

机器翻译句法错误分析

罗季美

(同济大学 外国语学院, 上海 200092)

摘要:以“汽车技术文献翻译语料库”为基础,对机器译文和人工译文进行了平行对比,描述了对机器译文中出现的名词短语、动词短语、介词短语、动词不定式短语、分词短语和从句等错译的标注和统计,归纳分析了各类句法错译的典型形式,缩小了对语言进行形式化处理的范围,为翻译系统在原有基础上补充建立形式化规则并最终提高机器翻译质量提供了语言分析基础。

关键词:机器翻译; 机器译文; 人工译文; 句法错误

中图分类号: H085

文献标识码: A

文章编号: 1009-3060(2014)01-0111-08

对机器翻译的研究从 20 世纪 40 年代第一台计算机诞生至今,无论在理论研究方面还是工程实践方面都已经积累了丰富的经验。随着计算机和因特网的普及,机器翻译已经得到广泛推

广,产生了巨大的社会效益和经济效益。中国的改革开放也给中国翻译服务提供了巨大的市场,市场对翻译的巨大需求再一次激起人们对机器翻译研究的强烈兴趣。

一、机器翻译的适用范畴

“计算机翻译不适用于文学性很强或文化味很浓的文本,而适用于科普文献、金融商业交易、行政管理备忘录、法律文件、说明书、农业及医学资料、工业专利、宣传册、报纸报道等”^{[2]5}。1976 年加拿大蒙特利尔大学和加拿大联邦政府翻译局联合开发的“天气预报”英—法机器翻译系统 TAUM-METEO 成为在受限语言内机器成功翻译最早的一个例证。为在全球范围内推销其产品和服务,让国外用户能够读懂其产品说明书和用户手册等技术文件,降低语言转换成本,欧洲航空航天工业协会也为其文件中使用的英语限定了 985 个术语和 60 条写作规则^[3]。

“十五”期间,中国科学院计算机研究所对我国现有的机器翻译系统进行了测评,结合机器翻译系统对诗歌、散文、小说、剧本和受限语言等不同体裁、不同风格的文本翻译,在对译文进行具体分析、对比和综合评价的基础上,确定了英汉机器翻译在受限语言处理中取得了最高的 5 级适用级别和 85% 以上的可翻译度^{[6]131-162}。机器翻译取得的成就与受限语言的五大基本特征密切相关,其常用性、有限性、简明性、单义性和规范性^{[1]10}可以大大简化基于某一受限语言范畴的机器翻译规则的制定。由此可见要提高机器翻译的译文质量,在受限语言范畴内展开机器翻译研究应该是行之有效的方法。

二、机器译文句法错译研究的方法和目的

基于对机器翻译适用范畴的研究成果,《机器翻译后译文编辑模式研究》项目选择了同济大

学汽车学院的“汽车技术文献翻译语料库”中某一知名品牌汽车操作手册作为研究对象,在这一

收稿日期: 2012-09-25

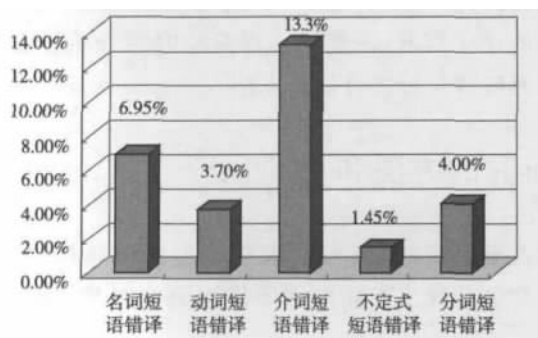
作者简介: 罗季美(1966-),女,安徽黄山人,同济大学外国语学院讲师。

限定专业范围内,由人工对英语原语、其机器译文和人工译文十万句对展开了句子一级的平行对比研究。语料中的机器译文由“华建机器翻译系统”(中国科学院计算机语言信息工程研究中心研发)生成获取,人工译文来源于同济大学汽车学院资料翻译研究所专业技术人员。

