口语翻译中的问题、方法与应用可能性

宗成庆

中科院自动化所 模式识别国家重点实验室 北京 100080

E-mail: cqzong@nlpr.ia.ac.cn

摘要 本文简要介绍了语音翻译技术的研究现状,分析了该项技术在 10 多年的发展过程中所取得的进展和存在的问题;然后侧重介绍和分析了目前语音翻译技术的基本方法,并对两种主要的实现技术 - 基于规则的分析翻译方法和基于统计模型的翻译方法的比较与结合问题进行了简要的讨论;最后本文介绍了在 973 项目“图象、语音、自然语言理解与知识挖掘”的支持下口语翻译关键技术研究的最新进展情况和取得的初步成果。

1 引言

语音翻译 (Speech-to-speech Translation) 就是让计算机实现从一种语言的语音到另一种语言的语音翻译的过程。其基本思想是,让计算机像人一样充当持不同语言的说话人之间翻译的角色。由于说话人使用的语言一般都是日常生活中的口语,而人们也正希望计算机翻译系统能够接受并实现任意口语语句的翻译,并且,这种希望随着语音识别技术和口语解析技术的快速发展和提高,已经不再是渺茫的设想。因此,现在的语音翻译又常被称作口语翻译 (Spoken Language Translation, SLT) [Kitano, 1994; 宗, 1999]。

集多项技术于一体的语音翻译系统一般由语音识别 (SR)、机器翻译 (MT) 和语音合成 (TTS) 等三个主要技术模块组成。从目前情况来看,尽管语音合成技术已在很多应用系统中得到实际应用,但语音识别技术和机器翻译技术则更多地停留在实验室研究阶段。因此,作为多技术集成的口语翻译系统仍然处于“艺术级的”(state-of-the-art) 研究现状也就不足为怪了。我们不回避在口语翻译研究中的实际困难,无论是语音识别,还是口语解析和机器翻译,都存在许多有待于深入研究的关键问题,但我们认为,一项新的技术和发明从实验室走向实用化是逐步完成的,是需要一个过程的,而不是等到这项新技术在理论上完全成熟后才寻找其实现可能和应用目标的。因此,我们认为目前的语音翻译技术之所以仍处于纯理论研究阶段可以归结为两个主要因素:一方面,相关技术本身确实不够成熟,许多理论问题和 技术实现问题都有待于进一步探索;另一方面是人的配合问题。无可厚非,作为用户,人们总是期盼理想化的、高性能的、可以广泛应用和操作方便的口语翻译系统,而且希望系统能够像人一样实现高正确率的自动翻译。而实际上这是不现实的,因为人在做翻译时要做到完全 100% 的翻译正确也并不是轻而易举的事情。那么,目前语音翻译的思想方法究竟存在什么样的问题?如何建立语音翻译的新方法?如何在目前这样一种状态下,尽早实现口语翻译的实用化?这就是我们本文想讨论的问题。

本文第二部分将简要介绍几个代表性的语音翻译实验系统,并分析该项技术在近 10 多年的发展中取得的进展和存在的问题;第三部分侧重分析目前语音翻译技术的基本方法,并对分析方法和统计方法的比较与结合问题进行简要的讨论;第四部分介绍本 973 项目“图象、语音、自然语言理解与知识挖掘”中口语翻译研究的进展情况;第五部分是本文的结束语。

2 语音翻译研究现状

世界上第一个语音翻译实验系统 SpeechTrans 于 1989 年由美国卡内基 - 梅隆大学 (Carnegie Mellon University, CMU) 开发成功[Kitano, 1994]。在过去的十多年里,尤其是近几年,随着相关技术和学科的迅猛发展,一批针对不同应用领域的语音翻译实验系统相继问世。表 1 列出了十几年来一些有代表性的语音(口语)翻译系统。实际上,除了表 1 中列出的代表系统以外,还有 AT&T Bell 实验室开发的 VEST (Voice English / Spanish Translator)系统[Roe et al., 1992], SRI International 开发的 SLT 口语翻译系统[Rayner et al., 1995; Agnas, 1995], 以及其它一些小型的语音翻译实验系统,这里不再一一赘述。另外,表 1 中的 JANUS-III 是指 JANUS 系统的第三个版本,实际上, JANUS 系统分为三个不同的版本,分别称叫作 JANUS-I [Waibel et al., 1991]、JANUS-II [Waibel, 1996] 和 JANUS-III¹。

系统名称	代表机构	研制时间	应用领域	翻译语种	翻译方法	识别词汇
SpeechTrans	CMU	1989	医生与病人对话	日-英	基于规则	-
SL-TRANS	ATR-ITL	1989	ATR 会议注册	日-英	基于规则	1035
JANUS-III	CMU	1997-	旅馆预订, 航空/火车订票, 旅游信息查询等	德、英、日、西班牙、韩、俄等	中间语言	开放
ATR-MATRIX	ATR	1998-2001	旅馆预订	日-英、韩等	基于事例	2000
Head Transducers	AT&T Labs	1996	航空旅游信息	英-汉/英-西班牙	基于统计	1200/1300
Verbmobil	Karlsruhe 大学、DFKI 等	1993-2000	会晤日程安排	德、英、日等	多策略结合	10000/2500
LodeStar	CAS-NLPR	1999	旅馆预订	汉-日、英	多策略	2000

