

# 机器翻译发展态势和质量分析

金璐钰<sup>1</sup>, 徐云娟<sup>1</sup>, 姚建民<sup>2,3</sup>

<sup>1</sup>苏州托普信息职业技术学院, 江苏 苏州

<sup>2</sup>苏州大学计算机科学与技术学院, 江苏 苏州

<sup>3</sup>苏州市科学技术情报研究所, 江苏 苏州

收稿日期: 2022年1月20日; 录用日期: 2022年2月14日; 发布日期: 2022年2月21日

---

## 摘要

机器翻译是通过特定的计算机程序将一种语言信息翻译为另一种语言, 是人工智能研究核心任务之一, 涉及语言理解、语法和语义信息获取、语言生成等多种语言信息处理的过程。在分析机器翻译任务、发展历程和现状基础上, 通过翻译实例揭示了机器翻译的难点问题和质量门槛, 展望了未来发展的趋势。

## 关键词

机器翻译, 人工智能, 翻译质量, 翻译实用化

---

# Development Prospect and Quality Analysis of Machine Translation

Luyu Jin<sup>1</sup>, Yunjuan Xu<sup>1</sup>, Jianmin Yao<sup>2,3</sup>

<sup>1</sup>Suzhou Top Institute of Information Technology, Suzhou Jiangsu

<sup>2</sup>School of Computer Science and Technology, Soochow University, Suzhou Jiangsu

<sup>3</sup>Suzhou Institute of Sci-Tech Information, Suzhou Jiangsu

Received: Jan. 20<sup>th</sup>, 2022; accepted: Feb. 14<sup>th</sup>, 2022; published: Feb. 21<sup>st</sup>, 2022

---

## Abstract

Machine translation is a key technology for artificial intelligence to transform information in one language into another with a computing program, which is a combination of multiple language information processing of language understanding, syntactic and semantic information acquisition, language generation, etc. The task, development and status quo are analyzed for machine translation, followed by machine translation cases of mainstream systems for manifestation of difficult points, quality thresholds, and future prospects.

## Keywords

Machine Translation, Artificial Intelligence, Translation Quality, Translation Utilization

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

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



Open Access

## 1. 引言

机器翻译是算法利用语言统计信息、世界知识图谱以及相关翻译资源，将一种语言信息翻译成另外一种语言信息计算机应用，随着互联网和智能机的普及，机器翻译逐渐成为跨文化交流、信息检索必不可少的利器[1]。机器翻译涉及理论研究、应用开发、市场推广，应用场景首先是最基本的单词或词组查询，比如单个字词、简单词组或常用结构。其他应用场景是辅助翻译，利用计算机整合翻译资料，为类似翻译提供帮助，Google Translator Toolkit、Microsoft LocStudio、Trados 等占据计算机辅助翻译软件主要市场份额[2]。最终目标是机器翻译，应用计算机把源语言转换为目标语言，百度翻译、谷歌翻译、有道翻译等均属于机器翻译软件[3]。

## 2. 机器翻译的发展

机器翻译从上世纪 50 年代提出想法，美国乔治敦大学和 IBM 协同尝试基于语言规则的英俄翻译实验，到 70 年代否定了翻译质量的可行性。随着 80 年代计算能力的提高和电脑使用的普及，基于规则的机器翻译被重新提起，并进一步奠定了计算语言学的基础工作，词法分析、句法分析、翻译词典、双语语料库等资源大规模投入应用[4]。以词法、句法分析为主、语义分析为辅的基于规则转换的方法、基于统计的翻译方法、以至于 2000 年之后基于神经网络的机器翻译大幅提升了翻译软件的商业化，语音翻译、在线翻译系统进入了实用阶段[5]。

中国的机器翻译始于上世纪 50 年代，80 年代进入繁荣期，译星、通译等翻译软件相继走向专业化和市场化。近几年国内机器翻译和互联网、大数据同步繁荣，机器翻译服务进入实用化。2011 年，百度机器翻译服务上线。2015 年，阿里巴巴机器翻译推动“让商业没有语言障”。2016 年，腾讯支持 15 种语言翻译并上线同声传译功能。搜狗上线机器同声传译、旅游翻译机。科大讯飞推出硬件翻译产品。网易上线神经网络翻译及有道翻译等产品[6] [7]。

