

基于皮尔斯符号学的人工智能翻译探析*

姚婷婷¹ 蒋蓉²

(1. 南京师范大学外国语学院; 2. 皖江工学院基础教学部)

摘要: 人工智能翻译的迅猛发展不但使传统人工翻译备受挑战,也进而极大地影响了整个语言服务人才教育产业。随着自然科学技术和哲学社会科学两个领域在方法论层面不同程度的融合,跨学科的研究思路和研究方法在这一过程中展现出无与伦比的优势。本文以皮尔斯符号学为理论框架,探查和分析人工智能翻译的发展,认为皮尔斯符号学以其缜密的逻辑推论和兼具理性与感性思维的意义理论,可以为人工智能翻译的完善提供思路和方法论指导。

关键词: 人工智能翻译;皮尔斯符号学;符号先行

0 引言

语言服务的迅猛发展对不同语种间的翻译效率和效应提出了更高的要求,人工智能辅助翻译便在此形势下迎来了革新,新技术新设备广泛应用于现今的语言服务产业中。近年来,人工智能翻译一直是语言服务领域和翻译研究领域讨论和研发的重要课题。世界翻译教育联盟翻译技术教育研究会多次以“AI时代翻译与研究”为主题召开国际研讨会,重点讨论人工智能翻译以及人工智能翻译对语言服务人才教育的影响等。由此可见,从图形识别到语音识别,从词条翻译到人工智能同声传译,人工智能翻译器似乎“越俎代庖”且展现出超越我们人类的“语言天赋”,人工智能已经在某种程度上扮演了“好译员”。分析传统人工翻译与人工智能翻译在语言转换层面的特征不仅仅涉及人类思维运作的生物机制和人工智能机器物理运作机制,二者还都在很大程度上关涉人类对其自身思维理解的哲学和符号学面相。这一领域随着科学技术和人文两大学科不同程度地融合,将受到越来越多地跨学科关注。皮尔斯早在一个多世纪前就去世了,他所处的年代还不可能有辅助翻译器,更谈不上人工智能的翻译。然而,皮尔斯的符号学思想具有超越时代的意义,在他的理论中有关符际翻译的固有特性以及符号逻辑的论述,可以为人工智能翻译提供研究思路和方法论的指导。

1 人工智能翻译:从统计机器翻译到神经网络机器翻译

当人工智能发展到今时今日,尤其是计算机硬件设备的革新,人们将感知器和传感器应用于加强计算机的计算能力,使得机器借助生命科学和电生物学的方法模拟人类神经元的传导机制,逐渐学会了思考和学习,机器翻译进入到认知智能阶段(张爱玲,杨子靖,刘晨璇等 2018)。目前我们所熟知的机器翻译软件如谷歌翻译(Google Translate)、微软翻译(Microsoft Translator)以及亚洲在线(Asia Online)等均采用了统计机器翻译(Statistical Machine Translation,以下简称SMT)模式。SMT是一种基于规则和知识的机器翻译,它通过对比分析大量平行语料库,区分性训练计算机以获得参数模型,解码预估的参数和给定的优化目标,获得待翻译语句的最佳翻译结果,同时构建统计翻译模型,在此模型基础上调序和查阅进行翻译输出。SMT系统依赖于所谓的语言模型(language models),或称目标语的单语语言模型(monolingual models),因此它首先需要解决一个单词在语言中的序列问题,而语言模式的训练须建立在一个目标语的平行语料库基础上,或是更大的目标语单语语料库的基础上(Kenny and Doherty 2014:278)。在统计机器翻译时代,语料库的丰富程度很大程度上决定了机器翻译的质量,然而SMT基于规则(rule-based)和知识(knowledge-based)的翻译方式对源语的处理往往囫圇吞枣,当源语语句意群较多较长时,其调序效果就会大打折扣,翻

* 基金项目:本文系江苏省高校哲学社会科学基金一般项目“钱钟书小说的英译和接受研究”(2023SJYB0255)阶段性成果,安徽省高等学校省级质量工程教学研究重点项目“应用型本科院校土木专业英语ESP课程设置与教学模式研究”(2020jyxm2067)阶段性成果。

