

甲骨文标注指南

一、标注费用

1.标注者薪酬

薪酬标准为每张甲片 0.6 元。为了获得薪酬，标注者至少需要标注 20 片甲片

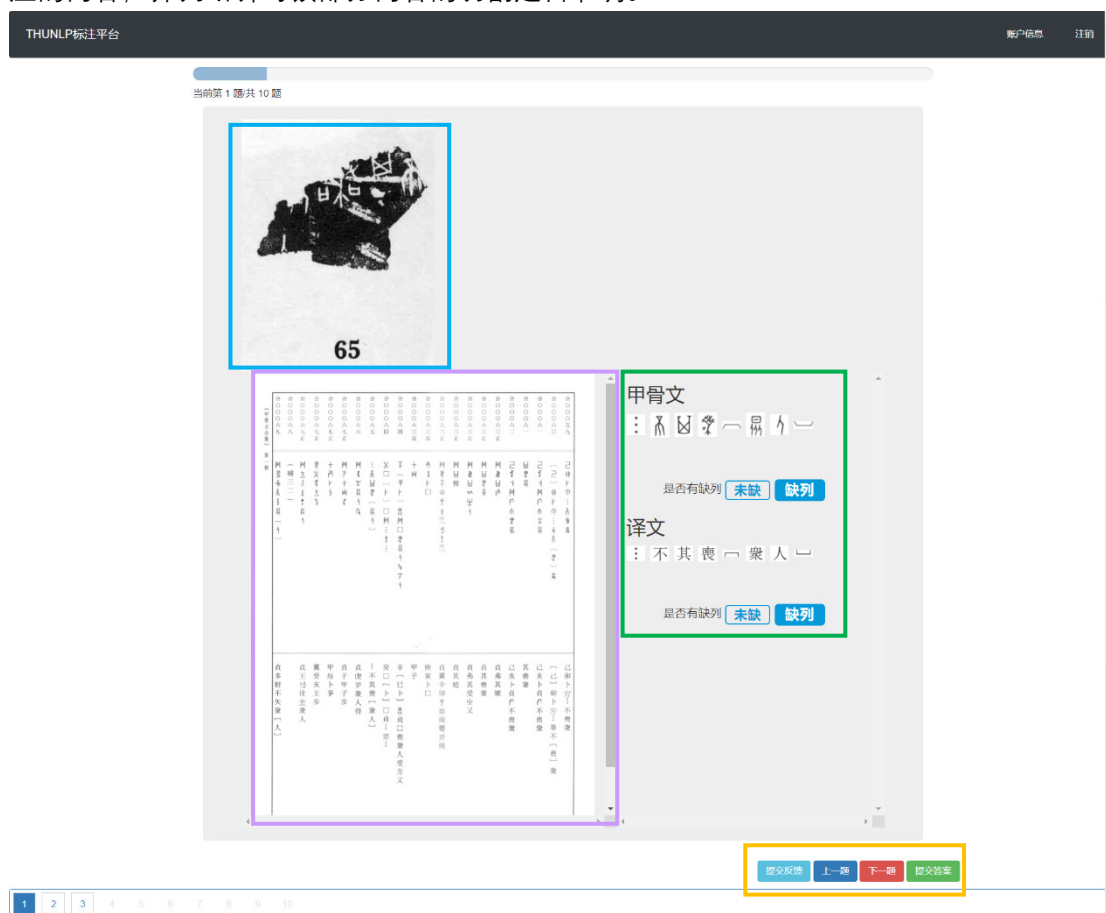
2.质量检测

在正式开始标注前，标注者必须仔细阅读学习本标注指南。在标注过程中，标注者须按照本标注指南中的标准进行标注。在完成标注后，我们会将您的标注结果与其他标注者的结果进行一致性检验。若检验结果异常，则会对您的标注进行人工核查。如果在人工核查后，我们认为您的标注存在明显的质量问题，我们将拒绝支付薪酬。因此，标注者需要始终确保自己的标注质量。

二、标注任务

1.概述

本次标注的结果将用于甲骨文数据库的建立以及其他针对甲骨文的计算机研究。标注任务为核查甲片拓片、《殷墟甲骨文字释全编》中对应的内容以及切割后得到的甲骨文信息是否匹配。对于每一片甲片，我们需要您根据甲片的编号在《殷墟甲骨文字释全编》中找到对应的内容，并判断针对该部分内容的切割是否准确。



如上图所示，甲片的标注页面由如下四部分组成：

- (1) 甲片拓片（蓝框部分）：最下方标注了甲片的编号，与《殷墟甲骨文字释全编》中

的编号对应。如本图中编号“65”，即对应书中编号“合〇〇〇六五”。部分甲片的编号可能带有“正”“反”“甲”“乙”“臼”等附加编号信息，属于正常现象，在书中也会有对应的附加信息出现。

