第 1 卷 第 1 期
2012 年 5 月

集 成 技 术
JOURNAL OF INTEGRATION TECHNOLOGY

Vol. 1 No. 1
May. 2012

机器翻译技术现状与展望

刘 群

（中国科学院计算技术研究所智能信息处理重点实验室 北京 100190）

摘 要 本文对机器翻译技术的研究现状进行了全面介绍，分析了亟待解决的核心问题，并对机器翻译的未来发展前景和趋势提出了自己的设想。

Machine Translation Research and Technology

LIU Qun

(Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190)

Abstract This paper gives a comprehensive introduction to the status of current machine translation research and technology, and analyzes the key problems to be resolved. Finally our idea of the future trends and prospects of machine translation are put forward.

1 机器翻译发展现状分析

从1954年美国乔治敦大学和IBM公司进行世界上第一次机器翻译实验以来，机器翻译的发展已经经历了半个多世纪的时间。

在这半个多世纪的时间中，特别是近十余年来，机器翻译的研究和应用都发生了巨大的变化。

从应用角度看：

（1）机器翻译的应用范围大大延伸，现在Google公司推出的互联网机器翻译系统已经可以支持60多种语言的互译，微软必应、百度、网易有道等互联网公司也都推出了自己的互联网机器翻译产品，机器翻译已经成为很多普通网民日常使用的工具，机器翻译在老百姓的日常生活（如教育学习、购物、旅游、甚至恋爱交友）中的应用已经非常普遍；

（2）计算机辅助翻译在市场上取得巨大成功，已经形成了完整的产业，自动翻译的结果也开始在计算机辅助翻译中得到了初步的应用，一些著名高校中已经开设了专门的计算机辅助翻译专业课程；

（3）机器翻译的应用形式更加多样化，云计算和移动终端的普及为机器翻译的应用带来了更多的应用形式，口语翻译、文字扫描翻译、照相翻译等都已经开始有了实际的应用。

应用方面发生的句法变化得益于机器翻译技术的进步。这些进步主要体现在：

（1）统计机器翻译取得巨大成功，从基于词的模型发展到了基于短语的模型和基于句法的模型；

（2）机器翻译的统计方法和规则方法走向融合；

（3）机器翻译系统开发效率大为提高，开发一个机器翻译系统的时间从数年缩短到数周；

（4）对大部分语言和领域，机器翻译系统的质量都有了明显的提高。

我认为，从整个机器翻译发展的历程看，机器翻译方法的演变可以主要体现为三个方面，我用三个“分离”来表示：

（1）程序与知识表示的分离；

（2）分析、转换与生成的分离；

（3）模型与搜索的分离。

其中，最后一个“分离”的思想，是IBM公司在1980年代末提出的，标志着统计机器翻译方法的诞生。模型与搜索的分离意义是非常重大的：独立的模型，意味着我们可以引入成熟的数理统计方法来对不同的机器翻译结果进行精细的定量化比较，而搜索算

刘群，研究员，博士生导师，中国计算机学会理事，中国中文信息学会常务理事，主要研究领域为中文信息技术、自然语言处理、机器翻译。E-mail：liuqun@ict.ac.cn。

法可以保证我们在一个庞大的搜索空间中尽可能找到合理的翻译结果。

我们仔细考察一下基于规则的机器翻译和统计机器翻译这两种范式的区别。

基于规则的机器翻译系统的开发流程如下图所示：

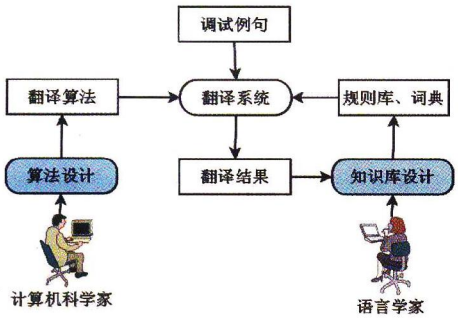

图1　基于规则的机器翻译研究范式

我们可以看到，在这种范式下，机器翻译系统的改进的循环主要依靠语言学家来推动。语言学家根据翻译结果不断地调整知识库，以提高机器翻译系统的性能。而对于计算机科学家来说，一旦完成了系统设计和实现，基本上就没有办法再对机器翻译的结果加以干涉了。