三、机器译文句法错误的统计和主要表现形式

《机器翻译后译文编辑模式研究》课题组在完成 1000 句抽样对比基础上归类设定了机器在词序、名词短语、动词短语、介词短语、被动语态、不定式短语和分词短语翻译处理上可能出现的七类句法错误^[4]。在对句法错误进行标注和统计中发现,词序错译表现最为复杂,错译率在各类句法错译中高居榜首,达到 29.35%^[5]。但因词序错译与机器对名词短语、动词短语、介词短语、被动态、不定式短语和分词短语等的错误处理存在很大的交错,以下不再将词序错译作为单独的错误形式进行讨论。此次译文对比也发现,机器对被动态的翻译处理达到了很高的准确性,对被动态的错译主要表现在机器对动词不定式中被动态的处理上,以下将在不定式短语错译中讨论被动语态错译。撇开词序错译和被动态错译,我们得到了表 1 所示的机器译文五类句法错译发生率。错译率为出现某类句法翻译错误的机器译文总句数/机器译文的总句数×100%。对同一句子中多次出现的同一类型句法翻译错误进行一次标注和统计;对同一句子中出现的不同类型句法错误分别加以标注和统计。

表 1 机器译文五类句法错译发生率



以下将根据机器译文句法错误的分类,结合统计结果,举例对语料对比中出现的机器译文典

本文将对机器译文中出现的句法错误加以描述和分析,从纷繁复杂的语言现象中寻找和归纳出机器在汽车技术文献这一受限专业领域内进行翻译处理时表现出的句法错误特征,为翻译系统补充建立形式化的句法规则提供语言基础,最终提高机器翻译的质量。

型句法错误加以描述、归纳和分析。在所举各例中,A句为机器译文,B句为人工译文。另外,为了使译文的对比更方便清晰,在所给例句中,每次只对所述的单一错译类型进行比较,并用下划线以示区分。

1. 名词短语错译

为了追求语言的简练,科技英语中大量使用名词短语,名词短语错译在整个语料中达到 6.95%。错译主要发生在对结构比较复杂的名词短语进行翻译时,如以下两例。

Install the new service hole cover.

A: 安装洞包括的新服务。

B: 安装新的检修孔盖。

此例中的名词短语“the new service hole cover”由形容词“new”+起修饰作用的名词短语“service hole”+名词“cover”构成,人工译文表现出了与原英语词序完全相同的顺序,将其译为“新的检修孔盖”。机器对“cover”词性的错误判断,对其前起定语作用的形容词“new”和名词短语“service hole”的错误切割,使得整个译文词序呈现出颠三倒四的乱象。

Remove the 4 bolts and engine moving control rod.

A: 除去移动操纵杆的 4 个螺栓和发动机。

B: 拆下 4 个螺栓和发动机移动控制杆。

对比人工译文可以发现,此例中“and”连接了两个名词短语“the 4 bolts”和“engine moving control rod”。机器将这两个并列名词短语错误判断为由现在分词短语“moving control rod”修饰的两个并列名词“the 4 bolts”和“engine”,由此产生错误译文。以上两例充分体现了对结构复杂的名词短语进行翻译时机器对词序处理的随意性。

2. 动词短语错译

出现动词短语错译的句子在整个十万句的机器译文中占 3.7%。错译率本身似乎并不高,而事实是并非每句句都会使用动词短语。因动词本身在句子中的重要性,出现错译往往会使原句语义大打折扣。对比发现动词短语错译的表现形式主要有以下四种类型:

(1) 由“动词+副词”构成的动词短语误译为由“动词+介词”构成的动词短语

Using SST and a hammer, tap in the oil seal.

A: 使用 SST 和一把铁锤, 在这个油封条内轻拍。

B: 使用 SST 和锤子, 敲入油封。

对比两句译文可以看到 A 句中机器误将原句中由“动词+副词”构成的动词词组“tap in”和宾语“the oil seal”分析成了动词“tap”和由“介词+宾语”构成的地点状语“in the oil seal”,使译文与原句表示的意义不相符。

(2) 由“动词+介词”构成的动词短语中介词与动词的分离

Align the thermostat jiggle valve with the upper stud bolt, and insert the thermostat in the water inlet housing.

