

语料库的多维特征及其应用于翻译实践的功能价值

裴海同

(黑龙江大学, 哈尔滨 150080)

提 要: 全球化的背景下, 翻译实践功能在跨语言、跨文化交流中发挥重要作用。本文通过系统探究通用单语语料库、平行语料库与可比语料库的多维特征, 以及三者功能互补, 共同构成译者的综合资源库, 阐释这三类语料库在翻译实践中的功能价值。研究表明: 通用单语语料库为源语和目标语的深度理解提供权威资源, 平行语料库提供特定语境下的等效表达范例, 可比语料库揭示语言单位的功能分布与原生表达差异。三类语料库功能互补: 单语语料库深化语言理解, 平行语料库提供翻译实例, 可比语料库检验译文可读性。译者可以根据需要策略性地选择和组合三类具有互补功能的语料库工具, 解决各类翻译难题, 实现有效翻译, 提高翻译质量。本文指出智能化融合、资源开放化专业化是未来关键方向。

关键词: 通用单语语料库; 平行语料库; 可比语料库; 翻译实践; 组合策略; 翻译质量

中图分类号: H315.9

文献标识码: A

1 引言

全球化进程不断加速的当下, 翻译活动在跨语言、跨文化交流中扮演着愈发关键的角色。翻译质量的高低直接影响着信息传递的准确性与有效性, 关系到不同文化间能否实现深入理解与交融。随着计算机技术和信息技术的迅猛发展, 作为一种强大的语言研究与应用工具, 语料库逐渐在翻译领域崭露头角并得到广泛应用。语料库可对海量真实语言数据进行系统收集、整理与分析, 从而为译者提供丰富且真实的语言参考。为更好地探究语料库对于翻译实践的功能价值, 本文从语料库的多维特征分析入手, 全面阐释各类型语料库用于翻译实践的功能价值, 进而提出语料库用于翻译实践的切实可行的组合策略, 旨在为译者提供更科学、高效的翻译实践指导, 提升翻译质量与效率, 推动翻译学科在新时代背景下借助语料库技术实现创新发展, 更好地服务于跨文化交流需求。

2 语料库的多维特征

语料库有多种类型, 确定类型的主要依据是它的研究目的和用途, 这一点往往能够在语料采集的原则和方式上有所体现。按照语料的语种, 语料库可以分成单语的、双语的和多语的。按照语料的采集单位, 语料库又可以分为语篇的、语句的、短语的。双语和多语语料库按照语料的组织形式, 还可以分为平行(对齐)语料库和比较语料库。我们主要介绍通用单语语料库、可比语料库与平行语料库的典型特征。

2.1 通用单语语料库

通用单语语料库, 指代表特定国家或地区语言整体面貌的大型、平衡、深度标注的单语电子文本集合。此类语料库的典型代表有: 英国国家语料库(BNC)、美国国家语料库(ANC)、

美国当代英语语料库 (COCA)、俄语国家语料库 (Национальный корпус русского языка)、北京大学 CCL 语料库和国家语委现代汉语通用平衡语料库, 等等。

通用单语语料库因规模宏大令人瞩目, 通常涵盖数亿词汇量, 犹如一座语言的宝库, 广泛收纳了各种语言使用场景下的文本。无论是日常的生活对话、校园里的学术讨论, 还是职场中的商务交流、专业领域的研究报告, 都能在其中找到对应的文本范例, 为语言研究提供了全面而丰富的数据基础。在标注方面, 通用单语语料库一般都进行了词性标注、语法分析等多维度标注。词性标注能够明确每个词汇的词性, 如名词、动词、形容词、副词等, 使研究者能迅速了解词汇在句子中的语法功能。语法分析则进一步剖析句子的结构, 标注出主谓宾定状补等成分, 将复杂的句子拆解为清晰的语法结构, 为深入的语言研究提供了结构化数据, 有助于研究者精准把握语言的内在规则。

