北京大学学报(自然科学版)
Acta Scientiarum Naturalium Universitatis Pekinensis
doi: 10.13209/j.0479-8023.2016.013

基于选择偏向性的统计机器翻译模型

唐海庆　熊德意†

苏州大学计算机科学与技术学院，苏州　215006; † 通信作者, E-mail: dyxiong@suda.edu.cn

摘要　针对基于短语的统计机器翻译使用有限的语义知识，导致长距离的动宾短语对翻译质量不高的问题，提出基于动词选择偏向性的翻译模型，引入动词对宾语的语义约束信息，为动词找到合适的宾语翻译。首先使用条件概率方法，训练动词对宾语的选择偏向性，然后将选择偏向性作为一个新特征，集成到基于短语的翻译系统中。在大规模测试数据集上完成汉语到英语的翻译，实验结果表明，基于选择偏向性的翻译模型能够很好地捕获长距离的语义依赖关系，从而提高译文质量。

关键词　语义知识；选择偏向性；语义约束；语义依赖

中图分类号　TP391

A Selectional Preference Based Translation Model for SMT

TANG Haiqing, XIONG Deyi †

School of Computer Science and Technology, Soochow University, Suzhou 215006; † Corresponding author, E-mail: dyxiong@suda.edu.cn

Abstract　The limited semantic knowledge is used in the phrase-based statistical machine translation (SMT) causing the translation quality of long-distance verb and its object is low. A selectional preference based translation model which inducts the semantic constraints that a verb imposes on its object to select the proper argument-head word for the predicate with long distance is proposed by the authors. First train the corpus to obtain the conditional probability based selectional preferences for verb. Then integrate the selectional preferences into a phrase-based translation system and evaluate on a Chinese-to-English translation task with large-scale training data. Experiment results show that the integration of selectional preference into SMT can effectively capture the long-distance semantic dependencies and improve the translation quality.

Key words　semantic knowledge; selectional preferences; semantic constraints; semantic dependencies

统计机器翻译作为自然语言处理研究领域中的一个热点问题，其研究方法从最初的基于单词的翻译，发展到基于短语的，再到基于句法的翻译，机器翻译性能逐渐得到优化。基于短语的翻译以短语(任意一串连续的单词)作为基本的翻译单位，可以很好地解决局部上下文依赖关系，却无法解决长距离依赖关系，从而导致翻译系统存在译文词汇选择不当的问题。图 1 的实例表明，基于短语的翻译系统在翻译源端动词“赢得”的宾语中心词“票”时，选择"NULL"作为其译文，译文单词选择明显错误。在基于短语的统计机器翻译系统中引入语义知识，可以有效地降低译文选词错误率[1]。

选择偏向性是指单词在其使用语境中所具有的语义限制[2]。更具体来讲，动词对宾语的选择偏向性可以体现该动词更偏向于将哪一类词语作为它的宾语。以英文动词“drink”为例，它的宾语更倾向于是可食用的且为液体的一类名词，不可食用的或是固体一类的名词一般无法成为其宾语。利用动词对其宾语的选择偏向性，能够很容易判断句中的动宾短语搭配是否符合语义约束。动词以及动词的参数构成句子的主要框架。对于统计机器翻译而言，译文的可读性取决于句子的主要框架是否翻译正确。

国家自然科学基金青年基金(61403269)、江苏省自然科学基金青年基金(BK20140355)资助
收稿日期: 2015–06–19; 修回日期: 2015–08–24; 网络出版时间: 2015-09-30 09:41:42

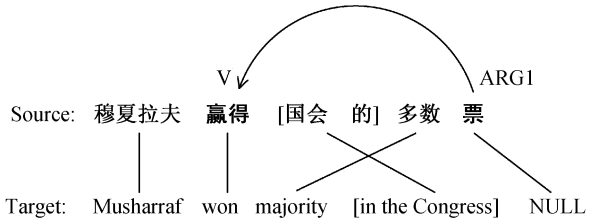

图 1　基于短语翻译系统的一个翻译实例

Fig. 1　An example of translation in the phrase-based SMT

因此在统计机器翻译中使用动词对其宾语的语义约束信息具有理论上的可行性。选择偏向性在自然语言处理领域有着相当广泛的应用，研究人员就如何从训练语料库中自动学习获得选择偏向性已开展一系列研究。

