附件 5：评测组织方发布的资源列表

如非特殊说明，评测提供的数据文件默认采用 UTF-8 无 BOM 编码

1 汉英/英汉新闻相关资源

1.1 训练数据

资源名称简写及 ChineseLDC 资源编号	资源描述	
Datum2015	名称	点通汉英平行语料库（2015）（部分）
	提供单位	点通数据有限公司
	语种	汉语—英语
	领域	综合领域，包括：语言教材、双语图书、技术文档、双语新闻、政府白皮书、政府公文和 Web 上双语资源等等
	规模	1000004 个句对
	说明	这是点通数据有限公司在 863 项目支持下开发的《双语/多语平行语料库》的部分内容。
CASICT2011 （CLDC-2010-001） （CLDC-2012-001）	名称	计算所 Web 汉英平行语料库（2013）
	提供单位	中国科学院计算技术研究所
	语种	汉语—英语
	领域	综合领域
	规模	1936633 个句对
	说明	该平行语料库是从互联网上自动挖掘获得的。双语平行网页的发现、确认，双语平行文本的获取，句子对齐等过程完全通过程序自动实现。语料库抽样评价的正确率在 95%以上。 该研究得到国家自然科学基金项目（编号：60603095）的支持。
CASICT2015	名称	计算所 Web 汉英平行语料库（2015）
	提供单位	中国科学院计算技术研究所
	语种	汉语—英语
	领域	综合领域
	规模	2036834 个句对
	说明	该平行语料库是从互联网上自动挖掘获得的。双语平行网页的发现、确认，双语平行文本的获取，句子对齐等过程完全通过程序自动实现。计算所在此基础上进行了大致的校对，语料库抽样评价的正确率在 99%以上。语料构成如下：网络语料占 60%，电影字幕语料占 20%，来自英汉辞书的例句语料占 20%。

CASIA2015	名称	中科院自动化所 Web 汉英平行语料库（2015）
	提供单位	中国科学院自动化研究所
	语种	汉语—英语
	领域	综合领域
	规模	1050000 句对
	说明	该平行语料库是从互联网上自动挖掘获得的。双语平行网页的发现、确认，双语平行文本的获取，句子对齐等过程完全通过程序自动实现。
Datum2017	名称	点通公司英汉平行语料库（2017）
	提供单位	点通数据有限公司
	语种	汉语—英语
	领域	
	规模	100 万句对，分为 20 个文件
	说明	
NEU2017	名称	东北大学英汉平行语料库（2017）
	提供单位	东北大学 自然语言处理实验室
	语种	汉语—英语
	领域	
	规模	200 万句对
	说明	
SSMT2007 MT Evaluation Data （2007-863-001）	名称	SSMT2007 机器翻译评测数据（英汉/汉英机器翻译部分）
	提供单位	中国科学院计算技术研究所
	语种	汉语—英语
	领域	新闻
	规模	该机器翻译测试语料包含 2 个翻译方向（汉英、英汉），语料为新闻领域。其中汉英机器翻译测试语料含 1,002 个汉语句子。英汉机器翻译测试语料含 995 个英语句子。每个测试句子包括 4 个人工翻译的参考译文。
	说明	
HTRDP(863)2005 MT Evaluation Data （2005-863-001）	名称	2005 年 863 机器翻译评测数据（英汉/汉英机器翻译部分）
	提供单位	中国科学院计算技术研究所
	语种	汉语—英语
	领域	包括两种评测语料，一种是对话语料，领域为奥运相关领域，包括体育赛事、天气预报、交通住宿、旅游餐饮等；一种是篇章语料，领域为新闻领域。
	规模	汉英对话句对：467 句，汉英篇章句对：489 句。

