口语自动翻译系统技术评析*

宗成庆　黄泰翼　徐　波
模式识别国家重点实验室　中科院自动化研究所　北京　100080

摘要 近几年来，随着信息技术的发展，口语自动翻译技术成为新的研究热点。目前国际上一些著名大学和研究机构甚至企业，都纷纷加入这一高技术的竞争行列，我国在相关技术方面也进行了卓有成效的研究。本文对目前自动口语翻译研究的技术现状进行了全面综述和分析，并对一些具体问题作了深入探讨。作者希望本文作出的分析和讨论的问题，能够对我国的自动口语翻译研究提供有益的参考。

关键词 口语翻译　语音翻译　对话处理　机器翻译　鲁棒性

The Technical Analysis on Automatic Spoken Language Translation Systems

Chengqing Zong, Taiyi Huang and Bo Xu
National Laboratory of Pattern Recognition, P. O. Box 2728, Beijing 100080

Abstract With development of information technology, the technique of automatic spoken language translation becomes a new research point. Recently, many famous universities and institutes in the world compete against with each other in this new technique field, and researchers of our country have made great progress in the related aspects. In this paper the techniques of automatic spoken language translation are summarized and analyzed and some concrete problems are discussed in detail. The authors hope that the paper will provide readers with valuable references.

Keywords Spoken language translation　Speech-to-speech translation　Dialogue processing　Machine translation　Robustness

一、引言

自 80 年代末期，人们开始致力于语音翻译的研究，由于该项研究不仅具有重要的科学意义，而且蕴涵着潜在的巨大社会和经济效益，因此，许多发达国家竞相投入经费开展全国性或多国性的联合攻关，美国的 CMU (Carnegie Mellon University)、日本的 ATR-ITL (Advanced Telecommunications Research Institute – Interpreting Telecommunications Research Labs)、德国联邦政府教育、科学、研究与技术部(Federal Ministry of Education, Science, Research and Technology, BMBF)和 Siemens 等世界著名大学、研究机构和企业都是自动语音翻译研究的重要的开拓者或参与者。为了进一步推动自动

* 本文于 1998 年 7 月 28 日收到。本课题得到国家自然科学基金资助（批准号：69835030）和国家 863 高技术项目资助（课题号：863-306-ZT03-02-2）。

语音翻译研究的更快发展，1991 年一些国际知名研究机构、大学和企业联合成立了国际口语翻译组织 C-STAR(Consortium for Speech Translation Advanced Research)，到目前为止，该组织已经发展成为 C-STAR II 并拥有包括中国科学院自动化研究所、声学研究所和哈尔滨工业大学在内的二十多个会员，其研究目标是形成特定领域的口语对话翻译系统，最终实现基于自动翻译的全球自由通讯。

然而，由于自动语音翻译技术涉及语言学、声学、信号处理、模式识别和认知心理学、计算机科学等各种学科的知识和技术，同时集语音识别、机器翻译和语音合成等理论难点和技术难关于一身，使得其研究历程步履维艰，语音翻译被认为是对现代科技重要的挑战之一。

本文将对自动语音翻译研究的现状、技术特点及其存在的问题等，进行全面综述和分析，作者希望本文提出的问题及所作的分析能对自动语音翻译研究提供有益的参考。

二、口语翻译技术特点

从原理上讲，一个单向的语音翻译系统主要由语音识别、机器翻译和语音合成三个模块构成，如图 1 所示。

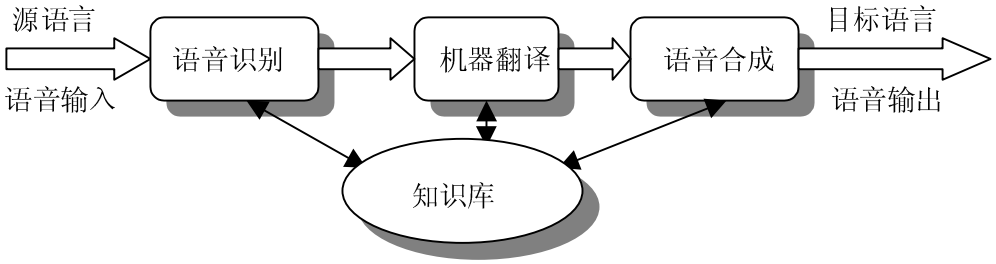

图 1 语音翻译系统组成原理

由于人们的日常交谈大多数是在情景对话中进行的，因此，针对情景对话的口语翻译比规范语言的语音翻译有更广泛的应用前景，目前的语音翻译系统基本上都是针对口语对话的。口语自动翻译与文本机器翻译相比，有许多自身的特点，主要表现在如下几个方面：