本文提出基于动词选择偏向性的翻译模型，在翻译过程中引入动词对宾语的选择偏向性，帮助翻译系统提高宾语候选词的选择正确率。本文使用基于条件概率的方法，从语料库中自动学习获得动词对宾语的选择偏向性。训练动词对宾语的选择偏向性可以按以下 3 个步骤实现: 1) 在大规模双语对齐语料集上分别对源端和目标端句子进行句法分析和语义角色标注，获得每个句子的 PAS (predicate-argument structure)信息; 2) 分别抽取出目标端动宾短语对和源端动词，以及目标端宾语的短语对; 3) 使用基于条件概率的方法，分别训练得到动词的单语义选择偏向性和跨语义选择偏向性，将动词的选择偏向性作为一个新特征加入到解码过程中。以 BLEU 值[3]作为评测指标，相比于基准系统可以提升 0.52 个点。同时对译文的分析也表明，本文提出的翻译模型能够有效地捕获长距离的语义依赖关系，从而在提高宾语候选词的选择正确率的同时，也能提高动词候选词的选择正确率。

1　相关工作

近年来，关于如何自动学习获取选择偏向性受到研究人员的广泛关注。获取选择偏向性的方法主要分为三类: 简单基于统计的、基于类的和基于主题模型的。基于统计的方法是统计语料库中动词与某一单词同现的频率，即条件概率方法，频率越高代表动词越倾向于将该单词作为其参数(如宾语、主语等)。该方法实现简单，但缺乏泛化能力。Resnik[4]最早开展基于类的方法研究，提出选择性联合的概念来刻画某一特定语义类作为某一动词的参数的语义适合程度，该方法需要一个语义类知识库或者带有明确分类信息的语料库。Clark 等[5]提出的基于类的方法是使用语义层次结构，为单词找到合适的语义类。近来，由于一系列相关的词语可以用一个主题来表示，基于主题模型的研究方法受到研究人员的关注。例如，Ritter 等[6]利用主题模型的方法生成每个单词的主题分布以获取单词的语义类。上述方法都依赖于一个大型的机器可读语料库。

译文单词选择是统计机器翻过程中的一项重要任务。现有的基于短语的统计机器翻译系统大都存在译文选词不当的问题。许多研究人员使用语义知识改善基于短语的统计机器翻译的性能。Carpuat 等[7]利用词义消歧(word sense disambiguation)，使得单词在具体语境中具有唯一的意义来提高译文选词的正确率。Xiao 等[8]使用主题建模方法发现文档的基本主题结构，确定单词的主题分布来计算源端和目标端翻译规则的主题相似度，该方法可以很好地解决文档级约束下的单词选择问题。Xiong 等[9]提出的谓词翻译模型和参数重排序模型，利用源端句子的 PAS 信息，可以显著提高译文的质量。Liu 等[10]提出根据源端和目标端语义角色的相对一致性，将源端的语义角色通过对齐信息映射到目标端，得到目标端的语义角色，选择与目标端语义角色一致的译文作为最终的译文。

本文主要针对动宾关系，研究动词对宾语的选择偏向性，并提出基于动词选择偏向性的翻译模型。选择偏向性已应用到自然语言处理的诸多方面，例如语义消歧[4]、语义角色标注[11]等，但将选择偏向性应用于统计机器翻译的研究十分有限。本文首次使用选择偏向性作为语义知识，辅助翻译系统在翻译长距离的动宾短语对时，为动词的宾语选择合适的译文单词。实验表明，本文的方法能够有效提高系统的性能。

2　选择偏向性学习

动词对宾语的选择偏向性能够很好地体现动词偏向于将哪些语义类作为宾语。使用基于类的方法获得动词的选择偏向性，关键是将语料库中观察到的宾语泛化成一个适当的语义类。但几乎所有基于类的方法都依赖于条件概率计算，所以本文考虑最简单的基于条件概率的方法，从语料库中自动学习动词的选择偏向性。基于条件概率的方法无需对动

词的宾语进行分类，只需要统计语料库中某一动词和某一宾语的同现频率。同现频率越高，表明动词更偏向于选择该单词作为宾语。本文从两方面使用条件概率方法训练动词的选择偏向性：1) 只在目标端计算动词对宾语的选择偏向强度，即获取单语义的选择偏向性；2) 通过计算源端动词对目标端宾语的选择偏向强度，获取跨语义的选择偏向性。

