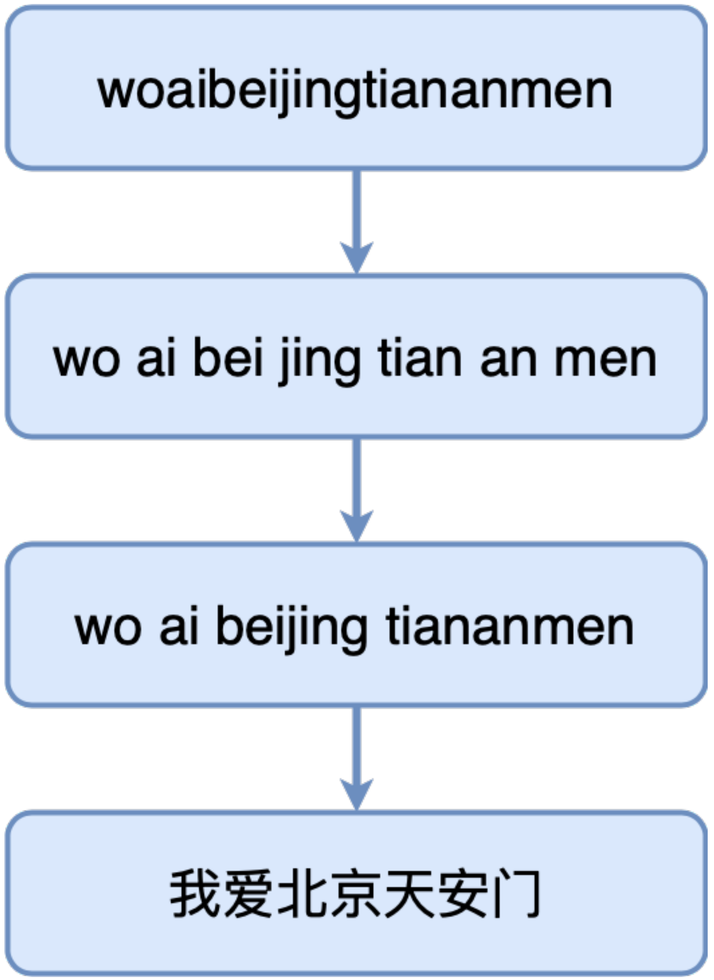
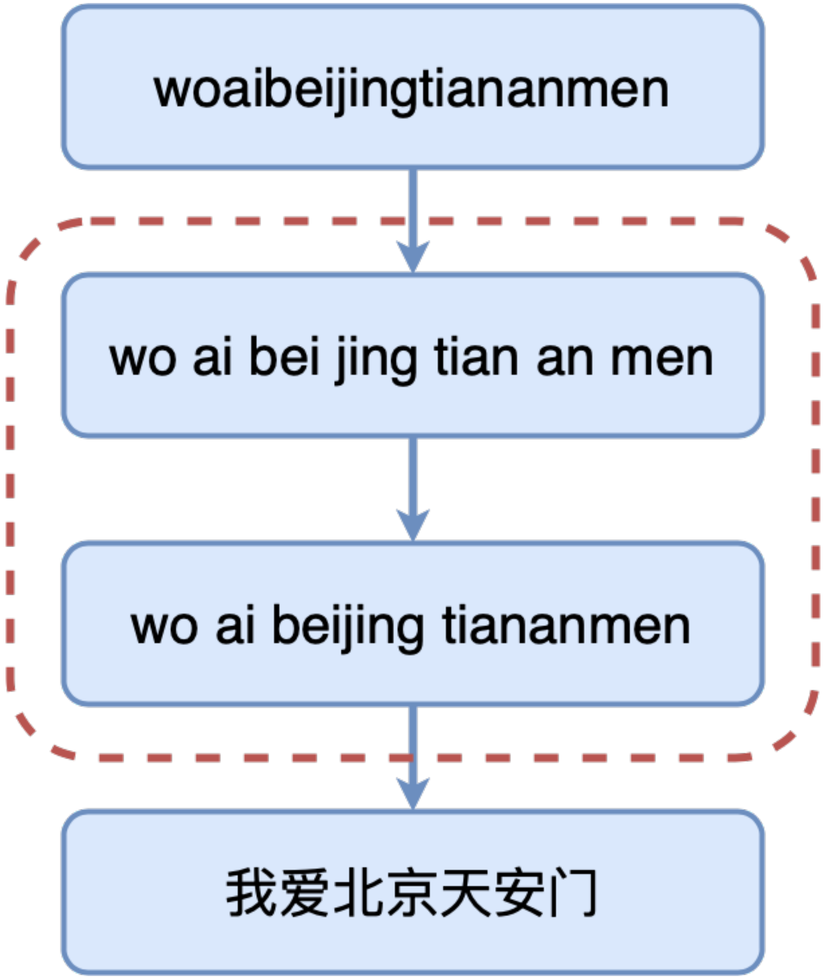