表 1 部分语音翻译系统

我国四川大学也曾于 1990 年左右研究开发了一个面向航空订票和信息查询领域的英汉语音翻译实验系统,限于当时的条件,系统可处理的词汇量只有 150 个英语单词,21 种句型,而且只能处理特定讲话人的规范语句[杨家沅等, 1992]。中国科学院自动化研究所模式识别国家重点实验室(NLPR)自 1996 年以来就开始口语处理技术的研究,并成为国际语音翻译研究联盟(Consortium for Speech Translation Advanced Research, C-STAR)的联系成员(Affiliate member), 2000 年 10 月正式成为 C-STAR 核心成员(Partner), 并作为我国唯一的核心成员参与 C-STAR 七国语言(英语、日语、德语、法语、意大利语、韩国语和中文)的语音翻译联合研究。在近几年中,先后开发成功中日口语翻译实验系统和汉英口语翻译实验系统,2000 年与日本松下电器产业株式会社先端技术研究所(京都)合作,成功地开发了国际上第一个面向旅游领域的中、日、英多语言语音翻译机原型系统。2002 年与韩国 ETRI 合作联合开发了基于普通手机的中韩语音翻译演示系统,该系统于 2002 年 3 月在 C-STAR 研讨会上成功地演示。

另外,中国科学院声学研究所、东北大学、哈尔滨工业大学和当时的先进人机通讯技术联合实验室(JDL)等单位也在该技术领域进行了富有成效的探索。

从国内外总的研究现状来看,语音翻译研究取得了长足的进步,关于其进展和目前存在的问题我们已在很多文章中有过详细的论述[宗, 1999; 宗, 2002a], 这里只是简单地归纳一下。语音翻译的主要进展主要体现在如下几个方面:(1)系统翻译的词汇量越来越大,输入语音逐步逼近口语化;(2)系统采用的翻译策略已经从单一的翻译引擎转向多翻译引擎的

¹ 参阅: <http://www.is.cs.cmu.edu/mie>

协作方式;(3)翻译语言从双语对开始扩展到多语言之间的互译,限定领域也从单一领域转向多领域。(4)语音识别的正确率和鲁棒性(Robustness)在过去的10多年里也有了很大的提高。

语音翻译中的困难也是多方面的,归结起来,我们将这些问题重点归咎于两个方面:一是语音识别的问题,从语音识别技术的现状来看,无论是语音识别器的正确率还是鲁棒性都远远不能达到真正实用化的程度。二是口语翻译的方法问题。在目前的语音翻译研究中,人们普遍采用的方法是将一个语音翻译系统划分成三个技术模块(SR, MT, TTS)分别实现,然后再统一集成,而实际上这种实现方法与人在完成对话翻译时的过程并不完全一致。尤其是机器翻译模块中所采用的翻译策略大都来自文本机器翻译,因此其翻译正确率和翻译模块的鲁棒性大打折扣。有人曾提出了“无缝联结(seamless connection)”和“快而粗(quick and dirty)”的实现方法[Seligman, 2000],但是如何实现真正的“无缝联结”,如何衡量“无缝”的标准,都有很多问题值得深入探索。

总的来说,目前的语音翻译研究中成绩与问题并存,一方面许多理论上的和实现技术上的问题都有待于长期的深入研究和探索,另一方面,我们认为在现有的技术状况下,针对具体的应用目标并确立适当的处理策略,开发实用的口语翻译系统仍然是可能的和可行的。

3 语音翻译方法

根据目前国际上研究开发的一些口语翻译实验系统所采用的不同的翻译机制,我们将其划分为三种主要类型[Zong et al., 2002]:传统翻译方法、非传统翻译方法和人机互助的混合翻译方法。

3.1 传统翻译方法

所谓的传统翻译方法主要指如下四种基本的翻译方法:(1)基于规则(rule-based)的语音翻译方法;(2)基于事例(example-based)和模板(template-based)的语音翻译方法;(3)基于中间语义表示的(inter-lingual)语音翻译方法;(4)基于统计模型的(statistical)翻译方法。由于这四种翻译方法是在文本机器翻译中提出来的,并切有的已被广泛地应用于商业化翻译软件和实验系统中,因此,我们将其称之为主流的翻译方法。关于这些方法的优缺点已经有很多论著中给予了评述,这里我们不再赘述。

3.2 非传统翻译方法

所谓的非传统翻译方法是指除上述四种基本的方法以外,近几年来提出和被尝试的其它各种新的翻译策略。尽管这些翻译方法不像传统的翻译方法那样被广泛地承认和采用,但它们毕竟是众多的研究者向着实用、高效的翻译目标逼近的见证,而且其中的许多思想和策略同样值得我们借鉴。我们将这些翻译方法称之为非主流的翻译方法。

