进入傅雷译作[1]

凯德金（Kim GERDES） [1]　　缪君 [2]
[1] 巴黎三大，新索邦，普通语言学音系学及应用语言学系学院 （ILPGA）
kim.gerdes@univ-paris3.fr
[2] 巴黎三大, 巴黎高等口译笔译学院 （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, corpus, Fu Lei, traductologie, lexicométrie
中文摘要：旨在提供有效捷径进入傅雷译作，本文描述如何运用词量学软件来构建其语料数据库和一个专门的浏览网站，这些软件的设计是符合中文翻译学需要的。我们在本文中展示了构建语料库的步骤，使用工具以及翻译学研究实例。这些新技术、工具为实现从整体上科学认识傅雷的翻译作品作开辟了新的道路。
关键词 ：数据库，语料库，傅雷，翻译学，词量学
Abstract: In order to provide an easier access to the work of Fu Lei, we describe the construction of a corpus and a web site adapted for Chinese translation studies, using lexicometrical tools. We outline the steps and tools used in the construction of this corpus, and we explain, with a concrete 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 Fu Lei's work.
Keywords: corpus, Database, Fu Lei, translation, lexicometrical study

傅雷的翻译作品对于翻译学以及比较文学的研究者来说都是一座巨大而珍贵的宝藏。但是要进入这个宝藏作深入研究，并非是件容易的事，因为傅雷译作多达 33 部，翻译字数总计为 500 余万字。

现有由法国驻中国大使馆承办的“傅雷图书数据库”[2]，可以查看到包括傅雷在内，自十九世纪以来，大陆所出版的法文书籍，但该网站没有提供直接访问这些图目的服务。我们认为，对于学习法语以及法汉翻译，甚至学习其他语种翻译的学生来说，建立真正的傅雷翻译作品数据语料库有着不仅必要而且迫切的需求，因为无须赘述，傅雷的翻译作品在法汉翻译领域，在整个中国翻译领域中，都占有重要的位置。

在本文，我们将根据已建立的部分语料库，来阐述建立傅雷译作数据库的可能性，困难性，我们还将列举实例来展示语料库建设的流程以及其实用性，特别突出词量学（Lexicométrie, 也称为 textométrie）方法对翻译学研究的重大作用。

语料的选择

作为首次实验，我们选取了罗曼·罗兰的《约翰·克利斯朵夫》作为法语原文。这部作品发表于 1904 年至 1912 年，得益于这部作品，罗曼·罗兰获得了 1915 年度的诺贝尔文学奖。

但真正选取这部作品作为语料，还是首先基于以下三点的考虑：一. 这部作品是傅雷译文中最宏大的一部，全文共有 10 卷，约合 50 万字。象这样大容量的文本，对于运用词量学来说是有必要的。二. 该法语作品原文可以在网站 http://www.ebooksgratuits.com/ ebooks.php 上获得免费的电子版，同时，傅雷的翻译也可以在网站 http://www.yifan.net/yihe/小说/外国/yhklsdf / klsdf.html 上获取。因此，这就解决了文档形式的问题。三. 罗曼·罗兰的作品最能体现傅雷的风格。正如其所说：“再提一提风格问题。我回头看看过去的译文，自问最能传神的是罗曼·罗兰，第一是同时代，第二是个人气质相近。”[3]

傅雷前后花费近六年时间来翻译和重译《约翰·克利斯朵夫》。他的首次翻译由商务印书馆于 1937 年至 1941 年出版。后因不满意自己的翻译风格，又将全部作品重新翻译。1952 年至

1　本文作者感谢安德烈·萨来姆（André Salem）及安娜·蒂斯戴尔（Anne Dister）对本文给予了热情的帮助。
2　http://www.fulei.org/
3　见 1951 年 10 月 9 日傅雷致宋淇函，《傅雷谈翻译》，2005，沈阳： 辽宁教育出版社，第 37 页。

1953 年，平明出版社印制、发行了这批重译作品。

为了确保电子文本的质量，我们根据纸制版本书籍对法语原文及翻译进行了手工校对。依据的版本，分别为 1966 年阿尔班·米歇尔出版社的法文版及 1998 年安徽人民出版社的《傅雷译文集》，后者是基于 1975 年人民文艺出版社印制的《约翰·克利斯朵夫》重译本。

数据库构建的困难

建立傅雷译作数据库首先面对的是语料可用性问题，这一点包括两方面： 1 ）涉及语料文档形式的问题。并非所有原作品及傅雷译作都有电子版本。假如无法找到某些作品的电子版本，那么必须先完成电子文档的创建立工作，而这部分工作则需要许多的人力和时间； 2 ）涉及语料合法使用问题。在对整个语料库免费开放前，必须考虑作者版权问题。

