- Millions to billions of words are easy to get (trillions of English words available on the web)

- But: huge language models do not fit into RAM