2.1　基于条件概率的单语义选择偏向性

本文首先仅在目标端计算动词对其宾语的选择偏向性强度，即获取单语义的选择偏向性。基于条件概率方法获取动词的选择偏向性的一般定义为：语料库中某一动词 v 在语义关系 r 下，名词 n 作为 v 的参数的可能性体现了 v 对 n 的选择偏向强度，可以用 $\hat{P}(n|v,r)$ 来估计，计算公式如下：

$$\hat{P}(n|v,r)=\frac{f(v,r,n)}{f(v,r)} \quad , \tag{1}$$

其中，$f(v, r)$ 表示语料库中动词 v 出现的次数，$f(v, r, n)$ 表示在语义关系 r 下动词 v 和名词 n 同现的次数。

根据条件概率方法获得动词的选择偏向性的一般定义，如式(2)所示。本文将语义关系 r 具体化为动宾关系，学习目标端动词对宾语的选择偏向性 SP_t：

$$SP_t=\frac{f(v_t,n_t)}{f(v_t)} \quad , \tag{2}$$

其中，在动宾关系下，$f(v_t)$ 代表训练语料库中动词 v_t 出现的次数，$f(v_t, n_t)$ 代表动词 v_t 与宾语 n_t 共同出现的次数。

2.2　基于条件概率的跨语义选择偏向性

跨语义选择偏向性利用跨语言知识，可以体现源端动词对其目标端宾语的选择偏向性。具体来说，在一个双语对齐的语料库上，观察到源端的一组动宾短语 (v_s, n_s)，利用对齐信息得到源端宾语 n_s 的目标端译文 n_t，那么源端动词对目标端宾语基于条件概率的选择偏向强度 $SP_{s\text{-}t}$ 可以通过式(3)计算得到：

$$SP_{s-t}=\frac{f(v_s,n_t)}{f(v_s)} \quad , \tag{3}$$

其中，在动宾关系下，$f(v_s)$ 代表源端动词 v_t 在源端语料库中出现次数，$f(v_s, n_t)$ 代表源端动词 v_s 与目标端宾语 n_t 共同出现的次数。

2.3　抽取动宾关系实例

使用条件概率方法获取动词对宾语的选择偏向性，首先需要从训练语料库中抽取出所有的(动词，宾语)关系实例，这需要借助于句子的 PAS 信息。对于目标端语料集，首先利用自然语言处理工具 SENNA[①]对所有句子进行词性标注、动词语义角色标注和句法分析，以获得每个句子的 PAS 信息。通常一个动词的宾语由多个单词组成，定义一组规则找到宾语中心词，从而完成目标端(动词，宾语)关系实例的抽取。

在抽取源端动词及其目标端宾语短语对时，首先要实现源端(动词，宾语)关系实例的抽取。首先使用 Berkeley Chinese[②]语法分析器对所有源端句子进行句法分析，再使用中文语义角色标注工具[12]为所有的动词标注出与其语义相关的角色。在获得源端每个句子的 PAS 信息后，同样定义另一组规则以找到源端宾语的中心词。在抽取源端(动词，宾语)关系实例的同时，使用对齐信息得到源端宾语中心词对应的译文。由于源端的一个单词可能对应于目标端多个单词，我们取目标端的第一个单词作为源端宾语中心词对应的译文，完成在双语对齐语料库上抽取(源端动词，目标端宾语)关系实例。

3　基于选择偏向性的翻译模型

本节主要给出基于动词选择偏向性翻译模型的定义，以及如何将通过条件概率方法计算得到的动词对宾语的选择偏向强度集成到对数线性模型的解码过程中。

3.1　模型定义

利用动词对其宾语的语义约束信息，我们把动词的选择偏向性作为一个新特征加入到基准系统的解码器中。解码时输入的源端句子经句法分析和动词语义角色标注后，带有 PAS 信息。给定一个翻译区间 (i, j)，得到该区间内存在的所有(动词，宾语)短语对的位置信息。如果使用的是单语义的选择偏向性，先根据对齐信息获得(动词，宾语)短语对的目标端翻译，然后计算目标端动词对宾语的选择偏向强度；如果使用的是跨语义的选择偏向性，则先根据动词的位置信息获得源端动词，根据对齐

