

国外计算机翻译的发展和近况

袁亦宁 (南京航空航天大学人文学院, 南京 210016)

摘要] 本文对国外计算机翻译的发展和近况进行了介绍和评述, 内容包括机器翻译和计算机辅助翻译的原理、功能和应用, 以及各自的特点和局限性。文章还对著名的计算机辅助翻译软件 TRADOS(塔多思)进行了简介。

关键词] 机器翻译; 机助翻译; 功能; 应用; TRADOS

中图分类号] TP3-05

文献标识码] B

文章编号] 1000-6141(2002)02-0058-02

人们通常用计算机翻译 (computer translation) 来统指机器翻译 (machine translation, MT) 和计算机辅助翻译 (computer-aided translation, CAT), 其实二者的概念有着很大的区别, 使用场合和效果都不同。前者指的是由计算机进行翻译, 而后者是由人工在计算机的辅助下进行翻译。

由于语言的语法、语义及文体等的复杂性, 目前机器翻译对大多数文本的翻译质量 (包括可读性和准确性) 离人们的期望还有相当差距, 其主要用途主要限于: (1) 企业部门对部分技术性文本进行翻译。(2) 帮助用户了解某些文本, 如产品说明书, 的大致内容。(3) 作为专业翻译人员对某些文本进行人工翻译前的参考翻译或者翻译后的校对参考。

机器翻译对文本语言的标准化的程度要求很高。某一文本是否可以被计算机正确翻译取决于它是否可以被计算机“理解”。也就是说, 原文必须是一种“受控语言 (controlled language)”, 即文本没有超出计算机的词典库范围的词汇和术语, 而且词汇和术语必须和词典库的词汇具有相同的释义。此外, 文本的歧义成分 (ambiguity)、印刷错误 (往往由不恰当地使用扫描仪造成) 和语法错误, 过分复杂的句法结构都会影响机器翻译的质量。因此, 机器翻译的对象一般是语言标准化和一致性程度相当高, 句法结构比较简单的技术性文本。例如, 国外的汽车、航空、国防等工业部门利用机器翻译对部分技术性文本, 如部件列表和规格列表等进行翻译。加拿大国家气象部门也使用机器释放系统对天气预报进行英-法语文本的翻译转换。国外的机器翻译的产品品种繁多, 主要包括: Globalink GTS Power Translator Transcent LOGOS Intelligent Translation System SYSTRAN Professional for Windows

鉴于计算机辅助翻译 (computer-aided translation, CAT) 与广大专业翻译人员关系比较密切, 以下加以述评。

一、机助翻译的原理和功能

尽管机器翻译的研究已经进行了几十年, 可是它的应用有很大的局限性。目前最先进的机器翻译系统也只能在非常有限的范围取代人工翻译。

专业翻译领域所涉及的某些技术性资料不仅专业性强, 而且篇幅长, 因此文本内容必然具有不同程度的重复。有关

统计数据表明, 部分技术性资料的重复率竟高达 70%。有些部分内容更新的文本的重复率可能会更高。计算机辅助翻译的基本原理利用翻译记忆 (translation memory, TM) 自动重复使用翻译过的语句或者语句成分, 使翻译人员免于翻译重复信息, 从而提高工作效率和译文的质量。

为了提高翻译效率和确保译文的一致性 (consistency), 计算机机器辅助翻译系统通常含有一系列的工具组件。其中两个最主要的工具组件是翻译记忆库 (translation memory) 和术语管理 (terminology management)。

翻译记忆库指的是计算机构建的原文和译文的语句等值数据库。当翻译人员使用计算机辅助翻译工具从事翻译工作时, 计算机便将翻译人员翻译的语句在后台的语言数据库中储存起来。在翻译过程中, 每当有相同或相近的原文语句重复出现时, 计算机自动显示翻译记忆库中的与之匹配的语句译文, 供用户重新使用。大多数翻译记忆软件都支持模糊匹配 (fussy match), 用户可以设定原文和译文之间的最低匹配质 (如 50% 或者 70%)。利用模糊匹配, 即使翻译记忆库中没有完全相同 (即 100% 匹配质) 的句子, 翻译人员仍然可以得到一个模糊匹配的句子, 只需对其中不同的部分加以翻译或者修改即可。翻译记忆库在后台不断更新和自动存储用户提供的新译文。随着翻译记忆库的不断充实, 随后的翻译工作的效率将变得越来越高。此外, 大多数翻译记忆产品具有网络共享功能, 使得多人合作从事某一翻译项目时可以通过局域网实时共享同一个翻译记忆库。

术语管理的工作原理是由计算机逐一扫描原文的术语, 并在计算机的专业词典库中进行查核。如果词典库中含有相应的术语, 计算机便自动为翻译人员提供对应译文。术语管理可以确保翻译术语的一致性和准确性, 同时也能提高翻译人员的工作效率。术语管理的功效取决于专业词典库的质量和规模。

计算机机器辅助翻译系统的其它工具组件的功能包括: 分析统计某一文本的字数和重复率; 文档格式转换; 利用已经翻译的文本建立翻译记忆库; 查找某一术语在不同语境的译文等等。

收稿日期] 2001-12-24

作者简介] 袁亦宁 (1956-), 男, 南京航空航天大学人文学院英语系副教授, 硕士, 研究方向: 语言教学和翻译。

二、机助翻译的益处

使用计算机机器辅助翻译可以比较显著地提高译文的质量和翻译的工作效率,降低成本。其益处包括:

