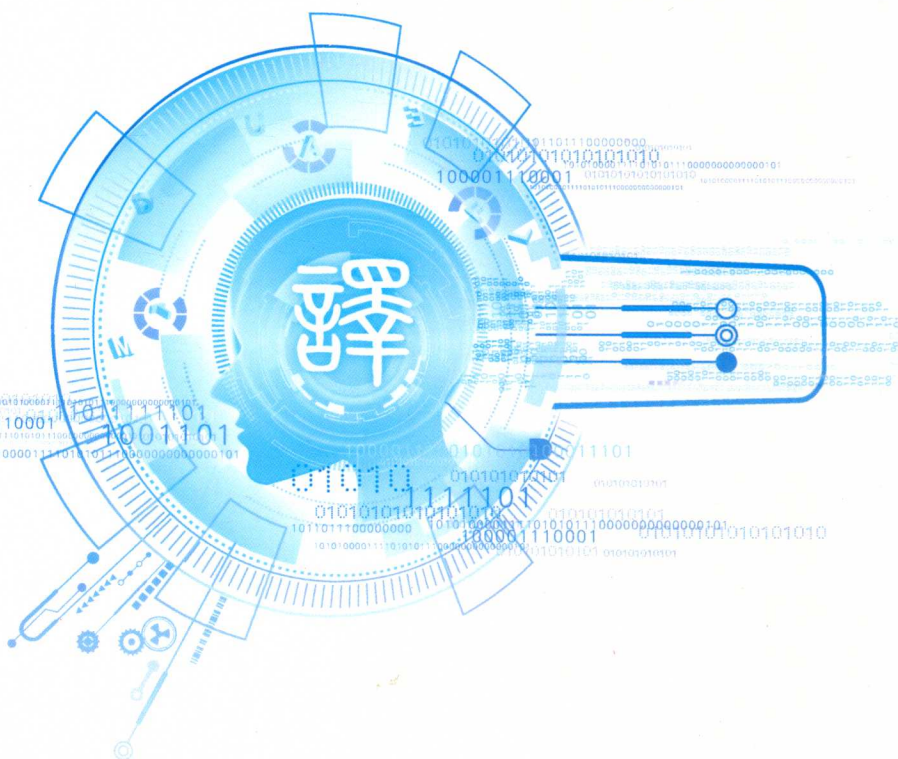


翻译技术100问

主 编 王华树

副主编 陈涅奥 叶梦轩



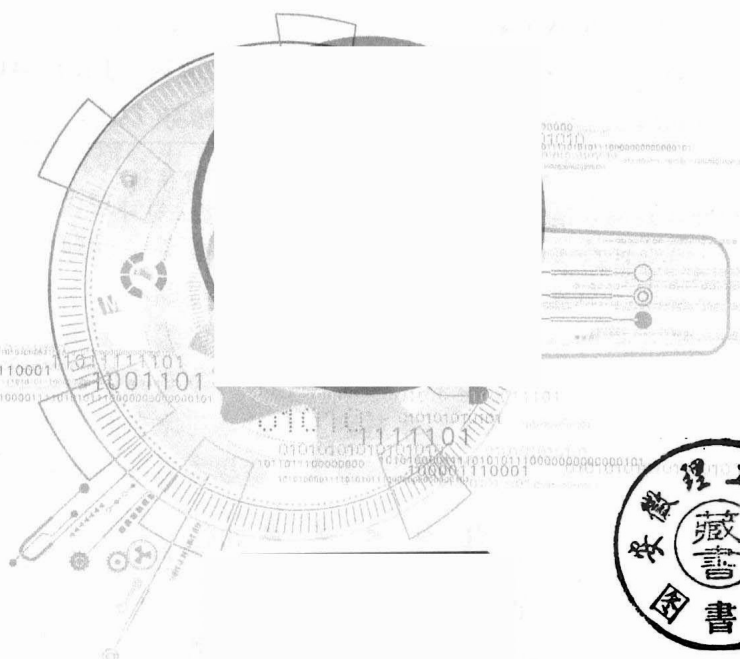
科学出版社

世界翻译教育联盟 (WITTA) 翻译技术教育研究会
广东外语外贸大学高级翻译学院

资助项目

翻译技术100问

主 编 王华树
副主编 陈涅槃 叶梦轩
顾 问 李德凤 李瑞林



科学出版社
北京

内 容 简 介

本书由国内知名翻译技术专家和语言服务企业技术专家编写，具有简明易懂、内容新颖、针对性强的特征。全书分为“基础篇”“进阶篇”“高级篇”三个部分，兼顾技术原理与技术实践，以项目案例为驱动，以解决问题为导向，以问答形式呈现技术解决方案，涵盖计算机辅助翻译技术、翻译与搜索技术、机器翻译与译后编辑、翻译项目管理技术、术语管理技术、本地化工程技术、字幕翻译技术等翻译实践中的常见技术，内容由浅入深，图解丰富直观，给读者带来真实的、全面的翻译技术学习体验。

本书的目标读者既包括外语、翻译相关专业的师生，也包括语言服务从业者和爱好者，旨在强化培养读者的技术思维，为培养懂翻译、懂技术的复合型语言服务人才提供借鉴和参考。

图书在版编目(CIP)数据

翻译技术 100 问/王华树主编. —北京: 科学出版社, 2020.3

ISBN 978-7-03-064440-4

I. ①翻… II. ①王… III. ①翻译学-问题解答 IV. ①H059-44

中国版本图书馆 CIP 数据核字(2020)第 026679 号

责任编辑: 张 宁 / 责任校对: 贾伟娟
责任印制: 李 彤 / 封面设计: 蓝正设计

科 学 出 版 社 出 版

