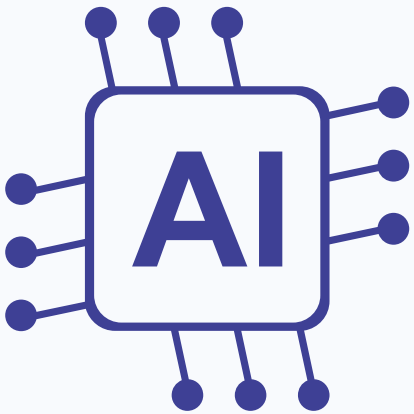
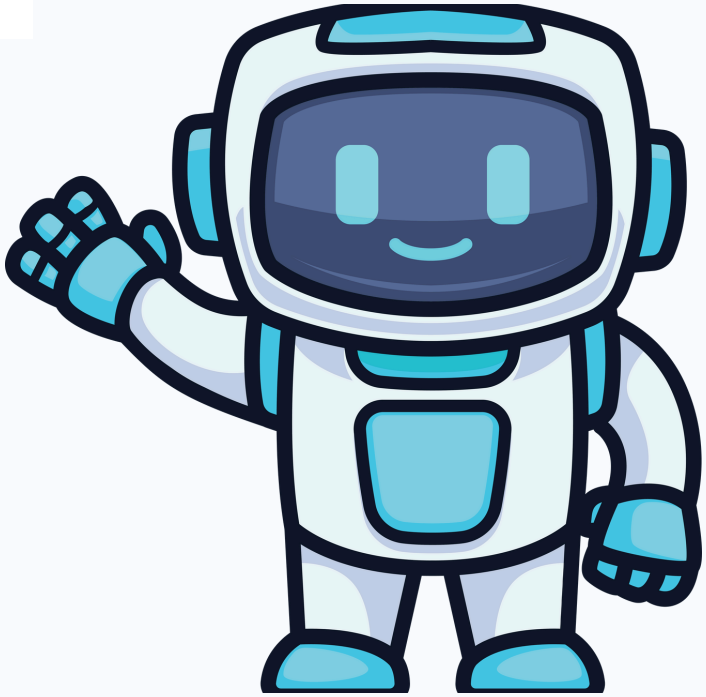
LiveAi

