DOI:10.16262/j.cnki.1000-8217.1997.03.022

我国智能型机器翻译研究获重要成果

陈肇雄 王英姿

（中国科学院计算技术研究所，北京 100080）

[关键词] 机器翻译，自然语言处理，人工智能

1 项目简介

机器翻译是利用计算机技术实现从一种自然语言（源语言）到另一种自然语言（目标语言）的转换（翻译），为此，就必须要有对源语言和对目标语言的了解，并要有一种好的处理机制，在这种机制中，有着包括人类常识在内的各种各样的知识[1]。可以说，高性能机译研究是涉及语言学、计算机科学、人工智能等多个学科的综合性研究课题，它几乎涉及了语言信息处理的所有研究领域，是当前语言信息处理技术发展的“瓶颈”和突破口之一[2,3]。机器翻译在技术实现上需要模拟人的复杂思维过程，理解各种语言学知识，并建立起复杂的推理计算模型，以实现两种自然语言之间的转换。因此，在现有的技术条件下实现高性能机译系统，被公认为是一个国际性的技术难题。由于机器翻译是未来语言信息处理产业的核心组成部分，许多发达国家不惜为此投资巨金，把机译研究提高到国家科技发展战略高度予以支持，并开展全国性和多国性的联合攻关。

中国科学院计算所机译课题组，在国家自然科学基金委员会和国家“863”专家组的大力支持下，组织了全国七家主要研究单位的几十位机译专家和工程技术人员，经过长达十年的艰苦努力，在机器翻译理论研究方面取得了重要的研究成果，创立了智能型机器翻译理论体系，并成功地开发了国际上第一个智能型英汉机译系统，该系统技术水平居国际领先地位。

该项目的参加单位有：中国科学院计算技术研究所、中国科技情报所、中国科健公司、北京科技大学、北京联大自动化工程学院、解放军后勤指挥学院、总参第61研究所等单位。

智能机器翻译研究项目获1995年国家科技进步奖一等奖，其中基金课题“智能机器翻译理论研究”，重点研究了智能型机器翻译中的深层次的理论问题，为该获奖项目的重要部分。

2 智能机译研究的主要内容和技术难点[4-6]

人们对机器翻译存在着许多误解，其中之一就是把机器翻译技术等同于电子词典、袖珍型电子翻译机，如快译通等。事实上，机器翻译是一门高难度的交叉学科，一方面，机译技术的发展会受到相关学科的限制；另一方面，机译技术的突破又可带动相关学科的发展。如，通过对智能型机器翻译的研究，可发现人工智能技术中可能存在的局限性，并可从应用角度为新一代高性能计算机系统结构的研究提供基本依据，以促进计算机系统朝着具备知识处理、

本文于1996年10月28日收到.

复杂推理和友好用户界面方向发展。

从语言学角度看，机器翻译要重点研究机器词典和语言文法等，包括词典中的语法信息、语义信息、以及与词条相关的一些语用特征的设置和选择等，以及与具体自然语言相关的短语、句子及文章的结构规律信息和这些信息的表示等。

从计算机角度看，自然语言处理研究的侧重点主要是如何建立友好的软硬件支撑环境，包括：为计算语言学工作者提供字典信息辅助或自动收集工具，文法规律辅助或自动发现工具，实现面向具体文法的和通用的语言分析处理机制，以及大型数据库和大型知识库等[4]。

从认知心理学角度看，主要是为自然语言的计算机处理建立各种认知模型，研究人类语言的本源特征，以便建立包括非语言学知识的日常生活常识处理的计算模型。

从数学角度看，主要是为自然语言处理提供数学基础，为各种文法分析机制、字典管理机制和常识库系统的计算机实现提供效率上可行的算法理论等。

总之，智能机译研究几乎包括了自然语言处理的所有核心技术，其处理对象是灵活多变的自然语言，不但需要用一种机器能理解的形式化语言对人类的认知进行模型化，对无规则的、无限的自然语言进行形式化描述，同时还需要建立高效、多功能的计算机算法，对所表示的知识进行推理。自然语言表示具有极大的模糊性，其所涉及的知识十分庞大，被公认为国际性难题。国外一个机器翻译系统的建立大都要经历数年的理论研究，而仅仅从模型到实用系统就需耗费数百万、千万甚至数亿美元，花 5 至 7 年的时间。

3 我国机译系统的技术水平

我们在国家自然科学基金和国家“863”计划的支持下，完成了理论上和工程实现上的重大突破，仅利用国家自然科学基金的 6 万元和“863”计划的 44 万元人民币资助，便完成了国外需数百万、千万以至数亿美元才能完成的系统开发工作，并形成了智能型机译理论体系。研制成功的智能型英汉机器翻译系统，在系统翻译正确率、译文可读性以及开发周期等方面均有突破性进展，在国内外机译界产生了很大反响。在国家科委组织并主持的鉴定会上，包括中科院院士在内的一些著名专家一致认为[5]：“智能型英汉机器翻译系统 IMT/EC 研制组创造性地提出了智能型英汉机器翻译理论体系，在系统的开放性与一致性保证、复杂多义区分、上下文相关处理、基于不完备知识的推理、多种知识的一体化分析、机译知识的获取和应用等方面均有重大突破。应用该理论所涉及和开发的 IMT/EC 系统具有软件独立于具体文种，智能化程度高，翻译速度快，占用空间小，准确率高，译文可读性好等优点。IMT/EC 系统在理论基础、总体设计、系统实现和应用效果等方面，总体上已经超过国外同类系统，处于国内外领先地位。”

