

AI翻译与人工翻译可读性对比研究

——以第三十七届韩素音国际翻译大赛材料为例

何苓淋

四川大学外国语学院, 四川 成都

收稿日期: 2026年4月20日; 录用日期: 2026年5月26日; 发布日期: 2026年6月5日

摘要

随着大语言模型持续迭代并加速进入翻译实践场景, 围绕AI翻译与人工翻译质量差异的讨论在学界日益升温。本研究以第三十七届韩素音国际翻译大赛的英译汉散文节选为案例, 对大型语言模型(以DeepSeek为代表)生成的AI译文与大赛官方发布的参考(人工)译文进行了文本可读性对比分析。研究采用量化方法, 使用“AlphaReadabilityChinese”可读性分析工具, 从词汇、句法和语义三个维度下的九项具体指标进行测量。研究发现, AI译文的整体阅读难度高于人工译文。在细分维度上, 二者各有优劣: 人工译文在语义精确度和清晰度上表现更佳, 词汇选择更简约, 话题推进更集中; 而AI译文则在句法丰富度和语义丰富度上具有优势, 语言形式更多样, 话题覆盖更广。研究旨在从量化视角为AI翻译与人工翻译的质量对比提供新的实证数据, 并为翻译质量提升和译后编辑实践提供参考。

关键词

人工智能, 人机翻译对比, 文本可读性

A Comparative Study of Readability between AI Translation and Human Translation

—A Case Study of the 37th Han Suyin International Translation Contest Material

Linglin He

College of Foreign Languages and Cultures, Sichuan University, Chengdu Sichuan

Received: April 20, 2026; accepted: May 26, 2026; published: June 5, 2026

Abstract

As large language models continue to evolve and become more frequently used in translation, scholarly debate on the quality differences between AI and human translation has intensified. This study takes an English-to-Chinese prose excerpt from the 37th Han Suyin International Translation Contest as a case to conduct a comparative readability analysis between AI-generated translation (exemplified by DeepSeek) and the officially released reference (human) translation. It applies the AlphaReadabilityChinese tool to measure the two translations across nine specific indicators under the three dimensions: lexis, syntax, and semantics. Results show that the AI-generated translation exhibits a higher overall reading difficulty than the human translation. At the sub-dimensional level, each translation has its own strengths: the human translation performs better in semantic accuracy and semantic clarity, as reflected in simpler word choices and more focused topic development. The AI-generated translation shows relative advantages in syntactic richness and semantic richness, featuring greater linguistic variability and broader topic coverage. This study aims to offer new empirical data from a quantitative perspective for the quality comparison between AI-generated translation and human translation, and provide suggestions for improving translation quality and post-editing practices.

Keywords

AI, Human-Machine Translation Comparison, Text Readability

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

当前, 人工智能翻译系统通常可按其技术架构与工作机理区分为两类: 其一为神经机器翻译(Neural Machine Translation, NMT), 其二为生成式人工智能翻译(Generative Artificial Intelligence Translation, GAIT)。GAIT 以大型语言模型为底层支撑, 依托自回归或自编码机制生成目标语文本, ChatGPT3.5 与 DeepSeek 是其中较具代表性的系统。近年此类系统迭代频率显著提升, 对翻译行业的业务形态与生产流程形成持续冲击; 但与之相伴的是更为密集的争论与追问: AI 翻译与人工翻译在译文质量上究竟何者更优? 二者差异主要落在何处? 这些问题一方面关乎行业未来的岗位分工与协作方式, 另一方面也会直接影响跨文化传播的实际成效与质量控制。

长期以来, 对译文质量的考察, 即翻译质量评估, 一直是翻译研究的重要课题。主流的翻译质量评估方法可归为人工评估、有参考自动评估与无参考自动预估三类, 各有其优势与适用场景[1]。其中, 人工评估主要通过直接评分(如忠实度、流利度等维度)或错误扣分(如 MQM 多维质量指标体系)来评判译文质量, 而“流利度”作为核心维度之一, 衡量的是译文是否自然、通顺、易于理解, 本质上即文本可读性。与之对应, 有参考自动评估(如 BLEU、TER、METEOR 等)以参考译文为基准计算相似度, 侧重于词汇与句法层面的匹配; 无参考自动预估(如 QuEst++、COMET)则借助机器学习或神经网络模型直接预测译文质量, 无需参考译文。尽管上述自动评估方法并未显式使用“可读性”一词, 但其核心算法均不同程度反映了译文的流畅性与可理解性。由此可见, 文本可读性既是人工评估的独立指标, 也隐含在自动评估算法的语义匹配过程中, 因此, 未来研究可以将可读性提升为更显性的翻译评估指标。