通用单语语料库还具有高度的代表性。它广泛收录多种体裁的文本, 文学作品中的经典小说、优美诗歌、散文, 新闻领域的时事报道、评论分析, 口语范畴的日常对话、演讲发言等均被囊括其中。不同体裁的文本具有各自独特的语言风格与特点, 通过对这些文本的综合分析, 能够真实地反映出语言在实际使用中的多样性与复杂性, 为语言研究与翻译实践提供贴近现实的语言样本。

因此, 通用单语语料库的核心特征可总结如下: 1) 规模宏大 (通常包含数亿词), 力求覆盖特定时期语言使用的广度; 2) 构成平衡, 尽可能按实际比例包含不同体裁 (文学、新闻、学术、口语等)、领域和时代的文本; 3) 标注体系完善, 通常包含词形还原、词性标注、语法标注等; 4) 开放性与代表性, 旨在成为该语言研究的权威资源。通用单语语料库的主要功能在于, 为语言本体研究 (语法、语义、语用)、词典编纂、语言教学等提供基于真实语言使用的权威数据基础。

以俄语国家语料库为例。俄罗斯语料库语言学研究虽起步晚于西方, 但发展迅速、内容广泛。随着俄语国家语料库的建成, 俄语语料库的建设水平已经有了大幅度提高, 开展基于语料库的俄语研究的条件已经基本成熟 (许汉成 2005: 22)。А. Мустайоки (2007: 158) 指出俄语国家语料库的核心特征在于其文本构成的平衡性, 这意味着该语料库尽可能涵盖特定语言中所有类型的书面与口头文本 (包含不同体裁的文学作品、多主题报刊文章、广告、专业文本、日记、书信等), 且所有文本均尽可能按其在该时期语言中的实际占比纳入。

2.2 平行语料库

平行语料库是指包含原文文本及其一种或多种译文, 并在句子或段落层面实现精确对齐的电子文本集合。平行语料库的结构与单语语料库存在本质差异, 其核心在于包含原文及其译成其他语言的一个或多个译本。

双语对齐是平行语料库的显著标志。在平行语料库中, 原文与译文在句子或段落层面实现了精确对应, 犹如精心编排的对照文本。每一个源语言句子或段落都能在目标语言中找到与之对应的翻译内容, 且这种对应关系通过先进的技术手段得以精准标注与呈现。这种双语对齐的特性为译者提供了直观的语言转换参照, 使译者能够清晰地看到源语言与目标语言在表达方式、结构组织上的异同, 为准确翻译提供了重要依据。当译者在翻译过程中遇到某个词汇、短语或句式时, 只需在平行语料库中进行查询, 便能获取大量该语言单位在不同语境下的翻译实例。以俄汉平行语料库为例, 若译者在翻译俄语词汇 *свобода* 时, 通过查询该语料库, 可发现其在中文中有“自由”“解放”等多种译法, 同时还能看到这些译法在不同文本语境中的具体使用情况。

由上, 平行语料库的核心特征可总结为: 1) 双语或多语构成; 2) 严格的翻译对应关系; 3) 精细的对齐结构 (句级/段级)。正如 T. McEnery, A. Hardie 强调的, 平行语料库展示了

“一种语言中表述的思想如何在另一种语言中得以传达”，从而为翻译实践提供特定语境下的等效表达范例，服务于翻译研究、翻译教学、双语词典编纂及机器翻译系统开发（McEnergy, Hardie 2008: 22）。

以俄语国家语料库的子库“俄汉平行语料库”为例。该库创建于2016年，目前收录文本逾千篇，总词量超350万，主体为19至21世纪俄中作家的文学作品，辅以新闻及公文文本。其网站支持俄语、中文及英语界面，具备完善的检索系统。该库作为典型的平行语料库，其核心价值在于为俄汉翻译实践和研究提供了大量精确对齐的等效表达实例。