A: 调整好这台恒温器轻摇有这上面双头螺栓的阀门,并且把恒温器插入这进水口住房中。

B: 将节温器跳阀与上双头螺栓对正,并将节温器插入进水口壳内。

对比两句译文可以看到在此例中机器没有识别“align X with Y”(将 X 与 Y 对正)这一固定动词词组,误将介词短语“with the upper stud bolt”当作名词“valve”的定语,造成错译。

When disconnecting the fuel line, cover it with a shop rag or a piece of cloth to prevent fuel from spraying or coming out.

A: 拆开这条燃料线时,与一块商店碎布或者一块布包括它防止燃料喷或者出来。

B: 断开燃油管路时,用棉丝抹布或一块布盖住,以防燃油喷出或涌出。

在此例中,机器没有识别“cover X with Y”(用 Y 盖住 X)这一固定动词词组,误将表示方式的介词短语“with a shop rag or a piece of cloth”

解释为表示伴随状态的状语“与一块商店碎布或者一块布(一起)”。

(3) 对动词短语的翻译不符合汉语表达语序

Install a new gasket to the No. 1 oil cooler bracket.

A: 对 1 号油更凉爽的方括号安装一个新垫片。

B: 将新衬垫安装到 1 号机油冷却器支架上。

此例中机器将动词短语“install X to Y”译为“对 X 安装 Y”。对比人工译文“将 Y 安装到 X 上”,A 句的译文显然不符合汉语的表达习惯。

Install the exhaust front pipe to the exhaust manifolds with the 2 nuts and 2 bolts.

A: 安装对排气歧管用 2 颗螺母和 2 个螺栓的排气前面管。

B: 用 2 个螺母和 2 个螺栓将前排气管安装到排气歧管上。

对比此例中的两句译文会发现,译员将动词短语“install X to Y with Z”译为“用 Z 将 X 安装到 Y 上”,汉语句义清晰明了;而机器没有按照汉语表达习惯解释动词短语,使得译文中由动词词组串联起来的三组名词短语关系混乱,严重影响译文的可理解性。

(4) “动词短语+多个并列宾语”的错译

The automatic headlight beam level control system mainly consists of the AFS ECU, rear height control sensor, and two headlight leveling motors.

A: 自动前灯梁水平控制系统主要由 AFS ECU 组成,后高度控制传感器并且两盏前灯使电动机成水平。

B: 前大灯光束高度自动控制系统主要包括 AFS ECU、后高度控制传感器和两个前大灯光束高度调整电动机。

在此例中,动词短语“consists of”后接三个作宾语的并列名词短语“the AFS ECU”、“rear height control sensor”和“two headlight leveling motors”,但机器译文表明机器未能准确识别动词短语后三个宾语的并列关系。

3. 介词短语错译

英语中的介词总数并不多,但一些常用介词如“in”、“on”、“for”、“with”和“as”等使用的频率

却非常高,其构成的介词短语在句子中可以表示多种句法功能。介词短语的使用频率也充分体现在介词短语机器译文 13.3% 的错误率上,这一比例在五类句法错译中位列第一。本次语料对比发现介词短语错译主要发生在介词短语在句子中作定语和状语时。

(1) 介词短语作定语

机器对作定语的介词短语的错误处理主要表现在以下两方面:一是介词短语作定语时序位处理不当;二是作定语的介词短语误译为状语。

Only for models without the intelligent AFS.

A: 只为模型没有聪明的 AFS。

B: 只用于不带智能 AFS 的车型。

在英语中作定语的介词短语后置于被修饰词,但汉语中的定语通常置于被修饰词前。在此例的机器译文中,作定语的介词短语“without the intelligent AFS”没有调整到被修饰词“models”之前,翻译处理时完全遵照原文词序,导致译文错误。

Calculate changes in the vehicle posture based on the signals from the height control sensor and each ECU.

