## 第十五届全国机器翻译大会机器翻译评测大纲

## （CCMT 2019 MT Evaluation）

### 1. 引言

　　第十五届全国机器翻译大会（CCMT 2019）将于2019年9月27日至29日在中国江西南昌举行。根据惯例，本次会议将继续组织统一的机器翻译评测。

　　CCMT 2019机器翻译评测包括：

* 新增语音翻译任务，由百度公司提供支持；
* 翻译质量估计任务中新增词汇级质量评估项目，由阿里巴巴（中国）有限公司提供支持；
* 翻译质量估计任务中现有的句子级质量评估项目进一步扩大了数据集的规模，仍由北京语智云帆科技有限公司提供支持。
* 其余翻译任务与CWMT2018保持一致，包括由CCMT与WMT2019合作组织的汉英、英汉新闻领域的翻译评测；维汉\*、蒙汉、藏汉的翻译评测；以及专利领域的日、汉、英多语言翻译评测等。

（\*该项目训练数据相对2018年有所减少）

　　与上次评测相同，本次评测不再设置统一发放数据的时间，各参评单位报名之后即可获取数据并进行系统训练，请有意向参与的单位尽快报名；本次评测的评测论文将与CCMT2019学术论文一同接受匿名审稿，并择优推荐与CCMT2019学术论文共同发表。

　　希望本次评测能够促进国内外科研单位、产业界相关单位之间的学术交流和联系，共同推动机器翻译研究和技术的发展。

　　本次评测的组织信息如下（三个以上的并列项以汉语拼音为序）：

　　评测主办机构：  
　　　　中国中文信息学会

评测组织单位：  
　　　　中国中文信息学会机器翻译专业委员会

评测资源提供单位：

阿里巴巴（中国）有限公司  
百度公司  
北京语智云帆科技有限公司  
点通数据有限公司  
东北大学  
南京大学  
内蒙古大学  
青海师范大学  
西北民族大学  
西藏大学  
厦门大学  
中国科学院合肥智能机械研究所  
中国科学院计算技术研究所  
中国科学院新疆理化技术研究所  
中国科学院自动化研究所

　　评测委员会主席：  
　　　　杨沐昀（哈尔滨工业大学）

评测委员会委员：  
　　　　才让加（青海师范大学）  
　　　　陈毅东（厦门大学）  
　　　　冯 洋 （中国科学院计算技术研究所）  
　　　　何中军（百度公司）  
　　　　侯宏旭（内蒙古大学）  
　　　　黄书剑（南京大学）  
　　　　骆卫华（阿里巴巴（中国）有限公司）  
　　　　那顺乌日图（内蒙古大学）  
　　　　魏勇鹏（北京语智云帆科技有限公司）  
　　　　肖 桐（东北大学）  
　　　　王 瑞（日本NICT）  
　　　　杨雅婷（中国科学院新疆理化技术研究所）  
　　　　伊力亚尔·加尔木哈买提（南京大学）  
　　　　张家俊（中国科学院自动化研究所）

有关评测的更多信息请参见： http://ccmt2019.jxnu.edu.cn/evaluation

有关会议的更多信息请参见： http://ccmt2019.jxnu.edu.cn

### 2. CCMT2019评测任务一览

　　本次机器翻译技术评测由双语翻译、多语翻译、语音翻译和翻译质量估计四个不同任务组成，我们将为每项评测项目的参评单位提供相应的训练语料、开发语料，测试语料、和免费的自动评测平台(稍后在评测网站上公布)，每个任务详细评测方法请参见大纲后续部分说明。

　　本次翻译技术评测的四个任务的具体评测项目如下所示：

表1. CCMT 2019任务1：双语翻译评测项目表

| **序号** | **项目代号** | **评测项目名称** | **翻译方向** | **领域** |
| --- | --- | --- | --- | --- |
| 1 | CE | 汉英新闻领域机器翻译 | 汉语->英语 | 新闻领域 |
| 2 | EC | 英汉新闻领域机器翻译 | 英语->汉语 | 新闻领域 |
| 3 | MC | 蒙汉日常用语机器翻译 | 蒙语->汉语 | 日常用语 |
| 4 | TC | 藏汉政府文献机器翻译 | 藏语->汉语 | 政府文献 |
| 5 | UC | 维汉新闻领域机器翻译 | 维语->汉语 | 新闻领域 |

