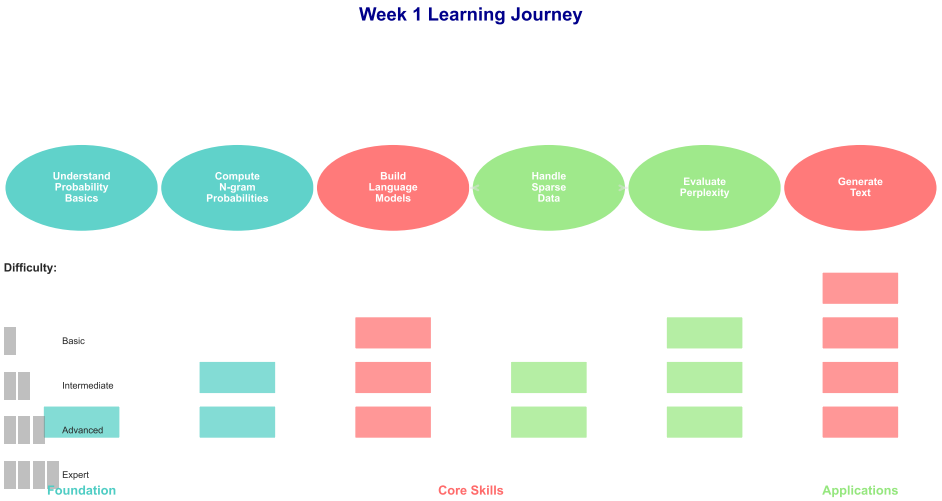
Week 1: Foundations & Statistical Language Models

Building the Groundwork for NLP

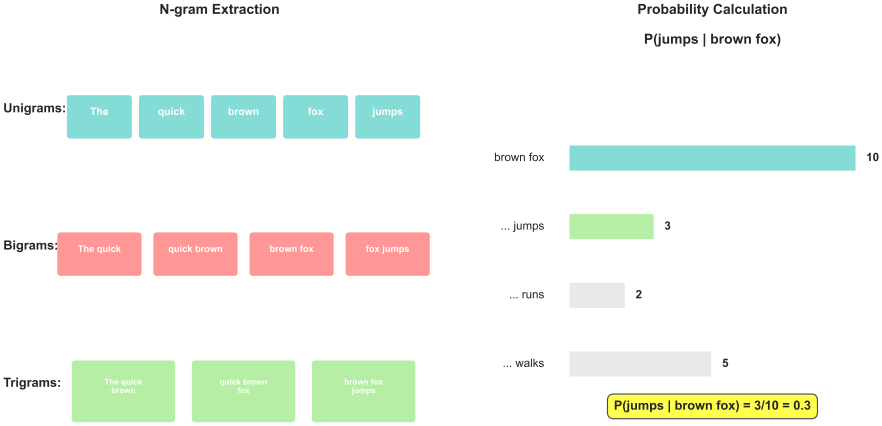
Natural Language Processing Course

BSc Computer Science

What We'll Learn This Week



Key Concept: N-gram Language Models



The Markov Assumption

- Future depends only on recent past
- $P(w_i \mid w_1 \ldots w_{i-1}) \approx P(w_i \mid$

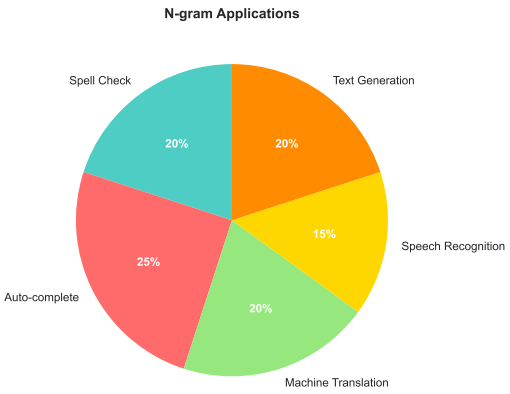
Probability Estimation

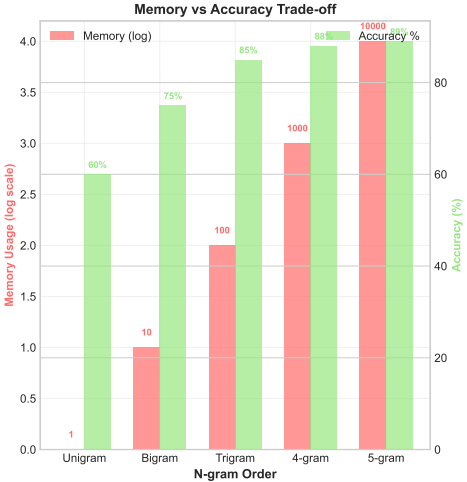
- Count n-grams in corpus
- Maximum likelihood:

Challenges

- Data sparsity
- Storage requirements

Applications & Real-World Impact





Where N-grams Excel:

- **Spell Checking**: Find likely corrections
- **Auto-complete**: Predict next words

Historical Significance:

- Dominated NLP for decades (1980s-2000s)
- Still used in hybrid systems