

面向受限领域的汉英语口语翻译

王海峰** 高文** 李生*

* 哈尔滨工业大学 计算机系

+ Motorola-NCIC 先进人机通信技术联合实验室

摘要: 口语语音翻译系统的应用不但有着显著的科学上和工程上的意义,而且有着重要的经济和文化上的影响。本文讨论了口语翻译相对于文本翻译的一些特有问題,提出了一种面向受限领域的基于实例的汉英语口语翻译方法,并介绍了一个面向会面安排领域的汉英语口语语音翻译的原形系统(EasySchedule)的口语翻译模块的实现。

Chinese - English Spoken Language Translation in Limited Domain

Haifeng Wang** Wen Gao** Sheng Li*

* Department of Computer Science and Engineering, Harbin Institute of Technology

+ Motorola-NCIC JDL for Advanced Computer and Communication Technologies

Abstract: The utility of spoken language translation will bring about not only obvious scientific and engineering significance but also major economic and cultural impacts. In this paper, we firstly discuss some important ways in which spoken language differs from its written counterpart. Then we propose an approach for Chinese-to-English spoken language translation on the basis of example-based machine translation method and describe the implementation of the translation module of EasySchedule — a prototype Chinese-English spoken speech translation system for meeting scheduling.

1. 引言

开发口语语音翻译(Speech-to-Speech Translation)系统是语音识别、机器翻译、自然语言处理、乃至整个人工智能界的终极目标之一。它的实现不仅需要准确的非特定人、大词表的连续语音识别,恰当表达话者意图的机器翻译,流畅的目标语语音合成,而且需要各个功能模块作为一个整体的有机结合。

口语语音翻译的应用除了显而易见的科学上和工程上的意义外,还将带来巨大的经济和社会效益。甚至一个小型的面向特定领域的语音翻译系统,如面向电话号码查询、宾馆客房预订、航空公司机票预订等小任务的系统都将为不会当地语言的人带来巨大方便,减少因语

言障碍而带来的不便。

虽然与口语语音翻译相关的领域如机器翻译、语音识别等已有很长的发展历史，但直到八十年代末，国际上才出现了有关语音翻译的研究[1,2]，而在国内，尚未见到有关口语语音翻译的文献。

1993年成立的国际语音翻译联合会（Consortium for Speech Translation Advanced Research, 简称 C-STAR）是一个以口语语音机器翻译为基本研究目标的国际合作组织，现在共有来自12个国家的共20个成员，包括美国的卡耐基-梅隆大学（CMU）、麻省理工学院（MIT），日本的ATR-ITL（Advanced Telecommunications Research Institute International - Interpreting Telecommunications Research Laboratories），德国的卡尔斯鲁厄大学（Karlsruhe Univ.）、西门子公司（Siemens）等。我国的哈尔滨工业大学和中科院也于1996年以观察员身份参加入了该组织的活动，目前主要以先进人机通信技术联合实验室（JDL）为基地承担汉英口语语音翻译系统的研究与开发，与CMU等单位相一致，目前将领域定为会面安排（Meeting Schedule），以便于以后各方系统的互连。

2. 口语翻译的困难

口语语音翻译系统需要包括语音识别、机器翻译和语音合成等几大功能模块。对于其中的机器翻译模块来讲，由于其输入是语音识别模块对口语语音进行识别的结果，所以它除了要面临一般文本翻译的共同困难外，还要处理一些口语翻译所特有的困难，具体来说主要包括以下几个方面[3,4]：

- 错误输入：口语会话中经常包括一些错误的输入，如错误的开始、重复、没有结尾等。例如：句子“哎，咱们...咱们中午吃什么？”，“明天，噢不，后天有个好电影。”

- 噪声输入：错误输入是由于说话人引起的，而噪声输入则是由于语音识别模块的错误识别所引起的。

- 严重的省略句：口语对话时，很多句子是高度省略的，只有考虑相应的上下文环境，才能对省略的部分给予恰当的恢复。

- 多种可选输入：语音识别模块所给出的往往不是一个唯一的句子，而是附有概率信息的词格形式（Lattice）或N-Best形式，口语翻译模块需要针对这种情况作出分析和选择。

