# Exercises Lecture 7

- In a project to analyze social media data, you encounter issues with Word2Vec embeddings for rare, slang, and compound words (e.g., "socio-economic" or "data-driven"), as well as hashtags. These terms are often split into individual words or poorly represented.

    - What preprocessing steps could you take to handle compound words, slang, and hashtags more effectively in Word2Vec training?
    - How would you modify the context window size or negative sampling to better capture the semantics of these rare or compound words?
    - Discuss the potential trade-offs of these adjustments in terms of generalizability to other types of vocabulary in the dataset.

- Describe how negative sampling works in Word2Vec. Why is it necessary, and how does it affect the training process?
- Word2Vec is known to have limitations in handling polysemous words (words with multiple meanings) - Explain
- Explain mathematically why using a distribution like P(w) (empirical word frequency) directly for negative sampling would be problematic.
- Explain why linear relationships in the embedding space can capture analogies. Derive a condition on the embedding vectors that supports this linearity.
- Explain the role of the chain rule in this derivation, and discuss how the backpropagation algorithm efficiently computes these gradients for all layers.