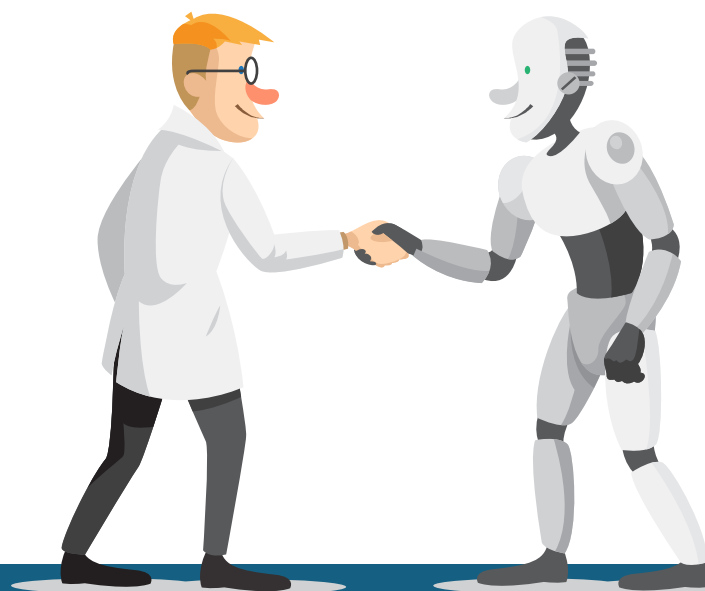


# 第11课 机器翻译的译后编辑 策略、方法与环境

黄婕

2024年11月27日/12月3日

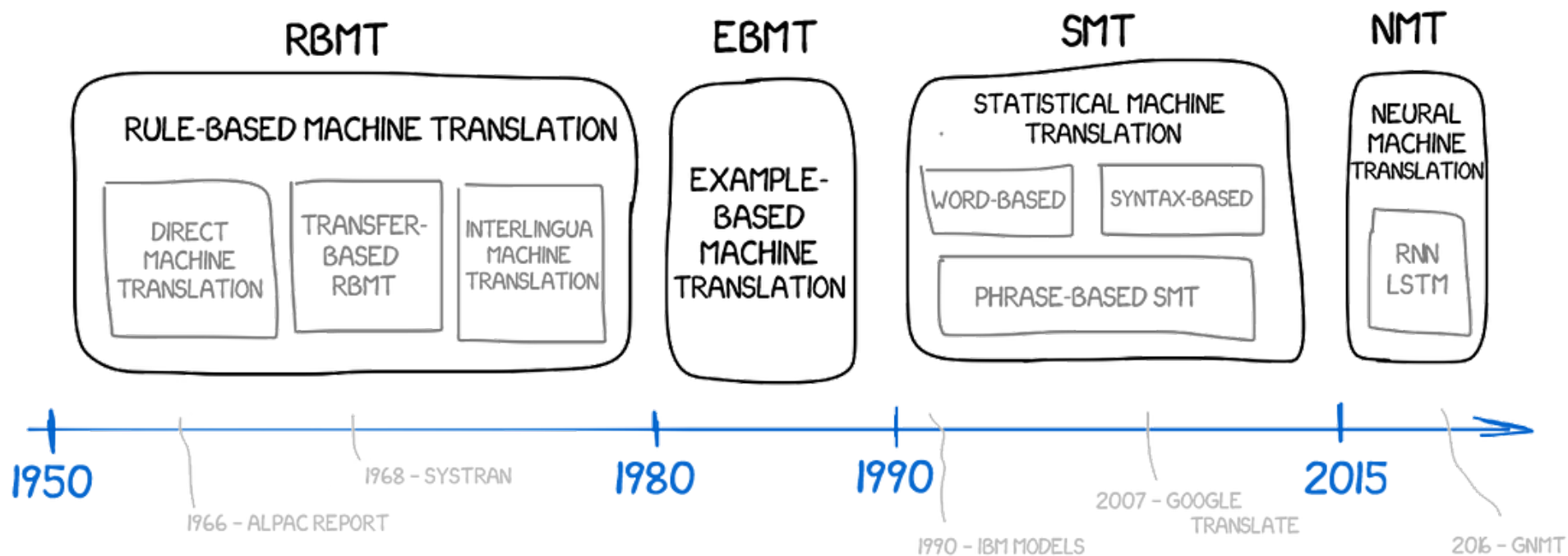


# 本节内容

- § 译者对机器翻译的态度如何?
- § 译后编辑些什么?
- § 开始吧, 译后编辑!
- § 机器翻译有哪些错误类型?
- § 译后编辑用什么工具?

# 机器翻译发展历程

## A BRIEF HISTORY OF MACHINE TRANSLATION



# Quiz

机器翻译的研究开始于哪个年代?

A.1970s

B.1950s

C.1960s

机器翻译的初始开发思路是什么？

A. 基于案例的机器翻译

A. 统计机器翻译

B. 基于规则的机器翻译

讨厌机器翻译，抢我的工作

# 1. 译者对机器翻译的态度如何？

好像挺方便的，想学一学

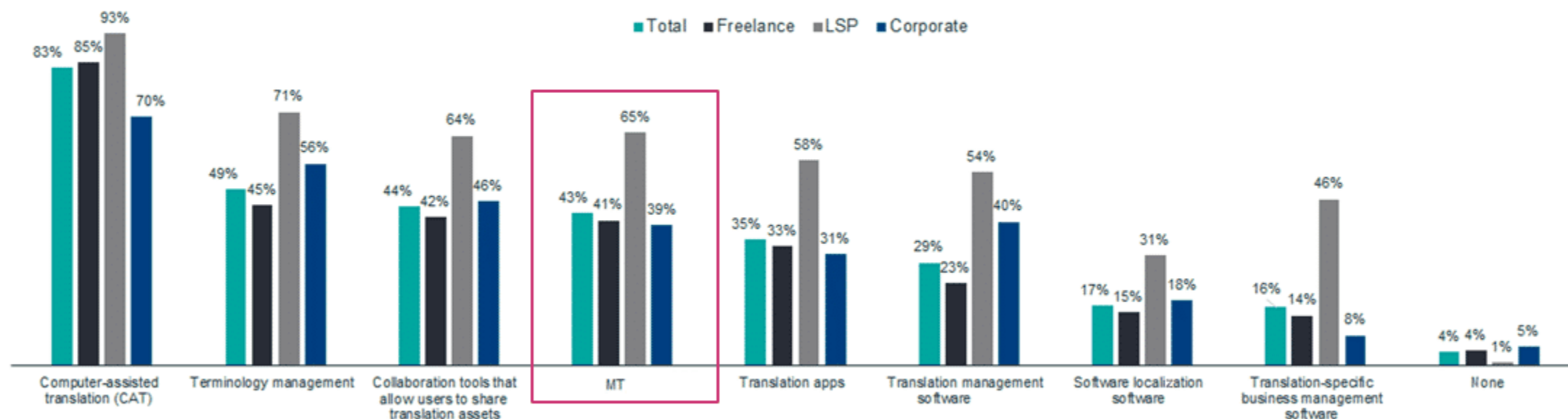
用了机器翻译，我翻得更快更好了！

2020年，SDL公司调查了语言服务行业对机器翻译的使用情况

发现使用机器翻译的 **LSP** 很多，达到总数的**65%**

不过自由译者的接受度就没那么高了

# Translation Technology Insights research: use of translation tools



What types of translation software do you/does your organization currently use?