从已有的翻译质量评估研究出发,越来越多学者开始关注到人工智能翻译性能以及人机译文差异,相关研究已形成一定规模。于蕾系统比较了 ChatGPT 译文、人工译文和 DeepL 译文的词汇多样性和句法复杂度[2]。还有的对比研究多以忠实度、流利度、术语一致性等维度作为评价重点[3][4],并据此指出 AI 翻译在语境识别、逻辑思维与中文特色表达处理方面存在相对固化的不足[5]。就研究方法而言,这类成果多围绕词汇与句法等表层特征展开质性、描述性分析,量化研究相对有限。也有学者将语言计量特征纳入分析框架[6]。在自动评价指标方面, BLEU 至今仍被广泛使用并构成机器翻译自动评价的重要工具。然而, BLEU 在同义词处理、权重分配以及召回率等方面的局限,使得研究者逐步转向多指标联合的综合评估思路。鉴于单一指标(如 BLEU)难以全面反映译文质量[7],在后续研究中,张曙康和赵朝永结合 BLEU、TER、BERTScore 等指标,对大语言模型与 NMT 的文学翻译质量进行了多维度评估[8]。

在上述背景下,为进一步引入更具操作性的量化视角,并在现有研究基础上补充综合评估指标,本研究拟从文本可读性维度切入,对 AI 翻译与人工翻译的译文质量进行对比分析。语料选取第三十七届韩素音国际翻译大赛英译汉节选材料作为原文,使用 AlphaReadabilityChinese1.0 可读性分析软件,对 DeepSeek 译文与人工译文进行测度与比较。研究聚焦两个问题:1) 以所选材料为例, AI 译文与人工译文在整体可读性水平上孰高孰低? 2) 在可读性的细分维度上,二者分别呈现怎样的差异?

2. 研究设计

2.1. 研究对象

本研究以第三十七届韩素音国际翻译大赛英译汉材料 *Scent Makes A Place* (节选)为例,原文长度为 455 个单词,并选取该材料的 DeepSeek 译文与人工译文作为对比分析对象(见附录)。之所以采用该材料,主要出于三方面的考虑。

首先,韩素音国际翻译大赛由《中国翻译》编辑部主办,在国内翻译学界被普遍视为具有权威性的赛事;其选材流程与评审标准较为成熟,因而在一定程度上能够为源文本及参评译文的质量提供保障。其次,该赛事材料通常具有一定的新颖性与独特性,往往并非已出版的经典名著或常见通用文本,网络环境中较难获得成熟且被广泛认可的“标准译本”。在此情形下,译文之间因既有译本传播而趋同的风险相对较低,有助于避免译文雷同,从而使对比分析更具客观性与真实性。再次,本届英译汉材料体裁为散文。与科技、法律等非文学文本相比,散文类文学文本对语言组织与表达的要求更高,也为译者的语言创造提供更大的空间;在此类任务中, AI 与人工译者的差异更容易显现,也更能直观反映两者在复杂语言处理上的不同表现。基于以上理由,本文将该赛事英译汉材料确定为原文,并在可读性维度上对 DeepSeek 译文与人工译文的翻译质量开展系统比较。

2.2. 研究方法

既往关于人机翻译质量差异的研究,通常围绕忠实度、流利度与术语准确性等维度展开,并借助人工评价或 BLEU 等自动化指标完成测评[3][9]。相较之下,译文作为目标语文本自身的可读性在相关研究中往往未被置于同等重要的位置。所谓译文可读性,主要指目标语读者理解译文所需付出的认知成本与难易程度;这一指标同样关系翻译质量判断,并会直接影响译文的传播效果及受众接受度。基于此,本文将文本可读性分析纳入研究框架,用以比较 AI 翻译与人工翻译在相关方面的差异。

本研究的数据测量采用雷蕾等开发的 AlphaReadabilityChinese1.0 (以下简称 ARC)汉语文本可读性分析工具[10]。ARC 面向汉语文本的可读性测度而设计,其研究思路在于回应传统可读性研究长期偏向英语语料、且多依赖词汇与句法等表层特征的倾向,从而对既有方法的适用边界作出补充。该工具从词汇、句法与语义三个维度搭建指标体系,最终输出九个维度的可读性指标值。

