# CS F469 Assignment 3

## Design Document

Kaustubh Welankar 2016A7TS0095H
Gaurab Das Gupta 2016A3PS0255H
Aman Kumar Jain 2016A7PS0009H

# Results for data1.json:

task 1 CPU time: 0.018326888000000004
******IBM Model 1*******
en_sentence: ['the', 'house']
fr_sentence: ['la', 'maison']
nltk alignment: 0-0 1-1
predicted alignment: [(0, 0), (1, 1)]

en_sentence: ['the', 'flower']
fr_sentence: ['la', 'fleur']
nltk alignment: 0-0 1-1
predicted alignment: [(0, 0), (1, 1)]

en_sentence: ['the', 'blue', 'house']
fr_sentence: ['la', 'maison', 'bleu']
nltk alignment: 0-0 1-2 2-1
predicted alignment: [(0, 0), (1, 2), (2, 1)]

en_sentence: ['the', 'blue', 'flower']
fr_sentence: ['la', 'fleur', 'bleu']
nltk alignment: 0-0 1-2 2-1
predicted alignment: [(0, 0), (1, 2), (2, 1)]

en_sentence: ['blue', 'apple']
fr_sentence: ['pomme', 'bleu']
nltk alignment: 0-1 1-0
predicted alignment: [(0, 1), (1, 0)]

******IBM Model-2*******

en_sentence: ['the', 'house']
fr_sentence: ['la', 'maison']
nltk alignment: 0-0 1-1
en_sentence: ['the', 'flower']
fr_sentence: ['la', 'fleur']
nltk alignment: 0-0 1-1
en_sentence: ['the', 'blue', 'house']
fr_sentence: ['la', 'maison', 'bleu']
nltk alignment: 0-0 1-2 2-1
en_sentence: ['the', 'blue', 'flower']
fr_sentence: ['la', 'fleur', 'bleu']
nltk alignment: 0-0 1-2 2-1
en_sentence: ['blue', 'apple']
fr_sentence: ['pomme', 'bleu']
nltk alignment: 0-1 1-0

# Results for data2.json:

task 1 CPU time: 0.0466361100000000064
******IBM Model 1*******
en_sentence: ['the', 'house', 'is', 'in', 'india']
fr_sentence: ['la', 'maison', 'est', 'en', 'inde']
nltk alignment: 0-0 1-1 2-2 3-3 4-4
predicted alignment: [(0, 0), (1, 1), (2, 2), (3, 3), (4, 4)]

en_sentence: ['india', 'is', 'green']
fr_sentence: ["l'", 'inde', 'est', 'verte']
nltk alignment: 0-1 1-2 2-3
predicted alignment: [(0, 1), (1, 2), (2, 3)]

en_sentence: ['intelligent', 'is', 'the', 'girl']

fr_sentence: ['la', 'fille', 'est', 'intelligent']
nltk alignment: 0-3 1-2 2-0 3-1
predicted alignment: [(0, 3), (1, 2), (2, 0), (3, 1)]

en_sentence: ['in', 'india', 'is', 'the', 'green', 'girl']
fr_sentence: ['la', 'fille', 'verte', 'est', 'en', 'inde']
nltk alignment: 0-4 1-5 2-3 3-0 4-2 5-1
predicted alignment: [(0, 4), (1, 5), (2, 3), (3, 0), (4, 2), (5, 1)]

en_sentence: ['the', 'girl', 'has', 'a', 'green', 'chair']
fr_sentence: ['la', 'fille', 'a', 'une', 'chaise', 'verte']
nltk alignment: 0-0 1-1 2-2 3-3 4-5 5-4
predicted alignment: [(0, 0), (1, 1), (2, 2), (3, 3), (4, 5), (5, 4)]

en_sentence: ['a', 'girl', 'is', 'on', 'the', 'green', 'chair']
fr_sentence: ['sur', 'la', 'chaise', 'verte', 'est', 'une', 'fille']
nltk alignment: 0-5 1-6 2-4 3-0 4-1 5-3 6-2
predicted alignment: [(0, 5), (1, 6), (2, 4), (3, 0), (4, 1), (5, 3), (6, 2)]

en_sentence: ['a', 'chair', 'is', 'in', 'india']
fr_sentence: ['dans', "l'", 'inde', 'est', 'une', 'chaise']
nltk alignment: 0-4 1-5 2-3 3-0 4-2
predicted alignment: [(0, 4), (1, 5), (2, 3), (3, 0), (4, 2)]

en_sentence: ['the', 'chair', 'is', 'green']
fr_sentence: ['la', 'chaise', 'est', 'verte']
nltk alignment: 0-0 1-1 2-2 3-3
predicted alignment: [(0, 0), (1, 1), (2, 2), (3, 3)]