① http://ml.nec-labs.com/senna/

② https://code.google.com/p/berkeleyparser/

信息得到源端宾语对应的目标端翻译，然后计算源端动词对目标端宾语的选择偏向强度。定义翻译区间(i, j)的选择偏向性特征值S_{sp}的计算方式如下:

$$S_{sp} = \log \prod_{i=1}^{N} P_i \,, \tag{4}$$

其中，N代表当前翻译区间中(动词，宾语)短语对的个数，P_i代表使用条件概率方法训练得到的单语义选择偏向性或跨语义选择偏向性。

3.2 解码过程

我们使用的基准系统是基于 BTG 的解码器。该解码器用的是 CKY 形式的解码算法，因此任何使用 CKY 形式解码的系统都能根据本节介绍的算法，将选择偏向性集成到系统的解码器中。对于一个带有 PAS 信息的源端句子，找到每个子翻译区间中存在的所有(动词，宾语)短语对，以计算每个翻译区间的选择偏向性特征值。

本文介绍的解码算法借鉴 Xiong 等[9]的解码思想。以图 2 为例说明本文的解码过程: 对于翻译区间为(i, j)的短语 f，如果在翻译短语表中有可用的翻译规则 R_k (k= 1, 2, …, K)，我们定义一个函数 $A(i, j)$来找到该区间内所有(动词，宾语)短语对的位置信息。例如所给例句中，$A(1,12)=\{(2,3),(8,12)\}$；而 $A(5,7)=\{\}$，即该翻译区间不存在(动词，宾语)短语对。那么对于每一条翻译规则得到的译文 e，如果是计算单语义的选择偏向性，我们根据对齐信息得到由函数 $A(i, j)$找到的所有(动词，宾语)短语对的翻译。如果是计算跨语义的选择偏向性，则利用对齐信息得到宾语的译文，$A(i, j)$位置信息取得源端动词，以获得(源端动词，目标端宾语)短语对。定义一个集合 M 来存放这些短语对。该短语区间(i, j)的选择偏向性特征值可用式(4)计算得到。

如果将翻译区间(i, j)拆分为两个子区间(i, k)和$(K+1, j)$分别进行翻译，再以正序或逆序的方式合并这两个子区间的译文，最终得到区间(i, j)的译文，则定义另外一个函数 $T(i, k, j)$，该函数实现将找到的两个相连短语子翻译区间(i, k)和$(k+1, j)$合并为(i, j)时产生的新(动词，宾语)短语对的位置信息，其数学定义可表示为 $T(i,k,j)=A(i,j)-(A(i,k)\cup A(k+1,j))$。如例句中的 $T(1,2,4)=\{(2,3)\}$，而 $T(1,3,5)=\{\}$。由于在动态规划解码过程中，两个翻译子区间的选择偏向性特征值已经计算过，所以在计算翻译区间(i, j)上的选择偏向性特征值时，只需要计算合并过程中新得到的动宾短语对的选择偏向性特征值，以避免重复计算。同样，如果是计算单语义的选择偏向性，则利用每个子区间的翻译短语的对齐信息，得到由函数 $T(i, k, j)$找到的(动词，宾语)短语对的译文；如果是计算跨语义的选择偏向性，则根据 $T(i, k, j)$得到的位置信息找到源端动词，利用翻译短语对齐信息得到宾语的目标端翻译，以获得(源端动词，目标端宾语)短语对。定义集合 N 来存放这些短语对，利用式(4)计算得到这些短语对的选择偏向性特征值，那么翻译区间(i, j)的选择偏向性特征值计算方式为: 两个子区间的选择偏向性特征值之和加上集合 N 上计算得到的选择偏向性特征值。

4 实验与分析

4.1 实验设置

本文使用 Xiong 等[13]等实现的基于短语的统计机器翻译系统，进行汉语到英语方向的翻译任务。实验在 NIST 03 上做最小错误率训练，得到最优线性模型的参数，并采用 NIST 04 和 NIST 05 两个评测语料作为我们的测试集。这两个测试集分别包含 919 句和 1082 句汉语句子，每个句子对应 4 个参考译文。采用大小写不敏感的 BLEU-4 作为系统翻译质量的评价指标。

实验中使用的双语训练语料集由 LDC 语料集的部分子集组成，包括 LDC2004E12, LDC2004T08, LDC2005T10, LDC2003E07, LDC2003E14, LDC2002E18, LDC2005T06 和 LDC2004T07，总计约有 400 万条中英文对齐的句对。首先，我们在汉语到英语方向和英语到汉语方向分别运行 GIZA++[14]工具，然后采用“grow-diag-final”[15]的启

