

FCGEC 标注手册（外部分发版）

（内部文档未经编者允许，请勿转发传播）

a) 标注准则及注意点

1. 标注之前，请首先浏览一遍完整的句子，对句义有一个大体的了解，标注时应强调直觉。并根据给出的修改解析确定该病句的核心错误。
2. 有时候，一个错误的句子可以对应多个正确的修改方案(解析往往不会给出所有的修改方案)，我们鼓励标注人员进行多样化的修改，例如：
 - i. 对于搭配不当类型的病句，如“他带领技术人员完成了一种造型美丽、性能可靠的新型吸油烟机。”，该错误是由于“完成”与“新型吸油烟机”搭配不当。那么我们可以把“完成”改为“制造”也可以是“设计”，“制作”等。
 - ii. 对于结构混乱类型的病句，如“近期，全国各地各种安全事故频频发生，造成大量的人员伤亡和巨大的财产损失，其主要原因是由于缺乏安全意识和监督管理不力造成的。”该错误是由于“主要原因是…”和“…造成的”两个结构杂糅，因此我们的改法可以删去其中之一，这样就有两种改法。
3. 由于有解析的存在，因此基本上大方向是确定的。标注只需要按照这个大方向去尽可能多样的对病句修改即可。另外在实际标注过程中要学会套用本手册下文中给出的例子，用最合适的方式标注扩展。
4. 如果在标注过程中发现解析和句子不符的情况，那么可以采用报告错误功能并跳过标注。
5. 对于手册中没有明确给出，但有相近例子的错误类型及纠正，应学会灵活地举一反三。
6. 标注过程中对于词语修改插入类型的错误，要善于利用已有的词典资源，先充分理解，再标注。千万不要贸然标注，影响质量。可以利用的资源如：在线新华字典 (<https://zidian.aics.cn/>)，近义词在线 ([近义词大全 - 在线近义词查询](#)) 等。

b) 错误类型详解

本项目的数据来源是历年来各地的病句试题。因此大家可能对里面的许多错

误类型不陌生。我们给定每个类型几个不同的例子来展示如何尽量多样化的进行标注。

1. 语序不当

语序不当涉及主要是两方面，其一是句法结构上，语序不当造成句子缺失成分(注意：这种情况下本手册统一认定其类型为语序问题而不是成分残缺)；其二则是语义表达中的递进关系错误。该类型错误基本上不需要多样化的标注，以下为一一些可参考的例子：

- i. **错误句子：**航空公司采用新技术后，飞机油耗不仅大大减少了，而且飞机飞行速度也有所提高。
正确句子：航空公司采用新技术后，不仅飞机油耗大大减少了，而且飞机飞行速度也有所提高。
解析：“不仅……而且”是固定句式，位置应该对应相同
- ii. **错误句子：**由于青少年好奇心强、辨别是非的能力相对较弱，比成年人更容易受不良信息的侵害。
正确句子：青少年由于好奇心强、辨别是非的能力相对较弱，比成年人更容易受不良信息的侵害。
解析：主语前后不对应，应将“青少年”和“由于”位置互换
- iii. **错误句子：**学校自从开展研究性学习以来，同学们踊跃参与，创新意识和创新能力得到很大的提升。
正确句子：自从学校开展研究性学习以来，同学们踊跃参与，创新意识和创新能力得到很大的提升。
解析：主语前后不对应，应将“学校”和“自从”位置互换
- iv. **错误句子：**参加奥运的选手们十分清楚，一场比赛的输赢，不仅关系到祖国的荣誉，而且关系到个人的尊严。
正确句子：参加奥运的选手们十分清楚，一场比赛的输赢，不仅关系到祖国的荣誉，而且关系到个人的尊严。
解析：递进关系错误，“祖国的荣誉”和“个人的尊严”根据语义来看应该是先“个人的尊严”再“祖国的荣誉”
- v. **错误句子：**经过老师的教导，我们改正并认识到了自己的错误。
正确句子：经过老师的教导，我们认识并改正到了自己的错误。
解析：递进关系错误，对于错误来说，根据语义应该是先“认识”到错误，再改正。因此需要互换“改正”和“认识”