******IBM Model-2*******
en_sentence: ['the', 'house', 'is', 'in', 'india']
fr_sentence: ['la', 'maison', 'est', 'en', 'inde']
nltk alignment: 0-0 1-1 2-2 3-3 4-4

en_sentence: ['india', 'is', 'green']
fr_sentence: ["l'", 'inde', 'est', 'verte']
nltk alignment: 0-1 1-2 2-3
en_sentence: ['intelligent', 'is', 'the', 'girl']
fr_sentence: ['la', 'fille', 'est', 'intelligent']
nltk alignment: 0-3 1-2 2-0 3-1
en_sentence: ['in', 'india', 'is', 'the', 'green', 'girl']
fr_sentence: ['la', 'fille', 'verte', 'est', 'en', 'inde']
nltk alignment: 0-4 1-5 2-3 3-0 4-2 5-1
en_sentence: ['the', 'girl', 'has', 'a', 'green', 'chair']
fr_sentence: ['la', 'fille', 'a', 'une', 'chaise', 'verte']
nltk alignment: 0-0 1-1 2-2 3-3 4-5 5-4
en_sentence: ['a', 'girl', 'is', 'on', 'the', 'green', 'chair']
fr_sentence: ['sur', 'la', 'chaise', 'verte', 'est', 'une', 'fille']
nltk alignment: 0-5 1-6 2-4 3-0 4-1 5-3 6-2
en_sentence: ['a', 'chair', 'is', 'in', 'india']
fr_sentence: ['dans', "l'", 'inde', 'est', 'une', 'chaise']
nltk alignment: 0-4 1-5 2-3 3-0 4-2
en_sentence: ['the', 'chair', 'is', 'green']
fr_sentence: ['la', 'chaise', 'est', 'verte']
nltk alignment: 0-0 1-1 2-2 3-3

# Parallel Corpus(data3.json) formed by us: (German-English)

fr: foreign sentence
en: english sentence


[
  {

```
        "fr": "das haus ist in indien",
        "en": "the house is in india"
    },
    {
        "fr": "indien ist grün",
        "en": "india is green"
    },
    {
        "fr": "ist das mädchen intelligent",
        "en": "intelligent is the girl"
    },
    {
        "fr": "in indien ist das grüne mädchen",
        "en": "in india is the green girl"
    },
    {
        "fr": "das mädchen hat einen grünen stuhl",
        "en": "the girl has a green chair"
    },
    {
        "fr": "ein mädchen ist auf dem grünen stuhl",
        "en": "a girl is on the green chair"
    },
    {
        "fr": "ein stuhl ist in indien",
        "en": "a chair is in india"
    },
    {
        "fr": "der stuhl ist grün",
        "en": "the chair is green"
    }
]
```

True Alignments:

1. 0-0 1-1 2-2 3-3 4-4
2. 0-0 1-1 2-2
3. 0-3 1-0 2-1 3-2
4. 0-0 1-1 2-3 3-3 4-4 5-5
5. Since “das” is absent “Auf dem” menas “on the”
6. 0-0 1-1 2-2 3-3 4-4
7. 0-0 1-1 2-2 3-3