A: 从高度控制传感器和每 ECU 计算在基于信号的车辆姿势方面的变化。

B: 根据高度控制传感器和各个 ECU 的信号计算车身姿态的变化。

此例中修饰“signals”的介词短语“from the height control sensor and each ECU”在机器译文中被错误地处理为句子的状语,导致对原文的误译。

(2) 介词短语作状语的错译

科技文本常常涉及到对方式、手段、速度、时间、步骤等的描述。用介词短语来描述方式、速度、时间、步骤等也是典型的科技英语文体特征。对比以下四例的两种译文,就会发现机器对这些常用介词短语的翻译并不到位,甚至存在重大错误。

Confirm that the belt has not slipped out of the grooves on the bottom of the crank pulley by hand.

A: 确认这条带子没亲手在曲柄滑车的底部从槽沟中滑掉。

B: 用手检查,确认皮带没有从曲轴皮带轮底部的凹槽中滑脱。

对比人工译文可以看到,此例的机器译文将原本说明主句谓语“Confirm”方式的“by hand”错误处理为修饰从句谓语动词“slipped”的方式状语,导致译文的重大错误。

However, reading the Data List as the first step of troubleshooting is effective to find malfunctions.

A: 当检修的第一步有效找到故障时,不过,读数据列举。

B: 然而,将读取数据表作为故障排除的第一步对于发现故障很有效。

英语中的“as”具有介词和连词的双重词性。对比译文可以看到,在此例中机器将介词短语“as the first step of troubleshooting”错误处理为由连词“as”引导的时间状语从句“as the first step of troubleshooting is effective to find malfunctions”,将介词“as”(作为)误译为连词(当……时),导致对整个句子主谓结构的错误处理。

Run the engine at 1,000 to 1,300 rpm for 5 seconds, and then check that the engine rpm returns to idle speed.

A: 5 秒 1,000 到 1,300 rpm 运转发动机,以及检查发动机 rpm 返回闲置的速度。

B: 以 1,000 至 1,300 rpm 的转速运转发动机 5 秒钟,然后检查并确认发动机转速恢复到怠速转速。

此句涉及用介词“for”和“at”构成的描述时间和速度的两个介词短语。这两个介词短语常常同时出现在“主语+(可以用速度来描述的)谓语动词+at the speed of … + for (time)”或“(可以用速度来描述的)谓语动词+宾语+at the speed of … + for (time)”的句型中,译文应该表述为“主语+以……速度+谓语动词+时间”或“使(让)宾语+以……速度+动词+时间”。A 句将“Run the engine at 1,000 to 1,300 rpm for 5 seconds,”译为“5 秒 1,000 到 1,300 rpm 运转发动机”,与这一句型的汉语表达方式不相符,导致译文语义不清。

Uniformly loosen and remove the 8 main bearing cap bolts and seal washers in several

steps and in the sequence shown in the illustration.

A: 均匀放松并且除去 8 主要部分忍受帽子螺栓和海豹洗衣机在几步内和在插图内表现的顺序内。

B: 按图示顺序, 分几个步骤均匀松开并拆下 8 个主轴承盖螺栓和密封垫圈。

此例中用“in several steps”、“in the sequence”和“shown in the illustration”三个短语对步骤、顺序和方式依次进行了描述。对比两句译文可以发现, 在汉语中针对同样的步骤、顺序和方式, 其描述顺序与英语是不同的。汉语遵照了定语在前被修饰词在后及在方式上由粗到细的次序, 译成了“按图示顺序, 分几个步骤……”, 而机器译文呈现出的混乱语序严重破坏了译文的可理解性。

英语科技文本中常出现含有“with”的独立主格结构。独立结构多用作状语, 在句子中位置灵活。含有“with”的独立主格结构在具体的上下文中可以表示原因、时间、方式、伴随状态、条件或结果等, 但其复杂的句法意义也使机器翻译留下许多遗憾。

With the driver side door open, turn the driver side door lock knob to LOCK, and then close the driver side door.

A: 由于司机侧门打开, 把司机侧门锁球形门柄转动到锁, 然后关上司机侧门。

B: 打开驾驶员侧车门, 将驾驶员车门的门锁按钮转至锁止位置, 然后关闭驾驶员侧车门。

对比此例的两句译文可以看到机器将原本表示状态的独立结构“With the driver side door open”误译成了表示原因的状语。

A “used belt” is a belt which has been used for 5 minutes or more with the engine running.

