

Text Data

Sequential Tokens

Words

Characters

Subwords

Key Characteristics

- Natural language structure
- Grammatical structure

- Variable length sequences
- Context-dependent meaning

✓ Applications

- Machine Translation
- Text Summarization
- Sentiment Analysis

⚠ Challenges

- Large vocabulary size
- Rare words handling
- Ambiguity (polysemy, homonyms)

💡 Real-World Examples

Machine Translation

Google Translate: "안녕하세요" →
"Hello"

Text Summarization

News aggregators condensing long
articles into key bullet points

Sentiment Analysis

Amazon reviews: "This product is
amazing!" → Positive (★★★★★)