译生硬,语言不通畅。

如果机器翻译发展的第一阶段是基于人类语言规则以尽量模仿人类语言规则为主,其第二阶段就是以人类语言实践为对象的经验学习,而第三阶段,机器通过受训,可以对人类语言中的元语法(n-grams)要素进行分析和重组,再进行类似人类神经网络构成方式的整合得到最为可能的翻译对应,即为神经网络机器翻译(Neural Machine Translation,以下简称NMT)。因此,在统计机器翻译的基础上,机器利用可实现的系统模拟人类脑神经的细胞结构和功能,利用源语言端和目标语言端的上下文信息构建神经网络联合模型。它将网络中一个序列映射到另一个序列的神经网络,输出为一个可以变长的或变短的序列,这有效改进了SMT中语序调整的为题,NMT通过训练的方式引入了全局注意力(global attention)和局部注意力(local attention)机制(Luong et al., EMNLP 2015),增强较长语句调序能力并提升词语表征。基于深度学习模型的NMT极大程度地提高了机器翻译的准确率,2016年谷歌正式推出了多语种神经机器翻译系统“GNMT”,其翻译错误率与谷歌之前的基于短语的SMT相比,翻译错误率下降了六成(Wu et al. 2016:20)。紧接着,2017年来自德国的DeepL神经网络翻译器一经问世就以优秀的算法和更精准的翻译在语言服务市场脱颖而出。在国内,腾讯“AI Lab”融合多项前沿技术,将神经网络机器翻译、统计机器翻译、输入法、语义理解、数据挖掘等有效结合,于2018年11月发布了旗下针对英汉互翻的交互式辅助翻译工具“Transmart”。腾讯的人工智能辅助翻译代表了我国目前NMT领域的较高水平,它基于自主实现的约束翻译解码技术,优化解码速度使之适应互联网环境中的实时交互要求,并通过预测用户动作,对原文建立换成来加快翻译速度。通过对比一下译文,我们可以得知人工智能翻译目前所能达到的翻译水平以及面临的挑战,以下两个例子分别是两款神经网络翻译软件对实用类文本和文学类文本的翻译结果:

例1. 英文: The absence of natural predators for the weed, coupled with the perfect temperature conditions of the lake, facilitated rapid spread of the plant. It thrived by gorging on open space, cutting off fishing routes and creating a conducive environment for disease-carrying mosquitoes. (中国日报网 *China Daily*, 2024-01-02)

“DeepL”翻译: 这种杂草没有天敌,再加上湖中适宜的温度条件,使得这种植物迅速蔓延。它在空地上大肆生长,切断了捕鱼的路线,为携带疾病的蚊子创造了有利的环境。

腾讯“Transmart”: 由于没有天敌,再加上湖泊完美的温度条件,促进了这种植物的迅速蔓延。它的繁荣是因为大量占用空地,切断了捕鱼路线,为携带疾病的蚊子创造了一个有利的环境。

例2. 中文: 清明时节雨纷纷,路上行人欲断魂。

“DeepL”翻译: The rain pours down during the Qingming Festival, and the pedestrians on the road want to break their souls.

腾讯“Transmart”: Qingming Festival rain one after another, pedestrians on the road to the soul.

作为两款采用神经网络技术的翻译软件,“DeepL”翻译和腾讯“Transmart”在翻译实践中基本可以达到实时翻译且译文质量优秀,不但基本解决了长句语序调序问题,而且对原文语义的也做到了还原。腾讯“Transmart”作为专注于英汉和汉英双语翻译的本土软件,对汉语行文习惯的把握更具优势,比如例1中“由于”表明原因,将主语置于结果状语从句中,避免使用“使字句”,译文更加凝练,更符合汉语日常表达习惯。而在文学文本翻译中(见例2),“DeepL”和腾讯的表现均差强人意,前者基本采取完全直译的方式,抛弃了诗歌中特有的格律和修辞;而腾讯“Transmart”的译本虽更为简洁,也试图以英文格律对应中文的平仄和押韵,但是句法结构和词组搭配上却不大符合英文的行文,且后半句“欲断魂”形容伤感极深,好像灵魂与身体要分开一样,被译为“to the soul”,在语义上与原诗有所出入。文学类文本和实用类文本在表达方式和使用场景方面有着巨大的差异。相较于实用类文本清晰、严密地阐述,文学类文本多运用形象思维,直觉思维、结合想象、联想等,同时搭载各种修辞手法进行表述,因此,文学类文本的翻译是对人工智能翻译器运作机制的极大挑战。随着人工智能的急速发展,它究竟是否可以完全模拟出人类独有的思维模式和情感模式呢?

