Large language model augmented exercise retrieval for personalized language learning

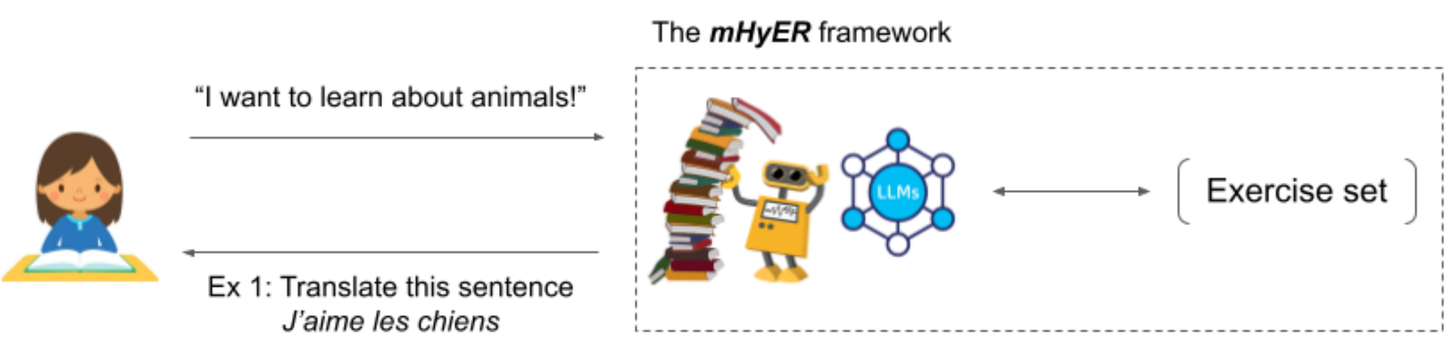
Austin Xu, Klinton Bicknell, Will Monroe

Contact: axu@gatech.edu

Self-directed learning for online language learning

How can we give learners the ability to request content in an online language learning setting?

- Learners should be able to tailor the online learning experience to fit needs


The *mHyER* framework
"I want to learn about animals!"
Exercise set
Ex 1: Translate this sentence
J'aime les chiens

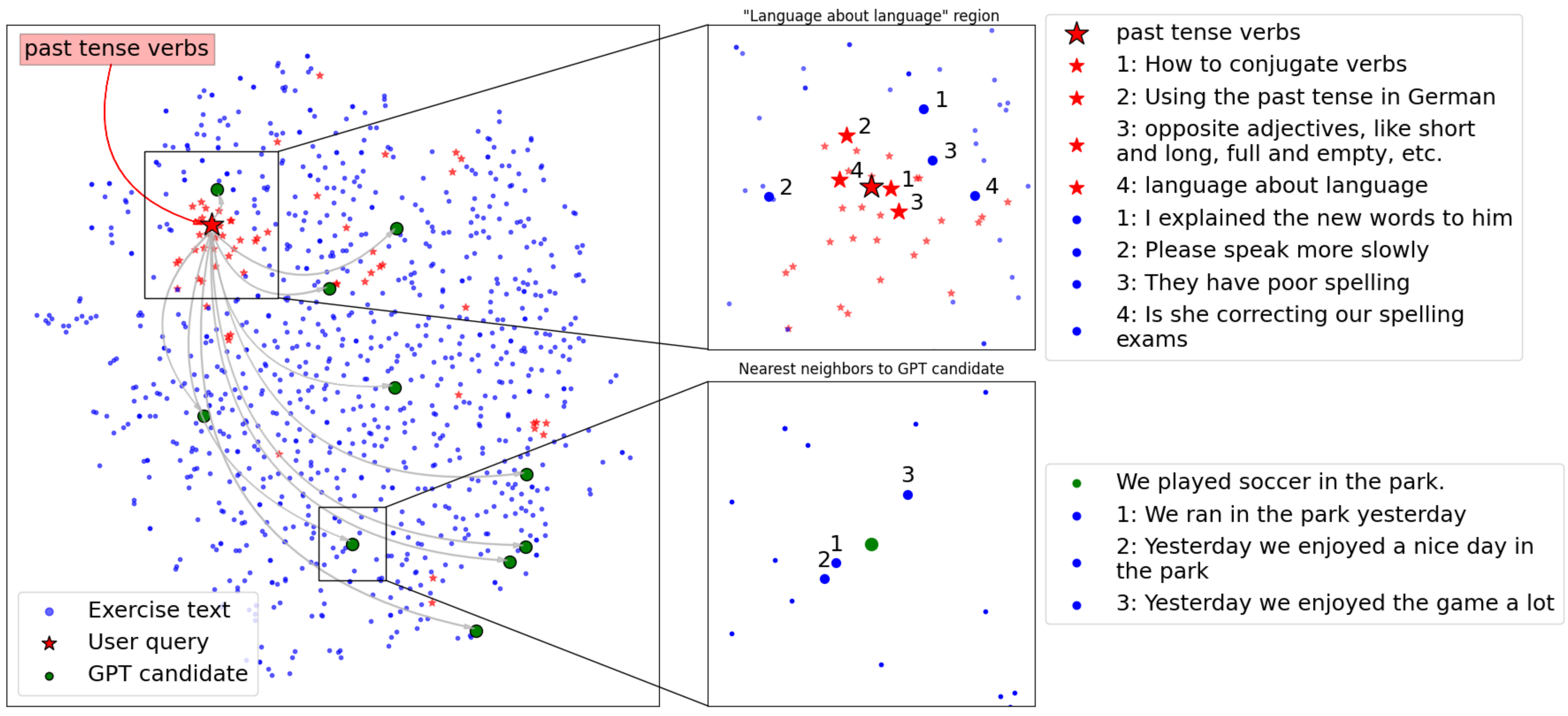
Can we just use direct similarity search / kNNs?