		英汉对话句对：459 句，英汉篇章句对：494 句。 每个翻译方向的每个测试句子各提供 4 个人工翻译的参考译文。
	说明	
HTRDP(863)2004 MT Evaluation Data （2004-863-001）	名称	2004 年 863 机器翻译评测数据（英汉/汉英机器翻译部分）
	提供单位	中国科学院计算技术研究所
	语种	汉语—英语
	领域	两种评测语料，一种是篇章语料，一种是对话语料。领域是通用领域和奥运的相关领域，其中奥运领域包括体育赛事、天气预报、交通住宿、旅游餐饮等。
	规模	汉英评测数据含 400 句对话语料，308 句篇章语料。英汉评测数据含 400 句对话语料，310 句篇章语料。每个翻译方向的每个测试句子各提供 4 个人工翻译的参考译文。
	说明	2004 年 863 机器翻译评测汉英、英汉部分测试数据。
HTRDP(863)2003 MT Evaluation Data （2003-863-004）	名称	2003 年 863 机器翻译评测数据（英汉/汉英机器翻译部分）
	提供单位	中国科学院计算技术研究所
	语种	汉语—英语
	领域	奥运相关领域，其中奥运领域包括体育赛事、天气预报、交通住宿、旅游餐饮等。
	规模	汉英评测数据含 437 句对话语料和 169 句篇章语料；英汉评测数据含 496 句对话语料和 322 句篇章语料。每个翻译方向的每个测试句子各提供 4 个人工翻译的参考译文。
	说明	2003 年 863 机器翻译评测汉英、英汉部分测试数据。
CWMT2008 Machine Translation Evaluation Data （CLDC-2009-001） （CLDC-2009-002）	名称	CWMT2008 机器翻译评测新闻语料（英汉/汉英机器翻译部分）
	提供单位	中国科学院计算技术研究所
	语种	汉语—英语
	领域	新闻
	规模	汉英评测数据含 1006 句对；英汉评测数据含 1000 句对。每个翻译方向的每个测试句子各提供 4 个人工翻译的参考译文。
	说明	
CWMT2009 Machine Translation Evaluation Data	名称	CWMT2009 机器翻译评测数据（英汉/汉英机器翻译部分）
	提供单位	中国科学院计算技术研究所
	语种	汉语—英语
	领域	新闻

	规模	汉英评测数据含 1003 句对；英汉评测数据含 1002 句对。每个翻译方向的每个测试句子各提供 4 个人工翻译的参考译文。
	说明	
CWMT2011 Machine Translation Evaluation Data	名称	CWMT2011 机器翻译评测数据（英汉机器翻译部分）
	提供单位	中国科学院计算技术研究所
	语种	英语—汉语
	领域	新闻
	规模	英汉评测数据含 3187 句对。每个测试句子各提供 4 个人工翻译的参考译文。
	说明	
NJU-newsdev2017-enzh （NJU-newsdev2017-zhen）	名称	南京大学 CWMT2017 汉英/英汉新闻语料开发集数据
	提供单位	南京大学
	语种	汉语—英语
	领域	新闻
	规模	共 2,002 句对
	说明	包含 1000 个汉语新闻句子及其英语翻译结果，以及 1002 个英语新闻句子及其汉语翻译结果。
NJU-newstest2017-enzh （NJU-newstest2017-zhen）	名称	南京大学 CWMT2017 汉英/英汉新闻语料测试集
	提供单位	南京大学
	语种	汉语—英语
	领域	新闻
	规模	共 2,001 句对
	说明	包含 1000 个汉语新闻句子及其英语翻译结果，以及 1001 个英语新闻句子及其汉语翻译结果。 该数据是 CWMT2018 的开发集

1.2 单语新闻数据

XMU-CWMT2017	名称	厦门大学 NLP 实验室新华网新闻汉语单语语料（2017）
	提供单位	厦门大学
	语种	汉语
	领域	新闻
	规模	现语料库共有 662,904 个文章，大约 1100 万词汇。
	说明	本资源由厦门大学 NLP 实验室收集，包括新华网 2011 年不同主题频道的新闻语料，例如：国内新闻，国际新闻，财经新闻，论坛，教育等。

		每篇文章包含：标题，日期，URL 和内容。

1.3　开发集数据

WMT 2019 的测试集*	名称	WMT 2019 汉英/英汉测试集数据
	提供单位	WMT 2019 发布（http://data.statmt.org/wmt19/translation-task/test.tgz）*
	语种	汉语—英语
	领域	新闻
	规模	共 3,981 句对
	说明	包含 2481 个汉语句子及其英语翻译结果，以及 1500 个英语句子及其汉语翻译结果；