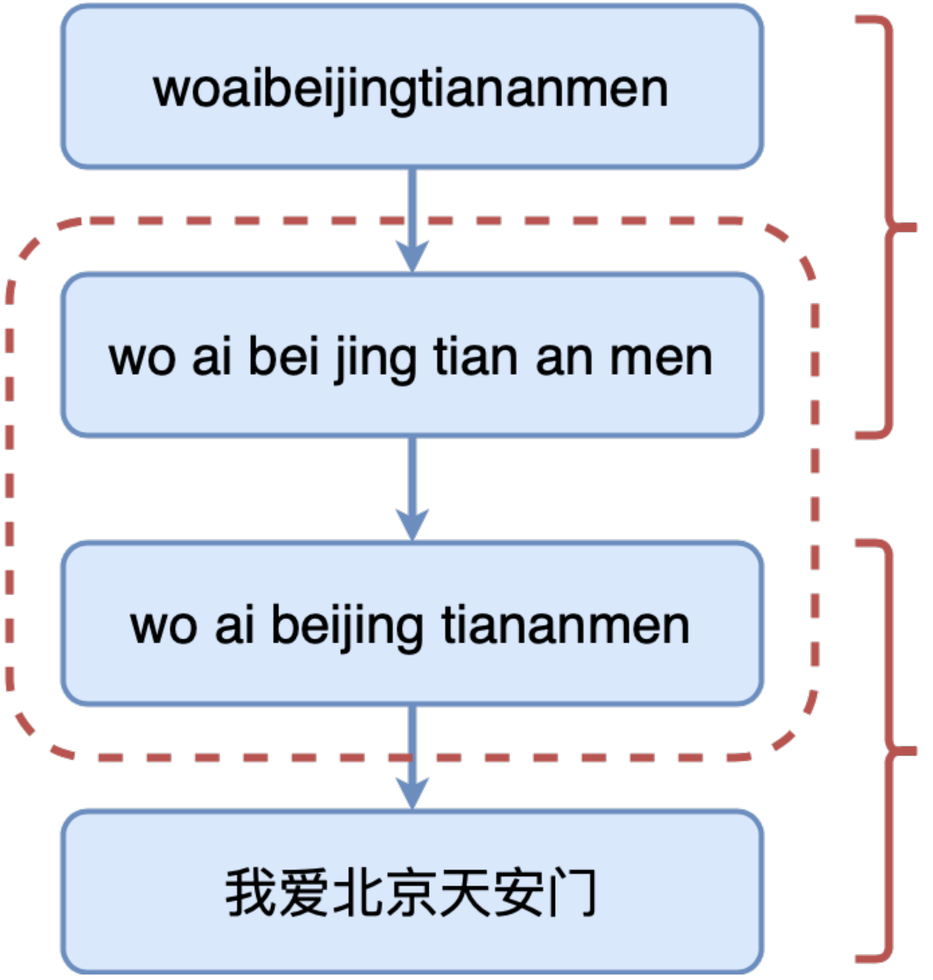
PY2HZ