ARC 包含词汇丰富度、句法丰富度、名词语义精确度、动词语义精确度、名词与动词语义精确度、实词语义精确度、语义丰富度、语义清晰度以及语义噪音九项指标，由此从不同层面对文本的阅读难度作出量化表征，并为汉语可读性研究提供相对明确的测量依据。将 ARC 引入人机翻译对比研究，可在既有评测框架之外提供新的分析入口。本文据此对两类译文在词汇丰富度、句法丰富度与语义清晰度等指标上的差异进行量化呈现，以揭示 AI 翻译与人工翻译在文本可读性维度上的差别，并为人工智能语境下的翻译质量研究补充相应的实证依据。

3. 研究过程

3.1. 材料选取

本研究以第三十七届韩素音国际翻译大赛英译汉节选材料为原文语料。考虑到分析片段需要保持基本的语义完整性，同时 ARC 软件对文本长度亦有最低要求，本文选取原文开篇连续 455 个单词作为分析对象。

人工译文的来源为大赛官方网站公布的参考译文。该译文由赛事评审专家遴选产生，在参赛译文中具有较高代表性，亦可视作较高水平译文的体现。AI 译文则由 DeepSeek 生成：研究过程中将原文以纯文本形式复制粘贴至 DeepSeek 对话界面，并输入指令“请将这篇英文散文翻译成中文”，据此获得系统输出结果。为尽量减少提示语对译文生成的引导效应，本研究未在指令中加入任何与翻译策略相关的附加信息，从而使 AI 译文的生成尽可能避免人为因素的介入。

两份译文获得后，先对文本进行统一化处理，包括将全部内容转换为 TXT 格式，并将两份译文的标点加以规范与统一。完成预处理后，将两份译文导入 ARC 软件，进行可读性指标测量。

3.2. 数据导出

ARC 软件基于哈尔滨工业大学 LTP 工具进行分词与依存句法分析，并依托清华大学 OpenHowNet 语义知识库计算实词义项数，最终输出九个维度的可读性指标值[10]。依照雷蕾等对 ARC 各项指标的含义及其计算方向所作说明，以下测量结果可先分别从词汇、句法与语义三个维度加以解释(表 1)。

Table 1. Readability indicators of AI translation and human translation (rounded to three decimal places)

表 1. AI 译文与人工译文可读性指标(保留三位小数)

指标维度	指标名称	指标代码	AI 译文	人工译文
词汇*	词汇丰富度	lexical_richness	5.226	5.270
句法*	句法丰富度	syntactic_richness	2.201	2.163
语义	名词语义精确度	semantic_accuracy_n	3.583	3.815
语义	动词语义精确度	semantic_accuracy_v	6.863	7.723
语义	名词与动词语义精确度	semantic_accuracy_n_v	5.118	5.609
语义	实词语义精确度	semantic_accuracy_c	5.288	5.692
语义*	语义丰富度	semantic_richness_n	0.218	0.191
语义	语义清晰度	semantic_clarity_n	0.025	0.081
语义*	语义噪音	semantic_noise_n	8.024	37.219

注：带*指标：数值越大，文本难度越大，文本越难读；无标记指标：数值越大，文本难度越小，文本越易读。

先看词汇维度。AI 译文的词汇丰富度为 5.226，人工译文为 5.270。该指标以文本中所有单词的熵值计算而得，用于刻画词汇使用的不确定性或变化程度；数值越大，通常意味着词汇变化更大，阅读难度也更高。由此可见，人工译文在该项指标上略高，提示其用词变化相对更为丰富，对应的文本难度亦稍高。

再看句法维度。AI 译文的句法丰富度为 2.201，人工译文为 2.163。句法丰富度同样以熵为基础，通过计算文本中全部依存关系的熵值得到，反映句法结构的不确定性或变化程度；数值增大往往意味着句法结构更为多变，阅读难度随之上升。就结果而言，AI 译文该项略高于人工译文，说明其句法结构变化更为丰富，因而文本难度相对略高。