* CCMT2020 继续与 WMT2020 合作，开发集设定也是使用同一个数据。

汉英与英汉评测项目的数据与 WMT2020 共享。因此 WMT 2020 提供的汉英双语数据也可以作为本次评测对应的汉英和英汉项目使用。

2 蒙汉日常用语项目数据

2.1 训练数据

IMU-CWMT2013 （CLDC-2010-005）	名称	内蒙古大学汉蒙平行语料库（2013）
	提供单位	内蒙古大学
	语种	汉语—蒙古语
	领域	政府文献和法律法规、日常对话、文学
	规模	共 104,975 句对 其中：CWMT2011 评测训练语料 67274 句对,领域包括: 日常对话、文学、政府文献和法律法规; CWMT 2013 新增训练语料:包括新闻语料 17,516 句对,政府文献语料 10,394 句对,课本语料 5,052 句对,蒙汉字典语料 4,739 句对;
	说明	
IMU-CWMT2015	名称	内蒙古大学汉蒙平行语料库（2015）
	提供单位	内蒙古大学
	语种	汉语—蒙古语
	领域	政府文献和法律法规、日常对话、文学
	规模	共 24,978 句对
	说明	
IIM-CWMT2015	名称	中国科学院合肥智能机械研究所蒙汉双语语料库（2015）
	提供单位	中国科学院合肥智能机械研究所
	语种	蒙古语—汉语
	领域	新闻
	规模	1,682 句对
	说明	
ICT-MC-corpus-CWMT2017	名称	中国科学院计算技术研究所蒙汉双语语料库（2017）
	提供单位	中国科学院计算技术研究所
	语种	蒙古语—汉语
	领域	新闻
	规模	30,007 句对
	说明	
IMU-corpus-CWMT2017	名称	内蒙古大学蒙汉双语语料库（2017）
	提供单位	内蒙古大学
	语种	蒙古语—汉语

	领域	综合，包括：政府文件，政府工作报告，国务院文件，法律法规等
	规模	100,001 句对
	说明	
IMU-dev-mnzh-CWMT2017	名称	内蒙古大学 CWMT2017 蒙汉开发集数据
	提供单位	内蒙古大学
	语种	蒙古语—汉语
	领域	政府文献和法律法规、日常对话、文学
	规模	共 1,000 句蒙古语，每句 4 个汉语参考译文
	说明	CMWT2017 蒙汉开发集与 CWMT2011、CWMT2013、CWMT2015 蒙汉开发集相同
IMU-test-mnzh-CWMT2017	名称	内蒙古大学 CWMT2017 蒙汉测试集数据
	提供单位	内蒙古大学
	语种	蒙古语—汉语
	领域	政府文献和法律法规、日常对话、文学
	规模	共 1,001 句蒙古语，每句 4 个汉语参考译文
	说明	CWMT2017 蒙汉测试集即 CMWT2018 蒙汉开发集数据

2.2　开发集数据

IMU-test-mnzh - CWMT2018	名称	内蒙古大学 CWMT2018 蒙汉测试集数据
	提供单位	内蒙古大学
	语种	蒙古语—汉语
	领域	政府文献和法律法规、日常对话、文学
	规模	共 1,000 句蒙古语，每句 1 个汉语参考译文
	说明	CWMT2018 蒙汉测试集被指定为 CCMT2019 蒙汉开发集

3 藏汉政府文献相关资源

3.1 训练数据

QHNU-CWMT2013	名称	青海师范大学藏汉平行语料库（2013）
	提供单位	青海师范大学
	语种	藏语—汉语
	领域	政府文献领域
	规模	33,145 句对
	说明	该平行语料库是通过录入、扫描、网页下载等方式获得的。双语平行的搜集、整理、确认、获取、句子对齐等过程是通过程序自动实现和人工干预实现的。语料库的正确率在 99%以上。 该研究得到国家自然科学基金项目（编号：61063033）和 973 前期研究专项（编号：2010CB334708）的支持。
QHNU-CWMT2015	名称	青海师范大学藏汉平行语料库（2015）
	提供单位	青海师范大学
	语种	藏语—汉语
	领域	政府文献领域
	规模	17,194 句对
	说明	该平行语料库是通过录入、扫描、网页下载等方式获得的。双语平行的搜集、整理、确认、获取、句子对齐等过程是通过程序自动实现和人工干预实现的。语料库的正确率在 99%以上。 该研究得到国家自然科学基金项目（编号：61063033）的支持。
XBMU-XMU	名称	央金藏汉平行语料库
	提供单位	厦门大学人工智能研究所 西北民族大学语言（技术）研究所
	语种	汉语—藏语
	领域	综合领域
	规模	52,078 句对
	说明	1)该藏汉平行语料库是用正式出版物、藏汉大词典和网络语料藏汉对照文本,经使用自主开发的“藏汉句子对齐工具”初步对齐之后,由人工逐句对齐。 2)5 万句对藏汉平行语料的对齐正确率为 100%。 3)该研究得到国家社科基金重点项目《藏语语料库建设研究》（批准号：05AYY001）和 863 重点项目《面向跨语言

