Errors in Coherence and Logic Embodied in the Microsoft Machine Translation Text— Based on the corpus of the Chinese-English translation texts taken from the 30th Han Suyin International Translation Contest

Min Yang[1,a], Fan Jia[2,b,*]

[1]English Language Institute, Xi'an International Studies University, Xi'an, Shaanxi, China

[2]School of English Languages, Southwest Jiaotong University, Chengdu, Sichuan, China

[a]yangmin1979@163.com

[b]jiafanfanjia@163.com

*jiafanfanjia@163.com

Keywords: Microsoft machine translation text, errors, discourse, coherence and logic.

Abstract. Based on the corpus of the original text, the Microsoft machine translation text, the human-translated text as well as the reference text taken from the 30th Han Suyin International Translation Contest, the paper establishes the error classification standard from three language levels and then focuses on the typical errors in coherence and logic embodied in the Microsoft machine translation text from the discourse level, aiming to provide the Microsoft machine translation system with reference and enlightenment for better translation quality and accuracy in the future.

微软机器翻译中行文连贯与逻辑思维问题研究
——以第三十届韩素音国际翻译大赛汉译英文本为语料

杨敏[1,a]，贾蕃[2,b,*]

[1]西安外国语大学英语实验基地，西安，陕西，中国

[2]西南交通大学外国语学院，成都，四川，中国

[a]yangmin1979@163.com

[b]jiafanfanjia@163.com

[*]jiafanfanjia@163.com

关键词：微软机器译本；错误；语篇；行文连贯；逻辑思维

摘　要. 本文以第三十届韩素音国际翻译大赛汉译英竞赛原文、微软机器译本、人工译本及参考译文为语料，从词汇、语法和句式三个层面建立了错误分类标准，并对微软机器译本中凸显的典型错误进行了标注与分析，在此基础上，从语篇层面重点探讨了微软机器翻译中的行文连贯与逻辑思维问题，以期为微软提高机器翻译译文质量提供参考与借鉴。

1. 引言

1989年，由中国翻译协会、《中国翻译》编辑部发起，英籍华裔作家韩素音支持创办了

Copyright © 2019, the Authors. Published by Atlantis Press.
This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

第一届韩素音青年翻译奖竞赛。第一届竞赛分为英译汉和汉译英两项，原文以散文和议论文为主，兼有小说和经贸类文本。竞赛自创办起，30年来从未间断，它激励和发掘了一批又一批优秀青年致力于翻译事业。从韩奖竞赛中脱颖而出的获奖者已成为各个行业的翻译精英或业务骨干，它在推动我国翻译专业教学、培养选拔高层次翻译人才方面发挥的作用不可小觑。俗语道，“三十而立”，2018年是韩素音青年翻译奖竞赛创办30周年。今年竞赛正式更名为“韩素音国际翻译大赛”，竞赛项目从英语与汉语的互译扩展为英语、法语、俄语、西班牙语、阿拉伯语五种语言与汉语的互译。

本竞赛以其优秀的选材、严格的匿名评审制度、对评审结果的详细解读，而受到来自全国各地及海外青年的积极参与。根据中国翻译协会和广东外语外贸大学的官方统计，本次大赛共收到有效参赛译文10 551份，参赛人数为历届之最。其中英译汉6 571份、汉译英3 106份。参赛选手除来自全国各地高等院校的师生外，还有工作在国家机关、企事业单位的翻译从业人员、自由职业的翻译爱好者，以及来自港澳台地区及海外的翻译爱好者。评审工作采取初评、复评和终评三审制。本届翻译大赛汉译英奖项的分布情况为：一等奖空缺，二等奖6名、三等奖11名、优秀奖87名，其中汉译英四个级别的奖项由104人斩获，占总汉译英有效参赛译文3 106份的3.3%。韩素音国际翻译大赛是目前我国组织时间最长、规模最大、影响最广的翻译比赛。其参赛选手行业领域之多样化以及其获奖之难度，在坊间被誉为“中国翻译界的奥斯卡奖”，因此，本文将以第三十届韩素音国际翻译大赛汉译英竞赛原文、微软机器译本、人工译本及参考译文为语料，重点对机器翻译中行文连贯与逻辑思维问题加以探讨。

