

2009 中国机器翻译研讨会计算所系统描述

米海涛, 刘洋, 冯洋, 夏天, 肖欣延, 熊皓, 郑达奇, 涂兆鹏, 吕雅娟, 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

E-mail: {htmi, yliu, lvyajuan, liuqun}@ict.ac.cn

摘要: 本文介绍了中国科学院计算技术研究所自然语言处理实验室参加 2009 年中国机器翻译研讨会机器翻译评测的情况。今年, 我们参加了汉英新闻领域单一系统及系统融合、英汉新闻、英汉科技和汉蒙日常用语五项评测任务, 使用了 Silenus、Bruin、Chiero 和 Moses 四个统计机器翻译单系统, 以及基于 TER、IHMM 的词级融合系统与句子级融合系统。在汉英新闻领域单系统评测中, 我们使用了 Chiero 与 Moses 系统; 汉英系统融合使用了两个词级、一个句子级构成的层叠式融合系统; 英汉新闻/科技使用了四个单系统以及句子级融合系统; 汉蒙日常用语评测我们使用了 Chiero 与 Moses 单系统。本文详细介绍了各个系统的模型框架和理论模型, 还对评测相关实验做了说明。

关键词: 机器翻译

The ICT System Description for CWMT 2009

*Haitao Mi, Yang Liu, Yang Feng, Tian Xia, Xinyan Xiao, Hao Xiong, Daqi Zheng,
Zhaopeng Tu, Yajuan Liu and Qun Liu*

Key Laboratory of Intelligent Information Processing

Institute of Computing Technology

Chinese Academy of Sciences

P.O. Box 2704, Beijing, China, 100190

{htmi, yliu, lvyajuan, liuqun}@ict.ac.cn

Abstract: *This paper presents a description for the ICT systems involved in the CWMT 2009 evaluation campaign. We participated in all 5 tracks including Chinese-English News Single System Translation, System Combination, English-Chinese Tech Translation, English-Chinese News and Chinese-Mongolian Dialogue Translation. And four statistical single machine translation systems were used: one linguistically syntax-based MT (Silenus), two formally syntax-based MT (Bruin and Chiero), and one phrase-based MT (Moses). We performed Chiero and Moses in ZH-EN-NEWS-SIGL track, a cascaded combination framework in ZH-EN-NEWS-COMBL track, four single systems and a sentence level combination system in EN-ZH-NEWS and SCIE-TRANS tracks, Chiero and Moses again in ZH-MN-DAIL-TRANS track accordingly. We will describe the framework and model of these systems and report the results on both the development and test sets.*

Key Words: *Machine Translation*

1 引言

2009 年中国机器翻译研讨会 (CWMT2009) 机器翻译评测共含五个项目, 分别为: 汉英新闻领域单一系统机器翻译、汉英新闻领域系统融合、英汉新闻领域机器翻译、英汉科技领

域机器翻译和汉蒙口语用语领域机器翻译。我们计算所参加了所有五个项目。

2 参评系统描述

这次机器翻译评测中我们使用了四个统计机器翻译单一系统：Silenus、Bruin、Chiero 和 Moses。其中，Silenus 是基于语言学句法的翻译模型，从压缩森林学习到的树到串规则，用句法森林来指导翻译；Bruin 是基于形式句法的翻译模型，在 BTG 规则上实现了基于最大熵的重排序；Chiero 是基于形式句法的翻译模型，使用层次短语；Moses 是基于短语的翻译模型，为开源系统。

2.1 Silenus

Silenus[Mi et al., 2008; Mi and Huang, 2008]为基于句法森林的树到串模型，衍生于单树到串系统 Lynx[Liu et. al, 2006, 2007]，最关键的不同之处在于，Silenus 在规则抽取和解码时使用共享压缩森林而不是 1-best 句法分析树。