表2. CCMT 2019任务2：多语翻译评测项目表

| **序号** | **项目代号** | **评测项目名称** | **语种** | **领域** |
| --- | --- | --- | --- | --- |
| 1 | JE | 日英专利领域多语言机器翻译 | 日语->英语 | 专利领域 |

表3. CCMT 2019任务3：语音翻译评测项目

| **序号** | **项目代号** | **评测项目名称** | **语种** | **领域** |
| --- | --- | --- | --- | --- |
| 1 | CE-SpeechMT | 汉英语音机器翻译 | 汉->英 | 计算机、互联网等领域 |

表4. CCMT 2019任务4：翻译质量评估评测项目表

| **序号** | **项目代号** | **评测项目名称** | **语种** | **领域** |
| --- | --- | --- | --- | --- |
| 1 | CE-QEW | 汉英词汇翻译质量估计 | 汉->英 | 电商、互联网来源 |
| 2 | EC-QEW | 英汉词汇翻译质量估计 | 英->汉 | 电商、互联网来源 |
| 3 | CE-QES | 汉英句子翻译质量估计 | 汉->英 | 时政、经济、文化、科技等领域 |
| 4 | EC-QES | 英汉句子翻译质量估计 | 英->汉 | 时政、经济、文化、科技等领域 |

### 3. 双语翻译评测项目概况

#### 3. 1 任务介绍

　　与往年一致，翻译任务主要评测参评单位在双语翻译任务上的性能。评测翻译语言对包括英汉、汉英、蒙汉、藏汉、维汉等。

#### 3.2 评测指标

* **自动评测**：自动评测是指利用自动评价工具对参评单位提交的最终翻译结果文件进行评价。

本次评测中的自动评测采用多种自动评价标准，包括：BLEU-SBP、BLEU-NIST、TER、METEOR、NIST、GTM、mWER、mPER以及ICT。

评测组织方进行自动评价时将采用如下设置：

* + 所有自动评测将采用大小写敏感（case-sensitive）的方式，评测结果中也包含部分大小写不敏感的评价作为参考；
  + BLEU-SBP作为主要的自动评价指标；
  + 英汉、藏汉、维汉和蒙汉四个方向将采用基于字符（character-based）的评价方式；
  + 英汉、藏汉、维汉和蒙汉四个方向中，评测组织方将对GB2312编码的A3区字符进行全角到半角的转换；
  + 汉英项目则采用基于词（word-based）的评价方式。

#### 3.3 评测数据

　　本次评测由主办方提供全部训练、开发、测试集数据，数据格式情况参见附件2，数据详细情况请参见附件6。 关于数据使用的要求，请参见后续系统要求部分。

#### (1). 训练数据

　　　本次评测训练数据的情况请见下表（提供单位排名不分先后，以汉语拼音为序）。

为方便比较，本次评测为汉英语言对的部分数据提供了统一的数据预处理，预处理后的数据随评测数据一起发放，推荐参评单位使用。

表5 CCMT 2019双语翻译任务训练数据情况

|  |  |  |  |
| --- | --- | --- | --- |
| 评测项目名称 | 训练规模（句) | 提供单位 | 说明 |
| 汉英-英汉新闻领域机器翻译 | 9,023,471 | 点通数据有限公司、东北大学、中国科学院计算技术研究所、中国科学院自动化研究所 | 平行语料 |
| 5,281 | 中国科学院计算技术研究所 | 汉语和四个英语参考译文 |
| 8,665 | 中国科学院计算技术研究所 | 英语和四个汉语参考译文 |
| 4003 | 南京大学 | CWMT2017开发集、测试集 |
| 约1100万词 | 厦门大学 | 汉语单语语料 |
| 蒙汉日常用语机器翻译 | 262,643 | 内蒙古大学、中国科学院合肥智能机械研究所、中国科学院计算技术研究所 | 平行语料 |
| 藏汉政府文献机器翻译 | 157,959 | 青海师范大学、西北民族大学、西藏大学、厦门大学、中国科学院计算技术研究所 | 平行语料 |
| 维汉新闻领域机器翻译 | 170,061 | 中国科学院计算技术研究所、中国科学院新疆理化技术研究所 | 平行语料 |

其中，汉英与英汉评测项目与WMT2019联合组织，因此WMT 2019 提供的数据也可以作为本次评测对应的汉英和英汉项目使用。

#### (2). 开发数据

　　原则上，本次评测开发数据基本依照惯例，将CWMT2018的测试集作为开发数据。具体数据说明参见附件6，下表简要给出各项目测试数据的规模，并向各数据提供单位表示感谢。