（1）在语言的规范性方面，口语句子含有大量的不规范语言现象，包括句子间长时间的停顿、重复、省略、修正和犹豫(嗯，哦，um, hmm 等)，汉语中大量同音字和词的存在以及语音识别技术的限制，使得单词识别本身难以达到 100%的正确，因此，与文本机器翻译不同，口语翻译系统面对的是有不少错误的输入文本，这样，一个成功的口语翻译系统就必须在语义层次上进行处理，同时忽略那些语义表达上并不重要的词语和语段，以达到对说话内容的理解。这些问题在声学和语言学层次上对现有的识别系统和语言解析机制都提出了更大的挑战。

（2）在人们的日常会话过程中，讲话者为了表示许多感情色彩，一般会使讲话的语速变化很大。在快速语速中，音素之间的协调发音现象会导致语音识别结果产生一些错误。同时在语音录入过程中，还有可能存在许多环境噪声，如咳嗽、电话铃声、关门声等，不对环境噪声进行合理的处理，识别系统会将它们作为词典中的一个词汇处理，从而导致翻译结果的严重错误。口语的这些特点既要求识别机制具有更高区分能力的建摸技术，又要求翻译模块必须能够处理识别机制的不确定性产生的多个候选结果，甚至具有纠错的能力。

（3）在口语翻译中，不存在传统意义上的句子，不象书面语有标点符号把它们区分开来，表达方式也相当简洁，含有较多的省略，甚至含混不清。发音是断断续续的，每一段话语可能包含多

个句子和主题。这样，即使声学层面上已经能够正确地切分发音并识别它们，把这些断续的发音翻译出来，其结果也是令人费解的。因此，口语翻译机制必须理解讲话者的意图，而不是简单的句子与句子之间的翻译，是在特定对话环境中综合理解语义和语用的基础上，进行的“意图翻译”。

（4）从语音合成角度看，在口语自动翻译中，人们更希望合成的口语带有同原讲话者接近的感情色彩，同时希望模拟讲话者的个人特征，这就对语音合成提出了更高的要求。

另外，在许多情况下口语翻译系统中不存在译后编辑，希望系统能够做到同声的即时翻译(Incremental Translation)，这对整个口语翻译系统的各个模块都提出了更高的要求。

总之，口语翻译比书面语翻译涉及到的问题更多，难度更大，只有把语音识别、机器翻译和语音合成各部分综合考虑，把声音和文字、识别和翻译作为一个整体处理，才有希望获得较好的翻译效果。

三、研究现状

1989 年美国 CMU 开发了被认为是国际上第一个的语音翻译原型系统 SpeechTrans，该系统因此而成为语音翻译研究的里程碑[1]。在随后的十几年里，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，系统由针对规范输入的语音翻译转向针对不规范输入的口语对话翻译，从而将语音翻译研究带入一个充满希望的新阶段。表 1 列出了近十年来一些有代表性的语音（口语）翻译系统。

系统名称	研制机构	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
JANUS2	CMU, Karls-ruhe 大学等	90 年代初	交通旅游、航空/火车订票、信息查询等	德, 英, 日, 西班牙, 韩等	中间语言	3000
JANUS3	同上	1997	旅馆预订，航空/火车订票，旅游信息等	德, 英, 日, 西班牙, 韩, 俄等	中间语言	开放
ATR-MATRIX	ATR-ITL	1998	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	BMBF 组织	90 年代初	会晤日程安排	德, 英、日等	多策略结合	2500

表 1 部分语音翻译系统

其中，1989 年完成的 JANUS 系统第一版本 JANUS1 只完成英语到日语的语音翻译，系统识别的词汇量只有 400 个，输入语音是句法结构比较规范的句子[2]。