而统计机器翻译的开发流程为：

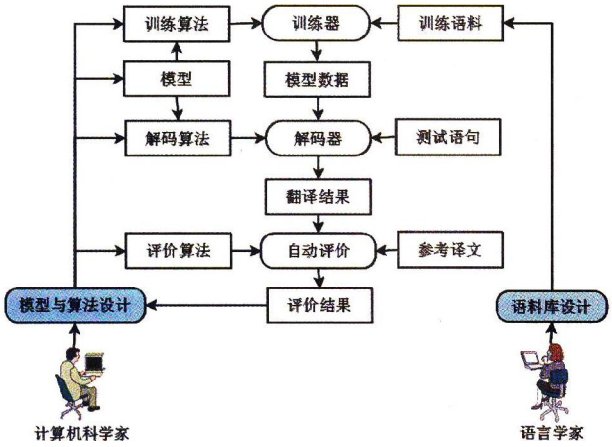

图2　统计机器翻译研究范式

我们看到，统计机器翻译的开发流程远比基于规则的机器翻译更复杂，模块的划分更加细致。最重要的一点区别是，机器翻译系统性能改进的循环不再靠语言学家来推动，而是主要靠计算机科学家来推动。语言学家一旦完成语料库设计和构建，就不再参与机器翻译系统的改进，而计算机科学家要通过不断调整模型、参数和算法来提高机器翻译系统的性能。这种方法保证了在给定语言数据的基础上，我们可以充分利用计算机强大的计算能力来达到最好的翻译效果。

不过，机器翻译的水平远没有达到理想的程度，下面我们对机器翻译面临的问题进行一些深入的分析，并试图提出我们的想法。

2　机器翻译发展面临的主要问题

机器翻译近年来虽然取得了长足的进步，但仍然面临很多的问题，从应用的角度看，这些问题主要体现在：

翻译质量还不够高：对于某些特定的语种和领域来说，机器翻译已经达到了可以接受的水平，用户借助于机器翻译系统已经基本可以了解原文的主要内容，如Google提供的法语、阿拉伯语到英语的翻译（新闻领域）。但就一般情况而言，机器翻译的质量还无法满足用户的需求。典型的例子是英汉和汉英翻译。汉英翻译几乎是机器翻译研究最多的语种，语料库的规模也极为庞大，达到了百万至千万句子对的数量级，但机器翻译系统的性能还是不能令人满意，翻译结果不通顺乃至意思完全无法理解的情况仍然比比皆是。

语种和领域的支持还不够多：虽然Google翻译支持的语种名义上达到了60多种，但实际上除了一些主要大语种到英语的机器翻译在口常用语和新闻领域达到了较高水平以外，其他大部分语种之间的翻译，尤其是非欧洲语言之间的互译，水平还远远不能令人满意；而对于日常用语和新闻以外的领域，目前也没有成熟的商业化机器翻译系统可供使用。

翻译结果还不够可信：对于很多用户来说，在机器翻译准确率不高的情况下，如果机器翻译系统能够准确说明哪些翻译结果是可信的，哪些不够可信，仍然可以为用户节省大量的时间和金钱。但现在的机器翻译系统在这方面还无能为力，导致机器翻译的结果正错混杂，机器翻译的可用性大大降低。另外，即使对一些比较窄的领域，机器翻译也还做不到高可信度。

系统使用还不够方便：虽然机器翻译质量还不能完全令人满意，但如果能找到有效的应用方式，机器翻译仍然可以被用户接受。目前的机器翻译总体上应用方式还没有做到非常友好。以跨语言检索为例，虽然研究人员普遍认为跨语言检索具有广阔的应用前景，也开展了深入的研究，Google搜索引擎也加入了相应的功能，但跨语言检索依然没有被用户普遍接受。语音翻译、扫描翻译、照相翻译等功能，虽然已经有很多产品出现，但也都还没有被用户普遍使用。