1997年Wakita等人曾提出了一种局部抽取可正确翻译成分的语音翻译方法[Wakita et al., 1997],其基本思想是以基于事例的翻译方法为基础,翻译模块首先对输入语句进行句法和语义解析,然后在解析结果中寻找那些具有一定的长度(词汇达到一定的个数)、结构上符合句法、与事例库中某一例句(或片断)的语义距离小于给定门限的那些片断或语块,最后通过基于事例的翻译方法只翻译那些符合条件的片断或语块。实际上这种方法采用的是一种“回避困难”的权宜之计。我们注意到,这种翻译方法并不是从整个输入语句的全局来进行语块或片断的切分[Furuse et al., 1998],而只是局部地计算片断与事例库中短语(或片断)的语义距离,这样有可能抽取出的翻译片断与原输入语句的含义并不相符,而且,有时候整个输入语句中抽取不出任何可翻译的片断,整个翻译系统也就无任何输出。另外,这种翻译方法与语义计算密切相关,因此,需要有一部高质量的语义词典和能够真正反映片断之间相似

性的语义距离测度模型，而这些对于中文计算来说，往往都是比较困难的。

1999年日本德岛大学(The University of Tokushima)教授任福继博士提出了一种基于超函数的翻译方法(Super-Function Based Machine Translation, SFBMT),这种翻译方法的基本出发点是:在目前机器翻译系统不能做到“信、达、雅”翻译目标的情况下,用户与其对系统翻译正确率和译文可读性等提出不现实的过高要求,还不如提高系统的运行速度、减少系统开销、提高系统可控性和易修改性等性能来得现实。例如,一个以英语为母语的机器翻译系统用户并不会在意系统把“吃药”翻译成“take medicine”还是“eat medicine”。因此,基于这种考虑,任福继教授提出了利用超函数(Supper-Function, SF)建立原语言输入与目标语言输出之间映射关系的实现思路。具体地讲,这种方法不像其它基于分析的机器翻译系统一样对原语言句子进行彻底的句法和语义分析,而是通过构造SF来建立原语言句子模板和目标语言句子模板的对应关系[Ren, 1999a]。问题的关键是如何获取SF。[Ren, 1999b]和[Ren et al., 2000]介绍了一种机器辅助的从双语对照语料中获取SF的基本方法。实际上,这种基于SF的直接翻译方法,在本质上与统计模型和模板方法是完全一致的。

日本ATR-SLT(Spoken Language Translation Laboratories)研究所的Yamamoto等人提出了一种基于输入语句改写的口语翻译模型,称之为Sandglass翻译模型[Yamamoto et al, 2001, 2002; Zong, 2001; 宗, 2002b]。这种翻译方法的一种解释是当翻译引擎不能实现输入句子的正确翻译时,让系统的原语言预处理模块自己实现输入句子的自动改写,并不断为翻译引擎提供原来语句另外可能的表达形式。其基本思想是借鉴和模拟人在进行口语翻译时的基本模式。另一种解释是,由于翻译模块能够处理的语言现象往往是有限的,如果系统原语言预处理模块能够生成输入语句的其它多种表达形式,那么这些表达形式中只要有一个落入到系统可以处理的有限的句型集之内,系统就可以得到原输入语句的正确翻译结果。Sandglass翻译模型的直接效果是把复杂的原语言解析任务从翻译模块中分离出来,让原语言本身来表达输入语句的含义,翻译模块可以采用简单的转换方法,例如:模板直接匹配等方法,实现有限集内原语言到目标语言的翻译转换。

实际上,Sandglass口语翻译模型面临许多新的问题和技术难点,而且,人们最容易提出这样一个问题:如果系统能够对原输入语句进行分析和改写,那么,为什么不直接将其翻译呢?