Verbmobil 又称 Verbmobil Demonstrator，该系统曾于 1995 年由当时的德国联邦研究部部长 Juergen Ruettgers 向公众展示。展示系统实现了德语语音输入到英语语音输出的翻译转换，可识别词汇量为 1292 个德语单词。1996 年底 Verbmobil 原型系统增加了日语到英语语音翻译（词汇量为 2500 个）[3]。目前，BMBF 已经开始对该系统实施第二阶段（1997~2000）的研究开发。在 1993 年至 1996 年 Verbmobil 系统研究的第一阶段中，BMBF 资助了 6490 万德国马克，参与研究的企业还赞助了 3100 万马克。共有 7 个公司（包括 Philips GmbH, Siemens 等）、22 所大学或研究所（包括

CMU, Stanford 大学的语言信息研究中心，Karlsruhe 大学等）参与了 Verbmobil 系统的研究和开发工作。

除了表 1 中列出的翻译系统以外，还有 AT&T Bell 实验室开发的 VEST (Voice English / Spanish Translator)系统[4]，SRI International 开发的 SLT 口语翻译系统[5,6]，日本 ATR-ITL 研制的 TDMT 系统[8-10]和 C-STAR II 正在组织研究的多国语音翻译系统，以及其它一些小型的语音翻译原型系统，在这里不再一一赘述。C-STAR II 正在组织研究的多国口语翻译系统至少包括英语、日语、德语、韩国语、意大利语、法语等 6 种语言，应用领域面向旅游计划安排和信息查询，识别词汇量初步定在 1000 ~ 2000 个。

此外，我国四川大学曾于 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，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[20]。1997 年先进人机通讯技术联合实验室也建立了一个小词汇量的面向会议日程安排的语音翻译实验系统[21]。

从目前的研究状况来看，实验系统都是面向特定领域的，系统能够识别的词汇量一般在 2000~3000 左右，句型一般都没有特别限制。

四、技术评析

4.1 语音翻译系统技术类型

从目前国际上研究开发的一些口语翻译实验系统来看，根据翻译机制所采用的不同方法，可以将其大致分为 4 种类型：①基于规则(rule-based)的语音翻译技术；②基于事例(example-based)的语音翻译技术；③基于中间语义表示的语音翻译技术；④基于统计模型的翻译技术。以下分别对这 4 种翻译技术进行简要的分析。

1．基于规则的语音翻译技术

基于规则的翻译技术多见于早期的语音翻译系统，这种翻译技术一般都是直接借鉴基于规则的文本机器翻译技术或在其基础上改造形成的。基本思想是，首先对源语言进行词法分析，然后利用 Earley 或 GLR 等经典的句法分析算法或在其基础上的改造算法，对源语言句子进行句法分析和语义分析，形成分析语句的句法分析树，最后根据句法分析树生成目标语言。

这类翻译技术的主流文法是上下文无关文法(Context Free Grammar, CFG)，扩充的短语结构文法(Augmented Phrase Structure Grammar, APSG)，中心驱动的短语结构文法(HPSG)，或词汇功能文法(Lexical Functional Grammar, LFG)等。`

代表系统有：JANUS1，SL-TRANS，VEST（早期版本）等。在 JANUS1 系统中，采用文法是 LFG，句法分析器采用 GLR 分析算法；SL-TRANS 系统采用 HPSG 和 LR 分析算法；VEST 系统的早期版本采用 APSG 和 Earley 句法分析算法。

基于规则的语音翻译技术的优点在于：可以较好地保持原文的结构，产生的译文结构与源文的结构关系密切，尤其对于语言现象已知的或句法结构规范的源语言语句具有较强的处理能力和较好的翻译效果。主要不足是：分析规则都要由人工编写，工作量大，规则的主观性强，规则的一致性难以保障，不利于系统扩充，尤其对非规范的语言现象缺乏应有的处理能力，这就限制了该方法在口语翻译系统中的运用。

2．基于事例的语音翻译技术

基于事例的机器翻译(Example-Based Machine Translation, EBMT)模型是 80 年代初期由 Nagao 首先提出的，由于该方法是建立在基于记忆的推理(Memory-Based Reasoning, MBR)技术之上的，因此，又称基于记忆的机器翻译。它的基本思想是通过类比实现机器翻译，这一思想建立在这样的假设之上：人类是根据以往的翻译经验从事翻译的。EBMT 的基本方法是：建立一个事例库用来存放成功翻译的事例，将分析句子与事例库中的所有例句比较，找出最相似的例句，然后，根据例句的译文得到分析句子的译文。该方法的核心问题有两个，一是如何计算分析句子与例句之间的结构和语义的相似性，确定合适的标准衡量哪个例句是与分析句子“最相似”的，二是如何提高事例库的检索速度，使其达到实时性要求。