# Self-Attention

$$Attention\ (Q, K, V) = Softmax\ (\frac{QK^T}{\sqrt{d_K}})V$$

Unlocking
Your Potential,
Unleashing
Your Success



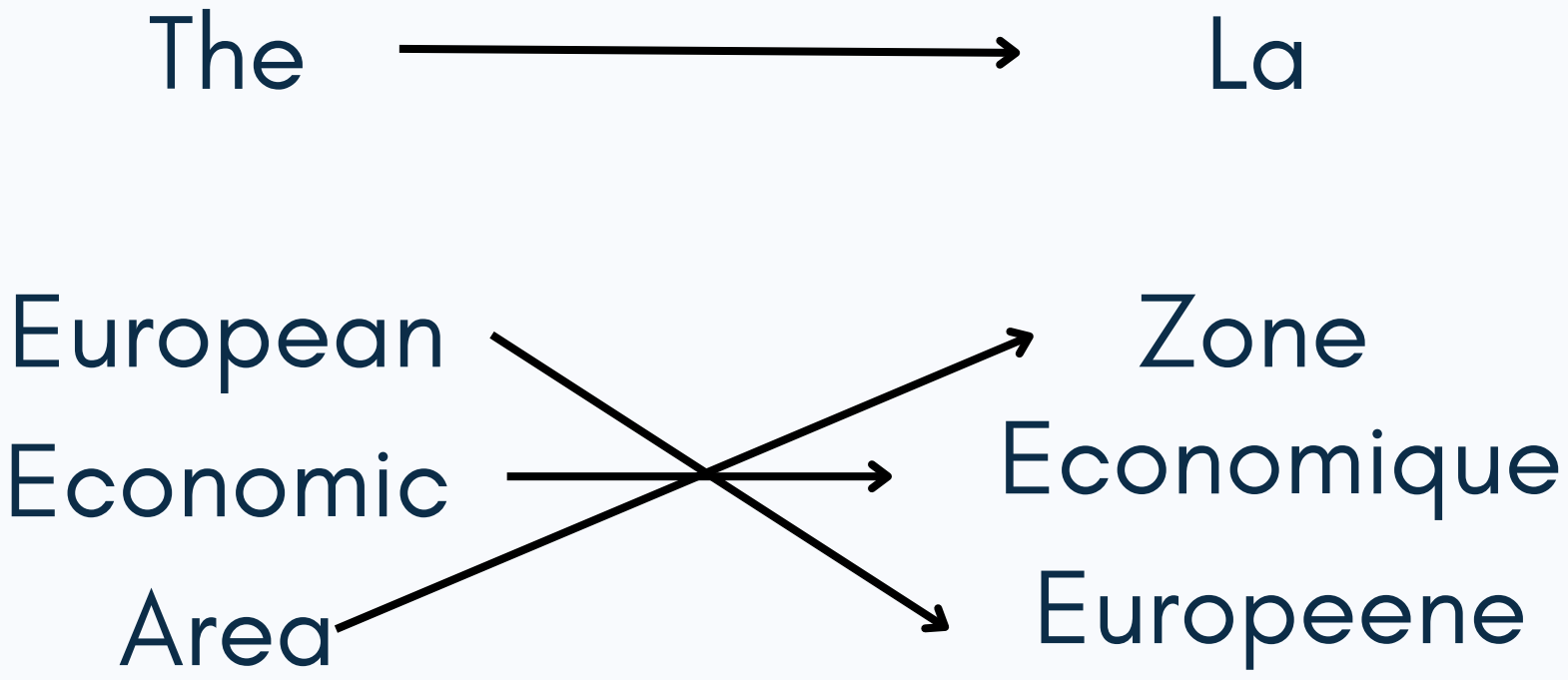



# Order

LiveAi