北京东黄城根北街 16 号

邮政编码: 100717

<http://www.sciencep.com>

北京中科印刷有限公司 印刷

科学出版社发行 各地新华书店经销

*

2020 年 3 月第 一 版 开本: 787×1092 1/16

2020 年 3 月第一次印刷 印张: 18 3/4

字数: 431 000

定价: 68.00 元

(如有印装质量问题, 我社负责调换)

王华树，博士、副教授，广东外语外贸大学“云山青年学者”，高级翻译学院硕士研究生导师，翻译技术教育与研究中心主任，外语研究与语言服务协同创新中心研究员，兼任世界翻译教育联盟（WITTA）翻译技术教育研究会会长，中国翻译协会本地化服务委员会副秘书长，中国英汉语比较研究会外语教育技术研究专业委员会副秘书长，UTH、ALSP、Translation Commons等技术顾问以及多所大学客座教授。在《中国翻译》《外国语》《外语电化教学》《上海翻译》等期刊发表论文六十余篇，主持十多项国家级、省部级及其他各类科研项目，出版《翻译技术教程》《计算机辅助翻译概论》《术语管理概论》《翻译与本地化项目管理》等十多部著作。研究兴趣：计算机辅助翻译、翻译与本地化、术语管理、语言服务管理。

陈涅槃，广东外语外贸大学高级翻译学院翻译与本地化管理方向研究生，笔译团队技术部负责人，翻译技术教育与研究中心工程师，世界翻译教育联盟（WITTA）翻译技术教育研究会技术专家。研究兴趣：本地化工程、网站本地化、机器翻译译后编辑。

叶梦轩，广东外语外贸大学高级翻译学院翻译与本地化管理方向研究生，高精尖翻译资源研发与认证中心工程师，世界翻译教育联盟（WITTA）翻译技术教育研究会技术专家。研究兴趣：计算机辅助翻译、语料加工、软件本地化。