(2) 《殷墟甲骨文摹释全编》对应页面（紫框部分）：该书为竖排页面，每页从上至下包括甲片编号、甲骨文摹写内容、对应翻译内容三部分。由于甲片上内容可能存在多列、分区等现象，因此可能会出现同一个甲片编号在书中出现多次（如果横跨多页则紫框部分将呈现多页内容，请注意不要遗漏），或甲片编号区中的一列对应甲骨文摹写内容区、翻译内容区中的多列等现象。

(3) 需要标注的切割得到的甲骨文信息（绿框部分）：此处展示的是将甲片拓片在《殷墟甲骨文摹释全编》中对应的甲骨文摹写内容与对应翻译内容切割为单字的信息，需要您来标注在切割时是否存在问题。书中每处甲骨文编号将对应绿框中的一组，如果编号重复出现多次则将呈现为多组。

(4) 工具栏（黄框部分）：包含“提交反馈”“上一题”“下一题”“提交答案”等 4 个按钮。“提交反馈”用于反映特殊问题，“上一题”“下一题”用于在不同甲片间切换（也可以点击页面左下角的数字序号来切换甲片），“提交答案”用于提交**整组**标注结果（每组标注共有 10 张甲片，请在 10 张甲片都标注完后再点击“提交答案”）。

2.标注规则

针对每一片甲片，您需要判断切割得到的信息是否存在以下问题：

(1) 缺列：同一个甲片编号在书中出现多次，但切割时遗漏了某些列；或书中出现甲片编号区的一列对应甲骨文摹写内容区的多列，但切割时却产生了漏切或列黏连（多列合并成一列）问题。

(2) 错字：书中的字在切割后出现了遗漏、不完整、冗余（将边框误认为文字，出现了多余的切割内容）、黏连（独立的两个字被认为是一个字）或割裂（一个字被切成了两部分）问题。

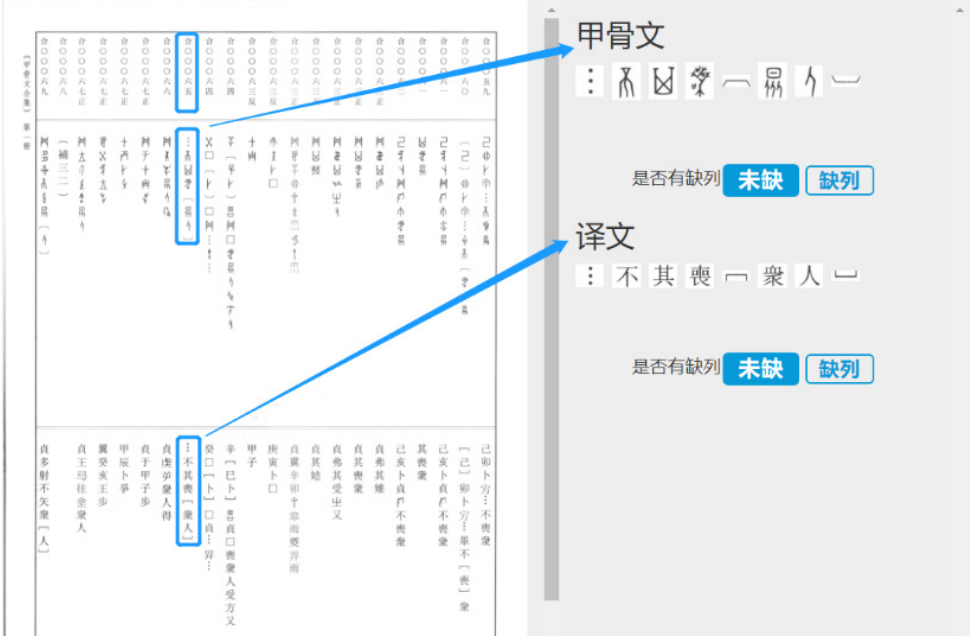
(3) 其他问题：如图片与书中内容不匹配、切割后的信息乱序等不包含于“缺列”“错字”的问题。

