

一个关于机器翻译的史料错误

冯志伟*

1

最近在《中文信息学报》第二十一卷第三期上看到了张钹院士的《自然语言处理的计算模型》一文，很受启发。张钹院士是清华大学资深教授，是著名的信息处理专家，是我们尊重的学术前辈，他在这篇明确指出自然语言处理是一个“强不适定问题”（Strongly Ill-posed Problem），只有通过大量的“约束”，才能使之成为“适定性的”、“可解的”问题。张钹院士的这个见解，不论在语言学上还是在计算机科学上都是非常深刻的，他的论证也是很有道理的。但是，在这篇文章中张钹院士引用了如下的材料却是错误的。张钹院士引用的材料如下：

“在机器翻译研究的初期（上个世纪60年代），美国人经常举以下的例子来说明机器翻译任务的艰巨性。

英文的原句是：

(1) The spirit is willing but the flesh is weak. (心有馀而力不足)

经过机器翻译成俄文之後（在文法分析、双语词典等支持下），再把它翻译回英文，得到的结果如下：

(2) The Voltka is strong but the meat is rotten. (伏特加酒是浓的，但肉却腐烂了)

这也许只是一则笑话，可是它充分说明了自然语言处理的困难所在。”

接着，张钹对于这个例子进行了分析，作为文章立论的重要材料。

据我所知，张钹院士引用的这些材料最早出自美国1962年8月号的《哈泼杂志》(Harper's Magazine)，这期《哈泼杂志》发表了题目为《翻译的困扰》(The trouble with translation) 的文章，作者叫做古温豪芬 (John A. Kouwenhoven)。这篇文章中编造了如下的故事：

有几个电子工程师设计了一部自动翻译机，这部机器的词典包含1500个基础英语词汇和相对应的俄语词汇。他们宣称这部机器可以马上进行翻译，而且不会犯人工翻译的错误。第一次试验时，观众要求翻译“Out of sight, out of mind” (眼不见心不烦) 这个句子，灯光一阵闪动之後，翻译出来的俄语句子的意思竟然是“看不见的疯子” (Invisible

* 冯志伟先生，教育部语言文字应用研究所，北京。

idiot)。他们觉得这样的谚语式的句子比较难于翻译，于是又给机器翻译翻译另一个出自圣经的句子“The spirit is willing, but the flesh is weak”（心有馀而力不足），机器翻译出来的俄语句子的意思却是“酒保存得很好，但肉已经腐烂”（The liquor is holding out all right, but the meat has spoiled）。

后来，在美国的某些出版物中，对于 John A. Kouwenhoven 编造的这个故事又做了添油加醋的如下描述：

美国的机器翻译系统曾经把英语翻译成俄语，然后又把俄语翻译成英语，最后翻译出来的英语是语无伦次的，结果闹出了大笑话。机器把英语的成语 The spirit is willing, but the flesh is weak（心有馀而力不足）翻译成俄语之后，这个句子的意思却变成了 The whiskey is alright, but the meat is rotten（威士忌没有问题，而肉却烂了）；机器把英语成语的 Out of sight, out of mind（眼不见心不烦）翻译成俄语之后，这个句子的意思变成了 Invisible idiot（看不见的白痴）。

显而易见，所有这些描述，都来自 John A. Kouwenhoven 编造的故事。

张钹院士在他的文章中，没有引用“Out of sight, out of mind”（眼不见心不烦）这个谚语式例子，而只引用了出自圣经中的“The spirit is willing but the flesh is weak”（心有馀而力不足）这个例子。

这样的故事显然是 John A. Kouwenhoven 凭空捏造的。稍微熟悉机器翻译发展历史的学者都知道，在机器翻译的早期（20世纪50-60年代），出于冷战的需要，美国要了解苏联的情况，急需把俄语翻译为英语，在那个时期，美国只研究过把俄语翻译成英语的机器翻译系统，并没有研究过把英语翻译为俄语的机器翻译系统，因此，在1962年，美国根本不可能使用计算机把英文句子翻译成俄文。这样荒谬绝顶的机器翻译例子显然是 John A. Kouwenhoven 本人挖空心思地捏造出来的，其目的在于攻击机器翻译，煽动人们对于机器翻译的不满情绪。

也许张钹院士没有考察他引用的例子的出处，也没有注意到当时美国并没有研制英语到俄语的机器翻译系统的历史事实。但是，从史料的来源上看，张钹院士引用的例子，最早是 John A. Kouwenhoven 挖空心思地捏造出来的。这样的史料是虚假的、是错误的。

应该指出，John A. Kouwenhoven 挖空心思地捏造的这些例子在我国其他学术著作中也引用过。例如，在石纯一、黄昌宁等编著的《人工智能原理》（清华大学出版社，1993年版）第3页，就引用了 John A. Kouwenhoven 挖空心思地捏造的“Out of sight, out of mind”（眼不见心不烦）这个例子，并且对于这样捏造的例子从正面给予肯定性的评价和引申。此后，在我国的一些学术著作中就不断地引用这些错误的例子，以至于像张钹院士这样的资深教授也难辨是非，上了 John A. Kouwenhoven 的当。

尽管 John A. Kouwenhoven 捏造事实对机器翻译进行了别有用心攻击，在美国学术界，也有正直的学者对于机器翻译的问题进行过严肃的思考和科学的探索。

早在机器翻译刚刚问世的时候，美国著名数理逻辑学家巴希勒 (Bar-Hillel) 在1959年就指出，“全自动高质量的机器翻译”(Fully Automatic, High Quality, MT, 简称 FAHQMT) 是不可能的。Bar-Hillel 说明，FAHQMT 不仅在当时的技术水平下是不可能的，而且，在理论原则上也是不可能的。他举出了如下简单的英语片段，说明要在上下文中发现多义词 pen 的正确译文是非常困难的事情。

John was looking for his toy box. Finally he found it. The box was in the pen.

John was very happy.

他的理由如下：

1. pen 在这里只能翻译为 play-pen（“游戏的围栏”），而绝对不能翻译为书写工具“钢笔”。
2. 要确定 pen 的这个正确的译文是翻译好这段短文的关键所在。
3. 而要确定这样的正确译文依赖于计算机对于周围世界的一般知识。
4. 但是我们没有办法把这样的知识加到计算机中去。

在机器翻译的早期，Bar-Hillel 就科学地预见到了机器翻译将会遇到的困难，显示了他的远见卓识。如果张钹院士注意到 Bar-Hillel 的这些研究，引用 Bar-Hillel 的这个例子作为他的论据，他的论文一定会更加具有说服力。

本人学识浅陋，不妥之处，请方家指正。

□