在这里，应当强调的是，建立此数据库的唯一目的，在于方便学者对傅雷译作进行学术研究，因而毫无商业利益。此外，我们所使用的语料库显示方式（将在下文作详细介绍），并非适合普通阅读。虽然读者可以很容易进入语料作品的某一具体章节段落，但是要获得整部作品的阅读，则需要无数次搜索、粘贴复制才能实践。因此，我们认为，此数据库的建立，不仅不会影响对原版书籍以及傅雷译作的购买，相反，它会为这些作品起到很好的广告作用促进更多的人来了解、研究这些作品。

除以上问题外，建立语料库还需解决两项技术难题： 一是关于中文切分；二是关于建立平行语料库时，法、中间文本的对齐问题。

1. 中文切分

中文的信息化要比其它使用字母的语系来得复杂得多的（Miao et Salem： 2008 ）。通过将中文切分成词，将极大提高词量学分析和研究的准确性（同上文）。众所周知，汉语是连续书写的语言，字与字（或词与词）之间相互接连没有空隙。假如希望能够迅速、准确进入中文语料库，然么第一步须将汉语句子切分成词。尽管近些几年，中文切分技术有了迅猛发展，但是细节上仍存在诸多问题。这不仅是因为技术上存有局限，而是因为，定义中文“字以及分析形态学上缺乏共识（Xiao：2000）。

就我们已建立的关于《约翰·克利斯朵夫》法汉平行语料库，所用的中文切分软件是由中国科学院计算技术研究所所研制的汉语词法分析系统 ICTCLAS[4] 。当然在市场上或者网络上存有多种免费开源的或收费的切分软件，如海量智能分词（研究版）、Lucene[5]等，但从词量学研究结果来看，ICTCLAS 所提供的分词效果最为优越。当然，现有的分词技术在不断完善当中，各种软件都无法到达百分之百的正确率，因此对于切分完毕的语料，必须经过人工的检查和校对。

在傅雷的《约翰·克利斯朵夫》语料中，我们修正了诸多由于外国人名以及少数歧义词组所造成的切分错误。此外，在语料库建设过程中，为了在下一步语料分析中能更好地对比法文原文与中文翻译，我们将中文中的标点符号调整为西欧式标点符号。但对于中文中特有的标点，我们给与予了保留（例如，顿号“ 、 ”）。

下表 1 展示的是中文切分前后的语料形式。

表 1 :《约翰·克利斯朵夫》中段落的切分前后对比

切分前	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。 老人嘴里是那么说，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。
切分后	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 那么 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 .

4 此软件开发于 2002 年，一直处于不断完善中，详细有关内容，请查阅： http://www.nlp.org.cn/project/project.php?proj_id=6

5 可分别查看：http://www.hylanda.com/server/ 及 http://lucene.apache.org/java/docs/index.html

最后，还有关于中文计算机信息编码的问题。尽管 Unicode 标准提供了一个很好的解决方法，但仍存在一定问题，比如，一些汉字并没有被收编，统一字形的单词被多个语言系统所共用（如中文与日文），但牵涉到的词义却大相径庭[6]，因此还需要时间来完善信息和技术。

2. 法汉文本对齐问题

法语和汉语两者语系差异悬殊，之间并没有同源词相联系[7]。因此，在这两个文本间进行自动对齐工作，要比在两个邻近语言文本间开展起来会难得多。此外，就语料文本性质上来看，涉及到文学翻译，傅雷的译文较原文在形式上做了较大的调整：有时他将两个段落二合为一；有时相反，将某一大段落改为两小节。但需指出，大段落的添加或删除并没有在傅雷译文中发现。

目前现有的自动对齐软件都难以适应法、汉语料的要求[8]，因此我们自行研制了一个对段落进行半自动对齐的软件：“Alignator”，此软件在 Python（计算）和 Javascript（界面图形）上实现编程的。

因没有同源词可以依托，所以 Alignator 采用“瞬时扩张”（dilatation temporelle）的动态计算方法来测拟语料中的“词语对”。其理论假设为“词语对”在两个语料文本中的分布大致是类似的。因为虽然它们的分布位置可能有偏差，翻译文本中难免有增、减词现象，但是从整体上来看，某些单词成对出现的几率要比其它单词出现的几率大得多，对于这些单词，我们将其定义为“词语对”。需要强调的是，“瞬时扩张”的计算方法不需要将连续书写的文本切分成词，该程序使用的是单个字。