# 译者在担心些什么？

NMT 才出现5年，  
太新了

NMT 模型的训练语  
料不同，不一定有  
我想要的效果

机器翻译来了，我  
挣的钱就少了

抢我的工作啊啊！



# Why are translators worried about machine translation?



## Still-new technology

NMT models have only been around for the last five years or so



## Trained differently

Not all NMT models are trained equally – they are trained in different ways, with different types of content



## Reduced rates

Rates are often cut when translators use MT



## Replacement for human translators

Translators are concerned that MT will start taking their jobs

# 对手很强，怎么办？

知己知彼，百战不殆。

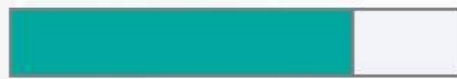
——《孙子兵法》

37%



of freelance  
translators believe  
MT output quality is  
good

81%



of freelance  
translators believe raw  
MT output varies from  
client to client

# 是对手，也是帮手

---

语言服务需求剧增，工作量超出人工处理能力范围

工作量

周转时间

交付周期变短，工作效率、流程势必优化

2016年以来，NMT质量显著提升，接受度提高

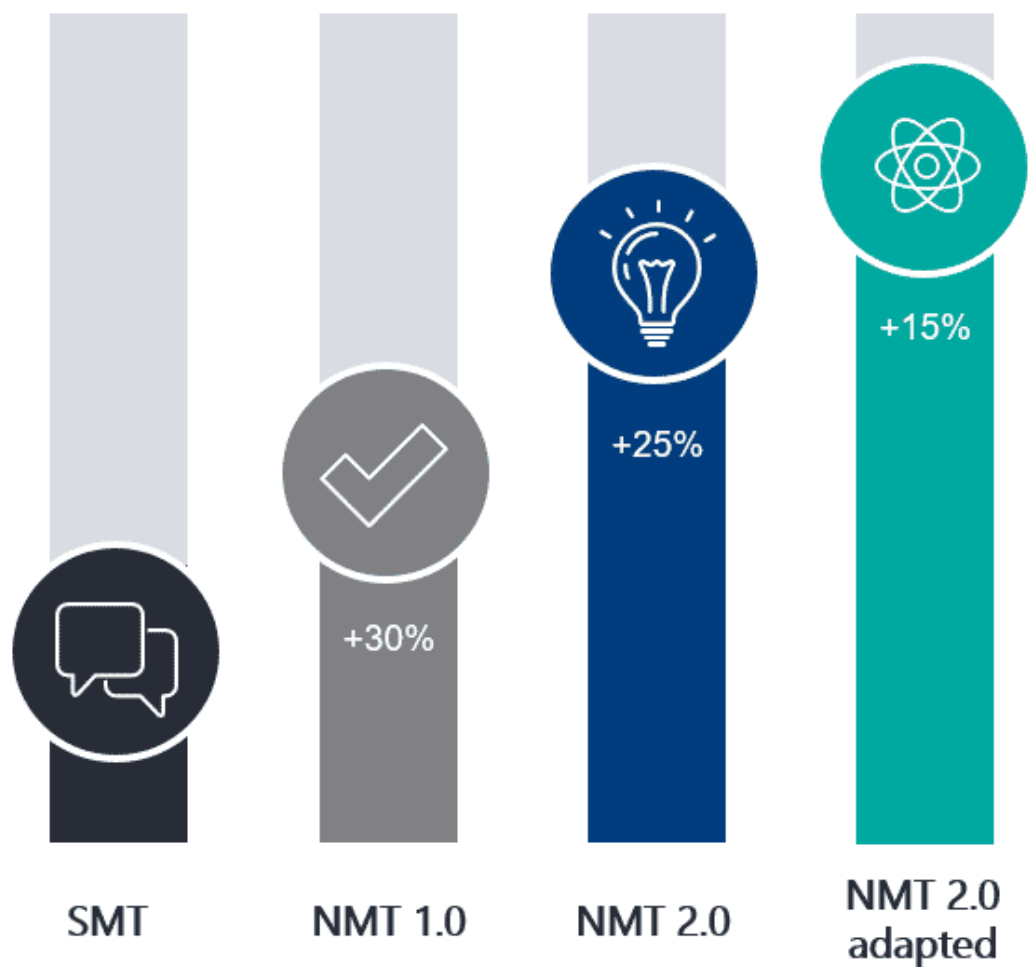
接受度

持续革新

在大语言模型、大数据加持下，技术持续革新

# MT质量的飞升

---



# 你，准备好了吗？

---



MT 不是翻译工作的威胁。



MT可以带来很多新机遇。神经网络机器翻译改变了翻译行业的游戏规则，突破了机器翻译的技术瓶颈。



译后编辑也是一种技能。和其他语言技能一样有着学习曲线，需要加以培养。



定期进行译后编辑训练，能帮助你得到更多工作机会，更了解人工翻译和译后编辑之间的差异。

## 机器翻译的需求从何而来?

- A. 缩短交付周期、降低成本
- B. 增加可完成的翻译任务量
- C. 人们对机器翻译的接受度提高
- D. 以上选项都对



不是所有项目，都可以译后编辑

