

机器翻译知识普及

为人工智能时代的用户赋能

Dorothy Kenny 编

Translation and Multilingual Natural
Language Processing 23



Translation and Multilingual Natural Language Processing

Editors: Oliver Czulo (Universität Leipzig), Silvia Hansen-Schirra (Johannes Gutenberg-Universität Mainz), Reinhard Rapp (Hochschule Magdeburg-Stendal), Mario Bisiada (Universitat Pompeu Fabra)

In this series (see the complete series history at <https://langsci-press.org/catalog/series/tmnlp>):

12. Nitzke, Jean. Problem solving activities in post-editing and translation from scratch: A multi-method study.
13. Vandevoorde, Lore. Semantic differences in translation.
14. Bisiada, Mario (ed.). Empirical studies in translation and discourse.
15. Tra&Co Group (ed.). Translation, interpreting, cognition: The way out of the box.
16. Nitzke, Jean & Silvia Hansen-Schirra. A short guide to post-editing.
17. Hoberg, Felix. Informationsintegration in mehrsprachigen Textchats: Der Skype Translator im Sprachenpaar Katalanisch-Deutsch.
18. Kenny, Dorothy (ed.). Machine translation for everyone: Empowering users in the age of artificial intelligence.
19. Kajzer-Wietrzny, Marta, Adriano Ferraresi, Ilmari Ivaska & Silvia Bernardini. Mediated discourse at the European Parliament: Empirical investigations.
20. Marzouk, Shaimaa. Sprachkontrolle im Spiegel der Maschinellen Übersetzung: Untersuchung zur Wechselwirkung ausgewählter Regeln der Kontrollierten Sprache mit verschiedenen Ansätzen der Maschinellen Übersetzung.
21. Frittella, Francesca Maria. Usability research for interpreter-centred technology: The case study of SmarTerp.
22. Prandi, Bianca. Computer-assisted simultaneous interpreting: A cognitive-experimental study on terminology.
23. Kenny, Dorothy (ed.). 机器翻译知识普及: 为人工智能时代的用户赋能.

机器翻译知识普及

为人工智能时代的用户赋能

Dorothy Kenny 编



Dorothy Kenny (ed.). 2025. 机器翻译知识普及: 为人工智能时代的用户赋能
(Translation and Multilingual Natural Language Processing 23). Berlin:
Language Science Press.

This title can be downloaded at:

<http://langsci-press.org/catalog/book/492>

© 2025, the authors

Published under the Creative Commons Attribution 4.0 Licence (CC BY 4.0):

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

ISBN: 978-3-96110-501-4 (Digital)

978-3-98554-130-0 (Hardcover)

ISSN: 2364-8899

DOI: 10.5281/zenodo.14886855

Source code available from www.github.com/langsci/492

Errata: paperhive.org/documents/remote?type=langsci&id=492

Cover and concept of design: Ulrike Harbort

Translation: 钟欣奕

Typesetting: Sebastian Nordhoff

Proofreading: 洪文杰, Wilson Lui

Fonts: Libertinus, Arimo, DejaVu Sans Mono

Typesetting software: Xe_{La}TeX

Language Science Press

Scharnweberstraße 10

10247 Berlin, Germany

<http://langsci-press.org>

support@langsci-press.org

Storage and cataloguing done by FU Berlin

目 录

致谢	iii
前言	
Dorothy Kenny	v
1 欧洲、多语制与机器翻译	
Olga Torres-Hostench	1
2 人工翻译和机器翻译	
Dorothy Kenny	19
3 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	
Caroline Rossi & Alice Carré	39
4 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 & Dorothy Kenny	65
5 如何处理机器翻译的错误：译后编辑	
Sharon O'Brien	83
6 伦理道德与机器翻译	
Joss Moorkens	95
7 神经机器翻译的工作原理	
Juan Antonio Pérez-Ortiz, Mikel L. Forcada & Felipe Sánchez-Martínez	111
8 定制化机器翻译	
Gema Ramírez-Sánchez	129
9 机器翻译之于语言学习者	
Alice Carré, Dorothy Kenny, Caroline Rossi, Pilar Sánchez-Gijón & Olga Torres-Hostench	147
Index	164

致谢

本书获得来自欧盟 Erasmus+ 战略伙伴计划的支持，是“MultiTraiNMT: 多语言公民机器翻译培训”项目（项目编号：2019-1-ES01-KA203-064245）的部分研究成果。

前言

Dorothy Kenny
都柏林城市大学

笔者在前言中阐述出版本书的缘由动机，并为读者提供使用建议。

1 写作背景

多语制是欧盟的基本价值观，实际上有赖于语言学习和翻译等多个支柱。近年来，这两大支柱都受到了机器翻译持续发展的深刻影响。如果使用得当，机器翻译不仅有助于语言学习，还有助于更多用户阅读更多语言的文本。条件适宜的情况下，机器翻译能显著提高职业译者的工作效率。但是，若不加批判地使用，这项技术不但有碍于个人学习语言，还会掩盖机器译文的潜在问题，导致用户无法意识到阅读的文本中包含机器翻译造成的误译或偏见。过度吹捧机器翻译，可能会吓退想要从事翻译工作的人。

为了利用好机器翻译，又无损多语制的其他支柱，我们认同Bowker & Ciro (2019)的观点，即优秀多语公民的身份和翻译职业必须以机器翻译素养为基础。针对不同群体，机器翻译素养的要求也不同。例如，Bowker & Ciro (2019)着眼国际学术界来探讨机器翻译素养。本书的目标读者包括两个群体：第一，偶尔使用机器翻译的用户，即利用这项技术获取信息，或无意中用到机器译文，或以语言学习为目的的用户；第二，职业译者或正在接受翻译培训的人。我们认为，所有机器翻译使用者都应基本了解机器翻译技术的重要性及其对维护多语言制度的作用，还应该对其工作原理有所认识，这样才能避免犯常见错误，从而有效利用这项技术。而希望进一步掌握这项技术的用户，可以通过阅读本书学习如何使机器翻译的价值最大化，例如如何进行译前编辑以获取更优质的译文。这类用户可能还对如何改进机器译文感兴趣。而已经或即将进入翻译产业的用户会特别关注机器译文评测，以衡量译文是否“适用”。他们还可能参与将机器翻译融入公司工作流程，或者需要了解如何定制机器翻译引擎，以满足特定客户群体的需求。此外，他们还会关注机器翻译如何影响自己的工作条件。这类读者需要更深入了解机器翻译技术及其辅助技术和工具。所有用户都应该对在机器翻译使用过程中，由不同



原因导致的各种道德伦理问题有所了解。有些用户担心滋生与机器翻译相关的作弊行为，例如学生在什么情况下使用机器翻译构成作弊？以专业译者为主的其他用户要考虑到，使用机器翻译可能造成违约。另外，如今大家都必须保护他人的隐私和数据权利。和许多其他技术一样，机器翻译也可能损害自然环境。而且众所周知，机器译文会带有偏见，例如阳性变体多于阴性变体。和所有通信技术一样，机器翻译技术也具有两面性。以上都是我们关心的问题。

2 本书使用指南

本书将引导读者探究上述所有问题。阅读本书不要求读者事先对普通翻译或专业机器翻译有所了解。涉及机器翻译的技术问题，特别是神经机器翻译（基于人工神经网络的最先进的翻译技术）时，我们会对这些技术进行描述，避免使用读者可能不熟悉的数学概念。本书尝试提供形象化的解释，特别是隐喻，以帮助读者理解这个领域的常见概念。这样一来，我们就能帮助读者轻松入门机器学习，展现这一领域的丰富内涵。总的来说，读者会觉得相比前几章，后面的章节及其后半部分的专业性更强。有些读者可以跳过其中一些章节，但即便机器翻译专家，或许也能从“非技术性”章节中受益，如讨论机器翻译与道德伦理的章节。

3 本书结构

本书以 Olga Torres-Hostench 在第一章中对多语制的讨论开篇，阐述其意义和实施方式，尤其是在欧盟内部的实施。她说明了为什么要将机器翻译整合到语言学习和翻译之中。第二章为 Dorothy Kenny 对翻译定义的概述，尤其是机器翻译，旨在对翻译进行祛魅。随后，她用简单易懂的方式介绍当代机器翻译方法的基本概念，如人工智能和机器学习。第三章由 Caroline Rossi 和 Alice Carré 共同撰写，主要介绍机器翻译测评这一具有重要学术和经济价值的领域。在第四章中，Pilar Sánchez-Gijón 和 Dorothy Kenny 阐述了如何使文章更适用于机器翻译。Sharon O'Brien 则在第五章介绍如何改进机器译文，即译后编辑。第六章由 Joss Moorkens 撰写，讨论了机器翻译使用过程中出现的道德伦理问题。第七章是本书专业性最强的一章。由 Juan Antonio Pérez-Ortiz, Mikel Forcada 和 Felipe Sánchez-Martínez 解释了神经机器翻译的工作原理，涵盖目前机翻系统最常用的基本技术。Gema Ramírez-Sánchez 在第八章中介绍如何定制化神经机器翻译。最后一章则专门讨论机器翻译与语言学习，由 Alice Carré, Dorothy Kenny、Caroline Rossi、Pilar Sánchez-Gijón 和 Olga Torres-Hostench 共同撰写。

配套资源

本书每章都设有互动练习,可访问 MultiTraiNMT 网站<http://www.multitrainmt.eu>/获取。这些活动的永久链接为<https://ddd.uab.cat/record/257869>。大多数活动可以自主完成,有些则在教师指导下进行为佳。

MultiTraiNMT 项目还创建了 MutNMT 教学平台,旨在帮助用户学习如何训练、定制和测评神经机器翻译系统。用户可通过 MultiTraiNMT 网站进入该平台,以便更好学习本书的第七章和第八章。

References

Bowker, Lynne & Jairo Buitrago Ciro. 2019. *Machine translation and global research*. Bingley: Emerald Publishing.

第一章

欧洲、多语制与机器翻译

Olga Torres-Hostench
巴塞罗那自治大学

本章将介绍作为欧盟基本原则的多语制，说明如何将其付诸实践，以及如何通过语言学习和翻译来支持多语制。作者以高等院校为例，认为机器翻译能促进多语制在高校的实践。

1 前言


欧盟以“和而不同”(united in diversity) 为创立理念，寓意“语言多样性和语言学习对欧洲项目具有重要意义”(European Commission 2021)。但是，欧盟的多语制政策大多基于语言学习和人员流动，两者都相当耗时。而且人类的语言学习颇具挑战性。毕竟，欧盟普通公民能够学习的语言数量有限。本章将为解决以上问题提供建议，指出机器翻译有助于促进欧洲的多语制，从而促进欧洲的语言多样性。

2 多语并用的欧盟

欧盟所谓基于人道主义的多语制不过只是幻想。对此，大家都心照不宣。(House 2003: 561)

ISO639-3 是由国际标准化组织 (International Organization for Standardization, ISO) 开发的一组代码，为所有已知人类语言都分配了由三个字母组成的识别码。截至 2020 年 1 月 30 日，该标准收录了 7868 种语言 (Wikizero 2020)，其中约 600 种在欧洲使用，24 种为欧盟官方语言。欧盟官方语言包括：荷兰语、法语、德语、意大利语 (始于 1958 年)、丹麦语、英语 (始于 1973 年)、希腊语 (始于 1981 年)、葡萄牙语、西班牙语 (始于 1986 年)、芬兰语、瑞典语 (始于 1995 年)、捷克语、爱沙尼亚语、匈牙利语、拉脱维亚语、立陶宛语、马



Olga Torres-Hostench. 2025. 欧洲、多语制与机器翻译. In Dorothy Kenny (ed.), 机器翻译知识普及: 为人工智能时代的用户赋能, 1-17. Berlin: Language Science Press. DOI: 10.5281/zenodo.14922285 

耳他语、波兰语、斯洛伐克语、斯洛文尼亚语（始于 2004 年）、保加利亚语、爱尔兰语、罗马尼亚语（始于 2007 年）和克罗地亚语（始于 2013 年）。

语言多样性是欧洲文化遗产的一部分。在欧洲，有些语言具有国家级官方地位，而有些则是区域性土著语言和/或少数民族语言，受认可程度各有不同。《欧洲区域或少数民族语言宪章》（以下简称《宪章》）是于 1998 年通过的欧洲条约，旨在保护和促进传统少数民族使用的语言。2019 年，《宪章》在监管机制指导下进行了改革与加强。《宪章》涵盖 201 个少数民族或语言群体的 79 种语言 (Council of Europe 2020)。这 79 种语言如表 1 所示表 1。

表 1: 《欧洲区域或少数民族语言宪章》所涵盖的语言

阿尔巴尼亚语	法兰克-普罗旺斯	兰科语	低地苏格兰语
阿拉贡语	语	莱昂语	苏格兰-盖尔语
阿兰语	法语	林堡语	塞尔维亚语
亚美尼亚语	弗里斯兰语	立陶宛语	斯科尔特萨米语
亚述语	加告兹语	低地德语	斯洛伐克语
阿斯图里亚斯语	加利西亚语	下萨克森语	斯洛文尼亚语
巴斯克语	德语	下索布语	南萨米语
贝亚斯语	希腊语	吕勒萨米语	瑞典语
白俄罗斯语	匈牙利语	马其顿语	鞑靼语
波斯尼亚语	伊纳里-萨米语	曼岛语	土耳其语
保加利亚语	爱尔兰语	梅安语	乌克兰语
布涅瓦茨语	伊斯特拉-罗马尼	摩尔多瓦语	阿尔斯特苏格兰
加泰罗尼亚语	亚语	北弗里斯兰语	语
康沃尔语	意大利语	北萨米语	上索布语
克里米亚鞑靼语	卡拉伊姆语	波兰语	巴伦西亚语
克罗地亚语	卡累利阿语	罗姆语	弗拉赫语
塞浦路斯马龙	卡舒比语	罗马尼亚语	威尔士语
派阿拉伯语	克里姆查克语	罗曼什语	叶尼什语
捷克语	库尔德语	俄语	雅兹迪语
丹麦语	克文语/芬兰语	鲁塞尼亚语	意第绪语
芬兰语	拉定语	萨特弗里斯兰语	

根据《宪章》，其中一些语言只会在一个国家受到保护，如芬兰的斯科尔特萨米语，而其他语言则应在多个国家受到保护，如斯洛文尼亚语在奥地利、波黑、克罗地亚和匈牙利都受到保护。除了《宪章》提及的语言，还有一些语言得到不同程度的认可。例如，意大利的撒丁岛自治区承认撒丁语为官方语

言,而在意大利北部山区的某些市镇使用的罗曼什语、拉定语、辛布里语和莫其诺语也得到地方认可。

然而,《宪章》只保障地区少数群体,而非移民群体的权利。此外,《宪章》还遗漏了某些语言,如法国西北部的布列塔尼语,尽管布列塔尼地区为促进该语言的日常使用,于2010年创建了一所布列塔尼语语言机构。

欧洲范围内的移民和人口流动也强化了多语制。例如,欧洲国家之间的移民使得有人在安道尔讲葡萄牙语,在爱尔兰讲波兰语,甚至讲欧盟以外的语言,如汉语(普通话)或阿拉伯语。《多语种城市计划》(Extra & Yagmur 2005)对比利时布鲁塞尔、德国汉堡、法国里昂、西班牙马德里、荷兰海牙和瑞典哥德堡这几座城市的中小学学生开展了家庭语言使用调查。收集到的语言清单如下:罗姆语、土耳其语、乌尔都语、亚美尼亚语、俄语、塞尔维亚语/克罗地亚语/波斯尼亚语、阿尔巴尼亚语、越南语、汉语、阿拉伯语、波兰语、索马里语、葡萄牙语、柏柏尔语、库尔德语、西班牙语、法语、意大利语、英语和德语。该研究的作者提出了清晰但具有挑衅性的结论:

在参与调查城市的20种主要语言中,10种源自欧洲,其他10种源自其他国家。调查结果表明,我们应该重新考量并扩大传统认知中的欧洲语言多样性(Extra & Yagmur 2005)。

可是,什么叫做语言多样性的“传统认知”呢?

2.1 象征欧洲语言多样性的24种官方语言

欧盟将其语言多样性视为宝贵资产。《欧洲区域或少数民族语言宪章》(成员国,2012年)第22条规定:“欧盟应尊重文化、宗教和语言多样性”。然而,欧盟成员国独有界定和承认国家和地区少数民族语言的权利,且语言政策也极具争议性。

可与此同时,欧盟也以其主要机构使用24种官方语言来支持语言多样性而感到自豪。从实践的角度来看,这种做法带来一个值得深思的重大挑战。

例如,欧洲议会的所有文件都以所有官方语言出版,“因为欧盟公民必须能够以本国语言阅读切身的立法”(European Parliament 2020),而且欧洲议会议员有权以任何官方语言发言和撰写文件。《欧洲议会议事规则》第167条与语言相关,规定:(一)所有议会文件应以官方语言编写;(二)所有议员均有权自行选择官方语言在议会发言;(三)应提供口译服务;(四)议会议长应该对不同官方语言版本之间提出存在的任何差异作出裁决(European Parliament 2021)。

至于欧盟公民,根据《欧盟运作条约》(Treaty on the Functioning of the European Union, TFEU¹),所有欧洲公民都有权以欧盟的任何官方语言向欧

¹《欧盟运作条约》的最新版本请参考<https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:12012E/TXT&from=EN>。除非另有说明,本章中所有网址的最后访问日期均为2022年1月。

盟机构提出问题,并得到以同种语言写就的答复。这是为了使欧盟机构更加民主,使欧盟公民更容易接触。《欧盟运作条约》中与多语制有关的其他规定可参考第 20 条、第 24 条和第 342 条。

有人认为 24 种官方语言太多,有人则认为不够。有些国家尝试采取其他办法。例如,西班牙人使用的加泰罗尼亚语、巴斯克语和加利西亚语被欧盟认为是“附加语言”(和西班牙语一样,这些语言在各自使用的地区均属于官方语言)。这种地位意味着,在西班牙,欧盟公民用上述语言进行的任何沟通都必须翻译成欧盟的“议事语言”,而欧盟机构的答复也必须从议事语言翻译成附加语言。所产生的翻译成本由西班牙承担。

使用英语、法语和德语这三种议事语言旨在简化欧盟的多语种沟通——对欧盟来说,24 种官方语言就意味着存在 552 种翻译语言对组合,“因为每种语言都可以翻译成其他 23 种语言”(European Parliament 2020),这样一来,文件翻译的工作量十分惊人。因此,有规定指明哪些文件需要翻译成其他 23 种语言,哪些文件只需翻译成上述三种议事语言。

欧盟委员会翻译总司(European Commission's Directorate-General for Translation, DGT)为欧盟各机构和公民翻译文本。截至 2022 年,每年翻译的文件页数超过 275 万页,其中 91% 为英语原文,2% 为法语原文,西班牙语原文不足 1%,最少的是德语原文。其他源语言合计约占翻译服务的 5%。在所有翻译文件中,63% 由翻译总司的内部翻译完成,37% 外包给其他公司。翻译内容涉及多个领域,包括欧盟法律制定(约 55%)、对外沟通和网络文本(22%)、与其他欧盟机构和成员国议会的沟通(12%)、与欧盟公民的沟通(5%)、其他官方文件(4%)以及欧盟政策公众咨询(2%)。2022 年的翻译预算为 3.55 亿欧元,占整个欧盟预算的 0.2%(European Commission, Directorate-General for Translation 2022)。

2.2 机器翻译的加入

翻译总司拥有 2000 名内部员工,包括语言专家和技术支持人员,并与数千名选定外部翻译合作(European Commission, Directorate-General for Translation 2022)。所有语言的译文都存于 Euramis 系统(EUROpean Advanced Multilingual Information System,“欧洲高级多语言信息系统”),如欧盟 24 种官方语言的立法文件汇编语料库《共同法律总汇》(*Acquis Communautaire*)。为了提高生产率和降低成本,欧盟委员会翻译总司在其部分工作流程中引入了机器翻译,目前使用的机器翻译系统名为 eTranslation²。启用该系统的初衷是为欧盟节省时间和资金,但并不仅限于此。最后,随着机器翻译的进一步发展,它不仅能满足文件翻译需求的增长,翻译从未提上翻译日程的文件,将来还能增加翻译的语言对,从而更好地体现欧洲语言的多样性。不过

²https://ec.europa.eu/info/resources-partners/machine-translation-public-administrations-ettranslation_en

短短的几年时间，机器翻译的发展让这一美好愿景不再是天方夜谭。对于原住民或地区少数民族、“非领土”或移民少数民族的语言来说，这是它们与国家官方语言一样，在欧盟机构获得代表性的唯一途径。

2.3 多语制对欧盟意味着什么？

多语制政策是维护和保障前文所述的语言多样性的一种方式。在欧盟，多语制也被视为加强社会凝聚力和促进劳动力流通的一种手段：“[1] 语言能力有助于提高欧洲公民的流动性，就业能力和个人发展” (Council of the EU 2014)。

之所以采用多语制的语言政策，是为了在多语通用的国家和组织促进语言发展。多语制有多种形式，如：

- 这是一种多语言政策，即组织、公司或机构在对内和对外沟通中使用两种以上语言的策略。
- 也可以指使用多种语言的欧盟，即不同的语言在欧盟共存。
- 能够使用多种语言的多语种公民。
- 使用多种语言的医疗体系，即为新居民提供更好的医疗服务而融合语言多样性的医疗体系。

欧盟主要的多语制政策可参考相关文件 (European Council 2002, European Commission 2008, Council of the EU 2008a,b, 2011, 2014, European Council 2017)。

欧盟的多语制政策和声明经常提到语言学习，特别是“母语加两种外语”政策。根据该政策，欧盟公民“从小应至少学习两种外语” (European Council 2002)。但是，这一政策的实施就真的足以培养出多语公民吗？我认为，有必要将更多的内容纳入欧盟的多语政策。

多语制关乎语言政策，但并不仅限于此。Cenoz (2013)提出了多语制的广泛定义，提醒我们，多语制是多维的，例如涉及个人与社会层面、能力与使用层面、双语制与多语制等。多语制也适用于地理区域或社会领域。它也可以应用于地理区域或社会领域。此外，多语制还可以从不同角度进行研究，包括认知、社会构建、身份、语言实践、多模态和技术等。而我在本章中使用的既不是维基百科的简单定义（“一个或多个发言者使用一种以上的语言”），也不是Cenoz (2013)的复杂多维定义，而是欧盟委员会对多语制的定义，即“社会、机构、团体和个人在日常生活中经常使用一种以上语言的能力”（欧盟委员会，2007年）。“接触”一词使我们能够结合本书中有用的细微差别。通过撰写本章，我谨请读者思考是否可以从技术角度来定义多语制，从而调整欧盟委员会使用的上述定义。

话题讨论

是否存在“技术层面的多语制”，即在日常生活中，企业、机构、团体和个人借助多语翻译工具经常接触一种以上语言的能力？

有趣的是，我们使用的多语制定义出自多语制高级小组的上述报告，提到将“多语电子工具或能辅助非专家使用者学习第二和第三外语”作为研究领域 (European Commission 2007)。同样，欧盟委员会在《多语制：既是财富，又是承诺》(欧盟委员会，2014 年) 的函件中称，“欧盟的语言差距可以通过媒体、新型技术和翻译服务来缩小”。本书旨在为这一领域作出贡献。

2.4 欧盟的语言多样性促进行动

欧盟关于语言多样性的网页³上提到了以下促进语言多样性的举措：欧洲语言日、“伊拉斯谟计划”交换项目、“欧洲文化之都”活动和“创意欧洲计划”：

- 欧洲语言日⁴由欧洲委员会于 2001 年设立，于每年 9 月 26 日th 举行。这一天，欧盟国家会开展促进语言多样性和提高其他语言能力的活动。
- “伊拉斯谟计划”交换项目。2014 年至 2020 年期间，该计划为 400 多万个交流项目拨款 147 亿欧元，其中包括 200 万个大学生交流项目。
- “欧洲文化之都”活动强调包括语言在内的欧洲文化多样性。例如，2020 年的“欧洲文化之都”为克罗地亚的里耶卡市 (Rijeka，通用语言为克罗地亚语) 和爱尔兰的高威市 (Galway，通用语言为爱尔兰语和英语)。
- 创意欧洲计划⁵由欧盟委员会推出。这项计划支持包括文学翻译在内的文化和音像行业。具体而言，该计划为欧洲语言之间的文学作品翻译提供资助。
- 还有一个有意思的语言多样性促进倡议，那就是欧盟资助项目下的“欧洲语言标签”⁶。这些项目虽然大多面向语言学习，但也有为语言多样性而专门设立的。

³https://ec.europa.eu/education/policies/language-diversity_en

⁴<https://edl.ecml.at>

⁵<http://www.creativeeuropeuk.eu/funding-opportunities/literatory-translation-0>

⁶https://ec.europa.eu/education/initiatives/label/label_public/index.cfm

讨论

就目前看来，欧盟在谈论语言多样性和多语制时，对机器翻译还未给予足够的关注。如何将机器翻译纳入语言学习项目？

将机器翻译纳入多语制举措，或能帮助欧洲公民增加可以掌握的语言数量。还能帮助他们提升兴趣，克服畏难心理，轻松进入不熟悉的语言环境，尊重当地语言。此外，机器翻译还可用于帮助使用者理解未学习过的语言文本。

2.5 欧盟的多语制和语言学习

根据欧洲委员会语言政策门户网站：⁷

语言学习者/使用者是语言政策方案工作的核心。无论其地位，所有语言都包括在内：外语、主要的教学语言、家庭使用语言、少数民族或地区语言，以及关于移民和难民的语言融合的特别计划。

促进多语制的倡议数不胜数，但语言学习值得密切注意，特别是考虑到上述的欧盟“母语加两种外语”政策。

欧盟近期的语言技能提升措施包括成立欧洲现代语言中心 (European Centre for Modern Languages, www.ecml.at；欧洲教育系统和政策网络报告 (Eurydice 2019))，该中心提供有关给予欧洲学校教授地区或少数民族语言的政策支持的信息，介绍了欧洲给予学校教授地区或少数民族语言的政策支持)、在线语言支持 (OLS) 平台⁸；欧洲语言共同参考框架 (CEFR)；“伊拉斯谟计划”交换项目。

这些为促进语言学习欧洲受资项目值得我们特别关注。项目所涉及的方法、语言和国家都各不相同。具体案例请参考表 2。

欧盟统计局的官方网站 Eurostat⁹提供了欧盟国家不同年级学生的第二语言和外语学习统计数据。该网站 2019 年的数据显示，初中生学习最多的外语是英语（将近 86.8%），其次是法语 (19.4%)、德语 (18.3%) 和西班牙语 (17.5%)(Eurostat 2022)。

⁷<https://www.coe.int/en/web/language-policy/home>

⁸<https://erasmusplusols.eu/en/about-ols/>

⁹<https://appsso.eurostat.ec.europa.eu>

表 2: 关注语言学习的欧洲项目案例

项目	语言	特点
iTongue: Our Multilingual Future (2013)	未指定	用于外语学习的副语言学数字工具
Massive open online courses with videos for palliative clinical field and intercultural and multilingual medical communication (2014)	德语、英语、法语、意大利语、西班牙语、罗马尼亚语	姑息疗法的 20 项基本程序的多语版本
Crafting Employability Strategies for HE Students of Languages in Europe (2015).	未指定	将就业能力纳入语言教学
LMOOCs for university students on the move (2018)	法语、西班牙语	面向大学生的公开教育资源
Gamifying Academic English Skills in Higher Education: Reading Academic English App (2016).	未指定	基于游戏的应用程序, 以提高大学生的学术英语阅读技能
E-LENGUA: E-Learning Novelties towards the Goal of a Universal Acquisition of Foreign and Second Languages (2015).	英语、阿拉伯语、西班牙语、法语、德语、意大利语、葡萄牙语	将数字能力纳入语言教学的最佳实践
EULALIA: Enhancing University Language courses with an App powered by game-based Learning and tangible user Interfaces Activities (2019).	意大利语、波兰语、西班牙语、马耳他语	基于移动学习模式和游戏式学习方法以及有形用户界面 (TUI) 应用的包容性学习工具

话题讨论

提到欧洲的第二语言学习,“多语言”是否暗指“英语”?

还有一个有趣的问题,有多少学生依照欧洲理事会European Council (2002)的建议,学习两种或两种以上的外语?据了解,2019年,有89.9%的中学生(近1400万人)学习了一门以上的外语(Eurostat 2022)。其中700多万人(48.1%)学习了两种或两种以上的外语

总的来说,有7亿多来自欧盟或欧盟以外国家的人(World Bank 2020)在欧洲使用的语言多达600种,而大多数欧盟国家学生学习的第一或第二外语主要有四种:英语、法语、德语或西班牙语。

2.6 外语水平

欧洲语言能力调查(European Commission 2019)选择在16个教育系统中最广泛教授的两门外语,对其54000名初中生(14至15岁)进行了水平测试。测试内容包括写作、阅读理解和听力理解,不设口语测试。调查的主要结论是,只有42%的受试学生能自主运用其第一外语(即达到欧洲语言共同参考框架的B1和B2级别),而第二外语达到同等水平的学生只占25%。此外,还有很多受试学生甚至没有达到初级水平:其中第一外语未到初级水平的占14%,第二外语未到初级水平的占20%(European Commission 2019)。

该调查报告还指出,2018年《欧洲晴雨表快讯》(flash Eurobarometer)对15至30岁人群进行调查的数据显示,85%的受访者表示希望能提高自己的外语水平(主要是英语):

这表明,受访者对自己在义务教育阶段的外语水平并不满意,或无法继续保持在校水平。接受调查的欧洲年轻人中,有三分之一表示不能使用除了在校使用语言(即通常是母语)以外的语言学习。(European Commission 2019: 102)

2.7 机器翻译能助力语言学习吗?

我们已经看到,欧盟国家中的语言学习往往集中在少数常见语种上,而学习者常常对自己的外语水平并不满意。这些情况表明,需要对语言学习提供进一步的支持,而我们有责任研究机器翻译在其中的作用。由于神经机器翻译的学习速度比任何外语学习者都快,理论上,它可以用来帮助学习者理解复杂文本,提高他们的第二语言书写技能。他们可以了解如何充分利用机器翻

译来学习第二语言，并依靠自己的语言水平来识别和改正机器翻译的错误。虽然有关机器翻译如何运用于语言课程的实证研究仍然不足，但还是有一些资料介绍了这方面的研究路径。Carré 等人 Carré et al. (2025 [本卷]) 讨论了相关研究，其中包括可用于语言学习课堂的进一步想法和策略。还有一些被列入 MultiTraiNMT 项目的活动数据库 (MultiTraiNMT 2020)。我认为，在语言课上使用机器翻译的方法有很多。如果有意识地加以批判使用，就没有必要禁止。

然而，有时确实没有足够的时间来培训第二语言的学生。其实，在机器翻译的发展历程中，很多研究的部分起因是找不到会说某种外语的人。冷战时期的俄英机器翻译研究正是如此 (Gordin 2016)。最近，2020 年日本东京奥运会的组织者（由于新冠爆发，实际上于 2021 年举行）意识到，不可能让大多数外国人学习日语，因此需要一个更快的方法来克服奥运会期间的语言障碍。因此，日本内务省拨款 13.8 亿日元用于机器翻译研究，以提高实时语音翻译技术的质量，目的是覆盖 90% 的奥运团队和游客的语言需求，希望他们能在奥运期间前往日本 (Murai 2015)。日本政府为东京奥运会专用机器翻译系统的研究提供资金，私营公司则负责研发运行该系统的设备和移动应用程序。该计划是，公司将通过向用户出售设备和应用程序订阅来收回投资成本。在这种情况下，机器翻译便成为日本快速营造多语环境的方法，而非语言学习。

3 案例研究：多语言大学

3.1 国际化与多语制

欧盟委员会关于“全球欧洲高等教育”的通报确定了欧洲大学国际化的重点是 访学交流、数字学习和加强战略合作。通报中有关语言的内容如下：

对于语言学习者、教师和机构来说，英语水平的确是任何国际化战略的一部分。部分欧盟成员国已经或正在引入有针对性的英语课程（尤其是在硕士阶段），作为其人才吸引战略的一部分，否则他们不会来欧洲。(European Commission 2013)

同时，多语制是欧洲的重要资产：它受到国际学生的高度重视，应该在整个高等教育课程的教学和研究中得到支持 (European Commission 2013)。

事实上，欧盟仍然致力于在大学校园内推广多语制实践。原因有三。首先，因多语言校园反映了欧洲的语言多样性；其次，这为学生提供了更多的出国交流和就业机会；第三，这能促进与不同文化和学习方法的交流。与此类似，Gao (2019) 列出了大学参与国际化的不同原因，包括国际化可以帮助学生准备好与来自不同文化的人互动，以此来促进文化理解，减少国家间的敌意。

然而,国际化给大学带来了挑战(同上),特别是与多语制有关的挑战。首先,翻译资源(包括人力资源)的缺乏有碍于大学成为完全使用英语的场所。其次,国际化可能涉及到母语地位的下降和语言多样性的丧失。不过,我认为机器翻译有助于应对这些困境。鉴于对国际理解的转变,这项技术似乎颇具前景:一直以来,大学都会制定“国内国际化”(吸引外国学生)和“国外国际化”(将学生送往国外)的计划。在此基础上,Mittelmeier et al. (2020)纳入了“远程国际化”的概念,为以校园为基础的机构开发在线国际远程学习模式。新冠疫情的爆发无疑加速了第三种方式的发展。技术可能会改变国际化的概念,而机器翻译是一种有可能对国际化作出贡献的技术。

3.2 英语通用语(ELF)、本地语言(LLS)与机器翻译

高校的国际化战略多种多样,但最常见的可能是访学交流、开设英语或双语课程。后者有不同的名称,例如,英语通用语、英语化和全英语教学(EMI)。它被定义为“在非英语国家的高等教育机构中使用英语作为教学媒介语言”(Multilingual Higher Education 2016)。

英语的使用便于很多国际学生听课。但是,如果本地大学出于经济原因(因为国际学生的学费比本地学生的高)而更照顾国际学生的话,英语水平不够好的本地学生可能会有被辜负的感觉。有些大学在慕课(译者注:大规模开放在线课堂)上提供英语授课,想以此吸引国际学生,但最近研究表明,开设慕课对于吸引国际学生收效甚微。(Zakharova 2019)。

为了克服作为通用语的英语和本地语言之间紧张关系,House (2003)将其区分为“交流语”和“认同语”。豪斯House (2003: 560)认为,交流语有助于母语不同的人进行交流。而认同语

通常是地方语言,特别是个人的母语,这可能是决定身份认同的主要因素,意味着在界定母语群体及其成员的集体语言文化资本中发挥作用……以及认同中涉及的情感-情绪素质类型。(House 2003: 561)

根据这一分类,英语是母语不同的人的交流语,是一种“专业知识”(House 2003: 561),而认同语在母语相同的人群内部使用。

豪斯House (2003)在德国汉堡大学对使用英语作为教学语言进行了一项案例研究,观察英语如何与当地语言互动,以及国际学生如何看待和应对这种“双层语言(diglossic)场景”(House 2003: 570)。结果显示,人们并不认为英语和德语存在竞争关系。英语“自成一类”,是一种超国家的辅助交流手段。在该研究中,“(目前)没有迹象表明当地语言(德语)和多语制受到威胁”(House 2003: 574),而且学校鼓励国际学生留学期间学习德语。

3.3 一所真正的多语大学

目前的多语大学是指为国际学生提供英语授课的大学，以及边境地区或使用两种以上官方语言的地区的大学。这些大学使用多种语言可能是出于历史、政治、地理或经济等原因，因此难以在语言政策上找到平衡点。国际化和多语言政策都不可能一蹴而就。Knight (1994: 12) 提出了机构国际化进程的六个阶段：

- 认清需求、目的和益处
- 所有大学行为者承诺参与
- 规划：确定资源、目的和目标，以及优先事项和战略
- 实施：开展学术活动和服务
- 评审：评估和提高质量、影响和进度
- 强化：设置激励、认可和奖励方法

这项与多语制特别相关的计划要求我们回答以下问题：英语授课是否仅限于国际学生或课程？如果只面向国际学生，这所大学是否能称为多语大学？如果国际学生在“多语”大学里只使用英语，那他们还算是多语使用者吗？如果一所大学只在网站上使用官方语言和英语，但用当地语言授课，这样就够了吗？授课使用多种语言，还是只一种语言？学生使用的本地语言（可能不是大学的官方语言）能够融入课堂吗？授课语言是判断“多语言大学”的唯一标准吗？提供哪些语言的教学资料？语言能力是否与非语言内容一同评估？本地学生如何为英语授课做准备？多语种是否意味着使用英语授课？多语种是否意味着使用当地语言和英语？若是，分别占比多少？要有多少种语言才足以成为多语种大学？需要融入哪些语言来保证大学生进一步融入地理区域或当地经济？国际学生会生活在只有英语的舒适圈里吗？

在一个真正的多语言校园里，大学既欢迎来自国际学生的语言，也欢迎来自社会或文化边缘化群体的语言。许多地方语言在学术界历来被低估，有人对其学术词汇的丰富程度表示怀疑。一个极端的例子是克丘亚语（Quechua）：第一篇完全用克丘亚语撰写的博士论文于 2019 年进行答辩，这年离秘鲁第一所大学成立大约过去了 468 年¹⁰。

作为“伊拉斯谟计划”多语言高等教育 Multilingual Higher Education (2016) 的一部分，网络课程《多语制和教育中的多语言》(Multilingualism and plurilingualism in education) 介绍了不同多语言大学的语言政策。例如，

¹⁰<https://www.theguardian.com/world/2019/oct/27/peru-student-roxana-quispe-collantes-thesis-inca-language-quechua>

瑞士弗里堡大学设有法语和德语课程；芬兰赫尔辛基大学设有芬兰语、瑞典语和英语授课；意大利博尔扎诺自由大学设有德语、意大利语和拉定语授课，并以英语作为通用语；卢森堡大学至少 20% 的课程同时以法语、英语和德语这三种语言授课。

Gao (2019: 89) 提出了用以区分国际化的战略愿望和现实的衡量维度和指标。针对多语制，我们对这些建议做出了调整。

在大学治理方面，促进采用多语制的行动可包括：（一）支持性的多语种政策框架/组织结构；（二）语言办公室/翻译部门；（三）机器翻译基础设施；（四）大学内的多语种展示/标识；（五）培养教工的多语种意识和技能；（六）用于支持多语制的预算；（七）多语情况监测/评价系统。

从学术角度来看，促进采用多语制的行动可包括：（一）多语课程（为什么课程必须使用一种且只有一种语言？）；（二）多语教学，规范多语课堂；（三）多语研究和多语会议；（四）班上多语学生，包括国际学生和本地学生之间的互动；（五）多语访问学者；（六）多语教学大纲；（七）多语研究期刊；（八）多语课外活动和（九）文化多样性的可见性。

最后，大学可以提供多语言定向计划、多语言支持和多语言图书馆。

3.4 在多语言高校教学、科研、管理中使用机器翻译的思考

多语言大学可以使用英语和当地语言互译最多的机器翻译系统。第一种策略是使用免费在线机器翻译服务，但用户应该识别并能够纠正机器翻译错误，或自行纠错，或求助专业的译后编辑服务。关于译后编辑的更多信息可参阅O'Brien (2025 [本卷])。

第二种策略是开发大学定制的本地语言和英语之间的高质量机器翻译资源，详见Ramírez-Sánchez (2025 [本卷])。如果资源允许，欧洲和欧洲以外的大学可以共享定制的机器翻译系统。

在未来真正的多语言校园里：

- 国际学生能和当地学生打成一片，且能获取机器翻译资源，以便使用任何一种本地语言听课，因为会有：（一）不同语言的教学材料（假设版权问题已经解决）；（二）使用多语种词汇表和专门术语数据库；（三）使用语音识别、转录和机器翻译功能等对课程进行录音。
- 学生将学习译后编辑的技能，以查看英语和/或其他语言的机器翻译结果，以便能够正确使用机器翻译输出。
- 大学将为高质量的教学指南和教材提供译后编辑服务。
- 将鼓励多语种研究传播和多语种出版物，提供嵌入式机器翻译功能。

- 希望用母语发言的客座教授将会得到翻译服务，无论是人工翻译（若资金允许的话），还是机器翻译，而不是用蹩脚的英语费力地完成演讲和会议论文。
- 将在校园内组织多语言活动，涉及音乐、戏剧、烹饪、政治、文学、团结、社会服务等各个领域。

在一所真正的多语言大学里，当地语言、英语和其他由留学生带来的语言应该共存。这一策略有助于国际学生融入多语环境。

4 结语

本章支持机器翻译作为推进欧洲多语制的工具。如本章所示，欧盟发布了宪章、条约和会议文件，宣传多语制是欧洲必须培养和维护的核心价值观。然而，尽管在语言学习项目上投入了很多努力和资源，“母语加两门外语”的学习目标依然很难达到。一方面，现实中大多数欧盟公民的外语只学英语；另一方面，语言学习是个漫长的过程。因此，机器翻译似乎为那些没有时间或资源继续学习更多语言的人提供了一些支持。

此外，本章还以小型多语言社群——大学——为案例，讨论了大学校园如何制定语言政策推行多语制。语言政策可能会出于多种原因在校园内产生矛盾，但如今大多数校园实际上都是多语言的，或是通过大学的国际化/英语化，或是由于留学生的到来。在此背景下，本章探讨了设计语言政策的必要性，这些政策应当承认机器翻译推进多语制的潜力，同时不忘机器翻译带来的挑战，特别是译文质量和道德问题。在此，我想借用西班牙语当中的一个表达——“打开机器翻译的甜瓜”（*abrir el melón*），“打开甜瓜”的意思是解决一个迟早需要处理的问题，尽管没有人愿意这样做，因为后果未知。换句话说，没有人知道这个瓜够不够甜，但只有打开才能知道。即使现有的机器翻译系统无法表达出这个短语的隐喻意义，能看懂机器翻译字面意思的读者还是能学到一个有用的西班牙语隐喻。说不定这个比喻甚至可以传播到新的语言和文化中，因为它仅用三个词就能表达出复杂的含义。而这就是多语制的具体体现。

References

- Carré, Alice, Dorothy Kenny, Caroline Rossi, Pilar Sánchez-Gijón & Olga Torres-Hostench. 2025. 机器翻译之于语言学习者. In Dorothy Kenny (ed.), 机器翻译知识普及: 为人工智能时代的用户赋能, 147–163. Berlin: Language Science Press. DOI: 10.5281/zenodo.14922301.
- Cenoz, Jasone. 2013. Defining multilingualism. *Annual Review of Applied Linguistics* 33. 3–18. DOI: 10.1017/S026719051300007X.

