

第五届全国机器翻译研讨会（CWMT2009）评测报告

（公开版 2009-9-18）

CWMT2009 机器翻译评测组

中国科学院计算技术研究所
中国科学院智能信息处理重点实验室

1 概述

为了全面了解国内外机器翻译技术的现状，促进机器翻译技术的研究，按照惯例，第五届全国机器翻译研讨会（CWMT2009）继续组织了统一的机器翻译评测，以推进参评单位的实质性交流和机器翻译技术的发展。

本次评测由中科院计算所主持，评测设置的项目如下：

评测项目代号	评测项目名称	语种	领域	技术
ZH-EN-NEWS-SINGL	汉英新闻领域单一系统	汉→英	新闻领域	单一系统
ZH-EN-NEWS-COMBI	汉英新闻领域系统融合	汉→英	新闻领域	系统融合
EN-ZH-NEWS-TRANS	英汉新闻领域机器翻译	英→汉	新闻领域	机器翻译
EN-ZH-SCIE-TRANS	英汉科技领域机器翻译	英→汉	科技领域	机器翻译
ZH-MN-DAIL-TRANS	汉蒙日常用语机器翻译	汉→蒙	日常用语	机器翻译

本文给出了此次评测的组织、准备过程及结果。为国内外研究单位在机器翻译方面的进一步研究提供了数据。

本报告以公开版本（public version）和内部版本（internal version）两种形式发布。内部版本仅发送给所有的参评单位，而公开版本将会公开发表。在公开版本中，将会列出所有参评单位名称，但在评测结果中，不会给出对应的单位名称，而是代之以单位的匿名代号。

本报告内容仅供研究使用，可以在研究论文中引用，但不可用于任何出于商业目的的宣传活动。在研究论文中引用时，如果没有得到其他单位的许可，不得公开其他单位的评测结果。

2 参评单位和语种

本次评测共有 18 个单位报名参加，其中国内单位 14 家，国外单位 4 家。其中 17 个单位在不同的项目和语言方向共提交了 102 个系统的翻译结果。

表 2.1 给出了本次评测的参评单位和系统的数量。

表 2.1 参评单位和系统数量表

	参评单位/主系统	参评系统
汉英新闻领域单一系统	15	32
汉英新闻领域系统融合	7	14

英汉新闻领域机器翻译	9	21
英汉科技领域机器翻译	10	24
汉蒙日常用语机器翻译	7	11
合计	48	102

表 2.2 给出了参评单位及参加项目情况。其中√表示报名并提交结果，○表示报名但没提交结果。

表 2.2 参评单位及参加项目情况表

报名单位	汉英新闻领域单一系统	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	汉蒙日常用语机器翻译
富士通研究开发中心有限公司	√		√	√	√
北京赛迪翻译技术有限公司	√		√	√	
南京大学	√				
中国科学院软件研究所	√	√			
Department of Human Language Technology Institute for Infocomm Research	√				
中国科学院自动化研究所	√	√	√	√	
东北大学自然语言处理实验室	√				
中科院合肥智能机械研究所	√	√	√	√	√
微软亚洲研究院	√	√			
内蒙古大学蒙古学学院&内蒙古师范大学计算机学院					√
厦门大学人工智能研究所	√	√	√	√	√
SYSTRAN SOFTWARE, INC	√		√	√	
中国科学院计算技术研究所	√	√	√	√	√
北京航空航天大学计算机学院智能信息处理研究所	√			√	
哈尔滨工业大学计算机学院语言技术研究中心	√	√	√	√	
内蒙古大学计算机学院					√
HKUST	√		√	√	√
中科院声学所	○	○	○	○	

3 评测过程

3.1 测试数据准备

本次机器翻译评测语料包含 3 个语言方向（汉英、英汉和汉蒙）、3 个领域（新闻、科技和日常用语）。

根据国外相关评测及具体分析，我们制订了相应的语料规模。在评测中输入输出文件仍采用 UTF-8 作为标准编码方式，与往年不同的是，本次评测的文件格式均采用严格的 XML 格式。

(1) 训练语料和开发语料

语料情况请见表 3.1

表 3.1 CWMT2009 评测训练和开发语料情况

评测项目	训练语料		开发语料
	双语语料	单语语料	
汉英新闻 单一系统 英汉新闻 英汉科技	380 万句对	路透社的 RCV1 语料和搜狗全网新闻语料库 SogouCA	CWMT2008 的相应评测语料(分别为 1,006、1,000、1,008 句对)
以上双语语料由点通公司、中信所、计算所、万方数据公司、北京大学、自动化所、厦门大学、哈工大等提供，单语语料由路透社、搜狗实验室提供。			
汉蒙日常 用语翻译	6.8 万句对	6.2 万句 (1 百万词)	400 句对
以上双语和单语语料均由内蒙古大学提供。			
汉英新闻 系统融合			12 个提供了 n-best 翻译结果的 CWMT2009 汉英新闻单一系统测试的参评系统在 SSMT2007 汉英测试集上的翻译结果 (1,102 句对)

(2) 测试语料

本次机器翻译评测的数据包括新闻语料、科技语料和日常用语语料。语料的规模见表 3.2。在评测中，英汉-汉英方向发送给被测单位的语料大约是真实测试集规模的 4 倍，汉蒙方向发送给被测单位的语料大约是真实测试集规模的 11 倍，测试语料包括真实测试集及干扰集。干扰集的结果在评判时被舍弃。

表 3.2 机器翻译测试语料规模

机器翻译评测语料	句子数	汉字数/单词数
汉英新闻领域单一系统	1003	39,828
英汉新闻领域机器翻译	1002	21,553

英汉科技领域机器翻译	1116	21,045
汉蒙日常用语机器翻译	400	4903

新闻测试语料采集自国内外新闻网站，由中科院计算所制作；科技测试语料由中国科学技术信息研究所（以下简称中信所）提供，汉蒙日常用语的测试语料由内蒙古大学和计算所联合提供。

在 CWMT2009 汉英新闻单一系统的评测中，共有 6 家单位的 12 个系统提供了 n-best 翻译结果，我们将这 12 个翻译结果作为本次汉英新闻系统融合项目的测试语料。

为了确保训练数据和测试数据不重叠，本次评测的数据原则上要求按照评测大纲规定的分割日期要求采集和提供，即评测组织方提供的测试数据都是在分割日期（2009 年 1 月 1 日）之后（含分割日期）产生的数据，而所有的训练数据和开发数据，包括评测组织方提供的数据和参评单位自己收集的数据，都要求是在分割日期之前（不含分割日期）产生的数据。

今年我们在汉英方向进行了 WoodPecker 评测，在 WoodPecker 检测点测试数据的制作过程中进行了人工干预，制作完成后检测点的原文抽查正确率为 98%，译文抽查正确率为 99%。该制作流程如下：

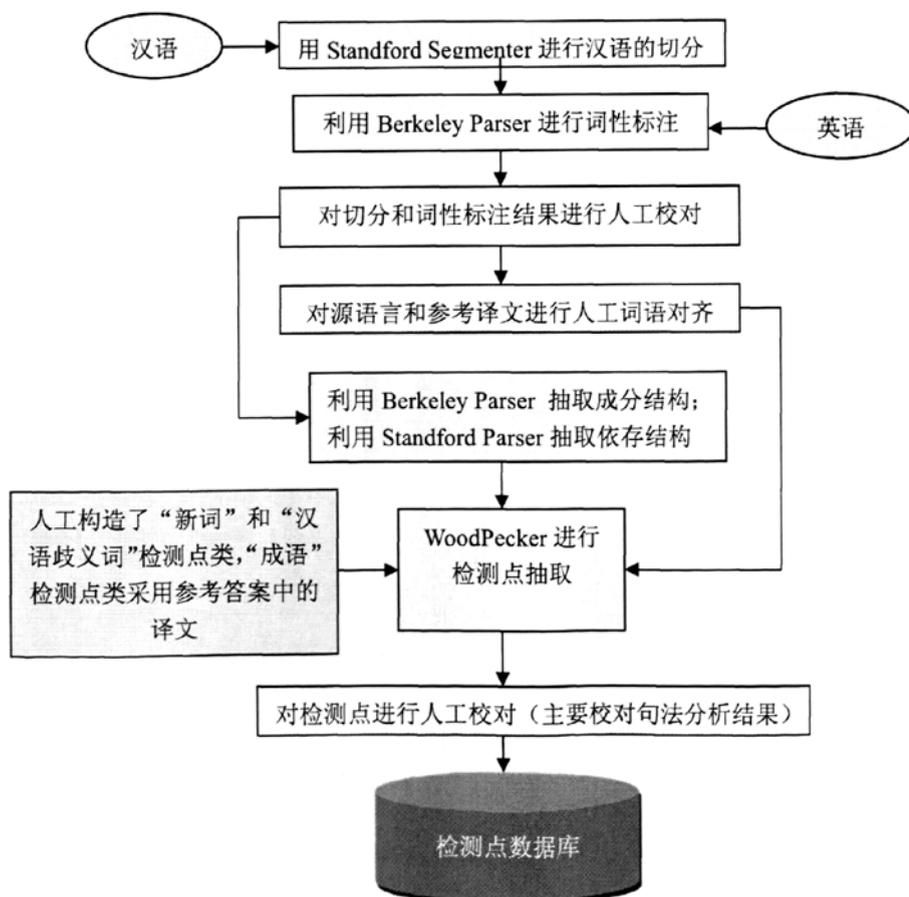


图 3.1 WoodPecker 检测点测试数据制作流程

另外，在 WoodPecker 检测点测试集的制作中，我们淘汰了少数意义不明显（如源语言端的“collocation”

