计算机科学1996 Vol. 23 No. 3

机器翻译研究的现状与发展方向*)

赵铁军 李 生 高 文
(哈尔滨工业大学计算机系 哈尔滨150001)

摘 要 This paper outlines a picture of current research and future directions in machine translation. It also introduces some views and proposals held by famous researchers in MT field. Author's points about several issues are proposed in the paper.

关键词 Machine Translation, Corpus, Evaluation.

1 机器翻译的沿革

机器翻译(MT)的历史是曲折而饶趣的。计算机刚一发明,就有人想到用它来进行自然语言的翻译。1949年美国人 Weaver 的著名备忘录第一次点燃了人们对 MT 的热情。1954年1月 Georgeton 大学和 IBM 公司的合作成功又把各界引入了对 MT 充满信心的时代。此后十来年的 MT 研究繁荣期却因1965年8月美国科学院著名的 ALPAC 报告而一蹶低谷,研究经费几乎全被削减,竟然出现了十年的沉寂时期[4]。

从七十年代中期开始,MT 研究重焕生机,发展了 MT 的第二代(以下简称 G2)技术,这些技术今天仍发挥着重要作用。一些 MT 系统的应用成功也激发了人们的兴趣和投资,例如加拿大在七十年代末安装的 METEO 系统,将天气预报从英语译成法语,在限定句子范围内几乎不需要人工后编辑。国际上八十年代的大型翻译工程有欧共体的 Eurotra 系统(该系统据报道已被当今国际软件巨人 Microsoft 买下)和日本的基于中间语言的 MT 系统等。AI 技术也在不断地渗透和应用到 MT 领域,基于知识的 MT 体系结构应运而生,如美国 Garnegie-Mellon 大学 KBMT-89系统。

G2MT 系统的主导思想是面向句法、基于规则的转换方法,研究者们追求用良好的形式化方法来处理自然语言现象,因而称为理性主义方法。尽管在 G2MT 体系结构上花了大量精力和经费,但批评者们认为 G2系统的译文质量与第一代系统相比并没有明显的改善。

随着计算机硬件环境的不断增强,MT 研究者开始寻求新的方法:从大规模的原始语言素材—语料库(corpus)中抽取翻译规律并获得较满意译文。这类新方法的出现标志着 MT 新体系的诞生,被称为经验主义方法。根据英国学者 Hutchins 的说法[3],由于基于实例的(以下简称 EBMT)和基于统计的(以下简称 SBMT)语料库方法先后出现,1989年可以看作第三代 MT 的开端。

在 MT 研究的不断深入发展的情况下,人们发现现有的 MT 系统距离人类的要求还相去甚远,远没有得到广泛使用,造成这一现状既有技术上的困难,也有人们对 MT 的认识问题。正是因为 MT 研究的现状不尽人意,所以引发了国际 MT 界的一场讨论。国际杂志《Machine Translation》第七卷第四期(1993)作为讨论专刊,发表了许多国际知名学者和专家的观点。这里简要介绍一下专家们的看法,有助于我们对 MT 现状的了解。

讨论是围绕英国学者 H. Somers 的一篇综述文章而展开的。文章作者对 G2MT 体系结构提出了尖锐批评,指出了"经典的第二代体系结构的错误是什么"的问题。其他学者和专家从不同方面讨论了当前 MT 研究中的主要问题和发展方向。不论反对或赞同 Somers 论点的人都承认,MT 研究确实存在很多困难,并阐明了要从不同角度客观地评价 MT 成果的意见。

按照 Somers 的说法,G2体系的特点可以概括为层次性(stratificational)和模块性。层次性是指 MT 采取由词到句的不同层次的分析过程,建立一种中间表示,然后完成由句到词的不同层次的转换和生成过程。这样的语言学处理过程和系统实现的模块性相辅,语言学过程和处理程序分开,源语言(SL)和目标语言(TL)模块的描述分开。那么 MT 的困难有哪些?

(1)认为 G2系统的特点造成了两个问题,一是系统的细节实现采取了自底向上方法,并且中间表

*)本文工作受国家863计划(863-306-03-06-04)支持。

示多少决定了目标语的生成；二是把保留结构的翻译作为MT的首选，因此还会造成不合乎习惯的、充满机器味的译文。

(2)人工智能(AI)技术的引入也不可能彻底解决MT的技术问题，因为对AI研究来说，从原型系统到实用化的中间还有很大的差距。

(3)八十年代以来，MT界出现了不少新的语言学理论，同时也出现了许多基于这些理论的MT系统。这些理论可以很快地开发出一个"玩具"系统，但在真正的应用压力下不可避免地脆弱。著名学者Wilks甚至从这种情况下得出两点结论：任何理论不管多么"愚蠢"，都可以成为某些有效的MT系统的基础；成功的MT系统很少将它们所宣称的理论贯彻到底。