- Council of Europe. 2020. *Languages Covered by the European Charter for Regional or Minority Languages*. <https://www.coe.int/en/web/european-charter-regional-or-minority-languages/languages-covered>.
- Council of the EU. 2008a. *Council conclusions of 22 May 2002 on multilingualism (2008/c140/10)*.
- Council of the EU. 2008b. Council resolution of 21 November 2008 on a European strategy for multilingualism. *OJ C 320* 16.12.2008. 1–3.
- Council of the EU. 2011. *Council conclusions on language competences to enhance mobility 2011*. <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2011:372:0027:0030:EN>.
- Council of the EU. 2014. *Conclusions on multilingualism and the development of language competences*. https://www.consilium.europa.eu/uedocs/cms_data/docs/pressdata/en/educ/142692.pdf.
- European Commission. 2007. *Final report from high level group on multilingualism*. europea.eu/en/publication-detail/-/publication/b0a1339f-f181-4de5-abd3-130180f177c7. ISBN: 978-92-79-06902-4 <https://op>.
- European Commission. 2008. *Multilingualism: An asset for Europe and a shared commitment*. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX>.
- European Commission. 2013. *European higher education in the world*. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. COM (2013) 499. <https://ec.europa.eu/transparency/regdoc/rep/1/2013/EN/1-2013-499-EN-F1-1.pdf>.
- European Commission. 2019. *Education and training monitor*. <https://op.europa.eu/webpub/eac/education-and-training-monitor-2019/en/>.
- European Commission. 2021. *Why is multilingualism important?* https://europea.eu/education/policies/multilingualism/about-multilingualism-policy_en.
- European Commission, Directorate-General for Translation. 2022. *Translation in figures 2022*. <https://data.europa.eu/doi/10.2782/253419>.
- European Council. 2002. *Presidency conclusions of the Barcelona European Council meeting 15 and 16 March 2002*. https://ec.europa.eu/commission/presscorner/detail/en/PRES_02_930.
- European Council. 2017. *European council conclusions on security and defence, social dimension, education and culture, and climate change*. 14 December. <https://>

- [//www.consilium.europa.eu/en/press/press-releases/2017/12/14/european-council-conclusions/](http://www.consilium.europa.eu/en/press/press-releases/2017/12/14/european-council-conclusions/).
- European Parliament. 2020. *Multilingualism in the European Parliament*. <https://www.europarl.europa.eu/about-parliament/en/organisation-and-rules/multilingualism>.
- European Parliament. 2021. *Rules of Procedure of the European Parliament*. https://www.europarl.europa.eu/doceo/document/RULES-9-2021-01-18-RULE-167_EN.html.
- Eurostat. 2022. *Pupils by education level and number of modern foreign languages studied - absolute numbers and % of pupils by number of languages studied*. Online data code: EDUC_UOE_LANG02. https://ec.europa.eu/eurostat/databrowser/view/educ_uae_lang02/default/table?lang=en.
- Eurydice. 2019. *The teaching of regional and minority languages in schools in Europe*. <https://op.europa.eu/s/n8TH>. Luxembourg: Publications Office of the European Union. DOI: 10.2797/472090.
- Extra, Guus & Kutlay Yagmur. 2005. Multilingual cities project: Crossnational perspectives on immigrant minority languages in Europe. *Noves SL. Revista de Sociolingüística* 1. 1–9.
- Gao, Catherine Yuan. 2019. *Measuring university internationalization*. Cham: Palgrave Macmillan. DOI: 10.1007/978-3-030-21465-4.
- Gordin, Michael. D. 2016. The Dostoevsky machine in Georgetown: Scientific translation in the Cold War. *Annals of Science* 73(2). 208–223.
- House, Julianne. 2003. English as a lingua franca: A threat to multilingualism? *Journal of Sociolinguistics* 7(4). 556–578.
- Knight, Jane. 1994. *Internationalization: Elements and checkpoints*. CBIE Research No. 7. Ottawa: Canadian Bureau for International Education (CBIE)/Bureau canadien de l'éducation internationale (BCEI).
- Mittelmeier, Jenna, Bart Rienties, Ashley Gunter & Parvati Raghuram. 2020. Conceptualizing internationalization at a distance: A “third category” of university internationalization. *Journal of Studies in International Education* 25(3). 266–282.
- Multilingual Higher Education. 2016. *Cooperation for innovation and exchange of good practices*. Erasmus+ Programme of the European Union. https://sisu.ut.ee/multilingual/book/6_2_Internationalization-domination-of-English.
- MultiTraiNMT. 2020. *Machine translation training for multilingual citizens*. Erasmus+ Project. www.multitrainmt.eu.
- Murai, Shusuke. 2015. *Translation tech gets Olympic push*. The Japan Times [Mar 31 2015]. <https://www.japantimes.co.jp/news/2015/03/31/reference/translation-tech-gets-olympic-push/>.

- 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.
- 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.
- Wikizero. 2020. ISO-639-3. https://www.wikizero.com/en/ISO_639-3.
- World Bank. 2020. *Total population –European union*. <https://data.worldbank.org/indicator/SP.POP.TOTL?locations=EU..>
- Zakharova, Ulyana S. 2019. Online course production and university internationalization: Correlation analysis. In Mauro Calise, Carlos Delgado Kloos, Justin Reich, Jose A. Ruiperez-Valiente & Martin Wirsing (eds.), *Digital education at the MOOC crossroads. Where the interests of academia and business converge*, 102–107. Cham: Springer.

第二章

人工翻译和机器翻译

Dorothy Kenny
都柏林城市大学

本章旨在向读者介绍何为人工翻译和机器翻译。作者不仅试图拨开翻译的迷雾，还强调译者在创造原文与译文之间对等关系中发挥的重要作用。本章的根本目标是帮助读者理解人工译文如何成为机器翻译的训练数据。此外，本章还从系统类型和输出用途两个方面区分机器翻译。作者还特别指出，当代机器翻译是人工智能的分支——机器学习，更确切地说，是深度学习的应用。

1 何为翻译？

本书以机器翻译为题，简要定义为由计算机程序完成的翻译。但这一定义仍然无法回答“何为翻译”。其实，已有大量翻译学学者对此进行深入研究，并探究翻译在文化、科学和政治等领域中的作用。翻译学研究内容丰富，无法在此详述，欲深入了解的读者可参考Baker & Saldanha (2020)。本章采用的是大多数评论家都认可的定义，即翻译是基于一种语言（即源语）的文本而产生的另一种语言（即目标语）的文本。“文本”的概念很重要，指的是真实的口头或书面语言的使用。一般来说，文本应符合以下标准：连贯一致且正确“组合”；用途明确，即便只是一句用于问候的“你好”。此外，文本还应符合特定语言和交际场景的需求。例如，本章作为一本供教学使用的英语论文集章节，应该满足读者的相关期待。本章内容围绕特定主题或领域（即机器翻译）展开，并遵照特定体裁（即教材）惯例撰写。

在翻译界人士看来，“翻译关乎文本”这一观点已是老生常谈，无需多言。但对于不太了解翻译的人群，我们想提请注意，翻译的对象是文本，而非语言。语言是个庞大且复杂的抽象系统，在人类交流和表达中的使用有无限可能；文本则是语言使用的具体实例，通常有可识别的起止点。即便脱离语境而孤立使用的语言看似潜力无限，会出现无法预知的含义或较大歧义，但这种潜力在特定文本中会大大减弱。例如，英语中的 *shower* 一词有以下



含义: (1) 阵雨; (2) 淋浴设备; (3) 送礼会, 因此在法语等语言中会有不同译法。但如果我们正在为浴室配件制造商翻译淋浴器安装指南, 就不用担心一词多义的问题。这种情况下, 该词应取第二种含义, 除非作者在玩文字游戏(但不可能出现在这种体裁的文本中)。关注文本而非语言, 可以让翻译更贴近实际情况且易于处理。

上述翻译定义的第二个要素是, 翻译的文本产出基于另一个已存在的文本。这说明, 翻译涉及源语文本和目的语文本之间的关系¹ 还有些学者就该问题展开了更深入的论述, 认为这种关系还意味着具有“相同意义”(same meaning), 但许多哲学家和语言学家对“意义”的理解诚然更复杂深入, 他们避免称翻译具有“相同意义”。这是因为, 文本的意义难以与创造和使用的情境剥离开来。我们可以认为意义就是作者或说话人想表达的意思, 但我们常常无法明确他们的真正意图; 也可以将意义与自己对文本的阐释联系起来, 但不得不承认, 对同一文本, 其他人的阐释会有所不同。翻译还会产生的问题是, 一个完全合理的译文, 其意义可能比源语文本更多或更少, 而这只是目的语的要求使然。

下面来看个例子。最近出版的一本回忆录 (Tammet 2006) 的开篇如例 (1) 所示:

- (1) I was born on 31 January 1979 –a Wednesday. (我出生于 1979 年 1 月 31 日, 一个星期三。)

其法语译文 (Tammet 2009) 见例 (2):

- (2) Je suis né le 31 janvier 1979. Un Mercredi.

尽管这两句话几乎词词对应, 但实际上, 法语译文传达的信息要比英语原文多。从法语译文可以看出, 原文作者 *I* 是男性, 因为如果是女性, 那么例 (2) 中的正确用词应该是 *née*, 而不是 *né*。法语中一些动词在特定时的变位形式必须标注主语的性别。

可是, 法语译者又从何得知, 原文中的“我”是男性呢? 毕竟, 这是该书开篇的第一句话。不过, 这是本回忆录, 根据此类体裁的惯例, 叙述者即作者, 而译者也知道该书作者是谁, 这在翻译合同和书的封面上都写着。因此, 尽管英语原文有时并未说明某人的性别, 译文还是得明确指出性别信息, 不过, 这对译者而言并非难事。虽算不上是问题, 但这个简单的例子还是反映了两个重点: 第一, 如上文所述, 译文的信息有时比原文更多; 第二, 有时无法从原文语句中获取翻译该句子所需的信息。相反, 译者应该关注 (1) 范围更大

¹显然, 该定义的第三个要素是, 源语文本和目的语文本为两种不同的语言。因此, 我们关注的是语际翻译 (interlingual translation)。有些学者, 最著名的是雅柯布逊 (Jakobson 1959), 认为还存在其他类型的翻译, 如语内翻译 (intralingual translation) 和符际翻译 (intersemiotic), 但这些翻译类型不在本章的讨论范围内。

的文本,有时也称为共现语篇 (*co-text*),即与给定语段相关的文本,例如封面;(2) 语境 (*context*) (此处理解为与文本相关的更大范围的语境),以找到解决办法。

还有些时候,译文包含的信息比原文的多,其原因不在于目的语,而在于体裁。在一项针对计算机辅助设计工具的用户界面研究中,Moorkens (2012)发现,英语标题 *Selecting* (选择) 的译文通常明确指出所选择的对象,因此该词有不同译法。以下例子为译文到英语的回译:

- (3) selection of polygon (选择多边形)
- (4) selection of line (选择线条)
- (5) selection of ellipse (选择椭圆形)
- (6) selection of rectangle (选择矩形)

等等。

这种译文表达信息比原文多的情况并不少见。当然,反之亦然,若译文无法表达原文所有信息,译者可以选择省略部分信息。这种情况的出现有时与语言类型相关。例如,英语经常使用动词来描述某物或某人移动的方式 (*manner*),而西班牙语则倾向于用动词来描述行动的路径 (*path*),使用状语短语句来表示移动方式,但西语译者有时选择不翻译这种移动方式,以免刻意且生硬的表达。具体可参考Slobin (2003)给出的例句 (7) 和 (8)。英语动词 *stomped* 意指“迈着沉重步子”的走路方式,而西语动词 *salió* 只表示该人物离开房子的这个动作。

- (7) He stomped from the trim house (他踩着重步走出整洁的房子)
- (8) Salió de la pulcra casa (走出整洁的房子)
'exited from the trim house.'

例 (8) 的西语句子比例 (7) 的英语句子还少了一处信息,即前者没有使用与英语单词 *he* 对应的主语代词。这是因为西班牙语属于代词脱落语言 (*pro-drop language*),可省略主语代词,而这些代词所含的大部分信息都可根据动词词尾获知。如例 (8) 的 *salió* 为动词过去式的第三人称单数,与例 (7) 相比,例 (8) 缺少了主语性别。但西班牙语读者可以从前文获知主语的性别信息 (男性),而不是对此浑然不知。所以,省略代词的译法既遵照了目的语语言规范,又没有造成小说读者的阅读障碍。

上述论点和例句旨在说明,为什么很多学者不赞同译文与原文的意思相同。我们更认同的说法是,译文接近原文。出于各种各样的原因,译者必须决定翻译时优先考虑什么,需要表达什么,以及应该让读者自己去弄清楚什

么。²译者帮助译文读者自行构建的意义,可能在很大程度上与源语文本的读者构建的意义相一致,但在许多情况下,二者并不完全相同。一般来说,这不是什么问题。

可是,如果源语文本和目的语文本之间(或更可能只是文本中的语段)并不是“意思相同”的关系,那又是什么呢?有人称之为“对等(equivalence)”关系。翻译学对“对等”这一术语的研究历程堪称跌宕起伏。但如果它被认为是译者决策使然的关系(即两个文本语段之间的关系),因为译者认为两个文本在各自的共同文本和语境中具有同等价值,那么完全可以用“对等”来描述原文和译文的关系。也就是说,例(7)的“salio”与例(8)的“he stomped”对等。显然,这种对等关系不会一成不变,也不可能用于“stomped”一词的所有其他语境,但如果我们认为在该语境下,“salio”与“he stomped”对等,这点便无关紧要。³

2 翻译难在何处?

关于机器翻译的书通常以讨论翻译的难点开篇,聚焦于上文提及的单语歧义和语言系统的差异。语际差异还可见于在两种语言的对等句子中,句义在词中分布不同。如摘自欧洲议会会议记录的例句(9)和(10),英语用动词“like”来表达“喜欢”,而喜欢的对象则使用“working with you”作为动词“like”的补语来表述。但德语用动词“kooperiere”来表达喜欢的对象,而“喜欢”则用副词“gern”来表达。

(9) I like working with you. (我喜欢跟你合作。)

(10) Ich kooperiere gern mit Ihnen. (我高兴地跟你合作。)

‘I cooperate happily with you.

这类例子展示了英语和德语表达心理状态的不同方式,但该差异并非翻译难点,至少对人类来说是如此。具备初级英语水平的德语使用者可以毫不费力地翻译出例句(9)。语言之间的非同构性(Non-isomorphism),即语言的结构方式不同,本身不会给译者带来翻译困难。

另一种很棘手的语言现象是不连续的依存关系(discontinuous dependencies),即由两个单词组成的短语被一个或多个单词隔开。例如,例(11)中“Send”和“back”相当于“return”单个词项的意义。不过,掌握基础英语知识的读者通常能够正确理解。

²我们所示的例句主要源于语言系统之间的差异,但译者可能考虑到文化因素而省略或更改原文的细节,以免读者看到不熟悉的引用感到困惑,甚至是为了避免冒犯读者或审查人员。译者还可能受到篇幅限制,如字幕翻译。

³翻译学者 Anthony Pym 在其研究中提出了基于交换价值的对等概念,可参见 Pym (2010)。

- (11) Send your certificate of motor insurance back. (把你的汽车保险凭证寄回来。)⁴

另一个常见的翻译难题是习语。在本文中,“习语”是指无法根据其构成词语来推断意思的短语。换句话说,习语具有非组合特性(non-compositional),如本章1 第一节中的例子“old hat”,意指所论之事为人熟知,已是老生常谈。这个短语与帽子毫不相干。和其他类型的修辞语一样,习语有时会给不了解该表达的读者造成困惑,不能望文生义。但是,即使阅读本章第一次看到“old hat”时并不理解其意思,你或许还是能推断出,讨论内容与帽子无关,不应按字面意思来理解这个短语。如果遇到不懂的习语,译者认为它应该是收录在词典中的常用表达,只需查一查在线词典即可。换句话说,虽然习语具有非组合特性和修辞作用,但仍属于约定俗成的表达。尽管译者遇到翻译难题,也就是说其翻译节奏被打断,但问题并不难解决。而找到解决方法,很可能会给译者带来极大的满足感(和所有语言学家一样,译者也乐于学到工作语言的新知识。)

但有时,就连经验丰富的译者也坦言会在翻译技术性强且了解不足的文本时犯难。法律专业出身或从事法律工作的译者可能喜欢法律翻译,而汽车工程师对此则望而却步。此外,即便是自己熟悉的领域,译者也会遇到质量低下、不完整或极难理解的原文。他们或许能够理解原文,但要在目的语中找到恰当的术语来表示原文的专业概念,这却是个很大的挑战。交稿日期太紧迫或软件程序故障,也都是可能让专业译者犯难的因素。不过,我们很少听到专业译者抱怨歧义、非同构性、不连续依存现象或非组合性的语言现象对他们造成困扰。

当然,在讨论机器翻译时,这些现象之所以被常提起,是因为它们在某些情况下会给机器带来问题。

3 译者通常如何解决翻译问题?

前文讨论了专业译者有时会在现实场景中遇到的问题。现在,我们一起来看看译者在翻译过程中还会有哪些困难。在无法理解原文,或者不知道如何用目的语表达某个术语,又或者难以用目的语来阐述某个意思的时候,专业译者通常先将手头的文本放一边,转而查看资料。例如,译者想要了解废水处理的细节时,可能会浏览各个地方政府网站,看看他们如何解释相关技术,可能在开放术语库找到特定术语的译法,还可能会查阅客户公司的其他文件或咨询工程师。译者还可以请教同事,或者在翻译论坛求助。最重要的是,大多数译者都意识到自己在专业知识上有所欠缺或需要灵感,从而认真研习以查漏补缺,努力解决问题,继续完成翻译任务。

⁴在本例句中,动词短语“send back”中间有五个单词,是英国国家语料库(British National Corpus)中最长的不连续短语动词例子之一。

你可能会问，为什么一本介绍机器翻译的书，却着力叙述人工翻译过程。原因有二：首先，真实的人工翻译过程（详见Rossi & Carré 2025 [本卷]）可为机器翻译设定评估标准，一切有助于提升人工翻译质量的方法也能用于改善机器翻译质量。同样，人工翻译过程可以突显机器翻译偶尔出现的缺陷。换句话说，人工翻译有助于机器翻译的质量评估和错误诊断。其次，也是更重要的一点，目前大多数机器翻译系统都以人工翻译为基础学习如何翻译。下文将对此进行详述。

在结束对人工翻译过程的讨论前，我们还要介绍一项对许多译者来说不可或缺的技术——翻译记忆库。

4 翻译记忆库

20 世纪 90 年代，蓬勃发展的软件本地化行业的翻译人员发现，他们翻译的文本要么自身重复率高，要么有大段内容与同一文档的早期版本相同，如软件使用手册，只要软件推出新版本，手册也必须随之更新。译者并不会像翻译全新文本那样，从头开始翻译每个句子，而是开发出被称为“翻译记忆”的工具，用于存储翻译过的文本，以便重复使用。该工具被称为“翻译记忆库”，将新导入的原文分成句段（句子或其他有意义的翻译单元，如标题或表格的单元格），然后将这些句段与记忆库中的源语言句段进行比较。如果找到完全匹配或非常相似的句段，那么相应的目的语句段会被调出供重复使用，译者可以对其进行或不做修改。译者在翻译新文本时，可以使用翻译记忆库，接受、拒绝或编辑现有译文，并在翻译过程中更新记忆库，为没有匹配结果的源语句段添加自己的译文。随着时间的推移，翻译记忆库变得越来越大。有些较早开始使用该技术的公司创建的翻译记忆库含有数十万甚至数百万“翻译单元”（translation units），即源语和目的语对齐的平行句段。例 (12) 为某个标题的翻译单元（英语-德语），摘自由欧洲议会官网数据组成的翻译记忆库，格式为 tmx（translation memory exchange，翻译记忆库交换）。<tu> 和 </tu> 元素表示翻译单元的起始，<tuv> 和 </tuv> 元素表示翻译单元变体（variant），⁵<seg> 和 </seg> 元素则表示相应语言的句段或文本字符串。

(12) <tu>
 <tuv xml:lang="EN">
 <seg>A common blacklist for unsafe airlines</seg>
 </tuv>
 <tuv xml:lang="DE">
 <seg>Unsichere Luftfahrtunternehmen kommen auf eine schwarze
 Liste</seg>

⁵本例中的第一个变体是英语（“EN”），第二个是德语（“DE”）。

```
</tuv>
</tu>
```

私营翻译企业也积累了大规模的翻译记忆库，这些宝贵的语言资产有利于控制翻译成本和提高竞争力。欧盟机构等国际组织也采用这项技术，创建了庞大的多语言翻译记忆库，并免费提供给计算机科学家，用于自然语言处理研究。

虽然翻译记忆库最初的作用之一是提高人工翻译的工作效率，但最终也有助于提高翻译行业的自动化。一方面，翻译记忆库可产出海量翻译数据，且数据格式便于机器翻译研发使用（见下文），另一方面，用于管理记忆库的工具还提供编辑环境，译者不仅可以编辑记忆库的人工翻译，还可以编辑机器译文。

翻译记忆库可以算是特殊的平行语料库，即句对齐的原文和译文构成的语料集合。即便在翻译过程中并未使用翻译记忆库，也可以在翻译之后再再将译文与原文对齐。例如，使用多种语言的欧洲议会的会议记录译文可从网站上获取，对齐后创建多语言 Europarl 语料库 (Koehn 2005)，这继而又极大推动了机器翻译研究。对齐的平行语料库不必是 tmx 格式，而是通常采用具有数千（甚至数百万）行的文档形式，每行一个句段。源语句段在源语文档中的位置，与其译文在目的语文档中的位置一一对应，因此，目的语文档的第 x 行为源语文件第 x 行的翻译。

5 何为机器翻译？

根据本章开篇给出的定义，机器翻译是指在源语文本的基础上，自动生成目的语文本的过程。与其他类型的翻译一样，我们也期待该目的语文本的阐释与源语文本的大致相符。不得不承认的是，人工翻译可能会与原文意思有些许差异，所以我们也应该允许机器翻译出现类似情况。重要的是，源语文本和目的语文本之间的明显差异，例如，日语比英语提供的信息更多，应该是由语言对、体裁或其他合理的原因引起的。

机器翻译是出现在二战后的、数字计算机的首批非数字应用之一。以今天的标准来看，早期的自动化翻译似乎很简陋，但必须承认的是，在 20 世纪 50 年代和 60 年代，研究者能使用的资源极其有限 (Hutchins 2000)。到了 20 世纪 60 年代末和 70 年代，自动翻译系统主要用于国防、政府和国际组织，直至 20 世纪末才开始大规模商用。1997 年，美国搜索引擎 AltaVista 推出免费在线机器翻译工具“宝贝鱼” (Babel Fish)，当时的用户达数百万人。随后的几十年里，互联网发展迅速，目前用户约有 46.6 亿人 (Johnson 2021)。据报道，截至 2016，最著名的免费在线机器翻译系统“谷歌翻译”的用户超过 5 亿人，

每天翻译超过 1000 亿字, 支持 103 种语言 (Turovsky 2016)。⁶例如, 与谷歌搜索或微软必应等搜索引擎相结合, 机器翻译可用来扩大搜索范围, 然后将相关的外语网页译成用户的母语。

但机器翻译并不仅用于网页, 还可以与自动语音识别和语音合成, 或光学字符识别和数字图像处理等技术结合使用, 让用户能使用两种以上的语言进行口语对话, 或看懂不熟悉的语言体系的路标。通常, 用户只需使用安装在手机里的翻译应用软件即可, 这些软件如今甚至能离线使用, 可以说是名副其实的口袋式机器翻译系统。而机器翻译也越来越多地用于以前被认为超出其技术能力范围的领域, 如视听翻译, 将外语电影和电视剧的字幕翻译成其他市场受众的语言。事实上, 订阅视频流媒体服务的蓬勃发展是基于一种模式, 这种模式将鲜为人知的“长尾”标题展示给新观众, 而这些标题之所以不那么为人熟知, 是因为最初用外语书写。因此, 视听内容正成为众多商业产品中的最新产品, 可以通过机器翻译来扩大其市场。自诞生以来的大约 70 年里, 机器翻译已经从政府和国际组织的专享技术, 变成了大众消费品。

尽管机器翻译在上述情况下的作用毋庸置疑, 并且在人道主义领域 (Nurminen & Koponen 2020) 等场景中也有用武之地, 但仍存在一些不足。首先, 和人工翻译一样, 机器翻译也会出错。这些错误可能无伤大雅, 也可能后果严重 (如医疗保健、新闻翻译或国际外交等领域)。因此, 大量研究致力于“评估”机器翻译系统的质量, “测评”机器译文, 通过“译后编辑”来提高译文质量, 或在进行机器翻译之前预先对原文进行加工, 使其更易翻译, 从而改善译文质量, 这一过程称为“译前编辑”。这些研究领域将在本书的第 3 章到第 5 章中详细讨论。机器翻译也引出了大量的道德和法律问题, Moorkens (2025 [本卷]) 讨论了伦理问题, Carré et al. (2025 [本卷]) 也略微涉及机器翻译对语言学习者的影响。

许多临时用户可能认为, 这些方面的考虑对自己的机器翻译需求来说并非必要的。这在某些使用场景下确实如此, 例如只想借助机器翻译理解文本主旨或了解网页的基本内容。在机器翻译的范畴内, 此类用途通常称为“同化” (assimilation), 一般涉及低风险且供个人使用的翻译文本, 几乎不存在名誉受损等风险。但是, 如果想借助机器翻译达到“传播” (dissemination) 目的, 例如用第二语言来发布博客、进行商业推广, 那么明智的做法是了解潜在风险, 甚至有必要采取降低风险的措施。这种能力是“机器翻译素养 (machine translation literacy)”的要素之一 (Bowker & Ciro 2019)。其他要素还包括基本了解机器翻译的运行原理及其使用对社会、经济和环境具有哪些更广泛的影响。这些知识看似深奥, 但实际上是可迁移的技能, 因为现代机器翻译与许多其他技术都基于相同的原则, 这些技术深刻改变了现代生活的很多方面, 尤其是我们的工作方式。简而言之, 目前的机器翻译在很大程

⁶截至 2022 年 5 月, 谷歌翻译支持 133 种语言, 但不同语言的支持深度和翻译质量有所不同。(Caswell 2022)。

度上是机器学习的应用,更具体地说是深度学习的应用。我们会在下文详细解释这些概念,Pérez-Ortiz et al. (2025 [本卷])也会深入介绍神经机器翻译的运行原理。如果你是翻译专业的学生、职业译者或在翻译行业中担任其他职务,那么你可能迫不及待想要打开机器翻译系统的“黑匣子”。或想了解如何通过定制,以最大化地让机器翻译系统为己所用。这些方面的讨论参见Ramírez-Sánchez (2025 [本卷])。接下来,我们一起了解机器翻译为何被称作“引领语言学家进入机器学习这一奇妙世界的‘大门’”。

6 人工智能、机器学习与机器翻译

如今的机器翻译经常伴随很多其他相关概念出现,如人工智能、机器学习、人工神经网络和深度学习,非专业人士很难区分其中某些概念。Goodfellow et al. 2016等借助维恩图来展示这些概念之间的关系。在上述概念中,人工智能包含的范畴最广,用最大的圆圈表示,通常被定义为计算机科学的分支,旨在创造机器,更具体地说是计算机程序,来解决通常需要人类智能的问题。这种机器不一定要像人类一样“思考”,但要像人类一样“行动”。它们可能只专注于特定问题,比如识别人脸,被称为“狭义人工智能”,或说得不好听,就是“弱人工智能”。而所谓的“强人工智能”则让人更加充满期待。它涉及“通用人工智能”(即机器会拥有类人智能、自我意识以及学习和规划未来的能力),或“超级智能”(即超过人类能力的智能)。可以说,产出比肩人类译者的翻译需要依赖强人工智能,但机器翻译系统目前并不具备这种智能。

6.1 基于规则的机器翻译

应对人工智能开发的挑战的一种方法是,给计算机程序输入解决特定问题所需的一切知识,并运用这些知识的规则。以机器翻译为例,给程序输入源语和目的语的所有单词列表,以及连词成句的规则;然后,规定两种语言的单词和结构如何相互匹配,并向机器提供如何使用这些信息来翻译句子的逐步指令(即“算法”)。这种方法被称为“基于规则的机器翻译”(rule-based machine translation, RBMT),在本世纪初之前一直占据主导地位,如1997年首次推出的免费在线机器翻译正是基于RBMT(Joscelyne 1998)。然而,RBMT存在很多问题,如需要高水平的语言学家为每个语言对编写规则,因此开发成本昂贵。此外,和其他基于知识的人工智能方法一样(Goodfellow et al. 2016),RBMT也受到知识瓶颈的制约:在许多情况下,根本无法预测使RBMT系统如愿运行所需的一切知识,无论是语言知识,还是与更广大世界有关的知识,即所谓的“现实世界知识”。⁷

⁷在撰写本书时,尽管RBMT已经不再占据主导地位,但仍用于少数机翻系统,尤其针对密切相关的语言之间的翻译,例子可参见Apertium (Forcada et al. 2011)。

6.2 数据驱动的机器翻译

此类机器翻译运用到了机器学习技术。机器学习的基本原理是，与其向计算机程序输入所需知识，不如让其自行获取。要做到这点，机器可以通过观察待解决问题以往的解决方案。我们已经了解到，如何借助翻译记忆库和其他平行语料库来存储翻译单元的语段，从而实现待译文本的已存译文抓取。这些翻译单元组成了“训练数据”，供现代机器翻译系统学习。因此，这类通常被归为“数据驱动型”系统。“从数据中学习”正是机器学习与其他类型人工智能的不同之处。

数据驱动的机器翻译还可进一步划分为“统计机器翻译”(statistical machine translation)和“神经机器翻译”(neural machine translation)，对此下文均有详细介绍。

6.3 统计机器翻译

统计机器翻译(SMT)系统依靠训练数据可构建两大类统计模型。⁸第一类是双语“翻译模型”，即训练数据的源语单词和短语及其目的语译文一一对照形成表格，以及每种译法的概率。这种表格叫做“短语表”。表1为短语表的摘选例子。⁹

表 1: 选自短语表，展示了短语 *a me piace* 在 Europarl 语料库中的不同译法及其出现概率

	英语	概率
a me piace	I like	0.78
a me piace	I should like to	0.11
a me piace	I admire	0.11

然而，在这里使用“短语”一词有些不恰当，因为所讨论的字符串不一定对应于语言学领域的“短语”，而是 n 元组(gram)，即在训练数据中连续出现的一个、两个、三个或 n 个单词的字符串。例如，在前面的句子中，“appear contiguously”是二元组，而“appear contiguously in”是三元组。

SMT 需要从数据中学习的第二类统计模型是“语言模型”，即目的语的单语模型(或模型组合)，同样基于 n 元组。假设已知前两个单词，那么三元组的目的语语言模型可以算出第三个特定词出现的概率。例如，基于 Europarl

⁸统计模型是被观察数据的数学表征。

⁹这个例子被大大简化了，因为它只展示了从意大利语到英语的合理译法。实际上，SMT 系统会学习包含许多无意义配对的翻译模型，不过其中大多数的概率很低。此外，还要给从未见过的译法保留一些概率。

语料库训练的三元组模型，可以算出“gorgonzola”出现在“I like”之后的概率为 0.024，这说明尽管“I like gorgonzola”确实出现在训练数据中（共 4 次），但除“gorgonzola”之外的许多词出现在“I like”之后的概率更大。¹⁰

在 SMT 系统中，翻译模型主要捕捉有关如何将单词和 n 元组译成目的语的知识，而语言模型则首先告诉你哪些是目的语的常见用词或短语。从目前的角度来看，这种方法的核心在于语言学家不必自行构建这些模型。模型由机器在训练阶段直接从数据中学习而得。到了第二阶段，即“调优”阶段，系统开发人员计算出应该分配给每个模型的权重，以获得最佳输出。一旦训练和调优完毕，系统就可以翻译训练数据以外的源语语句。在 SMT 系统中，翻译（相对于“训练”）被称为“解码”，在这一过程中，系统会为输入的句子生成数千个候选译文，并基于某个源语语句、系统已学习的模型以及系统分配权重的情况下，计算哪个译文出现的概率最大。

截至 2015 年，SMT 作为最先进的机器翻译技术的地位至少保持了十年之久。与之前的 RBMT 系统相比，SMT 取得了技术上的飞跃，但也存在许多缺陷，主要原因是使用了相对较短的 n 元组来建模，且系统将一个句子拆成孤立的 n 元组后分开翻译。SMT 尤其不擅长处理黏着语和高度屈折性语言。其他不足还包括单词丢失（word drop，即某种语言体系无法翻译某个单词）、译法不一致（即同一个源语单词在译文中对应两种译法，有时甚至在同一句子中出现不同译法）。到了 2015 年，SMT 已经被同为数据驱动的另一方法所取代，也就是上述的“神经机器翻译”，这一技术转向的完成只花了两年时间。

你可能觉得奇怪，既然 SMT 已经过时了，为什么还要在此提及呢？其实，我们是想借此将读者引入机器学习领域。SMT 表明了，从数据中学习的机器翻译系统比其他系统的性能更好，从而为机器学习方法在机器翻译中的运用铺平了道路。另外，SMT 开发者为机器翻译研究做出了重要贡献，比如通过提倡新方法和共享程序代码，以及从双语和多语的议会、国际组织、万维网等渠道收集翻译数据，并与全球的研究者共享数据，为机器翻译研究做出了显著贡献。还应注意的是，尽管范围有限，但 SMT 在翻译行业中仍有运用。例如，机器翻译服务供应商可能会先创建 SMT 系统来检验项目的可行性，再看看是否值得继续投入时间和精力开发神经网络。

然而，我们之所以讨论 SMT，主要是想说明从数据中学习的方法不止一种¹¹，更重要的是，就本章的目的来说，这些数据不止有一种表征方法。我们已在前文看到，SMT 借助短语表来表示翻译知识，还借助单独的 n 元组模型来表示目的语知识。在这些模型中，单词（和单词串）仍保留自身形态，但最重要的是，它们通过概率相互关联。正是这些概率让机器翻译系统得以运

¹⁰ 此处和表 1 中使用的 Europarl 版本可通过点击 Sketch Engine 的界面链接 sketchengine.eu 访问。

¹¹ 其实，我们几乎没有讨论 SMT 系统的具体学习算法。感兴趣的读者可参阅 Koehn (2010)。

行。某个目的语句子是某个源语句子的译文的概率，只需在短语表中找到组成该句子的各个 n 元组的翻译概率，然后将概率相乘即可。此外，某个目的语句子出现的概率也可以通过计算组成该句子的各个 n 元组的概率乘积得到，如从单语语言模型找出每个 n 元组出现的概率并计算乘积。然后，用一个方程将不同的模型组合在一起，以计算出概率最大的译法。¹²

本文介绍 SMT 的另一个原因是方便引出一些概念，如“ n 元组”。如 Rossi & Carré (2025 [本卷]) 所述，“ n 元组”这一概念对自然语言处理的其他领域，特别是机器翻译测评来说极其重要。

6.4 神经机器翻译

2004 年至 2014 年是 SMT 发展的全盛时期。机器翻译的大多数主要用户和供应商，包括谷歌翻译（始于 2007 年）和欧盟委员会（始于 2010 年），在此期间都使用这项技术，并且在“共享任务评估”（shared task evaluations）¹³ 中，SMT 不断取得喜人的成果。直到 2015 年，斯坦福大学开发了神经机器翻译（neural machine translation, NMT），不仅以明显优势击败了许多 SMT 系统，还能应用于难以处理的语言对，如英语-德语（Bentivogli et al. 2016）。斯坦福大学的成功宣告 Bentivogli et al. (2016) 所谓“NMT 新时代”的到来。研究人员和媒体对此的兴奋之情溢于言表。例如，媒体大肆宣扬这项新技术，称其可与职业译者相提并论，已经达到了“与人类同等的水平”。¹⁴ 还有人声称，NMT 不仅可以学习“惯用表达和隐喻”，还可以“在另一种语言中找到文化对等表达，而不是直译”（Marking 2016）。¹⁵ 尽管这么说确有几分道理，但不应被过度解读。NMT 系统的确可以翻译惯用表达，但这通常是因为它的学习数据中含有数百甚至数千个该翻译的例子。NMT 系统（本例使用了谷歌翻译）在正确翻译德语的习语时，并不“知道”这是惯用表达，也不知道自己使用了文化对等表达：

(13) Ich habe die Nase voll.

‘I have the nose full.’ (我鼻子塞满了。)

as

¹² 此处所指的方程基于贝叶斯定理，SMT 有助于翻译学者初步了解“贝叶斯优化”的机器学习方法。

¹³ 在共享任务评估中，计算机科学家选定某个语言对并使用不同类型的训练数据，从而对比看哪个系统的表现最好。

¹⁴ 这一说法由一直致力于中文-英语语言对的微软研究人员提出（Marking 2016）。包括 Toral et al. (2018) 在内的许多评论家都对此提出了质疑。

¹⁵ 这些评论是脸书的艾伦·帕克（Alan Packer）在 2016 年发表的（Marking 2016）。

(14) I'm sick of it. (我受够了。)

该系统不过是在输出从数据中学到的表达。¹⁶

但如果只是简单从数据中学习, 为什么 NMT 要比 SMT 好得多呢? SMT 不也是学习数据吗? 原因在于 NMT 系统使用的表示方式及其学习模型。

6.4.1 神经机器翻译的模型

先从模型说起。计算机“模型”是一些现实生活的事件、系统或现象的抽象数学表示。这类模型的用途之一是为新问题预测答案。例如, 用于翻译的计算模型应能将新的源语语句翻译成目的语。¹⁷

我们已经了解到, SMT 系统使用翻译和目的语的概率模型, 这些模型以短语表和 n -gram 概率来表示。相比之下, NMT 系统使用的模型受到人脑启发, 尽管神经网络只是基于对人类神经网络的简化设计。这些模型使用人工神经网络, 其中成千上万的独立单元(即“人工神经元”, 以下简称“神经元”)相互连接。在神经网络中, 来自其他神经元的刺激以及神经元相互连接的强度(或称“权重”)决定每个神经元的激活状态。正如Forcada (2017)所说, 单个神经元的激活状态本身意义不大。表示单个单词及其与其他单词之间关系的, 是大量相互连接的神经元的激活状态。训练 NMT 系统的关键在于, 精确学习会生成最佳性能翻译模型(即激活状态能预测最佳译文的模型)的权重。

如何才能做到这一点呢? 和所有机器学习一样, NMT 系统也要从数据中学习。神经网络翻译模型是通过让学习算法处理大量平行数据, 逐步建立而成。在不断处理数据的过程中, 算法会学习并不断调整权重, 使模型的预测越来越接近预期的“正确”答案。算法学习的详细过程可参阅Pérez-Ortiz et al. (2025 [本卷]), 而更全面的 NMT 技术讨论可参考Koehn (2020)。阅读本章只需了解, 数据驱动的机器翻译是典型的机器学习, 因为它涉及到为解决人类已知答案的问题而开发的技术, 且人类其实已经给出了至少一个正确答案。正确答案可出现在训练数据中, 或基于训练数据进行推断而得出。要检验某个机器翻译系统在训练期间的翻译质量是否有所改善, 或在训练结束后将其与另一系统进行比较时, 我们也可以通过让系统解决已经知道答案的问题来测试。比如, 我们通常会要求系统输出几个新句子的翻译, 但这些句子都已经有专门用来进行对比的优质人工译文。

NMT 系统被训练到令我们满意的状态时, 便可用于实际的翻译任务。这时谈论的不再是“测试”, 而是“使用”系统。大多数人都认为, 实际使用中的 NMT 系统是在进行“翻译”。与 SMT 系统一样, 计算机科学家也使用术语“解码”来描述 NMT 系统产生目的语的输出过程。

¹⁶这一翻译示例也已由谷歌翻译的用户社区验证。

¹⁷提及数学模型, 常会说到“预测”答案。但就本章目的而言, “预测”(predicting)和“输出”(outputting)答案并无实际区别, 都是机器翻译系统的实际运用状态。

6.4.2 神经机器翻译的词表示

我们已经说过，数学模型是一些系统、事件或现象的表示。关于数学和其他科学模型的地位存在很多争论，但不在在此展开叙述。我们认为，模型所表示的内容很复杂，包括很多相互联系的部分。如果要讨论更简单或颗粒度更小的实体，如数字 5 或“苹果”，只需用泛指术语“表示”(representation) 来指代描述该实体的方式。

表示方式很重要，因为正如 Goodfellow et al. (2016) 所述，在计算机科学和日常生活中，概念的表示方式会影响我们对其的处理方式。阿拉伯数字和罗马数字之间的差异对人类数学运算能力的影响，就是一个很好的例子。大多数人会发现计算 125 除以 5，比 CXXV 除以 V 容易得多，即使 CXXV 和 125 (以及 V 和 5) 所表示的数值相等。

词语也可视为概念的表示。例如，可以用“苹果”这个词来表示某类水果，也可以用一幅画来表示。当然，词语和图片的属性不同，也就意味着用处不同。例如，我们可以对词语进行拼写检查，但图片则不行。

NMT 系统还使用了另一种表示方式——向量 (vector)，即固定大小的数字列表。例如，“苹果”一词可用向量 [1.20, 2.80, 6.10] 来表示。对许多人来说，这似乎不可思议。很难理解如何用数字列表来表示一个单词。¹⁸如果说向量能很好地表示单词之间的关系，事情就开始变得合理。例如，单词“梨”可用向量 [1.20, 2.80, 5.50] 来表示，该向量与表示“苹果”的向量只有最后一个数字之差。若把向量中每个数字看作想象的三维空间中的一个维度，那么“苹果”和“梨”这两个词的位置则非常接近。而且还可以推测出，它们与不太相关的词语也离得较远，如“直升机”或“非常”。向量还具备其他有意思的属性，引起计算机科学家的极大兴趣，例如多个向量可以相加或相乘。但表示“苹果”和“梨”的词语或图像都不能进行这些运算！

那么，在上述例子中，“苹果”和“梨”的向量为何如此相似呢？其实，我们只是瞎编的。在真实的 NMT 系统中，我们会让计算机程序直接从语料库中“学习”所有单词的所有实例的合适向量。（记住，在机器学习中，无论是否有人工监督，计算机程序必须自行解决问题。）供机器学习的基于向量的单词表示称为“词嵌入”(word embeddings)。相关单词的词嵌入相似，是因为它们都根据特定单词在训练数据中的位置而定。若两个单词总是出现在相同或相似的共同文本中，如“苹果”和“梨”都很有规律地出现在单词“树”之前，且在“去皮”、“切片”和“切丁”之后，则它们最终的词嵌入也会相似。

Pérez-Ortiz et al. (2025 [本卷]) 指出，词嵌入的生成并非一步到位，而是通过连续的多个“神经层”才构建完成。在其外层之间夹有多层结构的人工神经网络被称为“深度神经网络”(deep neural network)。

进一步而言，“深度学习”(deep learning) 只是机器学习的分支，使用多个神经层来建立词语表示。在深度神经网络中，两个外层对应网络的输入和输

¹⁸ 请注意，我们在此的讨论从思想的表示，转移到了词语的表示。训练语料库包含数百万个可识别的词语，我们的目标是，在 NMT 系统中将这些词语表示出来。

出,且对人类分析师可见。而中间层或“隐藏”层却一直很难观察得到,这让“深度学习”给人以神秘莫测的感觉,导致一些评论者误导性地使用“魔法”一词来描述深度神经网络的内部运作。当大型科技公司宣告自己成功建立多语言翻译模型,有时不仅涉及数百种语言,还可以处理没有“直接”双语训练数据的语言之间的翻译,NMT系统便更添一份神秘感。¹⁹然而,人工智能的研究人员意识到了系统透明度不足所引发的问题。在“可解释的人工智能”(explainable AI, XAI)和“可诠释的人工智能”(interpretable AI)这两个领域,研究者正努力打开深度学习的“黑匣子”,以使用户更容易理解其内部运作,为输出提供解释,从而改进系统(示例可参见Vashishth et al. (2019))。

7 神经机器翻译的优缺点

NMT系统被普遍认为是迄今为止性能最好的机器翻译系统。例如,NMT系统之所以优于SMT系统,是因为它可以在考虑整个源语言句子,而不仅仅考虑n元组的情况下,为源语文本的单词建立非常丰富的词表示。产出译文时,NMT系统还可以同时考虑到这些丰富的表示和新产生的译文。正是由于NMT系统处理完整的句子,因此它更擅长处理棘手的语言特征(如不连续的依赖关系),且在处理各种一致现象方面的表现也比SMT系统更好。

尽管当前的NMT系统能考虑完整语句,但仍只限于正在处理的语句。这意味着这类系统还无法使用前一句子的信息,来判断代词(如it)在当前句子的具体所指,也无法获知某个西班牙语动词的主语是阴性(如西班牙语的空主语现象)。句级处理的这些缺点会导致许多其他问题,而这些问题只在用户翻译全文,而非单独语句时才显现出来。不过,目前已有“文档级机器翻译”的研究人员正努力解决这些问题(详见Bao et al. (2021))。此外,NMT系统还会输出实际上不存在于目的语中的单词。更严重的不足是,NMT系统的译文可能很流畅,但并不准确。当译文看上去和听起来都不错时,人们可能会忘了检查它是否与原文相符。再者,正如在大量现有文本上训练的其他技术一样,NMT也会放大训练数据中的偏见。已有大量研究讨论过性别偏见问题。以一个没有主语代词的西班牙语句子为例,如示例(8),许多NMT系统的英语译文都默认使用阳性主语代词。NMT开发者正在努力解决这个问题。现在有些系统同时输出阳性和阴性代词,供用户选择。本书也会在第4章和第5章中介绍用户如何才能充分利用某个NMT系统。

NMT系统的其他潜在不足与输出译文无关,而是关乎环境和社会问题。例如,NMT系统比以往系统的训练时间更长、算力要求更高,因此训练所耗费的能源更多。NMT系统不仅需要使用昂贵的专用图形处理器(graphical

¹⁹参见 <https://ai.googleblog.com/2016/11/zero-shot-translation-with-googles.html> 和 <https://about.fb.com/news/2020/10/first-multilingual-machine-translation-model>.

