

## Report on Machine Translation

### Problem 1 :

implement the direct MT system, where you will need to collect your own data.

### Problem Statement

Implement the direct MT system..

### Language chosen

I chose **Spanish**. So will be translating sentences from Spanish to English.

Spanish has a two-gender noun system and is primarily an SVO (Subject Verb Object) language. However, it is quite common to leave out the subject entirely if the subject can be deduced from the sentence's context or the verb

conjugation. Spanish has a two-gender noun system and is primarily an SVO (Subject Verb Object) language. However, it is quite common to leave out the

subject entirely if the subject can be deduced from the sentence's context or the verb conjugation. That is particularly confusing when translating from Spanish to English because the translator has to infer the subject. Another exception to the SVO structure is that when pronouns are used as subjects, Spanish uses the SOV (Subject Object Verb) order e.g. "Raj lo vi" means "Raj loves her," but the direct translation is "Raj her loves."

### Data Used

#### Test Corpus

A corpus of 15 sentences in Spanish language.

Nicolás Maduro critica a Israel y a la canalla mediática que lo apoya.

El costo de mantener estas tendencias demográficas es insostenible, pero los servicios digitales podrían ser parte de la solución que busca la sociedad.

Aunque esto parece ciencia ficción, está a punto de volverse una realidad.

En un futuro no muy distante, recibirás un diagnóstico y cura completa desde tu smartphone, incluso antes de notar que estás enfermo.

La tecnología digital está lista para transformar radicalmente al sector de la salud y el bienestar.

En el camino, nos ayudará a superar algunos de los retos más significativos que enfrentamos.

A medida que las personas mayores representan una mayor parte de la población, la prevalencia de problemas de salud de largo plazo aumentará.

Esto causará una mayor carga de costos y presionará a los sistemas de salud a acomodar a una mano de obra que envejece.

Una tendencia que ha capturado la imaginación de muchos es el "hacking del cuerpo" o el entendimiento del "ser cuantificado".

Las canciones de Brown han sido sampleadas por multitud de artistas de hip-hop.

Muchos moldes de letras para imprimir que te servirán para hacer carteles o utilizar en logos, letreros o cualquier otro motivo para utilizar moldes de letras.

El médico solo necesita una pequeña muestra de saliva para hacer la prueba de ADN.

Cuba de duelo por el Chávez que acompañó a Fidel "como un hijo verdadero"

Probamos toda clase de vinos en la feria gastronómica del pasado fin de semana.

envíe una muestra de sus ilustraciones

Sources:

<https://www.vix.com/es/tecnologia/181819/el-superlaser-de-star-wars-deja-de-ser-ciencia-ficcion-para-convertirse-en-una-brillante-realidad>

<https://dguinstation.wordpress.com/2011/11/12/la-ciencia-ficcion-y-la-divulgacion-cientifica/>

<http://pulsodelospueblos.com/la-canalla-mediatica-y-la-violencia-en-venezuela-por-atilio-boron/>

<http://www.hispantv.com/noticias/venezuela/231787/maduro-condena-canal-antena-3-espana-guerra-rajoy-propaganda>

<http://www.periodistadigital.com/inmigrantes/gobiernos/2013/03/05/nicolas-maduro-el-cancer-de-hugo-chavez-es-un-ataque-de-los-enemigos.shtml>

### **Methodology:**

#### **POS Tagging:**

##### **Reorder Noun1 of Noun2:**

To create compound noun in Spanish is to put two nouns together using some form of the word 'de,' which means 'of.'

For example, "problemas de salud" literally translates to

"problems of health," though it simply means "health problems." However, in English, compound

nouns are created by just putting the two parts together without anything in between. The first

part is typically a descriptor and the second part usually identifies the object itself e.g "health

sector." Thus in order to fix this, let's rearrange non-proper nouns that have the word 'of' in between, so in my text, "hack of body" became "body hack." Although this rule works for a

lot of cases, there are some instances for which it does not work, such as "pocket of sunshine",

"plate of food", or "bale of hay". When NOUN2 refers to what's contained in NOUN1, the of stays.