(4)由于理论和技术上的问题，使得G2MT的翻译质量并没有比第一代好到哪里，除了少数系统外，几乎所有的商业MT系统仍然采用第一代MT技术。而G2系统的改善在很大程度上归功于好的软件工程和好的软件支持。

原因何在？他们的看法可以归纳为：

(1)MT缺乏一个普遍接受的理论基础，只是从AI和语言学理论借用了不少东西。MT不应是外部理论的简单应用，也不应定义为人类翻译的模拟。如何从许多个别系统中得到一种普遍抽象是当前MT研究的一个任务。有观点认为今后一段时间的MT进展不是理论的进展而是工具的进展。

(2)对MT系统期望过高。大规模真实文本中存在大量的词形变化、错误拼写、生词；不合乎语法规范的句子；语言的变化，旧词具备了新意；句法歧义；人工也难以辨认的句子结构；两种语言的差别所造成的TL信息的缺乏等等现象，MT系统尚不具备处理真实文本中所包含的所有这些现象的能力。另一方面，对于真实文本本身的无限性质，有限的MT系统不可能进行穷尽处理。实际上现存的任何MT系统都不可能处理从报刊杂志上任意剪下来的文章。因此，通用的MT至少目前是不可能的。

大部分专家都呼吁要对MT抱有更加现实和更加灵活的态度。MT界应该清醒地知道自己的局限，并且要让公众也知道。特别要避免对MT的某些不负责任的承诺，以免造成人们对MT的不信任感。

这两个问题，一个是MT本身的体系问题，一个是MT的实用问题。研究者们可能更关心前者，但后者更是大家所普遍关注的，甚至决定了MT的生命力。这两个问题相互联系着，相互制约着，不仅是MT的问题，也是AI中普遍存在的问题。关于MT现状的讨论很容易让人联想起1991年AI界的那场关于人工智能基础的争论。研究者们总想从个别中找到一种普遍有效的原理，并且建立起一个可以推而广之的形式体系，然而现实却远远不尽人意，经验主义方法的兴起，正反映了人们的一种新探索。

2 如何评价MT的成果

专家们从不同角度看待MT现状，得出的结论也各不相同。对MT评价的研究和对MT技术本身的研究一样重要，怎样能更客观公正地评价迄今为止MT所取得成果？如何为今后的正确评估建立一个标准？下面分别就这两个问题加以讨论。

对MT现状的评价，主要有下述一些观点。

(1)把MT商品系统与MT研究区分开来。商品系统的成功不能代表其技术的先进，因为商品系统必然要采用成熟的技术，而技术的成熟滞后于研究，从实验室到市场，MT系统要经过多年的努力，所以不能批评商品系统的技术就是过时的。

(2)多从用户关心的角度评价MT系统。用户的主要标准是降低翻译成本，不管是采用哪一代技术，MT系统的应用都要提高翻译速度，减少人工劳动，所以应该尽量满足用户的特殊要求，如专业文献翻译的特殊风格，大量的技术词汇，更为友好的前后编辑功能等。将用户反馈引入MT系统的建立过程，使其不再只是研究者主观上的一相情愿。

但也有人指出，如果把MT中的人工因素降低到毫无兴趣的前后编辑状态，在使用中也会遇到阻力。这实际上指出了交互式或计算机辅助翻译的某种未来的潜在危险。我们从实际的用户反馈中也认识到，尽管我们一再宣称全自动高质量的MT不现实，但用户还是希望一下子就能用上这样的系统。

(3)从需求和结果相适应的角度看待MT质量。将MT结果细化，分为若干层次，不同的需求对MT结果的评价是不同的。例如，70%的接受率对于将结果用于可出版的需要显然是不行的，但对于仅仅估价文献的可用价值的需要则已能满足。

不论对MT现状持悲观主义还是乐观主义的看法，都既要正视MT的困难，又不能一概抹煞G2取得的成果。不利用现有的词典和规则库是不行的。不可能一切从头开始，必须在现有的范围内进行改善。以实用主义作为MT的基础并不意味着要放弃理想，否则一切科学上的探索都会失去真正的内在动力。正如有的专家指出的那样，在未来，MT研究者不得不扮演科学的和商业的双重角色，以便随时在语言这个无底洞和其使用者之间作出正确的妥协。

尽快为MT系统的正确评估建立一个客观的面向实践的标准，已经成为当前MT界的当务之急，甚

至许多专家认为MT评价标准的缺乏已经影响了MT的进步。在MT评价方面要做的工作包括：

1)建立基准测试的源语言文本和目标语言文本的标准语料库和词典。对于不同的系统在一个相同的测试环境下测试，才能对其性能作出准确公正的评价。

2)给出客观实际的评价标准。有的专家已经对诸如"可读的"、"可懂的"、"保留意义和风格的"等主观标准提出了异议。但是如何才能建立客观标准，还未给出具体的内容，当然要考虑综合因素，如语言学的标准、计算技术的标准、经济的标准等。在大规模的统计基础上制订标准也许会更客观些。

