# Lexical Semantics

Lexical Semantics คือ การศึกษาความหมายของคำศัพท์

# Lexical relation ความสัมพันธ์ของคำศัพท์

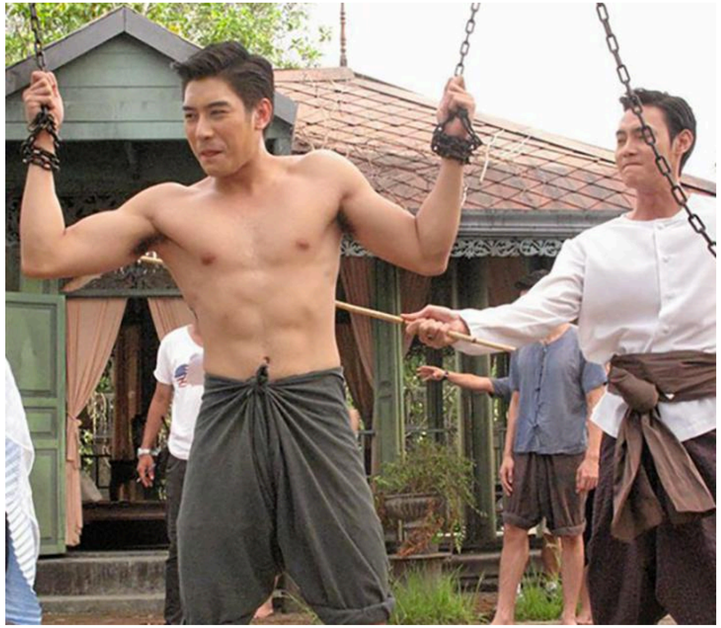
- Synonym: {dirty, filthy, dusty, septic} {clean, sterile}
- Antonym: dirty vs clean, easy vs hard

king

antonym?

queen

slave



# Lexical relation ความสัมพันธ์ของคำศัพท์


- Hyponym and hypernym:

  dog is an animal <—> *dog* is a hyponym of *animal* AND

  *animal* is a hypernym of *dog*

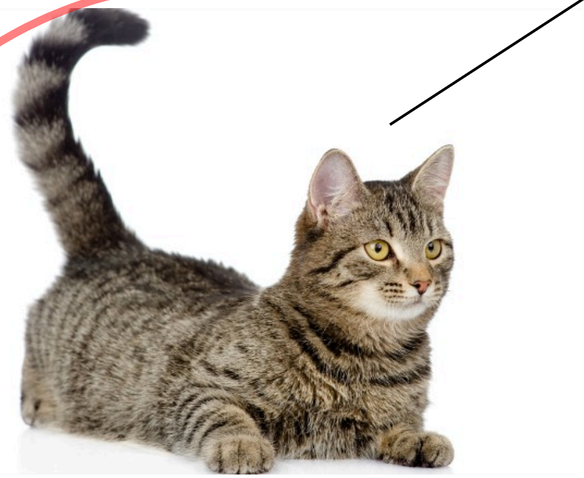
  cat is an animal <—> *cat* is a hyponym of *animal*

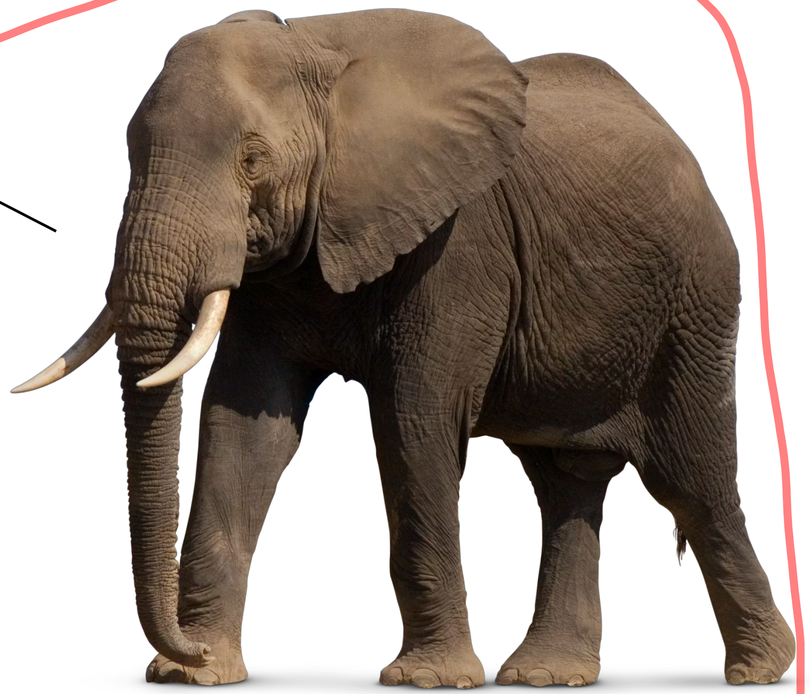
- Cohyponym:

  dog and cat have the same hypernym.

Small animal

# Animal





Lexical Semantics คือ การศึกษาความหมายของคำศัพท์

Computational Lexical Semantics คือ การสร้างโมเดลการคำนวณเพื่อการศึกษาความหมายของคำศัพท์

# Lexical Relation ไม่เพียงพอในการอธิบายความหมายของคำ

*"You shall know a word by the company it keeps."*

John Rupert Firth
(1957)

*"Die Bedeutung eines Wortes ist sein Gebrauch in der Sprache"*

Ludwig Wittgenstein
Philosophische Untersuchungen (1953)

# Lexical relation: Polysemy and Homonymy

- Bank
  - The economy urged the **bank** to raise the interest rate.
  - A new restaurant opened between the **bank** and the store.
  - The west river **bank** corroded due to the tide.

# Co-occurrence Statistics

- ลองนับดูว่าคำว่า bank เกิดขึ้นใกล้ (เช่น ห่างกันไม่เกิน 3 คำ) กับคำไหนบ้าง กี่ครั้ง

| | river | corroded | interest | rate | economy | between | next | ATM |
|---|---|---|---|---|---|---|---|---|
| bank (ที่ทำการธนาคาร) | 3 | 2 | 20 | 10 | 20 | 13 | 10 | 12 |
| bank (ธนาคาร) | 0 | 0 | 50 | 30 | 45 | 3 | 1 | 10 |
| bank (ฝั่งแม่น้ำ) | 20 | 30 | 1 | 3 | 2 | 5 | 6 | 0 |

