

## 第四章

# 机器翻译的文本选择和准备：面向全球读者的译前编辑和写作

Pilar Sánchez-Gijón

巴塞罗那自治大学

Dorothy Kenny

都柏林城市大学

与以往的技术相比，神经网络机器翻译（NMT）产出的译文越来越流畅、错误越来越少。因此对于很多语言来说，神经网络机器翻译正成为提升翻译速度的实用工具。但是，要获得最佳的目的地语机器译文，且确保该译文适于目的地语读者，这不仅取决于机器翻译系统的质量，还取决于源语文本是否适用于机器翻译。本章将介绍译前编辑的概念，即如何对源语文本进行编辑，使其更适于机器翻译和全球的目的语受众。

## 1 引言

简而言之，译前编辑 (*pre-editing*) 即重写部分源语文本，以改善该文本的机器翻译质量。<sup>1</sup> 译前编辑遵循一套正式规则，或称“约限语言” (*controlled language*)，用来规定文本允许使用和不可使用的特定词语或结构（参见 O'Brien 2003）。译前编辑也可能只涉及对原文的“修修补补”，如拼写改错或规范标点符号的使用。根据具体情况，译前编辑可能同时运用上述两种做法。但无论如何，其主要目的都在于提高机器译文质量。若同一源语文本被译成多种目的地语，按理说，译前编辑带来的质量提升能够体现在每一个目的地语译文中。因此，译前编译一贯来都被建议纳入多语翻译工作流程中。

---

<sup>1</sup>如 Rossi & Carré (2025 [本卷]) 中所讨论，质量并非固定不变的概念，而是取决于许多因素，如翻译目的。欲深入了解这一多变的概念，可参阅 Drugan (2013) 和 Castilho et al. (2018)。



确保文本适于机器翻译的另一种方法是，在写作伊始就带着这个目的进行写作。因此，作品会被译成多种语言的作者通常被要求写作时要心怀全球读者。此外，写作者还需遵循“简明易懂”的写作原则，如尽量避免具有文化特异性的元素。这一原则也适用于另一类作品的作家，这些作品的原著读者群并非写作语言的母语者。

基于这些共同目标，译前编辑的原则、约限语言和面向全球读者的写作指南之间存在相同之处也就不足为奇了。我们将在本章中概述其中常见指南的类型，但未必详尽。要提请读者注意的是，这些指南都因语言而异，如关于时态的使用建议只适用于有语法时态的语言，而许多语言都没有时态。另外，指南也可以针对特定的语言对和机器翻译类型或引擎。例如，某个语言结构可能给基于规则的机器翻译 (RBMT) 造成翻译难题，但对神经网络机器翻译 (NMT) 却不成问题。再比如，某个指南是专门针对用法律文本训练的神经网络机器翻译引擎可能产生的错误，那么就不适用于用医学文本训练的引擎。若写作时考虑到机器翻译的问题，哪些建议会有帮助着实要视具体情况而定。

神经网络机器翻译的出现，启发我们重新思考译前编辑和约限写作的相关建议的有用性（参见下文的Marzouk & Hansen-Schirra 2019 和 节 2），但从机器翻译的发展历程来看，译前编译在大多时候都对提升机器翻译质量大有裨益。无论是基于规则的机器翻译，还是统计机器翻译，充分了解机器翻译技术有助于预测源语言或源语文本中有可能导致机器译文出现错误的因素。然而，神经网络机器翻译的特点之一恰恰是不存在系统性的错误，即很难有把握地预测错误类型，因此无法预防特定错误的出现。此外，神经网络机器翻译大大提升了译文的流利度和忠实度，这也可能表明，为提高译文质量所做的译前编辑收效甚微。换句话说，在神经网络机器翻译的背景下，译前编辑似乎略显多余。但是，机器翻译技术的改进并没有消减所有类型的译前编辑的益处。如下文所述，虽然译前编辑的某些传统方法可能不再适用，但还是有些方法必不可少，特别是在缺少译后编辑环节的翻译工作流程中加入译前编辑，或译后编辑的准则只是产出“足够好”的译文（参见O'Brien 2025 [本卷]: §2）。况且，虽然译文质量有所提高必定意味着过去机器翻译常出现的错误已经大幅减少，但正如下文所述，错误并没有完全消除，同时还有新的错误出现，而到目前为止，机器翻译的新错误还未进行评测。这些都与翻译任务的性质（见下文）、源语文本的功能和作者的意图有关，正因如此，译前编辑还会继续在优化使用和提高机器翻译的有效性方面发挥作用。

下文首先讨论在近期和当前的神经网络机器翻译运用中，译前编辑的背景和用途。接着，我们会描述机器翻译的源语文本的选择策略，以及英语作为源语言对机器译文的影响。然后，我们会举例说明如何以全球读者为主进行写作。最后，我们会介绍常用的译前编辑指南，以及相关的资源和工具。