经过上述步骤计算获得的“词语对”，然后再用于标准的依靠同源词来对齐的检测方法，确切地说，通过计算段落间矩阵来获得最短的对角线路径从而实现语料间段落的对齐。试验结果显示，对于篇幅大的或对等率较高（即没有过多段落的添加，或大幅度段落次序的调整）的双语本，对齐精确率比较明显。上述结论，在对《约翰·克利斯朵夫》语料进行对齐时也是被证实了的。

Alignator 基本的切分思路并非是将文本中某一段落分割为二，而是需要在段落间寻找最为准确的切分边界。其次，所运用的计算法是基于以下的假设：文本 A 中的每一个段落都能在文本 B 中找到一个（或多个）匹配段落。我们知道，但这种假设与现实情况存在较大的差距。例如，某译者在文中添加了译者序，那么这部分内容就无法在原文中找到匹配段落。因此必须在自动对齐完成后进行手动纠正。

Alignator 为手动纠正步骤设计了诸多便利、快捷的方法，可以直接在在线的界面上完成。在此，我们对纠正过程中的软件的操作界面作一简单介绍。

所有的操作均可以由按钮或鼠标拖放来实现。

I. 工作区块边框上的按钮：

指示箭头向上： 意为能对区块进行所有向上的操作。

^：区块向上一格，进入一个空块 （情况是此区块在对等文本中没有匹配）；

^^：融入前一个区块；

^^^：融入前一个区块并重新计算所有以下的区块，但不触动已完成的结果。

指示箭头向下 :意为能对区块进行所有向下的操作。

操作规则与向上按钮的一致。

6 在欧洲语言方面也存在同样问题。如“哥特式字母‘A’与在标准字体中的‘A’是否为同一个字符？”或者“希腊大写字母是否与“A”为同一个字符，两者仅仅是因为形态上相等同？”参见：http://fr.wikipedia.org/wiki/Unicode

7 同源词（Cognat）是指形态上等同或类似的单词，在近邻的语系中，通常情况下，它们被视翻译中的对等词。

8 大多数自动对齐软件都建立在同源词的基础上，因此适用于相邻近的语言对（见 Simard et al. 1992）。另外，这些软件使用了大量的词汇列表，这对于中文文本来说也并非合适。

II. 鼠标拖放：

点击选定的工作区块（段落）并用鼠标将其拖至匹配区块位置，即可实现直观的动态对齐。在操作过程当中，一个包含所选匹配段落开头字句的窗口将会弹出，方便操作者的控制。操作完毕后，匹配区块间将被一链条而锁定，意为在之后的重新计算中，将不考虑此匹配的区块。

在下图 1 可以看到《约翰·克利斯朵夫》法汉文本对齐的手动纠正步骤的过程和结果。操作 I 使用了按钮功能，操作 II 则使用了鼠标拖放功能，二者的结果相同（除：在操作 I 完成时没有链条显示）。

一旦完成手动纠正，对齐结果可以 XML 方式输出，输出的形式可根据研究者要求来选择：分隔标记（自行选定），标记地点（对齐段落之前或之后），以单个对齐的文本还是双文本交叉来输出。

表 2 显示的是在段落后添加“＃”为分隔标记的交叉对齐文本。

表 2:《约翰·克利斯朵夫》法汉对齐语料（部分）

- 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 . # 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. # 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 . " #