目前，该智能型英汉机器翻译系统已经拥有 10.5 万条基本词汇，25 万个汉语对应词，1500 条通用规则，15 万条特殊规则和成语规则。表 1 列出了智能型机器翻译与传统机器翻译在理论和技术上的比较情况[7,8]。

4 重要创新

依据智能机译理论，我们设计和开发的 IMT/EC 系统，与当前国际上比较成功的机译系统相比，具有以下理论和实现方面的重要创新和优点[4]：

（1）总体设计。提出并实现了：开放式的总体结构；独立于具体文种的软件环境；适用于语言分布式存贮的多包机译知识库结构；强有力的例外处理能力；高效的实现算法。

（2）语言工程。研究如何把语言学知识和用于机器翻译的一些非语言学常识进行归纳和形式化描述，以便适合于计算机处理。其中，语言学知识包括翻译过程需要用到的词法、语法、语义以及语用等知识；而非语言学知识包括机译过程中常常涉及的学科分类、背景文化知识以及专业知识等。具体创新工作有：适用于智能型机器翻译的SC文法体系；与SC文法相适应的可变换层次词典结构；高度形式化的多级语言特征体系和固定结构规则。

（3）翻译处理环境。研究如何应用形式化的语言学知识和非语言学常识，实现从源语言输入到目标语言输出的转化。这一过程包括词法分析算法、结构分析算法、上下文相关处理以及目标语言生成等分析和推理机制的实现技术[9]。我们的工作有：独立于具体语言的词法分析算法；综合运用语法、语义、常识的一体化分析技术；基于超前与反馈分析的上下文相关处理技术；基于不完备知识的智能推理技术；动态多路径选择技术；以及主特征相容合一匹配技术。

（4）知识处理环境。研究如何提供一套有效的软件工具环境，帮助语言学家归纳语言知识和简单的非语言学常识，实现这些知识的形式化描述，供翻译处理软件使用。我们的工作包括：面向多用户操作的知识一致性保护机制；分散应用特殊规则的处理技术；单一形式的知识表示形式；高效快速的知识信息压缩技术；面向对象的多包知识库结构；开放式的知识获取环境；规则精炼及知识重组。

（5）系统开发环境。提出了高效的分散式语言工程实现方法；开放的系统软件环境；独立于具体文种的软件支撑环境；高效处理算法的实现技术。

表1　传统机器翻译理论和技术与智能型的比较

技术类型	语法型	语义型	知识型	智能型
应用知识类型	语法知识为主	语义知识为主	常识知识为主	语法、语义知识及简单常识一体化
优点	较好保持源文结构特点易于实现	便于多义区分	理解原文实现意译	有机综合语法语义型的优点，可实现多路径交叉分析
不足	不便于多义区分	规则不易总结系统很难实现	难以实现实用系统	新理论
应用范围	早期系统	实用系统很少	无法推广实用	已实际应用
典型系统	EUROTRA	ODAMT	KBMT	IMT/EC

5　项目的产品化推广

项目组开发成功的智能型英汉机译系统不仅开展了跨专业的全文翻译服务，还解决了机器翻译系统在袖珍机的时空限制下难以实现的国际性技术难题，并成功地开发了世界上第一部袖珍电子翻译机——“快译通EC863”系列产品。目前产品有：（1）智能型跨专业全文机译系统：电力专业，海洋专业，商贸专业和计算机专业等；（2）多语种机器翻译系统：汉英翻译系统，俄汉翻译系统，德汉翻译系统，英日翻译系统；（3）袖珍电子翻译机产品系列：科

智电译机 863H，快译通英汉 863A，快译通英汉 863B，快译通英汉 863C，EC6000 和 EB8000。

6 科学意义及经济效益

该项目的开展，促进了机译技术和相关学科的发展，培养了一批从事语言信息处理的科研人才；发表专著 2 部，在国内外学术刊物及会议上发表论文 80 余篇；培养博士 8 名，硕士近 20 名。该项目创造出了显著的经济效益，为国家直接创汇 860 万美元，十年合同额为 2444 万美元。

7 今后的研究

计划在智能型机器翻译研究成果的基础上，利用我们的技术优势和已形成的人才队伍，开展以下三个方面的研究：(1) 自动电话翻译的研究，该项研究已经起步，并已取得阶段性成果；(2) 军事科技情报领域智能型机器翻译系统研究，首先开展航空、航天领域的机译研究；(3) 信息高速公路上智能查询和实时翻译系统的研究。

参 考 文 献

[1] 刘涌泉．中国的计算机翻译．情报科学，1980.
[2] 陈肇雄，梁南元．自然语言处理理论研究发展战略．计算机科学技术，自然科学学科发展战略调研报告，北京：科学出版社，1994 年．
[3] 陈肇雄．智能型机器翻译研究进展．见机器翻译研究进展．北京：电子工业出版社，1992.
[4] 陈肇雄，高庆狮．智能化英汉机译系统 IMT/EC. 中国科学，A 辑，1989 (2)．
[5] 智能型机译系统研究小组．智能型英汉机器翻译系统 IMT/EC. 申报国家科技进步奖材料，1995.
[6] 陈肇雄，陈强．智能型机器翻译与语言信息处理产业．高技术通讯，1991 年．
[7] 冯志伟．国外实用化的机器翻译系统．中国计算机用户，1989 (9)．
[8] 黄河燕．中国机器翻译研究现状．计算机世界月刊，1994 (2)．
[9] 姚天顺等．自然语言理解导论．东北大学计算机科学与工程系，1993 年 2 月．

RESEARCH ON AN INTELLIGENT MACHINE TRANSLATION SYSTEM

Chen Zhaoxiong Wang Yingzi
(*Institute of Computing Technology, CAS, Beijing* 100080)

Key words machine translation, natural language processing, artificial intelligence