

来华留学生课程作文语料标注规范

一. 作文语料的基本结构:

文本型作文语料的基本结构如下:

[作文]	
[作文 ID] [/作文 ID]	//可以空着不填
[姓名] [/姓名]	//作文作者名称
[学号] [/学号]	//作文作者学号, 如不清楚, 可空着不填
[国籍] [/国籍]	//如“韩国”“泰国”“印尼”
[要求]	
[命题]	
[类型] [/类型]	//“自拟、命题、无题、自选”; 如不明确, 可以留空
[标题] [/标题]	//原文标题, 这里不要对标题中的偏误进行标注
[/命题]	
[体裁] [/体裁]	//“记叙文、说明文、事理文、议论文、调查报告、读后感”等
[主题] [/主题]	//对作文主题的一句话简要说明。
[时长] [/时长]	//该篇作文要求的写作时间长度
[字数] [/字数]	//如“400~600”, 或“不少于x字”或“x字左右”或“不超过X字”或“无要求”, 如不清楚, 可不填
[辅助] [/辅助]	//是否借助词典或手机等工具等完成
[具体要求] [/具体要求]	//填写作文要求的全部内容(包括阅读材料)
[/要求]	
[指导]	
[有无指导] [/有无指导]	//“有指导”或“无指导”
[指导过程] [/指导过程]	//事前指导/事后指导(修改、讲评、重写), 注: 部分考试作文有事前指导
[指导内容] [/指导内容]	//指导内容包括: 阅读范文; 分析体裁特点; 说明文章结构; 训练相关语言表达; 训练观点提炼; 训练论证方法; 检查写提纲等
[/指导]	
[课型] [/课型]	//如“商务汉语写作”“经贸汉语阅读与写作”
[来源] [/来源]	//课堂作文、课下作文、期中考试作文、期末考试作文、平时测验等
[教材] [/教材]	//如实填写教材名称或填自编
[学期] [/学期]	//如“2015-2016-1”“2015-2016-2”
[班级] [/班级]	
[源文] [/源文]	//原文的存放路径, 可以空着。
[日期] [/日期]	//不清楚可以空着不填
[文稿类型] [/文稿类型]	//“手写版”or“电脑版”
[版本] [/版本]	//一稿 or 二稿 or 三稿。

[正文]

[题目] [/题目]

//如题目中有字词等层面偏误，也要标注

.....

//标注正文中的字、词、短语层面的各种偏误

[/正文]

[批语] [/批语]

[得分] [/得分]

//如实填写，如没有该作文得分情况，不用填写

[原文] [/原文]

//原文内容，拷贝或录入到此处。

[提供] [/提供]

//作文提供者

[教师] [/教师]

//该门课任课教师名称

[录入] [/录入]

//录入与标注者名称，如“2022 研. 李嘉奇”

[录入日期] [/录入日期]

//录入与标注日期，如“2023. 05. 20”

[一校] [/一校]

//第一遍语料校对入名称，如“2022 研. 陈康倩”

[一校日期] [/一校日期]

[二校] [/二校]

//第二遍语料校对入名称

[二校日期] [/二校日期]

[三校] [/三校]

[/作文]

注：

1. 以上内容中，如没有相应内容，则空着不填。如：没有批语，则填为：

[批语] [/批语]；不清楚字数限制，则填为“[字数] [/字数]”

2. 以上内容中“//”后为注释内容，

仅供理解标注规范时参考。

3. 正文中的标点符号应为中文标点符号，如“。，？！”等，而不是英文的标点符号，如“., ?!”等。

4. 标点符号的标注暂未作要求

二. 正文字词语的标注：

在正文中标注字、词、熟语层面的偏误或精彩表达，分类如下：

1. 表达不妥或不规范的标注

标注 代码	代码 名称	代码 含义	示例	备注
[BD]	表达偏误	语句表达不规范或语法结构有问题	为什么写作是头疼[Cd-吗]的 [Cq-的][BD-是.....的] 如果我们坚持[Cq-坚持][BD-有一意都]肯定会更顺利。	不是字词或短语层面的问题，是语法层面错误的标注。如“比”字句、“是.....的”偏误。