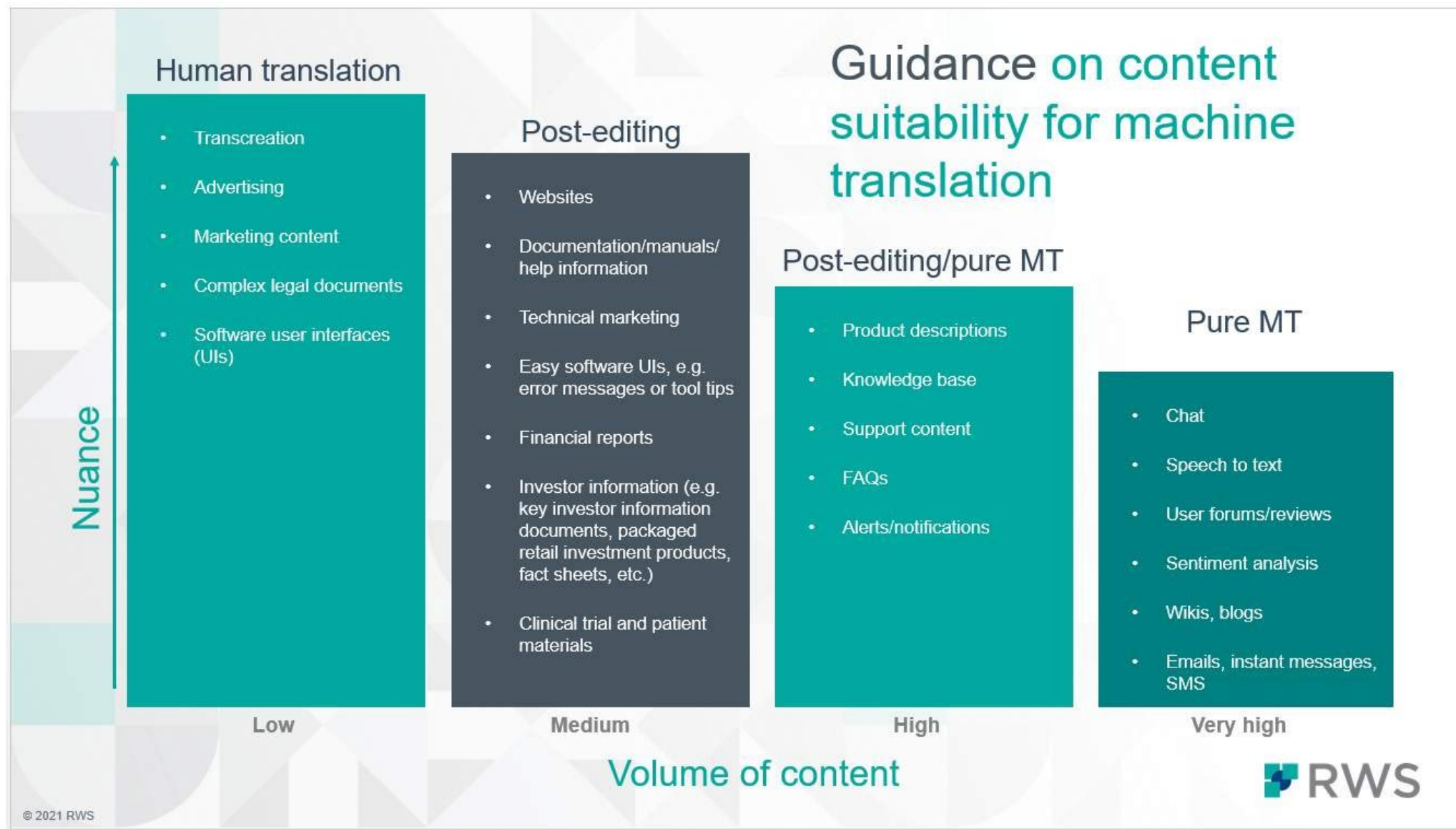
## 2. 译后编辑些什么？

有的.....机器翻译就够啦！

有的.....只能靠我们人类！



# 译后编辑适用的范围



# 适合译后编辑的具体场景

1

General content suitability and the chances of success based on language, vertical and content type

文本题材

2

Quality requirements

翻译质量要求

3

Volume requirements

工作量要求

4

Proposed future content (required when architecting a long-term solution)

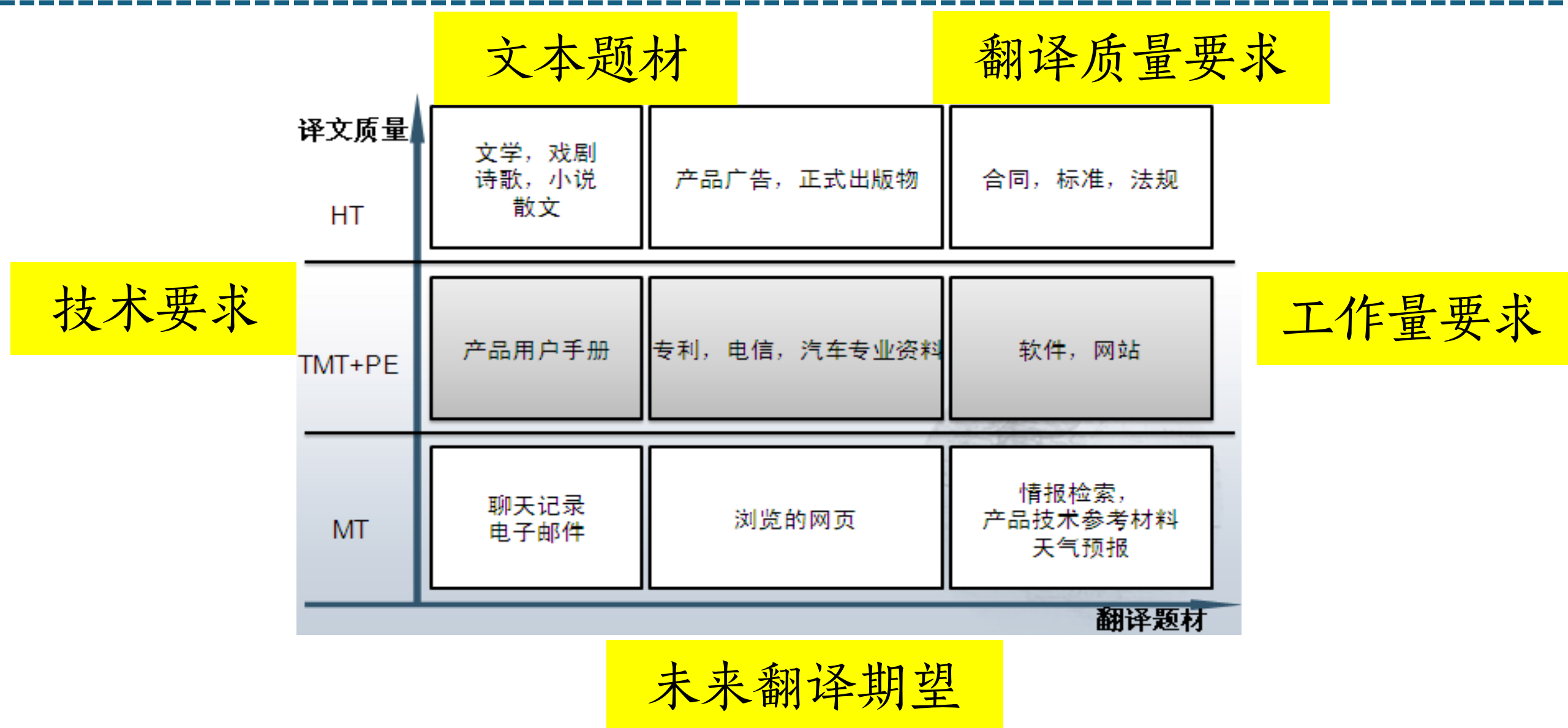
未来翻译期望

5

Technical integration requirements for the production workflow (CMS, CAT tool, TMS)

技术要求

# 适用于译后编辑的题材



# 机器翻译模型：通用 vs. 定制

## Generic neural machine translation models



Developed by Language Weaver for 130+ language pairs



Trained with hundreds of millions of words of bilingual data



High-quality data mined from reliable sources in the public domain



Applicable across language pairs, verticals and content types

# 机器翻译模型：通用 vs. 定制

## Customized neural machine translation models



Optimize the output for specific content



Trained based on specific bilingual data



Usually improved by additional relevant and high-quality data



Adhere to specific terminology and style

# QUIZ

通用 NMT 模型的理想基础是什么？

A. 双语词典

B. 一个包含数亿词汇的双语语料库

C. 一套 200-300 条转换规则

# Quiz

哪种 MT 解决方案能最大程度地遵循特定术语和风格?