2.3 可比语料库

可比语料库是指汇集不同语言中相同领域、相同时期、相同体裁的非翻译原生文本的电子文本集合。可比语料库一般收录了相同领域、体裁的原生文本，这些文本可能分别来自不同语言，但围绕着相同的主题展开。这种主题可比性为研究者提供了一个理想的平台，能够在相同主题背景下深入对比不同语言在词汇运用、句式结构、篇章组织等方面的差异与共性。

通过功能对比，可比语料库能够清晰地揭示不同语言对相同概念的表达差异。例如，在对比俄语和英语新闻中“кризис”（危机）一词的搭配模式时，利用可比语料库进行分析可发现俄语中“кризис”更常与“экономический”（经济的）连用，形成“экономический кризис”（经济危机）这一常见搭配，强调经济领域的危机状况。而在英语新闻中，“crisis”（危机）则多与“financial”（金融的）搭配，如“financial crisis”（金融危机），凸显金融层面的危机问题。这种对比分析不仅有助于译者准确把握不同语言在特定领域的表达方式，避免翻译中的“中式英语”或“欧式中文”等不地道表达，还能为语言学习者提供深入了解不同语言文化内涵的途径，促进跨语言交流与理解。

可比语料库的核心特征可总结为：1）双语或多语构成；2）文本间无直接翻译关系，但具有主题、领域、体裁可比性；3）文本为各自语言的原生表达。可比语料库主要是为对比语言学研究提供基础，揭示不同语言对相同概念或主题的原生表达方式差异，探究语言单位在各自语言系统中的功能分布，服务于翻译理论研究、跨语言信息检索及术语提取。与平行语料库提供直接的翻译映射和转换模式不同，可比语料库的核心价值在于：通过同源真实语境中的原生文本（排除平行语料库译文特有的转换干扰），实现语言或语言单位的功能对比，展示不同语言独立发展出的、针对相同主题的原生表达方式。其技术实现形式本质上是分布式单语语料库的引用集合。

以英语维基百科语料库（CEW）为例。英语维基百科语料库的建设工作于2014年正式启动，基于维基百科开源文本集合构建。英语维基百科语料库（CEW）包含440万篇文章中的19亿词（经词形还原处理以支持形态分析）。核心功能模块包括：词语速写（Word Sketch）、同义词库（Thesaurus）、词频列表（Word Lists）、N元语法（n-grams）、索引（Concordance）。该语料库凭借其大规模、多语言、原生文本等特性，是研究不同语言在百科领域原生表达习惯和功能分布的理想可比资源。

3 语料库应用于翻译实践的功能价值

作为翻译工具，单语国家语料库、平行语料库与可比语料库功能互补，共同构成译者的综合资源库。单语国家语料库可提供源语和目标语的深度语言知识，帮助理解原文背景、主题和语言习惯，并为目标语表达提供丰富的原生范例和搭配信息，是理解语言本身的基础；平行语料库提供直接的跨语言映射参考，展示特定词汇、结构或语篇在目标语中的实际等效译法，是解决具体转换问题的直接工具；可比语料库揭示不同语言对相似内容或功能的原生表达习惯差异，帮助理解语言间的系统性区别，避免“翻译腔”，追求更地道的目标语表达，并为翻译理论提供实证基础。

3.1 通用单语语料库：深化语言理解

通用单语语料库是译者不可或缺的基础工具。通用单语语料库可提供权威、真实的语言环境，帮助译者精准把握原文词汇的精确含义、高频搭配、语用色彩（如情感倾向）以及特定表达的文化内涵与背景知识。

以俄汉翻译为例。借助俄语国家语料库，译者能够深入探究俄语词汇的丰富内涵与使用规律。例如，对于俄语词汇“душа”（心灵），通过在俄语国家语料库中进行检索与分析，可发现其常与“русская”（俄罗斯的）连用，形成“русская душа”（俄罗斯的心灵）这一搭配。这一搭配不仅仅是简单的词汇组合，更深刻地体现了俄罗斯文化中对“душа”这一概念的独特认知与情感寄托，反映了俄罗斯民族的精神特质与文化底蕴。俄罗斯学者的实验表明，通用单语语料库（如 BNC）虽无法直接解析文化指涉，但能提供源语使用规范的核心参照（如搭配、频率、语域分布），这是双语词典无法替代的；而领域专用单语语料库（如《时代》杂志语料库）的历时分析，则通过揭示“a new frontier”在 1960 年代的使用峰值，为识别其文化内涵（肯尼迪“新边疆计划”）提供关键语言证据，凸显单语语料库在源语深度理解中的基础性作用（Тарнаева, Осипова 2016: 207）。

