Attention Intuition Exercise

Sentence

"The dog chased the cat."

Scenario

Query (Q) = embedding of "dog"

We want to understand which words in the sentence should receive high vs. low attention weights.

Words with HIGH Attention Weights

1. "chased" - HIGHEST attention

Reasoning:

- "dog" is the **subject** performing the action
- "chased" is the **verb** describing what the dog does
- Strong syntactic relationship (subject-verb dependency)
- Most semantically relevant to understanding what the dog is doing

2. "cat" - HIGH attention

Reasoning:

- "cat" is the **object** of the action
- Completes the meaning: dog chased *what*?
- Important semantic relationship (agent-patient)
- Necessary to understand the full action context

3. "dog" (self-attention) - MODERATE to HIGH

Reasoning:

- The word attends to itself
- Helps reinforce its own representation
- Common in transformer models (diagonal of attention matrix)
- Important for capturing the token's own context

Words with LOW Attention Weights

1. "The" (first occurrence) - VERY LOW

Reasoning:

- Determiner with minimal semantic content
- Doesn't add meaningful information about "dog"
- Grammatical function word
- Typically filtered out in traditional NLP

2. "the" (second occurrence before "cat") - VERY LOW

Reasoning:

- Same as above - just a determiner
- No semantic relationship with "dog"
- Purely grammatical marker

3. "." (period) - VERY LOW

Reasoning:

- Punctuation mark
- No semantic content
- Indicates sentence boundary but doesn't describe "dog"

Expected Attention Distribution

If we represent attention weights as percentages:

Word	Attention Weight	Category
The	~2%	Very Low
dog	~20%	Moderate-High (self)
chased	~45%	Highest
the	~2%	Very Low
cat	~30%	High
.	~1%	Very Low

Total: 100%

Reasoning Summary

Why High Attention?

Words receive **high attention** when they:

- Have **syntactic dependencies** with the query word (subject-verb, verb-object)
- Are **semantically related** (actors, actions, objects in same event)
- Provide **critical context** for understanding the query word's role
- Form the **core meaning** of the sentence

Why Low Attention?

Words receive **low attention** when they:

- Are **function words** (determiners, conjunctions, prepositions)
- Have **purely grammatical** roles without semantic content
- Are **punctuation marks**
- Don't directly relate to the query word's meaning or role

Key Insight

Attention learns to focus on semantically and syntactically relevant words, not just nearby words. In this example:

- "dog" → "chased": captures who is doing the action
- "dog" → "cat": captures the relationship between agent and patient
- "dog" → "the", ".": minimal attention to non-informative tokens

This is why attention is more powerful than fixed context windows - it dynamically focuses on what matters for understanding each word.