语义维度方面，ARC 共包含六项指标。其一为语义精确度系列指标，包括名词语义精确度、动词语义精确度、名词与动词语义精确度以及实词语义精确度。上述四项指标均以实词义项数的平均值为计算基础：义项数越大，通常对应词语更简单、更高频，文本更易读；义项数越小，则说明所用词语更难、更精确，阅读难度更高。数据表明，人工译文在四项语义精确度指标上均高于 AI 译文，这意味着人工译文更多使用简单、高频词语，其易读性更强。

其二为语义丰富度。该指标通过名词出现概率之和计算，用以反映文本话题的丰富程度；数值越大，话题越丰富，阅读难度也越高。本研究中，AI 译文语义丰富度为 0.218，人工译文为 0.191，显示 AI 译文的话题更为丰富，文本难度相对更高。

其三为语义清晰度。语义清晰度由名词出现概率的偏度计算，反映话题集中程度；数值越大，话题越集中，语义表达越清晰。人工译文该项为 0.081，明显高于 AI 译文的 0.025，说明人工译文在话题组织上更为集中，语义表达相对更清晰。

其四为语义噪音。该指标以名词出现概率的峰度计算，反映话题分布偏离重要话题的程度；数值越大，话题分布越倾向于不重要的话题，语义噪音越大。人工译文语义噪音为 37.219，显著高于 AI 译文的 8.024。该差异不宜忽略，后续章节将结合文本细读进一步分析其可能原因。

综合各项结果，两类译文在 ARC 九项可读性指标上均存在差异。人工译文在词汇丰富度、语义精确度系列指标以及语义清晰度方面数值更高；AI 译文在句法丰富度与语义丰富度上更为突出。语义噪音的差距尤为明显，或可提示两类译文在话题聚焦方式上呈现不同特征。

4. 研究结果

4.1. 整体可读性比较

可读性体现的是文本语言特征的综合结果，若仅凭单一指标，往往难以对其作出完整判断。ARC 从九个维度对文本进行测量，而不同指标的解释方向并不一致：有的指标随数值增大而对应更高的阅读难度，例如词汇丰富度、句法丰富度、语义丰富度与语义噪音；另一些指标则呈相反方向，数值越大通常意味着文本更易理解，包括语义精确度系列指标与语义清晰度。因而，在讨论整体可读性时，需要将各项指标的指向纳入同一分析框架之中，综合加以权衡。

结合本研究的测量结果，两类译文在多项指标上呈现出并行而分化的特征。在词汇丰富度和语义噪音两个维度人工译文数值较大，文本难度大；其余七个义项均显示 AI 译文较为难读。因此，总体来看，AI 译文文本难度较大，可读性较低。为更好地理解两类译文在不同语言层面的具体表现，后文将逐一分析九个指标的数值差异。

4.2. 各维度差异分析

4.2.1. 词汇维度

词汇丰富度用于刻画文本用词的变化幅度。结果显示，人工译文该项数值略高于 AI 译文，提示人工

译文在词语选择上更具多样性，可能借助同义表达的调配及词形层面的变动来提升语言层面的丰富度。但词汇变化的增加也意味着读者需要处理更多替代表达，认知负荷相应上升。与之相比，AI译文的用词更趋集中，变化较少，因此在词汇层面呈现出更明显的简约特征。

4.2.2. 句法维度

句法丰富度以依存关系的熵值反映句法结构的变化程度。对比可见，AI译文在句子组织上更为多变，具体表现为长短句的交错以及句式的转换等；人工译文整体走势相对平稳，句法复杂度较低。这种差异可能与AI模型训练中所吸收的多样句式资源有关，但结构变化的增强亦可能削弱译文的连贯与流畅。

4.2.3. 语义维度

语义维度由六项指标构成，用以同时呈现用词精确性与话题分布特征。就语义精确度系列指标而言，人工译文在四项指标上均高于AI译文：名词语义精确度(3.815 > 3.583)、动词语义精确度(7.723 > 6.863)、名词与动词语义精确度(5.609 > 5.118)、实词语义精确度(5.692 > 5.288)。这一结果表明，人工译文在实词选择上更倾向于采用常见义项，从而降低理解门槛；AI译文相对更多使用低频或多义词语，可能提升表达的精确性，但也会增加阅读难度。

语义丰富度以名词概率之和衡量话题丰富程度。AI译文该项高于人工译文，意味着其话题覆盖更为多元、信息密度更高，但也可能使读者注意力更易分散。语义清晰度用于反映话题集中程度，人工译文显著高于AI译文，说明其核心话题更突出，文本主线更易被读者把握。

