

# 句型转换式汉英机译的一些技术处理探讨

杨宪泽

(西南民族大学计算机科学与技术学院, 成都 610041)

**摘要:** 句型转换式机器翻译涉及的技术很多, 主要工作有四部分: 第一部分给出一种自动分词边界问题处理方法; 第二部分探讨一般句与疑问句的句型结构; 第三部分研究汉英机器翻译的译文生成问题; 第四部分是一种“是”动词“am; is; are”的处理方法。

**关键词:** 机器翻译; 边界问题; 句型转换; 译文生成

中图分类号: TP391.2

文献标识码: A

文章编号: 1003-4271(2014)02-0271-03

对于两种不同的自然语言(例如英语和汉语)句子在计算机上用程序实现自动翻译, 这就是机器翻译。机器翻译研究课题除了其相当高的理论价值外, 还由于巨大的社会效益和经济效益, 被确定为21世纪要解决的科技难题之一<sup>[1-2]</sup>。当然, 研究目前还面临许多困难, 因此, 各个方面的探讨就显得很有意义。

在计算机上用程序实现机器翻译目前仅有三种实用方法: 类比机器翻译方法; 概率统计机器翻译方法和转换式机器翻译方法。转换式机器翻译方法的程序处理就是两种语言间单词、短语和句型的匹配, 借助于词法分析、句法分析、语法分析、语义分析、语用分析和语境分析, 最后调序及目标语言的生成完成自动机器翻译。

我们在其它文献中介绍了我们所做汉英机器翻译研究的一些阶段成果<sup>[3-5]</sup>, 本文的工作是进一步研究的探讨。第一部分给出汉语句自动分词边界问题解决; 第二部分探讨一般句与疑问句句型表达式的处理。第三部分研究汉英机器翻译的自动调序问题。第四部分是一种“是”动词“am; is; are”的处理方法。

## 1 汉语句自动分词边界问题处理

在机器翻译过程中, 要把汉语句子翻译成英语句子, 汉语句子首先要自动分词。即把汉语句子切分成一个个单词再作下一步处理。关于自动分词方法, 文献[6]已经做过详细介绍, 本文不再赘述。现在要介绍的是自动分词遇到的边界问题: 文献[6]介绍的分词方法是一种正向分词方法, 其思路是把待翻译的汉语句子取前三个字查找单词库, 若不是一个单词, 则删除三个字的最后一个字再实施查找, 这样一直循环查找, 至找到一个单词为止。待翻译的汉语句子剩余部分重复此工作, 直到把所有的待翻译的汉语句子的词都分出为止。这里可能就出现这样的情况, 自动分词到最后, 待翻译的汉语句子剩下的不是三个字, 算法越界死机。为避免这种情况, 我们作如下处理:

- 步骤 1: 待翻译的汉语句子的字数 $\geq 3$ ? 是, 转步骤 7.
- 步骤 2: 待翻译的汉语句子的字数=2? 是, 转步骤 6.
- 步骤 3: 待翻译的汉语句子的字数=1? 是, 转步骤 5.
- 步骤 4: 已经完全分词, 进入下一阶段处理.
- 步骤 5: 进入查找单字词阶段.
- 步骤 6: 进入查找两字词阶段.
- 步骤 7: 自动分词算法正常运行.

收稿日期: 2014-01-08

作者简介: 杨宪泽(1954-), 男, 四川成都人, 教授, 研究方向: 自然语言处理与数据结构.

基金项目: 国家自然科学基金面上项目(61379019); 中央高校科研平台项目(11NPT02).

## 2 一般句与疑问句句型表达式的处理

汉英机器翻译采用句型转换的方法就是把汉语句子的单词和句型结构映射为英文句子相对应的单词和句型结构,能够程序实现翻译的核心是句型结构的表达式相同.但是,一般句与疑问句的句型表达式可能相同,但它们翻译出的英语句子中的单词顺序却不相同.例如:“你们是学生.”与“你们是学生?”句型表达式都是“pvn”,但译文应该是不同的两种,You are students.和 Are you students? 为此,处理的方式根据待翻译的句子结尾是句号还是问号来区别是一般句与疑问句.即,一般句与疑问句的句型表达式建立在不同的模块中,处理步骤如下:

- 步骤1: 求待翻译汉语句子字符长度,记为x;
- 步骤2: 根据x求出待翻译汉语句子最后一个字符,记为fh;
- 步骤3: fh是“?”? ,是转步骤5;
- 步骤4: 一般句的句型表达式方式处理;
- 步骤5: 疑问句的句型表达式方式处理.

显然,在不同的模块中,一般句与疑问句的句型表达式即使相同,它们的译文输出表达式却不相同.例如,“你们是学生.”与“你们是学生?”句型表达式都是“pvn”,但在不同的模块处理后,译文输出表达式为yy(1)yy(2)yy(3)和 yy(2)yy(1) yy(3).

## 3 汉英机器翻译的自动调序问题

调序指的是要翻译的英语句子与汉语句子单词表示的顺序不同,当然这就需要程序自动处理,完成调序.如果设一个英语单词由yy(i)表示,汉英句型转换的机器翻译句子的顺序有下列情况:

- (1) 要翻译的汉语句子单词顺序与生成的英语句子单词顺序相同.

