

doi 12 3969/j issn 1672-0598 2010 04. 020

《霍小玉传》两个英译本的翻译风格

——基于语料库的研究*

张燕, 陈建生

(天津科技大学 外国语学院, 天津 300222)

[摘要] 使用语料库语言学方法对《霍小玉传》两个英译本的译者风格进行了研究。使用语料库检索分析软件 WordSmith, 对两个英译本以及作为对比使用的英语本族语语料库中的平均词长、词汇密度和文本可读性等多方面数据做了统计和对比。同时, 还将定量研究与定性研究相结合, 运用相关翻译理论对两个译本中的翻译实例进行对比分析, 得出了两个译本译者在翻译风格上的特点。基于语料库的译者风格实证性研究具有科学性, 因此其研究结果可以为翻译批评提供可靠的依据, 也可用于翻译教学和译者培训中。

[关键词] 语料库语言学; 翻译风格; 霍小玉传; 词汇密度; 名词化

[中图分类号] H315.9 **[文献标志码]** A **[文章编号]** 1672-0598(2010)04-0111-06

一、引言

语料库语言学是上世纪五十年代后期发展起来的全新研究领域。近几十年来, 基于语料库的语言研究取得了长足发展, 与之相关的语言研究方法也得到了学术界的广泛认同。语料库不仅为语言研究提供了广泛的言语文本, 而且使传统语言研究由内省式的抽样方法转变为客观的数据调查和分析, 材料真实可靠。利用语料库进行翻译研究是新时期翻译学研究的有效手段, 语料库语言学的研究方法为描述式的翻译研究提供了一种定量分析的数据, 翻译研究者可以依据这些数据做进一步的定性分析。这样就保证了翻译研究成果的科学性。此外, 翻译语料库检索软件支持多方面和多层次的检索, 使翻译研究者有可能进行深入和细化的翻译研究。将语料库运用于翻译研究始于 Mona Baker, 1995年, 她筹建了翻译英语语料库 (Translational English Corpus), 对翻译的性质和特征进行描述。Baker等人在这一领域的开拓性研究取得了令翻译界瞩目的成就 (陈伟, 2007)。

风格是某一内容与形式在其表现或外化时所反映、流露和依赖的某种特殊的风味与方式。风格所指的部位时而贴近内容, 例如当它反映某种风情感调神态气度时, 又时而和更多的时候贴近于更实际的形式因素, 比如声调、节奏、词汇、句法等等, 而不停地游离于两者之间更是常事 (高健, 1995)。以往对文学翻译作品的研究涉及文体风格的不多, 而且主要是对某一作家作品中的某些篇章或句子的翻译风格的描述性分析, 然而这种传统的经验式和内省式的研究方法并不能对整个翻译作品的风格进行客观和全面的研究。

Biber (1994) 指出, 语料库能提供大量真实自然的语料, 研究者可以对语言中真正起作用的结构用法进行分析, 使研究建立在对自然语言结构及模式的分析基础上, 而非仅凭研究者的直觉或感受。语料库方法的优势在于可提供数据以及研究结果的可观测性和可验证性。Baker (2000) 从类符/形符比、平均句长和叙述结构三方面利用附码后的双语语料库来描述和分析两位译者的文体风格。用

* [收稿日期] 2010-06-10

[基金项目] 天津市“十五”社科规划资助项目 (TJYW06-3-028) “小规模英汉双语平行语料库的建设与应用研究”

[作者简介] 张燕 (1981—), 女, 河南省舞钢市人; 研究生, 在天津科技大学学习, 主要从事语言学、应用语言学和语料库语言学研究。

陈建生 (1953—), 男, 上海人; 教授, 在天津科技大学外国语学院任教, 主要从事语言学、应用语言学、计算机辅助英语教学、语料库语言学研究。

语料库研究翻译作品的文体风格是一种新的手段和方法,而且基于语料库的文体和文本研究可以与其他语言学方法、统计方法和翻译理论相结合,帮助读者全面的了解文本中的词汇、语篇、风格等多方面的特点。本文根据由《霍小玉传》的两个不同英译本建立的小型语料库以及英语本族语对比语料库中得出的数据,以定量和定性相结合的研究方法,从词汇和语篇的角度考察这两个英译本的翻译风格。