2 人类思维与机器语言:符号先行

许多学者认为人类能够翻译,机器只是看起来会翻译,它们做的是“伪翻译”。在人工翻译中,人类智力和语言的运用表现出译者和文化社群之间的相互影响和动态交流,而在机器翻译中,机器将一个翻译任务拆解为一系列可以执行的二进制逻辑任务,并且根据处理器的运算能力在极短的时间内进行数以万次的计算。一个语言符号的解释项内容通常包含了两层意义,主要是由语言的概念结构所支撑的,潜在的,系统的意义,以及在一个特殊文本中被表述地且具有一定效果的上下文意义。机器翻译一般只能译出第一层意义,而忽略语言符号的上下文意义,也就是语用层面的意义。

为了使机器更高效地执行施事者下达的指令,人

类有意识地发明了一种作为人类和机器沟通的且抽象的中间语言或中介语(intermediate language),从而将机器翻译分成两个步骤,即将源语文本译为中介语,再将中介语译为目标语文本。中介语显然在机器翻译过程充当了媒介作用,但是它与作为人类交际媒介的人类语言又有着本质的不同。人类语言的特殊性在于语言学家强调的被称之为“句法”(syntax)的那种东西,即语言单个要素的排列和运作规则。句法虽然一定程度上规定了语言要素的排列,但是语言能够以数量无限的方式对数量有限的零件进行组装(佩特丽莉 2014:15)。而机器对语言的再加工必须严格遵循句法,它接受的指令使得自身完全无法脱离句法的范畴,哪怕是一丝丝的偏离,因此,机器翻译现阶段对源语文本的加工在于用有限的零件组装成唯一限定的单词、片语、句子和篇章,而几乎不考虑源语句法层面以外的要素,如语言系统构造、语用、语境、语域等。

如果机器翻译也存在着自有的思维模式和翻译策略,那它与人类翻译思维的本质是有巨大差别的。皮尔斯把思维看作是一种符号,并提出了“思维符号”(thought-sign)的概念,他说,“当我们思考时,就在思考的那个时刻,我们表现为一个符号……当我们向意识表达自己的感觉、印象、概念,或其他符号再现体时,这个过程即为符号”(1935:283)。思维符号同样具有连续性,也受到符号第一性的制约,符号的第一性包含了符号的品质,品质为符号提供解释项信息,使得思维符号具备可推导性和连续性。因此,人工翻译的思维模式表明了用语言符号对思维符号进行解释的过程,在这一过程中语言符号充当了中介语,它将源语作者思维、译者思维以语言符号的形式按照语言的内在逻辑性和语言符号的解释项信息,用另一套符码表述出来,成为一系列具有思维上连续关系的文本。

思维的符号化使得建立在现象观察基础上的思维,具有逻辑性和可解释性。在人工翻译过程中,思维符号大多数情况下会以外部语言或言语为对象,那么它本身就是一个思维之外连接的符号,必然会开始遵循社群交流中符号体系的结构原则,比如,语言的翻译会按照句法展开,而在思维内部,思维符号会根据归纳、演绎并以联想的方式无限延伸。即便一个目标语文本只是以某一种确定的表达方式呈现给读者,但是在人工翻译的过程中,译者的翻译思维早已经历了无数个符号意指过程,无数个思维博弈过程,以及无数个意义抉择过程。

机器翻译的思维构造远远比人工翻译要简单得多。机器翻译中一个极大的障碍是机器,即电脑,无法像人类译者那样理解源语文本,机器借助于人类指令运转,自身无法获得语言文字投射于译者心灵中的那种感觉。面对语言文化中一些新奇的、异域的、复杂的

观念,一些在读者心灵中激起波澜的想法在机器看来不过是一连串符码,毫无触动。在语言文化中理解一个言语表达,通常需要区别三个方面的含义,句法、词法和语用(Enkvist 1987:337-350)。即便在当前最先进的人工智能翻译中,机器也无法完全理解语用层面的含义,需要借助更多的数据支撑。比如,

他给她帮了个大忙,所以她想意思意思。

腾讯“Transmart”:He did her a big favor, so she wanted to say something.

