一、#2标注原则：按照句子结构划分#2。

1、第一步：判断字数

整句或分句超过七个字的，进行第二步划分。不超过七个字的，则不用标注#2。

2、第二步：划分连词

如果连词加上后面的内容超过七个字，连词后面要标注#2。一个字的连词，初始标注时在连词后标#2；如果最后需要合并，那就合并。

连词举例：

并列关系连词：和、跟、与、既、同、及、而、况、况且、何况、乃至等。

承接关系连词：则、乃、就、而、便、于是、然后、至于、说到、此外、像、如、一般、比方、接着等。

转折关系连词：却、虽然、但是、然而、而、偏偏、只是、不过、至于、致、不料、岂知等。

顺承关系连词：那么、原来、因为、由于、以便、因此、所以、是故、以致等。

选择关系连词：或、或者、还是、亦、非…即、不是…就是等。

假设关系连词：若、如果、若是、假如、只要、除非、假使、倘若、即使、假若、要是、譬如等。

比较关系连词：像、好比、如同、似乎、等于；不如、不及；与其…不如、若…则、虽然…可是等。

让步关系连词：虽然、固然、尽管、纵然、即使等。

递进关系连词：不但、不仅、而且、何况、并、且等。

条件关系连词：不管、只要、除非等。

目的关系连词：以、以便、以免、为了等。

如：虽然我没有那么好。——虽然=我没有那么好。

又慢慢徘徊了几步——又=慢慢徘徊了几步。

从=二零一六年的爆发=到=二零一七年降温，再到=二零一八年的冰点，北京土地市场=变化很大。（从……到，算不算连词）

3、第三步：划分结构

如果第二步结束，句子或分句，仍然超过七个字，那么优先把主语划分出来。主语划分完毕后，仍然超过7个字，接下来划分宾语，宾语划分后，仍

然超过7个字，按照句子结构，从大到小进行细分（当句子缺少主语，就直接划分宾语，宾语也没有，就直接细分其他结构）

例句1：就等=丈夫李天宇回来了

a、句子没有主语，所以跳过，直接划分谓语“丈夫李天宇回来了”

b、丈夫李天宇=回来了

主语

丈夫李天宇回来了，超过7个字，在该结构内进行划分，先划分出主语“丈夫李天宇”