代表系统有：ATR-MATRIX[22]，$\Phi D_M D_{IALOG}$[1]，TDMT[7~10]等。$\Phi D_M D_{IALOG}$ 系统将基于事例的方法和基于规则的方法结合起来运用，但控制整个翻译过程的是基于记忆的推理机制。在 TDMT 系统中，引入了从上下文、对话情景和环境中抽取的部分超语言(extra-linguistic)信息，辅助实现基于事例的翻译过程[9,10]。

基于事例的机器翻译技术的主要优点在于：不要求源语言句子必须符合语法规定，翻译机制也不需要对其进行深层次的句法分析。另外，翻译机制所依赖的事例库是对翻译历史客观的记录，从而避免了人为因素（如基于规则的方法中人工标注字典、编写规则等）的主观性。其不足之处在于：两个不同的句子之间的相似性（包括结构相似性和语义相似性）往往难以把握，尤其在口语中，句子结构一般比较松散，成分冗余和成分省略都较严重，这更增加了分析句子与事例句子的比较难度。另外，系统往往难以处理事例库中没有记录的陌生的语言现象，而且当事例库达到一定规模时，其事例检索的效率较低。

3．基于中间语义表示的语音翻译技术

考虑到口语对话中，大量地存在非规范语言现象，包括严重的省略、重复和停顿，以及由于语音识别机制造成的字词识别错误等各种原因，使得翻译机制几乎无法对源语言句子进行完整的句法分析，因此，人们提出了通过抽取源语言句子的语义信息，然后根据语义信息实现“意译”的处理思想，即中间语义表示的处理方法。在这种翻译方法中，一般事先设计好一个语义框架(semantic frame)，通过对话活动(dialogue act)或若干语义子信息描述整个句子的含义。JANUS2 系统的研究者 Waibel 曾将此语义表示称作语义“中间语言”(semantic interlingua)[2]，当然是一种狭义意义上的中间语言，而 C-STAR 则将其称为中间转换格式(Interchange Format, IF)[23]。在语法分析过程中，分析器只需有目的地从源语言句子中发现和抽取需要的各种信息，然后填充相应的语义框架，语言生成机制根据中间语义表示生成目标语言句子。

由于在抽取语义信息时，往往只对源语言句子进行局部分析，包括归约、概念替换或合一等，因此，其语法分析过程又被称作局部的分析(partial parsing 或 partial processing)[11]。代表系统有：JANUS2，JANUS3 等。

基于中间语义表示的翻译方法的优点在于：分析器不需要对分析句子的任何成分都完全理解，源语言句子可以是任意形式的口语语句，因此，该方法比较适合于口语理解和翻译。另外，中间语言独立于具体的分析语言，从理论上讲，任何一种语言的句子都可以影射到中间语义表示，因此，该方法适合于多种语言的互译，实现的系统易于扩展，而且在语义抽取过程中可以方便地与基于规则的分析方法和 HMM 统计模型等方法相结合。其不足之处在于：无论是句子结构还是语义表达，

不同的自然语言之间毕竟有一定的差异，这一方面要求中间语义表示必须满足所有处理语言（包括源语言和目标语言）的要求，另一方面，由中间语义表示生成的目标语言容易造成千篇一律，有时难以表达源语言的准确含义及不同表达方式之间的差异。

4. 基于统计模型的翻译技术

基于统计模型的翻译技术一般可以采用两种方法。一种方法是，首先建立包含以往成功分析的语言经验的语料库，把经过标注的语料库看作一个语法，当系统处理新的语言现象时，从语料库中抽取片段单元来构造新的语言分析过程，即语义树，然后利用 HMM 计算 n-best 语义树的概率，根据确定的语义树生成无歧义的语义框架，最后生成目标语言。

另一种方法是，根据已知的双语语料库知识和源语言句子中每个单词所处的上下文语境，直接将每个单词影射到对应的目标语言单词，然后对生成的目标语言句子进行词序调整，这种方法的代表系统是 AT&T 开发的 Head Transducer 系统[19]。SLT-2 系统采用基于 HMM 模型的统计方法为主，同时结合了基于规则的翻译方法[12]。

另外，IBM 的研究人员也曾对基于统计方法的翻译技术进行了研究和实验[14,15]，ARPA ATIS 的自然语言接口系统中的语言理解技术采用了一种称为隐理解模型(Hidden Understanding Model, HUM)的统计方法[13]。

