Modern Linguistics 现代语言学, 2026, 14(8), 970-977
Published Online August 2026 in Hans. https://www.hanspub.org/journal/ml
https://doi.org/10.12677/ml.2026.148863

Hans汉斯

基于语料库的《呼兰河传》人机英译词汇风格比较

周佳玲

中国石油大学(北京)外国语学院，北京

收稿日期：2026年7月17日；录用日期：2026年8月14日；发布日期：2026年8月27日

摘　要

大语言模型进入文学翻译实践后，人机译文比较已从准确性评价拓展至风格研究，但现有研究多依赖节选或个别译例，难以判断局部差异是否构成稳定倾向。本文以《呼兰河传》全部正文为源文本，建立中文原文、葛浩文英译本与GPT英译文三方语料库，采用词汇多样性、近似词汇密度和词长指标进行全书及分章比较，并结合文化负载词作语境分析。结果显示，葛译词汇多样性较高，GPT译文近似词汇密度较高，两者词长构成接近。文化负载词案例表明，葛译会根据具体语境改变译法，GPT较多保留源语词汇构成和文化形式。研究认为，人机译文的词汇风格各自表现为相对稳定但并非固定的用词倾向，其具体表现会随文本内容和文化词项变化。全书指标、分章复核与译例细读相结合，可为人机文学翻译风格研究及译者主体性讨论提供语言证据，也有助于思考文学翻译中的人机协作模式。

关键词

语料库翻译研究，大语言模型，译者风格，《呼兰河传》，文化负载词

A Corpus-Based Comparison of Lexical Style in Human and GPT English Translations of *Tales of Hulan River*

Jialing Zhou

School of Foreign Languages, China University of Petroleum (Beijing), Beijing

Received: July 17, 2026; accepted: August 14, 2026; published: August 27, 2026

Abstract

With large language models increasingly used in literary translation, human–machine translation

文章引用: 周佳玲. 基于语料库的《呼兰河传》人机英译词汇风格比较[J]. 现代语言学, 2026, 14(8): 970-977.
DOI: 10.12677/ml.2026.148863

comparison has expanded beyond accuracy assessment to stylistic analysis. However, existing studies often rely on excerpts or isolated examples, making it difficult to determine whether local differences constitute stable tendencies. Using the complete text of *Tales of Hulan River* as its source, this study constructs a tripartite corpus comprising the Chinese original, Howard Goldblatt's English translation, and a GPT-generated translation. Lexical diversity, approximate lexical density, and word length are compared at both whole-book and chapter levels, supplemented by contextual analysis of culture-specific items. The results show that Goldblatt's translation has greater lexical diversity, whereas the GPT translation has higher approximate lexical density; the two translations display similar word-length profiles. The culture-specific items examined show that Goldblatt adjusts his lexical choices across contexts, whereas GPT more often retains source-language lexical components and cultural forms. The findings indicate that human and machine lexical styles represent relatively stable but not fixed patterns of lexical choice whose specific manifestations vary across textual contexts and culture-specific items. By combining whole-book measures, chapter-level validation, and close reading of translation examples, this study provides linguistic evidence for research on human-machine stylistic differences in literary translation and discussions of translator agency, while also informing approaches to human-AI collaboration in literary translation.

Keywords

Corpus-Based Translation Studies, Large Language Models, Translator Style, *Tales of Hulan River*, Culture-Specific Items

1. 引言

大语言模型进入翻译实践后，人机译文比较逐渐从准确性和流畅度评价扩展到语言特征、译文风格与文化信息处理。对于文学翻译而言，词汇选择是译文风格中较为直接且可重复观察的层面之一，因其不仅构成叙事节奏和文体色彩，也承载地方物项、民俗称谓与文化意象。当前相关讨论已由技术能力延伸至文学翻译的创造性、文化内涵以及人机分工与协作等问题[1] [2]。因此，对机器译文的考察不能只统计局部错译，还需要分析反复出现的词汇选择，判断其是偶发现象还是相对稳定的用词倾向。