Handle adjective nouns:

In Spanish, adjectives typically come after the noun that they describe, while in English, adjectives come before the noun. This rule fixes this order by switching nouns and adjectives e.g

“implications dramatic” becomes “dramatic implications.” It works for the majority of cases, but would not for a few Spanish adjectives that do come before the noun, such as “primera” (first).

Re-order in and adverbs:

In Spanish, the description of a particular noun or verb usually comes afterwards, not before. To fix this, I wrote a rule that moves the adverb clause before the subordinating conjunction and its subsequent clause. Looking at the phrase “in a future not very far,” the subordinating conjunction is “in,” the following clause is “a future,” and the adverb clause is “not very far.” this phrase becomes “not very far in a future.”

Remove article before proper noun:

A common practice in Spanish is to refer to proper nouns, usually people or places, using a definite article e.g “por el Chavez” literally translates to “for the Chavez,” which is unwieldy in English. So, I decided to remove the definite article before a proper noun, transforming “for the Chavez” to “for Chavez,” which is a much more fluent English phrase. Although there are situations for which this rule does not work (e.g I would want to keep the “the” in “the United States”), the net gain in faithfulness and fluency is greater than the net loss.

Switch adverbs and verbs:

In Spanish, adverbs often come after the verb that they describe, while in English, adverbs tend to come before the verb. Therefore, I wrote a rule that swaps verbs and adverbs. For example, “understanding immediately” should become “immediately understanding.” In general, based on Spanish grammar, this rule is a good one. However, this rule did not actually have any effect on my text because my POS tagger often mistagged verbs as nouns e.g “understanding” was tagged as a noun although in context, it was definitely a verb

swap reflexive verbs:

Reflexive verbs in Spanish are common, and usually refer to the object that is having the verb done to it. For example, lavarse is the verb to bathe, and when used in context, it says who is bathing who: “me lavo” means I wash myself. This is different from English, because I usually say the verb and then the object, “I pushed Sally”, “she loves chocolate”, etc. This rule remedies the difference by switching the order of the words to reflect the standard in English.

Handle an a:

For a particular noun in Spanish, which singular indefinite article to use depends on the noun's gender whereas in English, the choice of which singular indefinite article to use depends on whether or not the noun begins with a vowel. However, "una" and "un", Spanish's singular indefinite articles, both directly translate to "a," which sometimes results in grammatically incorrect phrases such as "a application." To fix this error, I implemented a rule that checks each instance of "a" or "an," its corresponding noun, and corrects the indefinite article if necessary. With this rule, "a application" becomes "an application."

Eliminate same words:

Sometimes, the output of a direct translation accidentally contains consecutive words that are the same, such as "for for." Exactly why this occurs depends on the specific text and context. One possible situation is that for a particular verb, the preposition that follows it is included in the dictionary entry but then is also included in the actual text e.g "is used for for." Most of the time, the second instance of the same word is extraneous and should be removed. Although I do not have any consecutive words that are the same in my text, I implemented this rule anyways because it is a situation that I have both often encountered while translating.

