Mnemonic Generator [Data]

Keyword mnemonic devices can help students remember vocab. Mnemonics take a vocab term, link it to a similarly-sounding keyword, and explain how the keyword and definition of the term are linked. For example, to learn that “benevolent” means “kind”, you might use the mnemonic:

* ***Benevolent*** *sounds like benefit, and a boss who gives their employees benefits is* ***kind***

There are a few works on using LLMs to automatically generate mnemonics, but most of them are in English, and only focus on vocabulary + definitions:

* [A SMART Mnemonic Sounds like “Glue Tonic”: Mixing LLMs with Student Feedback to Make Mnemonic Learning Stick](https://arxiv.org/abs/2406.15352)
* [Smartphone: Exploring keyword mnemonic with auto-generated verbal and visual cues](https://arxiv.org/abs/2305.10436)

But there are several use cases for mnemonics beyond English vocabulary:

* Multi-lingual vocab: *Baño sounds like Banjo. Imagine a banjo player in a bathroom*
* Country capitals: *Yeah Man (Yemen), it’s like a Sauna (Sanaa) in here; I’d be happier (Apia) if I had some more (Samoa)*
* Chemical element symbols: *It’s a Sin (Sn) to not remember the symbol for Tin*

This project will involve collecting data so we can train a mnemonic generator that can handle use cases beyond English vocabulary. We have recently found a website that contains several high-quality mnemonics for diverse domains: [https://mammothmemory.net/.](https://mammothmemory.net/)

[A bird on a car

AI-generated content may be incorrect.](https://mammothmemory.net/)

[*What to submit*: Submit a link to a Github repo that takes in a URL on](https://mammothmemory.net/) Mammoth memory and scrape the page using BeautifulSoup, Selenium, etc. This should yield a dataset with:

* Words/concept pair (e.g. vocab term + definition, chemical element + symbol)
* The mnemonic for this pair
* The highlighted text if present (e.g. “Ch” and “Cr” above)
* Images (Optional, if possible)

You are not expected to find every page on this website with mnemonics, but setting up some code that can work on 2-3 pages with diverse mnemonics would be a good starting point

*How this would translate into a full project:* We would use this dataset to train an mnemonic generator, and run a study (we have a flashcard app) with students to see how good they are.  You would help adapt the Flashcard app to display the generated multimodal mnemonics and to collect and analyze feedback from the generated mnemonicsmnmenonics.