similarity(king, queen) = 0.74

queen

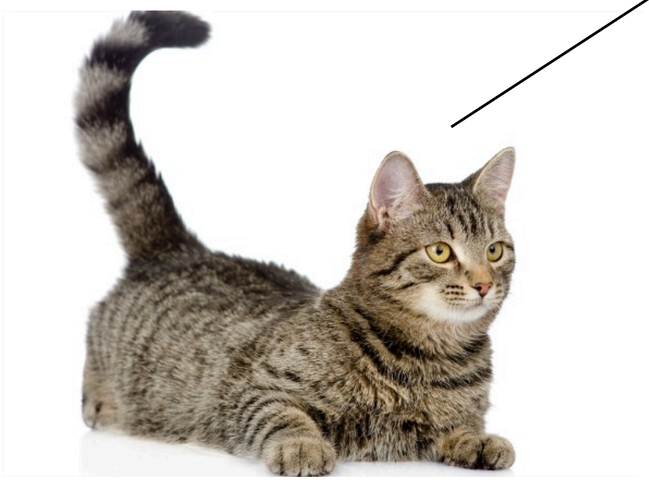
king

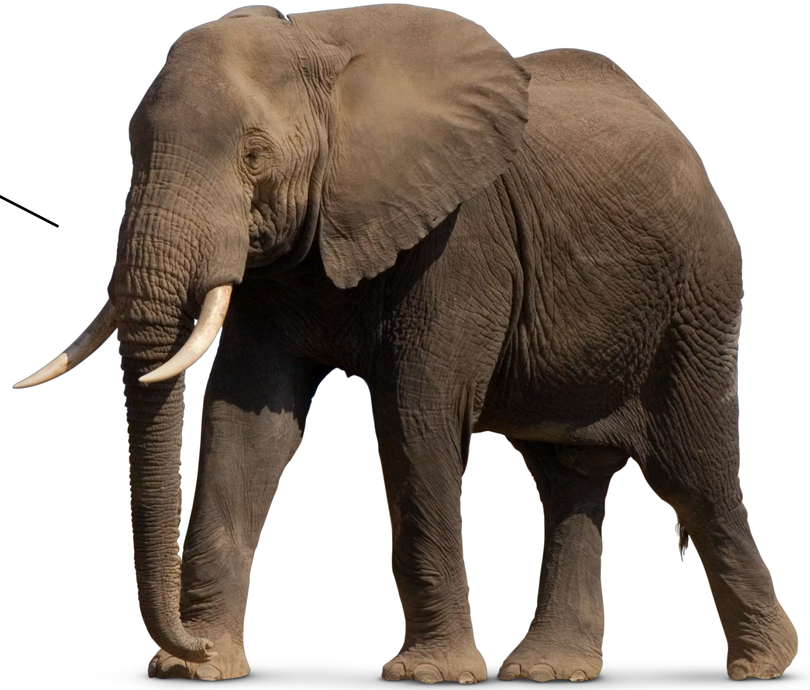
antonym?

slave

similarity(king, slave) = 0.34

# Animal





similarity(dog, cat) = 0.57

similarity(dog, elephant) = 0.34

# Distributional Semantics Model

Distributional Hypothesis posits that words that occur in the same contexts tend to have similar meanings.
(Zellig Harris, 1954)