表6 CCMT 2019双语翻译任务开发数据

| **评测项目名称** | **规模**  **（单位：句）** | **提供单位** | **说明** |
| --- | --- | --- | --- |
| 汉英-英汉新闻领域机器翻译 | 2481/1500 | WMT 2018测试集 | 单参考译文 |
| 蒙汉日常用语机器翻译 | 1000 | 内蒙古大学 | 单参考译文 |
| 藏汉政府文献机器翻译 | 1000 | 青海师范大学 | 单参考译文 |
| 维汉新闻领域机器翻译 | 1000 | 中国科学院新疆理化技术研究所 | 单参考译文 |

#### (3). 测试数据

　　本次评测测试数据的设计见下表，具体以测试数据发放时的情况为准。

表7 CCMT 2019双语翻译任务测试数据情况

| **评测项目名称** | **规模**  **（单位：句）** | **提供方** | **说明** |
| --- | --- | --- | --- |
| 汉英新闻领域机器翻译1 | 1000 | WMT 2019 | 单参考译文 |
| 英汉新闻领域机器翻译1 | 1000 | WMT 2019 | 单参考译文 |
| 汉英新闻领域机器翻译2 | 不少于1000 | 东北大学 | 单参考译文 |
| 英汉新闻领域机器翻译2 | 不少于1000 | 东北大学、 | 单参考译文 |
| 蒙汉日常用语机器翻译 | 1000 | 内蒙古大学 | 单参考译文 |
| 藏汉政府文献机器翻译 | 1000 | 青海师范大学 | 单参考译文 |
| 维汉新闻领域机器翻译 | 1000 | 中国科学院新疆理化技术研究所 | 单参考译文 |

#### 3.4 系统训练要求

　　对于每个评测项目，参评单位可以自由选择所采用的机器翻译技术（如：基于规则的机器翻译技术、基于实例的机器翻译技术、统计机器翻译技术及神经网络机器翻译技术等）。参评单位也可以使用系统融合技术，但在系统描述中必须做出明确说明，并在技术报告中说明进行系统融合的各个单系统的性能。此处，系统融合技术指使用两个及两个以上单系统的翻译结果进行字、词、短语、句子级别的重构或选择的技术。没有明确产生两个或两个以上单系统翻译结果的技术，如统计机器翻译中的协同解码、神经网络机器翻译的输出层ensemble、单个系统结果的重排序等，本次评测不认定为系统融合技术。评测组织方在发布评测结果时，将对使用系统融合技术的系统进行标注说明。

　　对于以基于平行数据进行训练的机器翻译技术（如统计机器翻译、神经网络机器翻译等）为主的参评系统，可以以“受限”和“非受限”两种方式参与评测。下面对两种方式进行详细说明：

* **受限训练**：受限训练是指只可以使用评测组织方指定范围的数据进行训练。具体说明如下：
  + 参评单位提交的“主系统”必须采用受限训练，以便于在尽可能一致的条件下对不同参评系统所采用的技术进行比较；
  + 对于以基于规则的机器翻译技术为主的参评系统，允许采用通过人工方式构造的翻译知识（如规则、模板、词典等），但要在系统描述和技术报告中对所使用的翻译知识的规模、构造和使用方式等给出清晰的说明。
  + 单语分析工具可以使用外部数据，如词法分析、句法分析及命名实体识别工具等可以使用外部数据；涉及双语翻译的工具不能使用外部数据，包括命名实体翻译、音字转换工具等（数词和时间词翻译不受此约束）；
  + 每个评测项目只允许使用评测组织方发布的该项目相关的训练数据，不可以使用其他评测项目的数据。即对于参加多个评测项目的单位，不同项目提供的数据不可以联合使用。
  + 与WMT联合组织的汉英、英汉领域评测项目的受限训练语料包括附件6列表中的CCMT方提供的数据；也包括由WMT组织提供的数据。为便于比较，请参评单位提交汉英、英汉领域系统的评测报告时说明是使用CCMT数据还是WMT数据还是两者皆有，评测组织方将在发布评测报告时对相应的系统结果予以标识。
* **非受限训练**：非受限训练是指可以使用评测组织方指定范围的数据之外的数据进行训练。具体说明如下：
  + 参评单位提交的 “对比系统” 可以采用非受限训练。
  + 采用非受限训练方式的系统，需要在系统描述和技术报告中对使用的数据进行说明（如数据规模和领域类型、是否为可公开获取的数据等。若为可公开获取的数据，则需说明数据出处）。
  + 欢迎参评单位使用自有的在线系统参与评测。在线系统一般认定为非受限系统，需要在系统描述和技术报告中对技术做简要说明。在线系统的结果仅作为参考，不参与非受限训练排名。