人工智能无法准确判断一些多义词,尤其是文化负载词的特定使用场景,因此它未能理解“意思”在源语句子中意为“给予相应报酬”,而是仅仅按句法和词法的规则按字面意思去对应。如果我们把源语文本看作是一个命题,要使翻译的结果符合期望,并非只是使命题符号符合它的对象,而更是最终被再现的对象相符于命题所展现出来的有效推断,相符于命题之推断结果中所存在的方法,或相符于与该命题相关的那些论证,并由此汇聚成某种结果(皮尔斯 2014:214)。

对于人工智能翻译而言,将现有知识“形式化”不可避免。科学家们不辞辛劳地改进机器的运算速度,扩大语料库,使得人工智能具备某些“准人类行为”,比如自主学习能力。但是知识的隐性本质在于其可预测,一些不言而喻的知识如果对人工翻译轻而易举,那么对于人工智能翻译而言则需要另外的程序加以备注。毕竟人工智能翻译过程是纯粹机器语言符号对源语作者思维符号的解读,在皮尔斯的符号三元关系中,可被看作用符号对象来解释符号自身。因此人工智能翻译多大程度上能解读源语文本的意义,在于语言符号解释项信息的确定性程度,正如句法和词法相较于语用而言,其规则更为明确。

无论是在人工翻译过程,还是在人工智能翻译过程,都包含了一个共同的事件,即翻译施事者,译者或人工智能,对源语文本语言符号按照各自的经验知识进行识别和阐释,只不过各自的翻译工具大相径庭,译者使用人类大脑,机器需要依靠中央处理器的运算和中介语。然而,需要特别提出的是,翻译中强调的“译后审查”(post-editing)是机器翻译永远无法做到的。机器却相反,它基于外界环境变化的自动反馈,其翻译模式和翻译经验一旦固定,除非人工干预否则不可能发生改变。从皮尔斯的符号学观来看,机器思维逊于人类的最重要一点在于机器无法将某种特定甚至动态可变的复杂目的结合经验观测来指导推理(陈曦 2018:109)。皮尔斯符号学重视人类这一特殊的能力,因此他构建批判逻辑学时,提出首先应关注“合理推断”(valid inference),即“优良的推理”(good reasoning)(皮尔斯 2014:215)。

3 皮尔斯符号学对人工智能翻译的启示

从人类工业革命至今日,尽管互联网信息和人工智能领域的技术变革日新月异,机器取代人类的预言不绝于耳。然而,机器即便可以阅读人类思维符号,也无法有效地模拟人类思维符号的意指和再现过程。无论是统计机器翻译还是神经网络机器翻译,在解读人类语言符号时,往往都能够在极其短的时间内对该语言符号包含的解释项信息中句法层面的内容进行高效翻译,也就是在目标语系统中进行逐字逐词的语言对应,而在语用层面,由于其几乎无法考虑该语言符号所在文化社群的影响因素,因此人工智能翻译无法进行精确地译后审查(post-editing),这就需要人工译员的干涉。许多非专业译员时常会盲目服从机器翻译的结果,摒弃译后阶段的人工校准,他们把经人工智能机器翻译的译文奉为权威,从而忽略人类语言知识获取的理性和经验都是建立在人类思维符号活动的基础上。正如皮尔斯所言,所谓权威,跟所有理性一样,都不是绝对无误的(2016:46)。而人类理性对于机器思维的优势在于“推论”,“从对已知之物的考察,发现我们尚未知的另一事物”(皮尔斯 2016:186)。在翻译实践中,尤其是在精细化程度相对较高的专业领域或对语言逻辑要求较高的领域,比如,法律、译学、建筑等,需要提倡机器辅助人工翻译(Machine-aided human translation),或人工辅助机器翻译(human-aided machine translation),并强调译后审查的重要性。