举例:要把“现在我们详细介绍计算机”,这一汉语句子翻译成英语句子,首先建立句型表达式,zs="dpdvn".其中,第一个d是“现在”这一单词的词性标注符;p是“我们”的词性标注符;第二个d是“详细”的词性标注符;v是“介绍”的词性标注符;n是“计算机”的词性标注符.程序自动调序结果单词顺序任然是:yy(1)yy(2) yy(3)yy(4)yy(5).

- (2)要翻译的汉语句子单词顺序与生成的英语句子单词顺序不同.

举例:要把“计算机对生产过程带来根本变化”,这一汉语句子翻译成英语句子,首先建立句型表达式,zs="njnvan".其中,第一个n是“计算机”这一单词的词性标注符;j是“对”的词性标注符;第二个n是“生产”的词性标注符;第三个n是“过程”的词性标注符;v是“带来”的词性标注符;a是“根本”的词性标注符;最后的n是“变化”的词性标注符.程序将自动调序,结果机器翻译句子单词顺序成为:yy(1)yy(5) yy(6) yy(7)yy(2) yy(3)yy(4).即:Computer brought complete change to production process.

## 4 关于“am; is; are”的处理

在机器翻译的汉译英中,“是”动词会遇到“am; is; are”的选择,人工翻译按照人称规则很容易选择.为使翻译程序实现自动处理,我们设计了如下子算法:

若翻译的汉语句动词为“是”,转入子程序SDCCL  
<SDCCL>

- yxz1: 要翻译生成的英语句“是”相邻单词出现“I”,“是”动词选择“am”,转yxz10;
- Yxz2: 要翻译生成的英语句“是”相邻单词出现“she”,“是”动词选择“is”,转yxz10;
- Yxz3: 要翻译生成的英语句“是”相邻单词出现“he”,“是”动词选择“is”,转yxz10;
- Yxz4: 要翻译生成的英语句“是”相邻单词出现“this”,“是”动词选择“is”,转yxz10;

- Yxz5: 要翻译生成的英语句“是”相邻单词出现“her”, “是”动词选择“is”, 转yxz10;  
 Yxz6: 要翻译生成的英语句“是”相邻单词出现“you”, “是”动词选择“are”, 转yxz10;  
 Yxz7: 要翻译生成的英语句“是”相邻单词出现“they”, “是”动词选择“are”, 转yxz10;  
 Yxz8: 要翻译生成的英语句“是”相邻单词出现“we”, “是”动词选择“are”, 转yxz710;  
 Yxz9: 其它情况, “是”动词使用“is”;  
 Yxz10: 子程序退出.

## 5 结束语

汉英机器翻译的研究和软件开发很难, 主要原因有许多基本问题的研究必须先解决. 例如英语字母大小写译文中怎么加入; 被动语气和主动语气; “the”怎么加入和复数问题等等都是开发软件需要面临的问题. 当然, 最难的还是单词的多义、汉语句自动分词的歧义问题和句子语义问题. 我们充分认识到, 由于不同国家、不同民族的自然语言发展有上千年历史、他们的思维理念和传统的一些文化方式肯定有不同点, 这就使机器翻译两种语言对于许多不同的部分, 不仅仅是单词和短语, 也不仅仅是汉译英, 英译汉也会遇到同样的问题. 例如, How do you do ?

人可以克服这个障碍, 计算机软件怎么办? 如果不特殊处理, 只能翻译成“你怎么样?”这类需要研究和解决的问题不少.

既然机器翻译被列为本世纪要解决的科技难题之一. 我们相信: 只要相关科技人员努力去做, 随着研究的深入, 随着科技的进一步发展, 随着时间的推移, 许多问题都会解决, 机器翻译的最终目标一定会达到.

### 参考文献:

- [1] WANG, CHAO, MICHAEL, COLLINS, et al. Chinese syntactic reordering for statistical machine translation[C]. In Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL), 2007:737-745.
- [2] XIONG DEYI, MIN ZHANG, AITI AW, et al. A linguistically Annotated Reordering Model for BTG-based Statistical Machine Translation[C]. In Proceedings of ACL-08:HLT, Short papers (Companion Volume), USA: Columbus, Ohio, 2008: 149-152.
- [3] 杨宪泽. 汉英机器翻译中短语处理与译文生成研究[J]. 西南民族大学学报: 自然科学版, 2013: 39(3): 456-460.
- [4] 杨宪泽. 一种混合式机器翻译方法的分析研究[J]. 计算机工程与科学, 2012: 34(2): 168-171.
- [5] 杨宪泽. 机器翻译的一些难点探讨[J]. 西南民族大学学报: 自然科学版, 2011: 37(3): 447-451.
- [6] 杨宪泽. 混合式机器翻译中单词处理的一些技术探讨[J]. 科技通报, 2011: 27(1): 101-104.

## Study on some technique processing of Chinese-English syntactic transfer machine translation

YANG Xian-ze

(School of Computer Science and Technology, Southwest University for Nationalities, Chengdu 610041, P.R.C.)

**Abstract:** The syntactic transfer machine translation involves many technologies. This paper comprises four parts. The first part gives a boundary problem treatment methods of automatic words segmentation. The second part studies the general and interrogative sentence construction. The three part is a research on some treatment methods of Chinese-English machine translation target-language generation. The four part presents a kind of processing method of verbs such as *am*, *is*, *are*.

**Key words:** machine translation; boundary problem; syntactic transfer; target-language generation