基于统计模型的翻译技术的主要优点在于：将复杂的句法、语义分析转化为概率数值计算，概率计算是以通过对大量的语言现象统计建立起来的语料库为基础的，具有较强的客观性，系统不需要做复杂的句法和语义分析。其主要不足表现在于：错误的翻译往往是由于系统不具备语言学知识而引起的，当一个词的翻译依赖于其它词的翻译时，可能引起更多的问题[15]，另外，系统统计的语料库规模也难以把握，语料库规模太小，难以覆盖绝大多数语言现象，语料库规模过大，系统就要付出太大的开销。

实际上，目前国际上正在研究的许多口语翻译系统都在寻求多种处理技术相结合的翻译方法，如 Verbmobil 系统和 C-STAR II 正在组织研究的多国口语翻译系统等。C-STAR II 正在组织研究的多国口语翻译系统尽管是以基于中间语义表示为主体框架的，但由于各成员国分别进行与各自国家语言相关的翻译研究，其翻译方法各有特色，往往是几种方法的集成，因此，很难将它们准确地归结为某一种特定的翻译类型。

4.2 已经取得的进展

在过去的十几年里，尽管语音翻译研究举步艰难，但无论在理论上，还是在工程实践方面，都取得了一定的进展。由于语音翻译系统涉及到语音识别、机器翻译和语音合成三个相对独立的研究分支，因而，这些进展很大程度上体现在这三个具体的研究分支里，在此作者不想具体到每一个研究方向去归纳它们各自的进展，只是从语音翻译研究的整体上来总结其取得的进步，概括起来，我们将其归纳为如下几点：

- 系统翻译的词汇量逐步扩大。随着语音识别技术的提高，语音翻译实验系统识别和翻译的词汇量已经由初期系统的几百个(如 JANUS1)或 1000 个左右(如 ST-TRANS, SLT1 等)，增长到 2000 到 3000 多个(如 Verbmobil)，甚至更多(如 JANUS3)。
- 系统处理的句子已经转向口语化。在初期的实验系统中，采用的分析方法往往是针对规范语言的(如 SpeechTrans, JANUS1 等)，而在目前的研究系统中都是针对日常口语会话的，分析算法提出了 Partial parsing 等多种处理思想。

- 翻译方法开始走向多元化和集成化。目前许多系统往往不是采用单一的一种翻译方法，而是以某种方法为主，同时结合其它翻译方法的处理思想，如 Verbmobil，SLT-2 等。
- 更多的世界知识和对话环境知识被引入到口语翻译系统中。为了提高口语解析的正确率，一些系统开始尝试将讲话人的社会角色、对话情景等环境知识引入到翻译系统，甚至配备电视设备[2,9,10]，以辅助听话人理解翻译结果，这符合人与人之间对话翻译的特点。
- 开始研究多语种翻译，实现双向口语翻译系统。在早期的实验系统中，一般都是研究从一种语言到另一种语言的单向语音翻译(如 SL-TRANS, SpeechTrans 等)，而目前的研究系统都是针对多种语种的双向口语翻译。

4.3 进一步研究的课题

尽管语音翻译研究已经在探索中取得了阶段性成果，但仍存在大量的问题有待于进一步研究，尤其结合到与中文相关的（汉-外、外-汉）口语翻译研究，不仅汉语口语分析和理解存在大量问题，而且面向口语对话的语音识别和语音合成中存在的问题也严重地限制了口语翻译研究的进展，归纳起来，应该进一步研究的课题包括如下几个方面：

