarenko A. and Salem A. (1997). *Les linguistiques de corpus*. Paris, Armand Colin/Masson.
- Liu K. L. and Mu L. (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)*. Vol.(27): 59-64.
- Olohan M. (2004). *Introducing Corpora in Translation Studies*. London and New York, Routledge.
- Zanettin F., Bernardini S. and Stewart D. (eds.) (2003). *Corpora in Translator Education*. Manchester, St. Jerome Publishing.
- Zhou Q. and Duan H.M. (周强, 段慧明) (2007). *Traitement de segmentation et de marquage des mots dans le corpus chinois moderne (现代汉语语料库加工中的切词与词性标注处理)*. Disponible au <http://hi.baidu.com/jagard/blog/item/dcdb653844fd842097ddd8ec.html>
- Zimina M. (2005). Topographie bi-textuelle et approches quantitatives de l'extraction de ressources traductionnelles à partir de corpus parallèles. In *Actes des 7es Journées scientifiques du Réseau de chercheurs 'Lexicologie, Terminologie, Traduction'*, Institut supérieur de traducteurs et interprètes (ISTI), Bruxelles.