

科技文本机器翻译句层错误分析

刘 玉

中南大学, 外国语学院, 湖南 长沙
Email: liuyu0309@csu.edu.cn

收稿日期: 2021年6月21日; 录用日期: 2021年7月26日; 发布日期: 2021年8月3日

摘 要

机器翻译借助计算机对文本进行自动翻译, 因翻译速度较快, 现被普遍用于翻译实践中。由于中西文化和思维方式的不同, 英汉语在句子层面存在明显的差异, 这些差异对于机器翻译译文的影响是显而易见的。机器译文与人工译本相比, 译文质量仍有一定的差距。因此, 透析英汉语句子层面的差异, 分析机器译文句层错误, 有利于改进机器翻译的质量。本文以英汉语句层的差异为切入点, 选取科技文本作为语料, 将其机器译文与人工译本进行对比, 从语法转换因子和逻辑关系两个方面归纳机器译文出现的句层错误, 揭示机器翻译中应注意的几个问题, 为提升机器翻译系统的形式规则提供语言基础, 为译后编辑提升译文质量提供方向。

关键词

英汉对比, 机器翻译, 错误分析

Error Analysis at the Sentence Level in Machine Translation of Sci-Tech Texts

Yu Liu

School of Foreign Language, Central South University, Changsha Hunan
Email: liuyu0309@csu.edu.cn

Received: Jul. 21st, 2021; accepted: Jul. 26th, 2021; published: Aug. 3rd, 2021

Abstract

Machine translation (MT) is the process of translating texts from one language into another by computer. Due to its fast translation speed, MT has been widely used in translation practice. For different logical thinking and cultural background, differences can be found between English and

Chinese at the sentence level. It's obvious that these differences will have an impact on machine translation. There is a certain gap between the quality of machine translation and human translation, and errors can be found in the machine translation. Thus, mastering the differences between English and Chinese at the sentence level is very important for improving the quality and readability of machine translation. Based on the difference between English and Chinese at the sentence level, this paper intends to compare the machine translation and human translation of scientific and technological texts, and summarize the errors in machine translation at sentence level grammatically and logically, aiming to provide a language basis for improving the formal rules of machine translation system.

Keywords

Difference of Chinese and English, Machine Translation, Error Analysis

Copyright © 2021 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

近年来,在英汉对比研究中,学者们在新兴领域进行了有益尝试,其中一个领域是英汉对比与机器翻译的研究(吴慧珍,2010) [1]。机器翻译的广泛使用会提高翻译效率,大力节省人力,对于人工翻译造成了冲击。然而,目前的英汉机器翻译都只能处理一些简单的句子,对于一些复杂结构的句子,机器翻译就力不从心,翻译一些常见的句法现象时会出现错误(罗季美、李梅,2012) [2]。究其原因,由于文化背景和思维方式的不同,英汉语在句子层面呈现出较大差异,但机器并不能完全识别出英汉之间存在的差异。有学者从英汉对比的角度进行机器翻译研究,如李刚、梁茂成(1999) [3]采用对比语言学的方法,利用计算机翻译系统,分析英汉语言中代词的异同。也有不少学者围绕机器翻译的译文质量进行研究,如崔启亮(2015) [4],殷姣(2017) [5],孙逸群(2018) [6]等学者从词汇、句法层面对于机器翻译的错误进行了剖析总结,其中孙逸群还从符号层面分析了机器翻译中出现的错误。但对于句法层面的错误分析,学者们大多是从短语切分的角度,讨论机器翻译中的短语错译或长难句的错译,鲜有学者涉及到句层的语法转换因子如语态、情态、语气等的错误分析,对逻辑关系如语序和连接词的讨论也比较少见。机器不是人脑,不具备思维能力,在英汉语差异的影响下,机器在理解源语和表达阶段难免会出现一定的转换错误。本文拟以科技文本为例,采用谷歌这款常用的机器翻译软件进行汉英互译,将其与人工翻译译文进行比较。文章所选例子来自于科技翻译相关教材,其人工译文经过推敲呈现于教材之中,具有一定的可信度。基于英汉语句层的差异分析,文章从语法转换因子和逻辑关系两个方面分析机器翻译在汉英转换中出现的句层错误,分析错误出现的原因,为译后编辑工作和机器翻译算法改进提供方向。

2. 语法转换因子

对于一些常见的语法现象如语态、情态、语气等,英汉语存在着明显的差异。在英汉转换的过程中,这些常见的语法转换因子会对译文的质量产生影响。

2.1. 语态不当

英汉语在语态方面存在着差异。英语中的语态可分为主动态和被动态,汉语中没有语态的概念,但有句子的被动意义和主动意义。英语中的被动语态往往体现于动词的形态变化,但汉语里的被动意义通

过虚词体现。机器在进行英汉语转换的过程中，有时会识别不出句子中所表达的被动意义，从而造成语态使用不当。

例 1: 各类科学家要分析数据，都需要有统计学知识。

机器译文: All types of scientists need statistical knowledge to analyze data.