前 言

随着人工智能技术的快速发展，翻译技术突飞猛进，翻译活动呈现出信息化、智能化、协作化、专业化等职业化特征，翻译实践正在发生巨变，翻译技术能力成为现代译者的基本能力。根据相关调查，国内外主要的翻译院校都意识到翻译技术能力的重要性。如欧盟的欧洲翻译硕士（European Master's in Translation）项目，将“翻译技术”作为核心能力之一；美国明德大学蒙特雷国际研究学院（Middlebury Institute of International Studies at Monterey）的“翻译与本地化管理”硕士项目开设了20多门翻译技术相关的课程；英国、爱尔兰、德国、加拿大、澳大利亚等许多国家的大学也开设了翻译技术相关的课程。目前，全国已有253所翻译硕士学位（MTI）院校、281所翻译与口译学士学位（BTI）院校以及大量培养不同层次翻译专业人才的高职高专院校。全国翻译专业学位研究生（MTI）教育指导委员会公布的示范性教学大纲中，计算机辅助翻译已经成为核心课程，越来越多的翻译院校将会开设翻译技术课程。

作为翻译技术领域的参考读物之一，《翻译技术100问》具有脉络连贯清晰、文字简明扼要、内容由浅入深、图解丰富直观等特征。本书兼顾技术原理与技术实践，以项目案例为驱动，以解决问题为导向，以问答形式呈现技术解决方案。全书内容涵盖翻译技术概论、翻译与搜索技术、计算机辅助翻译技术、机器翻译与译后编辑、翻译项目管理技术、术语管理技术、本地化工程技术、字幕翻译技术等翻译实践中的常见技术，旨在强化培养读者的技术思维，为培养懂翻译、懂技术的复合型语言服务人才提供借鉴和参考。

全书分为“基础篇”“进阶篇”“高级篇”三个部分。“基础篇”涵盖了语言服务领域的基本概念问题，便于读者快速查阅，了解相关知识；“进阶篇”和“高级篇”提出了翻译和本地化领域常见的技术问题，每个问题均带有案例，让读者体验到翻译技术在行业中的真实应用场景，并学习技术问题的解决方案。

读者既可以将本书作为案头工具书，在学习与应用翻译技术的过程中随时翻看，启迪翻译技术问题的解决思路；也可根据书中提供的案例资料，进行相关技术工具的配置与实操，不仅能达到解决问题的目的，更能达到提高翻译技术素养的效果。本书附录还对常见的计算机辅助翻译工具、术语管理工具、机器翻译系统、正则表达式、主要的翻译技术教材等资料进行了归纳整理，方便读者进一步深入了解及学习相关知

识。此外，本书出版纸质图书的同时，配套相关线上课程，并根据需求进行及时更新，为读者提供多元化的学习方式和与时俱进的学习内容，需要学习线上课程的读者，可以扫描本书封底的二维码，扫码后观看。

在本书编纂过程中，我们得到了中国翻译协会本地化服务委员会的大力支持，也得到了来自学界和业界专家的热心帮助！他们是广东外语外贸大学黄忠廉教授、陕西师范大学MTI中心副主任李海亭老师、业界资深译者刘瑞杰老师、SDL大中华区教育事业部经理贺娅君、SDL产品解决方案顾问刘鹏、CSOFT亚太区资深商务拓展经理于占国、上海创凌翻译服务有限公司总经理杨颖波、文思海辉副总裁陈海燕、华为翻译中心经理江燕飞。衷心感谢他们在百忙之中贡献智慧，无私分享，为本书的顺利完成提供了宝贵的经验。

