译介与原生汉语癌症健康信息沟通质量分析

赵梦丹 黄日柳

成都理工大学外国语学院，中国·四川 成都 610059

摘要：互联网健康科普是癌症知识传播的重要载体，但现有互联网医疗沟通质量研究多将沟通质量视作可测量的结果属性，较少从话语－语言层面揭示原生与译介健康文本的语言实现差异。本文选取权威平台原生汉语、国际机构译介汉语癌症科普文本各 100 篇，自建可比语料库，融合语料库辅助话语分析与翻译普遍性理论，从词汇、句法、语篇维度对比词汇密度、被动结构、连接词、文化负载成语的特征差异，探究语言选择对信息可及性、逻辑显化、文化适配三大沟通质量维度的影响。研究表明，译介文本未出现翻译普遍性所指的词汇简化，反而词汇密度更高，在保障术语精确性的同时加重受众认知负担；两类文本被动句法特征趋同，译者实现句法规范化；译介文本依靠高频连接词实现逻辑显化，但大量删减文化成语，提升清晰度的同时弱化文化亲和力。据此构建网络癌症健康信息话语评价模型，从分层内容供给、医学翻译规范、自动化检测工具三方面提出优化路径。本研究拓展翻译普遍性在中文健康传播的应用，证实沟通质量差异可通过量化语言特征开展经验识别，为网络健康信息编审、翻译及平台治理提供实证支撑。

关键词：互联网医疗沟通；沟通质量；语料库话语分析；癌症健康信息；翻译普遍性

Analysis of the Quality of Cancer Health Information Communication in Translated and Native Chinese Materials

Zhao Mengdan, Huang Riliu

School of Foreign Languages, Chengdu University of Technology, China Sichuan Chengdu 610059

Abstract: Online popular-science health resources constitute a vital vehicle for cancer-related knowledge dissemination. Nevertheless, existing research on internet-mediated medical communication quality mostly treats communication quality as a measurable outcome attribute and rarely reveals the linguistic disparities between native and translated health texts from a discourse-linguistic perspective. This study constructs a self-built comparable corpus consisting of 100 native-Chinese cancer popular-science texts from authoritative domestic platforms and 100 translated-Chinese counterparts sourced from international institutions. Integrating corpus-assisted discourse analysis with the theory of translation universals, it compares lexical density, passive constructions, connectives and culture-loaded idioms across lexical, syntactic and textual dimensions, and explores how linguistic choices influence three communication-quality dimensions: information accessibility, logical explicitation and cultural adaptability. The findings indicate that translated texts do not exhibit the lexical simplification postulated by translation universals; instead, they carry higher lexical density, which guarantees terminological precision yet increases audiences' cognitive load. The two text types converge in passive-syntactic features, demonstrating translators' achievement of syntactic normalization. Translated texts realize logical explicitation via high-frequency connectives, yet drastically cut culture-specific idioms, enhancing textual clarity at the expense of cultural affinity. Accordingly, a discourse-evaluation model for online cancer health information is established, with optimization strategies proposed from three aspects: tiered content provision, medical translation norms and automatic detection tools. This research extends the application of translation universals to Chinese-language health communication, verifies that disparities in communication quality can be empirically identified through quantifiable linguistic features, and provides empirical support for the editing, reviewing, translation and platform governance of online health information.

Keywords: Internet-mediated medical communication; Communication quality; Corpus-assisted discourse analysis; Cancer health information; Translation universals

0 引言

互联网医疗已经成为我国公共卫生服务体系的重要补充。截至 2023 年 6 月，我国在线医疗用户规模达到 3.64 亿人，互联网平台成为公众获取癌症防治知识的首要渠道。癌症属于高敏感、高风险病种，公众对准确、可理解的癌症健康信息存在刚性需求；国际权威医学知识经由翻译流入中文网络，译介文本与本土原生文本共同构成网络癌症信息生态。但现有研究大多将互联网健康信息质量视作一