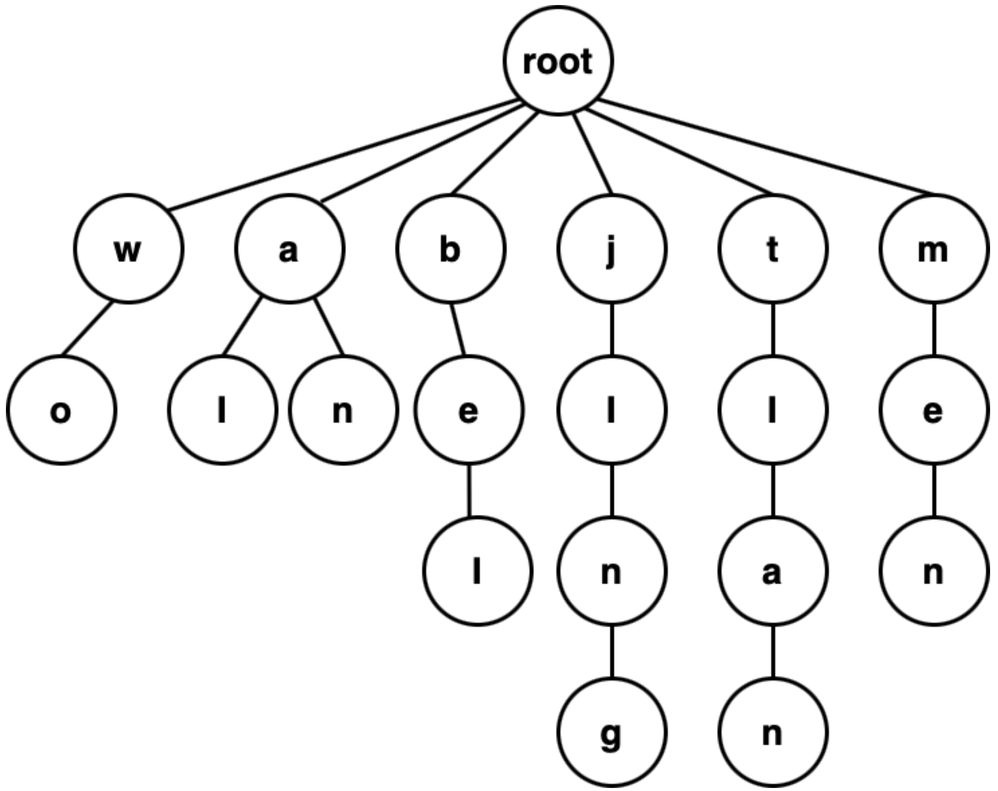
从拼音到汉字的中文输入法

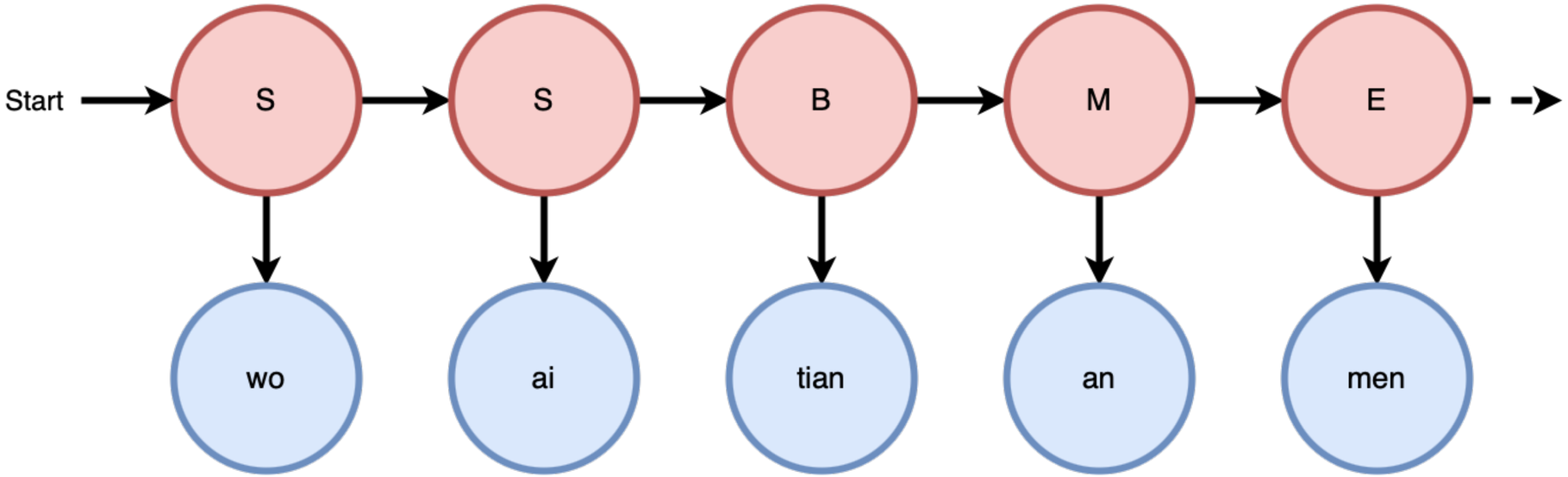
黄涛 林晗 吴大千 俞子逸 郑瑜婷

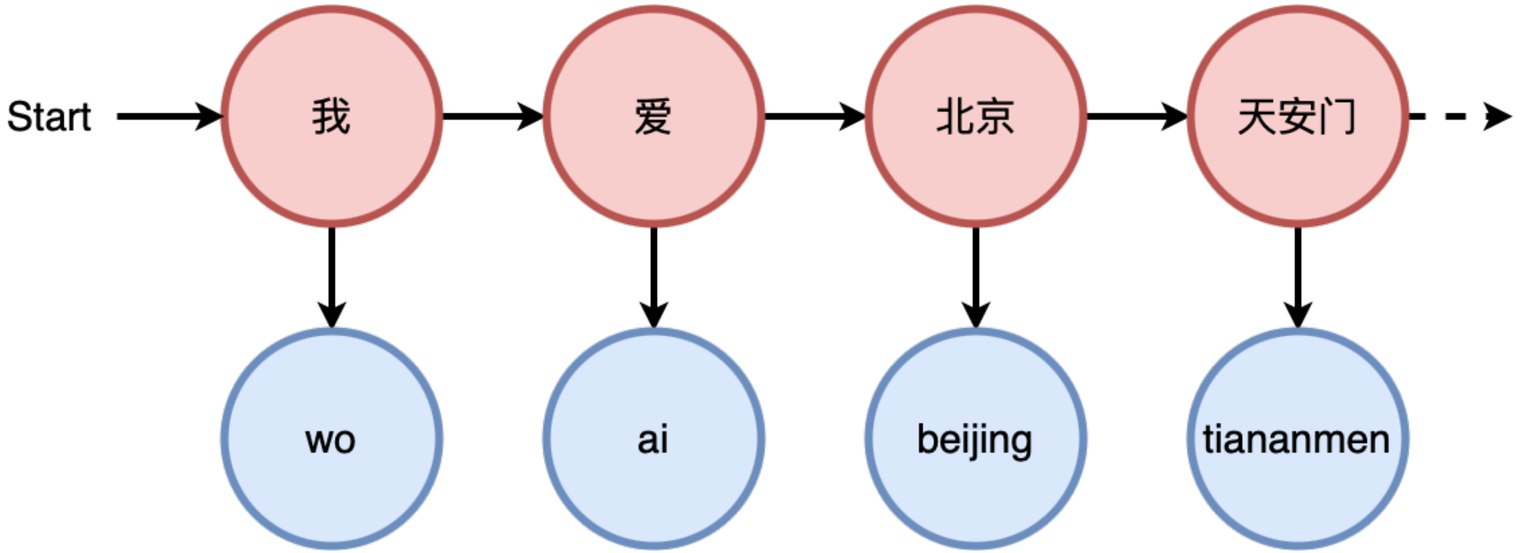

woaibeijingtiananmen
wo ai bei jing tian an men
wo ai beijing tiananmen
我爱北京天安门