检测点类)或出现频率低的检测点类,其中源语言检测点类中仅选取检测点出现频次大于10的检测点类,目标语言检测点类中仅选取检测点出现频次大于40的检测点类。

(3) 参考译文

本次评测的每个句子都包含4个参考译文。每份参考译文的原始文本均由4名经验丰富的专业翻译人员各自独立翻译而成。新闻部分的参考译文由中科院计算所提供,科技部分的参考译文由中信所提供,汉蒙日常用语部分的参考译文由内蒙古大学提供。汉英新闻系统融合项目的参考译文与CWMT2009汉英新闻单一系统的参考译文相同。

本次评测中,“汉英新闻领域单一系统”和“英汉新闻领域机器翻译”两个项目的参考译文不提供给参评单位,而是留到下次评测时继续使用,以便了解各参评单位在这一段时间间隔内的技术进步。

3.2 评测流程

本次CWMT评测采用了网上评测的方式,表3.3给出了此次评测的流程。

表3.3 CWMT2009 评测流程表

1	2009年7月15日	报名截止日期(原定6月30日报名截止,后来延长至7月15日)
2	2009年7月1日	评测组织方发放训练数据
3	2009年8月10日 上午9:00	评测组织方发放“汉英新闻领域单一系统”项目测试数据
4	2009年8月14日 上午9:00	“汉英新闻领域单一系统”项目参评单位提交运行结果和系统描述
5	2009年8月17日 上午9:00	评测组织方发放“英汉新闻领域机器翻译”和“英汉科技领域机器翻译”项目测试数据
6	2009年8月21日 上午9:00	“英汉新闻领域机器翻译”和“英汉科技领域机器翻译”项目参评单位提交运行结果和系统描述
7	2009年8月24日 上午9:00	评测组织方发放“汉英新闻领域系统融合”项目的测试数据(即“汉英新闻领域单一系统”项目参评单位提交的n-best结果的汇总)和“汉蒙日常用语机器翻译”项目的测试数据
8	2009年8月28日 上午9:00	“汉英新闻领域系统融合”和“汉蒙日常用语机器翻译”项目参评单位提交运行结果和系统描述
9	2009年9月10日	评测组织方完成结果评估,向所有参评单位通知评测结果
10	2009年9月20日	所有参评单位提交参加评测的技术报告
11	2009年10月 16-17日	在研讨会上进行研讨

3.3 结果评判

本次评测提供了评测提交结果的格式检查程序。

8月14日收到提交结果后开始进行自动评测,9月9日自动评测全部完成。

4 评测结果

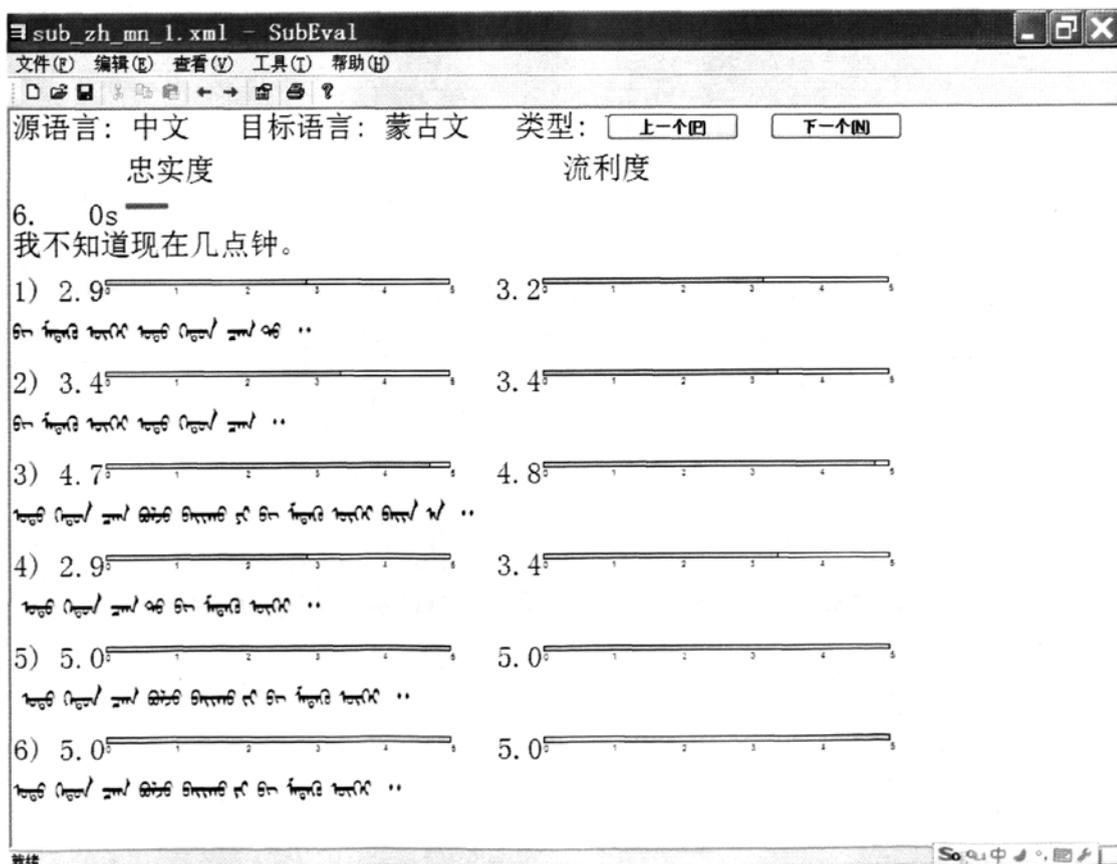
“机器翻译”项目的自动评测采用多种自动评价标准,主要评测指标为 BLEU-SBP (Chiang et al., 2008),其它自动评测指标包括: BLEU、NIST、GTM、mWER、mPER、ICT,汉英方向还采用了 Woodpecker 评测 (Zhou et al., 2008)。

自动评测的算法(包括 WoodPecker)都是大小写敏感的,中文的评测是基于字的,而不是基于词的。

在语料预处理中,汉语的 A3 区全角半角转换、英语的 tokenization 均采用中科院计算所自然语言处理研究组自己编写的程序。关于此次 Woodpecker 中出现的检测点名称的解释以及评测工具下载地址,请参见本报告的附录部分。

我们在 BLEU-SBP 的基础上,针对各主系统的翻译结果,进行了结果之间差异的显著性检验---符号检验 (Collins et al.,2005),总的做法是:分别以每个主系统为基准系统,测试了所有其它主系统与基准系统结果差异的显著性程度,以此构造了所有主系统翻译结果的差异显著性矩阵。

汉蒙评测除了自动评测外,还采用了人工评测,人工评测采用忠实度 (Adequacy) 和流利度 (Fluency) 指标,我们邀请了五位以蒙古语作为母语、汉语熟练的人类专家进行评测。下图为人工评测的工具界面:



评测结果中除了标记有“不受限”之外,其它所有系统均为受限系统,少数规则式系统也视为受限系统。

各项目具体评测结果如下:

4.1 汉英新闻单一系统

4.1.1 汉英新闻单一系统主系统评测结果

表 4.1.1 汉英新闻单一系统主系统评测结果 (15 家)

participants	BLEU4-SBP	BLEU4	NIST5	GTM	mWER	mPER	ICT
site9	0.2345	0.2477	7.4024	0.6923	0.6845	0.5030	0.3433
site7	0.2261	0.2382	7.3085	0.6976	0.6803	0.5055	0.3423
site12	0.2260	0.2348	7.9608	0.7140	0.7151	0.4908	0.3136
site6	0.2223	0.2371	7.7822	0.7127	0.7104	0.4938	0.3249
site5	0.2206	0.2365	7.5790	0.7043	0.7075	0.5024	0.3169
site13	0.2116	0.2255	7.4833	0.6730	0.7164	0.5146	0.3073
site1	0.2106	0.2243	7.4973	0.6933	0.7109	0.5058	0.3224
site15	0.2088	0.2252	7.4843	0.6899	0.7314	0.5186	0.3026
site3	0.2054	0.2145	6.8450	0.6748	0.7039	0.5190	0.3253
site11	0.1956	0.2074	7.2700	0.6820	0.7211	0.5129	0.3117
site8	0.1847	0.1926	6.4913	0.6517	0.7210	0.5405	0.3068
site14	0.1787	0.1913	7.2241	0.6831	0.7441	0.5242	0.2926
site2	0.1785	0.1801	6.9426	0.6868	0.7183	0.5171	0.2390
site4	0.1541	0.1600	5.7935	0.6053	0.7420	0.5773	0.2764
site17	0.1488	0.1592	6.2476	0.6527	0.8340	0.5515	0.2825

表 4.1.2 汉英新闻单一系统各主系统 BLEU4-SBP 差异显著性检验结果表
(显著标志●, 不显著标志○, p<0.05)

base contrast	site9	site7	site12	site6	site5	site13	site1	site15	site3	site11	site8	site14	site2	site4	site17
site9	—	○	○	●	●	●	●	●	●	●	●	●	●	●	●
site7	○	—	○	○	○	●	●	●	●	●	●	●	●	●	●
site12	○	○	—	○	○	●	●	●	●	●	●	●	●	●	●
site6	●	○	○	—	○	●	●	●	●	●	●	●	●	●	●
site5	●	○	○	○	—	○	○	●	●	●	●	●	●	●	●
site13	●	●	●	●	○	—	○	○	○	○	●	●	●	●	●
site1	●	●	●	●	○	○	—	○	●	●	●	●	●	●	●
site15	●	●	●	●	●	○	○	—	○	○	●	●	●	●	●
site3	●	●	●	●	●	○	●	○	—	○	●	●	●	●	●
site11	●	●	●	●	●	○	●	○	○	—	●	●	●	●	●
site8	●	●	●	●	●	●	●	●	●	●	—	○	●	●	●
site14	●	●	●	●	●	●	●	●	●	●	○	—	●	●	●
site2	●	●	●	●	●	●	●	●	●	●	●	●	—	○	●
site4	●	●	●	●	●	●	●	●	●	●	●	●	●	—	●
site17	●	●	●	●	●	●	●	●	●	●	●	●	●	●	—

4.1.2 汉英新闻单一系统各主系统 WoodPecker 评测结果

表 4.1.3 汉英新闻单一系统各主系统 WoodPecker 评测结果

	site12	site6	site5	site15	site9	site1	site13	site7	site11	site14	site3	site2	site8	site17	site4
GeneralScore	0.2981	0.2934	0.2862	0.2862	0.2835	0.2798	0.2773	0.277	0.2696	0.2679	0.2672	0.2616	0.2506	0.2412	0.2256
Words	0.507	0.4899	0.4744	0.4714	0.4735	0.4729	0.4707	0.4677	0.4624	0.4582	0.453	0.4562	0.4305	0.4162	0.3987
Phrases	0.2786	0.2751	0.2686	0.2689	0.2658	0.2619	0.2593	0.2592	0.2516	0.2502	0.2499	0.2435	0.2338	0.2248	0.2094
Source_Words	0.5207	0.4941	0.4682	0.4678	0.475	0.4751	0.4777	0.4719	0.4684	0.4604	0.4498	0.4743	0.425	0.4012	0.3934

Source_Special Words	0.3981	0.3099	0.2964	0.3161	0.2984	0.297	0.3099	0.2784	0.3307	0.3045	0.2828	0.3783	0.2712	0.241	0.2508
S:AmbiWord	0.4863	0.383	0.439	0.4618	0.4215	0.4	0.4033	0.3862	0.3554	0.395	0.4124	0.3793	0.374	0.3288	0.3324
S:NewWord	0.4118	0.395	0.1562	0.1397	0.1333	0.1579	0.3866	0.1667	0.4219	0.2162	0.1029	0.472	0.1111	0.125	0.15
S:Idiom	0.2605	0.1319	0.1404	0.1463	0.114	0.1599	0.1297	0.1204	0.2326	0.1579	0.1451	0.2932	0.15	0.1415	0.147
S:OverLapWord	0.4592	0.418	0.395	0.4804	0.4635	0.4162	0.3608	0.3788	0.3927	0.4355	0.3889	0.4702	0.3596	0.3281	0.3297
Source_Substan Words	0.5183	0.4948	0.4668	0.4686	0.473	0.4746	0.4778	0.4721	0.4689	0.4591	0.4491	0.4708	0.4238	0.3993	0.3904
S:Noun	0.5863	0.5573	0.507	0.4943	0.5297	0.5204	0.5316	0.5355	0.5396	0.4977	0.5027	0.5269	0.4642	0.4194	0.4077
S:Verb	0.3899	0.3696	0.3586	0.3835	0.3441	0.3687	0.36	0.3406	0.3348	0.3547	0.3278	0.3483	0.3252	0.3199	0.3044
S:Adj	0.5853	0.5757	0.563	0.6014	0.5838	0.6007	0.5764	0.5277	0.5683	0.5912	0.5626	0.4943	0.56	0.4966	0.532
S:Adverb	0.4712	0.467	0.4811	0.4835	0.4585	0.4465	0.4627	0.4458	0.4342	0.462	0.4341	0.4759	0.3971	0.4297	0.4273
S:Pron	0.7629	0.7506	0.7344	0.7191	0.7647	0.7584	0.7292	0.7533	0.6769	0.7368	0.741	0.725	0.7042	0.6608	0.6997
Source_Functional Words	0.6977	0.6538	0.6812	0.5935	0.6843	0.6499	0.6355	0.6475	0.592	0.6356	0.6288	0.6486	0.592	0.6013	0.6
S:Prep	0.6977	0.6538	0.6812	0.5935	0.6843	0.6499	0.6355	0.6475	0.592	0.6356	0.6288	0.6486	0.592	0.6013	0.6
Source Phrases	0.3761	0.3719	0.3658	0.3652	0.3628	0.3581	0.3568	0.3549	0.3423	0.3426	0.3407	0.3233	0.3188	0.2998	0.2865
S:NP	0.4146	0.4136	0.3986	0.3987	0.4004	0.3895	0.3986	0.393	0.3928	0.3748	0.3746	0.3666	0.3471	0.3239	0.3182
S:VP	0.33	0.3274	0.325	0.3261	0.3211	0.3165	0.314	0.3141	0.2952	0.3032	0.3002	0.2776	0.2813	0.2637	0.2486
S:PP	0.4316	0.418	0.4109	0.3946	0.4122	0.4056	0.3925	0.4029	0.3821	0.3853	0.3879	0.3834	0.3646	0.3438	0.3256
S:QP	0.6486	0.6658	0.6511	0.6444	0.5788	0.6458	0.6171	0.577	0.6298	0.6366	0.6341	0.553	0.5824	0.5427	0.6016
S:Locate_Phr	0.4582	0.4213	0.448	0.4483	0.4416	0.4463	0.4362	0.421	0.3922	0.3951	0.4	0.4059	0.3737	0.3735	0.3332
S:Sub_Predi	0.4606	0.4289	0.4089	0.3931	0.4164	0.3997	0.4081	0.4169	0.4163	0.3777	0.3994	0.3748	0.3658	0.3397	0.34