划分结果：就等=丈夫李天宇=回来了

例句2：两颗硕大的泪珠=从眼角泌了出来

主语

主语为“两颗硕大的泪珠”，所以在主语后面划分#2。划分完后，“两颗硕大的泪珠”与“从眼角泌了出来”，都为7个字，所以不再进行划分。

划分结果：两颗硕大的泪珠=从眼角泌了出来

例句3：王硕这有理有据的论述从哲学观点阐发了爱的真谛。

a、王硕这有理有据的论述=从哲学观点阐发了爱的真谛。

主语

主语“王硕这有理有据的论述”划分完后，可以看到“王硕这有理有据的论述”与“从哲学观点阐发了爱的真谛”均超过7个字，所以要进一步划分

出宾语

b、王硕这有理有据的论述=从哲学观点阐发了=爱的真谛。

宾语

划分完宾语“爱的真谛”，我们发现“王硕这有理有据的论述”、“从哲学观点阐发了”仍然超过7个字，接下来划分按照句子结构再细分

c、王硕这有理有据的论述=从哲学观点=阐发了=爱的真谛。

“从哲学观点阐发了”划分完后，剩下的主语“王硕这有理有据的论述”超过7个字，再进行细分

d、王硕=这有理有据的=论述=从哲学观点=阐发了=爱的真谛。

1、主语 2、宾语

“王硕这有理有据的论述”，根据步骤，先划分主语“王硕”，超过7个字。再划分宾语“论述”。直到句子中没有超过7个字存在

划分结果：王硕=这有理有据的=论述=从哲学观点=阐发了=爱的真谛。

4、第四步：合并

当句子划分完成后，句中及句末出现单节奏时（1-2个字的词，以及3个字且无法再划分的词：1、为实词+虚词，2、为专有名词，如：感觉上，漂亮

的，矜持点，预测了，社科院，李小红 等等），视与前后部分语义的紧密程度和节奏平衡原则进行合并，当紧密程度相当时，考虑节奏平衡原则，

优先向节奏较少的合并（可以超过7个字）。

注意：

当单节奏出现在句首时，不作合并；例句：决心=不计一切代价，要得到力量。

当句首连续出现两个单节奏时，考虑节奏平衡原则，两个单节奏合并。例句：如果这时候=苏糖还没明白=自己穿越到了=星际时代的话

单节奏举例

想要吃：双节奏，想要+吃（两个实词）

有气势：双节奏，有+气势（两个实词）

你父亲：双节奏，你+父亲（两个实词）

去影响：双节奏，去+影响（两个实词）

云南的：单节奏，云南+的（实词+虚词）

世界上：单节奏，世界+上（实词+虚词）

每一天：单节奏，每+一天（虚词+实词）

第八届：双节奏，第+数词+量词

第八：单节奏，第+数词

例句1：就等=丈夫李天宇=回来了

单节奏 单节奏

合并后：就等=丈夫李天宇回来了

分析：句中单节奏为“就等”与“回来了”，“回来了”在句末，需要向前合并，“就等”在句首，我们可以不做处理。

最终结果：就等=丈夫李天宇回来了

例句2：两颗硕大的泪珠=从眼角泌了出来

分析：可以明显的看到句中没有单节奏，所以不需要合并

最终结果：两颗硕大的泪珠=从眼角泌了出来

例句3：王硕=这有理有据的=论述，从哲学观点=阐发了=爱的真谛。

单节奏 单节奏 单节奏

合并后：王硕=这有理有据的论述，从哲学观点阐发了=爱的真谛。

分析：因为“王硕”在句首，所以可以不作合并，“论述”和“阐发了”，是在句中的单节奏，故合并。

最终结果：王硕=这有理有据的论述，从哲学观点阐发了=爱的真谛。

例句分析

她喜欢彪德西的钢琴曲。

分主语：她=喜欢彪德西的钢琴曲。

分宾语：她=喜欢=彪德西的钢琴曲。

合并：她喜欢=彪德西的钢琴曲。 分析：“喜欢”为句中单节奏，在合并时，与前后的关系都十分紧密，这时，我们优先向节奏较少的“

她”合并

南松与贵州恒丰俱乐部之间已经就其买断事宜有了基本共识。

分主语：南松与贵州恒丰俱乐部之间=已经就其买断事宜有了基本共识。

分宾语：南松与贵州恒丰俱乐部之间=已经就其买断事宜有了=基本共识。

细分1：南松=与=贵州恒丰俱乐部=之间=已经就其买断事宜有了=基本共识。

细分2：南松=与=贵州恒丰俱乐部=之间=已经=就其买断事宜=有了=基本共识。

合并：南松=与贵州恒丰俱乐部之间=已经就其买断事宜=有了基本共识。

据美常驻联合国代表黑莉的话称。

没有主语，分宾语：据=美常驻联合国代表黑莉的话=称。

细分1：据=美常驻联合国代表黑莉的=话=称。

细分2：据=美=常驻联合国代表=黑莉的=话=称。

合并：据=美常驻联合国代表=黑莉的话称。

通过一个将死之人的嘴说出来，似乎更显得沧桑厚重，更有说服力。

分连词：通过一个将死之人的嘴说出来，似乎=更显得沧桑厚重，更有说服力。

分主语：通过=一个将死之人的嘴=说出来，似乎=更显得沧桑厚重，更有说服力。

没有宾语，细分：通过=一个将死之人的=嘴=说出来，似乎=更显得沧桑厚重，更有说服力。

合并：通过=一个将死之人的嘴=说出来，似乎更显得沧桑厚重，更有说服力。 分析：“嘴”、“通过”、“似乎”为三个单节奏，因为“

嘴”、“似乎”不在句首，所以需要合并。

计白是被民政局的工作人员叫醒的。

分主语：计白=是被民政局的工作人员叫醒的。

分宾语：计白=是被=民政局的工作人员=叫醒的。

细分：计白=是被=民政局的=工作人员=叫醒的。

合并：计白=是被民政局的=工作人员叫醒的。 分析：“是被”在语义上与“民政局的工作人员”更近，所以往后合并。之前会议讨论结果

为：计白=是被民政局的=工作人员=叫醒的。按照现在规则，“叫醒的”需要合并。20190617：没问题，合并。

班中只有稀稀拉拉的几个孩子没人接。

分主语：班中只有=稀稀拉拉的几个孩子=没人接。