相关实证研究主要沿两条路径展开。一类以翻译质量为中心，借助人工量表评估大语言模型处理文学文本的准确性与文学性[3]；另一类转向词汇多样性、句法复杂度和可读性等语言特征，考察人工译文、大模型译文与神经机器译文的风格差异[4] [5]。此外，也有研究借助语料库比较人机译文，从语言表现讨论人类翻译能力的优势[6]。语料库翻译研究为从语言分布而非零散例句识别反复出现的特征提供了方法基础[7]。Baker 将译者风格理解为译者语言产出中具有辨识度的反复选择[8]；Saldanha 进一步指出，译者风格研究需要寻找一致、连贯的语言特征[9]。这些研究共同说明，人机差异不只体现为个别译例，也可能表现为译文中重复出现的语言现象。

将“风格”用于大语言模型，需要限定其解释范围。人类译者的风格通常与持续的创作身份、审美取向和翻译经验相关。模型输出虽可呈现可统计的词汇规律，但这里的“风格”仅是语料层面的操作性概念，指特定模型、提示词、上下文范围和生成批次共同形成的用词倾向，不能据此推断模型具有统一人格、审美意图或跨情境不变的个体风格。与“词汇风格”不同，“主体性”不能直接、对称地用于大语言模型。模型能够生成不同译项，但仅凭译项之间的差异，尚不足以证明其具有审美判断能力。因此，

本文以“词汇风格”或“用词倾向”描述人机译文中均可观察的语言现象，不讨论模型主体性，涉及“译者主体性”时，仅指人工译者选择的稳定性及语境调节在译文中的语言表征。

在上述概念边界下，现有研究仍有三个问题值得考察。首先，以节选文本或不同应用场景语料为单位的研究能够识别指标差异，却较难判断这些差异在一部长篇作品内部是否跨章节保持一致[3] [4]。其次，词汇多样性、词汇密度和词长反映的是词汇使用的不同方面，若以单项指标或笼统的“词汇复杂度”概括译文风格，可能掩盖各指标方向并不一致的情况。最后，宏观统计能够显示差异是否存在，却不能单独解释差异由哪些具体词汇选择形成。将全书指标与文化负载词的语境分析结合起来，可以进一步观察具体译例中两译本的差异与共同点。

基于此，本文选择《呼兰河传》作为研究对象。小说以东北地方社会为叙事空间，包含大量民俗物项、地方称谓和口语表达，既适合观察整部译文的词汇分布，也为考察文化负载词的具体处理提供了丰富语境。本文以中文原文、葛浩文英译本和 GPT 英译文建立三方语料库，拟回答两个问题：第一，葛译与 GPT 译文在词汇多样性、近似词汇密度和词长方面呈现何种差异，这些差异是否在不同章节中保持一致？第二，两部译文在文化负载词的具体处理上有何差异，这些差异与全书词汇指标所显示的用词特点有何联系？

为回答上述问题，本文先从全书及分章层面比较两部译文的词汇指标，再结合文化词译例分析宏观差异的具体表现及其适用边界。本文不以某一指标的高低判断人工译文与 GPT 译文的优劣，而将结合全书统计、分章结果和文化负载词译例，考察两部译文用词差异的稳定性及其在具体语境中的表现。通过比较，可以更清晰地呈现 GPT 译文的用词特点及葛浩文在不同语境中的翻译策略。在此基础上，本文进一步讨论上述发现对文学翻译人机协作模式的启示。