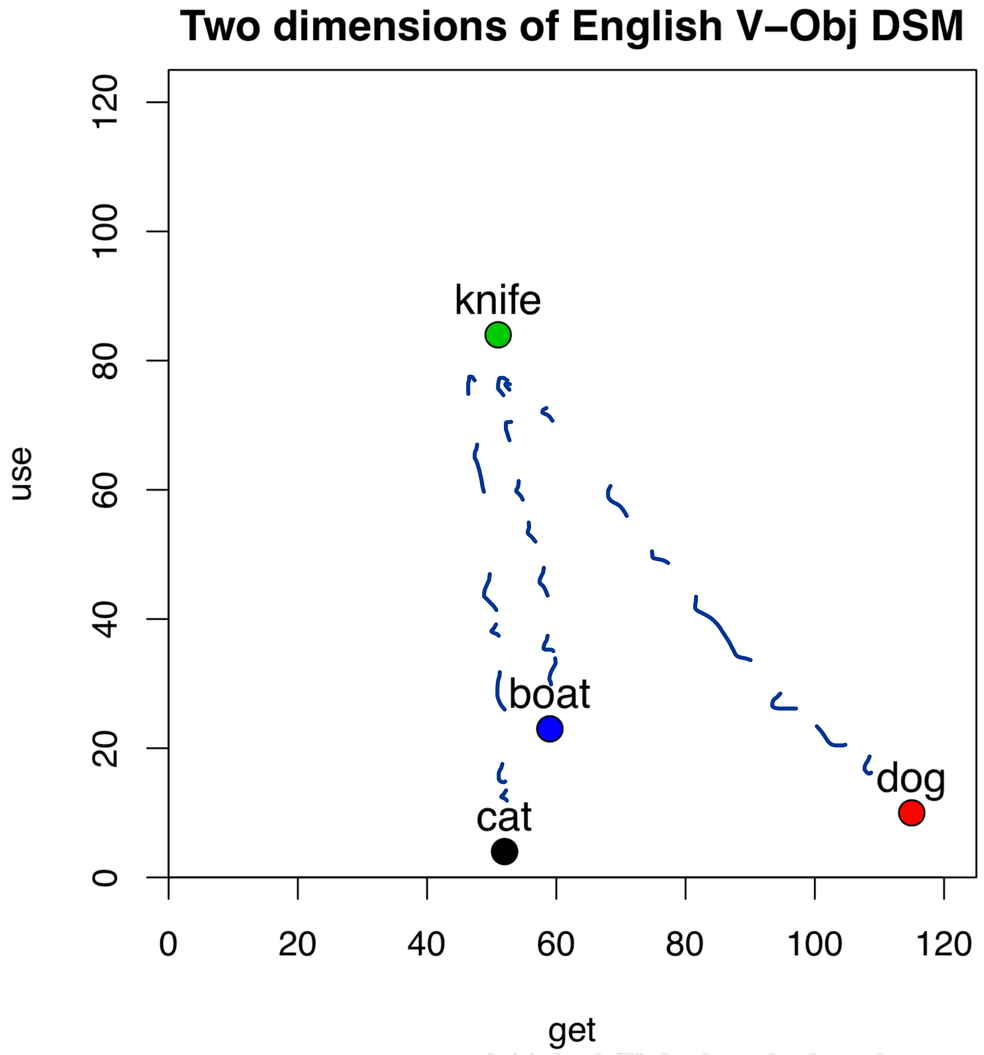
# word-context matrix

| | get | see | use | hear | eat | kill |
|---|---|---|---|---|---|---|
| knife | 51 | 20 | 84 | 0 | 3 | 0 |
| cat | 52 | 58 | 4 | 4 | 6 | 26 |
| dog | 115 | 83 | 10 | 42 | 33 | 17 |
| boat | 59 | 39 | 23 | 4 | 0 | 0 |
| cup | 98 | 14 | 6 | 2 | 1 | 0 |
| pig | 12 | 17 | 3 | 2 | 9 | 27 |
| banana | 11 | 2 | 2 | 0 | 18 | 0 |

