Paper Seminar

Analogies Explained: Towards Understanding Word Embeddings

Allen and Hospedales, 2019, ICML

Myeongsup Kim

Integrated M.S./Ph.D. Student
Data Science & Business Analytics Lab.
School of Industrial Management Engineering
Korea University

Myeongsup_kim@korea.ac.kr

Introduction

- Analogies in Word Embeddings

<What This Seminar Does Not Cover>

- **Details of Word2Vec**

  Mikolov et al., Efficient Estimation of Word Representations in Vector Space, ICLR Workshop, 2013

  Mikolov et al., Distributed Representations of Words and Phrases and their Compositionality, NIPS, 2013

<Analogies in Natural Language>

“man is to boy as woman is to girl”

<Analogies in Natural Language>

"man is to boy as woman is to girl"

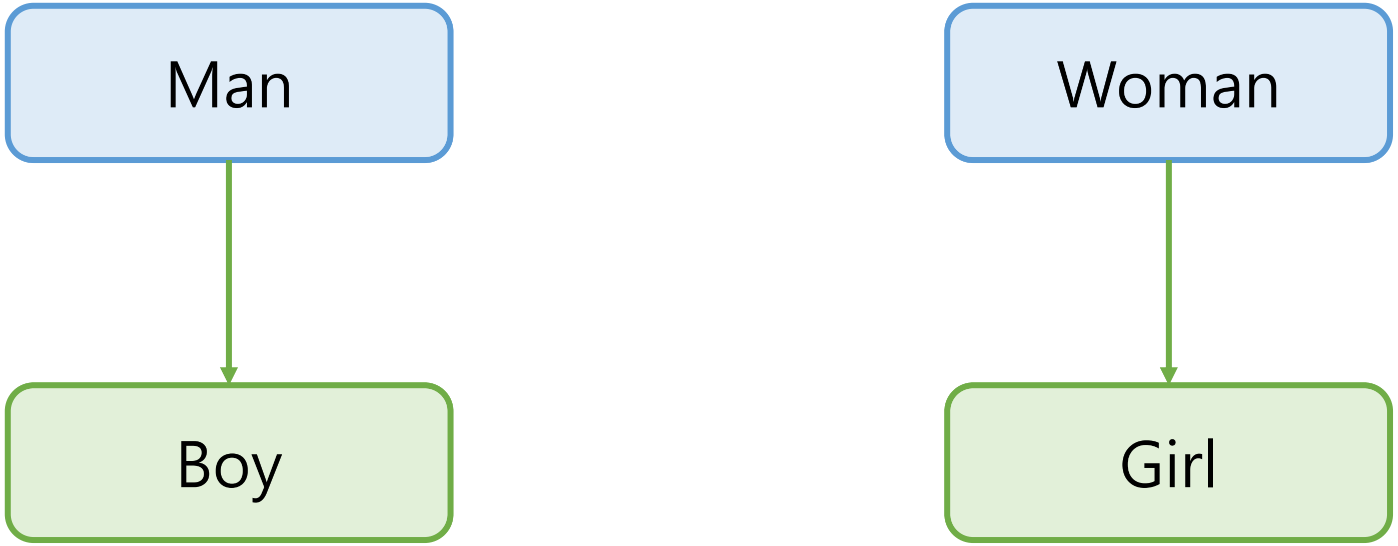
man : boy = woman : girl
(In Mathematical Expression)

<Analogies in Natural Language>

Man

Woman

<Analogies in Natural Language>



<Analogies in Natural Language>

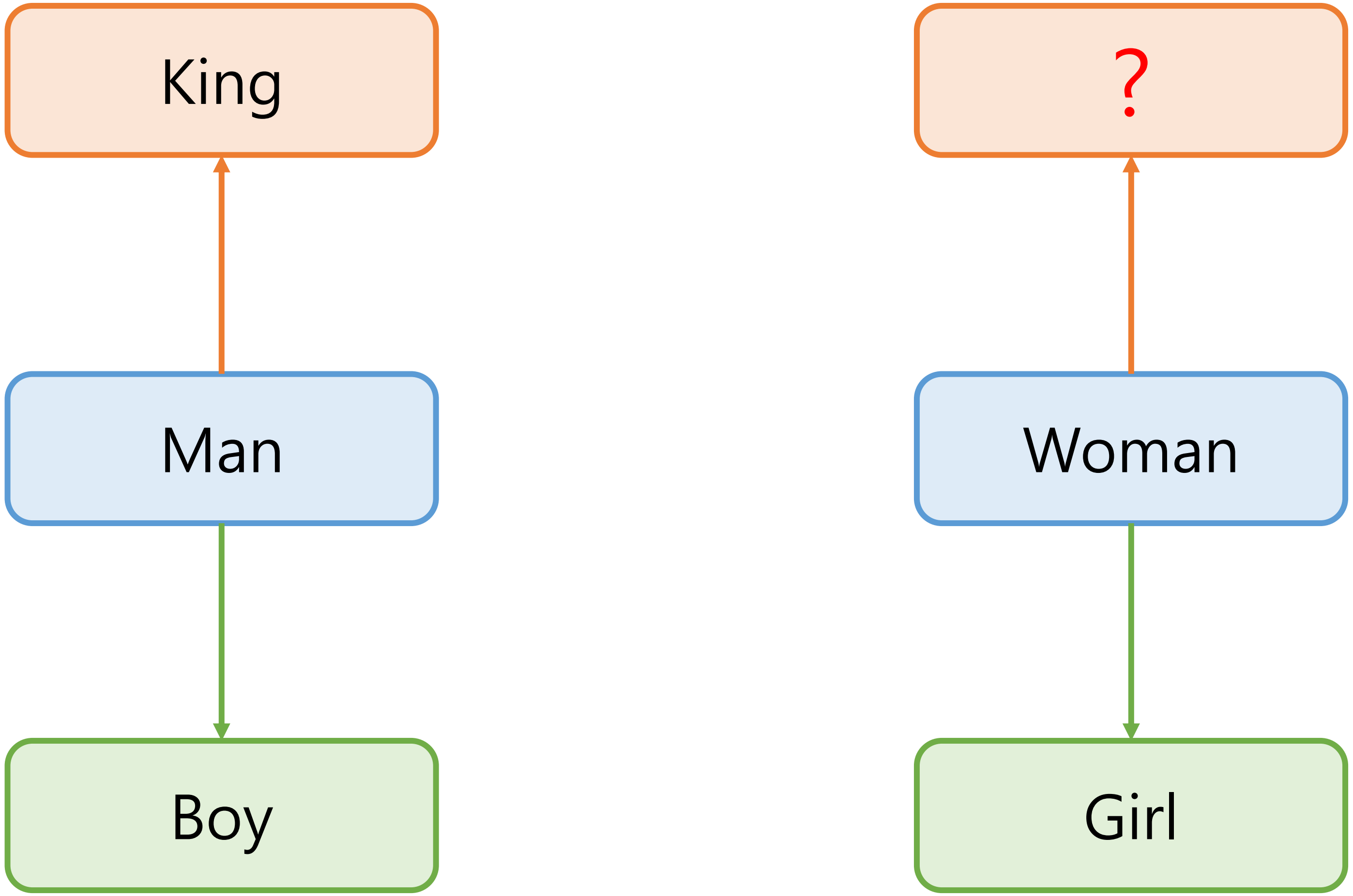
"man is to king as woman is to ...?"

man : king = woman : ?
(In Mathematical Expression)