English

French



# Size Mismatch

LiveAi

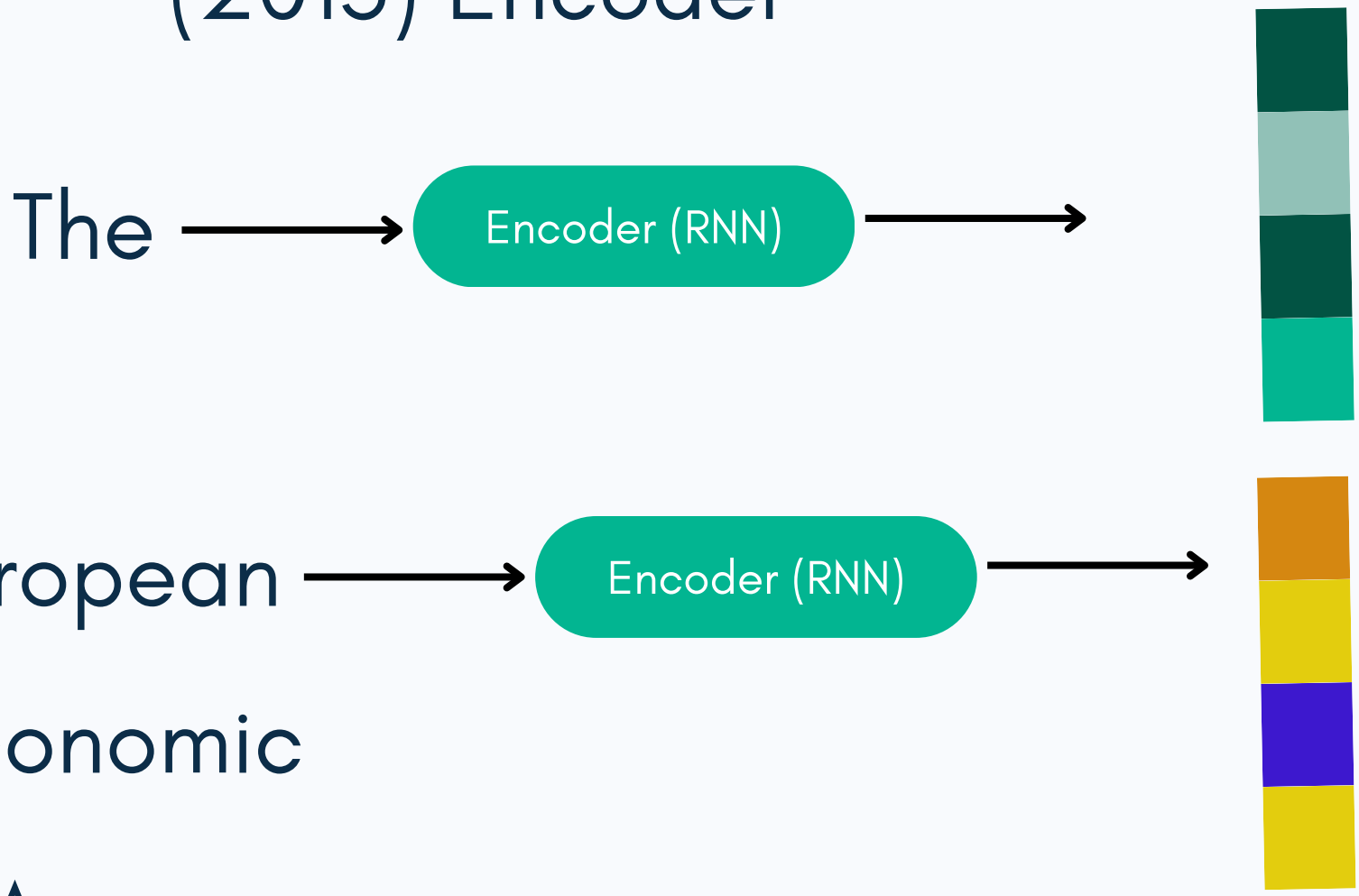
English

French

The → La

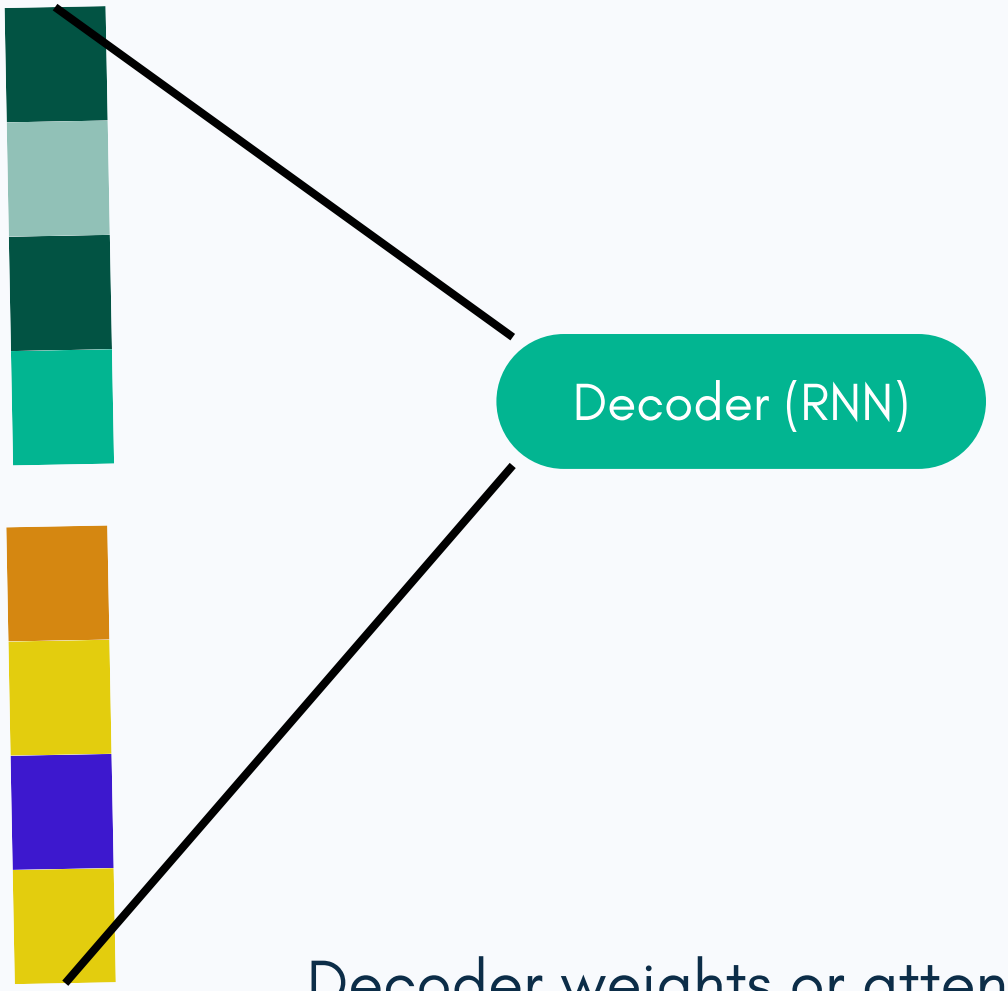
European Economic Area → Europeene Economique Zone

