What You Should Know

- What is the **sparsity problem** in NLP?
- What is the **distributional hypothesis**?
- How does **Brown clustering** work? What is it trying to maximize?
- What is **agglomerative clustering**?
- What is a **word embedding**?
- What is the difference between the **CBOW** and **Skip-Gram** models?
- What is **negative sampling** and why is it useful?
- What is the **learning objective** for **SGNS**?
- What is the matrix that **SGNS is implicitly factorizing**?
- What is the **learning objective** for **GloVe**?
- What are some examples of **hyperparameters for word embedding methods**?