

工业品外观设计国际注册海牙体系法律发展工作组

第十一届会议

2022年12月12日至14日，日内瓦

翻译技术可用性和有用性研究

国际局编拟的文件

一、背景

1. 在2021年举行的第十届会议上，工业品外观设计国际注册海牙体系法律发展工作组（下称工作组）审议了文件H/LD/WG/10/4和H/LD/WG/10/5，分别为“关于海牙体系引入中文和俄文所涉费用问题和技术可行性的进一步研究报告更新”和“选择其他语言引入海牙体系的标准更新”。工作组请秘书处考虑技术解决方案，特别是机器翻译，提供关于海牙体系引入新语言所涉费用问题和技术可行性的进一步研究报告的修订版，供工作组在下届会议上审议（见文件H/LD/WG/10/6第18(ii)段）。

2. 本文件详细介绍了产权组织不同翻译部门使用的翻译技术和做法。所提供的信息旨在为工作组提供便利，以审议翻译技术对海牙体系中引入其他语言的相关费用可能产生的缓解作用。

二、产权组织使用的翻译技术

3. 在产权组织，翻译工作由保护知识产权的各全球服务部门和语言司进行。翻译过程一般包括两个步骤：首先，将需要翻译的文本与适用的翻译记忆库中的条目进行匹配；其次，如果在该领域和语言组合中有可用的模型，则由笔译员人工要么对神经机器翻译（NMT）模型（WIPO Translate）生成的译文建议进行译后编辑，要么将剩余文本译成目标语言。最后，一定比例的译文由内部笔译员审校进行质量控制。

WorldServer

4. WorldServer 是一个翻译管理系统，包括一个计算机辅助翻译工具，它通过让翻译任务实现自动化来管理翻译工作流程。它首先从翻译记忆库中检索到的目标语言语段填充译文，然后用神经机器翻译建议填充剩余语段。马德里注册部的翻译和术语科以及 PCT 翻译司都使用 WorldServer。¹

翻译记忆库

5. 每种翻译实践的支柱都是翻译记忆库，它是特定领域（海牙、马德里、PCT 或语言司）先前翻译文本的储存库。翻译记忆库中的所有条目都经过人工验证以确保质量。每个部门和每种语言组合²都有独立的翻译记忆库。各个翻译记忆库中包含的条目数量根据先前翻译的文本数量而不同。例如，2016 年以来，用英文提交的海牙申请所占比例超过 86%，法文约 12%，西班牙文约 1%。³因此，英文为源语言的海牙翻译用翻译记忆库比法文为源语言的翻译记忆库的条目多得多，而西班牙文为源语言的翻译记忆库几乎没有条目。翻译记忆库被整合到翻译管理系统（WorldServer）中。

WIPO Translate

6. WIPO Translate 是一种基于人工智能的机器翻译工具，由产权组织内部开发，针对特定主题进行培训，采用了最新的神经机器翻译技术。神经机器翻译是一种基于神经网络模型的技术，它从大量先前翻译的文本⁴中“学习”。WIPO Translate 最初用 10 种 PCT 公布语言⁵的专利文献⁶进行训练，并侧重于这些文件。WIPO Translate 已在 Patentscope 上向用户公开提供，提供从这些语言到英文和从英文到这些语言的译文建议。

7. 已经为产权组织各翻译部门中的每一个开发了 WIPO Translate 的定制模型。⁷但是，为了产生高质量的译文，WIPO Translate 必须用相关领域的大量文本对特定语言组合进行训练。⁸

8. 此外，神经机器翻译模型为语言学上距离较远的语言组合生成的译文建议，通常比语言学上较近的语言组合的质量要低。因此，对于遥远的语言组合，拥有足够的训练数据以确保足够的翻译质量更为重要。⁹

翻译、译后编辑和质量控制

9. 上面介绍的工具有不能实现翻译过程的完全自动化，因此即使使用了机器翻译技术，也始终需要人力资源来进行译后编辑和质量控制。这些笔译员需要具备相应语言的适当语言技能和翻译技能。在这方面要回忆，最好只由第一语言是目标语言的笔译员来进行译入该语言的译后编辑和质量控制。质量控制由内部笔译员/审校进行。