S:Predi_Obj	0. 416	0. 4123	0. 4021	0. 4094	0. 3935	0. 395	0. 3878	0. 379	0. 3808	0. 3943	0. 3729	0. 3598	0. 3578	0. 3323	0. 3139
S:Prep_Obj	0. 6249	0. 564	0. 5657	0. 5253	0. 5841	0. 5622	0. 5282	0. 5368	0. 5189	0. 528	0. 5134	0. 5397	0. 4992	0. 4843	0. 4498
S:Adv_MOD	0. 41	0. 4027	0. 4063	0. 4157	0. 3904	0. 4006	0. 4018	0. 3735	0. 366	0. 388	0. 3608	0. 3489	0. 3552	0. 3427	0. 3346
S:Adj_MOD	0. 5514	0. 5373	0. 5092	0. 5339	0. 5425	0. 5346	0. 5254	0. 4915	0. 5493	0. 5263	0. 4662	0. 4919	0. 5293	0. 462	0. 4455
S:Noun_MOD	0. 4925	0. 4993	0. 477	0. 4725	0. 475	0. 4725	0. 4844	0. 475	0. 4732	0. 4449	0. 4631	0. 4429	0. 4184	0. 3966	0. 3809
Target_Words	0. 5029	0. 4887	0. 4763	0. 4724	0. 4731	0. 4723	0. 4687	0. 4664	0. 4607	0. 4575	0. 4539	0. 4508	0. 4322	0. 4207	0. 4002
Target_Substan Words	0. 4549	0. 4397	0. 4123	0. 4111	0. 4172	0. 4237	0. 4312	0. 4149	0. 4192	0. 4052	0. 4012	0. 3894	0. 3816	0. 3476	0. 3522
T:Noun	0. 4911	0. 4723	0. 4252	0. 4172	0. 4463	0. 4399	0. 4566	0. 4464	0. 4582	0. 4085	0. 4192	0. 4309	0. 3878	0. 3455	0. 3464
T:Verb	0. 2777	0. 2606	0. 2593	0. 2735	0. 2382	0. 2641	0. 2552	0. 2344	0. 2193	0. 2475	0. 2342	0. 1669	0. 2306	0. 2144	0. 2108
T:Adj	0. 4809	0. 4582	0. 4513	0. 4591	0. 4487	0. 4705	0. 4703	0. 4496	0. 4642	0. 4918	0. 4384	0. 4163	0. 4395	0. 3998	0. 4003
T:Adverb	0. 373	0. 3753	0. 3751	0. 3926	0. 3786	0. 3838	0. 3891	0. 3674	0. 3769	0. 3836	0. 3672	0. 3737	0. 3356	0. 3284	0. 348
T:Pron	0. 5583	0. 5429	0. 5319	0. 5191	0. 5183	0. 5321	0. 5247	0. 5247	0. 4732	0. 5065	0. 5237	0. 51	0. 5309	0. 4927	0. 5219
T:Quantity	0. 7403	0. 7546	0. 7164	0. 6891	0. 6856	0. 717	0. 7096	0. 6634	0. 68	0. 6908	0. 7062	0. 6068	0. 68	0. 6612	0. 6993
T:AdjTrans	0. 4428	0. 5065	0. 5229	0. 4837	0. 4493	0. 4853	0. 5327	0. 3873	0. 4673	0. 5196	0. 4526	0. 4601	0. 4869	0. 4199	0. 4232
T:AdvTrans	0. 6262	0. 6589	0. 6449	0. 6822	0. 6449	0. 7103	0. 7336	0. 6262	0. 6682	0. 7477	0. 6589	0. 5421	0. 5794	0. 5561	0. 6121
Target_Functio nalWords	0. 6146	0. 6025	0. 6249	0. 6149	0. 603	0. 5851	0. 5557	0. 5862	0. 557	0. 579	0. 5764	0. 5935	0. 5496	0. 5906	0. 5118
T:Prep	0. 5163	0. 5042	0. 5335	0. 5321	0. 4958	0. 4795	0. 4746	0. 4631	0. 4285	0. 4713	0. 4674	0. 5169	0. 4259	0. 483	0. 4162
T:ModalVerb	0. 2542	0. 2783	0. 2708	0. 2917	0. 2925	0. 2725	0. 2675	0. 2683	0. 24	0. 2708	0. 2617	0. 2362	0. 2617	0. 2517	0. 2317

T:Article	0.7623	0.7466	0.7642	0.7413	0.7559	0.7364	0.6767	0.7578	0.7348	0.7323	0.7317	0.7163	0.719	0.7467	0.6485
Target_Phrases	0.2614	0.258	0.2514	0.2518	0.2487	0.245	0.2423	0.2424	0.2357	0.2339	0.2338	0.2292	0.2188	0.2114	0.1959
T:NP	0.3011	0.2998	0.284	0.2805	0.2864	0.2761	0.2759	0.2808	0.2806	0.2634	0.2693	0.2671	0.2506	0.2385	0.2234
T:VP	0.2186	0.2151	0.2145	0.2168	0.2097	0.2078	0.2038	0.2045	0.1929	0.1979	0.196	0.19	0.1837	0.1791	0.1613
T:PP	0.2978	0.2961	0.289	0.2876	0.2844	0.2784	0.273	0.2726	0.2635	0.2643	0.2671	0.2684	0.2481	0.2476	0.2257
T:QP	0.6116	0.6318	0.6331	0.5602	0.5277	0.5127	0.5992	0.4652	0.527	0.5569	0.5394	0.4833	0.5088	0.499	0.4776
T:ADVP	0.2856	0.2749	0.2706	0.2656	0.2773	0.2529	0.2598	0.2801	0.2596	0.2527	0.252	0.244	0.2436	0.2434	0.2126
T:ADJP	0.2585	0.2634	0.2581	0.2664	0.2496	0.2606	0.2512	0.2387	0.2527	0.2466	0.2414	0.2299	0.2291	0.2117	0.2077
T:WH_Prep_Phr	0.1605	0.2346	0.2222	0.2099	0.2407	0.2407	0.1975	0.1543	0.1481	0.1728	0.1914	0.1501	0.1481	0.216	0.2099
T:WH_Noun_Phr	0.1893	0.1778	0.2294	0.2218	0.2772	0.2199	0.2887	0.2027	0.1415	0.1855	0.2103	0.2307	0.1683	0.2314	0.1893
T:CONJ_Phr	0.441	0.4028	0.4444	0.4236	0.4236	0.4062	0.4444	0.3333	0.4653	0.3299	0.3715	0.3865	0.3299	0.4167	0.3229
T:Sub_Predi	0.3272	0.3059	0.2866	0.2858	0.2874	0.2898	0.2984	0.2885	0.2796	0.2791	0.2822	0.2606	0.2627	0.2356	0.2605
T:Predi_Obj	0.2584	0.2403	0.2315	0.2442	0.2286	0.2474	0.2407	0.219	0.2255	0.2306	0.2155	0.2151	0.2075	0.1912	0.1866
T:Adv_MOD	0.256	0.2507	0.2517	0.2537	0.2514	0.2602	0.2623	0.2406	0.2344	0.2522	0.2273	0.2277	0.2222	0.2015	0.2171
T:Adj_MOD	0.3884	0.3784	0.3676	0.3723	0.3621	0.3787	0.3756	0.3611	0.3812	0.3878	0.353	0.3326	0.3451	0.3226	0.3225
T:Be_Predi	0.2108	0.2097	0.1963	0.2304	0.1821	0.203	0.2025	0.1901	0.1746	0.1781	0.1786	0.1656	0.1735	0.1743	0.1606