注 1：“[BD-***]”标在字词偏误标注的后边。

注 2：“[BD-***]”标注中尽量不要出现句子的标点符号，如句号、问号、感叹号等。

注 3：尽量寻找偏误原因，减少 BD 类偏误的标注数量

2. 汉字偏误标注：

标注 代码	代码 名称	代码 含义	示例	备注
[Zb]	字别	别字	我不能做我现在[Zb-再]的自己。 苍蝇比蜜蜂更聪[Zb-总]明。 他把这件事安[Zb-按]排好了。	汉语中存在该字，但用在该处不对
[Zc]	字错	错字	他脾[Zc-左竖心旁右“啤”的右边再]气不好。 祖父[Zc-写得像个“氏”]很惊讶地问。 看中国的电影[Zc-影字右半部分四撇]等。	汉语中没有该字，类似生造字。在“-”后描述具体的书写形式
[Zf]	字繁	繁体字	我爷爷[Zf-爺爺]80 岁了。	
[Zp]	字拼	拼音字	我们关系[Zp-xi4]很好。	为不自由语素注音，或双字与多字词中的单字注音。如果是整词都用拼音，则用 Cp 标记。
[Zy]	字异	异体字	他是个人才[Zy-材]。 经济凋[Zy-雕]落。	

3. 词汇标注偏误

标注 代码	代码 名称	代码 含义	示例	备注
[Cjy]	词混用	词的混淆	在沙漠里[Cjy-上]他们喝光了水。 这个非常重要[Cjy-重大]的人是妈妈。	不仅限于近 义词的混淆
[CZjy]	词组 混用	词组混淆	过了好多天连一只兔子也没捡到[CZjy-捡不到]。 亲身经历 [CZjy-直接来看]后，韩国和中国并不是完全不一样	
[Cq]	词缺	缺词	我每次都[Cq-都]很开心。	
[CSq]	词素缺	缺少语素	他看起来很骄傲[CSq-傲]。 人民大会堂很辉煌[CSq-煌]。	缺少构词成分
[CZq]	词组缺	缺词组	怕我朋友的经历发生在我身上[CZq-在我身上]。	
[Cd]	词多	多余的词	我可以说我是非常[Cd-很]喜欢。 我的朋友是美国[Cd-的]人。	
[CZd]	词组多	多余短语或 词组	[CZd-没想到这个]我在超市里选了我需要的东西。	
[Csd]	词素多	多语素	人民大会堂很辉煌[Csd-煌]。	
[Cp]	词拼	拼音词	他们让我去中国锻炼[Cp-duan lian]一下自己。	
[CSx]	词素序	词素(语素) 的顺序错误	我和他之间有代沟[CSx-沟代]。 警察抓了这几个罪犯[CSx-犯罪]。	词素序颠倒 后无论是否 成词，均标注
[Cx]	词实序	实词的顺序 错误	今天是我的运气好 [Cx-好运气] 她非常愉快，心情喜悦[Cx-喜悦心情]	
	词虚序	虚词的顺序 错误	他没想到兔子竟[Cx-竟兔子]死了 自己的田[Cx-田的]荒废了 我准备看关于经济的东西[Cx-东西关于经济]	
[Cz]	词造	生造词	免费下载歌曲引起了人们不再买昂贵的 CD，而导致了光碟 CD 的销量[Cz-卖量]逐渐下降的后果情况。	
[Cw]	词外	外文词	因为我不是中国人，我一点儿也不懂中国的文化 [Cw- culture]。	
[Cl]	词离	离合词错用	他蹲下来看兔子喘了一会儿气 [Cl-喘气一会儿]。 我想毕业后就结婚[Cl-和她结婚]她。	

Ccd	词重叠	词语重叠	他今天愉愉快快[Ccd-不能重叠]地来学校。 他总是说要研研究究[Ccd-研究研究]。	
[Cy]	词疑	自己拿不准的词语错误	[Cy-消化]	

4. 熟语偏误标注

标注代码	代码名称	代码含义	示例
[SYcy]	成语偏误	成语用错（1. 成语本身写错；2，成语用得恰当。）	鸡飞狗跳[SYcy--鸡飞猫跳] 每个人蠢蠢欲动[SYcy-成语不当]地找到一份有高工资的工作。 大家将来要自食其力[SYcy-成语不当]地找工作。 你会生活在一个伸手不见五指[SYcy-伸手不见指]黑暗的世界里。
[SYgy]	惯用语偏误	惯用语用错	开后门[SYgy-打后门]、跑龙套[SYgy-跑套子] 泼冷水[SYgy-泼热水]、碰钉子[SYgy-碰到钉子]
[SYxh]	歇后语偏误	歇后语用错	狗咬耗子——多管闲事[SYxh-猫咬耗子—多管闲事]
[SYyy]	谚语偏误	谚语用错	朝霞不出门，晚霞行千里。[SYyy-晚霞不出门，朝霞行千里]
[SYfixed]	待嵌格式偏误	固定格式用错	他喜欢东跑西颠[SYfixed-东颠西跑]