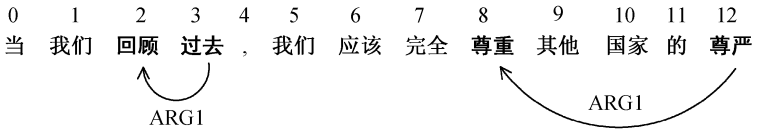

图 2 一个带有 PAS 信息的源端句子

Fig. 2 A sentence with predicate-argument structure

发式方法，获得单词对齐信息。使用 SRILM[16]工具在新华英文语料集 Gigaword 上训练我们的五元语言模型，并使用 KN 方法进行平滑。最小错误率训练过程中各个特征的权重值通过 MERT[17]来调整，取测试集 3 次 BLUE 值的平均值作为最终实验结果。

4.2　结果及分析

本文进行 3 组实验，用以验证基于动词选择偏向性的翻译模型是否可以提高翻译系统的译文选词正确率。表 1 给出在测试集 NIST 04 和 NIST 05 上的实验结果。观察表 1 中实验结果，可以得到以下 3 个结论。

1) 与基准系统相比，加入动词选择偏向性特征后的翻译系统的 BLEU 值在两个测试集上均有提高，系统性能有所改善 。

2)与基准系统相比，跨语义的选择偏向性 BLEU 值平均可以提高 0.37 个点，而单语义的选择偏向性 BLEU 值只能提高 0.28 个点。显然，跨语义的选择偏向性在提高译文选词正确率方面性能优于单语义的选择偏向性。

3) 翻译系统同时利用单语义和跨语义的动词选择偏向性时，系统性能优于单独使用这两种选择偏向性。

基准系统同时使用单语义的选择偏向性和跨语义的选择偏向性时，我们对测试集 NIST05 上得到的译文进行分析，发现动词的选择偏向性特征不仅能帮助翻译系统为源端动词的宾语找到合适的译文，同时在正确翻译动词的宾语前提下，也能提高动词的选词正确率。下面对表 2 和 3 给出的两个翻译例子进行具体分析。

基准系统在翻译表 2 中源端句子的动宾短语(防止，爆发)时，动词翻译正确，但宾语“爆发”的译文选为“NULL”；加入选择偏向性特征后的翻译系统将“爆发”的译文单词“outbreak”正确地选择出来。

在翻译表 3 中源端句子的(建立，系统)动宾短语时，基准系统在正确翻译宾语的前提下，动词并没有翻译正确，“establishment”虽然也有建立的意思，但却是名词性的单词；基准系统使用动词的选择偏向性特征后将“建立”正确地翻译为动词词性单词“build”。

5　总结与展望

本文提出一种基于动词选择偏向性的翻译模型，利用动词对其宾语的选择偏向强度来提高译文的选词正确率。首先，从训练语料库中抽取(动词，宾语)关系实例；然后，利用条件概率方法，获得动词对其宾语的选择偏向性；最后，将选择偏向性作

表 1　基于动词选择偏向性的测试集 BLEU 值

Table 1　The experimental results of verbal selectional preferences based on test sets

实验	开发集	测试集	
	NIST 03	NIST 04	NIST 05
基准系统	34.20	35.52	33.80
+单语义的选择偏向性	34.74	35.72*	34.15*
+跨语义的选择偏向性	34.75	35.76*	34.29*
+单语义的选择偏向性 +跨语义的选择偏向性	34.86	35.86*	34.32*

注: *表示相对基准系统的显著性测试中 $p<0.05$。

表 2　宾语翻译结果对比实例

Table 2　The comparison examples of translation results for object

源端句子	参考译文	基系统译文	+单语义的选择偏向性 +跨语义的选择偏向性
世卫组织主要关切的问题是**防止**诸如霍乱、痢疾等疫情**爆发**	The WHO’s main concern was to **prevent** the **outbreak** of diseases such as cholera and dysentery	The WHO's main concern is to **prevent** such as cholera , dysentery and other epidemic	The WHO’s main concern is to **prevent** such as cholera , dysentery and other epidemic **outbreak**

说明: 加粗单词表示句中的动宾短语。

表 3　动词翻译结果对比实例

Table 3　The comparison examples of translation results for verb

源端句子	参考译文	基系统译文	+单语义的选择偏向性 +跨语义的选择偏向性
来自 国外 的 救援 部队 正 **建立** 净水 供应 **系统**	Overseas rescue troops are **building** clean water supply **systems**	Rescue forces are from abroad with the **establishment** of the water supply **system**	From abroad rescue forces are **build** water supply **systems**