针对口语翻译的特有特点和各种语言本身的固有特点，各国研究者采取了各自不同的策略进行分析和翻译。例如：

CMU的JANUS系统采用了扩展的广义LR算法（GLR*）和基于概念的分析（Phoenix）相结合的策略[5, 6, 7]；

SRI International的SLT（Spoken Language Translator）采用了一种生成多个可选的中间表示形式，再在一定的语法和统计信息制导下进行选择的方法[8,9]；

ATR采用了基于实例的方法与基于规则的方法的混合策略[10]。

3. EasySchedule 中口语翻译的实现

3.1 系统概述

现在，JDL 已实现了一个面向会话安排领域的汉英口语语音翻译的原形系统——EasySchedule。其中的口语翻译部分采用了基于实例的基本翻译策略，并辅以一些基于规则的处理，其基本结构如下：

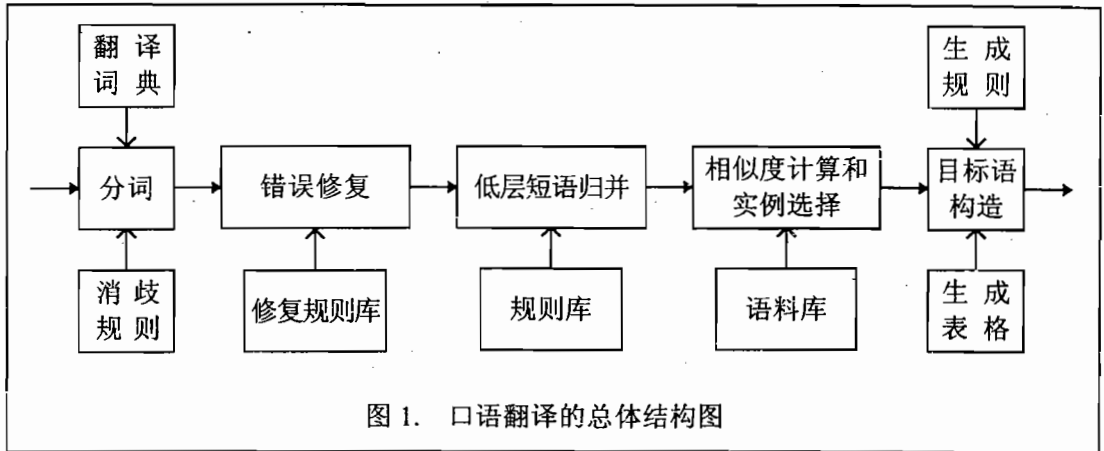


图 1. 口语翻译的总体结构图

3.2 面向口语翻译的汉语自动分词

正如陈力为先生指出的那样，汉语中什么是词，到现在并没有一个公认的定义^[11]。一般说来，词的划分是与具体的任务域密切相关的。对于汉语，语音识别和机器翻译对词的需求就有所不同。虽然二者都需要词的划分，但在语音识别中词的单位不能太大，否则将造成搜索空间的过大；而在机器翻译中，为了能得到准确的译文，一些在语音识别时被分开的词必须被合并而作为一个整体进行处理。例如：在语音识别时，“星期一”被作为两个词“星期”和“一”进行处理，而由于在英语中“星期一”有着一个单独的词“Monday”，所以在机器翻译时，它必须作为一个词处理。为此，在语音识别之后，我们根据机器翻译的需要对其结果进行了进一步的分词处理。在 EasySchedule 中，我们采用了一种能够容忍并忽略一些干扰信息的强健的分词算法。算法本质上是基于最大匹配的，此外，引入了生词的识别和干扰信息的略过机制。

3.3 口语中错误的修复

口语与书面语相比，一个重要特点就是由于说话者的失误而存在错误，如错误的开始、重复、没有结尾等。如果不作合适的处理，将对翻译的质量产生严重影响。

以下是一些需要修复的句子的实例：

- 1) 明天明天上午我有会。
- 2) 这个这个时间可以。
- 3) 明天上午九点上午十点吧。
- 4) 下午两点哦不三点吧。
- 5) 我星期二上午十点上八到十点有会。