人工译文: A knowledge of statistics is required by every type of scientist for the analysis of data.

例 1 中，“有统计学知识”属于陈述的方法，是句子中的信息重心，译成英语时采用被动态表述为宜，能够突出句子的受事要求和关注焦点。但机器译文中没有将这句话译为被动态，而是顺着中文的语序选用主动态直译，这不能体现被动转换的需求。

2.2. 情态错误

有时候为了表达自身的观点和认知，人们往往会用一些情态化的词汇来表示其主观意义。英语中表示情态意义的是情态动词，在汉语中，我们通常将表示情态意义的词称为能愿动词。英汉语对于情态意义的词汇也有不同的分类，机器翻译过程中，对于汉语能愿动词和英语情态动词的程度会出现识别不清的情况，从而造成翻译错误。

例 2: ① 这个千斤顶能支撑两顿的重量。

② 这个千斤顶可支撑两顿的重量。

机器译文: This jack can support the weight of two meals.

人工译文: ① The jack can support 2 tons.

② The jack could support 2 tons.

例 2 中“能”和“可”都属于汉语中表示主观判断的能愿动词，但在英语中，同样表示“可以”的情态意义会有不同的情态动词对应来显示程度的差异。机器对于两句话的翻译结果中都选用了“can”这一个情态动词，并不能显示出程度的差异。人工的译文里，将“能”译为“can”，将“可”译为程度更弱的“could”，程度差异一目了然。

2.3. 语气错误

汉语中认为语气即态度，一般认为，汉语中共有四种语气：陈述、疑问、祈使、感叹，句子的语气往往通过语调变换或标点实现(齐沪扬，2002) [7]。英语中认为语气是形式化的体现，即通过语法特征说明态度，英语中语气也可分为四类，分别为陈述、疑问、祈使、虚拟(魏在江，2003) [8]。通过对语气的大致划分，我们发现英汉语的语气差异主要在于虚拟语气，这也是机器进行英汉转换时的易错点。

例 3: 人类如果能把所有的太阳能都利用起来，那么就不需要煤和其他燃料了。

机器译文: If human beings can use all solar energy, then coal and other fuels will not be needed.

人工译文: If man made use of all the sunlight, he would have no need for coal or other fuels.

例 3 中的“如果”表示句子是个假设的概念，句子表达的意义与现在事实相反，在转为英语时，可以选用虚拟语气，表达与现在事实相反的概念。虽然机器译文出现了 IF-clause，但并没有翻译出虚拟的语气，这说明机器翻译对于语气的转换仍需改进。

3. 逻辑关系

英汉语句子对于逻辑关系的体现有很大的不同，英语句子主要的逻辑关系有因果、转折、条件、让步等，通常通过连接词或分句等语言结构显示逻辑关系。而汉语句子是意合语言，逻辑关系常常隐于句子之间，可通过一定的语序来体现(谷青松，2015) [9]。英汉语转换过程中，应注意到英汉语逻辑关系表达的差异性，从而正确译出句子中显现或隐含的逻辑关系。逻辑关系表达是思维活动，机器并没有思维

的能力，翻译软件产出的译文也表明机器翻译尚不能完全识别句子逻辑关系，主要表现为句子连接词输出错误和语序不当。逻辑关系不当容易让读者对于句子意思产生误解，应引起学者们的重视。

3.1. 语序不当

语序是思维习惯在句法上的直接表现。一般认为，中国人更为含蓄，属于螺旋式思维，习惯于整体，重悟性，所以汉语中通常把次要信息居前，在句子末尾呈现主要信息。但英美人的思维方式相反，直线式的思维要求他们在表达过程中开门见山，直接点题(孙晓凌、汪北华, 2002) [10]。所以为了使译文逻辑清晰、语义明确，语序的调整在英汉语转换过程中至关重要。

例 4: 虽然铝的强度不及钢，但是当铝与少量的钢、锰和镁制成合金，并经过热处理后，其强度可以接近钢的强度。

机器译文: Although aluminum is not as strong as steel, when aluminum is alloyed with a small amount of steel, manganese, and magnesium, and after heat treatment, its strength can be close to that of steel.

人工译文: Aluminium, though much less strong than steel, can be given a strength approaching that of steel when it is alloyed with small quantities of copper, manganese and magnesium, and subjected to heat treatment processes.