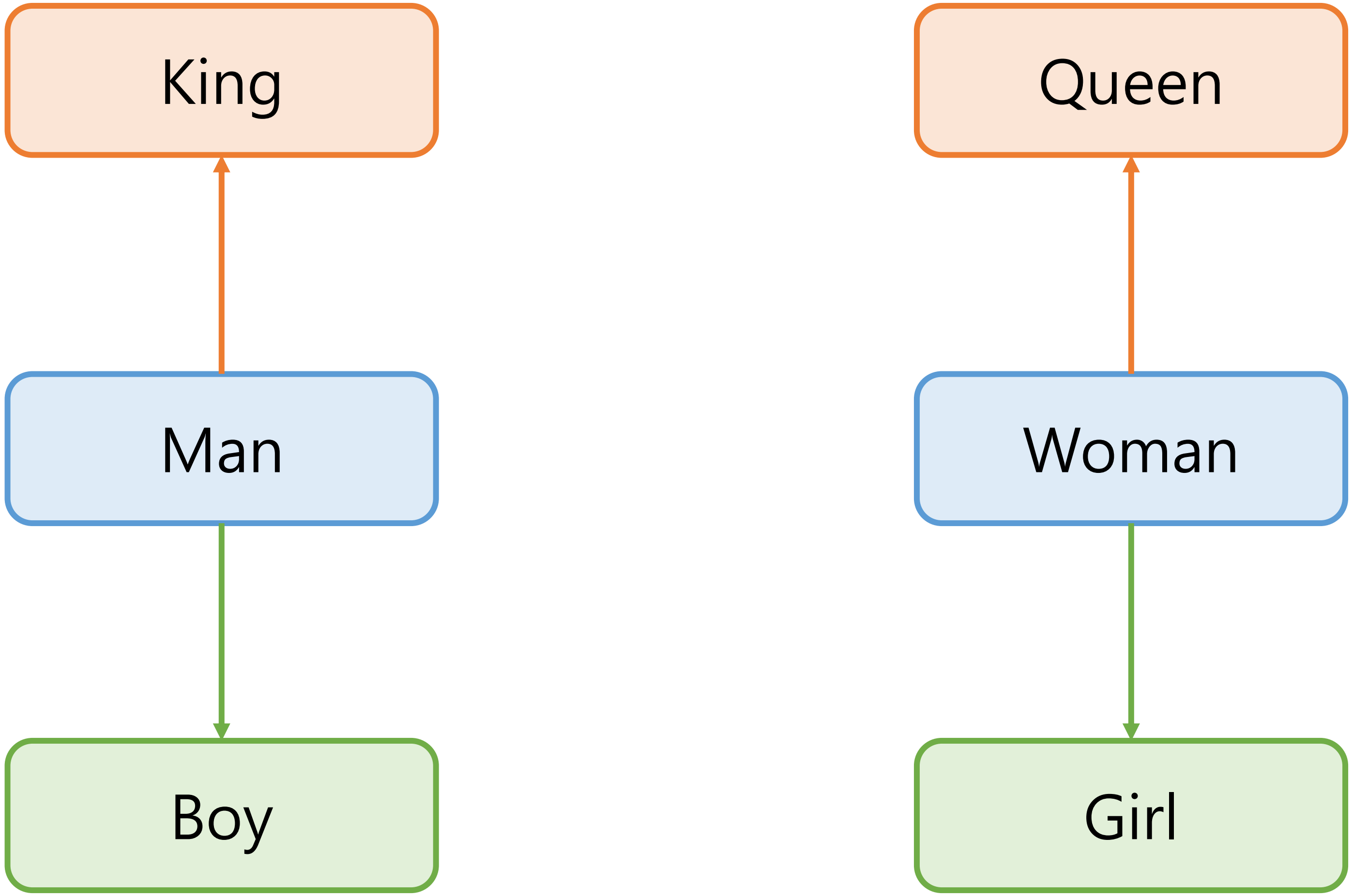
Introduction

-Analogies in Word Embeddings

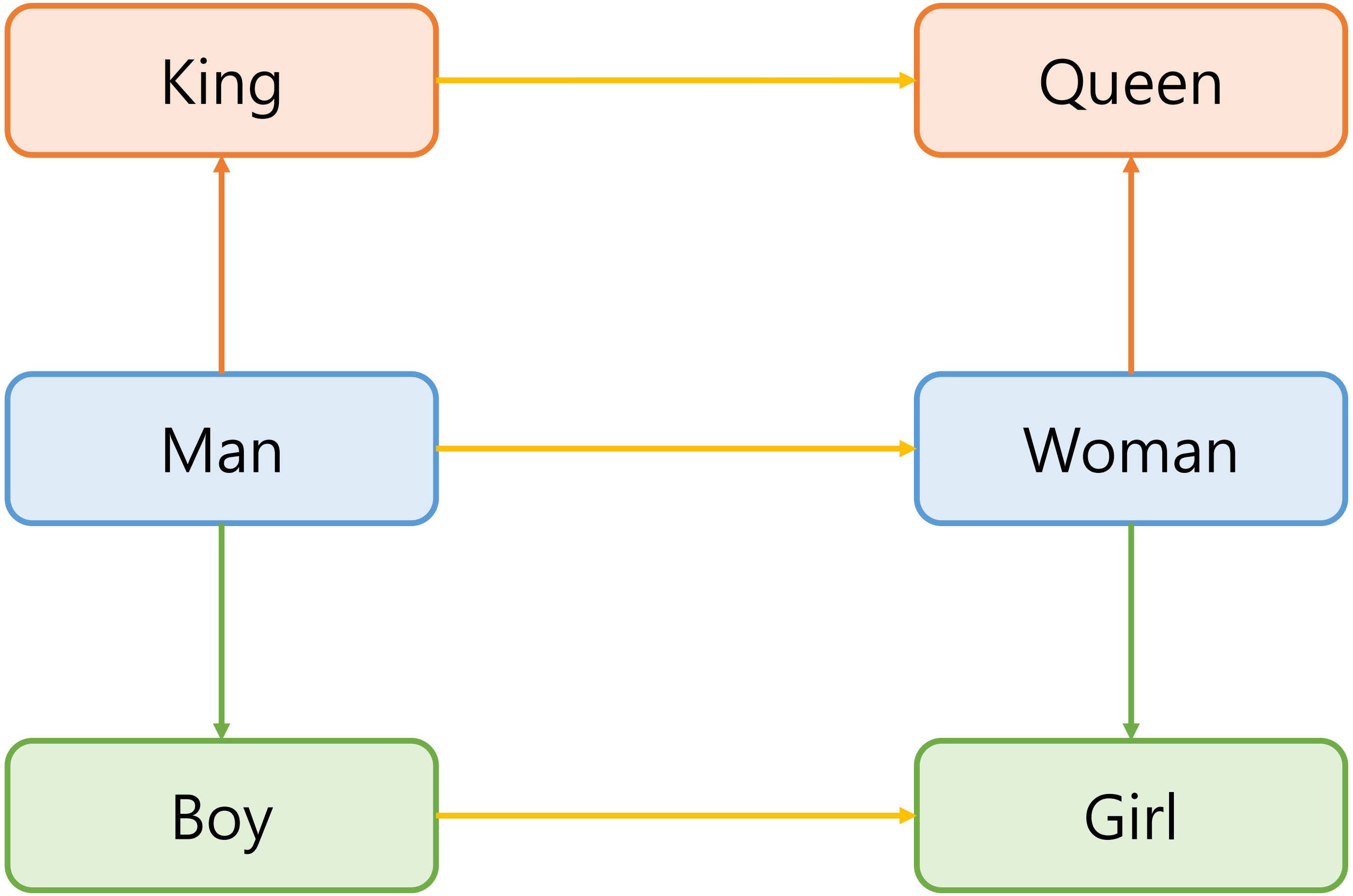
<Analogies in Natural Language>



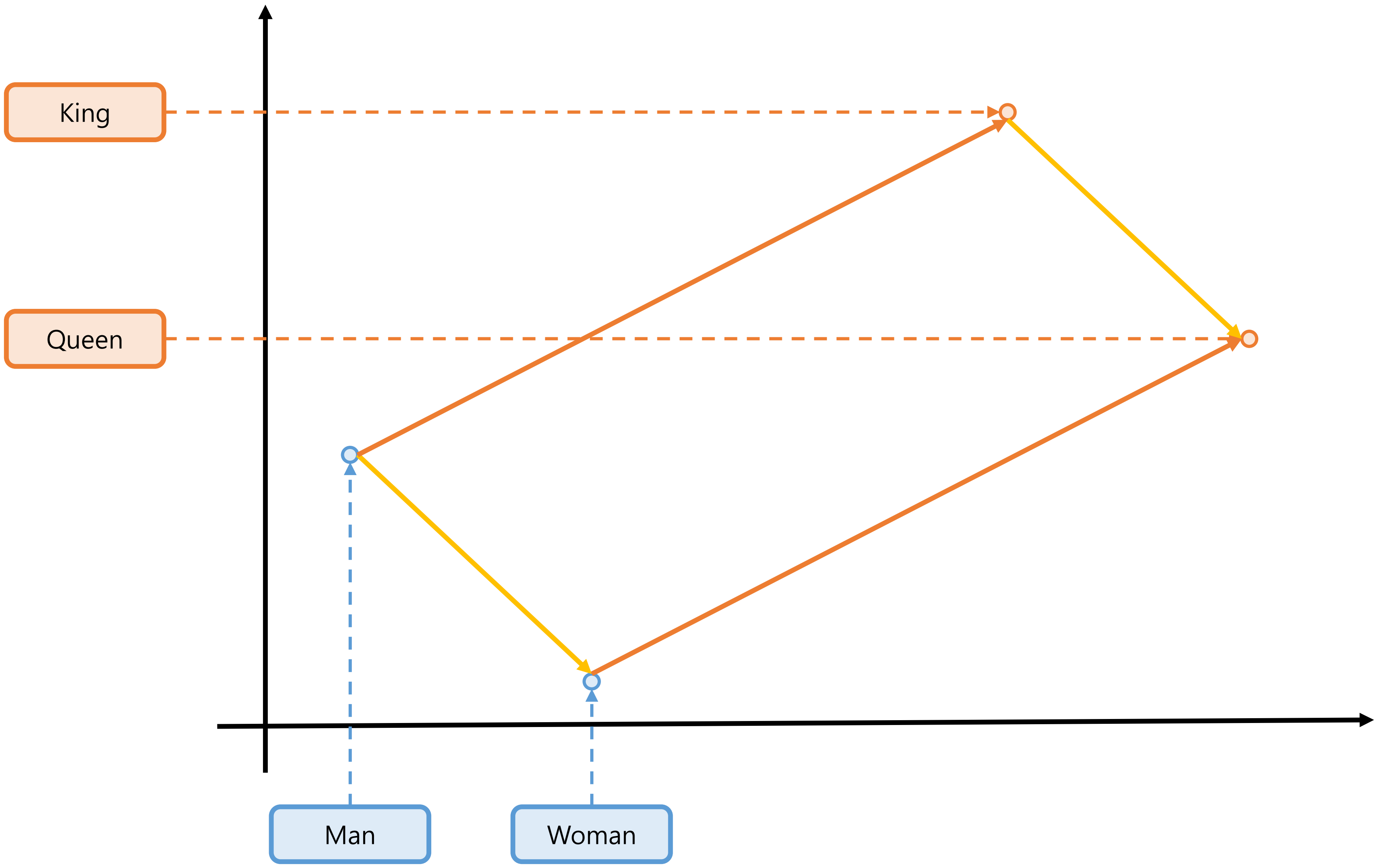
<Analogies in Natural Language>



<Analogies in Natural Language>

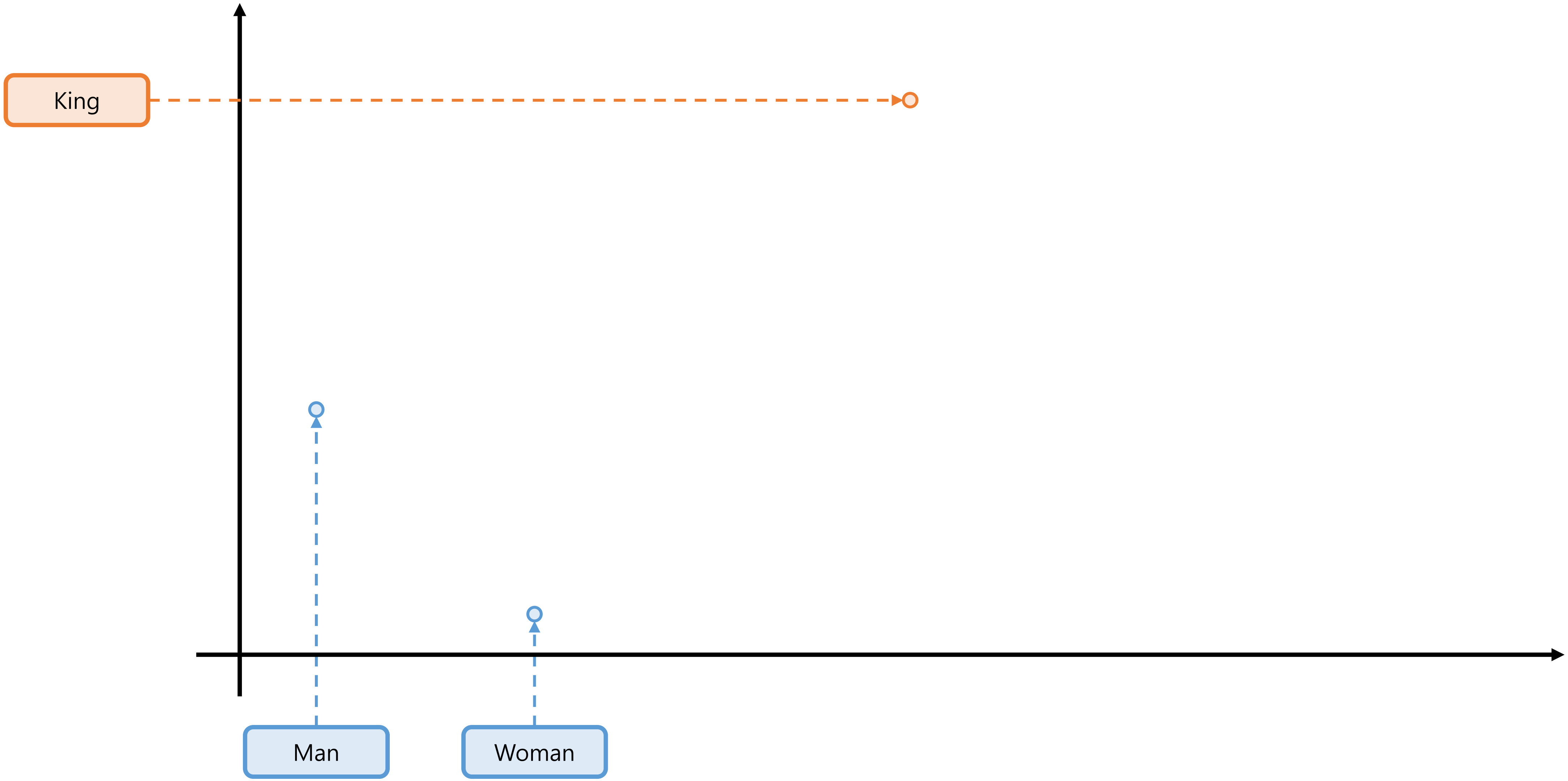


<Analogies in Vector Representation>



Introduction
-Analogies in Word Embeddings

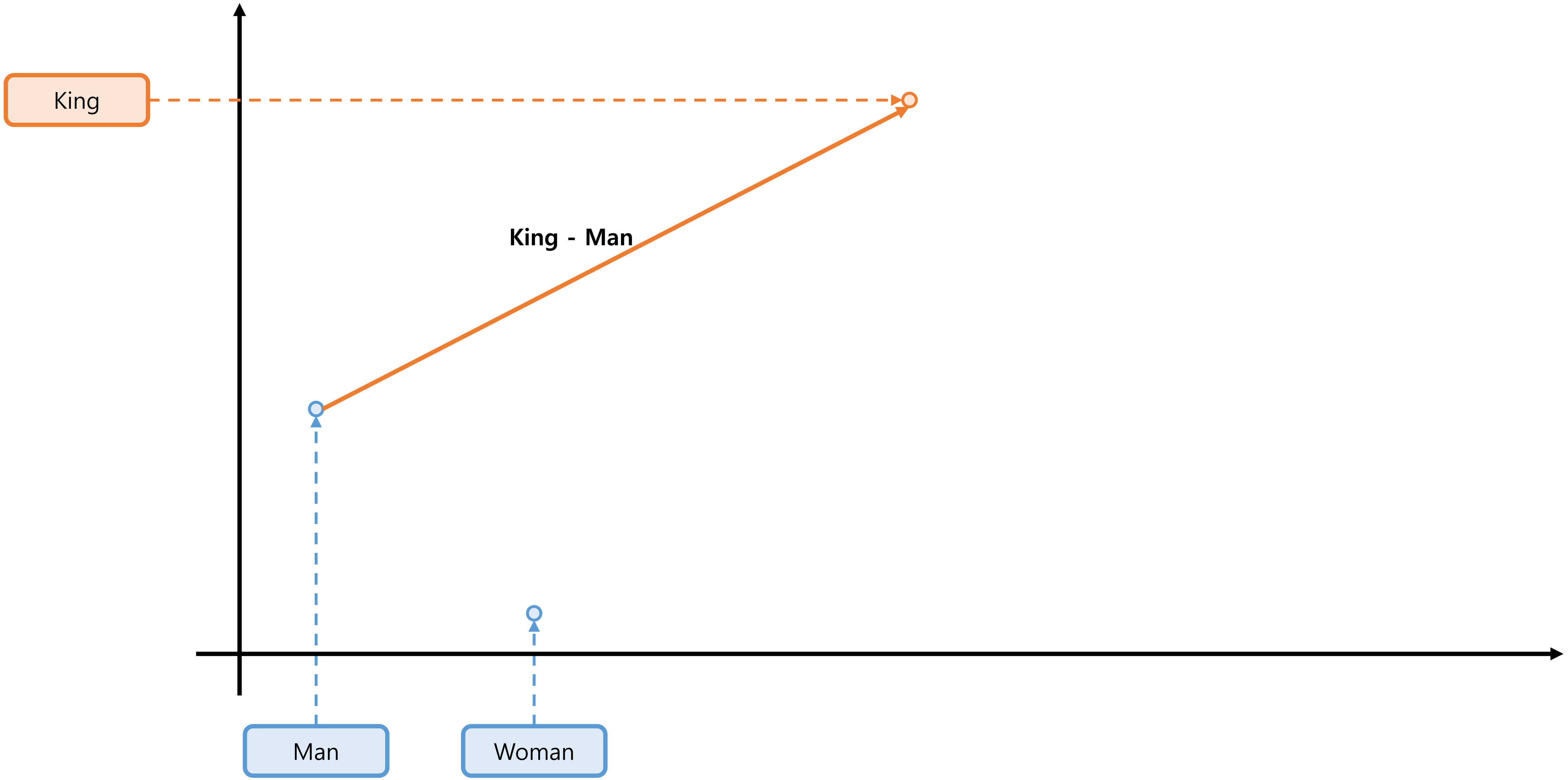
<Analogies in Vector Representation>



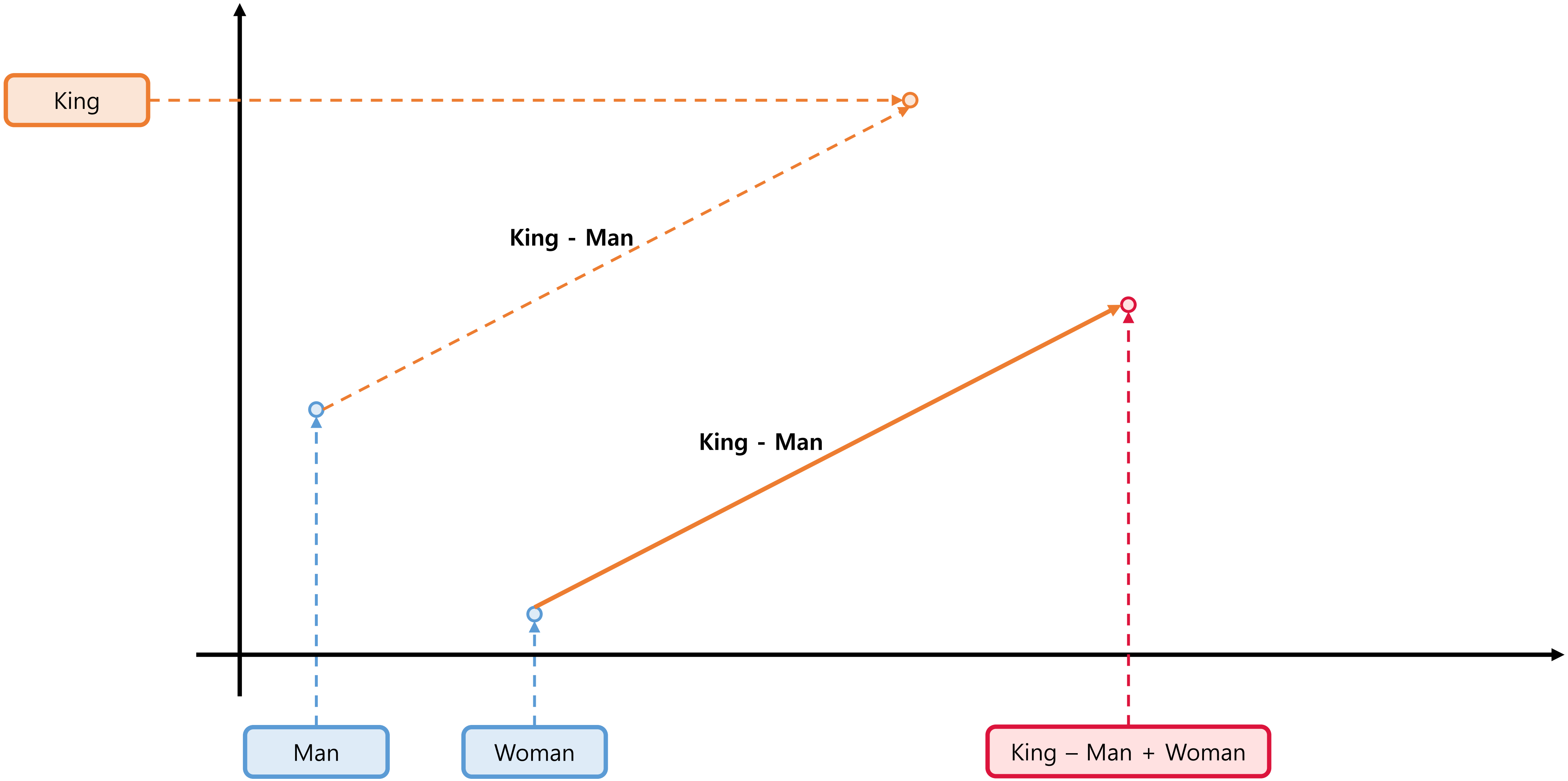
Introduction

-Analogies in Word Embeddings

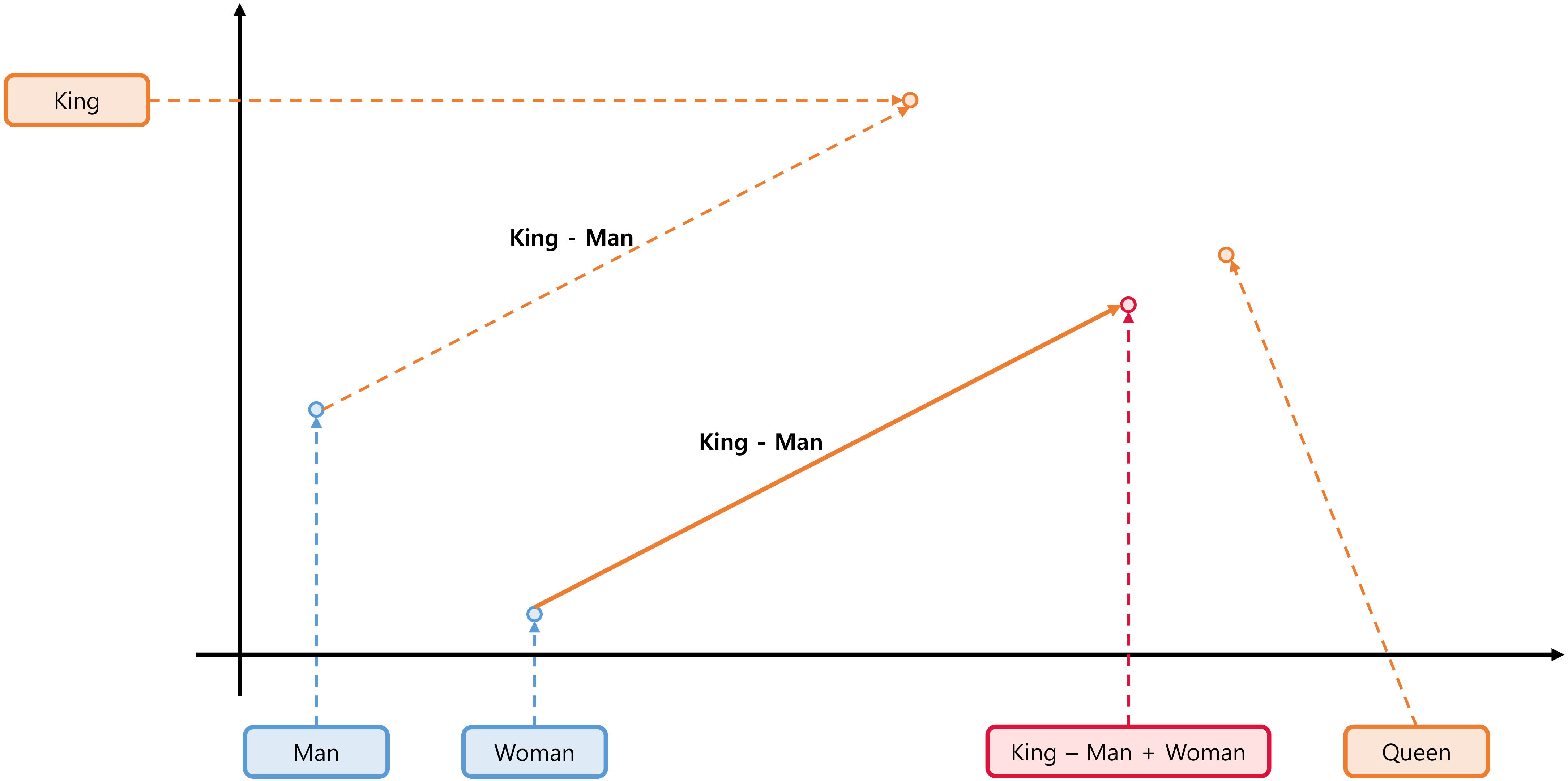
<Analogies in Vector Representation>



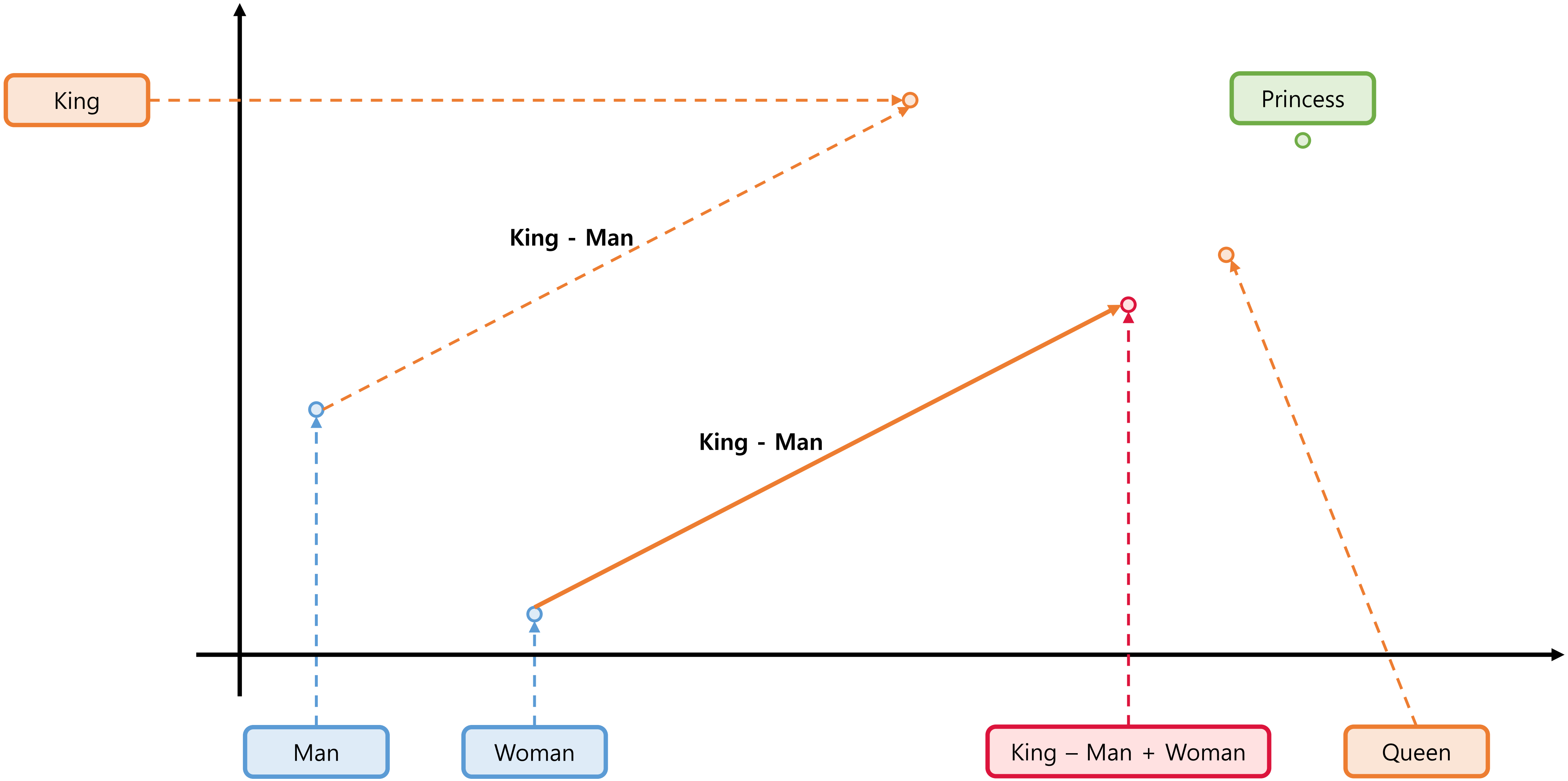
<Analogies in Vector Representation>



<Analogies in Vector Representation>



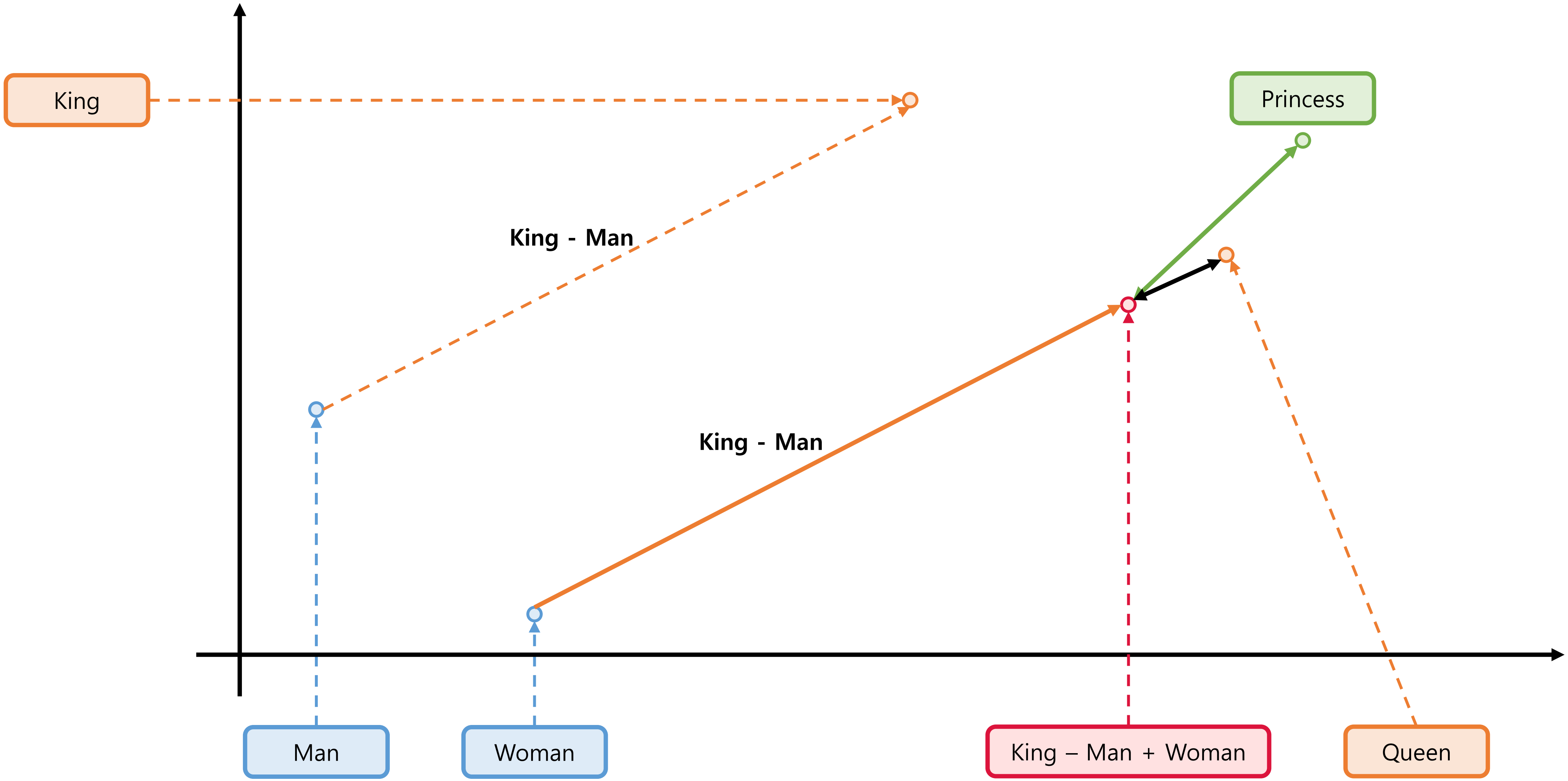
<Analogies in Vector Representation>



Introduction

-Analogies in Word Embeddings

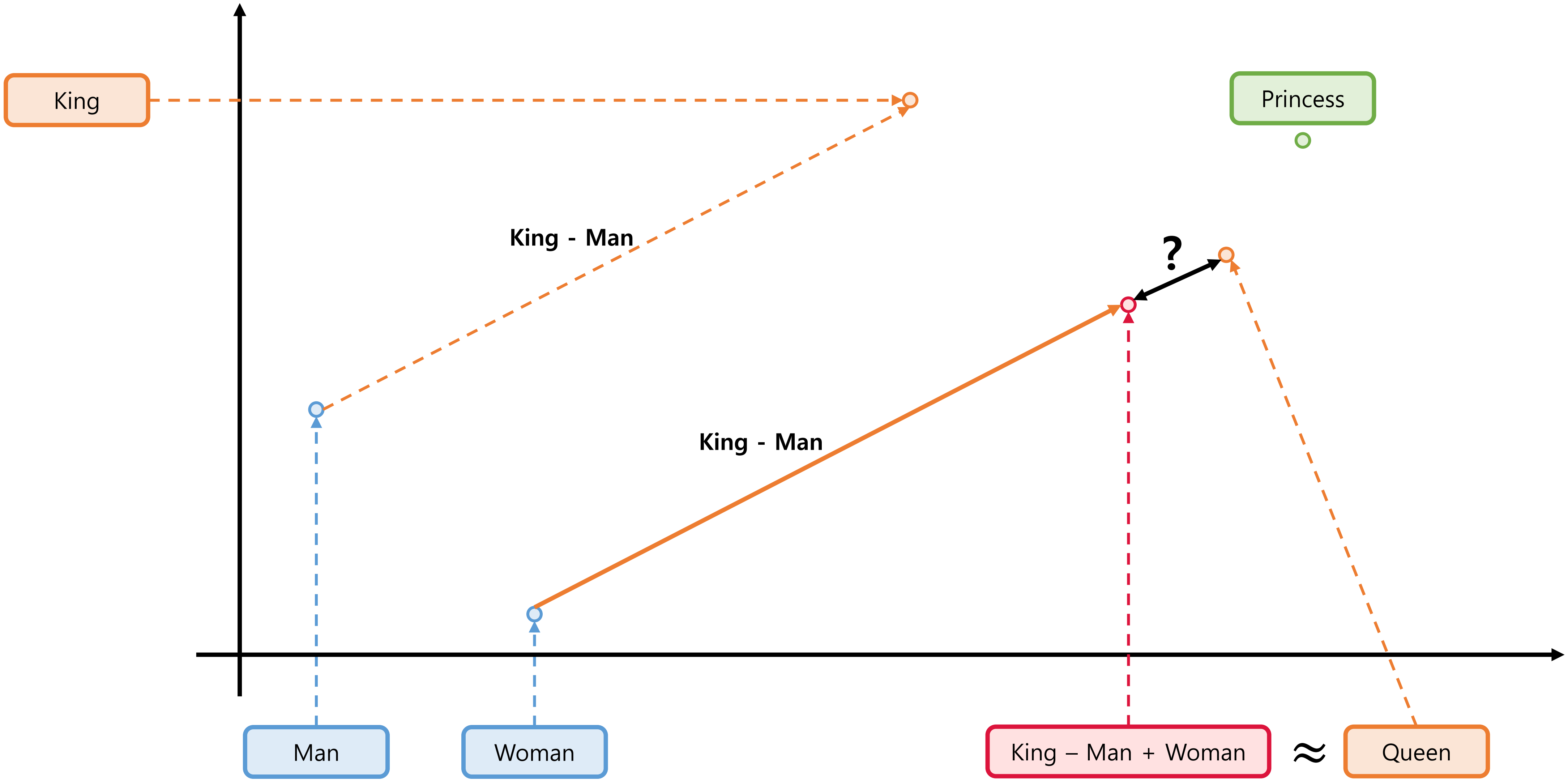
<Analogies in Vector Representation>



Introduction

-Analogies in Word Embeddings

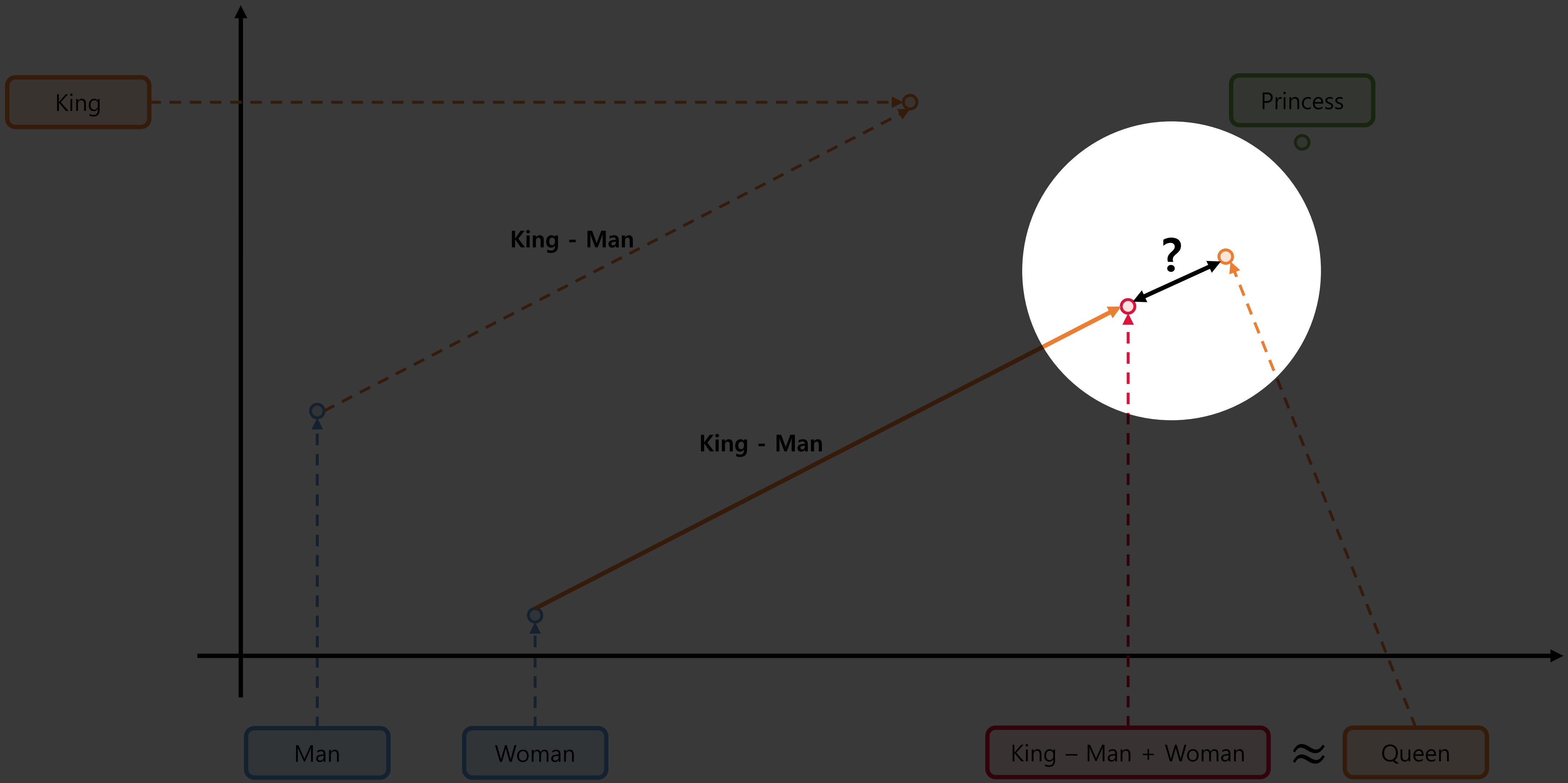
<Analogies in Vector Representation>



Introduction

-Analogies in Word Embeddings

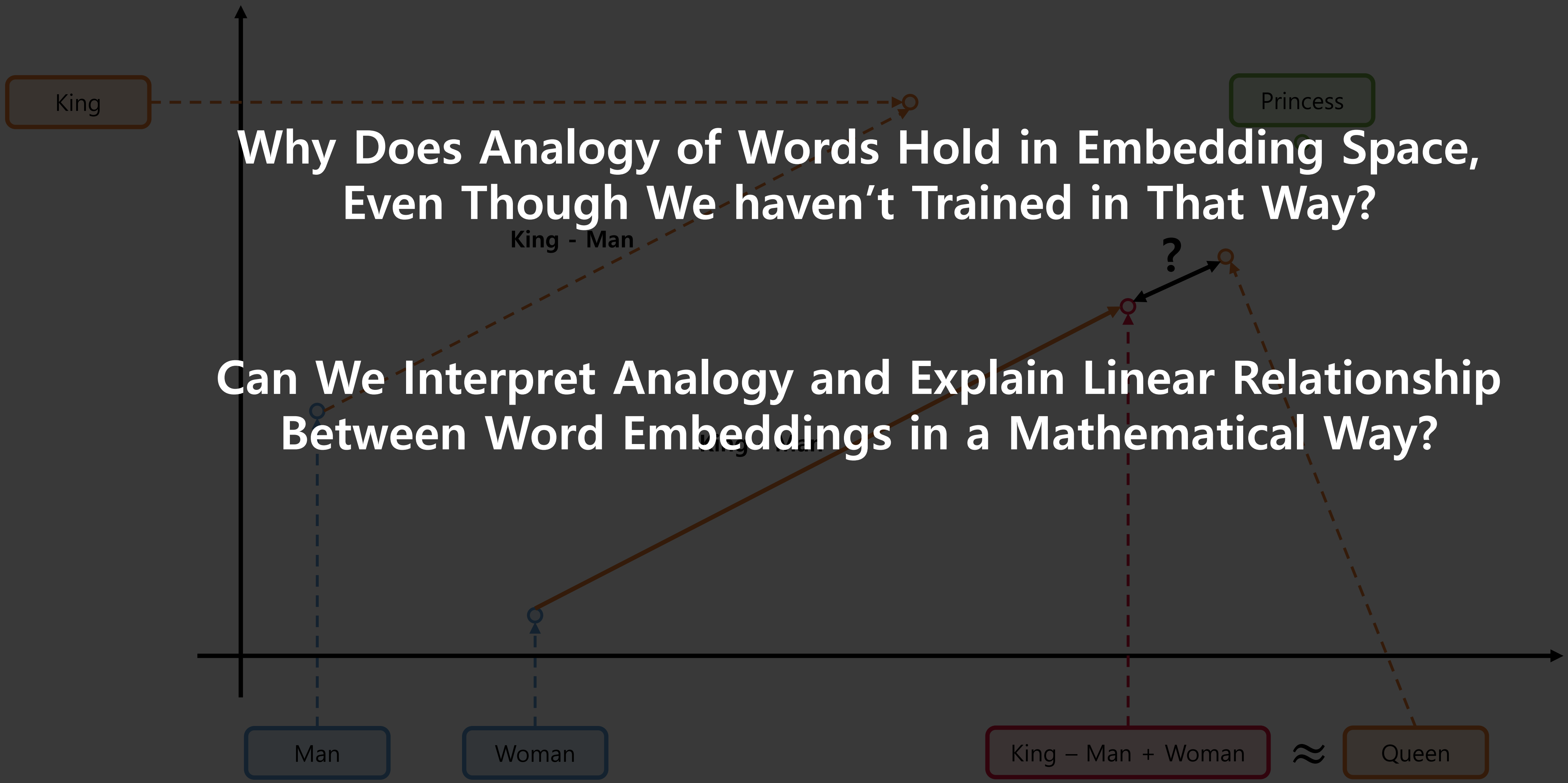
<Analogies in Vector Representation>



Introduction

-Analogies in Word Embeddings

<Analogies in Vector Representation>



Why Does Analogy of Words Hold in Embedding Space, Even Though We haven't Trained in That Way?

Can We Interpret Analogy and Explain Linear Relationship Between Word Embeddings in a Mathematical Way?

Analogies Explained: Towards Understanding Word Embeddings

Allen and Hospedales, 2019, ICML

"ICML Best Paper Honourable Mention"

Our key contributions are:

To provide the *<u>first rigorous proof</u>* of the linear relationship between word embeddings of analogies …

Analogies Explained: Towards Understanding Word Embeddings

Allen and Hospedales, 2019, ICML

“ICML Best Paper Honourable Mention”

Pre-requisites

-Word2Vec

<Word2Vec>

The | quick | brown | fox | jumps | over | the | lazy | dog

the	1	0	0	0	0	0	0	0
quick	0	1	0	0	0	0	0	0
brown	0	0	1	0	0	0	0	0
fox	0	0	0	1	0	0	0	0
jumps	0	0	0	0	1	0	0	0
over	0	0	0	0	0	1	0	0
lazy	0	0	0	0	0	0	1	0
dog	0	0	0	0	0	0	0	1

<Word2Vec>

The | quick | brown | fox | jumps | over | the | lazy | dog

(the, quick)
(the, brown)

<Training Samples>

[1, 0, 0, 0, 0, 0, 0, 0]

<Target Word>

[0, 1, 0, 0, 0, 0, 0, 0]
[0, 0, 1, 0, 0, 0, 0, 0]

<Context Word>

Pre-requisites
-Word2Vec

<Word2Vec>

The	quick	brown	fox	jumps	over	the	lazy	dog

(quick, the)

(quick, brown)

(quick, fox)

<Training Samples>

[0, 1, 0, 0, 0, 0, 0, 0]

<Target Word>

[1, 0, 0, 0, 0, 0, 0, 0]

[0, 0, 1, 0, 0, 0, 0, 0]

[0, 0, 0, 1, 0, 0, 0, 0]

<Context Word>

<Word2Vec>

The | quick | brown | fox | jumps | over | the | lazy | dog

(dog, the)

(dog, lazy)

<Training Samples>

[0, 0, 0, 0, 0, 0, 0, 1]

<Target Word>

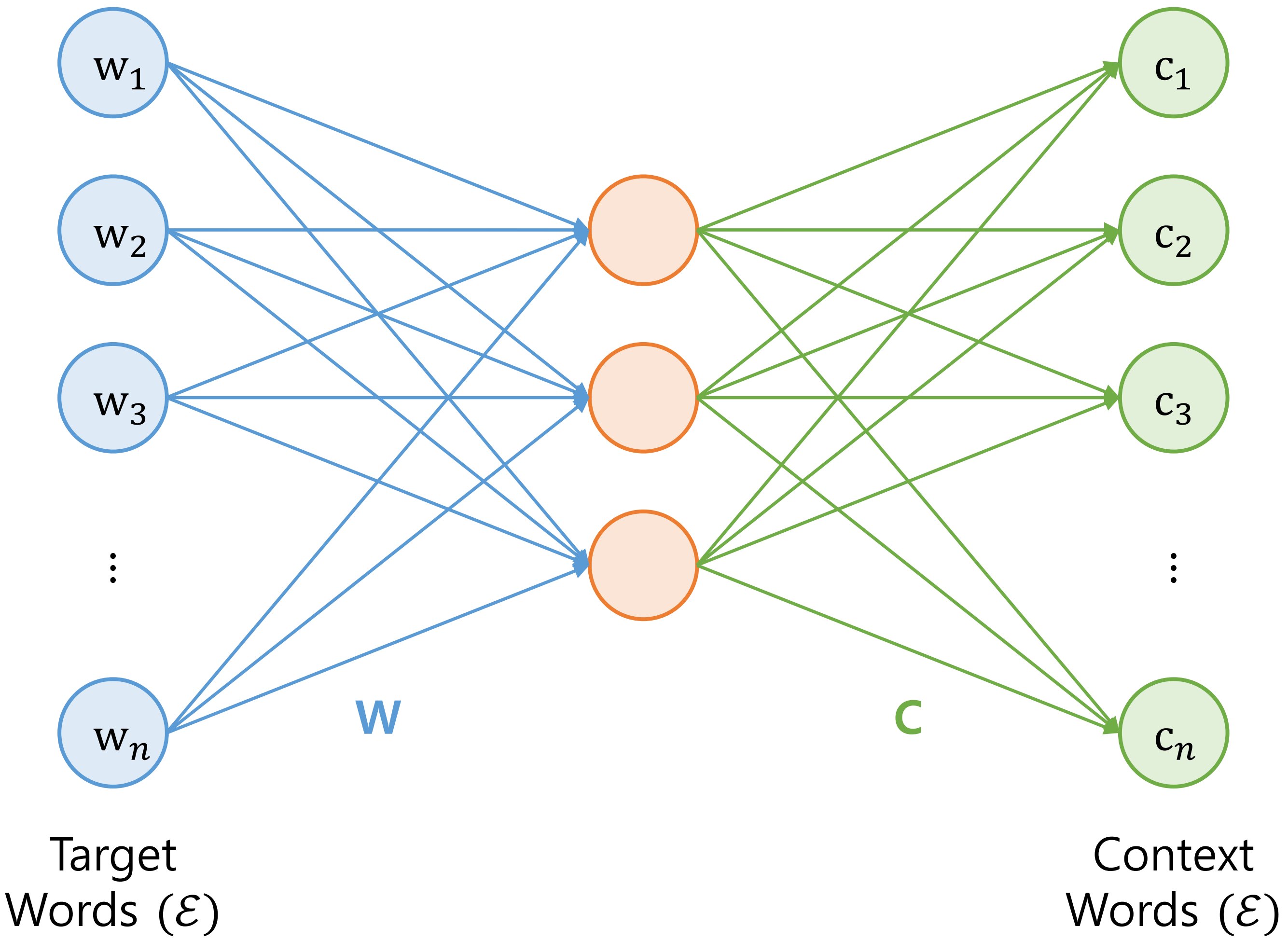
[1, 0, 0, 0, 0, 0, 0, 0]

[0, 0, 0, 0, 0, 0, 1, 0]

<Context Word>

Pre-requisites
-Word2Vec

<Word2Vec>



<Word2Vec>

0	0	0	1	0	0	0	0

<One-hot Encoding of “fox”>

×

0.1	0.4	0.3	1.5
0.1	0.2	1.6	0.3
0.8	0.3	0.4	1.4
1.4	0.4	0.6	0.9
0.6	1.4	0.9	1.4
0.6	0.2	0.5	1.6
1.0	1.4	1.4	0.9
1.5	1.2	0.7	0.7

<Matrix W^T>

=

1.4	0.4	0.6	0.9

<Embedding of “fox”>

Levy and Goldberg, Neural Word Embedding as Implicit Matrix Factorization, NIPS, 2014