3)快速和可靠的评价方法。

4)将MT评价作为一个单独的专门的研究方面。MT评价研究有自己的权利，而不必附属MT研究本身。

5)特别考虑系统扩大化的问题。许多研究者往往认为原型系统可以简单地通过扩大词典和知识库规模来扩大系统，系统的规模和成本呈线性增长的关系。而实际上其扩大化产生的问题远远不是那么简单，对于RBMT尤其如此，系统的实用化困难就在于此。扩大化问题可以看成是系统的鲁棒性问题，它和系统的模型、方法及词典规模等都有关系。正是因为存在系统的扩大化问题，所以不能把原型系统的功能或指标看成是系统潜在的能力。

6)不同的MT项目之间共享一部分资源(如词典)。由于词典的建造需要花费巨大的人力和财力，如果能够实现共享，则不仅可以节约资金投入，也可以为MT系统提供一定的共同基础，有助于MT的评测。

现在有关国际组织和国家已经组织了专门讨论会，出版了有关MT评价标准的指南。国际MT杂志亦出版了有关MT评价的专刊，但是MT的真正评价标准还有待完善。

3 MT近期发展方向

1988-89年IBM公司的Brown等人提出的MT统计方法，被认为是对MT正统方法的冲击。九十年代后，MT界掀起了基于实例(EBMT)方法的研究热潮。传统的MT方法在不同程度上也受到了SBMT和EBMT的影响。总的来说，MT研究有四个方向[5]：(1)传统的基于转换(以下简称TBMT)的方法；(2)将语言的理解集成进MT研究，使用更多的外部世界知识。这就是一般称之为基于知识(以下简称KBMT)的方法；(3)EBMT方法；(4)SBMT方法。下面我们简要地说明一下各种方法。

IBM的SBMT工作受语音识别研究的启发，应用了类似的方法，以大规模语料库(3百万句对)为基础，对源语言和目标语言词汇的对应关系进行统计，根据统计规律输出原文句子的译文句子。该方法根本没有使用语言知识，却也取得48%的正确率。

尽管大家都认为IBM的工作是一种新的方法，可是Wilks和台湾学者Su对"纯粹的"统计方法提出了异议。他们认为不能鼓励这种方法，必须引入高层语言模型，应该建立基于语料面向统计的方法。并且认为这种方法不一定能很好地作用于另一对语言(IBM工作的对象是英语和法语)，也不能产生高质量的MT，除非所有基于知识的MT观点全错。实际上IBM的研究者也承认这种方法不能解决语言中的长程依赖问题，他们当前的工作亦在词汇级的统计中结合某些规则的东西[7]。

EBMT的基础同样是大规模双语语料库，其基本思想很简单。对于输入的任何一个句子S，通过一定的评分机制在语料库中匹配一个最相近的句子S′，这样S′在语料库中的译文T′就可以作为S的译文。需要的话，可以对T′进行必要的修正，使之更接近或完全等同于S的译文T。可能与S相近的句子不止一个，那么选取最佳候选的算法便是EBMT的关键技术之一了。在基于实例的方法中，除了大规模的语料库以外，还要建立大规模的同义词词典。在词汇一级进行语义相似度的距离计算，并最终算出句子间的语义"距离"。

语料库方法中的一个重要研究内容是语料库的加工。语料库在应用中不断完善、扩展，越用越好。通过语料库来学习MT所需的语言知识也被看作是解决知识获取瓶颈的一个最富有潜力的方式。

传统的TBMT的处理过程分成分析、转换、生成三个步骤，核心是语言加工规则。以往的研究重点放在分析上，认为只要将源语言分析彻底了，就可以转换成好的译文。现在这种看法有所改变，目标语的生成也作为一个重要的研究内容来看待了，因为MT的结果最终要落实到目标语译文，特别是在那些要求译文质量很高的场合下。

KBMT的根本也是以规则为基础，但要加深对语言的理解。从人进行自然语言的翻译过程来看，许多情况下是先理解了语言的意思，然后再用目标语叙述。这个过程不是直接从源语言到目标语言的，而是要经过以语义的细粒度知识为基础的某种中间形式。基于中间语言的MT方法就是一种典型的KBMT。另一方面，在翻译过程中，总有一个附加的知识库在起作用，这个知识库中可以包括篇章级的上下文(语境discourse)，不同背景下的用法(语用

pragmatic)，以及文章所属的特定领域(领域知识 domain knowledge)等等。此外，使用语义映射器(semantic mapper)将句子映射为某种形式的语义网络，再从语义网络生成句子，也是一种 KBMT 方法[1]。