后来,人们提出了基于简化表达式的语音翻译方法[Zong et al., 2000b]等多种方法。总起来说,非传统方法基本上都是在传统方法的基础上进行的改造。

3.3 人机互助的混合翻译方法

人机互助的混合翻译方法是指多种翻译策略并用的多引擎交互式口语翻译方法。目前从事的语音翻译研究都是针对推进式的对话方式(push-to-talk dialogue),即两个持不同语言的说话人之间的对话过程是按每人一句进行的,也就是说当一方说完以后,系统把该说话人的语句翻译完毕并发送给对方,对方听完来自彼端说话人的翻译语音后才开始发音。在这种情况下,翻译人员(翻译器)可以有机会与说话人沟通,在实施翻译之前翻译(器)可以主动地与说话人对话,弄清说话人使用的词汇,确认说话者的意图。我们认为,在这种对话翻译系统中,中间翻译角色必须具备与说话人对话的能力,即翻译者与说话人之间应该是交互式的,而不是单向的。但从目前研究情况来看,除了JANUS系统外,几乎都不是人机交互式的,人们只是在原语言识别和分析技术上下功夫,一厢情愿地希望能够通过改进识别技术和翻译策略来提高系统翻译的正确率和鲁棒性,但实际上这有点儿徒劳无功,因为毕竟人说话是无法控制的,说话人愿意怎么讲就怎么讲,想到哪里就说到哪里,根本不像书面语那样经过深思熟虑,并且有表达符号标记,而且说话人可能随时对前面说过的话进行修正,在这种情况下让系统无论“听懂”还是“听不懂”硬去翻译,根本不给系统提问的机会,这有点儿不合情理,好像翻译器和说话人也都是不负责任的,说话人不管系统是否理解,只管说,系

统则不管理的对与错,只管翻。以这种方式实现的语音翻译系统的正确率和鲁棒性至今没有根本性的改进和提高也似乎是情理之中的事情。

近几年中人们提出了一些交互式翻译的设想[Waibel, 1996; Boitet, 1996; Blanchon, 1996; Seligman, 1997, 1998],但通过对这些交互式口语翻译系统的分析我们发现,其交互过程都是针对如何校对语音识别器的错误结果进行的,这种实现思想的出发点都是认为机器翻译模块只能接受 100%正确的输入语句。而实际上这是不现实的,一方面,语音识别器不可能做到无错误识别,尤其说话人在复杂环境下;另一方面,说话人与系统之间的交互过程势必影响系统的实时性,更重要的是说话人是否愿意使用这样一个繁琐的需要不停地校对错误的用户界面,用户所能够忍受的极限有多大?进一步分析,即使机器翻译模块接收到的输入是完全正确的,系统就一定能够得到正确的输出吗?显然未必。那么,问题就在于系统是否能够真正“理解”说话人的意图,如果系统一旦发现输入语句中存在无法消除的歧义,系统如何与说话人进行交互,向说话人获取系统想要得到的知识。因此,基于这样的考虑我们提出了基于对话管理的交互式语音翻译方法[Zong et al., 2002],在这个交互式翻译框架下,我们提出了基于填充槽的翻译方法(Slot-based translation)。

我们认为,基于对话管理的交互式口语翻译方法基本上模拟了人的翻译过程,该方法不但可以方便地集成传统翻译方法和非传统翻译方法中的多种翻译策略,而且可以将语音-语音翻译和文本-语音翻译有效地统一起来,有利于推进口语翻译系统的尽早实用化。

3.4 分析方法与统计方法的比较

分析方法与统计方法一直是自然语言处理中采用的两大主流方法,实际上,在我们上面介绍的三种翻译方法中,分析方法与统计方法都是不可缺少的两种基本方法。在过去自然语言处理研究中,尤其是机器翻译研究的几十年中,关于这两种方法的优劣比较和结合策略等问题的争论从来都没有停止过。2002年日本 ATR-SLT 的 Sumita 博士介绍,在 ATR 进行的语音翻译实验中,基于事例的翻译方法的正确率要远远高于基于统计的翻译方法[Sumita, 2002],参见下表。

不同方法的翻译器	A	A+B	A+B+C
日英统计翻译系统	25%	46%	64%
英日统计翻译系统	41%	48%	57%
基于事例的日英翻译系统 - 1 (D ³) ²	47%	66%	77%
基于实例的日英翻译系统 - 2 (HPAT) ³	50%	61%	71%

表 2. ATR-SLT 对不同翻译方法的比较

在德国的 Verbmobil 语音翻译实验系统中,却得出了完全相反的比较结果,在该实验系统中统计方法的性能却明显优于基于格的翻译器和基于语义转换的翻译器等其它 4 种翻译器[Wahlster, 2002]。参见表 4。

翻译器	翻译正确率(≥50%)	翻译正确率(≥75%)	翻译正确率(≥80%)
基于格的翻译器	37%	44%	46%

² Dp-match Driven transDucer. 参见: Sumita, 2001, Example-based Machine Translation Using DP-matching between Word Sequences. *Proc. of DDMT(ACL)*, pp. 1-8.]

³ Hierarchical Phrase Alignment (HPA) based Translation (HPAT). 参见: Imamura, 2002, Application of Translation Knowledge Acquired by Hierarchical Phrase Alignment, *Proc. of TMI*.