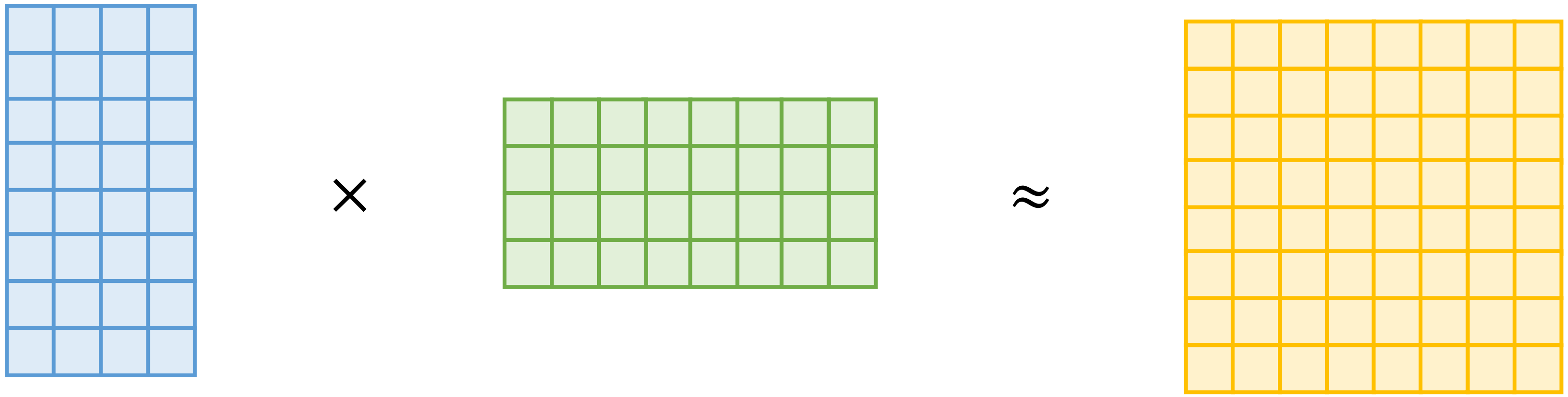
<Word Embedding as Matrix Factorization>

$$\mathrm{W}^T\mathrm{C} \approx \mathrm{PMI} - \log k = \mathrm{SPMI}$$

$$\mathrm{PMI}_{i,j} = \log \frac{p(w_i, c_j)}{p(w_i)p(c_j)}$$

$$w_i, c_j : column\ of\ \mathrm{W}, \mathrm{C}$$

$$where, k\ is\ chosen\ number\ of\ negative\ smaples$$



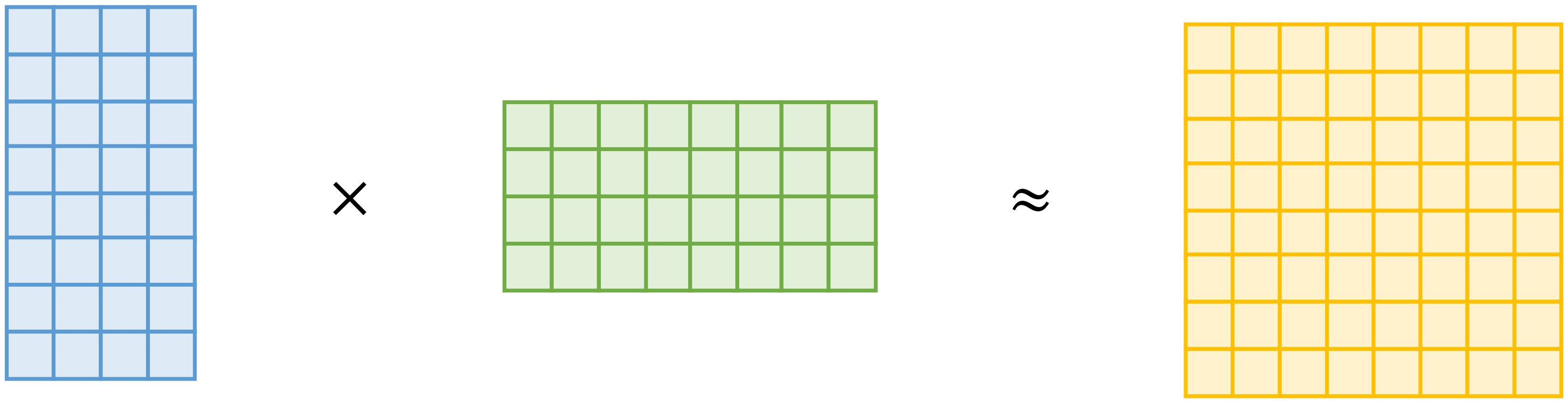
<Matrix W^{T}> <Matrix C> <SPMI>

Levy and Goldberg, Neural Word Embedding as Implicit Matrix Factorization, NIPS, 2014

<Word Embedding as Matrix Factorization>

$$\mathrm{W}^T\mathrm{C} \approx \mathrm{PMI} - \log k = \mathrm{SPMI}$$

$$\mathrm{PMI}_{i,j} = \log\frac{p(w_i, c_j)}{p(w_i)p(c_j)}$$



<Matrix W^T> <Matrix C> <SPMI>

"We analyze skip-gram with negative-samples and show that it is implicitly factorizing a word-context matrix, whose cells are the pointwise mutual information (PMI) of the respective word and context pairs, shifted by a global constant."
(Levy and Goldberg, 2014)

Preliminaries

- Impact of the Shift
- Reconstruction Error
- Zero Co-occurrence Counts

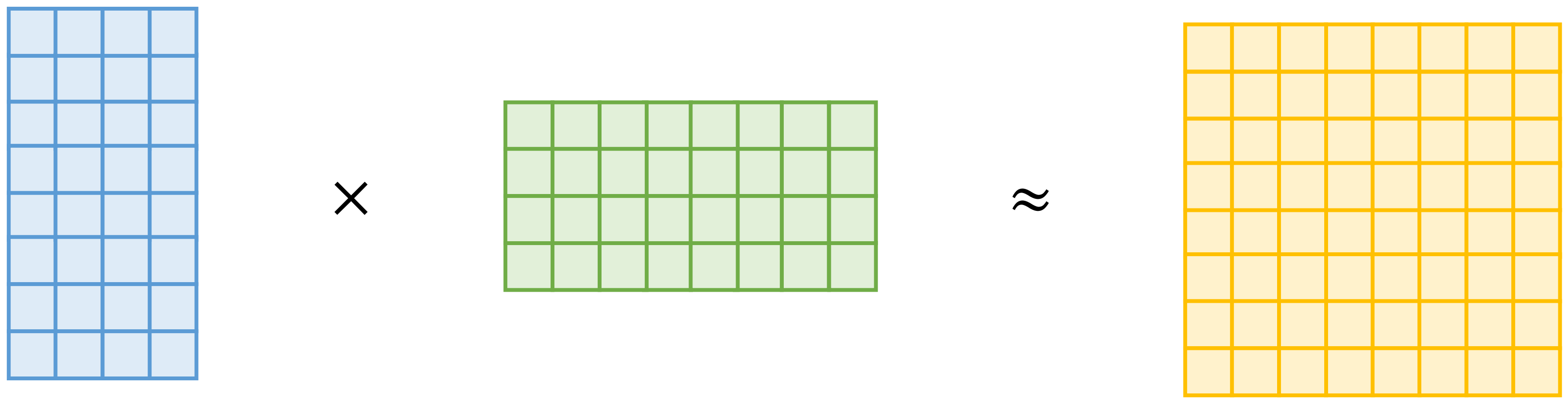
<Impact of the Shift>

$$\mathrm{W}^T\mathrm{C} \approx \mathrm{PMI} - \log k = \mathrm{SPMI}$$

$$\mathrm{PMI}_{i,j} = \log \frac{p(w_i, c_j)}{p(w_i)p(c_j)}$$

$$\mathrm{w}_i, \mathrm{c}_j : column\ of\ \mathrm{W}, \mathrm{C}$$

$$where,\ k\ is\ chosen\ number\ of\ negative\ smaples$$



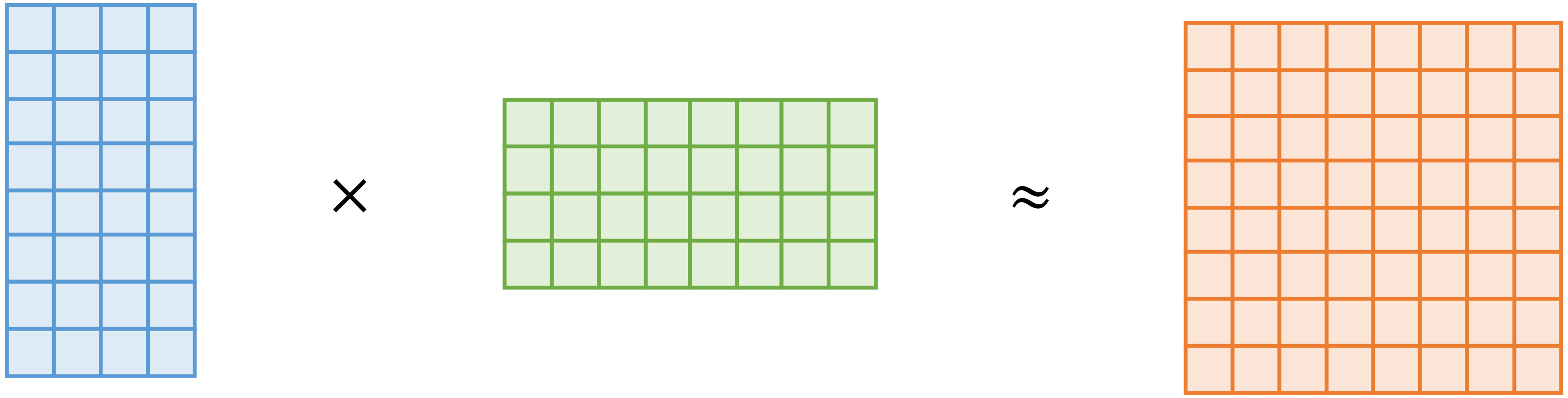
<Matrix W^T> <Matrix C> <SPMI>

<Impact of the Shift>

$$\mathrm{W}^T\mathrm{C} \approx \mathrm{PMI} - \log k = \mathrm{SPMI}$$

$$\mathrm{PMI}_{i,j} = \log \frac{p(w_i, c_j)}{p(w_i)p(c_j)}$$

$$w_i, c_j : column\ of\ \mathrm{W}, \mathrm{C}$$



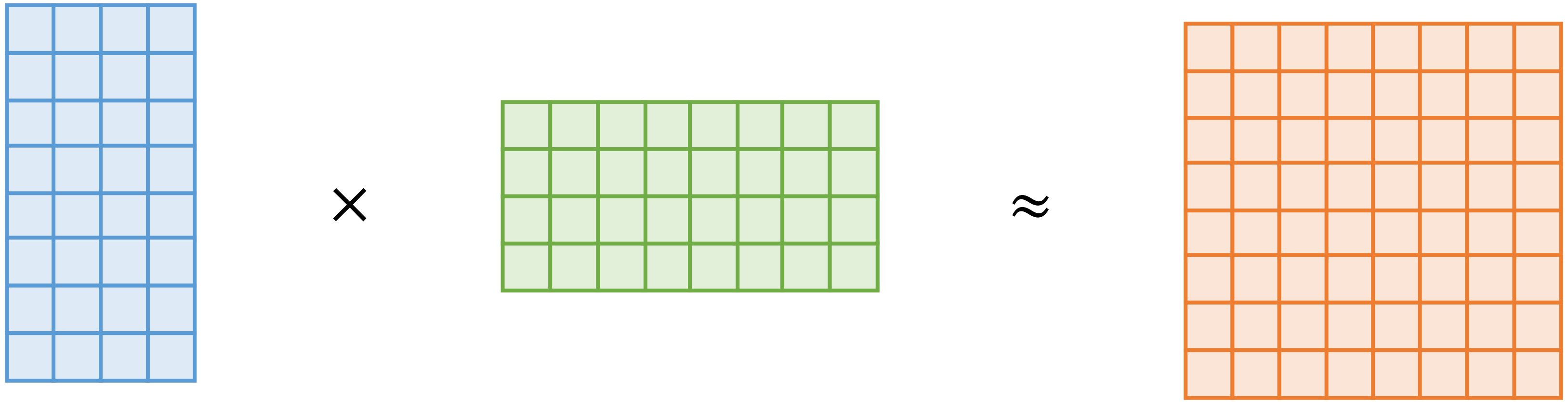
<Matrix W^T> <Matrix C> <PMI>

"It is observed that adjusting the Word2Vec algorithm to avoid any direct impact of the *shift* improves embedding performance."
(Le, 2017)

<Reconstruction Error>

$$W^T C \approx \mathrm{PMI}$$

$$w_i, c_j : column\ of\ \mathrm{W}, \mathrm{C}$$



$<Matrix\ \mathrm{W}^T>$ $<Matrix\ \mathrm{C}>$ $<\mathrm{PMI}>$

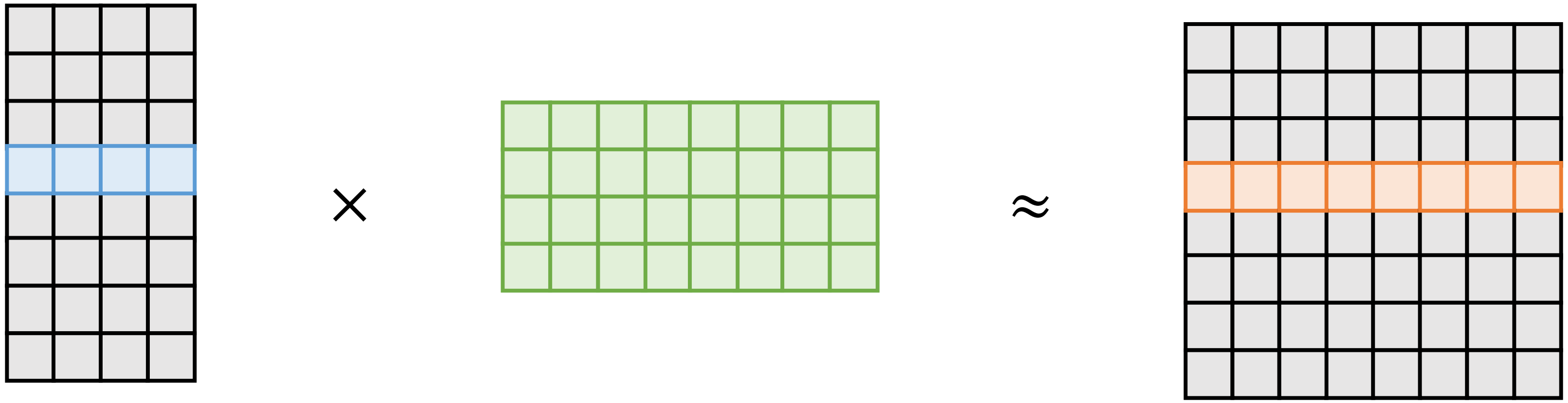
<Reconstruction Error>

$$\mathrm{W}^T\mathrm{C} \approx \mathrm{PMI}$$

$$\mathrm{w}_i, \mathrm{c}_j : \textit{column of } \mathrm{W}, \mathrm{C}$$

$$\mathbf{w}_i^T\mathbf{C} \approx \mathbf{PMI}_i$$

$$\textit{where } \mathrm{PMI}_i : \boldsymbol{row} \textit{ of } \mathrm{PMI}$$



$< \mathrm{w}_i^T >$ $\times$ $< \textit{Matrix } \mathrm{C} >$ $\approx$ $< \mathrm{PMI}_i >$

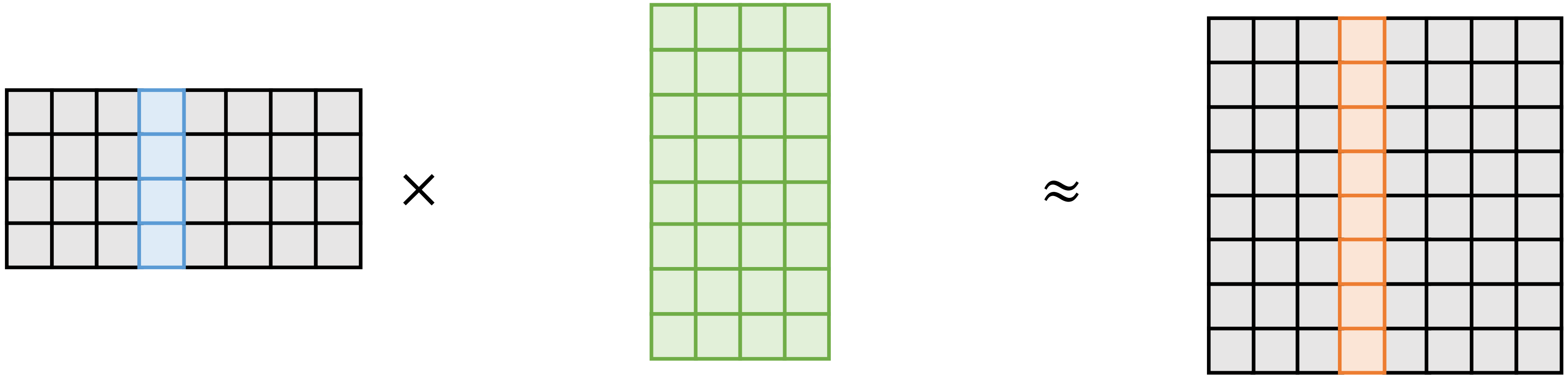
<Reconstruction Error>

$$\mathrm{W}^T\mathrm{C} \approx \mathrm{PMI}$$

$$\mathrm{w}_i, \mathrm{c}_j : \textit{column of } \mathrm{W}, \mathrm{C}$$

$$\mathbf{w}_i\mathbf{C}^T \approx \mathbf{PMI}_i$$

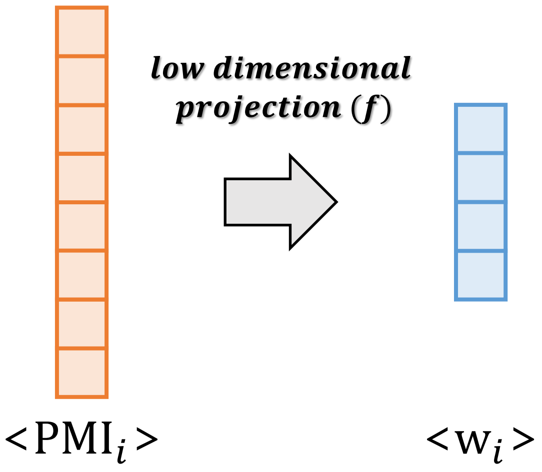
$$\textit{where } \mathrm{PMI}_i: \boldsymbol{column} \textit{ of } \mathrm{PMI}$$

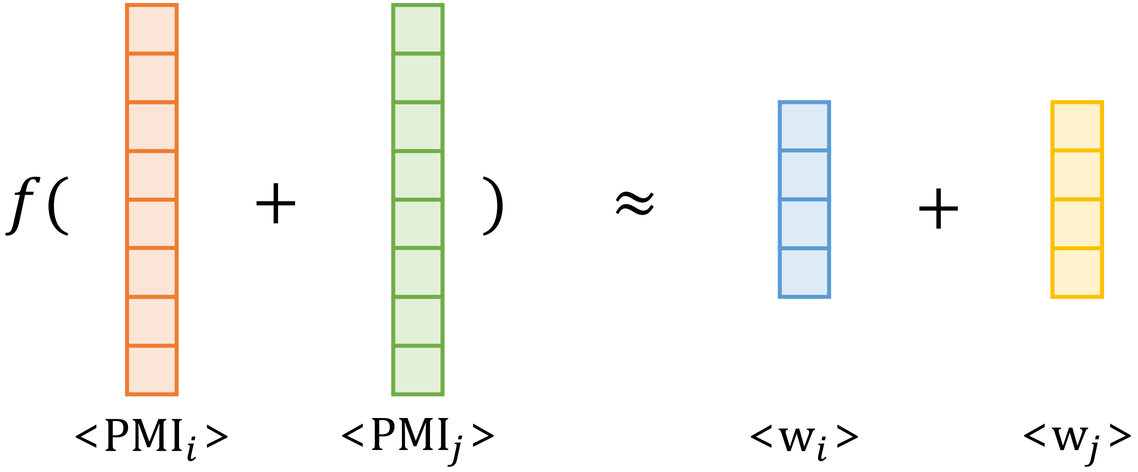


$<\mathrm{w}_i>$ $\quad$ $<\mathit{Matrix}\ \mathrm{C}^\mathrm{T}>$ $\quad$ $<\mathrm{PMI}_i>$

<Reconstruction Error>

Assumption A2. Letting PMI_k denote the k^{th} *column* of $\mathrm{PMI} \in \mathbb{R}^{n \times n}$, the projection $f\colon \mathbb{R}^n \to \mathbb{R}^d$, $f(\mathrm{PMI}_i) = \mathrm{w}_i$ is approximately homomorphic with respect to addition, $i.e.$ $f\left(\mathrm{PMI}_i + \mathrm{PMI}_j\right) \approx f(\mathrm{PMI}_i) + f(\mathrm{PMI}_j)$

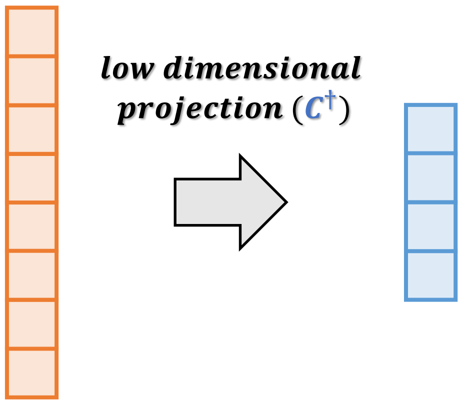


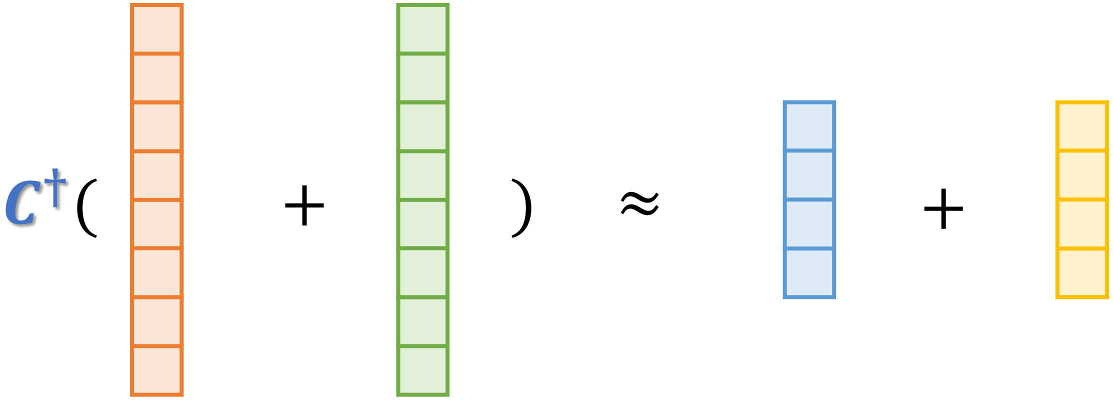


"***A2*** means that whatever factorization method used, linear relationships between columns of PMI are sufficiently preserved by columns of W"