2. 研究方法

2.1 语料来源

本研究选取的语料包括第三十届韩素音国际翻译大赛汉译英竞赛原文、微软机器译本、人工译本及在《中国翻译第三十届韩素音国际翻译大赛特刊》上公布的汉译英参考译文。

笔者之所以选取微软机器译本，是因为在2018年3月14日，微软的一组研究团队宣称，其研发的机器翻译技术取得重大突破，研发出了首个能够媲美人工翻译准确度的机器翻译系统，这也就意味着微软翻译系统代表了目前世界上最先进的翻译技术之一。人工译本由笔者提供，笔者亲身参与了第三十届韩素音国际翻译大赛汉译英比赛，译文提交凭据为4135726278，笔者将不对已提交的人工译本做任何的修改与变动。由于此人工译本并未获奖，相对于韩素音杯3.3%的获奖几率，它对人工译者平均翻译水平的代表性很高，用它去比对微软机器译本，鉴别优劣，很可能会得到正确的结论。汉译英参考译文，是在各评审组对大赛原文的内容、形式、语言、风格等进行逐字逐句分析讨论后形成的，之后又进行数次修改、完善，并由国内一流的翻译界专家教授欧阳利锋、鲍川运、赵军峰三位联合执笔，更增加了参考译文的信度。笔者将怀着探究的心态将微软机器译本、人工译本比照参考译文作为本研究的语料。

2.2 研究工具与程序

本文首先提取第三十届韩素音国际翻译大赛汉译英竞赛的原文，输入微软机器翻译软件，然后保存其机器译本，笔者将不对机器译本做任何的编辑与修改，完全保留其最原始的风貌，然后比对获取的人工译本，以参考译文为参照物，从词汇、语法、句式层面进行分析，建立错误分类标准，以此为基础，在语篇层面重点研究机器翻译中凸显的行文连贯与逻辑思维问题，并得出结论。

2.3 文本解读与结果分析

本文对微软机器译本、人工译本及参考译文将从词汇、语法、句式、语篇层面进行标注与分析，把机译常见错误分门别类并明确其成因，以期为提高机器翻译译文质量提供参考与

借鉴。需要说明的是，有的例子中机器译文的错误不止一处，但我们仅标注出涉及所述错误类型的地方，以提高文本解读的效度。

2. 3. 1 词汇层面

例1竞赛原文：地名中的传统文化含量

微软机译：Traditional cultural content in place names

人工译本：Traditional Culture Embodied in Geographical Names

参考译文：Place Names and Traditional Chinese Culture

这是文章的标题，选自清华大学历史系张国刚教授写的一篇小品文，阐述了地名和中国传统文化的关系。微软机译使用了“content”一词表达“含量”，这种词在汉语中叫做范畴词（category word），是一些具有概括意义，表示人或事物分类范畴的词语，例如“情景、态度、局势”等，常用在具体词的后面，对其进行分类总括，到了英语里，这些具体词语直接就能构成搭配关系，而原文中的范畴词便成了冗赘。人工译本为“地名中所展现或蕴藏的传统文化”，而参考译文则认为，作者在文中表达的是中国地名和传统文化的关系，应该当并列结构看比较合适，故译为“Place Names and Traditional Chinese Culture”。

例2竞赛原文：年轻的时候读书，有机会到河南、陕西、山西、甘肃等地考察……

微软机译：When I was young, I had the opportunity to study in Henan, Shaanxi, Shanxi, Gansu and other places…

人工译本：When I was young, I kept reading and got chances to visit places like Henan, Shaanxi, Shanxi, Gansu and so forth.

参考译文：When I was a college student, I had the chance to visit places in Henan, Shaanxi, Shanxi, Gansu, and other provinces.

