A Substitution Cipher is one in which letters are substituted for a different letter, to hide the information being sent. Since letters have a natural frequency, it is helpful to look at the natural letter frequency and the letter frequency of the encrypted text. This way you can get a start at what the possible substitution might be. To calculate for entropy, I used Shannon’s equation on each character to figure out how many bits they carried, and then summed all the entropies of the characters of a language together. All these calculations are shown in my excel file.

I decided to take a look at the letter frequencies for English, German and French. I picked German and French because they are geographically next to each other, yet their languages come from different origins French coming from Latin, and German and English coming from Germanic descent. Letter frequencies in different languages tend to differ, except in the Latin languages they appear to be relatively similar. Because their words are similar and their alphabets are as well then it makes sense that their frequencies are similar. Even though German has 4 additional letters and French has 16 additional letters compared to English, their entropies are within .01 of each other. Their entropies are 4.184417, 4.194538, and 4.184204, for English, German, and French respectively. I thought there would be a little bit more of a difference in the entropy, given the extra letters that French and German have. The additional letters are not actually base letters, as all three languages have the same 26 letter alphabet, German and French have special characters with added accents or umlauts or multiple letters together like the French Ae or the German β, written as ss in English. The first thing I noticed about the individual letter frequencies was that the most frequent letter for all three was e, at 12%, 16%, 14%, for English, German, and French. I was not surprised to see that many of the extra letters from German and French were not very high except for é, which was at 2.46%. It's interesting that most of the lower frequency letters are the same in all three, such as j, k, q, x, y and z. I would expect that at least some of the letters would be used more in other languages, however they aren’t. I find it interesting how a, I, n, r, s, and t, are all around the same for all three languages and are about the middle to higher frequency letters.

Overall the 3 languages are very similar. Even though they have different origins, over time they have evolved similarly, influencing each other to cause the similarity in letter frequencies. The differing alphabets played little to no role in the overall entropy of a language, as the extra letters are no more than 3.5% of the total French letters and no more than 1.5% of the total German letter. If the frequencies of all the letters were even then substitution ciphers would be much harder to crack, however no natural language does this. Since the entropies are all similar, the three languages are equally susceptible to a substitution cipher being cracked.

Website used for letter frequencies:

<http://practicalcryptography.com/cryptanalysis/letter-frequencies-various-languages/>