在语法层面，通用单语语料库同样为译者提供了有力支持。例如，俄语中存在多种复杂的语法结构，如动词体的运用、复杂句式的构建等。通过分析语料库中的实例，译者能够清晰地看到在不同语境下，何时使用完成体，何时使用未完成体，从而在翻译中准确转换语法结构，使译文符合目标语的语法习惯与表达逻辑。此外，通用单语语料库还可帮助译者了解语言的地域变体、社会变体等。不同地区的俄语可能在词汇、发音、语法等方面存在差异，不同社会阶层、职业群体使用的俄语也会有其独特之处。通过对通用单语语料库中多样化文本的研究，译者能够更好地应对这些语言变体，在翻译中根据原文的背景信息，选择合适的目标语变体进行表达，提高翻译的准确性与适应性。

3.2 平行语料库：提供翻译实例

在翻译实践中，平行语料库的价值突显，其句/段级对齐特性使译者能够直观观察特定语境下词汇、结构、篇章的等效翻译模式，为解决具体转换难题（如习语、文化专有项、复杂句式）提供直接参考和多样化方案。当译者面对俄语习语“вешать лапшу на уши”时，这一习语具有独特的文化内涵，直译为“把面条挂在耳朵上”，显然不符合中文的表达习惯。此时，通过在俄汉平行语料库中进行查询，译者可找到多种对应的中文译法，如“糊弄人”“忽悠”等。这些译法并非简单的字面对应，而是在充分考虑了习语的文化内涵与目标语表达习惯的基础上得出的。平行语料库支持对文本的定量与定性分析，学习者通过查询平行语料库，能通过动态语境下的真实译例（如“上火”在情绪语境中译为“get angry”，在生理语境中译为“get inflamed”），帮助学习者探索跨语言转换规律，解决一词多义等翻译难题（陈昕 2020: 84）。此外，还可以帮助学习者了解核心词汇的含义，实现精准翻译，如确认“社会主义核心价值观”的标准译法“the core socialist values”（肖海燕 2024: 70）。

对于一些复杂的句子结构或长难句的翻译，平行语料库同样能提供宝贵的参考。例如，在翻译包含多个从句的俄语长句时，译者可以在平行语料库中查找类似结构的句子及其对应的中文译文，观察目标语是如何对句子进行拆分、重组，以符合中文的表达习惯。通过学习这些转换范例，译者能够掌握有效的长难句翻译技巧，提高翻译的流畅性与可读性。此外，平行语料库还可用于翻译风格的模仿与学习。不同的译者或翻译作品往往具有独特的翻译风格，通过对比平行语料库中同一原文的不同译文，译者可以分析出不同译者在词汇选择、句式运用、修辞手法等方面的差异，从而学习到多样化的翻译风格，并根据翻译任务的要求，灵活运用这些风格元素，提升自己的翻译水平。

3.3 可比语料库：检验译文可读性

可比语料库在验证译文的自然度与地道性方面具有重要作用。译者可通过其验证译文搭配的地道性，使译文更符合目标语的语言习惯与文化背景。

在词汇选择方面，可比语料库能够帮助译者筛选出最符合目标语语境的词汇。例如，在翻译俄语中的“друг”（朋友）一词时，汉语中有“朋友”“友人”“伙伴”等多个近义词。通过对比可比语料库中不同语境下“друг”的翻译实例以及汉语中这些近义词的使用频率与搭配习惯，译者可以确定在具体的翻译语境中，哪个词汇最为恰当。如果是在描述日常亲密关系的语境中，“朋友”可能是最合适的选择；而在较为正式、书面的语境中，“友人”或许更能体现文本的风格。