此例中“考察”本意既指“实地观察调查”，也可指“细致深刻地观察”，一般可以译为“inspect; make an on-the-spot investigation”，但此处作者表达的是“年轻的时候读书，有机会到过河南、陕西、山西、甘肃等地”，是一位教授追忆自己的青葱岁月时，使用的一种轻松愉悦的口语体。此处用“visit”表达“考察”，在翻译技巧上属于一种词语的淡化法或虚化法（understatement）。如果将“考察”译为“inspect”在语体色彩上显得过于正式，若译为“make an on-the-spot investigation”不但非常正式，语义还过于丰满，这势必造成译语冗长、行文阻滞，译文的可读性降低；微软机译用“study”也有些过于明直，有简单粗暴之嫌。人工译本和参考译文均采用了“visit”一词，以增加词语翻译的可读性，译得不多不少，恰到好处。

例3竞赛原文：……油然升起一种穿越历史时空的苍茫感。

微软机译：… and oily raised a sense of crossing the historical time and space. China’s long history and culture seem to be around, not so far away.

人工译本：… I couldn’t help sighing with deep emotion at the vast expanse of the history in which I had transcended time and space.

参考译文：… I had the feeling of travelling back through time and space, as if centuries of Chinese history and culture were coming to life right before my own eyes.

此例中“油然”形容“思想感情自然而然地产生”，微软机译用了“oily”，这是典型的不求甚解，望文生义。由此可以看出翻译的译前准备工作是极其必要的，理解是表达的前提，表达是理解的延续和归宿。

放眼微软机译版全文，其措辞方面的失误不胜枚举，例如“给自己新的居住地，安上故乡

的旧地名”用“install”表达“安”，根据《牛津高阶英汉双解词典第（8）版》，“install”指的是“to fix equipment or furniture into position so that it can be used”，在文中“安上”是指“用故乡的旧地名给自己新的居住地命名、起名字”；还有诸如“联合国的非物质遗产申遗，大约就是鼓励这种行为”把“行为”译为“behavior（the way that sb. behaves, especially towards other people）”，而原文意思是说“联合国的非物质遗产申遗，大约就是为了鼓励那种想方设法地保存历史记忆的做法。”笔者认为用“practice（a way of doing sth. that is the usual or excepted way in a particular organization or situation）”更确切一些；把“阳历新年”译为“the Gregorian calendar（格列高利历）”，因为较生僻，更是让读者一头雾水，不知所云。因此，翻译中应特别重视选词，要对原文和译文所用词语在词义的范围大小、轻重缓急、分寸、语体色彩、感情色彩及搭配习惯方面仔细琢磨，反复推敲，力求译文用词精当、贴切，这一切显然是目前微软机器翻译达不到的。

2. 3. 2 语法层面

例4竞赛原文：读万卷书，行万里路。

微软机译：Read thousands of books and walk miles.

人工译本：As the saying goes, “It is of equal importance to travel ten thousand miles as to read ten thousand books”.

参考译文：People often say, as in a well-known Chinese saying, that “to be a learned person, you must read ten thousand books and travel ten thousand miles”.

例5竞赛原文：于是，设置了很多侨置的州县，沿用北方故土的名字。

微软机译：As a result, set up a lot of overseas Chinese counties, follow the name of the northern native land.

人工译本：Consequently, they established many “overseas” provinces and counties, and named them by borrowing the old geographical names from their hometown in the North.

参考译文：… they used place names from the north for their new settlements, i.e. counties and prefectures that were established for the immigrants, as a way to refresh memories of their former hometowns…

例6竞赛原文：值得我们反思。

微软机译：Worthy of our reflection.

人工译本：All these questions call for our reflection.

参考译文：Not available（漏译或略译）

例4、例5及例6被放在一起讨论，是因为其都属于微软机器译本在语法层面的错误，而且其错误极其类似，都在于缺少了主语。汉语句子的根据在语义，所以汉语句子被称为“意合型”或“语义型”。由于汉语的“意合”特点，其主语可由诸多不同类别的词语充当，主语隐含不显或无主语句的情况时常可见；而英语句子的根据在语法，所以英语句子被称为“语法型”或“形合型”。“语法型”的英语句子必须含有具备一致关系的“主——谓”结构，英语句子的主语和动词不可或缺，具有这种主谓结构的句子，才给人一种语法的完整感。否则，即便表达了明确的意思，也不能算作句子，不能加句点。刘宓庆先生也认为，任何英语语句都可以完形为主谓统携全句的基本态势。在例4、例5及例6竞赛原文中，在上下文语义关系明确的前提下，汉语中主语隐含或省略了。译成英语，则必须依照英语的规则，将主语增补起来。例如人工译本和参考译文都增补了相应的主语，如“As the saying goes/ People often say, as in a well-known Chinese saying”、“they”以及“All these questions”。