机器翻译系统的应用模式还有待扩展。

应用方面的问题归根结底还是技术上的问题。以下我们对目前机器翻译技术上面临的主要问题有进行一些深入的分析。

2.1 深层次语言知识的运用

机器翻译问题本质上还是个语言问题，机器翻译问题的最终解决也必须依靠语言知识的运用。现有的统计机器翻译框架下，语言知识主要体现在语言模型和翻译模型中[1]。对于语言模型而言，现有的机器翻译系统所使用的语言模型还是n-gram模型为主，完全没有利用任何语言知识。也有一些工作[10]使用了目标语言依存语言模型，不过这种语言模型理论上还比较粗糙，而且也只是作为n-gram模型的补充，效果也比较有限。从翻译模型来看，基于词的翻译模型[2]和基于短语的翻译模型[7]所使用的语言知识都还只局限于词汇层面（如英语的词例化知识和汉语的词语切分知识），基于句法的翻译模型引入了句法知识[5,8,15-17]，如句法成分的结合关系（句法成分树）或依存关系（依存树）。但从理论上说，机器翻译所需要的语言知识应该包括：

（1）语音层面：韵律知识

（2）词语层面：屈折变化（屈折语或粘着语）、词语切分（汉语、日语、泰语等）、构词模式（如汉语重叠、离合、缩合等）

（3）句法层面：句法结构、句法约束、位移

（4）语义层面：知识本体、语义偏向、语义角色

（5）篇章层面：指代消解、话题结构

（6）语用层面：情境语义、情感

可以看到，目前的统计机器翻译系统最多还只利用了词语层面屈折变化和词语切分知识以及句法层面的句法结构知识，对于其他的语言知识几乎都还没有涉及。

从统计翻译模型发展的趋势看，目前基于句法的翻译模型已经比较成熟，下一步研究的重点很可能是基于语义的翻译模型。

其实，把上述语言知识应用到机器翻译中去，实现起来并不复杂，在基于规则的机器翻译方法中，上述很多语言知识都已经得到应用。但是在统计机器翻译框架下，简单地把上述语言知识应用到机器翻译中来，并不能被得到认可，关键是要在大规模的数据集上使得机器翻译性能得到显著的提高，而不是解决个别例句的翻译问题。这方面现在还缺乏有说服力的工作。这方面的主要困难在于：

（1）评价问题：某些语言现象在实际语言中出现并不普遍，即使花大力气解决了，机器翻译结果的总体BLEU值并不能显著提高，以至于研究人员对解决这类问题失去兴趣，这跟现有机器翻译的评价方法有关；

（2）语言资源缺乏：应用更深层次的语言知识，需要有标注这些知识的语料库，而这一类资源标注往往需要较大的工作量；

（3）模型与算法支持不够：把这么多种类型的语言知识引入机器翻译模型，必定会导致机器翻译模型的复杂度大大增加，而现有统计模型的表示形式和机器学习算法还比较简单，还不足以处理如此复杂的语言知识。

2.2 复杂形态语言翻译建模

复杂形态语言包括粘着语（如芬兰语、日语、韩语、蒙古语、维吾尔语）和部分形态变化比较复杂的屈折语（如德语、法语、阿拉伯语、俄语）。目前机器翻译研究界关注最多的汉语和英语都属于形态变化比较简单的语言。汉语基本上没有形态变化，英语形态最丰富的动词也只有四种变化形式。现在的机器翻译系统对于简单形态语言研究较多，对于复杂形态语言研究较少。复杂形态语言给机器翻译带来的问题主要有：

（1）在现有的统计机器翻译方法中，都不对动词的变化形式进行还原（stemming），而是把每种变化形式作为单独的词语来处理。但对于形态复杂的语言来说，这样做就不合适了。形态复杂语言动词的变化形式从几十种到上千种之多，如此众多的变化形式，可能大部分形式都是没有在语料库中出现过的，因此会导致数据稀疏问题极为严重。因素化模型（factored model）[18]为这类复杂形态语言的机器翻译提供了一种可行的办法，但因素化模型本身过于复杂，实现起来效率低，性能提高也很有限，因此并不是理想的解决办法。

（2）在复杂形态语言中，句子的时态、体态、模态、人称、单复数、阴性和阳性、否定、疑问等句法特性都通过动词的形态来表达，而在形态简单的语言中，这些语法特性大部分都通过特定的词语来表达，这就导致这两类语言的句法同构性非常差，而现有的机器翻译模型对于这种结构差异较大的语言之间的翻译效果都不理想。