processing units, GPU), 还需要大量的训练数据, 但并非每个语言对都有如此大量的数据。

NMT 技术的进步也引起人们对外语学习的质疑: 如果机器能把其他语言的语音或文本翻译成自己的语言, 为什么还要大费周章去学习这门外语呢? 不过, 这种观点体现出对第二语言或外语学习益处的狭隘认知, 也忽略了机器翻译只适用于世界上少数语言。他们还倾向于将机器翻译与语言学习对立起来, 而不认为前者是后者的辅助手段。机器翻译在语言学习和生活方面的使用还引发了很多伦理和社会问题, 本书的第 1、6 和 9 章对此均有论述。

8 系统、引擎和定制化神经机器翻译

本章到目前为止, 我们试图概括性地解释何为翻译, 何为机器翻译, 以及不同类型机器翻译的运作模式。最后, 我们还想简要介绍一下特定的机器翻译系统以及机器翻译引擎的相关概念。

在常见用法中, 机器翻译“系统”(system) 通常是指由单个供应商或开发商提供的机器翻译产品或服务。因此, 谷歌翻译 (Google Translate) 可以说是谷歌的机器翻译系统, 而微软翻译器 (Microsoft Translator) 则是微软的机器翻译系统。用户可以在不同平台访问这些系统, 如安装谷歌翻译的手机应用程序, 或者通过网络浏览器在线使用谷歌翻译, 又或者在第三方软件中借助“应用程序接口”(application programming interface, API) 来访问。²⁰

在专业性更高的语境中, 如学术论文或职业翻译, 通常听到的概念是机器翻译“引擎”(engine)。这种语境下的“引擎”基本上是个机器翻译程序 (甚至可以说是“模型”), 训练过程使用特定语言对, 且通常针对某个领域或体裁。例如, 某家商业机器翻译公司以财务报表的平行语料库为训练数据, 为客户提供英语-法语的机器翻译引擎, 或者以医学领域文本为训练数据, 提供汉语-德语的机器翻译引擎。客户甚至可以使用自己的数据, 来创建或定制自己的机器翻译引擎。KantanAI 一度是提供这类服务是的领先企业。²¹ Ramírez-Sánchez (2025 [本卷]) 对定制化机器翻译进行了更深入的讨论。此外, MultiTraiNMT 项目也为此专门开发了教学界面, 让学生可以自行训练 NMT 引擎。²²

9 关于机器翻译, 你需要知道的最后四要点

许多读者可能只使用免费在线机器翻译, 因此只会用到为其感兴趣的语言对而构建的通用型引擎。但即便如此, 这些读者也应当了解:

²⁰我们在此以谷歌翻译为例, 只因它可能是最为人们熟知的机器翻译服务商。和专门的机器翻译供应商一样, 所有科技巨头也都提供不同的机器翻译“解决方案”。

²¹<https://www.kantanai.io>

²²<http://www.multitrainmt.eu>

- 不同的系统可能输出不同的译文；
- 同一系统中的不同引擎可能输出不同的译文；
- 即便输入内容相同，系统也可能因共现语篇不同而产出不同的译文；
- 某个系统的输出可能会随着时间的推移而有所不同。

例如，在撰写本文时，DeepL 的“法语-英式英语”机翻引擎将原文 (15) 译成译文 (16)，该法语表达 *mon petit doigt me dit* (字面意思是“我的小指告诉我”) 的含义是“我有预感”或“有人告诉了我”。读者可能也注意到，英式英语译文 (16) 使用了意思相似的恰当修辞手法。

- (15) *Mon petit doigt me dit que tu es marié.* (意为“我的小指告诉我你结婚了”)
- (16) *A little birdie tells me that you are married.* (意为“一只小鸟告诉我你结婚了”，译自 DeepL UK)

相反，出自谷歌翻译的译文 (17) 采用了不恰当的直译。

- (17) *My little finger tells me you're married.* (意为“我的小手指告诉我你结婚了”，译自谷歌翻译)

同样在撰写本文时，DeepL 的法语-美式英语机翻引擎输出译文 (18)，但如果改变原句中的一个单词，如原文 (19)，那么 DeepL 的法语-美式英语机翻引擎的性能要好得多，如译文 (20) 所示。

- (18) *My little finger tells me that you are married.* (意为“我的小手指告诉你已经结婚了。”，译自 DeepL US)
- (19) *Mon petit doigt me dit que tu es parti.* (意为“我的小指告诉我你走了”)
- (20) *A little birdie tells me that you've left.* (意为“一只小鸟告诉我你已经走了”，译自 DeepL US)

不过，读者阅读到这儿时，这两个系统的输出可能已经完全改变了，因为模型会被重新训练，且用户可以纠正错误译文。

10 结语

尽管崛起势头强劲, NMT 也不过是众多自动化翻译技术的先进代表。它的成功让政策制定者和普通公民对外语学习或译者培训的价值产生质疑。但是, 这类立场忽略了一个事实, 那就是 NMT 仍然依赖人工翻译, 或至少依赖经过人工审核的翻译作为训练数据。和其他类型的机器翻译一样, NMT 并非完美无缺, 其译文仍然需要熟练掌握源语和目的语的专业译者来测评或改进。提升用户机器翻译素养迫在眉睫, 即便对于一般用户亦是如此, 这样他们才能少走弯路。在合适的条件下, NMT 能成为促进和维护多语制、语言学习和由人类完成或监督的持续翻译的重要支柱。本书其余章节都正致力帮助创造这些条件。

References

- Baker, Mona & Gabriela Saldanha (eds.). 2020. *The Routledge encyclopedia of translation studies*. 3rd edition. London/New York: Routledge.
- Bao, Guangsheng, Yue Zhang, Zhiyang Teng, Boxing Chen & Weihua Luo. 2021. G-transformer for document-level machine translation. In *Proceedings of the 59th annual meeting of the Association for Computational Linguistics and the 11th international joint conference on Natural Language Processing (volume 1: Long papers)*, 3442–3455. Association for Computational Linguistics. <https://aclanthology.org/2021.acl-long.267>.
- 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.
- Carré, Alice, Dorothy Kenny, Caroline Rossi, Pilar Sánchez-Gijón & Olga Torres-Hostench. 2025. 机器翻译之于语言学习者. In Dorothy Kenny (ed.), *机器翻译知识普及: 为人工智能时代的用户赋能*, 147–163. Berlin: Language Science Press. DOI: 10.5281/zenodo.14922301.
- Caswell, Isaac. 2022. *Google Translate learns 24 new languages*. <https://blog.google/products/translate/24-new-languages/>.
- Forcada, M. L., M. Ginestí-Rosell, J. Nordfalk, J. O'Regan, S. Ortiz-Rojas, J. A. Pérez-Ortiz, F. Sánchez-Martínez, G. Ramírez-Sánchez & F. M. Tyers. 2011. Apertium: A free/open-source platform for rule-based machine translation. *Machine Translation* 24(1). 1–18.
- Forcada, Mikel. 2017. Making sense of neural translation. *Translation Spaces* 6(2). 291–309.

- Goodfellow, Ian, Yoshua Bengio & Aaron Courville. 2016. *Deep learning*. Cambridge, MA: MIT Press.
- Hutchins, John (ed.). 2000. *Early Years in Machine Translation. Memoirs and Biographies of Pioneers*. Amsterdam/Philadelphia: John Benjamins.
- Jakobson, Roman. 1959. On linguistic aspects of translation. In Reuben A. Brower (ed.), *On Translation*, 232–239. Cambridge, MA: Harvard University Press.
- Johnson, Joseph. 2021. *Worldwide digital population as of January 2021*. statista.com/statistics/617136/digital-population-worldwide/.
- Joscelyne, A. 1998. AltaVista translates in real time. *Language International* 10(1). 6–7.
- Koehn, Philipp. 2005. Europarl: A parallel corpus for Statistical Machine Translation. In *Proceedings of Machine Translation Summit X*, 79–86. <https://aclanthology.org/2005.mtsummit-papers.11>.
- Koehn, Philipp. 2010. *Statistical Machine Translation*. Cambridge: Cambridge University Press.
- Koehn, Philipp. 2020. *Neural Machine Translation*. Cambridge: Cambridge University Press.
- Marking, Marion. 2016. *Facebook says Statistical Machine Translation has reached end of life*. <https://slator.com/facebook-says-statistical-machine-translation-has-reached-end-of-life>.
- Moorkens, Joss. 2012. *Measuring consistency in translation memories: A mixed-methods case study*. Unpublished PhD thesis. Dublin: Dublin City University. <https://doras.dcu.ie/17332/>.
- 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.
- Nurminen, Mary & Maarit Koponen. 2020. Machine translation and fair access to information. *Translation Spaces* 9(1). 150–169.
- 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.
- Pym, Anthony. 2010. *Translation and text transfer*. Tarragona: Intercultural Studies Group.
- 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.
- Slobin, Dan. 2003. Language and thought online: Cognitive consequences of linguistic relativity. In Dedre Gentner & Susan Goldin-Meadow (eds.), *Language in mind: Advances in the investigation of language and thought*, 157–191. Cambridge, MA: MIT Press.
- Tammet, Daniel. 2006. *Born on a Blue Day*. London: Hodder & Stoughton.
- Tammet, Daniel. 2009. *Je suis né un jour bleu*. Paris: J'ai Lu.
- Toral, Antonio, Sheila Castilho, Ke Hu & Andy Way. 2018. Attaining the unattainable? Reassessing claims of human parity in neural machine translation. In *Proceedings of the third conference on machine translation: Research papers*, 113–123. Brussels: Association for Computational Linguistics. DOI: 10.18653/v1/W18-6312. <https://aclanthology.org/W18-6312>.
- Turovsky, Barak. 2016. *Ten years of Google Translate*. <https://www.blog.google/products/translate/ten-years-of-google-translate/>.
- Vashishth, Shikhar, Shyam Upadhyay, Gaurav Singh Tomar & Manaal Faruqi. 2019. Attention interpretability across NLP tasks. *CoRR* abs/1909.11218. <http://arxiv.org/abs/1909.11218>.

第三章

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

Caroline Rossi

格勒诺布尔-阿尔卑斯大学

Alice Carré

格勒诺布尔-阿尔卑斯大学

机器翻译发展迅速，但无法提供放之四海而皆准的解决方案。为了给特定项目选择恰当的解决方案，用户需要进行多种比较和评测。这从来都并非易事，尤其当机器翻译输出的质量看似越来越好，错误越来越难识别。我们如何才能最恰当地定义和评测神经机器翻译解决方案的质量，从而做出正确的选择呢？第一步当然是尽可能明确需求。本章先从语用角度给质量下定义，然后介绍人工评测和自动评测机器翻译质量的关键概念，并概述译者可以如何运用这些概念。

1 引言

神经机器翻译（NMT）被大肆宣传的同时，用户也的确注意到了机器翻译输出质量的提升。本章旨在说明，机器翻译输出虽看似更流畅，但并不一定更易处理。此外，神经机器翻译的输出译文总在变化之中，其使用应考虑具体情况和最终用户的需求。接下来，我们会提出了对质量和衡量标准的定义，这些定义和衡量标准并不局限于神经机器翻译输出表面上的易用性和流利度。

本章首先要回答的问题是：如何以可靠且有效的方式来评测神经机器翻译解决方案？根据使用场景和文本类型，答案可能会有所不同。接下来，我们将介绍机器翻译评测的关键问题，希望能帮助用户选择适合其特定需求的机器翻译引擎。



1.1 何为机器翻译质量评测？

广义的翻译质量包括产品和过程：“翻译质量既包括最终产品（译文）的质量，也包括交易（所提供服务）的质量”（Gouadec 2010: 270）。此外，翻译质量评测在很大程度上取决于完成翻译时的场景，而且翻译教师和有特定使用目的的客户对翻译质量的期望肯定有所不同。换句话说，“质量的概念是相对的”（Grbić 2008: 232）。在翻译研究领域，翻译质量众所周知地难以定义且千变万化。因此，许多综述研究便根据不同的翻译理论来定义翻译质量（参见例如Drugan 2013、House 2015）。在机器翻译质量评测中，也存在对“质量”的不同定义。“翻译技术（特别是机器翻译）的发展和广泛运用，催生出太多不明晰且操作方法不同的质量的定义及其相应的衡量标准”（Doherty 2017: 131）。就机器翻译而言，质量评测往往被视作实现某个目标（即改进系统）的手段，因此语用维度的考量便占据主导地位，通常结合人工评测和自动评测（Doherty 2017: 133）。在介绍现有的评测方法以及如何将其结合使用之前，我们先来解释语用维度在机器翻译质量评测中的必要性。

对于机器翻译的用户来说，机器翻译输出的质量评测过于复杂。虽然输出质量主要由系统本身决定，但翻译场景和用户需求也是质量评测需要考量的关键因素。举个简单的例子：你或许觉得适应用户手册中翻译不恰当的使用说明并不难，因为你已经清楚知道如何使用自己购买的产品，或者图示说明就足够了。在这种情况下，我们很可能对机器翻译错误有相对较高的容忍度（Castilho & O'Brien 2016）。现在，我们切换到另一个完全不同的场景，也涉及技术文本，但增加了法律层面的考量：专利证书译文的使用者需要获得精确的相关信息，因此对机器翻译错误的容忍度会低得多。提及专利领域的神经机器翻译，Castilho et al. (2017: 113) 的研究表明，神经机器翻译时常漏译原文信息，而在这种翻译场景中的机器译文信息缺失会造成严重后果。在上述两个例子中，质量评测的语用维度考量也就意味着采取可测量的有用性指标，如用户满意度评分、译后编辑效率的提高或机器翻译的产品描述带来的销售额增幅。

总的来说，翻译质量评测绝非小事。无论是人工评测还是自动评测，都会受到多重因素的影响。首先，合理的译文通常并非只有一个，也就是说原文存在多种同样可以接受的译法。此外，人工评测往往比较主观——事实上，评测者对统一译文质量水平存在不同看法的情况并不少见。但是，基于评测者处理译文的过程来计算评测效率，却不失为一种的客观方法。总之，为了减少主观影响，我们必须明确界定评测的目标和指标。人工评测的另一个缺点是过程耗时且占用大量资源。作为人工评测的替代方案，基于算法的自动评测尽管肯定比人工评测的成本更低、速度更快，但有时关联度低，因为自动评测无法评估译文在具体应用场景的可用性。因此，这两类评测方法各有利弊，应根据翻译项目和需求来进行选择。

1.2 质量恰好：得再三考虑！

如上所述，若从语用维度来进行机器翻译质量评测，那么显然并非所有错误的严重程度都一样。我们常以食谱翻译为例，因为译文质量很容易检验（“尝”得出来）。其实，食谱还常被用来展示机器翻译如何闹笑话。例如，以下为用机器将法语译成英语的千层面食谱，试问，单看译文，你能做出千层面吗？

表 1: 食谱译文的神经机器翻译错误示例（划线部分）（法译英）

法语原文 ^a	英语译文（神经机器翻译）
Préchauffez votre four à 180°C. Dans un plat à gratin beurré, versez un peu de béchamel. Déposez une couche de pâte et poser une couche de farce a la viande. Déposez à nouveau des pâtes, béchamel et viande.	Preheat your oven to 180°C. In a buttered gratin dish, pour a little béchamel. Put a layer of dough and a layer of meat stuffing. Put a layer of pasta, béchamel and meat filling on top.
Terminer par une couche de pâte avec de la béchamel et saupoudrez de fromage râpé. Laissez cuire 35 à 40 minutes.	Finish with a layer of dough with béchamel and sprinkle with grated cheese. Leave to cook for 35 to 40 minutes.

^a奶奶千层面（法语食谱）：<https://www.750g.com/lasagnes-r66998.htm>

你想必会惊讶地发现，原文中 *pâte*（面团；面条）出现了单数和复数形式，但均指意大利面，而机器翻译将单数形式译为面团。聪明谨慎如你或许能猜到正确词义，但一味按照通顺的机器翻译食谱来操作，你可能会做成另一道菜，而问题仅仅源于一个字母（表示复数的词尾）。尽管机器翻译译文总体流畅准确，稍作修改即可提升质量，但只要存在一个严重问题，就会让译文的功能失效。食谱内容相对简单，再加上如果对常见菜肴有所了解的话，读者可以自行纠正译文错误，理解起来并不困难。但是，对于更复杂的其他文本类型或专业领域，缺乏背景知识的读者就没这么好理解了。另外，这里提出的评测方法涉及到人、配料和厨房：测试成本或许很高，且几乎从未被使用过。

除了这类错误，大多数机器翻译用户很可能会遇到抽象概念和隐喻表达的翻译问题。表 2 展示了一本法语原著的简介摘录，以及神经机器翻译系统输出的英语译文。请问英语读者知不知道 *a veritable pie in the sky*（法语原文为 *véritable tarte à la crème*）意为 *well-trodden path*（常有人走的路），还是 *prefabricated subject*（预制对象）？

表 2: 习惯用语的误译示例 (带下划线)

法语原文 ^a	英语译文 (神经机器翻译)
L' <u>indépendance</u> du <u>parquet</u> , <u>véritable</u> <u>tarte</u> à la <u>crème</u> remise sur le plateau à chaque campagne présidentielle, était aussi une proposition du candidat Macron. Il ne l' a pas tenue.	The independence of the public prosecutor's office, a veritable pie in the sky put forward in every presidential campaign, was also a proposal by candidate Macron. He did not keep it.

^a法语原文来源: <https://www.grasset.fr/livres/ministere-de-linjustice-9782246827504>

你如果经常使用机器翻译, 或许能识别出上述或其他的翻译错误。经验很重要! 机器翻译译文越是通顺流畅, 就越需要谨慎对待——最近有研究表明, 学生对神经机器翻译译文的修改率低于流畅度较差的其他类型的机器翻译系统 (Yamada 2019)。熟悉某个语言对或领域的神经机器翻译译文的常见错误, 有助于使用者更有效地识别和改正这些错误。

总之, 尽管神经机器翻译的质量确实越来越高, 这点无可否认, 但它可能并不比其他类型的机器翻译更容易处理, 而机器翻译也从来都不是唾手可得的灵丹妙药。相反, 我们需要注意到隐藏在流畅的机器翻译译文中的微小错误, 并在选择合适的机器翻译解决方案前, 仔细思考自身需求。

2 为专业翻译选择合适的机器翻译引擎

上一节主要概述了神经机器翻译的一般使用和评测, 而为专业翻译选择合适的机器翻译引擎则意味着要考虑很多其他因素。关键在于, 适合专业用途的机器翻译引擎不仅必须满足客户的隐私要求、完美融入工作流程、支持所需语言对, 还能输出只需最小的译后编辑工作量, 就能提供满足客户需求的译文。机器翻译输出的质量将取决于专业领域和文本类型、引擎的可训练性以及译前编辑和译后编辑的工作量。

2.1 隐私和保密

机器翻译系统是如何处理用户数据的? 尽管这个问题对你、你的业务员或客户来说都至关重要, 但大多数机器翻译解决方案并没有给出明确回答, 或没有在最开始就隐私问题发出警示。相反, 你必须仔细阅读隐私声明, 并确保在使用前做出正确选择。

例如, 欧盟的神经机器翻译系统 eTranslation 的隐私声明中警示道: “用户向任何在线服务 (包括 eTranslation) 提交潜在敏感文件时, 应自行做出判

断”。即使数据不会被保存在机器翻译供应商的服务器中，这类警示也是正当且值得关注的。其实，就算已上传的数据不久后便会删除，也有可能造成隐私问题。出于保密考虑，欧盟法院的听证会甚至禁止使用 eTranslation 等内部机器翻译系统，因为没有哪个机器翻译系统能够达到此类听证会的严格保密要求¹ (C. Lenglet, 个人交流)。

当然，免费在线机器翻译存在的隐私问题风险要高得多，因为数据会被一直保存和重复使用。因此，无意中输入免费机器翻译系统的文本的某些机密信息，可能会以意想不到的方式泄露 (请参阅 Moorkens 2025 [本卷] 深入了解如何合乎道德地、安全地处理数据)。

2.2 比较机器翻译输出

你或许想用同一篇原文来测试几种机器翻译工具，然后通过比较其输出以判断哪种最符合自己的需求。

根据你想解决的问题，机器翻译输出的比较方法有以下几种。例如，可以通过人工评估、自动评测和/或译后编辑工作量给译文打分，然后比较得分情况。译后编辑工作量的计算方法详见 O'Brien (2025 [本卷])。可以说，译后编辑工作量小或不需要进行译后编辑的机器翻译输出，一般认为比译后编辑工作量大的“更好”；还可以说，如果合理推断机器翻译文本会用于传播目的，则译后编辑工作量可用于评测译文质量 (可参阅 Kenny (2025 [本卷]) 详细了解以同化和以传播为目的的机器翻译之间的区别)。我们将在下文重点介绍机器翻译输出的人工评测和自动评测。

2.3 人工评测

人工评测需要依靠评测者来评估一个或多个系统的输出，通常逐句 (或逐个语段) 进行，但也可以进行文档级别的评估 (例子可参见 Castilho 2020)。评测者通常从两个方面来评分。第一是忠实度 (*adequacy*)，即衡量机器翻译译文对原文意思表达的完整程度。忠实度的衡量通常采用范围从 1 (指机器翻译完全没有表达出原文含义) 到 4 (机器翻译完整表达出原文含义) 的四级量表。有时也使用五级量表，不过由于量表级别数为单数，评测者可能会过度倾向于选择中间项。第二是流利度 (*fluency*)，即衡量“译文在多大程度上遵循了目的语的规则和规范” (Castilho et al. 2018: 18)。理论上，评测者对流利度的判断无需借助原文。其测量通常也使用四级量表，1 表示机器翻译译文“不流畅”，4 则表示该译文的流利程度达到母语水平 (Castilho 2020: 1152-1153)。忠实度和流利度的评测非常耗时，因此操作成本很高。

另一个简单快捷的机器翻译系统输出的比较方法是，只需评测者对译文的质量进行排序，也就是说，判断哪种“更好”即可，无需说明理由。目前已

¹详细描述和解释可参考: <https://eur-lex.europa.eu/legal-content/en/TXT/PDF/?uri=CELEX:32013D0488>

有很多机器翻译供应商采用这种方法，以便迅速获得在线用户的反馈。例如，Moorkens (2018)指出，2017 年，微软采用该方法获得了用户对其统计机器翻译和神经机器翻译系统输出的评测结果。

其他机器翻译供应商已经开发出更复杂的工具来协助机器翻译输出的人工评测。例如，Kantan AI 提供了名为 KantanLQR (Language Quality Review, 即语言质量审查) 的工具，允许用户按需自行选择质量评测标准 (如忠实度、流利度和术语使用等)，然后根据这些标准来比较多达四种不同的机器翻译输出。² 这类工具非常实用，不仅能通过饼图和条形图为评测者对每个句段的质量得分进行视觉化呈现，还能计算不同机器翻译引擎或系统的总体得分。此外，这类工具还具备为翻译公司的项目经理以及评测者而设计的功能。PET(Aziz et al. 2012) 等非商业工具也可用于帮助评测者评估机器翻译输出，同时备受研究人员的青睐。

其他支持人工评测的常用工具还包括电子表格程序，使用者可以先将分数手动输入表格，如表 3所示，然后用内置函数来计算质量指标的平均得分。各种免费在线表格也可以用来进行人工评测。³ 线上表单尤其适用于调查问卷，也和电子表格一样能够自动计算结果以及其他统计学指标。

表 3: 用于比较机器翻译系统的电子表格示例

原文	NMT 译文 1	NMT 译文 2	NMT 译文 3	偏好译文	注解
语段 1					
语段 2					

2.4 错误类型

有时，评估者不仅要使用上述标准对机器翻译的段落或文档进行评分，还要根据错误分类法 (*error typology*) 将译文中的每个错误归类，具体说明译文的问题所在。对错误进行分类是识别机器翻译输出问题的重要步骤，通常是为了向系统开发人员提供反馈意见。

但是，错误类型往往非常复杂。例如，多维度质量评估 (Multidimensional Quality Metrics, MQM) 框架列出了详尽的错误类型 (Mariana et al. 2015: 140)。为便于操作，评测时可选择其中的常见类型，例如 Moorkens 为课堂翻译评测练习选择的错误类型 (Moorkens 2018: 380) 包括：

- 词序错误 (短语或单词级别的词序不正确)

²详见 <https://kantanmt.zendesk.com/hc/en-us/articles/115003644483-What-is-KantanLQR> 和 <https://twitter.com/i/status/1466392446552657927>.

³最有名的或许要数谷歌表单 (Google Sheet)。可参见 <https://support.google.com/docs/answer/6281888?hl=en&co=GENIE.Platform%3DDesktop>

- 误译（单词、性别、数字或大小写的翻译有误）
- 漏译（原文中的单词在目标文本中没有译出）
- 增译（原文中没有的词却在目标文本中译出）

有可能用于评测机器翻译输出的样本太小，不足以说明每个机器翻译引擎的水平。理想的情况应该是，用不同的样本进行重复比较，然后找出性能最佳的机器翻译引擎或系统。不过，大型机构可能有条件进行大规模评测，而较小的翻译服务提供者和自由译者可能更倾向于采取自动评测和计算译后编辑工作量的方法。

2.5 自动评测指标的使用

自动评测指标（automatic evaluation metrics, AEM）比人工评测的速度更快且成本更低，因此机器翻译用户可根据需要频繁对机器翻译输出进行评测。例如，在机器翻译引擎的训练过程中，用户可以每调整一次参数便进行评测，看看引擎的性能是否有所改善。如果不是训练引擎，而是为某个项目选择一个合适的引擎，那么自动评测指标也能用于评测同一原文的多种机器翻译输出。

对于同一原文，人工翻译的译文经常是千差万别。因此，我们不能指望机器翻译系统与人工翻译达到完全一致。但是，与人工翻译越相似，机器翻译的质量可能越好，反之则越差，因此许多自动评测指标都是基于相似性原则，即向评测工具同时输入人工翻译作为“黄金标准”（或称“参考译文”）和系统输出（或称“候选译文”，有时也称为“假设 (hypothesis)”）。然后，将候选译文与参考译文进行对比，并计算相似性 (similarity) 或相异性 (dissimilarity)。考虑到会存在多种参考译文，一些评测工具甚至可以采用多种参考译文。⁴

过去的二十年里出现了大量自动评测指标及其变体。不过在此，我们只会根据章节内容会提到的评测工具，如 KantanMT 和 MutNMT，有选择性地为读者介绍几种自动评测指标。想了解更多有关自动评测指标信息的读者，可参阅 Koehn (2010) 和 Koehn (2010)。为保持一致，我们将采用 Koehn 提出的术语和标记方法，并在适用的情况下，在以下例子中运用 Koehn 的解释来评测神经机器翻译的输出。

以下例子摘录自一种无线电收发器的用户手册⁵，如图 1 所示。

手册中逐点列举了确保该收发器防水的前提条件，上述原文便是其中一点 (参见图 2)。

⁴在这种情况下，需要决定如何计算参考译文的长度。例如，同时采用多种参考译文的 BLEU (双语替换评测) 采用的是“长度最接近候选译文”的参考译文。(Qin & Specia 2015: 114)

⁵Vertex Standard 的 VX-450 系列现已停产。

原文:	Battery pack is attached to the transceiver.
-----	--

图 1: 本节主要示例的原文

<p>重要提示</p> <p>只有在下列条件下, 才能保证收发器的防水性能 (IP57: 1 米/30 分钟) (<i>sic</i>):</p> <ul style="list-style-type: none">• 电池组附着于收发器;• 天线连接到天线插孔;• MIC/SP 帽安装在 MIC/SP 插孔中。

图 2: 收发器用户手册摘录

虽然该手册是面向大众而撰写的, 但原文属于技术文本, 因此其翻译可视为专业领域翻译。文本涉及无线电通信领域, 因此必须采用该领域的术语和措辞, 而作为用户手册, 译文也应保留这类体裁的特点。例如, 每个概念都应只用一个术语来表示 (即不应使用同义词), 且每个术语都只对应一个概念, 即“单义性”(monosemy)。手册说明应简洁易懂, 写作模式统一。(更多关于专业领域和文本体裁的介绍, 请参见Kenny 2025 [本卷]。)在我们上述的翻译任务中, 目标语是法语, 译文用途与原文的一致, 即供收发器的用户使用。

接下来, 我们就这段摘录为原文, 对比三种机器翻译工具的输出。首先是欧盟的机器翻译系统 eTranslation (以下简称“系统 A”, 其输出简称“候选译文 A”)。⁶ 其次是谷歌翻译 (以下简称“系统 B”, 其输出简称“候选译文 B”)。⁷ 最后是 DeepL (以下简称“系统 D”, 其输出简称“候选译文 D”)。⁸ 在撰写本书时, 这些系统均对公众免费开放, 不过 eTranslation 要求用户进行注册, 且身份类别为中小企业、公共服务官员或公共服务提供商。

本章节介绍的自动评测指标中, 有些简单方法可以手动计算。但对于复杂的方法, 我们会借助 MutNMT 来计算。⁹ 我们会向读者举例, 帮助了解这些指标的运算过程, 但还想说明的是, 分数的精确计算会根据评测指标的具体操作细节有所不同——如果使用不同的工具来计算似乎相同的 AEM (例如 BLEU), 您很可能会得到不同的结果。¹⁰ 评测结果出现差异的原因可能有

⁶<https://webgate.ec.europa.eu/etranslation/translateTextSnippet.html>

⁷<https://translate.google.com/?hl=en>

⁸<https://www.deepl.com/en/translator>。请注意, 我们将 DeepL 称为“系统 D”, 而不是“系统 C”, 是为了避免在使用字母 C 来指代候选译文 (candidate translation) 的情况下出现混淆。

⁹<https://mutnmt.prompsit.com/index>

¹⁰不过, 最近有开发者对努力实现自动评测指标的规范化和参考应用, 如 Matt Post 的 sacrebleu 等软件 (Post, 2018)。

以下几个：评测工具在计算前处理引号、连字符、普通空格和不断行空格等的方式不同；对于词符 (token) 的定义方式不同 (是否考虑到撇号、连字符、标点符号或语言学信息，如词目 (lemmas) 或多词单位)；对大小写是否敏感；度量参数细节 (例如，具体操作中 *n*-grams 的顺序)。¹¹ 在上述例子中，我们将候选译文 D 中的撇号改为参考译文所使用的撇号。这样一来，自动评测指标的结果就不会受到不同撇号的影响，我们便可以只关注机器翻译输出本身。此外，我们为了展示而手动计算自动评测指标时，将连字符和撇号视为单词“分隔符”，也就是说参考译文 (参见图 3) 共包含 8 个词。

参考译文：	La batterie est installée sur l' émetteur-récepteur.
-------	--

图 3: 参考译文示例

图 4 列出了我们将在下文讨论的原文、候选译文和参考译文。

原文：	Battery pack is attached to the transceiver.
参考译文：	La batterie est installée sur l' émetteur-récepteur.
候选译文 A(eTranslation)：	Le bloc-batterie est fixé à l' émetteur-récepteur.
候选译文 B(谷歌翻译)：	La batterie est fixée à l' émetteur-récepteur.
候选译文 D(DeepL)：	Le bloc-piles est fixé à l' émetteur-récepteur.

图 4: 本节的主要示例——原文、参考译文和候选译文

对于上述例子，人工评测者会给出什么反馈呢？第一点，除非客户另有说明，术语 battery pack (电池组) 应译为 batterie。而候选译文 D 中译成 bloc-piles 显然是错的，因为这种收发器使用的并非 piles (不可充电的一次性电化学电池)，而是 batterie (可充电的电池组)。在这种情况下，收发器使用的是锂电池。候选译文 A 中译作 bloc-batterie 本质上没错，但这不符合译入语表达习惯，而是仿造词 (calque)，即对英语原文的逐字翻译。第二点，动词词组 is attached to 可以根据个人喜好译为 est installée sur 或 est fixée à。第三点，transceiver (收发器) 是 transmitter-receiver (发射接收机) 的缩写，理想情况下应译为 émetteur-récepteur，如图 4 中列出的所有译文所示。不过，对这个翻译任务来说，译作 radio (无线电)，甚至 appareil (设备) 也是可以的 (关于“翻译”和“对等”的更多内容，请参见Kenny 2025 [本卷])。现在，让我们深入了解自动评测指标是如何评测候选译文的。

¹¹ 本书作者要对 Gema Ramírez-Sánchez 的解释致谢。

2.5.1 核心概念： n -grams、精确率、召回率和 F 值

我们将在本节介绍四个概念： n -grams, 精确率, 召回率 and F 值。这些概念有助于我们理解下文涉及的复杂的自动评测指标。

2.5.1.1 n -grams

在翻译领域, n -grams (参见Kenny 2025 [本卷]) 通常被理解为 n 词序列。如上述例子中, battery 是 1-gram 或一元分词 (unigram), battery pack 是 2-gram 或二元分词 (bigram), battery pack is 是 3-gram 或三元分词 (trigram)。此外, 还有 4-gram、5-gram 等, 如 battery pack is attached 为 4-gram。

n -grams 通常用于语言建模, 例如, 一个 3-gram 的概率表示假设已知某个词前面的两个词, 则该词出现的概率。

在自动评测指标中, n -grams 指的是参考译文中与候选译文匹配的 n 词序列。近来, 有研究者提出了基于字符序列, 而非词序列的自动评测指标。那么, N -grams 就可以理解为 n 个字符的序列, 而非 n 个单词的序列。

讨论 BLEU 时, n -grams 指的是 n 词序列 (参见 see 2.5.4.), 而讨论 ChrF3 时, n -grams 指的是 n 字符序列 (参见2.5.5)。

2.5.1.2 精确率与召回率

精确率是自然语言处理许多分支都会用到的基本概念。我们借用一个简单的例子来解释: 假设老师让学生用英语说出一个星期的每一天, 学生回答“星期一和星期二”, 则他给出了两个正确答案, 没有错误答案。由于“精确率”指的是给出的正确答案与答案总数之比, 因此该学生的得分是 $2/2$, 即精确率达到惊人的 100%。

但了解正确答案的老师知道该学生的回答并不完整, 因为他漏掉了其他五天。因此, 老师可以说该学生回答的“召回率”低。“召回率”指的是给出的正确答案与正确答案总数 (理想答案) 之比。在这种情况下, 该学生回答的召回率为 $2/7$, 相当于略低于 29%。

对机器翻译译文进行自动评测时, 精确率为候选译文中的正确词数 (即同时出现在参考译文中的词) 与其总词数之比:

$$C \text{ 的精确率} = \frac{C \text{ 的正确词数}}{C \text{ 的总词数}} \quad (1)$$

C 表示候选译文。

一起来看看上述的例子。在图 5 中, 不同的候选译文对比参考译文。“正确”用词 (即同时出现在参考译文的词) 用下划线标出, 而“错误”用词 (即没有出现在参考译文的词) 则用删除线标出。

原文:	Battery pack is attached to the transceiver.
候选译文 A(eTranslation):	Le bloc - <u>batterie</u> <u>est fixé</u> -à l' <u>émetteur-récepteur</u> .
候选译文 B(Google Translate):	La <u>batterie</u> <u>est fixée</u> -à l' <u>émetteur-récepteur</u> .
候选译文 D(DeepL):	Le bloc - <u>piles</u> <u>est fixé</u> -à l' <u>émetteur-récepteur</u> .
参考译文:	La batterie est installée sur l' <u>émetteur-récepteur</u> .

图 5: 原文、参考译文和候选译文

现在来计算每个候选译文的精确率。候选译文 A 的正确用词为 5/9, 即精确率为 0.56 或 56%。¹² 候选译文 B 的正确用词为 6/8, 即精确率为 0.75 或 75%。最后, 候选译文 D 的正确用词为 4/9, 即精确率为 0.44 或 44%。由以上结果得出, 候选译文 B 优于候选译文 A 和候选译文 D。

同理, “召回率”指的是候选译文的正确词数与参考译文的总词数之比:

$$C \text{ 的召回率} = \frac{C \text{ 的正确词数}}{R \text{ 的总词数}} \quad (2)$$

C 表示候选译文, R 表示参考译文。

换句话说, 召回率不仅考虑候选译文用了哪些词, 还考虑应该用哪些词。回到以上的例子 (图 6)。

原文:	Battery pack is attached to the transceiver.
候选译文 A(eTranslation):	Le bloc - <u>batterie</u> <u>est fixé</u> -à l' <u>émetteur-récepteur</u> .
候选译文 B(Google Translate):	La <u>batterie</u> <u>est fixée</u> -à l' <u>émetteur-récepteur</u> .
候选译文 D(DeepL):	Le bloc - <u>piles</u> <u>est fixé</u> -à l' <u>émetteur-récepteur</u> .
参考译文:	La batterie est installée sur l' <u>émetteur-récepteur</u> .

图 6: 原文、参考译文和候选译文

¹²在本节中, 所有在 0 到 1 区间的结果将保留两位小数, 而百分比则保留两位数。

参考译文共有 8 个词。候选译文 A 有 5 个词出现在参考译文中, 则召回率为 0.63 或 63%。候选译文 B 有 6 个词出现在参考译文中, 则召回率为 0.75 或 75%, 而候选译文 D 有 4 个词出现在参考译文中, 则召回率为 0.5 或 50%。根据召回率的计算结果, 候选译文 B 依然优于候选译文 A 和候选译文 D。

2.5.1.3 F 值

上述例子中的学生可优先考虑精确率, 只回答“星期一、星期二”即可, 而不必冒险给出错误答案; 也可能优先考虑召回率, 一口气说出几十个答案, 希望能尽量命中正确答案。所以学生可能会回答“星期一、星期二、星期三、星期四、星期五、星期六、星期日、一月、二月、三月、四月、五月、六月、七月、八月、九月、十月、十一月、十二月”。此时, 召回率突然升到 100%, 因为他们的回答包括了所有正确答案; 但准确率竟不足 37%, 因为在 19 个回答中, 只有 7 个是正确的。在老师看来, 这两种策略都不可取。老师想要的是, 让学生同时最大化精确率和召回率。他们需要一个同时兼顾二者的指标, 而这就是所谓的 F 值。

用数学术语来说, F 值是精确率和召回率的调和平均值。计算方法如下:

$$F = 2 \cdot \frac{\text{精确率} \cdot \text{召回率}}{\text{精确率} + \text{召回率}} \quad (3)$$

还可以表示为:

$$F = 2 \cdot \frac{C \text{ 的正确词数}}{C \text{ 的总词数} + R \text{ 的总词数}} \quad (4)$$

C 表示候选译文, R 表示参考译文。

下面一起来计算表 4 中三个候选译文的 F 值。

表 4: 候选译文 A、B 和 D 的精确率、召回率和 F 值

评测指标	候选译文 A	候选译文 B	候选译文 D
精确率	56%	75%	44%
召回率	63%	75%	50%
F 值	$2 \cdot \frac{56 \cdot 63}{56+63} = 59$	$2 \cdot \frac{75 \cdot 75}{75+75} = 75$	$2 \cdot \frac{44 \cdot 50}{44+50} = 47$

候选译文 A 的 F 值为 59%, 候选译文 B 的 F 值为 75%, 而候选译文 D 的译文 F 值为 47%。以上结果依然得出, 候选译文 B 优于候选译文 A 候选译文 D。

精确率、召回率和 F 值这三个指标的得分越高, 则机器翻译的质量越好。然而, 这些评测指标只考虑用词, 并不考虑词序。

2.5.2 翻译错误率 (TER)

翻译错误率 (*translation error rate, TER*), 也称作“翻译编辑率”(translation edit rate, TER), 将词序考虑在内。

这种方法基于使用 Levenshtein 距离的单词错误率 (word error rate, WER)。Levenshtein 距离计算不同序列 (此处指单词序列) 之间的差异, 其定义是“匹配两个序列所需的编辑步骤 (包括插入、删除和替换) 的最小值” (Koehn 2010: 224)。“单词错误率”根据参考译文 的长度将 Levenshtein 距离进行归一化 (Koehn 2010: 225):

$$WER = \frac{\text{替换词数} + \text{插入词数} + \text{删除词数}}{R\text{的总词数}}$$

(5)

R 表示参考译文。

可是, 当词序列或整个从句被移到句子的其他地方, 每个单词的移动都被视为两个错误 (即一次删除和一次插入), 从而导致单词错误率非常高。

为了解决这个问题, “翻译错误率”增加了一次额外操作——移位, 即移动词序列只算一个错误:

$$TER = \frac{\text{移位词数} + \text{替换词数} + \text{插入词数} + \text{删除词数}}{R\text{的总词数}}$$

(6)

R 表示参考译文。

一起回到上述例子。将候选译文 A、B 和 D 与参考译文进行对比 (图 7)。把候选译文 A、B 和 D 改为参考译文, 最少需要多少步?

原文:	Battery pack is attached to the transceiver.
候选译文 A(eTranslation):	Le bloc-batterie est fixé à l' émetteur-récepteur.
候选译文 B(Google Translate):	La batterie est fixée à l' émetteur-récepteur.
候选译文 D(DeepL):	Le bloc-piles est fixé à l' émetteur-récepteur.
参考译文:	La batterie est installée sur l' émetteur-récepteur.

图 7: 原文、候选译文 (A、B、D) 和参考译文

“翻译错误率”是启发式 (或迭代) 过程, 这种算法试图通过对比译文的词序列来找到最佳答案 (把一个序列转换为另一个序列所需的最少步骤数)。我们可以借助矩阵来手动计算翻译错误率。但为了方便解释, 我们采用更简便但可能不完美的方法¹³: 比较每个候选译文, 并计算匹配、移位、替换、添加和删除的词数。记住, 匹配的词数不计算在内。如前所述, 连字符和撇号被视作单词的“分隔符”。不将这些符号视作分隔符的工具把 l' émetteur-récepteur 当成 1 个单词, 而不是 3 个单词, 因此计算结果也会不一样。

¹³注意, 我们的例子没有移位情况, 因此这里的翻译错误率与单词错误率相等。

2.5.2.1 候选译文 A

表 5: 候选译文 A 变成参考译文所需的操作步骤

操作步骤	被编辑的词	编辑步骤数
匹配	batterie, est, l' , émetteur, récepteur	5
移位		0
替换	le/la, fixé/installée, à/sur	3
插入		0
删除	bloc	1

候选译文 A 的翻译错误率计算如下:

$$TER_A = \frac{0 + 3 + 0 + 1}{8} = 0.5 = 50\% \quad (7)$$

2.5.2.2 候选译文 B

表 6: 候选译文 B 变成参考译文所需的操作步骤

操作步骤	被编辑的词	编辑步骤数
匹配	la, batterie, est, l' , émetteur, récepteur	6
移位		0
替换	fixée/installée, à/sur	2
插入		0
删除		0

候选译文 B 的翻译错误率计算如下:

$$TER_B = \frac{0 + 2 + 0 + 0}{8} = 0.75 = 75\% \quad (8)$$

2.5.2.3 候选译文 D

表 7: 候选译文 D 变成参考译文所需的操作步骤

操作步骤	被编辑的词	编辑步骤数
匹配	est, l' , émetteur, récepteur	3
移位		0
替换	le/la, bloc/batterie, fixé/installée, à/sur	4
插入		0
删除	piles	1

候选译文 D 的翻译错误率计算如下:

$$TER_D = \frac{0 + 4 + 0 + 1}{8} = 0.63 = 63\% \quad (9)$$

候选译文 A 的翻译错误率为 50%，候选译文 B 的翻译错误率为 75%，候选译文 D 的翻译错误率为 63%。因为“单词翻译错误率”和“翻译错误率”都是计算错误所占的百分比，所以这些指标考虑的是不匹配的情况。也就是说，这两者与精确率、召回率和 F 值的解读方式相反，数值越低，则机器翻译的质量越好。因此，由计算结果得出，候选译文 A 为最佳译文。

2.5.3 人工翻译编辑率 (HTER)

有些候选译文虽然与参考译文差异较大，但还是可以接受的，因此以参考译文为准来评测候选译文的方法，可能对机器翻译系统来说严苛且不公平。这时我们可以采用人工翻译编辑率 (*human translation edit rate*, HTER)，即要求评测者对某个候选译文进行译后编辑，然后计算将该候选译文变为译后编辑版本所需的编辑步骤数 (Snover et al. 2006)。

再来看看上述例子。将候选译文 A、B 和 D 与译后编辑版本进行比较，评测者可任选某个候选译文进行编辑 (图 8): 从候选译文 A、B 和 D 到译后编辑版本最少需要多少步?