在语篇层面，可比语料库有助于译者调整译文的篇章结构与逻辑关系，使其更符合目标语的思维方式。不同语言在篇章组织上存在差异，如英语和汉语在段落划分、主题句的位置、逻辑连接词的使用等方面都有所不同。通过对比可比语料库中相同主题的源语和目标语文本，译者可以学习目标语的篇章构建方式，在翻译过程中对译文的篇章结构进行优化，使译文的逻辑更加清晰、连贯，易于目标语读者理解。例如，基于英汉期刊论文可比语料库（CCEAP）的研究显示：“有趣”这一属性只伴随评价对象，而从不赋予行为动作，“it is interesting to note that”需译为“值得注意的是”（或“有趣的是”）而非直译为“有趣地注意到”，且“it is easy to see that”在汉语期刊中更倾向以“不难发现”和“不难看出”实现功能对等（张乐 2021: 57）。

4 各类型语料库用于翻译实践的组合策略

面对翻译实践中的具体问题，译者需根据问题性质（理解、转换、生成）和所需信息类型（语义内涵、转换模式、表达规范），策略性地选择和组合三类具有互补功能的语料库工具。我们可针对性地运用不同语料库工具来解决各类翻译难题。

4.1 术语翻译

在术语翻译过程中，通用单语语料库、平行语料库和可比语料库相互配合，能够帮助译者准确确定术语的翻译。对于俄语术语“инновация”，首先利用单语语料库，如俄语国家语料库，通过分析该术语在大量俄语文本中的使用情况，确定其精确含义为“创新”。接着，使用平行语料库，查找“инновация”在俄英、俄汉等平行语料库中的已有译法，发现其常译为“innovation”（英语）和“创新”（汉语）。最后，借助可比语料库，验证“innovation”在英语科技领域的使用习惯和搭配模式，确保译法符合目标语的专业表达规范，从而实现术语翻译的准确性和专业性。

4.2 文化专有项翻译

文化专有项的翻译是翻译实践中的难点之一，三类语料库的结合使用为解决这一问题提供了有效途径。以俄语文化专有项“матрёшка”为例，单语语料库可以帮助译者深入分析其文化内涵，通过在俄语国家语料库中检索相关文本，了解到“матрёшка”象征着俄罗斯的传统手工艺，承载着丰富的俄罗斯文化元素。在平行语料库中，译者可以参考“матрёшка”的多种翻译策略，如直译为“matryoshka”或意译为“Russian nesting doll”，并分析不同译法在不同语境下的应用效果。最后，利用可比语料库，寻找英语中类似的文化符号，如“teddy bear”（泰迪熊），通过对比分析，确定最能传达“матрёшка”文化内涵且符合目标语读者理解习惯的翻译方式，实现文化专有项的有效翻译。

4.3 译后校验

译后校验是保证翻译质量的重要环节，单语语料库和可比语料库在这一过程中发挥着关键作用。在检查语法和搭配方面，单语语料库可以提供标准的语言使用范例。例如，对于俄语短语“принять решение”，通过在俄语国家语料库中查询，确定其正确的英语译法应为

“make a decision”，而不是“take a decision”，从而避免语法和搭配错误。在避免“翻译腔”方面，可比语料库具有独特优势。通过对比俄语和目标语（如英语）相同领域文本的语篇结构和表达方式，译者可以发现译文与原生目标语文本在表达习惯上的差异，并对译文进行调整，使译文更加自然流畅，符合目标语的表达习惯，提高译文的质量和可读性。

