

French Canadian Dataset for MRL 2025

I brainstormed a bunch of topic ideas with LLMs. After that, I just began writing sentences in those topics. I ended up choosing 10 topics and wrote 20 for each one and did 100 examples as label 1 being correct and making 100 as label 0 being correct. Then I just wrote the examples following all of the rules, which were: variable length, two candidate solutions should be as similar as possible, one being unambiguously correct and the other incorrect. I tried not to start all examples the same way. Lastly, I varied the length between each topic for diversity