¹ 语言司将在不久的将来改用 WorldServer。

² 例如，目前海牙体系的三种工作语言有六个独立的翻译记忆库（英-法、英-西、法-英、法-西、西-英和西-法）。

³ 见《2022 年海牙年鉴》，图 13。

⁴ 例如，WIPO Translate 用于翻译专利文件的训练数据集通常有数百万份文件。

⁵ 阿拉伯文、德文、俄文、法文、韩文、葡萄牙文、日文、西班牙文、英文和中文。

⁶ 对 WIPO Translate 的培训使用了 PCT 申请和 Patentscope 上提供的国家和地区局专利集的文本。

⁷ 目前，海牙翻译、马德里翻译、PCT 的部分语言组合以及语言司的某些文件采用定制模型。

⁸ 虽然 WIPO Translate 可以只针对一个翻译方向进行训练，但通常是针对两个方向，因为用平行数据（相同内容的两种语言）进行训练可以大大提高翻译质量。

⁹ 例如，英文和法文是语言学上相似的语言，而英文和阿拉伯文、中文或韩文是遥远的语言。英文和俄文都是印欧语系的一部分，但它们在语法上有差异，这就是为什么自由文本翻译需要更多的训练数据。

10. 神经机器翻译输出的质量对译后编辑的成本有影响。如果机器翻译输出的质量很差，译后编辑的端到端成本可能接近甚至超过直接人工翻译的成本。

三、产权组织的机器翻译实践

海牙体系

11. 要回顾的是，海牙体系国际申请的翻译工作目前由马德里注册部的翻译和术语科负责。¹⁰与海牙申请有关的翻译业务涉及产品名称、工业品外观设计特征的说明书、权利要求书和图例。¹¹产品名称、权利要求书和图例通常少于 10 个单词，但说明书为自由文本，2022 年平均为 64 个单词。¹²据观察，含有说明书的国际申请数量从 2012 年的 17%增加到 2022 年的 65%。¹³

12. 翻译和术语科采用直接翻译做法，依靠海牙体系三种工作语言的六个海牙翻译记忆库以及 WIPO Translate。以前翻译过的海牙申请文字内容的翻译记忆库目前共有约 28 万个条目，大部分都包含在英文为源语言的翻译记忆库中。国际局估计，平均而言，对于英译法或英译西，约 35%的文本语段在翻译记忆库中找到了匹配，其余文本（65%）需要对 WIPO Translate 生成的译文建议进行译后编辑。相比之下，对于法文或西班牙文为源语言的翻译，在翻译记忆库中找到匹配的情况要少得多。¹⁴因此对于这类翻译，大部分文本是用 WIPO Translate 生成的有限译文建议进行译后编辑的。

13. 此外据观察，虽然翻译记忆库不断增长，但匹配率的增长速度却慢得多。国际局估计，对于海牙文字内容，鉴于工业品外观设计领域的不断发展以及海牙申请中包含的相关文本（特别是自由文本说明书），海牙翻译记忆库的匹配率可能不会超过 40%。¹⁵因此，国际局假定，约 60%的翻译将始终为 WIPO Translate 生成译文建议的人工译后编辑。

马德里体系

14. 要回顾是的，马德里体系的语言制度与海牙体系相同，为相同的三种语言。马德里体系进行的大部分翻译工作涉及国际申请中商品和服务名称的翻译。这些商品和服务通常是一个标准化词语的清单，因此不同于海牙申请的说明书。

15. 马德里注册部翻译和术语科依靠以前翻译过的马德里文本翻译记忆库¹⁶和 WIPO Translate。马德里翻译记忆库共有 600 多万条条目，存在六个马德里翻译记忆库中。马德里翻译记忆库中马德里文本的翻译匹配率约为 65%¹⁷，其余文本需要对 WIPO Translate 生成的译文建议进行译后编辑。

¹⁰ 目前，海牙文本的所有翻译（译后编辑）都在内部进行。

¹¹ 见《日内瓦文本》（1999 年）第 5 条第(2)款(b)项第(ii)和(iii)目，《〈海牙协定〉1999 年文本和 1960 年文本共同实施细则》（下称《共同实施细则》）第 7 条第(3)款第(iv)项、第(4)款(b)项和第(5)款(a)项以及第 11 条第(2)和(3)项，《适用〈海牙协定〉的行政规程》第 403 条和第 405 条(c)项。