2. 搭配不当

搭配不当主要来源于三方面，其一是主谓、动宾结构词语搭配不当；其二则是修饰词使用不当；除此之外还有一类是由于两面对一面造成的。错误的原因经常是由于出现长句时，容易顾头不顾尾，顾近不顾远。搭配不当往往是可以多种修改方式的，因此标注者在标注这类错误的时候应该尽量进行多样化的标注。

以下为一些可参考的例子：

- i. **错误句子：**在 2008 年花样滑冰世锦赛上，张丹、张昊两人的表演，无论在艺术上和技术上都达到了很高的境界，获得了最高分。
正确句子：在 2008 年花样滑冰世锦赛上，张丹、张昊两人的表演，无论在艺术上还是在技术上都达到了很高的水平，获得了最高分。
解析：有两处错误，一是“无论…还是”是固定关联词搭配，二是“境界”不能修饰“技术”，应该用“水平”
- ii. **错误句子：**为贯彻中央关于加强税收对居民收入分配调节的要求，降低中低收入者税收负担，加大对高收入者的调节力度，有必要修订个人所得税法。
正确句子：为贯彻中央关于加强税收对居民收入分配调节的要求，减轻中低收入者税收负担，加大对高收入者的调节力度，有必要修订个人所得税法。
解析：错误在于“降低”无法用于修饰“负担”，而应该是“减轻”，为了多样化标注，同理还可以改为“减少”
- iii. **错误句子：**我国农村正步入“第三步”改革的新阶段。这一阶段将以促进农村上层建筑改革为核心，实行农村综合改革，改进农村上层建筑与经济基础不相适应的某些深层问题。
正确句子：我国农村正步入“第三步”改革的新阶段。这一阶段将以促进农村上层建筑改革为核心，实行农村综合改革，解决农村上层建筑与经济基础不相适应的某些深层问题。
解析：错误在于“改进”无法用于修饰“问题”，而应该是“解决”，为了多样化标注，同理还可以改为“处理”等词
- iv. **错误句子：**省气象台的资料显示，我省今年 7 月份每天的平均气温为 31℃，创下了历史平均最高气温。
正确句子：省气象台的资料显示，我省今年 7 月份每天的平均气温为 31℃，创下了历史平均最高气温记录。
解析：错误在于“气温”不能用“创下”来修饰，所以要在“气温”后加上“记录”。当然我们也可以转换思路，也可以对“创下”进行修改，将其改成“突破”，这样就实现了多样性的标注
- v. **错误句子：**是牺牲环境发展经济，还是争取发展经济和环保双赢，这是关系到保证

我市社会健康发展的重大问题。

正确句子：是牺牲环境发展经济，还是争取发展经济和环保双赢，这是关系到我市社会健康发展的重大问题。

解析：错误在于两面对一面，可以删去“保证”。同时为了多样性的标注，也可以改为“这是关系到能不能保证我市社会健康发展的重大问题”

3. 结构混乱

结构混乱是在语言语法中常见的一种错误，常见的错误原因是把两种不同的句法结构混杂在一个表达式中，结果造成语句结构混乱、语义纠缠。其还有暗换主语，藕断丝连等其他情况。由于结构混乱往往涉及两种结构，修改方式大多为删去其中的一种，所以很多结构混乱都至少会有两种修改方式，以下为一些可参考的常见例子：

- i. **错误句子：**很久以后，香雪才知道它所以能自动合上的原因，是因为铅笔盒里包藏着一块不大不小的吸铁石。
正确句子：很久以后，香雪才知道它所以能自动合上的原因，是铅笔盒里包藏着一块不大不小的吸铁石。
解析：错误在于“…的原因，是因为…”两个重复结构杂糅，一种改法是去掉“的原因”，另一种则是删去“是因为”
- ii. **错误句子：**有关专家指出，当前房价过高，关键是房子升值过快在起着重要作用，以至于大量的人都把房子当成了谋利的手段。
正确句子：有关专家指出，当前房价过高，房子升值过快在起着重要作用，以至于大量的人都把房子当成了谋利的手段。
解析：错误在于“关键是…”以及“在起着重要作用”两个重复结构杂糅，一种改法是去掉“关键是”，另一种则是删去“在起着重要作用”
- iii. **错误句子：**百叶窗并非舶来品，湖北云梦县曾出土了一件陶土的楼房模型，据鉴定，属于东汉晚期，中国人已经在享受百叶窗之益了。
正确句子：百叶窗并非舶来品，湖北云梦县曾出土了一件陶土的楼房模型，据鉴定，模型属于东汉晚期，中国人已经在享受百叶窗之益了。
解析：“东汉晚期”是“属于”的宾语，又被当做了后一分句的状语，因此后一句主语被暗换了，改法是“据鉴定，**模型**属于东汉晚期，这一时期中国人已经在享用百叶窗了”。在多样化标注中，这里可以在局鉴定后加入“模型”，“该模型”或是“楼房模型”
- iv. **错误句子：**寒流侵入我国的路径，不是每次都一样的，这要看北极地带和西伯利亚的冷空气哪一部分气压最高，哪一部分气压最低所决定的。
正确句子：寒流侵入我国的路径，不是每次都一样的，这要看北极地带和西伯利亚