## 2 译前编辑和神经网络机器翻译

过去, 基于规则的机器翻译系统经常在忠实度和流利度这两方面出现明显且系统性的错误 (参见Rossi & Carré 2025 [本卷]), 因此为了最大化发挥机器翻译的价值, 译前编辑往往必不可少。即便发展到统计机器翻译阶段, 研究人员仍认为译前编辑非常有用。例如, Seretan et al. (2014)在研究英法、英德和法英这些语言对时发现, 对于用户产生的技术和健康领域内容, 适当的译前编辑能提高机器翻译的质量。在一项同样针对统计机器翻译技术的相关研究中, Gerlach (2015)发现, 使用经过译前编辑的英语原文提高了法语机器译文的译后编辑效率, 尽管对整个翻译流程的效率影响并不明确。同样, Miyata & Fujita (2017)发现, 使用经过译前编辑的日语原文可以提高英语、汉语和韩语的译文质量。在他们看来, 这一结果证实了译前编辑在统计机器翻译的多语言环境中发挥了重要作用 (同上: 54)。这两位学者随后又研究了译前编辑对两个神经网络机器翻译系统输出的影响, 但这一次却发现, 对于从日语译到英语、汉语和韩语的机器译文来说, 译前编辑与译后编辑的工作量几乎不具有相关性 (Miyata & Fujita 2021)。Miyata 和 Fujita (同上) 还研究了不同类型的译前编辑的影响, 发现机器翻译的传统编辑方法较少运用在神经网络机器翻译的工作流程中:

缩短和简化原文句子的操作 (在神经网络翻译 workflow 中) 并不常见, 这一点与译前编辑惯常做法截然不同。。而更为重要的是使内容、句法关系和词义更加清晰明了, 即便相关操作会让原文句子变得更长 (Miyata & Fujita 2021: 1547)。

还有其他研究也表明, 译前编辑对神经网络机器翻译系统来说根本不是有效策略。例如, Marzouk & Hansen-Schirra (2019)发现, 在德语到英语的技术文本翻译中, 译前编辑提高了基于规则的机器翻译、统计机器翻译以及结合两者的机器翻译系统性能, 但所测试的神经网络机器翻译系统除外。<sup>2</sup> Hiraoka & Yamada (2019)是为数不多热衷于研究神经网络机器翻译的译前编辑的学者。他们对日语 TED 演讲的字幕只采用 3 条译前编辑规则, 即:

- 补充遗漏的标点符号,
- 补充遗漏的语法主语和 (或) 宾语,<sup>3</sup>

---

<sup>2</sup>和其他作者一样, Marzouk & Hansen-Schirra (2019)细心指出, 其研究基于所谓的“黑匣子”系统, 即内部运作无法被观察分析的现成系统。

<sup>3</sup>和西班牙语一样, 日语也是“主语脱落”(pro-drop) 语言 (参见Kenny 2025 [本卷]: §1), 也就是说某些代词可以省略但不妨碍理解。在日语中, 这些代词可以是主语代词, 也可以是宾语代词。

- 用目的语写出专有名词。

Hiraoka 和 Yamada (同上) 认为, 这 3 条译前编辑规则提高了使用现成的神经网络机器翻译系统所产出的英语字幕译文。不过, 在极少数情况下, 这些规则反而会导致机器译文的质量下降。

鉴于缺乏明确科学依据来证实译前编辑为神经网络机器翻译的工作流程提供有力支持, 建议神经网络机器翻译的行业用户在生产环境中推广使用译前编辑前, 先仔细测试译前编辑的效果。如本章引言所述, 用户可能会发现, 因为文本体裁、神经网络机器翻译引擎和所使用的训练数据会有所不同, 某些译前编辑操作仅适用于特定语言对。

### 3 基于体裁和领域的建议

在专业译者(和翻译行业)看来, 翻译项目中的机器翻译使用通常与特定的文本体裁和领域相关联(参见Kenny 2025 [本卷]: §1)。过去, 关注可预测的、重复的或受限的词语和结构的体裁和领域表明, 基于约限语言的译前编辑准则行得通, 也就是说, 在需要成本低且速度快的翻译时(如大众无法获得的内部技术文件), 源语文本中的词语和结构被简化, 以确保机器译文的质量。在数据驱动翻译的时代, 对体裁和领域的关注仍然重要, 因为行业中的统计机器翻译和神经网络机器翻译引擎所使用的训练数据, 也是来自特定的体裁和领域, 抑或是这些引擎至少可以按照特定的体裁和领域进行定制化(参见Ramírez-Sánchez 2025 [本卷])。

因此, 许多语言服务提供商根据自己的机器翻译使用经验, 建议将机器翻译(也可以引申至神经网络机器翻译)的使用限制在以下范围:

某些类型的技术文档: 这类文件通常已经是标准化文本, 其术语使用严格, 文体简洁明了, 语言使用非常接近约限语言的应用规则。此外, 这类技术文档的内在概念框架在源语言和目的语的“区域和语言环境”(locales)中也是一致的。例如, 在爱尔兰和法国销售的个人电脑的技术规格基本相同, 因此将这些规范从英语译到法语时, 会发现存在很多共同点; 无需对译文进行大幅调整以适应法语目的语读者或新的概念框架。若使用神经网络机器翻译系统来翻译这类文本, 使用约限语言的规则来处理选词或代词的使用(如 *it*)或许尚有潜在价值, 但句法规则的用处可能不大。

技术规范简单易懂, 与同一产品相关的营销文本和法律材料截然不同。后者可能需要进行调整, 以使其更容易被潜在消费者接受, 或符合目的语的法律框架。其实, 有些领域的文本在翻译时需要完全“重新思考”, 以便符合新的概念系统, 而法律翻译就是最好的例子。