由于水平和时间的限制，书中难免会有瑕疵和遗漏，恳请业界同仁不吝赐教。

王华树

广东外语外贸大学高级翻译学院

翻译技术教育与研究中心

2019年6月16日

本书中英对照术语

语言服务	Language Service
技术能力	Technological Competence
翻译技术	Translation Technology
口译技术	Interpreting Technology
计算机辅助翻译	Computer-Aided Translation / Computer-Assisted Translation, CAT
翻译记忆	Translation Memory, TM
模糊匹配	Fuzzy Match
完全匹配	Full Match
上下文匹配	Context Match
预翻译	Pre-translation
伪翻译	Pseudo Translation
机器翻译	Machine Translation, MT
基于规则的机器翻译	Rule-based Machine Translation, RMT
基于统计的机器翻译	Statistical-based Machine Translation, SMT
神经网络机器翻译	Neural Machine Translation, NMT
译后编辑	Post-Editing, PE
术语库	Termbase
术语管理	Terminology Management
术语管理系统	Terminology Management System
语料库	Corpus
光学字符识别	Optical Character Recognition, OCR
语音识别	Speech Recognition

众包翻译	Crowdsourced Translation
字幕翻译	Subtitle Translation
桌面排版	Desktop Publishing, DTP
质量保证	Quality Assurance, QA
质量控制	Quality Control, QC
翻译管理系统	Translation Management System, TMS
内容管理系统	Content Management System, CMS
技术写作	Technical Writing
达尔文信息类型化体系结构	Darwin Information Typing Architecture, DITA
辅助写作	Computer-Aided Writing / Computer-Assisted Writing
全球化	Globalization, G11n
国际化	Internationalization, I18n
本地化	Localization, L10n
翻译	Translation, T9n
本地化工程	Localization Engineering
本地化测试	Localization Testing
语言测试	Linguistic Testing
功能测试	Functionality Testing
界面测试	Cosmetic Testing / User Interface (UI) Testing
敏捷本地化	Agile Localization

目 录

前言

本书中英对照术语

基础篇

1. 什么是语言服务行业?	3
2. 语言服务行业的标准和规范有哪些?	3
3. 语言服务行业有哪些发展趋势?	3
4. 新时代译者需要具备什么样的能力?	4
5. 什么是翻译技术?	4
6. 翻译技术有哪些作用?	4
7. 翻译技术有哪些主要标准?	5
8. 什么是口译技术?	5
9. 什么是计算机辅助翻译?	5
10. 常见的计算机辅助翻译工具有哪些?	6
11. 什么是翻译记忆?	6
12. 什么是预翻译?	6
13. 什么是伪翻译?	7
14. 什么是译前准备?	7
15. 什么是机器翻译?	7
16. 机器翻译是否会取代人类译者?	8
17. 什么是机器翻译译后编辑?	8
18. 译后编辑对于译者有哪些技能要求?	9

19. 什么是术语管理?	9
20. 什么是术语管理技术?	10
21. 什么是术语管理系统?	10
22. 什么是语料库?	11
23. 什么是语料技术?	11
24. 常见的语料库有哪些?	12
25. 什么是光学字符识别技术?	12
26. 什么是语音识别技术?	12
27. 常见的语音识别产品有哪些?	13
28. 什么是众包翻译?	13
29. 常见的众包翻译平台有哪些?	13
30. 什么是字幕翻译?	14
31. 字幕翻译的相关工具有哪些?	14
32. 什么是本地化桌面排版?	14
33. 本地化桌面排版涉及哪些工具?	15
34. 什么是翻译管理系统?	15
35. 常见的翻译管理系统有哪些?	16
36. 什么是质量保证?	16
37. 常见的翻译质量保证工具有哪些?	16
38. 什么是技术写作?	17
39. 什么是达尔文信息类型化体系结构?	17
40. 常见的技术写作工具有哪些?	17
41. 什么是辅助写作?	18
42. 什么是 GILT?	18
43. 本地化行业涉及哪些机构和角色?	18
44. 本地化需要解决哪些问题?	19
45. 常见的本地化项目类型有哪些?	19
46. 什么是本地化工程?	20
47. 常见的本地化工程工具有哪些?	20
48. 什么是本地化测试?	20