<Reconstruction Error>

Assumption A2. Letting PMI_k denote the k^{th} $column$ of $\mathrm{PMI} \in \mathbb{R}^{n \times n}$, the projection $\boldsymbol{C}^{\dagger}\colon \mathbb{R}^n \to \mathbb{R}^d$, $f(\mathrm{PMI}_i) = \mathrm{w}_i$ is approximately homomorphic with respect to addition, $i.e.\, \boldsymbol{C}^{\dagger}(\mathrm{PMI}_i + \mathrm{PMI}_j) \approx \boldsymbol{C}^{\dagger}(\mathrm{PMI}_i) + f(\mathrm{PMI}_j)$



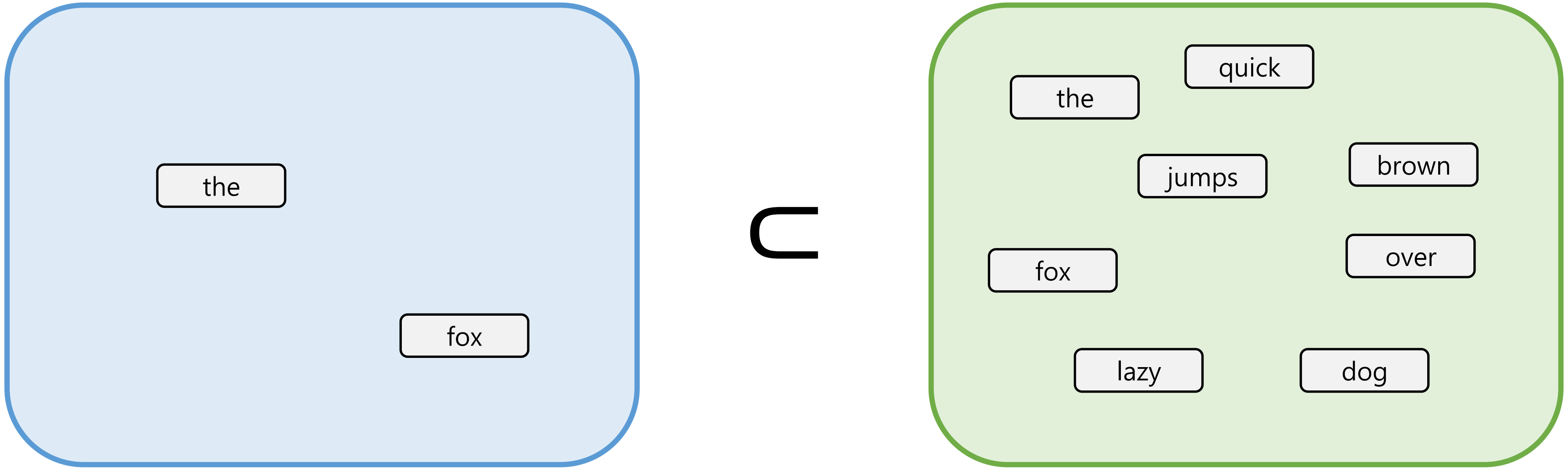


"For example, minimizing a **least squares loss function** gives linear projection $\mathrm{w}_i = f_{LSQ}(\mathrm{PMI}_i) = C^{\dagger}\mathrm{PMI}_i$
Where $\boldsymbol{C}^{\dagger} = \left(\boldsymbol{C}\boldsymbol{C}^{\boldsymbol{T}}\right)^{-\boldsymbol{1}}\boldsymbol{C}$. We write $\boldsymbol{f(\cdot) = C^{\dagger}(\cdot)}$ to emphasize linearity of the relationship"

<Zero Co-occurrence counts>

$\mathcal{W}$: *small word set*

$\mathcal{E}$: *fixed size dictionary*

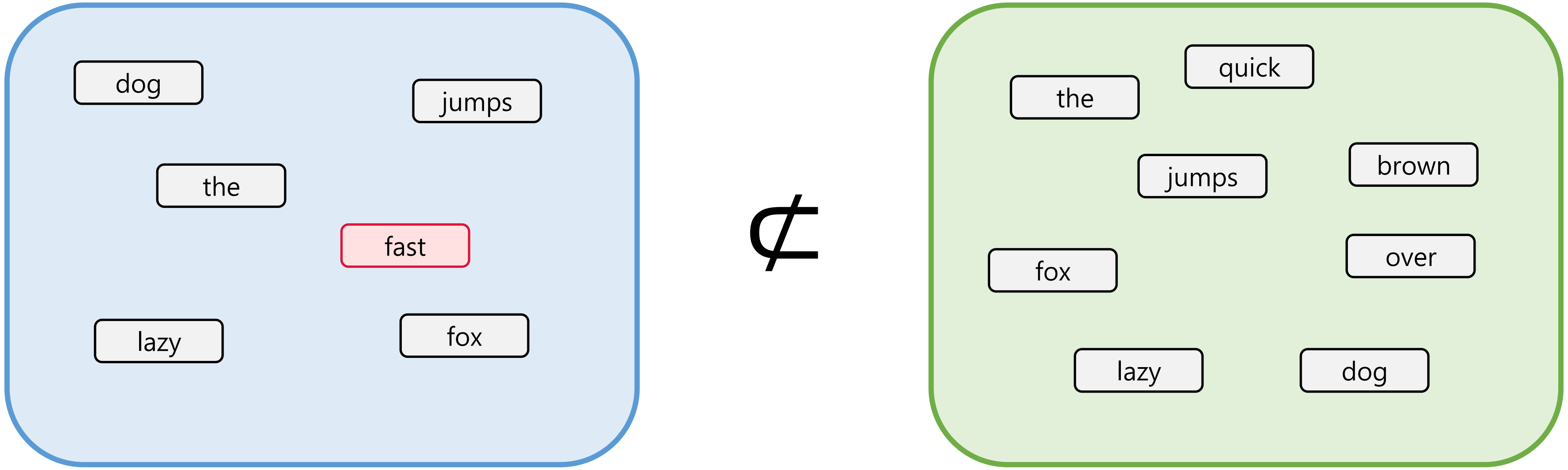


$< wordset\ \mathcal{W} >$ $< dictionary\ \mathcal{E} >$

<Zero Co-occurrence counts>

$\mathcal{W}$: *small word set*

$\mathcal{E}$: *fixed size dictionary*

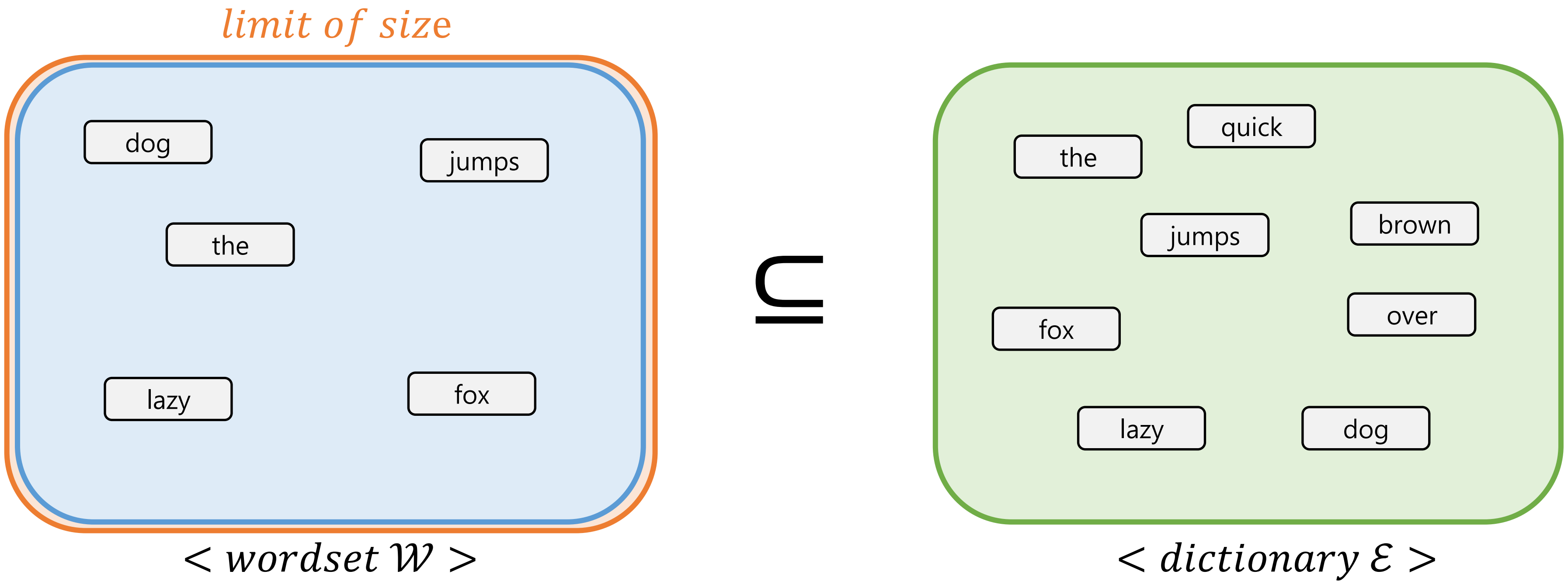


$<$ *wordset* $\mathcal{W}$ $>$ $<$ *dictionary* $\mathcal{E}$ $>$

<Zero Co-occurrence counts>

$\mathcal{W}$: *small word set*

$\mathcal{E}$: *fixed size dictionary*

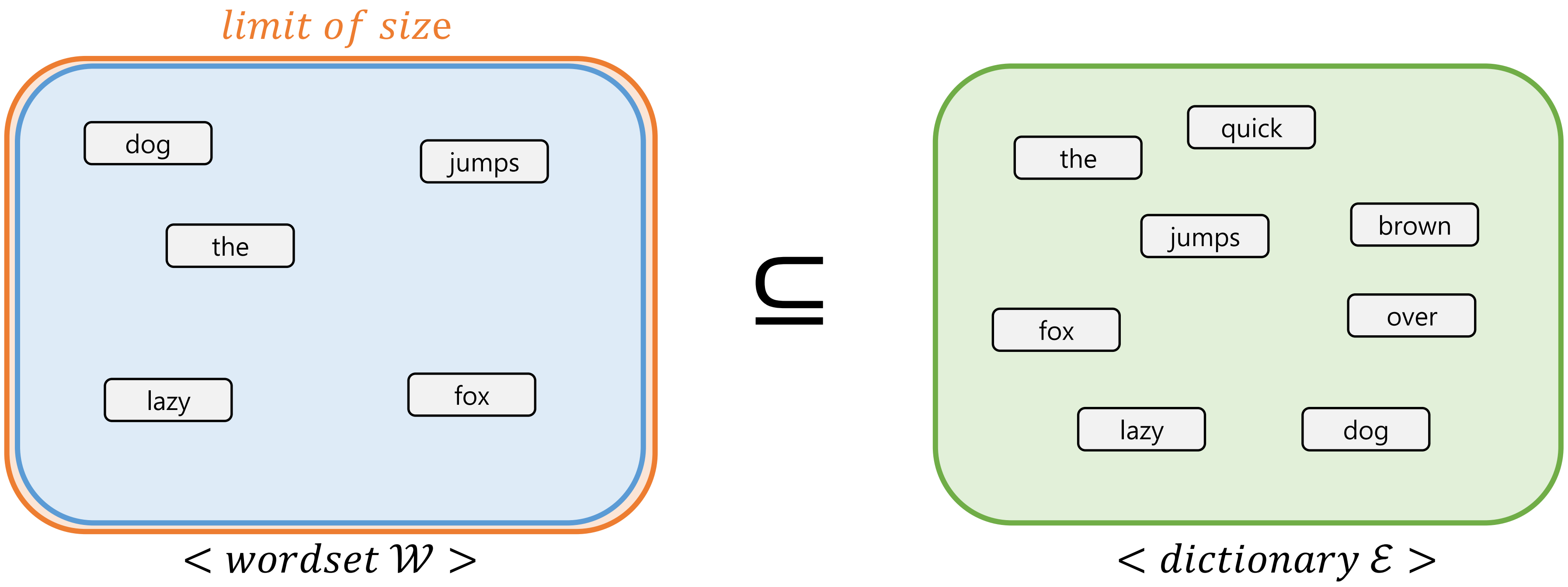


<Zero Co-occurrence counts>

Assumption A3. $p(\mathcal{W}) > 0, \forall \mathcal{W} \subseteq \mathcal{E}, |\mathcal{W}| < l$

$\mathcal{W}$: *small word set*

$\mathcal{E}$: *fixed size dictionary*



Proof

- Paraphrase
- Analogy

Proof

- Paraphrase
- Analogy

(Slides) Analogies Explained: Towards Understanding Word Embeddings, ICML, 2019

<Routemap>

"man is to king as woman is to queen"

$\Updownarrow$

man *transforms to* king as woman *transforms to* queen

$\Updownarrow$

{woman, king} *paraphrases* {man, queen}

$\Downarrow$

$$\mathrm{PMI}_{king} - \mathrm{PMI}_{man} + \mathrm{PMI}_{woman} \approx \mathrm{PMI}_{queen}$$

$\Downarrow$

$$\mathrm{w}_{king} - \mathrm{w}_{man} + \mathrm{w}_{woman} \approx \mathrm{w}_{queen}$$

Proof

-Routemap

<Routemap>

"man is to king as woman is to queen"

$\Updownarrow$

man *transforms to* king as woman *transforms to* queen

$\Updownarrow$

{woman, king} *paraphrases* {man, queen}

$\Downarrow$

$$\mathrm{PMI}_{king} - \mathrm{PMI}_{man} + \mathrm{PMI}_{woman} \approx \mathrm{PMI}_{queen}$$

$\Downarrow$

$$\mathbf{w}_{king} - \mathbf{w}_{man} + \mathbf{w}_{woman} \approx \mathbf{w}_{queen}$$

semantic

geometric

Proof
-Routemap

<Routemap>

"man is to king as woman is to queen"

$\Updownarrow$

man *transforms to* king as woman *transforms to* queen

$\Updownarrow$

{woman, king} *paraphrases* {man, queen}

$\Downarrow$

$$\mathrm{PMI}_{king} - \mathrm{PMI}_{man} + \mathrm{PMI}_{woman} \approx \mathrm{PMI}_{queen}$$

$\Downarrow$ $\quad \mathbf{PMI}_i \approx \mathbf{w}_i^T \boldsymbol{C}$

$$\mathrm{w}_{king} - \mathrm{w}_{man} + \mathrm{w}_{woman} \approx \mathrm{w}_{queen}$$

semantic

geometric

Proof
-Routemap

<Routemap>

"man is to king as woman is to queen"

$\Updownarrow$

man *transforms to* king as woman *transforms to* queen

$\Updownarrow$

{woman, king} *paraphrases* {man, queen}

$\Downarrow$

$$\mathrm{PMI}_{king} - \mathrm{PMI}_{man} + \mathrm{PMI}_{woman} \approx \mathrm{PMI}_{queen}$$

$\Downarrow$ $\quad \mathbf{PMI}_i \approx \mathbf{w}_i^T \boldsymbol{C}$

$$\mathrm{w}_{king} - \mathrm{w}_{man} + \mathrm{w}_{woman} \approx \mathrm{w}_{queen}$$

semantic

geometric

<Paraphrase>

$$\mathcal{W} = \{w_1, \dots, w_m\} \subseteq \mathcal{E}$$

$w_* \in \mathcal{E}$ $\boldsymbol{paraphrases}$ $\mathcal{W}$, if w_* and $\mathcal{W}$ are semantically interchangeable within the text

<Paraphrase>

$$\mathcal{W} = \{w_1, \ldots, w_m\} \subseteq \mathcal{E}$$

$w_* \in \mathcal{E}$ $\boldsymbol{paraphrases}$ $\mathcal{W}, if$ w_* and $\mathcal{W}$ are semantically interchangeable within the text

"The need for profit **may** push up prices"

<Sentence 1>

$\approx$

"The need for profit **is likely to** push up prices"

<Sentence 2>

<Paraphrase>

$$\mathcal{W} = \{w_1, \dots, w_m\} \subseteq \mathcal{E}$$

$w_* \in \mathcal{E}$ $\boldsymbol{paraphrases}$ $\mathcal{W}$, if w_* and $\mathcal{W}$ are semantically interchangeable within the text

"The need for profit **may** push up prices"

<Sentence 1>

$\approx$

"The need for profit **is likely to** push up prices"

<Sentence 2>

may $\approx$ **be likely to**

Semantically
Interchangeable

<Paraphrase>

$$\mathcal{W} = \{w_1, \dots, w_m\} \subseteq \mathcal{E}$$

$w_* \in \mathcal{E}$ ***paraphrases*** $\mathcal{W}$, *if* w_* and $\mathcal{W}$ are semantically interchangeable within the text

"The need for profit **may** push up prices"

<Sentence 1>

$\approx$

"The need for profit **is likely to** push up prices"

<Sentence 2>

may $(\boldsymbol{w_*})$ paraphrases **be likely to** $(\boldsymbol{\mathcal{W}})$

Proof
-Paraphrase

<Paraphrase>

$$\mathcal{W} = \{w_1, \dots, w_m\} \subseteq \mathcal{E}$$

$w_* \in \mathcal{E}$ $\boldsymbol{paraphrases}$ $\mathcal{W}$, if w_* and $\mathcal{W}$ are semantically interchangeable within the text

"**The need for profit** may **push up prices**"

<Sentence 1>

$\approx$

"**The need for profit** is likely to **push up prices**"

<Sentence 2>

may (w_*) paraphrases be likely to ($\mathcal{W}$)

$$\boldsymbol{p(c_j|w_*)} \approx \boldsymbol{p(c_j|\mathcal{W})}$$

<Defining a Paraphrase>

$C_{\mathcal{W}} = \{c_{j_1}, \dots, c_{j_t}\} : sequence\ of\ words\ (with\ repetition)\ observed\ in\ the\ context\ of\ \mathcal{W}$

$w_* \in \mathcal{E}: which\ best\ explains\ the\ observation\ of\ C_{\mathcal{W}}$

<Defining a Paraphrase>

$C_{\mathcal{W}} = \{c_{j_1}, \dots, c_{j_t}\}$: $\textit{sequence of words (with repetition) observed in the context of } \mathcal{W}$

$w_* \in \mathcal{E}$: $\textit{which best explains the observation of } C_{\mathcal{W}}$

"The need for profit is likely to push up prices"

<Defining a Paraphrase>

$C_{\mathcal{W}} = \{c_{j_1}, \ldots, c_{j_t}\}$: *sequence of words (with repetition) observed in the context of* $\mathcal{W}$

$w_* \in \mathcal{E}$: *which best explains the observation of* $C_{\mathcal{W}}$

"The need for profit is likely to push up prices"

"The need for profit may push up prices"

Proof
-Paraphrase

<Defining a Paraphrase>

$C_{\mathcal{W}} = \{c_{j_1}, \dots, c_{j_t}\}$: *sequence of words (with repetition) observed in the context of* $\mathcal{W}$

$w_* \in \mathcal{E}$: *which best explains the observation of* $C_{\mathcal{W}}$

$$w_* = \underset{w_i \in \mathcal{E}}{argmax}\, p(C_{\mathcal{W}}|w_i)$$

"The need for profit may push up prices"

<Defining a Paraphrase>

$C_{\mathcal{W}} = \{c_{j_1}, \ldots, c_{j_t}\}$: *sequence of words (with repetition) observed in the context of* $\mathcal{W}$

$w_* \in \mathcal{E}$: *which best explains the observation of* $C_{\mathcal{W}}$

$$w_* = \underset{w_i \in \mathcal{E}}{argmax}\, p(C_{\mathcal{W}} | w_i)$$

assuming $c_j \in C_{\mathcal{W}}$ *to be independent draws from* $p(c_j | \mathcal{W})$

$$w_* = \underset{w_i}{\operatorname{argmax}} \prod_{c_j \in \mathcal{E}} p(c_j | w_i)^{\#j}$$, $\#j$: *count of* c_j *in* $C_{\mathcal{W}}$

"**The need for profit may push up prices**"

<Defining a Paraphrase>

$C_{\mathcal{W}} = \{c_{j_1}, \ldots, c_{j_t}\}$: *sequence of words (with repetition) observed in the context of* $\mathcal{W}$