语义噪音衡量话题分布中不重要话题所占比例。人工译文该项高达37.219，明显高于AI译文的8.024。如此差距需要特别关注。结合文本细读发现，人工译文中出现了大量修饰性、描写性的名词，如“馥郁”“馨香”“荒原”“喧嚣”“暗香”等。这些词汇虽增强了文学性，但大量低频名词的出现使得话题分布更为分散，语义噪音升高。AI译文则更聚焦于核心名词，如“沙漠”“气味”“植物”“嗅觉”等，语义噪音较低，但也因此少了几分文学韵味。就可读性而言，语义噪音过高容易干扰关键信息的提取，使文本呈现出更强的离散与杂散感。

5. 结论

本研究以第三十七届韩素音国际翻译大赛英译汉材料为语料，借助AlphaReadabilityChinese1.0可读性分析工具，从词汇、句法与语义三个层面对DeepSeek译文与人工译文进行量化对照。围绕研究问题，本文获得两方面结论。

第一，在整体可读性层面，AI译文可读性较低，表明当前大语言模型在处理文学散文时，产出的译文在阅读流畅性与理解便捷性方面仍不及人工译者。第二，在分项指标上，两类译文各有优劣：人工译文在语义精确度与语义清晰度方面更占优势，用词较为简明，话题组织亦更集中；AI译文在句法丰富度与语义丰富度方面数值更高，语言形式的变化性更强，话题呈现也更为多样。

两类译文在ARC九项指标上表现出的差异折射出人工翻译与AI翻译的不同取向。从翻译理论的角度审视，人工译文在语义精确度与清晰度上的优势，正契合了奈达提出的“功能对等”原则，即译者不仅传递信息，更关注译文读者的理解与接受，力求在目标语中产生与源语相近的阅读体验。人工译者通过简化用词、聚焦话题，主动降低读者的认知负荷，这正是对“动态对等”理念的实践。相比之下，AI译文在句法丰富度与语义丰富度上的高值，反映出其倾向于呈现训练数据中多样化语言模式，而非主动服务于读者的理解便利。这种现象或可视为一种隐性的“异化”策略，AI译文通过保留更多句法变化与话题跳跃，无意中呈现出一种“陌生化”的文本特质，尽管这种陌生化并非出于文化政治意图，而是源于统计学习对语言多样性的模仿。

然而, AI 的翻译模式在理论上也带来了新的挑战。传统翻译理论无论是功能对等还是异化/归化, 均预设了译者的主体性与意向性。而 AI 译文的“策略”并非有意识的选择, 而是数据驱动的概率输出。这意味着, 当机器翻译在句法丰富度等指标上“超越”人工译文时, 我们无法简单地用既有理论框架赋予其“异化”或“归化”的标签。AI 的翻译行为更应被理解为一种去主体化的语言转码, 它既不追求功能对等, 也不刻意异化, 而是以最大化训练数据中的语言模式匹配为目标。

同时需要说明, 本文仍存在一定局限。其一, 材料来源仅为单一散文文本, 样本规模有限; 其二, 结论的普适性仍需在更多文本类型中进一步检验。尽管如此, 本文仍具有一定意义: 可读性分析被用于人机翻译质量对比, 有助于在评估方法上补充量化视角, 并为相关研究提供新的实证例证。研究结果亦可为译文质量提升提供参考方向: 人工译者在保持语义清晰度的同时, 可适度控制语义噪音; AI 开发者则可针对语义精确度与话题集中性进一步优化模型表现。与此同时, 相关发现也可为译后编辑中的查漏补缺提供参照。

从行业发展看, 人工智能与人工翻译更可能走向深度融合与人机协同, 而非简单的替代关系。在此基础上, 翻译行业才可能在技术变革的持续推进中获得相对稳定的可持续发展, 并在跨文化交流与人类文明互鉴中发挥更大作用。