#### 3.5结果提交

　　参评单位收到测试数据后，应在规定时间内提交最终翻译结果文件。对于每个评测子项，参评单位可以提交一个主系统翻译结果（Primary Result）及最多三个对比系统的翻译结果（Contrast Result）。提交的每个结果文件都应包含详细的系统描述。具体数据格式请参见附件2。

　　汉英、英汉新闻领域机器翻译项目的参评单位可以选择向CCMT2019，或WMT19，或同时向WMT19和CCMT2019，提交测试数据的翻译结果。向WMT19提交的结果应满足WMT19对结果提交的要求，向CCMT2019提交的结果应满足CCMT2019对结果提交的要求。

### 4. 多语言翻译任务评测方法

#### 4.1 任务介绍

多语言翻译任务尝试发掘不同语言之间的对应和转换关系，2019年这一任务与CWMT2018基本相同，提供专利领域的英汉、日汉双语平行数据作为训练数据，任务目标为评价日英翻译方向的翻译性能。

#### 4.2 评测指标

　　请参见翻译任务评测方法中的英文为目标的评价方法。

#### 4.3 评测数据

本次评测的指定训练数据是由北京语智云帆科技有限公司提供日汉和英汉的训练数据（各300万句对），开发集数据日汉（6000句对）、英汉和日英（各3000句对），以及日英方向的测试数据（1000句）。

数据设定与2018年评测一致，具体请参见附件6中的相关信息。

（注：该任务的2018年测试集未正式发布，请不要在2019年的评测中使用）。

#### 4.4 系统要求

本任务在2019年只接受“受限”方式参与评测，即只可以使用4.3节中提供的数据进行训练。

#### 4.5 提交结果

　　请参见3.5双语翻译任务评测方法中的结果提交方法。

### 5. 语音翻译任务评测（新任务）

#### 5.1 任务介绍

语音翻译（同传场景）近年来成为学术界和工业界的热点。本任务面向会议、演讲场景的语音翻译（同声传译），将演讲语音翻译为文字，翻译方向为“中到英”。

#### 5.2 评测指标

自动翻译评价：采用客观的BLEU评价指标。为了最大程度模拟同传场景，测试集reference的构建，采用同传译员边听边译，再人工转成文字。在此基础上计算BLEU得分作为系统得分。

人工评价：采用人工译员进行翻译质量评估，译员将从流利度、忠实度两个层面进行人工评估。和传统的文本翻译评估略有不同，在同声传译场景中，允许忽略演讲者部分非关键信息，只需保证核心信息正确翻译。

翻译时延：MAT（Maximum Awaiting Token），计算翻译一篇演讲稿时，最长的等待Token数量。即开始翻译时，需要等待演讲者输出多少Token。

#### 5.3 评测数据

数据来自中文普通话学术报告和演讲，领域包括科技、人文、经济等，组织方提供不少于50小时的训练数据、1个小时的开发数据（包含原始语音、转写后中文、译文）。开发及测试集根据演讲内容、演讲者等分为不同的文件（doc）。

除了上述训练数据，本任务可使用CCMT2019中英新闻翻译任务的训练语料（不含WMT2019数据），但需要在提交结果中对数据使用情况予以说明。

*测试集规模将在确定后公布。*

评测训练集、开发集和测试集由百度公司提供，需单独签署数据授权协议（http://ai.baidu.com/broad/）；

训练数据发布时间预计是2019年3月底，请关注上面的网站和评测网站发布的最新信息。

#### 5.4 系统要求

组织方提供语音识别、自动断句、翻译的开源工具及benchmark（即将发布）。本任务要求翻译系统都在受限训练条件下训练（可使用CCMT2019中英新闻翻译任务的训练语料），鼓励参赛者使用额外的语音识别训练语料提高语音识别性能（需要在提交结果中备注使用了哪些额外语料）。参赛者可以选择实现自己的语音识别、自动断句、机器翻译等模块，组织方鼓励参赛者实现创新模型，例如端到端实时语音翻译模型。