¹² 说明书的篇幅差异很大。例如，2021 年以及 2022 年前三季度，15%的说明书超过 100 个单词，1%超过 500 个单词。2021 年，589 份海牙申请的说明书篇幅在 100 至 500 个单词之间，30 份海牙申请中的说明书超过 500 个单词。

¹³ 这一发展是由于一些缔约方的加入，这些缔约方的主管局是审查局，要求或建议列入说明书。在这方面要指出，中国作出了声明，要求国际申请包含对外观设计特征的简要说明，而俄罗斯联邦局（ROSPATENT）强烈建议提供工业品外观设计特征的简要说明。

¹⁴ 通常只在同一申请人为非常相似的外观设计提出多件申请的情况下才会发现匹配。在这些情况中，可以为产品名称或图例找到匹配，但说明书的匹配较少。

¹⁵ 这一发展的原因在于工业品外观设计保护的性质，多数国家法律都要求新颖性和创造性。外观设计注册源于新产品开发导致的新外观设计，以及为新产品和现有产品创造的新外观设计。

¹⁶ 由于马德里体系与海牙体系有相同的三种工作语言，马德里翻译和术语科为相同的语言组合使用六个翻译记忆库。

¹⁷ 见文件 MM/LD/WG/19/7。

WIPO Translate 生成的译文建议的质量输出，特别是英文为源语言的翻译，使国际局能够减少 50%的译后编辑费用。¹⁸

16. 商标国际注册马德里体系法律发展工作组（下称马德里工作组）目前正在探讨一些可能的措施，作为与可能扩大语言制度有关的讨论的一部分，如是否可能建立包含大量阿拉伯文、中文和俄文对等词语的翻译记忆库，培训 WIPO Translate 使其能够处理这些语言的马德里翻译，以及收集各缔约方主管局可能拥有的商品和服务清单语言对。¹⁹

17. 但是，由于海牙申请和马德里申请中包含的文字内容类型不同（自由文本说明书相对于标准化词语）、工业品外观设计领域的不断发展、海牙翻译记忆库和马德里翻译记忆库的条目数量不同（28 万相对于 600 多万）造成的两者翻译记忆库匹配率不同，以及每个 WIPO Translate 模型潜在的训练数据量不同，在海牙体系中探索类似措施的希望不大。

PCT 体系

18. PCT 翻译司负责将 PCT 申请的标题和摘要从 10 种 PCT 公布语言译成英文和法文（如果它们不是这两种语言），并将国际检索报告和国际可专利性初步报告译成英文（如果它们不是英文）。²⁰

19. 发明标题是对发明的简短、准确的描述，建议长度 2 至 7 个单词。²¹PCT 申请的摘要为 50 至 150 个单词²²，包含发明公开的摘要²³。国际检索报告和国际可专利性初步报告是专利局制定的与发明可专利性有关的技术文件，根据制定报告所用的语言，国际检索报告平均为 50 至 200 个单词，国际可专利性初步报告 550 至 950 个单词。这些报告中的文字内容与海牙申请中的有很大不同。

20. PCT 翻译司依靠 10 种 PCT 公布语言的 18 种语言组合的 PCT 翻译记忆库²⁴、术语数据库（“WIPO Pearl”）²⁵和 WIPO Translate。由于多年来以各种语言提交的 PCT 申请量的差异²⁶，因此在中文、德文、日文和韩文到英文的语言组合中，翻译记忆库有大量条目，而相比之下，阿拉伯文或俄文到英文语言组合的条目不多²⁷。因此，阿译英和俄译英是由笔译员人工完成的（使用非常有限的现有翻译记忆库和术语库）。

21. 同样，由于可用于培训的数据集数量不同，WIPO Translate 目前只在内部用于中译英和英译法。目前正在评估 WIPO Translate 在其他语言组合中的有用性（尤其是法、德、日、韩译英）。