2. 研究设计

2.1. 语料构成

本文建立由中文原文、葛浩文英译本和 GPT 英译文构成的三方语料库。中文原文采用浙江人民美术出版社 2017 年版《呼兰河传》[10]，保留七章、尾声等全部文学正文；人工译文采用葛浩文 1988 年英译本 *Tales of Hulan River* [11]，删除前言、目录和附录等副文本。经清洗，中文正文共 2120 个段落，葛译和 GPT 译文分别为 79,200 词和 72,726 词。全书指标以两部完整英译文为统计对象，段落对应主要用于 GPT 译文生成、文化负载词定位和语境核查。GPT 译文通过 Codex 命令行入口生成，界面模型标签为“gpt-5.5”。生成时采用不含直译、归化或模仿特定译者等风格要求的中性提示词，并保留原文段落编号；葛浩文译本未提供给模型。正式分析前，为排除模型训练时接触过葛译本的情况，本文通过 Python 程序计算 N 元组覆盖并检索长连续匹配，未发现 20 词及以上的连续复现，说明 GPT 译文并未大段照搬葛译本，可作为独立译本纳入后续比较。

2.2. 分析方法

本文词汇指标由 Python 自编程序计算，主要从词汇多样性、近似词汇密度和词长三个方面比较两部译文。词汇多样性采用标准类符/形符比(STTR)和移动平均类符/形符比(MATTR)，本文将两项指标的统计窗口均设为 1000 词；其中 MATTR 通过移动窗口减少原始类符/形符比受文本长度影响的问题[12]。近似词汇密度以 Python 自然语言处理库 NLTK (Natural Language Toolkit)所提供的英语停用词表为参照，计算非停用词占全部词形的比例，用于观察两部译文在功能词与非停用词构成上的差异，不等同于基于词性标注的严格词汇密度。为复核近似词汇密度的差异来源，研究另使用 spaCy 进行词性标注，统计两部译文的实词比例及主要功能词的每万词频率。词长则考察平均词长及不同长度词的比例。研究先比较全

书指标，再以七个主要章节复核差异方向是否一致。

在此基础上，本文参照 Nida 关于文化差异领域的划分及其翻译讨论[13] [14]，结合小说中的地方物项、民俗称谓和口语表达筛选文化负载词，通过四个代表性译例考察两部译文的归化与异化取向[15]，以及直译、意译、音译和解释等具体处理，比较文化信息的保留与变化。文化负载词分析主要用于把全书指标与具体译例联系起来，考察两部译文在文化词处理上有哪些差异，以及这些差异在不同语境中如何变化。

3. 研究结果与讨论

3.1. 词汇指标对比

表 1 显示，葛译共 79,200 词，GPT 译文共 72,726 词，后者篇幅约短 8.2%。按停用词口径分解，两部译文的非停用词分别为 37,778 词和 37,671 词，仅相差 107 词；停用词则相差 6367 词，占总词数差额的 98.3%，主要集中在介词及小品词、冠词和人称及物主代词。分章数据进一步表明，这一差异章节分布并不均匀。第一至六章的 GPT 译文均比葛译短约 8.5%~13.4%，第七章却多出 296 词(约 3.2%)。段组核查显示，第七章若干位置存在葛译压缩或合并原文信息的情况，说明人类译者倾向于随章节语境调节信息密度，相比之下，本次模型译文表现出较稳定的源文跟随倾向。

葛译的 STTR 和 MATTR 分别为 40.332%和 40.294%，均高于 GPT 译文的 37.869%和 37.891%。按 MATTR 换算，在 1000 词窗口内，葛译平均比 GPT 译文多使用约 24 个不同词形。这表明，在 1000 词窗口内，葛译使用的不同词形较多，GPT 译文则倾向于重复已经出现的表达。这一现象在例 1 中得到体现。

Table 1. Whole-book lexical measures of the two translations
表 1. 两部译文的全书词汇指标

指标	葛浩文译文	GPT 译文	GPT − 葛译
总词数(Tokens)	79200	72726	−6474
STTR-1000 (%)	40.332	37.869	−2.462
MATTR-1000 (%)	40.294	37.891	−2.402
近似词汇密度(%)	47.700	51.799	+4.099
平均词长(字母)	4.224	4.246	+0.022
1 至 5 字母词(%)	77.218	77.135	−0.084
6 字母以上词(%)	22.782	22.865	+0.084

例 1：

原文：大地满地裂着口……它们毫无方向地……只要严冬一到，大地就裂开口了。

葛译：the earth’s crust begins to *crack and split*…the *cracks run* in every direction…the earth’s crust *opens up*.