**POS Tagging for each of the English words:**

[('Nicolás', 'NNP'), ('Maduro', 'NNP'), ('criticism', 'NN'), ('to', 'TO'), ('Israel', 'NNP'), ('and', 'CC'), ('to', 'TO'), ('the', 'DT'), ('despicable', 'JJ'), ('media', 'NNS'), ('that', 'IN'), ('that', 'DT'), ('which', 'WDT'), ('is', 'VBZ'), ('support', 'NN')]  
 [('the', 'DT'), ('cost', 'NN'), ('from', 'IN'), ('maintain', 'NN'), ('be', 'VB'), ('tendency', 'JJ'), ('demographic', 'JJ'), ('is', 'VBZ'), ('unstable', 'JJ'), ('but', 'CC'), ('the', 'DT'), ('service', 'NN'), ('finger', 'NN'), ('be', 'VB'), ('able', 'JJ'), ('to', 'TO'), ('being', 'VBG'), ('part', 'NN'), ('from', 'IN'), ('the', 'DT'), ('solution', 'NN'), ('that', 'WDT'), ('search', 'VBZ'), ('the', 'DT'), ('society', 'NN')]  
 [('although', 'IN'), ('this', 'DT'), ('appear', 'JJ'), ('science', 'NN'), ('fiction', 'NN'), ('this', 'DT'), ('to', 'TO'), ('dot', 'VB'), ('from', 'IN'), ('come', 'VBN'), ('back', 'RP'), ('a', 'DT'), ('reality', 'NN')]  
 [('in', 'IN'), ('a', 'DT'), ('future', 'NN'), ('not', 'RB'), ('very', 'RB'), ('far', 'RB'), ('you', 'PRP'), ('will', 'MD'), ('receive', 'VB'), ('a', 'DT'), ('diagnosis', 'NN'), ('and', 'CC'), ('priest', 'VB'), ('complete', 'JJ'), ('from', 'IN'), ('your', 'PRP\$'), ('smartphone', 'NN'), ('even', 'RB'), ('before', 'IN'), ('from', 'IN'), ('notice', 'NN'), ('that', 'IN'), ('be', 'VB'), ('sick', 'JJ')]  
 [('the', 'DT'), ('technology', 'NN'), ('finger', 'NN'), ('this', 'DT'), ('list', 'NN'), ('for', 'IN'), ('transform', 'NN'), ('radically', 'RB'), ('to', 'TO'), ('sector', 'NN'), ('from', 'IN'), ('the', 'DT'), ('health', 'NN'), ('and', 'CC'), ('the', 'DT'), ('well-being', 'NN')]

[('in', 'IN'), ('the', 'DT'), ('route', 'NN'), ('us', 'PRP'), ('help', 'VBP'), ('to', 'TO'), ('lead', 'VB'), ('some', 'DT'), ('from', 'IN'), ('the', 'DT'), ('challenge', 'NN'), ('more', 'RBR'), ('meaningful', 'JJ'), ('that', 'IN'), ('put', 'VBD'), ('face', 'NN'), ('to', 'TO'), ('face', 'VB')]  
 [('to', 'TO'), ('measure', 'VB'), ('that', 'IN'), ('the', 'DT'), ('persons', 'NNS'), ('ancestors', 'NNS'), ('represent', 'VBP'), ('a', 'DT'), ('older', 'JJR'), ('part', 'NN'), ('from', 'IN'), ('the', 'DT'), ('settlement', 'NN'), ('the', 'DT'), ('prevalence', 'NN'), ('from', 'IN'), ('problem', 'NN'), ('from', 'IN'), ('health', 'NN'), ('from', 'IN'), ('long', 'JJ'), ('period', 'NN'), ('grow', 'NN')]  
 [('this', 'DT'), ('cause', 'NN'), ('a', 'DT'), ('older', 'JJR'), ('freight', 'NN'), ('from', 'IN'), ('cost', 'NN'), ('and', 'CC'), ('put', 'VB'), ('pressure', 'NN'), ('on', 'IN'), ('to', 'TO'), ('the', 'DT'), ('system', 'NN'), ('from', 'IN'), ('health', 'NN'), ('to', 'TO'), ('seat', 'VB'), ('to', 'TO'), ('a', 'DT'), ('hand', 'NN'), ('from', 'IN'), ('work', 'NN'), ('that', 'IN'), ('grow', 'JJ'), ('older', 'JJR')]  
 [('a', 'DT'), ('tendency', 'NN'), ('that', 'WDT'), ('be', 'VB'), ('capture', 'VBP'), ('the', 'DT'), ('imagination', 'NN'), ('from', 'IN'), ('many', 'JJ'), ('is', 'VBZ'), ('the', 'DT'), ('hack', 'NN'), ('of', 'IN'), ('the', 'DT'), ('body', 'NN'), ('or', 'CC'), ('the', 'DT'), ('understanding', 'NN'), ('of', 'IN'), ('the', 'DT'), ('being', 'VBG'), ('quantified', 'VBD')]  
 [('the', 'DT'), ('songs', 'NNS'), ('from', 'IN'), ('brown', 'VBN'), ('have', 'VBP'), ('been', 'VBN'), ('sampled', 'VBN'), ('by', 'IN'), ('crowd', 'NN'), ('from', 'IN'), ('artists', 'NNS'), ('from', 'IN'), ('hip', 'NN'), ('hop', 'NN')]  
 [('many', 'JJ'), ('molds', 'NNS'), ('from', 'IN'), ('letter', 'NN'), ('for', 'IN'), ('print', 'NN'), ('that', 'IN'), ('you', 'PRP'), ('serve', 'VBP'), ('as', 'IN'), ('for', 'IN'), ('do', 'NN'), ('posters', 'NNS'), ('or', 'CC'), ('use', 'NN'), ('in', 'IN'), ('logo', 'JJ'), ('signs', 'NNS'), ('or', 'CC'), ('any', 'DT'), ('other', 'JJ'), ('reason', 'NN'), ('for', 'IN'), ('use', 'NN'), ('molds', 'NNS'), ('from', 'IN'), ('letter', 'NN')]  
 [('the', 'DT'), ('doctor', 'NN'), ('alone', 'RB'), ('needs', 'VBZ'), ('to', 'TO'), ('a', 'DT'), ('little', 'JJ'), ('samples', 'NNS'), ('from', 'IN'), ('saliva', 'NN'), ('for', 'IN'), ('do', 'VBP'), ('the', 'DT'), ('proof', 'NN'), ('from', 'IN'), ('DNA', 'NN')]  
 [('Cuba', 'NNP'), ('from', 'IN'), ('duel', 'NN'), ('by', 'IN'), ('the', 'DT'), ('Chavez', 'NNP'), ('that', 'WDT'), ('accompany', 'VBZ'), ('to', 'TO'), ('Fidel', 'NNP'), ('like', 'IN'), ('a', 'DT'), ('son', 'NN'), ('true', 'JJ')]  
 [('I', 'PRP'), ('sampled', 'VBD'), ('whole', 'JJ'), ('class', 'NN'), ('from', 'IN'), ('wines', 'NNS'), ('in', 'IN'), ('the', 'DT'), ('fair', 'JJ'), ('gastronomic', 'NN'), ('of', 'IN'), ('the', 'DT'), ('past', 'JJ'), ('end', 'NN'), ('from', 'IN'), ('week', 'NN')]  
 [('send', 'VB'), ('a', 'DT'), ('samples', 'NNS'), ('from', 'IN'), ('their', 'PRP\$'), ('illustrations', 'NNS')]

