

doi:12.3969/j.issn.1672-0598.2013.02.021

TM 还是 MT?

——谈 CAT 软件设计中的认知理据

陈友勋

(重庆文理学院 外国语学院,重庆 永川 402160)

摘要:本文从比较分析机器翻译和翻译记忆的不同现状及其背后所隐藏的认知原理入手,指出 CAT 软件是人类在解决语义问题之前,利用计算机对语言转换之间的形式特征进行描写计算所取得的辉煌成果。其中,尽管机器翻译和翻译记对语义障碍采取了不同的态度,前者试图直接针对当前翻译本身进行分析处理,后者则借助过去积累的翻译经验对语言转换进行形式计算,但从认知策略上它们都体现了人类共同的认知经济倾向。所以我们在设计 CAT 软件时,应当创新途径,同时容纳机器翻译和翻译记忆,让二者能够优势互补、配合使用,从而取得更好的翻译辅助效果。

关键词:机器翻译;翻译记忆;计算机辅助翻译;认知经济;语义障碍;认知资源有限性

中图分类号:IP311 **文献标志码:**A **文章编号:**1672-0598(2013)02-0143-08

引言:MT 与 TM? 厚此薄彼的选择

真正从事职业翻译的人员,一般都对 MT (Machine Translation: 机器翻译)嗤之以鼻,而对 TM (Translation Memory: 翻译记忆)情有独钟。在他们看来,MT 完全是糊弄外语盲的花哨噱头,没有任何实用价值,用它翻译出来的文章全是错误,不忍卒读,去检查和修正这些错误所花的时间和精力,甚至远远超出译者自己独自翻译源文所需投入的时间和精力,因此绝大多数译者宁愿全靠自己人工翻译也不愿求助于机器翻译。但对于 TM,以及以 TM 为核心而建立起来的 CAT (computer aided translation: 计算机辅助翻译)软件,绝大多数译者都是非常器重,认为 TM 给翻译带来了真正意义上的技术革命,大大提高了译者翻译的效率和准确性,具有极为宝贵的实用价值。

其实无论 MT 还是 TM,都是人类在认知经济

原则驱动下,试图减轻翻译难度而进行的一种积极探索或有益尝试,其背后都体现了人类追求效益的认知取向。这些探索或尝试本身,不论成败,都是一种积极而有益的努力:因为成功的探索固然让我们享受到了它对翻译实践的促进效果,但失败的尝试至少也给我们揭示了提高翻译效率将要努力的方向。

一、MT: 通天塔的幻想

关于翻译,圣经中有一个著名的故事叫做 The Story of Babel (见《圣经·旧约》创世纪第十一章)。对于其中的 Babel,汉语中有两种译法,分别是“通天塔”和“巴别塔”。笔者认为前者是一种意译的结果,强调的是人们对翻译所寄予的一种厚望,即希望借翻译之功重修“通天”之塔——

* [收稿日期]2012-12-27

[作者简介]陈友勋(1975—),男,重庆永川人;英语硕士,重庆文理学院外国语学院,副教授。

完成传说中人类未竟之业(Come, let us build ourselves a city, with a tower that reaches to the heavens, so that we may make a name for ourselves and not be scattered over the face of the whole earth:让我们修建一座通天之塔,以此扬名立万并永不分离)。这也间接反映出人们在翻译实践中追求的理想境界是像故事源文所宣扬的那样“(if) as one people speaking the same language …, then nothing they plan to do will be impossible for them”,即通过翻译帮助人们达到像说同一种语言的效果,可以沟通无障碍,做任何事情都无往而不利。而在本文中我们要谈论的 MT,可以说就是人们在改造翻译实践时产生的这种“重建通天塔”似的努力。