细分：班中只有=稀稀拉拉的=几个孩子=没人接。 分析：由于划分结束后，没有1-2个字的词以及三个字单节奏，所以不需要进行合并。之

前会议讨论结果为：班中=只有稀稀拉拉的=几个孩子=没人接。

20190617：“班中”如果是状语，可以这样划分：班中只有=稀稀拉拉的=几个孩子=没人接。

二、#3标注规则(#3用$表示）

第一步：

按照字数，当一个句子小于20个字，则不需要标注#3

第二步：

当句子大于或等于20个字，则需要根据以下情况，把句子中的某些#2，标注成#3（一般每20个字，可以标注一个#3，#3倾向少标，模糊地

带不标注，#3两边的字数尽量匀称。）

1、在句子中插入逗号、冒号，顿号等标点符号而不影响句意的地方，可以标注#3

如：研究院院长杨现领=也认为=租赁市场=是未来的=发展方向

改成逗号不影响：研究院院长杨现领=也认为，租赁市场=是未来的=发展方向

标注结果：研究院院长杨现领=也认为$租赁市场=是未来的=发展方向

一群距离咱们=上万公里的外国人=居然从=近两千年历史的=重重迷雾里=准确把握住了=刘备这个人物的精髓，

改成逗号不影响:一群距离咱们=上万公里的外国人，居然从=近两千年历史的=重重迷雾里，准确把握住了=刘备这个人物的精髓，

标注结果：一群距离咱们=上万公里的外国人$居然从=近两千年历史的=重重迷雾里$准确把握住了=刘备这个人物的精髓，

一年四季=随时随地=都会有人=动脑筋=要弄死你的狗=偷掉你的狗！

插入顿号不影响：一年四季=随时随地=都会有人=动脑筋=要弄死你的狗$偷掉你的狗！

插入逗号不影响：一年四季=随时随地=都会有人=动脑筋$要弄死你的狗$偷掉你的狗！

标注结果：一年四季=随时随地=都会有人=动脑筋$要弄死你的狗$偷掉你的狗！

2、当句子中存在两个可升级为#3的#2时，先看成分，如果成分划分导致一部分小于七个字或者严重不均衡，则按照字数均衡标注

如：美国媒体警告=美对华贸易政策=将严重影响地方经济。》美国媒体警告美对华贸易政策$将严重影响地方经济。

朴成华曾说=韩国球队逢中国队必胜的日子=一去不复返了。》朴成华曾说韩国球队逢中国队必胜的日子$一去不复返了。

注：句子中出现用小括号表示的解释说明文字，数据筛选时去掉此类数据，后期针对此情况讨论具体的标注方法

身为大学同学的罗昊生，康雪妍（男女主角）等六人

《白鹿原》里有两个女人最为悲惨

字数均衡和语法之间需要衡量

**中英混韵律标注规范**

**标注**宗旨：以发音人的实际发音为准，对文本进行标注。

**1、韵律词标注：**用“#1”表示。

韵律词解释：口语中，紧密连在一起发音的几个音节的组合。

**韵律词的标注规则：**

当一个双音节或三音节词没有前后缀，也没有相邻的单音节词时，它会自成韵律词。如果有前缀或后缀，会与其共同组成一个韵律词。

单音节词往往会跟相邻的双音节词共同组成一个韵律词，除非该单音节词被明显拉长自成音步，或者与前后之间有明显的停延边界。

包含超过三个音节的词，往往会被分解成多个双/三音节韵律词，如果他有前后缀，或有相邻的单音节词，往往会融合进最前或最后的韵律词。

2**、语调短语：**用#4表示，即句末。

**3、韵律短语标注：**（一级韵短用“#3”，二级韵短用“#2”）

\*\*下列表格三列特征条件是与的关系。

|  |  |  |
| --- | --- | --- |
|  | **停顿时长** | **前后音高特征** |
| **韵律词边界** | **不停顿或从听感上不可察觉停顿** | **无** |
| **二级韵短边界**  **（两种条件其一成立即可）** | **较短的明显停顿、明显的静音段** | **音高不下倾或稍下倾，韵末不可做句末** |
| **可以感知停顿，但无明显的静音段。** | **音高不下倾或稍下倾，韵末不可做句末** |
| **一级韵短边界** | **有较长停顿** | **音高下倾比较完全，韵末可以作为句末** |

**（1）** **一级韵律短语：**用“**#**3”表示。明显停顿，可标为一级韵律短语。

**（2）** **二级韵律短语：**用“**#**2”表示。又叫复合韵律词，即 应在一口气说出的音节串。

**注意：**前后连标记、一级韵短标记、二级韵短标记的位置分为两种情况：

（a）如果标记处前后无标点，标记紧连前词。

例如：“中国#2文化#1悠久#4。”

（b）如果标记处有标点，标记需标在标点之前。

例如：“哪#1来的#3, 回#1哪去#4！”

**特例情况：**文本中“春天里的#1歌（#2 拍俊）”为错误现象，需更正为“春天里的 歌#2 （拍俊）”。

特例情况尤其注意：双引号、书名号、括号。

例如：读#1《论语#4》。

中英混韵律标注规则与以上中文韵律标注规则一致（英文单词可视为对应语境下的汉字）

标注样例：

外墙#2有#1很明显的#3no coffee#3 no work#3玻璃#1墙贴#4。

wai4 qiang2 you6 hen3 ming2 xian3 de5 / N OW1 / K AA1 . F IH0 / N OW1 / W ER1 K / bo1 li5 qiang2 tie1

可理解为：

外墙#2有#1很明显的#3不要咖啡#3 不要工作#3玻璃#1墙贴#4。

歌曲#1無限#1loop到#1ending#4。

可理解为：

歌曲#1無限#1循环到#1ending#4