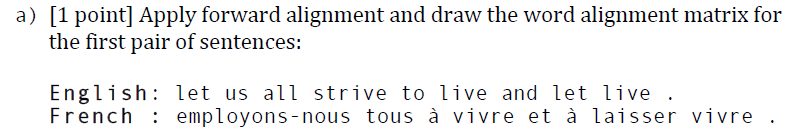
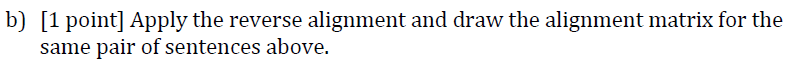
**NAME: XIANG LI, GWid: G47754486**

**Bonus Homework #4**

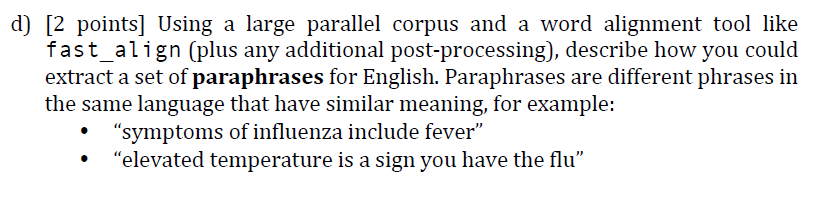
# Instructions:

This assignment is due on **Tuesday November 19, 2019, by 11:59 pm**.



***A: 0-0 2-1 4-2 5-3 6-4 4-5 7-6 8-7 9-8***

*A: 0-0 1-0 2-1 3-0 4-3 5-3 6-4 7-6 8-7 9-8*

*A:* *0-0 1-0 2-1 3-0 4-2 5-3 6-4 7-6 8-7 9-8*

*A:* Considering the functionality of tool fast\_align, “it can produce output in the widely-used i-j “Pharaoh format” where a pair i-j indicates that the *i*th word (zero-indexed) of the left language (by convention, the source language) is aligned to the *j*th word of the right sentence (by convention, the target language).”

Then, a very simple and straightforward idea is to pick up sentences from another language and using a large parallel corpus in our processing English text for word alignment. And, collecting all the good alignment of these parallel corpus because they are highly likely to be the **paraphrases**. And we can use minimum distance to quantify the goodness of alignment between your processing English text and one sentence from other language.

*Reference:*

1. [Chris Dyer](http://www.cs.cmu.edu/~cdyer), [Victor Chahuneau](http://victor.chahuneau.fr/), and [Noah A. Smith](http://www.cs.cmu.edu/~nasmith). (2013). [A Simple, Fast, and Effective Reparameterization of IBM Model 2](http://www.ark.cs.cmu.edu/cdyer/fast_valign.pdf). In Proc. of NAACL.