A. 通用 MT 模式

B. 无 MT 模式

C. 定制 MT 模式

### 3. 开始吧，译后编辑！





# 什么是机器翻译的译后编辑

## 机器翻译 ( Machine Translation )

- 使用计算机系统将文本或语音从一种自然语言自动翻译为另一种语言
- ISO 17100-2013:  
Automated translation of text or speech from one natural language to another using a computer system

## 译后编辑 ( Post-editing )

- ISO 17100-2013:  
检查和修正机器翻译的输出  
( To check and correct MT output )

# 译后编辑的标准

---

- ISO 18587-2017: Translation services — Post-editing of machine translation output — Requirements
- GB/T 40036-2021 翻译服务：机器翻译结果的译后编辑要求
  - 译后编辑定义
  - 译后编辑流程
  - 译后编辑能力与资质
  - 完全译后编辑需求

# 译后编辑策略: light vs. full

## 轻度译后编辑 ( Light post-editing )

- 尽量多的采用机器翻译的原始输出，结果应可理解并准确，但不要求风格适当。
- 确保未添加或遗漏信息；修改任何不恰当的内容；在意思不正确或不清楚的情况下重组句子结构。
- 重点是修改错译、文化差异的内容、重新组织句式结构、不修改译文风格。

## 深度译后编辑 ( Full post-editing )

- 结果应准确、可理解、风格运当，句法、语法和标点符号使用正确.目的是产生与人工翻译结果效果相同的译文。
- 重点是确保未添加或遗漏信息；修改任何不恰当的内容；在意思不正确或不清楚的情况下重组句子结构；生成语法、句法和语义均正确的目标语言内容；遵循客户和/或领域专业术语规范；遵循拼写、标点符号和断字规则；确保使用适合文本类型的风格，并遵守客户提供的风格指南；遵循格式规则。

# MT + Full PE + review

## Quality matrix for full post-editing + review

Criteria	Full post-editing + review
High-quality deliverable comparable to human translation + review	✓
Correct grammar, spelling and punctuation	✓
Appropriate style	✓
Correct and consistent terminology	✓
Deliverable reads well and is suitable for its intended purpose	✓
Post-editors use the same references and processes as for conventional human translation	✓

# MT + Full PE (No review)

## Quality matrix for full post-editing *(no review)*

Criteria	Full post-editing (no review)
Accurate deliverable comparable to translation-only service	✓
Self-checks for grammar, spelling and punctuation	✓
Appropriate terminology is observed but a degree of inconsistency is possible	✓
Style requirements are observed but a degree of inconsistency is possible	✓
Natural fluency of all sentences is guaranteed	✗
Consistency between post-editors is guaranteed	✗

# MT + light PE (No review)

## Quality matrix for light post-editing (*no review*)

Error category	Issue	LPE	Comment
Mistranslation		✓	Correction of critical and major errors only
Terminology	In-context terminology choices	✓	Automated checks for client-preferred terms at end possible only if glossary provided in MultiTerm format
Accuracy	Omissions/additions	✓	Correction of critical and major errors only
Language	Grammar	X	Not corrected unless meaning is affected
	Spelling	X	Not corrected unless meaning is affected
	Punctuation	X	
Consistency		X	
Style	General style	X	
Country	Country standards	X	
	Register and tone	X	

# Pure MT

## Quality matrix for pure machine translation

Criteria	Pure MT
Completely automated translation process	✓
Instant turnaround	✓
Suitable where speed of information transfer is key	✓
Language quality control and compliance with existing terminology or style	✗
Free of mistranslations, untranslated parts and nonsensical translations	✗



# 译后编辑策略取决于质量要求

## Service levels – quality matrix

Service level	Workflow	What to expect
MT + full post-editing + review (machine-assisted publishable) <b>Focus on quality</b>	Apply translation memory – apply MT – full post-editing – review – customer delivery	<ul style="list-style-type: none"><li>✓ Comparable to full human translation and review</li><li>✓ Lower cost and faster turnaround than human translation and review</li><li>✓ Terminology compliant</li><li>✓ Style-guide compliant</li></ul>
MT + full post-editing (no review) (machine-assisted) <b>Focus on accuracy</b>	Apply translation memory – apply MT – full post-editing – customer delivery	<ul style="list-style-type: none"><li>✓ Comparable to human translation with no review</li><li>✓ Lower cost and faster turnaround</li><li>✓ Terminology is observed but a degree of inconsistency is possible</li><li>✓ Style requirements are observed but a degree of inconsistency is possible</li><li>✗ Sentences may lack natural fluency</li><li>✗ Consistency between post-editors cannot be guaranteed</li></ul>
MT + LPE <b>Focus on understandability</b>	Apply translation memory – apply MT – LPE – customer delivery	<ul style="list-style-type: none"><li>✓ MT output is edited to an understandable level</li><li>✓ Focus is on meaning and usability of the output</li><li>✗ Style is not adjusted and may lack natural fluency</li><li>✗ Register and tone are basic, not adapted to text type/target audience</li><li>✗ Sentence structures are readable, but not perfect in terms of spelling and grammar</li></ul>
Pure MT (raw MT) <b>Focus on speed</b>	Apply MT – customer delivery	<ul style="list-style-type: none"><li>✓ Completely automated translation process, at very low cost</li><li>✓ Suitable where speed of information transfer is key</li><li>✗ No language quality control and no compliance with existing terminology or style</li><li>✗ Mistranslations, untranslated parts and nonsensical translations are likely</li></ul>