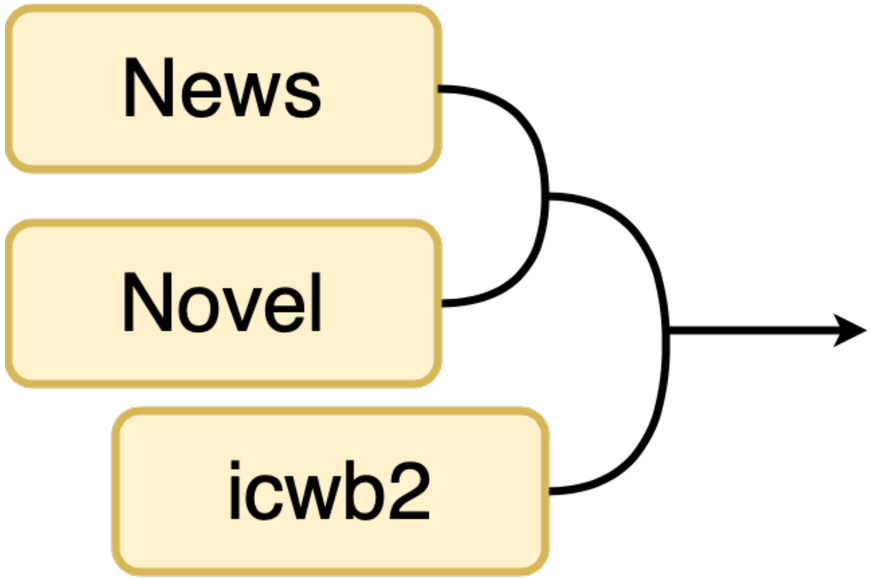

woaibeijingtiananmen
wo ai bei jing tian an men
wo ai beijing tiananmen
我爱北京天安门


woaibeijingtiananmen
wo ai bei jing tian an men
wo ai beijing tiananmen
我爱北京天安门

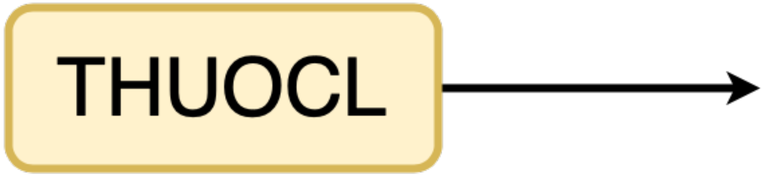

root
w
a
b
j
t
m
o
l
n
e
l
l
e
l
n
a
n
g
n


Start
S
S
B
M
E
wo
ai
tian
an
men


Start
我
爱
北京
天安门
wo
ai
beijing
tiananmen



我 爱 北京 天安门
理由 是 中 日 关系 正常化 了
她 松 了 一 口 气
服务 质量 得 不 到 保证
他 对 国际 高压 科学 与 技术 的 卓越 贡献
随后 两 人 在 我 后面 跟着
深入 学习 贯彻 党 的 十五大 精神