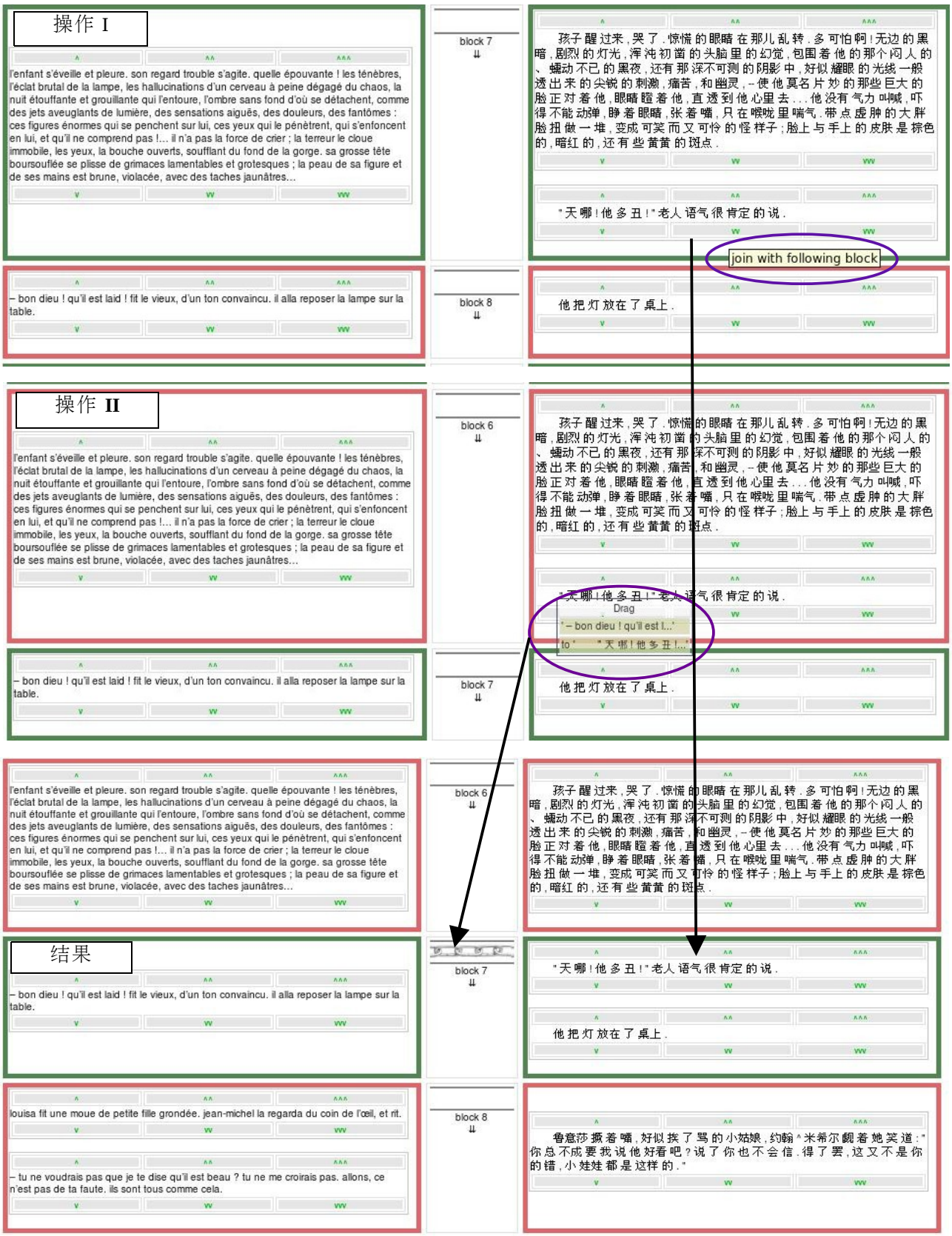

图 1: Alignator 中的法汉文本对齐手动纠正的操作和结果（部分）

语料分析

以上述方式完成的对齐语料可以直接导入词量学软件 Alignoscope 。Alignoscope 也是我们根据翻译学研究需要所自行设计的软件，虽然总体上是一项语汇索引工具，但是它能够观测，在整个文档范围内（以图表表征形态显示），查询字词／组的分布情况，并提供特征词和“包含”、“不包含”混合搜索的功能。目前，我们以将部分《约翰·克利斯朵夫》的法汉语料库放在网站：http://miaojun.net/alignoscope 上，读者可免费查阅 ，其它的语料在仍准备阶段。

事实上，语汇索引这一技术目前被广泛应用于各种计算机辅助的语料库研究软件中，大多数软件都支持寻找查询词以及查询词扩展语境的功能，但要实现自由观测上下文语境，确定查询词在整体语料中的分布情况，却没有多少软件能够实现这样的操作。在 Alignoscope 中的同步显示功能，则满足这两项要求。此外，对 Alignoscope 的使用，也无需在本地电脑中安装特殊的软件，仅需要一个符合国际标准的浏览器即可（推荐使用 firefox 浏览器）。

在下文我们将以“honnête”（诚实）及“amour”（爱）为例，介绍如何运用 Alignoscope 对傅雷译文进行翻译学研究。

1. 查询词（组／串）

首先我们将考察傅雷是如何将法语“honnête”翻译成中文的。在法语原文语料部分的“包含”（contient）搜索项中输入“honnête”，点击“搜索”（chercher），即可获得下图 2。（有关软件的具体操作将在下文中谈到）