1. 确保译文(包括术语、短语和语句的一致性),尤其是在多人共同合作完成某一大型翻译项目时;
2. 翻译效率可以提高 30~50%;
3. 翻译成本可以降低 15~30%;
4. 便于翻译项目的质量控制(quality assurance)管理;
5. 缓解截止时限与翻译质量之间的矛盾;
6. 缩短急需进行翻译的产品,如需要本地化的软件产品,的上市时间;
7. 在某些文本最后定稿前就可以开始着手翻译。

三、机助翻译的使用范围

计算机机器辅助翻译的原理决定了它的使用效率取决于原文的重复率。要充分发挥计算机机器辅助翻译的作用,原文必须具有下列特点:

1. 术语的一致性:文本内的术语总是具有一种释义,同样的概念和事件等总是用同样的术语表达;
2. 短语的一致性:同样的概念和动作等总是用同样的短语来表达;
3. 措词用句简明性:使用简单明了的表达方式,以减少歧义成分;
4. 文本篇幅较长,部分内容经常更新:提高文本重复率。

显而易见,计算机机器辅助翻译对使用周期长,部分内容经常更新的技术性文本的帮助特别大。对于采用多种修辞手法的文学作品,如诗歌、散文、小说,以及广告、报刊文章等的翻译并不实用。计算机机器辅助翻译并非解决所有翻译问题的万灵药,是否采用计算机机器辅助翻译要对原文文本的特点作具体分析。

四、语言覆盖面和电子文档的格式处理能力

计算机机器辅助翻译支持多语种之间的双向互译。例如,德国塔多思(Trados)公司的计算机机器辅助翻译产品可以支持多达62种语言(包括主流亚洲语言,如中文、日文及韩文)间的双向互译。另外,计算机机器辅助翻译产品支持大多数流行的文档格式,如DOC、RTF、HTML、SGML、XML、FrameMaker等。

五、投资和回报

需要说明的是国外的计算机机器辅助翻译的产品价格比较昂贵,有的高达上千美元,其投资的回报并非像有的厂商广告所宣称的那样会立竿见影。其中原因包括:(1)翻译记忆库和专业词典库的建立都需要一个积累过程。(2)除了上文已经提到的文本重复率的因素,还要考虑译文是否是电子文本。尽管现在扫描仪的识别率已经大有提高,但是仍然需要化大量的时间进行人工核对和校正。

六、TRADOA产品及其他

目前国外市场上主要的计算机机器辅助翻译工具产品有:TRADOS、IBM Translation Manager、Bj Vn SDL Sdk、STAR Transit、TSS Joust、SL8 TransPro和Eu-

rolang Optimizer等。不同的产品各有其优点和不足,性能和价格也有较大的差异。有的产品,如TRADOS,可以从公司网站下载免费的试用版本。

德国塔多思(Trados)公司的计算机机器辅助翻译产品为专业翻译工作提供了较为完整的翻译解决方案,是目前世界上最好的计算机辅助翻译软件之一,占有该领域70%以上的市场份额。目前TRADOS已经基本垄断了国际大型翻译和本地化公司,如Berlitz、Lionbridge等。此外,其用户还包括Microsoft、Cisco、Intel、HP、Compaq、Dell、Oracle、Yahoo、德国大众、西门子、联合国、欧盟、国际货币基金组织等全球四万多企业和国际组织。TRADOS公司的最新版本TRADOS 5主要含有以下组件:

1. WorkSpace(工作空间):TRADOS的核心组件和主界面。跨应用程序的集成能力使用户可以在浏览器风格的界面中集中控制所有TRADOS组件,对所有的翻译项目和 workflow 进行管理。具有项目经理(Project Manager)、技术专员(Specialist)、译员(Translator)三种操作模式。

2. Translator's Workbench(翻译者工作台):与MS Word 97/2000/XP无缝集成,支持62种语言之间的双向互译。用户可以在Word环境中进行交互翻译。Workbench利用后台强大的神经网络数据库,为用户储存并管理原文和译文,实现对译文的部分自动翻译,动态提示相关译文语句,自动保留原文格式,并与Tag Editor有很好的兼容性。

3. MultiTerm(术语管理):专业翻译中术语的一致性确保翻译质量的关键。MultiTerm与Workbench结合工作,可以在用户翻译的过程中动态提示已经定义好的术语译文,确保术语的一致性。

4. ExtraTerm(术语抽取):自动抽取翻译项目所需的潜在的术语,自动生成单词或双语术语表。在ExtraTerm界面中,用户可以方便地核对和编辑术语表,并且在必要时加上相关的句子。ExtraTerm的导出格式可以将术语表导入到MultiTerm的术语库中。

5. Win Align(匹配建库):帮助用户回收过去已经翻译好的资料,通过对原文和译文的比较和匹配,以语句为单位将原文和译文一一对应匹配,建立起翻译记忆库,供Workbench使用。

6. Tag Editor(标记格式编辑器):对网页(Web)格式文件(如HTML/ASP/JSP/SGML/XML/XSL)和桌面出版(DTP)格式文件(如Adobe FrameMaker/PageMaker/QuarkX Press/Interleaf/Quicksilver)提供支持。此外,Tag Editor还集成了Microsoft Word字典功能,可以为所有语言提供了更先进的拼写检查功能。

从长远来看,智能翻译的出路应该是机器翻译和计算机机器辅助翻译二者的集合。近年来有些机助翻译软件的厂商已经成功地在自己的产品中集成机器翻译的界面,使得翻译过程成为机器翻译和计算机机器辅助翻译的互补。

参考文献:

- [1] Esselink, Bert. *A Practical Guide to Localization*. John Benjamins Pub Co., 2000.