机器翻译(Machine Translation:简称为 MT)是探索如何借助计算机程序将文字或语音从一种自然语言翻译成另外一种自然语言的技术。人们曾怀着美好的愿望,期待有一天所有的翻译问题都能交给计算机自动完成,从此人类可以从语言交流的障碍中解脱出来,更加自如地专注于自己的事业。但早在机器翻译问世之初,美国著名数理逻辑学家 Bar-Hillel 就指出,除非计算机可以自动根据语料习得语义,否则全自动高质量的机器翻译(Fully Automatic and High Quality MT)就是不可能的。而美国科学院语言自动处理咨询委员会在 1964 年曾专门调查过机器翻译的研究情况,并在报告中对机器翻译也采取了否定的态度,指出机器翻译研究遇到了难以克服的“语义障碍”(semantic barrier)问题。更有甚者,他们嘲笑说 MT 不是 machine translation 的缩写,而是指 mad translation(疯子的翻译)。在这方面,德国翻译学者 Snell-Hornby 的观点具有代表性:“毫无疑问,技术的产物,无论多么复杂,都不能和人脑的创造能力相媲美。”(Snell-Hornby, 1997:66)事实果然如此。尽管人类进入了二十一世纪,尽管计算机技术也突飞猛进,同时我们对语言的研究也日益深入,但机器翻译离真正的实用阶段还差得太远。

所以,尽管人类提出 TM 的初衷是良好的,是希望借此一劳永逸地解决翻译中的所有问题,从

而可以重返远古社会人类交流的“黄金时代”——(the whole world had one language and a common speech:人们使用同样的语言,说着同样的口音)。但是,正如圣经中修建通天塔的努力最终夭折一样,MT 在现实生活中也遭受了同样的命运,被证明只是一种不切实际的乌托邦,是翻译中的最终幻想(final fantasy)。

二、TM:巴别塔的辉煌

前文提到,汉语中对 Babel 第二种流行译法是“巴别塔”,这主要强调现实生活中人类语言千差万别的实际状态,而这也反映出我们对翻译的探索已经从理想的高空回归到现实的大地上,开始脚踏实地地探索翻译现象。于是在机器翻译(MT)裹足不前的时候,翻译记忆(TM)的概念应运而生,并且,正如下文即将论及的那样,这一转变在翻译实践中被证明是极为成功的。所以,如果 MT 代表的是重建“通天塔”的幻想,那么 TM 则代表了实践中人类利用“巴别塔”所取得的辉煌。

翻译记忆(Translation Memory, 翻译记忆,简称为 TM)是“译者运用计算机程序部分参与翻译过程的一种翻译策略。”(Shuttleworth & Cowie, 1997:98)它是通过计算机软件来实现的专业翻译解决方案。形式最简单的翻译记忆只是一个数据库,译者可以记录(通常是半自动化地)以前的翻译结果以便于将来再次使用以及进行方便有效的检索。换言之,翻译记忆的工作原理是计算机利用用户已有的原文和译文,建立起一个或多个翻译记忆库,在翻译过程中,电脑将自动搜索翻译记忆库中相同或相似的语料资源(如句子、段落等),给出参考译文,使用户避免无谓的重复劳动,只需专注于新内容的翻译。并且,对于给出的参考译文,译者可以完全照搬,也可以修改后使用,如果觉得不满意,还可以弃之不用。正是在这个意义上,翻译记忆实际起到了辅助翻译的作用,所以叫做“Computer-Aided Translation”(计算机辅助翻译:CAT)。

国际上开发翻译记忆(TM)软件的厂家有很多,比较著名的有 Trados, SDL X, Dejavu, Transit, WordFisher, Wordfast 等。国内的有雅信 CAT、通

译等,其中最为知名和广泛使用的是德国的塔多思(Trados)。

目前在职业翻译领域,CAT 软件的使用是极为广泛的,不少翻译公司甚至同时使用好几套不同的 CAT 软件,这样可以取长补短,配合使用,以此满足译者的不同要求,最大程度地提高翻译的效率。其实,这些软件尽管具体形式有所变化,但都是基于 TM 而设计开发的,TM 才是他们的核心概念。因此,限于篇幅,我们在本文中无暇论及这些 CAT 软件的表面差异,只是希望针对 MT 和 TM 做进一步的探讨,寻找其中的认知根源,加深我们对它们的认识,进而促进我们对 CAT 软件的改进和革新。