注：熟语标注中也可以标注字、词层面的偏误。如：他喜欢东跑西[Zb-西]颠[SYfixed-东颠西跑]，但标注符号应顺序标注，不能嵌套。

5. 用得精彩的词语、熟语标注代码

标注代码	偏误名称	示例
[Gw]	词用得好的	他给了我心灵的慰藉[Gw-慰藉]
[Gcy]	成语用得好的	闻鸡起舞[Gcy-闻鸡起舞]
[Ggy]	惯用语用得好的	跑龙套[Ggy-跑龙套]、泼冷水[Ggy-泼冷水]、碰钉子[Ggy-碰钉子]
[Gxh]	歇后语用得好的	狗咬耗子——多管闲事[Gxh-狗咬耗子—多管闲事] 大热天穿棉[Zb-绵]袄——不是时候[Gxh-大热天穿棉袄—不是时候]
[Gyy]	谚语用得好的	朝霞不出门，晚霞行千里。[Gyy-朝霞不出门，晚霞行千里]
[Gxcxy]	新词新语用得好的	我是一个月光族[Gxcxy-月光族]
[Gsg]	诗歌用得好的	有一句名言“纸上得来终觉浅，绝知此事要躬行[Gsg-纸上得来终觉浅，绝知此事要躬行]”
[Gmy]	名人名言用得好的	莎士比亚：“爱是一种甜蜜[Zb-蜜]的痛苦[Gmy-爱是一种甜蜜的痛苦]”。
[Gfixed]	固定格式用得好的	在场的人连连称赞[Gfixed-连连称赞]道

注：词语熟语用得精彩，但不表示其中不会出现字或词的偏误。

特殊情况处理：

1. 电子版作文在打字输入过程中出现的错误

原句：值得注意的是，租房从 1990 年到 2000 年一直在增长，愿意很值得调查。

标注：值得注意的是，租房从 1990 年到 2000 年一直在增长，原因[Cjy-愿意（电）]很值得调查。

2. 留学生作文写作中出现的知识性、常识性错误

原句：昨天，我看了一部很有名的中国电影——“中国合伙人”。这部电影上映于 2003 年。

标注：昨天，我看了一部很有名的中国电影——“中国合伙人”。这部电影上映于 2013[Cjy-2003（知）]年。

3. 字词语的标注，不能嵌套，只能按顺序先后标注。如

“爱是一种甜蜜[Zb-蜜]的痛苦[Gmy]”

三. 语料标注样例:

标注样例一:

[作文]

[作文 ID] [/作文 ID]

[姓名] 朴韩朔[/姓名]

[学号]201404157 [/学号]

[国籍] 韩国[/国籍]

[要求]

[命题]

[类型]自拟[/类型]

[标题]目标之力[/标题]

[/命题]

[体裁]议论文[/体裁]

[主题]关于目标的讨论[/主题]

[时长] 30 分钟[/时长]

[字数]不少于 400 字[/字数]

[辅助]不能借助词典[/辅助]

[具体要求]

1. 请从刚才阅读的三个故事中选择一篇,作为写作的基本事例,并联系你个人的经历或阅读过的相似故事,写一篇两事一理的短文。题目自拟;

2. 作文的观点与材料能够总结出的观点一致;

3. 按“事 1—理一事 2—理”或“理一事 1—理一事 2—理”的结构写作;

4. 写作中引用阅读材料作为基本事例时用“材料 n”表示,不必重复;

【探险队(材料 1),农场主(材料 2),剑客(材料 3)】

5. 除阅读材料以外,自己写作的内容不少于 400 字;

[/具体要求]

[/要求]

[指导]

[有无指导] 有指导[/有无指导]

[指导过程]事前指导 [/指导过程]

[指导内容] 分析体裁特点;说明文章结构;训练相关语言表达;训练观点提炼[/指导内容]

[/指导]

[课型]经贸汉语阅读与写作[/课型]

[来源]期中考试作文[/来源]

[教材]自编教材 [/教材]

[学期] 2014-2015-2 [/学期]

[班级]2015 同传 1 班[/班级]

[源文]2015 春季学期学生作文\期中考试作文[/源文]

[日期] 2015.05.10[/日期]

[文稿类型]电脑版[/文稿类型]

[版本]二稿[/版本]

[正文]

[题目]目标之力[/题目]

今天的大家，在[Cq-在]在我看来，都没有活力，都没有目标。学生们没有了梦想，人们都安于[Cz-安住]自己[CSd-的]现在的生活，我也差不多，以前有很大的梦想，但现在那个都没有了。我们到底[Zb-低]怎么回[Zb-会]事呢[Zb-尼]？