以上例句中的斜体部分为待修复项（即需要被修复的部分），加黑的字为修复项，带下划线的部分为修复标志。在修复过程中需要识别出这三者，然后只保留修复项，而删除待修复项和修复标志。在上述例句中，1）、2）中的修复项是待修复项的直接重复，3）中修复项是待修复项的部分替换，4）中有明显的修复标志，5）可以认为是在待修复项中间插入了一些新内容而形成修复项（插入了“八到”）。

但在汉语中，并不是所有符合上述形式的都需要修复，例如：

- 6) 有变化给我打电话，我的分机是**五五六**。
- 7) 让我**考虑考虑**。
- 8) 你**想不想**去。

例句6)属于数字的重复，7)、8)则是汉语的重叠现象，它们并不在被修复的范围之内，在实现修复功能时须严格加以区分。

综合以上各种情况，我们设计了一种基于打分机制的修复算法，算法给每一个候选的待修复项一个打分。对于得分较高的，说明其需要修复的可信度较高，系统将进行自动修复；对于得分较低的，则说明其需要修复的可信度也较低，系统将和用户进行交互，人工决定是否修复。

3.4 低层短语分析

在汉语中，一些汉字必须被合并为一个整体来加以考虑，而由于它们无法被一一列举在词典中，所以也就无法在基于词典的分词过程中将它们合并。例如数词“二十一”，在进行句子分析或相似度计算时，应把它作为一个整体进行考虑，但要把这类数词都放入字典又显然是不合适的。在会面安排领域，这类表示时间和数字的短语是大量存在的，为此，我们在查找相似例句之前引入了一个专门的基于规则的处理模块来处理这类现象。对于象“二十一”、“一九九六年十一月二十六日”、“九点二十五”、“下周二”这类现象可以正确的归并，并动态地给出恰当的译文。

在低层短语分析时，我们还对汉语中的离合词、重叠词等特殊现象采用基于规则的方法进行了一些处理。

3.5 基于实例的机器翻译

基于实例的机器翻译方法最早由日本的长尾真教授在1984年提出[12]，国内也有学者作过将这种方法用于机器翻译实践的尝试[13]。这种方法的基本思想是比照相似句子的翻译实例来进行翻译，它的最主要特点就是把翻译实例作为它的主要翻译知识源。基于实例的机器翻译方法有很多优点，对于口语翻译来讲，有一点是尤为重要的，那就是根据实例，而不局限于严格的规则约束，这种灵活性正适合口语语言灵活多变的特点。

3.5.1 语料的处理

为了计算输入句子与语料库中例句的相似度，必须先对语料库进行适当地处理。我们根据需要首先对双语语料库的汉语部分进行了分词以及词性和语义类的标注，并对双语句对作了短语一级的对齐。

3.5.2 相似度计算与实例选择

在基于实例的文本翻译中,通常都需要参照多个例句来生成一个输入句子的各个部分的译文,然后再进行片段的合并[14]。但对于受限领域的口语翻译,鉴于口语句子一般都很短,而且很少象书面汉语那样多层嵌套,因而我们采用了整句匹配的方法。实例的选取是基于句子相似度的计算的,通过对输入句子与实例库中的例句的相似度的计算,最后选出一个相似度最大的例句。相似度的计算过程主要包括三步:词的相似度、上下文环境的环境相似度以及全句的相似度。计算时主要考虑了以下因素:

- 词之间的相似度,包括词、词性和语义类的相似;
- 词所在的上下文;
- 词在句中的位置;
- 句子的语气;
- 句子的长度;等。

经计算后,每一个候选例句都会得到一个 0 ~ 2.0 之间的值,如果存在相似度大于 1.2 的例句,系统将选取其中具有最大值的例句作为最相似实例,否则认为库中的实例皆不合适,这时系统将对输入的源语言句子的各个词的译文做一些简单处理,然后顺序输出作为译文。