欲进一步了解特定领域和体裁的翻译, 请参见Olohan (2015)和Šarcevic (1997)。

低风险的内部文件: 这类文本的传阅率低, 即便翻译质量欠佳也不会造成严重后果(欲进一步了解机器翻译使用的风险, 请参阅Canfora & Ottmann 2020 和 Moorkens (2025 [本卷])), 甚至仅在用户或客户的公司内部使用。因此, 与其他应用场景相比, 目的语文本是否通顺流畅没那么重要(尽管神经网络机器翻译的输出通常十分流畅), 但很多公司仍希望能约限和降低词汇丰富度。

低风险的外部文件: 这类文本仅供用户偶尔参阅或用作帮助数据库或类似功能的文本, 通常由服务或产品的用户群体, 而非客户生成。在许多类似的情况下, 机器翻译供应商往往会明确拒绝对翻译错误造成的任何损失承担责任。

有些文本的功能不仅限于提供信息或指示说明, 而具有“感召功能”, 对此通常不推荐使用机器翻译。例如, 唤起读者对某个品牌的兴趣, 或引导做出某种行为。换句话说, 文本传递信息的功能越突出, 就越局限于其字面意义, 其隐含信息就越少, 对读者的文化或社会现实的相关参照就越少, 那么机器翻译的质量就会越高。

## 4 英语在控制领域的影响

对于非比喻性或创造性, 而是贴近字面意思且体裁规则明晰的文本, 基于规则的机器翻译和后来出现的统计机器翻译的表现都尤其突出。在许多情况下, 源语言无疑都是英语。其他主要语言成为目的语, 这意味着在一些场景下, 翻译受到英语的高度制约, 由此产生了语言同质化的文本, 以求简化译文, 便于终端读者理解。正如产品用户手册等体裁, 根据我们以及Navarro (2008) 和 Aixelá (2011)等人的研究, 无论是从宏观篇章角度(如文本结构和文本论证的推进), 还是从微观篇章角度(如词汇、形态和句法借用), 西班牙语等语言都深受英语原文的影响。

约限领域的交际目的是基于无歧义的精准用词促进对文本的阅读和理解, 从而原文不仅简单易读, 且便于翻译。有些行业采取更进一步的措施, 使用约限语言来确保文本没有歧义。在这些情况下, 英语作为源语言对其他语言的影响则更加明显。<sup>4</sup> 如航空工业先使用英语建立约限语言规则, 再应用于目的语 (Ghiara 2018)。

这些例子表明, 某些领域需要约限源语言, 以确保获得快速准确的翻译。在这些情况下, 译文的准确度(尤其是使用机器翻译时)优先于目标文本的

---

<sup>4</sup>Seoane Vicente (2015) 详述了英语作为约限语言在不同领域的使用。

任何其他交际层面。然而，神经网络机器翻译的到来意味着，机器翻译的使用如今已经超出了仅限于特定受众的领域，我们会在下一节对此进行详述。

## 5 为全球读者写作

有时，译前编辑的目的不在于减少错误，而是确保译文不仅能传达原文意思，还尽可能地在目的语读者中重现原文在源语言读者中引起的效果。问题在于，要如何提供面向全球读者的文本，并尝试在每种目的语的读者中产生同样的效果。

从交际的角度看来，无论该文本是否会使用机器翻译，作者在写作过程中要具有翻译思维一直是值得提倡的做法。其实，为将来的翻译进行文档准备也是技术写作人员的培训内容之一 (Maylath 1997)。

在过去的 50 年里，翻译行业以及从译者到主要技术开发人员和分销商的所有相关利益方都认识到，最佳的翻译策略需要适当的产品国际化 (Fry 2003: 14)。而要让在某个地区设计和开发的产品为其他地区用户所接受，最佳做法就是剔除产品中所包含的原产地独有的元素。这样一来，任何数字产品都可以进行本地化，并在目标语言环境中使用，同时适用于任何设备或平台，而无需调整原始设计。计划以不同语言发布的文本也已经采取了类似的写作策略。

语言服务提供商和本地化数字产品开发商都发现，他们全球传播策略的关键在于源语文本的译前编辑。许多语言服务企业都在其官网上宣称，出色的多语交际策略始于创作恰当的源语文本。此外，数字产品开发人员也发现了，与其用户和潜在客户沟通的最佳策略是心怀全球用户。在创作任何文本内容时，都应该考虑制定一套体现这种策略的指南。例如，谷歌的文档风格指南特别突出“为全球受众写作”为基本原则，制定了一系列英文写作指南，这便于将文档翻译成任何目的语。其中包括一般规则和禁忌，如使用现在时态、提供语境、尽量不使用否定结构、使用短句、使用清晰准确且无歧义的语言，以及确保写作的一致性和包容性 (Google 2020)。

如今，随着翻译技术的发展，我们能够结合使用计算机辅助翻译工具（如翻译记忆库工具，参见Kenny 2025 [本卷]: §4）和机器翻译系统。就此看来，机器翻译的局限性不在于技术层面，而决定于机器翻译原文的质量（原文是否出错）和译文是否符合目标交际语境（语域、语调、体裁惯例，以及与译文交际功能相关的其他问题）。