三、TM 和 MT 的共性:认知经济原则

黑格尔曾说,存在的都是合理的,而合理的都是自然的。于是人们对周围的事物往往因见惯不惊而疏于去探询其中的究竟。这种情况同样体现在我们对 MT 和 TM 的态度上。我们想当然地只是把 MT 和 TM 当成人类翻译实践中两种努力方向,很少有人去研究其背后是否隐藏着某种认知规律,更别说是具体探究这些规律是否存在相同之处。这种情况,被美国认知心理学家 Fodor (1983,107;2002,89)恰当地总结为“认知科学不存在的第一条法则”,即对于越是常见的认知现象,人们越是习以为常,鲜有研究。或者做个通俗的类比,常见现象背后隐藏的认知规律,就好比人们生活中的盐:作为基本的调味品,虽然家家必备、顿顿不缺,但平时很少有人对它加以关注,遑论研究。可这种现象若任其发展下去,显然也不利于我们对它们进行深入的探索和改进。就像人们一直不对生活用盐加以研究,也就不会促进食盐工业的发展,于是今天的超市柜台上也就不会出现低钠盐、调味盐、麻辣盐、胡椒盐、汤料盐、保健盐等五花八门的含盐调料,从而丰富饮食结构,提高生活质量。因此,本文将职业翻译中的这类“盐”现象——即翻译实践中的 MT 和 TM 概念,做一个类似的深入研究,以期找出其中的认知规律,提高我们对翻译认知本质的认识,并在此基础上对 CAT 软件的设计提供改进意见。

其实,在翻译中无论是应用 MT 还是 TM,从

最基本的认知层面上却是相通的,其中都体现了人类一以贯之的认知策略——省力原则(the Principle of Least Effort)(Zipf,1949)。在认知研究中,“省力原则”又被称为认知的经济原则(the economy principle),可以简单地概括为:以最小的代价换取最大的收益,或者从反面解释,人类总是倾向于在完成任务时采用偷懒的方法。法国语言学家马尔丁内(Martinet,1962)认为认知经济原则是语言“运转的基本原理”(Principe de fonctionnement),是指导人类行为的一条根本性原则。姜望琪(2005)则撰文指出,经济原则体现了人类行为的一种惰性趋向,却是“指导人类行为的一条根本性原则”。

例如,人类研究机器翻译,其目的是想一劳永逸地解决困扰我们的语言沟通问题。但由于遭遇到难以克服的语义障碍,MT 研究至今裹足不前,事实上成为了一个翻译研究中的通天塔幻想。这种情况,正如有的学者所概括的那样“MT 做了人类想做的事情,但没有做好。”

而 TM 借鉴了 MT 失败的教训,不再好高骛远,而是脚踏实地,主要是代替人类老老实实在地完成翻译中那些机械的重复和繁琐的查询工作。而做这方面的工作正好是人类的弱项、电脑的长项,因此人类在翻译实践中应用 TM 可谓是让人脑与电脑达到了取长补短、珠联璧合的效果。根据前面的介绍,我们知道翻译记忆就是充当译者的扩展记忆库,为他积累以前的翻译经验,让译者对同一个句子不必翻译两遍,从而达到节约时间、提高效率、同时保证译文前后一致的积极作用。这正如有的学者评价的那样“TM 做了人类不想做的工作,但是做得相当好”。