4.1.3 汉英新闻单一系统所有参评系统评测结果

表 4.1.4 汉英新闻单一系统所有参评系统评测结果 (32 个)

systems	BLEU4-SBP	BLEU4	NIST5	GTM	mWER	mPER	ICT
zh_en_news_trans_site6_contrast_systemd.result	0.2409	0.2559	8.0473	0.7280	0.6900	0.4838	0.3339
zh_en_news_trans-site9-primary-systems.result	0.2345	0.2477	7.4024	0.6923	0.6845	0.5030	0.3433
zh_en_news_trans-site9-contrast-systemd.result	0.2330	0.2472	7.3670	0.6899	0.6841	0.5082	0.3242
zh_en_news_trans_site6_contrast_systeme.result	0.2317	0.2489	7.8652	0.7140	0.6915	0.4904	0.3326
zh_en_news_trans-site7-contrast-systemc.result	0.2300	0.2444	7.6165	0.6989	0.6808	0.4985	0.3363
zh_en_news_trans-site7-contrast-systemd.result	0.2268	0.2392	7.3987	0.6965	0.6808	0.5038	0.3389
ce-news-test-site12-constrst-systemb.result	0.2262	0.2348	7.9218	0.7097	0.7152	0.4939	0.3089
zh_en_news_trans-site7-primary-systema.result	0.2261	0.2382	7.3085	0.6976	0.6803	0.5055	0.3423
ce-news-test-site12-primary-systema.result	0.2260	0.2348	7.9608	0.7140	0.7151	0.4908	0.3136
zh_en_news_trans_site6_contrast_systemsc.result	0.2251	0.2421	7.8081	0.7176	0.7064	0.4913	0.3294
zh_en_news_trans-site9-contrast-systemh.result	0.2231	0.2357	7.2634	0.6913	0.6840	0.5046	0.3363
zh_en_news_trans_site6_primary_systema.result	0.2223	0.2371	7.7822	0.7127	0.7104	0.4938	0.3249
site5-ce-news-tst_primary.result	0.2206	0.2365	7.5790	0.7043	0.7075	0.5024	0.3169
zh_en_news_trans-site7-contrast-systemb.result	0.2206	0.2337	7.2914	0.7032	0.6875	0.5050	0.3430
site13_primary_chiero_cwmt09_nbest_f.result	0.2116	0.2255	7.4833	0.6730	0.7164	0.5146	0.3073
zh_en_news_singl-site1-primary-jianzhen.result	0.2106	0.2243	7.4973	0.6933	0.7109	0.5058	0.3224
zh_en_news_trans_site15_ltrc-primary-systema.result	0.2088	0.2252	7.4843	0.6899	0.7314	0.5186	0.3026
site13_contrast_moses_cwmt09_nbest_f.result	0.2065	0.2158	7.1201	0.6748	0.7010	0.5157	0.3303

zh_en_news_trans-site3-primary-systema.result	0.2054	0.2145	6.8450	0.6748	0.7039	0.5190	0.3253
zh_en_news_trans-site3-contrast-systemb.result	0.2006	0.2079	6.7343	0.6693	0.6983	0.5223	0.3273
zh-en-news-singl-site11-primary-pbtm.result	0.1956	0.2074	7.2700	0.6820	0.7211	0.5129	0.3117
zh_en_news_trans-site3-contrast-systemd.result	0.1951	0.2024	6.6795	0.6624	0.7016	0.5276	0.3240
zh_en_news_trans-site3-contrast-systemc.result	0.1934	0.1991	6.3633	0.6549	0.7000	0.5325	0.3269
zh_en_news_trans_site15_ltrc-contrast-systemb.result	0.1901	0.1970	7.3208	0.7032	0.7514	0.5219	0.2610
zh_en_news_trans-site8-contrast-systemb.result	0.1867	0.1948	6.5732	0.6550	0.7218	0.5371	0.3068
zh_en_news_trans-site8-primary-systema.result	0.1847	0.1926	6.4913	0.6517	0.7210	0.5405	0.3068
zh_en_news_trans-site14-primary.result	0.1787	0.1913	7.2241	0.6831	0.7441	0.5242	0.2926
zh_en_news_trans-site2-primary-systema.result	0.1785	0.1801	6.9426	0.6868	0.7183	0.5171	0.2390
zh-en-news-singl-site11-contrast-moses.result	0.1754	0.1825	6.4747	0.6414	0.7226	0.5428	0.3015
zh_en_news_trans-site4-primary-sys_main.result	0.1541	0.1600	5.7935	0.6053	0.7420	0.5773	0.2764
site17-ce-news-test.submission_primary.result	0.1488	0.1592	6.2476	0.6527	0.8340	0.5515	0.2825
zh_en_news_trans-site2-contrast-systemb.result	0.1197	0.1286	6.0791	0.5908	0.7964	0.5888	0.2427

4.2 英汉新闻

4.2.1 英汉新闻主系统评测结果

表 4.2.1 英汉新闻主系统评测结果(9家)

participants	BLEU5-SBP	BLEU5	BLEU6	NIST6	NIST7	GTM	mWER	mPER	ICT
site13	0.3563	0.3727	0.3070	10.0201	10.0328	0.7975	0.6131	0.3630	0.4284
site6	0.3352	0.3478	0.2834	9.6229	9.6325	0.7876	0.6310	0.3726	0.4259

site2	0.3350	0.3411	0.2774	9.9471	9.9576	0.8087	0.6147	0.3670	0.3829
site1	0.3345	0.3445	0.2810	9.5447	9.5554	0.7757	0.6271	0.3799	0.4272
site15	0.3179	0.3268	0.2645	9.2265	9.2363	0.7771	0.6333	0.3823	0.4351
site12	0.3138	0.3275	0.2626	9.5463	9.5557	0.7779	0.6716	0.3881	0.3679
site8	0.2997	0.3115	0.2477	9.1022	9.1105	0.7600	0.6565	0.4026	0.3835
site11	0.2781	0.2870	0.2277	8.5056	8.5130	0.7415	0.6534	0.4225	0.3935
site17	0.2410	0.2543	0.1925	8.5770	8.5815	0.7454	0.7936	0.4195	0.3377

表 4. 2. 2 英汉新闻各主系统 BLEU5-SBP 差异显著性检验结果表
(显著标志●, 不显著标志○, $p < 0.05$)

base contrast	site13	site6	site2	site1	site15	site12	site8	site11	site17
site13	—	●	●	●	●	●	●	●	●
site6	●	—	○	○	●	●	●	●	●
site2	●	○	—	●	○	●	●	●	●
site1	●	○	●	—	●	●	●	●	●
site15	●	●	○	●	—	●	●	●	●
site12	●	●	●	●	●	—	○	●	●
site8	●	●	●	●	●	○	—	●	●
site11	●	●	●	●	●	●	●	—	●
site17	●	●	●	●	●	●	●	●	—

4.2.2 英汉新闻所有参评系统评测结果

表 4.2.3 英汉新闻所有参评系统评测结果 (21 个)

systems	BLEU5-SBP	BLEU5	BLEU6	NIST6	NIST7	GTM	mWER	mPER	ICT
ec-news-site13-primary-syscomb.result	0.3563	0.3727	0.3070	10.0201	10.0328	0.7975	0.6131	0.3630	0.4284
ec-news-site13-contrast2-chiero.result	0.3422	0.3556	0.2908	9.7978	9.8090	0.7910	0.6253	0.3669	0.4234
ec_news_site6_contrast_systemb.result	0.3406	0.3518	0.2886	9.5470	9.5571	0.7852	0.6225	0.3756	0.4350
ec-news-site13-contrast4-silenus.result	0.3404	0.3567	0.2911	9.9430	9.9544	0.7962	0.6318	0.3705	0.4054
ec-news-site13-contrast3-moses.result	0.3382	0.3529	0.2869	9.8168	9.8275	0.7914	0.6377	0.3683	0.4153
ec_news_site6_primary_systema.result	0.3352	0.3478	0.2834	9.6229	9.6325	0.7876	0.6310	0.3726	0.4259
ec-news-site2-primary-systema.result	0.3350	0.3411	0.2774	9.9471	9.9576	0.8087	0.6147	0.3670	0.3829
ec-news-site1-primary-jianzhen.result	0.3345	0.3445	0.2810	9.5447	9.5554	0.7757	0.6271	0.3799	0.4272
ec-news_site15_ltrc-contrast-systemb.result	0.3217	0.3311	0.2688	9.2729	9.2833	0.7783	0.6300	0.3804	0.4361
ec-news_site15_ltrc-primary-systema.result	0.3179	0.3268	0.2645	9.2265	9.2363	0.7771	0.6333	0.3823	0.4351
ec-news-site12-contrast-systemb.result	0.3166	0.3312	0.2659	9.5856	9.5956	0.7786	0.6697	0.3862	0.3712
ec-news-site8-contrast-systemc.result	0.3163	0.3277	0.2652	9.2178	9.2275	0.7649	0.6453	0.3972	0.3975
ec-news_site15_ltrc-contrast-systemc.result	0.3157	0.3249	0.2630	9.2191	9.2288	0.7775	0.6355	0.3828	0.4338
ec-news-site12-primary-systema.result	0.3138	0.3275	0.2626	9.5463	9.5557	0.7779	0.6716	0.3881	0.3679
ec-news-site8-primary-systemd.result	0.2997	0.3115	0.2477	9.1022	9.1105	0.7600	0.6565	0.4026	0.3835
ec-news-site8-contrast-systemb.result	0.2992	0.3091	0.2459	8.9962	9.0043	0.7539	0.6528	0.4035	0.3917
ec-news-site8-contrast-systeme.result	0.2948	0.3048	0.2436	8.8050	8.8127	0.7413	0.6533	0.4155	0.3783
ec-news-site11-primary-memt.result	0.2781	0.2870	0.2277	8.5056	8.5130	0.7415	0.6534	0.4225	0.3935

ec-news-site11-contrast-moses.result	0.2580	0.2624	0.2053	8.1694	8.1756	0.7281	0.6557	0.4210	0.3940
ec-news-site17_primary.result	0.2410	0.2543	0.1925	8.5770	8.5815	0.7454	0.7936	0.4195	0.3377
ec-news-site11-contrast-pbtm.result	0.2400	0.2553	0.1961	8.5045	8.5095	0.7351	0.7154	0.4452	0.3209