## 3. 机器翻译的应用

从不同的信息媒介看来，机器翻译可以在文本、语音和图像翻译等角度得到应用。

文本翻译是将源语言文字翻译到目标语言文字的信息转化过程，比如常见的百度翻译，可以帮助用户理解外语文本的信息。谷歌翻译、微软必应、百度翻译、有道翻译等在线系统都为用户提供了优质的免费多语言翻译系统，为不懂外语的人了解信息、外语水平一般的人获取快速高质量的翻译结果提供了便利的工具。文本翻译广泛应用在不同的领域，对不同行业、不同外语水平的用户都提供了便利、可靠的翻译工具，兼具了翻译词典、双语例句库、以及初步翻译助手的功能。但是，机器翻译仍然是更便利的双语词典和句对查询工具，翻译结果会出现很多语义、语法质量不足甚至根本的翻译错误，需要依靠人的语言和世界知识进行选择 and 补充[8] [9]。

语音翻译对文本翻译进行了信息载体上的扩充,更契合日常生活中的需求场景,可以在 PDA,桌面系统和智能手机中实现。语音翻译为异国旅游的人提供了方便。VoxTec 的 Phraselator 产品可以在数百种语言间翻译语句,美国军队应用在与异国场景沟通上,其中包括阿富汗和伊拉克。亚马逊 Alexa、苹果 Siri、微软 Cortana、百度翻译、科大讯飞等都提供语音翻译功能[10]。

图像翻译在计算机视觉领域是根据源领域的图像生成目标领域的对偶图像,和文本翻译类似,它将输入图像的表达转化为另一种表达,应用在图像创作、图像风格化、图像修复、黑白照到彩色照的转换等领域。同时,翻译系统识别图像中的文字,并翻译到目标语言,也是机器翻译在图像领域中的应用之一。

#### 4. 机器翻译的机遇和存在问题

深度学习和大数据技术推动了 AI 的进步,也带来了机器翻译的机遇和泡沫。谷歌研究人员发表论文,推出的神经机器翻译系统较之传统方法“英-西班牙翻译错误率下降了 87%;英-汉下降了 58%;汉-英下降了 60%”,已接近人工翻译的水平。2016 年 11 月,搜狗机器同传系统亮相世界互联网大会,CEO 王小川现场演讲、实时文字转录同时屏幕显示,对同声传译行业造成一定的震动。但一位备注为北京语言大学高级翻译学院教师的网友表示,搜狗对外声称的 90% 准确率是让人怀疑的,机器翻译无法像人一样体会上下文,以至于翻译出语焉不详的句子,这样的效果何以取代同传?百度自然语言处理部吴华在媒体采访中表示:机器翻译有两大问题造成翻译效果不好。一是训练语料本身是有噪音的。第二个是模型本身的不完美性[11]。

随着互联网的普及和神经网络翻译模型的性能提升,机器翻译在数据、算法和算力方面都大幅提升并在实际应用中获得用户的认可。在产品质量认可之外,我们从学术角度对机器翻译质量进行分析,希望能找到机器翻译进一步提升的方向。根据通用领域翻译的需求,我们把质量较好的百度翻译、谷歌翻译和必应翻译比较普遍存在的问题进行了分析,在本文中列举部分问题,作为机器翻译存在问题的典型案例,以备算法设计或数据加工参考。

**实例 1:** A brewery made headlines recently when its prototype drone delivered a small case of beer to **ice fishermen** on a frozen lake.

百度翻译:一家啤酒厂在最近的一个头条新闻时,它的原型无人机交付了一个小的情况下,啤酒**冰上的渔民**在一个冰冻的湖。

人工翻译:最近一家酿酒厂上了头条,他家的样品无人机为一片结冰的湖面上进行**冰钓的渔民**送了一箱啤酒。

行百里者半九十,实例 1 虽然看似简单的问题,却反映出机器翻译建模存在关键不足。翻译不能靠武断映射处理,翻译是再创作的过程。人工翻译知道“冰上的渔民”、“冰渔民”是在做“冰钓”。机器翻译系统却没有应变能力。在系统的词典中有 ice fishing,没有 ice fisherman,这个人工轻易可以解决的问题,却是机器翻译的卡脖子技术。

**实例 2:** Are we going to **brave the elements** and go for a walk?

百度翻译:我们要**勇敢的元素**,去散步吗?

人工翻译:我们要**冒着风雨**去散步吗?