若文本属于严格遵守形式惯例且主要具备信息性或指示性交流功能（如技术文档等）的体裁，高性能机翻引擎可以产出良好或优质的译文，这要取决于语言对等因素。在此情况下，“译前编辑”仅限于对源语文本进行拼写检查，因为这类体裁的文本通常不涉及使其超出标准和原文非复杂使用范围的文体或指代特征（见下文）。

然而,使用机器翻译来处理具备多种交际功能的体裁并非易事,例如最近流行的技术小工具“拆箱”视频,通常既有指导性(信息性)又有娱乐性(呼吁性和表达性)。

而其他体裁的文本可能包含对源语群体的社会、经济或文化生活的参考内容,这些参照能唤起源语读者的认同感,但或许无法对目的语读者产生同样影响(参照元素可见节 6.4)。对有些机器翻译引擎来说,其他翻译障碍可能还包括源语文本的修辞和文体手法(如缩合、缩写、新词、不完整的句子等),不过源语读者能够识别这些信息。

神经网络机器翻译的译文不仅越来越通顺准确,且相对人工翻译而言,翻译速度也有所提高。此外,很多不同体裁的神经网络机器翻译译文质量似乎都很高。但是,语法正确和翻译无误的目的语文本仍不能称作“恰如其分的”翻译。而译前编辑可以在确保翻译适当性的同时,考虑到全球读者。可目前,翻译行业很少采用这一步骤。过去,一些使用统计机器翻译或基于规则的机器翻译的国际企业会对原文进行译前编辑,以免在使用其机翻系统时出现重复错误。有了神经网络机器翻译,译前编辑或许会作为翻译策略的一部分普遍应用于翻译行业,这不仅可以减少翻译错误,还可以使机器译文更加符合目的语译文的使用语境。

## 6 译前编辑指南

### 6.1 写在前面

在处理面向全球读者或约限领域的文本时,译前编辑主要通过应用一系列特定策略来提高机器翻译的质量。译前编辑有助于确保针对全球读者的约限领域的交流清晰。这种语境主要涉及信息性文本类型,即语言使用不具有创造性或审美意义,而是明确的字面意思,旨在为读者提供信息或作出指引。以下是面向全球读者的最常用的交流指南,也是译前编辑策略的基础。指南的主要目的不仅在于通过重建传达原文“信息”且语法正确的译文来提高机器翻译的有效性,还在于根据文本功能和语境来获得适合读者交际情景的译文。指南具体分为三类:

1. 选词
2. 结构与风格
3. 参照元素 (Referential elements)

无论如何,译前编辑的成功取决于两个因素。第一是(源语和目的语)文本的功能,也就是说,文本的信息或指引功能大于交际或美学功能,则对原文进行译前编辑就越有意义。第二,所选机器翻译系统会出现的翻译错误类型,以及依靠对原文进行译前编辑可避免或减少的机器译文错误类型。

对原文进行译前编辑有两个目标，第一是获得尽可能无误的机器译文，第二是获得适于全球读者的机器译文。本节介绍的译前编辑指南正是对这两个目标的回应。

6.2 词汇指南

正如在Pérez-Ortiz et al. (2025 [本卷])中所看到的，神经网络机器翻译所处理的每个单词或意义单位都取决于其语境，反之亦然。词汇选择与其所在文本和语境的范围相关联。我们来看一个例子，假设要使用机器翻译在最短时间内将文本翻译成几种语言并发布。原文选词恰当不仅可以避免翻译错误，还可以根据文本的功能和发布缘由更有效地遵循语言用法。如表 1所示为词汇相关的常见准则。

表 1: 译前编辑的常见准则（词汇）

指南	解释
避免词汇的语域转换	避免使用会改变文本风格或表达方式的词语。这有助于文本理解，并规范化面向读者的表达方式。
避免不常见的缩略形式	只使用常见的缩略形式。避免使用那些不易于根据当下语境翻译的缩略形式。
避免冗余的词语	避免使用冗余的词语来传达所需信息。冗余的词语意味着神经网络机器翻译系统需要处理更多的词语组合，导致出错概率提高。
保持一致	术语使用要保持一致。避免不必要的词汇变化形式（即避免同义词）。

6.3 结构与风格

就便于理解而言，文本的总体结构和造句与用词同样重要。一个文本中的观点在句子层面、文本层面，甚至文本间层面互相关联，呈现出特定的顺序，有助于读者的理解和阐释文本。对于神经网络机器翻译来说，原文的遣词造句会激活或阻止译文的遣词造句。客观来说，过于复杂且含糊不清的文本结构容易产生不同解释，这样的结构会增加神经网络机器翻译系统生成微观结构元素（术语、短语或句法单位）的正确翻译的可能性。但是，这些元素在同一文本中连接起来，就会产生缺乏内部连贯性的文本，表达的意思或与原文有别，或根本难以理解。



如表 2 所示为文本风格和结构的译前编辑准则。其主要目的不仅是为了优化神经网络机器翻译系统的使用，也是为了使译文更加便于理解和准确。

表 2: 译前编辑的常见准则 (结构与风格)