（3）复杂形态语言的生成也是一个比较难处理的问题。如果要将形态简单的语言翻译成形态复杂的语言，就需要为形态复杂的语言词语加上各种形态变

化，而这些形态变化在源语言中并没有直接的对应物，如何生成这一类语言的词语形态，目前还没有好的做法。

复杂形态语言在自然语言中占有不小的比例，其带来的问题是普遍性的，不仅仅具有实践意义，更具有非常大的理论意义。要能够很好地实现复杂形态语言和简单形态语言之间的机器翻译，必须在语言的更深层次上实现两种语言的映射，这实际上是触及了机器翻译最本质的问题。因此复杂形态语言机器翻译的研究意义是极为巨大的。

2.3　资源缺乏语言机器翻译

现有最成熟的统计机器翻译方法都是建立在大规模双语语料库的基础上。如果没有大规模双语语料库，现有的统计机器翻译方法的优势，如开发成本低、周期短等等，就都不存在了，因为大规模双语语料库，如果要通过人工翻译得到，其成本也是极其高昂的。我们可以设想一下，在完全没有平行语料库的情况下，如何才能尽可能低成本、快速构造一个高质量的机器翻译系统？是采用基于规则的方法更好，还是采用统计机器翻译方法更好呢？似乎并不能轻易得出结论。我们认为这是非常值得探讨的问题，也许某种二者结合的方法才是最好的方法。

目前解决这个问题的一种方法是采用中枢语言，也就是说，找另外一种语言，跟原来的两种语言都存在大规模平行语料库。但这不是彻底的解决办法，一方面不一定存在这样合适的中枢语言，另一方面，采用中枢语言方法得到的机器翻译系统性能也很难超过采用平行语料库直接训练的系统。

语言资源缺乏的另一种情况是领域资源的缺乏。这种情况下我们需要解决机器翻译模型的领域自适应问题。这方面已经有不少研究工作，但效果还不够理想。

2.4　机器翻译自动评价

机器翻译的自动评价是研究人员的重要工具，对机器翻译的研究具有重要的引导作用。目前机器翻译的自动评价普遍采用以BLEU[9]为代表的基于n-gram字符串匹配的方法，以及一些改进的方法，如METEOR[19]、TER[20]等。这一类自动评价方法在机器翻译总体质量较差的时候起到了非常好的作用，大大推动了机器翻译研究的进展。但随着机器翻译水平逐步提高，这一类方法的局限性也逐步凸显出来，越来越不能满足机器翻译研究的需要。这类方法的本质问题在于只能从翻译结果的局部来进行评价，无法了解整个句子的结构是否合理。这种情况为机器翻译研究的进一步进展带来了很大的困扰，实际上也是最近机器翻译研究很难取得较大突破的主要原因之一。

2.5　互联网大规模数据的应用

互联网上存在的大规模数据为机器翻译提供了海量的资源，但如何才能利用这些数据来改进机器翻译效果并不是一件简单的事情，主要问题有：

（1）平行语料库抽取：机器翻译需要的是句子对齐的平行语料库，而互联网上的平行语料库的存在形式非常复杂，要从各种形式的网页中抽取出可用的平行语料库，并不容易；

（2）语料库的过滤和整理：网络数据正错混杂，内容非常不规范，文体、领域、风格分布都非常多样化，如果不能进行细致的整理和过滤，很可能对机器翻译会带来更大的干扰；

（3）用户创造的数据：由最终用户提交自己创造的双语数据，对机器翻译翻译来说是极其有吸引力的想法，如果能够实现，必将会把机器翻译的应用水平带到一个新的台阶。但关键是如何能够吸引用户提交他们的数据，这方面虽然有一些商业上的尝试[28]，但还没有获得大规模的成功。

3　发展前景与趋势展望

根据前面的分析，我们试图对机器翻译今后的发展前景和趋势提出一些自己的想法，着重说明我们认为比较有前景的一些研究方向。

3.1　基于复杂特征的统计建模和机器学习方法

到目前为止，最复杂的统计翻译模型都是建立在句法树的基础上的，这种树可以是短语结构树，也可以是依存树。相应的语法形式是同步上下文无关语法[5]，其变化形式有反向转录语法（ITG）[21][22]、同步树替换语法[15][23]、同步树粘结语法[24]等。