$w_* \in \mathcal{E}$: *which best explains the observation of* $C_{\mathcal{W}}$

assuming $c_j \in C_{\mathcal{W}}$ *to be independent draws from* $p(c_j|\mathcal{W})$

$$w_* = \underset{w_i}{\operatorname{argmax}} \prod_{c_j \in \mathcal{E}} p(c_j|w_i)^{\#j}$$, $\#j$: *count of* c_j *in* $C_{\mathcal{W}}$

$$\rightarrow w_* = \underset{w_i}{argmax} \sum_{c_j \in \mathcal{E}} p(c_j|\mathcal{W}) \log p(c_j|w_i)$$

"The need for profit may push up prices"

"Analogies Explained" … Explained, Carl Allen: Machine Learning Blog

<Defining a Paraphrase>

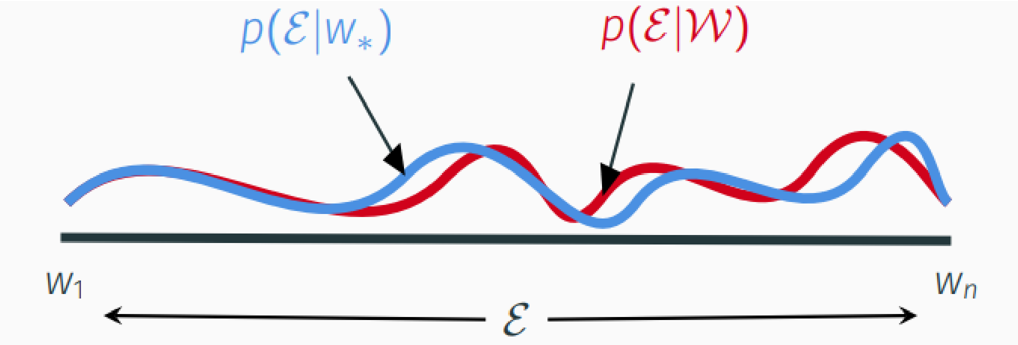
$$w_* = \underset{w_i}{\operatorname{argmax}} \prod_{c_j \in \mathcal{E}} p(c_j | w_i)^{\# j}$$

$$\rightarrow w_* = \underset{w_i}{argmax} \sum_{c_j \in \mathcal{E}} p(c_j | \mathcal{W}) \log p(c_j | w_i)$$

Minimize
KL-Divergence

$$\Delta_{KL}^{\mathcal{W}, w_*} = D_{KL}[P(c_j | \mathcal{W}) \,||\, P(c_j | w_*)]$$

$$= \sum_j p(c_j | \mathcal{W}) \log \frac{p(c_j | \mathcal{W})}{p(c_j | w_*)}$$



"the KL divergence lower bound (zero) is achieved iff the induced distributions are equal"

$$\Delta_{KL}^{\mathcal{W}, w_*} = 0 \leftrightarrow p(c_j | w_*) = p(c_j | \mathcal{W})$$

<Defining a Paraphrase>

$\boldsymbol{Definition\ D1}$. We say word $w_* \in \mathcal{E}$ $\boldsymbol{paraphrases}$ word set $\mathcal{W} \subseteq \mathcal{E}, |\mathcal{W}| < l$, if the $\boldsymbol{paraphrase\ error}$ $\rho^{\mathcal{W},w_*} \in \mathbb{R}^n$ is (element-wise) small, where:

$$\rho_j^{\mathcal{W},w_*} = \log\frac{p(c_j|w_*)}{p(c_j|\mathcal{W})}, c_j \in \mathcal{E}$$

"The need for profit **may** push up prices"

<Sentence 1>

$\approx$

"The need for profit **is likely to** push up prices"

<Sentence 2>

may ($\boldsymbol{w_*}$) paraphrases **be likely to** ($\boldsymbol{\mathcal{W}}$)

$$\boldsymbol{w_*} \approx_{\boldsymbol{P}} \boldsymbol{\mathcal{W}}$$

<Paraphrase = Embedding Sum + Error>

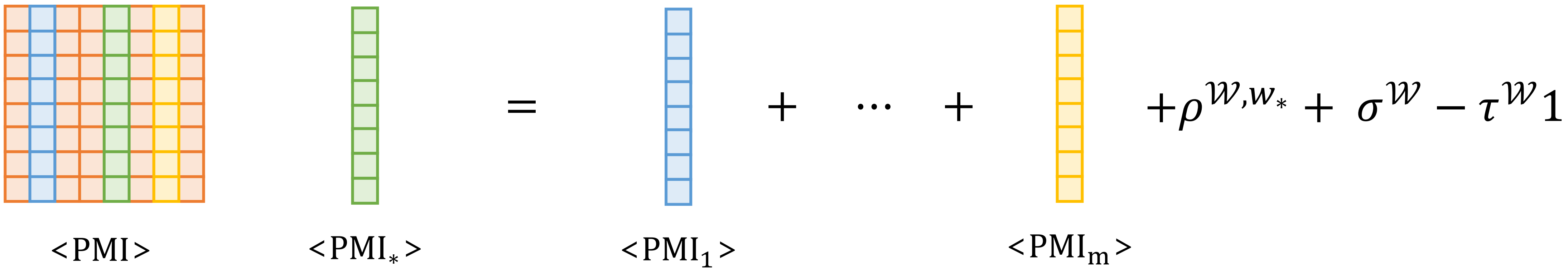
***Lemma* 1**. for any word $w_* \in \mathcal{E}$ and word set $\mathcal{W} \subseteq \mathcal{E}, |\mathcal{W}| < l$:

$$\mathrm{PMI}_* = \sum_{w_i \in \mathcal{W}} \mathrm{PMI}_i + \rho^{\mathcal{W},w_*} + \sigma^{\mathcal{W}} - \tau^{\mathcal{W}} 1$$

$\mathrm{PMI}_\bullet$: *column of* PMI *corresponding to* $w_\bullet \in \mathcal{E}$

$1 \in \mathbb{R}^n$: *vector of* 1s

$\sigma_j^{\mathcal{W}} = \log \frac{p(\mathcal{W}|c_j)}{\prod_i p(w_i|c_j)}, \tau^{\mathcal{W}} = \log \frac{p(\mathcal{W})}{\prod_i p(w_i)}$: *error terms*



<Paraphrase = Embedding Sum + Error>

***Proof of Lemma* 1.**

$$PMI(w_*, c_j) - \sum_{w_i \in \mathcal{W}} PMI(w_i, c_j)$$

$$= \log \frac{p(w_* | c_j)}{p(w_*)} - \log \prod_{w_i \in \mathcal{W}} \frac{p(w_i | c_j)}{p(w_i)}$$

$$= \log \frac{p(w_* | c_j)}{\prod_{\mathcal{W}} p(w_i | c_j)} - \log \frac{p(w_*)}{\prod_{\mathcal{W}} p(w_i)} + \log \frac{p(\mathcal{W} | c_j)}{p(\mathcal{W} | c_j)} + \log \frac{p(\mathcal{W})}{p(\mathcal{W})}$$

$$= \log \frac{p(w_* | c_j)}{p(\mathcal{W} | c_j)} - \log \frac{p(w_*)}{p(\mathcal{W})} + \log \frac{p(\mathcal{W} | c_j)}{\prod_{\mathcal{W}} p(w_i | c_j)} - \log \frac{p(\mathcal{W})}{\prod_{\mathcal{W}} p(w_i)}$$

$$= \log \frac{p(c_j | w_*)}{p(c_j | \mathcal{W})} + \log \frac{p(\mathcal{W} | c_j)}{\prod_{\mathcal{W}} p(w_i | c_j)} - \log \frac{p(\mathcal{W})}{\prod_{\mathcal{W}} p(w_i)}$$

$$= \rho_j^{\mathcal{W}, w_*} + \sigma_j^{\mathcal{W}} - \tau^{\mathcal{W}}$$

<Paraphrase = Embedding Sum + Error>

***Lemma* 1.** for any word $w_* \in \mathcal{E}$ and word set $\mathcal{W} \subseteq \mathcal{E}, |\mathcal{W}| < l$:

$$\mathrm{PMI}_* = \sum_{w_i \in \mathcal{W}} \mathrm{PMI}_i + \rho^{\mathcal{W},w_*} + \sigma^{\mathcal{W}} - \tau^{\mathcal{W}} 1$$

$$C^{\dagger}(\mathrm{PMI}_*) = C^{\dagger}\left(\sum_{w_i \in \mathcal{W}} \mathrm{PMI}_i + \rho^{\mathcal{W},w_*} + \sigma^{\mathcal{W}} - \tau^{\mathcal{W}} 1\right)$$

$$\mathrm{w}_* = \mathrm{w}_{\mathcal{W}} + C^{\dagger}\left(\rho^{\mathcal{W},w_*} + \sigma^{\mathcal{W}} - \tau^{\mathcal{W}} 1\right)$$

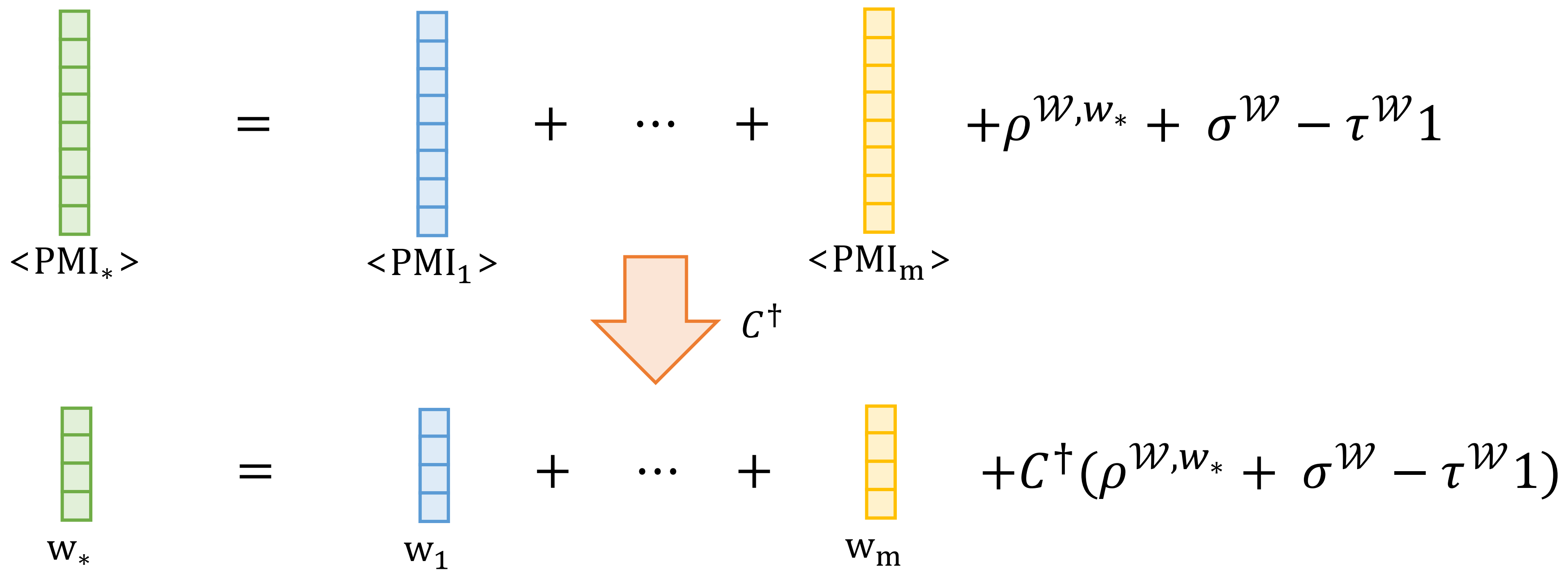
$$where, \mathrm{w}_* = \sum\nolimits_{w_i \in \mathcal{W}} \mathrm{w}_i$$

<Paraphrase = Embedding Sum + Error>

***Theorem* 1.** (Paraphrase). for any word $w_* \in \mathcal{E}$ and word set $\mathcal{W} \subseteq \mathcal{E}, |\mathcal{W}| < l$:

$$\mathrm{w}_* = \mathrm{w}_{\mathcal{W}} + C^{\dagger}\left(\rho^{\mathcal{W},w_*} + \sigma^{\mathcal{W}} - \tau^{\mathcal{W}}\mathbb{1}\right)$$

$$where, \mathrm{w}_{\mathcal{W}} = \sum_{w_i \in \mathcal{W}} \mathrm{w}_i$$



<Paraphrase = Embedding Sum + Error>

***Theorem* 1.** (Paraphrase). for any word $w_* \in \mathcal{E}$ and word set $\mathcal{W} \subseteq \mathcal{E}, |\mathcal{W}| < l$:

$$\mathrm{w}_* = \mathrm{w}_{\mathcal{W}} + C^{\dagger}\left(\rho^{\mathcal{W},w_*} + \sigma^{\mathcal{W}} - \tau^{\mathcal{W}}\mathbf{1}\right)$$

$$where, \mathrm{w}_{\mathcal{W}} = \sum\nolimits_{w_i \in \mathcal{W}} \mathrm{w}_i$$

$$\underset{\mathrm{w}_*}{\square} = \underset{\mathrm{w}_1}{\square} + \cdots + \underset{\mathrm{w}_m}{\square} + \underset{\substack{paraphrase\\ error}}{C^{\dagger}(\rho_j^{\mathcal{W},w_*}} + \underbrace{\underset{\substack{conditional\\ independence\\ error}}{\sigma^{\mathcal{W}}} - \underset{\substack{independence\\ error}}{\tau^{\mathcal{W}}\mathbf{1})}}_{no\ w_*}$$

Proof
-Paraphrase

<Paraphrase = Embedding Sum + Error>

***Theorem* 1.** (Paraphrase). for any word $w_* \in \mathcal{E}$ and word set $\mathcal{W} \subseteq \mathcal{E}, |\mathcal{W}| < l$:

$$\mathrm{w}_* = \mathrm{w}_{\mathcal{W}} + C^{\dagger}\left(\rho^{\mathcal{W}, w_*} + \sigma^{\mathcal{W}} - \tau^{\mathcal{W}} 1\right)$$

$$where, \mathrm{w}_{\mathcal{W}} = \sum_{w_i \in \mathcal{W}} \mathrm{w}_i$$

$$\sigma_j^{\mathcal{W}} = \log \frac{p(\mathcal{W}|c_j)}{\prod_i p(w_i|c_j)} : conditional\ independence\ error$$

$p(\mathcal{W}|c_j)$

"**The need for profit is likely to push up prices**"

$\prod_i p(w_i|c_j)$

$\sigma_j^{\mathcal{W}} = 0\ iff$ all $w_i \in \mathcal{W}$ are conditionally independent given each $c_j \in \mathcal{E}$

<Paraphrase = Embedding Sum + Error>

***Theorem* 1.** (Paraphrase). for any word $w_* \in \mathcal{E}$ and word set $\mathcal{W} \subseteq \mathcal{E}, |\mathcal{W}| < l$:

$$\mathrm{w}_* = \mathrm{w}_{\mathcal{W}} + C^{\dagger}\left(\rho^{\mathcal{W},w_*} + \sigma^{\mathcal{W}} - \tau^{\mathcal{W}}1\right)$$

$$where, \mathrm{w}_{\mathcal{W}} = \sum_{w_i \in \mathcal{W}} \mathrm{w}_i$$

$$\tau^{\mathcal{W}} = \log \frac{p(\mathcal{W})}{\prod_i p(w_i)} : independence\ error$$

$p(\mathcal{W})$

"The need for profit **is likely to** push up prices"

$\prod_i p(w_i)$

$\tau^{\mathcal{W}} = 0\ iff$ all $w_i \in \mathcal{W}$ are mutually independent

<Paraphrase = Embedding Sum + Error>

$\boldsymbol{Theorem\ 1.}$ (Paraphrase). for any word $w_* \in \mathcal{E}$ and word set $\mathcal{W} \subseteq \mathcal{E}, |\mathcal{W}| < l$:

$$\boldsymbol{w_* = w_{\mathcal{W}} + C^{\dagger}\left(\rho^{\mathcal{W},w_*} + \sigma^{\mathcal{W}} - \tau^{\mathcal{W}}\mathbf{1}\right)}$$

$$where, \mathrm{w}_{\mathcal{W}} = \sum_{w_i \in \mathcal{W}} \mathrm{w}_i$$

$\boldsymbol{Definition\ D1.}$ We say word $w_* \in \mathcal{E}$ $\boldsymbol{paraphrases}$ word set $\mathcal{W} \subseteq \mathcal{E}, |\mathcal{W}| < l$, if the $\boldsymbol{paraphrase\ error\ \rho^{\mathcal{W},w_*}}$ $\in \mathbb{R}^n$ is (element-wise) **small**, where:

$$\rho_j^{\mathcal{W},w_*} = \log\frac{p(c_j|w_*)}{p(c_j|\mathcal{W})}, c_j \in \mathcal{E}$$

king $\boldsymbol{paraphrase}$ **{man, royal}** ⇒ $\boldsymbol{w_{man} + w_{royal} \approx w_{king}}$

$\boldsymbol{\rho, \sigma, \tau}$

Proof

-Paraphrase

<What We Want to Prove>

{woman, king} *paraphrases* {man, queen}

$$\Downarrow$$

$$\mathrm{w}_{king} - \mathrm{w}_{man} + \mathrm{w}_{woman} \approx \mathrm{w}_{queen}$$

<What We Have Proven>

king *paraphrases* {man, royal}

$$\Downarrow$$

Dependency Error

$$\mathrm{w}_{man} + \mathrm{w}_{royal} \approx \mathrm{w}_{king}$$

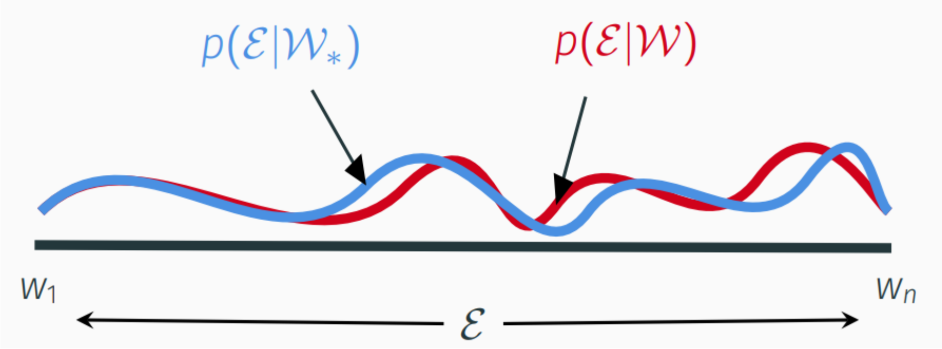
<Paraphrasing Word Sets>

Definition D1. We say word $w_* \in \mathcal{E}$ ***paraphrases*** word set $\mathcal{W} \subseteq \mathcal{E}, |\mathcal{W}| < l$, if the ***paraphrase error*** $\rho^{\mathcal{W},w_*} \in \mathbb{R}^n$ is (element-wise) small, where:

$$\rho_j^{\mathcal{W},w_*} = \log \frac{p(c_j|w_*)}{p(c_j|\mathcal{W})}, c_j \in \mathcal{E}$$

one word w_*
↓
word set $\mathcal{W}_*$

$$\rho_j^{\mathcal{W},\mathcal{W}_*} = \log \frac{p(c_j|\mathcal{W}_*)}{p(c_j|\mathcal{W})}, c_j \in \mathcal{E}$$



<Paraphrasing Word Sets>

$\boldsymbol{Definition\ D2}$. We say word set $\mathcal{W}_* \subseteq \mathcal{E}$ $\boldsymbol{paraphrases}$ word set $\mathcal{W} \subseteq \mathcal{E}, |\mathcal{W}|, |\mathcal{W}_*| < l$, if the $\boldsymbol{paraphrase\ error}$ $\rho^{\mathcal{W},\mathcal{W}_*} \in \mathbb{R}^n$ is (element-wise) small, where:

$$\rho_j^{\mathcal{W},\mathcal{W}_*} = \log \frac{p(c_j|\mathcal{W}_*)}{p(c_j|\mathcal{W})}, c_j \in \mathcal{E}$$

"The need for profit **will probably** push up prices" $\approx$ "The need for profit **is likely to** push up prices"

<Sentence 1> *<Sentence 2>*

will probably ($\boldsymbol{\mathcal{W}_*}$) paraphrases **be likely to** ($\boldsymbol{\mathcal{W}}$)

$$\boldsymbol{\mathcal{W}_*} \approx_{\boldsymbol{P}} \boldsymbol{\mathcal{W}}$$

Proof

-Analogy

<Paraphrasing Word Sets>

Lemma 2. for any word sets $\mathcal{W},\ \mathcal{W}_* \subseteq \mathcal{E}, |\mathcal{W}|,\ |\mathcal{W}_*| < l$:

$$\sum_{w_i \in \mathcal{W}_*} \mathrm{PMI}_i = \sum_{w_i \in \mathcal{W}} \mathrm{PMI}_i + \rho^{\mathcal{W},\mathcal{W}_*} + \sigma^{\mathcal{W}} - \sigma^{\mathcal{W}_*} - \left(\tau^{\mathcal{W}} - \tau^{\mathcal{W}_*}\right)\mathbf{1}$$