套可打分的静态属性，依靠可读性公式、问卷评价完成质量测评，对“沟通质量如何通过语言选择得以实现”这一基础问题关注不足。换言之，学界已经确认优质健康信息对健康素养、健康行为的积极作用，却对专业精确性与公众可读性之间的语言张力缺少精细化考察，对译介文本带来的话语变异缺少系统的语料库证据。这一理论缺口造成沟通质量研究长期停留在结果评价，难以解释：专业准确与通俗易懂之间的语言边界在哪里？翻译过程如何重塑健康文本的信息负荷、逻辑表达与文化适配性？

上述问题在癌症健康信息场景下尤为突出。癌症受众包含大量患者、家属，群体健康素养参差不齐，文本的词汇负荷、句法选择、修辞表达直接决定信息能否被有效接收，甚至影响风险感知与治疗依从性。网络科普文本以书面文字为唯一载体，缺少面对面沟通的副语言补偿，语言选择几乎承载全部沟通质量的实现功能。因此，从话语 – 语料库视角揭示原生与译介健康文本的语言机制，既具备健康传播的理论价值，也对网络健康内容编审、医学翻译实践、平台质量管控具备直接实践意义。

基于该问题意识，本文以网络癌症健康科普文本为研究对象，整合语料库辅助话语分析与翻译普遍性理论，考察原生汉语与译介汉语文本在词汇、句法、语篇层面的话语特征，揭示语言选择如何生成沟通质量，构建面向癌症网络健康信息的话语评价模型，并提出可落地的提升路径。

1 相关研究现状

1.1 互联网健康信息研究：外部影响因素积累丰富，话语内部机制考察有限

互联网健康信息研究早期集中于信息质量评估、信息搜寻行为、用户采纳与满意度等外部变量层面。已有研究证实信息可信度、文本可读性、平台权威性显著影响公众的信息接受与健康行为。国内研究进一步指出，网络健康信息存在术语晦涩、表述不规范等现实问题，信息不平等现象客观存在。

翻译是跨境健康知识流动的重要路径，大量国际癌症防治材料经翻译进入中文互联网。现有医学翻译研究多聚焦术语一致性、翻译技巧，肯定翻译对健康知识传播的价值。但多数研究停留在宏观评述，较少进入文本内部：哪些语言特征发生系统性偏移？这些偏移如何作用于沟通质量？已有成果回答了哪些因素影响信息效果，但尚未充分回答沟通效果在语言实践中如何生成。这正是本研究试图填补的缺口：不再止步于验证可读性与受众感知的相关性，而是追问质量差异背后具体的词汇、句法、语篇选择。

1.2 沟通质量评价研究：测量取向成果丰厚，话语实现机制存在缺位

健康信息沟通质量评价主要依托两大脉络：一是可读性测量传统，依赖 Flesch–Kincaid、FOG 等可读性公式，从句子长度、词汇难度开展量化评估；二是以受众为中心的沟通框架，强调信息可理解性、行动可操作性、文化适配性。国内学者也开发本土化的健康材料评价工具。

既有评价范式共同特征是测量取向：沟通质量被当作可以打分、评级的静态属性，而非在文本生成（含翻译）过程中建构出来的话语成就。这就造成沟通质量研究积累大量“质量高低”的判断，却缺少“质量何以形成”的语言学解释。对于完全依托书面文本的网络科普，该缺陷尤为明显：文本就是沟通本身，语言选择直接定义沟通质量。本文试图突破单一测量取向，将沟通质量界定为可通过可观测语言特征识别的话语产物。

1.3 翻译普遍性研究：通用语域证据较多，健康专业语域拓展不足

翻译普遍性是描写翻译学的核心命题，包含简化、显化、规范化三大经典假说。简化指译文倾向使用更为简单的语言；显化指译文倾向把隐含逻辑关系显性化；规范化指译文趋向贴合目标语典型表达习惯。该理论已经在文学、新闻等通用语域积累大量可比语料证据。

部分汉语可比语料研究发现，译介汉语普遍表现出连接词增多、部分词汇简化等特征。但面向医学健康这类高度专业化语域，翻译普遍性是否依然成立，尚缺少充分实证。健康翻译的首要目标是医学准确性，译者会优先保障术语严谨，这很可能改写通用语域下的翻译普遍特征。现有健康翻译的语料库研究大多聚焦临床病历、试验材料，面向公众的癌症网络科普的对比研究仍然薄弱。