4.3 英汉科技

4.3.1 英汉科技主系统评测结果

表 4.3.1 英汉科技主系统评测结果（10 家）

participants	BLEU5-SBP	BLEU5	BLEU6	NIST6	NIST7	GTM	mWER	mPER	ICT
site13	0.4905	0.4985	0.4313	11.5429	11.5669	0.8800	0.4971	0.2808	0.5348
site12	0.4828	0.4876	0.4204	11.3342	11.3579	0.8814	0.4902	0.2884	0.5145
site6	0.4767	0.4900	0.4216	11.4158	11.4394	0.8736	0.5175	0.2828	0.5543
site8	0.4619	0.4756	0.4061	11.1680	11.1889	0.8605	0.5328	0.2955	0.5331
site1	0.4567	0.4640	0.3965	11.1424	11.1642	0.8723	0.5314	0.2891	0.5145
site15	0.4497	0.4575	0.3875	11.1445	11.1640	0.8724	0.5457	0.2906	0.5234
site14	0.4427	0.4507	0.3810	11.0264	11.0462	0.8639	0.5378	0.3006	0.5116
site11	0.3973	0.4095	0.3389	10.4956	10.5098	0.8524	0.5747	0.3247	0.4779
site17	0.3883	0.3924	0.3237	10.3937	10.4077	0.8774	0.6340	0.2938	0.4874
site2	0.3826	0.3847	0.3179	10.1776	10.1901	0.8422	0.5226	0.3342	0.4193

表 4.3.2 英汉科技各主系统 BLEU5-SBP 差异显著性检验结果表
(显著标志●, 不显著标志○, $p < 0.05$)

base contrast	site13	site12	site6	site8	site1	site15	site14	site11	site17	site2
site13	—	○	●	●	●	●	●	●	●	●
site12	○	—	○	●	●	●	●	●	●	●
site6	●	○	—	●	●	●	●	●	●	●
site8	●	●	●	—	○	●	●	●	●	●
site1	●	●	●	○	—	○	●	●	●	●
site15	●	●	●	●	○	—	○	●	●	●
site14	●	●	●	●	●	○	—	●	●	●
site11	●	●	●	●	●	●	●	—	○	●
site17	●	●	●	●	●	●	●	○	—	○
site2	●	●	●	●	●	●	●	●	○	—

4.3.2 英汉科技所有参评系统评测结果

表 4.3.3 英汉科技所有参评系统评测结果 (24 个)

systems	BLEU5-SBP	BLEU5	BLEU6	NIST6	NIST7	GTM	mWER	mPER	ICT
ec-tech-site13-primary-syscomb.result	0.4905	0.4985	0.4313	11.5429	11.5669	0.8800	0.4971	0.2808	0.5348
ec-tech-site13-contrast4-silenus.result	0.4838	0.4902	0.4246	11.4175	11.4409	0.8764	0.5009	0.2854	0.5204
ec-tech-site12-primary-systema.result	0.4828	0.4876	0.4204	11.3342	11.3579	0.8814	0.4902	0.2884	0.5145
ec-tech-site12-contrast-systemb.result	0.4826	0.4875	0.4201	11.3393	11.3631	0.8807	0.4900	0.2885	0.5155
ec-tech-site13-contrast3-moses.result	0.4801	0.4919	0.4221	11.4840	11.5062	0.8760	0.5237	0.2769	0.5577
ec-tech-site6-primary-systema.result	0.4767	0.4900	0.4216	11.4158	11.4394	0.8736	0.5175	0.2828	0.5543
ec-tech-site6-contrast-systemc.result	0.4733	0.4830	0.4151	11.3711	11.3938	0.8747	0.5312	0.2847	0.5367

ec-tech-site6-contrast-systemb.result	0.4688	0.4817	0.4130	11.3120	11.3343	0.8705	0.5265	0.2859	0.5471
ec-tech-site6-contrast-systemd.result	0.4656	0.4838	0.4136	11.2657	11.2882	0.8668	0.5215	0.2875	0.5582
ec-tech-site13-contrast2-chiero.result	0.4656	0.4809	0.4113	11.2803	11.3034	0.8695	0.5236	0.2895	0.5470
ec-tech-site8-contrast-systemb.result	0.4625	0.4786	0.4085	11.1447	11.1653	0.8560	0.5297	0.2953	0.5422
ec-tech-site8-primary-systema.result	0.4619	0.4756	0.4061	11.1680	11.1889	0.8605	0.5328	0.2955	0.5331
ec-tech-site1-primary-jianzhen.result	0.4567	0.4640	0.3965	11.1424	11.1642	0.8723	0.5314	0.2891	0.5145
ec-tech-site14-contrast2.result	0.4500	0.4617	0.3903	11.1405	11.1610	0.8641	0.5448	0.2958	0.5321
ec-tech_site15_ltrc-primary-systema.result	0.4497	0.4575	0.3875	11.1445	11.1640	0.8724	0.5457	0.2906	0.5234
ec-tech-site14-contrast1.result	0.4443	0.4520	0.3823	11.0544	11.0738	0.8624	0.5358	0.2992	0.5095
ec-tech-site14-primary.result	0.4427	0.4507	0.3810	11.0264	11.0462	0.8639	0.5378	0.3006	0.5116
ec-tech_site15_ltrc-contrast-systemb.result	0.4418	0.4492	0.3791	11.0529	11.0714	0.8699	0.5562	0.2941	0.5154
ec-tech-site11-contrast-moses.result	0.4126	0.4313	0.3607	10.5208	10.5366	0.8386	0.5478	0.3110	0.5380
ec-tech-site11-primary-memt1.result	0.3973	0.4095	0.3389	10.4956	10.5098	0.8524	0.5747	0.3247	0.4779
ec-tech-site17-primary.result	0.3883	0.3924	0.3237	10.3937	10.4077	0.8774	0.6340	0.2938	0.4874
ec-tech-site2-primary-systema.result	0.3826	0.3847	0.3179	10.1776	10.1901	0.8422	0.5226	0.3342	0.4193
ec-tech-site11-contrast-memt2.result	0.3678	0.3723	0.3044	9.9743	9.9865	0.8518	0.5720	0.3374	0.4214
ec-tech-site11-contrast-pbtm.result	0.3304	0.3330	0.2656	9.3968	9.4045	0.8423	0.5967	0.3532	0.3822

4.4 汉蒙日常用语

4.4.1 汉蒙日常用语主系统评测结果

表 4.4.1 汉蒙日常用语主系统自动评测结果 (7 家)

participants	BLEU4-SBP	BLEU4	NIST5	GTM	mWER	mPER	ICT
zh_mn_dail_trans-site1-primary-jianzhen.result	0.2166	0.2269	6.1648	0.6455	0.5683	0.5048	0.4521
zh_mn_dail_trans-site13-primary-chiero.result	0.2157	0.2312	6.1213	0.6280	0.5801	0.5073	0.4695
zh_mn_dail_trans-site8-primary-systema.result	0.2077	0.2227	5.9730	0.6221	0.5877	0.5091	0.4757
zm_dial_test_site17_submission_primary.result	0.1445	0.1564	4.9564	0.5717	0.6591	0.5593	0.4594
zm_dial_test_site10_result_primary.result	0.1432	0.1517	4.8815	0.5185	0.6759	0.5954	0.3838
zh_mn_dail_trans-site11-primary-moses.result	0.1054	0.1128	4.4310	0.5032	0.7009	0.6236	0.3877
zh_mn_dail_trans-site16-primary-systema.result	0.0521	0.0521	2.6096	0.3868	0.7484	0.7189	0.1311