同理，匹配词数不计算在内。

原文: Battery pack is attached to the transceiver.
候选译文 *A*(*eTranslation*): Le bloc-batterie est fixé à l' émetteur-récepteur.
候选译文 *B*(*Google Translate*): La batterie est fixée à l' émetteur-récepteur.
候选译文 *D*(*DeepL*): Le bloc-piles est fixé à l' émetteur-récepteur.
译后编辑: La batterie est fixée à l' émetteur-récepteur.

图 8: 原文、候选译文 (A、B、D) 和译后编辑

2.5.3.1 候选译文 A

表 8: 候选译文 A 变成译后编辑版本所需的操作步骤

操作步骤	被编辑的词	编辑步骤数
匹配	batterie, est, à, l' , émetteur, récepteur	6
移位		0
替换	le/la, fixé/fixée	2
插入		0
删除	bloc	1

候选译文 A 的人工翻译编辑率计算如下:

$$\text{HTER}_A = \frac{0 + 2 + 0 + 1}{8} = 0.38 = 38\% \tag{10}$$

2.5.3.2 候选译文 B

表 9: 候选译文 B 变成译后编辑版本所需的操作步骤

操作步骤	被编辑的词	编辑步骤数
匹配	la, batterie, est, fixée, à, l' , émetteur, récepteur	8
移位		0
替换		0
插入		0
删除		0

候选译文 B 的人工翻译编辑率计算如下：

$$\text{HTER}_B = \frac{0 + 0 + 0 + 0}{8} = 0 = 0\% \quad (11)$$

2.5.3.3 候选译文 D

表 10: 候选译文 D 变成译后编辑版本所需的操作步骤

操作步骤	被编辑的词	编辑步骤数
匹配	est, à, l' , émetteur, récepteur	4
移位		0
替换	le/la, bloc/batterie, fixé/fixée	3
插入		0
删除	piles	1

候选译文 D 的人工翻译编辑率计算如下：

$$\text{HTER}_D = \frac{0 + 3 + 0 + 1}{8} = 0.5 = 50\% \quad (12)$$

候选译文 A 的人工翻译编辑率为 38%，候选译文 B 的人工翻译编辑率为 0%，候选译文 D 的人工翻译编辑率为 50%。注意，“翻译错误率”和“人工翻译编辑率”均为错误率，数值越低，则机器翻译的质量越好。由计算结果可得，候选译文 B 的质量最好。

表 11: 候选译文的 TER 值和 HTER 值

评测指标	候选译文 A	候选译文 B	候选译文 D
TER	50%	25%	63%
HTER	38%	0%	50%

现在，比较这个例子 (表 11) 中候选译文的翻译错误率和人工翻译编辑率：所有候选译文的人工翻译编辑率都比翻译错误率值低。这个例子证实了“机器翻译及其译后编辑版本之间的编辑率，大大低于机器翻译和纯人工翻译的参考译文之间的编辑率” (Koehn 2020: 52)。这种差异提醒我们要警惕译后编辑者可能由于时间压力过大，从而导致译后编辑不足的风险。¹⁴

2.5.4 双语替换评测 (BLEU)

双语替换评测 (*bilingual evaluation understudy*, BLEU) 指候选译文和参考译文共有的 *n*-grams。¹⁵因此，这种方法同时考虑了匹配词数和词序。其 *n*-grams 的设置范围从 1-gram 到 4-grams，且可以赋予不同权重。

图 9 为 René Magritte 的画作《图像的背叛》(*The Treachery of Images*) 的名句的候选译文和参考译文。若以单词 (1-gram) 计算，参考译文含有 6 个单词，其中 5 个出现在候选译文 (若句末的标点符号也算作单词匹配的话)；若以双词 (2-gram) 计算，参考译文含有 5 个双词序列，其中 4 个出现在候选译文 ([is not], [not a], [a pipe], [pipe .]); 若以三词 (3-gram) 计算，参考译文含有 4 个三词序列，其中 3 个出现在候选译文 ([is not a], [not a pipe], [a pipe .]); 最后，若以四词 (4-gram) 计算，参考译文含有 3 个四词序列，其中 2 个出现在候选译文 ([is not a pipe],[not a pipe.]), 第一个四词序列如图 9 所示。

原文:	Ceci n' est pas une pipe.
候选译文 (假设):	That is not a pipe .
参考译文:	This is not a pipe .

图 9: 句子 Ceci n' est pas une pipe 的候选译文和参考译文, 示例完全一致的四词序列。

BLEU 值计算的是候选译文与参考译文所匹配的 *n*-grams 数量和候选译文的 *n*-grams 数量的比值，因此该值为精确率。表 12 所示为候选译文不同

¹⁴本章对分享此观点的 Mikel Forcada 致以谢意。
¹⁵请注意，如前文所述，BLEU 允许使用多个参考译文。

n -gram 的精确率 (以比率和小数来表示)。¹⁶

表 12: 候选译文 That is not a pipe 的精确率 (从 1-grams 到 4-grams)。

评测指标		
精确率 (1-gram)	5/6	0.83
精确率 (2-gram)	4/5	0.80
精确率 (3-gram)	3/4	0.75
精确率 (4-gram)	2/3	0.66

候选译文的总体 BLEU 值便是通过计算不同 n -grams 的精确率的几何平均值 (一种特殊的“平均值”) 而得出, 结果略低于 0.76 或 76%。¹⁷ (其实, 这一分值非常高, 不过例子也很简单。)

我们还应该注意的是, BLEU 值的计算通常基于整个语料库, 而非单个句子。有些系统只对有把握的词进行翻译, 从而变相提高精确率 (就如同上文例子中, 害怕犯错而只给出两个回答的学生), 因此 BLEU 值还设置了长度惩罚 (*brevity penalty*), 即候选译文和参考译文的词数比 (详见 Koehn 2020: 227), 当候选译文短于参考译文时, 该系数就能发挥出作用。不过, 图 9 的例子没有受到长度惩罚, 因为候选译文与参考译文的长度一致。

虽然这一指标通常被称为“BLEU 值”, 但是其计算过程涉及很多不同的参数 (Post 2018), 非专业人士难以理解某个自动测评工具如何计算出 BLEU 值。对于想要借助自动评测工具来评估机器译文的译者来说, 重要的是, 不同译文的评测方法保持一致。简言之, 就是确保使用相同的机器翻译评测工具并了解其设置; 如果有用户可自定义的设置, 则使用该工具时, 确保使用相同的自定义设置来对比不同译文。这样, 译者就能得到具有可比性的分数。

如表 13 所示为图 4 候选译文的 BLEU 值, 分别由 MutNMT 和 Tilde 的工具计算得出。¹⁸

由计算结果得出, 候选译文 B 的优于候选译文 A。这与截至目前得到的结果一致。不过在实际情况中, 不同评测工具算出的 BLEU 值可能出现很大差异, 因此用户需要分析导致这种差异的原因。

¹⁶表 12 的表述方式源自 (Koehn 2010: 227)。

¹⁷读者可采用熟练的电子表格软件来计算几何平均值。专门的 BLEU 分数计算器可能还需要考虑其他权重问题, 如较长的 n -grams, 且应该能够对 n -grams 精确率为 0 的情况使用平滑方法 (smoothing)。详见 Post (2018)。

¹⁸<https://mutnmt.prompsit.com/index>; MutNMT 使用 SACREBLEU 算法 (Post 2018)。Tilde 的“交互式 BLEU 值计算器”网址为<https://www.letsmt.eu/Bleu.aspx>。

表 13: 借助 MutNMT 和 Tilde 计算的候选译文 A、B 和 D 的句级 BLEU 值

BLEU 值计算器	候选译文 A	候选译文 B	候选译文 D
MutNMT	15%	31%	15%
Tilde	50%	61%	47%

2.5.5 ChrF3

字符匹配度 (ChrF) 是基于字符 n -grams 的 F 值。因此, 该数值以精确率和召回率为基础。回顾一下 F 值的计算公式:

$$F = 2 \cdot \frac{\text{精确率} \cdot \text{召回率}}{\text{精确率} + \text{召回率}} \quad (13)$$

而 ChrF 值的计算公式是:

$$\text{ChrF}\beta = (1 + \beta^2) \cdot \frac{\text{ChrP} \cdot \text{ChrR}}{\beta^2 \cdot \text{ChrP} + \text{ChrR}} \quad (14)$$

其中,

- ChrP 表示字符 n -gram 的精确率, 即候选译文的正确字符 n -grams 的数量除以其 n -grams 的总数,
- ChrR 表示字符 n -gram 的召回率, 即候选译文的正确字符 n -grams 的数量除以参考译文的 n -grams 的总数,
- β 是一个权重因子, 即表示召回率的影响力是精确度的 β 倍。假设 $\beta = 1$, 则召回率和精确率同等重要; ChrF1 是字符 n -gram 的精确率和召回率的调和平均数 (harmonic mean) (Popović 2015)。

19

ChrF3 值则是 $\beta = 3$ 时的 ChrF 值的变体, 也就是说, 召回率的重要性是精确率的三倍。Popović (2015) 认为, 评测试验表明 ChrF 值, 尤其 ChrF3 值, 是很有前景的机器翻译自动评测方法。

和 BLEU 值一样, 我们不会在此计算 ChrF3 值。不过, 可以借助自动评测工具来计算。

表 14 展示了通过 MutNMT 计算的候选译文 A、B 和 D 的 ChrF3 值。

由计算结果再次得出, 候选译文 B 的优于候选译文 A 和候选译文 C。

¹⁹注意, 这也适用于 F: 到目前为止, 所有的 F 值都是 F_1 值, 也就是 $\beta = 1$ 。 β 的值可以更改。

表 14: 候选译文 A、B 和 D 的 ChrF3 值

评测指标	候选译文 A	候选译文 B	候选译文 D
ChrF3	64%	69%	49%

2.5.6 提请注意：自动评测指标的结果解读注意事项

在本章节末，我们想提请大家注意一点，那就是确保正确解读自动评测指标的结果。有些结果用小数或百分数来表示（如 0.8 或 80%）。在 0 到 1（或 0% 到 100%）的区间内，对于某个指标（如翻译错误率）来说，0 可能表示最佳，1 可能表示最差；而对于另一个指标（如 BLEU 值），1 可能表示最佳，0 可能表示最差。

此外，在对比不同工具得出的评测指标计算结果时，也应当谨慎，因为看似相同的自动评测指标的算法的不同之处，或许难以被非专业人士察觉。

最后，用户应该知道，若要好好利用各种评测指标，不仅要思考这些结果对其目的的意义，还要懂得如何解读。对比不同的自动评测指标，并将其与人工评测相结合会有所帮助，即使可能存在差异 (Doherty 2017: 134)。在处理非随机顺序呈现的语段时，人工评测对语境的把握会更好，例如，评测者能识别出代词错误，而自动评测指标通常只能基于单词和句子级别来评估选词或句子流畅度。还有一种有意思的方法，是把人工翻译编辑率（即译后编辑的工作量）与译后编辑的工作时长相结合。²⁰ O'Brien (2025 [本卷]) 简要介绍了如何测量译后编辑工作量。

2.6 类符形符比

和上述指标不同，类符与形符之比 (*type-token ratio*, *TTR*) 并非用于评测翻译质量，而是全面展现文本的词汇多样性。之所以在此介绍这种方法，是因为已有研究者用该指标对比机器译文与同语言的人工翻译或自然生成的文本，看两者之间的差异程度有多大（参见 Toral 2019）。MutNMT 等在线评测工具目前支持使用这一指标。

基本来说，类符形符比用于测量文本或语料库中的词汇多样性 (Williamson 2009)。文本的总词数指形符 (*token*) 的数量。但同一个词可能在文本中重复出现，例如，若某个单词在文本中出现三次，则计为 3 个形符，但只能计为 1 个类符 (*type*)。形符与类符之间的数量关系被称为类符形符比，计算方法如下：

$$\text{类符形符比} = \frac{\text{类符数}}{\text{形符数}} \tag{15}$$

²⁰虽然采用两者的平均值说明不了问题，但寻找相关性可能有助于识别最严重的错误。

或

$$\text{类符形符比} = \frac{\text{类符数}}{\text{形符数}} \cdot 100 \quad (16)$$

第一种计算方法的结果范围是 0 到 1, 而第二种方法的结果为百分比, 范围是 0% 到 100%。类符形符比越高, 则文本的词汇就越多样化。

不过, 关于类符形符比还要注意以下几点。首先, 这种指标对文本的长度非常敏感。其实, 文本篇幅越长, 限定词和冠词等的重复频率就越高。此外, 由于文本都围绕某个主题展开, 尤其是专业领域的文本, 因此相关术语会重复出现。这样一来, 篇幅越长, 其类符形符比就越低。所以, 为减弱长度对结果的影响, 使用者应根据任务将语料划分为固定长度 (如 1000 个形符) 的片段, 分别计算 TTR 再进行标准化。这种标准化的方式可用于对比机器译文语料库与不同长度的同一目标语文本语料库的类符形符比。

其次, 虽然在比较同一语言的文本时, 词形还原 (lemmatization) 并不重要, 但在使用类符形符比来比较两种或两种以上的语言时, 必须进行词形还原, 因为有些语言的屈折形态比其他语言的更丰富, 因此词汇更多样化, 而这不过是由于某个动词具有更多形态。不过, 如果只是简单使用标准化的类符形符比来比较机器译文和同一语言的其他文本的词汇多样性, 那就没有必要进行词形还原。

最后, 请记住, 类符形符比的数值越大 (即词汇更多样化), 并不等同于复杂性更高。例如, 对比一下 *The girl saw a fire.* (女孩看见了火。) 和 *The lexicographer observed the conflagration.* (词典编纂者观察到了这场大火。) 这两个句子。它们都包含 5 个单词 (形符), 但前者有 5 个类符, 而后者只有 4 个 (因为形符 *the* 出现了 2 次)。因此, 第一句的类符形符比为 1 (或 100%), 而第二句的则为 0.8 (或 80%)。可是, 虽然第二句的词汇不比第一句的多样化, 第二句却更复杂。²¹

如前所述, 比较句段级的类符形符比可能没多大意义, 但对于文档级或语料库级来说, 同一语言文本的标准化类符形符比的比较, 则可以为我们提供有用信息, 而这也要取决于所处理的文本类型。本章重点关注专业领域翻译。但不同的专业领域需要遵循不同的惯例约定。例如, 文学或营销领域翻译的词汇更多样 (即类符形符比更高) 则质量可能更好, 从而提高目标语文本读者的阅读愉悦度; 与此相反, 技术文本的翻译通常必须遵守某些惯例约定, 这往往倾向于降低词汇多样性, 以便于用户阅读使用。本章使用的主要例子来自用户手册, 也就是说, 它应该遵循惯例, 如一个术语只表示一个概念, 不得有所变化, 还应尽量遵循统一的写作范式。例如, 若 *transceiver* 时而译为 *émetteur-récepteur*, 时而译为 *radio* 和 *appareil*, 会导致类符形符比变高, 且对最终用户造成困惑。

以上就是希望读者能注意的方面。

²¹作者在此向 Dorothy Kenny 提出的看法致谢。

3 结语

在本章中，我们试图说明从语用维度来评测机器翻译对专业译者或翻译学习者的意义。之所以说语用维度，是因为这种方法不仅将评测当成实现目标的手段，还意味着要根据情况选择不同的方法，通常结合使用人工评测和自动评测。

虽然机器翻译输出的比较贯穿本章内容，但值得注意的是，在实际情况下，专业译者很少有机会选择使用评测指标。相反，他们通常需要迅速判断某个机器翻译解决方案是否适用，或者对机器翻译质量做出总体评价。

因此，我们介绍了如何综合使用人工评测和自动评测来评估机器翻译输出。本章之所以详细介绍了自动评测，是因为我们认为尽管存在局限性，但若理解得当并结合人工评测，自动评测就可以得到有效利用。

References

- Aziz, Wilker, Sheila Castilho & Lucia Specia. 2012. PET: A tool for post-editing and assessing machine translation. In *Proceedings of the eight international conference on language resources and evaluation (LREC'12)*, 3982–3987.
- Castilho, Sheila. 2020. On the same page? Comparing inter-annotator agreement in sentence and document level human machine translation evaluation. In *Proceedings of the 5th conference on machine translation (WMT)*, 1150–1159. <https://aclanthology.org/2020.wmt-1.137.pdf>.
- 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.
- Castilho, Sheila, Joss Moorkens, Federico Gaspari, Iacer Calixto, John Tinsley & Andy Way. 2017. Is neural machine translation the new state of the art? *The Prague Bulletin of Mathematical Linguistics* 108. 109–120. DOI: 10.1515/pralin-2017-0013.
- Castilho, Sheila & Sharon O'Brien. 2016. Evaluating the impact of light post-editing on usability. In *10th international conference on language resources and evaluation (LREC)*, 310–316. May 2016, Portorož, Slovenia. ELRA.
- Doherty, Stephen. 2017. Issues in human and automatic translation quality assessment. In Dorothy Kenny (ed.), *Human issues in translation technology*, 131–148. London: Routledge.
- Drugan, Joanna. 2013. *Quality in professional translation: Assessment and improvement*. London: Bloomsbury.

- Gouadec, Daniel. 2010. Quality in translation. In *Handbook of translation studies. Volume 1*, 270–275. John Benjamins Publishing Company.
- Grbić, Nadja. 2008. Constructing interpreting quality. *Interpreting* 10(2). 232–257.
- House, Juliane. 2015. *Translation quality assessment: Past and present*. London: Routledge.
- 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.
- Koehn, Philipp. 2010. *Statistical Machine Translation*. Cambridge: Cambridge University Press.
- Koehn, Philipp. 2020. *Neural Machine Translation*. Cambridge: Cambridge University Press.
- Mariana, Valerie, Troy Cox & Alan Melby. 2015. The multidimensional quality metric (MQM) framework: A new framework for translation quality assessment. *The Journal of Specialised Translation* 23. 137–161.
- Moorkens, Joss. 2018. What to expect from neural machine translation: A practical in-class translation evaluation exercise. *The Interpreter and Translator Trainer* 12(4). 375–387.
- 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.
- 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.
- Popović, Maja. 2015. Chrf: Character n-gram f-score for automatic MT evaluation. In *Proceedings of the tenth workshop on statistical machine translation*, 392–395. Association for Computational Linguistics. 10.18653/v1/W15-3049.
- Post, Matt. 2018. A call for clarity in reporting BLEU scores. In *Proceedings of the third conference on machine translation (WMT), Volume 1: Research papers*, 186–191. Association for Computational Linguistics. DOI: 10.18653/v1/W18-6319.
- Qin, Ying & Lucia Specia. 2015. Truly exploring multiple references for machine translation evaluation. In *Proceedings of the 18th annual conference of the European Association for Machine Translation*, 113–120. <https://aclanthology.org/W15-4915/>.
- Snover, Matthew, Bonnie Dorr, Rich Schwartz, Linnea Micciulla & John Makhoul. 2006. A study of translation edit rate with targeted human annotation. In *Proceedings of the 7th conference of the Association for Machine Translation in the Americas: Technical papers*, 223–231. Cambridge, Massachusetts: Association

- for Machine Translation in the Americas. <https://aclanthology.org/2006.amta-papers.25/>.
- Toral, Antonio. 2019. Post-editsese: An exacerbated translationese. In *Proceedings of machine translation summit XVII*, 273–281. EAMT. <https://www.aclweb.org/anthology/W19-6627/>.
- Williamson, Graham. 2009. *Type-token ratio*. Last retrieved 5 Dec. 2020. <https://www.sltinfo.com/wp-content/uploads/2014/01/type-token-ratio.pdf>.
- Yamada, Masaru. 2019. The impact of Google neural machine translation on post-editing by student translators. *The Journal of Specialised Translation* 31(2019). 87–106.

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。¹ 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。

¹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。² 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,³

²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。

³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。⁴ 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

⁴Seoane 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，DeepL⁵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

⁵<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。⁶ 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

⁶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.
- 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.
- 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.
- 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.

第五章

如何处理机器翻译的错误：译后编辑

Sharon O'Brien
都柏林城市大学

机器译文可能出现需要修改的错误，尤其是如果译文用于发布或绝对不能出错。识别并修改这些错误的工作就称为“译后编辑”（post-editing, PE）。我会在本章中借助学术界和业界资料，概述译后编辑的过程，还会解释它大致分为“轻度”译后编辑和“完整”译后编辑，并分别介绍对应的操作指南，同时还会关注应用中出现的问题。本章还概述了译后编辑使用的各类界面（包括文字处理软件、电子表格软件和专业的计算机辅助翻译工具），以及交互模式（传统、自适应或交互式）。最后介绍译后编辑研究者所使用的概念和工具，并重点关注如何测量时间、技术和认知努力。

1 定义

机器翻译技术还不完善，有时能产出意思准确、符合语境的译文，有时却出现误译、漏译、增译或文体不当等严重错误。如果使用机器翻译只为了理解文本大意，或许无需修改这些错误。但是，如果要使用机器翻译来生成可供出版或在组织内外广泛传播的译文，通常则需要修改文中的错误。这种错误的识别和改正被称为“译后编辑”（post-editing）。在机器翻译系统开发的早期阶段，就已经有人开始使用这个术语。比起现在，当时的技术发展更落后，肯定也也不能随用随取，即时输出译文。文本先以电子版发送到机器翻译系统进行翻译，然后将译文返回给发送者，再由发送者对自动翻译好的文本进行编辑。“译后编辑”这一术语一直沿用至今，甚至也用来指代修改机器译文错误的过程（参见节 3.1）。

译后编辑是双语处理任务，通常由经验丰富的专业译者来完成。担任此工作的人常被称为“译后编辑者”（post-editor）。然而，这个职业名词还存在争议，尤其是因为专业译者的日常工作也会使用计算机辅助翻译工具，这类工具结合了翻译记忆库、术语管理和机器翻译等技术（参见下文的Kenny 2025



[本卷]: §4 和 节 3)。采用这种工作模式的译者可能会先处理某个句子的模糊匹配, 然后翻译下一句, 再接着对下一句进行译后编辑。在这个过程中, 译者的角色是不是从审校, 变成译者, 再变成了译后编辑者? 其实不然! 因为这项工作从根本上来说, 仍然是翻译和审校。主要区别在于译者在处理不同的句子时, 使用了不同的技术支持和输入来源。

进行译后编辑时, 译者必须同时理解原文和机器译文, 然后识别错误, 思考并制定校正策略, 最后进行修改。从根本上说, 译后编辑就是审校工作。

机器译文可能会出现各类错误, 如语法错误、句法错误、非必要增译漏译、词汇错误、术语不符、搭配错误或风格不当等。机器译文的错误类型和数量受很多因素影响, 如语言对和翻译方向、内容类型以及机器翻译引擎的训练数据或技术。

以下以英文的机器译文及其修改版本为例介绍几种错误类型。

(1) 语法错误

- a. The cat is very protective of her kittens. She scratches anyone *which* tries to touch them.
- b. The cat is very protective of her kittens. She scratches anyone *who* tries to touch them.

(2) 词汇错误

- a. The cat is very protective of her *pups*. She scratches anyone who tries to touch them.
- b. The cat is very protective of her *kittens*. She scratches anyone who tries to touch them.

(3) 句法错误

- a. The new-born cygnets on the lake *swam*.
- b. The new-born cygnets *swam* on the lake.

(4) 搭配错误

- a. The house had no *flowing water*.
- b. The house had no *running water*.

2 译后编辑的不同程度和指南

机器翻译的主要目标之一是, 尽可能在提高效率和降低成本的同时, 实现更多语言之间的信息转换。以这些目标为导向, 译后编辑主要分为“轻度译后编辑”或“完整译后编辑”。

进行轻度译后编辑只需修改关键错误，追求速度；而完整译后编辑则需要修改机器译文的所有错误，比轻度译后编辑更耗时。但无论是哪种程度的译后编辑翻译，译后编辑的译文产出速度都应该比不借助计算机辅助翻译工具的翻译速度更快。国际标准化组织（ISO）制定了一份译后编辑标准（ISO 2017），编号为 ISO18857:2017。根据该标准，轻度译后编辑被定义为“只需产出意思正确的文本，无需追求可比拟人工翻译的译后编辑过程”（第 2 页）。完整译后编辑被定义为“以产出可比拟人工翻译的译文为目标的译后编辑过程”（同上）。

然而，从概念上讲，这些定义是有问题的。首先，除了以上非常笼统的描述方式，我们很难明确两者的区别。到底什么是“关键”错误？这个问题视需求因人而异。“意思正确”是什么意思，如何衡量？轻度译后编辑真的能够不以人工翻译的质量为对标吗？此外，很难算清深度译后编辑要比轻度译后编辑所需时间差值。这些问题促使翻译行业更规范地描述不同程度的译后编辑。例如，翻译自动化用户协会（TAUS）制定了操作指南，提出完整译后编辑包括文体层面的修改，而轻度译后编辑则不然。

表 1 列出了翻译自动化用户协会关于轻度和完整译后编辑的操作指南（TAUS 2010），其中并列内容可进行比较，空白单元格表示该套指南中无可比较内容：

根据 ISO18857 号标准（第 6 页），译后编辑的目标是确保：

- 经译后编辑的译文可以理解；
- 源语和目标语的内容相符合；
- 遵从 TSP 对译后编辑的要求。

TSP 指“翻译服务提供商”（translation service provider），意为“提供翻译服务的语言服务提供商”（第 4 页）。

要实现这些目标，译后编辑需遵从以下标准（第 6 页）：

- 术语/词汇必须保持一致，且符合专业领域的术语表达；
- 使用目标语的标准句法、拼写、标点符号、音调符号、特殊符号和缩写等拼写惯例；
- 遵守任何适用标准；
- 格式正确；
- 适用于目标受众和目标语内容；
- 遵从客户与 TSP 之间的协议。

表 1: TAUS 译后编辑操作指南

轻度译后编辑	完整译后编辑
确保译文的语义正确。	确保译文的语法、句法和语义正确。 确保正确翻译关键术语，且未翻译的术语包含在客户的“无需翻译”(Do Not Translate) 术语列表中。
确保无增译或漏译信息。	确保无增译或漏译信息。
修改所有含冒犯之意或文化不容的不当内容。	修改所有含冒犯之意或文化不容的不当内容。
尽量保留机器译文。	尽量保留机器译文。
遵从基本的拼写规则。	遵从基本的拼写、标点符号和连字符规则。
无需修改文体问题。	
无需仅为了提高文本流畅度而重组句子。	
	确保格式正确。

这两套指南虽然侧重点不同，但有部分内容重叠，共同组成了译后编辑的常用指南。TAUS 指南鼓励尽量保留机器译文，而 ISO 指南更关注协议、标准和对目标受众的适用性。尽量保留机器译文是译后编辑的重要方面。译者往往很容易忽视机器译文，直接弃用并重新翻译。其实，很多译者都有这么做的倾向，认为自己自己翻得更好更快，而且比译后编辑更高效。虽然在“更好”的问题上，过去机器译文质量的确不尽如人意，但神经机器翻译的发展已经大大提高了机器翻译的整体质量，使其更加便于使用。而在“更快”的问题上，人工翻译是否比译后编辑用时更少，仍未有定音。即使译者持相反看法，但有研究表明译后编辑肯定比人工翻译更快（如Guerberof Arenas 2014）。我们可以通过实践找到如何快速评测机器译文质量，从而决定其可用性和所需修改的方法。

从概念上讲，译后编辑程度与质量水平相关联，尽管下文例子说明这种联系存在问题。经过轻度译后编辑的译文“质量恰好”或“可以理解”，也就是说，要准确传达原义，但不要求通顺流畅或文体使用考究。另一方面，经过

完整译后编辑的译文“质量堪比人工翻译”。但问题又来了，这种观点的内在逻辑是人工翻译的质量总是很高，可坦白讲，实际情况并非如此。

译后编辑的程度和译文质量水平的关系有时关系复杂，让人捉摸不透，为了便于理解，我们来看个例子：

- (5) a. In a new report on the quality of teaching practice, inspectors said pre-school teachers could be *training* to integrate languages such as French, German and Polish in early *learn* settings.
 b. In a new report on the quality of teaching practice, inspectors said pre-school teachers could be *trained* to integrate languages such as French, German and Polish in early *learning* settings.

快速改正例句(a)的两处错误后得到例句(b)。一方面，我们可以说对例句(a)进行了“轻度”编辑，因为只涉及两处快速修改。但如果根据轻度译后编辑指南的建议，我们无需进行任何改动。另一方面，按照完整译后编辑指南的要求，为了确保译文的语义、句法和语法正确，这两处错误必须改正。这么说来，这两处简单的错误修改，究竟是轻度译后编辑，还是完整译后编辑呢？再看另一个例子。

- (6) a. In addition, it *reiterates the instinct* of modern *language* into the primary school curriculum *as a fitting step*.
 b. In addition, it says *re-instating modern languages* to the primary school curriculum *would be a timely move*.¹

例句(a)表意模糊，需要进行大幅修改才能至少准确表达原义。从根本上说，译后编辑的程度视机器译文（即机器翻译系统产出的每句译文质量）和目标质量而定。译后编辑者会根据每句译文和目标质量，在简单编辑和深入编辑之间切换。无论每一句质量如何如何，最重要的是终稿的质量一致，且符合翻译委托人和最终用户的期望。

轻度译后编辑和完整译后编辑的最后一个概念问题是，专业译者通常不愿意只提供“易于理解”的译文质量，也没有翻译专员愿意承认自己只会选择轻度译后编辑。

对这些概念及其指南的了解非常重要。显然，这些方面都还面临一些挑战。

3 译后编辑的操作界面

基本上，只要能在阅读原文的同时编辑机器译文，译后编辑就可以借助任何文本编辑器完成，甚至只需通过能并排显示原文和译文的电子表格就能进

¹<https://www.irishtimes.com/news/education/foreign-languages-could-be-taught-in-preschool-and-primary-department-1.4270886>

行编辑。不过目前专业译者常用计算机辅助翻译 (CAT) 工具, 尤其是翻译记忆库 (TM)。Kenny (2025 [本卷]: §4) 指出, 翻译记忆库是个存储已翻译过的文本句段的数据库。翻译记忆库软件是用于访问、编辑和更新该数据库中文本的软件应用程序。从本书其他章节可以看出, 翻译技术和机器翻译显然是不同的技术, 尽管二者关系密切, 因为当代数据驱动的机器翻译系统常用存储于翻译记忆库的数据作为机器学习的重要输入。此外, 由于译者普遍使用翻译记忆库, 机器翻译技术就算还没完全嵌入, 也已经与很多翻译记忆库工具相关联, 如 Trados Studio、MQM、MateCat 等等。从实践的角度来看, 这意味着译后编辑工作常常在翻译记忆库编辑环境中完成。

机器翻译可以用不同方式嵌入翻译记忆库环境。例如, 如果记忆库中没有可用的匹配句段, 那么译者就可以调用机器翻译系统。或者, 译者可以自定义翻译记忆库设置, 在在特定条件下 (如无法在记忆库中找到匹配句段), 则自动显示机器译文。

大家普遍认为, 如果翻译记忆库维护良好并持续更新, 其数据库提供的精确匹配比机器翻译系统对译者更有价值, 因为后者可能包含错误, 而前者 (理想情况下) 则不然。同理, 以往的业内共识是, 相似度为 75% 以上的翻译记忆库“模糊匹配”要优于机器翻译。然而, 近来的机器翻译技术更加先进, 生成的译文可用性甚至大于相似度超过 75% 的翻译记忆模糊匹配, 更加便于译者使用。因此, 译者可以根据翻译记忆库、机器翻译系统、语言对和文本主题来自定义翻译记忆库的软件设置, 以便同时看到来自翻译记忆库的匹配结果和机器译文。其实, 译者还可以同时获得多个机器翻译系统的译文, 择优选用, 然后根据需要进行编辑。

结合使用翻译记忆库和机器翻译的好处是译者在翻译时可以获得更多的选择和帮助。但是, 这种工作模式也不乏挑战。这样的设置将大量信息呈现给译者, 意味着他们必须处理所有这些信息 (记忆库匹配及其匹配度、机器翻译建议, 可能还有术语和元数据, 如记忆库匹配的创建者和创建时间等 (参见Teixeira & O'Brien 2017)), 同时从语言的角度决定使用哪个匹配以及需要如何修改。这可能导致很高甚至超负荷的认知需求, 或许正因如此, 才会有译者表示译后编辑比其他形式的翻译和修改工作的要求更高且更耗费精力。

这种混合型交互界面带来的另一个挑战就是前文提到的译后编辑指南。译者在处理机器译文时需要牢记这些指南, 但如果处理的是记忆库的匹配结果, 那么他们编辑的则是 (与机器翻译系统相对的) 人工翻译, 因此不能算作真正的“译后编辑”。这种结合了译后编辑、审校和翻译的工作模式可能会使翻译任务变得相当复杂, 更别提译者总是顶着时间压力执行翻译任务!

3.1 传统、自适应和交互式译后编辑

在上文中，我们已经概述了译后编辑的常用交互界面，在传统翻译工作环境中，译者使用翻译记忆工具的惯用翻译设置，机器翻译匹配静态显示，供选择编辑或弃用，以支持翻译记忆库匹配或原文句子的完整翻译。与所有技术一样，新发明的出现——自适应机器翻译和交互式机器翻译——使传统的默认设置有了变化。

“自适应 MT” (Adaptive MT) 是翻译记忆库/机器翻译集成者开发的一项功能。有了这项功能，机器翻译系统可以实时学习译者的编辑结果。这解决了一个曾让译者为之苦恼的机器翻译缺陷，即系统无法学习译者已经修改过的机器译文错误。因此，如果同一句话再次出现，译后编辑者必须再修改一次。结合翻译记忆库能在一定程度上解决这个问题——在记忆库环境中修改的某个句子会被保存到记忆库的数据库中。如果同一句话再次出现，译者看到的会是这句话在翻译记忆库的数据库中的精确匹配。但是，如果没有使用精确匹配的数据库的翻译记忆库环境，那么机器翻译系统就会出现同样的错误。使用自适应机器翻译系统能“学习”修改结果，并且理论上应该能自行适应译者的修改，以确保不再重蹈覆辙。Trados Studio 这款工具便很好地发挥了这项功能。

“交互式机器翻译” (Interactive MT) 可以看作是自适应机器翻译的特殊形式。这种形式与机器翻译的默认结合方式如上所述——机器翻译系统先预翻译整个句段，然后完整呈现给译者，而交互式机器翻译则会根据译者所选的每个词或短语调整后续输出。当译者接受或确认一个词时，机器翻译系统会即时调整输出。这种与机器翻译输出交互的方式独具特色，不过，它类似许多人已经习以为常的预测文本的概念。Lilt 平台便以交互式机器翻译为主要特点。其实，Lilt 的定位就是自适应和交互式机器翻译系统界面。²（其他交互式机器翻译的例子可参见 Torregrosa Rivero 2018。）

交互式机器翻译令人对“译后编辑”提出了质疑。如前所述，“译后编辑”一词出现在几十年前，当时是机器翻译系统翻译整篇文本，译者进行编辑，再返回给翻译委托人/客户。但是，在交互模式下，机器翻译随着译者当下的决定实时产出译文。因此，编辑似乎并不发生在翻译之“后”，称之为“交互式机器翻译”更加准确。和许多领域一样，“译后编辑”一词已普遍为人接受，所以不会短时间内被弃用，但或许会逐渐消失。而与机器译文交互并修改其错误这项工作在不久的将来消失的可能性很小。

3.2 研究界面

上文介绍了不同的译后编辑主流界面以及交互模式。还有很多其他可供使用的界面，但功能基本一致。由于译后编辑对专业译者来说是一项相对较新

²<https://lilt.com>, 最后访问于 2022 年 6 月

的任务，因此相关学术研究方兴未艾。研究主题涉及译后编辑的常见类型、是否比翻译省时、产品质量及其认知过程。

为了收集以上主题的数据，有研究人员自行开发了译后编辑界面。这样做主要是因为，对于研究项目来说，商业工具可能过于昂贵且功能相对复杂，有时反而阻碍研究目标的实现，又或者这些工具不利于控制实验条件。我们要介绍的两款为译后编辑实验开发的研究工具分别是 Casmacat 和 Translog。Casmacat 是一项欧盟资助研究项目的成果。³ 该项目旨在使用交互式自适应机器翻译，搭建下一代译者的工作平台，并建立交互式机器翻译的认知过程模型。Translog 最初由哥本哈根商学院为研究翻译过程而开发。Translog II 为最新版本，支持对翻译和译后编辑的研究。⁴ 这两款工具都可以结合其他有利于研究的技术，如键盘记录（记录键盘使用）和眼动追踪（记录翻译或译后编辑过程中的眼球运动和认知负荷）。这些工具的使用有助于深入理解译后编辑这项任务。不过，这些工具虽然适用于科研目的，但其功能与标准的商业级翻译记忆库工具一比较就相形见绌了。因此，它们无法完整呈现真实的编辑工作过程。

4 译后编辑工作量的测量

技术取得进展后，会从实验室走向公共领域，以检验其有用性。最初，新技术可能会是破坏因素——多年来使用和所接受的流程被打乱，因此引发质疑、担忧或恼怒，这是可以理解的。当习惯了某种工作方式，我们会发现很难快速接受新的方式。而且，如果新技术带来新问题，我们就不太可能完全接受它。从职业角度来看，机器翻译和译后编辑便是如此。因此，人们对译后编辑工作量的研究热度大增，以证实或反驳“译后编辑比人工翻译更高效”的说法。

约十五年以来，致力于探讨这一问题 (Koponen 2016) 的学术研究数不胜数，在此无法一一列举。但是，有一篇开创性文章非常值得一提，因为它为译后编辑工作量的测量设定了标准。该文章作者是汉斯-彼得·克林斯 (Hans-Peter Krings)，原文为德语，其英文版于 2001 年出版 (Krings 2001)。克林斯对译后编辑和人工翻译的工作量进行了对比研究。他在进行这项研究时，机器译文的质量远比现在的差得多，而且当时的实验环境非常简易——实验任务用纸笔完成，克林斯用摄像机记录整个过程。不过，克林斯这项研究的重要性并不体现在实验条件或研究发现，而在于他提出了包含三个维度的译后编辑工作量衡量法——时间、技术和认知。

³受到欧盟第七框架计划项目 287576(ICT-2011.4.2, Seventh Framework Programme Project) 的共同资助-<http://www.casmacat.eu>

⁴<https://sites.google.com/site/centretranslationinnovation/translog-ii?authuser=0>

译后编辑工作量的测量通常只关注时间维度，即这项任务对比其他任务所花费的时长。时间维度当然是最重要的维度之一，尤其是在商业环境中。而且时间维度也相对容易测量，因此当商业机构在测量译后编辑工作量时主要考量时间维度。

然而，克林斯的研究表明，还需要考虑其他两个维度。技术维度测量键盘和鼠标的操作，即删除或添加的词或字符数量，选择的短语数量，文本剪切和粘贴的词数等。前文提到的 Translog 就是支持键盘日志记录的工具。译后编辑就是文本修订，因此了解技术操作的工作量非常重要。不仅如此，译后编辑这种审校工作模式涉及删除、重新键入、复制和粘贴等操作，需要大量使用键盘和鼠标，使身体产生疲惫感，甚至导致手和手腕劳损。另一方面，如果机器译文质量相对较好，就可以减少译者的打字量。

除了键盘记录，技术维度也可以通过“编辑距离”(edit distance) 衡量。简单来说，编辑距离计算将文本的字符串变成另一个字符串所需的最少操作步数。这种“操作”包括词的删除、插入或位置移动。编辑距离的测量指标分为几种，计算操作步数的方法略有不同。基础的测量指标之一是“莱文斯坦距离”(Levenshtein distance)，计算不同单词、短语或句子之间转换所需的字符插入、删除或替换的最少次数。

以单词 drink 为例，要改变多少个字符才能将其变成 drunk？答案是一个——只需将 i 替换成 u。举个更复杂一点的例子，把短语 He drinks 变成 He is drinking 的莱文斯坦距离是 6（在 He 后插入字母 i 和 s，以及一个空格符，然后将 drinks 结尾的 s 换成 i，并插入 n 和 g）⁵我们还可以采用更复杂的编辑距离指标，其中常用于测量译后编辑的编辑距离的指标是“翻译编辑率”(Translation Edit Rate, TER)(Snover et al. 2006)。该指标的结果范围是 0-1 或 0% 到 100%，分数越低，则译后编辑工作量就越少，如结果为 30% 表示需要对机器译文内容的 30% 进行编辑才能获得译后编辑版本对应的字符串。如何更好地计算编辑距离仍面临挑战，因此存在几种不同的方法，经常使用不同的指标。

时间和技术这两个维度相对容易测量，而第三个维度——认知努力——的测量要复杂得多。认知努力是隐藏的认知过程，包括阅读、理解、比较原文和机器译文的意思、决策，同时还要考虑编辑指南和客户的期待，边修改边在编辑的同时监控文本质量。这些过程发生在大脑中，无法直接观察或测量。尽管如此，认知努力仍然是需要考虑的重要方面。有人说，译后编辑有时比人工翻译的认知负荷更高。这可能是因为译后编辑要顾及上述提及的一系列认知过程，且有些人对此还不太熟悉。即便译者借助机器翻译可以提高工作效率，但译者可能会觉得译后编辑比手动翻译更让人疲惫。提高工作效率虽然符合商业生产逻辑，但是造成译者过度疲劳的话未必是好事。正因如此，为什么在测量译后编辑工作量时，认知努力是重要的考虑因素。

⁵可以使用在线计算工具，如<https://planetcalc.com/1721/>。

但是,如何测量认知努力呢?其实,这是所有希望量化认知努力的人都想回答的问题。有时,认知努力可以通过执行者在完成任务时的“有声思维”来估算。这种方法能让执行者指出自己的认知困难。当然,在过程中进行有声思维会对任务造成干扰且降低速度,这也是该方法的缺点。第二种方法是对任务过程进行电脑录屏,待完成后让执行者边观看回放视频,边回想当时遇到的问题。这样做的好处是不会影响完成速度,但缺点是执行者可能想不起所有问题。第三种方法是研究者试图采用眼动追踪 (eye tracking) 技术来测量译后编辑的认知努力。这项技术用于记录眼睛注视屏幕的位置,并记录眼睛在文本上停留的时间(称为“注视时间”),甚至测量“瞳孔放大”(即测量瞳孔大小)。这几种都是公认的认知努力测量的有效方法。但是,这些方法也面临许多挑战,如眼动设备花费昂贵,具体操作和数据分析复杂;还要控制好数据收集环境,避免受试者过度频繁地移动头部,或者确保光线不会有大幅变化,以免引起瞳孔大小改变等。认知努力的测量困难重重,因此不难理解为什么很少有人从这个维度来测量译后编辑的工作量。尽管如此,认识到认知努力对译后编辑工作必不可少。

关于译后编辑工作量,还有最后一点要补充。译后编辑工作量可以间接体现某个机器翻译系统针对某个语言对或主题而产出的译文质量。因此,译后编辑工作量可以用作机器翻译质量测评。机器翻译系统的质量越低,则译后编辑所需操作和时间就越多。机器翻译质量还可以用其他方法来测量,如通过识别、分类和计算错误数量,这些方法虽然行之有效,但译后编辑工作量能提供的信息或许更多,因为它反映了想要在机器翻译的基础上获得达到质量要求的译文的难易程度。

5 译后编辑者的资质与培训

优秀的译后编辑者应该具备什么特质?应为其提供哪些培训?随着机器翻译逐渐发展成主流技术,这两个问题一直困扰着语言行业和学术界(参见Nitzke & Hansen-Schirra 2021)。

关于第一个问题,普遍的建议是,优秀的译后编辑者首先应该是优秀的译者。我们凭直觉就知道,有些人做得好翻译,但做不好审校,反之亦然。同理可得,并非人人都能胜任译后编辑的工作。

但怎样才能称作优秀的译后编辑者?这个问题已经引起了一些关注。例如,de Almeida & O'Brien (2010)认为优秀的译后编辑者应该能够:

1. 识别并修改机器译文的问题;
2. 以合理的速度进行后期编辑的能力,以满足这类工作的日产量要求;
3. 遵从指南准则,尽量减少“偏好性”或严格来说不必要的修改,且修改要保持在正常的译后编辑范围内。

具体来看, 第(1)点要求译后编辑者首先必须掌握翻译技能。第(2)点和第(3)点表明, 译后编辑者需要能够高效工作并遵从指南准则, 避免过度编辑。归根结底, 能否成为“优秀的”译后编辑者与其对机器翻译技术的态度密切相关(深入讨论可参见Guerberof Arenas 2013)。如果对机器翻译技术很反感, 那译者可能会删除或直接忽视机器译文。这反过来会导致翻译任务更加耗时, 对委托人造成更高的成本投入。这样的译者就不能算是“优秀的”译后编辑者。不过, 这在很大程度上也要取决于具体情况和机器翻译质量。

现在我们来讨论第二个问题。在过去十年里, 人们越来越重视培训工作。随着机器翻译逐步融入其他计算机辅助翻译工具和翻译流程, 专业译者需要继续接受职业发展培训, 如机器翻译和译后编辑的专题工作坊。机器翻译和译后编辑已被纳入大学的翻译教学计划, 但教学方法各有不同。有些大学单独开设译后编辑课, 有些大学将译后编辑嵌入审校课程, 还有些大学则是将其并入翻译技术课(参见O'Brien & Vázquez 2019)。

培训的核心重点都是确保翻译专业学生了解机器翻译的最新方法、优势和局限、如何评测其质量以及如何进行译后编辑。无论是对于翻译专业学生, 还是对于没有接受过翻译培训的人, 培训的重点都在于了解何时以及如何使用机器翻译(关于“机器翻译素养”的讨论可参见Bowker & Ciro (2019))。

References

- Bowker, Lynne & Jairo Buitrago Ciro. 2019. *Machine translation and global research*. Bingley: Emerald Publishing.
- de Almeida, Gisele & Sharon O'Brien. 2010. Analysing post-editing performance. Correlations with years of translation experience. In *Eamt 2010. Proceedings of the 14th annual conference of the European Association for Machine Translation*. EAMT. <http://www.mt-archive.info/10/EAMT-2010-Almeida.pdf>.
- Guerberof Arenas, Ana. 2013. What do professional translators think about post-editing? *Journal of Specialised Translation* 19. 75–95. https://www.jostrans.org/issue19/art_guerberof.php.
- Guerberof Arenas, Ana. 2014. Correlations between productivity and quality when post-editing in a professional context. *Machine Translation* 28. 165–186.
- ISO. 2017. *ISO 18857:2017. Translation services – post-editing of machine translation output: Requirements*. <https://www.iso.org/standard/62970.html>.
- 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.
- Koponen, Maarit. 2016. Is machine translation post-editing worth the effort? A survey of research into post-editing and effort. *The Journal of Specialised Translation* 25. 131–148. https://www.jostrans.org/issue25/art_koponen.pdf.