四、TM 胜出的原因:语义现象的瓶颈效应和认知资源的优化配置

(一)语义现象的瓶颈效应

到目前为止,在这场 MT 和 TM 的较量之中,TM 能够胜出的最主要原因可以归结为语义现象产生的瓶颈效应。

我们知道,目前无论是 MT 还是 TM,其计算根据都是人类对语义现象的外部形态进行描写、总结、并进而转换成电脑程序而实现的,因此翻译

研究中的语义问题构成了 MT 和 TM 发展过程中的一个瓶颈。

语言表达与语言意义之间是形式和内容的关系,从哲学上讲,形式是表现属性,内容才是变化的根本。从语言学的角度,二者相当于乔姆斯基提出的语言运用(linguistic performance)与语言能力(linguistic competence),或表层结构(surface structure)和深层结构(deep structure)之间的关系。在翻译过程中,如果要排除语义的影响而不影响表达结果的准确性,除非我们能够事先穷尽人类所有可能的语义-表达的配对形式,并把它们全部转化为计算机可以识别的程序语言,而这显然是不现实的想法。因为人类语言无论是其形式还是其意义本身,都是在无法预料的动态的因素(如语境)下综合作用而产生的,具有创造性是其鲜明的特征,即既可以用有限的形式表达无限的意义,也可以把同样的意义用无限的形式来表达。而人类是不能超越无限的,这就决定了语言的形式和意义之间绝不可能存在某种确然的一一对应关系。既然语言形式-意义之间的配对关系不存在确然性和有限性这两个基本的前提,那么翻译问题从逻辑上就不可能单单依靠其形式特征,即指望通过频率统计从而简化为机械的匹配过程。换言之,计算机要解决翻译问题,必须另辟蹊径。

乔姆斯基认为人类有一套天生的语言习得机制(LAD),可以靠有限的输入获得无限的语言语用能力。所以这个语言习得机制的核心就是人脑对语言现象的自我适应和学习。而目前电脑之所以不具备处理语义的基本能力,也就在于电脑不具备自我学习的能力,没有这方面的任何主动性和智能性。也就是说,在习得语言方面,电脑完全是被动的、静态的,无法处理新出现的语言现象,因此 MT 在发展的道路上被挡在了“此路不通”的牌子前,而 TM 正好是明智地避开对语义现象的直接处理而把电脑的长处诉诸对语言之间的转换特征进行形式计算,所以才发展出一片广阔的天空。所以,正是对翻译过程中所涉及的语义现象采取了不同的态度(面对或回避),才导致 MT 和 TM 今天不同的命运。

基于同样的原因,笔者也相信,除非将来有一

天,计算机可以要采用发展的、可以自我修正、更新和扩展的语言处理系统,即智能化语言程序,摆脱一切都只能靠外部输入的被动局面,否则我们还是不要指望 MT 能够有比今天更好的作为,而应当继续关注和发展 TM 的务实道路。

(二) 语言工具性与翻译题材的专业化

此外, TM 胜出的原因还在于职业翻译所涉及的材料具有单一性和专业性,具有相对固定的术语范围和表达模式。从语言学的角度分析,人类语言具有两种不同的表达功能:工具性和艺术性。语言的艺术性主要表现于文学作品之中,它尽量以新奇的表达和怪异的结构来凸显作者的个性,其追求的表达效果是“存异”,所以才会出现“一百个读者心目中有着一百个哈姆雷特”,而像《红楼梦》等文学名著也才会被电影导演们乐此不倦地反复翻拍。而语言的工具性恰恰相反,它的目的是“求同”,即尽量排除主观的干扰,消除语言表达中的歧义,强调语言指称对象的客观性和准确性,使不同的读者读同一篇文章能够获得基本相同的理解效果。出于这个目的,强调工具性的文体往往都有相对固定的表达术语和语言模式。这在科技和商务文体中表现得尤为明显。

而这就使得这类文体特别适合于用电脑进行数学意义上的统计分析。此外,电脑代替人脑完成这部分工作,解放人脑专门从事对艺术性语言的翻译,这样各尽其责,优势互补,又正好体现了认知经济原则。事实也的确如此。比如职业翻译所涉及的翻译资料往往数量巨大,但范围相对狭窄,主要集中于某个或某几个专业,如政治、经济、军事、航天、计算机、通讯等专业,这就必然带来翻译资料在不同程度的重复。据统计,在不同行业和部门,这种资料的重复率达到 20%~70%。这就意味着如果不使用 CAT 软件,译者至少有 20% 以上的工作是无谓的重复劳动。而翻译记忆技术就是从这里着手,首先致力于消除译者的重复劳动,从而提高工作效率。

职业翻译涉及的一个巨大挑战来源于专业术语。这些术语在商务、科技领域的资料中大量而频繁地出现,而且专业化的程度很高,有时太新的术语根本找不到相关的查询资料。而对术语的翻

译要求是“忠实”“地道”和“统一”(张新红、李明,2003:14)。所以,可以肯定地说,任何个人都无法独自胜任这样高标准的科技或商务翻译。但 CAT 软件中通过专门的工具处理 Glossary(如塔多思是用 Multiterm 来储存这些专业术语),并且还可以通过在线连接,分享别人或网络中储存的术语库,从而成功地解决了以往让人头疼不已的术语难题。