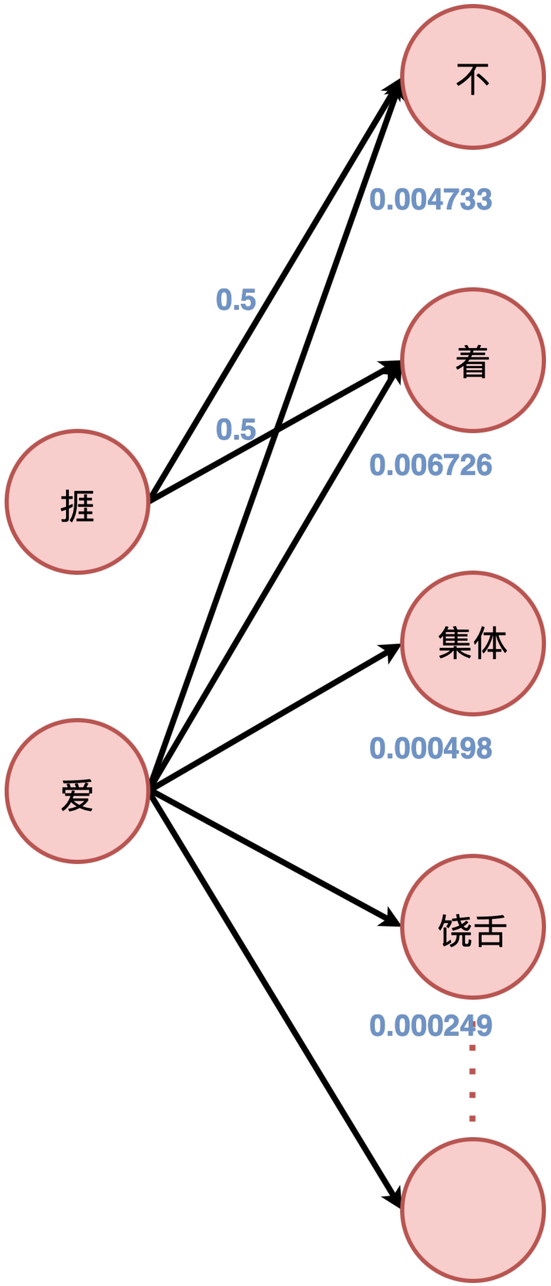


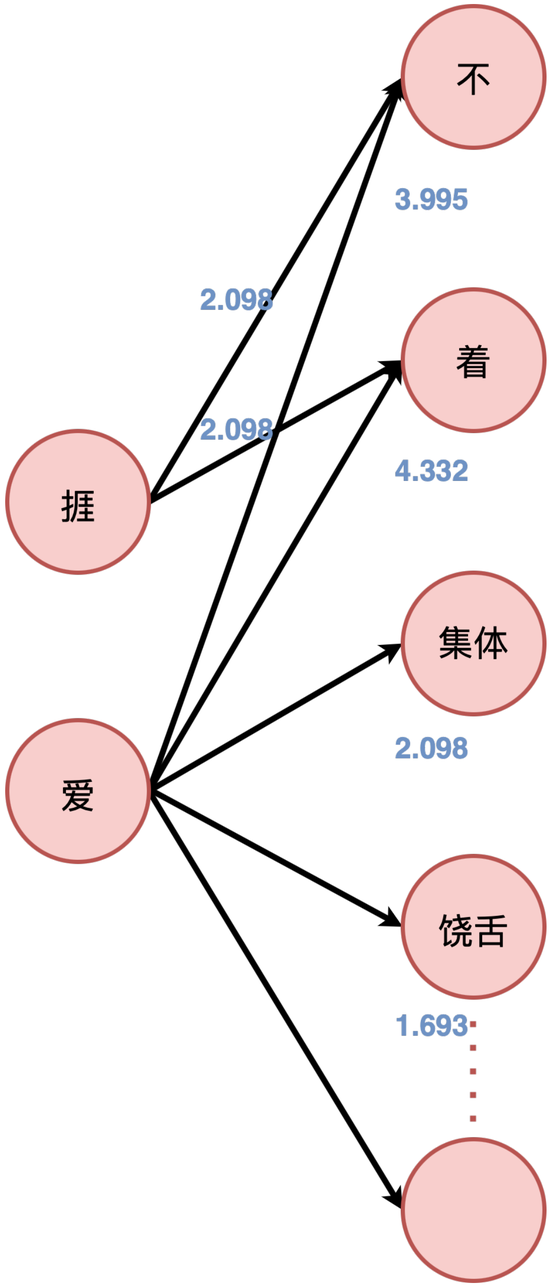
财政部 74243
比亚迪 147538
呼吸道 23457
小蜜蜂 22680

['我', '捱', '贝京', '天安门']

['我', '捱', '贝京', '天安门']

['在', '喙', '鞑', '缙', '糁', '馓', '你安', '谧', '勂', '婧', '芰', '砝', '搌', '掮', '婧', '的', '问', '题', '是']





$$P = log(Count + \lambda) + 1$$

Stage	Acc Type	Accuracy (%)		
		3-5	6-8	≥9
Seg	W	0.999	0.998	0.992
	S	0.995	0.980	0.935
Tok	W	0.623	0.653	0.655
	S	0.265	0.140	0.060
Trs	W	0.924	0.938	0.957
	S	0.770	0.700	0.655

Stages	Acc Type	Accuracy (%)		
		3-5	6-8	≥9
Seg+Tok+Trs	W	0.782	0.827	0.824
	S	0.310	0.230	0.085
Seg+Trs	W	0.786	0.793	0.801
	S	0.365	0.220	0.095

Number of tokens	Accuracy Type	Accuracy (%)		
		Top1	*Top3*	*Top5*
3-5	W	0.782	0.826	0.837
	S	0.310	0.375	0.395
6-8	W	0.827	0.855	0.863
	S	0.230	0.285	0.295
$\geq$9	W	0.824	0.854	0.864
	S	0.085	0.115	0.135

['在', '喙', '鞑', '缙', '糇', '馓', '你安', '谧', '劼', '婧', '芰', '砝', '搌', '掮', '婧', '的', '问', '题', '是']

['在', '会', '大', '进', '后', '三', '年', '世', '界', '经', '济', '发', '展', '前', '景', '的', '问', '题', '是']

['在', '回答', '今后', '三', '年', '世界', '经济', '发展', '前景', '的', '问题', '是']