二、语料库的建立和研究方法

本研究选用的语料是由外文出版社 2003 年出版的《中国古代短篇小说选》和《唐代传奇选》中《霍小玉传》的两个英译文,译者分别为颜惠庆(以下简称颜译)和杨宪益、戴乃迭(以下简称杨译)。两个英译本分别在句子水平上与原文对齐。在选择本族语参照语料库时,考虑到文体的一致性,本研究选用了英语本族语语料库 BROWN、LOB、FROWN 和 FLOB 中的 Category P Romance and love story 部分作为参照语料库(以下简称参照库),共计 234 533 词。为了对文本做词汇密度和名词化方面的研究,需要对三个语料库进行词性附码。本研究采用的是 CLAWS 词性附码器和分类详细的 CLAWS7 词性附码集^①。

本研究使用语料库检索和分析工具 WordSmith 4.0 对语料库中的语料进行比较和分析。通过(标准化)类符/形符比、平均句长、平均词长等方面的统计数据,并结合相关的统计学方法来分析两个不同译本与英语本族语语料库之间存在差异以及两个译本文之间的异同,得出《霍小玉传》两个英译文的风格特征。

三、结果的讨论与分析

(一) 词汇分析

1. 词频概况

一个文本的形符数是指该文本中所有的词数。相同的形符称为类符,即文本中不同的词形的总数。类符/形符比(type/token ratio)可以反映语料中用词的变化情况。比值越大,说明该文本所使用的词汇越丰富。一般情况下,如果一个文本的形符数超过 1 000 时,需要使用标准化类符/形符比,以减少长文本中功能词词频过高造成的误差。平均词长是指文本中词(即形符)的平均长度,以字

母数为单位。平均词长较长说明文本中用的长词较多,一般文本的平均词长为 4 个字母左右。词长反映了用词的复杂程度。词长标准差反映出文本中各单词长度与文本平均词长的差异,标准差越大,文本中各单词词长之间的差异就越大。

表 1 三个语料库中的词频概况对比

	颜译	杨译	参照库
形符数	4 342	4 019	234 533
类符数	1 375	1 074	16 026
类符/形符比	31.67	26.72	6.83
标准化类符/形符比	46.78	42.13	44.29
平均词长	4.33	4.06	4.26
词长标准差	2.34	2.04	2.14

根据表 1 的数据可知:两个译本的词频概况与参照语料库的词频概况略有差异。与参照语料库相比,颜译的标准化类符/形符比略高,而杨译略低。这反映出颜译中用的类符更多。在平均词长和词长标准差方面,颜译高出参照语料库,而杨译则低于参照语料库。两者的平均词长分别与参照语料库中的平均词长做独立样本 t 检验,结果表明其差异都具有不同程度的显著性。这说明颜译中($t = 2.1319, p < 0.0330$)长词较多,用词比较复杂,甚至超过参照语料库。而杨译($t = 5.8792, p < 0.0001$)则很少使用长词。

表 2 三个语料库中每 1 000 词中 6 个字母以上的长词数量比较

	颜译	杨译	参照库
7-letter words	72.3	67.5	69.1
8-letter words	56.8	40.0	40.3
9-letter words	37.3	16.0	24.2
10-letter words	14.3	5.7	12.1
11(+)-letter words	18.8	9.8	10.9
合计	199.5	139.0	156.6

这一情况在词长表中也能体现。常见的文本中分别由 2~6 个字母组成的单词居多。所以表 2 中仅列出 6 字母以上的词长表。WordSmith 软件在进行词频统计时,会自动统计出各长度词在语料库中的使用频率。但如果语料库的容量不一样,语料

① CLAWS 词性附码器和 CLAWS7 词性附码集的详细情况见以下网址: <http://ucrel.lancs.ac.uk/claws7tags.html> 和 <http://ucrel.lancs.ac.uk/claws/trial.html>

库的实际词长出现次数就不具可比性。所以采用每千词不同长度的单词在语料库中出现的次数统计。这种方法可以更客观地对各长度单词在不同语料库中的分布情况做出比较。

从表 2 的数据可以看出,除了 10 个字母的词以外,颜译中 6 个字母以上的词的出现次数均大于杨译以及参照库。这也说明了颜译在词的使用上比较偏重复杂的长词。对比两个译文可知:

原文 1 因谓生曰:“素闻十郎才调风流,今又见仪容雅秀,名下固无虚士。”