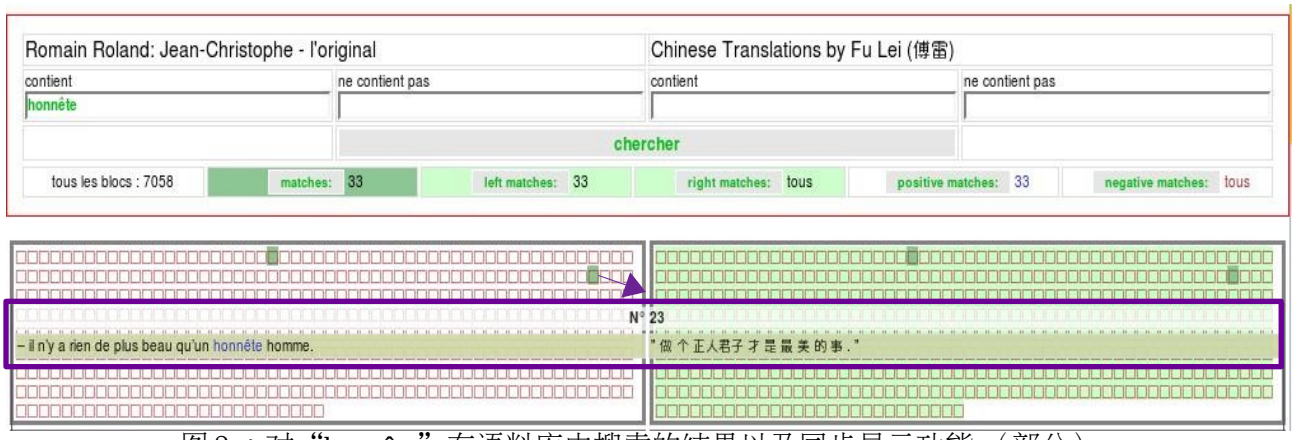

图 2 ：对“honnête”在语料库中搜索的结果以及同步显示功能 （部分）

整个语料库共有 7058 工作区块（方格），每一工作块包含一个对齐单位（段落）。搜寻结果表明（图 2），在法语语料中共有 33 个符合我们搜索条件的区块，这些区块以暗绿色的方格显示。将鼠标移到此暗绿色方格上，包含双语料的工作区块即以瞬间的方式显示在屏幕上。而将鼠标点击此暗绿色方格，同步显示则以恒定的方式展开。

这种“按需显示”的模式不仅保证了软件的高速运转，并且还确保了电脑内存低的用户的操作。反过来，假设在软件使用前需要下载 10 大卷的《约翰·克利斯朵夫》语料，不仅下载需要时间，而且之后在对全语料库进行运算上也需要花费时间，像这样的操作，会显得繁琐迟缓。同步显示窗口中间的数字代表此工作区块在整个语料中的位置。通过点击窗口中空白处（白色）或者窗口上的数字，即可以将显示窗关闭。另外，所搜索到的查询词在显示窗中以颜色着重标出，方便研究者的查询工作。

在上图 2 中，窗口左侧显示的是法语原文，可以看到“honnête”被标记为蓝色，包含在“– il n’y a rien de plus beau qu’un honnête homme.”句子段落中。在其右侧，则是对等的傅雷翻译： “做个正人君子才是最美的事。” 很明显，在此例中，傅雷运用了中国习语 “正