例7竞赛原文：又如，县级机构的设立是秦汉创建郡县制国家体制的基础要素之一。

微软机译：For example, the establishment of county-level institutions is one of the basic elements of the national system for the creation of county system in Qin and Han dynasties.

人工译本：Then let me take another example, the establishment of the county-level organs is one of the essential elements in setting up the prefecture-county system as the state system during the Qin and the Han dynasties.

参考译文：Take the county-level division as another example. It was the foundation for the system of prefectures and counties established in the Qin and Han dynasties.

此例微软机译在语法层面的错误较单一，在“秦汉（朝代）”前漏掉了定冠词“the”。按照绝大多数语法书的分类，汉英词类的数量大致相等，类别基本上对应。两种语言词类划分最显著的不同在于英语的冠词和汉语的量词分别为各自所独有，而没有直接对应。译成英语，在表示“历史朝代、宗教事务、学校、星座、船只、舰队、铁路、飞机、三军”等的名词前必须加上定冠词“the”，如“宋朝the Song Dynasty”、“春秋时代the Spring and Autumn Period”、“石器时代the Stone Age”等。

2. 3. 3 句式层面

例8竞赛原文：地名是故乡的第一记忆。

微软机译：Place names are the first memories of hometown.

人工译本：Geographical names are the quickest reminders of hometown.

参考译文：A name is the first thing people remember about their hometowns.

此例中，乍一看，微软机译似乎没有问题，“Place names”、“are”、“the first memories”从形式上貌似达到了“语法型”的英语句子对主谓一致、单复数一致的要求。实则从句式层面分析，其英文表述由于套用母语模式或规则而产生了错误的或不合适目的语的形式，即负向迁移（negative transfer），亦称干扰（interference）；人工译本采用了“the quickest reminders”，至少呼应了后文“用地名保留故乡记忆是人类共同的感情”，只要人们一看到或听到故乡的地名，就能立即想到自己的故乡；参考译文是在透彻理解原文意思的基础上，跳出了原文结构，通过意译传达了原文的真谛。

例9竞赛原文：对于各地的历史地名，进行一次拨乱反正，也许正当其时。

微软机译：For historical place names everywhere, make a good point, perhaps just in time.

人工译本：It is probably the right opportunity to make corrections to the names of historical places throughout the country and eventually bring them order out of chaos.

参考译文：Now is the time to give these place names their due!

此例中，不考虑语法问题，从句式层面看，微软机译文本不周密、不严谨，似乎“直译”出了原文，但表意不清，易给读者造成误解；人工译本虽然表意完整，但比对参考译文，确实显得累赘冗长（redundant）。参考译文的译者在深谙两种语言和文化的基础上，简洁地译为“give sb. his due”，非常忠实地再现了原文的蕴含意义，真正做到了达意传神。

3. 机器翻译中行文连贯与逻辑思维问题

以上部分从词汇、语法、句式三个层面探讨了微软机器译本出现的问题，以下部分将从语篇层面，重点对微软机器翻译中行文连贯与逻辑思维问题加以探讨。行文连贯与逻辑思维问题相辅相成，要做到行文连贯，就必须运用逻辑思维，注意逻辑关联，有时候适当加词，

以保证思路的连续和线索的清晰；用好了逻辑思维，才能透过现象看本质，才能从整体上把握弥漫在篇章中的作者的真实意图，通过调整语序，调节句式结构等手段用适当的方式翻译，才能避免句际衔接突兀，上下文不贯通。逻辑明了了，译文自然行文连贯。

例10竞赛原文：时间是流动的，“逝者如斯夫，不舍昼夜”。空间却相对的固化，“人生代代无穷已，江月年年只相似。”（张若虚《春江花月夜》。

微软机译：Time is flowing, “lament, generate electricity around.” Space is relatively solidified, “life generation Infinite has been, the river month is only similar.” (Zhang Ruoxu “Spring River Flower Moonlight”).