参考文献

- [1] 王均松, 庄淙茜, 魏勇鹏. 机器翻译质量评估: 方法、应用及展望[J]. 外国语文, 2024, 40(3): 135-144.
- [2] 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.
- [3] 李奉栖, 张云, 丁丽杰. 大语言模型与神经网络机器翻译系统专业文本翻译质量对比——以法律汉英翻译为例[J]. 上海翻译, 2025(6): 62-67.
- [4] 张文煜, 赵壁. 生成式人工智能开创机器翻译的新纪元了吗?——一项质量对比研究及对翻译教育的思考[J]. 北京第二外国语学院学报, 2024, 46(1): 83-98.
- [5] 蔡力坚. AI 工具在翻译中扮演什么角色[J]. 中国翻译, 2025, 46(4): 182-187.
- [6] 蒋跃. 人工译本与机器在线译本的语言计量特征对比——以 5 届韩素音翻译竞赛英译汉人工译本和在线译本为例[J]. 外语教学, 2014, 35(5): 98-102.
- [7] 王金铨, 文秋芳. 国内外机器自动评分系统评述——兼论对中国学生翻译自动评分系统的启示[J]. 外语界, 2010(1): 75-81, 91.
- [8] 张曙康, 赵朝永. 大语言模型之于文学翻译的适切性研究——基于多指标评估的《边城》多模型译文质量对比[J]. 外语教学与研究, 2025, 22(4): 85-95.
- [9] 李奉栖. 人工智能时代人机英汉翻译质量对比研究[J]. 外语界, 2022(4): 72-79.
- [10] 雷蕾, 韦瑶瑜, 刘康龙. AlphaReadabilityChinese: 汉语文本可读性工具开发与应用[J]. 外语教学, 2024(1): 83-93, 149.

附录

原文:

In the late spring, the desert smells like chocolate. It's fleeting, and it isn't everywhere in New Mexico, but sometimes, walking in the scrubland, it suddenly hits: a sweetness shimmering through the air. At first, I didn't know how to read this olfactory information, but now I can look for the source: yellow-petaled flowers with dark centers—chocolate daisies—blooming in the sun.

The American Southwest smells unlike anywhere I've lived before. It's better than the woods of Maine; it's far more fragrant. I didn't expect that when I moved here. I'm a perfume collector, so I had smelled the desert through art before I smelled it in person. Through my experience sampling "Mojave Ghost," "Arizona," and "Desert Eden"—perfumes designed to evoke cactus flowers and conifers—I thought the mesas would smell dusty and musky, with a little green cypress thrown in. I was wrong. Here, the plants hold their essences close to the stem out of necessity, only letting their oils free when it's safe to do so, when they're ready to be fertile.

Here, the sand bakes under the sun and the fragile soil releases its secrets with each step. Here, even my dog's urine is more potent, more fragrant on the wind, a louder yellow than I ever witnessed during our walks in Maine, blending uneasily with the grey rabbitbrush. It's wetter and stranger than I ever anticipated—complex, elusive, fecund.

After a year in Santa Fe, I've finally started to scratch the surface of knowing this landscape. But the learning is slow and requires all my senses, including the one most often forgotten. Unable to see or hear, smell became her primary way of reading the wider world; she lamented how that "most important" sense had been "neglected and disparaged" by the general populace, though she found it hard to communicate this knowledge to others. "It is difficult to put into words the thing itself," she wrote. "There seems to be no adequate vocabulary of smells, and I must fall back on approximate phrase and metaphor."

Perhaps due to our trouble translating scents into language, it was once common wisdom that human noses were weak, shoddy things compared to our animal friends. Time and research have challenged that paradigm. Although we don't have the smart wet noses of dogs or the large nasal chambers of reptiles, humans can discriminate between an estimated 1 trillion different odors; the myth of our poor olfaction is rooted in the Victorian distaste for all things scented and the eschewing of all things bodily. In other eras people assumed that odd smells were evidence of impending illness, ghostly presences, or moral failings, rather than taking them for what they were.

人工翻译:

暮春时节，沙漠里弥漫着巧克力般的气味。这种气味转瞬即逝，即便在新墨西哥州也并非在哪里都能闻到。然而，当我穿行在灌木林地中，空气中有时会突然有一缕香甜的气息扑面而来。起初，我并不知道这缕香气从何而来，如今却已寻得它的源头——几株黄色花瓣深色花心的巧克力雏菊正在阳光下悄然绽放。

美国西南部的的气味，与我以往所居住过的地方都大不相同。它比缅因州的森林更胜一筹，这里的香气更加馥郁。初来此地时，我未曾料想会是如此。作为香水收藏家，在来此之前，我曾闻过沙漠的气味，只是那香气出自调香师的巧手。我品鉴过一些以仙人掌花和针叶树为主调的香水，比如“荒漠孤魂”“亚利桑那”和“沙漠伊甸”，便让我误以为该台地会是尘土气与麝香交织，间或夹杂着些许翠柏的气味。我

