

## **Riley Skief**

### **Project 1**

**URL of github:** <https://github.com/rileyskief/csci335>

**URL of kaggle:**

<https://www.kaggle.com/code/riskief/hidden-markov-model-language-analysis/edit>

### **Accuracy:**

After making adjustments, this program performed very efficiently. With some sentences being quite short and including names of places, there are a few mistakes, but 13/16 sentences accurately guessed is moderate. Since only four languages could even be recognized (Spanish, English, French, and German), it is understandable that the four other languages the sentences were in (Icelandic, Italian, Turkish, and Portuguese) were all calculated incorrectly. 3 out of 4 of the Portuguese sentences were guessed to not be Spanish, which I found intriguing since Portuguese and Spanish are very similar languages.

Markov chains could be used to predict the probability of an event recurring when given events that happened at a particular event. When it detects those events, it could provide the likelihood of that initial event happening again.

The Hidden Markov Model creates a state for vowels and a state for consonants. Using the Kaggle notebook as a guide, it can be determined that the English, Spanish, French, and German languages have the same vowels. The exception is that the Hidden Markov Model thought that Spanish uses q as a vowel, and German uses c as a vowel- though I know this is not an accurate assumption by the HMM.

The fact that German was the probable language for English sentences is not a surprise given the similarities between the languages. Though it is interesting that the German sentences' probable languages were 50/50 with French and English.

Language	Sentence	Probable Language	French %	Spanish %	English %	German %
Spanish	Puedo ir al baño por favor	Spanish	3.641	0.99	1.599	1.042
	Yo soy de	French	0.713	0.299	0.002	3.895

	<b>Louisiana</b>					
	<b>Yo tengo hambre</b>	<b>Spanish</b>	1.65	0.964	0.035	1.121
	<b>Yo juego softball</b>	<b>English</b>	2.233	0.001	0.934	0.064
<b>English</b>	<b>I love sandwiches.</b>	<b>English</b>	0.009	0.003	0.976	0.0104
	<b>My favorite color is blue.</b>	<b>French</b>	0.991	2.39	0.009	8.318
	<b>I live in the United States.</b>	<b>English</b>	4.17333	1.1234	0.9993	6.541
	<b>I speak French.</b>	<b>English</b>	0.04829	0.0042	0.9469	4.4686
<b>French</b>	<b>Quelle heure est-il?</b>	<b>French</b>	0.998	0.0067	0.001	0.002
	<b>J'ai deux chats.</b>	<b>French</b>	0.99999	1.531	4.07423	1.471
	<b>Je déteste mon école.</b>	<b>French</b>	0.9999	6.7145	6.9946	7.9173
	<b>Hendrix est très belle.</b>	<b>French</b>	0.9997	2.3205	2.50377	5.652
<b>German</b>	<b>Mein Lieblingsfarbe ist hellrosa.</b>	<b>German</b>	9.519	2.565	0.021	0.98

	<b>Jeden Tag singe ich gern.</b>	<b>German</b>	3.61	5.29	0.001	0.997
	<b>Am Wochenende bin ich zur Party gegangen.</b>	<b>German</b>	French:4.964511	Spanish:2.3009	English:8.14583	German:0.9999
	<b>Ich habe zwei Brüder.</b>	<b>German</b>	1.806	1.3898	1.1080	0.999
<b>Icelandic</b>	<b>KEMUR ALLT MEÐ KALDA VATNINU</b>	<b>German</b>	3.38	6.57	0.03	0.97
	<b>RÚSÍNAN Í PYLSUE NDANUM</b>	<b>French</b>	0.373	0.0424	0.3246	0.2593
	<b>Hvað heitir þú?</b>	<b>French</b>	0.7175	2.7289	0.16300	0.11941
	<b>Hvernig hefur þú það?</b>	<b>English</b>	0.0015	1.865	0.977	0.031
<b>Turkish</b>	<b>Adınız nedir?</b>	<b>French</b>	0.93994	0.0038	0.05007	0.0060
	<b>Kaç yaşındasınız?</b>	<b>English</b>	0.1200	1.8622	0.85747	0.0223
	<b>Bir bira isterim</b>	<b>Spanish</b>	0.09814	0.5772	0.03844	0.2862

	<b>Yakında görüşmek üzere!</b>	<b>Spanish</b>	0.01714	1.3635	0.98275	German:1.0834
<b>Portuguese</b>	<b>Bom dia.</b>	<b>French</b>	0.7573	0.0425	0.1290	0.0710
	<b>Seu sorriso é lindo</b>	<b>French</b>	0.93170	0.0567	0.00726	0.0042
	<b>Vai pentear macacos !</b>	<b>French</b>	0.96325	0.0362	4.83172	7.375
	<b>Ficar com um melão</b>	<b>Spanish</b>	0.04666	1.7656	0.95040	0.0027
<b>Italian</b>	<b>la pesca e la mia frutta favorita</b>	<b>French</b>	0.88948	0.11051	1.44780	4.0869
	<b>mi piace molto a guidare la mia macchina bianca</b>	<b>Spanish</b>	8.64983	0.9999	1.16386	1.6450
	<b>la classe di matematica e molto difficile</b>	<b>French</b>	0.877	0.1224	3.265	4.21
	<b>ciò un sacco di</b>	<b>Spanish</b>	0.465	0.474	0.599	6.285

	<b>amici qui a là universit à</b>					
--	---	--	--	--	--	--