* A comment on the language F that you chose. You should make a brief statement of particular challenges in translating your choice of F language to English (relative to other possible choices for F), and key insights about the language that you made use of in your strategies to improve your baseline MT system.
* If you are implementing the statistical MT system,
  + Your strategy to improve the baseline IBM model 1 system
* Your error analysis on the test set, including specific reference to what your code does and ideas for how further work might fix your remaining errors.
* The output of Google Translate.
* A comparative analysis commenting on your system's performance compared to Google Translate's. Show where the systems agree, what your system does better than Google Translate, and what Google Translate does better than your system.

For this assignment, we implemented a statistical machine translation system to translate Spanish to English. Luckily, English is very similar to Spanish seeing that English is a creole of semi-Romance ancestry (and Spanish is a Romance language), however, there are some key properties of Spanish that make it distinct from English that we took heed of.

For starters, there are predicate argument divergences (meaning there are differences in the ways the predicates and the arguments are related). One of these divergences is called head swapping. In English, the manner of motion is denoted on the verb and the direction of motion is described in the adverb/satellite. This makes English a Satellite-framed language. Yet in Spanish the verb typically marks the direction of motion, and the adverb/satellite generally denotes the manner – making Spanish a Verb-framed language. We can see this in the English example: “he ran in,” where the verb “ran” describes the manner of motion and the satellite “in” describes the direction of motion. In Spanish the correct translation would be, “entró corriendo,”which literally means in English “he entered running.” Note that the direction of motion is described with the verb, and manner with the satellite/adverb. Head swapping does not always occur though because English is a creole of romantic, Verb framed languages. For instance, “He danced off the stage” translates to “él bailó fuera del escenario”, where the verb in both English and Spanish denotes the manner of motion (danced), and the satellite marks direction (off the stage). This is important to our machine translation project because we must be aware that sometimes the parts of verb meaning are going to be moving around since we are translating from a Verb-framed language to a typically Satellite-framed language.

Another key property of Spanish that makes it distinct from English is argument swapping, another form of predicate argument divergences. In this case, the subject argument and object arguments swap in sentences of preference between English and Spanish translations. An example of this is how in English when I say, “I love homework,” the subject argument is myself, “I”, and the object argument (thing I love) is “homework”. But in Spanish, that same sentence translates to “Me encanta la tarea” – literally meaning “Homework is delightful to me”. “Homework” becomes the subject argument, and “me” becomes the object argument. Once more we must be careful of not only head swapping, but also argument swapping. Yet, this is a minor problem in comparison to lexical divergences.

There is a whole array of lexical divergences where the literal translation between Spanish and English doesn’t really flow fluently. There are Part of Speech divergences like the phrase “Tengo sed”, which in Spanish literally translates over to English as “I have thirst”. However, we would translate “Tengo sed” to “I am thirsty” (thereby switching the noun “thirst” to an adjective “thirsty”) so that it is more fluent in the English language. These Part of Speech divergences are most common. Additionally there are the lexical specificity divergences between Spanish and English where, for instance, Spanish attributes gender to plural pronouns and English does not. More specifically, English uses “they” as its plural pronoun, and Spanish uses both “ellos” and “ellas” depending on the gender of the overall group involved. Furthermore, beyond lexical specificity there is also semantic specificity, where depending on the context or how the word is used, different terms would be more appropriate. For example, whereas in English we say, “try” when we attempt to do some action or even taste a new food, Spanish speakers say “probar” for tasting new things, and “tratar” for attempting to do some action. To increase the fluency of our machine translation project, we must take these divergences into account and perhaps make special cases.