A: 一条“使用的带子”是已经用于有运转的发动机的 5 分钟或更多的一条带子。

B: “旧皮带”是指在发动机运转的情况下使用时间超过 5 分钟的皮带。

The hub bearing may be damaged if it is subjected to the vehicle's full weight, such as moving the vehicle with the drive shaft removed.

A: 如果它受到车辆的充分的重量, 轮毂轴承

可能被损坏, 例如用除去的驱动轴移动车辆。

B: 如果轮毂轴承承受车辆的全部重量, 则它可能会损坏, 例如在驱动轴已拆下的情况下移动车辆时。

对比以上两例中的机器译文和人工译文, 可以发现机器未能准确处理含有“with”的独立主格结构, 将两例中原本表示状态的独立主格结构状语分别错误理解为表示修饰关系的定语和表示方式手段的状语。

4. 动词不定式短语错译

动词不定式是英语中常见的句法现象, 有自身的时态和语态变化, 有着丰富的句法功能。译文对比结果中 1.45% 动词不定式错译率表明机器对动词不定式的处理已经达到相当高的准确性。对比中发现的错译主要发生在动词不定式短语作目的状语和一般现在时被动语态动词不定式短语作定语时。

(1) 作目的状语的动词不定式短语错译

Always use a fully charged battery to obtain an engine speed of 250 rpm or more.

A: 总是使用完全把电池记在账上获得一发动机 250 rpm 或更多的速度。

B: 务必使用充足电的蓄电池使发动机转速达到 250 rpm 或更高。

在此例中可以看到译员在翻译表示目的的动词不定式短语时添加了“使”字。在汉语中缺少此类连接词是无法准确传达“目的”之义的; 机器译文因没有增加相应的词汇, 未能忠实传达原文表示目的的语义。

(2) 作定语的一般现在时被动语态动词不定式短语错译

Line up the two parts of the pipes to be connected.

A: 排列成行管的两个部分被联结。

B: 对准两个将要连接的油管部位。

在此例中, 动词不定式短语紧随名词之后, 成为限定名词的定语。对比译文可以发现机器准确识别了“被动语态”, 但对“一般现在时动词不定式”可以表示“将来发生的动作”这一语法意义还未能做出准确处理。

5. 分词短语错译

分词短语是英语中常见的句法形式, 因其具

有比从句更简洁的特征,在科技英语中得到广泛应用。和动词不定式一样,分词也有自己的时态和语态,兼具动词和形容词的性质,在句子中可以充当表语、定语、状语、补语和独立结构等。这些复杂的句法功能给机器翻译增加了不少困难。本次译文对比中,分词短语错译在整个语料中占了 4% 的比例,其错译形式主要表现在以下几方面:

(1) 分词短语作定语时词序处理不当

The AFS ECU receives signals indicating the height of the vehicle from the rear height control sensor.

A: AFS ECU 收到信号从后高度控制传感器注明车辆的高度。

B: AFS ECU 接收后高度控制传感器发送的车辆高度信号。

All ECUs and sensors connected to the CAN No. 5 J/C are not displayed for 1 minute.

A: 全部 ECUs 和传感器连接罐头 5 号 J/C 没被展示 1 分钟。

B: 所有连接到 CAN 5 号接线连接器上的 ECU 和传感器均持续 1 分钟未显示。

在汉语中定语通常放置在被修饰词之前,但在上两例中,机器译文遵从原文词序,未对以现在分词短语和过去分词短语形式出现的定语做出必要的前置调整,生成的译文不符合汉语表达习惯,导致了译文和原文的语义差异。

(2) 分词短语作状语时位序处理不当

The AFS ECU determines whether the vehicle is moving forward or backward based on this signal.

A: AFS ECU 确定车辆根据这个信号是向前行驶或者向后的。

B: AFS ECU 根据该信号判断车辆是在前进还是后退。

用作状语的去分词短语“based on …”在科技英语文本中频繁出现,其位置非常灵活,可置于句首、句中或句末,短语可以与句子的其他部分连为整体,也可以用逗号将其与句子的其他部分分隔。在此例中可以看到,译员在翻译此类句式时,通常会遵从汉语表达习惯,在译文中将“based on …”短语进行必要的位置调整,而机器的调整确实还不够到位。

(3) 分词句法功能的混淆

Leveling motor LH ground

A: 使电动机 LH 地成水平

B: 左侧光束高度调整电动机搭铁

Check A/F sensor and heated oxygen sensor operation.