规则抽取：

在给定上下文无关语法下，一个句子的**句法森林**是其所有可能推导（即，句法分析树）的紧凑表示形式[Billot and Lang, 1989]。压缩森林可以形式化定义为一个四元组 $H_p = \langle V, E, t, R \rangle$ ，其中： V 为有限的节点集合， E 为有限的超边集合， $t \in V$ 是句法分析根节点， R 为权重集合。给定一个句子 $w_{1:l} = w_1 \dots w_l$ ， $v \in V$ 的每个节点形式化表示为 $X_{i,j}$ （表示已经识别的跨度为 $i \sim j$ 的非终结符）。每个超边 $e \in E$ 为一个三元组 $e = \langle T(e), h(e), f(e) \rangle$ ，其中 $h(e) \in V$ 为超边的头节点， $T(e) \in V$ 为超边的尾节点向量， $f(e)$ 为 $R^{T(e)/h(e)}$ 到 R 的权重函数。图 1 给出了一个中文森林和英文串的例子。

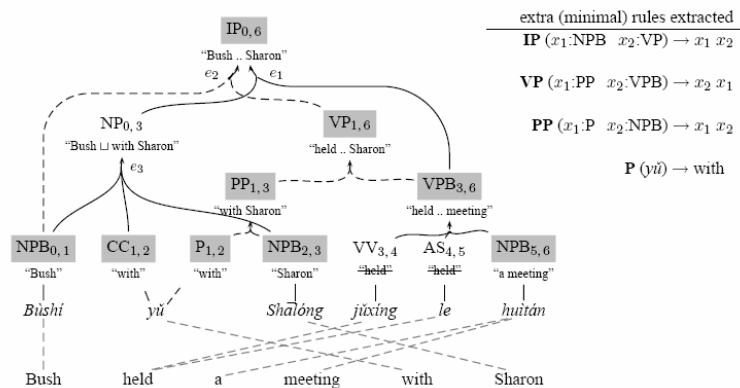


图 1 源语言端压缩森林和目标语言端串

为了在三元组（源语言句法森林、目标语言串和词语对齐）中抽取翻译规则，我们扩展了 GHKM[Galley et. al, 2004]抽取算法。首先抽取最小规则，进一步获得组合规则。在最小规则抽取中，与 GHKM 相似，我们也分为两步：边缘节点集合计算（何处切）和划分（如何切）。其中边缘节点集合计算可以直接使用 GHKM 算法，图 1 中的边缘节点使用灰色表示；而划分步骤变得复杂，因为对于每个节点，我们将面临多个超边的选择问题。

为从压缩森林中抽取树到串规则，我们开发了一种宽度优先的抽取算法。基本思想是：对于每个边缘节点 v ，我们首先创建一个 *open* 队列用于存放以 v 为根的句法片段（由超边与节点构成）。然后不断扩展 *open* 中的不完整片段（叶子节点含有非边缘节点），当得到一个完整的片段（所有叶子节点都是边缘节点）时，则根据目标端的语序创建一条翻译规则。图 1 列示了抽取出来的最小规则，这些规则在基于 1-best 树串对齐中是抽取不到的。

在基于树的规则抽取中，对于每个句对，抽取的每条规则的数量为 1。然而压缩森林是

大量树的紧凑表示，因此应对从非 1-best 句法分析树中抽取的规则进行惩罚。对于一条规则 r ，我们使用相应树片段 $t=lhs(r)$ 的后验概率来进行惩罚。后验概率可以由以下三个部分的乘积得到：树片段根节点的外向概率、叶子节点的内向概率和片段中超边的概率。

$$\alpha\beta(t) = \alpha(\text{root}(t)) \times \prod_{e \in t} P(e) \times \prod_{v \in \text{leaves}(t)} \beta(v) \quad (1)$$

其中， $\alpha(\cdot)$ 和 $\beta(\cdot)$ 分别为节点的外向和内向概率， $\text{root}(\cdot)$ 为树片段的根节点， $\text{leaves}(\cdot)$ 为树片段的所有叶子节点。