- Krings, Hans P. 2001. *Repairing texts: Empirical investigations of machine translation post-editing processes*. Ohio: Kent State University Press.
- Nitzke, Jean & Silvia Hansen-Schirra. 2021. *A short guide to post-editing*. (Translation and Multilingual Natural Language Processing 16). Berlin: Language Science Press. DOI: 10.5281/zenodo.5646896.
- O'Brien, Sharon & Silvia Rodríguez Vázquez. 2019. Translation and technology. In Sara Laviosa & Maria González-Davies (eds.), *Routledge handbook of translation and education*, 264–277. London: Routledge.
- Snover, Matthew, Bonnie Dorr, Rich Schwartz, Linnea Micciulla & John Makhoul. 2006. A study of translation edit rate with targeted human annotation. In *Proceedings of the 7th conference of the Association for Machine Translation in the Americas: Technical papers*, 223–231. Cambridge, Massachusetts: Association for Machine Translation in the Americas. <https://aclanthology.org/2006.amta-papers.25/>.
- TAUS. 2010. *MT post-editing guidelines*. <https://www.taus.net/academy/best-practices/postedit-best-practices/machine-translation-post-editing-guidelines>.
- Teixeira, Carlos S. C. & Sharon O'Brien. 2017. Investigating the cognitive ergonomic aspects of translation tools in a workplace setting. *Translation Spaces* 6(1). 79–103.
- Torregrosa Rivero, Daniel. 2018. *Black-box interactive translation prediction*. PhD Thesis. Alicante: University of Alicante. https://rua.ua.es/dspace/bitstream/10045/77110/1/thesis_daniel_torregrosa_rivero.pdf.

第六章

伦理道德与机器翻译

Joss Moorkens
都柏林城市大学

神经机器翻译比以往的机器翻译范式更利于促进交流，但也会导致一些后果。和任何技术的发展一样，机器翻译并非道德中立，而是反映了其背后开发者的价值观。本章围绕机器翻译的伦理问题，以数据的收集和重复使用为切入点，探究机器翻译如何才能契合译者的价值观和准则。若能反映价值体系，机器和系统能明确地“好”，并消除译文中的偏见吗？机器翻译对可持续性和多样性的讨论有何贡献？本章的目的不是推广不假思索地遵循指令来运用技术的方法，而是强调在设计数据驱动的机器翻译工作流程时，有意识的决策过程的重要性。

1 何为伦理规范？

伦理学领域研究道德规范、善恶对错，并回答如何才能更好地生活。对这类主题的讨论最早可以追溯到埃及、巴比伦和印度。苏格拉底等希腊哲学家引入了“美好生活”（good life）的概念，这是一种有价值且令人向往的生活。亚里士多德提出了一套具化这种生活的道德准则，人类如果遵照这些准则，就能繁荣发展。不过，这些准则还是比较抽象，不总能帮助我们判断行为的对错。之后的哲学家和伦理学家也提出了一些判断行为是否正确或合乎道德的方法，例如，是否能为大多数人谋福利或只出于良好纯粹的动机。

而问题在于，对某个群体来说动机善良或带来益处的事情，并不一定能给另一个群体带来同样的正面效果。对此，两派观点争执不下。有人认为，行动可以是普遍有益且合乎道德的，如维护正义或诚实，但还有人认为，价值观可能因个人或群体而异，这两种观点之间存在矛盾。关于如何根据主体性、关系或周边叙述来理清某个行为是否合乎道德，已经有很多这方面的建议。理论伦理学就此迈入应用伦理学，试图为我们提供具体场景的行动指导。

在职场领域，应用伦理学通常包括指导职业行为的准则或标准。这些准则如果过于严格，可能阻碍潜在发展或社会效益。僵化的准则还可能导致陷入



困境，因为道德决策很少是非黑即白的，选择可能受到决策者所面临的特殊情况和压力的影响。正因如此，不同领域的应用伦理学便应运而生，根据特定职业场景来思考共同的问题和困境。本章将围绕与机器翻译关系最密切的领域，如计算机和信息伦理、数据伦理，讨论人类在系统开发过程中，如何合乎道德地使用机器翻译。本章第三节节 3 借鉴商业伦理和越来越多的翻译伦理相关研究，阐述了如何在翻译工作流程中合乎道德地使用机器翻译。第四节节 4 借鉴机器伦理和计算机与信息伦理，探讨计算机如何成为伦理主体。而最后几个小节则借鉴各项最新相关研究，聚焦伦理和人工智能的可持续性和多样性。

技术逐渐成为人类生活必不可少的一部分，且计算机的运用也进入到很多领域，因此技术领域的道德准则也受到越来越多的关注。在设计、开发或使用技术的过程中，我们需要意识到自己所做选择的影响。常有人认为，技术是道德中立的，只在人类使用该技术时才会产生偏见。然而，伦理学家和科学哲学家都一致认为，技术并非道德中立，而是反映了设计者的价值观。这些价值会决定技术要解决的问题、创造技术的决策、落实方法、预期用户、所使用的参考或训练数据、数据处理、数据存储的地点和安全性，以及可能基于成本或地理位置的技术获取限制。

科技发展之快且规模之大，都让监管措施无可避免地稍有滞后，因此，我们只好依靠工程师和开发商能遵照道德准则行事。我们也或多或少依赖有政治影响力和财力的大型科技公司能以人类集体的最大利益来行事。但近年来的一系列报告和披露表明，这些科技巨头辜负了公众的信任。技术虽然有助于开辟途径和增加收益，但也会导致公众面临风险。本章将讨论机器翻译系统开发和使用过程中的选择和风险，旨在指引用户做出明智且合乎伦理道德的决定。贯穿始终的重点主要（但不完全）是以传播为目的的机器翻译，即机器译文并非生产流程的最后一步。

2 机器翻译系统开发：人类合乎伦理道德的机器翻译使用

2.1 数据使用的案例研究

本节着重探讨机器翻译系统开发过程中，翻译数据使用的法律和道德问题。正如Pérez-Ortiz et al. (2025 [本卷])所证，数据驱动的机器翻译，尤其是神经机器翻译，需要大量数据进行训练。¹ 尽管重复使用翻译数据来训练机器翻译系统完全合法，但这是否合乎道德呢？思考以下例子，或许有利于理解本节要讨论的问题。

¹ “数据”指任何形式的记录信息，通常以数字形式存储。我们将可大量获取并规模处理的数据，称为“大数据”。

译者 A 自愿与经常合作的客户签署了一份合同,以自由职业者的身份,在专门的网络平台完成翻译任务,并明确准许其翻译数据可用于训练机器翻译系统。该雇主使用译者 A 等人的译文来训练机器翻译系统。随着时间的推移,译者 A 的语言对的神经机器翻译系统质量得到了提高,于是,该公司将翻译任务改为译后编辑,并单方面把按词计算的翻译费用降低了 30%。之所以按此幅度下调,是因为该公司通过平台收集到的翻译活动数据获知,翻译效率提高了约 30%。为了增加收入,该公司决定对外出售机器翻译服务,其中还包括为武器制造商提供的翻译内容。

译者 B 出于个人原则而反对机器翻译。该译者的合作公司希望译者提交包含目标文本的翻译记忆库,以便日后供其他译者使用。但是,译者 B 并不知道自己的译文已经由自动化项目管理系统自动分配,既没有翻译任务说明,也没有与公司直接沟通。此外,译者 B 也不知道,该公司不久后会被一家大型企业收购,该企业会用所有手头数据来训练机器翻译系统,然后出售。数据在共享前应删除所有个人信息(参见节 2.5),但当数据传给某个买家时,这些信息都意外保留了。该公司试图不做声张,以逃避责任。

就以上两个案例,你能从中发现哪些道德问题?如果雇主或译者做出不同的道德决定,会出现什么变化?我们会在下文探讨数据的所有权、使用权限、流通、隐私,以及数据共享的法律框架。希望这些内容能引导你对上述问题的思考。

2.2 数据的所有权

数据常被比作石油,让人误以为大数据是自然形成的,而实际上,它是由人类创造而成。近年来,数据呈指数级增长,这意味着如今无论是可用于训练机器翻译系统的数据,还是对翻译的需求,都比以往任何时候都多,远远超过了人工翻译的生产能力。机器翻译训练数据通常以人工翻译的平行或对齐的双语文本语段的形式,存储在翻译记忆库中(尽管机器翻译译文有时也用于系统训练)。这些翻译数据可能来源于公共数据库(如欧盟委员会总司翻译数据,该数据库可“出于商业或非商业目的,免费复制和传播……”,可参见Steinberger et al. 2012: 457)、个人的翻译数据库或从网上爬取的平行数据。

1886 年首次颁布的《伯尔尼公约》为翻译版权提供了法律基础。根据《公约》,译文属于衍生作品,“应作为原创作品受到保护,但不损害原创作品的版权”(World Intellectual Property Organization 1979 第二条)。《公约》赋予原创作品的作者授权翻译的专有权,但并没有清晰界定何为“原创作品”或“原创性”,导致不同管辖权对此有不同的解释。Troussel & Debussche (2014)认为,“创造性翻译”可视为具有原创性,不过这还有待法庭检验。他们还认为,根据欧洲数据库指令,如果“对内容的获取、查证或呈现进行了大量投资”,就可以主张对翻译记忆库的所有权(European Parliament 1996 第七条)。但在

实践中,无论译者是否在合同中约定放弃记忆库所有权,翻译记忆库通常都会发送给客户。

在大规模范围内,翻译大数据已经成为机器翻译和机器学习系统训练的宝贵资源 (Moorkens & Lewis 2019a)。然而,这并不意味着要降低译者的报酬,而且机器翻译训练中的数据细粒化重复使用意味着,训练数据通常无法溯源。通过网页爬取收集平行文本数据也是如此。在我们的案例研究中,译者 A 和 B 除了交出自己的翻译数据并接受随之而来的后果之外,别无选择,特别是大多数译者都是自由职业者,没有太多雇主争论的余地。我们有理由认为,更公平的数据所有权制度会有助于翻译行业的可持续发展(还可参见节 5)。

2.3 数据的使用权限

在一些司法管辖区域,支付费用的雇主被认为是翻译成果的合法所有人,而在另一些区域,所有权可以转让,并准许他人重复使用。我们认为,那些愿意与其他译者,甚至陌生人,共享资源的译者之间存在“一定程度的协作关系”(Moorkens & Lewis 2019b: 8)。然而,若翻译数据用于训练机器翻译系统,译者对自己的翻译数据重复利用的接受度就会降低,尤其对于那些认为机器翻译技术的进步会损害人工翻译利益的译者来说。一些翻译合同可能明确规定,翻译数据会用于人工或机器翻译工作流,但少有译者能左右自己的译文会如何被重复使用。同样地,通过爬虫获取的互联网数据也从未考虑到使用权的问题。

这意味着译者 A 和 B 对今后的项目有所贡献,但他们并不了解这些项目的目的和最终用途。译者 A 可能出于道德因素,反对为武器制造商工作,但仍会在不知情的情况下参与其中。对于数据被搜集并重复使用,以及为大型技术项目做贡献的人来说,这种不透明性更加普遍存在,他们并没有机会询问:“我的工作的最终应用和用途是什么?”以及“我会因参与实现该用途感到满意,还是羞愧?”(参见 Moorkens 2020 和 Weizenbaum 1986)。

翻译过程的数据创建,会根据保存和交换格式记录下许多属性,通常包括译者的姓名或 ID (参见节 2.5)、创建日期和时间、语言代码、使用的软件和项目 ID。这些信息不仅有助于按照译者、项目或创建日期来决定何时何地重用数据,还可以记录译者的翻译任务数据,包括更详细的计时、编辑动作和每一次键盘活动记录,特别是译者 A 在项目中使用的专门网络翻译平台。这些数据可用于监督译者的工作,但通常在训练机器翻译系统时删除,只保留平行句对。一旦元数据(关于数据的数据)被删除,就无法保存未来使用或重复使用的偏好,也就无法衡量个人贡献,即使协议有追溯性变更,这意味着贡献应获得版税。另一方面,抹去元数据将改善翻译数据的匿名化,该操作对于数据的共享或交换来说非常重要。

2.4 数据的流通

Topping (2000)写道,在翻译记忆共享的早期阶段,独立译者、本地化机构和本地化客户之间共享翻译记忆库,只有译者认为这种做法合乎道德规范。而到了 2021 年,这种观点似乎变得怪诞,因为有些公司的商业模式就是基于搜集并转售为机器翻译和其他机器学习技术所用的数据。²这种商业模式切实可行,因为训练数据量和数据管理水平都会影响到 NMT 系统的质量和价值。

正如节 2.2 所提到的,某些数据集可以通过开放平行语料库 Opus 等项目免费提供和流通 (<http://opus.nlpl.eu>, Tiedemann 2012),因为这些数据集都有许可协议,或受到欧盟关于公共部门信息再利用指令 (2003/98/EC) 的保护。不过,这并不一定意味着译者明确授权所有可能的重复利用形式,但他们知道数据将与公众共享。

数据可能会根据公司之间的协议,或因某公司被其他公司收购而流通,这在语言服务行业中司空见惯 (Moorkens 2020)。这些数据可出于科研或慈善目的被买卖或捐赠,而无需事先获得数据创建者的批准。这么做完全合法,除非涉及到限制使用的个人数据。

2.5 私隐与个人数据

可识别个人身份的数据被视为个人数据,其中包括含有创建者姓名或编码(匿名)的翻译记忆库。在欧盟内部,《一般数据保护条例》(General Data Protection Regulation, GDPR)³自 2018 年起限制个人数据的共享和再利用,为国家立法提供了指导方针,对数据泄露行为予以高额罚款。这不仅提高了网络安全,还限制了服务器在非欧盟地区的使用。除研究目的等例外情况,任何个人数据只有初次使用许可包括允许再利用,才可以再次利用。欧盟以外,还有许多其他国家出台法规管理个人数据的使用。

公司应该报告任何数据泄露情况,但有不少揭露数据泄露事件的媒体报道反而被公司压了下来。原因之一是为了逃避 GDPR 规定的罚款(根据该条例第 83 条,罚款最高可达 2000 万欧元或年营业额的 4%,以金额较大者为准),但除此以外,公司也希望能避免影响企业形象、失去消费者的信赖,且如果是上市公司,还需要担心股价下跌的问题。理想情况下,透明度会提高公众对企业的信任度,但大型企业的数据安全问题并非总是清晰明了,其庞大规模使得数据保护颇有难度。此外,并非所有数据泄露的情况都一样,有可能是受聘于公司的“道德黑客”(ethical hackers)为了识别漏洞以保护网络安全或公众,也有可能是恶意黑客为了谋取私利而窃取数据。

²有关机器学习的更多信息,请参见本书的导言。就本章而言,我们对机器学习的理解是,使用计算机借助大数据进行推理,而非根据输入的明确命令来达到目的。

³欧盟法规 2016/679, 获取连接<https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN>

一旦清除个人属性并匿名化,那这些就仅仅是可分享的数据而已。⁴当然,即使没有元数据,如果内容或风格容易辨认或生物特征数据与个人有所关联,那么这些也是可识别的数据。如果数据被共享或汇集,来自个人或群体的数据可用于推断某些信息,如种族或性别等受 GDPR 特别保护的属性。由于这些推论基于综合数据得出,而非明确包含在任何单一数据集中,因此它们通常不在 GDPR 的范围内 (Wachter & Mittelstadt 2019)。这给“群体隐私”(group privacy)造成了风险,某个群体会因无法识别个人的数据内容而受到歧视 (Floridi & Taddeo 2016)。

例如,译者 A 和 B 可以汇集自己和他人的翻译数据,准许第三方对他们个人或作为群体的一员做出相关推断。随着在线翻译平台的使用越来越普遍,译者不仅更无法掌控其翻译数据,还要允许翻译过程受到监控。翻译平台会收集翻译活动数据,若平台捕捉到个人情况导致翻译效率或质量出现短暂下降,这可能会对他们未来的就业前景产生负面影响。如果这些记录翻译表现不佳的、可识别个人信息的数据被泄露出去,可能会对该译者产生深远的不良影响。这并不一定意味着监控质量或生产率的做法不符合道德规范。翻译中介或公司需要能够监督他们的翻译。然而,公司将雇佣决策或沟通过程自动化后,如译者 B 的项目管理,译者便没有机会解释翻译决策或建立长期的信任关系。即使在最理想的情况下,也无法保证译者或平台用户的行为符合道德规范,但研究表明,若只是纯粹的交易关系,双方的信任和假定诚信尤其会受到破坏,从而产生连带效应,影响满意度和绩效 (Whipple et al. 2010)。

2.6 机器翻译评测中的伦理问题

Rossi & Carré (2025 [本卷]) 探究了机器翻译质量的人工和自动评测方法,有许多与此相关的道德规范问题值得在此提及。大多数机器翻译系统会在训练过程中对译文进行自动评测,待训练完成后,以简单快速的低成本方式再次进行评测。在机器翻译共享任务竞赛中,开发团队带着各自的机翻系统一较高下,系统性能的衡量方法包括自动评测或大众评分。基于这些评测,如果句段级别的群体评分或自动评测得分与人工“参考译文”的相一致,那么机器译文就被认为已经能与人工翻译相媲美;如果机器译文的得分超过人工翻译,那么前者就被认为是“超越人工翻译”的译文。

这种说法值得商榷,尤其涉及到传播范围更广泛的学术出版物和市场营销文案。甚至新闻媒体还会报道这种观点,让人觉得机器译文完美无缺,毫无风险,不再需要人工翻译。然而,自动和人工评测结果相关性极其微弱,且众包评分存在几个问题,如对连贯的译文语段进行排名或打分的是匿名或未经培训的互联网用户。Freitag et al. (2021)发现,能获得完整原文和译文的

⁴在撰写本文时,有很多人正致力于实现翻译数据的自动匿名化,但很难取得可靠的结果。

专家（职业译者）评测人员在详细的翻译错误分析时，得到的结果与众包评分者的存在显著差异，且明显更青睐人工译文，而非机器译文。此外，报酬低廉、工作条件差、用户评级系统不透明，以及在无监督或道德审查的情况下利用人类（众包工作者）参与研究，这些都是众包形式的问题所在。尽管如此，基于自动评测和众包评分的测评结果大多也只是简略提及，既贬低了人工翻译的价值，还让翻译客户等大众对机器翻译产生了不切实际和不加批判的看法。这种看法使得机器翻译更有可能被纳入翻译工作流程。

3 翻译工作流程中合乎道德规范的机器翻译使用

3.1 翻译的利益相关者

节 2.1 中的案例研究不仅引发了我们对伦理和机器翻译系统开发的思考，同时也促使我们思考如何在翻译工作流程中，合乎道德规范地使用机器翻译。例如，雇佣译者 A 的公司单方面决定将翻译工作改为译后编辑，但理想情况下，将机器翻译嵌入翻译工作流程的前提是通过协商，得到所有利益相关者的一致同意。如节 2.1 的案例所示，利益相关者包括翻译代理机构或语言服务提供商，其中还包括多个内部角色，以及自由职业译者。此外，翻译客户应知晓翻译过程中会使用机器翻译，并应了解相关的利弊（参见节 3.2）。几乎所有关于译后编辑翻译效率的研究都显示，与从头开始翻译或使用翻译记忆库相比，译后编辑的翻译效率更高。然而，一般认为，机器翻译的使用应该与翻译文本的可用时限和风险程度相关。翻译客户依赖于中介的专业能力来选择合适且性价比高的翻译流程，而终端用户则依赖于翻译客户提供可靠的目标文本。此外，Pym (2012) 提出，即使低风险目标文本的机器翻译质量差，难以理解，但这种译文也有可能符合客户需求（通过轻度或完全译后编辑来降低成本），不过，最终读者需要花费更多精力理解文本。

在使用机器翻译的工作流程中，翻译软件开发人员也是重要的利益相关者，节 1 中提到的技术开发人员的价值和相关设计决策也适用于这一群体。所使用的翻译工具可以改变翻译活动及其意义。与机器翻译的互动可通过交互式机器翻译进行，其中机器翻译用于自动建议、动态编辑或译后编辑，作为额外的翻译建议或自动扩展显示在译文窗口。翻译工具注重使用性，例如兼容惯用的键盘快捷键、提供整洁的界面，以及尽可能多的个性化设置。翻译活动数据（参见节 3.2）若被收集，可以对用户公开，这样他们就能看到收集内容并自愿使用。译者也可以使用削弱译者自主能力的翻译平台，只会即时接收任务而看不到数据收集情况，交互界面的功能比较有限，且客户可以对译者表现进行评级，但没有提交反馈或讨论的选项。

译者可以根据文本领域和工作条件，来选择接受或拒绝某项翻译工作。但对于那些不了解机器翻译可以通过多种方式融入翻译工作流程的人来说，可能很难做出选择，尤其在翻译中介的信息不透明的情况下。同样重要的是，

译者要说明工具使用情况,尤其是机器翻译,以确保遵守保密协议,不会在无意中损害其他利益相关者的利益(或如节 3.2所述,有损自身利益)。对于译者来说,遵循职业道德准则,即指导合乎道德的行为规则,属于义务伦理学的范畴。尽管这些准则是对翻译中道德角色的狭义理解,但它们仍然在决策中起到了积极的作用。⁵和许多其他行业组织一样,翻译协会通常发布这类行为准则,以鼓励专业操守、公正诚实和遵守保密原则。这些准则还有助于建立与现有和潜在客户之间的信任关系。撰写本文时,我在查阅这些准则的过程中发现,准则并没有直接提及机器翻译,即使(正如我们刚刚讨论的)决定在翻译项目中是否使用机器翻译,也可能涉及到伦理问题。Chesterman (2001)撰写了很多有关翻译道德和信任的论述,提出了一种以服务伦理为重点的通用原则,强调要忠于客户提出的条款和质量要求、源文本及其作者以及目标文本读者。

使用机器翻译不一定会导致翻译质量变差,而且耗时费力的高成本人工翻译也不适于所有类型的文本,特别是那些时限短且风险低的文本。然而,对于误译后果较严重的重要文本来讲,使用机器翻译必须要谨慎考虑和反复审查。有证据表明,译者使用机器翻译时,一些项目经理会睁一只眼闭一只眼(Sakamoto 2019),但翻译客户有理由要了解机器翻译的使用情况,并在合同中明确说明是否可以使用。

3.2 风险和责任

如节 2.2所述,翻译合同可以规定翻译数据的所有权,或允许译者或客户保留翻译记忆库。而Canfora & Ottmann (2020)将在下文介绍有关机器翻译使用的其他两个合同规定范畴——法律责任和保密条款。

译者可能会因为疏忽或未能履行对客户的注意义务而违反合同。责任只能指人的行为,这意味着行为人必须对在翻译过程中,由于机器翻译导致的错误而造成的伤害或损失承担责任。暂且不论责任,机器翻译可能给最终用户带来风险的这一事实就是道德伦理问题。关乎安全的内容不应该使用原始的机器译文。目前的翻译和译后编辑标准没有提到最终用户的责任或风险。

不过,ISO 翻译标准确实强调了“安全和保密地处理所有相关数据和文件”的重要性(ISO 2015第 3.2.a 节)。免费在线机器翻译系统的用户授予服务供应商使用用户输入数据的权利,此外,还存在不经思考而使用免费在线机翻系统来翻译机密和敏感材料,从而导致数据泄露的情况。正因为机翻系统会造成这类网络安全风险,Canfora & Ottmann (2020)才认为,最好选购不会保留数据的机器翻译服务,并建议公司使用服务器架构不对公众开放的封闭平台,或为了保密,根本不把翻译任务外包出去。但正如第 2 节和第 3.1 节所

⁵Lambert (2018)和其他学者认为,许多道德准则强调的翻译中立性假设,与对技术的假设一样存在缺陷。

述,自由译者对封闭平台较为反感,因为无法自行管理翻译数据和译者活动数据。

ISO 译后编辑标准并未提及(ISO 2017)保密性或数据安全风险,这相当令人惊讶,因为译后编辑必然会用到机器翻译,从而可能引入风险。信任是降低风险的关键,因为只有当译者认为自己与合作伙伴相互信任时,标准、指南和合同才有价值。若无信任,他们可能试图合理化不道德的行为(参见Abdallah 2010)。

除了有风险顾虑,由于节 2 中描述的流程或人工智能(AI)对工作领域和可持续性的影响,译者和用户可能并不想使用机器翻译。下一节内容会从神经机器翻译的角度,来深入探讨人工智能对翻译领域和可持续发展的影响。

4 可持续性

4.1 译者/译后编辑者的报酬、条件和工作满意度

虽然翻译任务需要精湛的专业技能,但在译者 A 和译者 B 的例子中,部分翻译流程已经(在一定程度上)实现自动化,如自动分配任务、进行译后编辑,以及翻译数据使用目的发生转变,用于译者不知情的任务。越来越多的人认为,人工智能将对许多以往看来不受自动化影响的工作领域产生重大影响。虽然可能不会直接造成失业率上升,但这些变化可能会以难以预测的方式影响经济收入、工作组织和技能管理。这些在很多行业都是未来的考虑因素,但对翻译领域已经产生影响,原因如下。首先,自 2010 年左右以来,机器翻译的译后编辑一直是翻译市场增长最快的领域,早于从统计机器翻译向神经机器翻译的转变。自翻译记忆库在 20 世纪 90 年代早期出现以来,储存翻译数据的做法已经司空见惯,尽管以监控和自动化为目的的翻译活动数据收集才出现不久。其次,很大一部分译者为自由职业工作者意味着他们兼具灵活性和自主性,但要在一个接一个项目的时间框架内工作。这造成了权力的不平等——在一个个项目中,翻译中介和雇主可以单方面改变的翻译过程和工作条件,但译者对此几乎没有发言权。节 2 关于数据的讨论清楚展现了权力差异的影响。随着合并和收购的步伐加快以及大型翻译上市公司的成立,商业运营决策者与从事翻译、译后编辑、审校、标注、修订、字幕制作或其他与文本直接相关工作的众多自由职业者,两者脱节越来越严重。翻译行业提出将项目管理自动化,以及使用区块链来确定确定著作权,都不太可能改善这种情况。更概括来说,正如 Moorkens & Rocchi (2021)提到的,翻译行业对伦理道德问题一直未予以重视。

早在 1980 年, Kay 曾就翻译技术提出一个观点,即译者应掌握技术,以辅助完成机械重复且单调乏味的工作(Kay 1980)。但我们所见却正相反,许多译者的工作因为技术而受到限制。一些译者的工作内容已经简化为质量检查、标注或改正质量参差不齐的机器翻译中重复出现的错误。在改正机翻错

误时，机翻译文甚至可能切分成无语境的独立句子，有些自动通过质量检查，有些则标记为待审校。

有些译者喜欢进行译后编辑，即便薪酬降低（如节 2 第一个案例研究中的情况），他们也觉得这种工作收益好，值得做。但是，在技术化的翻译过程中，所有利益相关者都需要在短期效率和长期收益之间取得平衡。

从业者的满足感来自有意义的工作、对任务的掌控以及与同事共同奋斗。他们的工作动力是成就感和对这种成就的认可。如果在翻译过程中没有成就感，那更高水平的译者会离开这个行业，从而导致高水平的译者和译后编辑者的短缺。这种短缺会影响多语信息的可靠获取，以及机器翻译训练所依赖的高质量双语数据的收集。Docherty et al. (2008: 4) 认为，可持续的工作系统必须满足许多，而不是少数利益相关者的需求；不能只关注“短期的静态效率，如生产率和盈利能力，还必须注重长期的动态效率，如学习和创新”。联合国可持续发展目标 (UN Sustainable Development Goal) 第 8 点是提供体面的工作和经济增长，但该目标的第 13 点和第 15 点所涉及的环境可持续性也与机器翻译有关。

4.2 环境问题

我们有理由认为，经济增长目标与环境可持续性目标之间难以兼顾。Cronin (2017) 特别指出了本地化行业的增长依赖性 (growth dependency)。机器翻译所依赖的信息与通讯技术 (ICT) 行业需要开采稀有金属，且因资源回收不善和对环境造成污染而久负骂名。神经机器翻译的资源密集特点尤为突出，如需要性能强大的图形处理器 (GPU) 进行训练，且能源消耗极高。Strubell et al. (2019) 估计，训练一个 Transformer 神经网络大模型所产生的二氧化碳排放量，大约是一辆汽车整个生命周期（包括燃料）的 5 倍。⁶ 然而，大多数模型训练的资源密集型程度远低于本书所举的例子。此外，硬件的性能更加强大，设计和生产成本也越来越高，但随着能耗优化，且大量程序能够同时运行，训练所需能耗也相应减少。尽管如此，神经机器翻译系统的训练成本依然高昂，耗电量大。这对环境的影响将取决于能量的来源。发布机器翻译系统的详细信息时，目前还没有商定的能耗基准，尽管在可持续人工智能发展建议的背景下，已经提出了一些基准。Van Wynsberghe (2021) 强调，如果不关注人工智能（由此推及神经机器翻译）开发及运用的可持续性，人工智能开发本身将是不可持续的。

5 多样性

5.1 开发者和用户

开发神经机器翻译的巨大障碍在于成本和能耗。对数据的要求意味着早期的系统只能使用公开可用的数据（参见节 2），通常为主要的欧洲语言创建系

⁶我们还注意到，只有规模最大的公司才负担得起训练这类大模型的费用。

统。因此，无怪乎早期发表的神经机器翻译相关工作主要由北美和欧洲的资源丰富的学术研究小组主导。但这种情况有所改变，原因有二。首先，大型技术公司全力支持神经机器翻译的研究工作，建立了资源丰富的团队，这些团队在优化主要语言对的机器翻译系统方面处于领先地位。这意味着许多学术研究团体难以跻身主要的欧洲语言领域，只能转向低资源语言和少数民族语言系统开发这一“小众”领域。其次，通过将单语数据从预期目标语言通过机器翻译成预期源语言⁷来合成平行数据的能力，提高了低资源语言对的机器翻译质量。因此，第五届国际机器翻译大赛（WMT20）包括因纽特语和泰米尔语与英语的互译任务。然而，提高低资源语言翻译质量的另一种方法是构建大型多语言系统，不过通常是大型团队才有资源和财力这么做。

目前还没有关于机器翻译研究团队多样性的调查。不过，只需要搜索相关学术论文就会发现，有相当数量的论文使用简体中文、孟加拉语和印地语来发表，这些语言的使用者众多，但都不是欧洲语言，且孟加拉语和印地语在历史上一直是低资源语言。在研究团队带头人与研究议程决策者当中，多样性可能较低。当论及机器翻译之外的偏见，这种多样性的缺乏是一个突出的问题，导致了使用机器学习的系统中许多广为人知的错误和盲点，例如为识别疑犯而开发的人脸识别系统过度倾向于将少数族裔划分为罪犯。

Vieira et al. (2020: 13) 在关于机器翻译的社会影响的文章中指出，在用于医疗和法律领域的案例中，机器翻译的使用不当会对弱势群体造成伤害。他们发现，这可能会加剧不平等现象，因为“世界上相对较少的语言拥有的数据和资源有些不成比例”，加之机器翻译的使用者缺乏机器翻译素养。另一方面，机器翻译有助于促进交流，因为越来越多的免费机器翻译系统增加了低资源语言选项。在撰写本文时，谷歌翻译（Google Translate）共涵盖 109 种语言，包括齐切瓦语、苏格兰盖尔语、维吾尔语和鞑靼语等低资源语言。

5.2 机器翻译的输出

Jones 于 2016 年算出，在超过 6000 种非濒危语言中，只有 1% 的语言被机器翻译系统覆盖。得益于节 5.1 提及的研究，这种情况有所改善，百度和谷歌覆盖的语言数量越来越多也证明了这一点。但是，在现实世界中，这发生在一个信息倾向于从资源丰富的语言，流向资源贫乏的语言的世界里。由于机器翻译比人工翻译对源语言的干扰更大（参见 Toral 2019），人们担心从长远来看，资源匮乏的语言会愈加贫瘠。

所有能用机器翻译的语言都可能面临类似的处境，特别是如果人工翻译的数据不足意味着机器翻译系统很难跟上当代语言发展。Vanmassenhove et al. (2019) 展示了，当神经机器翻译引擎的训练至所谓的收敛点⁸，会导致词

⁷机器翻译研究人员称这一过程为“回译”。不要将“回译”与 Baker (2018) 等传统翻译研究资料中，用作注释技术的“反向翻译”混淆。

⁸即自动测评分数显示输出质量没有改善时，迭代的神经机器翻译系统训练停止的点。参见节 7.2。

汇多样性下降,这表明若训练在更早的时候停止,神经机器翻译输出则不那么同质化,且词汇更多样化。Vanmassenhove (2019)表明,这种输出译文的标准化解显出算法偏差,这种偏差加剧了训练数据的性别偏差,其中名词等形式与最常见的某个性别相关联(在两性制度中),而不太常见的性别则在输出数据中被标准化了。这导致输出译文加深社会偏见,性别使用不一致,甚至在单个句段中也是如此。下节内容将讨论目前如何消除输出中的偏见,以及作为伦理问题的隐藏主体,计算机还发挥着哪些作用。

6 伦理主体——计算机

有关人工智能道德问题的讨论主要涉及安全,并将提升机器的自主性以及将人工智能技术的应用拓展到更多任务。与此同时,机器道德问题的特点在于,它将“机器”视为具有“能动性”(即行动的意愿和能力)的主体,而不是客体。关于机器翻译的偏见和风险的研究虽不多,但呈现日益增长的趋势(参见节 5.2),此类研究的范围也囊括更宽泛的机器翻译和人工智能的内在价值观(参见节 5.1)。这便催生了计算机作为伦理主体的概念,因此道德决策隐含于计算机的设计之中。虽然已有这方面的努力正在进行,但我们经常在网上看到有偏见或歧视性的机器译文,其出现场景对于开发者而言难以预测。当机器似乎显露出智能,技术(尤其是人工智能)带来的意外后果启发了一系列探讨机器翻译技术使用有违伦理道德规范的书籍和文章。

机器翻译技术的问题可能涉及数据收集、用于从数据中归纳提取模式的算法,以及数据本身。技术带来的结果也可能是有意为之,也就是说,由此技术的“预设用途”(affordance,即工具所暗示的使用方式和互动方式)推动用户以某种方式行动,而这种方式可能对他们自己或他人带来负面影响。

在关于神经机器翻译偏见问题的少量研究中,研究重点仅限于性别偏见,而非种族或性向(sexuality)等其他属性,并提出了一些解决方案来“消除”机器译文的偏见。在神经机器翻译译文中,如何正确使用性别和消除性别偏见,建议方法包括使用性别标签(类似礼貌用语和语域的标签)、消除偏见的词嵌入,或者用类似领域调试(domain adaptation)的方法处理性别偏见,比如利用小型性别平衡数据集进行迁移学习(Tomalín et al. 2021)。2020年,谷歌针对某些语言推出了一个系统,对每个带性别标记的名词同时输出阴性和阳性两种形式的译文。用户可以自行评估不同的选项,选择合适的译文。虽然这种做法仅基于二元性别论,但还是提高了谷歌翻译处理性别信息的译文质量(Johnson 2020)。

未来,计算机有可能变成显性的伦理主体,有能力处理每种场景的信息,并自主决定最佳或最符合道德规范的行动方案。但目前的技术还未发展到这种先进程度,因此机器创造的内容还不能声明版权,机器也不能对损失或伤害负责。即便技术很成熟了,我们也不能假定计算机将按照道德规范来行

事。我们已经确定技术并非无害。虽然技术的使用能够造福个人和社会,但也存在一些使用之前难以预料的风险。

毫无疑问,免费在线机器翻译工具的使用促进了很多人的交流,但是无缝界面和质量更高的输出可能会让终端用户认为,该技术的使用不会造成伤害(通常不会)。节 2.1 所举的例子中,翻译工作流程的利益相关方可以被认为是具有语言技术方面的专业知识,但对于普通大众来说并非如此,他们可能期望机器翻译具有与(优质)人工翻译相同水平的连贯性和可理解性。他们甚至可能意识不到自己正在读的是机器译文,即使译文被标记为机器翻译生成,他们也不太可能意识到误译的风险。

7 小结

随着机器翻译质量的提高,无论是用于直接交流还是作为翻译工作流程的一部分,这项技术都促进了更多交流。然而,机器翻译开发者、翻译客户、翻译机构、译者和翻译消费者都需要考虑道德伦理问题。与所有技术一样,机器翻译开发和输出都不应被视为道德中立,而是传播开发人员或创造翻译数据的译者的观点,他们的观点体现在为与机器翻译交互而设计的工具以及机器译文之中。对评估结果良好的机器翻译不加批判的报道,淡化了公众对机器译文存在风险和偏见的意识,同时有可能贬低人工翻译。在使用和处理机器翻译时,读者不妨参考本章提出的问题,并思考相关过程是否符合自己的价值观、目标和原则。

本章先从翻译来源和译者数据、所有权、许可、版权和流通模式等方面来讨论道德问题。在翻译工作流程中,机器翻译的使用是否合乎道德规范可能取决于所有利益相关者的态度、保密规则以及机器翻译平台背后的设计决策。这与可持续性、人机互动模式以及译者对工作流程的自主掌握程度有关。我们对于能够确保多样性和消除偏见的系统和数据的方法或许最为欠缺,但毫无疑问,这需要进一步的讨论和调整。

References

- Abdallah, Kristiina. 2010. Translator's agency in production networks. In Tuija Kinnunen & Kaisa Koskinen (eds.), *Translator's agency*, 11–46. Tampere: Tampere University Press.
- Baker, Mona. 2018. *In other words*. London: Routledge.
- Canfora, Carmen & Angelika Ottmann. 2020. Risks in neural machine translation. *Translation Spaces* 9(1). 58–77.
- Chesterman, Andrew. 2001. Proposal for a hieronymic oath. *The Translator* 7(2). 139–154. DOI: 10.6509.2001.10799097.

- Cronin, Michael. 2017. *Eco-translation: Translation and ecology in the age of the anthropocene*. London: Routledge.
- Docherty, Peter, Mari Kira & A. B. (Rami) Shari. 2008. What the world needs now is sustainable work systems. In Peter Docherty, Mari Kira & A. B. (Rami) Shari (eds.), *Creating sustainable work systems: Developing social sustainability*, 1–22. London: Routledge.
- European Parliament. 1996. *Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases*. <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:31996L0009>.
- Floridi, Luciano & Mariarosaria Taddeo. 2016. *What is data ethics?* DOI: 10.1098/rsta.2016.0360.
- Freitag, Markus, George Foster, David Grangier, Viresh Ratnakar, Qijun Tan & Wolfgang Macherey. 2021. Experts, errors, and context: A large-scale study of human evaluation for machine translation. *Transactions of the Association for Computational Linguistics* 9. 1460–1474. DOI: 10.1162/tacl_a_00437.
- ISO. 2015. *ISO 17100:2015. Translation services – requirements for translation services*. <https://www.iso.org/standard/59149.html>.
- ISO. 2017. *ISO 18857:2017. Translation services – post-editing of machine translation output: Requirements*. <https://www.iso.org/standard/62970.html>.
- Johnson, Marvin. 2020. *A scalable approach to reducing gender bias in Google Translate*. <https://ai.googleblog.com/2020/04/a-scalable-approach-to-reducing-gender.html> (20 May, 2020).
- Kay, Martin. 1980. *The proper place of men and machines in language translation* (Report CSL- 80-11). Palo Alto, CA: Xerox Corporation.
- Lambert, Joseph. 2018. How ethical are codes of ethics? Using illusions of neutrality to sell translations. *Journal of Specialised Translation* 30. 269–290. https://www.jostrans.org/issue30/art_lambert.php.
- Moorkens, Joss. 2020. “A tiny cog in a large machine”: Digital Taylorism in the translation industry. *Translation Spaces* 9(1). 12–34.
- Moorkens, Joss & David Lewis. 2019a. Copyright and the reuse of translation as data. In Minako O’Hagan (ed.), *The Routledge handbook of translation and technology*, 469–481. London: Routledge.
- Moorkens, Joss & David Lewis. 2019b. Research questions and a proposal for the future governance of translation data. *Journal of Specialised Translation* 32. 2–25.
- Moorkens, Joss & Martha Rocchi. 2021. Ethics in the translation industry. In Kaisa Koskinen & Nike K. Pokorn (eds.), *The Routledge handbook of translation and ethics*, 320–337. 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.
- Pym, Anthony. 2012. *On translator ethics: Principles for mediation between cultures*. Amsterdam: John Benjamins.
- 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.
- Sakamoto, Akiko. 2019. Unintended consequences of translation technologies: From project managers' perspectives. *Perspectives* 27(1). 58–73.
- Steinberger, Ralf, Andreas Eisele, Szymon Kloczek, Spyridon Pilos & Patrick Schlüter. 2012. DGT-TM: A freely available translation memory in 22 languages. In *Proceedings of the 8th international conference on language resources and evaluation (LREC 2012)*. ELRA. <https://www.aclweb.org/anthology/L12-1481/>.
- Strubell, Emma, Ananda Ganesh & Andrew McCallum. 2019. *Energy and policy considerations for deep learning in NLP*. <https://aclanthology.org/P19-1355.pdf>.
- Tiedemann, Jörg. 2012. *Parallel data, tools and interfaces in OPUS*. In *Proceedings of the 8th International Conference on Language Resources & Evaluation (LREC 2012)*. 2214–2218. Luxembourg: ELRA. http://www.lrec-conf.org/proceedings/lrec2012/pdf/463_Paper.pdf.
- Tomalin, Marcus, Bill Byrne, Shauna Concannon, Danielle Saunders & Stefanie Ullmann. 2021. *The practical ethics of bias reduction in machine translation. Why domain adaptation is better than data debiasing*. 419–433. DOI: 10.1007/s10676-021-09583-1.
- Topping, Suzanne. 2000. Sharing translation database information: Considerations for developing an ethical and viable exchange of data. *Multilingual* 11(5). 59–61.
- Toral, Antonio. 2019. Post-edits: An exacerbated translationese. In *Proceedings of machine translation summit XVII*, 273–281. EAMT. <https://www.aclweb.org/anthology/W19-6627/>.
- Troussel, Jean-Christophe & Julien Debussche. 2014. *Translation and intellectual property rights*. Luxembourg: Publications Office of the European Union.
- Van Wynsberghe, Aimee. 2021. Sustainable AI: AI for sustainability and the sustainability of AI. *AI and Ethics* 1. 213–218. DOI: 10.1007/s43681-021-00043-6.
- Vanmassenhove, Eva. 2019. *On the integration of linguistic features into statistical and neural Machine translation*. PhD Thesis. Dublin City University.