22. 为 PCT 内部翻译定制的 WIPO Translate 模型不会对海牙文字内容的翻译有用，因为首先它目前只用于两种语言的组合，其次它训练所用的文字内容来自不同的技术领域（专利文献）。

¹⁸ 这是指译后编辑的费用是在无 WIPO Translate 协助下全文翻译费用的 50%。

¹⁹ 见文件 MM/LD/WG/19/7，第 39 至 50 段。

²⁰ 见《专利合作条约实施细则》（下称《PCT 实施细则》）细则 45.1、48.3(c)、62 之二.1、72.1 和 86.2。

²¹ 见《PCT 实施细则》细则 4.3。

²² 英文摘要或者译成英文的摘要。见《PCT 实施细则》细则 8。

²³ 见《PCT 实施细则》细则 8.1(a)(i)。

²⁴ 但是，由于 PCT 翻译司为译入英文，所以几乎没有其他翻译方向的条目，即从英文译成其他 PCT 公布语言（英-法除外）。

²⁵ PCT 术语数据库包含源自 10 种 PCT 公布语言专利文献的科技术语，目前包含约 23 万个术语，可以公开访问：<https://www.wipo.int/reference/zh/wipopearl/index.html>。

²⁶ 见《2022 年 PCT 年鉴》图 C1。

²⁷ 此外，用阿拉伯文或俄文制定的国际检索报告和国际可专利性初步报告没有机器可读格式，所以这些文件的译文不能纳入 PCT 翻译记忆库或用于训练 WIPO Translate。因此，PCT 翻译记忆库中阿拉伯文和俄文的现有数据仅来自以这些语言提交的 PCT 申请的标题和摘要译文。

语言司

23. 产权组织语言司提供以下方面联合国六种语言的译入译出：产权组织和 UPOV²⁸管理的条约和其他国际文书；国家立法文本；成员国大会、其他主要机构、委员会和工作组的文件，包括工作文件和会议报告；出版物、培训材料、产权组织网站及其他宣传内容。因此，语言司翻译的文字内容与海牙申请书的文字内容不同。²⁹

24. 产权组织语言司的多数翻译³⁰是从英文译成联合国其他正式语言，少数从法文和西班牙文译成英文，很小一部分文本从联合国其他正式语言以及德文和葡萄牙文等译成英文以及从英文译成这些语言³¹。所有的翻译都是在各自翻译记忆库的支持下由笔译员人工完成的。在这方面要指出，语言司使用的翻译记忆库有从英文到阿拉伯文、中文和俄文的数据，但没有现成的从这些语言到英文的数据。几年来，对 WIPO Translate 生成的机器翻译输出的译后编辑已被用于某些类别文件的翻译，如逐字报告。

25. 为语言司内部使用定制的 WIPO Translate 模型不会对海牙文字内容的翻译有用，因为首先它训练所用的文字内容来自不同的技术领域（会议文件），其次大部分翻译是从英文译成其他语言，所以没有其他语言译成英文的数据。

四、其他考虑

翻译质量

26. 必须回顾，海牙申请中包含的文字内容的译文由国际局制定，申请人没有任何投入或控制。³²特别是，某些被指定主管局接下来在对外观设计进行实质审查时考虑产品名称和说明书或其译文，它们在所有被指定缔约方就驳回、无效宣告或侵权进行的行政或司法程序中仍然相关。因此，国际局制定的译文对缔约方管辖区内涉及保护有效性或保护范围的任何实质性决定构成了重要因素³³，而与 PCT 不同³⁴的是，申请人在其中没有参与。因此，国际局的高质量译文对于保障注册人在被指定缔约方领土内的权利必不可少。

神经机器翻译工具评估

27. 此外，国际局还对神经机器翻译工具对翻译海牙申请文字内容的有用性进行了测试评估，包括在 Patentscope 上公开提供的 WIPO Translate。测试评估是针对中国国家外观设计注册中文字内容的翻译进行的，这些内容摘自全球外观设计数据库。³⁵内部笔译员的初步结论是，除非用具体领域文字内容对技术进行培训，否则质量输出不足以用于译后编辑。不过，在翻译技术术语时，WIPO Translate 确实比其他神经机器翻译工具显示出某些优势。