准则	解释
保持句子简短	避免会造成歧义的冗余复杂句，使得原文和译文更加便于理解。例如，神经网络机器翻译系统可能无法正确处理回指和下指的句法结构，导致漏译或误译。避免使用容易出现歧义的句子结构。
保持句子完整	避免省略或切分信息。源语中不言而喻的信息的补偿机制不一定适用于目的语。例如，句中的动词被动形式未指明施动者可能会造成误解。同样地，若句子补语以选项列表的形式出现（如项目符号列表），也可能造成误解。这时，补语被分割成孤立的短语，导致机器翻译无法正确处理。要注意，机器翻译系统通常以句子作为翻译单位（参见Kenny 2025 [本卷]: §7），即标点符号之间的文本，如句号或段落分隔符。
关联的句子使用平行结构	列表中或语境相同的句子使用同样的句法结构（如章节标题和直接指示）。这种“标志性链接”（参见Byrne 2006）通常更便于原文和译文的理解。此外，这还有助于发布后阶段的系统性错误识别。
主动语态	酌情优先使用主动语态或其他明确显示动作“参与者”的结构（需考虑文本体裁和所涉及语言的惯例）。
统一风格	保持风格统一。这有助于对原文和译文的理解，尤其当文本面向全球读者时。

表 2 中的大多数准则旨在提供便于原文读者理解的文本。若神经网络机器翻译引擎的训练使用以这些标准来翻译的数据集，那么对原文的译前编辑有助于获得质量最佳的机器译文。但要注意，若引擎的训练使用“域内”数据，即使用专门领域的同质数据集，且属于特定体裁并于特定专业领域相关（参见Ramírez-Sánchez 2025 [本卷]: §2.1），那么最好的译前编辑，如有需要，也应符合该体裁或领域的特点。除了以上的一般性建议外，大多数情况下还必须考虑针对源语或目的语的特定准则。比如避免使用非常含糊不清的表述，不仅对于机器翻译系统，对于读者也是如此。

以英语为例，避免歧义表达也就是避免隐含的复数。如名词短语 *the file structure*（文件结构）既可以指“多个文件的结构”，也可以指“某个文件的结构”。虽然读者通读文本后可以消除这个歧义，但这个名词短语的词义不清，无法避免译文出现歧义。另一个常见的歧义结构为否定式，除了英语，许多其他语言也有同样情况。如常见的歧义句子 “No smoking seats are available”，有不同的理解方式，导致产生翻译错误。

便于读者理解和避免出现误译的第三个方法是简化动词时态。尽管不同动词时态和语态的翻译不一定会给机器翻译造成障碍，但目的语的动词时态使用不当（尽管句子结构无误）会导致译文的意思有误。表 3所示为常见的动词形式相关准则。

表 3: 译前编辑的常见准则（动词形式）

准则	解释
使用主动语态	酌情尽量使用主动语态。
使用动词的一般时态；最好使用一般现在时或一般过去时。	根据语言对和机器翻译引擎，尽量避免使用动词的复合时态。虽然同一复合时态可能存在于源语和目的语，但其用法可能有所差别，从而产生不同的解释。
避免使用串联动词 (concatenated verbs)。	不必要的动词串联会增加文本的理解和翻译难度，应尽量避免使用。

6.4 指称元素

指称元素用于替代或参照其他元素，无论是在同一文本中（即文内指称），还是在文本外（即文外指称）。最直观的例子是代词，如 I、he、she、him、her 等，以及相关范畴的物主限定词，如 my、his、her 等。

若代词与其所指在同一个句子里，神经网络机器翻译通常不会犯错，因为同一个句子的所有格限定词和名词之间，一般不会出现性数配合不一致的错误（Bentivogli et al. (2016) 早期曾讨论过，随着神经网络机器翻译的出现，机器译文中的性数一致现象有所改善）。同样地，若文本中连续出现的代词均有同一所指，一般也不会有问题（如相继出现的句子有同一主语）。不过在其他情况下，代词的译法可能会受到训练语料库中的常见译法影响。当文本有不同的所指交替出现，代词的翻译就很容易出错。在此情况下，即便读者很清楚每个代词的所指为何人何物，机器翻译系统通常却无法做出同样的判断，更倾向于使用最常见的代词形式。

因此，对于通过性数一致等语法规则来反映所指和代词之间关系的语言，机器翻译可能无法实现代词和所指的一致关系。不过，我们可以通过使用简单句来尽量避免这类问题的出现。

如下文示例，例 (1) 中的所有格限定词“su”，若不考虑语境，可以表示“他的”或“她的”。但在例 (1) 中，如给出的注脚式译文 (gloss translation) 所示，

只能理解为“她的”。然而，DeepL<sup>5</sup>将单词“su”译为“his”，因为它没有弄清“su”和“María”之间的指称联系。

- (1) María llamó, pero Pepe no llamó. El sonido de *su* llamada me despertó.  
'Maria called, but Pepe didn't call. The sound of *her* call woke me up.'  
DeepL: 'Maria called, but Pepe didn't call. The sound of *his* call woke me up.'

