Kang Yi Da & Jashwant Gunaseelan **Words: 966**

1. Reintroduction of Project

What Dialect La has two main objectives, using GPT-4, we are looking to classify dialects based on written text and to regenerate sentences in one dialect into another dialect. While English is a universal language, there are regional dialects that are tied to the specific cultures for each region. By understanding who might be typing something, we are able to tailor responses that are most respectful to people of that culture and to ensure we do not tap on anything that may be taboo within that culture. By regenerating sentences from one dialect to another, people will have a tool to understand the nuances in other cultures to assist them with getting assimilated into their new culture.

2. Data Processing

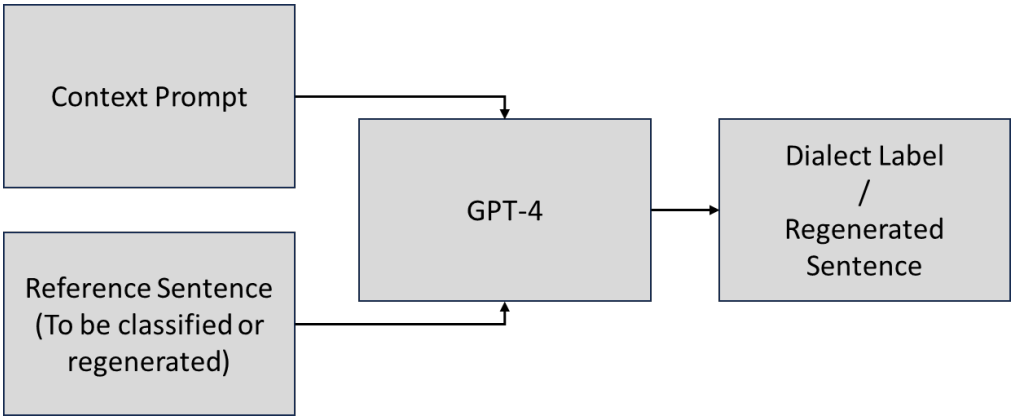
We got the datasets from online street interviews and picking out specific sentences that **we** thought were unique and might be telling of the dialect. Some of the sentences were very obviously of a certain dialect while other sentences are vaguer and can share vocabulary and sentence structures with other dialects.

We collected 30 sentences from 6 different dialects and 30 sentences that we felt were more neutral in nature. Here are sentences:

Dialect	Sentence
Australian	Yeah nah, not in the mood for it tonight
British	Aye it's not nice. Lots of druggies, scroungers, you know.
Indian	My cousin brother lives in Canada.
Irish	That film was brilliant, so it was.
Singaporean	I’m going to the market, you want come along?
South African	I'm after getting a new job in the city, so I am.
Standard	As a safety feature, a fail-safe feature is enabled by default.

3. Architecture

Below is describes the general architecture of the systems we want to implement.



For the classifier, the context prompt tells GPT-4 to be a classifier and the kinds of labels that are available. The sentence that we want GPT-4 to classify is the input prompt and GPT-4 will provide the classification label.

The context prompting will have 3 variants and we will test the effectiveness of each of these prompting styles: Zero-Shot, Single-Shot and Single-Shot with Chain of Thought Prompting.

For the generator, the context prompt tells GPT-4 to translate a specific dialect to another specific dialect. The sentence that we want GPT-4 to translate is the input prompt and GPT-4 will provide the translated sentence.

We will likely test the initial generations on how well it can generate in any given dialects. Unfortunately, Jashwant and Yi Da are only fluent in some dialects and we will be serverely limiting the dialects we choose to test.

4. Model effectiveness gauge

To test the effectiveness of the classifier, we will let GPT-4 process the datasets and provide the labels. We will test the accuracy of the output with our own labels and determine the accuracy. For the comparison, we will compare the system to the zero-shot prompt as the baseline to determine how much we have succeeded.

To test the generator, GPT-4 will take in the prompt and provide the translated sentence. Jashwant and Yi Da will assess how well the sentence has been translated using mean opinion score (1 to 5, base score is 1). The rubrics is as per this table:

Points	+0	+1	+2
Original meaning preserved	Meaning of original sentence was lost	Some parts of the meaning has been altered but is largely correct.	The meaning of the sentence got preserved entirely.
Sentence structure and vocabulary of the target dialect	The generated sentence does not contain any aspects of the target dialect.	Some aspects of the target dialect were shown but there were either grammatical or sentence structural mistakes	The generated sentence has the sentence structure and some vocabulary from the target dialect.

The generator would have succeeded if we got a rating of at least 4.3, or if it has improved it from the zero-shot variant.

5. Result

We ran the classifiers on the dataset. We are starting with zero-shot 2 class classification and evolved the prompt to ensured it functioned. Once we were comfortable with the accuracy, we increased the number of classes. This repeated until we classified 6 different classes.

We proceeded with the single-shot classification afterwards and evolved the prompt by changing the phrasing and the example provided.

The last method we tried was with single-shot chain of thought prompting. There were a couple of issues faced here especially when GPT-4 came up with labels that were outside of our dataset.

These tables reflect our results. The rows are the actual labels, the columns are the predictions.

Zero-Shot Confusion Matrix

	Standard	Brit	Sgp	Ind	Aus	SAfrican	Irish
Standar	30	0	0	0	0	0	0
Brit	2	26	0	0	0	0	2
Sgp	4	1	22	3	0	0	0
Ind	3	1	2	24	0	0	0
Aus	0	1	0	0	29	0	0
SAfrican	7	5	1	1	3	13	0
Irish	6	7	1	0	1	0	15

Accuracy: 75.7%

Single shot with Confusion Matrix

	Standard	Brit	Sgp	Ind	Aus	SAfrican	Irish
Standar	30	0	0	0	0	0	0
Brit	4	25	0	0	0	0	1
Sgp	9	1	19	1	0	0	0
Ind	3	0	0	27	0	0	0
Aus	0	1	0	0	29	0	0
SAfrican	7	5	1	1	2	15	0
Irish	6	7	0	1	2	0	14

Accuracy: 75.7%

6. Discussion

Clearly more evolving is required. Looking at the prompts that GPT-4 puts out, it could auto-complete sentences and infer using the auto-completed results, leading to drastically different results. There are also instances where GPT-4 would come up with its own classification that wasn't included in the original dataset.

We will need to come up with a more efficient way of testing how the prompt impacts the output without the need to rerun the whole system due to the cost of running the system. We would also need to relook at our dataset and redefine what makes a good dataset.

7. Teamwork and progress

We have been working well thus far and haven't had many issues. For coming up with the dataset, we set out the parameters and we worked on the datasets individually. As for the rest of the tasks, we met up to work and brainstorm on those together. The table below breaks down what we have done individually thus far:

Task	Primary taskman
Coming up with dataset	Jashwant & Yi Da
Prompting	Jashwant
Coding with the API	Yi Da
Report writing	Yi Da & Jashwant