- 49. 常见的本地化缺陷有哪些? 20
- 50. 什么是敏捷本地化? 21

进 阶 篇

- 1. 常见的搜索资源有哪些? 25
- 2. 搜索引擎的常用检索技巧有哪些? 25
- 3. 如何拆分及合并 PDF 文档? 29
- 4. 如何翻译字符型 PDF 文档? 32
- 5. 如何翻译扫描型 PDF 文档? 34
- 6. 如何使用工具统计翻译项目的字数? 37
- 7. 如何对翻译项目进行去重? 40
- 8. 如何创建并加载翻译记忆库? 45
- 9. 如何使用记忆库预翻译文件? 49
- 10. 如何创建术语库? 51
- 11. 如何转换术语表? 56
- 12. 如何加载术语库? 63
- 13. 如何添加术语? 64
- 14. 如何提取术语? 66
- 15. 如何在 CAT 工具中使用机器翻译? 73
- 16. 如何在 CAT 工具中进行译后编辑? 78
- 17. 如何保证翻译质量? 80
- 18. 如何评估翻译质量? 86
- 19. 如何在 CAT 工具中审校文档? 94
- 20. 如何导出和导入外部审校文档? 95
- 21. 如何将译文定稿导出为双语对照格式? 98
- 22. 如何拆分及合并翻译项目文件? 101
- 23. 如何在 CAT 工具中处理必要的标记? 105
- 24. 如何处理不必要的标记? 107
- 25. 如何进行语料回收? 111
- 26. 如何进行语料清洗? 114

27. 如何进行语料对齐?	116
28. 正则表达式在翻译中的作用有哪些?	120
29. 如何使用正则表达式处理文档?	122
30. 如何在 Microsoft Word 中录制宏?	125
31. 如何使用宏快速处理翻译文档?	133
32. 如何使用 TransTools 处理文档?	144
33. 翻译文档管理的作用有哪些?	149
34. 如何快速比较翻译文本?	149
35. 如何快速搜索、重命名文档?	154

高级篇

1. 如何通过翻译管理平台进行协作?	163
2. 如何在 CAT 工具中自定义解析规则?	182
3. 如何使用字幕翻译一体化工具?	194
4. 如何进行 CHM 文档本地化?	206
5. 如何进行 DTP 文档本地化?	216
6. 如何进行 CAD 文档本地化?	222
7. 软件本地化的基本流程是怎样的?	231
8. 如何进行软件本地化?	231
9. 网站本地化的基本流程是怎样的?	244
10. 如何进行静态网站本地化?	245
11. 如何进行动态网站本地化?	250
12. 安卓应用程序本地化的基本流程是怎样的?	260
13. 如何进行安卓应用程序本地化?	261
14. iOS 应用程序本地化的基本流程是怎样的?	266
15. 如何进行 iOS 应用程序本地化?	267
参考文献	277
附录	279