$\mathrm{PMI}_\bullet$: *column of* PMI *corresponding to* $w_\bullet \in \mathcal{E}$

$\mathbf{1} \in \mathbb{R}^n$: *vector of* 1*s*

$\sigma_j^{\mathcal{W}} = \log \frac{p(\mathcal{W}|c_j)}{\prod_i p(w_i|c_j)}, \tau^{\mathcal{W}} = \log \frac{p(\mathcal{W})}{\prod_i p(w_i)}$: *error terms*

<Paraphrasing Word Sets>

$\boldsymbol{Theorem\ 2.}$ (Generalised Paraphrase). for any word sets $\mathcal{W}, \mathcal{W}_* \subseteq \mathcal{E}, |\mathcal{W}|, |\mathcal{W}_*| < l$:

$$\mathrm{w}_{\mathcal{W}_*} = \mathrm{w}_{\mathcal{W}} + C^\dagger\left(\rho^{\mathcal{W},\mathcal{W}_*} + \sigma^{\mathcal{W}} - \sigma^{\mathcal{W}_*} - \left(\tau^{\mathcal{W}} - \tau^{\mathcal{W}_*}\right)1\right)$$

$$where, \mathrm{w}_{\mathcal{W}} = \sum_{w_i \in \mathcal{W}} \mathrm{w}_i$$

$\boldsymbol{Proof\ of\ Theorem\ 2.}$ Multiply $\boldsymbol{Lemma\ 2.}$ by $C^\dagger$

<Paraphrasing Word Sets>

Definition D2. We say word set $\mathcal{W}_* \subseteq \mathcal{E}$ ***paraphrases*** word set $\mathcal{W} \subseteq \mathcal{E}, |\mathcal{W}|, |\mathcal{W}_*| < l$, if the ***paraphrase error*** $\rho^{\mathcal{W},\mathcal{W}_*} \in \mathbb{R}^n$ **is (element-wise) small**, where:

$$\boldsymbol{\rho_j^{\mathcal{W},\mathcal{W}_*} = \log \frac{p(c_j|\mathcal{W}_*)}{p(c_j|\mathcal{W})}, c_j \in \mathcal{E}}$$

Lemma 2. for any word sets $\mathcal{W}$, $\mathcal{W}_* \subseteq \mathcal{E}, |\mathcal{W}|$, $|\mathcal{W}_*| < l$:

$$\sum_{w_i \in \mathcal{W}_*} \boldsymbol{PMI_i} = \sum_{w_i \in \mathcal{W}} \mathbf{PMI}_i + \boldsymbol{\rho}^{\mathcal{W},\mathcal{W}_*} + \boldsymbol{\sigma}^{\mathcal{W}} - \boldsymbol{\sigma}^{\mathcal{W}_*} - (\boldsymbol{\tau}^{\mathcal{W}} - \boldsymbol{\tau}^{\mathcal{W}_*})\mathbf{1}$$

Theorem 2. (Generalised Paraphrase). for any word sets $\mathcal{W}, \mathcal{W}_* \subseteq \mathcal{E}, |\mathcal{W}|, |\mathcal{W}_*| < l$:

$$\mathbf{w}_{\mathcal{W}_*} = \mathbf{w}_{\mathcal{W}} + \boldsymbol{C}^{\dagger}(\boldsymbol{\rho}^{\mathcal{W},\mathcal{W}_*} + \boldsymbol{\sigma}^{\mathcal{W}} - \boldsymbol{\sigma}^{\mathcal{W}_*} - (\boldsymbol{\tau}^{\mathcal{W}} - \boldsymbol{\tau}^{\mathcal{W}_*})\mathbf{1})$$

{woman, king} *paraphrases* {man, queen}

$$\Downarrow \quad (\boldsymbol{\rho}, \boldsymbol{\sigma}, \boldsymbol{\tau})$$

$$\mathrm{PMI}_{king} - \mathrm{PMI}_{man} + \mathrm{PMI}_{woman} \approx \mathrm{PMI}_{queen}$$

$$\Downarrow$$

$$\mathrm{w}_{king} - \mathrm{w}_{man} + \mathrm{w}_{woman} \approx \mathrm{w}_{queen}$$

Proof

-Routemap

<Routemap>

"man is to king as woman is to queen"

$\Updownarrow$

man *transforms to* king as woman *transforms to* queen

$\Updownarrow$

{woman, king} *paraphrases* {man, queen}

$\Downarrow$ ***Dependency Error***

$$\mathrm{PMI}_{king} - \mathrm{PMI}_{man} + \mathrm{PMI}_{woman} \approx \mathrm{PMI}_{queen}$$

$\Downarrow$ $\mathbf{PMI}_i \approx \mathbf{w}_i^T C$

$$\mathrm{w}_{king} - \mathrm{w}_{man} + \mathrm{w}_{woman} \approx \mathrm{w}_{queen}$$

semantic

geometric

Proof

-Routemap

<Routemap>

"man is to king as woman is to queen"

$\Updownarrow$

man *transforms to* king as woman *transforms to* queen

$\Updownarrow$

{woman, king} *paraphrases* {man, queen}

$\Downarrow$ ***Dependency Error***

$$\mathrm{PMI}_{king} - \mathrm{PMI}_{man} + \mathrm{PMI}_{woman} \approx \mathrm{PMI}_{queen}$$

$\Downarrow$ $\mathbf{PMI}_i \approx \mathbf{w}_i^T \boldsymbol{C}$

$$\mathrm{w}_{king} - \mathrm{w}_{man} + \mathrm{w}_{woman} \approx \mathrm{w}_{queen}$$

semantic

geometric

Proof

- Paraphrase
- Analogy

<Analogy>

$\boldsymbol{analogy}$: w_a $is\ to$ w_{a^*} as w_b $is\ to$ w_{b^*}

$where, w_a, w_{a^*}, w_b, w_{b^*} \in \mathcal{E}$

Proof
-Analogy

<Analogy>

$$\boldsymbol{analogy}: w_a \ is\ to\ w_{a^*}\ as\ w_b\ is\ to\ w_{b^*}$$

$$where, w_a, w_{a^*}, w_b, w_{b^*} \in \mathcal{E}$$

More Rigorously

$$\boldsymbol{analogy}: \mathfrak{A}$$

$$S_{\mathfrak{A}} \subseteq \mathcal{E} \times \mathcal{E} : set\ of\ ordered\ word\ pairs$$

$$(w_x, w_{x^*}) \in S_{\mathfrak{A}} \quad iff \quad "w_x\ is\ to\ w_{x^*}\ as\ [all\ other\ analogical\ pairs]\ under\ \mathfrak{A}$$

Proof

-Analogy

<Analogy>

$\boldsymbol{analogy}: w_a \text{ is to } w_{a^*} \text{ as } w_b \text{ is to } w_{b^*}$

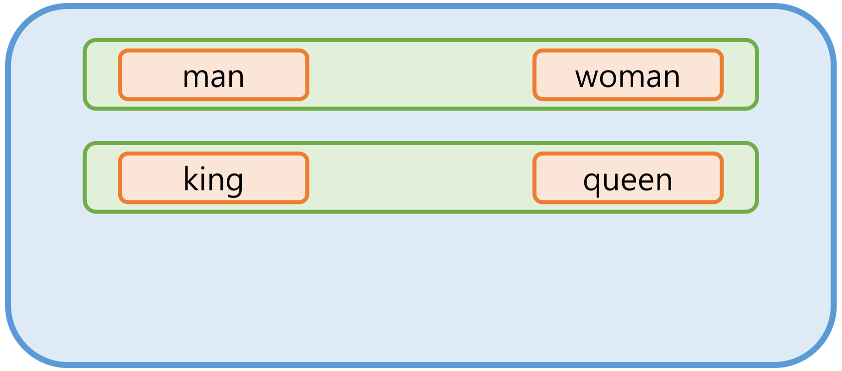
$where, w_a, w_{a^*}, w_b, w_{b^*} \in \mathcal{E}$

More Rigorously

$\boldsymbol{analogy}: \mathfrak{A}$

$S_{\mathfrak{A}} \subseteq \mathcal{E} \times \mathcal{E}$: *set of ordered word pairs*

$(w_x, w_{x^*}) \in S_{\mathfrak{A}} \quad iff \quad "w_x \text{ is to } w_{x^*} \text{ as [all other analogical pairs] under } \mathfrak{A}$



$\mathfrak{A}$

<Analogy>

$$\boldsymbol{analogy}: w_a \text{ is to } w_{a^*} \text{ as } w_b \text{ is to } w_{b^*}$$

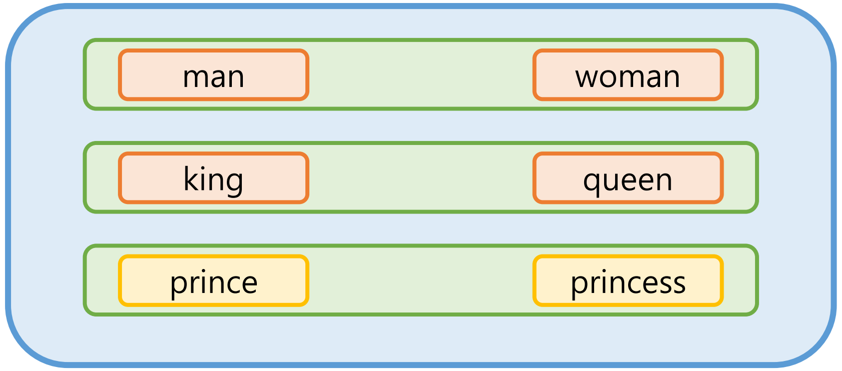
$$where, w_a, w_{a^*}, w_b, w_{b^*} \in \mathcal{E}$$

More Rigorously

$$\boldsymbol{analogy}: \mathfrak{A}$$

$$S_{\mathfrak{A}} \subseteq \mathcal{E} \times \mathcal{E} : set\ of\ ordered\ word\ pairs$$

$$(w_x, w_{x^*}) \in S_{\mathfrak{A}} \quad iff \quad "w_x \text{ is to } w_{x^*} \text{ as } [all\ other\ analogical\ pairs]\ under\ \mathfrak{A}$$



$\mathfrak{A}$

Proof
-Analogy

<Analogy>

$\boldsymbol{analogy}: w_a \text{ is to } w_{a^*} \text{ as } w_b \text{ is to } w_{b^*}$

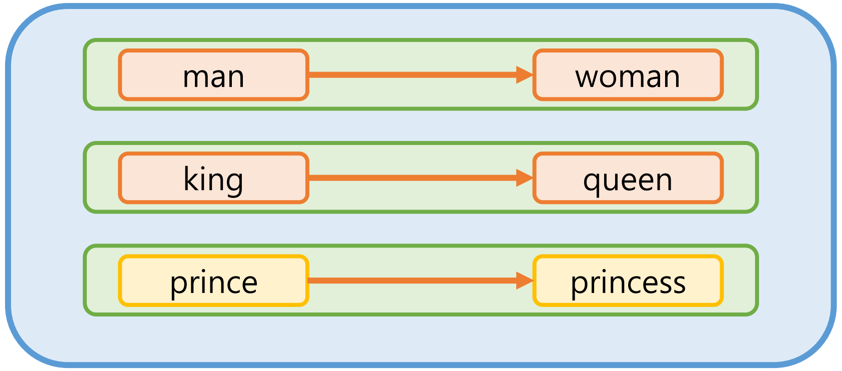
$where, w_a, w_{a^*}, w_b, w_{b^*} \in \mathcal{E}$

More Rigorously

$\boldsymbol{analogy}: \mathfrak{A}$

$S_{\mathfrak{A}} \subseteq \mathcal{E} \times \mathcal{E}$: $set\ of\ ordered\ word\ pairs$

$(w_x, w_{x^*}) \in S_{\mathfrak{A}}\ \ iff\ \ $"$w_x$ is to w_{x^*} as [all other analogical pairs] under $\mathfrak{A}$



$\mathfrak{A}$

<Analogy>

$\boldsymbol{analogy}: w_a \textit{ is to } w_{a^*} \textit{ as } w_b \textit{ is to } w_{b^*}$

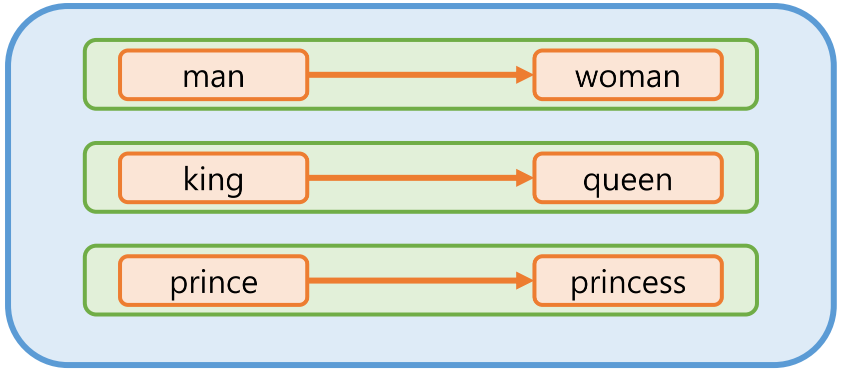
$where, w_a, w_{a^*}, w_b, w_{b^*} \in \mathcal{E}$

More Rigorously

$\boldsymbol{analogy}: \mathfrak{A}$

$S_{\mathfrak{A}} \subseteq \mathcal{E} \times \mathcal{E}$: $set\ of\ ordered\ word\ pairs$

$(w_x, w_{x^*}) \in S_{\mathfrak{A}} \quad iff \quad "w_x \textit{ is to } w_{x^*} \textit{ as } [all\ other\ analogical\ pairs]\ under\ \mathfrak{A}$



$\mathfrak{A}$

<Analogy>

$\boldsymbol{analogy}: \mathfrak{A}$

$S_{\mathfrak{A}} \subseteq \mathcal{E} \times \mathcal{E}$: *set of ordered word pairs*

$(w_x, w_{x^*}) \in S_{\mathfrak{A}}$ *iff* "w_x *is to* w_{x^*} *as [all other analogical pairs] under* $\mathfrak{A}$

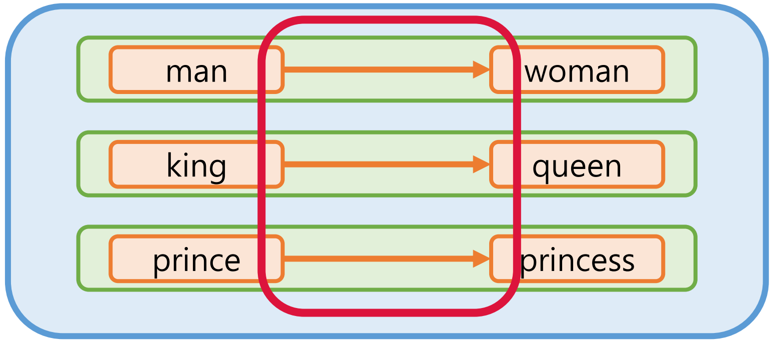
To Explain

$$\mathrm{w}_{b^*} \approx \mathrm{w}_{a^*} - \mathrm{w}_a + \mathrm{w}_b$$

In more general case

$$\mathrm{w}_{x^*} - \mathrm{w}_x \approx \mathrm{u}_{\mathfrak{A}}$$

$\forall (w_x, w_{x^*}) \in S_{\mathfrak{A}}$, $\mathrm{u}_{\mathfrak{A}} \in \mathbb{R}^n$: *specific vector to* $\mathfrak{A}$



<Word Transformation>

Theorem 2. (Generalised Paraphrase). for any word sets $\mathcal{W}, \mathcal{W}_* \subseteq \mathcal{E}, |\mathcal{W}|, |\mathcal{W}_*| < l$:

$$\mathrm{w}_{\mathcal{W}_*} = \mathrm{w}_{\mathcal{W}} + C^{\dagger}\left(\rho^{\mathcal{W},\mathcal{W}_*} + \sigma^{\mathcal{W}} - \sigma^{\mathcal{W}_*} - \left(\tau^{\mathcal{W}} - \tau^{\mathcal{W}_*}\right)1\right)$$

$$where, \mathrm{w}_{\mathcal{W}} = \sum\nolimits_{w_i \in \mathcal{W}} \mathrm{w}_i$$

$$\mathcal{W} = \{w_x\} \cup \mathcal{W}^+$$
$$\mathcal{W}_* = \left\{w_{x_*}\right\} \cup \mathcal{W}^-$$

Corollary 2.1. for any words $w_x, w_{x^*} \in \mathcal{E}$ and word sets $\mathcal{W}^+, \mathcal{W}^- \subseteq \mathcal{E}$, $|\mathcal{W}^+|, |\mathcal{W}^-| < l - 1$:

$$\mathrm{w}_{x^*} = w_x + \mathrm{w}_{\mathcal{W}^+} - \mathrm{w}_{\mathcal{W}^-} + C^{\dagger}\left(\rho^{\mathcal{W},\mathcal{W}_*} + \sigma^{\mathcal{W}} - \sigma^{\mathcal{W}_*} - \left(\tau^{\mathcal{W}} - \tau^{\mathcal{W}_*}\right)1\right)$$

$$where, \mathcal{W} = \{w_x\} \cup \mathcal{W}^+, \mathcal{W}_* = \left\{w_{x_*}\right\} \cup \mathcal{W}^-$$

Proof

-Analogy

<Word Transformation>

$$\mathcal{W} = \{w_x\} \cup \mathcal{W}^+, \mathcal{W}_* = \{w_{x_*}\} \cup \mathcal{W}^-$$

{be, likely, to} ($\mathcal{W}$) $\approx_P$ {will, probably} ($\mathcal{W}_*$)

(Slides) Analogies Explained: Towards Understanding Word Embeddings, ICML, 2019

<Word Transformation>

$$\mathcal{W} = \{w_x\} \cup \mathcal{W}^+, \mathcal{W}_* = \{w_{x_*}\} \cup \mathcal{W}^-$$

Proof
-Analogy

<Word Transformation>

$$\mathcal{W} = \{w_x\} \cup \mathcal{W}^+, \mathcal{W}_* = \{w_{x_*}\} \cup \mathcal{W}^-$$

{be, likely, to}
($\mathcal{W}$)

$\approx_P$

{will, probably}
($\mathcal{W}_*$)

+ {be, likely}
($\mathcal{W}^+$)

- {probably}
($\mathcal{W}^-$)

To
(w)

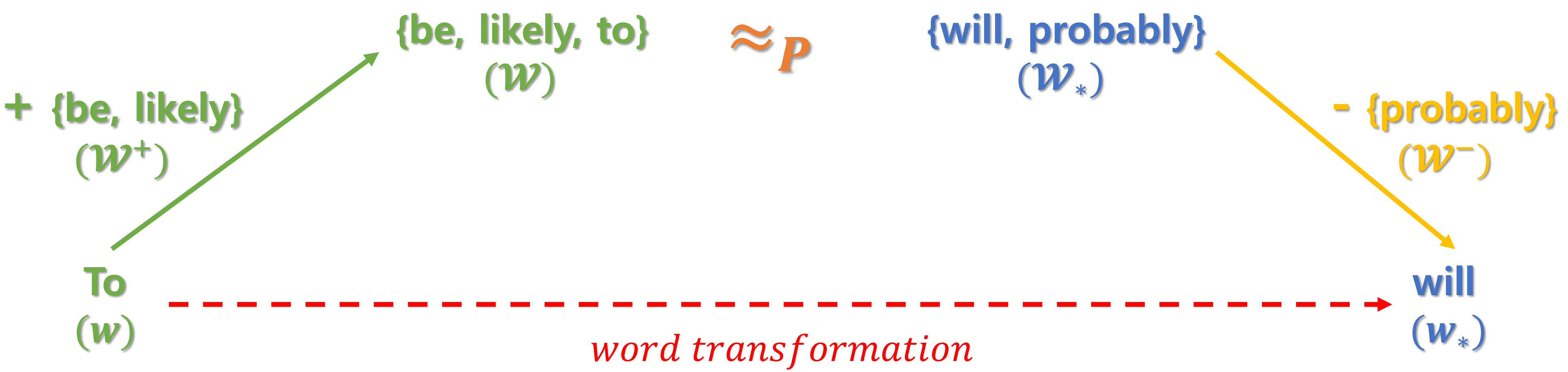
will
(w_*)