有天一个探险队去某个沙漠探险，热热的沙漠里[Cjy-上]他们[Cd 都]喝光[CSq-光]了水。他们非常非常渴，到了马上快渴死的地步。队员们说：“渴死了！”“我快死了[Cq-了]，我以后[Zb-候]不能走了[Cq-了]。但队长的包里还有一壶水，队长说：“这壶水是最后的水，过沙漠前谁都不能喝！但过了[Cq-了]沙漠的话，我让你们随便[Zb-边]喝！”

队员们相信队长，终于过去了沙漠，打开那壶水[CSd-了]，他们真惊了，因为壶里一点儿水也没有[Cx-也没有水]，而且有黄沙！然后他们就都醒悟[CSq-悟]了，是[Cq-是]队长救[Zb-求]了他们的生命。

如果他们没有“水”，那样的希望[Cjy-目标]的泪，他们在沙漠里面就[Cx-就在沙漠里面]会渴死，但他们有了“水”的希望[Cjy-目标]后继续追求那个目标，终于他们可以求生。那样，目标里还有[Cjy-在]“求生”的能力。

还有这里有关于蜜蜂的故事。蜜蜂在[Cq-在]生物学的观点看来 [Cx-看来生物学的观点]，它绝[Zb-继]对不能飞那么长的时间，绝[Zb-继]对不能飞去那么长的距离，蜜蜂却[Cx-却蜜蜂]飞得那么好。对生物学者来讲，它飞得好的原因就在它的目标性，才飞得那么长。

在人生当中，我们很多人[CZq-很多人] [Cd-太]没有那个“目标”。但其实，目标里有很大的潜力，甚至有你知道的求生的能力。如果你要某件事[Cd-的]成功，你必须[Zb-需]要立目标，为了目标不断努力，好像蜜蜂一样！

[/正文]

[批语] 题目 5/5 观点 8/10 略不同于故事 故事 8/10 结构 15/15 语言 26/30

[/批语]

[得分]62/70[/得分]

[原文]

今天的大家，在我看来，都没有活力，都没有目标。学生们没有了梦想，人们都安住自己的现在的生活，我也差不多，以前有很大的梦想，但现在那个都没有了。我们到底怎么会事尼？

有天一个探险队去某个沙漠探险，热热的沙漠上他们都喝了水。他们非常非常渴，到了马上快渴死的地步。队员们说：“渴死了！”“我快死，我以候不能走。但队长的包里还有一壶水，队长说：“这壶水是最后的水，过沙漠前谁都不能喝！但过沙漠的话，我让你们随边喝！”

队员们相信队长，终于过去了沙漠，打开那壶水了，他们真惊了，因为壶里一点儿也没有水，而且有黄沙！然后他们就都醒了，队长求了他们的生命。

如果他们没有‘水’，那样的目标的话，他们就在沙漠里面会渴死，但他们有了‘水’的目标后继续追求那个目标，终于他们可以求生。那样，目标里还在‘求生’的能力。

还有这里有关于蜜蜂的故事。蜜蜂看来生物学的观点，它继对不能飞那么长的时间，继对不能飞去那么长的距离，却蜜蜂飞得那么好。对生物学者来讲，它飞得好的原因就在它的目标性，才飞得那么长。

在人生当中，我们太没有那个”目标”。但其实，目标里有很大的潜力，甚至有你知道的求生的能力。如果你要某件事的成功，你必需要立目标，为了目标不断努力，好像蜜蜂一样！

[/原文]

[提供]吴晓峰[/提供]

[教师]吴晓峰[/教师]
[录入]2022 级对外汉语. XXX [/录入]
[录入日期] 2022. 05. 20 [/录入日期]
[一校] 2022 研. 李嘉奇[/一校]
[一校日期]2023. 春[/一校日期]
[二校][[/二校]
[二校日期][[/二校日期]
[三校][[/三校]
[/作文]