对于人工译员而言,人工智能翻译的介入已经是不可逆的趋势,也将会在未来为更多地非专业翻译人士提供高质量的服务。在机器辅助人工翻译或人工辅助机器翻译的过程中,任何一个具备基础设备操作知识的普通人人都可以胜任超出其语言能力和翻译水平的翻译工作。即便机器取代了一部分人工译员的工作,但是大多数机器所进行的翻译仍然维持在语义明确,语篇术语重复性高,认知难度较低的文字符号处理及符号替换阶段,人工智能翻译暂时无法掌握应对复杂不确定世界的符号思维操控能力。对于人工智能完全取代人类译者的断言,在一定层面上过分夸大了机器的用处,也低估了人类思维符号活动的创造性。

在皮尔斯的符号学中,人类思维活动不仅仅局限于人类自身身体内,思维不再被描述为身体的感觉或行为,思维可以体验外界实在,体现了一种符合实用主义原则的三元关系。皮尔斯将人类思维置身于符号意指过程,思维不是无边界的,其意义也必然不是任意的,需要如解释此思维符号的符号那样符合逻辑。翻译的思想也和其实践一样,体现了一种思维符号的三元关系,二者既由解释它们之间关系的符号相连接,也

与外界社群环境相连接。如果人工智能确实具备“类人类”思维,那也只不过是人类思维的一部分,人工智能享有自由却也饱受限制,它的自由越不过人类思维,限制也较人类思维更甚。此外,符号伦理在人类思维活动中占据制高点。皮尔斯的符号学关注并强调人类的思维能力的重要性,与此同时,它以人类自我控制为着力点着眼点,为人类的道德属性预留空间。皮尔斯的批判逻辑为符号伦理的具象化提供了方法和途径,使得符号的翻译在科学精神驱动其向外延伸和人文情感羁绊其向内聚合之间达到完美的平衡。这种和谐的均衡之态也将会成为未来人工智能翻译的准则之一。而机器对于人类思维及情感层面的阅读和模拟将会随着机器与人类之间多维度的交流交往,而向更智慧化更人性化的方向发展,届时会赋予语言符号翻译更蓬勃的生机和更广阔的前景。

参考文献

- [1] Enkvist, N. E. More about Text Strategies [M]. Lörcher, W., & R. Schulze. Perspectives on Language in Performance. Tübingen: Niemeyer, 1987.
- [2] Kenny, D. & S. Doherty. Statistical machine translation in the translation curriculum: Overcoming obstacles and empowering translators [J]. The Interpreter and Translator Trainer, 2014(2): 276-194.
- [3] Luong, T., Pham, H., & C. D. Manning. Effective Approaches to Attention-based Neural Machine Translation [C]. In Conference on Empirical Methods in Natural Language Processing (EMNLP), 2015: 1412-1421.
- [4] Peirce, C. S. Collected Papers of Charles Sanders Peirce, Volumes V and VI: Pragmatism and Pragmaticism and Scientific Metaphysics [M]. Hartshorne, C. & P. Weiss, eds. Cambridge, MA: Belknap Press, Harvard University Press, 1935.
- [5] Wu, Yonghui, et al. Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation [DB/OL]. (2016-09-26) [2016-10-08]. <https://arxiv.org/abs/1609.08144>.
- [6] 皮尔士. 皮尔士论符号[M]. 詹姆斯·胡普斯, 编. 徐鹏, 译. 上海: 上海译文出版社, 2016年.
- [7] 皮尔士. 论符号[M]. 赵星植, 译. 成都: 四川大学出版社, 2014年.
- [8] 佩特丽莉. 符号疆界: 从总体符号学到伦理符号学[M]. 周劲松, 译. 成都: 四川大学出版社, 2014年.
- [9] 张爱玲, 杨子靖, 刘晨璇等. 人工智能技术发展专业口笔译实践耦合机制路径初探[J]. 外语电化教学, 2018(3): 88-94.

[作者信息] 姚婷婷 南京师范大学外国语学院, Email: 66583567@qq.com; 蒋蓉 皖江工学院基础教学部, Email: 717707591@qq.com