A: 检查 A/F 传感器并且 加热 氧传感器操作。

B: 检查 A/F 传感器和 加热型 氧传感器的工作情况。

在上两例的机器译文中,作“定语”的现在分词“Leveling”和过去分词“heated”分别被机器错误分析为句子的“谓语”动词。错译也反映了机器对谓语动词时态和分词时态分析处理的不稳定性。

Securely support the automatic transaxle using a jack.

A: 可靠支持使用一台千斤顶的自动与变速器连成一体的驱动桥。

B: 用千斤顶牢牢地支撑住自动传动桥。

The AFS ECU determines the Lo beam angle based on the number of steps (position) of the step motor.

A: AFS ECU 确定基于台阶电动机的台阶的数量(位置)的卢梁角度。

B: AFS ECU 根据步进电动机的步数(位置)确定近光光束角度。

对比译文可以看到,在以上两例中机器分别将作方式状语的去分词短语“using a jack”和作为句子状语的“based on …”过去分词短语错误分析为“the automatic transaxle”和“the Lo beam angle”的定语。

6. 从句错译

在语料对比的过程中,也发现了一些在设置错误类型的抽样对比中被忽视的从句错译问题。在笔者承担的一万句对的对比中,对从句错译进行了补充标注和统计。在一万句对的语料中复合句占 13.15%,其涉及的从句主要包括状语从句(10.1%)、宾语从句(1.45%)和定语从句(1.1%)等。从句错译率在一万句对语料中为 5.35%,在所有复合句中错译率则高达 40.68%。数据表明对从句错译确实有展开研究的必要性。

(1) 定语从句错译

从句错译中,定语从句的错译率最高,错译的句数占有所有定语从句的 72.73%,错译主要表现在以下两方面:一是定语从句序位处理不当;二是定语从句引导词错译。

Disconnect the branch wire connectors, which are connected to the ECUs or sensors that never appear on the screen, from the CAN J/C.

A: 拆开分支电线连接器, 这连接从未在屏幕上出现的 ECUs 或者传感器, 从罐头 J/C。

B: 将连接到从未在屏幕上出现的 ECU 或传感器的支线连接器从 CAN 接线连接器上断开。

与汉语中定语通常置于被修饰词之前不同,在英语中,无论是限制性定语从句还是非限制性定语从句都置于被修饰词之后。处理定语从句的翻译时,译员一般都会对词序做出相应调整。但此处机器译文中的定语“这连接从未在屏幕上出现的 ECUs 或者传感器”保留了其原文的序位,表明机器在翻译中还缺乏调整定语从句语序的能力。

定语从句有关系代词和关系副词两类引导词,此次汽车操作手册语料涉及的定语从句主要由关系代词“that”和“which”引导。对比译文发现机器对定语从句引导词的处理极不稳定,错译主要表现在对关系代词的错误处理上。如在下两例的机器译文中,定语从句引导词“that”和“which”分别被处理成了代词“那”和疑问词“哪个”。

If voltage that is input to the light control ECU is approximately 20V or more, the ECU stops illuminating the headlights.

A: 那把 ECU 输入给光量控制的电压 20 V 或更多的大约如果, ECU 停止清楚前灯。

B: 如果输入灯控 ECU 的电压约为 20 V 或更高, ECU 会停止前大灯照明。

Check that the ECUs and sensors, which repeatedly appeared and disappeared from the “BUS CHECK” screen, are constantly displayed.

A: 检查 ECUs 和传感器, 哪个反复出现并且从屏幕, 被经常展示的“公共汽车检查”消失。

B: 检查并确认从“BUS CHECK”屏幕上反复时隐时现的 ECU 和传感器已经持续显示在屏

幕上。

(2) 状语从句错译

机器对状语从句的错译占有所有状语从句的 34.16%。本次语料中涉及的状语从句主要为方式状语从句、时间状语从句和条件状语从句,错译也主要发生在对这些从句的不当处理上,包括对方式状语从句语序的不当处理、对时间状语从句引导词“when”的错译和对条件状语从句引导词“if”的错译。

Using a piston ring expander, install the oil ring rail as shown in the illustration.