颜译:“I have heard a great deal of your scholarly attainments and romantic personality” she declared cordially, “and now I perceive that you are at so very handsome indeed you do not belie your reputation”

杨译:“We have heard of your brilliance as a scholar” she said to Li “and now that I see what a handsome young man you are too I am sure your fame is well deserved

例句中颜译中用到的 6 字母以上的单词有 7 个,杨译中只有 3 个。另外,就文体而言,《霍小玉传》属于唐代传奇,作为中国风格的小说形式,虽然篇幅较短,但通过作者有意识的主观创作,已经作为一种独立的文学体裁形成独特鲜明的风格特色。唐以前的文学作品大都用词华美,句式工整对仗,写景状物、记人叙事力求完美。但唐传奇作家开始追求生活化、个性化的语言,用词简洁凝练,表现力强,显然在颜译中长词的大量使用是与唐传奇的语言特点相悖的。

2 词汇密度

近年来,国内外一些学者使用词项密度研究语篇问题,但大都停留在对写作或口语语体特征的研究层面上,用词项密度研究翻译风格还为数不多。Halliday (1985)将句子中的词语分为语法和词汇两大类。词汇类指的是实义词,包括名词、动词、形容词、大部分副词等。语法类是指功能词,或虚词,包括冠词、代词、大部分介词、连词、某些副词、助动词等。词汇密度可以反映句子中实词的比例。在句中通常实词比虚词传达的信息要多,虚词主要起语法和文本的连接作用。一句话中的词汇密度越高,说明该句中的实词数越多,该句所承载的信息量也越大。

通常用于计算词汇密度的公式有两种,分别由 Ure (1971)和 Halliday (1985)提出。Ure 计算的公

式是运用较为广泛的计算方法,用语篇中的实词数除以单词总数再乘以 100% 而得出平均每百词中所包含的词项数量,即:

$$\text{实词数} / \text{单词总数} \times 100\% = \text{词汇密度}$$

Ure 指出,词汇密度是区别语体正式程度的一个标准,语体越正式,词汇密度越高,语体越非正式,越接近自然口语,词汇密度越低。Ure 发现口语的词汇密度大概在 40% 以下,书面语的词汇密度则在 40% 以上(吴瑾,邹青,2009)。

Halliday 的词汇密度公式是用词汇项的数除以语篇中的小句数总数,即:

$$\text{实词数} / \text{小句总数} = \text{词汇密度}$$

在 Halliday 看来,句子中的小句是一种具有弹性功能的单位。它可以同时表现三种功能:(1)作为现象的表征;(2)作为话语功能的表达;(3)作为信息的载体。这三种功能决定了小句可以作为一个独立的功能单位来加以分析(高彦梅,2003)。

Halliday 对英语口语和书面语作了对比分析,发现口语语篇的平均词汇密度为 1.5 至 2 书面语篇的平均词汇密度为 3 至 6(杨信彰,1995)。表 3 是利用这两种公式来计算颜译和杨译的词汇密度的结果对比。

表 3 颜译和杨译的词汇密度对比

计算方法	颜译	杨译
Ure 公式	63%	63%
Halliday 公式	5.3	5.6

由表 3 可知,虽然颜译的类符/形符比数大于杨译(分别为 31.67 和 26.72),但两种不同公式计算出的结果都是杨译的词汇密度略大于颜译。这说明从信息承载量方面,杨译更加丰富。高词汇密度也说明对同一个单词的重复次数少,语言形式的多样化。

书面语中词汇密度高主要是因为出现名词化的倾向,而口语词汇密度较低,是因为名词化倾向较弱,这是由它们不同的社会功能所决定的。书面语在语篇生成的过程中,对语境的依赖性不像口头语那么直接,就可以经过一番思考和推敲,就会显得更正式,更严谨。因此,书面语,尤其是文学翻译,译者对所使用的词汇语法越刻意雕琢、推敲就越容易出现名词化倾向,这是因为书面语在呈现现实世界的经验和现象时,把这些经验和现象作为物体来看待,所以多使用名词。译本的名词化现象

