



Language Access and Large Language Models

Tomonori Nagano <tnagano@lagcc.cuny.edu>

Education and Language Acquisition Dept.
LaGuardia Community College

May 9, 2024

What is Language Access? I

Language Access

Language access refers to efforts to ensure access to information, services, resources, and opportunities regardless of languages of preference.

The primary motivation for language access is compliance with legal requirements for limited English proficiency (LEP) who have a limited proficiency in English (cf. NYS EO26.1, NYC Local Laws 30 and 73). However, more broadly, language access can be implemented for other reasons such as social inclusion and marketing.

Familiarity Check

- Languages spoken at LaGuardia (The data was obtained from the Institutional Research; 89% didn't respond the language item in the CUNY survey. They were removed from the calculation.)

	LANGUAGE	FALL 2022		LANGUAGE	FALL 2022
1	English	60.5%	13	French	0.4%
2	Spanish	18.6%	14	Urdu	0.4%
3	Bengali	3.6%	15	Punjabi	0.4%
4	Chinese	3.5%	16	Portuguese	0.3%
5	Nepali	1.4%	17	Igbo	0.3%
6	Haitian Creole	1.2%	18	Hindi	0.2%
7	Tibetan	1.2%	19	Burmese	0.2%
8	Tagalog	1.1%	20	Pilipino	0.2%
9	Arabic	1.0%	21	Uzbek	0.2%
10	Korean	1.0%	22	Thai	0.2%
11	Polish	0.6%	23	Russian	0.1%
12	Albanian	0.4%	24	Japanese	0.1%

Languages at LaGuardia I

Languages at LaGuardia Community College

- Nearly half of students speak languages other than English (presumably at home)
- The number of speakers other than English is increasing.
- The number of languages spoken is also increasing.

My Project

- How can we provide better language access with AI/LLM?
- How can we use better language support for advisement & student outreach?

Using LLM for Language Access I

- Very little is known why LLMs can translate languages (since the data source for training models is not publicly shared).
- Supported languages by OpenAI's Whisper

1. Afrikaans	13. Dutch	25. Indonesian	37. Nepali	49. Swedish
2. Arabic	14. English	26. Italian	38. Norwegian	50. Tagalog
3. Armenian	15. Estonian	27. Japanese	39. Persian	51. Tamil
4. Azerbaijani	16. Finnish	28. Kannada	40. Polish	52. Thai
5. Belarusian	17. French	29. Kazakh	41. Portuguese	53. Turkish
6. Bosnian	18. Galician	30. Korean	42. Romanian	54. Ukrainian
7. Bulgarian	19. German	31. Latvian	43. Russian	55. Urdu
8. Catalan	20. Greek	32. Lithuanian	44. Serbian	56. Vietnamese
9. Chinese	21. Hebrew	33. Macedonian	45. Slovak	57. Welsh
10. Croatian	22. Hindi	34. Malay	46. Slovenian	
11. Czech	23. Hungarian	35. Marathi	47. Spanish	
12. Danish	24. Icelandic	36. Maori	48. Swahili	

Using LLM for Language Access II

Experimenting language translation with LLM 1

- Asked GPT4, GPT3.5, and Gemini the following question:
 - "¿How can you make 寿司 🍣 en casa"
- Responses are written in different language (sometimes in Japanese), but they never mixed languages in the same response

```
~ » myllm.py -m "gpt-4-turbo" -p "¿How can you make 寿司 🍣 en casa?" ]  
Making sushi at home can be a fun and rewarding experience. Here's a basic guide to help you get started with making sushi rolls (maki sushi), one of the most popular types of sushi.  
  
### Ingredients:  
- **Sushi Rice:** Specially prepared rice mixed with sushi vinegar.  
- **Nori:** Seaweed sheets used to wrap sushi rolls.  
- **Fillings:** Common fillings include fresh fish (like salmon or tuna), cucumber, avocado, and crab meat. Ensure any fish used is fresh and safe for raw consumption.  
- **Sushi Vinegar:** A mix of rice vinegar, sugar, and salt to season the sushi rice.  
- **Soy Sauce, Wasabi, and Pickled Ginger:** For serving.  
  
### Equipment:  
- **Bamboo Sushi Mat:** To roll the sushi.  
- **Sharp Knife:** To cut the rolls.  
- **Rice Cooker or Pot:** For cooking the rice.
```

Using LLM for Language Access III

Experimenting language translation with LLM 2

- Asked GPT4, GPT3.5, and Gemini the following question:
 - *"Using this text and other available resources/information, explain to [LANG] speakers who are not familiar with the American higher education why they should go to LaGuardia Community College. Use [LANG]."*
- Google Translate was also used to translate the same text.
- A few native speakers of languages other than English assesses the quality of outputs.
 - See the translated texts and audio files at <https://bit.ly/LG-DiveIntoAI2024>

Using LLM for Language Access IV

Experimenting language translation with LLM 2 (cont.)

- A few native speakers of languages other than English assesses the quality of outputs.
 - Gemini and GPT3.5 failed to produce translation in some languages (e.g., Burmese, Igbo, Tagalog etc.). Google Translate and GPT4.0 produced translation in all target languages (i.e., the top 24 languages at LaGuardia).
 - Google Translate was by far the least popular among all translation.
 - GPT3.5 and GPT4.0 produced audio files, which sounded very natural like a native speaker.
 - Overall, GPT4 was a lot better than GPT3.5 (in terms of both accuracy and fluency).
 - GPT3.5 and GPT4 produced very similar texts (but GPT4 was a lot more fluent). GPT3.5/GPT4 and Gemini developed very different outputs in all languages.