- 加强口语的声学特性分析，使声学-语音层 HMM 模型精细化。同书面语相比，口语的声学特性有一定的特殊性，这类语音的音段特征（即语音的谱特性）和超音段特性（包括语速、语调、音强等）随表达的内容、感情色彩等的不同，变化的范围比朗读语言大得多，同时还有非语言信号和噪声，充分研究这些特性，建立精细的声学模型非常重要。
- 研究口语的语言学特征，完善语言模型。在语言学层面，口语句子中含有大量的修正、重复、口头语、省略等非规范语言现象，研究这些特征，对语言模型进行完善，包括建摸、算法和训练等，将有助于提高语音识别的正确率。
- 提高语音识别的鲁棒性(robustness)。对于真正实用的口语翻译系统来说，讲话人往往是在较强的背景噪声或多讲话人环境下发音的，如果是电话自动语音翻译系统，还存在通讯干扰等其它因素的影响，因此，提高语音识别在不同说话人、不同声学环境及通道条件下的鲁棒性，在口语翻译系统中尤其重要。
- 提高系统的自适应能力。一个口语翻译系统往往要面对大量的、发音风格完全不同的讲话人，因而，系统必须能够快速而有效地适应这些不同的讲话人。
- 韵律引导的搜索和韵律控制生成。在口语中，除了句子间没有标点符号外，句法的某些结构、语义群的分割等，都可以以一定的形式在韵律变化中反映出来。在口语识别理解中提取并充分利用这些韵律特征，可以大大地缩减识别搜索空间，减少分析的歧义性。从合成角度看，建立和完善汉语的控制规则，可以提高合成语音的自然度。
- 加强受限领域语料库的研究。在以往的自然语言处理研究中，尤其在中文信息处理研究中，一般都侧重于规范的书面语言的研究，而口语语料的收集、语言学特性及统计规律分析和语料库建设等，都没有得到足够重视，而这恰恰是建立口语翻译系统的基础。
- 研究对话情景知识的表示和利用。这里所讲的情景知识是指除讲话人所处的物理环境以外的，在对话过程中能够通过语言表达出来的世界知识，包括讲话人的社会角色、历史背景、讲话人的情绪或态度等等，这些知识对于理解讲话人的讲话内容起着重要的参考作用，而如何表示、发掘和利用这些知识，这是口语翻译需要研究的问题之一。
- 构造鲁棒的口语解析器(parser)。正如前面提到的，口语中存在大量的不规范语言现象，

并且目前的语音识别器不可能具备 100%的识别正确率，因此，语法解析器必须具有较强的容错能力和分析、理解能力。

- 进一步加强翻译方法的理论研究。尽管目前提出并不断实践着各种口语翻译方法和理论，但其翻译效果离人的最终要求还有相当的距离，无论是短语边界的自动确定、歧义消解、指代问题，还是译文生成算法及译文可读(听)性修饰等诸多具体问题，都有待于进一步研究。
- 研究多风格、多发音人的汉语合成系统。口语翻译系统中的语音合成子系统应该能够反映不同讲话人的性别、年龄、音色等多风格的特点。
- 提高系统的扩展能力、知识获取能力和系统可移植性。一个理想的实用的知识处理系统应该具有知识自动获取能力，在系统使用过程中自动实现知识库扩充和知识更新，并能够很快地移植到不同的应用领域。
- 优化各种分析算法，提高系统的实时性。

从上述分析我们不难看出，语音翻译技术的多样化恰恰说明该技术还很不成熟，仍处于不断的探索中，而目前已有的各种翻译方法都有它们各自的优点和不足，要完成口语翻译这一复杂的任务，恐怕仅靠单纯的一种方法难以奏效。上述多种方法相结合，充分考虑口语的语言特性、统计规律以及人脑对口语的理解过程，在现有技术的基础上进一步发掘新的方法和技术，才有希望最终突破口语翻译的难关。

五、问题讨论

在过去的十多年里，尽管不断有新的语音翻译原型系统或实验系统相继问世，但是，语音翻译的关键技术并没有得到根本解决，尤其是语音翻译的理论研究进展缓慢，系统性能的提高在相当程度上仰仗于硬件技术的提高，如计算机运算速度的提高，内存容量的增大等因素。那么，语音翻译的“瓶颈”到底在哪里，人对语音翻译系统的要求到底有多高？在此，作者愿意就如下两个问题与读者共同探讨。

5.1 关于目前语音翻译的技术路线

在目前的语音翻译系统中，都是首先将源语言的语音信号转换成文字，然后再对文字进行分析、转换、生成，最后将译文转换成语音信号输出，如图 2 中的 L2。而实际上人在进行口语翻译时，是直接从语音到语音的，尤其当翻译者对源语音比较熟悉时，根本不需要从音到字、再从字到音的翻译过程，如图 2 中的 L1。