关于文外指称，除了原文固有的所有指称外（如法律文件中的具体立法），译前编辑需要考虑两类指称：1) 针对读者的指称，2) 读者认同的任何文化指称。

如前所述，从文体的角度来看，文本最好由简短句子组成，并使用直截了当的风格，这包括主动语态或包含施动者和受动者的被动语态。这种风格尤其适合指示性文本。对于这类文本，直截了当的风格和主动语态的使用能直接面向读者。在处理这类句子时，机器翻译往往体现出语料库中最常见的用法，因此，如果目的语存在多种面向读者的表达方式，则可能出现不同的译法（大体为意思明确的正式风格），导致文本前后不一致，不过这类错误可以通过译前编辑来避免。

与文化因素有关的文外指称很难笼统地处理，尤其是源语特有的文化指称。在许多情况下，文本的译前编辑就是要时刻记着面向全球读者，尽量用明晰的方式表述这些文化指称中的所有隐含信息。

对于这两类指称（针对读者的指称和文化指称），机器翻译的译前编辑都应考虑目标语读者的语言和背景。

## 7 译前编辑的工具和资源

如上节所述，译前编辑必须在翻译（或多语出版）项目的框架内进行，因此同样需要考虑翻译项目的制约因素。风格指南用于详细说明如何正确对原文进行译前编辑，不仅用结构化的方式列出每个层面的编辑方法，还给出译前编辑前后的例子对比。风格指南与译后编辑类似，不同之处在于示例通常仅来自源语文本。

译前编辑指南的目的是为文本内容准备提供语言使用的相关指导。这些写作指南旨在避免多种语言的机器译文出错，并确保为全球读者提供尽可能高质量的文本。因此，这些指南涉及写作的微观结构（推荐或要求使用的词汇或句法结构）和宏观结构两方面。后者旨在为原文和译文提供必要机制，以确实文本内外的一致性。当文本嵌入数字产品或成为消费品相关文档的

---

<sup>5</sup><https://www.deepl.com/en/translator> 访问于 2022 年 1 月

表 4: 译前编辑的质量保证

译前编辑的质量保证	解释
校对拼写和语法	确保原文没有会对读者造成理解困难的拼写错误或机器翻译错误。
使用既定词汇	检查词汇表的应用是否得当, 避免使用其他同义词, 导致不必要的用词变化。 这么做不仅为了保证系统地使用词汇表中的术语 (包括商品名称或任何专有名词), 还为了尽可能避免专业和非专业词汇的使用出现歧义。
针对读者的指称	若文本有明确的读者群, 则应检查指称风格是否贯彻一致。
风格	检查文本整体的语言风格是否一致, 避免文中出现风格转变。

一部分, 则在该消费品的所有相关文本中, 语言和指称元素 (如术语) 的使用一致有助于保持文外指称一致。

用于发布和翻译的文本内容准备指南还要说明具体做法和使用工具。与翻译中的质量保证 (QA) 环节相比, 文本内容的准备也必须遵循质量标准, 从而保证译文质量。这样一来, 不仅要求译文语法正确, 还要遵循客户的语言标准。

把控源语文本内容质量的常见具体做法如 表 4所示。

很重要的一点是, 要留意文本中旨在获得特定读者认可的方面。例如, 包容性语言有助于避免读者对原文和译文产生排斥感。与译后编辑一样 (参见O'Brien 2025 [本卷]: §2), 译前编辑应当避免文本出现有冒犯之意的粗言秽语。文本内容指南由作者或出版方制定, 因此某家公司的译前编辑指南可能会对语言使用的性别、种族、文化等包容性特点有所涉及和要求。这一点在有性别变化的语言中尤其重要。

以全球读者为目标的源语文本的准备或译前编辑可以通过写作辅助工具来完成。大多数文本编辑软件都包括执行译前编辑和质量保证所需的基本功能。其他功能只能通过专门的创作工具实现。表 5总结了辅助源语文本译前编辑的约限语言检查器的主要功能。

大多数编辑软件都会在不同程度上, 提供满足这些需求的功能。然而, 当译前编辑成为多语内容发布策略的一部分, 这些程序往往无法满足需求。基

表 5: Functions of controlled language checkers

计算机辅助译前编辑	解释
校对拼写和语法	使用语法和拼写检查器。
使用既定词汇	使用包含推荐或禁用词条的字典和词汇表。
针对读者的指称	根据译前编辑指南, 使用适于恰当用语的语法检查器。
风格/语域	通过文本的修改建议, 使用适于文本正式程度的语法检查器。

于译前编辑和使用特定语言的多语内容发布策略, 意味着使公司或机构的所有交流都同质化, 如发布于互联网、社交网站、常见问题模块等的内容。在这些情况下, 需要使用具有译前编辑功能的约限语言检查器, 这些功能可以整合入内容发布流程, 甚至翻译工作中。<sup>6</sup> 这类编辑工具通常不仅限于检查拼写和用词; 其实, 在某些情况下, 它们甚至会随时根据文本的正式程度提供修改建议。不过, 这些工具大多时候仅作为其附加菜单嵌入常用于制作内容的程序中, 如互联网内容管理器、电子邮件或社交网络管理等。这样, 内容创作者和译前编辑者都可以在创建和发布每条内容的流程中直接使用这些工具, 而不需要求助于外部工具。