可以从一个侧面反映出译者的文体风格(杨信彰, 1995)。

3 名词化

名词化是指动词、形容词等在功能上转化为名词或名词性句法结构的过程。一般来说,正式程度越高,句子结构越完整复杂,词汇选择也越讲究,发音也越充分清晰(侯维瑞, 1998)。王晋军(2003)在对不同语篇类型的语料进行分析后发现名词化比例与语篇类型的正式程度成正比,法律英语和科技英语的正式程度最高,新闻英语和英语小说次之,寓言和童话最低。其中小说中名词化的比例仅为 27.2%,这与小说的写作目的和读者群相关。小说主要是通过模仿生活来启迪大众,小说的语言多为一些日常用语,名词化程度较低。

语篇中名词化的形式很多,由于篇幅有限,本文仅就两译文中名词化的某些形式进行分析。通过 WordSmith 软件中的 Concord 工具对附码后的两个译文中具有名词化标志词缀的名词,例如 -tion, -ment, -ship 等进行检索统计,结果如下:

表 4 颜译和杨译的名词化结构使用频率对比

	颜译	杨译
小句总数	461	510
名词化总数	274	159
名词化比率	59.6%	31.2%

由表 4 可知,颜译的名词化比率将近杨译的两倍,名词化程度相对较高,这就使颜译显得比较正式,例如:

原文 2 太夫人素严毅,生逡巡不敢辞让;

颜译: As the old lady was very punctilious and conservative in such matters, he hesitated to express any objection, still less opposition.

杨译: His mother was so strict that Li, though hesitant, dared not decline.

原文虽然出自唐代作家蒋防之手,但文风简洁,结构紧凑,句子短小精悍,除去少量古代文言词之外,内容浅显易懂。颜译中高名词化现象与作为通俗文学的短篇小说的文体风格相悖。而且根据 Mänttari 的行为翻译理论,翻译过程被看成是由一系列参与者组成的过程,其中包括:发起者(initiator)、原文生产者(source-text producer)、译者(translator)、译文收受者(target-text addressee)和译文接收者(target-text receiver)等七个翻译过程

的参与者。在国内,两个译文的收受者和接收者主要是英语学习者、爱好者及翻译工作者。作为译者应综合考虑以上两方面的因素,在翻译过程中尽量使译文风格符合原文风格,更加贴近读者。因此,杨译在这些方面优于颜译。

(二)语篇分析

1. 句子特点

表 5 三个语料库中句长的对比

外	颜译	杨译	参照库
句子总数	198	234	19,075
平均句长	21.93	16.75	12.17
句长标准差	11.15	7.77	9.37

句子长度也可作为表现个人文体风格典型的标志之一。句子长度是指一个句子中单词的数量。由表 5 可知,与颜译的句子平均长度相比,杨译和参照库的平均句长差别较小。另外,原文是由 12 个自然段组成,颜译共有 63 段,杨译共分 46 段,这表明两个译文都对原文结构作出了调整。颜译中较多使用成分复杂的长句进行翻译,有时一句话就单独成段,使译文呈现出结构复杂、衔接紧密、逻辑严密的翻译风格,而杨译整体上使用了中等长度句进行翻译,译文结构简单明了,清晰易懂,杨译句长的使用上更接近本族语使用者,更符合译语语言的文类特点。例如:

原文 3 生自以孤负盟约,大愆回期,寂不知闻,欲断其望,遥托亲故,不遣漏言。

颜译: Realizing now that he had violated his oath to Jade and broken his promise to meet her soon, Li decided not even to correspond with her; so as to put an end once and for all to her hopes, and moreover cowardly requested his friends and acquaintances to keep everything about his doings secret from her.

杨译: Knowing that he had broken his promise to fetch Jade at the appointed time, he sent no message to her, hoping that she would give him up. He also asked his friends not to disclose the truth.

原文中这句话并不单独成段,颜惠庆在翻译时将其独辟一段,整段由一个句子构成,共用 56 个单词。长句的使用增加了句子的复杂性,在翻译时采用加译的策略,使用了多个起连接作用的单词或短语来构建句子间的关系和衔接,这就增加了译文的逻辑性和连贯性。杨译把原文译成两句,共用 37