基础篇

JICHUPIAN



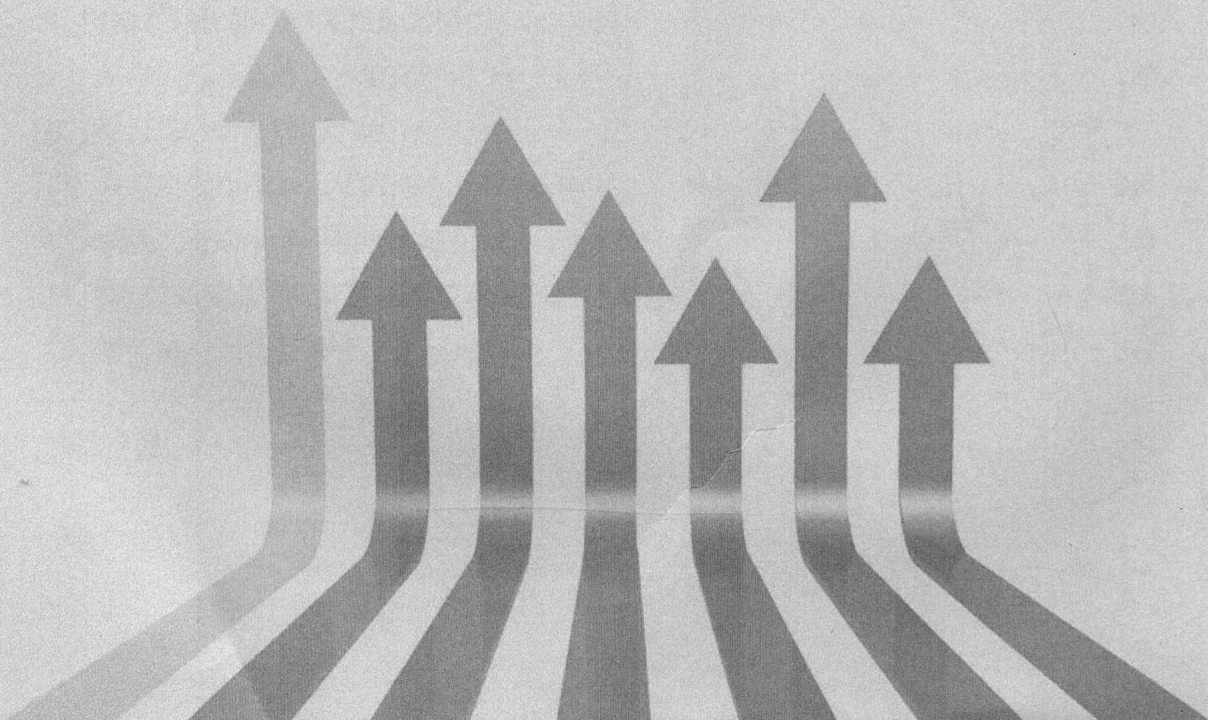
基础篇共 50 个问题，主要介绍语言服务领域的常识性概念，大致可分为以下几个模块：

（一）语言服务：该模块介绍了语言服务的定义，语言服务行业的产业链、标准规范和发展趋势等，并对新时代的语言服务从业人员所需要具备的能力素养作了概述。

（二）翻译技术：该模块首先从宏观角度出发，引入翻译技术、计算机辅助翻译、机器翻译、译后编辑等基本概念，随后对一些具体技术及其相关工具进行了介绍，如翻译记忆技术、术语管理技术、语料技术、光学字符识别技术、语音识别技术、质量保证技术、字幕翻译技术、桌面排版等。

（三）技术写作：该模块简要介绍了技术写作的定义和范畴、技术写作的实践原则、技术文档的特点，以及技术写作与学术写作的区别等。

（四）本地化：该模块包括 GILT（全球化、国际化、本地化、翻译）几个概念的定义与联系，以及本地化的基本内涵；同时还介绍了本地化工程、本地化测试以及敏捷本地化模式等概念。



1. 什么是语言服务行业?

语言服务行业 (Language Service Industry) 是指提供跨语言、跨文化信息转换服务和产品, 以及相关技术研发、工具应用、知识管理、教育培训等专业化服务的现代服务业 (中国翻译协会, 2014: 8)。

根据《中国语言服务业发展报告 2012》的描述, 语言服务业包括所有从事语言信息转换及关联服务的机构, 可分为三个层次: 核心层、相关层和支持层。

(1) 核心层是指经营或其业务的主要内容为提供语言间信息转换服务、技术开发、培训或咨询服务的企业或机构, 如翻译企业、本地化企业、翻译软件开发企业、翻译培训机构、多语信息咨询机构等。

(2) 相关层是指经营或业务部分依赖于语言间信息转换的机构或企业, 包括国家外事、外宣和新闻出版部门, 大型跨国企业以及旅游、对外贸易和信息技术等涉外行业的机构和企业。

(3) 支持层是指为语言服务提供支持的政府部门、机构和企业, 包括政府相关决策和管理部门、行业协会、高等院校、研究机构等。

2. 语言服务行业的标准和规范有哪些?