- Vanmassenhove, Eva, Dimitar Shterionov & Andy Way. 2019. Lost in translation: Loss and decay of linguistic richness in machine translation. In *Proceedings of machine translation summit XVII: Research track*. Dublin: EAMT, 222–232. <https://www.aclweb.org/anthology/W19-6622.pdf>.
- Vieira, Lucas Nunes, Minako O'Hagan & Carol O'Sullivan. 2020. Understanding the societal impacts of machine translation: A critical review of the literature on medical and legal use cases. *Information, Communication & Society* 24(11). 1515–1532. DOI: 10.118X.2020.1776370.
- Wachter, Sandra & Brent Mittelstadt. 2019. A right to reasonable inferences: Rethinking data protection law in the age of big data and AI. *Columbia Business Law Review* 2019(2). 1–130.
- Weizenbaum, Joseph. 1986. Not without us. *Computers and Society* 16. 2–7.
- Whipple, Judith M., Daniel F. Lynch & Gilbert N. Nyaga. 2010. A buyer's perspective on collaborative versus transactional relationships. *Industrial Marketing Management* 39(3). 507–518. DOI: 10.1016/j.indmarman.2008.11.008.
- World Intellectual Property Organization. 1979. *Berne convention for the protection of literary and artistic works*. (as amended on September 18, 1979). Geneva: WIPO. <http://www.wipo.int/wipolex/en/details.jsp?id=12214>.

第七章

神经机器翻译的工作原理

Juan Antonio Pérez-Ortiz
西班牙阿利坎特大学

Mikel L. Forcada
西班牙阿利坎特大学

Felipe Sánchez-Martínez
西班牙阿利坎特大学

本章介绍神经机器翻译系统的主要工作原理。我们会一一讲解这些系统的关键概念，帮助读者全面了解其内部运作和可能性。这些概念包括神经网络、学习算法、词嵌入、注意力和编码器-解码器架构。

1 引言

了解神经机器翻译 (NMT)，首先应该知道它把翻译看作一项数学系统的数字处理任务。这些被称作“人工神经网络”的数学系统把句子转换成一系列数字。系统会加上一些数字（通常是几千或几百万个数字），然后乘以及其他数字，再进行一些相对简单的数学运算，最终输出原句的译文。

你可能一直都从另一个角度来理解翻译，认为这是一项发生在人类大脑深层区域的智力任务，涉及的认知过程难以窥知。你的理解没错！不过，目前计算机模拟人工翻译的方法采取截然不同的路径：数百万次数学运算只需片刻便可生成译文，而译文的质量时好时坏。实践证明，过去几年里，生成恰当译文的概率有了显著提高。但从发展历程看来，人工神经网络是模拟自然神经网络（如人脑）运作方式设计的简化模型，来模拟人类大脑等自然神经网络的运作原理，其认知过程也是分布式神经计算过程的结果，这些过程与上述的数学运算并无太大区别。



Juan Antonio Pérez-Ortiz, 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

本章会讲解神经机器翻译技术的关键要素。首先要介绍的是人类大脑处理翻译与神经机器翻译之间的联系，在此过程中我们会引入理解“机器学习”和“人工神经网络”所需的基本概念。“机器学习”和“人工神经网络”是神经机器翻译的两大基石。然后探讨词的计算机化表示方法——“语境无关的词嵌入”(non-contextual word embeddings)的基本原则，这种方法如果伴随许多重要属性，若结合“注意力”机制，就可以得到实现神经机器翻译的关键因素——“语境单词嵌入”(contextual word embeddings)。以上所有内容能够帮助我们全面了解最常用的两种神经机器翻译模型——Transformer 模型和循环模型——的内部运作。本章最后还介绍了一系列次要主题，有助增进读者对系统内部运作的了解。

2 人工翻译与神经机器翻译的简单类比

为了简化讨论，我们姑且认为，翻译一篇文章就是翻译其中的每个句子。现在假设句子的翻译分两步：译者先确定整个原句的“阐释”或“意义”，然后一下子用目的语表达出差不多的意思。但译者每天都会遇到新的句子（如 *The pencil slipped from my hand, stood up, and started talking to me*），仍可以照常翻译。这是怎么做到的？语言学对此给出的解答是依靠“语义组合原则”(principle of semantic compositionality)，即把对每个单词的理解组合起来“构建”对句子的理解，单词的组合顺序由句法结构制约——单词组成短语、短语组成语段，直至得到完整的句子。译者分析源语句子的意义，然后进行上述过程的逆向操作，使用目标语生成对应的句子。当然，译者并非总把句子当成一个整体来处理，尤其是遇到长句时，他们可能采取简化方法来避免整个句子的复杂分析，不过我们暂且不考虑这些例外。

而神经机器翻译的工作原理与此类似。翻译句子时，系统会在“编码”阶段为原文的每个单词分配一个神经“表征”或“嵌入”。然后，将这些神经表示结合起来，为句子产生类似的表示。组合时，各个表示会根据语境进行调整，这可以视为理解或意义的语境化表现。到了“解码”阶段，句级表示逐步拆解，以逐个预测目标语句子的单词。完成“编码”和“解码”的是两个人工神经网络——“编码器”和“解码器”，它们相互连接组成单个复合神经网络。

与译者一样，在生成目标语的每个单词时，目前的神经网络架构也并非考量一整个句子，而是学习“注意”相关的源语单词和已经生成的目标语单词。

本章的余下部分会进一步解释神经表示的性质，描述使用选择性聚焦重点的“注意力”机制来构建和转变这些表示的人工神经网络（以下简称“神经网络”）的结构，以及通过翻译实例来介绍这些人工神经网络的训练方式。

3 人工神经网络

要想理解神经机器翻译，就得弄清实现这项技术的人工神经网络 (Goodfellow et al. 2016) 的构成、运作原理和训练方法。

“神经”一词很容易让人联想到神经元和动物的神经系统，尤其是人类大脑的运作方式。实际上，人工神经网络的确是由成千上万个人工神经元组成，这些神经元的“激活”（即“兴奋”或“抑制”的程度）取决于它们从其他神经元接收的信号以及携带这些信号的连接强度。

3.1 人工神经元

人工神经元是人工神经网络的主要组成部分。这些人工神经元（以下简称“神经元”）的状态更新或激活可分为两步。以图 1 的简单情况为例，了解神经元 S_4 如何在接收来自神经元 S_1 、 S_2 和 S_3 的刺激后被激活。

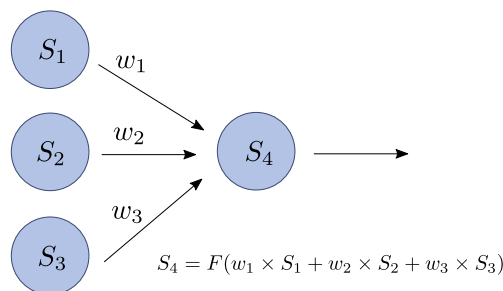


图 1: 神经元 S_4 接收来自神经元 S_1 、 S_2 和 S_3 的刺激后被激活。

如图所示，第一步，所有与神经元 S_4 相连接的神经元 (S_1 、 S_2 和 S_3) 都要乘以其对应的连接强度“权重” (w_1 、 w_2 和 w_3)，得到的神经元 S_1 、 S_2 和 S_3 的激活值进行加法运算；这些权重决定了神经元 S_1 、 S_2 和 S_3 对神经元 S_4 的实际刺激值。权重可为正值或负值。例如，若权重 w_2 为正且 S_2 的激活值高，则神经元 S_2 会激发神经元 S_4 (正刺激)；但若 w_2 为负，则会抑制神经元 S_4 (负刺激)。一般来说，通过正权重连接的神经元会同时受到激发或抑制，而通过负权重连接的神经元则激活状态相反。说回神经元 S_4 ，若把来自每个神经元的刺激相加，可得到“净刺激”：

$$x = w_1 \times S_1 + w_2 \times S_2 + w_3 \times S_3. \quad (1)$$

净刺激 x 可为任意的正值或负值，但还不是神经元 S_4 的实际激活值。第二步，神经元 S_4 “响应”这个刺激。在这个例子中，当刺激处于中间水平（即没有过度偏正或偏负），神经元 S_4 对它的响应便非常敏感。然而，当刺激变

大时 (无论正负), 刺激值的变化对结果的影响较小, 因为神经元在很大程度上被抑制或激发。

在该示例中, 神经元 S_4 的激活值范围在 -1 到 $+1$ 之间。图 2 表示神经元 S_4 如何对等式 (1) 中的刺激值作出反应。该反应以“激活函数” $F(\dots)$ 来表示, 用来计算刺激值; 得到的结果便是 S_4 的激活值:

$$S_4 = F(x) = F(w_1 \times S_1 + w_2 \times S_2 + w_3 \times S_3). \tag{2}$$

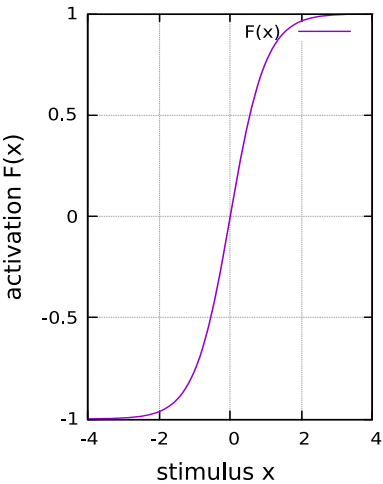


图 2: 神经元如何对接收到的总刺激值做出反应。

可以看出, 横轴取值为 0 左右的反应与刺激成比例, 但当刺激值过大或过小, 神经元会处于过度激发或抑制状态, 因此做出的反应要小得多。对于这种神经元, 无论总刺激有多强, 都不会达到 -1 和 $+1$ 的实际极值。如上所述, 示例中的神经元 S_4 是激活值范围在 -1 到 $+1$ 之间的特殊神经元。还有取值范围不同的激活函数, 但不在本章的讨论范围内。

3.2 从神经元到神经网络

上一节讨论过的神经元可以连接起来形成人工神经网络, 完成特定的计算任务, 以解决特定的问题。在神经网络中, 有些被称为“输入神经元”, 用于接收外部刺激作为神经网络的输入 (正如眼睛连接到大脑并向其输入画面), 表示一个需要解决的问题; 有些被称为“隐藏神经元”, 只接受来自其他神经元的刺激; 最后, 还有些被称为“输出神经元”, 表示问题的解决方案 (类似发送到手部肌肉的信号, 让手做出某个动作)。图 3 展示了含有五个神经元的

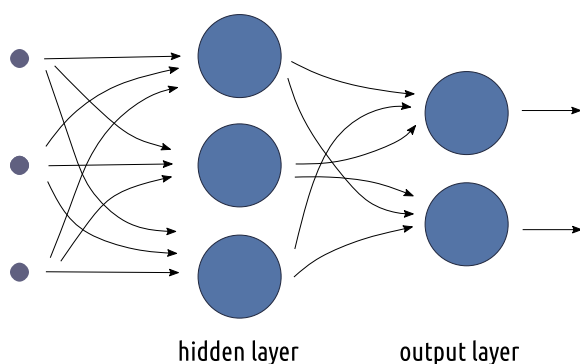


图 3: 含有三个隐藏神经元和两个输出神经元的人工神经网络。每个连接都对应一个权重, 但图中并未显示。最左边的三个小圆圈表示输入神经元, 直接传出外部输入值。与一般神经元不同, 不需要通过激活函数来计算输入神经元接受的刺激值。

神经网络, 其中三个输入向量将信号传递给三个隐藏神经元, 然后又刺激两个输出神经元。

要建立神经网络来解决特定问题, 首先得确定其“架构”, 即神经元的数量和连接方式、哪些是输入神经元或输出神经元。不过, 实际计算取决于神经网络所有连接的权重。第 3.5 节会解释如何得出这些权重。这里只需指出, 人工神经网络的优点是它们经“训练”后能执行示例中的任务, 也就是说, 可以通过观察一组已解决的示例, 使其将权重设为特定值, 每个示例都包括输入信号值 (表示问题的特征) 和期望的输出激活值 (表示解决方案)。

3.3 神经元层

假设你是零基础学习者, 想学一些风景油画的基本技巧。教材可能会告诉你有哪些步骤, 如素描起稿 (粗略构图)、色彩分布、细节微调和上色成图 (完成最后的润色)。这里的重点不在于步骤数量或每步特点, 而是整个过程层层递进, 每一个步骤的输出即为下一个步骤的输入。每一步优化了上一步的结果——第二步 (色彩分布) 的结果比第一步 (素描起稿) 的结果更像一幅风景画。同样, 从概念上讲, 我们可以认为第四步 (上色成图) 的结果优于前面所有步骤产出的结果。

事实证明, 神经网络计算也是类似的层层递进过程。早在 60 年代就有研究人员发现, 使用多层神经元可以处理更复杂的任务。多层神经网络中的每一层都细化上一层的输出, 并向最终解决方案迈出或大或小的一步。多层神经网络的最终架构类似图 3 所示, 但增加了一些隐藏层。图 3 简单的神经网络清楚展示了这种分层结构——计算分两步, 在两层神经元中进行。

由多层神经元组成的模型被称为“分层神经网络”。尽管理论结果表明，双层神经网络的算力足以执行几乎所有任务 (Hornik 1991)，但实际上，神经网络的算力似乎与神经网络的层数相关；含有多层神经元的模型通常被称为“深层神经网络”，相应的训练算法被称为“深度学习算法”。

深度神经网络内在结构极为复杂，例如，2020 年发布的自然语言生成领域最大的神经网络之一 GPT-3 (Brown et al. 2020) 含有 96 层，每层的神经元达数万个，其训练算法需要学习大约 1750 亿个权重。GPT-3 系统的训练需要借助超级计算机，训练过程长达数周甚至数月，但据估计，如果使用性能绝佳的游戏台式电脑来学习这类模型的权重，估计需要 350 多年。¹

3.4 神经机器翻译

如果能把一个源语句子表示为一组神经网络输入值，然后将神经网络输出解读成目标语句子，这就是“神经机器翻译” (neural machine translation, NMT) 系统。神经机器翻译首先处理源语句子的单词。神经网络的编码器每读取一个源语单词，网络中一组特定的神经元的激活状态就会改变。处理完整个源语句子后，神经网络的解码器便开始运作。训练过的编码器在给定已输出的目标语单词的情况下，逐步得出译文中每个可能的目标语单词的概率。这一过程类似于现代智能手机的预测键盘的工作方式，但是，正如我们将看到的，神经机器翻译的单词预测也依赖于源语句子，因为毕竟预测的结果是源语句子的翻译。

神经机器翻译系统是深层神经网络，其架构将在后面的第6节中讨论。它们有数以千计的神经元和数以百万计的权重（甚至更多），神经元和权重的训练需要借助从包含数百万源句及其翻译的平行语料库中获取的示例，来训练这些神经元和权重。源语句子单词的数学表示作为输入传送到神经网络，而对应的目标语句子单词用于表示期望的输出。正如你所料，在合理时间内训练一个大型神经网络对算力的要求很高——需要能够处理大量复杂计算任务的高性能硬件，一遍遍地展示例子来训练神经网络。每次迭代都会给神经网络的权重带来细微调整，以改进其对目标语单词的预测。

3.5 神经网络的训练

神经网络的训练过程就是确定神经元之间的连接权重，以便在给定输入-输出示例的“训练集”的情况下，得到的实际输出尽可能接近相关示例。

训练初始权重或随机选取，或从解决类似任务的神经网络中提取。在训练期间，权重不断调整，以达到最小误差函数值（又称“损失函数”），该函数测量实际输出相对于期望输出的偏离程度。“训练算法”（又称“学习算法”）反

¹“OpenAI’s GPT-3 language model: A technical overview” (2020). 检索自 <https://lambdalabs.com/blog/demystifying-gpt-3>.

复计算权重的微小调整（或更新），直到误差函数对于训练集的所有示例都是最小的或足够小，或者在不同“开发集”中观察到某种预期的性能，这也是预留开发集的目的所在（见第7.2节）。

训练算法的技术细节不在本章的讨论范围内，只需知道它通常基于计算当每个权重以固定但很小的程度（即误差函数的“梯度”）变化时，误差函数会出现多大变化，然后对每个权重进行调整以减小误差函数。² 这类训练被称为“梯度下降法”（gradient descent），它无法保证找到最佳权重，但很可能会找到不错的权重。权重变化的幅度由“学习率”（learning rate）这一参数来调节。训练算法执行之初，学习率通常较高，但是随着权重越来越接近最终值，学习率逐渐变小。注意，神经网络的训练过程很费力，因为需要大量实例而且多次呈现给系统学习才能成功。然而，这主要是由于训练算法的局限性，而非特定神经网络缺乏表示问题解决方案的能力。

权重一旦确定，训练随即停止（见第7.2节）。这时，可以向神经网络输入训练集以外的例子，得到相应输出。

3.6 神经网络中的泛化

“泛化”（generalisation）是人类和动物的基本认知过程，使我们已有关知识运用到新的场景，这些新情况与当初的学习场景类似，但又不尽相同。例如，一个原本就会开车的人并不需要为了驾车驶入陌生道路或驾驶新车而重新学习开车。同样地，已经通过某种方式对某种刺激做出反应的生物体，能够以类似方式对类似刺激做出反应，这就是“泛化”。“泛化”也是语言学习的关键，幼儿很快就能学会说出从未听过的话。

理想情况下，神经网络可以在机器翻译的场景实现泛化，通过相似的输入得到相似的输出，无论这些输入是否存在于训练集中。神经网络的一大特点就是“平滑性”，这意味着如果输入值略有变化，计算的结果不会有显著变化。

在广义上，为了实现泛化，相似的句子应有相似的表征，并且由于句子表征来自单词表征，可以推断出，用相似的数字来表示相似的单词，是神经网络语言处理中泛化的先决条件。

下一节，我们将深入研究如何利用神经网络的平滑性，得到句子中单词的神经网络表征，使经过训练的系统能正确地泛化新句子。

4 词嵌入：词向量表征形式

在上一节中，我们讲到神经元通常分层排列，上一层神经元的输出是下一层神经元的输入。值得注意的是，某层神经元组的输出恰为当前一步正在处理的信息的表征。

²看到这里，有些读者可能认出了“函数导数”这一数学概念。

在自然语言处理领域,如上所述,神经网络处理的信息由单词组成,而神经网络的单词表征常被称为“嵌入”(Mikolov et al. 2013)。使用词嵌入的重要特质在于,意义相近或在相同语境下共现的词嵌入也相近。为了更好地理解这一点,请在纸上画一个边长约 10 厘米的正方形。现在,根据意思的远近把以下单词填入正方形,意思越相近,单词距离越近。如果无法判断意思远近,可以根据单词在句子或段落中同时出现的频率来决定其位置。单词包括 restaurant (餐厅)、red (红色)、garden (花园)、fountain (喷泉)、flower (花)、tomato (番茄)、balloon (气球)、waiters (服务员)、knife (刀)、flowers (花)、menu (菜单)、cooked (煮熟)、chromosome (染色体) 和 consistently (始终如一)。请先完成这一步,再继续阅读。

受词义远近程度的限制意味着不能把单词随意放在正方形内。你可能会把“餐厅”、“菜单”和“服务员”分成一组,而“花园”、“花”和“喷泉”为另一组。但有些词的位置难以决断,如“红色”明显与“番茄”相近,但同时与“花”也相近,可以折中把它放在两者之间。如果我们认为与“花”相比,“红色”这一属性与“番茄”的关系更为紧密,则“红色”更接近“番茄”,而不是“花”。

你可能已经注意到正方形里出现了几个词群,有的表示餐厅及相关事物的语义场,有的则围绕花园和果园的概念。还有一些离群词,尤其是“始终如一”一词,不得不把它放在远离其他单词的位置。难以归类的离群词还有“染色体”,但由于“花”和“服务员”都具有携带遗传信息的染色体,因此它可以放在这两个词之间,但同时不会太接近“红色”。如图4所示为以上单词的可能分布情况,不一定与您给出的答案相同。³

为了用数学方式表示这些单词,我们用坐标来表示每个单词在正方形中的位置。在这个二维空间里,每个单词需要一对坐标:第一个数字表示该词到正方形左侧垂直边的距离,第二个数字表示到正方形底部水平边的距离。例如,单词“餐厅”的坐标分别为 0.25 和 1.1,单词“菜单”的为 0.6 和 1.3,接近“餐厅”,如图4所示。这些坐标值可用“向量记法”来表示,只需用逗号将两个数字分开,放在中括号里。因此“餐厅”和“菜单”对应的向量分别为 $[0.25 \times 1.1]$ 和 $[0.6 \times 1.3]$,有可能是代表这两个词的词嵌入之一。

虽然可能并不是非常直观,但使用两个数字而非一个数字来表示词嵌入,有助于解决单词远近位置的问题,因为我们有更多的自由来满足上述问题的限制条件。事实上,从二维增至更高维度,能更好地解决这个问题。一个词的五维表示可以是 $[2.34 \times 1.67 \times 4.81 \times 3.01 \times 5.61]$ 。神经机器翻译系统能处理包含数百个维度的词嵌入,而待翻译的句子便由这些规模庞大的词嵌入的集合来表示。

我们在第 refss:trainnn 节中介绍过神经网络权重的学习依靠算法实现,而同样的算法也可以用于词嵌入的学习。实际上,权重和词嵌入的学习同时进行。考虑到在神经网络机器翻译中,神经网络的输入层通常由输入句子的词

³我们特意将图4放在后几页,避免您做练习的时候看到。

嵌入组成，我们无需局限于固定的向量表示。相反，向量值在训练过程中不断更新，以便使误差函数的值最小化。。

4.1 泛化

如上所述，为使神经网络能适当“泛化”，即能学习翻译并能翻译新句子，则相似的句子应有相似的表征。由于句子表征来自词嵌入，我们可以推断出，用相似的数字表示相似的单词是神经自然语言处理中泛化的先决条件。例如，poured（倾注）、rained（下雨）、pouring（倾注）和 raining（下雨）的语义相似，所以它们的词嵌入应该相似；而 pouring 和 raining 的向量也应该更接近 driving 这类词，因为它们都是动名词，可能在相似语境中出现；poured 和 rained 应该处于邻近位置，因为它们都是过去式。高维向量表示的必要性由此可见：我们希望单词能同时根据不同判断标准彼此接近。

4.2 词嵌入的几何特性

词嵌入可以表示单词语义（或与语义相关的）特征。如前所述，词嵌入由实数组成（通常是成百上千个），每个数字似乎都表示词义的某个方面。例如，单词 Dublin 的词嵌入捕捉到的词义包括城市、爱尔兰首都和多家跨国公司的欧洲总部所在地等。

由于词嵌入不同维度具有专门含义，我们可以对词嵌入进行算术运算，并得到有意义的结果。两个词嵌入的对应向量值两两相加（或相减），例如 $[1.24 \times 2.56 \times 5.23] + [0.12 \times 1.12 \times 0.01] = [1.36 \times 3.68 \times 5.24]$ 。以下为神经机器翻译系统通常学习的两个词嵌入运算：

$$\begin{aligned} [\text{king}] - [\text{man}] + [\text{woman}] &\approx [\text{queen}] \\ [\text{Dublin}] - [\text{Ireland}] + [\text{France}] &\approx [\text{Paris}] \end{aligned}$$

其中，方括号指的是词嵌入，而 \approx 表示运算得到的词嵌入约等于符号右侧的词嵌入。这可以理解为 king 之于 man，正如 queen 之于 woman，即男君主或女君主；而 Dublin 之于 Ireland，正如 Paris 之于 France，即国家首都。

5 使用注意力机制的语境词嵌入

词语并非在所有句子中都具有相同含义。例如，letter 一词可以指“字母”，也可以指“信件”，因此应该使用不同的词嵌入。其实，对于神经机器翻译系统来说，更有趣的是，可以根据这个词是指“情书”，还是“投诉信”，用不同的

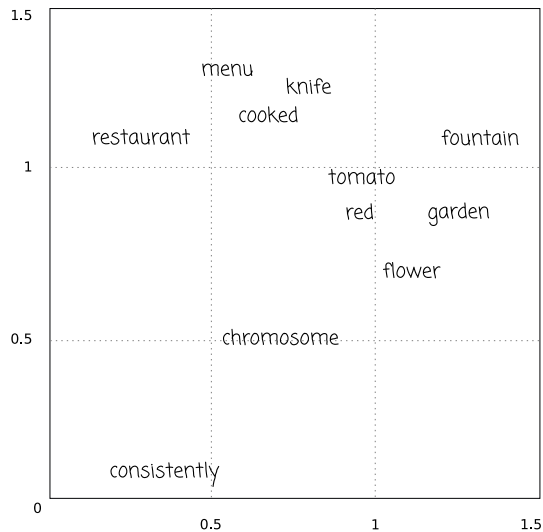


图 4: 在二维空间中, 意思相近的单词距离较近, 意思不近的单词距离较远。

词嵌入来表示。上文介绍的是“非语境”词嵌入：其计算主要根据词语是否经常同时出现在句子中，并没有考虑一词多义的情况。

在神经机器翻译领域，“注意力”机制起着重要的作用，使神经网络能计算“语境词嵌入”，也就是说，在不同句子中含义不同的词，其表征也随之改变，以此计算出词的向量表征。再次说明一下，“注意力”可以通过数学运算来实现，而训练算法能轻松学习这些运算。在这个语境下，“注意力”类似我们在日常生活中对某人某事的关注。

如以下示例，通过将“注意力”集中在某些单词上，season 一词在两个句子中的词嵌入向量有所不同：

1. The first episode will pick up right where the previous season left off.
2. Summer is the hottest season of the whole year.

原则上，这听起来好像语境词嵌入是为了让单词的不同含义有不同的表征，尽管通常可以这么理解，但并不仅限于此。在句子“Winter is the coldest season of the year in polar and temperate zones”、“Summer is the hottest season of the whole year”，甚至“Of the whole year, summer is the hottest season”中，season 一词的语境词嵌入都不同，尽管彼此可能比在句子“The first episode will pick up right where the previous season left off”中的表征更接近。这些差

异源自句子中的单词或其排列顺序不同。值得注意的是，单词 the 在上述两个例句中的语境向量也不同，因为其所在的语境不同。

如何使用“注意力”机制对语境词嵌入进行数学计算？以上述第二个句子 (Summer is the hottest season of the whole year.) 为例，计算过程从得到（第 4 节中介绍过的）“语境无关的词嵌入”开始。该句含有 9 个单词，可得到 9 个向量。基于此，为了计算 season 一词的语境词嵌入，神经网络以数学方式得出“注意力”向量。该向量包含 9 个百分比，用来表示句中每个单词所需的“注意力”程度，以便得到 season 一词的表征。向量元素与句子单词的“注意力”位置一一对应。例如，“注意力”向量 [25% 8% 10% 15% 25% 8% 2% 0% 7%] 表示，为了计算例句中 season 一词的语境向量表征，单词 summer 和 season 的词嵌入的重要程度都极高（共占 50% 的“注意力”）。这也合理，因为从语义的角度来看，它们都与气象季节的概念相关。请注意，season 前面的限定词也获得一定比例的注意力（10%），这可能是因为它有助于将 season 标记为名词。谓语动词（8%）对语境词嵌入的贡献在于把 season 标记为单数。所有百分比之和一定为 100%。

至于如何结合“注意力”向量和原始的非语境词嵌入，以获得新的词嵌入，这超出了本章的内容范围。简单来说，这个过程涉及特定的数学运算，得到新的词嵌入位于多个原始词嵌入之间的某个位置。

回到上述的例句，计算该例句的语境词嵌入需要计算 9 个注意力向量（每个单词都有一个向量），然后应用于原始的语境无关词嵌入，从而获得 9 个对应句子单词的新词嵌入。这些新的嵌入可以被认为是语境嵌入，因为它们受到句子其他单词不同程度的影响。

5.1 注意力层，多多益善

我们在本章的第 3.3 节讨论过利用多层神经网络的模型来接连优化计算的好处。类似地，为了得到更精确的表征，新的词嵌入也可以结合新的注意力向量，继续获得新的词嵌入。事实上，另一个发布于 2020 年的最大语言模型之一——图灵自然语言生成 (T-NLG) ——含有 78 个注意力层，可以“接力式”优化高达 4256 个维度的词嵌入。⁴ 回顾一下，这些运用多层网络习得的表征被称为“深层”表征。

5.2 注意力头，多多益善

在每一层神经网络中，每个单词都并不只有一个注意力向量。以句子 “My grandpa baked bread in his oven daily”（我的爷爷每天都在他的烤箱里烤面

⁴“Turing-NLG: A 17-billion-parameter language model by Microsoft”, 2020. 检索于 <https://www.microsoft.com/en-us/research/blog/turing-nlg-a-17-billion-parameter-language-model-by-microsoft/>

包)为例,单词 oven (烤箱)的词嵌入很有意思,它不仅和 grandpa (爷爷)一词有关,说明这个烤箱属于一名老年人,还和 bread (面包)一词有关,表明面包是用烤箱烘焙的。只使用一个注意力向量就会导致表示不同注意力的信息混杂其中,可能不利于准确找出词嵌入所表征的那个单词的译文。因此,有些神经机器翻译系统会在每一层神经网络中,给每个单词分配不同的注意力向量,以此算出每个单词的不同词嵌入。可以说,每个词嵌入都是使用不同“注意力头”的计算所得。T-NLG 模型每层含有 28 个注意力头。因此,在该模型的最后一层神经网络,每个单词得到 28 个 4256 维的词嵌入。

5.3 自然语言处理中的语境词嵌入应用

词嵌入不仅是神经机器翻译的基石,也能助力许多其他自然语言处理的应用,如情感分析和自动摘要。举例来说,有些系统能够将包含产品评价的文本中的句子自动分为两类——正面情感和负面情感。首先,计算句中每个单词的深层语境词嵌入,然后将这些词嵌入输入结构更简单的神经网络,计算出表示句子情感倾向评分的结果,结果范围为 0 到 1 之间(例如,0.95 表示句子包含正面情感,0.2 表示句子包含负面情感,0.51 表示句子情感为中性)。这些系统通常使用人工标注的句子语料库来进行训练。目前,很多语言都可以免费使用基于数百万个句子的“预训练”模型,因此模型的词嵌入计算训练并非必须使用特定的语料库。

6 压轴登场:神经机器翻译

下面,我们会简单描述神经机器翻译的基本原理,希望前文的铺垫能帮助您更好地理解这一点。我们将主要介绍 Transformer 模型和循环神经网络这两种架构。

6.1 Transformer 模型:基于注意力机制的编码器-解码器

简言之,使用 Transformer 模型的神经机器翻译系统由两个模块组成:第一个模块计算源句子中每个单词的语境词嵌入,而第二个模块接连预测目标语句子的每个单词。前者称为“编码器”,后者称为“解码器”。为了预测目标语的单词,解码器会关注源句子所有单词的词嵌入,以及已经生成的目标语单词的词嵌入。这一整个架构称作 Transformer 模型 (Vaswani et al. 2017)。图5展示了有三层神经网络的编码器以及计算第二层和第三层的词嵌入使用的注意力程度。图6则在此基础上加入了解码器,展示了整个 Transformer 模型架构。

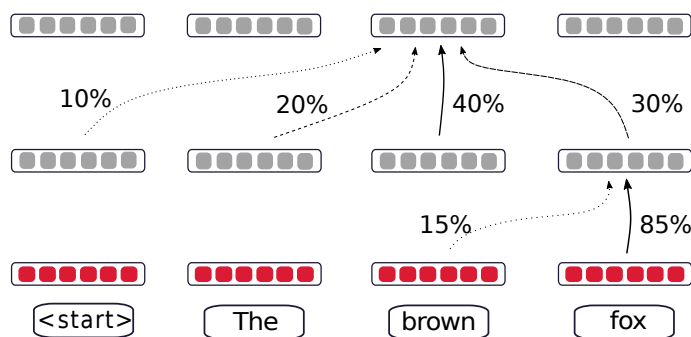


图 5: 基于 Transformer 模型的神经机器翻译系统的编码器。符号 *start* 通常为首标记句子的起点。从示意图中还可以看出, 单词 *brown* 和 *fox* 的第一层词嵌入对生成 *fox* 的第二层词嵌入的贡献程度不同; 同理, *brown* 在最后一层的词嵌入综合了第二层所有词嵌入的信息, 每个词的注意力程度各不相同。

学习算法使用平行语料库来获得 Transformer 模型所需的权重、词嵌入和注意力向量, 以便在一定程度上重现训练数据, 使机器翻译系统能够对训练集以外的句子实现泛化。

以句子“My grandpa baked bread in his oven daily”为例, 若要使用每层只有一个注意力头的 Transformer 模型将其译成西班牙语, 首先编码器会生成 8 个词嵌入向量。然后, 解码器算出一个八维的注意力向量, 如 [60%, 10%, 0%, 0%, 0%, 30%, 0%, 0%], 并利用此得到源语句子的信息, 从而获得目标语句第一个单词的词嵌入。假设机器翻译系统正确生成了西班牙语单词 *mi*。然后, 解码器算出一个九维的注意力向量, 如 [50%, 10%, 0%, 0%, 0%, 20%, 0%, 0%, 20%] (最后一个百分比为分配给目标语句第一个单词的注意力), 并基于该向量得到目标语句第二个单词的词嵌入。这个过程持续进行, 直至解码器生成表示句末的特殊标记。

解码器的每一步输出, 并不完全是对下一个词的词嵌入的估算。实际上, 解码器的末端会额外增加一层神经网络, 用以计算目标语词汇表中每个单词的概率或可能性向量。第 7.3 节将讨论如何使用这些概率来获得生成目标语句子的单词序列。

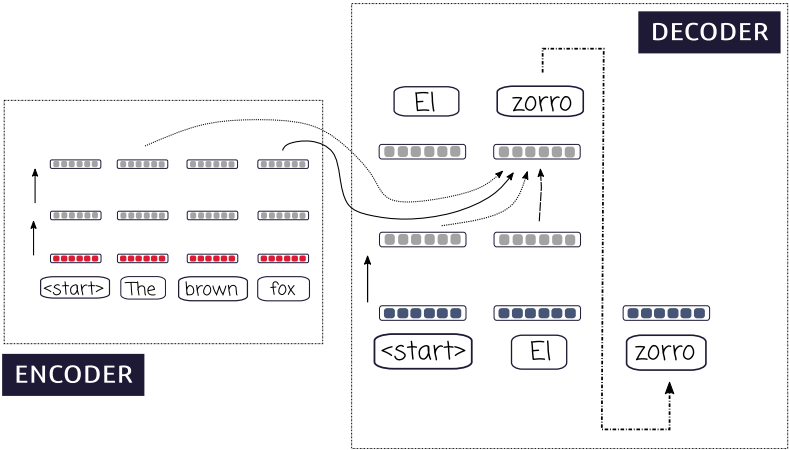


图 6: 基于 Transformer 模型的神经机器翻译系统的完整架构及其句子翻译过程。图5展示了放大的编码器。注意解码器是如何通过使用前一目标语单词的词嵌入, 以及编码器最后一层的一些源语单词的词嵌入, 来预测 zorro 一词。

6.2 循环神经网络

如上一节所述, Transformer 模型当前大多数用于商业神经机器翻译系统, 不过也有可以取而代之的其他神经网络模型。另一个顶级模型是“循环”编码器-解码器模型 (Bahdanau et al. 2015)。它和 Transformer 模型一样, 也有编码器和解码器, 前者为源语句子单词生成词嵌入, 后者通过整合源语单词和已生成的目标语单词的信息, 使用注意力机制来计算每个目标语单词的词嵌入。但是, 循环神经网络模型中的编码器和解码器以局部方式计算语境词嵌入, 例如第五个编码单词的词嵌入不仅基于前四个单词的词嵌入, 还基于下一个单词的词嵌入。这是通过从左到右、再从右到左遍历源语句子来实现的。图7所示模型只显示了从左到右的处理过程。

值得注意的是, 对于需计算语境词嵌入的单词附近的单词, 所使用的数学模型施加了相关性限制 (如例子中第五个单词), 因此该机制特别关注附近的单词, 而倾向于忽略较远的单词的表征。与 Transformer 模型类似, 解码器也会在最后一层网络算出一个向量, 该向量给出每个目标语单词是输出句子对应位置单词的概率。Forcada (2017)更详细地描述了循环神经网络编码器-解码器模型, 还讨论了神经机器翻译系统的输出类型。

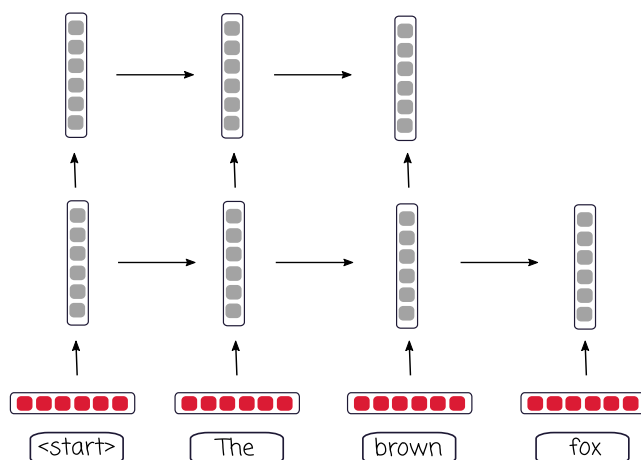


图 7: 处理 “<start>The brown” 之后, 即将处理单词 fox 的循环神经机器翻译编码器的子模型系统 (从左到右)。

7 其他设置

7.1 词与子词

根据本章以上内容, 不管是使用 Transformer 模型还是循环模型, 都可以在训练后得到每个单词的词嵌入。这是否意味着, 我们最终可以得到该语言所有单词的词嵌入? 不完全是。语言, 特别是高度屈折语和黏着语, 很可能有数不胜数的不同词形。为什么这对神经机器翻译系统来说是个难题? 因为词嵌入的数量 (即“词汇表”) 决定了神经网络的权重数量, 而且大型神经网络通常难以对未见过的数据实现泛化。只考虑出现在训练语料库的单词形式, 可以缩小词汇表, 但这通常仍意味着要考虑大量单词, 且还会衍生出新问题: 训练结束后, 神经机器翻译系统翻译训练集以外的新句子时, 这些未见过的单词会降低模型的性能和准确度, 因为在这种情况下, 每个未知单词都只有一个语境无关的词嵌入。

机器翻译算法工程师提出的解决方案是, 将单词分成所谓的“子词单元”。理想情况下, 这些字词单元应具有语言学意义并且能组合构成语义。例如, 将单词 demystifying 拆分为 “de-” + “-myst-” + “-ify-” + “-ing”, 肯定比拆分为 “dem-” + “-yif-” + “-yi-” + “-ng” 更具有语言学意义 (因此可能更有助于机器翻译)。但是, 以符合语言学规则的方式拆分单词的前提是该语言已存在拆分规则和步骤, 而很多语言可能都不满足这些条件。

常用的解决方法是通过历遍大量文本来自动学习拆分规则, 如训练集的所有源语句子和目标语句子。有种常见的方法⁵叫做“字节对编码” (byte-pair

⁵还有更先进的方法, 如 SentencePiece(Kudo & Richardson 2018), 它将整个文本视为一个字符序列, 并一次性进行单词切分 (“分词”) 和子词切分。

encoding, BPE), 从字母大小的单元开始, 若相连的某两个或更多字母经常出现在语料库中, 则视为一个单元。⁶ “字节对编码”的方法可能识别出许多频现的动词后缀-ing (如 *marching* 和 *continuing*), 然后将其删除, 即便是没有见过的形式 (如 *bartsimpsoning*)。就这样, 后缀-ing 会被转化为具有单元意义的语境词嵌入。

7.2 系统训练停止条件和译文质量评测指标

如第3.5节所述, 除了大型训练语料库, 通常还会预留一个不用于训练的小型“开发语料库”。该语料库的目的是监控神经机器翻译系统在训练过程中的翻译表现, 以决定何时停止训练。训练过程努力使误差函数最小化 (对于神经机器翻译, 实际上是使训练语料库的目标语句子的概率最大化)。还有可能出现的问题是, 在训练语料库上的过度训练, 降低了泛化能力, 因为这会导致神经网络最终“死记硬背”那些示例译文。这时, 开发语料库就有了用武之地: 在训练算法进行了一定次数的迭代或训练算法执行步骤后, 用神经网络来翻译开发语料库的源语句, 然后使用简单的近似自动评测指标 (参见 Rossi & Carré 2025 [本卷]), 自动对比机器译文与语料库的目标语句, 其中最常见的评测指标是 BLEU (Papineni et al. 2002)。BLEU 会算出机器译文有多少个一词、双词、三词和四词序列出现在参考译文中, 计算结果范围为 0% (零匹配) 到 100% (完全匹配)。在训练期间, 如果开发集的 BLEU 分数显示系统性能下降, 则可停止训练, 或者先保存当前的所有权重, 然后继续训练一段时间, 看看 BLEU 分数是否回升。当然, 除了 BLEU, 还可以使用很多其他的质量自动评测指标。

7.3 集束搜索

如第6.1节和第6.2节所述, 神经机器翻译系统中的解码器按顺序生成译文句子, 每次生成一个目标语单词。在每个时间步长, 神经网络为目标语词汇表中的每个单词得出一个概率 (范围在 0% 到 100% 之间)。获得这一概率后, 可以选择输出概率最大的目标语单词, 忽略其他可能性。值得注意的是, 这么做的话, 我们就完全确定了神经机器翻译系统的后续步骤, 因为当前的预测结果会作为下一步解码器的输入 (例子可见图6所示的单词 *zorro*)。还有一种方法或许可以考虑更多可能性, 例如, 考虑概率最高的三个单词, 并将系统“克隆”成三个系统, 分别由三个选项来决定, 并观察其运行过程。但是该操作不能无限重复进行下去, 因为每运行一步, 都会出现三个系统翻译同一个句子, 则系统数量会呈指数增长。为了避免这种情况, 只允许一定数量的系统“存活”, 即在近似计算中, 完整译文句子的概率最高的系统。这通常被称

⁶字节对编码最初是一种文本压缩算法: 频现的字母 (“字节”) 序列被存储一次并由短代码取代, 以减少所需的总存储空间。

为“集束搜索”(beam search), 这种近似方法也常见于人类语言处理(如语音识别)的其他概率模型中。

8 结语

训练一个神经机器翻译系统需要数百万个源语和目标语的句对。许多语言对、领域和文本体裁都不具备如此大量的资源, 限制了许多特定应用, 但是对于资源丰富的语言来说, 通用神经机器翻译已经成为现实且运用非常广泛, 用户不仅限于译者。此外, 近年来, 得益于多语言模型或无监督神经机器翻译等领域最新研究进展的出现, 低资源语言领域的研究也取得了可喜的成果。⁷

本章介绍了神经机器翻译系统的关键内容和技术细节, 并探讨其在“基于Transformer模型的神经网络”和“循环神经网络”这两种当前最流行的架构中的交互作用。在撰写本文时, 这一领域的研究非常火热, 几乎每个月都有新模型横空出世。如果有足够的平行语料库可用于训练, Transformer模型就是现有的最佳范式, 因为与循环神经网络相比, 前者所需的训练时间更短, 且能够细微优化质量。不过, 情况可能随时发生巨变。

References

- Bahdanau, Dzmitry, Kyunghyun Cho & Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In Yoshua Bengio & Yann LeCun (eds.), *3rd International Conference on Learning Representations, ICLR 2015*. DOI: 10.48550/arXiv.1409.0473.
- Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Dario Amodei, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner,

⁷ “多语言模型”是单一的神经网络, 以完成多个语言对之间的翻译为训练目的, 从而将资源丰富的语言的“知识”“转移”到资源匮乏的语言中。有趣的是, 多语言模型使“零样本翻译”(zero-shot translation) (Ko et al. 2021) 成为可能。例如, 即使没有西班牙语-上索布语的平行语料库可用, 我们可以先用德语-上索布语和西班牙语-德语这两个语料库训练出一个多语言模型, 然后机翻系统利用这个多语言模型模型来完成西班牙语-上索布语的翻译, 得到质量还不错的译文。“无监督的神经机器翻译”则更进一步, 只需学习单语语料库。

- Sam McCandlish, Alec Radford, Ilya Sutskever & Dario Amodei. 2020. Language models are few-shot learners. *CoRR* abs/2005.14165. <https://arxiv.org/abs/2005.14165>.
- Forcada, Mikel. 2017. Making sense of neural machine translation. *Translation Spaces* 6(2). 291–309.
- Goodfellow, Ian, Yoshua Bengio & Aaron Courville. 2016. *Deep learning*. Cambridge, MA: MIT Press.
- Hornik, Kurt. 1991. Approximation capabilities of multilayer feedforward networks. *Neural Networks* 4(2). 251–257.
- Ko, Wei-Jen, Ahmed El-Kishky, Adithya Renduchintala, Vishrav Chaudhary, Naman Goyal, Francisco Guzmán, Pascale Fung, Philipp Koehn & Mona Diab. 2021. Adapting high-resource NMT models to translate low-resource related languages without parallel data. In *Proceedings of the 59th annual meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, 802–812.
- Kudo, Taku & John Richardson. 2018. SentencePiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 66–71. Brussels, Belgium: Association for Computational Linguistics.
- Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg Corrado & Jeffrey Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems 30*, 3111–3119.
- Papineni, Kishore, Salim Roukos, Todd Ward & Wei-Jing Zhu. 2002. BLEU: A method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 311–318. Philadelphia, Pennsylvania, USA: Association for Computational Linguistics. DOI: 10.3115/1073083.1073135.
- 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.
- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser & Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems 30*, 5998–6008.