人工译本：Time is flying, as depicted in “The passage of time is just like the flow of water, which goes on endlessly day and night.” while space stands relatively still, as portrayed in the poem “Life has been infinite in a ceaseless succession generation after generation, whereas the moon at the riverside looks quite alike year by year (from Zhang Ruoxu’s At the Moonlit Riverside in Spring Blossom)”.

参考译文：Time flows, *“passing on just like a river, day and night”*, as observed by Confucius, a great Chinese philosopher. In contrast, space is static, as described by Zhang Ruoxu, a renowned Tang dynasty poet, in his poem *Moonlight, River, and Flowers: “One generation goes and another generation comes, but the moon and the river remain all the same”*.

此例中，汉语原文没有明确表示出语义对比，只简单用了一个“却”字，这是汉语语法隐含性（convertness）的一种表现。除去两句古文的翻译失误、英译标点符号滥用及英文大小写混乱外，微软机译显然还是对字面意义的简单堆砌，语篇不连贯，逻辑性不强，译文句式简单呆板，缺乏变化，因而表现力不佳。人工译本就注意到了这种结构衔接，运用了对照（contrast）的修辞手段来实现“Time is flying while space stands relatively still”。参考译文则明确使用“in contrast”突出了对比的意味，这也是英语的语法呈外显性特征（overtness）的表现。

例11竞赛原文：书中所见是历史的时间地图；路中所见是历史的空间地理。

微软机译：What is seen in the book is a time map of history; What is seen in the road is the spatial geography of history.

人工译本：What’s discovered in books is the timeline for when all the events happened in history, while what’s exposed to during the journey is the spatial geography for where all the events occurred in history.

参考译文：When we read, we learn history as it evolves in time, and when we travel, we get to know history as it unfolds in space.

此例中，微软机译较完整地传达了原文字面的意义，句子也合乎英语语法规则，但通篇不见任何句际衔接手段，结果就是语篇不连贯，逻辑性不强，读起来给人一跳一跳的感觉，译文可读性较差。人工译本借助逻辑思维，根据上下文语境，在深入分析的基础上，认为“书中所见是历史的时间地图；路中所见是历史的空间地理”指的是“我们通过读书（read），可以了解历史事件发生的时间（time/when），而通过行路（travel），可以到达当年历史事件发生的空间（space/where）”，这正与上文提到的“历史有时间与空间这两个维度”、“读万卷书，行万里路”相吻合。比对参考译文，译者通过适当加词“when we read”和“when we travel”保证了思路的连续和线索的清晰，达到了篇章衔接、语意连贯的效果。参考译文不但译出了原作的意图，而且保留了原作的对偶形式，达到了内容和形式的统一，堪称一例妙译。

4. 结论

本文以第三十届韩素音国际翻译大赛汉译英竞赛原文、微软机器译本、人工译本及参考译文为语料，从词汇、语法、句式层面对文本进行了标注与分析，并在语篇层面重点对机器翻译中行文连贯与逻辑思维问题进行了探讨，并得出如下结论：

（1）在词汇层面，微软机器译本的主要失误在于机械地“直译”原文，措辞单一、无变化，如表达“其实安徽最深刻的记忆是‘徽’”、“而‘秦’恰恰是陕西最深刻的记忆”，都用“the deepest memory”来表达，忽略了英语是一种忌重复（redundancy）的语言,措辞必须多样化；而且，微软机器译本措辞中规中矩、呆板、无生气，不能生动逼真地传达原文，如“欧洲人到了美洲”，只用“the Europeans went to America”表达,其效果远不及人工译本和参考译文用“the Europeans made their way to America”、“the Europeans found their way into America”；甚至个别词语翻译不求甚解、望文生义，如前文所述把“油然”译为“oily”等。