说明: 加粗单词表示句中的动宾短语。

为一个新特征, 集成到一个基于短语的统计机器翻译系统中。本文研究两种选择偏向性: 目标端动词对目标端宾语的单语义选择偏向性和源端动词对目标端宾语的跨语义选择偏向性。在大型语料集上进行汉语到英语的翻译, 实验结果表明: 基于选择偏向性的统计机器翻译模型能够有效地捕获长距离的语义依赖关系, 提高译文质量; 并且, 跨语义的选择偏向性在提高译文选词正确率方面性能优于单语义的选择偏向性。

值得注意的是, 本文自动学习选择偏向性的方法相对简单, 没有对观察到的宾语进行泛化, 得到这些宾语所属的语义类, 从而也无法为训练语料库中未出现过的动宾短语对做出合理预测。未来, 我们将从以下 3 个方面展开深入研究。

1) 本文只对动词与宾语这样一种语义关系进行研究, 未来将对动词与主语的语义关系进行研究, 并探讨动词的哪种语义关系能更有效地提高机器翻译的性能。

2) 本文采用的获取动词对宾语的选择偏向性方法相对简单, 且存在数据稀疏问题, 未来将使用主题模型的方法为动词的参数找到合适的语义类, 使用基于类的方法获得动词的选择偏向性, 以解决数据稀疏问题。

3) 本文在处理源端的一个单词翻译到目标端可能有多个单词的情况时, 采取选择目标端的第一个单词作为源端单词译文的做法。例如源端单词“吃”, 根据对齐信息得到的译文可能是“to eat”, 按照我们的处理方法, “吃”被翻译为“to”, 显然是不正确的, 这就导致翻译系统无法正确获得当前动词对其宾语的选择偏向性。未来, 我们将使用对齐概率信息, 找到源端单词最有可能的目标端译文单词, 更好地解决源端单词对应目标端多个译文单词的情况。

参考文献

[1] 刘群. 机器翻译研究新进展. 当代语言学, 2009, 11(2): 147−158

[2] Brockmann C, Lapata M. Evaluating and combining approaches to selectional preference acquisition // Proceedings of the 10th Conference on European Chapter of the Association for Computational Linguistics — Volume 1. Budapest, 2003: 27−34

[3] Papineni K, Roukos S, Ward T, et al. BLEU: a method for automatic evaluation of machine translation. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia, 2002, 311−318

[4] Resnik P. Selectional preference and sense disambiguation // Proceedings of the ACL SIGLEX Workshop on Tagging Text with Lexical Semantics: Why, What, and How. Madrid, 1997: 52−57

[5] Clark S, Weir D. Class-based probability estimation using a semantic hierarchy. Computational Linguistics, 2002, 28(2):187−206

[6] Ritter A, Etzioni O. A latent dirichlet allocation method for selectional preferences // Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics. Uppsala, 2010: 424−434

[7] Carpuat M, Wu D. Improving statistical machine translation using word sense disambiguation // EMNLP-CoNLL. Prague, 2007: 61−72

[8] Xiao X, Xiong D, Zhang M, et al. A topic similarity model for hierarchical phrase-based translation // Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers — Volume 1. Jeju Island, 2012: 750−758

[9] Xiong D, Zhang M, Li H. Modeling the translation of predicate-argument structure for SMT // Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers — Volume 1.

Jeju Island, 2012: 902−911

[10] Liu D, Gildea D. Semantic role features for machine translation // Proceedings of the 23rd International Conference on Computational Linguistics. Beijing, 2010: 716−724

[11] Zapirain B, Agirre E, Màrquez L, et al. Improving semantic role classification with selectional preferences // Human Language Technologies: the 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. Los Angeles, 2010: 373−376

[12] Li J, Zhou G, Ng H T. Joint syntactic and semantic parsing of Chinese // Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics. Uppsala, 2010: 1108−1117

[13] Xiong D, Liu Q, Lin S. Maximum entropy based phrase reordering model for statistical machine translation // Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics. Sydney, 2006: 521−528

[14] Och F J, Ney H. A systematic comparison of various statistical alignment models. Computational Linguistics, 2003, 29(1): 19−51

[15] Koehn P, Och F J, Marcu D. Statistical phrase-based translation // Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics. Edmonton, 2003: 54−58

[16] Stolcke A. SRILM — an extensible language modeling toolkit // Proceedings of the 7th International Conference on Spoken Language Processing. Denver, 2002: 901−905

[17] Franz J, Och. Minimum error rate training in statistical machine translation // Proceedings of ACL. Sapporo, 2003: 160−167