#### 5.5 提交结果

本任务需要参赛者按照附件3的格式提交符合要求的结果文件。鼓励参赛者提供能够运行解码测试的模型文件和可执行文件（包括语音识别、自动断句、机器翻译等模块），并配有完整的使用说明。组办方将根据参赛者提供的模型和可执行文件输入实时语音信号，获取参赛系统的翻译结果及翻译时延。

### 6. 翻译质量评估任务评测方法

#### 6.1 任务介绍

无参考译文条件下的翻译质量估计，是当前机器翻译质量评估的难点。本次翻译质量评估包含两项任务：词汇级翻译质量估计（新增）和句子级翻译质量估计。评估对象为汉英和英汉机器翻译结果。

**(1). 词汇级（汉←→英 ）翻译质量估计（新任务）**

对于给定的待评估译文，在没有参考译文的情况下，以词为单位，评估其翻译质量。具体需要预测：译文词本身、或者译文词词与词之间的OK/BAD 二分类标注。

BAD表示某译文词是个错译或者多译；或者表示某两个译文词之间存在（一个或多个词）漏译。

具体地，对于句长为n（句中标点也算译文词）的译文来说，需要对2n + 1个位置进行OK/BAD二分类标注：n个译文词位置，以及n+1个译文词词间位置（其中句首词前和句尾词后各计有1个位置）。

**(2). 句子级（汉←→英 ）翻译质量估计**

目前译文的翻译质量通过专业译员进行译后编辑的代价来评估。评测方请专业译员对训练集、开发集和测试集中的译文进行译后编辑，并参照译后编辑的结果对待评估的译文计算HTER值（Human-targeted Translation Edit Rate, Snover M, Dorr B, Schwartz R, et al. A study of translation edit rate with targeted human annotation[J]. Machine Translation Workshop North Bethesda Md, 2006(1):223--231.），用以衡量该译文的翻译质量。翻译质量评估的任务即为预测不同译文的HTER值。

#### 6.2 评测指标

**(1). 词汇级（汉←→英 ）翻译质量估计**

该任务采用自动评价方法对参评系统进行评价。因为待标记位置数量确定，本任务采用标注准确性进行评价：将测试集合上所有译文的评估标注结果与对应的真实标注结果进行比较，分别计算“OK”和“BAD”的F1分数；这两个F1分数的乘积作为最终系统评估指标。

具体评测脚本即将发布。

**(2). 句子级（汉←→英 ）翻译质量估计**

该任务采用自动评价方法对参评系统进行评价。评价指标为在测试集合上所有译文的评估得分与相应给出的真实HTER得分之间的皮尔森相关系数（Pearson’s correlation coefficient）。其他参考指标还包括Mean Absolute Error（MAE）和Root Mean Squared Error（RMSE）)等。

#### 6.3 评测数据

**(1). 词汇级（汉←→英 ）翻译质量估计**

本项目的数据包括：阿里巴巴（中国）有限公司提供的质量评估评测任务训练集和开发集，具体说明如下

• 数据来源：源语言句子部分来自阿里巴巴电商和IT场景，另一部分来自互联网经济、政治、科技、口语场景。总计6个场景。

• 生成过程：首先，每个原文的译文来自3个在线翻译引擎（通过在线翻译服务获得）和1个神经网络机器翻译引擎（由阿里巴巴提供），译文结果进行去重。然后，所有机器翻译译文由专业译员进行人工后编辑得到正确译文。采用TER工具 http://www.cs.umd.edu/~snover/tercom/ 计算译文和人工后编辑结果的对齐关系（采用如下设置：分词（tokenized），大小写不敏感（case insensitive），精准匹配 （exactly matching only），使用“-d 0 ” 选项禁用shift（disabling shifts by using the ‘-d 0’ option）），进而得到译文词和词间的OK/BAD标注。

• 数据规模：1）训练数据中，汉英和英汉方向分别包含4606和4105句原文，分别对应11039和10878句译文、编辑结果以及词级别标注结果；2）开发数据中，汉英和英汉方向分别各包含375句原文，分别对应1050和1129句译文、编辑结果以及词级别标注结果。