统计翻译器	69%	79%	81%
基于对话主题的翻译器	40%	45%	46%
基于语义转换的翻译器 (SeT)	40%	47%	49%
基于子串的翻译器 (Substring-based)	65%	75%	79%

表 3. Verbmobil 系统中对不同翻译方法的比较

事情看起来似乎很矛盾，但我们仔细分析却会发现，其中存在很多值得推敲的地方。第一，ATR 比较的是日英之间的翻译，而 Verbmobil 系统中比较的却是德英之间的翻译，日英之间的语言差异比德英之间的语言差异明显地要大得多。第二，在 ATR 实现的语音翻译系统中，基于事例的翻译引擎经历了 10 多年的调试和不断改进，几乎从 ATR 从事语音翻译研究的开始他们就在做基于事例翻译方法实验，系统中的翻译模板 (pattern)、规则和词典等都经历了很多人十几年的精心调试，而基于统计模型的翻译方法只是在近一年左右才开始进行的，而且训练语料也是有限的。而在 Verbmobil 系统中，统计翻译方法的研究时间要更久一些，研究基础也更好一些。尽管研究时间的长短并不能决定实现方法的优劣，但是，训练语料标注的精细程度和训练语料规模大小等却是起决定作用的。第三，在 ATR-SLT 的实验系统和 Verbmobil 系统中，基于统计的翻译方法采用的都是基于词对位的，那么，如果是基于短语的或者基于 chunk 的、甚至是基于结构的，翻译结果又会怎样呢？无论如何，如果翻译的语言对不同，训练和测试的语料不同，或者统计的基元不同等诸多因素不是处于统一起跑线，其比较结果是没有太大意义的。

在谈到分析方法和统计方法比较的时候，曾有人把统计方法比作“石头汤”⁴。所谓的“石头汤”典故指的是这样一个故事：很久以前在东欧某一个地方发生了一场饥荒，有个小贩到一个村庄去卖东西，他看到那里的人们都没有饭吃，就对大家说我给大家做一个石头汤吧，起初大家都不相信，石头怎么能做汤呢？于是小贩就架起了铁锅，然后倒入水，放进去石头就开始点火煮，锅里只有水和石头，水开以后小贩说，你们给我一点调料吧，于是有人给了他一些调料，后来又有人加上了卷心菜等其它可吃的东西，于是整个村庄就享用了这样一大锅美味的“石头汤”。显然，这锅汤的美味并非来源于石头，而是调料和其它蔬菜作用的结果。当时美国一批从事机器翻译的专家为了批判 IBM 的统计翻译方法而引用了这个典故。IBM 刚刚提出基于统计翻译模型的翻译方法时，大有完全抛弃其它方法的架势。那些批评统计翻译方法的人认为，单纯的统计模型不可能从根本上解决问题，统计方法可以作为一种辅助手段用于提取规则、获取语言学知识等，而最后起决定作用的还是要靠其它方法。我在这里引用这个典故却不是想批判哪一种方法，而是想说明无论是口语翻译还是自然语言处理的其它研究课题，任何一种方法都不可能从根本上解决问题，我们需要的是一种融合技术，至于这个融合技术中什么方法起主导作用并不重要，关键是如何融合。统计方法的崛起打破了几十年来分析方法一统天下的局面，但决不意味着分析方法的结束，很多时候这两种方法并不是完全对立的，两者之间存在一个绝对化与模糊化、唯一化与多选化、主观性与客观性、以及定性与定量的关系。概率上下文无关文法 (Probabilistic Context Free Grammar, PCFG) 是规则方法与统计方法相结合的一种，但这种结合方法仍然是外围的或者说是孤立的结合，那么，如何从更深层次、以更紧密的方式结合，真正实现“我中有你，你中有我”，以及在什么阶段和条件下结合等一系列问题，都有待于我们进一步研究和探索。

4. 基于多策略的汉语话语翻译研究

⁴ 参见 http://spanky.triumf.ca/www/fractint/stone_soup.html

话语翻译技术研究是国家重点基础研究发展规划项目(973项目)“图象、语音、自然语言理解与知识挖掘”的子课题之一(编号:G1998030504-01)。在该项目的资助下,课题组在数据资源建设、翻译方法研究、实验系统开发和国际合作等方面均确立若干主要的研究成果。

4.1 数据资源建设

大规模真实语料和数据是口语信息处理的基础。但国内真实口语语料资源十分缺乏,大规模的双语对照口语语料尤其匮乏。因此,我们把基础语料资源的建设作为一项首要的工作,有针对性地进行了口语语料、多语对照语料的收集和处理,建立了一个现场收录的自然口语语料库及一个通用的口语语料库,并建立了多个大规模多语口语对照语料库及专用语料库。主要包括:

(1) 收集旅游咨询真实场景下的口语对话约90段,3000个对自然口语句,并在此基础上对全部口语语料进行了标注和分析。

(2) 与国际语音翻译联盟(C-STAR)成员联合收集旅行过程中最常用的中-英-日三语对照句子各2万多句,其中每个中文语句又由3个意义相近的句子构成,约7万句的规模。另外,与C-STAR成员联合收集日常多领域多语言口语对照语句(每种语言)约20万句。

(3) 与国外有关研究机构联合收录中国人名地名及机构名约300多万个,经过分析处理后,用于识别人名地名、及机构名称。

4.2 基于多策略的口语翻译方法

在口语翻译方法研究中,我们先后提出了基于灵活模板的汉英口语直接翻译方法、交互式口语翻译方法,以及多翻译引擎的串行工作策略等。这些新方法在我们的实验系统中均取得了初步的成功。