错了。

在这里，植物出于生存的需要，往往将其精华存储于茎部，只有当它们准备繁衍且感觉安全时，才会释放出油脂。在这里，烈日炙烤着黄沙，每走一步，脚下松软的土壤都会轻声诉说深藏的秘密。在这里，连我家狗狗尿液的气味都更加浓烈，随风飘散，格外刺鼻，且颜色更深更黄，与周围灰绿色的金菟木格格不入，这与我在缅因州遛狗时所闻所见截然不同。总之，这里比我想象的更湿润，更奇特：复杂难解、变幻莫测，却又生机勃勃。

在(新墨西哥州首府)圣达菲生活了一年之后，我终于对这片土地的风貌有了一定的了解。不过，认知过程非常缓慢，且需要调动所有的感官，包括嗅觉。海伦·凯勒自幼因病失明失聪，因此嗅觉成为她感知广阔世界的主要方式；她时常悲叹这一“最重要”的感官竟被世人无端“忽视、贬低”，却又苦于难以向他人传递自己的体悟。“气味本身难以言表，”她写道，“英语中似乎又缺乏描述气味的词汇，因此，(描述气味时)我不得不借助一些模糊的词语和比喻。”

或许由于气味本身难以用语言描述，人们曾经普遍认为与动物相比，人类的嗅觉十分拙劣，不堪大用。但随着时间的推移，相关研究已颠覆了这一传统认知。人类虽然没有狗那样湿润灵敏的鼻子，也没有爬行动物宽大的鼻腔，但我们却能分辨出约1万亿种不同的气味。关于人类嗅觉迟钝的误解，实际上源自维多利亚时代人们对一切有香味的物品的厌恶，以及对所有感官享乐的排斥。在其他历史时期，人们则将异味视为疾病将至、幽灵缠身或道德败坏的征兆，所以无法客观地看待它们。

AI 翻译：

暮春时节，沙漠里会飘起巧克力香。这气息转瞬即逝，在新墨西哥州也并非处处可寻，但有时漫步灌木丛地，它会突然袭来：一丝甜意在空气中荡漾。起初我不知如何解读这嗅觉讯息，如今却能寻踪觅源——那些绽放在阳光下的深色花心、黄色花瓣，正是巧克力雏菊。

美国西南部的气息与我过往居住之地迥然不同。它胜过缅因州的森林，芬芳远胜于此。初至此地时我未曾料到。身为香水收藏家，在亲身体验沙漠之前，我已通过艺术嗅过它的气息。通过试闻那些以仙人掌花和针叶树为灵感的“莫哈韦幽灵”“亚利桑那”“沙漠伊甸园”，我曾以为这片台地会是尘土与麝香交织，间或夹杂丝柏的绿意——我错了。此处的植物不得不将精华紧锁茎中，只在安全时刻，在准备孕育之时，才让油脂释放。

这里的沙砾在烈日下炙烤，脆弱的土壤随脚步吐露秘密。就连我家狗狗的尿液都更浓烈，随风飘散着更张扬的黄色讯息，与灰蒙蒙的兔刷花形成不安的混响，远胜我们在缅因州散步时的所见所闻。这里比我想象的更湿润、更奇异——复杂难辨，难以捉摸，生机盎然。

在圣达菲住满一年，我终于开始触及这片土地的皮毛。但这认知过程缓慢，需要调动所有感官，包括最常被遗忘的那个感官。因失明失聪，嗅觉成了她解读世界的主要途径；她哀叹这“至关重要”的感官竟被大众“忽视贬低”，却又难以向他人传达这种认知。

“事物本身难以言说，”她写道，“仿佛没有足够的嗅觉词汇，我只能借助近似的措辞与隐喻。”或许正因难以将气味转译成语言，人们曾普遍认为人类鼻子相较动物朋友是孱弱粗劣的器官。时间与研究已颠覆此范式。虽无狗儿灵敏湿润的鼻子，也无爬行动物宽大的鼻腔，人类却能辨别约一万亿种不同气味；我们嗅觉低劣的迷思，实植根于维多利亚时代对一切芬芳之物的厌恶，以及对一切身体之物的回避。在其他时代，人们将异香视为疾病将至、幽灵现身或道德败坏的征兆，而非坦然接受其本真。