所以,通过使用翻译记忆,译者不必再囿于手工作坊似的传统翻译手段,而是利用认知规律并借助现代化的手段,大大地降低了职业翻译的难度,提高了翻译的效率,以较小的认知代价换取了较大的认知收益,体现了翻译活动中所遵循的经济原则。

(三) 认知资源总量有限性

如果说经济原则体现的是译者使用翻译记忆的一种基本的认知策略,属于翻译中的积极因素,那么译者认知资源总量的有限性则是翻译中的一种消极因素,从这个角度研究翻译记忆,则可以说使用翻译记忆是译者无奈之下的一种被动选择。

梁宁建(2003)在《认知心理学》中指出人类进行认知处理时都面临一个瓶颈障碍——认知资源总量的有限性。它是个人认知负荷的最大值,人们可以胜任在其范围内的认知操作,并且这种认知操作的强度是和它距这个最大值之间的长度成正比关系的:即越是远离这个最大值,越是感觉轻松自在、游刃有余,但越是逼近这个最大值,越是让人力不从心,感觉紧张和疲惫。此外,值得指出的是,个人认知负荷的最大值在短时期内是一个相对恒定的数值,基本上可以认为是稳定不变的。因此认知资源总量是一个全局性的常量,其对人类认知活动产生的制约作用无时不有、无处不在。出于以上原因,人类在完成认知任务时,如何制定活动策略,分配有限的认知资源、实现认知资源的最优化配置,就显得极为重要,事实上这也一直是认知研究中的一大热点。如法国学者丹尼尔·吉尔在《口笔译训练的基本概念与模式》(Basic Concepts and Models for Interpreter and Translator Training)曾专门开辟章节研究翻译过程中的认知资源分配方案,并进而提出了著名的

“精力分配模式”(Effort Model),指导译者在进行翻译转换时,通过动态的调整资源分配方式,从而带来最佳的翻译体验,实现最好的翻译效果。

当然,我们知道,进行这样的资源配置、优化组合,最重要的是知人善任,即让最适合的对象去完成他最擅长的部分。而 CAT 软件中使用的 TM 概念,通过分解笔译任务,把其中繁琐机械的重复翻译部分和术语提取工作交给电脑完成,让译者集中精力处理新出现的待译原文,可以说是给翻译带来了一种出色的“双赢”局面:既分解了笔译任务,有效地降低了翻译难度,同时也让电脑和人脑各得其所,各尽其能。所以,在翻译中使用 TM 其实也是一种“精力分配模式”,同样对译者有限的认知资源起到了优化重组的作用。

(四) 人脑和电脑的差异

从认知科学的角度,我们知道人类认知的一大特点是“喜新厌旧”,即人类喜欢接触从未见过的事情或任务,并善于从观察这些新奇事物中分析、提炼其特征规律并进而找出其对应的解决方案。这是一种“知难而上”的认知倾向,因为新奇事物所带来的新鲜感和挑战性能够让人兴致盎然而乐于参与,同时人们也能从解决这些新奇事物的过程中获得一种前所未有的满足感和成就感。换言之,人脑擅长的是面向新鲜事物的创造性思维,而不习惯于对以往经验的简单重复。失眠的人都有这样的生活经验,即通过数数字来让自己早点进入梦乡,因为这种简单乏味的机械重复是最缺乏挑战性和刺激感的任务,最容易让人丧失兴趣而晕晕欲睡。

但机械重复恰好就是电脑的长项。因为电脑按其设计原理就是面向过去经验的机械操作。电脑软件的一切操作其实只不过是按事先设定好的程序步骤去提取、比较数据并根据简单的二进制判断输出显示结果(即把电子管物理性质的开关操作转换成二进制中的 0 和 1 进行数值计算)。但正如人类无法准确地预见未来的情况一样,再智能的电脑程序都无法做到事先就把未来所有可能发生的结果都考虑进去并记录在案,因此再先进的电脑目前都无法独立处理从未见过的语言现象(在翻译中主要就是指 semantic barrier,语义障