基于灵活模板的直接翻译方法是多引擎口语翻译系统中采用的基本翻译策略,使翻译机制与语言保持较好的相对独立性,便于实现语言扩充和系统向多领域移植。我们还提出了灵活的“柔性”翻译模板匹配策略,既允许模板中出现关键词和词性变量,也允许多个候选关键词以逻辑“或”的形式出现,还允许跳过任意词汇,并且语义特征可作为限制条件,目标语言生成也采用了灵活的处理方式。这个策略有效地克服了一般基于模板的翻译方法对于翻译句型缺乏灵活性、翻译结果过于死板的弱点。测试结果表明基于模板的翻译器可以对完全正确的文本输入的语句得到70%以上的正确翻译结果,对于语音识别模块给出的前10个候选结果,可以获得68%的翻译正确率。

根据前面的分析,目前的语音翻译实验系统主要存在鲁棒性差、翻译正确率低的问题,其主要原因一方面源自于语音识别器有限的正确率和鲁棒性,另一方面翻译系统缺乏人机交互的能力,因此,我们提出了基于对话管理的交互式语音翻译方法。在这种翻译方法中,综合了基于模板的翻译方法、基于中间语义表示(IF)的翻译方法、基于统计的翻译方法和基于对话管理引导的槽填充翻译模式等多种翻译方法[Zong, 2002a],该方法在理论上体现了人充当翻译角色时的基本工作方式和思维过程。

4.3 实验系统开发

基于上述工作基础,我们先后开发了多个口语翻译实验系统,其中包括上面提到的LodeStar中日、中英语音翻译实验系统,基于C-STAR III翻译平台的中英双向电话语音翻译实验系统和与韩国电子通信研究院(ETRI)联合开发的基于C-STAR III翻译框架的中韩双向电话语音翻译实验系统。中韩双向电话语音翻译实验系统建立在旅馆预定领域,可识别韩语词汇约16000个,中文词汇约12000个,系统可以约2到3倍的实时速度运行。该系统于2002年3月在C-STAR III国际口语翻译联盟首次在我国召开的C-STAR研讨会上成功地演示。

值得提及的是,目前自动化所模式识别国家重点实验室正与北京首都信息集团公司等多家单位联合开展面向 2008 年北京奥运会的多语言网络信息服务系统的联合攻关研究,其中,语音翻译是重要的技术模块之一。

4.4 国际合作

在本项目的资助下,口语翻译研究在国际合作方面取得了重要进展,自动化所作为核心成员参与了国际七国语言电话语音自然口语互译合作计划,与国际上本研究领域最具实力和水平的大学和研究所共同合作。同时与日本 ATR、韩国 ETRI 建立了长期的合作关系,与日本松下公司实现了富有成效的合作开发联合实验;与国际著名的法国机器翻译研究所(GETA, IMAG)建立了良好的合作关系,双方联合开展中法口语翻译技术的研究,双方曾多次互派访问学者,并就双边联合培养硕士生、博士生一事达成协议;与 Nokia 中国研发中心建立了良好的合作关系,双方就口语翻译中的若干基础问题,包括语料收集,语音识别等方面,联合攻关,并在已有的工作基础上联合参与了欧共体项目“面向多语种口语翻译的词汇处理(LC-STAR)”(编号:IST-2001-32216)。

5 结束语

总的来说,本文想表达的观点可以归结为如下几点:(1)在过去 10 多年的语音翻译研究中已经取得了长足的进步,但是无论在理论上还是在技术上都还存在若干重要问题需要进一步深入研究;(2)在口语翻译研究中,统计方法和分析方法仍然是两大主流技术,两种方法的结合势在必行,研究和探索基于大规模真实数据的统计方法与分析方法的融合技术在语音翻译中同样具有重要的科学意义和实用价值;(3)在本 973 项目的资助下,口语翻译研究无论是在数据资源建设和翻译方法研究方面,还是在实验系统开发和国际合作与交流等方面,均取得了若干重要成果;(4)尽管在现阶段口语翻译中仍存在许多尚未解决的关键问题,但适当地定位口语翻译系统(或口语翻译机)在实际应用中的场合、领域和使用方式等,从实际需求出发,合理地利用已经成熟的技术,实现人机互助的有助于人们语言交流的语音翻译系统,仍然是可行的。

参 考 文 献:

- [1] Agnas, M. S., H Alshawi et al. 1995. Spoken Language Translator: First Year Report. Available from <http://www.cam.sri.com/tr/crc043/abstract.html>
- [2] Gao, Wen, Bo Xu, et al. 1997. Chinese-English Spoken Speech translation. In *Proceedings of CJSLP'1997*, pages 1~5.
- [3] Karger, Reinhard. 2002. VERBMOBIL: Translation of Spontaneous Speech. Available from <http://www.dfki.uni-sb.de/Projects/Verbmobil>
- [4] Kitano, Hiroaki. 1994. Speech-to-speech Translation: A Massively Parallel Memory-Based Approach. *Kluwer Academic Publishers*, Boston.
- [5] Laive, Alon, Alex Waibel et. al. 1997. JANUS-III: Speech-to-Speech Translation in Multiple Languages. In *Proceedings of ICASSP'1997*. Vol. I. Pages 99-102.
- [6] Mima, Hideki, Osamu Furuse and Hitoshi Iida. 1997a. Improving Performance of Transfer-Driven Machine Translation with Extra-Linguistic Information from Context, Situation and Environment. In *Proceedings of IJCAI*, Japan, vol. 2, pages 983~988.
- [7] Mima, Hideki, Osamu Furuse and Hitoshi Iida. 1997b. A Situation-based Approach to spoken dialogue Translation between Different Social Roles. In *Proceedings of the 7th International Conference on*

- Theoretical and Methodological Issues in Machine Translation*, USA. Pages 176~183.
- [8] Rayner, Manny, David Carter. 1997. Hybrid Language Processing in Spoken Language Translator. SRI Cambridge Technical Report. Available from <http://www.cam.sri.com/tr/crc064/paper/paper.html>
- [9] Ren, Fuji. 1999a. Super-function Based Machine Translation, in *Communications of COLIPS*, 9(1): 83-100.
- [10] Ren, Fuji. 1999b. SFBMT and Tool Development for Acquisition of Super-Functions. In *Proceedings of the IASTED International Conference Artificial Intelligence and Soft Computing (ASC)*. August 9-12, 1999, Honolulu, Hawaii, USA.
- [11] Ren, Fuji, Yoichi Takehara. 2000. Tool MKSF for Acquiring Super-Function in JAVA. In *Proceedings of IEEE International Conference on System, Man, and Cybernetic*. Vol. 4. Pages 2362-2368.
- [12] Roe, David B. Fernando C. N. Pereira, et al. 1992. Efficient Grammar Processing for a Spoken Language Translation System. In *Proceedings of ICASSP*, USA, vol. 1, pages 213~216.
- [13] Seligman, M. 2000. Nine Issues in Speech Translation. In *Machine Translation*. 15: 149-185.
- [14] Somers, Harold L. 1993. Current Research in Machine Translation. *Machine Translation*, 7:231~246.
- [15] Sugaya, Fumiaki, Toshiyuki Takezawa, Akio Yokoo, Seiichi Yamamoto. 1999. End-To-End Evaluation in ATR-MATRIX: Speech Translation System Between English and Japanese. In *Proceeding of EuroSpeech'99*.
- [16] Sumita, Eiichiro, Setsuo Yamada and Kazuhide Yamamoto. 1999. Solutions to Problems Inherent in Spoken-language Translation: The ATR-MATRIX Approach. In *Proceedings of MT Summit VII*, Sept. 1999. Pages 229-235.
- [17] Sumita, Eiichiro. 2002. Corpus-Centered Computation. In *Proc. ACL Workshop: Speech-to-speech Translation*. Philadelphia, USA. July 11, 2002. Pages. 1-8.
- [18] Vasconcellos, Muriel. 1993. What Do We Want from MT ? *Machine Translation*, 7(4): 293~301.
- [19] Wahlster, Wolfgang (Ed.). 2000a. *Verbmobil: Foundations of Speech-to-Speech Translation*. Springer-Verlag Berlin Heidelberg New York.
- [20] Wahlster, Wolfgang. 2002. *Verbmobil Multilingual Processing of Spontaneous Speech*. From www.dfki.de/~wahlster/VM-final
- [22] Wahlster, W. 2000b. Mobile Speech-to-Speech Translation of Spontaneous Dialogs: An Overview of the Final Verbmobil System. In *Verbmobil: Foundations of Speech-to-Speech Translation*. Springer Press. Pages 3-21.
- [23] Waibel, Alex. 1996. Interactive Translation of Conversational Speech. In *Proceedings of (C-Star II) ATR International Workshop on Speech Translation*, Sept. 10 - 11, 1996, Japan, pages 1~17.
- [24] Waibel, Alex, A. M. Jain, A. E. McNair et al. 1991. JANUS: A Speech-to-Speech Translation System Using Connectionist and Symbolic Processing Strategies. In *Proceedings of ICASSP'91*. 1991.
- [25] Wakita, Y., Jun Kawai, Hitoshi Iida. 1997. Correct Parts Extraction from Speech Recognition Results Using Semantic Distance Calculation, and Its Application to Speech Translation. In *Proceedings of a Workshop Sponsored by the ACL and by the European Network in Language and Speech (ELSNET)*. Pages 24-29.
- [26] Wu, Hua, Taiyi Huang Chengqing Zong and Bo Xu. 2000. Chinese Generation in a Spoken Dialogue Translation System". In *Proceedings of COLING*, July 27 - August 4, 2000. Germany. Pages 1141-1145.
- [27] Xie, Guodong, Chengqing Zong and Bo Xu. 2002. Chinese Spoken Language Analyzing Based on Combination of Statistical and Rule Methods. To appear in ICSLP'2002. USA.
- [28] Xu, Weiqun et. al. 2002. Bridging the Gap Between Dialogue Management and Dialogue Models. *Proc. ACL-02 Workshop: 3rd Sigdial*. pp. 201-210.
- [29] Yamamoto, Kazuhide, Satoshi Shirai, Masashi Sakamoto and Yujie Zhang. 2001. SANDGLASS: Twin Paraphrasing Spoken Language Translation. In *Proceedings of ICCPOL'2001*, Seoul, Korea. Pages 154-159.
- [30] Yamamoto, Kazuhide. 2002. Machine Translation by Interaction between Paraphraser and Transfer. In