总体而言，翻译普遍性理论多应用于独白式翻译文本，但较少用于健康传播场景下的沟通质量分析。本文将翻译普遍性与语料库话语分析相结合，考察原生、译介两类独白式健康科普文本，剖析专业语域中简化、显化、规范化的具体表现，弥补该领域的研究缺口。

2 研究设计

2.1 语料库构建

本研究采用可比语料库，分为原生汉语语料库、译介汉语语料库。原生汉语文本取自中国抗癌协会、中国健康教育网；译介汉语文本取自世界卫生组织中文网、默沙东诊疗手册中文版、国家癌症防控网。两类语料库各 100 篇面向普通公众的癌症科普文本，剔除科研论文、商业广告，原生语料总字数 29381 字，译介语料总字数 28725 字，样本规模基本均衡。

语料处理分为采集、清洗、分词三个环节。采集完成后开展人工文本清洗，剔除网页噪声，统一格式；之后使用 HanLP 完成中文分词与词性标注。语料检索、词频统计、索引行提取依托 AntConc4.3.1 开展；统计检验使用

SPSS 完成。全部文本均为网络公开科普材料，不涉及个人隐私信息。

2.2 分析框架与研究假设

本研究将翻译普遍性三假说（简化、显化、规范化）与互联网健康信息三大沟通质量维度（信息可及性、逻辑显明度、文化适配度）进行对接，形成分析框架。

表 1 原生 – 译介癌症健康信息整合分析框架

沟通质量维度	核心关注	理论来源	语言观测指标
信息可及性	信息负荷高低，专业精确性与受众理解难度的平衡	简化假说	词汇密度（实词 / 总词）
逻辑显明度	文本因果、条件等逻辑关系是否清晰外显	显化假说	各类连接词标准化频次
文化适配度	句法表达贴合汉语习惯；修辞表达与本土文化语境兼容	规范化假说	被字句频次；成语（文化负载词）标准化频次

据此提出研究假设：H1：译介汉语文本词汇密度低于原生汉语，体现翻译的简化特征；H2：译介汉语文本连接词频次高于原生汉语，体现翻译的显化特征；H3：译介汉语文本被字句频次高于原生汉语，受源语英语被动结构牵引；H4：译介汉语文本成语频次低于原生汉语，表现文化修辞层面的简化。

分析分为两个阶段。第一阶段：借助 AntConc 输出词表、词频，对词汇密度、被字句、连接词、成语做标准化统计，使用独立样本 t 检验开展组间对比。第二阶段，结合索引行做质性话语解读，将统计结果与沟通质量建立阐释关联。

3 分析与发现

3.1 词汇层面：译介文本词汇密度更高，简化假说在专业语域不成立

词汇密度即实词占总词的比重，反映文本信息负荷。统计与 t 检验结果显示：原生汉语词汇密度均值 0.532；译介汉语均值 0.568，两组存在统计学显著差异（$t=-4.663$，$p<0.001$）。译介文本词汇密度显著更高，H1 不成立。这与文学、新闻语域观察到的“译文简化”现象相悖。在癌症健康科普这一专业语域，译者优先保障医学术语的完整性与精确性，不愿为降低词汇密度删减专业概念，由此带来更高信息负荷。更高词汇密度可以强化专业可信度，但客观上抬升普通受众的认知负担，可能加剧健康信息获取鸿沟。由此可见，“词汇密度越低质量越好”这一通用逻辑不能直接套用于医学翻译沟通。

3.2 句法层面：被动（被）字句使用趋同，句法规范化得到实现

以每万字频次统计被动标记“被”：译介汉语每万字 17.06 次，原生汉语每万字 17.70 次，二者差异微弱。H3 不成立。英语源文本大量使用被动语态，但译者并未机械移植被动句式，主动转换为汉语惯用表达，实现翻译普遍性中的规范化。两类文本被动结构整体处于较低水平，契合健康传播“多用主动表达，提升可读性”的实践原则，这是译介癌症信息句法层面的积极特征。