个单词, 这样的处理即准确的表达了意思, 又显得简洁、流畅。

2 可读性

可读性研究始于上世纪二十年代的美国, 是指篇章的难易程度, 主要涉及语言表面层次的语义和句法特性, 如词长、句长以及语言表达的复杂性。可读性计算公式是检测阅读篇章难易程度的工具, 在预测可读性方面能起到有效的检测作用(姚国树, 2008)。常用的计算可读性的公式有四种: (1) 福格指数 (Gunning - Fog Index), (2) 福来士易读指数 (Flesch Reading Ease Score), (3) 福来士难度级 (Flesch-Kincaid Reading Grade Level Readability Test), (4) 自动可读性指数 (The automated readability index)。计算福格指数公式表示为:

$$\text{Fog Index} = 0.4 \times (L + H)$$

其中 L 为平均每句所含单词的数量, H 为平均每百个单词所含多音节单词数。福格指数测量的是你需要接受多少年的教育才能看懂阅读的内容。指数越低阅读内容就越容易, 6 代表容易, 20 代表很难。

福来士易读指数的计算公式为:

$$\text{Flesch Reading Ease Score} = 206.835 - (1.015 \times \text{ASL}) - (84.6 \times \text{ASW})$$

其中 ASL 表示句子的平均长度; ASW 表示每个单词的平均音节数。其对阅读篇章进行从 100 (容易) 到 0 (很难) 的评估。福来士把 65 作为一般难度英语的分值。

表 6 颜译和杨译的可读性指数对比

可读性指数	颜译	杨译
福格指数	12	9
福来士易读指数	51	65
福来士难度级	11	8
自动可读性指数	9.3	6.2

福来士难度级数的测量基于美国学校年级等级, 如 8.0 表示 8 年级的学生能看懂的篇章, 7.8 到 8.0 之间属于正常水平。其计算公式为:

$$\text{Flesch-Kincaid Grade Level} = (0.39 \times \text{ASL}) + (11.8 \times \text{ASW}) - 15.59$$

其中 ASL 和 ASW 所表示的内容与福来士易读指数中的变量相同。

自动可读性指数是由 Smith and Senter (1967) 首创, 最早是为了评估空军手册和技术性文件的文本可理解性。其公式为:

$$\text{The automated readability index} = 4.71 \times (c \div w) + 0.5 \times (w \div s2) - 21.43$$

其中 C 代表字母数, W 代表单词数, S 代表文本中的句子数, 自动可读性指数的平均水平为 7, 高于 7 则说明较难。

表 6 是两个译文的四种可读性指数对比。

德国功能派翻译理论学派学者 Vermeer 提出的目的论理论中有两个重要的法则, 连贯性法则和忠实性法则。连贯性法则即“语内连贯”, 要求译文必须符合译入语文化的表达习惯, 能让读者理解。可见可读性也是衡量译文的重要依据之一。虽然计算的公式不同, 但是数据所反映出的结果是一样的, 那就是总体而言, 杨译比颜译更浅显易懂, 更具可读性。例如:

原文 4 生闻之惊跃, 神飞体轻, 引鲍手且拜且谢曰: “生作奴, 死亦不悛。”

颜译: “I am willing to be her slave all my life” cried the young scholar, dancing with delight and seizing Mrs. Bao's hands to show his gratefulness

杨译: When Li heard this he leaped for joy and felt as if he were walking on air. Taking Bao's hands he bowed and thanked her, saying “I shall be your slave as long as I live.”

原文 5 自是长安中稍有知者。风流之士, 共感玉之多情; 豪侠之伦, 皆怒生之薄行。

颜译: The story of Li's scandalous treatment of his mistress became widely known among the literary and romantic circles of Changan. One and all sympathized with and lauded Miss Jade for her deep attachment to her lover and condemned Li for his heartlessness.

杨译: When the story became known in the capital all the young scholars were moved by the girl's love and all the young gallants were indignant at Li's heartlessness.

在颜译中有一些不妥之处。如“一生作奴”中到底是作谁的奴隶? 根据原文的语境可知, 这是李生为感谢鲍氏为其做媒说的客套话, “作奴”译为“be her slave”, 而“her”指的是谁? 这样翻译显然不可取。“风流之士”和“豪侠之伦”被译作“the literary and romantic circles of Changan”也不免让人产生疑惑。风流在中国古代文学作品中有多种理解, 如风韵, 才华出众, 风采, 风俗教化, 遗风等等。但是风流并不等同于爱情和浪漫, 用“romantic”一词容易使读者产生误解。相比之下, 杨译文中的用

