

Homework 1: Solutions and Notes on Grading

Junkun's Notes Revised from a Previous Rubric
CS 539

Total points possible: 5.

1. Language Model, Possible points: 3

We automatically ran your unigram, bigram, and trigram models on the held-out blind dataset.

We took the best entropy values from all of these models to grade (this usually means your trigram model, but if it wasn't implemented or threw an error, we would default back to bigram/unigram/etc).

For Language Models. we assigned points as follows:

states	points
< 2.95	3
< 3.5	2
< 4	1.5
< 5	1
> 5	0

2. Possible points: 2 After this, we gave you 2 points for answering the questions on your report:

- 1 point for part 1.
- 1 point for part 2.

For using eval2.py, you need to convert the file format to be the same as test.txt. Otherwise your test results will exceed 0.9. You can use

```
sed -e 's/<s>>//g;s/<\\s>>//g;s/ //g;s/-/ /g'
```

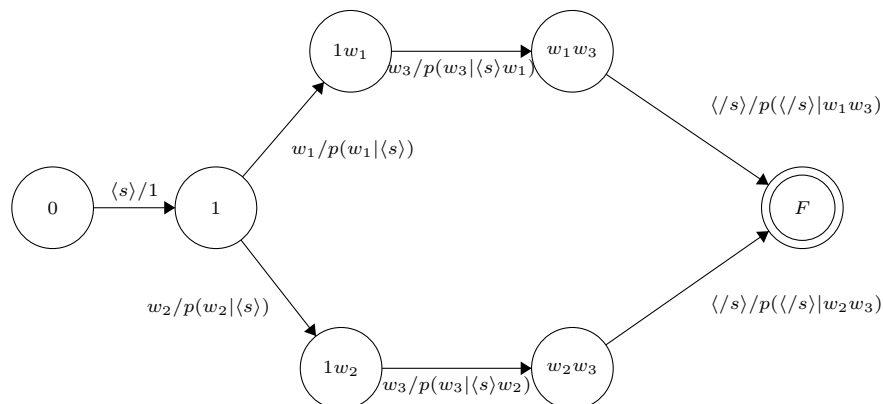


Figure 1: an example of trigram.