就国外而言, 2006 年, 欧洲标准化委员会正式发布了翻译服务标准 EN 15038: 2006, 统一了欧洲各国的相关语言服务标准。同年, 美国材料与试验协会发布了 ASTM F2575-06《翻译质量保证标准指南》, 是美国翻译服务行业的一项重要规范性文件。2008 年, 加拿大标准总署发布了 CAN/CGSB-131.10-2008《翻译服务》, 是在欧洲翻译服务标准的基础上结合加拿大的具体情况修订而成的标准。2015 年, 国际标准化组织技术委员会 ISO/TC37 新颁布了世界公认翻译质量认证标准 ISO 17100。这项标准基于 EN: 15038, 并按照最高级别和国际化水平标准进行了修订与改进, 涵盖了会对翻译质量和交付产生影响的各个翻译流程。

就国内而言, 2003 年以来, 中国翻译协会也陆续出台了《翻译服务规范第 1 部分: 笔译》(GB/T 19363.1-2008)、《翻译服务译文质量要求》(GB/T 19682-2005)、《翻译服务规范第 2 部分: 口译》(GB/T 19363.2-2006)、《本地化服务报价规范》(ZYF 001-2013)、《本地化服务供应商选择规范》(ZYF 001-2014)、《笔译服务报价规范》(ZYF 002-2014)、《口译服务报价规范》(ZYF 003-2014)、《本地化译文翻译和文档排版质量评估规范》(ZYF 001-2016)、《口笔译人员基本能力要求》、《翻译服务采购指南》、《语料库通用技术规范》、《翻译服务: 口译服务要求》、《翻译服务: 笔译服务要求》等规范。

3. 语言服务行业有哪些发展趋势?

语言服务业已成为全球化时代我国社会经济发展的基础性支撑性行业, 呈现出全球化、产业化、多元化、集群化、智能化、流程化、协作化和职业化的发展趋势。其

中，全球化表现为语言服务提供商不断向全球扩张，市场需求持续增长，服务标准与规范国际化；智能化表现为计算机与互联网技术带动下的语言技术迅猛发展，翻译技术和工具不断改进，云翻译平台和人工智能（Artificial Intelligence, AI）翻译平台逐渐流行。

此外，在我国“一带一路”倡议和“新一代人工智能发展规划”等大政方针的背景下，语言服务在国家战略中扮演的角色日益突出，中国语言服务企业的服务模式也在逐渐变革，语言服务生态逐步完善。

4. 新时代译者需要具备什么样的能力？

综合来看，新时代译者是懂语言、懂技术、懂市场的人才，需要具备语言能力、技术能力和职业能力。

（1）语言能力表现为具备双语理解和转换能力、跨文化交流能力以及跨行业的专业知识。

（2）技术能力表现为具备良好的计算机技能、信息检索能力、术语能力、计算机辅助翻译（Computer-Aided Translation, CAT）工具应用能力，熟悉现代翻译的基本流程和规则，熟悉不同格式文件的处理和翻译方法。

（3）职业能力表现为了解语言服务行业的动态和发展趋势，能够把握语言服务行业的市场需求，能够运用进度管理、成本管理、质量管理等项目管理知识，具备项目组织、协调与整合能力以及协作意识、沟通意识、服务意识和保密意识等。

中国翻译协会于 2017 年发布的《口笔译人员基本能力要求》也提出，口笔译活动要求口笔译人员除了具备外语能力、口译能力或笔译能力外，还应具备语言文字处理能力、领域能力、研究能力、信息获取和处理能力、文化能力及技术能力等专业能力（中国翻译协会，2017）。

