Homework 5

1. Word alignment

Log likelihood of 10 training iterations:

```
-180953.87966470644
-164750.30000882014
-158841.74109425058
-156524.87326182798
-155461.83106280334
-154904.7631101377
-154583.137481278
-154383.79320597925
-154253.312901054
-154164.06777532617
```

After training (10 iterations), t(f|e):

```
t(绝地, jedi) = 0.546989
t(机械人, droid) = 0.000082
t(原力, force) = 0.763994
t(原虫, midi-chlorians) = 0.000006
t(你, yousa) = 0.355140
```

Alignments for the first five lines of train.zh-en:

```
0-0 1-1
0-0 1-1 2-3 3-4
0-5 1-1 2-2 3-4 4-5
0-2 1-2 2-2 3-2 4-1 4-4
0-3 1-11 2-6 2-9 3-7 4-10 5-7 6-3 7-15 8-3
```

F1 score:

```
predicted alignments: 42413
true alignments:      56809
matched alignments:   26199
precision:            0.6177115507037937
recall:               0.4611769261912725
F1 score:             0.5280885287537038
```

2. Monotone decoding

First 10 lines of generated translations:

```
general the ship just before scanner
may droids with belly angle the then ship dissen i see been hah much
- roger sir
i the location show you have 0300 my caring aim
we're in commando rear been kenobi general
fighters capture battle
beware begin been let them from us comlink planet's
i've been post been - quick maybe ...
- i woolly help them
don't let them tie firepower us justice we have
```

The bleu score of this model was 1.57%.

```
BLEU: 0.015745322629675214
```

A possible problem with the model is the bigram approach, as sentences often need more context that just the previous word to accurately predict translations. An example would be with Jar Jar Bink's speech, in which the phrases are translated differently in english but appear the same in Chinese (ex: you —> yousa).

Additionally, only using the top ten translations ignores a lot of potential options for the model that could be useful, especially with languages that are so different as English and Chinese. Since a lot of context is lost in translation, a word in Chinese could have had many different representations in English. Some words could be translated to many different words, especially when considering the different dialects that appear in English speech that are not translated in Chinese, so keeping track of all the training data, especially in a sample size as small as the one we are working with in this particular assignment should improve the accuracy score.

My score using all the translations from the ibm_model1 (rather than just the top 10), as suggested in my second proposed modification was 2.40%, with the generated output in data/test.mod.out (as it took over an hour to run).

```
BLEU: 0.024000152582949807
```