# 译后编辑的行业应用

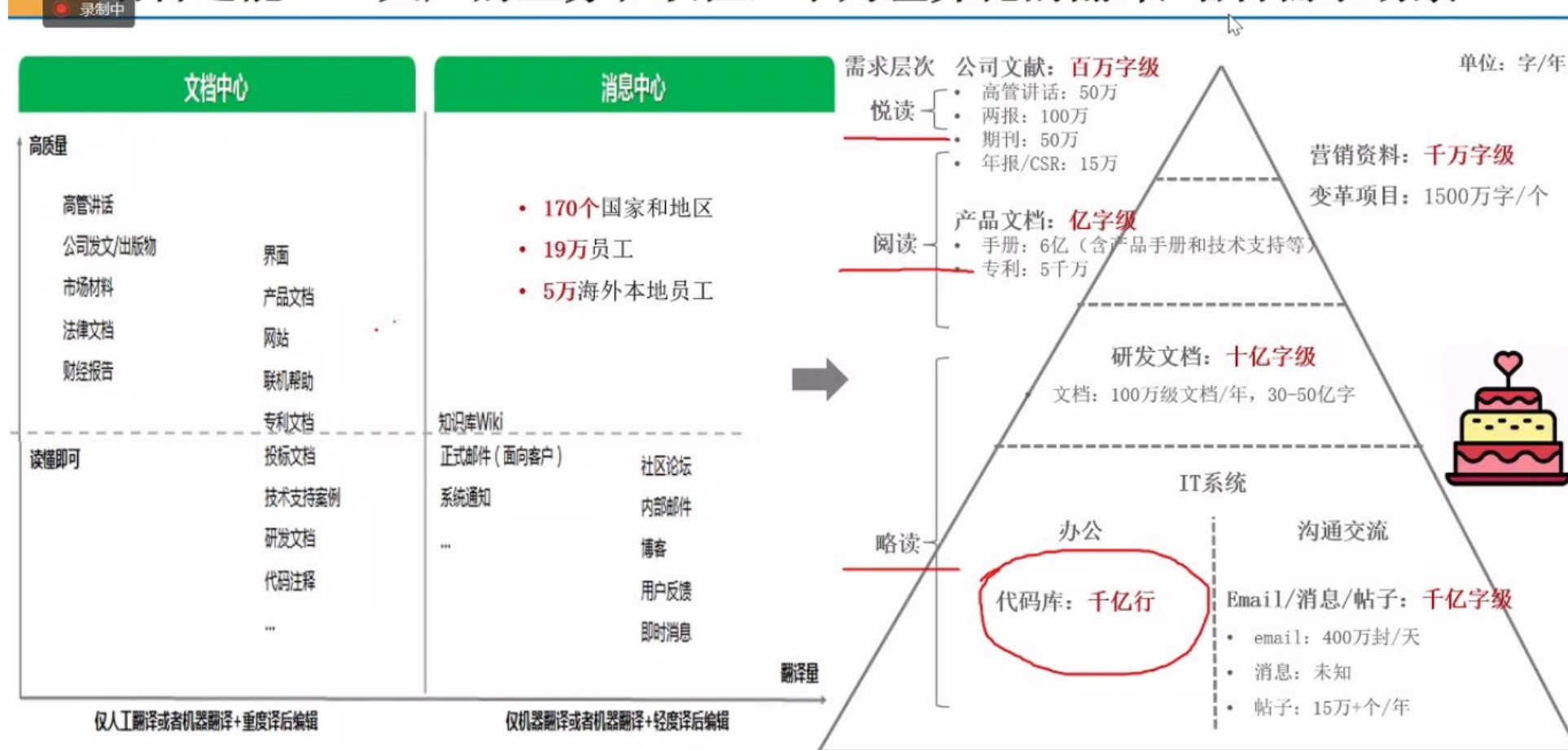
- 在参与调研的翻译企业实施的翻译项目中，2023年有52.9%的项目使用了机器翻译。
  - 30.4%的项目采用了 “机器翻译+译后编辑” 模式
  - 22.5%的项目采用了 “全机器翻译交付” 模式。

类型	年代	2017	2018	2019
提供MTPE服务的企业占比		29.86%	36.44%	36.53%
MTPE的服务占比		3.74%	3.64%	3.81%

# 华为公司的翻译方式与译后编辑

语言之能 - 大厂的业务和职位，华为差异化的翻译/语言需求场景

正在讲话: 吴永波



知己知彼.....错哪改哪

## 4. 机器翻译有哪些错误类型？

预判 MT 的预判

# Sustainable post-editing

## Requirements for sustainable post-editing

- Good understanding of the source and target language
- Subject-matter expertise
- Proficient use of translation tools
- Willingness to give MT and post-editing a chance
- Understanding of expected MT behaviours



# 译后编辑规则分类

## 语言类译后编辑规则

- 思维差异的译后编辑规则
- 词汇差异的译后编辑规则
- 语法差异的译后编辑规则
- 使用正确的惯用句式

## 增译和漏译译后编辑规则

- 语料训练的增译
- 词句不平行的增译
- 机器翻译局限的漏译
- 机器翻译语料训练的漏译

## 翻译风格译后编辑规则


- 粗体与斜体
- 丢失或多的余的空格
- 符号翻译错误
- 专有名词错误
- 缩略语错误
- 信息文本的无主语句子
- 使用规范的动宾搭配
- 去除冗余信息
- 确保一致性

# 机器翻译常见错误类型

---



理解错误



术语错误



表达错误



语态表达错误



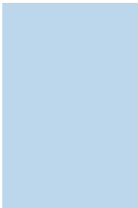
俚语错误



格式错误



过译



句子结构错误

# I. 理解错误

- Source: The average growth rate for the top 100 in our ranking this year was 12.9% in 2019. The top 20 grew at an average of 10.1% — 2.8% below the top 100 average.
- MT: 我们今年排名前100名的平均增长率在2019年为12.9%。前20名的平均增长率比前100名平均增长率低10.1%-2.8%。
- PE: 2019年前100名语言服务商的平均增长率为12.9%。前20名的平均增长率为10.1%——比前100名的平均增长率低2.8%。

## 2. 术语错误

- **Source:** The following section builds on the previous recommendations by the WHO (2006) and NICE (2018), Public Health England (2014) and Institute for European Environmental Policy (IEEP) (2016)...
- **MT:** 以下部分以世界卫生组织（2006年）、NICE（2018年）、英格兰公共卫生组织（2014年）和欧洲环境政策研究所（IEEP）（2016年）之前提出的建议为基础.....
- **PE:** 以下内容以世界卫生组织（2006）、英国国家卫生与临床优化研究所（2018）、英国公共卫生部（2014）和欧洲环境政策研究所（2016）之前提出的建议为基础.....



### 3. 表达错误

- Source: This text is intended to be a useful companion as you make your own transitions journey, a resource for you to reference while you teach reading and writing to children.
- MT: 这篇文章的目的是作为一个有用的伴侣，当你做你自己的过渡之旅，一个资源供你参考，当你教孩子阅读和写作。
- PE: 但本书的目的是能够成为您在教学方法转型之路上的有益伙伴和儿童读写教学中的可靠参考。

## 4. 语态表达错误

- **Source:** Urban green spaces are considered to be urban spaces covered by vegetation of any kind..
- **MT:** 城市绿地被认为是被任何类型的植被覆盖的城市空间。
- **PE:** 城市绿地是指植被覆盖的城市空间。