		搜索的机器翻译关键技术研究》（批准号：2006AA010107）的支持。
XBMU-XMU-UTibet	名称	西北民族大学、西藏大学与厦门大学藏汉语料（2012）
	提供单位	西北民族大学语言（技术）研究所 西藏大学 厦门大学人工智能研究所
	语种	汉语—藏语
	领域	政论，法律
	规模	24,159 句对
	说明	语料来源：2008 年和 2009 年全国最新法律文件和十八大报告、2011、2012 年政府工作报告等，政论类与法律类语料各占一半。系西北民族大学、西藏大学与厦门大学于 2012 年通过对原材料进行扫描、识别、校对并独立加工完成。
ICT-TC-corpus-CWMT2017	名称	中国科学院计算技术研究所藏汉双语语料库（2017）
	提供单位	中国科学院计算技术研究所
	语种	藏语—汉语
	领域	新闻
	规模	30,004 句对
	说明	
QHNU-dev-tizh-CWMT2017	名称	青海师范大学 CWMT2017 藏汉开发集数据
	提供单位	青海师范大学
	语种	藏语—汉语
	领域	政府文献
	规模	共 650 句藏语，每句 4 个汉语参考译文
	说明	CMWT2017 藏汉开发集与 CWMT2011、CWMT2013、CWMT2015 藏汉开发集相同
QHNU-test-tizh-CWMT2017	名称	青海师范大学 CWMT2017 藏汉测试集数据
	提供单位	青海师范大学
	语种	藏语—汉语
	领域	政府文献
	规模	共 729 句藏语，每句 4 个汉语参考译文
	说明	CWMT2017 藏汉测试集即 CMWT2018 藏汉开发集数据

3.2　开发集数据

	名称	青海师范大学 CWMT2018 藏汉测试集数据

QHNU-test-tizh-CWMT2018	提供单位	青海师范大学
	语种	藏语—汉语
	领域	政府文献
	规模	共 1000 句藏语，每句 1 个汉语参考译文
	说明	CCMT2019 藏汉开发集使用 CWMT2018 的藏汉测试集

4 维汉新闻相关资源

4.1 训练数据

XJU-CWMT2013 XJU-corpus-CWMT2017 XJU-dev-uyzh-CWMT2017 XJU-dev-uyzh-CWMT2018		*这几份数据 2019 年暂未进入训练语料；* (注 CMWT2017 维汉开发集与 CWMT2011、CWMT2013、CWMT2015 维汉开发集相同)
XJIPC-CWMT2015	名称	中国科学院新疆理化技术研究所维汉双语语料库（2015）
	提供单位	中国科学院新疆理化技术研究所
	语种	汉语—维吾尔语
	领域	新闻
	规模	59,990 句对
	说明	此语料在 CWMT 2013 年的基础上新增加约 3 万句对 语料比例：2007 年-2014 年媒体新闻类语料比例约占 95%，2012-2014 年政府报告类语料和法律法规类语料合计比例约占 5%。 语料来源：系中国科学院新疆理化技术研究所在 2012 年-2014 年间采集、标注、校对完成。
ICT-UC-corpus-CWMT2017	名称	中国科学院计算技术研究所维汉双语语料库（2017）
	提供单位	中国科学院计算技术研究所
	语种	维吾尔语—汉语
	领域	新闻
	规模	30,071 句对
	说明	
XJIPC-corpus-CWMT2017	名称	中国科学院新疆理化技术研究所维汉双语语料库（2017）
	提供单位	中国科学院新疆理化技术研究所
	语种	维吾尔语—汉语
	领域	新闻
	规模	30,000 句对
	说明	
XJIPC-corpus-CWMT2018	名称	中国科学院新疆理化技术研究所维汉双语语料库（2018）
	提供单位	中国科学院新疆理化技术研究所
	语种	维吾尔语—汉语

	领域	新闻
	规模	50,000 句对
	说明	

4.2 开发集数据

XJIPC-test-uyzh-CWMT2018	名称	中国科学院新疆理化技术研究所 CWMT2018 维汉测试集数据
	提供单位	中国科学院新疆理化技术研究所
	语种	维吾尔语—汉语
	领域	新闻
	规模	共 1000 句维语，每句 1 个汉语参考译文
	说明	CCMT2019 维汉开发集数据为 CWMT2018 维汉测试集数据