碍),这也是MT无法推广应用的主要原因。但我们人类也无法否认电脑在处理数据方面的巨大优越性:首先是电脑不存在人脑那样的遗忘机制,数据一旦存储进去,在理论上就可以永久存在。其次是电脑能高效快速地完成对海量数据的提取、分析和转换,并且只要电源充足,电脑运行起来就永不疲惫。因此,一切翻译现象,只要能够转换成电脑可以识别的数据格式,就可以在电脑中得到高效的运用。换言之,翻译记忆之所以可以在职业翻译中大行其道,就是因为电脑擅长于对以往翻译经验的复现。

从以上比较中,我们可以得出这样的结论,即翻译记忆正好是取电脑之长补人脑之短,二者可以扬长避短,各尽其能。因此译者使用CAT软件进行翻译是获得了一种相得益彰的互补搭配,从而保证了职业翻译的高效和准确。如果我们从“人”(译者)的角度来说,就认为翻译记忆在认知方面的一大贡献是用电脑解放了人脑,让译者从此可以集中精力,专注于翻译新出现的句子。

五、TM和MT的融合以及CAT软件的改进

(一)TM和MT的无缝结合

如果我们具体分析TM在CAT软件中的呈现方式,可以看出它对译者的辅助作用主要是依靠统计其出现概率的方式实现的。现在CAT用户一般都可以对软件进行自定义设置,规定计算机在检索翻译记忆库时,如果当前待译源文与记忆库中的TM片段在结构上的相似度达到或超过某一个阈值(通常为一个百分比数字)时,就在翻译编辑界面给出相应提示。可见,CAT软件中这种提取TM的方法其实和统计机器翻译(Statistical Machine Translation,SMT)的基本思想如出一辙,即认为如果翻译记忆库中存在和源文片段类似的表达结构,那么从源文到译文的转换只是一个概率问题。换言之,可能翻译记忆库中同时存在好几个TM片段,都和源文片段具有对应关系,但每对备选的TM片段在实际翻译中出现的频率肯定不一样,理论上出现频率大的TM片段应当是接受程度最高的表达,因此其中的翻译片段自然也应当是最佳的候选译文,于是提取TM的任务就

转变成了搜索翻译记忆库中对应概率最大的TM片段。冯志伟(2007)曾从语言哲学的高度对机器翻译进行了总结,认为统计机器翻译(SMT)是以大规模语料库的分析为基础,体现了科学研究中的“经验主义”,那我们也可以说CAT软件中对TM的实际应用其实也是这样的一种“经验主义”——是一种基于过去翻译经验的翻译实践。

但正如前文分析的那样,基于过去翻译经验的CAT软件,是响应用户当前选择的源文而提取以前类似的翻译例子,属于一种被动的再现,其最大的弊端就是缺乏创造性,不能应对从未遇见的源文表达,因此无法直接针对当前源文本身进行操作,在职业翻译中只能处于从属地位,起到一个辅助作用。

不过如今随着网络技术的进一步发展,许多主流的CAT软件,如SDL Trados 2009,已经意识到CAT软件中存在的这一弊端,于是在设计中引进了自动翻译机制,即当前选择的源文片段在TM中不存在符合设置要求的对应例子时,便自动联网登录在线翻译引擎(塔多思是以language weaver为默认翻译引擎,当然Google在线翻译也深受很多CAT软件的青睐)并把机器翻译的译文反馈回翻译编辑界面供译者参考。整个过程完全是在后台自动进行,不会干扰译者的思维流程,并且由于职业翻译中大多数内容属于格式和用语都比较固定、规范的科技或商务资料,所以这样获得的机器翻译译文也的确具有一定的参考价值,甚至不排除在个别句式简单、术语规范的情况下译者可以直接引用机器翻译的译文。

当然,值得指出的是,此处的机器翻译应当是基于规则的机器翻译RBMT(Rule-based Machine Translation),即将人类翻译时所采用的知识和规律总结、抽象出来,以特定的形式存入计算机。然后在翻译过程中,计算机结合源语言的输入,选择相应的知识进行推理或转换,生成目标语言的翻译结果作为输出。和先前的统计机器翻译的原理相反,冯志伟认为基于规则的机器翻译是以转换成语言学为基础的方法,体现的不是基于先前翻译语料的“经验主义”,而是科学演绎为主的“理性主义”。或者可以这样说,基于规则的机器