人知道元素不散步,机器判断不明。世界事物和语言符号的关联,是翻译需要的知识,包含常识、语言知识、专业知识等等。人知道冰上的渔民或者冰渔民是冰钓者,元素不会散步。从词典词条译文的组合,到通顺的目标语,中间需要语义知识,目标语言调整和语义知识对翻译是很重要的。

**实例 3:** During the 1990s, universities were faced with significant pressure to **produce** innovative results that could be exploited more effectively by industry.

百度翻译：在 20 世纪 90 年代，大学面临着巨大的压力，**产生**创新的结果，可以更有效地利用行业。

谷歌翻译：在 20 世纪 90 年代，大学面临着巨大的压力，要**产生**可以被工业更有效地利用的创新成果。

必应翻译：1990 年代，大学面临着巨大的压力**产生**创新的结果，可以更有效地利用产业。

人工翻译：二十世纪九十年代的大学面临着巨大的压力，要**研发**出可以更高效地被工业所利用的创新成果。

“产生”是词典词条，“研发”不在词条里，但是“研发……成果”明显比“产生……成果”更符合目标语习惯，目标语的生成比词典映射要复杂一些。

**实例 4:** Drones change **the way that we see**. They provide a **specular** technology.

百度翻译：无人机改变了**我们所看到的方式**。它们提供了一种**镜面反射**技术。

必应翻译：无人驾驶飞机改变**我们看到的方式**。它们提供了一种**镜面技术**。

人工翻译：无人机改变了**我们看世界的方式**，提供了一种**窥探**的技术。

在词典词条“specular adj. 镜的；窥器的；用窥器(检查)的；镜子似的”，“用窥器(检查)的”是不太可能成为真实译文的，它只是一个“意义解释”。机器翻译应该从真实语料中学习译文，搜集的语料也要覆盖这类意义解释的译文词条。要研发合适的翻译模型，充分利用知识为翻译服务，比如利用专业知识、世界常识、语言搭配等服务机器翻译。

## 5. 结语

全球化和互联网化给机器翻译提供了宝贵的产业发展机会，神经网络和大数据为机器翻译提供了支撑技术，机器翻译逐渐实用化，由网页端向移动端转移、从文字翻译向语音和图像翻译扩展，被应用到各种各样的生活场景中，为人类交流提供了新的工具。我们从主流翻译系统存在的问题中看到未来技术发展的潜力，在算法改进、数据加工等方面进一步提供更好的机器翻译支撑，为各种应用场景提供更优质的翻译服务。

## 基金项目

本论文部分研究获得江苏省“青蓝工程”优秀青年骨干教师培养对象项目支持。

## 参考文献

- [1] 刘洋. 神经机器翻译前沿进展[J]. 计算机研究与发展, 2017, 54(6): 1144-1149.
- [2] 刘玉. 科技文本机器翻译句层错误分析[J]. 现代语言学, 2021, 9(4): 955-959.
- [3] 崔启亮, 李闻. 译后编辑错误类型研究——基于科技文本英汉机器翻译[J]. 中国科技翻译, 2015, 28(4): 19-22.
- [4] 殷姣. 英汉机器翻译中的词汇和句法错误分析[D]: [硕士学位论文]. 成都: 电子科技大学, 2017.
- [5] 金鹏, 张春祥, 冯禹瑄, 等. 基于语言学资源的汉-英机器翻译[J]. 科学技术创新, 2021(4): 99-100.
- [6] 曹宜超, 高翊, 李淼, 冯韬, 王儒敬, 付莎. 基于单语语料和词向量对齐的蒙汉神经机器翻译研究[J]. 中文信息学报, 2020, 34(2): 27-32.
- [7] 丁立福, 蒋威. 论文化语境对机器翻译质量的影响暨应对机器翻译弊端的策略——以“有道翻译官”为例[J]. 安徽理工大学学报: 社会科学版, 2021, 23(2): 81-85.
- [8] 申芳. 融合翻译知识的机器翻译质量估计算法研究[J]. 数字通信世界, 2021(4): 243-244.
- [9] 秦颖. 基于神经网络的机器翻译质量评析及对翻译教学的影响[J]. 外语电化教学, 2018(2): 51-56.
- [10] 孙逸群. 基于问卷与数据分析的机器翻译质量评价方法研究[J]. 中国外语研究, 2018(1): 72-77.
- [11] Chen, M.X., Firat, O., Bapna, A., et al. (2018) The Best of Both Worlds: Combining Recent Advances in Neural Machine Translation. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers), Melbourne, 76-86. <https://doi.org/10.18653/v1/P18-1008>