表 4.4.2 汉蒙日常用语各主系统 BLEU4-SBP 差异显著性检验结果表

(显著标志●,不显著标志○,p<0.05)

base contrast	site1	site13	site8	site17	site10	site11	site16
site1	—	○	○	●	●	●	●
site13	○	—	○	●	●	●	●
site8	○	○	—	●	●	●	●
site17	●	●	●	—	○	●	●
site10	●	●	●	○	—	○	●
site11	●	●	●	●	○	—	●
site16	●	●	●	●	●	●	—

表 4.4.3 汉蒙日常用语各主系统人工评测结果

participants	Average Score of Adequacy	Average Score of Fluency
site1	2.83	2.92
site13	2.72	2.82
site8	2.65	2.80
site10	2.55	2.56
site16	2.32	2.41
site11	2.12	2.19
site17	2.09	2.33

4.4.2 汉蒙日常用语所有参评系统评测结果

表 4.4.4 汉蒙日常用语所有参评系统评测结果 (11 个)

systems	BLEU4-SBP	BLEU4	NIST5	GTM	mWER	mPER	ICT	受限情况
zh_mn_dail_trans-site8-contrast-systemb.result	0.2206	0.2339	6.1540	0.6282	0.5591	0.4993	0.4795	不受限
zh_mn_dail_trans-site1-primary-jianzhen.result	0.2166	0.2269	6.1648	0.6455	0.5683	0.5048	0.4521	
zh_mn_dail_trans-site13-primary-chiero.result	0.2157	0.2312	6.1213	0.6280	0.5801	0.5073	0.4695	
zh_mn_dail_trans-site8-primary-systema.result	0.2077	0.2227	5.9730	0.6221	0.5877	0.5091	0.4757	
zh_mn_dail_trans-site13-contrast1-moses.result	0.1988	0.2138	5.6610	0.6145	0.5802	0.5117	0.4941	
zh_mn_dail_trans-site8-contrast-systemc.result	0.1890	0.1989	5.5309	0.5870	0.6180	0.5359	0.4560	不受限
zm_dial_test_site17_submission_primary.result	0.1445	0.1564	4.9564	0.5717	0.6591	0.5593	0.4594	
zm_dial_test_site10_result_primary.result	0.1432	0.1517	4.8815	0.5185	0.6759	0.5954	0.3838	
zh_mn_dail_trans-site11-primary-moses.result	0.1054	0.1128	4.4310	0.5032	0.7009	0.6236	0.3877	

zh_mn_dail_trans-site16-primary-systema.result	0.0521	0.0521	2.6096	0.3868	0.7484	0.7189	0.1311	
zh_mn_dail_trans-site16-contrast-systemb.result	0.0476	0.0477	2.4725	0.3710	0.7560	0.7313	0.1262	

4.5 汉英新闻系统融合评测结果

4.5.1 汉英新闻系统融合主系统评测结果

表 4.5.1 汉英新闻系统融合主系统评测结果 (7 家)

participants	BLEU4-SBP	BLEU4	NIST5	GTM	mWER	mPER	ICT
site9	0.2427	0.2541	7.1467	0.7025	0.6570	0.5032	0.3635
site15	0.2422	0.2495	6.5776	0.7090	0.6477	0.5035	0.3967
site13	0.2407	0.2535	7.1814	0.7012	0.6656	0.5033	0.3540
site4	0.2341	0.2475	7.5110	0.7033	0.6766	0.4986	0.3438
site11	0.2334	0.2552	7.4724	0.7142	0.6991	0.5150	0.3332
site6	0.2327	0.2473	7.5500	0.7065	0.6788	0.4980	0.3418
site8	0.2294	0.2395	6.8788	0.6986	0.6658	0.5132	0.3543

表 4.5.2 汉英新闻系统融合各主系统 BLEU4-SBP 差异显著性检验结果表
(显著标志●, 不显著标志○, $p < 0.05$)

base contrast	site9	site15	site13	site4	site11	site6	site8
site9	—	○	○	○	●	●	●
site15	○	—	○	○	○	○	●
site13	○	○	—	●	●	○	●
site4	○	○	○	—	○	○	○
site11	●	○	●	○	—	○	○

site6	●	○	○	○	○	—	○
site8	●	●	●	○	○	○	—

表 4.5.3 汉英新闻系统融合各主系统 WoodPecker 评测结果

	site11	site6	site4	site13	site9	site15	site8
GeneralScore	0.2866	0.2836	0.2829	0.28	0.2789	0.277	0.2735
Words	0.4668	0.475	0.4746	0.4629	0.4626	0.4576	0.4541
Phrases	0.2698	0.2657	0.265	0.2629	0.2617	0.2601	0.2566
Source_Words	0.4607	0.4788	0.4779	0.4612	0.4649	0.4569	0.4522
Source_SpecialWords	0.274	0.2721	0.2873	0.2651	0.2818	0.2746	0.2847
S:AmbiWord	0.3979	0.367	0.3973	0.3865	0.4044	0.3887	0.393
S:NewWord	0.15	0.1667	0.1667	0.1417	0.1583	0.1333	0.175
S:Idiom	0.1176	0.1219	0.1323	0.1048	0.1144	0.1217	0.1344
S:OverLapWord	0.3717	0.4031	0.4053	0.3663	0.3929	0.3842	0.3795
Source_SubstanWords	0.4599	0.4799	0.4778	0.4607	0.4636	0.4553	0.4501
S:Noun	0.5179	0.5441	0.5395	0.5163	0.516	0.5107	0.5003
S:Verb	0.3385	0.349	0.3473	0.3386	0.341	0.3302	0.3335
S:Adj	0.5351	0.5412	0.5372	0.5266	0.5503	0.5295	0.5199
S:Adverb	0.4269	0.449	0.4576	0.4481	0.4569	0.4442	0.4357
S:Pron	0.7446	0.7665	0.7587	0.7449	0.7666	0.7467	0.7572
Source_FunctionalWords	0.6552	0.6487	0.6537	0.6574	0.6642	0.6614	0.6573
S:Prep	0.6552	0.6487	0.6537	0.6574	0.6642	0.6614	0.6573
Source_Phrases	0.3654	0.3634	0.3635	0.3587	0.3591	0.3575	0.3498
S:NP	0.4052	0.4028	0.4035	0.3916	0.3922	0.3886	0.381
S:VP	0.3224	0.3212	0.3212	0.319	0.3206	0.3183	0.3115
S:PP	0.4172	0.4118	0.4105	0.4108	0.4019	0.4091	0.4011
S:QP	0.636	0.5927	0.5962	0.5785	0.5771	0.5983	0.5507
S:Locate_Phr	0.4488	0.4464	0.4386	0.433	0.4232	0.4283	0.4304
S:Sub_Predi	0.4118	0.419	0.4213	0.4141	0.4118	0.4136	0.396

S:Predi_Obj	0.396	0.3927	0.392	0.3882	0.3961	0.3877	0.3788
S:Prep_Obj	0.5588	0.5361	0.5344	0.5652	0.5471	0.5535	0.538
S:Adv_MOD	0.3715	0.3827	0.3911	0.3838	0.3936	0.3913	0.374
S:Adj_MOD	0.4989	0.5097	0.4973	0.492	0.5088	0.4924	0.4871
S:Noun_MOD	0.4898	0.4869	0.4845	0.4647	0.4586	0.4574	0.4506
Target_Words	0.4687	0.4739	0.4736	0.4634	0.4619	0.4577	0.4547
Target_SubstanWords	0.4063	0.4203	0.4216	0.4026	0.4042	0.4033	0.3912
T:Noun	0.4303	0.4494	0.4499	0.4237	0.4171	0.4239	0.4071
T:Verb	0.2382	0.2411	0.2477	0.2401	0.2407	0.232	0.2334
T:Adj	0.4312	0.4549	0.4571	0.4326	0.4562	0.4416	0.4262
T:Adverb	0.3532	0.373	0.3761	0.3596	0.3707	0.3637	0.3612
T:Pron	0.5442	0.5403	0.5357	0.5278	0.5393	0.5337	0.5421
T:Quantity	0.697	0.6714	0.6674	0.6925	0.701	0.6811	0.6241
T:AdjTrans	0.4232	0.4363	0.4069	0.415	0.4265	0.4477	0.4036
T:AdvTrans	0.6121	0.6028	0.6262	0.6262	0.6542	0.6776	0.6682
Target_FunctionalWords	0.6135	0.5982	0.5943	0.6044	0.5957	0.5841	0.6022
T:Prep	0.5081	0.4817	0.4813	0.4924	0.4791	0.4645	0.485
T:ModalVerb	0.2575	0.2567	0.265	0.2625	0.2742	0.2525	0.2617
T:Article	0.7689	0.7647	0.7558	0.7658	0.7604	0.7533	0.7694
Target_Phrases	0.2526	0.2483	0.2475	0.2458	0.2444	0.2429	0.2399
T:NP	0.2931	0.288	0.2866	0.2828	0.2806	0.2791	0.2744
T:VP	0.2132	0.209	0.2081	0.2079	0.2075	0.2055	0.2038
T:PP	0.292	0.2808	0.2799	0.2814	0.2746	0.2759	0.2743
T:QP	0.5459	0.4847	0.473	0.5472	0.5107	0.5459	0.4574
T:ADVP	0.2725	0.2796	0.273	0.2844	0.2765	0.2737	0.2773
T:ADJP	0.2555	0.251	0.2513	0.2428	0.2469	0.2498	0.2481
T:WH_Prep_Phr	0.2037	0.216	0.179	0.2346	0.1975	0.2037	0.1852
T:WH_Noun_Phr	0.2218	0.2084	0.2161	0.2467	0.2409	0.2256	0.2314
T:CONJ_Phr	0.3889	0.4132	0.3715	0.3854	0.4132	0.375	0.4097
T:Sub_Predi	0.2824	0.2949	0.2951	0.2848	0.2861	0.2843	0.2709