Proof

-Analogy

<Word Transformation>

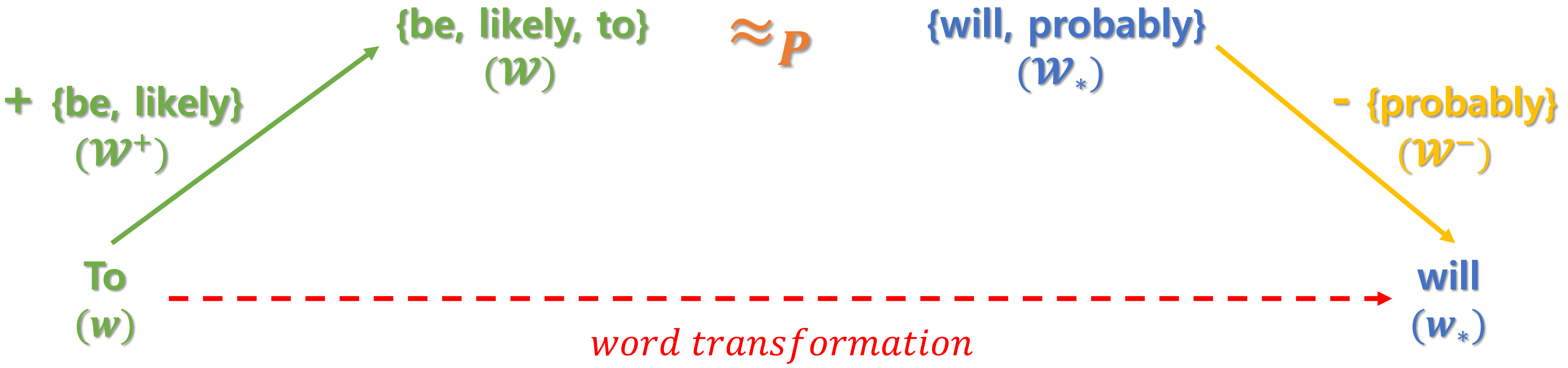
$$\mathcal{W} = \{w_x\} \cup \mathcal{W}^+, \mathcal{W}_* = \{w_{x_*}\} \cup \mathcal{W}^-$$



(Slides) Analogies Explained: Towards Understanding Word Embeddings, ICML, 2019

<Word Transformation>

Definition D3. There exists a ***word transformation*** from $w_x \in \mathcal{E}$ to $w_{x^*} \in \mathcal{E}$ with ***transformation parameters*** $\mathcal{W}^+, \mathcal{W}^- \subseteq \mathcal{E}$ *iff* $\{w_x\} \cup \mathcal{W}^+ \approx_P \{w_{x^*}\} \cup \mathcal{W}^-$



[(Slides) Analogies Explained: Towards Understanding Word Embeddings, ICML, 2019]

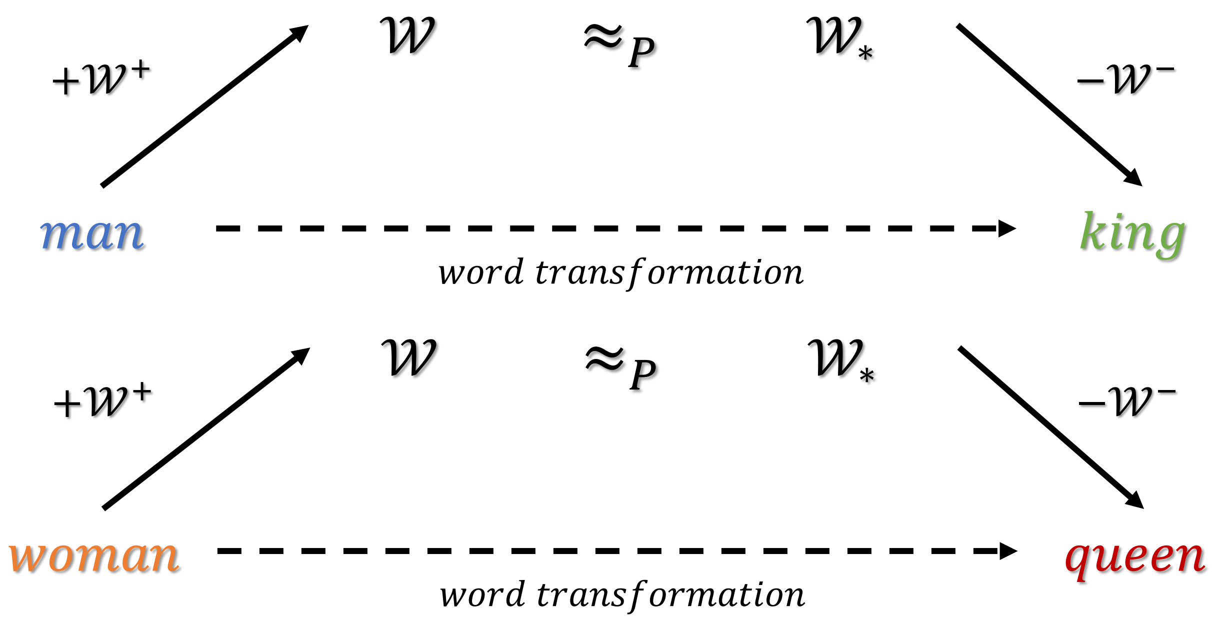
<Word Transformation to Analogy>

$$\begin{tikzcd} & \mathcal{W} \arrow[r, phantom, "\approx_P"] & \mathcal{W}_* \arrow[dr, "-\mathcal{W}^-"] & \\ man \arrow[ur, "+\mathcal{W}^+"] \arrow[rrr, dashed, "word\ transformation"'] & & & king \end{tikzcd}$$

Proof

-Analogy

<Word Transformation to Analogy>



Proof
-Analogy

<Word Transformation to Analogy>

$$\mathcal{W} \approx_P \mathcal{W}_*$$

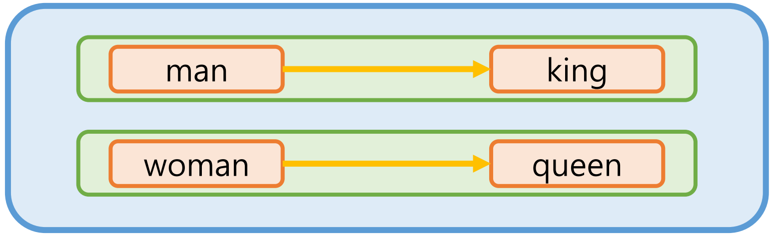
$+\mathcal{w}^+$ $\quad -\mathcal{w}^-$

$man \xrightarrow{\text{word transformation}} king$

$$\mathcal{W} \approx_P \mathcal{W}_*$$

$+\mathcal{w}^+$ $\quad -\mathcal{w}^-$

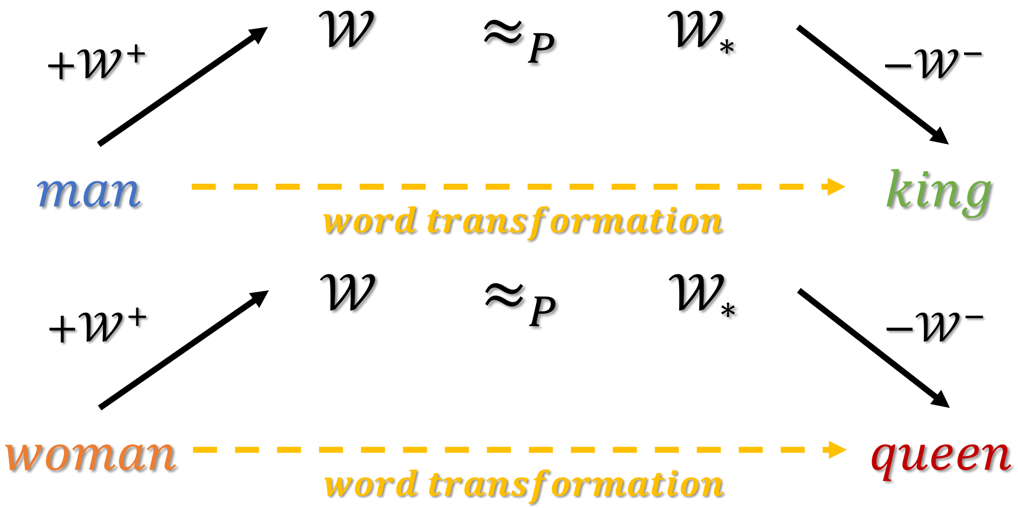
$woman \xrightarrow{\text{word transformation}} queen$

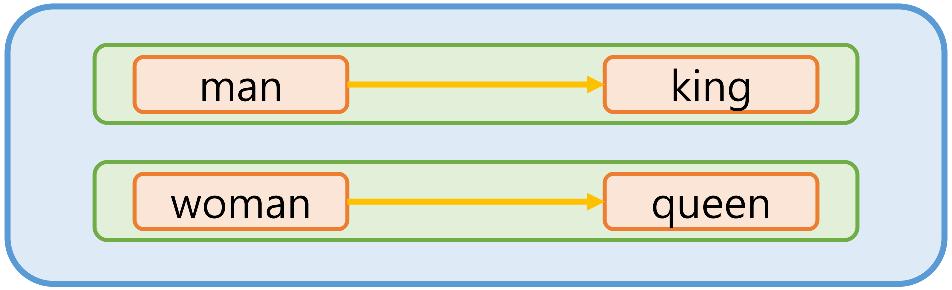


Proof
-Analogy

<Word Transformation to Analogy>

Definition D4. We say "$\boldsymbol{w_a}$ ***is to*** $\boldsymbol{w_{a^*}}$ ***as*** $\boldsymbol{w_b}$ ***is to*** $\boldsymbol{w_{b^*}}$" for $w_a, w_b, w_{a^*}, w_{b^*} \in \mathcal{E}$ iff there exist parameters $\mathcal{W}^+, \mathcal{W}^- \subseteq \mathcal{E}$ that **simultaneously transform** w_a to w_{a^*} and w_b to w_{b^*}





Proof

-Analogy

<Analogy to Paraphrase>

That is, we say:

"man is to king as woman is to queen"

iff there exist parameters $\mathcal{W}^+, \mathcal{W}^- \subseteq \mathcal{E}$ that simultaneously transform man to king and woman to queen

$$\begin{tikzcd} & \mathcal{W} \approx_P \mathcal{W}_* \arrow[dr, "-\mathcal{W}^-"] & \\ man \arrow[ur, "+\mathcal{W}^+"] \arrow[rr, dashed, "word\ transformation"'] & & king \end{tikzcd}$$

$$\begin{tikzcd} & \mathcal{W} \approx_P \mathcal{W}_* \arrow[dr, "-\mathcal{W}^-"] & \\ woman \arrow[ur, "+\mathcal{W}^+"] \arrow[rr, dashed, "word\ transformation"'] & & queen \end{tikzcd}$$

(Slides) Analogies Explained: Towards Understanding Word Embeddings, ICML, 2019

<Analogy to Paraphrase>

That is, we say:

"man is to king as woman is to queen"

iff there exist parameters $\mathcal{W}^+, \mathcal{W}^- \subseteq \mathcal{E}$ that simultaneously transform man to king and woman to queen

$$\begin{tikzcd} & \mathcal{W} \arrow[r, phantom, "\approx_P"] & \mathcal{W}_* \arrow[dr, "-\mathcal{W}^-"] & \\ man \arrow[ur, "+\mathcal{W}^+"] \arrow[rrr, dashed, "word\ transformation"'] & & & king \end{tikzcd}$$

Let $\mathcal{W}^+ = \{$king$\}$,
$\mathcal{W}^- = \{$man$\}$

$$\begin{tikzcd} & \mathcal{W} \arrow[r, phantom, "\approx_P"] & \mathcal{W}_* \arrow[dr, "-\mathcal{W}^-"] & \\ woman \arrow[ur, "+\mathcal{W}^+"] \arrow[rrr, dashed, "word\ transformation"'] & & & queen \end{tikzcd}$$

(Slides) Analogies Explained: Towards Understanding Word Embeddings, ICML, 2019

<Analogy to Paraphrase>

That is, we say:

"man is to king as woman is to queen"

iff there exist parameters $\mathcal{W}^+, \mathcal{W}^- \subseteq \mathcal{E}$ that simultaneously transform man to king and woman to queen

$$\begin{tikzcd}
 & \mathcal{W} \approx_P \mathcal{W}_* \arrow[dr, "-man"] & \\
man \arrow[ur, "+king"] \arrow[rr, dashed, "word\ transformation"'] & & king
\end{tikzcd}$$

Let $\mathcal{W}^+ = \{\text{king}\}$,
$\mathcal{W}^- = \{\text{man}\}$

$$\begin{tikzcd}
 & \mathcal{W} \approx_P \mathcal{W}_* \arrow[dr, "-man"] & \\
woman \arrow[ur, "+king"] \arrow[rr, dashed, "word\ transformation"'] & & queen
\end{tikzcd}$$

Proof

-Routemap

(Slides) Analogies Explained: Towards Understanding Word Embeddings, ICML, 2019

<Routemap>

"man is to king as woman is to queen"

$$\Updownarrow$$

man *transforms to* king as woman *transforms to* queen

$$\Updownarrow$$

{woman, king} *paraphrases* {man, queen}

$$\Downarrow \quad \textit{Dependency Error}$$

$$\mathrm{PMI}_{king} - \mathrm{PMI}_{man} + \mathrm{PMI}_{woman} \approx \mathrm{PMI}_{queen}$$

$$\Downarrow \quad \mathbf{PMI}_{i} \approx \mathbf{w}_{i}^{T} C$$

$$\mathbf{w}_{king} - \mathbf{w}_{man} + \mathbf{w}_{woman} \approx \mathbf{w}_{queen}$$

semantic

geometric

<Analogies Explained>

$\boldsymbol{Lemma\ 3.}$ If "$\boldsymbol{w_a\ is\ to\ w_{a^*}\ as\ w_b\ is\ to\ w_{b^*}}$" by $\boldsymbol{D4}$ with transformation parameters $\mathcal{W}^+, \mathcal{W}^- \subseteq \mathcal{E}$, then:

$$
\begin{aligned}
PMI_{b^*} = PMI_{a^*} - PMI_a + PMI_b \\
+ \rho^{\mathcal{W}^b, \mathcal{W}_*^b} - \rho^{\mathcal{W}^a, \mathcal{W}_*^a} \\
+ \left(\sigma^{\mathcal{W}^b} - \sigma^{\mathcal{W}_*^b}\right) - \left(\sigma^{\mathcal{W}^a} - \sigma^{\mathcal{W}_*^a}\right) \\
-\left(\left(\tau^{\mathcal{W}^b} - \tau^{\mathcal{W}_*^b}\right) - \left(\tau^{\mathcal{W}^a} - \tau^{\mathcal{W}_*^a}\right)\right) 1
\end{aligned}
$$

$where\ \mathcal{W}^x = \{w_x\} \cup \mathcal{W}^+, \mathcal{W}_*^x = \{w_{x^*}\} \cup \mathcal{W}^-\ for\ x \in \{a, b\}\ and\ \rho^{\mathcal{W}^b, \mathcal{W}_*^b}, \rho^{\mathcal{W}^a, \mathcal{W}_*^a}\ are\ small.$

$\boldsymbol{Proof\ of\ Lemma\ 3.}$ Let $\mathcal{W} = \mathcal{W}^x, \mathcal{W}_* = \mathcal{W}_*^x$ for $x \in \{a, b\}$ in instance of $\boldsymbol{Cor\ 2.1}$ and take the difference. $\mathcal{W}^x$ paraphrases $\mathcal{W}^x$ for $x \in \{a, b\}$ by $\boldsymbol{D3}$ and $\boldsymbol{D4}$

(Slides) Analogies Explained: Towards Understanding Word Embeddings, ICML, 2019

<Analogies Explained>

$\boldsymbol{Theorem\ 3}$. (Analogies) If "$\boldsymbol{w_a\ is\ to\ w_{a^*}\ as\ w_b\ is\ to\ w_{b^*}}$" by $\boldsymbol{D4}$ with transformation parameters $\mathcal{W}^+, \mathcal{W}^- \subseteq \mathcal{E}$, then:

$$
\begin{aligned}
\mathrm{w}_{b^*} = \mathrm{w}_{a^*} & - \mathrm{w}_a + \mathrm{w}_b \\
& + C^\dagger \left(\rho^{\mathcal{W}^b, \mathcal{W}^b_*} - \rho^{\mathcal{W}^a, \mathcal{W}^a_*}\right. \\
& + (\sigma^{\mathcal{W}^b} - \sigma^{\mathcal{W}^b_*}) - (\sigma^{\mathcal{W}^a} - \sigma^{\mathcal{W}^a_*}) \\
& \left. - ((\tau^{\mathcal{W}^b} - \tau^{\mathcal{W}^b_*}) - (\tau^{\mathcal{W}^a} - \tau^{\mathcal{W}^a_*}))\mathbf{1}\right)
\end{aligned}
$$

"man is to king as woman is to queen"

$$\Downarrow \quad (\boldsymbol{\rho}, \boldsymbol{\sigma}, \boldsymbol{\tau})$$

$$\mathrm{w}_{king} - \mathrm{w}_{man} + \mathrm{w}_{woman} \approx \mathrm{w}_{queen}$$

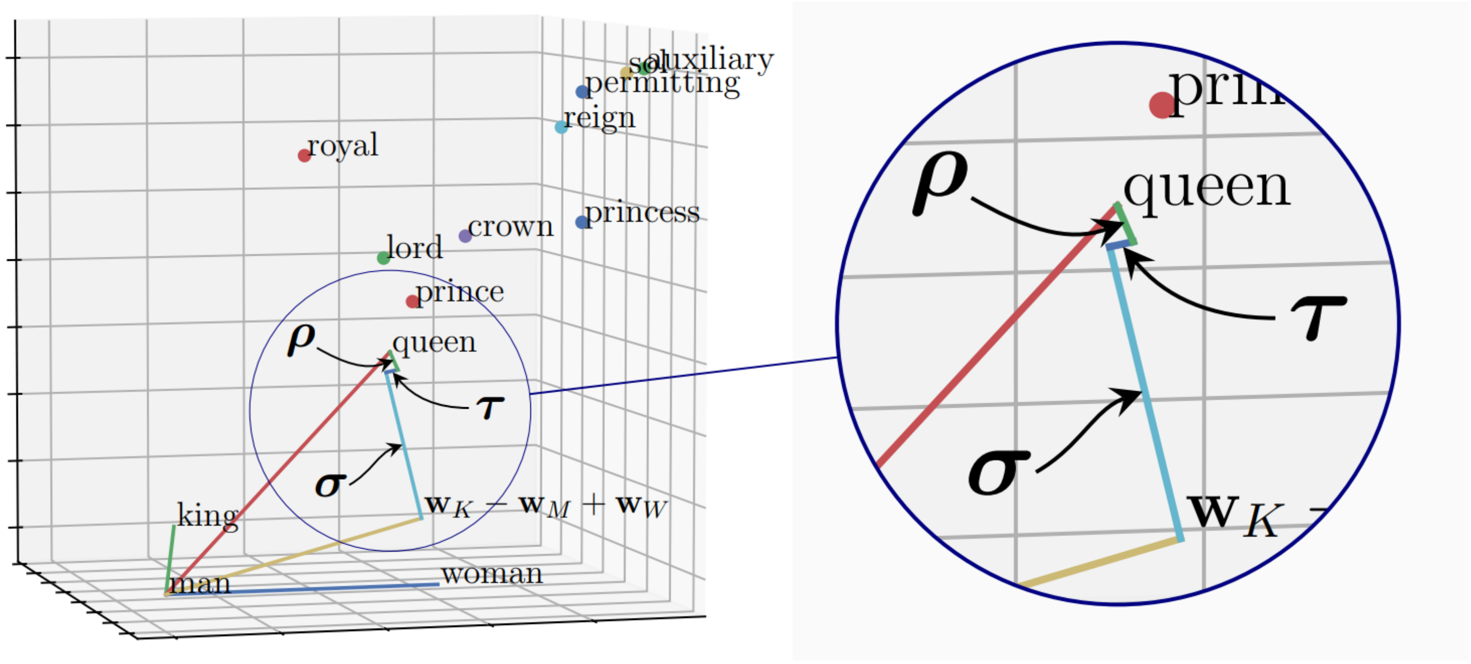
Proof
-Overall

<Analogies Explained>

"man is to king as woman is to queen"

$$\Downarrow \quad (\boldsymbol{\rho}, \boldsymbol{\sigma}, \boldsymbol{\tau})$$

$$\mathbf{w}_{king} - \mathbf{w}_{man} + \mathbf{w}_{woman} \approx \mathbf{w}_{queen}$$



Conclusion

<Conclusion>

- To derive a probabilistic definition of $\boldsymbol{paraphrasing}$ and show that it governs the relationship between one word embedding and any sum of others
- To show how paraphrasing can be generalized and interpreted as the $\boldsymbol{transformation}$ from one word to another, giving a mathematical formulation for "w_x is to w_{x^*}"
- To provide the first rigorous proof of the linear relationship between word embeddings of analogies, including explicit, interpretable error terms

Any Questions?

Thank You