5. 什么是翻译技术？

翻译技术（Translation Technology, TT）是一个动态发展且开放的技术体系，其内涵随着时代和技术的发展而发展。ISO 17100: 2015 认为，翻译技术涵盖一系列从业者用于辅助翻译、修订、校对等的工具，包括内容管理系统、写作技术、桌面排版、文字处理、翻译管理系统、翻译记忆工具和计算机辅助翻译、质量保证工具、修订工具、本地化工具、机器翻译、术语管理系统、项目管理软件、语音-文本识别以及其他现存的和未来将要出现的多种翻译技术。

6. 翻译技术有哪些作用？

翻译技术突飞猛进，在译前、译中、译后发挥日益重要的作用。例如，在译前，翻译技术可作用于源文档撰写与质量控制、可译资源提取、工作量估算、项目和工作流创建、预翻译、术语抽取及术语库创建等方面；在译中，翻译技术可作用于翻译记忆复用、自动文本输入、自动拼写检查、术语识别与插入、搜索与验证和译后编辑；在译后，翻译技术可作用于翻译质量保证、本地化排版、本地化编译、本地化测试、

语言资产管理和译文格式保留等方面。简而言之，翻译技术的问世与发展加快了翻译速度，优化了翻译流程，降低了翻译成本，提升了行业整体的生产效率。

7. 翻译技术有哪些主要标准？

在信息技术迅速发展的今天，翻译技术与工具的应用日益广泛，翻译流程日益复杂，翻译项目文件格式日趋多样化，翻译数据交换越来越频繁，而纯文本形式的数据交换已难以满足时代发展需要。因此，制定翻译技术标准的主要作用在于促进不同翻译技术工具之间的数据兼容、交换与共享，提高翻译与本地化效率。

翻译技术执行的常用标准涉及翻译记忆库、术语库、断句规则、本地化等领域及同一领域下不同服务提供商间各子库、子集的交换，常见的有翻译记忆交换标准（Translation Memory eXchange, TMX）、术语库交换标准（TermBase eXchange, TBX）、断句规则交换标准（Segmentation Rule eXchange, SRX）、全球信息管理度量交换标准（Global Information Management Metrics eXchange, GMX）、本地化交换文档格式标准（XML Localization Interchange File Format, XLIFF）、达尔文信息类型化体系结构标准（Darwin Information Typing Architecture, DITA）等。

其中，TMX 是一种基于可扩展标记语言（eXtensible Markup Language, XML）的翻译记忆数据交换标准。它能够在不同的 CAT 工具及本地化工具之间无损地传递翻译记忆数据，其中包含的信息主要有源语及译语句段、语言名称、创建日期和时间等。TBX 也类似，是一种开源、基于 XML 的术语数据标准格式，能够促进术语信息的流通和交换，提高术语的一致性。

8. 什么是口译技术？

口译技术（Interpreting Technology, IT）是指远程口译、电话口译、电视口译、视频会议口译、智能机器口译等技术，涉及语音识别软件、口译术语库、视频会议操作系统、智能笔等口译职业辅助工具。其中，机器口译（Machine Interpreting, MI）是指让计算机实现从一种语言的语音到另一种语言的语音自动翻译的过程。机器口译包含三项核心技术：语音识别、机器翻译和语音合成。

9. 什么是计算机辅助翻译？

关于计算机辅助翻译（Computer-Aided Translation/Computer-Assisted Translation, CAT）的概念，国内外诸多学者均有论述。归结起来，其大致可分为广义和狭义两个维度的概念。

广义的计算机辅助翻译相当于翻译技术，涵盖了译者在翻译过程中可能用到的提高翻译效率的信息技术，例如译前的编码处理、可译资源提取、字数统计、任务分析、术语提取等；译中的片段复用、搜索验证、术语识别、进度监控；译后的格式转换、模糊匹配、自动化质量保证、语言资产管理等；以及语料自动对齐、机器翻译、语音输入、语音翻译等技术。

狭义的计算机辅助翻译指利用翻译记忆来简化重复劳动的信息化技术。它利用计