对于以下类型的数据分词存在问题：

6WU91Q5LMC28EY47}JEJ$QC

新增方法1.如果数据是全中文且不包含分号，则根据空格分词，包含等的话将等删掉

[顾] [群] [郝晓弘] [巨圆圆]

0Y~$Q$}5PWHRAK`0CGRGR1I

祝乔;Yin, Yi-Xin;Cheng, Han-Qing;Chen, Xian-Zhong

新增方法2.首先用分号分词，如果分开后是纯中文，则不进行下一步分词

[祝乔] [Yin, Yi-Xin] [Cheng, Han-Qing] [Chen, Xian-Zhong]

如果分完后是全英文 + ‘,’ + ‘.’ + ’-’组成 则不进行下一步分词

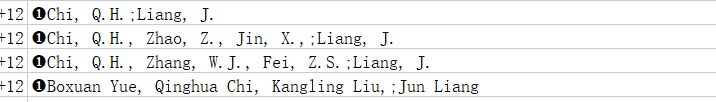
[祝乔] [Yin, Yi-Xin] [Cheng, Han-Qing] [Chen, Xian-Zhong]

[7QBM1BBJTT__}3~]LMB%ZR

如果存在括号，将括号内的内容删掉

H9{AL}@9G`PM~`93N`EJ$`I

使用新增方法2



???

4]5%)]]%]%P)3`%YS4C48LRVUO)IU1O_(K{RU7K6%JS)I6

使用新增方法2

分词方法:

data去掉’\u3000’

data是否是由中文+符号构成的？

是：data分词后全是中文

根据符号分词

否：data分词后由有中文也有英文

创建data\_pro

先将data用分号分为多个部分 data[i]

对每个data[i]单独操作

data[i]去空格

去除空格符号后字符串是否为全中文？

是

根据符号分词

否