# [x, y] = [#get, #use]

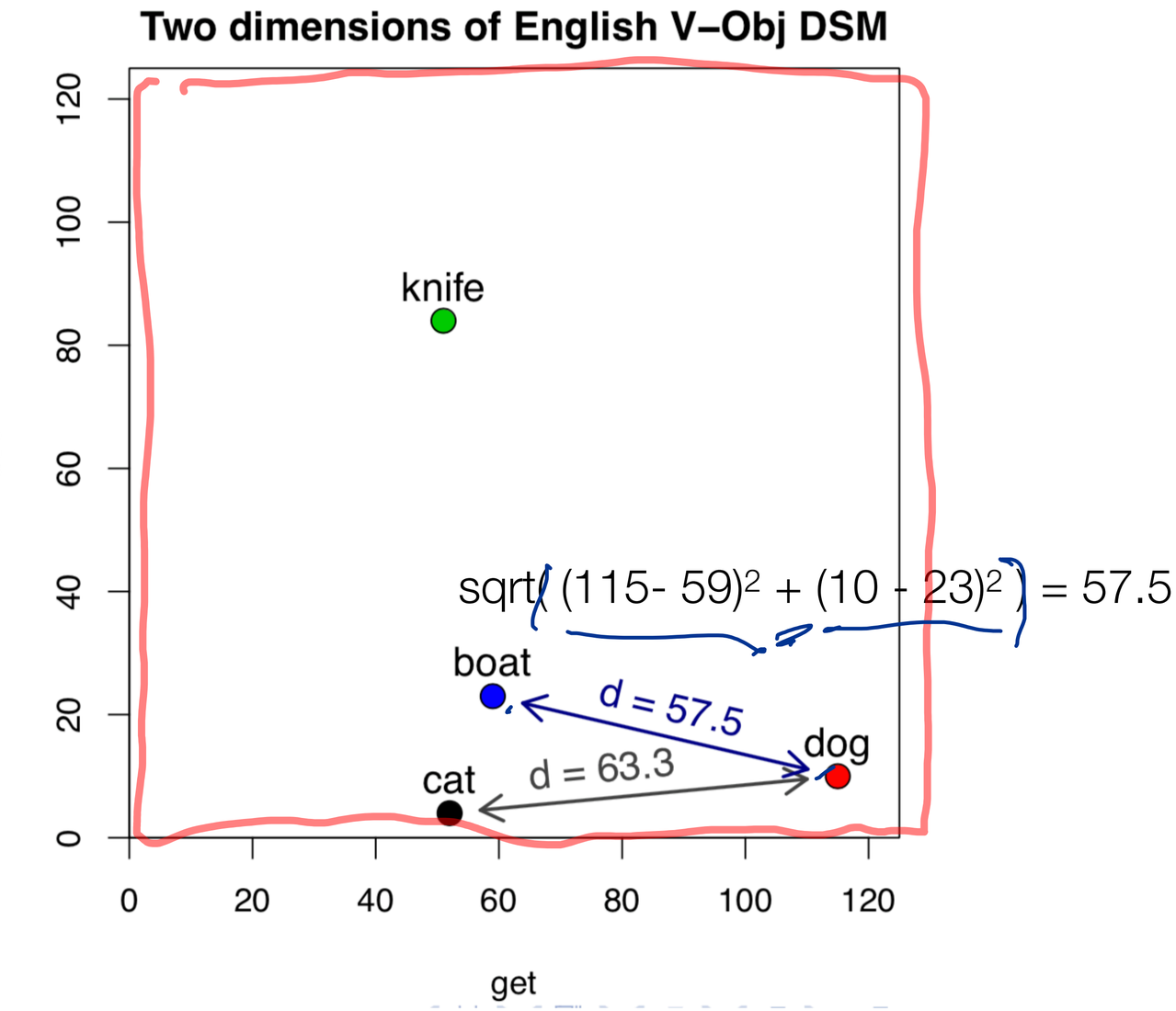
vector semantics

| | **get** | see | **use** | hear | eat | kill |
|---|---|---|---|---|---|---|
| knife | **51** | 20 | **84** | 0 | 3 | 0 |
| cat | **52** | 58 | **4** | 4 | 6 | 26 |
| dog | **115** | 83 | **10** | 42 | 33 | 17 |
| boat | **59** | 39 | **23** | 4 | 0 | 0 |
| cup | **98** | 14 | **6** | 2 | 1 | 0 |
| pig | **12** | 17 | **3** | 2 | 9 | 27 |
| banana | **11** | 2 | **2** | 0 | 18 | 0 |



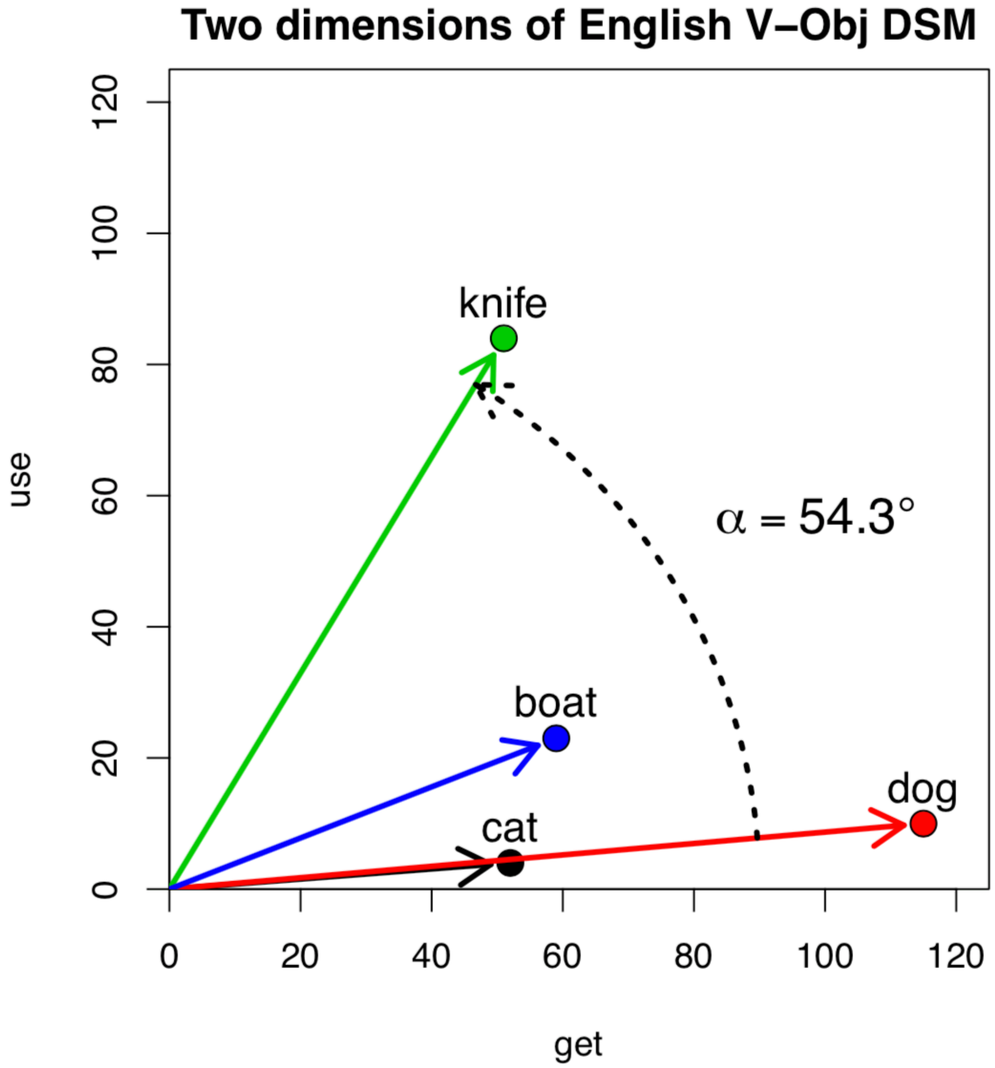
# [x, y] = [#get, #use]