# Machine Translation (2015) Encoder




Vectors which represent meaning of word or words in the context of sentences

# Machine Translation (2015) Decoder




La attend to 1st e -word

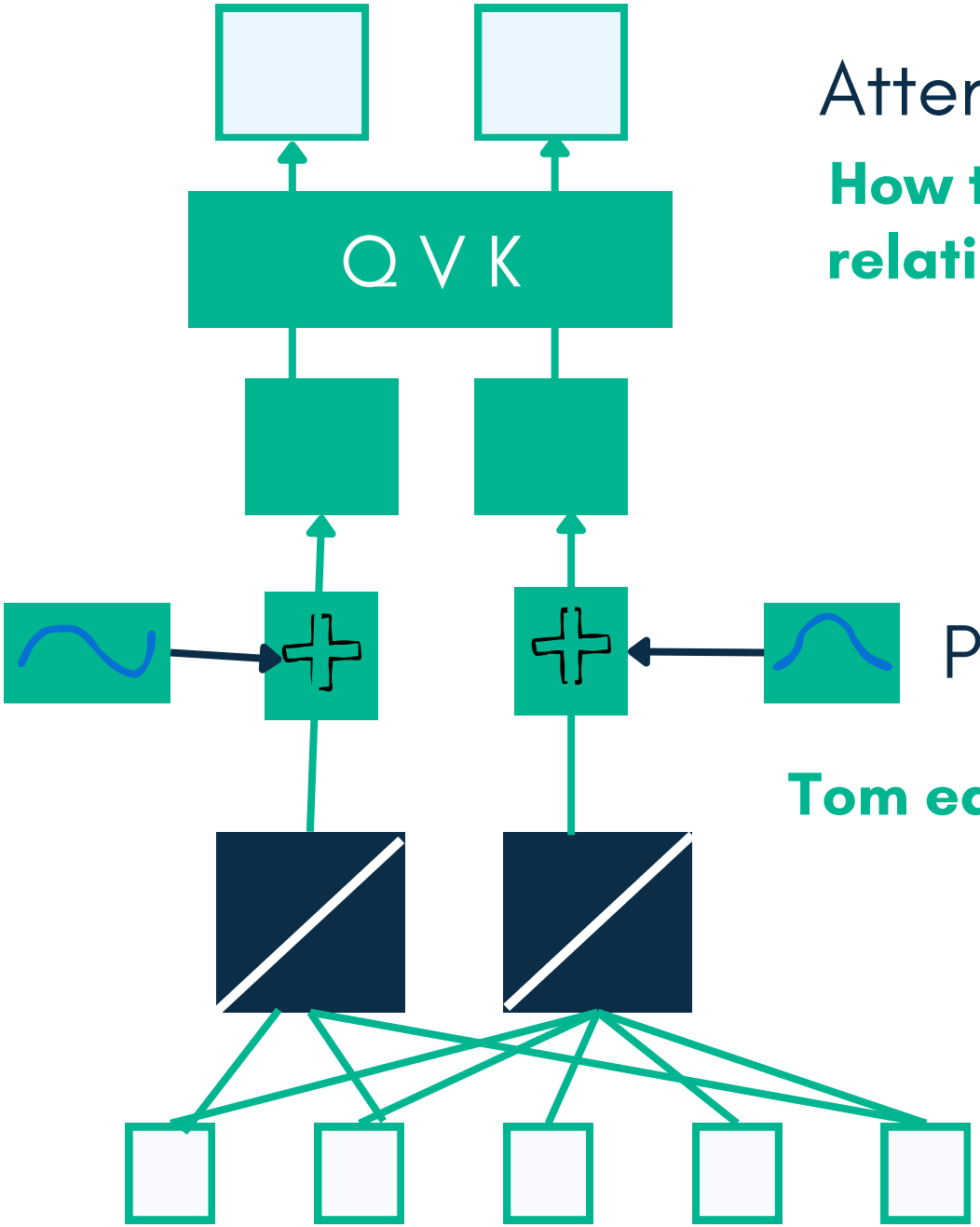
Zone attend to 4th e -word

Economique attend to 3rd e -word

Europeene attend to 4th e -word

Decoder weights or attends to the inputs based on the previous and current words being generated .

LiveAi




Attention

**How transformers establish relations among words with attention?**

Positional Embedding

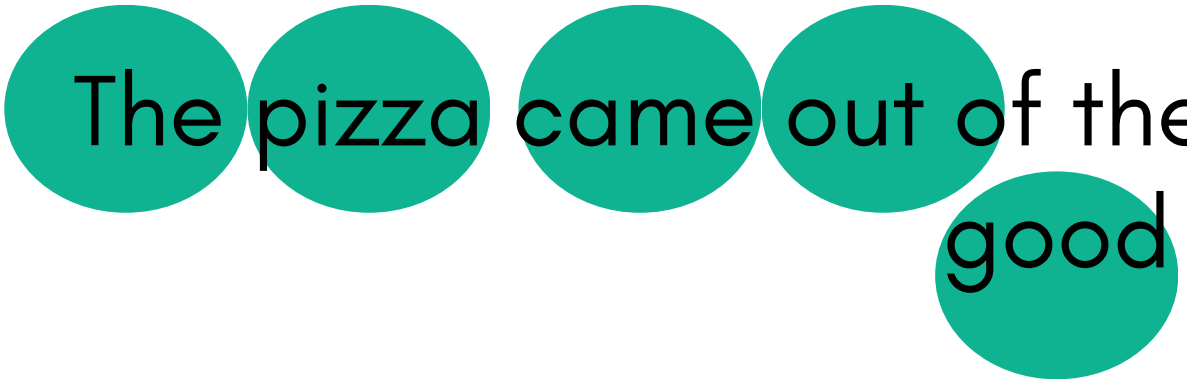
**Tom eats Pizza** **Pizza eats Tom**

Word Embedding

[0.1, 2.13, 3.14,5.0]

The pizza came out of the oven and **it** tasted good

Transformers have **attentions** to correctly associate the word it to pizza




Self attention calculates the similarity between **The** and all the words in the sentence.