5 结束语

本文系统探究了通用单语语料库、平行语料库与可比语料库的多维特征及其在翻译实践中的功能价值。研究表明，三类语料库在结构、功能和目标上呈现显著差异：单语国家语料库以其大规模、结构化、深度标注的文本集合，为语言理解和背景知识深化提供权威资源；平行语料库的核心价值在于其句级/段级对齐机制，直观展示跨语言转换模式，提供特定语境下的等效表达范例；可比语料库则通过汇集相同领域、体裁、时期的不同语言原生文本，揭示语言单位的功能分布与原生表达差异。三类语料库并非孤立工具，而是构成功能深度互补、协同支撑翻译全流程的有机工具链：通用单语语料库（源语/目标语）是夯实语言理解与地道生成的基础，平行语料库直接指导跨语言转换操作，可比语料库则揭示原生表达差异、规避“翻译腔”并为宏观风格匹配提供基准。

展望未来，语料库资源的发展与翻译理论及实践的深化将更加紧密交织。其有效利用的关键在于译者对各类语料库特性、互补机制及组合策略的精准把握。随着大数据和人工智能技术的发展，深化三类工具的智能化融合应用（如增强检索、智能推荐、动态组合），持续推动语料库资源的丰富化、开放化（尤其专业领域平行与可比语料库）与专业化建设，将是克服现有障碍、进一步释放语料库作为译者工具潜能的关键方向。语料库语言学方法必将继续深刻塑造翻译研究与实践的格局，助力提升翻译质量、效率与跨文化沟通效能。

参考文献

- [1]McEnery T., Hardie A. Corpus Linguistics: Method, Theory and Practice[M]. Cambridge University Press, 2011.
- [2]Мустайоки А. Роль корпусов в лингвистических исследованиях языков[A]. Национальный корпус русского языка и проблемы гуманитарного образования: материалы международной научной конференции[C]. Москва: ГУ-ВШЭ, 2007(12).
- [3]Тарнаева Л.П., Осипова Е.С. Использование ресурсов корпусной лингвистики при подготовке переводчиков в сфере профессиональной коммуникации[J]. Филологические науки. Вопросы теории и практики, 2016(9).
- [4]陈 昕. 平行语料库在翻译教学中的应用研究[J]. 文化创新比较研究, 2020(13).
- [5]肖海燕. 双语平行思政语料库在翻译教学中的应用探索[J]. 语言与文化研究, 2024(4).
- [6]许汉成. 俄语语料库的新发展[J]. 中国俄语教学, 2005(1).
- [7]张 乐. 基于英汉期刊论文可比语料库的对比短语学研究:背景、理念和方法[J]. 解放军外国语学院学报, 2021(1).

Multidimensional Features of Corpora and Their Functional Value in Translation Practice

Pei Hai-tong

(Heilongjiang University, Harbin 150080, China)

Abstract: In the context of globalization, translation practice plays a vital role in cross-linguistic and cross-cultural communication. By systematically exploring the multidimensional features of standard monolingual corpora, parallel corpora, and comparable corpora, as well as their functional complementarity in collectively forming a comprehensive resource repository for translators, this paper elucidates the functional value of these three types of corpora in translation practice. The research demonstrates that: standard monolingual corpora provide authoritative resources for in-depth understanding of the source and target languages; parallel corpora offer patterns of equivalent expressions in specific contexts; comparable corpora reveal the functional distributions of linguistic units and differences in native expressions. The three types of corpora are functionally complementary: monolingual corpora deepen linguistic understanding, parallel corpora provide translation examples, and comparable corpora test the readability of translations. Translators can strategically select and combine these three types of corpora tools with complementary functions according to their needs to solve various translation problems, achieve effective translation, and improve translation quality. This paper points out that intelligent integration and the opening-up and specialization of resources are key future directions.

Keywords: standard monolingual corpora; parallel corpora; comparable corpora; translation practice; combined strategies; translation quality

基金项目: 本文系黑龙江省哲学社会科学研究规划项目“认知语言学视域下汉俄民族世界图景比较研究”(24YYC006)的阶段性成果之一。

作者简介: 裴海同(1993—), 俄语语言学博士, 黑龙江大学俄语学院讲师, 硕士生导师, 研究方向: 认知语言学、语言文化学。

收稿日期: 2025-01-09

[责任编辑: 李 侠]