的冷空气哪一部分气压最高，哪一部分气压最低。

解析：类似错误(ii)，在这里“这要看…”以及“…所决定的”两个重复结构杂糅，一种改法是去掉“所决定的”，另一种则是将“这要看…”改为“这是由”

- v. **错误句子：**借着中秋小长假的东风，许多旅行社都推出主题为以弘扬传统文化为目的的“中秋赏月之旅”。各景区为满足游客需求，也推出了像“欢乐中秋游”等众多活动。

正确句子：借着中秋小长假的东风，许多旅行社都推出主题为弘扬传统文化的“中秋赏月之旅”。各景区为满足游客需求，也推出了像“欢乐中秋游”等众多活动。

解析：类似错误(ii)，在这里“以…为目的”以及“主题为…”两个重复结构杂糅，一种改法是去掉“以，为目的”，另一种则是删去“主题为”

- vi. **错误句子：**这个问题是值得我们所重视的。

正确句子：这个问题是值得我们重视的。

解析：类似错误(ii)，在这里“值得…”以及“…所”两个重复结构杂糅，一种改法是去掉“值得”，另一种则是删去“所”

4. 成分残缺

此类错误主要是因为句中的部分成分丢失，例如主语，宾语，介词等。该类

别的错误往往可以有多多样性的标注结果。以下为一些常见的例子：

- i. **错误句子：**从“民族英雄”一夜之间沦为“科学骗子”，这种大起大落，不仅韩国克隆专家黄禹锡本人是一次沉重的打击，而且让整个韩国科学界为之蒙羞，更让人类的克隆科学研究遭受了重创。

正确句子：从“民族英雄”一夜之间沦为“科学骗子”，这种大起大落，不仅韩国克隆专家黄禹锡本人是一次沉重的打击，而且让整个韩国科学界为之蒙羞，更让人类的克隆科学研究遭受了重创。

解析：介词缺失，应该在“不仅”后加上“对”

- ii. **错误句子：**首届中国大学生电视节设有最受大学生欢迎的主持人、最受大学生欢迎的电视剧和最受大学生欢迎的栏目。

正确句子：首届中国大学生电视节设有最受大学生欢迎的主持人、最受大学生欢迎的电视剧和最受大学生欢迎的栏目等奖项。

解析：宾语缺失，根据多样性标注要求，可以在最后加上“等奖项”，“等荣誉”等词语

- iii. **错误句子：**当一个人在现实生活中感到孤独和寂寞的时候，就会使他从甜美的回忆中得到难以言传的慰藉和愉悦。

正确句子：当一个人在现实生活中感到孤独和寂寞的时候，他就会从甜美的回忆中得到难以言传的慰藉和愉悦。

解析：主语缺失，“就会使他从”造成了后半句主语缺失，因此应改为“他就会从”

- iv. **错误句子:** 据有关资料披露, 近一年来, 山西省的艾滋病已由高危人群向普通人群扩散的快速增长阶段。
正确句子: 据有关资料披露, 近一年来, 山西省的艾滋病已进入由高危人群向普通人群扩散的快速增长阶段。
解析: 谓语缺失, 因此需要在“已”后面加上“进入”
- v. **错误句子:** 人的悲剧性实质, 在于走向前方、到处流浪时, 又无时无刻惦念着正在远去和久已不见的家、家园和家乡。
正确句子: 人的悲剧性实质, 在于走向前方、到处流浪时, 又无时无刻惦念着正在远去和久已不见的家、家园和家乡。
解析: 谓语缺失, 因此需要在“已”后面加上“进入”