# Results for data3.json:

task 1 CPU time: 0.027670662999999984
******IBM Model 1*******
en_sentence: ['the', 'house', 'is', 'in', 'india']
fr_sentence: ['das', 'haus', 'ist', 'in', 'indien']
nltk alignment: 0-0 1-1 2-2 3-3 4-4
predicted alignment: [(0, 0), (1, 1), (2, 2), (3, 3), (4, 4)]

en_sentence: ['india', 'is', 'green']
fr_sentence: ['indien', 'ist', 'grün']
nltk alignment: 0-0 1-1 2-2
predicted alignment: [(0, 0), (1, 1), (2, 2)]

en_sentence: ['intelligent', 'is', 'the', 'girl']
fr_sentence: ['ist', 'das', 'mädchen', 'intelligent']
nltk alignment: 0-3 1-0 2-1 3-2
predicted alignment: [(0, 3), (1, 0), (2, 1), (3, 2)]

en_sentence: ['in', 'india', 'is', 'the', 'green', 'girl']
fr_sentence: ['in', 'indien', 'ist', 'das', 'grüne', 'mädchen']

nltk alignment: 0-0 1-1 2-2 3-3 4-4 5-5
predicted alignment: [(0, 0), (1, 1), (2, 2), (3, 3), (4, 4), (5, 5)]

en_sentence: ['the', 'girl', 'has', 'a', 'green', 'chair']
fr_sentence: ['das', 'mädchen', 'hat', 'einen', 'grünen', 'stuhl']
nltk alignment: 0-0 1-1 2-3 3-3 4-4 5-5
predicted alignment: [(0, 0), (1, 1), (2, 2), (3, 2), (4, 4), (5, 5)]

en_sentence: ['a', 'girl', 'is', 'on', 'the', 'green', 'chair']
fr_sentence: ['ein', 'mädchen', 'ist', 'auf', 'dem', 'grünen', 'stuhl']
nltk ibm model 1 alignment failed
predicted alignment: [(0, 0), (1, 1), (2, 2), (3, 3), (4, 3), (5, 5), (6, 6)]

en_sentence: ['a', 'chair', 'is', 'in', 'india']
fr_sentence: ['ein', 'stuhl', 'ist', 'in', 'indien']
nltk alignment: 0-0 1-1 2-2 3-3 4-4
predicted alignment: [(0, 0), (1, 1), (2, 2), (3, 3), (4, 4)]

en_sentence: ['the', 'chair', 'is', 'green']
fr_sentence: ['der', 'stuhl', 'ist', 'grün']
nltk alignment: 0-0 1-1 2-2 3-3
predicted alignment: [(0, 0), (1, 1), (2, 2), (3, 3)]

******IBM Model-2*******
en_sentence: ['the', 'house', 'is', 'in', 'india']
fr_sentence: ['das', 'haus', 'ist', 'in', 'indien']
nltk alignment: 0-0 1-1 2-2 3-3 4-4
en_sentence: ['india', 'is', 'green']
fr_sentence: ['indien', 'ist', 'grün']
nltk alignment: 0-0 1-1 2-2
en_sentence: ['intelligent', 'is', 'the', 'girl']
fr_sentence: ['ist', 'das', 'mädchen', 'intelligent']
nltk alignment: 0-3 1-0 2-1 3-2

en_sentence: ['in', 'india', 'is', 'the', 'green', 'girl']
fr_sentence: ['in', 'indien', 'ist', 'das', 'grüne', 'mädchen']
nltk alignment: 0-0 1-1 2-2 3-3 4-4 5-5
en_sentence: ['the', 'girl', 'has', 'a', 'green', 'chair']
fr_sentence: ['das', 'mädchen', 'hat', 'einen', 'grünen', 'stuhl']
nltk alignment: 0-0 1-1 2-2 3-3 4-4 5-5
en_sentence: ['a', 'girl', 'is', 'on', 'the', 'green', 'chair']
fr_sentence: ['ein', 'mädchen', 'ist', 'auf', 'dem', 'grünen', 'stuhl']
nltk ibm model 2 alignment failed
en_sentence: ['a', 'chair', 'is', 'in', 'india']
fr_sentence: ['ein', 'stuhl', 'ist', 'in', 'indien']
nltk alignment: 0-0 1-1 2-2 3-3 4-4
en_sentence: ['the', 'chair', 'is', 'green']
fr_sentence: ['der', 'stuhl', 'ist', 'grün']
nltk ibm model 2 alignment failed

At times nltk ibm models failed to print an alignment, because a valid mapping wasn’t present. But our alignment code maps the words to the most likely available word of the target sentence.