而在语言学上，目前描述语言知识表达能力最强，也是被普遍接受的知识表示形式是复杂特征集，或者叫做特征结构[25][26]。与句法树相比，二者的区别在于：

（1）复杂特征集可以表示成无环有向图（DAG），图上每个结点可以有多个前趋结点，而不像树一样每个结点只有一个父节点；

（2）复杂特征集所表示成的无环有向图上，边都是有标记的，而句法树上边是不带标记的；

（3）从语言学角度看，复杂特征集表示的每个

语言成分都带多个特征，甚至可以用嵌套的特征来表示，而句法树上每个结点都是单一特征。

目前，复杂特征集已经成为大部分形式化的语法理论所采用的基本的知识表示方法。我们认为，要在机器翻译中引入更深层次的语言知识，必须解决基于复杂特征的统计翻译建模和机器学习方法问题。

3.2 基于语义的翻译模型

目前基于句法的翻译模型研究已经很多，很多学者都希望语义的引入能够进一步提高机器翻译系统的性能，但目前在这方面的工作还很少[11]，已有的工作基本上都是在现有的机器翻译模型上引入词义排歧[3][4]和语义角色标注[15]等语义信息，还不能称之为基于语义的翻译模型。

我们设想的基于语义的翻译模型应该满足以下两个条件：第一，这种模型应该采用独立的语义表示形式，而不是在句法树基础上加上局部的语义信息；第二，这种模型应该能够描述词语之间的语义关系，而不是仅仅进行词义排歧。

3.3 基于结构的语言建模

目前普遍采用的n-gram语言模型不能刻画语言的全局结构，无法保证获得的句子符合句法约束，基于结构的语言模型是未来语言模型发展的必然方向。目前研究者已经开始关注这个方向并开展了一些研究工作[12-14,27]，但总体上还没有被没有取得大的突破。

我们设想的基于结构的语言模型应该满足以下条件：第一，这种模型基于某种结构形式，如短语结构树、依存树，或者任何可以反映句子中长距离约束的语言结构形式；第二，这种模型应该可以利用大规模单语语料库进行训练，而不是只能依赖于少量的句法树库之类的标注语料库进行训练；第三，这种模型应该可以跟基于句法的解码算法有效结合，使用时的时间复杂度应该在可接受范围内。

3.4 混合机器翻译知识的主动学习

对于资源缺乏的语言或者领域来说，构造一个机器翻译系统，采用人工撰写规则库的方法和采用统计学习的方法各有优势，我们认为，有必要结合二者的优势，采用一种混合的机器翻译知识表示方法。同时，应采用主动学习方法，通过人机互助的方式，确定现有的机器翻译系统的知识盲点，有针对性地人工构造规则库和语料库，从而加快学习的进度，以最低成本和最快的速度构造一个特定语言或特定领域的机器翻译系统。

3.5 引入结构的机器翻译自动评价

机器翻译的自动评价必须引入结构信息，否则无法引导机器翻译研究走向深入。我们认为，这种结构应该可以刻划句子中词语的远距离搭配关系，而不是像现在的BLEU这样只能刻划近距离的词语搭配关系而无视句子的总体结构。另外，评价算法本身应该尽可能简单高效，不需要引入太多的语言知识，尽量避免对机器翻译的结果进行句法分析之类的操作，因为机器翻译的结果很可能是完全不符合语法的，对机器翻译的结果进行深层次的语言分析完全没有意义。

3.6 新的机器翻译人机交互模式

机器翻译的用户非常广泛，除了专业的翻译人员，大部分用户使用机器翻译的直接目的并不是为了翻译而翻译，有些用户可能是为了获取外语信息，有些用户是为了做外贸交易，有些是为了跟不同语言的人交友，等等。对于这些用户来说，机器翻译系统应该避免直接以机器翻译软件的形式出现，而应该尽可能以非常自然的形式嵌入到其他的应用软件中，以方便用户的使用。这就需要我们对机器翻译的人机交互模式进行深入的研究。在这方面，苹果公司的iPhone 4S手机中推出的Siri系统为我们提供了很好的启示。

