Modules used

1) nltk
2) numpy
3) pandas
4) json
5) collections (defaultdict)
6) nltk.translate (AlignedSent, Alignment, IBMModel, IBMModel1, IBMModel2)
7) nltk.translate.phrase_based.phrase_extraction

Data Structures and their relevance

Data Structure	Type	Relevance
fr	list	will contain a list of French sentences
en	list	will contain a list of English sentences
distinct_en	list	list of distinct English words
distinct_fr	list	list of distinct French words
t_val	dict	pair of words to translation probabilities
count	dict	counts of English word given a French Word
total	dict	stores the sum of translation probabilities corresponding to a French word in the sentence (normalized)
s_total	dict	normalization for an English word weighed by the prior translation probabilities
align	list	list of tuples where each tuple is a mapping from English to French word
ibm1	IBMModel1 object	Applies the IBM Model 1 algorithm to the bitext argument. Returns translation probabilities and alignment
ibm2	IBMModel2 object	Applies the IBM Model 2 algorithm to the bitext argument. Returns translation probabilities and alignment
trans_dict	dict	Map from word pairs to final translation probabilities for IBM Model 1

trans_dict2	dict	Map from word pairs to final translation probabilities for IBM Model 2
align_ibm1	list	Alignment returned by IBM Model 1 for any statement pair
align_ibm2	list	Alignment returned by IBM Model 2 for any statement pair
phrases	phrase_extraction object	Takes a pair of sentences (each translation of another) and their alignment as argument and returns a list of phrases and their translations.
count_fr_phrase	dict	count of French phrases
count_en_fr_phrase	dict	count of occurance of pairs of (English phrases, French phrases)
bitext	list	parallel corpus of English and French sentences
phrase_t	dict	pair of phrases to translations score
score_to_phrase_pair	list	list will contain tuples of the form (phrase_translation_score,(english_phrase, foreign_phrase))