根据中文表达“先轻后重”可知句子的主要信息为“其强度可以接近钢的强度”，但句子中间夹入了较多的状语成分。在转换成英语时，为了符合英语表达习惯，应对语句重心进行调整。但我们发现机器译文仍按照原文的顺序进行直译，并没有对句子的主次信息进行调整。相比之下，人工译文将 a strength approaching that of steel 放在从句之后，状语成分之前，这样一来，主次信息层次更为清楚。

3.2. 连接词缺失

由于分属不同的语系，英汉语在呈现逻辑关系时存在着很大的差异，这种差异的直接体现方式就是对于连接词的使用不同。一般来说，英语重形合，属形态型语言，句子之间会用一些语言结构如连接词或者分句连接句与句之间的关系，而汉语重意合，属语义型语言，句子结构简短而松散(程洪珍, 2003) [11]。汉语句子逻辑关系通常隐含于句子成分的意义中，层层推进，对于连接词的使用则少之又少。由于这种连接手段的差异，我们发现机器翻译还不能对连接词进行完全恰当的处理，从而易造成逻辑关系的混乱。

例 5: 活塞与气缸之间的配合间隙，对发动机的寿命影响很大。影响它的主要因素是活塞在工作状态下的变形，真实反映活塞变形情况，为此，有限元法在数值分析中得到了广泛的应用。

机器译文: The clearance between the piston and the cylinder greatly affects the service life of the engine. The main factor affecting the fit clearance is the deformation of the piston in the working state, which truly reflects the deformation of the piston. The finite source element method has been widely used for numerical analysis.

人工译文: As the joint gap between a piston and a cylinder has great influence on the service life of an engine, while the primary factor influencing the joint gap is the deformation of piston under working condition, it is important to indicate the actual deformation of piston and for this purpose the finite element method has been widely applied to analyze it numerically.

机器译文中，句子的单句都是符合语法的，但将这段话作为整体来看时，会发现机器翻译的译文句子之间的关系并不明确。分析汉语可知，句子想表达的意思是“活塞的变形是配合间隙的影响因素”，所以把整段话译为因果关系最为恰当，人工翻译时，我们可通过添加连接词 As 将句子之间的逻辑关系清晰地展现出来。可见，英汉转换的过程是形合和意合相互转换的过程。在英汉转换过程中，可以通过连

接词的增加或减少呈现逻辑关系。但逻辑与思维相关，机器并不具备人脑的思维，所以机器在英汉转换过程中如何清楚地展现逻辑关系对于后续机器翻译系统的改进是一个不小的挑战。

4. 结语

基于英汉语在句层差异的分析，通过对比机器译文和人工译文，我们发现机器译文在一些基本的语法因子如语态、情态、语气等方面会出现错误。对于句子之间逻辑关系的呈现手段如语序和连接词选择也不能令人满意。归根结底，英汉语属于不同的语系，二者之间存在着诸多差异。而机器不具备人脑的思维，不能根据英汉语的差异选择恰当译文。文章针对科技文本的汉英机器翻译进行了分析，虽语料有限，但对于机器翻译系统的提高与译后编辑工作都有着启发意义。

机器翻译系统的研发人员应注意到机器翻译的错误点，在改进机器翻译系统时，注重意义的切分，加强逻辑语义的识别，适当添加连接词。译后编辑人员也应以句法错误的几个方面为考察点，迅速地对机器译文进行调整，提高译后编辑效率，增强机器译文的可读性。就目前来看，机器翻译有着速度优势，而人工翻译在译文质量方面更胜一筹。在信息化时代，人们对于翻译的需求越来越大，这要求研发人员提高英汉翻译软件的正确率和可读性，不断提高翻译速度的同时保证译文质量，逐步地完善机器翻译技术。

基金项目

2020 年度湖南省研究生科研创新项目“人工智能时代信息型文本译后编辑研究”(项目号: CX20200372)。

参考文献

- [1] 吴慧珍. 英汉对比研究十年回顾(1999-2008) [J]. 长春理工大学学报: 社会科学版, 2010, 23(1): 79-82.
- [2] 罗季美, 李梅. 机器翻译译文错误分析[J]. 中国翻译, 2012, 33(5): 84-89.
- [3] 梁茂成, 李刚. 英汉代词对比分析及计算机辅助翻译[J]. 外语与外语教学, 1999, 12(6): 22-24.
- [4] 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.
- [5] 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.
- [6] 孙逸群. 材料类文摘机助翻译的错误剖析[J]. 中国科技翻译, 2018, 31(3): 15-18.
- [7] 齐沪扬. 论现代汉语语气系统的建立[J]. 汉语学习, 2002(2): 1-12.
- [8] 魏在江. 英汉语气隐喻对比研究[J]. 外国语(上海外国语大学学报), 2003(4): 46-53.
- [9] 谷青松. 英汉科技翻译中隐性逻辑的显性化[J]. 中国科技翻译, 2015, 28(3): 7-10.
- [10] 孙晓凌, 汪北华. 从思维方式差异看中西文化差异[J]. 河海大学学报(哲学社会科学版), 2003, 5(2): 68-70.
- [11] 程洪珍. 英汉语差异与英语长句的汉译[J]. 中国科技翻译, 2003, 16(4): 21-22.