No!

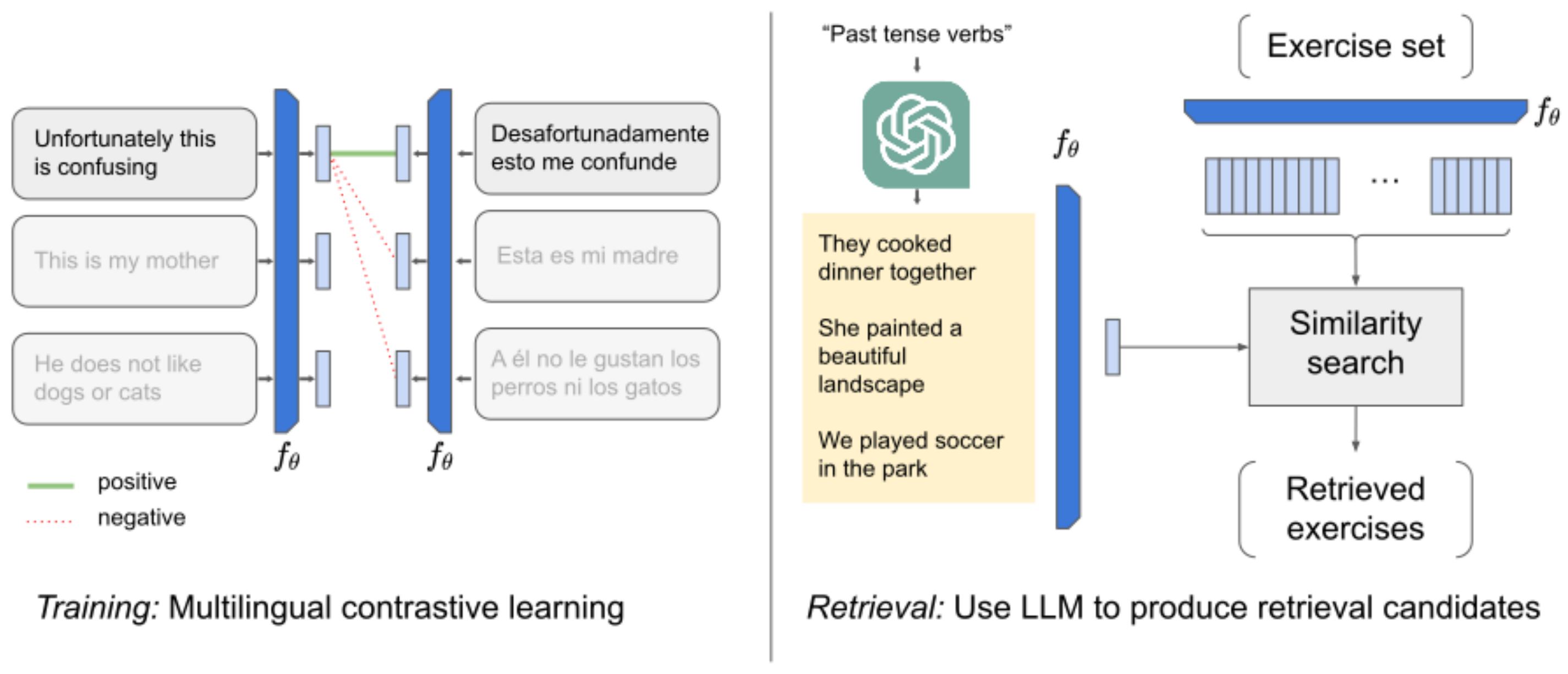
- Learners describe what they want to learn using "language about language", e.g., "I want to learn about **verbs**"
- kNN with embedded learner input returns exercises explicitly about language!
- Semantic gap between learner inputs and exercise content
- Cannot be overcome with large-scale pretraining, e.g., BERT

Problem setup:

- Learners provide text input describing what they want to learn
- Goal: Retrieve the most *relevant* exercises using a method that is
  - *Zero-shot:* no relevance labels are available for training
  - *Multilingual:* exercises are comprised of multilingual sentences



mHyER: Synthesize hypothetical exercises based on user inputs


Unfortunately this is confusing
This is my mother
He does not like dogs or cats
Desafortunadamente esto me confunde
Esta es mi madre
A él no le gustan los perros ni los gatos
f_θ
f_θ
positive
negative
"Past tense verbs"
They cooked dinner together
She painted a beautiful landscape
We played soccer in the park
f_θ
Exercise set
f_θ
Similarity search
Retrieved exercises

Training: Multilingual contrastive learning

Retrieval: Use LLM to produce retrieval candidates

Stage 1: Multilingual contrastive learning

- Multilingual exercises have inherent structure: sentences and translations should be "similar" in representation space
- Idea: Use multilingual contrastive learning [1] to optimize similarity space!

Stage 2: Use LLM to generate retrieval candidates

- Need to bridge semantic gap
- Idea: Generate sentences similar to exercises *conditioned* on learner input [2]
  - Use LLMs to align learner input and exercises in representation space

Experimental results

Retrieval on Tatoeba data

- mHyER outperforms supervised baselines in zero-shot retrieval

		English		English (L2) from Spanish (L1)				Spanish (L2) from English (L1)			
		AUC	P@15	AUC L1	AUC L2	P@15 L1	P@15 L2	AUC L1	AUC L2	P@15 L1	P@15 L2
Unsup. pretraining	mBERT	0.468	0.037	0.446	0.487	0.038	0.040	0.469	0.442	0.039	0.019
	mContriever	0.571	0.064	0.438	0.503	0.051	0.063	0.559	0.564	0.061	0.027
	SimCSE	0.646	0.115	0.535	0.559	0.069	0.054	0.635	0.610	0.127	0.068
	mHyER$_{\text{mBERT}}$+Duo-OOD	0.752	0.211	0.734	0.738	0.215	0.206	0.739	**0.757**	0.225	0.242
	mHyER$_{\text{mContriever}}$+Duo-OOD	0.729	**0.258**	**0.748**	0.723	**0.267**	**0.264**	0.713	0.744	**0.271**	**0.294**
Sup. pretraining	Contriever	0.541	0.164	0.491	0.492	0.120	0.086	0.530	0.492	0.180	0.105
	mContriever	0.575	0.104	0.548	0.510	0.126	0.108	0.560	0.581	0.112	0.101
	mHyER$_{\text{Contriever}}$+Duo-OOD	**0.775**	0.246	0.668	**0.797**	0.102	0.240	**0.760**	0.692	**0.268**	0.108
	mHyER$_{\text{mContriever}}$ +Duo-OOD	0.738	**0.255**	**0.761**	0.734	**0.260**	**0.264**	0.722	**0.752**	0.255	**0.280**

Ablations

- Both contrastive learning and generated retrieval candidates contribute to performance gains

		English		English (L2) from Spanish (L1)				Spanish (L2) from English (L1)			
		AUC	P@15	AUC L1	AUC L2	P@15 L1	P@15 L2	AUC L1	AUC L2	P@15 L1	P@15 L2
Unsup. pretraining	mContriever	0.571	0.064	0.438	0.503	0.051	0.063	0.559	0.564	0.061	0.027
	mContriever +GPT	0.676	0.237	0.613	0.663	0.213	0.213	0.643	0.602	0.245	0.217
	mContriever +Duo-OOD	0.665	0.096	0.670	0.665	0.119	0.106	0.656	0.657	0.090	0.077
	mHyER$_{\text{mContriever}}$+Duo-OOD	**0.729**	**0.258**	**0.748**	**0.723**	**0.267**	**0.264**	**0.713**	**0.744**	**0.271**	**0.294**
Sup. pretraining	mContriever	0.575	0.104	0.548	0.510	0.126	0.108	0.560	0.581	0.112	0.101
	mContriever +GPT	0.731	0.250	0.642	0.724	0.238	0.243	0.706	0.636	**0.263**	0.258
	mContriever +Duo-OOD	0.672	0.106	0.678	0.677	0.128	0.120	0.662	0.661	0.113	0.091
	mHyER$_{\text{mContriever}}$+Duo-OOD	**0.738**	**0.255**	**0.761**	**0.734**	**0.260**	**0.264**	**0.722**	**0.752**	0.255	**0.280**

References

[1] Yaushian Wang, Ashley Wu, and Graham Neubig. English contrastive learning can learn universal cross-lingual sentence embeddings. In *EMNLP 2022*

[2] Luyu Gao, Xueguang Ma, Jimmy Lin, and Jamie Callan. Precise zero-shot dense retrieval without relevance labels. In *ACL* 2023