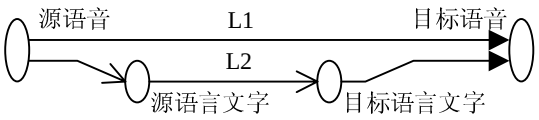

图 2 口语翻译技术路线

那么，在目前的语音翻译系统中，采用 L2 处理过程显然是一种不得已而为之的处理方案，而这种方案是在计算机硬、软件技术（包括内存、运算速度和知识表示方法等）受到一定局限的情况下诞生的，可以设想，当计算机的内存和运算速度足够大，语言知识包括语音知识在内的语义知识、语用知识和语境等知识得到充分表达和利用时，直接实现从语音到语音的翻译是完全有可能做到的事情。

在人们的日常对话中，情景知识是理解讲话者语音的重要因素。一个人理解对方的话语往往

不只是通过对方的语言本身，而也看对方的手势、动作、表情等行为特征和讲话人的社会地位、角色以及谈话背景等抽象特征，需要时，听讲人双方还要有一个问答的交互过程，在特定情况下，这些辅助信息对于帮助听众理解讲话人的内容起了决定性的作用。因此，一个面向口语的语音翻译系统应该是交互式的，系统并不一定要对讲话人的每一句话都完全“听懂”和“理解”，只要系统“理解”其中的一部分，就可以通过与用户交互来获得在字面上无法反映出来的对话情景知识，从而帮助系统对源语言句子理解和翻译。

总之，作者认为，语音翻译的技术路线不应该是单一的，必须充分体现人脑完成口语翻译过程的复杂性，并在可能实现技术的基础上，建立多种方法联合协作的语音翻译集成系统。

5.2 关于语音翻译系统的评估问题

从人的要求来看，语音翻译系统能够达到的翻译正确率越高越好，如果能达到 100%的翻译正确率，就可以完全替代人实现全自动翻译，但实际上，这是不可能的，至少在目前没有这种可能。那么，人是否一定要有 100%正确的翻译才能理解讲话人的意图呢？回答显然是否定的，人们在日常会话中，即使持两种不同语言的人之间的会话，也不需要清清楚楚地完全听明白对方讲的每一个字，有时只需要听清一部分内容，甚至一个词就可以完全明白对方的意思。实际上，作为一个翻译系统来说，人本身也是其中的一个组成部分（语音的接受者和理解者），而且是具有相当智能的一个组成部分，因此，从这一角度考虑，我们没有必要刻意要求系统必须达到一个目前无法实现的标准，似乎达不到某一标准就予以彻底否定，系统翻译的正确率也不应该是衡量系统好坏的唯一标准。

参照机器翻译评估的方法[16~18]，作者认为，除了系统所能实现的翻译正确率以外，评价一个口语翻译系统还应着重从如下几方面考虑：

- 系统的鲁棒性和自适应能力

系统的鲁棒性一方面指语音识别机制对环境噪声和其它各种因素的抗干扰能力，另一方面指翻译机制对来自语音识别模块的输入语句的容错能力和理解能力。自适应能力则是指系统对不同讲话人的适应性。

- 系统的扩展能力和可移植性

系统的扩展能力主要指系统在识别词汇量、能够处理的语言现象、系统知识库规模等各方面的扩展能力。可移植性则是指系统从一种应用领域转移到另一种应用领域或从一种平台移植到另一种平台的方便程度。

- 系统的交互能力和学习能力

系统的交互能力是对话翻译系统的重要功能之一，一方面系统可以通过交互操作获取理解句子的基本知识，以提高系统的解析正确率，另一方面可以通过学习机制实现系统的知识库完善和修改。

- 系统的响应速度

由于口语翻译系统是包含多个子系统的集成系统，尤其语音识别和机器翻译模块往往需要较多的分析时间，因此，如何使系统的响应速度达到实时性要求，也是衡量系统实用性的一个重要方面。

- 译文的可听性

在语音翻译系统中译文是通过语音信号播放的，而不是供阅读的，因此，评价译文的质量除了正确性以外，还要看译文的可听性。所谓的可听性主要指译文的流畅性、是否符合目标语言的表达习惯和表达方式，以及输出语音的清晰度和自然度等各方面与人的要求的符合程度等。

六、结束语

综上所述，口语翻译研究不仅是一个复杂的理论课题，而且是一个庞大的工程项目。从目前技术水平来看，要实现无任何限制的、完全人格化的口语翻译系统是不可能的，但是，针对某个特定领域和要求，实现受限领域的口语翻译系统却是非常现实的，而且同样具有重要的科学意义和实用价值。