*测试集规模将在确定后公布。*

阿里巴巴提供的数据授权协议需另行签署（附件1-1），和评测注册表同时提交。

**(2). 句子级（汉←→英 ）翻译质量估计**

　　数据来源：所有的源语言句子来源于语智云帆试译宝上的翻译训练题，覆盖时政、经济、文化、科技等领域。待评估译文选自试译宝上的人工答题译文和若干个机器翻译引擎的结果。

　　生成过程：人工译文来源于试译宝上的人工答题译文；机器译文来源于5个在线翻译引擎（通过在线翻译服务获得）和1个统计机器翻译引擎（由小牛翻译提供，2019年新增的数据不再提供SMT结果）。根据试译宝打分标准，从人工译文和机器译文分别筛选出不同等级的N个译文（N<=9）；并筛选结果合并得到最终的待评估译文集合。待评估译文由专业译员进行后编辑得到编辑结果。

最终数据：训练数据包含英汉方向3043句原文，14789句译文及其对应的编辑结果；汉英方向2503句原文，10070句译文及其对应的编辑结果。 开发数据包含英汉方向300句原文，1381句译文及其对应的编辑结果；汉英方向300句原文，1143句译文及其对应的编辑结果。

（注：本项目数据集相对于2018年已经整体更新，融入了2018年的测试集数据）

*测试集规模将在确定后公布。*

#### 6.4 系统要求

　　参评系统可以采用各种方法进行学习和训练，如需要平行语料或单语语料，可使用CCMT2019汉英、英汉项目对应的数据。所有数据资源的使用应该在系统说明和评测报告中予以说明。

#### 6.5. 提交结果

　　参评单位收到测试数据后，对于每个评测子项，应在规定时间内提交最终翻译质量评估结果文件和系统描述信息文件。具体数据格式请参见附件4。

### 六. 提交技术报告及参加评测会议

　　评测结束后，每个评测项目的参评单位应向CCMT 2019会议提交一份详细的技术报告，说明系统的架构、原理，使用的主要技术，以及数据使用的情况。评测报告将与CCMT2019投稿论文采用相同的匿名同行评审机制，并择优推荐与CCMT2019论文共同发表。参评单位应派至少一人参加CCMT 2019会议进行相应技术交流。技术报告相关要求请参见[附件5](http://nlp.nju.edu.cn/cwmt2018/report.html)。

### 七. 评测日程

| **日期** | **评测环节** | |
| --- | --- | --- |
| 2019.03.15 | 发布评测大纲，评测报名开始。评测组织方向报名单位提供训练集、开发集数据，以及BLEU-SBP打分程序、格式检查程序（通过ftp方式发放） | |
| 2019.04.15 | 报名截止，停止发放训练集、开发集数据 | |
| 2019.04.20 | CCMT评测平台上线（估计时间） | |
| 2019.06.1 10:00am | **评测组织方**发放多语言翻译任务的日英项目的测试数据；  **评测组织方**发放翻译质量评估任务的汉英、英汉机器翻译四个项目的测试数据。 | |
| 2019.06.7 10:00am | **参评单位**提交多语言翻译任务的日英项目的翻译结果；  **参评单位**提交翻译质量评估任务的汉英、英汉机器翻译四个项目的翻译结果。 | |
| 2019.06.8 10:00am | **评测组织方**发放翻译任务的汉英、英汉新闻领域机器翻译两个项目的测试数据；  **评测组织方**发放维汉新闻领域、蒙汉日常用语、藏汉政府文献机器翻译三个项目的测试数据； | |
| 2019.06.13 10:00am | **评测组织方**发放语音翻译评测项目的测试数据 | |
| 2019.06.15 10:00am | **参评单位**提交翻译任务的汉英、英汉新闻领域机器翻译两个项目的翻译结果；  **参评单位**提交维汉新闻领域、蒙汉日常用语、藏汉政府文献机器翻译三个项目的翻译结果； | |
| 2019.06.20 10:00am | **参评单位**提交语音翻译评测项目的测试数据 | |
| 2019.07.01 | 参评单位提交评测技术报告（参照CCMT学术论文投稿要求和方法） | |
| 2019.07.20 | | 评测组织方向参评单位通知初步评测结果 |
| 2019.08.15 | 评测组织方返回评审结果 | |
| 2019.08.31 | 评测技术报告终稿提交 | |
| 2019.9.27-9.29 | 会议召开，会上正式报告评测结果并进行研讨 | |

### 八. 附件

附件1：报名表及评测协议

附件1-1：CCMT2019词汇级翻译质量估计评测数据协议  
附件2：机器翻译任务数据  
附件3：语音翻译任务数据格式  
附件4：翻译质量评估任务数据格式  
附件5：技术报告要求

附件6：评测数据一览