GPT 译：the whole ground *split open*…*cracks*…*opened*…the earth *split open*.

原文在同一段中反复使用“裂口”“裂开口”。葛译分别使用 *crack and split*、*cracks run* 和 *opens up*，GPT 则两次使用 *split open*，中间以 *opened* 对应。该例显示，在处理同一段落中的重复表达时，葛译会调整英文词形和搭配，GPT 较多复用相同表达。该段呈现的局部差异与全书 MATTR 的方向一致，使总体指标在具体语境中得到体现。

GPT 译文的近似词汇密度比葛译高 4.099 个百分点。词性标注的复核结果与此方向一致，葛译和 GPT

译文的实词比例分别为 46.12%和 50.22%。进一步统计发现，GPT 译文中冠词、代词和介词的合计频率每万词比葛译少约 477 个，且这三类词在七个主要章节中均低于葛译，体现出 GPT 译文功能词使用较少、非停用词占比相对较高的词汇构成特点。

两部译文的平均词长和长短词比例十分接近，说明葛译较高的词汇多样性并非通过增加长词形成。分章结果进一步显示，七个主要章节中，葛译的 MATTR 均高于 GPT 译文，差值为 1.589 至 3.155 个百分点；GPT 译文的近似词汇密度均较高，差值为 3.549 至 5.000 个百分点。总体来看，葛译的突出特点是词形变化较多，GPT 译文的突出特点是冠词、代词和介词占比较低，且两种差异均贯穿七个主要章节。

3.2. 文化负载词译例分析

全书指标能够显示两部译文在词形变化和功能词比例上的差异，却不能说明具体文化词在语境中如何得到处理。以下四例结合上下文和相关表达在作品中的使用情况，考察两部译文如何选择和组织文化词的对应表达。

例 2：

原文：假若他是正在牙痛，他也绝对的不去让那用洋法子的医生给他拔掉，也还是走到李永春药店去，买二两黄连，回家去含着算了吧！

葛译：Instead he would go over to the Li Yongchun Pharmacy, buy two ounces of bitter herbs, take them home and hold them in his mouth, and let that be the end of that!

GPT 译：He would still go to Li Yongchun’s medicine shop, buy two liang of coptis root, go home, and hold it in his mouth.

从归化与异化的角度看，两部译文呈现出不同的词汇选择取向。葛译将“二两”译为 *two ounces*，以英语读者熟悉的计量单位替换了中国传统计量单位；又将具体药名“黄连”译为含义较宽的 *bitter herbs*，突出其味苦和药物属性。这些处理减少了译文中的陌生文化成分，使英语读者能够直接理解该物项，体现出归化取向，但“两”和“黄连”所包含的具体文化信息有所减少。GPT 译文采用 *two liang of coptis root*，*liang* 保留了源语计量单位的名称，*coptis root* 则保留了“黄连”的具体指称，因而体现出异化取向。与葛译相比，GPT 译文保留的文化信息更多，但译文的陌生程度也更高。该例表明，葛译倾向于选择英语读者熟悉、含义较宽的词汇，GPT 则更重视文化负载词具体指称的保留。

例 3：

原文：“那算完，长的是一身穷骨头穷肉，那穿绸穿缎的她不去看，她看上了个灰秃秃的磨官。真是武大郎玩鸭子，啥人玩啥鸟。”

葛译：“She’s done for now,” Second Uncle You lamented. “She was born to poverty, right down to her bones. Instead of falling for someone wearing silks or satins, she had to get hooked up with a grimy miller. Well, to each his own.”