5 多语言机器翻译（日英专利领域）相关资源

5.1 训练数据

Lingosail-train-zhjp-CWMT2018	名称	北京语智云帆科技有限公司日汉专利平行语料库（2018）
	提供单位	北京语智云帆科技有限公司
	语种	日语—汉语
	领域	综合
	规模	3,000,000 句对
	说明	该数据更新了 2017 版的日汉专利平行数据
Lingosail-train-enzh-CWMT2018	名称	北京语智云帆科技有限公司英汉专利平行语料库（2018）
	提供单位	北京语智云帆科技有限公司
	语种	英语—汉语
	领域	综合
	规模	3,000,000 句对
	说明	

5.2 开发集数据

Lingosail-dev-jpzh-CWMT2017	名称	北京语智云帆科技有限公司日汉双语开发集数据（2017）
	提供单位	北京语智云帆科技有限公司
	语种	日语—汉语
	领域	综合
	规模	3000 句日语，每句含一个汉语参考译文
	说明	CWMT2017 日汉专利领域翻译开发集数据
Lingosail-dev-enzh-	名称	北京语智云帆科技有限公司英汉双语开发集数据（2018）

CWMT2018	提供单位	北京语智云帆科技有限公司
	语种	英语—汉语
	领域	综合
	规模	3000 句英语，每句含一个汉语参考译文
	说明	
Lingosail-dev-jpzh-CWMT2018	名称	北京语智云帆科技有限公司日汉双语开发集数据（2018）
	提供单位	北京语智云帆科技有限公司
	语种	日语—汉语
	领域	综合
	规模	3000 句日语，每句含一个汉语参考译文
	说明	
Lingosail-dev-enjp-CWMT2018	名称	北京语智云帆科技有限公司英日双语开发集数据（2018）
	提供单位	北京语智云帆科技有限公司
	语种	英语—日语
	领域	综合
	规模	3000 句英语，每句含一个日语参考译文
	说明	

5.3 汉语专利数据

Lingosail-cn_for_lm-CWMT2017	名称	北京语智云帆科技有限公司汉语专利语料（2017）
	提供单位	北京语智云帆科技有限公司
	语种	汉语
	领域	综合
	规模	7,114,700 句对
	说明	CWMT2017 日汉专利领域翻译汉语单语数据

6 语音机器翻译评测相关资源

2020 年该项评测数据有所扩大，具体评测设置（包括评测时间）与 The 1st Workshop on Automatic Simultaneous Translation（AutoSimTrans 2020 @ ACL 2020，https://autosimtrans.github.io/shared）一致,注册链接 https://aistudio.baidu.com/aistudio/competition/detail/18?lang=en）。

6.1 训练数据

AutoSimulTrans-zhen-train	名称	同传中英语料库-训练集
	提供单位	百度
	语种	汉-英
	领域	综合

	规模	约 60 小时
	说明	包含中文语音、中文文本、中文识别结果、英文文本翻译
AutoSimulTrans-enes-train	名称	同传英西语料库-训练集
	提供单位	百度
	语种	英-西
	领域	新闻
	规模	约 50 小时
	说明	包含英文文本、西班牙语文本

6.2 开发集数据

AutoSimulTrans-zhen-dev	名称	同传中英语料库-开发集
	提供单位	百度
	语种	汉-英
	领域	综合
	规模	3 小时
	说明	包含中文语音、中文文本、中文识别结果、英文文本
AutoSimulTrans-enes-dev	名称	同传英西语料库-开发集
	提供单位	百度
	语种	英-西
	领域	新闻
	规模	3 小时
	说明	包含英文文本、西语文本

7 汉英/英汉多领域机器翻译质量评估相关资源

7.1 词汇级翻译质量估计

7.1.1 训练数据

Alibaba-train-enzh-qe-word-level-CCMT2019	名称	阿里巴巴英汉多领域机器翻译质量评估语料库（2019）
	提供单位	阿里巴巴（中国）有限公司
	语种	英语—汉语
	领域	经济，政治，科技，口语，电商，IT 等领域
	规模	10878 句对
	说明	训练集由四个文件组成，包括由逐行对应的源语言文件 train.source、译文文件 train.target、对译文进行人工译后编辑的文件 train.pe，每行为