²⁸ 也向 UPOV 提供笔译和口译服务，它使用四种官方语言：英文、法文、德文和西班牙文。

²⁹ 虽然语言司负责翻译与海牙有关的文本内容，如工作组文件，但这些文件中的文本与海牙申请中的文本不同。

³⁰ 产权组织语言司估计，这一比例约为 90%。

³¹ 德文和葡萄牙文的译入译出是外包的。

³² 见《共同实施细则》第 6 条第(4)款。在这方面要指出，虽然申请人可以随国际申请提交国际申请中任何文字内容的建议译文，但实践中国际局从未收到过这种译文。

³³ 缔约方可以依靠提交的说明书原文或译文。

³⁴ 由国际局翻译的 PCT 申请摘要只作为在特定技术中进行检索的扫描工具，不为任何其他目的纳入考虑，特别是不为所寻求保护的权利要求书所考虑（PCT 第 3 条第(3)款和《PCT 实施细则》细则 8.3）。保护范围由权利要求书确定，而说明书则是对发明的披露。然而，权利要求书和说明书在主管局进入国家阶段时由申请人进行翻译，国家/地区局的实质审查根据该译文进行。

³⁵ 测试评估对文字内容的中译英进行，是因为 PCT 中这种语言组合的翻译使用 WIPO Translate。

28. 可以进一步评估 WIPO Translate 在翻译中文和其他语言的海牙文字内容方面的有用性。但是，为此国际局将需要其他语言的外观设计注册样本，涵盖多种多样的产品名称、图例和说明书。³⁶

外部供应商提供的机器翻译服务

29. 如果有的话，另一种可能性是依靠外部供应商提供的机器翻译服务。但是，翻译输出的质量可能没有使用 WIPO Translate 时高，因为质量输出取决于用相关领域大量数据进行训练。在这方面要回顾，海牙翻译的 WIPO Translate 模型是用海牙翻译记忆库的语料库来训练的。此外，使用外部机器翻译服务也可能比内部开发的 WIPO Translate 费用更高。最后，由于海牙申请的保密性，需要对可能的供应商进行安全和数据保护评估。

五、机器翻译工具在引入其他语言方面的有用性

30. 目前，国际局没有任何翻译记忆库用于现有三种工作语言以外任何语言海牙申请文字内容的翻译。此外，WIPO Translate 也没有进行过其他语言的海牙术语培训。根据上文对翻译技术可用性和有用性的分析，在可预见的未来，将海牙文字内容译入译出任何其他语言将依靠于笔译员人工翻译。

31. 此外，这些工具对其他语言的有用性将取决于若干因素，特别是以该语言提交的申请量（这决定了翻译记忆库中的条目数量和 WIPO Translate 的训练数据）、用于训练 WIPO Translate 的每种语言组合其他相关训练数据的可用量，以及每种语言组合的相对相似度/距离。

与建设翻译记忆库有关的考虑

32. 如果在海牙体系中引入其他语言，国际局将需要为每种新的语言组合建立和建设翻译记忆库。

33. 在这方面，在引入其他语言之前建设翻译记忆库语料可能是提高其有用性的手段，如果能够预测到有足够数量的申请以这些语言提交，还可能减少这些语言的翻译费用。因此，将翻译记忆库中的现有条目预先译成其他语言，就像马德里工作组目前正在为商品和服务名称考虑的那样，可能是一种选择。但是，虽然创建其他语言对等术语可能对较短的文本语段（特别是产品名称或图例）有用，但对说明书的翻译没有同样的帮助³⁷，而说明书占了翻译工作的大部分。

34. 此外，翻译记忆库中条目的翻译将产生费用。³⁸因此，将现有翻译记忆库译成其他语言的成本，需要与直接翻译其他语言海牙申请文字内容、逐渐建设翻译记忆库的成本进行比较。如果申请量低于一定水平，翻译现有翻译记忆库的成本将大大高于翻译以其他语言提交的海牙申请中的文字内容。

培训 WIPO Translate

35. 对翻译记忆库未涵盖的海牙文字内容，国际局的机器翻译目前依靠 WIPO Translate 进行。为新的语言组合使用 WIPO Translate，将是一个可供选择的办法。