人君子”来翻译法文“honnête”一词。

另外，Alignoscope 还提供搜索结果的输出，见图 3。

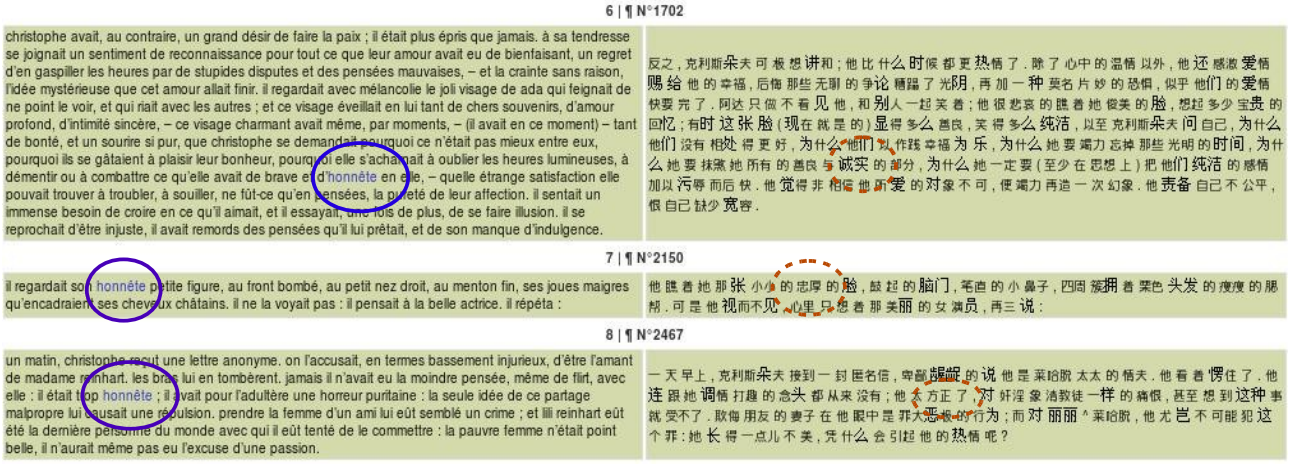

图 3： 数据库“honnête”搜索结果的输出 （部分）

对图 3 法汉部分作一简单对比，即可看到傅雷根据不同的语境，来使用不同的汉语词汇：“诚实的”，“忠厚的”，“方正”等。

虽然通过查询“honnête”一词的，我们看到了傅雷的中文翻译，但是我们还没有考虑到此词在法语语料中其它形式的使用情况。为此，须搜索“honnête”一词的各形态及派生，这项要求可以通过搜索词语串“honnête|honnêtes|honnêteté|honnêtement” 来完成。（“|”意为“或”，因此这表达式可以搜索符合要求项之一的词语。）

根据上上述方法，我们可以获得对“honnête”一词的所有中文翻译，但工作还没有结束，接下来是，需要翻译研究者对研究结果进行分析和阐释，并期待能进一步提出假设，剖析傅雷对“诚实”的理解。

2. 特征词

在这里我们引入另一个概念：特征词。输入所要查询的单词（或词组），Alignator 将进行超积累分布统计计算（英文：Hypergeometric Cumulative Distibution, 又称“精确费雪测试”，法文：le test exact de Fisher），即对于每一查询词，软件将计算此词出现在（由符合搜索条件所组成的）子语料中概率的总和。假如某一单词的出现次数超出均衡分布，那么意味着这个单词在子语料中具有显著特征。当然，查询词总为特征词，因为子语料正是通过搜索查询词而汇总的。但假如其它单词也同样出现在特征词列表中，即这些词的出现次数高于预期，那么可以断定这些词和所查询词之间，必定在句法上、语义上、或者文本上有关联。即使只按照单语种的要求来搜索，但通过计算对齐了的工作区块（子语料），我们同样可以获得第二种语种语料的特征词列表，此列表将包含各种关于查询词的“翻译”的列表。

相反情况，我们还可以获得出现次数低于预期的单词列表。这些词与所查询词之间，在广义句法、语义上，往往带有非相关联的“反义”。（参见 Lebart et Salem）。

在本文语料库中，在原文语料一侧输入“honnête|honnêtes|honnêteté|honnêtement”[9]，原文和译文两边都将显示高特征词和消极特征词列表。由于文章篇幅有限，仅在这里比较原文和译文的高特征词列表。

9 等同于表达式 “honnêt.*”，意为搜索所有以“honnêt”为起首的单词。在 Alignoscope 中也可使用其他的表达式，具体内容在其网站上有说明。

表 3: 法汉平行语料库中“honnête”一词的特征词列表 （局部）

spec	freq sub	freq tot	token (texte original)
∞	35	35	honnête
∞	24	24	honnêtes
∞	16	16	honnêteté
∞	3	3	villiers
∞	3	3	blondine
∞	2	2	écœurée
∞	2	2	traduits
∞	2	2	tannhäuser
∞	2	2	lucie
∞	2	2	imitateurs
∞	2	2	fromage
∞	2	2	effronté
∞	2	2	difformités
∞	2	2	affectent
∞	1780	53289	,
∞	1218	38746	’
∞	848	24673	de
∞	696	28498	.
∞	474	19078	il
∞	458	13658	et
∞	419	14671	la
∞	325	12092	à
∞	302	11489	le
∞	300	10436	l
∞	298	8632	les
∞	282	7523	d
∞	254	8535	elle
∞	243	7492	un
∞	243	6289	qui
∞	238	7302	qu
∞	234	6194	des
∞	211	6584	en
∞	206	7008	ne
∞	198	6806	;
∞	190	6355	était
∞	173	6131	pas
∞	168	6513	se
∞	166	6081	–
∞	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
5	2	3	vertueuse

spec	freq sub	freq tot	token (texte traduit)
∞	5	5	维廉哀
∞	4	4	坦华塞
∞	3	3	阿迦玛农,
∞	3	3	巡礼
∞	3	3	守本分
∞	3	3	君子
∞	2	2	鲍东
∞	2	2	阿丹丽
∞	2	2	贪心
∞	2	2	苏阿勃
∞	2	2	眉开眼笑
∞	2	2	特维廉哀
∞	2	2	正派
∞	2	2	戏谑
∞	2	2	前夫
∞	2	2	信奉
∞	2	2	作料
∞	2	2	何谓
∞	2	2	任意
∞	1673	48512	的
∞	1640	49890	,
∞	854	34292	.
∞	452	17791	他
∞	376	10714	是
∞	281	9442	不
∞	275	13680	了
∞	176	8133	在
∞	169	7607	着
∞	94	8727	"
39	19	21	诚实
13	26	87	’
10	11	27	规矩
9	6	11	依斐百尼
9	4	6	饼
8	26	114	老实
8	3	4	伯伯
6	4	9	帝王
6	3	6	神幻