## 8 什么时候, 该由谁来进行译前编辑?

在大型企业里, 需要使用机器翻译的文本通常要进行译前编辑, 这是技术写作的一部分。译前编辑需要熟练掌握源语言, 但对目的语的了解并不重要。

如今, 神经网络机器翻译的使用方便且质量大有改善, 使得很多主流语言的译者都考虑将机器翻译纳入翻译工作流程。“具备机器翻译素养”的译者 (Bowker & Ciro 2019) 可以根据自身经验, 或甚至通过尝试使用来确定机器翻译是否有所帮助。神经网络机器翻译的常用评估方法是分析机器译文。译文若只需进行有限的译后编辑, 则可被认为适合使用神经网络机器翻译。然而, 极少有人思考原文的用词。考虑到其功能和传阅度, 有些文本应该不难使用神经网络机器翻译来完成。但如果文本过于复杂、不连贯, 或不符合既定的编辑标准, 使用神经网络机器翻译只会使上述问题变得更为突出。在这

<sup>6</sup>可供使用的约限语言检查器和写作辅助工具有很多。商用工具包括 acrolinx(<https://www.acrolinx.com/>) 和 ProWritingAid(<https://prowritingaid.com/>)。

种情况下,译前编辑就大有可为,可以使得原文的内容连贯、风格明确和修辞恰当,以尊重作者意图和文本功能,同时保证其便于理解和翻译。尽管译前编辑并非翻译行业的常用做法,但内容要以多语发布的需求,可能会使这一环节在不久的将来变得更加普遍。

若客户从未为制作文本内容制定语言策略,则译者可以凭借自己所掌握的技能胜任母语和第一外语的文本译前编辑工作。译者熟悉其工作语言对在语法和用词方面的细微差异,这使他们能做出必要的调整,提供读者可以理解的源语言文本,从而减少神经网络机器翻译译文的错误。此外,他们基于对语言对的社会和文化知识的了解,能够判断哪些指称元素在源语和目的语文本中都能够被读者理解,也知道该如何使这个元素更加明晰易懂。译者的双语技能以及对两种文化和社会的了解,意味着他们是准备用于发布的双语或多语文本的最佳人选。他们的主要工作语言仍会是其母语,但在这种情况下,他们产出的不是最终文本,而是目标语读者能够理解的机器译文。因此,译后编辑尽管无论何时都不能省略,但可以减少到最低程度。

作为翻译服务的一部分,译前编辑是有意义的,无论内容是为了生成译文而写,还是也要以源语言发布。在这两种情况下,译前编辑都可以保证原文的质量,并优化神经网络机器翻译的使用。

## 9 结束语

作为翻译项目中的一种可用资源,机器翻译旨在提高翻译效率,从而缩短产出高质量翻译的所需时间。从这个意义上说,译前编辑是在设法优化原文,尽量避免译文出错(当机器翻译用于同化时),并减少达到预期翻译质量所需的编辑工作(当机器译文要进行译后编辑或供人工翻译使用时)。

在翻译信息性或指示性文本时,如果神经网络机器翻译几乎没有出现流利度或忠实度的错误,那么机器翻译面临的挑战就不仅限于这些文本类型了。然而,翻译具有不同交际功能的文本,如游戏或呼吁功能更强的文本,翻译的目标就不仅仅是规避错误。机器翻译需要产出符合原文意图的译文,使目标语读者能产生像源语读者一样的认同感。在这种情况下,译前编辑的附加价值就在于,有助于准备适合以多种语言发布的文本。

作为一种策略,译前编辑可以在外语学习中发挥一定作用。但它的主要应用环境还是多语内容发布。虽然译前编辑最初是技术文档等翻译工作流程的一部分,但神经网络机器翻译的广泛应用可能会使译前编辑应用于更复杂的文本,甚至使译者最终将其技能应用于原文,而不是像几个世纪以来的翻译发展历程那样,主要专注于目标语文本。

## References

- Aixelá, Franco. 2011. An overview of interference in scientific and technical translation. *The Journal of Specialised Translation* 11. 75–88. [https://www.jostrans.org/issue11/art\\_axela.pdf](https://www.jostrans.org/issue11/art_axela.pdf).
- Bentivogli, Luisa, Arianna Bisazza, Mauro Cettolo & Marcello Federico. 2016. Neural versus Phrase-Based Machine Translation quality: A case study. In *EMNLP 2016*. arXiv:1608.04631v1.
- Bowker, Lynne & Jairo Buitrago Ciro. 2019. *Machine translation and global research*. Bingley: Emerald Publishing.
- Byrne, Jody. 2006. *Technical translation. Usability strategies for translating technical documentation*. Dordrecht: Springer.
- Canfora, Carmen & Angelika Ottmann. 2020. Risks in neural machine translation. *Translation Spaces* 9(1). 58–77.
- Castilho, Sheila, Stephen Doherty, Federico Gaspari & Joss Moorkens. 2018. Approaches to human and machine translation quality assessment. In Federico Gaspari Joss Moorkens Sheila Castilho & Stephen Doherty (eds.), *Translation quality assessment: From principles to practice*, 9–38. Cham: Springer.
- Drugan, Joanna. 2013. *Quality in professional translation: Assessment and improvement*. London: Bloomsbury.
- Fry, Deborah. 2003. *The localization industry primer*. 2nd edition. Updated by Arle Lommel. Féchy: LISA. <https://www.immagic.com/eLibrary/ARCHIVES/GENERAL/LISA/L030625P.pdf>.
- Gerlach, Johanna. 2015. *Improving statistical machine translation of informal language. A rule-based pre-editing approach for french forums*. Doctoral thesis. University of Geneva. <https://archive-ouverte.unige.ch/unige:73226>.
- Ghiara, Silvia. 2018. *El lenguaje controlado. La eficacia y el ahorro de las palabras sencillas*. <https://qabiria.com/es/recursos/blog/lenguaje-controlado>.
- Google. 2020. *Writing for a global audience. Google developer documentation style guide*. <https://developers.google.com/style/translation>.
- Hiraoka, Yusuke & Masaru Yamada. 2019. Pre-editing plus neural machine translation for subtitling: Effective pre-editing rules for subtitling of TED talks. In *Proceedings of machine translation summit XVII: Translator, project and user tracks*, 64–72. Dublin: European Association for Machine Translation. <https://aclanthology.org/W19-6710>.
- Kenny, Dorothy. 2025. 人工翻译和机器翻译. In Dorothy Kenny (ed.), 机器翻译知识普及: 为人工智能时代的用户赋能, 19–38. Berlin: Language Science Press. DOI: 10.5281/zenodo.14922287.