A: 使用一台活塞环扩展器, 安装油环横栏如插图中所示。

B: 使用活塞环扩张器, 如图所示安装油环刮片。

Install the 2 knock control sensors with the 2 bolts as shown in the illustration.

A: 安装这 2 阵敲打如这张插图中所示与这 2 个螺栓一起传感器。

B: 如图所示, 用 2 个螺栓安装 2 个爆震控制传感器。

“as shown in the illustration”是科技英语中常见的的方式状语从句,在汉语中通常置于表示动作的动词前。两例机器译文中,一句保持了其原文语序,一句对其语序做出了不当调整。两种错误的处理方式展现了机器对这一方式状语从句语序处理的随意性;人工译文也体现了方式状语在汉语句子中位置的灵活性,但遵从了汉语表达习惯。

本次语料所涉及的时间状语从句引导词主要有“when”、“before”和“after”。错译主要发生在对引导词“when”的错误处理上。

When person carrying electrical key enters detection area, door will enter unlock standby mode and front map light will illuminate.

A: 传播电钥匙的人加入察觉地区什么时候, 门将进入开锁备用方式和前面地图光将亮。

B: 当有人携带电子钥匙进入检测区域, 车门将进入解锁待命模式并且前阅读灯将亮起。

The AFS ECU performs the medium-to-high speed control when all the following conditions are fulfilled.

A: 这 AFS ECU 执行这种媒介给高速度控

制到那时全部下列状况都被履行。

B: 当符合所有下列条件时, AFS ECU 执行中高速控制。

在以上两例中, 机器将两句状语从句的引导词“when”分别处理成了“什么时候”和“到那时候”, 与从句中的主谓结构完全分离, 造成译文的重大错误。另外, 机器译文也表现出对状语从句引导词“when”在句子中语序处理的不稳定性。

本次语料所涉及的条件状语从句主要由“if”引导, 错译主要发生在对“if”序位的不当处理上, 错误形式基本相同, 即机器将从句的引导词“if”错误置于整个从句的句尾, 如下例 A 句所示。

If there is a problem in this intercommunication, the active control engine mount ECU sets a DTC.

A: 有在这相互联系内的一问题如果, 积极控制器骑在马上 ECU 确定 DTC。

B: 如果互相通信有故障, 则主动控制发动机支座 ECU 设置一个 DTC。

(3) 宾语从句错译

语料中涉及的宾语从句主要为由“that”、“if”和“whether”引导的宾语从句, 错译占有所有宾语从句的 56.41%。机器对宾语从句的错误处理主要表现在其对“that”引导的宾语从句的处理上, 包括以“that”为引导词的宾语从句错译为定语从句及宾语从句引导词“that”的词义错译。

Check that the striker can engage with the hood lock smoothly.

A: 罢工者能顺利与罩锁接合的检查。

B: 检查并确认锁扣能够与发动机盖锁顺利接合。

“that”既可以用作定语从句的引导词, 也可用作宾语从句的引导词, 显然机器在此处将两者混淆, 错译应该与机器无法准确判断从句前“check”的动词词性有关。

Check that all functions in the smart entry and start system cannot be operated.

A: 检查那全部在这聪明进入内的功能并且启动系统不能被经营。

B: 检查并确认智能上车和起动系统中的所有功能都不能操作。

在此例中, 机器错误将宾语从句引导词“that”误译为指示代词“那”, 导致对宾语从句的错译。对比以上两例的机器译文也可发现, 机器在对这两句同类句型的处理上, 表现出其对“check”词性的判断和对宾语从句引导词“that”的分析呈现不稳定性。