		一个句子；以及待评估译文的 Tags 文件 train.tags。
Alibaba-train-zhen-qe-word-level-CCMT2019	名称	阿里巴巴汉英多领域机器翻译质量评估语料库（2019）
	提供单位	阿里巴巴（中国）有限公司
	语种	汉语—英语
	领域	经济，政治，科技，口语，电商，IT 等领域
	规模	11039 句对
	说明	训练集由四个文件组成，包括由逐行对应的源语言文件 train.source、译文文件 train.target、对译文进行人工译后编辑的文件 train.pe，每行为一个句子；以及待评估译文的 Tags 文件 train.tags。

7.1.2 开发集数据

Alibaba-dev-enzh-qe-word-level-CCMT2019	名称	阿里巴巴英汉多领域机器翻译质量评估开发集数据（2019）
	提供单位	阿里巴巴（中国）有限公司
	语种	英语—汉语
	领域	经济，政治，科技，口语，电商，IT 等领域
	规模	1129 句对
	说明	开发集由四个文件组成，包括由逐行对应的源语言文件 dev.source、译文文件 dev.target、对译文进行人工译后编辑的文件 dev.pe，每行为一个句子；以及待评估译文的 Tags 文件 dev.tags。
Alibaba-dev-zhen-qe-word-level-CCMT2019	名称	阿里巴巴汉英多领域机器翻译质量评估开发集数据（2019）
	提供单位	阿里巴巴（中国）有限公司
	语种	汉语—英语
	领域	经济，政治，科技，口语，电商，IT 等领域
	规模	1050 句对
	说明	开发集由四个文件组成，包括由逐行对应的源语言文件 dev.source、译文文件 dev.target、对译文进行人工译后编辑的文件 dev.pe，每行为一个句子；以及待评估译文的 Tags 文件 dev.tags。

7.2 句子级翻译质量估计

7.2.1 训练数据

Lingosail-train-enzh-qe-CCMT2019	名称	北京语智云帆科技有限公司英汉多领域机器翻译质量评估语料库（2019）
	提供单位	北京语智云帆科技有限公司
	语种	英语—汉语
	领域	时政、经济、文化、科技等领域
	规模	14789 句对
	说明	子训练集由四个文件组成，包括由逐行对应的源语言文件 train. source、译文文件 train. target、对译文进行人工译后编辑的文件 train.pe，每行为一个句子；以及待评估译文的 HTER 值文件 train.hter，每行为一个区间[0,1]的数值。
Lingosail-train-zhen-qe-CCMT2019	名称	北京语智云帆科技有限公司汉英多领域机器翻译质量评估语料库（2019）
	提供单位	北京语智云帆科技有限公司
	语种	汉语—英语
	领域	时政、经济、文化、科技等领域
	规模	10070 句对
	说明	子训练集由四个文件组成，包括由逐行对应的源语言文件 train.source、译文文件 train.target、对译文进行人工译后编辑的文件 train.pe，每行为一个句子；以及待评估译文的 HTER 值文件 train.hter，每行为一个区间[0,1]的数值。

7.2.2 开发集数据

Lingosail-dev-enzh-qe-CCMT2019	名称	北京语智云帆科技有限公司英汉多领域机器翻译质量评估开发集数据（2019）
	提供单位	北京语智云帆科技有限公司
	语种	英语—汉语
	领域	时政、经济、文化、科技等领域
	规模	300 句英语，每句含多个汉语参考译文，共计 1381 句译文
	说明	开发集由四个文件组成，包括由逐行对应的源语言文件 dev. source、译文文件 dev. target、对译文进行人工译后编辑的文件 dev.pe，每行为一个句子；以及待评估译文的 HTER 值文件 dev.hter，每行为一个区间[0,1]的数值。

Lingosail-dev-zhen-qe-CCMT2019	名称	北京语智云帆科技有限公司汉英多领域机器翻译质量评估开发集数据（2019）
	提供单位	北京语智云帆科技有限公司
	语种	汉语—英语
	领域	时政、经济、文化、科技等领域
	规模	300 句汉语，每句含多个英语参考译文，共计 1143
	说明	开发集由四个文件组成，包括由逐行对应的源语言文件 dev. source、译文文件 dev. target、对译文进行人工译后编辑的文件 dev.pe，每行为一个句子；以及待评估译文的 HTER 值文件 dev.hter，每行为一个区间[0,1]的数值。