### Bi-gram Model:

In custom model I have used all the features of Bigram language model along with the tri-gram language model. The motivation is to use the best possible feature of different model to increase the accuracy. The logic of the custom model is explained below.

$$P(w_i | w_{i-2} w_{i-1}) = \frac{\text{count}(w_{i-2} w_{i-1} w_i) + 1}{\text{count}(w_{i-2} w_{i-1}) + |V_2|}$$

```
if count3 > 0: #Biigram
    value += math.log(count3)
    value -= math.log(self.bigramDict[(word_1, word_2)])
```

```

elif count2 > 0: # no trigram, but bigram exists
    value += math.log(0.4) + math.log(count2)
    value -= math.log(self.unigramDict[word_2])
else: # no trigram or bigram
    value += math.log(0.4) + math.log(self.unigramDict[word_3]+1)
    value -= math.log(self.total + (len(self.unigramDict)))

```

### Tri-gram Model:

To find **Tri-gram probability**. I fetch the frequency of tri-char from the training model. Along with the frequency of starting character from the training model. Then, evaluate the probability of tri-character. Using this tri-character probability I find the probability of the word to determine its language.

For eg: Let's take the same test word "equal"

$P(\text{equal}) = P(q | <s>e)P(u | eq)P(a | qu)P(l | ua)P(</s> | al)$

### Error Analysis

The errors I ran into while translating fell into four categories: those caused by a bad dictionary, those caused by POS tagging, those caused by differences in Spanish and English that I weren't able to account for, and those caused by oversimplified rules. The first place my translation had some errors was in the dictionary. Looking up the words was solely based on the first translation that showed up in WordReference, so some words that could've been translated better in context had glaring problems. For example, 'cura' was translated to 'priest', even though it was obvious that the intended meaning was 'cure'. Another example was 'digital', which was translated to 'finger' instead of the correct translation of 'technology', and 'duelo', which was translated to 'duel' instead of 'hurt'. All in all, I can't do much about the dictionary unless I take into account context or use our own knowledge of the context to write the word meanings.