表 3 左侧显示词“honnête”，“honnêtes”，“honnêteté”在法语语料中分别出现 35、24、16 次。通过此表，我们注意到“honnête”的一系列词语通常出现在评论人的描述中：例如：“écoeurée”（厌恶的），“affectent”（使痛苦），“moralité”（道德）以及“vertueuse”（品德高尚的）等等（表 3 中紫色标记为笔者所加）。在这些段落中，罗曼·罗兰描写了人的性格，这也解释了为什么在特征词列表当中出现人名，如：“villiers”，“lucie”（Lucie de Villiers 为一个人），“christophe”。此外，列表显示了与查询词共现的标点符号以及功能词，但同时也包含了一些干扰词如“fromage”（奶酪）－这些词出现在列表中，仅仅因为它们随机出现在子语料库当中。

表右侧，直眼观察（没有察看原、译文上下），即可推断出词语“守本分”，“君子”，“正派”，“诚实”非常有可能是傅雷翻译“honnête”一词所用的中文词汇。此外一些词语，如“规矩”（特征率 18）及“老实”（特征 8），认为它们在语义上非常接近“诚实”，所以我们推测这些词语同样也是傅雷用来翻译“honnête”的。

需要强调一点，尽管可以证实特征词提供了诸多关于翻译特点及倾向的信息，但要准确把

握译文词汇，则必须回归到原文和译文中逐一验证。因此，特征词列表并不能替代对平行语料的具体研究。借助软件中的同步显示功能，我们完成此了此项语境检测工作。研究结果列表于下（表4）。

表 4 :傅雷翻译“honnête”所使用的中文词汇

原文词形	原文词形	中文翻译	出现次数	中文翻译	出现次数
honnête	形容词单数	正人君子	1	笃厚的	1
		善	1	规矩诚实	2
		老实的	1	良家妇女的	2
		老实	6	规矩	4
		真正的	1	清清白白的	1
		守本分的	1	胆小	1
		诚实的	6	道德	1
		忠厚的	2	安分守己的	1
		方正	1	认真的	1
		本色的	1	老实人的	1
		真诚	1	完全的	1
honnêtes	形容词复数	老实的	4	诚实的	4
		泰然的	1	正派的	1
		老实话的	1	简单的	1
		老实人	4	懦弱的好人	1
		胆小	1	好好先生	1
honnêteté	名词	道德	2	正人君子	1
		大贤大德	1	诚实	5
		老成	1	奉公守法	1
		老实	3	守本分	1
honnête-ment	副词	老老实实	1		

由于文章字数有限制，我们在这里不对表 4 中内容作一一具体分析，仅强调以下三点 ： 1. 前一步骤中的特征词计算具有可靠性，能够为翻译研究提供思路。因为我们的语境考察中都遇到特征词列表中所出现的单词。 2. 回归语境的考察非常有必要，它能够为辨别词汇间的细微差别提供依据。 3. 某些汉语词汇只能在回归语境中观测到，这与使用切分了的中文语料有关。如，词组“懦弱的好人”被是分割成三个部分：“懦弱”，“的”和“好人”。

3.非常规翻译

对于文学或翻译研究者来说，通常关切的是译文中的非一般翻译方法，或者是译文较原文的变动情况，在本章节中，我们将尝试在 Alignoscope 中自动提取傅雷所使用的非常规翻译，以希望由此能够进一步展示傅雷在翻译方面的技巧和创造性。

Alignoscope 提供两大项搜索功能：正项搜索（“包含”）和负项搜索（“非包含”）。正项“包含”搜索比较容易理解，而对于负项“非包含”搜索，意思为将一定字词组“踢出考察范围。（下文将举例说明）