另外，对于专业用户来说，人机交互模式的设计也是极为重要的。目前计算机辅助翻译（computer-aided translation，CAT）技术取得了很大的成功，但在CAT软件中，即使嵌入了机器自动翻译的功能，大部分翻译人员还是不喜欢在机器翻译的结果上进行修改，而是宁愿自己看懂原文重新翻译。这一方面是由于目前机器翻译的结果还远不能令人满意，另外，机器翻译的人机交互模式过于简单也是导致用户不愿意使用机器翻译的重要原因。我们认为在这方面有必要开展深入的研究。理想的机器翻译人机交互方式，应该不仅仅让用户愿意使用机器翻译系统，还应该让系统能够自动收集用户人工编辑或者构造的译文和用户的使用习惯，从而实现机器翻译系统与用户的良性循环，从而大大改进机器翻译的质量和用户体验，让机器翻译系统的应用更上一个台阶。

4 结 语

近二十年来，机器翻译研究和应用都发生了深刻的变化。在研究方面，机器翻译研究范式发生了明显的迁移，研究水平有了大幅度提高，但离全自动高质量的理想还有很大差距。在应用方面，机器翻译应用的广度和深度都达到了前所未有的水平，但离用户的

需求依然还相差很远。本文对机器翻译的发展现状做了深入的分析，提出了目前机器翻译发展面临的五个主要问题，并给出了我们认为比较有发展前景的六个研究方向。

概况地说，我们认为，机器翻译技术的发展将向更深、更大、更好三个方向发展。

更深，指的是机器翻译将采用更深层次的语言知识，自然也需要引入更加复杂的机器学习技术。在这方面，我们认为有必要对机器翻译的研究范式再次进行调整，重新引入语言学的循环。也就是说，机器翻译的改进不能仅仅依靠计算机科学家，也要让语言学家参加进来，对机器翻译的结果进行分析，重新设计和标注训练语料库，从而在机器翻译中引入新的语言学知识。新的研究范式如下图所示：

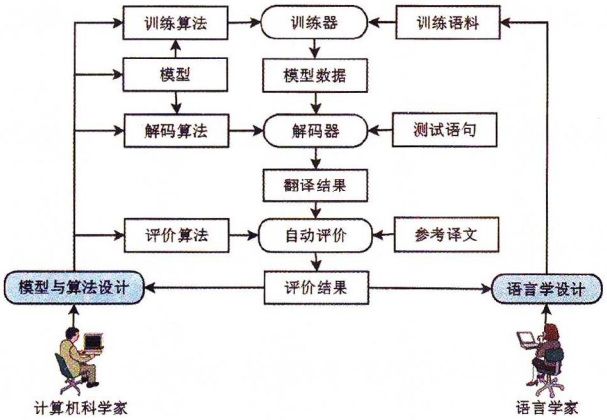

图3　机器翻译研究范式的改进建议

更大，指的是机器翻译将利用更大规模的数据，尤其是互联网数据和用户产生的数据。

更好，指的是机器翻译将越来越贴近用户的需求，将越来越好用。

我们相信，未来机器翻译的技术还将取得更大的突破，机器翻译的应用前景也将更加美好！

致谢：本文的主要思想在“中国中文信息学会成立30周年学术会议”上口头报告过。原报告是本人代表中国中文信息学会机器翻译专委会做的综述报告，报告题目是“机器翻译技术的进展与展望”。原报告在成文过程中，与机器翻译专委会的王海峰、王惠临、宗成庆、赵铁军、史晓东、朱靖波、陈家俊、张民等老师进行了邮件沟通和讨论。原报告把他们列为了共同作者。本人在与他们的交流中获得了许多有益的启示，这些启示有些也体现在本文的一些观点中，在此向以上各位老师一并表示感谢。

参 考 文 献

[1] Brown P F, Cocke J, Pietra S A D, et al. A statistical approach to machine translation [J]. Computational Linguistics, 1990, 16(2): 79-85.

[2] Brown P F, Pietra S A D, Pietra V J D, et al. The mathematics of statistical machine translation: parameter estimation [J]. Computational Linguistics, 1993, 19(2): 263-311.

[3] Carpuat M, Wu D. How phrase sense disambiguation outperforms word sense disambiguation for statistical machine translation [C] //11th Conference on Theoretical and Methodological Issues in Machine Translation (TMI 2007). Skövde(Sweden), 2007: 43-52.

