

仿人汉英口语机译系统

王永成

(上海交通大学电脑应用技术研究所)

改革开放,国内外交流日趋频繁,汉外对译需求已成了推动机器翻译研究的强大动力。据报道:目前世界语言有5651种,3000多种文字,500多种有影响的语言,但是,其中只有十二种最受到重视。而联合国的工作语言也只有英、法、西、俄、中、阿六种。虽在一次大战前,法语是世界通行的工作语言;一次大战后法语与英语已经并驾齐驱;二次大战后则随美国国力的增强,美国英语则成了霸主。据报道:虽然目前世界上还只有10%的人日常皆使用英语,但是,国际文献中恰有的80%的用英语(另外还有15%的用法语,4%的用西班牙语,1%的用俄语),国际邮件中恰有70%的用英语。因此,作为以世界上使用人数最多的中国人,抓紧汉英对译的研究与系统的开发,乃具有极大的科学与实用意义。

几十年从事电脑软件研究与开发的实践曾告诉我们:“仿人算法是研制开发电脑上高级算法的捷径”(虽然它不一定是最好的,但往往是最容易找到的较好的算法)。

“仿人机译”算法的核心就是模仿人的翻译方法。人在翻译过程中常用的手段有:

- 1) 对很多的常用的简单句,最常用检索来代替翻译。
- 2) 对常用句,可对译句进行模糊匹配。
- 3) 在常用句句型不变时,应能进行部分词或句意辞的置换。
- 4) 翻译的过程又是一个不断将有关的翻译过程转化为检索的过程。
- 5) 翻译不了的少数词(如人名、地名及一些专有名词)可记下原文并给出其译音。
- 6) 对一时难用检索代翻译的复杂句或歧义句,可借助于语法分析或语义分析、上下文理解来协助翻译或给出多种翻译方案(可按使用或然率排序)提供用户选择。

仿照这些做法,我设计了所谓的仿人机译算法的实现流程图,从1992年开始就让研究生先后对“中英文常用口语对译系统”进行了试开发,并首先从开发者较少的汉译英开始,现已获得了一定的成功。易顺明等的研究开发报告讲:对英语中的一万多常用句(库中句)已能做到句句“翻译”得快速、准确、优美;对测试的非库中句也有80%的正确率。其实,因为对前者它实际上不是翻译而是检索现成的答案,只要检索系统成熟,结果当然是必然的;对后者,因其时间与条件的限制,加之程序还待大大地改进,不仅还离理想的状态远,而且对现有统计结果也有一定的偶然性。

The Imitating Human Oral Chinese-English MT System

Wang, YongCheng

(The Application Technology Institute of Shanghai Jiao-Tong University)

Abstract In this paper, the basic idea of so-called "Imitating human MT algorithm" and some realize produced in developing the model system process are introduced.