英语具有“形合”的特征, 句法是其“形合”实现的重要方式, 是组织零散词汇构成句子的手段。机器在对名词短语、动词短语、介词短语、不定式短语、分词短语和从句的翻译处理中表现出的错误可以说是纷繁复杂, 严重的句法错误甚至使整个句子变得凌乱不堪, 不知所云, 这也充分证实了对机器译文句法错误展开研究、进行归纳的必要性。我们对“汽车技术文献翻译语料库”中机器译文和人工译文进行的精确对比、标注、统计、分析和归纳也确实发现了各类句法错译中存在的特征和共性。

四、结 语

在汽车技术文件这一限定领域内的译文精确对比表明生成本次汉语译文的华建机器翻译系统具备了一定的英汉翻译能力, 对英语中大部分主要时态、被动语态、大部分的动词不定式的句法作用及某些特殊句型等的处理都达到了相当高的准确性, 确保了机器译文基本的可懂度, 保证了机器译文的一定价值, 为译后编辑, 特别是建立译后编辑模式打下了一定的语言基础。另一方面, 译文对比也发现, 句子结构稍一复杂, 修饰成分和限定成分一多, 机器处理就开始力不

从心。在句法上要为每一句特例编制形式化的语法可能还不现实; 语言的形式化本身也是一项艰深的难题。对机器译文和人工译文进行句法处理的对比将有助于在人力技术有限的情况下缩小对语言进行形式化处理的范围。在对语言规则形式化的编程过程中, 应该以解决机器翻译目前存在的典型句法错译为首要目标, 减少不必要的浪费, 提高机器翻译的可懂度和忠实度, 最终提高机器翻译的鲁棒性。

(以下“参考文献”、“英文摘要”转第 124 页)

[16] Lyons, J. *Semantics 2*[M]. Cambridge: Cambridge University Press, 1977.

[17] 任绍曾. 《话语标记》导读[M]//希夫林话语标记. 北京:世界图书出版公司, 2007:90.

The Pragmatic Functions of “*ni zhidao*” as a Discourse Marker

— A Study from the Perspectives of Cognition, Psychology and Social Interaction

SHAN Yi

(School of English Language and Culture, Zhejiang International Studies University,
Hangzhou, 310023, China)

Abstract: The pragmatic functions of “*ni zhidao*,” which are dynamic and contextual, call for a multi-perspective study rather than an account from a single perspective. This paper focuses on the analysis of the functions of construction of cognitive context, projection of psychological attitudes, and ostension of relevant inference of “*ni zhidao*”, from the angles of cognition, psychology and social interaction, thus revealing its underlying pragmatic motivations. The present study holds that a comprehensive and accurate mastery of the pragmatic functions and motivations of discourse markers results from a multi-perspective research in dynamic interactive contexts.

Key words: discourse marker; “*ni zhidao*”; construction of cognitive context; projection of psychological attitudes; ostension of relevant inference

(责任编辑:谢 闽)

(上接第 118 页)

参 考 文 献

[1] Brockmann, D. Controlled Language and Translation Memory Technology: A Perfect Match to Save Translation Cost[C]. TC-Forum, 1997:4-97.

[2] Hutchins, W. J., and Somers H. L. *An Introduction to Machine Translation*[M]. San Diego: Academic Press, 1992.

[3] 胡清平. 机器翻译中的受控语言[J]. 中国科技翻译, 2005, (8):24-27.

[4] 李梅. Identifying Error Patterns in MT Post-editing[R]. 上海:第十八届世界翻译大会, 2008.

[5] 罗季美, 李梅. 机器翻译译文错误分析[J]. 中国翻译, 2012, (5):84-89.

[6] 张政. 计算机翻译研究[M]. 北京:清华大学出版社, 2006.

Analyses of Syntactic Errors in Machine Translation

LUO Ji-mei

(College of Foreign Languages, Tongji University, Shanghai 200092, China)

Abstract: This research marks the syntactic errors in the machine translation(MT) of noun phrases, verb phrases, prepositional phrases, infinitive phrases, participial phrases and various clauses on the basis of a sentence-level comparative study of MT and human translation from Automobile Document Translation Corpus. The paper details the statistic results of each type of errors and describes the typical patterns of syntactic errors in MT output, which aims to narrow down language rules in need of optimization or formalization in MT and finally improve the robustness of MT.

Key words: machine translation (MT); MT output; human translation; syntactic errors

(责任编辑:周淑英)