The **pizza** came out of the oven and **it** **tasted** good

If you have a lot of examples where the word **pizza** is related to **it** and **taste**

Then the similiarity score between pizza, it and taste will be more

$$Attention\ (Q, K, V) = Softmax\ (\frac{QK^T}{\sqrt{d_K}})V$$

Key

| Last Name | Room Number |
|---|---|
| Smith | 200 |
| Summer | 201 |
| Smeeth | 202 |

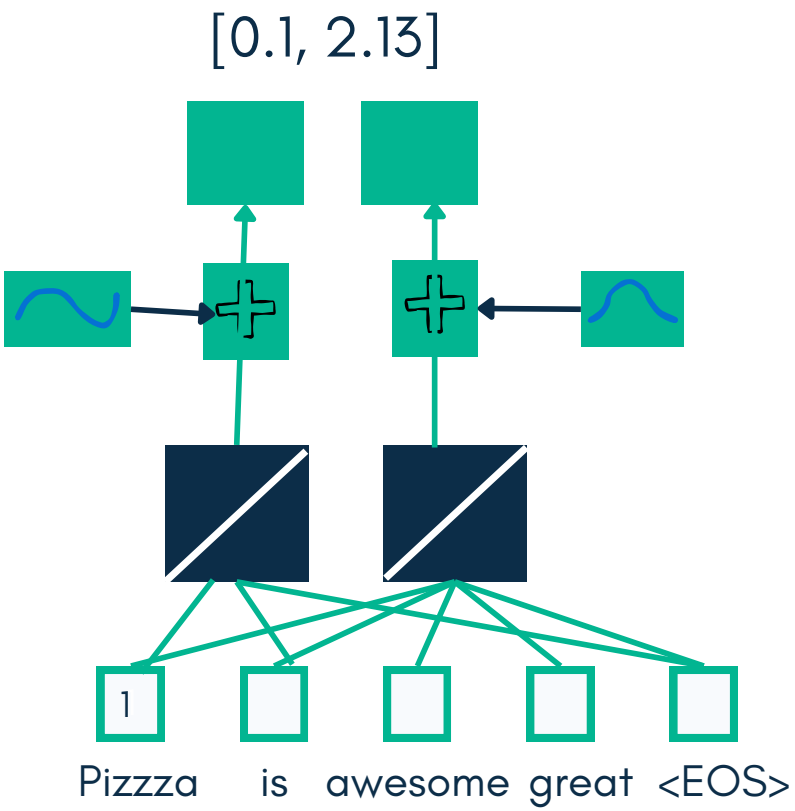
Summeth

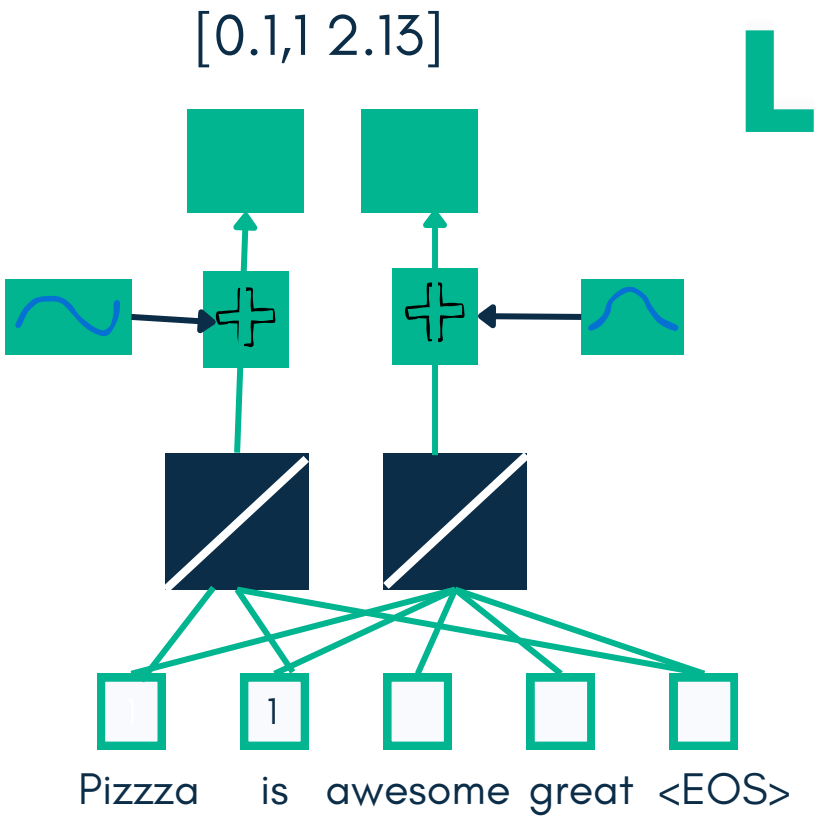
Query

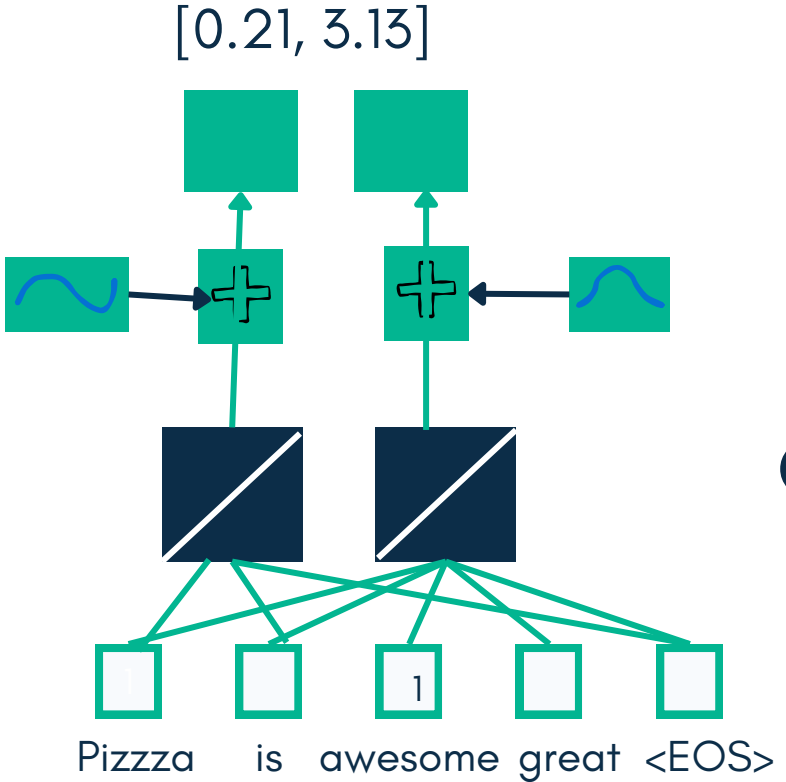
200

Value

LiveAi




Pizza
is
awesome
<EOS>

| **0.1** | **2.13** |
|---|---|
| 0.11 | 2.13 |
| 0.21 | 3.13 |
| 0.8 | 0.9 |

**Value**

Value

| | |
|---|---|
| **0.1** | **2.13** |
| 0.11 | 2.13 |
| 0.21 | 3.13 |
| 0.8 | 0.9 |

Query Weights T

| | |
|---|---|
| **0.78** | **2.0** |
| 0.9 | 1.7 |

=

Query

| | |
|---|---|
| .. | .. |
| .. | .. |
| 0.8 | 0.8 |
| .. | .. |

Pizza
is
awesome
<EOS>