（2）在语法层面，微软机器译本最明显的失误在于对汉语无主句的处理。在上下文语义关系明确的前提下，汉语中无主语的句子比比皆是，而译成英语则必须增补主语，以满足英语语法对形式的要求，这就要求译者必须运用逻辑思维，在透彻理解原文的基础上增补主语。增补主语的原则，一是要推敲语境，二是要考虑英语语法习惯和行文需要。由于机器翻译并不会思考，所以在汉语主语隐含或省略的情况下，如何增补恰当的主语，是目前机器翻译需要克服的难题之一。

（3）在句式层面，微软机器翻译还是倾向于“直译”，对于简单句、并列句的把控较强，但对于翻译复合句，特别是一些用来阐述复杂的情感或思想，或一些逻辑严密、语义复杂的长句时，其翻译效果极其不理想。机器翻译系统由于无法厘清复合句内在的逻辑结构，所以不能对长句进行正确地切分，因而，也不可能采取相应的翻译策略，如“拆分法”来翻译长句。有的即使生硬译出，但由于表意不清，易给读者造成误解，译文可读性差。

（4）在语篇层面，由于微软机器翻译系统没有做“译前准备”，所以它虽然能基本上传达原文字面的含义，但由于缺少了逻辑介入，词句之间在语言形式上不具备粘连性（cohesion），在语义上欠缺连贯性(coherence)，语篇行文连贯性较差，译本逻辑思维含混不清，译文显得不容易理解，可读性差。

由此可以看出，微软研究团队宣称的其研发出的首个能媲美人工翻译准确度的机器翻译系统还有待于进一步优化，特别是在语篇层面，微软机器翻译系统如何运用其“逻辑思维”介入原文，理解原文、揣摩作者真实意图，在深入其“境”中，得其“境”之妙，捕捉到“弦外之音”，如何译出原作语义内涵，体现原作作品风格等都是当前微软机器翻译系统需要克服的难题。

致谢

本论文受2017年陕西省社会科学基金项目：“英、汉、西语商务语篇话语策略的跨语言、跨文化研究”（项目编号为2017K019）；2018年西安外国语大学科研计划项目：“中国政府记者招待会的跨文化语用差异分析：以”一带一路“相关议题为例”（项目编号为18XWD02）；2018年陕西省教育厅科研项目：“基于教材语料库的商务英语词汇计量研究”（项目编号为18JK0638）资助。

作者简介：

杨敏 (1979--) 陕西汉中人。现为西安外国语大学英语实验基地在职教师，讲师职称，美国加州洪堡州立大学访问学者。本科毕业于西安外国语大学，英语教育专业；硕士毕业于西北工业大学，外国及应用语言学专业，研究方向：二语习得、英汉互译。联系电话：13379013191；电子邮箱：yangmin1979@163.com；通讯地址：陕西省西安市友谊西路376号自在广场，邮编

710068。

References

[1] Ouyang Lifeng, Bao Chuanyun, Zhao Junfeng, On the Translator's Logical Intervention and Textual Coherence—Analysis of the Chinese-English Translation Texts Submitted to the 30th Han Suyin International Translation Contest [J]. *Chinese Translators Journal* (Special Issue of the 30th Han Suyin International Translation Contest, 2018), pp. 56-65, 2019.

[2] Wang Zhikui. *A Collegiate Course in Chinese-English Translation (4th Edition)* [M]. Jinan: Shandong University Press, 2005.

[3] A. S. Hornby. *Oxford Advanced Learner's English-Chinese Dictionary (8th Edition)* [Z]. Beijing: The Commercial Press and Oxford University Press, 1997.

[4] Du Ruiqing. *New Century Chinese-English Dictionary (2nd Edition)* [Z]. Beijing: Foreign Language Teaching and Research Press, 2016.

[5] Chen Hongwei, Li Yadan. *A New Coursebook on Chinese-English Translation (2nd Edition)* [M]. Shanghai: Shanghai Foreign Language Education Press, 2013.

[6] Sun Yiqun. An Analysis of Mistakes on Computer-Assisted Abstract Translation of Material Science [J]. *Chinese Science & Technology Translators Journal*, Vol. 31, 2018 (3).

[7] https://cn.bing.com/translator/