3.标注方法

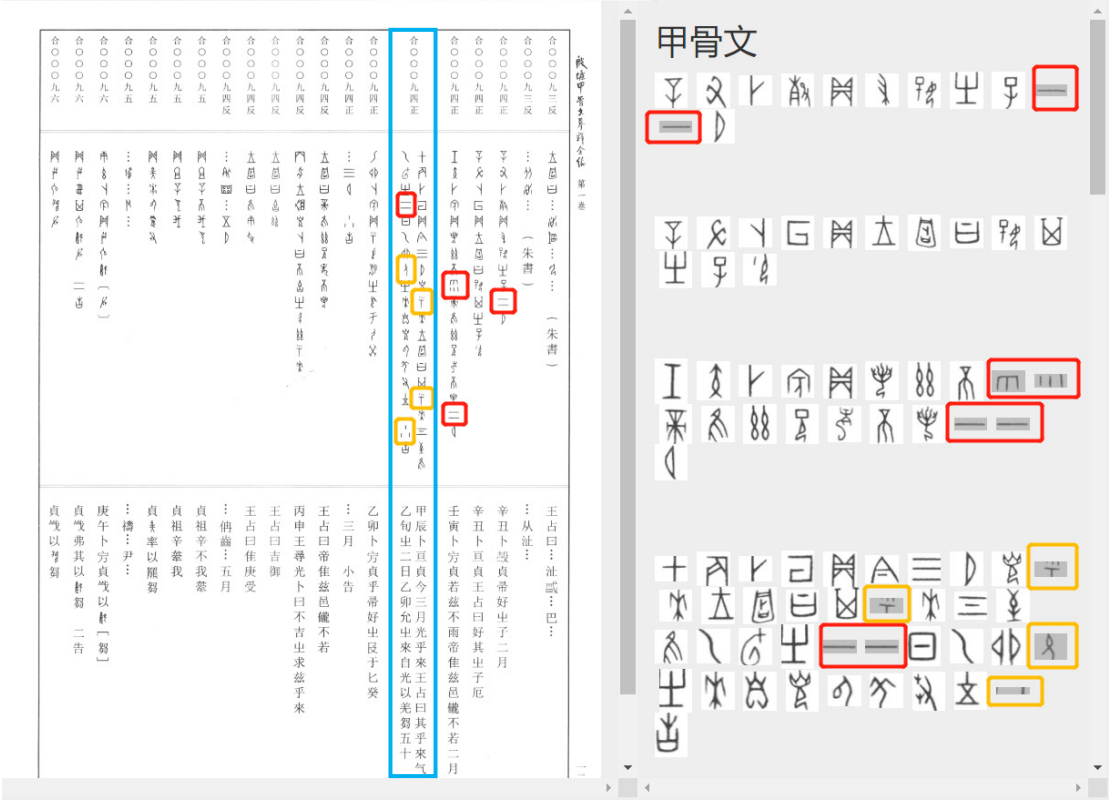
总体而言，对于每片甲片，如果甲骨文或译文出现“缺列”问题，则在标注区（上文绿框部分）选择“缺列”标签，否则选择“未缺”（请注意甲骨文和译文的**默认选项均为“缺列”**）；如果出现了“错字”问题，则直接点击标注区中具体的甲骨文/汉字（如果出现了某些字缺失的问题，请点击位于缺失字前后的两个字并提交文字反馈），被选中的文字将由白底变为灰底；如果出现其他问题，请点击工具栏中的“提交反馈”，以文字形式描述具体问题。

下面通过几个例子，具体讲解标注方法。

【例 1】以上文的标注页面为例，甲片编号为“65”，因此寻找《殷墟甲骨文摹释全编》页面中所有名为“合〇〇〇六五”的列，共有一列（见下图蓝色部分）。



对比后发现切割后的甲骨文出现了较多“错字”现象，将有问题的字选中（下图以前三列为例，其中红框为“错字”中的割裂问题，黄框为“错字”中的不完整问题，这些都归为“错字”问题，需要点击文字使之变成灰底）。需要注意的是，当甲片编号区中的一列对应甲骨文摹写内容区、翻译内容区中的多列时（如下图中蓝框部分）更容易出现“错字”和“缺列”等现象，需格外关注。



【例3】



切割内容中误将边框辨识为文字，因此多出了红框内容的部分，应将其选中标出。

【例 4】

甲骨文

译文

是否有缺列 **未缺** **缺列**

是否有缺列 **未缺** **缺列**

译文切割内容中遗漏了“二”字，因此选择位于其上下“羌”和“告”两字，并提交文字反馈“缺字”。

【例 5】



图中甲片编号为“199”，因此寻找《殷墟甲骨文摹释全编》页面中名为“合〇〇一九九”的列，为红框所示。对比发现右侧切割出来的甲骨文与译文均为名为“合〇〇二〇〇”的列的内容，并没有包含“合〇〇一九九”列的内容。因此勾选“缺列”，并提交文字反馈“甲片编号与切割内容不匹配，甲片编号为 199，切割内容为编号 200 的内容”。（如果切割后的结果既包含编号为 199 的内容又包含编号 200 的内容，则不勾选“缺列”，只提交文字反馈“甲片编号与切割内容不匹配，甲片编号为 199，切割内容为编号 199 和 200 的内容”）