GPT 译：“That settles it. She was born with a whole body of poor bones and poor flesh. She didn’t look at those wearing silk and satin; she took a fancy to a gray, bald-looking mill hand. Truly, Wu Dalang plays with ducks: each person plays with his own kind of bird.”

“武大郎玩鸭子，啥人玩啥鸟”并非对择偶偏好的一般说明，而是有二伯对王大姑娘与磨官的嘲讽。他借武大郎的形象和“鸭子”的比喻，将两人的结合说成同类相配。单看 *to each his own*，葛译删去了典故和动物意象，语气也有所缓和。但从整句来看，*born to poverty*、*get hooked up with* 和 *a grimy miller* 已经承担了贬低和调侃的作用。葛译并未完全删除原句的评价，而是将其分散到前面的词语和短语中，再以英语熟语收束。GPT 译文保留了 *Wu Dalang*、*ducks* 和 *birds*，源语意象较为完整，但没有补充武大郎的

身份和形象，典故所包含的嘲讽意义因而未在译项中得到充分说明。

同一句俗语在第六章有二伯谈论自己的蝇甩子时已经出现。葛译在该处保留 *Wu Dalang* 和 *duck*，并加入 *the ugly ne'er-do-well Wu Dalang* 等解释，使典故参与塑造有二伯爱说俗语、带有自嘲意味的说话方式；到第七章评价王大姑娘时，则改用 *to each his own*。这表明葛译并非一律以英语熟语替换文化表达，而是根据俗语在具体语境中的作用调整译法。当文化表达主要参与人物塑造时保留并解释，当它主要承担评价功能时，则将评价意义分散到整句的词汇选择中。

GPT 在两处均基本保留人名、动物和原有句式，前后文形式一致性较高，没有出现葛译那样明显的语境化调整，本例说明，大语言模型可能缺乏语境理解能力，倾向于维持同一文化表达的形式对应，对译文在人物塑造和人物评价中的不同作用区分得不够明显。葛译在两处对同一俗语采用了不同译法，说明其用词会随语境调整，这与全书较高的词汇多样性相一致。

例 4：

原文：那家的老太太终年生病，跳大神都是为她跳的。

葛译：The old woman of this family was sick the year round, and the dance of the sorceress was performed for her benefit.

GPT 译：The old lady of that family was sick all year round, and the spirit-inviting rites were all performed for her.

“跳大神”既指特定民间仪式，也包含施术者、动作和召神功能等多个可被凸显的语义侧面。葛译的 *the dance of the sorceress* 突出了施术者及舞蹈动作，便于读者形成场景图像；GPT 译文的 *the spirit-inviting rites* 则把仪式功能显性化，使读者直接理解其与召神治病的关系。该例说明，人机译文的差异并非简单表现为归化或异化的策略选择，而在于对文化信息的选择性凸显；同时，两译都采用解释性短语，显示其在降低理解障碍这一功能上存在趋同。

例 5：

原文：有二伯说话的时候，把“这个”说成“介个”。

葛译：When Second Uncle You spoke he pronounced “this” as “dis.”

GPT 译：When Second Uncle You spoke, he pronounced “this” as “dis.”

“介个”是“这个”的方言发音，文化特征主要体现在语音形式而非词义。两部译文均以 *dis* 替代标准形式 *this*，利用英语非标准拼写表现人物发音偏离规范的特点。这种写法没有直接复制东北方音，而是借助英语读者能够识别的发音差异，保留“人物带有口音”这一特征，但“介个”所指向的东北地域身份没有得到再现。两部译文采用相同的核心译项，表明在人机译文可利用的目标语形式较为有限时，可能出现表层趋同。