Because we stareted with 2 encoded values we multiplies with 2-D weight matrix . If we start with 512-encoded value we will have a 512X512 weight

Value

| | |
|---|---|
| **0.1** | **2.13** |
| 0.11 | 2.13 |
| 0.21 | 3.13 |
| 0.8 | 0.9 |

Key Weights T

| | |
|---|---|
| **0.78** | **2.0** |
| 0.9 | 1.7 |

=

Key

| | |
|---|---|
| .. | .. |
| .. | .. |
| 0.18 | 0.81 |
| .. | .. |

Pizza
is
awesome
<EOS>

Because we stareted with 2 encoded values we multiplies with 2-D weight matrix . If we start with 512-encoded value we will have a 512X512 weight

Value

| | |
|---|---|
| **0.1** | **2.13** |
| 0.11 | 2.13 |
| 0.21 | 3.13 |
| 0.8 | 0.9 |

Value Weights T

| | |
|---|---|
| **0.78** | **2.0** |
| 0.9 | 1.7 |

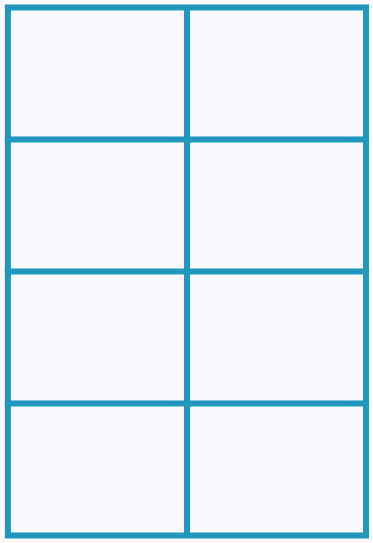
=

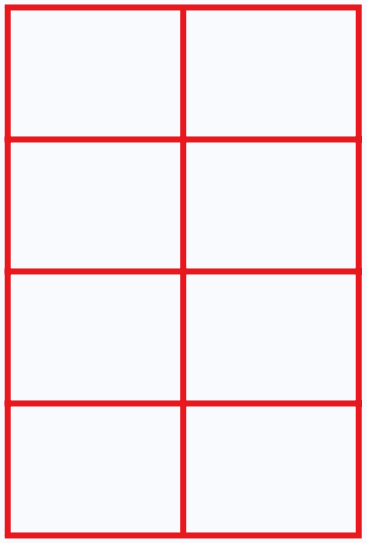
Value

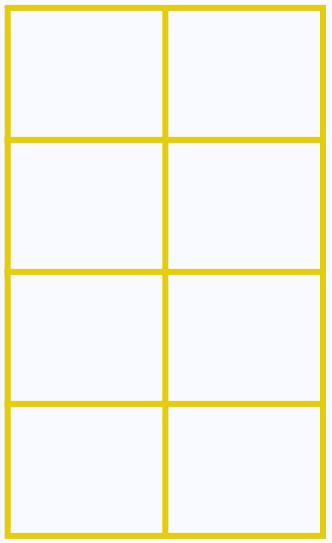
| | |
|---|---|
| .. | .. |
| .. | .. |
| 0.18 | 0.81 |
| .. | .. |