3.5.3 目标语句子的构造

当得到一个最相似的实例后,就可以参照这个实例构造输入句的译文了。这时的的工作主要是关键词的替换和译文的形态处理。在形态处理过程中,引入了一些英语的形态变化规则和不规则变化表。

3.5.4 匹配速度的加快策略

为了加快匹配速度,使系统尽量满足实时的需要,采用了对实例对分组的策略,在匹配过程中,若某组中的一个例句得到了一个较低的相似度的值,则该组中的其它句子将不再被继续计算。

4. 结束语

现在,我们已根据上述方法初步实现了一个口语翻译系统,并实现了与语音识别模块的联调,完成了一个能初步运行的面向会面安排领域的口语语音翻译的原形系统。以下是系统的部分运行结果:

有点儿事想与你谈谈	I have something to talk about.
让我看一看	Let me see.
今天整个下午都不行	It doesn't work all the afternoon.
明天上午九点半到十点怎么样	What about 9:30 to 10:00 tomorrow morning?
九点半可以	9:30 is OK.
那就这样	That's it.

对于口语翻译模块,我们将继续在以下几方面多做工作:

- 对于口语对话翻译,句间的上下文信息对一个句子的理解,尤其是对省略句的理解

有着举足轻重的作用，我们将在系统的进一步完善中考虑将其引入；

- 引入对语音识别模块输出的 Lattice 或 N-Best 形式的结果的选择能力；
- 引入更强有力的消歧手段（如基于统计的方法）来支持歧义的消除和译文的选择；
- 引入实例的自动学习模块，以将曾经翻译正确的输入句子加入实例库，从而使系统越用越好。

参考文献

- [1] M. Tomita, M. Kee, H. Saito, et al. *Towards a Speech-to-Speech Translation System*. Journal of Applied Linguistics, 3.1, (1989)
- [2] K. Ogura, I. Sakano, J. Hosaka, et al. *Spoken Language Japanese-English Translation Experimental System (SL-TRANS)*. TR-I-0102, ATR Interpreting Telephony Research Laboratories, (1989)
- [3] Alon Lavie, Donna Gates et al., *Multi-lingual Translation of Spontaneously Spoken Language in a Limited Domain*. in Proceedings COLING-96, (1996)
- [4] Hiroaki Kitano, *Speech-to-Speech Translation: A Massively Parallel Memory-Based Approach*. (Kluwer Academic Publishers, Boston, 1994)
- [5] Alex Waibel, *Interactive Translation of Conversational Speech*, IEEE Computer, 29(7), (1996, 7)
- [6] Alon Lavie, *GLR*: A Robust Grammar-Focused Parser for Spontaneously Spoken Language*. PhD's thesis of CMU, (1996. 5)
- [7] L. J. Mayfield, M. Gavalda, Y-H. Seo, et al. *Parsing Real Input in JANUS: A concept-based Approach to Spoken Language Translation*. Proceedings of TMI-95, (1995)
- [8] M-S Agnas, H Alshawi, I. Bretan, et al. *Spoken Language Translator: First year Report*. SRI technical report CRC-043, (1994)
- [9] M. Rayner, D. Carter, *Fast Parsing using Pruning and Grammar Specialization*. Proceedings of ACL-96, (1996.4)
- [10] Yoshinori Sagisaka, *Recent Advances in Speech Translation Research at ATR-ITL*. Proceedings of C-STAR II 96: ATR International Workshop on Speech Translation, (1996.9)
- [11] 陈力为, *汉语书面语的分析问题——一个有关全民的信息化问题*. 中文信息学报, Vsol.10, No.1, (1996.1), 11-13
- [12] M. Nagao, *A Framework of a Mechanical Translation between Japanese and English by Analogy Principle*, in: Artificial and Human Intelligence, (Amsterdam, Netherlands, 1984), 173-180
- [13] 陈利人、陈群秀, *基于实例的日汉机器翻译方法中的句子相似度计算研究*. 全国第三届计算语言学联合学术会议论文集《计算语言学进展与应用》, 清华大学出版社, (1995), 257-263
- [14] Satoshi Sato, *MBT2: A Method for Combining Fragments of Examples in Example-Based Translation*, Artif. Intell. 75 (1995), 31-49