这样，规则 r 的相对频次定义为

$$c(r) = \frac{\alpha\beta(\text{lhs}(r))}{\alpha\beta(t)} \quad (2)$$

其中 t 表示森林的根节点。

在给定规则频次之后我们就可以计算相应的词汇化与非词汇化翻译概率。

解码：

在给定句法分析森林 H_b 下，Slienus 首先使用匹配转换算法将句法分析森林转换成翻译森林 H_t ，然后利用树替换语法将源语言的翻译森林转换为目标语言串搜索最优推导（翻译规则序列）。

给定一个句法森林 H_b 和一个翻译规则集合，转换算法最基本的想法是自顶向下遍历整个句法森林的每一个节点 v ，然后对每一条规则尝试当前以 v 为根节点的子森林，如果匹配成功则生成相应的翻译超边。相应的翻译森林如图 2 所示。

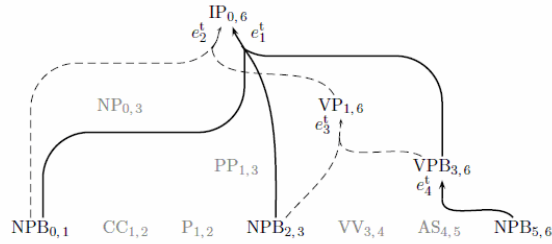


图 2 相应的翻译森林

解码要在给定的源语言翻译森林 H_t 的所有可能推导 D 中搜索最好的一个推导 d^* 。

$$d^* = \arg \max_{F, d \in D} \lambda_0 \log P(d | H_t) + \lambda_1 \log P_{lm}(\tau(d)) + \lambda_2 |d| + \lambda_3 |\tau(d)|$$

其中，第一个特征为翻译特征，第二个为语言模型，第三个是使用规则数量，第四个为翻译词数。翻译特征 $P(d|H_t)$ 可以进一步分解成该推导 d 相对应的树 T 概率与每个规则的连乘：

$$P(d | H_t) = \arg \max \{ T(d)^{\lambda_4} \cdot \prod_{r \in d} P(r) \} \quad (3)$$

$P(r)$ 又可以进一步分解成五个概率的乘积：

$$P(r) = P(r | \text{lhs}(r))^{\lambda_4} \cdot P(r | \text{rhs}(r))^{\lambda_5} \cdot P(r | \text{root}(r))^{\lambda_6} \cdot P_{lex}(r | \text{lhs}(r))^{\lambda_7} \cdot P_{lex}(r | \text{rhs}(r))^{\lambda_8} \quad (4)$$

其中， $P(r|\text{lhs}(r)), P(r|\text{rhs}(r))$ 为基于频次的翻译概率； $P_{lex}(r|\text{lhs}(r)), P_{lex}(r|\text{rhs}(r))$ 为词汇化翻译概率，由 IBM 模型 1 中的词翻译概率得到； $P(r|\text{root}(r))$ 为根节点选择概率。

在给定翻译森林与语言模型的前提下，解码主要完成两个任务：1-best 翻译与 k-best 翻译列表，后者用于最小错误率训练。

1-best 翻译搜索主要应用了 cube pruning 技术[Huang 2007b; Chiang 2007]。具体算法为：后续遍历翻译森林中的节点 v ，先使用翻译超边集合 $BS(v)$ 初始化当前的 cube，然后使用最优先扩（best-first expansion）的策略近似搜索 k-best 最佳翻译假设，直到生成 k-best 个+LM 的翻译假设。遍历到根节点则生成了 1-best 全句翻译结果。