翻译直接面向当前译文本身进行分析和翻译,在一定程度上克服了 TM 单纯面向过去翻译经验的弊端,因此可以说 TM(Translation Memory)和 MT(Machine Translation)的结合是 CAT 软件未来发展的方向。

(二) CAT 软件的改进意见

1. 切实加强目前的成功机制,切实加强和发挥 TM 在形式计算方面的功能

以前的 CAT 软件主要是靠译者个人积累的翻译资料,进行本地化存储。这些软件除了自动把译者在软件中进行的翻译保存为 TM 资料之外,也提供相应的扩展功能可以把以前的翻译资料转换为 TM 文件(如 Trados 2007 中的 WinAlign)。后来随着翻译业务的扩展,很多翻译任务必须依靠多人协作才能尽快完成,于是又出现了通过 CAT 软件在特定成员内部共享 TM 资源,这从量和质两方面都大大扩展了 TM 的使用价值。但这仍然只是局限于少数人员之间,没有充分发挥 TM 概念的完整意义。而当前海量网络资源的免费共享才是 TM 发展的终极之道,才可以从真正意义上实现共享全人类的翻译经验。

当前的 CAT 软件已经体现了 TM 的发展趋势是从个人到集体、从本地到网络,逐渐包括尽可能多的网络资源(如在线词典和在线句库等)。国内的“雪人”翻译和 Transdog(翻译狗)都直接内置了好几个著名的翻译网站,其他的软件也可以通过自定义功能扩展使用这些网络资源。以英汉翻译为例,Cnki 翻译助手、中国译典、金山词霸、海词在线、句酷网等都是国内比较著名的在线网络,提供了优秀的词级、句级的 TM 资源。

2. 可以引进“预翻译”的概念

“预翻译”除了指电脑自动把 TM 库中和原文匹配程度达到或超过某一阈值的进行自动翻译之外(如塔多思的翻译工作台的“100% 匹配”和“翻译至下一匹配处”的功能),还可以指对术语的自动处理:即电脑直接针对原文材料进行处理,自动对待译原文进行扫描统计。对其中出现频率达到一定阈值的词汇自动认定为专业词汇,因此需要固定译法,从而达到译文前后一致的效果。如果在当前的本地术语库中找不到相应的译文,则可

以联网搜索相应的在线词典,并自动下载译文供译者参考。目前 SDL 2009 Mutiterm Extract 可以在以前的 TM 文件中提取术语表,而 Transdog 则可以直接联网在线词库并提取当前源文的术语表。

3. 继续优化 TM 和 MT 在翻译过程中的配合使用

除了我们设定 TM 的优先级别高于 MT 之外,即搜索不到相应的 TM 就自动联网进行机器翻译,同时整个辅助功能应当全在后台运行,并且是以一种自然的方式把搜索结果呈现给译者作为翻译的参考。

此外, TM 和 MT 还应可以根据源文的类型自动选择最具参考价值的网络资源进行搜索和翻译,这样既提高响应速度,同时也提高了搜索结果的准确性。其中重要的一点是,整个搜索过程不影响到前台工作的译者。这除了是为了最大程度地减轻译者工作强度之外,还主要是体现了一种流畅性思维。翻译过程中流畅性思维极为重要。这种流畅,从认知上是指译者在翻译过程中的工作思维不受其他不相干因素的干扰,能够始终专注于翻译工作本身,并且以一定的节奏有条不紊地进行下去。在这种流畅思维的作用下,翻译工作才可能如行云流水一样高效而准确,同时也给译者自身带来舒适的翻译体验并进而提高他对翻译工作的成就感。这就要求我们的 CAT 软件应当体现一种人性化的设计思想,即真正做到“以人为本”:使用软件本身不是干扰,而是要符合人们的翻译习惯、提高翻译的舒适度、准确度和速度。设计优秀的 CAT 软件在使用过程中,应当使译者忘记自己是在使用软件翻译,产生的感觉仿佛就是完全靠自己在翻译,达到了所谓物我两忘、浑然一体的最高境界。