3.3 语篇层面：连接词显化与成语修辞简化并存

3.3.1 连接词：译介文本逻辑显化得到验证

标准化频次显示，译介文本中“和”“并”“以及”等连接词频次整体高于原生文本。其中“和”译介每万字 127.41 次，原生仅 77.58 次。高频连接词把源文隐含的逻辑链条显性标记，验证显化假说，H2 成立。显化策略有利于厘清复杂医学因果关系，降低受众梳理逻辑的心智成本，对信息可及性形成正向贡献。但连接词过度堆砌也会造成文本臃肿，是医学翻译实践需要警惕的问题。

3.3.2 成语（文化负载词）：译介文本大幅简化

原生汉语语料识别 26 个成语，每万字 8.85 次；译介汉语仅识别 3 个成语，每万字 1.04 次。H4 成立。原生科普善用本土成语完成风险警示、行为倡导，增强文本感染力；译介文本则对文化负载修辞采取规避策略，追求直白无歧义。该简化策略降低跨文化理解障碍，但也牺牲文本的文化亲和力，造成话语表达的中性化、疏离化。

3.4 语言特征与沟通质量的话语生成路径

综合以上结果，可以归纳癌症网络健康信息的话语生成路径，如表 2。

表 2 网络癌症健康信息沟通质量的话语生成路径

文本类型	话语 – 语言特征	生成的沟通质量维度	潜在局限
原生汉语科普	中等词汇密度；中等连接词；较多成语修辞	信息可及性；文化适配度	部分成语提升理解门槛；部分表述专业严谨度不足
译介汉语科普	高词汇密度；高频连接词；低成语频次；被动句法规范化	逻辑显明度；句法适配性	信息负荷偏高；文化亲和力缺失

沟通质量不是单一指标得分，而是多重话语特征博弈形成的综合结果。译介文本可以通过连接词显化优化逻辑理解，但高词汇密度又会削弱信息可及性；句法上实现规范化，但文化修辞层面发生简化。这种内在张力，是互联网癌症健康信息沟通的核心矛盾。

4 互联网癌症健康信息的话语模型与提升路径

4.1 话语模型

基于上述分析，本文提出网络癌症健康信息的话语模型。其核心命题：网络健康信息沟通质量并非静态的服务属性，而是词汇、句法、语篇多层语言特征相互博弈形成的话语成就。

信息可及性主要由词汇密度塑造。译介文本出于专业精确性诉求推高词汇密度，形成“精确性 – 可读性”的内在冲突；专业语域下翻译简化并不必然体现为词汇简化，更多体现为文化修辞层面的简化。逻辑显明度依托显化策略实现。译介文本依靠增加显性连接词把隐含因果关系外化，降低受众逻辑推理负担，是健康翻译的有益策略，但要把握适度原则。文化适配度由规范化与修辞简化共同塑造。在句法层面，译者能够顺应汉语表达习惯完成规范化；但在文化修辞层面，为规避歧义大量舍弃成语，造成文本亲和力流失。

三大质量维度之间相互牵制。追求术语精确可能损伤可及性；追求逻辑显化可能带来语句冗余；追求无歧义会牺牲本土修辞感染力。沟通质量提升不是单向追求某一项指标最优，而是多重约束下的动态平衡。

4.2 实践提升路径

立足于话语模型，从内容供给、翻译规范建设、质量评价三个层面提出提升路径。第一，内容供给层面，推行分层信息发布机制。区分专业版与公众版两套文本。专业版保留完整术语与高词汇密度，面向医护、高健康素养受众；公众版在保证医学事实准确前提下，适度控制词汇密度，审慎使用成语等文化负载修辞。平台应当明确标注两套版本的适用人群，满足差异化信息需求。

