Exercise 2.1

This language appears to be head-final, SOV. The alignment for the sentence pair would be $< 2, 1, 5, 4, 3 >$

Exercise 2.2

After first Iteration: $n_{\text{'eat','mange'}} = 0.\bar{6}$, $_{\text{'eat','mange'}} = 0.25$
After second Iteration: $n_{\text{'eat','mange'}} = 1.\bar{3}$, $\tau_{\text{'eat','mange'}} = 0.2\bar{7}$

(Did not use the null word.)
(Screenshots of the Excel file I used to arrive at these numbers are at the end of this document.)

Exercise 2.3

For a word aligned corpus, $n_{e,o}$ will only be the same as the word count of e if for every instance of e there is a word that aligns with it. One obvious reason that this is unlikely is that it presupposes that every pair of E/F sentences has equal length. We can see from exercise 2.2 that this is not always the case from French to English ('du' would align with nothing), and there are probably analogous examples from English to French.

Exercise 2.4

Totally. This is good because it allows the translation of idioms; i.e. a phrase like "kick the bucket" may best translate to a 3 word phrase that does not contain "kick," "the," or "bucket," but is a colloquial euphemism for death.

Exercise 2.5

Initially all τ's are equal. Let's call this value c. Then p_k for some arbitrary k will be a sum over τ's, so $p_k = dc$ for some k and some integer d. (p_k then is the value that *does* vary depending on the initial value of c) Then $n_{e_j f_k} = \frac{\tau_{e_j f_k}}{p_k} = \frac{c}{dc} = \frac{1}{d}$. You can see that at this point in the algorithm, the initial τ value cancels out; because k is arbitrary, this will apply to any $n_{e,f}$ value. The new τ values are derived from the newly incremented $n_{e,f}$ values, so they are likewise unaffected by our initial τ values.

Exercise 2.6

"The" is one of the most common English words, so any French word will align with 'the' very often. As a result, while the frequency of "the" aligning with "pain" is high, the *relative* frequency of "the" aligning with "pain" is more or less negligible; if "the" did show up in figure 2.6 as a translation for "pain", its values would all be incredibly close to zero.

Exercise 2.7

Equation 2.24 is false because our IBM Model 1 has gone through training with the EM algorithm, which yields word alignment probabilities for any e/f word pair. Therefore since the probabilities in 2.24 are conditioned on an E/F sentence pair, (i.e. since we know which French words need to be aligned with our English words) then not all sentence alignment probabilities will be equal. 2.23 is still true though, because given only an English sentence and no French, then IBM model 1's e/f word-pair alignment probabilities are useless, leaving all alignments equally likely.

tao1

	she	eats	bread	he	beef
Elle	1	1	1	1	1
mande	1	1	1	1	1
du	1	1	1	1	1
pain	1	1	1	1	1
boef	1	1	1	1	1
il	1	1	1	1	1

nef2

	she	eats	bread	he	beef
Elle	0.3333333333	0.3333333333	0.3333333333	0	0
mange	0.3333333333	0.6666666667	0.3333333333	0.3333333333	0.3333333333
du	0.3333333333	0.6666666667	0.3333333333	0.3333333333	0.3333333333
pain	0.3333333333	0.3333333333	0.3333333333	0	0
boef	0	0.3333333333	0	0.3333333333	0.3333333333
il	0	0.3333333333	0	0.3333333333	0.3333333333
n_eo	1.333333333	2.666666667	1.333333333	1.333333333	1.333333333

pk1S1

Elle	3
mande	3
du	3
pain	3

pk1S2

il	3
mange	3
du	3
boef	3

tao2

	she	eats	bread	he	beef
Elle	0.25	0.125	0.25	0	0
mande	0.25	0.25	0.25	0.25	0.25
du	0.25	0.25	0.25	0.25	0.25
pain	0.25	0.125	0.25	0	0
boef	0	0.125	0	0.25	0.25
il	0	0.125	0	0.25	0.25

pk2S1

Elle	0.625
mande	0.75
du	0.75
pain	0.625

pk2S2

mange	0.75
du	0.75
boef	0.625
il	0.625

nef3

	she	eats	bread	he	beef
Elle	0.7333333333	0.5333333333	0.7333333333	0	0
mange	0.6666666667	1.333333333	0.6666666667	0.6666666667	0.6666666667
du	0.6666666667	1.333333333	0.6666666667	0.6666666667	0.6666666667
pain	0.7333333333	0.5333333333	0.7333333333	0	0
boef	0	0.5333333333	0	0.7333333333	0.7333333333
il	0	0.5333333333	0	0.7333333333	0.7333333333
neo	2.8	4.8	2.8	2.8	2.8

tao3

	she	eats	bread	he	beef
Elle	0.2619047619	0.1111111111	0.2619047619	0	0
mange	0.2380952381	0.2777777778	0.2380952381	0.2380952381	0.2380952381
du	0.2380952381	0.2777777778	0.2380952381	0.2380952381	0.2380952381
pain	0.2619047619	0.1111111111	0.2619047619	0	0
boef	0	0.1111111111	0	0.2619047619	0.2619047619
il	0	0.1111111111	0	0.2619047619	0.2619047619