例 4 中，两种解释性短语都降低了文化词的理解障碍，表现为功能趋近；例 5 中，两部译文均以 dis 对应“介个”，构成表层趋同。这些共同点并不意味着两部译文具有相同风格。语言学上，英语中可识别的非标准拼写选项有限；翻译学上，原文功能与目标语可接受性会压缩译文范围；从模型生成看，高频、常规化的表达也更可能成为输出结果。因此，趋同可能由形式资源、翻译规范和生成偏向共同造成。人机风格比较既要考察差异，也应把共同选择作为判断边界。

这些译例也为全书指标所呈现的用词特点提供了具体参照。葛译随语境改变词形、搭配和译项，与其较高的词汇多样性相呼应；GPT 对部分表达的重复使用，也与其词形复现程度较高的结果一致。不过，两译本之间的不同或趋同的处理方式说明，宏观指标与具体译例之间并非一一对应。全书指标反映大量词汇选择累积形成的总体特点，译例分析则说明这些特点如何在具体语境中出现或发生变化。二者结合，既可以避免根据个别译例概括整部译文，也可以避免依据指标高低直接判断翻译质量。

3.3. 人机协作启示

以上分析为后续文学翻译中的人机协作模式提供了参考。文学翻译中的人机协作不能只依赖固定的文化词对照表。同一文化表达在不同场景中可能分别承担指称事物、评价人物、表现口语或塑造形象等功能，因而未必需要始终采用同一译法。文化词译项表除了记录源语词和基本译法，还应标明首次出现时的解释方式、具体语境中的功能及允许采用的变体。大模型可以据此提供候选译项，人工译者则需判断某一处应保持前后统一，还是应随语境改变表达。

语料库方法可以用于人机协作中的译后检查，但相关指标更适合作为问题筛查工具，而不是译文优劣的评分标准。例如，词汇多样性可以帮助发现同一表达是否被过度重复，功能词统计可以提示译文是否存在冠词、代词或介词使用偏少的倾向，重复文化词检索则可以呈现同一称谓在不同章节中的译法变化。在发现这些现象后，仍需回到原文和上下文，判断相关重复或变化是语言使用问题，还是具体语境所要求的合理选择。

Baker 将译者风格理解为译者语言产出中具有辨识度的反复选择[8]。从人机分工的角度看，大模型适合承担初步翻译、译项生成、重复表达检索和备选方案整理，人工译者则需要统筹人物语言、文化解释和全书风格。大模型能够随输入语境改变译项，但这种变化还可能受到提示词、上下文范围的影响；人工译者的作用在于判断这些变化是否符合人物身份、叙事情境和作品整体表达。人机协作的重点因此不是使机器译文完全模仿人工译文，而是利用模型提高译项生成和信息检索的效率，由译者完成语境判断和整体风格控制。这种分工也表明，机器输出可以呈现风格特征，却不因此获得与译者对称的主体性；对文化解释、审美取舍和风险后果的责任仍由个人承担。

不过，上述人机分工并非固定模式，其具体实施还需根据文本的风格特征与译者能力进行调整。对于文化负载词或其他风格标记较为集中的文本，如地域文学作品、诗歌等，译者需特别留意模型在词汇选择上的规范化与趋同倾向，并结合具体语境进行辨析与调整；对于术语规范、表达相对固定的文本，如法律条文、操作指南等，模型生成在术语统一和用词一致性方面可能具有一定辅助价值，但仍需专业审校。在译者层面，经验较少者应警惕对模型常规化表达的依赖，具有较强文体意识和翻译经验的译者则可将模型用于词汇方案的比较与筛选。

4. 结论

本文以《呼兰河传》中文原文、葛浩文英译本和 GPT 英译文建立三方语料库，通过全书统计、分章比较和文化负载词译例分析，考察两部英译在词汇使用上的差异及其在具体语境中的表现。

研究发现，葛译在等长窗口内使用的不同词形较多，GPT 译文的非停用词比例较高，而且两个差异方向在七个主要章节中保持一致；两部译文的平均词长和长短词比例则十分接近。词性统计进一步表明，GPT 较高的近似词汇密度主要与冠词、代词和介词占比较低有关。在文化负载词的处理中，葛译较多根据文化表达在具体场景中的作用调整译项，GPT 较常保留源语词汇构成或直接说明具体所指，但两者在部分译例中也呈现出表层同形或功能趋同。