2.2 Bruin

Bruin 是基于最大熵括号转录语法（MEBTG）的统计机器翻译模型，属于形式化基于句法的翻译模型。详细介绍请参见[Xiong et.al 2006]。

MEBTG 的最基本思想是将 BTG[Wu, 1997]中选择保序（straight）或者逆序（inverted）问题看成一个用最大熵分类问题。

在 BTG 语法下，翻译过程实际上更像是一个单语分析过程，该过程使用以下 3 条规则：

$$A \xrightarrow{[] } (A^1, A^2) \quad (5)$$

$$A \xrightarrow{\langle \rangle } (A^1, A^2) \quad (6)$$

$$A \longrightarrow (x, y) \quad (7)$$

解码时，源语言句子首先被分割成一个短语序列，然后采用规则(7)将源语言短语 x 翻译为目标语言短语 y 并行成一个块 A 。所有短语翻译完成之后，开始使用保序规则(5)和逆序规则(6)将两个相邻的块合并成一个更大的块。两条规则的不同之处在于，保序规则合并时要求两个块的顺序在源语言和目标语言中保持一致，逆序规则则要求两个块在源语言和目标语言中正好相反。这两条规则统称为合并规则，解码器不断使用这两条规则，直到源语言句子全部被覆盖。

在采用不同的特征函数时，这三个规则的得分可以写成如下形式：

对于两个合并规则：

$$P^r(r) = \Omega^{\lambda_\Omega} \Delta_{LM}^{\lambda_{LM}}(A^1, A^2) \quad (8)$$

其中 Ω 为块 A^1 和 A^2 的重排序分数， λ_Ω 是其相应权重。 $\Delta_{LM(A^1, A^2)}$ 为两个块合并后的 N-gram 语言模型分数的增量， λ_{LM} 为其相应权重。对合并规则中的重排序模型进行详细分解：

$$\Omega = f(o, A^1, A^2), o \in \{straight, inverted\} \quad (11)$$

其中， o 为 AP^r 和 AP^p 的顺序，即保序（straight）或者逆序（inverted）。在此框架下，将最大熵模型用于重排序。

对于词汇规则：直接由短语翻译概率，语言模型的特征分数得到。

训练一个最大熵重排序模型，需经过 3 个步骤：1) 重排序实例的抽取； 2) 重排序特征的提取； 3) 最大熵分类器的构建。

解码器建立在 CYK 算法基础上，为了避免搜索所有的推导，我们采用了 beam search 策略，在搜索过程中删除一些不好的推导，以减少搜索空间。

2.3 Moses

Moses [Koehn et al., 2007] 是由英国爱丁堡大学、德国亚琛工业大学等 8 家单位联合开发的一个基于短语的统计机器翻译系统。

它的主要特点包括：

- (1) 基于 beam search 的解码；
- (2) 基于短语的翻译模型；
- (3) 基于要素 (factor)。

此外，Moses 还支持词汇化重排序模型。Moses 支持的功能众多，参赛系统主要使用如下配置：

- (1) 短语长度限制为 7。
- (2) 使用基于 msd-bidirectional-fe 词汇化重排序模型。此模型包含三种重排序的情况，Mono/Swap/Discontinuous；排序考虑与前后 (bidirection) 短语的关系；排序依赖于源端 (f) 及目标端 (e)。

2.4 Chiero

Chiero 是基于层次短语的模型。层次短语模型可以认为是基于短语的模型的扩展，其可以抽取源句子中非连续的部分，并将其翻译为目标句子的非连续部分。该模型可以形式化为同步上下文无关文法 (SCFG)，其规则形式为：

$$X \rightarrow \langle \gamma, \alpha, \sim \rangle$$

其中， X 为非终结符， γ 和 α 为终结符和非终结符组成的串， \sim 为 γ 和 α 中非终结符的一一对应关系。

Chiero 如实地模仿了 David Chiang [Chiang, 2007] 的工作，唯一不同在于 cube pruning 算法的终止条件。在 David Chiang 的实现版本 Hiero 中，cube pruning 算法的退出条件是下一个候选项的分数超出 beam ϵ 。试验中，我们发现这一条件下仍然会得到大量的候选项。为了解决这一问题，我们进一步对从堆中取出的候选项的数量进行了限制。