第八章

定制化机器翻译

Gema Ramírez-Sánchez
Prompsit Language Engineering

本章概述了以满足特定要求为目的的定制化机器翻译的理论意义和实践意义。本章的目标读者为机器翻译的初识者, 不过对于想要向外行人解释机器翻译的专业也会有所帮助。本章假设的机器翻译范式是神经网络机器翻译范式。

1 引言

1.1 通用机器翻译

毫无疑问, 大多数偶尔使用机器翻译的用户都依赖于“通用”机器翻译, 即机器翻译引擎的训练涵盖广泛的主题、文体和体裁, 而非专门针对特定领域。

通用型引擎虽然很适合一般用途, 但它很难用于词汇范围小、风格特殊, 或受特定体裁惯例限制的文本。这种情况多见于法律或医学等专业化程度高的领域的文本, 不过在日常生活中, 我们也会见到这种文本特点。例如, 食谱的特殊行文结构和用词有别于使用说明书等其他“日常”文本。例如, 与使用说明书不同, 食谱很少使用疑问句。这两类文本都经常被译成其他语言, 或供日常使用(如搜索引擎翻译“如何做巧克力饼干?”、“要节能的话, 推荐使用哪种灯泡?”), 或有专业用途(如出版商要翻译食谱, 或者制造商翻译消费品的技术规格说明)。在下文中, 我会以食谱(苹果酥)和消费者指南(灯泡)为例, 来展示通用机器翻译引擎如何处理这两类常见文本的术语。

食谱通常包括三部分: 标题、食材和做法。对于通用型机翻引擎来说, 光是简简单单的标题就足以叫人犯难。表 1 展示了三个通用型机翻系统给出的“apple crumble”(苹果酥)在西班牙语、法语和意大利语三种语言中的译法。从表中可以看出, 机翻 1 给出的法语和意大利语译文尚佳, 但西班牙语译文较差; 机翻 2 的三种译文都不尽如人意; 机翻 3 的法语译文不错, 但西班牙语和意大利语的译文都不理想。



表 1: 食谱标题的机器翻译: apple crumble

	机翻 1	机翻 2	机翻 3
西班牙语	migas de manzana 'crumbs of apple'	se desmorona la manzana 'the apple falls apart'	Desmoronamiento de la manzana 'falling apart of the apple'
法语	crumble aux pommes 'apple crumble'	Crumble d' apple 'crumble of apple'	Crumble aux pommes 'apple crumble'
意大利语	Crumble di mele 'apple crumble'	La mela si sbriciola 'the apple crumbles'	Crumble di mela 'crumble of apple'

不懂西班牙语、法语或意大利语的话，也没关系。你可以选一个喜欢的在线机翻引擎，把“apple crumble”译成自己懂的语言。得到的译文可能类似表表 1 中的 collapsing apples 或 apples that fall apart。其他译文都很好，如 crumble aux pommes 和 Crumble di mele（这就是为什么我们在表 1 中只简单将其注释为 apple crumble）。机翻引擎输出的译文取决于其训练数据。在此基础上，我们还可以添加一些步骤，如在训练过程或译后编辑阶段给引擎提供正确的术语。但是，如果不给引擎提供术语，通用机翻引擎往往就会产出如上文机翻 2 的低质输出。

下面，我们一起来具体分析灯泡的例子。对于要想了解市面上各式各样灯泡的消费者来说，有大量相关的专业信息可供参考。假设你是生活在英语国家的非英语母语者，有盏灯坏了，邻居想给你送个 twisted fluorescent lamp。你听不懂邻居说的话，只好用笑容回应。于是，你拿出手机，使用一些通用型机翻系统来查看翻译，得到的结果如表 2 所示。

看完这些译文，你以为邻居会给自己一盏造型奇特的常见灯泡或台灯。这一猜测由何而来？原来，是因为译文表达模糊不清：所有机翻引擎都没有使用表示“灯泡”这一含义的词来翻译“lamp”，而都采用该词的另一个含义 lighting equipment（照明设备）。

通用机器翻译翻错了，不过定制化机器翻译引擎应该能翻对。那么，什么是定制化机器翻译呢？请看下节内容。

表 2: 某种类型的灯泡: 螺旋形荧光灯泡。

	机翻 1	机翻 2	机翻 3	正确译文
西班牙语	Lámpara fluorescente retorcida	Lámpara fluorescente retorcida	Lámpara fluorescente retorcida	Bombilla fluorescente en espiral
法语	Lampe fluorescente torsadée	Lampe fluorescente tordue	Lampe fluorescente torsadée	Ampoule spirale fluorescente
意大利语	Lampada fluorescente contorta	Lampada fluorescente contorto	Lampada fluorescente attorcigliata	Lampadina fluorescente spirale

2 定制化机器翻译

与通用机器翻译相反,“定制化机器翻译”是为满足特定目的而设计的机器翻译。

假设你在一家生产大牌汽车的公司工作。和汽车行业的其他制造商一样,你的公司也要用几十种语言来产出大量技术手册、用户指南和营销资料。任何有助于公司提升与内部员工的沟通效率、培训汽车销售人员或吸引客户等,都是关键事项,在像这样的多语言环境中,机器翻译能够大显身手。因此,你公司使用机器翻译来翻译几乎所有资料的初稿,然后由审稿人,也就是所谓的译后编辑者(参见O’Brien 2025 [本卷])来改善译文。

你的公司最开始先使用通用机器翻译系统,然后通过译后编辑来提高机器译文质量。译后编辑者很快便发现,他们需要一遍遍重复修改相同的术语、体裁和风格错误,这一过程不但枯燥无味且效率低下。然后,考虑到这样的资料翻译已经进行了几十年,公司想知道,是否有办法能利用现有的翻译数据来改善翻译流程。

答案是肯定的。但是怎么做呢?首先,公司要用自己以往的翻译数据来训练机翻引擎,也就是说,用自己的训练数据来打造定制化机翻引擎。定制化机翻引擎翻译的初稿更接近公司过去的翻译,术语和风格方面的错误也少得多,译后编辑者也更高兴。

但是,事情真的就这么简单吗?是的,不过前提是你有足够多的数据(数百万个翻译过的句子),数据格式正确(对齐的平行语料;参见Kenny 2025 [本卷])、保持内部一致(否则产出译文不一致),且为所需的语言对。你还需要机器翻译算法工程师或外部供应商来训练系统,将其融入公司的翻译工作

流程以及合适的硬件和软件中。这一切还只是开始。然后,如果想继续利用新的翻译资料,就需要为系统重新训练进行规划:若采用自适应式机器翻译,这可以即时完成;若翻译产出量巨大,那么每六小时就需要重新训练一次系统;如果只为了更新或保持系统一致,则每六个月或一年一次。

所以,这一过程或许不能用“简单”来形容,你也许会想是否有必要如此大费周章?下面我们来看看对定制化机器翻译有哪些合理期待。

2.1 我们对定制化机器翻译的期待是什么?

定制化机器翻译曾专属于机器翻译专家,但如今已普遍为各种用户所用。我们甚至能看到定制化机器翻译输出的未经修订的原始译文,只需点击“获取翻译”便可一键生成。酒店的在线预订网站、专业软件的在线技术支持、招聘职位列表、教育应用程序中的教师信息等都可以看到定制化机器翻译的踪影。

作为普通用户,我们的主要目的通常是理解从某个渠道(通常是网站)检索到的信息。在这种情况下,我们可能希望至少在术语和习语方面,定制化机器翻译比通用机器翻译更准确,也希望能保持文本风格一致。

例如,上网查找疗养院的相关信息时,我们希望导航菜单中的“home”译为“主页”,而网页内容中出现的“home”译为“老年人照护机构”。而在浏览棒球网站时,我们期待定制化机器翻译能够区分“home”的不同含义。在其他语境中,“home”这个单词根本不应该翻译,如作为品牌名称的一部分。我们希望定制化机器翻译能更好地处理特定语境下的翻译。

对于语言专业人士来说,其他看似微小的细节尤为重要。他们希望当定制化机器翻译引擎可以正确输出单词的大小写形式、处理项目符号列表等格式惯例,或者恰当表示数字时,他们会感激不尽。反之(如表3),译后编辑者就需要仔细检查机翻引擎译文,纠正这些“小”错误,这既令人恼火又耗时。在专业环境中,我们也希望机翻译文能符合文本规范和问题要求。

如果训练力度足够且资源合适,那么定制化机器翻译的输出不会出现表3所示的错误。所需的资源包括合适的人力资源,我将在下一节讨论这一点。

2.2 谁来定制机器翻译?

我们会从这一小节了解到,机器翻译的确是个跨学科领域,需要计算机工程师与语言学家的共同努力。

在基于规则的机器翻译为领先技术的阶段,计算机工程师和语言学家都在机翻系统的搭建中发挥着积极的作用:语言学家编写语法规则和词典,计算机工程师则负责编写程序来执行规则。

表 3: 机器翻译输出的细节很重要

输入	输出: 通用机器翻译		
	西班牙语	法语	意大利语
For this match, the following players will be excluded	Para este partido, los siguientes jugadores serán excluidos:	Pour ce match, les joueurs suivants seront exclus:	Per questa partita, saranno esclusi i seguenti giocatori:
a) One	uno	un	un solo
b) Four	b) cuatro	b) Quatre	b) Quattro
c) 6	c) seis	c) Six	c) Sei

后来,到了统计机器翻译时代,工程师几乎负责整个机器翻译系统的搭建工作。语言学家偶尔参与输出评估,但很少参与错误分析,也参与制定改进机器翻译输出的具体措施。

然而,在统计机器翻译时代,这项技术仍然依赖于译者,因为他们提供的翻译为翻译模型提供了训练数据。¹但我可以肯定地说,在统计机器翻译的全盛时期,语言学家或多或少被排除在机器翻译系统的实际搭建工作之外。

如今,精通语言学的工程师逐渐不仅开始关注参数调整和硬件或自动评测指标,还开始关注机翻系统在高细粒度级别输出的译文质量。与此同时,精通技术的语言学家也已经开始参与评测和处理训练引擎所需的数据,使用训练工具包和评测系统,并制定策略改善系统,思考如何将其融入客户公司的翻译工作流程中。

鉴于机器翻译的本质,理想的情况是这两个领域的专业人员开展合作,有兴趣了解彼此的领域并为之做出贡献。这对于定制化机器翻译特别有效,因为有了语言学家评测训练数据的有用性、调控(或至少理解)训练过程遵循的策略,以及分析输出,对定制特定领域的机翻系统大有裨益。定制化机器翻译还受益于工程师对他们正在使用的文本和语言具体细节的理解,以便创造性地找到解决方案来推进和解决译文存在的主要问题,例如是否需要新模块、预处理或后处理步骤来处理多样的词语形态、产品名称或字母数字编码?

越来越多的学校针对语言学家和工程师增设机器翻译技术或语言学方面的培训课程,但这类培训主要还是出现在需求更加灵活变动的职业场景。

在本节结束前,还有一个群体值得提及,他们对定制化机器翻译的贡献与日俱增,但不一定会被称作“语言学家”,那就是“译者”。在这一章中,我主

¹下次听到有人说“机器翻译正在实现人类平等而‘无需人类干预’”时,请记住这一点。若真是如此,用来训练引擎的文本从何而来呢?

要将定制化机器翻译作为一种离线活动来介绍，只是偶尔需要人工干预。然而，值得注意的是，目前有些“自适应式”机器翻译的设置（参见O’Brien 2025 [本卷]）可以实时进行定制化（或至少比其他设置更快且更频繁）。在这种情况下，译者及其译文正成为定制化机器翻译的基石，而定制化实则为实时进行的用户模仿。当译者提供新的平行句对（源语-目标语句对），新的翻译会自动保存到已有系统中，成为译者可用的首选译文。

3 如何定制机器翻译引擎

3.1 一则寓言

假设你降落在新的星球上，没有生命迹象，但有座宏伟的图书馆，里面似乎藏有大量不同语言的文本：有些用这种语言（L1），有些用另一种语言（L2），还有一些是双语文本（L1-L2）。但是，由于没有关于语法、拼写和系统的语言学知识的书，所以无从知晓这些语言的使用方式，眼前只有大量句子或句对。

你还偶然发现了双语词汇表。此外，在查看这些文本时，你发现有些带印章，印章还各不相同，还有些不带任何印章。² 你感到庆幸，因为很可能只有通过这些文本，才可以了解这个新发现星球失落的生命遗迹。

但事情并非如此！等等！就在即将离开图书馆时，你发现地球上还有生命——两个星球居住者恶狠狠地盯着你，但你也注意到他们之间也相互怒视。过了一会儿，你发现他们敌意主要源自无法相互理解，一个说 L1 语，另一个说 L2 语，且都不知道图书馆的存在。你需要帮助他们！作为一个通晓多种语言的人，你也曾提供这样的帮助。现在，你别无选择，只能教会他们如何互译这两种语言。

你要怎么做呢？先单独学习这两种语言，再深入研究怎么进行互译？（这么做很耗时。）还是直接从学习双语文本和单词表开始？（这种方法似乎更好。）你可以从单词表开始，仔细观察其他单词、短语和更长的语块之间的关系。或者，可以把文本分为带印章和不带印章两类，例如将带有相似印章的文本放在一起。又或者，可以同时使用所有文本。如何从这些数据中学习两种语言的互译，你有很多方法可供选择。

在这一点上，学习如何翻译的机器翻译系统和你面对同样的处境：都有双语（和单语）文本，可能还有术语表，但仅此而已。这些是唯一的学习资源，不过学习过程多种多样。

回到新发现的星球，你首先发散思维，尝试从双语文本寻找学习思路，然后通过已编制好的单词表来验证自己的假设，接着在这些假设的基础上做出新的假设。你的观察对象很快就从单词发展到更长的语块。在这一阶段，你并不考虑文本是否带有印章，而是同时使用所有资源。

²在这个比喻中，不同印章代表不同领域，而没有印章的文本可被视为非特定领域的文本。

同样地，要新建一个机器翻译系统，通常先不分领域地把所有双语数据都串联起来，使用默认的软件设置进行初始训练。

在第一次训练后，你开始把 L1 语言使用者表达的信息译成 L2 语言，再让 L2 语言讲者来验证译文，然后在另一个翻译方向 (L2->L1) 执行同样的操作。你可以通过这两种语言使用者的面部表情来丰富语言知识。有时他们会哈哈大笑，但大多时候会点头示意，有时甚至看起来好像理解了你的译文。你从他们的反馈中学习并继续改进。

在机器翻译系统开发过程中，评测通常不是基于人类（或外星人）的评估。相反，我们使用自动评测指标来对比机翻系统译文和职业译者译文，从而计算出质量得分（参见 Rossi & Carré 2025 [本卷]）。在大多数情况下，机器译文与人工翻译越相似，质量就越高。如果机器译文的自动评测结果不错，且快速审阅后没发现严重的翻译错误，则该系统便可视为功能基准。否则，我们需要继续训练系统，比如增加一些预处理或后期处理。每轮训练结束后，都要用自动评测指标来检查，直至结果达到我们满意的水平，才会停止训练。

回到这个新发现的星球，你的学习取得了进展，还发现对于同一语言组合，有些单词虽然不止一种译法，但在带有相同印章的文本中的译法始终保持一致。因此，你决定按照印章将文本归类，并为每个类别分别制作译法列表。接着，你开始先后观察不带印章和带印章的文本，这两类文本略有不同，例如有些带章文本往往使用长句，而不带章文本则使用短句。

在这种情况下，根据拥有的数据量和系统的最终使用目的，我们可以只使用带相同印章的文本（印章表示领域）来训练机器翻译引擎。我们的域内系统可以用一般领域和特定领域的文本，也可以只用特定领域的文本。我们一定会充分利用最先进的机器翻译技术，尽量打造针对特定领域的系统。这正是定制化意义所在——玩转数据和技术。我们将在下文用简单易懂的方式来解释这些方法。有关神经网络机器翻译的“领域自适应”的更全面的研究，可参考 Saunders (2021)。

3.2 基于数据的定制

机器翻译系统可使用特定文本，来为特定用途进行调整，例如我们可以为移动电话的描述开发优质的机器翻译系统，只要拥有足够多的描述移动电话的文本和对应的目标语译文。我们也可以使用相关领域的单语文本或双语词汇表。这就是我们所说的“域内数据” (in-domain data)。理想的情况是，拥有语句平行对照³ 的双语域内数据。

³即源语-目标语的翻译句对，最好按照在原文出现的顺序排列，以便利用邻近句子的语境（参见 Kenny 2025 [本卷]）。

3.2.1 我们需要多少数据？

很难说要多少数据才算够。对于通用机器翻译系统来说，能够获取的数据越多越好，然后基于某些质量标准进行筛选，例如删除多次重复的句子（可以学习的内容太少）、内容杂乱的句子（如大多为数字的句子），以及太长的句子（学习难度太大）。而对于定制化机器翻译系统，答案可能相同，但还要考虑到域内数据必须占整个训练数据集的很大一部分，否则系统无法学习如何产出域内翻译。

这里特地只宽泛地说“很大一部分”，是因为我们知道很少有足够的域内数据来训练系统。毕竟，我们至少需要几百万个句对。这与通用引擎所需数据相比较少，但依然是相当庞大的数据量。因此，我们通常会结合使用域内数据和通用或域外数据。

定制化机器翻译系统的第一步如果是把可用的域内数据添加到域外数据，根据不同的语言组合，我们通常会遇到不同情况：数据要么太多，要么太少。而数据的规模很重要。

3.2.1.1 数据“过多”

数据过多时，系统训练会耗费大量时间，需要运行更多服务器，这样的操作有些不切实际。至于多少数据算多，很难给出确切的数字，假设你发现减少数据后结果相同，数据就算过多了。减少数据的好处在于，所需的计算资源和训练时间都更少。（进一步了解开发人员如何使用 BLEU 等指标来判断是否让引擎停止学习，请参见Pérez-Ortiz et al. (2025 [本卷])，尤其是7.2）。

开发定制化机器翻译系统，要优先使用所有能获取的域内数据。而对于域外数据，我们则选择可用的平行句对子集。通常以域内数据作为理想模板来筛选数据。数据的自动筛选方法有很多种，如：

- 根据与域内句子的语篇、语义或句法的相似度，对域外句子进行评分
- 根据在域内数据中发现的主题，对域外数据进行分组
- 根据优质或低质的域内句子，把域外句子重新分类为域内句子

3.2.1.2 数据“过少”

相反，数据过少时，如果无法获取更多的数据，则会有损机器翻译系统的性能。系统可能无法翻译很多词语，输出质量很差的译文。在这种情况下，首先要基于已有域内数据进行数据扩展。我们可以免费获取或考虑购买现成的数据库或定制数据。自动数据扩展的方法包括：

- 通过爬取包含目标语的多语言网站来获取更多双语数据，要格外留意域内数据所含词汇
- 使用更多的单语数据，最好是目标语，并通过第三方机器翻译系统将其回译至源语
- 使用中转语来翻译单语数据，首先将单语数据译为中转语，再译入我们所需语言对的目标语或源语
- 利用现有数据自动合成新句子（将单词替换成同义词或频率相同的词、采用自动改写等），从而创造出新的源语和目标语句子

经验表明，不仅双语数据，而且单语和多语数据，以及通用、域内和多域数据，都在帮助机器翻译系统学习方面发挥了作用（参见 (Saunders 2021)）。更重要的是，自适应或增量机器翻译都开始考虑使用少量数据（参见O'Brien 2025 [本卷]）。机器翻译的开发日新月异，但有一点是肯定的：只要有数据，就能学习，机器翻译系统会充分利用这些数据。

3.2.2 数据质量

数据的质量也会在机器翻译系统的定制中发挥作用，直接影响到最终输出的质量。尤其随着神经网络机器翻译的兴起（参见Kenny 2025 [本卷] 和 Pérez-Ortiz et al. 2025 [本卷]），数据的质量成为热门话题，因为有研究表明神经网络机器翻译对训练数据的噪声非常敏感 (Khayrallah & Koehn 2018)。大多数致力于解决这个问题研究包括结合使用模式和规则来过滤明显的噪声、句子质量评分，以及根据内容质量将数据分为优质和低质两类。此外，还有研究采用删除重复内容的方法 (Khayrallah & Koehn 2018)。

3.2.3 数据的结构

最后，数据的结构也是热议话题 (Mohiuddin et al. 2022)。有些研究使用长度相似的句子作为训练数据，以提高训练和翻译速度，有些则通过给模型输入由易到难的句子来提高质量，还有些使用成篇文档，而不是打乱句子，以便利用更大的语境来改善机器翻译质量。

3.3 基于技术的定制

整合清理了所有数据后，接下来可以进入域内机器翻译系统的训练环节。该如何使用这些数据？哪类系统架构最适用于我们的目的？我们有不同选项可用于约束输出吗？这些就是基于技术的定制所要讨论的内容。这些技术可能

涉及神经网络架构的调整、训练期间的参数调整，以及训练或翻译期间（也称为推理）不同系统的组合。

我会在下文介绍一些开发特定领域机器翻译系统的最常见技术。更多讨论可参考Koehn (2020: Ch. 13)。

3.3.1 自学系统

对于自学系统，我们最初仅用通用型数据来训练，然后以此为起点，仅用域内数据进行第二次训练。这种方法训练出来的微调系统既有一般性语言知识，也有域内词汇和结构等专业知识。

3.3.2 受导系统

除了使用通用和域内数据来训练机器翻译系统之外，还可以使用基于目标语域内文本训练而成的语言模型，不过起码是高质量的单语文本。该语言模型有助于生成更符合域内语言特点的文本。此外，也可以使用域控制器或鉴别器，给训练数据进行单词、句子甚至词嵌入级别的标记。具体来说，就是先精准识别哪些通用数据与域内数据相似或不同，然后利用这些信息进行训练。

3.3.3 合作系统

有些系统由若干部分组成（如特定领域的子神经网络）或甚至可以包含多个完整系统。它们结合所有系统的知识可以生成比单个系统质量更好的文本。

4 实践中的定制

理论和实践往往不可分割。本节会介绍神经网络机器翻译引擎定制化的实践细节，内容主要针对初学者，不涉及复杂的技术知识。

4.1 可用工具

机器翻译定制化的专业工具已有好几种可供选择，但大多只允许用户改变数据，不可以调整技术。因此，虽然很多语言专业人士都有能力进行基于数据的定制化，但基于技术的定制化仍然只能是机器翻译研究者和开发者的专业领域。

大多数机器翻译供应商以定价方式提供对预训练通用或特定领域机翻引擎的远程访问，有些还提供定制服务。⁴ 定制服务通常包括：

⁴在撰写本文时，有 40 多家供应商提供机翻服务，约 20 家提供一些定制服务。来源：<https://inten.to/mt-landscape/>，最后访问于 2022 年 6 月 26 日。

- 添加用户的语料库
- 添加用户的术语
- 利用用户的数据训练新模型
- 测试定制化模型

以上内容都可以半自动方式完成，即人工处理数据和相关流程，或以全自动方式完成，即定制时无需人工干预。

还有专为教授语言专业人士如何使用机器翻译而设计的机器翻译测试环境。这些测试环境作为教学工具，用于翻译技术课堂或职业场景，通常提供定制化选项，以观测系统在使用通用或域内数据训练时有何变化，如 MutNMT 就是一个很好的例子。⁵

4.2 策略制定的关键因素

以下是机器翻译系统定制策略的最重要因素：

- 语言对
- 领域
- 可用数据
- 系统用途
- 定制期限
- 硬件特性和可用性

下面，我们来一一了解。

你可以根据语言组合来决定使用哪种工具。例如，形态变化丰富的语言适合使用支持预处理和后期处理的工具。

有些领域的语篇特点突出，需要特定训练工具包的支持。例如，某个特定训练工具包可能非常适合处理很长或很短的句子、数字表达，以及通常需要在译文中保留的专有名词。如果该领域还有其他特点，也需要一并考虑。

定制系统时，可用数据的数量和质量不仅决定训练的进度，还决定是否需要增加数据或提高数据质量。此外，如果想添加平行句对以外的数据，必须确保有技术支持。

⁵参见 <http://www.multitranmt.eu/index.php/en/neural-mt-training/mutnmt>，最后访问于 2022 年 6 月 26 日。

大多数神经网络机器翻译系统都致力于产出最好的译文，但是这通常还需要满足其他要求。选择服务供应商前，可能需要先考虑其语言模型是否支持远程访问，或是否真的需要在本地进行访问。该系统是否可以与其他用户同时在线使用，还是需要排队？该系统是否能翻译文本字符串？还是需要不同文件格式来获取翻译？是支持 API、网络应用程序，还是通过第三方工具的连接器进行访问？诸如此类问题。

机器翻译系统的定制规划可能需要权衡质量和交付期限。训练出最佳系统所需要的时间，可能远超你的预期。

最后，无论是系统的训练，还是其使用，硬件都是关键因素。综合考虑其他因素，你可能需要全天候的服务，且同时使用多个 GPU 或 CPU。

4.3 获取正确数据

机器翻译系统的定制在很大程度上依赖于域内平行数据的可用性和质量，且这些数据要接近需要翻译的原文。在3.2.1中，我讨论了数据规模，以及如何根据场景来选择或扩展数据的方法。在此，我想讨论一个很基础的问题：在哪里可以找到适合系统的平行数据？

首先，可以获取现有的免费平行数据。有很多公开的免费数据库提供可用于机器翻译的双语语料库。这类数据库或语料库大多可在由赫尔辛基大学维护的 OPUS 网站获取，该网站提供 200 多种语言组合的语料库。⁶

其次，可以购直接买平行数据（出售供机器翻译使用的平行数据也是一门生意）。待售的数据规模大小不一，使用权、商业模式和定价也不尽相同。TAUS 是最大的机器翻译数据集供应商，提供覆盖 600 多种语言对的平行语料。⁷

此外，还可以自行构建平行数据。虽然难以搜集到大量数据，但还是可以通过以下方式构建平行语料库：

- 爬取多语言网站：可以通过包含双语内容的 URL 下载所需语言对的双语文本，并自行对齐（对齐文本的例子可参见Kenny 2025 [本卷]），或使用第三方服务进行对齐。
- 对齐自己翻译过的文本：如果有翻译过的文档及其源语文本，可以使用标准的开放获取或专有工具进行句级对齐。⁸

⁶参见 <https://opus.nlpl.eu/index.php>，最后访问于 2022 年 6 月 26 日。

⁷参见 <https://www.taus.net/>，最后访问于 2022 年 6 月 26 日。

⁸免费对齐工具有 LF Aligner (<https://sourceforge.net/projects/aligner/>，最后访问于 2022 年 6 月 26 日)。有些为人熟知的付费对齐工具还提供翻译记忆库工具（参见 Kenny 2025 [本卷]）。

4.4 正确处理数据

获取正确数据后，我们需要在训练系统前进行数据准备，确保：

- 每个文件只含单语（或语言对）语料，因为机器翻译训练常使用纯文本文件。理想的情况是每个文件都是单语的，尽管有些系统可以处理分列对照的电子表格形式（每列一种语言），甚至 TMX 文件（参见Kenny 2025 [本卷]）。
- 每行一句：文本文件为每行一句，也就是说一句话单独成段。通常不允许出现多句或不成句的情况⁹，不过标题也可视为句子，独占一行。
- 句子行行对齐：无论哪种格式，源语文件与目标语文件都必须行行对齐。
- 数据清理干净：无重复、拼写错误或含噪音的句子（如句子仅含数字、出现乱码或出现其他语言）。
- 数据经过匿名化处理（如有必要）：要删除训练语料库中的任何敏感数据，尤其是个人数据。
- 数据结构清晰：语料库需按训练的不同阶段分为三个数据集，通常命名为“训练集”（training set，简称 train）、“验证集”（validation set，也称为“开发集”（development set，简称 dev）和“测试集”（test set，或简称为“test”）。

建议在整理数据时应该：

- 必须避免这三个数据集出现重叠。其实，如果可能的话，每个数据集应使用来源不同的句子，以保证其平衡和独立性。
- 训练集的规模从几千到几百万个句子不等，但验证集和测试集通常不会超过 5000 句。
- 训练数据可能包含通用型和定制化数据，但验证集和测试集应尽量使用域内数据，以便测试训练模型是否能够成功翻译域内文本。
- 若训练集同时含有通用数据和域内数据，应尽量提高域内数据的比例，否则模型将主要从通用数据中获取知识。

至此，我们讨论过的内容都并非为定制化引擎特有。这些准备工作均适用于通用型和定制化机器翻译的训练。下面将要介绍的预处理相关步骤也是如此，但可能因特定语言或语言组合而异：

⁹此处的训练单位为句子，而非文档。

- 文本分词：需要为训练提供分词准确的文本。分词单元 (token) 指文本分成的不同单位，包括单词、空格或标点符号。分词主要是识别词的边界，有时需要判断词的起止位置，在笔语中通常以空格为界。但是，泰语等语言的词语之间没有空格，这就增加了分词的难度。有时，这类语言的分词器只是将句子切分为看似无明显词界限的字符序列。这种方法虽不考虑“意义单位”，但优点是适用于任何语言。
- 文本大小写的处理：我们或许希望系统通过训练学习到，在训练数据中，句首首字母大写单词 (如 The) 与该词在句子其他位置的小写形式 (如 the) 是同一个词。因此，我们使用大小写转换器 (truecaser) 将 (在英语等语言中的) 专有名词之外的所有单词转为小写。¹⁰ 大小写转换仅适用于区分大小写的语言，因此不适用于中文、阿拉伯语、希伯来语等语言。
- 子词拆分：根据文本分词情况，还可以进一步将单词拆分为子词、字符或其他语块。这种拆分可以基于频数或具有语言学意义的单元，如语素、词干和附加的形态学信息等。但是，并非所有定制化系统都拥有支持这种方法的配置环境，因为它要求神经网络能够处理特殊类型的输入和输出，而预处理和后处理通常只支持一种方法。

那么，如果你想使用的训练数据格式不对，该怎么办呢？别着急，大多情况下，分词、大小写转换和子词拆分 (若适用) 均为标准训练工具包预训练、预处理和后期处理步骤的默认环节。之所以在此提到这些环节，是为了让用户对训练工具包有所了解。还有很多独立工具可以实现这些步骤。这类工具可通过 Hugging Face ¹¹ 或 Github ¹² 等平台获取，或直接在搜索引擎上查找“句子切分器” (sentence splitter)、“句子对齐工具” (sentence aligner)、“平行语料库过滤” (parallel corpus filtering)、“匿名化工具” (anonymizer)、“分词器” (tokenizer)、“大小写转换器” (truecaser) 等。要进行精准搜索，还可以添加源语和目标语。通过以上方法，你会发现很多工具。

4.5 训练定制化模型

获取格式正确的适用数据后，在目前的很多训练环境中，训练机器翻译模型只需点击一个按钮。有些环境允许用户调整许多参数，以充分利用数据训练环境和系统架构。有时，用户可以根据教育或研究目的来调整设置。

在训练前，用户最常调整的参数如下：

¹⁰注意，在德语中，所有名词的首字母都要大写，因此这些名词跟英语专有名词一样，不应该进行大小写转换。

¹¹<https://huggingface.co/>

¹²<https://github.com/>，最后访问于 2022 年 6 月 26 日

- “词表大小” (Vocabulary size) 规定从训练语料库中计算得出的不同单词或子词 (也称子词单元、类型或单词类型) 的数量。
- “批量大小” (Batch size) 指每个训练步骤一并处理的分词单位¹³的数量。这一步很有必要, 因为不可能将训练集的所有数据一次性输入神经网络。
- “波束大小” (Beam size) 是在翻译一个单词时, 需要考虑的翻译假设 (即候选译文) 的数量。翻译假设产生于训练过程和系统的实际翻译过程。
- “训练时长” (Duration) 是指训练轮次的数量。一“轮” (epoch) 指完整训练一次训练集的所有句子。每一轮都包含历遍全部数据所需的所有训练步骤。¹⁴
- “验证频率” (Validation frequency) 指每次评估训练情况前的训练步骤数量。通常来看, 每一轮训练都会进行多次验证。
- “停止条件” (Stopping condition) 规定在机翻引擎性能没有改善的情况下, 所允许的最大验证执行次数。若达到预设的最大验证次数, 无论轮次初始设置为多少, 系统都会停止训练。训练效果可通过任何单个或多个自动评测指标来衡量。常见的自动评测指标有 BLEU、chrF1 和困惑度 (perplexity)。¹⁵

这些参数通常都设置了默认值, 这些默认值是开发者在为特定环境优化训练过程后设置的。

设置好参数或采用默认参数后, 训练流程如下: 每次迭代都将一批训练数据输入神经网络, 计算该批数据每个句子的输出译文, 计算误差损失, 更新权重, 然后再重复上述过程处理下一批数据! 完成预先设定的迭代数 (“验证频率” 参数设置) 后, 评测机翻引擎性能, 然后继续训练。若更多的训练未能改善引擎性能或性能变差, 则训练停止。模型训练完成!

¹³近几年, 以“分词单位” (参见Rossi & Carré 2025 [本卷]) 代替句子来计算批量大小已成为最常用的批量类型, 以使不同批量的大小更接近。

¹⁴考虑到神经网络机器翻译需使用大量数据, 通过轮次来衡量训练时长可能不切实际。除了轮次, 还可以使用与特定批量大小有关的步骤数, 来单词计算模型、语言对或数据量来的训练时长。

¹⁵在自然语言处理中, 具体到机器翻译, 困惑度用于衡量翻译模型在翻译时预测下一个单词的不确定程度。若翻译模型为给定目标语句子中的每个单词/分词单位赋予高概率值, 则为低困惑度。欲进一步了解 BLEU 和 chrF1, 请参见Rossi & Carré (2025 [本卷])。

4.6 测试定制化模型

机器翻译系统训练结束后,需要对其进行测试或评测。测试方法总结如下:

- 自行(或请别人)测试!如果你会说正在研究的语言,了解训练系统的用途,且有足够的时间,不妨用系统来翻译一些句子,看看质量如何!使用正确的工具,不仅可以查看到系统的最佳译文,还可以得到它为每个句子提供的 n 个候选译文列表。
- 评测质量!自动评测指标可以计算系统的性能得分(参见Rossi & Carré 2025 [本卷])。大多数指标都采用对比系统译文与专业译者的“参考”译文的方法。这些指标的意义有时差别较大:有的虽然适用于比较两个不同的系统,但数字本身没有太大意义,例如基于 n -gram 或字符的评测指标,包括 BLEU(Papineni et al. 2002)、METEOR(Denkowski & Lavie 2014) 和 CHRF1-3(Popović 2015)。还有的指标有助于测量将机器译文变为专业译文所需的工作量,如 WER(Popović & Ney 2007) 或 TER(Snoover et al. 2006) 等用来测量译后编辑工作量的指标。此外,还有一些指标能够体现机器译文的特点。文本层面的测量指标涉及词汇的多样性或密度。最后,有些指标用于对系统进行排名,指出你的系统是否优于其他系统。
- 从实际使用中获得反馈!根据系统用途,你可以评估该系统在真实的专业或日常使用场景中是否有用。你训练系统是为了帮人写电子邮件吗?是的话,那就让人用它来写邮件,然后向你反馈使用体验。你训练系统是为了帮人看懂食谱吗?是的话,那就让人按照系统翻译的食谱做菜,然后给你反馈。你训练系统是为了翻译法律文件吗?是的话,那就让人用来翻译,然后跟你分享使用心得,如系统是否能节省翻译时间等。通过收集这些反馈,你不仅能了解当前的系统状况,还能提供改进系统的思路。

最后,历尽艰辛训练出定制化机器翻译系统,你可能想通过前文所述的测试方法,来对比自己的定制化系统和通用系统的译文。如果前者优于后者,那就算成功了!否则,再接再厉吧!

5 结语

本章概述了如何定制开发机器翻译系统。作者首先区分定制化和通用机器翻译系统,强调了管理定制预期的重要性,然后介绍了定制开发神经网络机器翻译系统所涉及的专业角色,并思考机器翻译如何融入翻译工作流程。此外,本章还讨论了基于数据和技术的定制方法,建议将其与现实生活中的学

习过程进行类比。最后, 作者从实践角度出发, 介绍了相关工具、定制策略、数据搜集和准备、系统训练和结果测试, 以帮助读者获得机器翻译系统定制化的实践经验。

References

- Denkowski, Michael & Alon Lavie. 2014. Meteor universal: Language specific translation evaluation for any target language. In *Proceedings of the ninth workshop on Statistical Machine Translation*, 376–380. Baltimore: Association for Computational Linguistics. DOI: 10.3115/v1/W14-3348.
- 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.
- Khayrallah, Huda & Philipp Koehn. 2018. On the impact of various types of noise on neural machine translation. In *Proceedings of the 2nd workshop on neural machine translation and generation*, 74–83. Melbourne: Association for Computational Linguistics. <https://aclanthology.org/W18-2709.pdf>.
- Koehn, Philipp. 2020. *Neural Machine Translation*. Cambridge: Cambridge University Press.
- Mohiuddin, Tasnim, Philipp Koehn, Vishrav Chaudhary, James Cross, Shruti Bhosale & Shafiq Joty. 2022. *Data selection curriculum for neural machine translation*. DOI: 10.48550/ARXIV.2203.13867.
- 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.
- Papineni, Kishore, Salim Roukos, Todd Ward & Wei-Jing Zhu. 2002. BLEU: A method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 311–318. Philadelphia, Pennsylvania, USA: Association for Computational Linguistics. DOI: 10.3115/1073083.1073135.
- 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.
- Popović, Maja. 2015. Chrf: Character n-gram f-score for automatic MT evaluation. In *Proceedings of the tenth workshop on statistical machine translation*, 392–395. Association for Computational Linguistics. 10.18653/v1/W15-3049.

- Popović, Maja & Hermann Ney. 2007. Word error rates. In *Proceedings of the Second Workshop on Statistical Machine Translation* (StatMT '07), 48–55. Prague, Czech Republic. DOI: 10.3115/1626355.1626362.
- 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.
- Saunders, Danielle. 2021. Domain adaptation and multi-domain adaptation for neural machine translation: A survey. *CoRR* abs/2104.06951. <https://arxiv.org/abs/2104.06951>.
- Snover, Matthew, Bonnie Dorr, Rich Schwartz, Linnea Micciulla & John Makhoul. 2006. A study of translation edit rate with targeted human annotation. In *Proceedings of the 7th conference of the Association for Machine Translation in the Americas: Technical papers*, 223–231. Cambridge, Massachusetts: Association for Machine Translation in the Americas. <https://aclanthology.org/2006.amta-papers.25/>.