词就更加贴切,从语义和形式上都更加忠实原文。

四、结语

通过以上的比较分析可见,杨译的翻译风格更加简洁流畅,忠实原文。在表情达意方面,能充分发挥译语的特点;遣词造句形式多样,富于变化,多种翻译手法的交互使用,张弛有度,使译文更加贴近读者。颜译能做到再现原文的特点,将原作者的意图传递给了译文读者。用词丰富,译文丰满,并极具情感色彩,句子结构紧凑,逻辑性强。长词的大量使用,使译文略显复杂,同时也增加了阅读的难度;但过多直译使译文风格稍显呆板,有欠灵活性。

利用语料库研究译文风格,通过分析所获得的数据,探讨不同译文的翻译风格。这种方法把定量研究与定性研究相结合,使客观地把握原文风格和科学地分析译文风格成为可能。同时,这种研究方法更系统、更有针对性、可比性更强,为翻译研究提供可靠的量化依据,具有很强的客观性和说服力。这样,既避免了传统研究方法的主观性和随意性等缺陷,又能从整体上客观地把握文本,译者风格研究的科学性和实证性特点,使其可以为翻译批评提供可靠的依据,也可以运用到翻译教学和翻者培训中,因此不失为翻译研究的一条新路。

[参考文献]

- [1] Baker, Mona. 2000. Towards a methodology for investigating the style of a literary translators[J]. *Target* 12.
- [2] Biber, D., S. Conrad, R. Reppen. 1994. Corpus-based approaches to issues in applied linguistics[J]. *Applied Linguistics* 15: 169-189.
- [3] Halliday, M. A. K. *Spoken and Written English* [M].

Victoria DUP, 1985.

- [4] Ure, Jean N. *Lexical density and register differentiation* [A]. *Selected Papers of the Second International Conference of Applied Linguistics* [C]. Cambridge: Cambridge University Press, 1971.
- [5] 陈伟. 翻译英语语料库与基于翻译英语语料库的描述性翻译研究 [J]. *外国语*, 2007(1): 67-73.
- [6] 侯维瑞. *英语语体* [M]. 上海:上海外语教育出版社, 1988.
- [7] 高健. 翻译中的风格问题 [J]. *外国语*, 1995(3): 7-11.
- [8] 高彦梅. 运用 LD 评估语篇能力 [J]. *外语教学*, 2003(5): 65-68.
- [9] 李文革. *西方翻译理论流派研究* [M]. 北京: 中国社会科学出版社, 2003.
- [10] 吴瑾, 邹青. 中国学生英语口语笔语语体特征研究: 词汇密度与词频 [J]. *山东外语教学*, 2009(1): 8-13.
- [11] 王晋军. 名词化在语篇类型中的体现 [J]. *外语学刊*, 2003(2): 74-78.
- [12] 颜惠庆. *中国古代短篇小说选* [M]. 北京: 外文出版社, 2003.
- [13] 杨惠中. *语料库语言学导论* [M]. 上海: 上海外语教育出版社, 2002.
- [14] 杨宪益, 戴乃迭. *唐代传奇选* [M]. 北京: 外文出版社, 2003.
- [15] 杨信彰. 英语书面语体中的词汇密度特征 [J]. *解放军外国语学院学报*, 1995(3): 14-18.
- [16] 姚国树. 英语篇章的可读性检测 [J]. *内江科技*, 2008(11): 28.
- [17] 朱耕. 论译者的翻译策略 [J]. *重庆工学院学报(社科版)*, 2009(3): 142.

(责任编辑:朱德东)

Corpus-based Study of Translation Style of Two English Versions of The Story of Huo Xiaoyu

ZHANG Yan, CHEN Jian-sheng

(School of Foreign Languages, Tianjin University of Science and Technology, Tianjin 300222, China)

Abstract The present research makes an attempt at an investigation of the translators' styles of the two English versions of *Huo Xiaoyu Zhuan* in a corpus-based approach. With the help of WordSmith, a corpus retrieval and analysis suite, the data about the word length, lexical density, readability and so on are collected from the two translated texts and a comparable native English speaker corpus, and then these data and the samples of translation are compared and analyzed both quantitatively and qualitatively, which leads to the findings of the translation styles of the two translators. The corpus-based study of translator's style is both empirical and scientific, and its findings, therefore, can provide a solid foundation for translation criticism and find application in translation teaching and translator training.

Key words corpus linguistics; translation style; *Huo Xiaoyu Zhuan* (The Story of *Huo Xiaoyu*); vocabulary density; substantivization