2.4 系统融合

2.4.1 句级系统融合

句级融合的过程类似于 Macherey 和 Och [Macherey and Och, 2007] 的工作，使用 global linear 模型对合并后的 n-best 结果进行重排序：

$$\hat{y} = \arg \max_{y \in GEN(x)} f(x, y) \cdot W$$

其中， x 为源语言句子， y 为翻译结果， $f(x, y)$ 为特征向量， W 为权重向量， $GEN(x)$ 为可能的候选译文结果的集合。

所采用的特征类型包括：(1) 与其他候选译文的相对 BLEU 值；(2) 语言模型分数；(3) 译文结果的长度。各个特征的权重经由最小错误率训练得到。

2.4.2 基于 TER 的词级系统融合

本系统主要模仿了 [Rosti, Matsoukas et. al, 2007] 的工作，但在实现细节上略有不同。任意选取一个系统，从不重复的 N-best (具体取 10) 中选取第一个翻译作为混淆网络的骨架，然后将所有系统输出的 N-best 均与该骨架对齐，构造基于当前系统的混淆网络。此过程依次作用于每个系统。原文的方法为了在 N-best 中选择一个合适的骨架，需要多次计算所有翻译与候选骨架之间的 TER 对齐，这在实际使用中比较消耗时间，而本文直接选择第一个翻译结果，实际使用中速度较快，可忽略损失。

我们采用线性模型，使用的特征包括

- (1) 构成当前候选翻译的词语的置信度分数之和，具体请参考 [Rosti et. al, 2007]。

- (2) 语言模型的权重。
- (3) 空词数惩罚。
- (4) 非空词个数惩罚。

其中参数调整的过程，本文方法采用 MERT 算法，目标函数是 BLEU-sbp，原文采用鲍威尔算法。采用 MERT 算法的原因也是基于速度的考虑，同时将四个系统权重设置为 1，为了在实际中获得适中且稳定的提高。而采用 BLEU-sbp 最大化作为训练目标的原因是为了消除开发集中句子的长度不一而对参数的训练的影响，请参考[Chiang DeNeefe et. al, 2008]，我们实验证明采用 BLEU-sbp 比采用 BLEU 效果更好。

2.4.3 基于 IHMM 的词级系统融合

词级融合主要模仿了 Rosti 和 Matsoukas 等人[Rosti et. al, 2007]的工作。

首先会针对每个系统构建一个混淆网络，所有单系统的混淆网络构成一个多混淆网络，再采用一些特定的特征搜索多混淆网络，分值最高的路径对应的译文便作为基于 IHMM 的系统输出。

对于某个单系统对应的单混淆网络，构建过程如下：

- (1) 用 Minimum Bayes Risk 从该单系统的 nbest 中选择一个代价最小的 hypothesis 作为 skeleton，选取方法如下：

$$E_s = \arg \min_{E \in E_i} \sum_{j=1}^{N_s} \text{TER}(E_j, E_i)$$

- (2) 用基于 Indirect-HMM 的对齐方法将所有单系统的 nbest 与 skeleton 进行对齐。
- (3) 以 skeleton 的单词顺序为基准，将所有的 hypothesis 根据与 skeleton 的对齐进行 normalize，如果对齐到空，则要在适当的位置添加空词。
- (4) 逐句的将 normalize 后的句对加入好混淆网络中。
- (5) 分别添加一个头结点和尾结点，将所有单混淆网络合成一个多混淆网络。

搜索多混淆网络时，采用如下特征计算分数：每个词的概率，语言模型，非空词个数，非空词个数。采用 Powell 算法进行调参。

2.4.4 层叠式系统融合

将所有的单系统、基于 TER 的系统和基于 IHMM 的系统，作为句级融合输入系统，将句级融合的结果作为最终的输出。

3 实验结果

3.1 数据使用

此次评测我们参加了所有项目的评测，汉英新闻、英汉新闻、英汉科技翻译使用了主办方提供的新闻和科技领域公共训练语料，共 378 万双语句对，其中中科院计算所和自动化所中英句子级对齐双语语料库 252,327 句对，北京大学汉英/汉日双语语料库 200,082 句对，厦门大学英汉电影字幕平行语料库 176,148 句对，哈工大信息检索组英汉句子级对齐语料库 100,000 句对，哈工大机器翻译组英汉句子级对齐语料库 52,227 句对，点通汉英平行语料库 1,000,004 句对，计算所汉英平行语料库 1,076,313 句对，863 评测数据约 6,000 句对，中信所英汉科技文献句子级对齐语料库 600,000 句对以及万方汉英中文科技期刊论文摘要语料库 32,0985 句对；汉蒙日常用语翻译使用了主办方提供的 6.8 万双语句对；汉英新闻系统融合使用了主办方提供的汉英新闻单系统参评系统提交的 n-best 翻译结果。