第二，医学翻译实践层面，建立中文健康信息翻译指南。明确三条原则：①适度显化：在逻辑跳跃处增补连接词，拒绝无意义堆砌；②句法规范化：将英文被动句式转化为汉语主动或无标记被动；③文化负载词弹性处理：对高歧义成语予以替换，适度保留低歧义、通俗易懂的本土修辞，兼顾清晰度与文化亲和力。该指南可以作为互联网平台采购翻译服务的质量标尺。

第三，质量评价层面，推动评价逻辑从单纯“可读性测量”走向话语取向评价。在传统可读性指标之外，增加语料库可自动提取的话语指标，构建混合评价体系。

表 3 网络癌症健康信息话语取向评价指标方案

评价维度	核心指标	数据来源	测量方式
信息可及性	词汇密度；专业术语占比	全文本	分词统计 + 词频测算
逻辑显明度	连接词标准化频次；冗余连接词占比	全文本	标准化词频统计 + 索引行人工核验
文化适配度	被动（被）字句频次；成语等文化负载词密度	全文本	标准化频次统计

话语取向指标不会取代传统可读性工具，而是形成互补。传统工具回答“信息好不好读”；话语指标回答“为什么不好读，问题出在哪些语言层面，可以如何修改”，实现可诊断、可干预的质量评价。可将上述指标集成，开发自动化检测工具，嵌入平台内容编审流程，实现文本上线前的话语质量反馈。

5 结语

本文以各 100 篇原生、译介癌症网络健康科普文本构建可比语料库，结合语料库辅助话语分析与翻译普遍性理论，考察两类文本的语言特征及其沟通质量效应。主要发现如下：第一，在癌症健康这一专业语域，翻译普遍性的简化假说并不体现为词汇层面简化，译介文本词汇密度反而更高；简化更多表现在文化修辞（成语）层面；显化假说在连接词维度得到验证；句法层面实现规范化。第二，互联网癌症健康信息沟通质量是多重语言特征博弈的话语成就，译介文本存在“专业精确性 – 公众可读性”的内在张力，连接词带来逻辑增益的同时，高词汇密度、修辞缺失会损害信息可及性与文化亲和力。第三，面向网络健康信息，应当建立话语取向的混合评价框架，推动分层内容供给、医学翻译规范建设、自动化话语检测工具落地。理论层面，本研究把翻译普遍性拓展到中文互联网健康传播的独白科普文本，证明沟通质量差异可以依托语料库语言特征开展经验识别；实践层面，为网络癌症科普的编审、翻译、平台质量管控提供语言学依据。

本研究仍存在局限：语料来源局限于若干权威网站，结论普适性有待进一步检验；研究聚焦文本语言特征，尚未纳入受众阅读反应的实证数据。未来研究可以扩大病种与平台范围，结合受众实验，将文本话语特征与受众理解、风险感知、健康行为开展三角互证，进一步完善互联网医疗沟通质量的话语模型。

参考文献：

[1] 黄立波，王克非. 基于语料库的翻译研究发展阶段与前沿趋势分析[J]. 外语教学与研究，2023,55(5): 764−776.

[2] 朱长富，李茹. 基于语料库的庞德英译中国典籍译者风格研究[J]. 外语教育，2023,44(4): 75−82.

[3] Baker, M. Corpus linguistics and translation studies: Implications and applications[A]. In M. Baker, G. Francis & E. Tognini−Bonelli (Eds.), Text and Technology: In Honour of John Sinclair[C]. Amsterdam: John Benjamins, 1993: 233−250.

[4] Baker, M. Corpus linguistics and translation studies: Implications and applications[A]. In K. H. Kim & Y. Zhu (Eds.), Researching Translation in the Age of Technology and Global Conflict: Selected Works of Mona Baker[C]. London: Routledge, 2020a: 9−24.

[5] Baker, M. Corpus−based translation studies: The challenges that lie ahead[A]. In K. H. Kim & Y. Zhu (Eds.), Researching Translation in the Age of Technology and Global Conflict: Selected Works of Mona Baker[C]. London:

Routledge, 2020b: 44-54.

[6] Baur, C. & Prue, C. The CDC Clear Communication Index is a new evidence-based tool to prepare and review health information[J]. Health Promotion Practice, 2014, 15(5): 629-637.