这样看来，MT 研究有两大类：一类以规则作为翻译的"引擎"，可称为基于规则的方法(RBMT)，包括 TBMT 和 KBMT；一类以语料库作为翻译的基础，可称为基于语料库的方法(CBMT)，包括 SBMT 和 EBMT。当然，在实际系统中还有许多其他的工作方式，但基本思路不超出这两大类。

Hutchins 认为：不管 MT 系统的类型如何，其基本思想仍然是如下两点：MT 从文本开始又以文本结束；MT 的核心处理是意义对等条件下的词汇和结构转换。一句话，离不开翻译本身。

4 混合策略

许多专家认为，MT 的真正进展大多来自混合方法[7]，不论采取何种途径进行 MT 研究，建造 MT 系统，单一的方法都很难达到预期的效果，这也许就是"一个 MT 系统很少把它的理论贯彻到底"的原因。目前，由于 MT 的 SB 和 EB 方法尚处于建立实验系统或很难满足自然语言约束的阶段[5]，因此还不能直接为大规模的翻译服务。有鉴于此，将经验主义方法与传统的 RBMT 体系结合起来，就成为国际上许多 MT 研究者的共识[1,2,6]。

如何把各种 MT 方法结合起来，正是当前国际 MT 界的一个努力方向。混合策略之一是将多种 MT 方法集成在一个 MT 环境之下，各个 MT 引擎同时或分别工作。这就是多引擎 MT 体系[5]。这种混合策略的目标就是要改善系统的结果，是一种面向结果的策略。该体系有两种工作方式，一种是译后判定，称之为 Best Out Segment(简称 BOS)；一种是译前判定，称之为 Dispatcher-Based(简称为 DB)。BOS 方式的工作过程是：对于同一篇输入，让各个 MT 引擎(如 KBMT、EBMT、TBMT)同时工作，在各个子系统的输出译文中挑选最佳的语段加以组合，生成最终的译文。DB 方式的工作过程是：对于同一篇输入，先将其拆成合适的语段，判断各语段适合哪个 MT 引擎，然后发送给相应的引擎，最后组合各个引擎的输出。

混合策略之二是面向 MT 过程，也可分为两种形式。一是以一种方法为主，辅之以其他方法来改善系统的译文。例如，语料库分析用于校正分析和转换语法；在标准的 TBMT 的转换阶段使用双语语料库；统计信息用于语言消歧和 KBMT 的某些阶段等[5]。第二种是将不同的 MT 思想方法融合在一起，形成新的 MT 方法。我们正在进行的课题研究就是把传统的 TBMT 与新近流行的 EBMT 思想混合起来，构成了一种基于模式(pattern-based)、面向实例(Example-Oriented)的 MT 方法，其具体作法是：在词汇和短语一级的源语言分析上仍采用规则方法，在句子一级的翻译上以双语语料库中的目标语句子为牵引，比照实例建立两种语言句子间的转换模式。这样做的好处是：(1)翻译知识拟人化(或具体化)，帮助系统实现者克服目标语语感不足的困难；(2)转换在句子一级作一定的抽象，避免了语料加工的一些困难；(3)兼顾实例和语法，生成高质量的译文。

多方面地探索 MT 的实现方法，综合考虑涉及 MT 系统实现的各个因素，有下述一些问题需要注意。

(1)放弃 GPMT 目标，选择受限的子语言作为处理对象。

(2)建立各种各样的 MT 系统，以满足用户的各种不同需求。

(3)将机器学习引入 MT，更多地考虑知识获取问题，把它放到比知识表达更严重的地位。

(4)人机合作交互式进行 MT，如对话(dialogue)MT，在一个固定模式的文本基础上由人机交互填写。

(5)加强 MT 环境方面的研究。如大规模的面向 MT 的词典的建设；更加友好的用户界面；基于语言学的面向 MT 的编辑环境等。

(6)继续 AI 技术应用于 MT 的努力。如加深对语境的理解；更合理地集成世界知识；引入黑板技术，进行并行处理；采用联接主义方法等。

参 考 文 献

[1]Carbonell, J. G. et al., The KANT Perspective: A Critique of Pure Transfer (And Pure Interlingua, Pure Statistics…), TMI-92

[2]Grisman, R., Kosaka, Michiko, Combining Rationalist and Empiricist Approaches to MT, same to [1]

[3]Hutchins, J., Latest Developments in MT Technology: Beginning a New Era in MT Research, MT Summit IV, 1993. 7

[4]Makoto Nagao, Machine Translation: How Far Can It Go?, Oxford University Press, 1989

[5]Nirenburg, S., et al., Two Types of Adaptive MT Environmemts, COLING-94

[6]Su, Kelh-Yih, Chang, Jing-Shin, Why Corpus-Based Statistics-Oriented MT, same to [1]