36. 在这方面，为了使海牙文字内容的翻译获得足够的质量，WIPO Translate 将需要用每种新语言组合相关领域的大量文本进行培训。因此，分享任何相关语言组合的术语，如产品名称、图例、权利要求书和说明述，将有所帮助。但不确定的是，缔约方主管局自己是否掌握还是可从其他地方获得大量的此类培训数据。还要强调，虽然 WIPO Translate 可以只用一个翻译方向的条目进行训练（如英译

³⁶ 这方面要指出，全球外观设计数据库不包含任何阿拉伯文或俄文数据。

³⁷ 如上所述，说明书为自由文本，相互之间差异很大。另见第 17 段。

³⁸ 将马德里翻译记忆库译成阿拉伯文、中文和俄文的费用估计为 236 万瑞士法郎至 431 万瑞士法郎。见 M/LD/WG/19/7，第 42 段。

中），但将其用于相反翻译方向的翻译（如中译英）不保证同样的翻译质量。³⁹此外，在任何情况下，如果由国际局来做，相关文字内容的前期翻译都是有成本的。⁴⁰

37. 此外要回顾，尽管有大量的潜在训练数据可用，但 PCT 系统目前仅在两种语言组合中使用 WIPO Translate⁴¹，而语言司目前仅在非常有限的范围内使用 WIPO Translate⁴²。因此，在现阶段，WIPO Translate 在翻译目前三种语言之外海牙文字内容方面的潜在有用性似乎相当有限。

38. 最后还要指出，包括 WIPO Translate 在内的神经机器翻译正在进步，因此可以建议对用于翻译海牙文字内容的现有神经机器翻译工具进行定期评估。

六、对成本估算的影响

39. 文件 H/LD/WG/10/4 中估算的翻译海牙文字内容的经常性费用，是基于 PCT 适用于翻译外包的费率。⁴³这些费率基于每种语言组合的翻译文本量、所译文本的技术性质以及 PCT 每种语言组合的具体情况，因此，如果引入其他语言，适用于每种语言的费率将在海牙体系的整体情况下发生变化。在这方面，本分析的重点是目前可用的翻译技术对这种与其他语言有关的经常性翻译费用的潜在缓解作用。

40. 该文件中的成本估算已经考虑了通过逐步编纂翻译记忆库而获得的效率。即使国际局将翻译记忆库语料译成其他语言，也不大可能提高说明书翻译的匹配率，而且前期翻译也有成本。⁴⁴事实上，鉴于工业品外观设计领域的不断发展和海牙申请中包含的相关文本，国际局观察到海牙翻译记忆库的匹配率增长放缓，因此估计大约 60%将始终需要对神经机器翻译工具生成的译文建议进行人工译后编辑。⁴⁵

41. 此外，根据目前的分析，在可预见的未来，目前可用的神经机器翻译工具不会提高其他语言海牙申请文字内容翻译的生产率或效率。翻译和质量控制的成本不可能在不影响翻译质量的情况下降低。在这方面要强调，对翻译质量的任何妥协都会损害用户的权利⁴⁶。因此，技术分析没有在海牙体系中引入其他语言方面导致对成本估算的任何修订。

42. 最后要指出，使用翻译技术只能对实际翻译的费用产生影响。如果在海牙体系中引入其他语言，所有其他费用，如支持任何额外翻译工作量和质量控制的服务器费用，无论是否使用翻译技术都会发生。

43. 请工作组讨论本文件的内容并发表评论意见。

[文件完]

³⁹ 见脚注 8。

⁴⁰ 见第 34 段。

⁴¹ 见第 21 段。

⁴² 某些类别的文件，如产权组织会议的逐字报告。见第 24 段。

⁴³ 按工作组的要求，费用估算显示了引入中文和俄文的额外翻译费用（见文件 H/LD/WG/8/8，第 29 段）。虽然可以假设引入其他语言的同样经常性翻译费用也可以根据 PCT 适用的费率来计算，但如果工作组核可，可以进行进一步评估。

⁴⁴ 见上文第 34 段和第 36 段。

⁴⁵ 见上文第 13 段。

⁴⁶ 见上文第 26 段。