5. 成分赘余

成分赘余与成分残缺相反, 是句子中有部分要素重复出现, 因此需要删去其一。同样地, 大部分样本都有两种修改方式。需要注意的是, 如果是在修改类型标签时需要注意其和结构混乱的区别, 结构混乱是两种特定的句式结构的杂糅, 而这里的错误往往不是结构上的, 具体的常见例子如下所示:

- i. **错误句子:** 面对着越来越多标榜自己是经典的作品的出现, 我们需要思考一下: 到底什么样的作品可以堪称经典?
正确句子: 面对着越来越多标榜自己是经典的作品的出现, 我们需要思考一下: 到底什么样的作品堪称经典?
解析: 错误在于“可以”和“堪”的意思是重复的, 一种改法是去掉“可以”, 另一种则是将“堪称”改为“称为”
- ii. **错误句子:** 学生应重视课前预习, 在预习中发现问题, 上课时就可以向老师提出质疑。
正确句子: 学生应重视课前预习, 在预习中发现问题, 上课时就可以向老师质疑。
解析: 错误在于“提出”和“质疑”的意思是重复的, 一种改法是去掉“提出”, 另一种则是将“质疑”改为“疑问”
- iii. **错误句子:** 据《光明日报》报道, 中国青少年近视趋于低龄化, 全国青少年近视率平均达 40%, 大学生近视率已逾 70% 以上。
正确句子: 据《光明日报》报道, 中国青少年近视趋于低龄化, 全国青少年近视率平均达 40%, 大学生近视率已逾 70%。
解析: 错误在于“逾”和“以上”的意思是重复的, 一种改法是去掉“以上”, 另一种则是将“逾”改为“达”

6. 不合逻辑

不合逻辑属于语用学上的错误, 其有多种的表现形式。例如并列关系错误,

因果关系颠倒，否定词连用造成多重否定等，需要反复推敲。这里展示部分常出现的例子及修改方式作为参考：

- i. **错误句子：**本田汽车社长福井威夫认为欧洲是汽车文化的发源地，也是对品质要求最为严格的市场，品质能否得到欧洲顾客的认可是出口事业成功的关键。
正确句子：本田汽车社长福井威夫认为欧洲是汽车文化的发源地，也是对品质要求最为严格的市场，品质得到欧洲顾客的认可是出口事业成功的关键。
解析：错误是因为两面对一面，为了多样性的改法，可以在“成功”前加上“能否”或者是去掉“品质”后的“能否”
- ii. **错误句子：**所谓象征，就是通过一特定的具体形象间接地表现某种概念、思想和情感，它要求象征物和被象征物之间有某种相似的特点，能使人引起由此及彼的联想，反而给读者留下深刻的印象以及广阔的思考空间。
正确句子：所谓象征，就是通过一特定的具体形象间接地表现某种概念、思想和情感，它要求象征物和被象征物之间有某种相似的特点，能使人引起由此及彼的联想，反而给读者留下深刻的印象以及广阔的思考空间。
解析：按照逻辑来说，这里的“反而”是错误的，应该改为“从而”
- iii. **错误句子：**这件事以后，大家立即召开支委会，互相作了深刻的自我批评。
正确句子：这件事以后，大家立即召开支委会，作了深刻的自我批评。
解析：“互相”与“自我批评”是矛盾的，两种改法分别是删去“互相”以及删去“自我”
- iv. **错误句子：**防止煤矿不出事故的最好办法，就是加强安全工作。
正确句子：防止煤矿出事故的最好办法，就是加强安全工作。
解析：“防止”与“不”构成了双重否定，造成了逻辑的错误。应该将“防止”改为“使”或者是删去“不”
- v. **错误句子：**市委要求，各学校学生公寓的生活用品和床上用品由学生自主选购，不得统一配备。
正确句子：市委要求，各学校学生公寓的生活用品由学生自主选购，不得统一配备。
解析：该错误属于并列关系不当，“生活用品”包含“床上用品”，因此改法是：删去“和床上用品”或者是把“和”改为“如”等词来实现多样化的标注