从目前我国相关技术的研究状况来看，无论是语音识别、机器翻译，还是语音合成，都已具备了一定的基础和水平，并且已经走向或正在走向实用化，中文信息处理技术也正在向着工程化和实用化的进程发展，因此，我国研究和开发与中文相关的口语翻译系统的条件已经成熟，我们有理由相信，在不远的将来开发出实用的中-外文口语互译系统。

致谢 作者衷心地感谢本文审阅者提出的宝贵意见。

参考文献

[1] Hiroaki Kitano. Speech-to-speech Translation: A Massively Parallel Memory-Based Approach. *Kluwer Academic Publishers*, Boston, 1994

[2] Alex Waibel. Interactive Translation of Conversational Speech. In *Proceedings of ATR International Workshop on Speech Translation*, Sept. 1996, Japan, pages 1~17

[3] Reinhard Karger. The Verbmobil Project. Available from *http://www.dfki.uni-sb.de/ verbmobil*

[4] David B. Roe, Fernando C. N. Pereira, et al. Efficient Grammar Processing for a Spoken Language Translation System. In *Proceedings of ICASSP'92*, USA, vol. 1, pages 213~216

[5] Manny Rayner, David Carter. The Spoken Language Translator Project. SRI Cambridge Technical Report, 1995. Available from *http://www.cam.sri.com/tr/crc057/paper/paper.html*

[6] M-S Agnas, H Alshawi et al. Spoken Language Translator: First Year Report, 1995. Available from *http://www.cam.sri.com/tr/crc043/abstract.html*

[7] Eiichiro SUMITA, Kozo OI, Osamu FURUSE and Hitoshi IIDA. Example-Based Machine Translation on Massively Parallel Processors. In *Proceedings of IJCAI-93*, France, vol. 2, pages 1283~1288

[8] Eiichiro SUMITA, Hitoshi IIDA. Experiments and Prospects of Example-Based Machine Translation. In *Proceedings of ACL-91*, USA, pages 185~192

[9] Hideki Mima, Osamu Furuse and Hitoshi Iida. Improving Performance of Transfer-Driven Machine Translation with Extra-Linguistic Information from Context, Situation and Environment. In *Proceedings of IJCAI-97*, Japan, vol. 2, pages 983~988

[10] Hideki Mima, Osamu Furuse and Hitoshi Iida. 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, 1997, pages 176~183

[11] W. Eckert, H. Niemann. Semantic Analysis in a Robust Spoken Dialog System. In *Proceedings of ICSLP-94*, Japan, vol. 1, pages 107~110

[12] Manny Rayner, David Carter. Hybrid Language Processing in Spoken Language Translator. SRI Cambridge Technical Report, 1997. Available from *http://www.cam.sri.com/tr/crc064/ paper/paper.html*

[13] Richard Schwartz, Scott Miller. Language Understanding Using Hidden Understanding Models. In *Proceedings of ICSLP-96*, USA, vol. 2, pages 997~1000

[14] M. Epstein, K. Papineni, et al. Statistical Natural Language Understanding Using Hidden Clumpings. In *Proceedings of ICASSP'96*, USA, vol. 1, pages 176~179

[15] Harold L. Somers. Current Research in Machine Translation. *Machine Translation*, 7:231~246, 1993

[16] Muriel Vasconcellos. What Do We Want from MT ? *Machine Translation*, 7(4): 293~301, 1993

[17] Doug Arnold, Louisa Sadler et al. Evaluation: An Assessment. *Machine Translation*, 8(1): 1~24, 1993

[18] Pamela W. Jordan, Bonnie J. Dorr et al. A first-Pass Approach for Evaluating Machine Translation Systems. *Machine Translation*, 8(1): 48~58, 1993

[19] Hiyan Alshawi and David Berkley. Spoken Language Translation at AT&T Labs. *Report of AT&T Labs*, March 1998

[20] 杨家沅，林道发，罗万伯等．连续英汉语音翻译系统的设计和实现. 声学学报. 1992, 17(5): 327 ~ 333

[21] Wen Gao, Bo Xu, et al. Chinese-English Spoken Speech translation. In *Proceedings of CJSLP'97*, pages 1~5

[22] Akio Yokoo. System Integration and Technology in Support of Research. *ATR Journal*, vol. 1, June, 1998, pages 30~31

[23] The C-STAR Interchange Format. Available from *http:/arrow.boltz.cs.cmu.edu:8090*