Pizza
is
awesome
<EOS>

Because we stareted with 2 encoded values we multiplies with 2-D weight matrix . If we start with 512-encoded value we will have a 512X512 weight



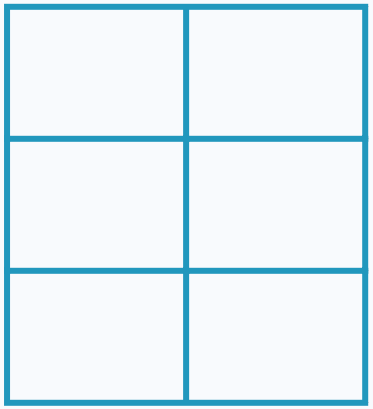


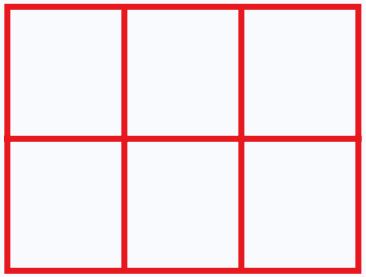


V

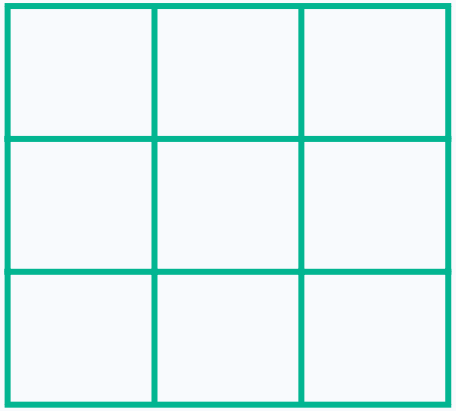
Q

K^T



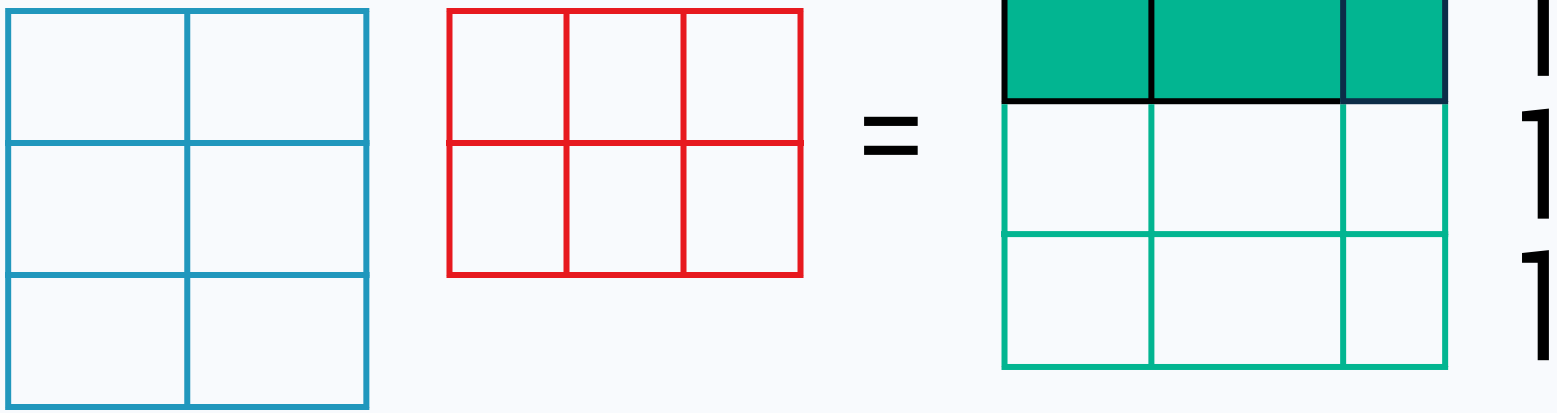


=



unscaled dot product and scale each dot product similarity by sqrt(2) -- encoded word dimnesion size

