CS 585, oct 29 2015 - Alignments
From Kevin Knight 1997, AI Magazine

Name: _____

These are sentence pairs in the (Centauri, Arcturan) made-up languages.
Learn the translation dictionary and word alignments.
The translation dictionary is mostly nonambiguous,

```
1a.  ok-voon ororok sprok .
1b.  at-voon bichat dat .
2a.  ok-drubel ok-voon anok plok sprok .
2b.  at-drubel at-voon pippat rrat dat .
3a.  erok sprok izok hihok ghirok .
3b.  totat dat arrat vat hilat .
4a.  ok-voon anok drok brok jok .
4b.  at-voon krat pippat sat lat .
5a.  wiwok farok izok stok .
5b.  totat jjat quat cat .
6a.  lalok sprok izok jok stok .
6b.  wat dat krat quat cat .
7a.  lalok farok ororok lalok sprok izok enemok .
7b.  wat jjat bichat wat dat vat eneat .
8a.  lalok brok anok plok nok .
8b.  iat lat pippat rrat nnat .
9a.  wiwok nok izok kantok ok-yurp .
9b.  totat nnat quat oloat at-yurp .
10a. lalok mok nok yorok ghirok clok .
10b. wat nnat gat mat bat hilat .
11a. lalok nok crrrok hihok yorok zanzanok .
11b. wat nnat arrat mat zanzanat .
12a. lalok rarok nok izok hihok mok .
12b. wat nnat forat arrat vat gat .

Translation dictionary:

ghirok - hilat              ok-yurp - at-yurp
ok-drubel - at-drubel       zanzanok - zanzanat
ok-voon - at-voon
```

Figure 2. Twelve Pairs of Sentences Written in Imaginary Centauri and Arcturan Languages.

An initial dictionary is given on the bottom left.

<u>New entries in the translation dictionary:</u>

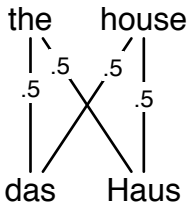
EM for Model 1

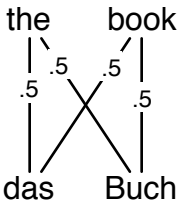
Here there are 4 words in both the foreign and English vocabularies. There are 3 sentences in the training data. Assume no NULLs. Initialize the translation parameters to be uniform:

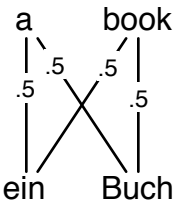
	das	ein	Buch	Haus
the	0.25	0.25	0.25	0.25
a	0.25	0.25	0.25	0.25
book	0.25	0.25	0.25	0.25
house	0.25	0.25	0.25	0.25

t(fIe)
Translation probs
Every row is one t(fIe) prob dist.

1a. E-step: Given t(f|e), calculate posterior alignments over the training data.
Each English word came from one German word in the sentence. Which?







p(Buch from "book") =

$$\frac{t(\text{Buch} \mid \text{book})}{t(\text{Buch} \mid \text{book}) + t(\text{Buch} \mid \text{a})}$$

1b. M-step: Given these posterior alignments,

(1) calculate fractional translation counts ……... (2) normalize into a new translation probability table.

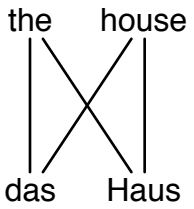
tcount(fIe): Translation COUNTS

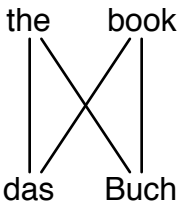
	das	ein	Buch	Haus
the				
a				
book				
house				

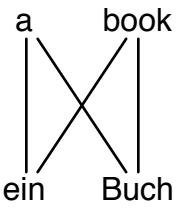
t(fIe): Translation PROBS

	das	ein	Buch	Haus
the				
a				
book				
house				

2a. E-step







2a. M-step

tcount(fIe): Translation COUNTS

	das	ein	Buch	Haus
the				
a				
book				
house				

t(fIe): Translation PROBS

	das	ein	Buch	Haus
the				
a				
book				
house				