T:Predi_Obj	0.2182	0.223	0.2254	0.2187	0.223	0.2161	0.2097
T:Adv_MOD	0.23	0.2474	0.2508	0.2388	0.2423	0.2394	0.2367
T:Adj_MOD	0.3508	0.3667	0.3695	0.3516	0.3679	0.3554	0.351
T:Be_Predi	0.1976	0.1936	0.2043	0.1982	0.2017	0.1891	0.1984

4.5.2 汉英新闻系统融合所有参评系统评测结果

表 4.5.4 汉英新闻系统融合所有参评系统评测结果（14 个）

systems	BLEU4-SBP	BLEU4	NIST5	GTM	mWER	mPER	ICT
zh_en_news_comb_site15_ltrc-contrast-systemc.result	0.2475	0.2608	7.7291	0.7123	0.6648	0.4895	0.3612
zh_en_news_comb_site15_ltrc-contrast-systemd.result	0.2473	0.2608	7.7336	0.7129	0.6667	0.4873	0.3618
zh_en_news_comb_site15_ltrc-contrast-systemb.result	0.2448	0.2524	6.7240	0.7142	0.6468	0.4980	0.4071
site9-syscombination_primary.result	0.2427	0.2541	7.1467	0.7025	0.6570	0.5032	0.3635
zh_en_news_comb_site15_ltrc-primary-systema.result	0.2422	0.2495	6.5776	0.7090	0.6477	0.5035	0.3967
zh_en_news_comb-site13_primary-system.result	0.2407	0.2535	7.1814	0.7012	0.6656	0.5033	0.3540
zh_en_news_combi_site6_contrast_systemb.result	0.2349	0.2518	7.5135	0.7136	0.6801	0.5029	0.3302
zh_en_news_combi_site6_contrast_systemc.result	0.2342	0.2497	7.6359	0.7091	0.6793	0.4944	0.3397
systemcombined-site4-primary-systema.result	0.2341	0.2475	7.5110	0.7033	0.6766	0.4986	0.3438
zh_en_news_trans-site8-contrast-systemb.result	0.2335	0.2435	7.2322	0.6984	0.6636	0.5004	0.3648
zh_en_news_combi-site11-primary-memt.result	0.2334	0.2552	7.4724	0.7142	0.6991	0.5150	0.3332
zh_en_news_combi_site6_primary_systema.result	0.2327	0.2473	7.5500	0.7065	0.6788	0.4980	0.3418
systemcombined-site4-contrast-systemb.result	0.2318	0.2438	7.3551	0.6997	0.6729	0.5015	0.3479
zh_en_news_trans-site8-primary-systema.result	0.2294	0.2395	6.8788	0.6986	0.6658	0.5132	0.3543

参考文献

- David Chiang, Steve DeNeefe, Yee Seng Chan, and Hwee Tou Ng. 2008. Decomposability of translation metrics for improved evaluation and efficient algorithms. In Proc. EMNLP 2008, pages 610-619.
- Michael Collins, Philipp Koehn, and Ivona Kučerová. 2005. Clause restructuring for statistical machine translation. In Proc. ACL 2005, pages 531-540.
- LDC. 2005. Linguistic data annotation specification: Assessment of fluency and adequacy in translation. Revision 1.5.
- Ming Zhou, Bo Wang, Shujie Liu, Mu Li, Dongdong Zhang and Tiejun Zhao. 2008. Diagnostic Evaluation of Machine Translation Systems Using Automatically Constructed Linguistic Check-Points. In Proc. Coling 2008, pages 1121-1128.

附录

I. Definition

FreeCode versus Encoded

Based on the generation style, check-points are also classified into *FreeCode* check-points and *Encoded* check-points, where *FreeCode* check-points are generated according to tags and dependencies in syntax parse trees and *Encoded* check-points are generated by linguistic knowledge that depends on manually compiled rules and extra resource files.

II. Woodpecker Toolkit Download

Woodpecker – A Linguistic Diagnostic Evaluation Platform for Machine Translation Systems:
<http://research.microsoft.com/research/downloads/details/ad240799-a9a7-4a14-a556-d6a7c7919b4a/details.aspx>

III. Check-point taxonomy for Chinese-English evaluation

Check-point Name	Meaning	Origin	Type
S:AmbiWord	Ambiguous words	Source	Encoded
S:NewWord	New words	Source	Encoded
S:Idiom	Chinese Idiom	Source	FreeCode
S:OverLapWord	Overlap words	Source	Encoded
S:Noun	Noun	Source	FreeCode

S:Verb	Verb	Source	FreeCode
S:Adj	Adjective	Source	FreeCode
S:Prep	Preposition	Source	FreeCode
S:Adverb	Adverb	Source	FreeCode
S:Pron	Pronoun	Source	FreeCode
S:NP	Noun phrase	Source	FreeCode
S:VP	Verb phrase	Source	FreeCode
S:PP	Preposition phrase	Source	FreeCode
S:QP	Quantity phrase	Source	FreeCode
S:Locate_Phr	Location phrase	Source	FreeCode
S:Sub_Predi	Subject-object relationship	Source	FreeCode
S:Predi_Obj	Predicate-object relationship	Source	FreeCode
S:Prep_Obj	Preposition-object relationship	Source	FreeCode
S:Adv_MOD	Adverb modified relationship	Source	FreeCode
S:Adj_MOD	Adjective modified relationship	Source	FreeCode
S:Noun_MOD	Noun modified relationship	Source	FreeCode
T:Noun	Noun	Target	FreeCode
T:Verb	Verb	Target	FreeCode
T:Adj	Adjective	Target	FreeCode
T:Prep	Preposition	Target	FreeCode
T:Adverb	Adverb	Target	FreeCode
T:Pron	Pronoun	Target	FreeCode
T:Quantity	Quantity	Target	FreeCode
T:AdjTrans	Comparative and superlative forms of adjective	Target	FreeCode
T:AdvTrans	Comparative and superlative forms of adverb	Target	FreeCode
T:ModalVerb	Modal verb	Target	FreeCode
T:Article	Article	Target	FreeCode
T:NP	Noun phrase	Target	FreeCode

T:VP	Verb phrase	Target	FreeCode
T:PP	Preposition phrase	Target	FreeCode
T:QP	Quantity phrase	Target	FreeCode
T:ADVP	Adverb phrase	Target	FreeCode
T:ADJP	Adjective phrase	Target	FreeCode
T:WH_Prep_Phr	“WH” used as preposition phrase	Target	FreeCode
T:WH_Noun_Phr	“WH” used as noun phrase	Target	FreeCode
T:CONJ_Phr	Conjunction phrase	Target	FreeCode
T:Sub_Predi	Subject-predicate phrase	Target	FreeCode
T:Predi_Obj	Predicate-object phrase	Target	FreeCode
T:Adv_MOD	Adverb modified relationship	Target	FreeCode
T:Adj_MOD	Adjective modified relationship	Target	FreeCode
T:Be_Predi	Copula-predicate relationship	Target	FreeCode