Proceedings of Coling'02. August 2002, Taipei.

- [31] Zong, Chengqing, Taiyi Huang and Bo Xu. 2000a. An Improved Template-Based Approach to Spoken Language Translation. In *Proceedings of the International Conference on Spoken Language Processing (ICSLP'2000)*. October 16-20, 2000. Beijing. Pages 440-443.
- [32] Zong, Chengqing, Yumi WAKITA, Bo Xu, Kenji Matsui and Zhenbiao Chen. 2000b. Japanese-to-Chinese Spoken Language Translation Based on the Simple Expression. In *Proceedings of the International Conference on Spoken Language Processing (ICSLP'2000)*. October 16-20, 2000. Beijing. Pages 418-421.
- [33] Zong, Chengqing, Taiyi Huang and Bo Xu. 2000c. Design and Implementation of a Chinese-to-English Spoken Language Translation System. In *Proceedings of the International Symposium of Chinese Spoken Language Processing (ISCSLP'2000)*, October 13-15, 2000. Beijing. Pages 367-370.
- [34] Zong, Chengqing, Yujie Zhang, Kazuhide Yamamoto, Masashi Sakamoto, Satoshi Shirai. 2001. Approach to Spoken Chinese Paraphrasing Based on Feature Extraction, In *Proceedings of the 6th Natural Language Processing Pacific Rim Symposium (NLPRS'2001)*. November 27-30, 2001. Tokyo, Japan. Pages 551-556.
- [35] Zong, Chengqing, Bo Xu, and Taiyi Huang. 2002a. Interactive Chinese-to-English Speech Translation Based on Dialogue Management. In *Proceedings of ACL-02 Workshop on Speech-to-Speech Translation: Algorithms and Systems*. July 11, 2002. Philadelphia, USA. Pages 61-68.
- [36] Zong, Chengqing. 2002b. Research Activities for C-STAR III at NLPR. *Report in C-STAR III Workshop*. March 2002. Guilin, China.
- [37] 程葳, 徐波. 2002. 一种面向口语的译文质量自动评价方法. 《中文信息学报》, 16(2):47 - 53. 韩兆兵, 徐波. 2001. C-STAR III 中文系统平台建设, 第七届全国青年通信学术会议, 南京. p44-47.
- [39] 杨家沅, 林道发, 罗万伯等. 1992. 连续英汉语音翻译系统的设计和实现. 声学学报. 17(5): 327 ~ 333
- [40] 宗成庆, 黄泰翼, 徐波. 1999. 口语自动翻译系统技术评析. 中文信息学报, 13(2):56 - 65.
- [41] 宗成庆. 2002a. 语音翻译技术研究现状. 第二届中日自然语言处理专家研讨会论文集, 2002 年 10 月 31 日至 11 月 2 日, 北京大学. 第 32-39 页.
- [42] 宗成庆, 张玉洁, 山本和英, 坂本仁, 白井谕. 2002b. 面向口语翻译的汉语语句改写方法. In *Journal of Chinese Language Computing*, 2002,12 (1): 63-77. Singapore.

作者简介

宗成庆 男, 1998 年 3 月毕业于中国科学院计算技术研究所, 获博士学位。1998 年 5 月至 2000 年 4 月在中科院自动化研究所模式识别国家重点实验室从事博士后研究, 出站后在模式识别国家重点实验室工作至今。1999 年和 2001 年两次在日本国际电气通信基础技术研究所(ATR)做访问学者, 现为模式识别国家重点实验室中文信息处理研究组组长, 担任学术期刊 *Journal of Chinese Language and Computing* (新加坡) 编委、*Journal of Information* (日本) 编委、中国人工智能学会自然语言处理专业委员会委员和 2003 IEEE 自然语言处理与知识工程大会(NLP-KE) 程序委员会协作主席(co-chair) 等多种学术职务。近几年来, 他主要从事口语解析方法、口语语音翻译和人机对话等技术的研究, 作为项目负责人和技术骨干承担国家自然科学基金项目、国家 863 项目、国家重大基础研究项目(973)和国际合作研究项目等 10 余项, 发表学术论文 30 余篇。