The second category of errors was caused by bad POS tagging (I used the NLTK POS tagger), and generally made my rules less efficient, or included words that I did not want to be affected by the rule. For example, one of the rules I tried to implement was replacing 'of' + a verb with the "ing" form of the verb (e.g. of notice → noticing). Unfortunately, the POS tagger gave us some difficulty - many words that were verbs in context were marked as nouns (e.g. notice). What I then decided to do was ignore the original part of speech of the word, and instead add "ing" to the end of the word, and check if that was a verb. Unfortunately, adding "ing" apparently makes lots of things verbs. I had 'beinging', 'healthing', and 'lifing'

as verbs, and some of the nouns that I didn't want to change became verbs (e.g. long → longing). Overall, if the POS tagger was a little better, I could've used this rule, since it applied in many cases.

A second effect of a variable POS tagger was that some of my rules were diluted. For example, rule #9 for switching adverbs and verbs had a few examples that weren't picked up: "transform radically", for example, didn't make the cut because "transform" was marked as a noun by the POS tagger.

The last subset of problems were those caused by exceptions to the rules, which was a bit too simplistic in some cases. For example, the first rule (NOUN1 of NOUN2 → NOUN2 NOUN1) doesn't apply in cases where noun1 contains noun2 (e.g. bale of hay, plate of food, pocket of sunshine). Another overly basic rule was "be VERB" → "VERB", because it fails to take into account tense - I could "be running" or Jamie "has been running", and in those cases my code would just drop the first form of be.

### **Results:**

#### **Translated text : Spanish → English**

Nicolás Maduro criticism to Israel and to the despicable media that which is support  
the cost from maintain be tendency demographic is unstable but the service finger be able  
to being part from the solution that search the society  
although this appear science fiction this to dot from come back a reality  
not very far in a future will receive a diagnosis and priest complete from your smartphone  
even before from notice that be sick  
the technology finger this list for transform radically to sector from the health and the well-  
being  
more in the route help us to lead some from the challenge meaningful that put face to face  
to measure that the persons ancestors represent an older part from the settlement the  
prevalence from problem from health from long period grow  
this cause an older freight from cost and put pressure on to the system from health to seat  
to a hand from work that grow older  
a tendency that capture the imagination from many is the hack of the body or the  
understanding of the being quantified  
the songs from brown have been sampled by crowd from artists from hip hop  
many molds from letter for print that serve you as for do posters or use in logo signs or any  
other reason for use molds from letter  
the doctor alone needs to a little samples from saliva for do the proof from DNA  
Cuba from duel by Chavez that accompany to Fidel like a true son  
sampled I whole class from wines in the fair gastronomic of the past end from week  
send a samples from their illustrations



**Problem2:****IBM Model 1****Language choosen: French**

I chose French will be translating French to English using IBM model 1.

Re-arranging based on POS is additional feature implemented.

An IBM-M1 model consists of a finite set  $E$  of English words, a set  $F$  of French words, and integers  $M$  and  $L$  specifying the maximum length of French and English sentences respectively. The parameters of the model are as follows:

$t(f|e)$  for any  $f \in F, e \in E$ . The parameter  $t(f|e)$  can be interpreted as the conditional probability of generating French word  $f$  from English word  $e$ .

**Pseudocode:**

initialize  $P(f|e)$

for a number of iterations:

    set  $tcount(f, e)$  to 0 for all  $f, e$

    set  $total(e)$  to 0 for all  $e$

    for each sentence pair  $(F, E)$  in training corpus:

        for each unique word  $f$  in  $F$ :

$denom\_c = 0$

        for each unique word  $e$  in  $E$ :

$denom\_c += P(f|e) * F.count(f)$

        for each unique word  $e$  in  $E$ :