其次,我们在设计 CAT 软件时应当扩展新的途径。这方面我们可以通过划分翻译单位来充分发挥 MT 和 TM 的作用。运用翻译单位最基本的操作就是把源文划分为一个个更小的表达结构,这其实也起到了认知经济的效果,即借助翻译单位我们可以在翻译实践中化繁为简、化长为短、化难为易。如果从 TM 配对上来说,越小的源文单

位找出匹配译文的几率也更大。因为大的语言结构都是由更小的语言结构组合起来的,所以越小的语言结构使用频率也就越高,越是容易找到匹配的对象。从 MT 实用性来说,语义的复杂性主要表现在语言组合结构之间的逻辑关系上,而逻辑关系主要是一个纯粹的语义判断问题,既然目前的计算机是无法直接针对语义现象进行运算的(所以 MT 翻译长句总是错误百出),所以运用翻译单位也就降低了翻译过程中进行语义判断的难度(该部分工作已经被人类译者在划分翻译单位时完成了),从而可以间接提高了 MT 翻译的精度,提升了其实用性。

从以上论述可以看出,在翻译过程中,能否对源文进行结构分析显得相当重要。虽然结构分析主要是一个语义问题,目前只能靠人工完成,但必要的形式辅助也可以大大降低其中的操作难度、提高分析的效率。特别是在科技英语中,出现长句的几率相当大,包含几十或甚至上百个单词的句子屡见不鲜,有时甚至一个句子就是一个段落。实践之中译者往往是借助纸笔在源文稿件上进行勾画分析,但目前采用电子稿件的翻译形式也极为普遍,这就决定了传统的以纸笔方式进行结构分析的方式已经落伍过时,我们应当与时俱进,采取更科学的方式。如果我们的 CAT 软件能够在这方面设计并提供相应的功能,辅助译者对源文

方便而直观地进行结构分析,则无疑可以提高软件的实用性和亲和力。而目前的软件都没有提供这种功能,不能不说让人大为遗憾。

[参考文献]

- [1] 冯志伟. 自然语言处理中的哲学问题[J]. 心智与计算, 2007(03): 333-353.
- [2] 姜望琪. Zipf 与省力原则[J]. 同济大学学报(社会科学版), 2005(1): 87-95.
- [3] 梁宁建. 当代认知心理学[M]. 上海: 上海教育出版社, 2003.
- [4] 张新红, 李明. 商务英语翻译[M]. 北京: 高等教育出版社, 2003.
- [5] Fodor, J. A. (1983), The modularity of mind. MIT Press, Cambridge, MA.
- [6] J. A. 福多. 心理模块性[M]. 上海: 华东师范大学出版社, 2002.
- [7] Martinet, A. A Functional View of Language[M]. Oxford: Oxford University Press, 1962.
- [8] Shuttleworth, Mark & Cowie, Moria. Dictionary of Translation Studies[M]. St. Jerome publishing, 1997.
- [9] Zipf, G K. Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology[M]. Cambridge, Mass.: Addison Wesley Press, INC., 1949.

(责任编辑:朱德东)

TM or MT?

—On the Cognitive Principles in CAT Software Design

CHEN You-xun

(School of Foreign Languages, Chongqing Art and Science College, Chongqing Yongchuan 402160, China)

Abstract: By comparing and analyzing the status quo of and the hidden cognitive principles in Machine Translation and Translation Memory, this paper concludes that CAT software is a great achievement by using a computer to describe formal features in inter-lingual transmission as semantic barriers remain unsettled in translating practice. Although Machine Translation and Translation Memory take different attitudes towards semantic barriers, the former tries to directly analyze and process current translation itself, the latter conducts formal computation on inter-lingual transmission by using previous cumulated experience, however, they both embody common human cognitive economic trend on the basis of cognitive strategy, therefore, when CAT software is designed, the method should be innovated and Machine Translation and Translation Memory should be simultaneously integrated to let the two support each other so that better translation-aided effect can be obtained.

Key words: Machine Translation; Translation Memory; computer aided translation; cognitive economy; semantic barrier; limited cognitive resources