7. 表意不明

此类错误是因为歧义引起的，也即文字表达的语义不唯一，因此相对较难修改，但同样的，这也是数量最少的一类错误，一些例子如下所示：

- i. **错误句子：**这一天街上很热闹，工人免费修自行车，医生免费检查身体。
正确句子：这一天街上很热闹，工人免费为群众修自行车，医生免费为群众检查身

体。

解析：该错误的产生是因为“工人免费修自行车”无法表达是工人帮别人免费修还是工人可以享受免费修自行车的待遇，“医生免费检查身体”也是同理。

- ii. **错误句子：**油条只要不是人工添加含铝物质，也没有含铝容器和含铝包装污染的问题，其含量通常足够低，到了不产生健康威胁的程度，就无需担心了。

正确句子：油条只要不是人工添加含铝物质，也没有含铝容器和含铝包装污染的问题，铝含量通常足够低，到了不产生健康威胁的程度，就无需担心了。

解析：该句语义相对明确，错误的产生是因为“其”指代不明，不能确定“其”指的是“油条”还是“铝”，根据句意来判断应该是指“铝”

- iii. **错误句子：**农业部 25 日表示，康菲公司提供的 10 亿元将用于解决河北省和辽宁省部分区县养殖生物以及渤海天然渔业资源损害赔偿、补偿问题；地方政府部门将组织做好资金发放工作。

正确句子：农业部 25 日表示，康菲公司提供的 10 亿元将用于解决河北省和辽宁省部分区县养殖生物以及渤海天然渔业资源损害赔偿、补偿问题；地方政府部门将组织做好资金发放工作。

解析：“河北省和辽宁省部分区县”表意不明，有歧义。该类情况为了多样化标注可以改为“河北和辽宁两省的部分区县”或“部分河北省和辽宁省区县”

- iv. **错误句子：**他决定背着妈妈去医院检查身体。

正确句子：他决定瞒着妈妈去医院检查身体。

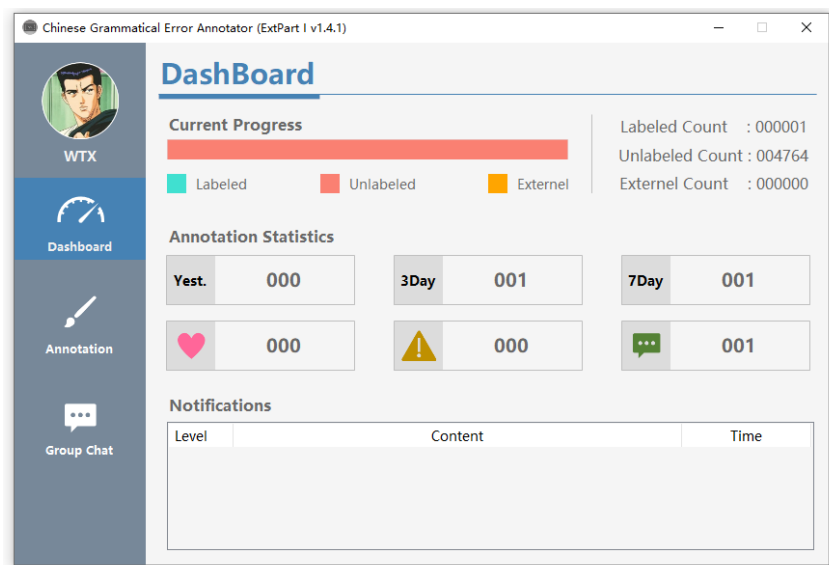
解析：该句“背”是产生歧义的原因，可以将“背”改为“瞒”或者在身体前加“她的”

(望您在阅读以下部分前已仔细阅读以上准则和示例)