第九章

机器翻译之于语言学习者

Alice Carré^a, Dorothy Kenny^b, Caroline Rossi^a, Pilar Sánchez-Gijón^c & Olga Torres-Hostench^c

^a 格勒诺布尔-阿尔卑斯大学 ^b 都柏林城市大学 ^c 巴塞罗那自治大学

机器翻译在二语和外语的学习中一直存在争议,但在某些情况下,有策略地使用机器翻译可能有利于语言学习。我们会在本章讨论语言学习过程中如何运用机器翻译,提出机器翻译的数字替代方案,并举例说明机器翻译如何为语言学习者提供帮助。

1 引言

在二语和外语学习中¹,机器翻译的使用一直存在争议,部分学者认为这会助长抄袭行为、造成更多错误或影响正常的学习轨迹。但是,机器翻译有时却能帮助学习者完成某些学习任务,可以视作当代语言学习者的众多数字资源之一。机器翻译与语言学习的成功结合要求我们了解这项技术的基本工作原理、机器译文质量测评方法、如何通过译前和译后编辑来提高输出质量,以及机器翻译使用的伦理问题等。这些知识和技能通常属于“机器翻译素养”(Bowker & Ciro 2019)的范围,已在本书的第2章到第6章中做了详细阐述。与此同时,Torres-Hostench (2025 [本卷])还就“为什么机器翻译和语言学习一样,都应被视为多语言社会的重要组成部分”提出了具有说服力的观点。到目前为止,本书还未深入探究如何在语言学习中使用机器翻译,而本章的目的正是填补这一空缺。本章先介绍翻译在语言学习中的作用,然后讨论机器翻译是否也能用于语言学习,以及这种做法有何好处。接着,我们会根据使用场景,为语言学习者提出机器翻译适用和不适用的语言学习场景,

¹请注意,本章以通用术语“语言学习”(language learning)和“语言学习者”(language learner)来指代外语学习、第二语言和随后的语言学习。学习者的第一语言为L1,其他语言则相应为L2、L3、Ln。



并介绍更适用的数字资源。最后，我们会举例说明如何在语言学习环境中使用机器翻译。

2 语言学习与翻译

在讨论机器翻译对语言学习的作用之前，值得一提的是，这个话题长期以来饱受争议。“语法翻译法”（即要求学习者先学习词汇和语法规则，再翻译没有给出语境的句子）备受批评，人们认为借助翻译进行语言教学具有局限性，但这种看法本身对翻译的理解就很狭隘。与此同时，Cook (2010)重新激发了学界对“语言教学中的翻译”（Translation in Language Teaching, TILT）的关注，而且“在过去十年，各学科中有关翻译在语言课堂上的运用的研究成倍增加”（Pintado Gutiérrez 2018: 12）。研究发现，借助翻译学习语言不仅有助于提高“多语言、多文化和语言交流的能力”、写作技能、语言意识和掌控能力（同上：13），还能减轻语言学习者的焦虑和认知负荷（Kelly & Bruen 2017）。更多有关 TILT 的运用案例可参阅Noriega-Sánchez et al. (2021)，作者介绍了如何将翻译相关活动融入语言教学的最新动态。

3 语言学习与机器翻译

3.1 运用机器翻译学习语言可取吗

2010 年左右的研究表明，机器翻译完全被排斥在当时的语言课堂之外。Pym et al. (2013)对全球教师的调查结果发现，极少教师会在语言课上使用机器译文。其中一名研究对象表示，机器翻译与不愿使用翻译进行语言教学之间似乎也存在因果关系——在中学，不使用翻译的原因往往是“害怕学生依赖机器翻译工具”（Pym et al. 2013: 93）。

不过，无论老师怎么想，使用免费在线机器翻译（Free Online MT, FOMT）的语言学习者越来越多，其使用效果值得研究。Lee (2021)和Jolley & Maimone (2022)等学者对机器翻译在语言教学中的使用进行了综述，发现很多人将使用机器翻译视为作弊行为。于是便有学者撰写如何“检测”学生是否在 L2 写作中使用机器翻译。以前，机器翻译的质量一般较差，可以通过识别机翻译文的典型错误轻松判断学生是否使用了机器翻译。但如今，研究人员认为：

随着机器翻译技术的不断进步，对于语言教师来说，识别翻译“错误”可能愈发困难。不过，帮助识别机器翻译使用的，正是这项技术的成功，而非缺点。（Ducar & Schocket 2018: 787）

换句话说，被发现使用机器翻译的学生露出马脚并不是因为他们的写作错误百出，而是因为机器译文质量超出了学生的实际水平。例如，若法语初学者使用了高阶课程才学到的虚拟动词形式，他就有可能使用了机器翻译。

但一个人是否作弊的判断标准不在于所使用的技术,而在于“游戏规则”。如果在 L2 写作中不允许使用机器翻译,但学习者还是偷偷使用,这就是作弊行为。即便没有明确禁止使用机器翻译,但学生还是在老师不知情的情况下使用,并将机器译文当做自己的写作成果,这也是投机取巧的做法。事实上,将他人的观点据为己有属于“剽窃”行为, Mundt & Groves (2016)就机器翻译在语言学习的使用这一背景探讨了这个问题。然而,许多研究(如Correa 2011, Clifford & Munné 2013, Ducar & Schocket 2018)表明,根据机器翻译的使用程度等诸多因素,学生和教师对在语言学习中使用机器翻译持不同态度,因此情况并不是非黑即白。

如果你是在正规的环境中学习语言,最好先向老师了解清楚哪些是作弊行为。如果你是语言老师,最好先跟学生明确学习要求。无论如何,明确在完成语言类学习任务时是否能够使用机器翻译,避免作弊只是需要考量因素中的一个。其他需要考虑的因素请参阅本章节 3 的“情境参数”和 Moorkens (2025 [本卷])有关道德规范的阐述。我们接下来的讨论重点在于,无论老师是否同意,使用机器翻译会有哪些好处。

3.2 在语言学习中使用机器翻译的好处

有证据表明,使用机器翻译来完成特定任务的学习者能在短期内取得进步。例如,在 O'Neill (2019)的研究中,310 名具备中级水平的美国大学生在不同的实验条件下用法语和西班牙语进行写作,分为“可使用谷歌翻译且事先经培训”、“可使用谷歌翻译但事先无培训”、“可使用一款在线词典且事先经培训”、“可使用一款在线词典但事先无培训”和“无任何技术支持”。其中可使用谷歌翻译且事先经培训的学生的作文得分最高,其次是可使用一款在线词典且事先经培训的学生。这些学生在一周和三到四周后又分别进行了两次测试,期间不能使用上述工具,结果显示“谷歌翻译 + 培训”组的学生得分并不高于其他组。这说明他们凭借工具取得的优势只能短暂存在,且还要取决于能继续使用工具。

在另一项研究中, Fredholm (2019)对 31 名外语为西班牙语的瑞典高中生进行了为期一学年的跟踪调查,研究其作文的词汇多样性,其中大约一半的学生使用纸质词典,另一半使用谷歌翻译。他发现,机器翻译的使用与更高的词汇多样性有关,只要学生能够继续使用谷歌翻译,得分就更高,可一旦不能使用,相对优势就消失殆尽。同样,机器翻译的益处似乎要靠一直使用才能获得。

这是否意味着学习语言没必要使用机器翻译呢?不完全是。在上述两项研究中,从长远来看,使用机器翻译工具对学生没有坏处,只不过一旦不能使用工具,使用和不使用机器翻译的学生在写作方面的差别就不明显了。不过在短期内,使用机器翻译的学生的文章写得更好。因此,使用机器翻译是否

有好处,似乎取决于是从短期还是长期的角度来看,以及使用目的是为了完成某个特定任务,还是持续学习语言。

通过O'Neill's (2019) 的研究还发现,(至少在短期内)工具使用的培训很重要。受过机器翻译工具使用训练的学生,即便只是短暂的训练,也比没有受过训练的写作更好。

3.3 机器翻译对语言学习到有何益处

我们已经看到,使用机器翻译可以帮助一些学生提高整体写作水平,或者有助于提高写作中的词汇多样性。有关使用机器翻译对二语(或三语)写作的影响的研究重点关注学习者对词汇、语法和句法的使用。Lee (2021)指出:

大量研究证实,机器翻译有助于学生减少拼写、词汇和语法错误,更专注于内容,因此,使用机器翻译的学生的修改成功编辑次数更多,二语写作质量更高。(Lee 2021: 4)

但应该注意的是,个别研究的结果却截然不同。例如, Fredholm (2015)发现,使用免费在线机器翻译的瑞典学生在进行西班牙语写作时,比起没有使用的学生,出现的拼写和冠词/名词/形容词配合方面的错误更少,但句法和动词变化方面的错误更多。

其他研究还关注学生的“元语言意识”(metalinguistic awareness),即:

将语言当做客体,关注、思考并评价语言本身的能力。(Thomas (1988: 531) in Thue Vold 2018: 67)

Enkin & Mejías-Bikandi (2016)在一项研究中设计了一个练习(但并未进行测试),在这些练习中,学生接触到将英语容易出错的特定句法结构,翻译成西班牙语的机器翻译,而由于这两种语言的对比差异,机器翻译的效果通常较差。例如,非限定从句翻译成西班牙语时,最好使用限定从句。这个练习的目的是,学生可以思考机器翻译的错误,从而强化元语言意识,尤其是对比意识。然而,作者指出,随着机器翻译质量的提高,“教学资料或许也要跟上步伐”(Enkin & Mejías-Bikandi 2016: 145)。不过客观来讲,英语-西班牙语这类语言对的神经网络机器翻译引擎已经达到了很高的水平,已经不太可能出现误译重要句法结构的情况,因此,上述学者建议的练习还需重新审视,鼓励学生反思机器翻译的优点,而非缺点。

让学习者给机器译文纠错的研究也得出同样的结论。不仅接触“低质模型”(Niño 2009)在语言学习中会引发争议,而且一眼识别出神经网络机器翻译的错误可能会越来越困难(Castilho et al. 2017, Looock & Léchauguette 2021),这使得译后编辑之类的任务(参见O'Brien 2025 [本卷])与以往相比不再适合在某些语言学习环境中使用(参见Zhang & Torres-Hostench 2019)。²

²话虽如此, Looock & Léchauguette (2021)等的最近研究可能更关注语言学习者机器翻译素养的培养,而不是元语言意识本身。教师指导下的机器译文错误分析也是出于这一目的。

Thue Vold (2018)开展了另一项关于元语言意识的研究,对象是挪威一所高中将法语作为第三语言的学生。他们要阅读同一文本的两种机器译文(分别来自谷歌翻译和微软必应翻译),然后说出哪种译文更好,并解释原因。虽然作者(Thue Vold 2018: 73)表示这些学生的法语水平不高于欧洲共同语言参考框架的 B1 等级³,但他们的具体水平并不确定。Thue Vold 总结道,虽然使用机器译文来培养学习者的元语言意识“大有可为”,但“培训、技术支持和教师的引导至关重要”(同上: 89),因为如果听之任之,学习者可能无法作出富有成效的分析,学生群体之间交流甚至可能强化对语言的误解(同上)。

3.4 在语言学习中使用机器翻译的技巧

在语言学习中使用机器翻译具体有哪些好处,对此还没有定论。然而,这种好处很可能取决于一系列因素,包括学生的语言水平、二语写作中的文本体裁(参见Chung & Ahn 2021),以及相关的语言对(参见下文节 4.1)。然而,目前越来越多研究都一致认为:

- 语言学习者使用机器翻译,与其试图禁止其使用,不如根据机器翻译的有用程度进行区别化教学,这可能取决于使用的程度和场景。
- 受过适当训练的语言学习者能更好地使用机器翻译。
- 如果语言学习者已经熟练掌握外语,他们通常会从机器翻译中受益更多(O'Neill 2012, Resende & Way 2021)。

如果学习语言时的确使用了机器翻译,有几点并不显而易见的事项需要老师和学生注意(还可参见 Bowker 2020)。

首先,学生能使用的机器翻译工具不止一种。许多人似乎只知道谷歌翻译(可参见Dorst et al. 2022),但学生也可以通过比较不同的免费在线机器翻译工具,如必应翻译、DeepL 或百度翻译,学生可以从中获益不浅。

其次,由于机翻引擎在使用过程中会反复训练或改进,机翻译文也会随之改变。这意味着不应仅凭一次使用就否定某个机翻系统,也意味着在文章中使用免费在线翻译工具的研究人员应该说明生成机器译文的确切时间,但少有学者做到这一点。

最后,如果你用免费在线机器翻译工具来创建语言教学练习、阐述关于机器翻译技术的观点,或者辅助写作,就要给这个系统好好发挥的机会。输入文本的微小变化可能会对输出产生巨大影响。一项关于机器翻译在二语学习中的运用的研究中,DucarSchocket2018 举了谷歌翻译的一个误译例子,如图图 1所示,该译文基于“milk 这个词的常用字面意思,而非隐喻意义”。

这里值得注意的是,输入句子首字母小写,且句尾无句号。但是,一如图 2所示,若补充这些标准书面语言的特点,输出译文也有所改变——变得更好。

³<https://www.coe.int/en/web/common-european-framework-reference-languages>

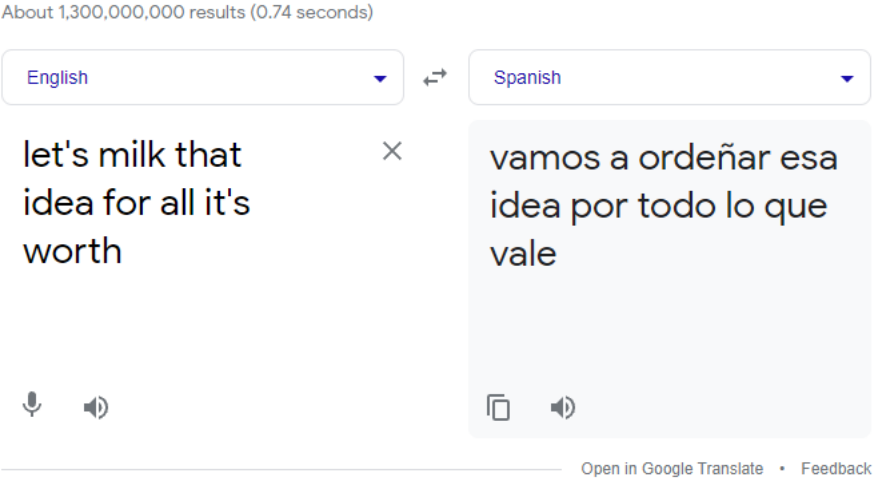


图 1: : 输入谷歌翻译的句子首字母小写且句尾无标点符号, 译文生成于 2021 年 10 月 19 日

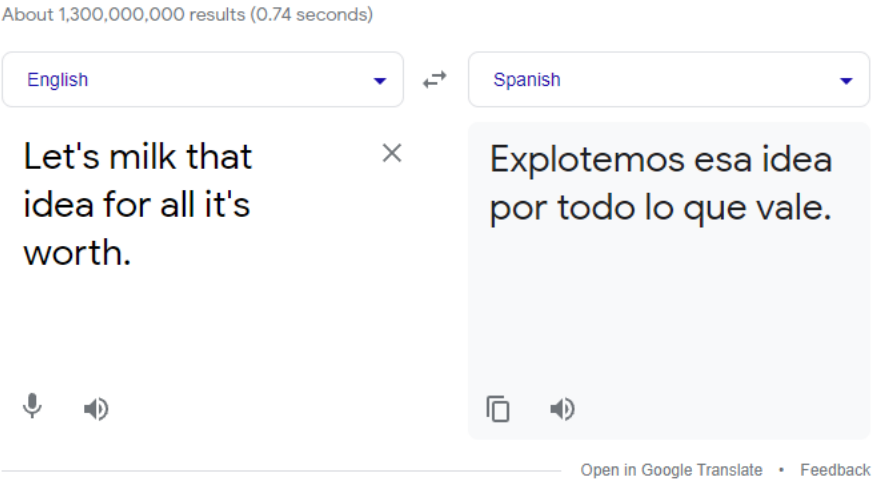


图 2: 输入谷歌翻译的句子首字母大写且句尾有标点符号, 译文生成于 2021 年 10 月 19 日

学生将文本复制粘贴到免费在线机器翻译工具时，也观察到了类似的问题。他们没有意识到复制的文本每行行末都有一个换行符，因此工具识别的是独立的行，每行都单独进行翻译。最先进的机器翻译引擎以句子为单位来训练翻译，只在能够识别和翻译完整的句子时的表现最好。因此，确保不要将带有换行符的文本“输入”到机器中，这点很重要。

- 正如不同的机器翻译引擎或系统的输出可以有效对比，机器翻译的有用性也可以与语料库或在线词典等工具进行对比。详细阐述请看下一节。

4 何时使用机器翻译：语言和情景要素

4.1 我只要一点帮助

如果你正在学外语，或者遇到自己不懂的语言，机器翻译是你最好的帮手吗？机器翻译不仅看似方便快捷，甚至可以帮你蒙骗老师或交际对象，让对方以为你的外语水平精湛，以至于现在的语言类学生经常分享对使用免费机器翻译引擎的依赖，例如承认这个工具能让他们做“一小时的双语者”。⁴但是，你能使用机器翻译来提高外语的听、说、读和写吗？上述的研究表明，借助机器翻译提高外语技能的条件有二：第一，良好的外语水平；第二，扎实的机器翻译知识和“机器翻译素养” (Bowker & Ciro 2019)。前者只能靠反复练习，但后者的培养可以参考本章提出的准则和建议。

4.2 语言对与体裁

首先，要知道机器翻译在不同语言对上的表现有好有坏。事实上，因为神经网络机器翻译系统是基于语料库的（如第 2 章和第 7 章所示），若训练数据不足，系统的输出质量会更差。本节内容给出的例子为法英语言对（包括英法双向的翻译），目前这个语言对的机器翻译质量不错，但我们欢迎读者举出自己语言对的翻译例子，与本节示例进行对比。

体裁也会造成差异。例如，你可能会发现，免费在线机器翻译系统更擅长翻译散文，而非诗歌或你喜欢的歌词。这可能是因为用于训练机器翻译系统的数据与前者更相似，并且训练数据中罕有歌曲和诗歌译文。诗歌和歌曲的翻译要求严苛，译文必须也能够朗诵或吟唱，需要特定的押韵或韵律。虽然可以训练机器撰写甚至翻译诗歌 (Van de Cruys 2018, 2019, 2020)，但通用型免费在线机器翻译系统可能无法胜任这项任务。不过，试试也无妨，你可以

⁴这只是其中一名学生的机器翻译使用经历：<https://mtt.hypotheses.org/our-students-mt-stories>

选一首母语或二语的流行歌曲、诗歌或童谣，找到优质的人工译文⁵尝试用目标语来翻出有韵律和节奏的译文。然后，借助免费在线机器翻译系统来翻译原文，将其与人工翻译进行比较。对比结果可能有助于你思考机器翻译的优点，以及人工翻译的出彩之处。

4.3 请教语言学家

其次，要了解你自己、老师和/或对话者的期望。不妨问问他们是否认为机器翻译可取，以及这种工具是否干扰学习语言。在某些情况下，老师使用的材料或许不能输入免费在线机器翻译引擎，因为其中包含个人或机密数据（参阅Moorkens 2025 [本卷]）。语言老师可能还会说，机器翻译不利于你学习语法规则，因为它让你失去了主动学习句子结构和措辞的积极性。如果需要熟练掌握语法知识才能修改机器翻译的输出，那么使用机器翻译不一定能提高你的语法水平。恰恰相反，学生暴露在错误中却不自知，这会带来不利影响。然而，至少在某些特定情况下，通过反复使用流畅的机器翻译输出来学习外语是可以实现的：在最近的一项研究中，Resende & Way (2021)证明了，部分学生可以通过神经网络机器翻译的输出间接学习句法。但学习者是否也会受到机器翻译错误的影响还有待观察。如果没有良好的机器翻译素养和足够好的目的语知识，他们很可能根本识别不出错误。（例子可参见Loock & Léchaugnette 2021。）

因此，我们建议学生时常请教语言学家——先请教老师，因为他们不仅了解你的二语水平，还知道使用免费在线机器翻译工具来翻译某个语言对的特定内容有何优缺点。

4.4 情景参数

表表 1简要列出了建议考虑的情景决定参数，以帮助语言学习者判断何时可以或不应使用机器翻译。我们的建议基于免费在线神经网络机器翻译工具的使用。（注意：若表表 1中参数 1 的答案为否，则不要使用机器翻译。）

5 机器翻译与同类数字资源

免费在线机器翻译工具使用便捷，还可以全文输入，你可能对结果满意，不想尝试其他方法了。但通常来说，机器翻译是不够的，甚至某些情况不适合使用这种工具。机器翻译与其他数字资源相比如何？接下来，我们会依次讨论下列的四个问题，解释机器翻译指的是什么。讨论机器翻译和其他在线工具的使用，不妨先从回答以下问题开始：

⁵你也许可以在网上找到已出版的诗歌、歌曲和童谣的人工翻译。

表 1: 何时使用或不使用机器翻译

情景	决定参数 (根据重要程度排序)
二语文本理解	1、文本无个人或机密数据 2、神经网络机器翻译输出的质量 (无需完美, 够好就行)
二语写作	1、教师同意 2、文本无个人或机密数据 3、二语水平够好 (CEFR 的 B1 或 B2 水平) 4、神经网络机器翻译输出的质量
翻译作业	1、教师同意 (不太可能, 因为使用机器翻译的话, 翻译作业就变成了译后编辑) 2、文本无个人或机密数据 3、神经网络机器翻译输出的质量
二语口头汇报筹备	1、教师同意 2、内容不含个人或机密数据 3、二语水平够好 (CEFR 的 B1 或 B2 水平) 4、神经网络机器翻译输出的质量以及能够实现文本转语音输出

- 你用不用在线词典, 用的话, 是哪些词典?
- 你如何定义语料库?
- 你用过在线语料库吗?
- 你知道什么是索引 (concordance) 吗?

下文将机器翻译与其他工具进行对比, 目的在于帮助二语学习者和一般的外语使用者了解机器翻译除即时翻译以外的特性。我们想通过对比说明的关键点之一是, 虽然机器翻译使用便捷, 可进行即时翻译, 但这并不意味着神经网络机器翻译工具在任何时候都是首选。

5.1 机器翻译与在线词典

一般来说, 字典是“按字母顺序列出词语及其含义, 或用另一种语言给出对应词语的书” (Cambridge advanced learner's dictionary and thesaurus, 2020)。

这个定义可延伸至电子词典，如在线词典或应用程序，这些产品通常为语言学习者所熟知。在线词典包括传统词典，如牛津英语词典（Oxford English Dictionary）⁶，以及至少部分信息由群体创作的新形式，如维基词典（Wiktionary）⁷或市井词典（Urban Dictionary）。⁸

不管是否经常使用这些在线词典，在如此丰富的资源面前，每个用户都会遇到的问题是：如何判断一本词典是否可靠？对此，辞书学（即词典的编纂）提供了最佳答案，并不断发展以考虑到新词典形式和格式的变化。现在许多词典都是基于大量可用作参考的文本集合（也称为语料库），但词典编纂者永远不会满足于仅仅引用语料库中的内容。语料库反而经常被用来检查语言的运用和寻找相关的例子。在这之前，词典编纂者的工作主要包括为给定的词典找到相关的词条，并给出恰当的释义。要更好地理解这一点，看看为学习者词典就知道了，里面包含了精挑细选的词条、定义和示例。现在，大多数词典都有免费在线版本，你应该很容易就找到一两本学习词典，查看最近学过或特别喜欢的单词。

还有一点很重要：Linguee 网站声称内含一本词典，但具体是哪类词典呢？我们可以区分为“基于语料库”和“由语料库驱动”这两类词典。前者仍然依赖于词典编纂者的直觉，并根据词典编纂的方法创建，而后者则自动提取自语料库。Linguee 内含的是由语料库驱动的双语词典，并不提供单词的定义或精心挑选的例子。

在查看在线语料库之前，我们先总结一下神经网络机器翻译译文和词典词条之间的主要区别：

- 词典词条基于单个单词，而神经机器译文的文本长度不限。神经网络机器翻译引擎远不如字典有用，因为它针对单个单词的翻译通常不可靠。脚注之所以特别指出这一点是因为，实际上，语言学习者和大学生经常使用免费在线机器翻译工具来查找单个单词的翻译（例子参见 Jolley & Maimone 2022, Dorst et al. 2022）。
- 词典提供定义，这可能是确保你理解词义的唯一可靠方法。
- 词典词条基于人类的直觉，并且（大多时候）由词典编纂者设计和/或检查。
- 神经机器译文基于语料库，但并非照搬训练语料库的内容（如 Pérez-Ortiz et al. 2025 [本卷]所解释）。下节内容会更详细地阐述这种差异。

⁶<https://www.oed.com/>

⁷<https://en.wiktionary.org/>, 最后访问于 2022 年 6 月 20 日

⁸<https://www.urbandictionary.com/>, 最后访问于 2022 年 6 月 20 日。

5.2 机器翻译与在线语料库

在开始解释机器翻译和在线语料库的区别之前，至少要弄清楚两个定义。

首先，平行语料库是原文和译文的对齐文本集合。这种平行或对齐意味着每个片段（通常是一个句子）与其译文相对照。

索引工具用于在语料库中查找数据并显示结果。表表 2 的示例选自英国议会议事录语料库 (Hansard)，即加拿大议会参议院和众议院辩论的平行语料库（英语和法语）。

首先要注意，索引是一种搜索功能。例如，我们查找一个有英语译文的法语短语，先出现的是法语（位于左侧），尽管语料库的这部分内容原文是英语（在线语料库不一定提供这一信息）。同样值得一提的是，双语索引通常会在原文语段中高亮搜索词（见表表 2 加粗部分）。很多双语索引还会高亮与原文语段相对应的目的语语段，但这种基于概率的识别通常并不准确。事实上，我们已经“清理”了表表 2 的目的语一栏，使得索引结果“对等”（参见Kenny 2025 [本卷]: §1）部分更明了。简而言之，句子层面的对齐或能精准，但双语或平行语料索引中更精细的语言单元对照有时无法自动识别。

还要注意，在线索引通常会显示每个索引所在的完整段落或文本，如表表 2 所示，或者链接到语料库的相关部分。

相比之下，机器翻译只会提供输入内容的建议译文，没有其他语境元素。

表 2: 语段 *nous allons faire le nécessaire* 在英国议会议事录语料库中的索引结果

法语译文	英语原文
Nous vous disons que <i>nous allons faire le nécessaire</i> , mais aidez-nous à nous assurer que tout le monde respecte les règles.	But we're saying, hey, <i>we'll do it, we'll set it up</i> , but help us to make sure everybody abides by the rules.
Monsieur le Président, nous avons promis que <i>nous allons faire le nécessaire</i> pour ratifier l'accord.	Mr. Speaker, we promised that <i>we were going to do what was required to</i> ratify the agreement.

表表 3 比较两次检索，前者比后者更短，且表意更模糊。例子表明神经网络机器翻译引擎能根据句子语境作出调整，即有明确的目的状语时（*pour ratifier l'accord*），英语机器译文使用不同的句子结构，其中 *faire le nécessaire* 翻译成 *take the necessary steps to*。

总的来说，神经网络机器翻译引擎更适用于完整的句子（参见Kenny 2025 [本卷]: §7）或文本；而查找单词、搭配或短语时，词典和/或语料库能提供更

表 3: 语段 *nous allons faire le nécessaire* 的神经机器译文示例 (机器译文于 2021 年 11 月 1 日出自 <https://www.bing.com/translator>)

法语检索	英语神经机器译文
Nous allons faire le nécessaire.	We will do what is necessary.
Nous allons faire le nécessaire pour ratifier l' accord.	We will take the necessary steps to ratify the agreement.

有效可靠的受控结果。平行语料库能提供多种语境易于获取的译文;而机器译文则基于对训练数据的复杂计算,但用户有时无法访问这些数据(通常对免费在线机器翻译工具的用户不可见)。这使得用户很难确定建议译文是否真的可靠。

6 错误分析

利用机器翻译学习第二语言或外语所需的关键技能之一是能敏锐地识别错误。为培养这种技能而设计的活动有很多,但因为建议还很少,因此我们在下文为读者提供一个带有评论的例子。⁹

学生给一段文字及其译文列出自己能够识别并改正的错误类型,然后将列表交给老师。老师对此给出反馈,并指出学生没有识别出的所有错误,给出解释的同时,提出改进建议和有用的例子。如果目的语文本(机器译文)为第二语言,这项练习会有难度,因此我们建议老师从母语的机器译文开始。虽然如本章开头所述,此类任务很长一段时间都不用于语言课堂,但最近有学者提出将翻译融入情境任务,以期将学生成为“在日益多语言的世界中自我反思、具有跨文化能力和有责任的意义构建者”(Laviosa 2014: 105)。

表表 4所示为英翻法的神经机器译文示例。¹⁰机器译文中的错误为加粗部分,注释见下方。¹¹

例句的翻译错误包括过度直译,如 *une journée typique de ma vie*。法语译文最好使用复数形式,且 *typique* 应译成 *mes journées de lycéen* (字面意思是

⁹有关如何将机器翻译融入语言教学的更多最新观点,请参阅Vinall & Hellmich (2022)。
¹⁰文本选自法国学习者的英语教科书 (Joyeux 2019:22)。为了将此活动变成情境任务,学习者要给一个几乎不懂英语的法国人(例如特地来课堂观摩的人)提供合适译文。他们尽量不要从机器译文中获取信息,也不要对其进行修改。
¹¹机器译文的错误分析通常基于“错误类型”(error typologies),其中包括机器译文会出现的各类错误。这些错误通常包括准确度错误(如目标语段的意思与源语段的意思不一致)和影响目的语的流畅度或语言规范的错误(如语法一致、词序、搭配等错误)。更多信息请参阅Rossi & Carré (2025 [本卷])对机器翻译评测,以及O'Brien (2025 [本卷])对译后编辑的阐述。

表 4: 教材片段摘录的 NMT 输出示例

英语原文	法语神经机器译文
Hi there!	Bonjour à tous !
My teacher asked me to write and tell you what a typical day in my life looked like so I' ll do my best to give you an idea! I get up around 7:00 am, I have breakfast (two slices of toast and a cup of tea) then I get ready and put on my uniform. I make my lunch, and I double-check that my bag is packed. I leave my house at around 8:00 am. I' ve only got a 15-minute walk to school, so I arrive early. I usually chat with my friends, or listen to music with my headphones. Classes begin at 8:45 am.	Mon professeur m'a demandé de vous écrire et de vous raconter à quoi ressemblait <i>une journée typique de ma vie</i> , alors je vais faire de mon mieux <i>pour vous donner une idée</i> ! Je me lève vers 7 heures du matin, je prends mon petit déjeuner (deux tranches de pain grillé et une tasse de thé) puis je me prépare et je mets mon uniforme. Je prépare mon déjeuner, et je vérifie que mon sac <i>est bien emballé</i> . Je quitte <i>ma maison</i> vers 8 heures. <i>Je n'ai que 15 minutes de marche pour me rendre à l'école, donc j'arrive tôt</i> . J'ai l'habitude de discuter avec mes amis ou d'écouter de la musique avec mes écouteurs. Les cours commencent à 8h45.

“我当学生的日常”)。你可能注意到, a typical day 这个短语不适合直译, 因为 a typical day 在源语中可算作固定表达。对于习语 to give you an idea, 也要在目的语中找个习语, 如 *pour vous en donner un aperçu*。

还有些语法和语言使用的错误。法语需要 en 这类接语代词 (clitic pronouns), 如可以将 *donc j' arrive tôt* 变成 *donc j' y arrive en avance* (意为“所以我提前到”)来改进译文。另外, 在英语中, 即便文中的所属关系不明显, 也会使用所有格, 但这种情况下, 法语更喜欢使用限定词 (如 *je quitte la maison* 更倾向于译成 I leave the house, 而不是 my house)。语言使用也与词汇有关, 虽然英语通常指移动方式 (a 15-minute walk to school, 意为“15 分钟步行到学校”), 但法语的用法更中性 (15 minutes de trajet pour l' école, 意为“到学校要 15 分钟的路程”), 只有在必要时才添加细节 (如 à pied, 意为“步行”)。

类似的例子不胜枚举, 但希望以上例子足以说明, 尽管法语机器译文看起来够好了, 没有严重的语法或词汇错误, 但仍有很大的改进空间。找出学习者能够和不能纠正的错误, 肯定会对教师有所启发 (关于这点可参阅 Loock & Léchauguet (2021))。

我们鼓励读者获取自己语言对的机器译文，先译成母语。在情景任务中使用第二语言的神经机器译文的的活动，可以用于后期的语言学习，尤其是识别和纠正译文错误的练习。

7 结语

本章介绍了目前关于机器翻译在语言学习中的应用的主要研究发现，重点关注自神经网络机器翻译出现以来，有关机器翻译技术进步的研究成果。在建立Rossi & Carré (2025 [本卷])提出的质量测评的语用方法之前，我们还提供了一些在语言学习中使用机器翻译的基本技巧，重点关注它对第二语言 and 外语学习者的作用。与在翻译任务中可能无法选用哪种工具的专业译者不同，语言学习者有多种工具可供选用，但他们首先得决定是否和何时使用机器翻译。为此，我们提供了情景参数列表，帮助这些学习者做出选择。我们还将机器翻译与词典和语料库等在线辅助工具进行了对比，重点介绍两者的优点。最后，我们提出了如何利用神经网络机器翻译，并将其纳入第二语言或外语课堂的活动建议。

References

- Bowker, Lynne. 2020. Machine translation literacy instruction for international business students and business English instructors. *Journal of Business & Finance Librarianship* 25(1-2). 25–43. DOI: 10.1080/0896.1794739.
- Bowker, Lynne & Jairo Buitrago Ciro. 2019. *Machine translation and global research*. Bingley: Emerald Publishing.
- Cambridge University Press. 2020. *Cambridge advanced learner's dictionary and thesaurus*. <https://dictionary.cambridge.org/dictionary/english/>.
- Castilho, Sheila, Joss Moorkens, Federico Gaspari, Iacer Calixto, John Tinsley & Andy Way. 2017. Is neural machine translation the new state of the art? *The Prague Bulletin of Mathematical Linguistics* 108. 109–120. DOI: 10.1515/pralin-2017-0013.
- Chung, Eun Seon & Soojin Ahn. 2021. The effect of using machine translation on linguistic features in L2 writing across proficiency levels and text genres. *Computer Assisted Language Learning*. DOI: 10.1080/09588221.2020.1871029.
- Clifford, Lisa Merschel, Joan & Joan Munné. 2013. Surveying the landscape: What is the role of machine translation in language learning? @tic. *Revista D'innovació Educativa* 10. 108–121.
- Cook, Guy. 2010. *Translation in language teaching*. Oxford: Oxford University Press.

- Correa, Maite. 2011. Academic dishonesty in the second language classroom: Instructors' perspectives. *Modern Journal of Language Teaching Methods* 1(1). 65–79.
- Dorst, Lettie, Susana Valdez & Heather Bouman. 2022. Machine translation in the multilingual classroom. *How, when and why do humanities students at a Dutch university use machine translation? Translation and Translanguaging in Multilingual Contexts* 8(1). 49–66. DOI: 10.1075/ttmc.00080.dor.
- Ducar, Cynthia & Deborah Houk Schocket. 2018. Machine translation and the L2 classroom: Pedagogical solutions for making peace with google translate. *Foreign Language Annals* 51. 779–795.
- Enkin, Elizabeth & Errapel Mejías-Bikandi. 2016. Using online translators in the second language classroom: Ideas for advanced-level Spanish. *LACLIL* 9(1). 138–158. DOI: 10.5294/lacil.2016.9.1.6.
- Fredholm, Kent. 2015. Online translation use in Spanish as a foreign language essay writing: Effects on fluency, complexity and accuracy. *Revista Nebrija de Lingüística Aplicada a la Enseñanza de las Lenguas* 18. 7–24.
- Fredholm, Kent. 2019. *Effects of google translate on lexical diversity: Vocabulary development among learners of Spanish as a foreign language*, vol. 13. 26378/rn-lael1326300. 98–117. DOI: doi:10.26378/rnlael1326300.
- Jolley, Jason R. & Luciane Maimone. 2022. Thirty years of machine translation in language teaching and learning: A review of the literature. *L2 Journal* 14(1). 26–44. <http://repositories.cdlib.org/uccllt/l2/vol14/iss1/art2>.
- Joyeux, Maël. 2019. *Fireworks, anglais*. <https://www.lelivrescolaire.fr/>.
- Kelly, Niamh & Jennifer Bruen. 2017. Using a shared L1 to reduce cognitive overload and anxiety levels in the L2 classroom. *The Language Learning Journal* 45(3). 368–81.
- 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.
- Laviosa, Sara. 2014. *Translation and language education: Pedagogic approaches explored*. London/New York: Routledge.
- Lee, Sangmin-Michelle. 2021. The effectiveness of machine translation in foreign language education: A systematic review and meta-analysis. *Computer Assisted Language Learning* 33(3). 157–175. <https://www.tandfonline.com/doi/full/10.1080/09588221.2021.1901745>.
- Loock, Rudy & Sophie Léchauguette. 2021. Machine translation literacy and undergraduate students in applied languages: Report on an exploratory study. *Revista Tradumàtica: tecnologies de la traducció* 19. 204–225. DOI: 10.5565/rev/tradumatica.281.

- 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.
- Mundt, Klaus & Michael Groves. 2016. A double-edged sword: The merits and the policy implications of Google Translate in higher education. *European Journal of Higher Education* 6(4). 387–401. DOI: 10.8235.2016.1172248.
- Niño, Ana. 2009. Machine translation in foreign language learning: Language learners' and tutors' perceptions of its advantages and disadvantages. *ReCALL* 21(2). 241–258. DOI: 10.1017/S0958344009000172.
- Noriega-Sánchez, María, Ángeles Carreres & Lucía Pintado Gutiérrez. 2021. Introduction: Translation and plurilingual approaches to language teaching and learning. *Translation and Translanguaging in Multilingual Contexts* 7(1). 1–16.
- 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.
- O'Neill, Errol. M. 2012. *The effect of online translators on L2 writing in French*. University of Illinois at Urbana-Champaign. (Doctoral dissertation). <http://hdl.handle.net/2142/34317>.
- O'Neill, Errol. M. 2019. Training students to use online translators and dictionaries: The impact on second language writing scores. *International Journal of Research Studies in Language Learning* 8(2). 47–65.
- 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.
- Pintado Gutiérrez, Lucía. 2018. Translation in language teaching, pedagogical translation, and code-switching: Restructuring the boundaries. *The Language Learning Journal* 49(2). 219–239.
- Pym, Anthony, Maria del Mar Gutiérrez-Colón Plana & Kirsten Malmkjaer. 2013. *Translation and language learning: The role of translation in the teaching of languages in the European Union*, vol. 3. Luxembourg: Publications Office of the European Union.
- Resende, Natália & Andy Way. 2021. Can Google Translate rewire your L2 English processing? *Digital* 1(1). 66–85.
- 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.

- Thomas, Jacqueline. 1988. The role played by metalinguistic awareness in second and third language learning. *Journal of Multilingual and Multicultural Development* 9(3). 235–246. DOI: 10.1080/01434632.1988.9994334.
- Thue Vold, Eva. 2018. Using machine-translated texts to generate L3 learners' metalinguistic talk. In Åsta Haukås, Camilla Bjørke & Magne Dypedahl (eds.), *Metacognition in language learning and teaching*, 67–97. Routledge. DOI: 10.4324/9781351049146.
- Torres-Hostench, Olga. 2025. 欧洲、多语制与机器翻译. In Dorothy Kenny (ed.), *机器翻译知识普及: 为人工智能时代的用户赋能*, 1–17. Berlin: Language Science Press. DOI: 10.5281/zenodo.14922285.
- Van de Cruys, Tim. 2018. *Il pleure dans mon processeur multi-cœur*. Poésie automatique. <http://www.timvandecruys.be/media/charles2018.pdf>.
- Van de Cruys, Tim. 2019. *La génération automatique de poésie en français*. Proceedings of TALN 2019. Toulouse. 113 – 126. <https://aclanthology.org/2019.jeptalnrecital-long.8.pdf>.
- Van de Cruys, Tim. 2020. *Articuler ou trahir*. Translating Europe Workshop. Université Toulouse Jean Jaurès. 7 February 2020. http://timvandecruys.be/media/presentations/pres_tradlit_070220.pdf.
- Vinall, Kimberly & Emily Hellmich (eds.). 2022. *Machine translation & language education: Implications for theory, research, & practice*, vol. 14. 1–3. https://escholarship.org/uc/uccllt_l2/14/1.
- Zhang, Hong & Olga Torres-Hostench. 2019. Cómo enseñar posesión de traducción automática a una segunda lengua: Una propuesta didáctica para el aprendizaje de lenguas. *Revista Tradumàtica: traducció i tecnologies de la informació i la comunicació* 17. 153–161. DOI: 10.5565/rev/tradumatica.237.

人名索引

- Abdallah, Kristiina, 103
Ahn, Soojin, 151
Aixelá, Franco, 69
Aziz, Wilker, 44
- Bahdanau, Dzmitry, 124
Baker, Mona, 19, 105
Bao, Guangsheng, 33
Bentivogli, Luisa, 30, 74
Bowker, Lynne, v, 26, 77, 93, 147, 151, 153
Brown, Tom B., 116
Bruen, Jennifer, 148
Byrne, Jody, 73
- Canfora, Carmen, 69, 102
Carré, Alice, 10, 24, 26, 30, 65, 67, 100, 126, 135, 143, 144, 158, 160
Castilho, Sheila, 40, 43, 65, 150
Caswell, Isaac, 26
Cenoz, Jasone, 5
Chesterman, Andrew, 102
Chung, Eun Seon, 151
Ciro, Jairo Buitrago, v, 26, 77, 93, 147, 153
Clifford Joan, Lisa Merschel, 149
Cook, Guy, 148
Correa, Maite, 149
Council of the EU, 5
Cronin, Michael, 104
- de Almeida, Gisele, 92
- Debussche, Julien, 97
Denkowski, Michael, 144
Docherty, Peter, 104
Doherty, Stephen, 40, 59
Dorst, Lettie, 151, 156
Drugan, Joanna, 40, 65
Ducar, Cynthia, 148, 149
- Enkin, Elizabeth, 150
Eurydice, 7
Extra, Guus, 3
- Floridi, Luciano, 100
Forcada, M. L., 27
Forcada, Mikel, 31, 124
Fredholm, Kent, 149, 150
Freitag, Markus, 100
Fry, Deborah, 70
Fujita, Atsushi, 67
- Gao, Catherine Yuan, 10, 13
Gerlach, Johanna, 67
Ghiara, Silvia, 69
Goodfellow, Ian, 27, 32, 113
Gordin, Michael. D., 10
Gouadec, Daniel, 40
Grbić, Nadja, 40
Groves, Michael, 149
Guerberof Arenas, Ana, 86, 93
- Hansen-Schirra, Silvia, 66, 67, 92
Hellmich, Emily, 158
Hiraoka, Yusuke, 67

- Hornik, Kurt, 116
 House, Juliane, 1, 11, 40
 Hutchins, John, 25
- Jakobson, Roman, 20
 Johnson, Joseph, 25
 Johnson, Marvin, 106
 Jolley, Jason R., 148, 156
 Joscelyne, A., 27
 Joyeux, Maël, 158
- Kay, Martin, 103
 Kelly, Niamh, 148
 Kenney, Dorothy, 43, 46–48, 67, 68, 70,
 73, 83, 88, 131, 135, 137, 140,
 141, 157
 Khayrallah, Huda, 137
 Knight, Jane, 12
 Ko, Wei-Jen, 127
 Koehn, Philipp, 25, 29, 31, 45, 51, 56,
 57, 137, 138
 Koponen, Maarit, 26, 90
 Krings, Hans P., 90
 Kudo, Taku, 125
- Lambert, Joseph, 102
 Lavie, Alon, 144
 Laviosa, Sara, 158
 Léchaugnette, Sophie, 150, 154, 159
 Lee, Sangmin-Michelle, 148, 150
 Lewis, David, 98
 Loock, Rudy, 150, 154, 159
- Maimone, Luciane, 148, 156
 Mariana, Valerie, 44
 Marking, Marion, 30
 Marzouk, Shaimaa, 66, 67
 Maylath, Bruce, 70
 Mejías-Bikandi, Errapel, 150
 Mikolov, Tomas, 118
- Mittelmeier, Jenna, 11
 Mittelstadt, Brent, 100
 Miyata, Rei, 67
 Mohiuddin, Tasnim, 137
 Moorkens, Joss, 21, 26, 43, 44, 69, 98,
 99, 103, 149, 154
 Mundt, Klaus, 149
 Munné, Joan, 149
 Murai, Shusuke, 10
- Navarro, Fernando A., 69
 Ney, Hermann, 144
 Niño, Ana, 150
 Nitzke, Jean, 92
 Noriega-Sánchez, María, 148
 Nurminen, Mary, 26
- O'Brien, Sharon, 13, 40, 43, 59, 65, 66,
 76, 88, 92, 93, 131, 134, 137,
 150, 158
 O'Neill, Errol. M., 149–151
 Olohan, Maeve, 69
 Ottmann, Angelika, 69, 102
- Papineni, Kishore, 126, 144
 Pérez-Ortiz, Juan Antonio, 27, 31, 32,
 72, 96, 136, 137, 156
 Pintado Gutiérrez, Lucía, 148
 Popović, Maja, 58, 144
 Post, Matt, 57
 Pym, Anthony, 22, 101, 148
- Qin, Ying, 45
- Ramírez-Sánchez, Gema, 13, 27, 34,
 68, 73
 Resende, Natália, 151, 154
 Richardson, John, 125
 Rocchi, Martha, 103

人名索引

Rossi, Caroline, 24, 30, 65, 67, 100,
126, 135, 143, 144, 158, 160

Sakamoto, Akiko, 102

Saldanha, Gabriela, 19

Šarcevic, Susan, 69

Saunders, Danielle, 135, 137

Schocket, Deborah Houk, 148, 149

Seoane Vicente, Ángel Luis, 69

Seretan, Violeta, 67

Slobin, Dan, 21

Snover, Matthew, 53, 91, 144

Specia, Lucia, 45

Steinberger, Ralf, 97

Strubell, Emma, 104

Taddeo, Mariarosaria, 100

Tammet, Daniel, 20

Teixeira, Carlos S. C., 88

Thomas, Jacqueline, 150

Thue Vold, Eva, 150, 151

Tiedemann, Jörg, 99

Tomalin, Marcus, 106

Topping, Suzanne, 99

Toral, Antonio, 30, 59, 105

Torregrosa Rivero, Daniel, 89

Torres-Hostench, Olga, 147, 150

Troussel, Jean-Christophe, 97

Turovsky, Barak, 26

Van de Cruys, Tim, 153

Van Wynsberghe, Aimee, 104

Vanmassenhove, Eva, 105, 106

Vashishth, Shikhar, 33

Vaswani, Ashish, 122

Vázquez, Silvia Rodríguez, 93

Vieira, Lucas Nunes, 105

Vinall, Kimberly, 158

Wachter, Sandra, 100

Way, Andy, 151, 154

Weizenbaum, Joseph, 98

Whipple, Judith M., 100

Wikizero, 1

Williamson, Graham, 59

Yagmur, Kutlay, 3

Yamada, Masaru, 42, 67

Zakharova, Ulyana S., 11

Zhang, Hong, 150

机器翻译知识普及

语言学习与翻译素为欧盟多语言政策的核心支柱，二者相辅相成。正如意大利作家、符号学家翁贝托·埃科（Umberto Eco）所言：“欧洲的语言就是翻译”。然而，随着机器翻译和生成式人工智能等技术的不断发展和广泛运用，传统的语言学习和翻译模式正面临挑战。如今，语言学习者可以借助免费在线机器翻译工具来帮助理解和生成外语文本。但如果学习者盲目使用这些技术，反而可能弊大于利。尽管机器翻译被视为颠覆翻译行业的革命性技术，但其现有的工作原理犹如“黑匣子”，难以理解。此外，机器翻译还涉及法律和道德等问题，使得职业译者难以融入机器翻译的工作流程。

在这一背景下，本书面向语言学习者、语言教师、翻译学习者、翻译教师以及职业译者等广泛读者群体，旨在推动机器翻译技术在教学与实践中的应用。全书由多位作者共同编撰，首先阐明了学习机器翻译的理论基础，接着介绍了机器学习的基本原理，并深入分析了神经网络机器翻译的相关内容。书中还讨论了机器翻译技术所引发的道德问题，并提出了其在语言学习中的应用建议。此外，作者还探讨了通过译前编辑、译后编辑和定制化机器翻译等方法，帮助用户最大化地发挥机器翻译技术的潜力。