标注样例二：

[作文]
[作文 ID] [/作文 ID]
[姓名]金润成[/姓名]
[学号]201404162[/学号]
[国籍]韩国[/国籍]
[要求]
 [命题]
 [类型]自拟[/类型]
 [标题]为什么写作是头疼吗[/标题]
 [/命题]
 [体裁]调查报告[/体裁]
 [主题]关于留学生汉语写作的调查[/主题]
 [时长]约需 30 到 50 分钟 [/时长]
 [字数]300 字左右[/字数]
 [辅助] 可以借助词典或手机 [/辅助]
 [具体要求]
 以下数据来自《对外经贸大学留学生汉语写作现状调查》，调查对象为对外经济贸易大学一年级留学生，请说明并分析下面两张图所提供的信息，并得到调查结论。
 要求：1. 采用结论式小标题；2. 包含图表说明和分析；3. 300 字左右。
 [/具体要求]
[/要求]
[指导]
 [有无指导]有指导[有无指导]
 [指导过程]事前指导 [/指导过程]
 [指导内容]分析体裁特点；训练观点提炼[/指导内容]
[/指导]
[课型]经贸汉语阅读与写作[/课型]
[来源]期末考试作文[/来源]
[教材]自编教材 [/教材]
[学期] 2014-2015-2 [/学期]
[班级]2015 级国贸某班[/班级]
[源文]2015 春季学期学生作文\期末考试作文[/源文]
[日期] 2015. 07. 01[/日期]
[文稿类型]电脑版[/文稿类型]

[版本]二稿[/版本]

[正文]

[题目]为什么写作是头疼[Cd-吗]的[Cq-的][BD-“是”字句偏误][题目]

现在在世界上用汉字的国家[Cd-是]只有两个国家一个当然是[Csx-是当然]中国还有一个是日本。韩国也是以前使用[Cd-了]，所以除了[Cq-除了]东北亚国家其他[Cjy-另外]国家是非常难的[Cq-的]。我的一个美国朋友说他自己不是写[Cd-了]字，他感觉是[Cq-是][Cd-一种]画画儿。[Cd-现在]很多留学生认为[Cq-认为]最难的课目[Cd-是]应该[Zb-刻]是写作，先看图 1，不知道写什么的[Cq-的]学生是 29%，不知道怎么写的[Cq-的]61%，根本不想写的[Cq-的]是[Zc-写得像“晟”] 10%，汉字的种类[CSq-类]特别多，所以学习汉语的时候一定要努力学习！如果是[Cq-是]西方国家学生[Cd-是]肯定需[Zb-须]要很长时间。图 2 也一样，很多方面学写作有[Cq-有][Cd-的]难的理由，文章结构方面 28.6%，词汇方面 36.7%，语法方面 16.3%，汉字方面 16.3%，其他是 2%。各[Zb-个]种各[Zb-个]样的理由，这样的问题[Cd-是]当然能克服，天天努力学习[Cd-这个]是最好的方法。

[/正文]

[批语] 无[/批语]

[得分]11/20[/得分]

[原文]

如图所示，对外经贸大学留学生汉语写作最头疼的问题如下，不知道怎么写占 61%，不知道学什么占 29%，根本不想写占 10%。首先在学习汉语上，留学生也找到汉语写作的困难。最大的困难是不知道怎么写，在写作文的时间大学生必将忘怎么写汉字或语法，因为一年级留学生是刚刚学汉语的汉字和语法。汉字有一点难写因为错一个字比如“我”和“找”，一个表示我自己，一个表示找人或找东西。语法也是不一样的对留学生的国家，因此汉语写作的语法有一点难。再次，不知道写什么是词汇方面文章结构方面。在学写作上留学生需要记得多词汇。对留学生的相思。这种问题难学。文章结构也是一个原因为什么该情况变成一个问题，结句是不简单的写。我以为该学生不认真学习的。

最中，在学汉语写作的方法，一定在课常上认真听老师，问老师。到家是时候有时间再学习老师教的。

[/原文]

[提供]吴晓峰[/提供]

[教师]吴晓峰[/教师]

[录入]2022 级对外汉语. XXX[/录入]

[录入日期] 2023. 05. 20 [/录入日期]

[一校] 2023 研. 王馨雨[/一校]

[一校日期]2023. 05. 20[/一校日期]

[二校]2022 研. 李嘉奇[/二校]

[二校日期][[/二校日期]

[三校][[/三校]

[/作文]

作文语料库结构;

序号 n,
作文 ID c(12),
作者 c(16),
学号 c(10),
国籍 c(16),
题目类型 c(4),
标题 c(30),
体裁 c(10),
主题 m,
时长要求 c(10),
字数要求 c(15),
辅助 c(2),
作文要求 m,
有无指导 c(2),
指导过程 c(30),
指导内容 m,
课型 c(20),
来源 c(80),
教材 c(50),
学期 c(13),
班级 c(16),
日期 c(10),
文稿 c(6),
版本 c(4),
正文 m,
题目 m,
批语 m,
得分 c(5),
原文 m,
提供 c(8),
教师 c(8),
录入 c(30),
录入日期 c(10),
一校 c(30),
一校日期 c(10),
二校 c(30),
二校日期 c(10),
三校 c(30),
源文 c(40))