本文的主要价值在于将全书指标、分章复核和译例细读纳入同一分析过程。全书统计用于识别总体差异，分章数据用于检验差异是否贯穿作品，译例分析则考察具体词汇选择及其语境依据。这一方法可以减少节选语料和个别译例对研究结论的影响，也为人机文学译文的词汇风格比较和译者主体性研究提供了可复核的语言证据，并为文学翻译中的人机协作提供参考。人机协作方式应随文本特征与译者经验调整，风格标记密集的文本更依赖人工语境判断，程式化文本则可更多发挥模型在术语一致性方面的辅助作用；经验较少的译者应避免依赖模型的常规化表达，经验较丰富者可将其用于译项比较与检索。

本文仅考察一部文学作品的单一大模型译文，文化负载词分析也以代表性案例为主。后续研究可扩大作品、译者和模型的范围，对不同类别的文化负载词进行系统编码，并比较提示词、上下文长度和生成批次等条件对大模型译文词汇选择的影响。

参考文献

[1] 葛颂, 王宁. 人工智能时代的文学翻译: 挑战与机遇[J]. 外语与外语教学, 2024(1): 94-101, 149-150.

[2] 王克非. 智能时代翻译之可为可不为[J]. 外国语, 2024, 47(1): 5-9, 13.

[3] 赵衍, 张慧, 杨祎辰. 大语言模型在文本翻译中的质量比较研究——以《繁花》翻译为例[J]. 外语电化教学, 2024(4): 60-66, 109.

[4] 于蕾. ChatGPT 翻译的词汇多样性和句法复杂度研究[J]. 外语教学与研究, 2024, 56(2): 297-307, 321.

[5] 朱亚菲, 张继东. 《三体》人机英译风格对比研究——以刘宇昆、GPT-4 与火山翻译译本为例[J]. 外语电化教学, 2025(4): 88-96, 113.

[6] 梁君英, 刘益光. 人类智能的翻译能力优势——基于语料库的人机翻译对比研究[J]. 外语与外语教学, 2023(3): 74-84, 147-148.

[7] Baker, M. (1995) Corpora in Translation Studies: An Overview and Some Suggestions for Future Research. *Target. International Journal of Translation Studies*, **7**, 223-243. https://doi.org/10.1075/target.7.2.03bak

[8] Baker, M. (2000) Towards a Methodology for Investigating the Style of a Literary Translator. *Target. International Journal of Translation Studies*, **12**, 241-266. https://doi.org/10.1075/target.12.2.04bak

[9] Saldanha, G. (2011) Translator Style: Methodological Considerations. *The Translator*, **17**, 25-50. https://doi.org/10.1080/13556509.2011.10799478

[10] 萧红. 呼兰河传[M]. 杭州: 浙江人民美术出版社, 2017.

[11] Xiao, H. (1988) *Tales of Hulan River*. Goldblatt, H., Trans., Joint Publishing (H.K.) Co., Ltd.

[12] Covington, M.A. and McFall, J.D. (2010) Cutting the Gordian Knot: The Moving-Average Type-Token Ratio (MATTR). *Journal of Quantitative Linguistics*, **17**, 94-100. https://doi.org/10.1080/09296171003643098

[13] Nida, E. (1945) Linguistics and Ethnology in Translation-Problems. *Word*, **1**, 194-208. https://doi.org/10.1080/00437956.1945.11659254

[14] Nida, E.A. (1964) Toward a Science of Translating: With Special Reference to Principles and Procedures Involved in Bible Translating. E.J. Brill. https://doi.org/10.1163/9789004495746

[15] Venuti, L. (2008) The Translator's Invisibility: A History of Translation. 2nd Edition, Routledge.