[7] Chapple, A., et al. "Getting the words right": A qualitative study of language translation in clinical trials[J]. Sociology of Health & Illness, 2021, 43(5): 1187-1203.

[8] Chen, J. J. Explicitation through the use of connectives in translated Chinese: A corpus-based study[D]. University of Manchester, 2006.

[9] HanLP: Han language processing[EB/OL]. GitHub, 2024. https://github.com/hankcs/HanLP.

[10] Kreps, G. L. & Bonaguro, E. W. Health communication as applied communication inquiry[A]. In L. R. Frey & K. N. Cissna (Eds.), Routledge Handbook of Applied Communication Research[C]. London: Routledge, 2009: 410-429.

[11] Kruger, H. Explicitation in the translation of health texts: A corpus-based study[J]. Journal of Specialised Translation, 2017(28): 22-40.

[12] Lau, T. Chinese readability analysis and its applications on the internet[D]. The Chinese University of Hong Kong, 2006.

[13] Laviosa, S. Core patterns of lexical use in a comparable corpus of English narrative prose[J]. Meta: Translators' Journal, 1998, 43(4): 557-570.

[14] Li, X. & Wang, Y. Readability assessment and cultural adaptation of online health information: A comparative study[J]. Journal of Medical Internet Research, 2022, 24(4): e34567.

[15] Mauranen, A. Corpora, universals and interference[A]. In A. Mauranen & P. Kujamäki (Eds.), Translation Universals: Do They Exist?[C]. Amsterdam: John Benjamins, 2004: 65-82.

[16] McEnery, T. & Xiao, R. Parallel and comparable corpora: What is happening?[A]. In G. Anderman & M. Rogers (Eds.), Incorporating Corpora[C]. Clevedon: Multilingual Matters, 2007: 18-31.

[17] Olohan, M. Introducing corpora in translation studies[M]. London: Routledge, 2004.

[18] Pang, S. A diachronic and multi-genre corpus-based investigation of "translation universals" [J]. Foreign Languages and Their Teaching, 2024(5): 132-144, 150-151.

[19] Pöchhacker, F. Introducing interpreting studies (2nd ed.)[M]. London: Routledge, 2016.

[20] Shoemaker, S. J., Wolf, M. S. & Brach, C. Development of the Patient Education Materials Assessment Tool (PEMAT): A new measure of understandability and actionability for print and audiovisual patient information[J]. Patient Education and Counseling, 2014, 96(3): 395-403.

[21] Stubbs, M. Lexical density: A computational technique and some findings[A]. In M. Coulthard (Ed.), Talking about Text: Studies Presented to David Brazil on His Retirement[C]. Birmingham: English Language Research, University of Birmingham, 1986: 27-48.

[22] Wu, J., et al. A contrastive study of syntactic complexity in translated and original Chinese[J]. Foreign Language Teaching and Research, 2023, 55(2): 263-275, 320-321.

[23] Xiao, R. How different is translated Chinese from native Chinese? A corpus-based study of translation universals[J]. International Journal of Corpus Linguistics, 2010, 15(1): 5-35.

[24] Xiao, R., He, L. & Yue, M. In pursuit of the "third code": Using the ZJU corpus of translational Chinese in translation studies[A]. In R. Xiao (Ed.), Using Corpora in Contrastive and Translation Studies[C]. Newcastle: Cambridge Scholars Publishing, 2010.

[25] Xu, J. & Guo, T. Simplification as a translation strategy in public health communication: A corpus-based study of English-Chinese translation[J]. Translation and Interpreting Studies, 2023. Advance online publication.

[26] Zhu, C. & Li, R. A corpus-based analysis of Ezra Pound's translator style in English translations of Chinese classics[J]. Foreign Language Education, 2023, 44(4): 75-82.

基金项目：教育部人文社会科学研究青年基金项目"医疗改革背景下互联网医疗沟通质量的综合评价与提升路径研究"（项目编号：22YJCZH258）。

作者简介：赵梦丹（1992-），女，中国湖北黄石人，博士，讲师，研究方向：健康传播、语料库语言学、翻译研究。