$tcount(f, e) += P(f|e) * F.count(f) * E.count(e) / denom\_c$

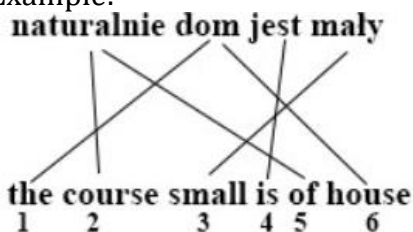
$total(e) += P(f|e) * F.count(f) * E.count(e) / denom\_c$

for each  $e$  in  $domain(total(:))$ :

    for each  $f$  in  $domain(tcount(:,e))$ :

$P(f|e) = tcount(f, e) / total(e)$

Example:



Sample  $P(f|e)$  probabilities:

Format{English text French Text Probability}



c&apos; it 0.297240553961447  
 c&apos; is 0.067335653122196  
 c&apos; already 0.0000000000000001  
 c&apos; history 0.0000000000000001  
 charges higher 0.0000000000000001  
 decision diamandouros 0.02210610441962

### **Sample alignemt numbers of translated Text:**

As one example, consider a case where  $l = 6$ ,  $m = 7$ , and

e = And the programme has been implemented  
 f = Le programme a ete mis en application

In this case the length of the French sentence,  $m$ , is equal to 7; hence I have alignment variables  $a_1, a_2, \dots, a_7$ . As one alignment (which is quite plausible), I could have  
 $a_1, a_2, \dots, a_7 = (2, 3, 4, 5, 6, 6, 6)$

Le  $\Rightarrow$  the  
 Programme  $\Rightarrow$  program  
 a  $\Rightarrow$  has  
 ete  $\Rightarrow$  been  
 mis  $\Rightarrow$  implemented  
 en  $\Rightarrow$  implemented  
 application  $\Rightarrow$  implemented

Note that each French word is aligned to exactly one English word. The alignment is many-to-one: more than one French word can be aligned to a single English word (e.g., mis, en, and application are all aligned to implemented). Some English words may be aligned to zero French words: for example the word And is not aligned to any French word in this example.

Note also that the model is asymmetric, in that there is no constraint that each English word is aligned to exactly one French word: each English word can be aligned to any number (zero or more) French words. We will return to this point later

### **Improving the IBM Model1 :**

Rearranging translations based on POS

Model	BLEU-1 scores	BLEU-2 score
IBM Model1	18.77	2.30
Re-arranging based on POS	43.21	7.14

Screenshot of BLEU score:

### IBM\_Model-1

```
C:\Users\Admin\Desktop\NLP\Assignment6\Raj_IBM1>py bleu_score.py IBM1_output.txt fr-en/dev/newstest2012.en
2997 sentences evaluated.
BLEU-1 score: 18.779430
BLEU-2 score: 2.301425

C:\Users\Admin\Desktop\NLP\Assignment6\Raj_IBM1>
C:\Users\Admin\Desktop\NLP\Assignment6\Raj_IBM1>
C:\Users\Admin\Desktop\NLP\Assignment6\Raj_IBM1>
```

After implementing the feature Rearranging translations based on POS:

```
C:\Users\Admin\Desktop\NLP\Assignment6\Raj_IBM1>py bleu_score.py my_output1.txt fr-en/dev/newstest2012.en
2995 sentences evaluated.
BLEU-1 score: 43.216025
BLEU-2 score: 7.145726

C:\Users\Admin\Desktop\NLP\Assignment6\Raj_IBM1>
```

### Problem3:

Analysis of Google Translate and my implementation

#### Direct MT

##### Input Test Corpus:

Nicolás Maduro critica a Israel y a la canalla mediática que lo apoya.

El costo de mantener estas tendencias demográficas es insostenible, pero los servicios digitales podrían ser parte de la solución que busca la sociedad.

Aunque esto parece ciencia ficción, está a punto de volverse una realidad.

En un futuro no muy distante, recibirás un diagnóstico y cura completa desde tu smartphone, incluso antes de notar que estás enfermo.

La tecnología digital está lista para transformar radicalmente al sector de la salud y el bienestar.