[4] Chan Y S, Ng H T, Chiang D. Word sense disambiguation improves statistical machine translation [C] //Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL 2007). Prague(Czech Republic), 2007: 33-40.

[5] Chiang D. A hierarchical phrase-based model for statistical machine translation [C] //Proceedings of the 43rd Annual Meeting of the ACL. Ann Arbor, 2005: 263-270.

[6] Galley M, Graehl J, Knight K, et al. Scalable inference and training of context-rich syntactic translation models [C] //Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics. Sydney(Australia), 2006: 961-968.

[7] Koehn P, Och F J, Marcu D. Statistical phrase-based translation [C] //Proceedings of the Human Language Technology and North American Association for Computational Linguistics Conference. Edmonton(Canada), 2003: 127-133.

[8] Marcu D, Wang W. Abdessamad echihabi and kevin knight, SPMT: statistical machine translation with syntactified target language phrases [C] //Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing. Sydney(Australia), 2006: 44-52.

[9] Papineni K, Roukos S, Ward T, et al. BLEU: a method for automatic evaluation of machine translation [C] //Proceedings of the 40th Annual Meeting of the ACL. Philadelphia, PA, 2002: 311-318.

[10] Shen L B, Xu J X, Weischedel R. A new string-to-dependency machine translation algorithm with a target dependency language model [C] //Proceedings of ACL-08: HLT. Columbus(USA), 2008: 577-585.

[11] Wu D, Fung P. Can semantic role labeling improve SMT? [C] //Proceedings of the 13th Annual Conference of the EAMT. Barcelona, 2009: 218-225.

[12] Schuler W, AbdelRahman S, Miller T, et al. Broad-coverage incremental parsing using human-like memory constraints [J].

Computational Linguistics, 2010, 36(1).

[13] Schuler W. Parsing with a bounded stack using a model-based right-corner transform [C] //Proceedings of NAACL. Boulder, Colorado, 2009: 344-352.

[14] Schuler W, AbdelRahman S, Miller T, et al. Toward a psycholinguistically-motivated model of language processing [C] //Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008). Manchester, 2008: 785-792 .

[15] Liu Y, Liu Q, Lin S X. Tree-to-string alignment template for statistical machine translation [C] //COLING-ACL 2006. Sydney(Australia), 2006: 17-21.

[16] Mi H T, Huang L, Liu Q. Forest-based translation [C] // Proceedings of ACL-08:HLT. Columbus(USA), 2008: 192-199.

[17] Xie J, Mi H T, Liu Q. A novel dependency-to-string model for statistical machine translation [C] //Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP2011). Edinburgh(UK), 2011: 216-226.

[18] Koehn P, Hoang H. Factored translation models [C] // Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Prague, 2007: 868-876.

[19] Lavie A, Denkowski M J. The meteor metric for automatic evaluation of machine translation [J]. Machine Translation, 2009, 23(2-3): 105-115.

[20] Snover M, Dorr B. A study of translation edit rate with targeted human annotation [C] //Proceedings of the 7th Conference of the Association for Machine Translation in the Americas. Cambridge, 2006: 223-231.

[21] Wu.D K Stochastic inversion transduction grammars and bilingual parsing of parallel corpora [C] //Computational Linguistics. Cambridge: MIT Press, 1997, 23(3).

[22] Xiong D Y, Liu Q, Lin S X. Maximum entropy based phrase reordering model for statistical machine translation [C] // COLING-ACL. Sydney(Australia), 2006:17-21.

[23] Huang L, Knight K, Joshi A. Statistical syntax-directed translation with extended domain of locality [C] //Proceedings of AMTA. 2006.

[24] Liu Y, Liu Q, LüY J. Adjoining tree-to-string translation [C] // Proceedings of ACL/HLT. Portland(USA), 2011: 1278-1287.

[25] Pollard C J, Sag I A. Head-driven phrase structure grammar [M]. Chicago: University of Chicago Press, 1994.

[26] Falk Y. Lexical functional grammar [M]. CSLI Publications, 2001.

[27] Tan M, Zhou W L, Zheng L, et al. A large scale distributed syntactic, semantic and lexical language model for machine translation [C] //Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics. Portland(Oregon), 2011: 201-210.

[28] Duolingo [EB/OL].http://duolingo.com/