语言模型使用了主办方提供的路透社的 RCV1 语料和搜狗全网新闻语料库 SogouCA 以及 6.2 万句蒙语单语语料。

汉英新闻、英汉新闻和英汉科技翻译的开发集为 CWMT2008 的测试集；汉蒙日常用语翻

译为主办方提供的 400 句对话料；汉英新闻系统融合使用了主办方提供的 SSMT2007 各单位汉英测试集上的 n-best 翻译结果作为开发集。

3.2 模块说明

- 汉英翻译前处理
使用 ICTCLAS 进行分词。
- 汉英翻译后处理
对翻译结果做句首字母大写，去掉未登录词汇及对特殊标点处理等。
- 英汉翻译前处理
Tokenization, 首字母进行 truecase。
- 其他技术
GIZA++、SRILM、MERT 等

3.3 特殊说明

3.3.1 Silenus

Silenus 参加了英汉方向的评测任务，为了获得共享压缩森林，我们使用英文句法分析器[Charniak et al., 2005]产生整个句法森林，然后使用内向外向算法进行减枝。

3.3.2 系统融合

在英汉新闻和科技领域翻译中，参评的系统融合采用了句级融合技术，使用了 Silenus, Bruin, Chiero 和 Moses 的 n-best 翻译结果。

在汉英新闻领域系统融合项目中使用了层叠式系统融合技术，主要使用了两个词级系统融合与一个句子级系统融合系统。我们选择的单系统结果为 Sys1-5，在开发集上最好的结果为 Sys-2，BLEU 值为 0.2780。

3.4 在开发集上的结果

训练过程中，汉英新闻单系统、英汉新闻和科技领域翻译我们将 CWMT2008 语料作为开发集；汉蒙日常用语翻译使用了 400 句对主办方提供的语料作为开发集；汉英新闻系统融合使用了主办方提供的 SSMT2007 各单位提交的 n-best 作为开发集，各系统经过训练后在其上进行测试，测试结果见表 1。

表 1 各系统在开发集上的结果

测试项目	参评系统	BLEU4
汉英新闻单系统	Chiero	0.2529
	Moses	0.2777
英汉新闻	句级	0.4419
	Bruin	0.4070
	Chiero	0.4231
	Moses	0.4252
	Silenus	0.4209
英汉科技	句级	0.5298
	Bruin	0.4972
	Chiero	0.5074
	Moses	0.5094
	Silenus	0.5143
汉蒙日常用语	Moses	0.2540

	Chiero	0.2482
汉英新闻系统融合	Sys1	0.2777
	Sys2	0.2780
	Sys3	0.2771
	Sys4	0.2746
	Sys5	0.2723
	TER	0.2887
	IHMM	0.2979
	层叠式系统融合	0.3043

3.5 实验结果

表 2 给出了我们参评的 6 个系统的 BLEU 值，结果为 BLEU4-SBP。

表 2 各系统在 CWMT2009 测试集上的结果

测试项目	参评系统	BLEU4-SBP
汉英新闻 单系统	Chiero(primary*)	0.2116
	Moses	0.2065
英汉新闻	句级融合(primary*)	0.3563
	Chiero	0.3422
	Moses	0.3382
	Silenus	0.3404
英汉科技	句级融合(primary*)	0.4905
	Chiero	0.4656
	Moses	0.4801
	Silenus	0.4838
汉蒙日常用语	Moses	0.1988
	Chiero(primary*)	0.2157
汉英新闻系统融合	层叠式系统融合(primary*)	0.2407