Softmax



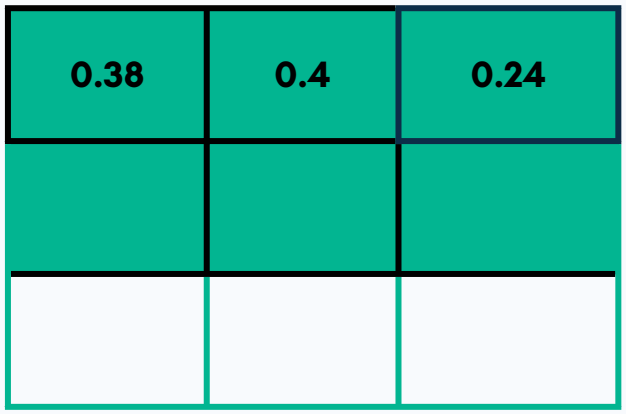

Pizza is awesome

Pizza
is
awesome

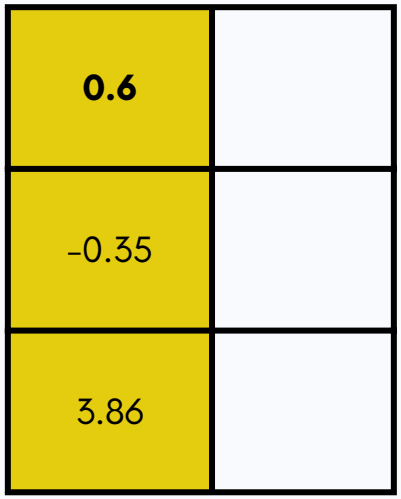
| | | |
|---|---|---|
| 0.38 | 0.4 | 0.9 |
| | | |
| | | |

Pizza is 0.38% similar to Pizza, 0.4% similar to is etc...

Pizza is awesome




Value Matrix

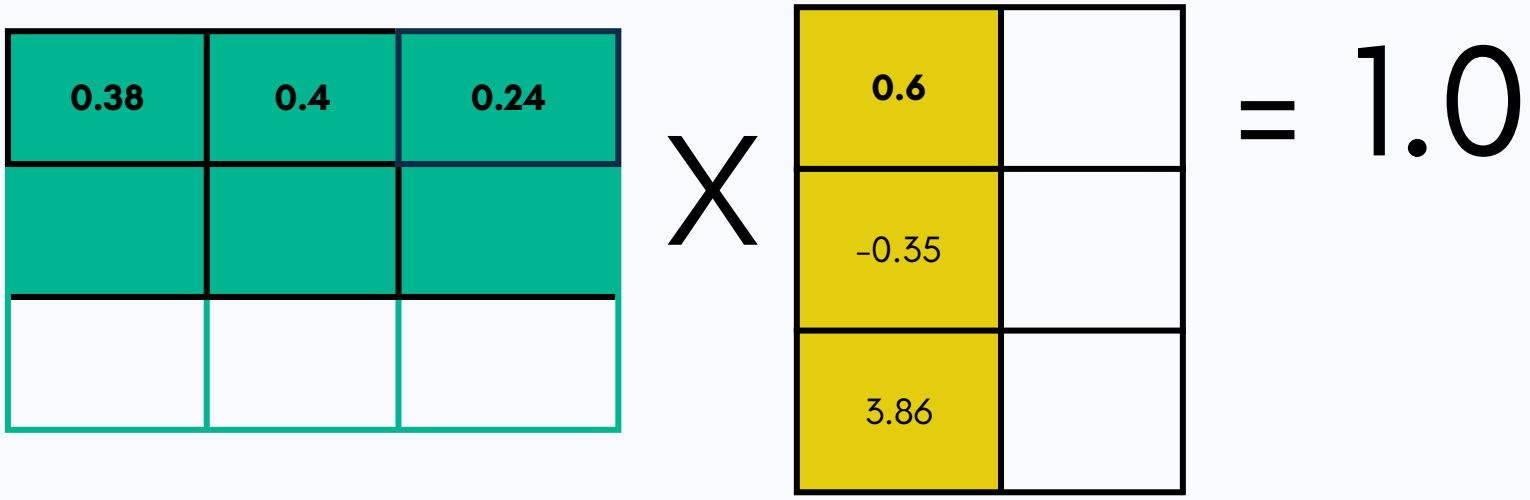



X = 1.0

In other words the percentages that comes out of the softmax .... tells us how much influence each word should have on the final encoding for a given word

Pizza is awesome

Value Matrix




we calculate 36% of the first value for Pizza

we calculate 40% of the first value for is

we calculate 24% of the first value for awesome

Pizza is awesome

| | | |
|---|---|---|
| **0.38** | **0.4** | **0.24** |
| | | |
| | | |

X

Value Matrix

| | |
|---|---|
| **0.6** | |
| -0.35 | |
| 3.86 | |

=

Self- Attention Score

| | |
|---|---|
| **1.0** | **1.9** |
| 0.2 | 0.4 |
| 3.86 | 2.2 |

Pizza
is
awesome

| | |
|---|---|
| **1.0** | **1.9** |
| 0.2 | 0.4 |
| 3.86 | 2.2 |