- Marzouk, Shaimaa & Silvia Hansen-Schirra. 2019. Evaluation of the impact of controlled language on neural machine translation compared to other MT architectures. *Machine Translation* 33. 179–203. DOI: 10.1007/s10590-019-09233-w.
- Maylath, Bruce. 1997. Writing globally: Teaching the technical writing student to prepare documents for translation. *Journal of Business and Technical Communication* 11(3). 339–352.
- Miyata, Rei & Atsushi Fujita. 2017. Dissecting human pre-editing toward better use of off-the-shelf machine translation systems. In *Proceedings of the 20th annual conference of the european association for machine translation (EAMT)*, 54–59. [https://ufal.mff.cuni.cz/eamt2017/user-project-product-papers/papers/user/EAMT2017\\_paper\\_42.pdf](https://ufal.mff.cuni.cz/eamt2017/user-project-product-papers/papers/user/EAMT2017_paper_42.pdf).
- Miyata, Rei & Atsushi Fujita. 2021. Understanding pre-editing for black-box neural machine translation. In *Proceedings of the 16th conference of the european chapter of the association for computational linguistics*, 1539–1550. <https://aclanthology.org/2021.eacl-main.132.pdf>.
- Moorkens, Joss. 2025. 伦理道德与机器翻译. In Dorothy Kenny (ed.), 机器翻译知识普及: 为人工智能时代的用户赋能, 95–110. Berlin: Language Science Press. DOI: 10.5281/zenodo.14922295.
- Navarro, Fernando A. 2008. La anglización del español: Mucho más allá de bypass, piercing, test, airbag, container y spa. In Luis González & Pollux Hernández (eds.), *Traducción: Contacto y contagio. Actas del III congreso internacional « el español, lengua de traducción »*. 12-14 July 2006, 213–132. Puebla: ESLEtRA. [https://cvc.cervantes.es/lengua/esletra/pdf/03/017\\_navarro.pdf](https://cvc.cervantes.es/lengua/esletra/pdf/03/017_navarro.pdf).
- O'Brien, Sharon. 2003. Controlling controlled English. An analysis of several controlled language rule sets. In *Controlled language translation*. Dublin City University. 15-17 May 2003. EAMT/CLAW. <https://aclanthology.org/2003.eamt-1.12.pdf>.
- O'Brien, Sharon. 2025. 如何处理机器翻译的错误: 译后编辑. In Dorothy Kenny (ed.), 机器翻译知识普及: 为人工智能时代的用户赋能, 83–94. Berlin: Language Science Press. DOI: 10.5281/zenodo.14922293.
- Olohan, Maeve. 2015. *Scientific and technical translation*. London: Routledge.
- Pérez-Ortiz, Juan Antonio, Mikel L. Forcada & Felipe Sánchez-Martínez. 2025. 神经机器翻译的工作原理. In Dorothy Kenny (ed.), 机器翻译知识普及: 为人工智能时代的用户赋能, 111–128. Berlin: Language Science Press. DOI: 10.5281/zenodo.14922297.
- Ramírez-Sánchez, Gema. 2025. 定制化机器翻译. In Dorothy Kenny (ed.), 机器翻译知识普及: 为人工智能时代的用户赋能, 129–146. Berlin: Language Science Press. DOI: 10.5281/zenodo.14922299.



- Rossi, Caroline & Alice Carré. 2025. 如何选择合适的神经机器翻译解决方案: 机器翻译质量评测. In Dorothy Kenny (ed.), 机器翻译知识普及: 为人工智能时代的用户赋能, 39–63. Berlin: Language Science Press. DOI: 10.5281/zenodo.14922289.
- Šarcevic, Susan. 1997. *New approach to legal translation*. The Hague: Kluwer Law International.
- Seoane Vicente, Ángel Luis. 2015. *Lenguaje controlado aplicado a la traducción automática de prospectos farmacéuticos*. handle.net/10045/53587. Doctoral Thesis. URI: <http://hdl.handle.net/10045/53587>.
- Seretan, Violeta, Pierrette Bouillon & Johanna Gerlach. 2014. A large-scale evaluation of pre-editing strategies for improving user-generated content translation. In *Proceedings of the 9th edition of the language resources and evaluation conference (LREC)*, 1793–1799. [http://www.lrec-conf.org/proceedings/lrec2014/pdf/676\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2014/pdf/676_Paper.pdf).