事实上，通过“包含”和“非包含”项可以获得在原、译文间交织查询的结果 。查询结果在 Alignoscope 以颜色组合显示，从而便利了两大层次信息的传递。

II. 方格背景颜色（或填充色）： 显示满足或部分满足搜索条件。

- 深绿色 : 完全符合对两种语言的搜索；
- 浅绿色 : 符合相关区域的搜索，但并不满足另一种语言中的搜索；
- 无填充色 : 不符合相关区域的搜索。

III. **方格边框颜色：显示满足交织间的搜索条件。**

- 蓝色： 符合两种语言中对正项的搜索；
- 红色： 符合两种语言中对负项的搜索；
- 黑色： 在两种语言中都没有与正、负项搜索相匹配的。

例如：一个填充为深绿色的蓝框方格，它将表示此工作区块中含有符合所有搜索条件的字词。 同样，搜索到的查询字在同步显示窗口中也用颜色来明显标记出：

- 蓝色：被搜索词 （“包含”）；
- 红色：被排除词（“非包含”）。

举例来说，假如想要知道关于“aimour”（爱），而非“aimbale”（可爱）这一概念的非常规翻译。那么，首先在原文一侧的“包含”搜索项中输入“amour|aim.*”。这样一来，法语中“aimer”（爱）一动词的各形式变位都纳入研究范围之内。同时我们在“非包含”搜索项中输入“aimable.*”，换句话说，通过这一操作，包含有单词“aimable(s)” 和“aimablement”的段落都被排除在“真正”搜索范围之内。在中译文一侧，众所周之，“amour”和“aimer”通常被翻译成 “爱”、“喜欢”、“爱情”，因此这些单词不是我们研究的兴趣点，故将其列入“不包含”搜索项 。与此同时，我们空留出“包含”项，因为我们不知道傅雷会用什么单词来翻译“amour|aim.*”。

图 4：傅雷译文中对“amour”的非常规翻译（局部）

图 4 显示，整个平行语料库中共有 170 个区块满足上述搜索要求，也就是说，在 170 个段落单位中，傅雷都没有使用“爱”，“爱情”以及“喜欢”来翻译“amour|aim.*”。这些区块由深绿色填充的蓝框方格加以显示。例如，在显示窗口编号为 19 的区块中，傅雷选择了“我多疼你”来翻译“je t'aime”。

在编号 409 的显示窗口中，可以看到原文中的“aimerai”标注为蓝色，而中文中的“喜欢”被标注为红色，因为这个单词为负搜索的查询词。由此，也看到在原文侧和译文侧的方格颜色并非一致，前者为黑框浅绿色，后者为蓝框无色。

鉴于我们在译文侧的正项搜索项上空留着，因此只要译文段落区块中不含有负搜索的单词，那么这个区块就以浅绿色填充色呈现。比如区块 767，左侧原文中，没有包含任何正、负搜索词，因此以黑框无填色显示，右侧译文中，虽包含了正项搜索（因为搜索为空却），但却没有负搜索词，因此以黑框浅绿色显示。

区块 687 中，单词“aimable”和“喜欢”出现在各自语言范围内，但是这两个词均为所负搜索单词（“不包含”），因此它们呈红色。

简而言之，就上文所列举的各种情况，对与翻译研究者来说，重要的是考察完全符合搜索条件的区块，即深绿色蓝框方格，而其它情况则可以作参考。所涉及的具体的研究情况在这里一一列举，但是我们已看到 Alignascope 对研究的重大意义，集中的色彩的组合功能大大方便了研究者对信息的把握，并对查询词在原、译文处的整体状况有了直观的了解。

结论

语料库研究方法不仅对翻译研究甚至对普遍的人文研究中都具有重大意义。虽然有研究者[10]自 80 年代起就注意中国语料库的建设问题，但他们的首要目标是促进中国英语的教学改革，以及为大学英语课程和大学英语等级考试提供数据信息（卫乃兴，2005 ：1）。虽然近些年，语料库建设[11]在中国突飞猛进，但是在法汉平行语料库方面仍然还是空白。另外，由于缺乏足够的适用中文语料的工具，大多数文章和研究只停留在语料库建设的探讨层面上。

目前，翻译学呈现巨大趋势向实证方向上发展。数据库的建设和词量学的研究方法必将会为翻译学开辟更为科学、有效的道路。我们期待，傅雷翻译作品数据库的建立将为研究者挖掘和探索这一翻译、文学领域巨藏提供捷径，并不断促进翻译学这本学科往科学的道路上发展。

参考书目:

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

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, Montréal. pp. 67-81

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 (Consulté le 14 février 2008).

刘康龙, 穆雷（2006）“语料库语言学与翻译研究”，中国翻译， Vol. 27， N° 1, 59-64 页。

卫乃兴，李文中，濮建忠（2005）语料库应用研究，上海：上海外语教育出版社。

10 杨惠中在 80 年代建立了中国第一个语料库“交大科技英语语料库”（JDEST：http://corpus.sjtu.edu.cn/），其又于桂诗春在 90 年代创建了“中国学习者英语语料库”，（CLEC，http://www.clal.org.cn/corpus/ChiSearchEngine.aspx），另外还可以看到何安平的英语教育教学语料库 CEEC 等等。

11 参见刘康龙, 穆雷（2006）“语料库语言学与翻译研究”，中国翻译， Vol. 27， N° 1, 59-64 页.