## 5. 俚语错误

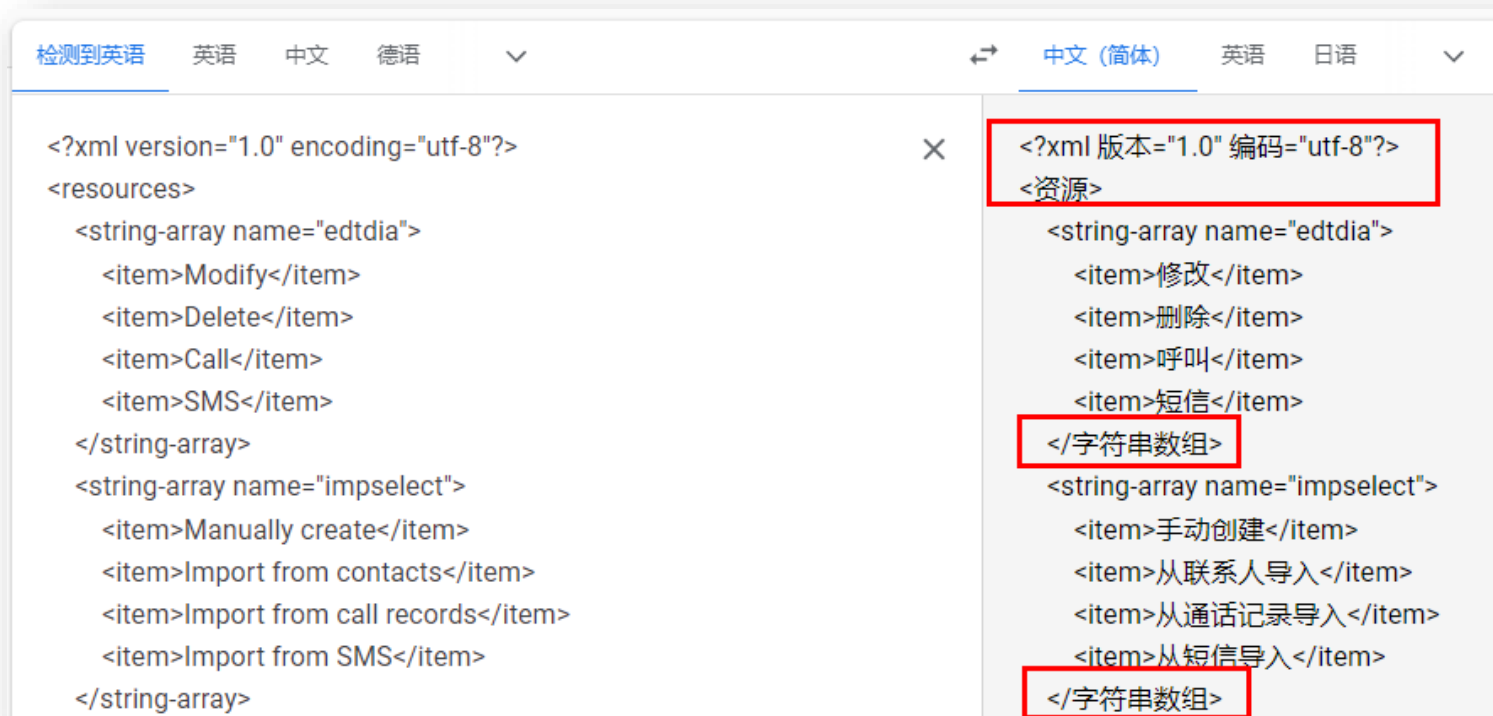
- **Source**: As has been said so well, “Once you see it in action, you can’t go back”
- **MT**: 正如已经说得那么好, “一旦你看到它在行动, 你就不能回头”。
- **PE**: 俗话说得好: “开弓没有回头箭”。

## 6. 格式错误

- Source: Gets the localized name of the `< ui > Toolbox < /ui >` tab with which this toolbox item is associated.
- MT: 获取被相关与该工具箱项 `< ui >` `< / ui >` 的标签的本地化的名称。
- PE: 获取被相关与该工具箱项 `< ui >` `< /ui >` 的标签的本地化的名称。

## 7. 过译 ( Over translation )

- Source: See also Hide ( ) and Show ( ).
- MT: 另请参见隐藏 ( ) 和显示 ( )。
- PE: 另请参见Hide ( ) 和Show ( )。



## 8. 信息操作文本无主句翻译错误

- Source: 此时需要重新挂载SUSE Linux 12 SP2镜像文件。
- MT: In this case, **you need to** mount the SUSE Linux 12 SP2 image file again.
- PE: In this case, mount the SUSE Linux 12 SP2 image file again.

## 9. 其他错误

- Source: In terms of volume translated from Chinese into foreign language **X**, the top five languages are: English, French, Japanese, German and Russian.
- MT: 从中文译成外文**X**的数量来看, 前五位的语言是: 英语、法语、日语、德语和俄语。
- PE: 就中文翻译成外语的数量来看, 排名前五位的语种是: 英语、法语、日语、德语和俄语。

# 译后编辑指南



1. 考虑目标受众、客户和内容类型。
2. 一定要先阅读原文。
3. 找出原文中不确定的任何元素。
4. 现在查看 MT 译文，看看它与你心目中的译文有什么差异。
5. 以 MT 译文为基础，根据需要修改译文。
6. 仔细重读该语段，并与原文进行比较。
7. 做完最终修改后，重读该语段。



# 案例练习

- 任务：
1. 在审校模式，标注当前机器翻译中出现的错误类型
  2. 在翻译模式，对机器翻译的译文进行译后编辑

- Case：
1. Source: Failure is glorious.docx
  2. NMT: Failure is glorious.docx.sdlxliff