##### Output of my Implementation:

Nicolás Maduro criticism to Israel and to the despicable media that which is support the cost from maintain be tendency demographic is unstable but the service finger be able to being part from the solution that search the society although this appear science fiction this to dot from come back a reality not very far in a future will receive a diagnosis and priest complete from your smartphone even before from notice that be sick

the technology finger this list for transform radically to sector from the health and the well-being

**Output of Google Translate:**

Nicolás Maduro criticizes Israel and the media rabble that supports him.

The cost of maintaining these demographic trends is unsustainable, but digital services could be part of the solution society seeks.

Although this looks like science fiction, it is about to become a reality.

In the not too distant future, you will receive a complete diagnosis and cure from your smartphone, even before you notice that you are sick.

Digital technology is ready to radically transform the health and well-being sector.

**Statistical MT:****Test Corpus:**

Est-ce que sans occupation du sol , les sociétés pétrolières auront facilement accès à l'or noir libyen .

Peut-être pas autant et l'occident pourra finalement se vanter d'une protection désintéressée des droits de l'homme .

Pour peu que nous en empêchent ces adeptes de la charia .

Une nouvelle ère de crise commence .

La France et le reste de l'Europe sont dans le viseur des investisseurs .

**Output of my Implementation: (Statistical MT)**

without land cover oil companies may have good access to land libyan gold black

May be not so much west will finally boast it protection of the rights of the person man concerned

if we do not allow these followers of sharia

a new era of begins crisis

france and rest the of the country europe are in sights of investors

**Output of Google Translate:**

Without land cover, oil companies will have easy access to the land. Libyan black gold.

Maybe not so much and l & e; West will finally boast of it. protection of the rights of the person concerned man.

If we do not allow these followers of sharia.

A new era of crisis begins.

France and the rest of the country Europe are in the sights of investors.

**Where does Google do Better:**

Google Translate seems to do well with some of the areas of text mentioned above in the error

analysis, mostly because I haven't implemented as many features as they have. Google Translate also seems to have better rules.

One example is when Spanish has the word "be" + verb in it. Currently, my system just takes out the be, but Google Translate figures out how to convert the phrase into a present perfect one (

e.g. "be acquire" in my system is "acquire" and in Google is "has acquired").

In many cases Google translate does better job but in some case my translation model does fine job.

### **Where Google Translate and my implementation ties:**

my system's rules seems to match up to Google's pretty well. Here are examples of each rule

where both Google and I got around the same result:

original output → my output ("Google output")

1. noun1 'of' noun2 → noun2 noun1:

"style of life" → "life style" ("lifestyle")

2. 'be' verb → verb:

"be capture" → "capture" (" has captured")

3. noun adjective → adjective noun:

"implications dramatic" → "dramatic implications" ("dramatic implications")

4. 'in' word+ adverb+ → adverb+ 'in' word+:

"in a future not very far" → "not very far in a future" ("in a not too distant future")

5. 'at' proper\_noun → proper\_noun:

"criticism at israel" → "criticism israel" ("criticized israel")

6. article proper\_noun → proper\_noun:

"the Nicolás Maduro" → "Nicolás Maduro" ("Nicolás Maduro")

7. fixing a/an:

"a application" → "an application" ("an application")

8. removing consecutive words that are the same\*:

"for for" → "for" ("for")

\*note: this rule didn't occur after i changed the dictionary, so this is an older case.

9. verb adverb → adverb verb\*:

"transform radically" → "radically transform" ("radically transform")

\*note: this is dependent on how accurate my pos tagger is.

10. personal\_pronoun verb → verb personal\_pronoun - reflexive verbs  
“you serve” → “serve you” (“will serve”)

**Where my Implementation does better:**

There are two main place where my code is more clear than Google’s. The first is in the last line, which I translate as:

“for Nicolás Maduro that accompany at Fidel like a son true”

and Google Translates as:

“for Nicolás Maduro that accompany at Fidel like a son tru”

my translation is better because it differentiates Fidel and Nicolás. It is more readable, since Google has two proper nouns right next to one another. It also doesn’t have an extraneous “the”.

The second translation that I performed better on than Google is “being quantified” (they had

“be quantified”) because it makes much more sense in the context.

\*\*\*\*\***END**\*\*\*\*\*

---