4. 总结

通过评测结果我们可以发现，系统融合技术对于改善翻译结果有较大的帮助。此外，由于使用共享压缩森林而不是 1-best 句法分析树，Silenus 可以很大程度上减少不良的句法分析结果带来影响，从而得到较好的翻译结果。

参考文献

- S. Billot and B. Lang. 1989. "The structure of shared forests in ambiguous parsing," in *Proceedings of ACL 1989*, pages 143–151.
- E. Charniak and M. Johnson. 2005. "Coarse-to-fine n-best parsing and maxent discriminative reranking," in *Proceedings of ACL 2005*, Ann Arbor, Michigan, pages 173–180.
- D. Chiang. 2007. "Hierarchical phrase-based translation," *Computational Linguistics*, vol. 33, no. 2, pages 201–228.
- M. Galley, M. Hopkins, K. Knight, and D. Marcu. 2004. "What's in a translation rule?" in *Proceedings of HLT/NAACL 2004*, Boston, Massachusetts, USA, pages 273–280.
- Z. He, Q. Liu, and S. Lin. 2008. "Partial matching strategy for phrase-based statistical machine translation," in *Proceedings of ACL/HLT 2008 (Short Paper)*, Columbus, Ohio, pages 161–164.
- L. Huang. 2008. "Forest reranking: Discriminative parsing with non-local features," in *Proceedings of ACL 2008*, Columbus, Ohio, pages 586–594.
- L. Huang and D. Chiang. 2005. "Better k -best parsing," in *Proceedings of IWPT 2005*, Vancouver, Canada, pages 53–64.
- L. Huang and D. Chiang. 2007. "Forest rescoring: Faster decoding with integrated language models". In *Proceedings of ACL*, pages 144–151, Prague, Czech Republic, June.
- P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, C. Dyer, O. Bojar, A. Constantin, and E. Herbst. 2007. Moses: Open source toolkit for statistical machine translation. In *Proc. of ACL 2007 (demonstration session)*.
- Y. Liu, Q. Liu, and S. Lin. 2006. "Tree-to-string alignment template for statistical machine translation," in *Proceedings of COLING/ACL 2006*, Sydney, Australia, pages 609–616.
- Y. Liu, Y. Huang, Q. Liu, and S. Lin. 2007. "Forest-to-string statistical rules," in *Proceedings of ACL 2007*, Prague, Czech Republic, pages 704–711.
- H. Mi, L. Huang, and Q. Liu. 2008. "Forest-based translation," in *Proceedings of ACL/HLT 2008*, Columbus, Ohio, USA, pages 192–199.
- W. Macherey and F. J. Och. 2007. "An empirical study on computing consensus translations from multiple machine translation systems," in *proceedings of EMNLP-CoNLL2007*, Prague, pages 986–995.
- F. J. Och. 2003. "Minimum error rate training in statistical machine translation," in *Proceedings of ACL 2003*, pages 160–167.
- F. J. Och and H. Ney. 2002. "Discriminative training and maximum entropy models for statistical machine translation," in *Proceedings of ACL 2002*, pages 295–302.
- A. Rosti, S. Matsoukas, and R. Schwartz. 2007. "Improved Word-Level System Combination for Machine Translation", in *Proceedings of ACL 2007*, pages 312–319.
- D. Wu. 1997. "stochastic inversion transduction grammars and bilingual parsing of parallel corpora," *Computational Linguistics*, vol. 23, pages 377–404, 1997.
- D. Xiong, Q. Liu, and S. Lin. 2006. "Maximum entropy based phrase reordering model for statistical machine translation," in *Proceedings of COLING/ACL 2006*, Sydney, Australia, pages 521–528.
- D. Xiong, S. Li, Q. Liu, and S. Lin. 2005. "Parsing the penn chinese treebank with semantic knowledge," in *Proceedings of IJCNLP 2005*, pages 70–81.