## 5. 译后编辑用什么工具?



# 译后编辑方法与工具

---

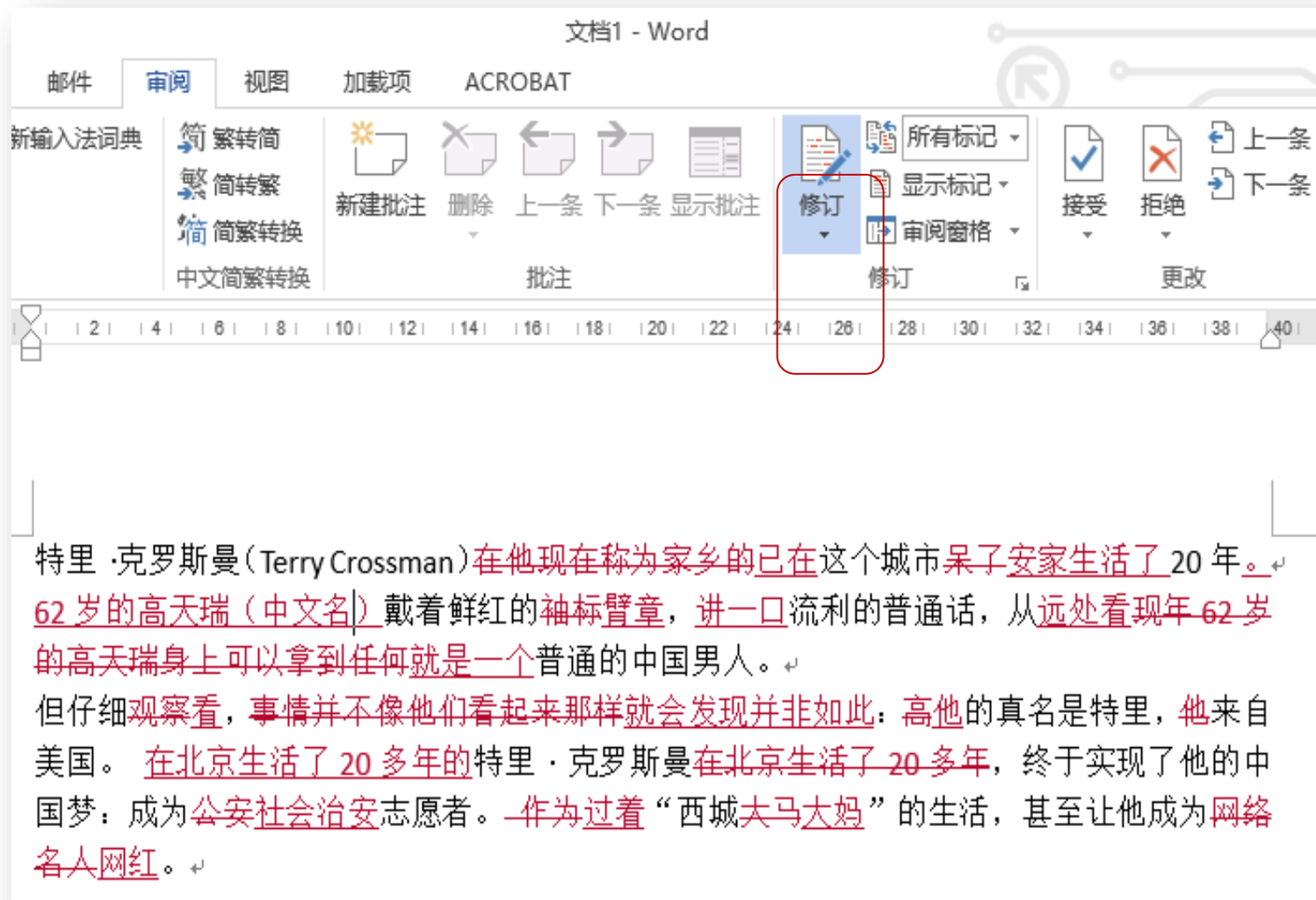
- 译后编辑方法

- 人工译后编辑（HPE）
- 自动译后编辑（APE）

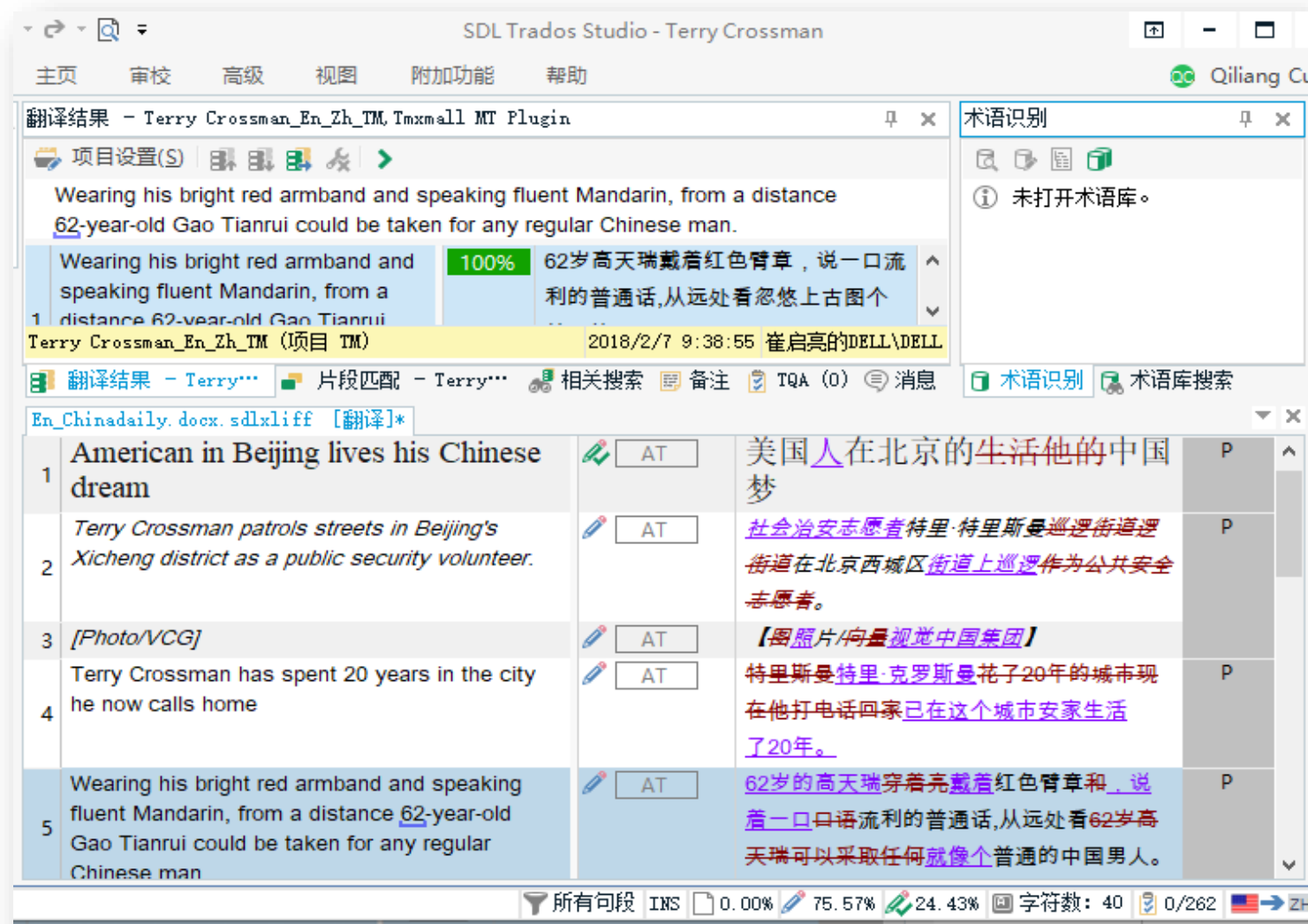
- 译后编辑工具

- 文字处理软件中的译后编辑
- 桌面版计算机辅助翻译工具中的译后编辑
- 在线计算机辅助翻译工具中的译后编辑

# 工具I：文字处理软件中的译后编辑（“修订”模式）



# 工具2：计算机辅助翻译工具中的译后编辑 ( Trados studio )



# 工具3：在线CAT软件的译后编辑（YiCAT）

原文	译文	Aa	^	v	
1	American in Beijing lives his Chinese dream	美国人在 <del>在</del> 的北京 <del>实现他的</del> 中国梦	✓	AT	译后编辑
2	Xinhua/chinadaily.com.cn	新华社/ chinadaily.com.cn	✓	AT	译后编辑
3	2017-08-31 08:54:06	2017-08-31 08:54:06	✓		译后编辑
4	Terry Crossman has spent 20 years in the city he now calls home	特里·克罗斯曼(Terry Crossman) <del>在这个他</del> <del>现在称之为家的城市</del> 在北京安家生活了20年	✓	AT	译后编辑
5	Wearing his bright red armband and speaking fluent Mandarin, from a distance 62-year-old Gao Tianrui could be taken for any regular Chinese man.	62岁的高天瑞戴着鲜红的臂章袖标，说着流利的普通话，从远处看，他和任何一个 就像一个普通的中国男人没什么区别。	✓	AT	译后编辑
6	But on closer inspection, things are not quite	但仔细一看，事情并不像看上去的那样非	✓	AT	译后编辑

# 工具4: ChatGPT译后编辑

请校对国际软件测试认证委员会(ISTQB)基础级考试大纲第3章第一段的译文。下面第一段是英文原文，第二段是中文译文。

In contrast to dynamic testing, in static testing the software under test does not need to be executed. Code, process specification, system architecture specification or other work products are evaluated through manual examination (e.g., reviews) or with the help of a tool (e.g., static analysis). Test objectives include improving quality, detecting defects and assessing characteristics like readability, completeness, correctness, testability and consistency. Static testing can be applied for both verification and validation.