c) 标注程序基本介绍以及不同程序的标注要求

1. Dashboard 界面

该界面用于显示当前的总览信息，包括标注进度等信息，可供标注员对速度进行把控参考。



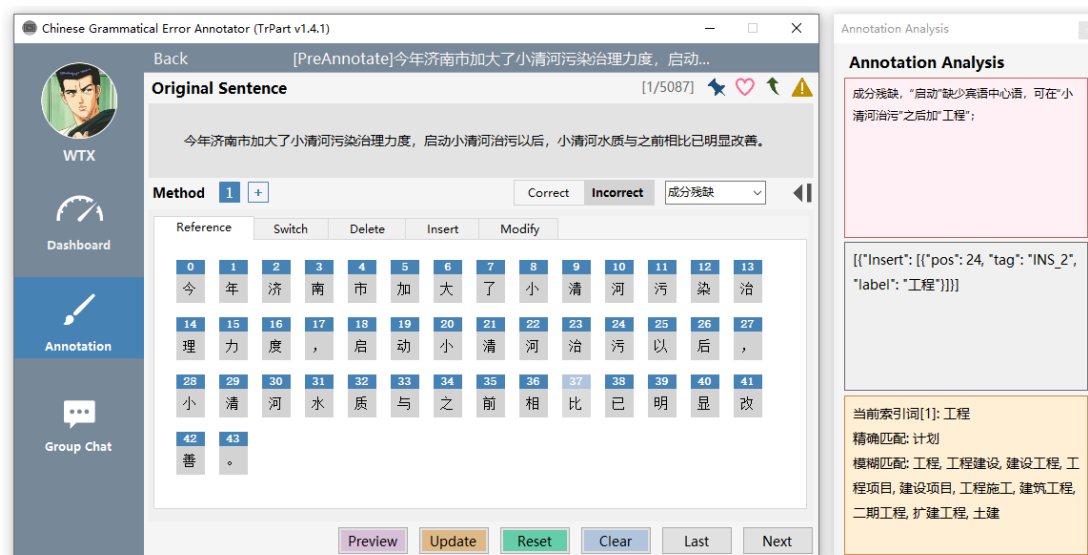
2. 标注导航界面

标注导航界面可以快速选择跳转到对应的样本位置。其中推荐标注者从上方导航栏中的“Unlabeled”栏目下进行标注，该栏目下仅显示未标注的样本，因此可以直接进入标注而不需要再去翻找上一次标到的位置。



3. 标注界面

标注界面分为两个窗体，左侧窗体为标注窗体，右侧区域则是提示窗体，如下图所示：



其中各部分区域及各部分作用解释如下所示：



- (1) **原句/修改后的句子区域**：该区域会显示病句文本。当按下“6 按钮操作区”中的“Preview 预览按钮”时该区域会变成按当前操作修改后的句子
- (2) **增加改错方案区域**：该区域点击加号可以增加新的改错方案(多样化标注要求)，并通过点击不同方案数据进行操作的切换

(3) **特殊按钮区域：**该区域从左到右的四个按钮分别代表：

- i. 图钉按钮：窗体置顶(程序默认开启置顶)
- ii. 爱心按钮：类似收藏功能，收藏后的句子可以在标注导航界面的“Liked”下查看，可以把不确定或者先放放的句子用这个功能码住
- iii. 绿色箭头：发起讨论，如果你觉得这个的改法或是解析拿捏不准，可以点击后输入讨论内容，之后我们会进行查看和回复
- iv. 黄色感叹号：如果句子本身存在问题(句子不全，句子中有奇怪的字母或是存在错别字)

(4) **错误类型区域：**该区域有一个多选下拉框，需要核查病句的错误类型并进行改正和选择。

(5) **文本操作区：**该区域用于文本操作，“5.1 操作选择”分别对应更换顺序、插入、删除、修改四种操作，“Reference”仅提供索引参考，可忽略。

操作方式主要为：

- i. 鼠标点击：选择
- ii. 鼠标右键：多选(确定左右边界后会自动选择所有在边界内的字符，主要用于“更改顺序”以及“修改”这两个操作)

我们会在培训中给大家演示具体的标注方法。

(6) **按钮操作区域：**该区域从左到右的按钮分别代表：

- i. “Preview”按钮是预览当前操作后的文本，会显示在区域1中
- ii. “Update”按钮是更新当前的操作到服务器上(为了防止一些情况下数据没有自动上传，请务必在每条样本标完后按一下，避免)
- iii. “Reset”按钮时恢复操作为初始情况(该按钮一般不需要使用)

iv. “Clear” 按钮用于清空所做操作，其中短按是清空当前操作页，长按是清空所有操作

v. “Last” 和 “Next” 按钮用于切换标注样本

(7) **解析提示区**：句子的修改解析，标注者可以进行参考

(8) **标签转化提示区**：该区域标注者可忽略

(9) **近义词提示区**：对于部分样本，为了更好的帮助标注者进行多样化标注，我们提供了多种的近义词项供标注者参考，标注者可选择推荐的近义词中合理的词语增加自己的修改方案。

4. 讨论界面

讨论界面中可查看讨论的消息，其中创建一个讨论需要在标注界面使用“3 特殊区域”中的绿色剪头按钮或是黄色感叹号按钮，具体讨论界面如下所示：