与动态测试相反，在静态测试，测试统治下的软件不需要执行。代码、过程规范、系统架构规范或其他工作产品通过人工检查（例如，评审）或借助工具（例如，静态分析）进行评估。测试的目标包括改进质量、检测缺陷和评估可读性、完整性、正确性、易测试性和一致性等特征。验证和确认都可以申请静态测试。

# 工具4： ChatGPT译后编辑

与动态测试相比，静态测试不需要执行被测试的软件。通过手动检查（例如评审）或使用工具（例如静态分析），对代码、流程规范、系统架构规范或其他工作产品进行评估。测试的目标包括提高质量、检测缺陷以及评估可读性、完整性、正确性、可测试性和一致性等特征。静态测试可用于验证和确认。

原来的译文存在哪些问题？

原来的译文相对准确，但存在一些微小的问题：

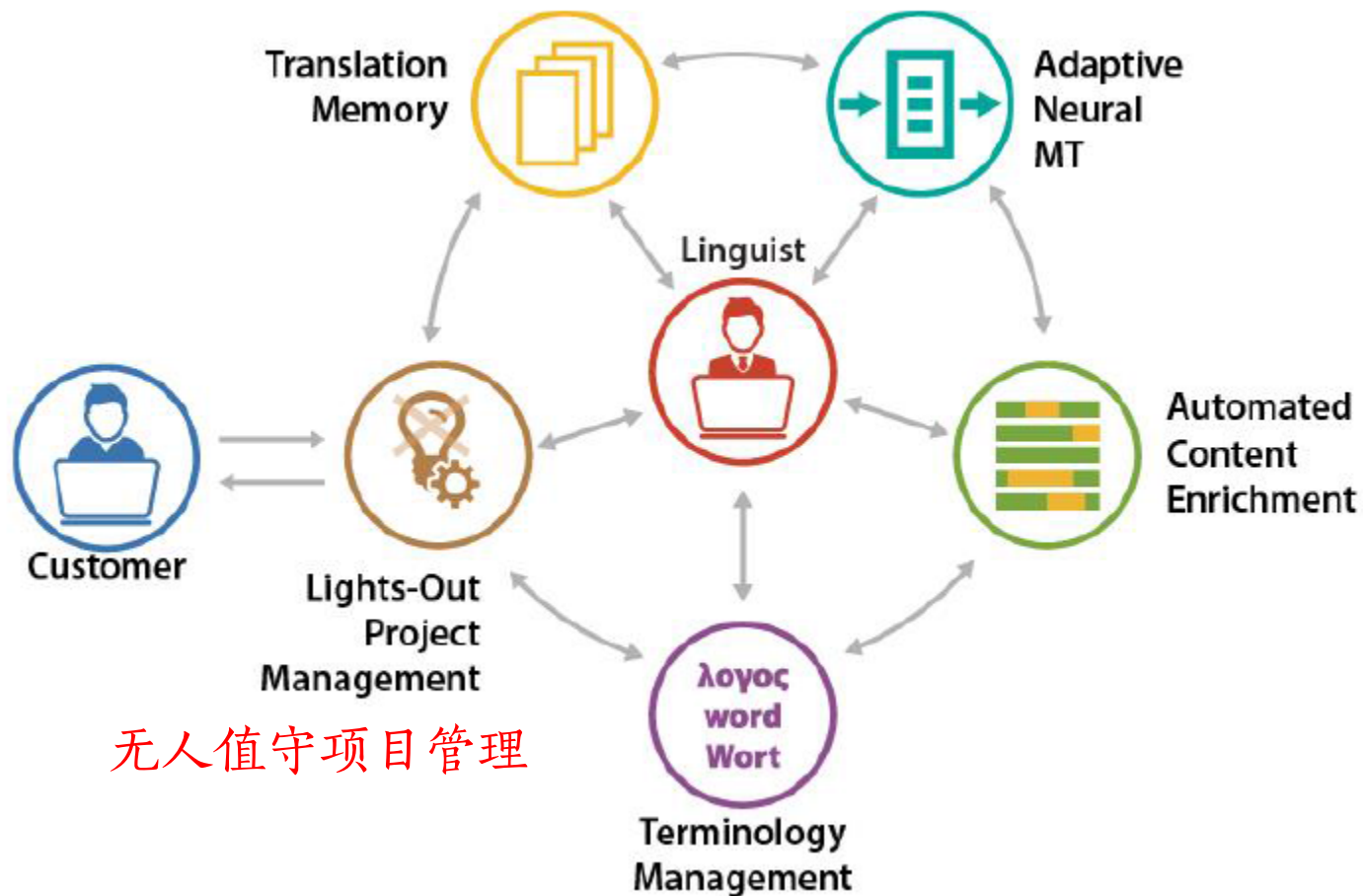
“测试统治下的软件”的翻译不够准确，应该翻译为“被测试的软件”或“测试对象的软件”。

“验证和确认都可以申请静态测试”的翻译不准确，应该翻译为“静态测试可用于验证和确认”。

总体而言，译文的意思基本传达了原文的信息，但在部分词汇和短语的翻译上还有改进的空间。



# 语言专家为中心的增强翻译模式



# 机器翻译的译后编辑能力

- 翻译能力
- 源语言和目标语言与文字能力
- 信息研究、获取与处理能力
- 跨文化交际能力
- 技术应用能力
- 行业领域能力

# 提高译后编辑能力的建议

1. 打好语言、翻译基础
2. 增强跨文化交际能力
3. 专注特定行业领域
4. 善于应用翻译技术与工具
5. 加强译后编辑实践

# 课外阅读材料

1. GBT40036-2021翻译服务机器翻译结果的译后编辑要求
2. ISO\_DIS\_18587-机器翻译输出的译后编辑标准.pdf
3. 论机器翻译的译后编辑\_崔启亮-中国翻译-2014(6).pdf
4. 译后编辑白皮书-SDL.pdf
5. RWS 译后编辑培训课程(1小时, 免费):

<https://www.trados.com/training/post-editing-machine-translation/>

*END*