| | get | use |
|---|---|---|
| knife | 51 | 84 |
| cat | 52 | 4 |
| dog | 115 | 10 |
| boat | 59 | 23 |
| cup | 98 | 6 |
| pig | 12 | 3 |
| banana | 11 | 2 |



# [x, y] = [#get, #use]

| | get | use |
|---|---|---|
| knife | **51** | **84** |
| cat | **52** | **4** |
| dog | **115** | **10** |
| boat | **59** | **23** |
| cup | **98** | **6** |
| pig | **12** | **3** |
| banana | **11** | **2** |



# Cosine distance and similarity

| | get | use |
|---|---|---|
| knife | **51** | **84** |
| cat | **52** | **4** |
| dog | **115** | **10** |
| boat | **59** | **23** |
| cup | **98** | **6** |
| pig | **12** | **3** |
| banana | **11** | **2** |

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}},$$

dot product

math.cosine

similarity(knife, dog) = 0.59

distance = 1 - similiarity

distance(knife, dog) = 1 - 0.59 = 0.41

similarity(cat, dog) = 1

distance(cat, dog) = 1 - 1 = 0

$$\frac{51 \times 115 + 84 \cdot 10}{\left(\sqrt{51^2 + 84^2}\right)\left(\sqrt{115^2 + 10^2}\right)}$$

# word-context matrix

| | get | see | use | hear | eat | kill |
|---|---|---|---|---|---|---|
| knife | 51 | 20 | 84 | 0 | 3 | 0 |
| cat | 52 | 58 | 4 | 4 | 6 | 26 |
| dog | 115 | 83 | 10 | 42 | 33 | 17 |
| boat | 59 | 39 | 23 | 4 | 0 | 0 |
| cup | 98 | 14 | 6 | 2 | 1 | 0 |
| pig | 12 | 17 | 3 | 2 | 9 | 27 |
| banana | 11 | 2 | 2 | 0 | 18 | 0 |

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}},$$

norm

```
In [46]: import numpy as np

In [47]: dog = np.array([115,83,10,42,33,17])

In [48]: cat = np.array([52,58,4,4,6,26])

In [49]: dog.dot(cat) /
(np.linalg.norm(dog) * np.linalg.norm(cat))

Out[49]: 0.9229773468286717
```

# word-context matrix สร้างยังไง

syntactic context

- ใช้คำประเภทไหนบ้างที่จัดว่าเป็น context
- Context กว้างแคบแค่ไหน
- word vector ดึงออกมาจาก word-context matrix ได้ยังไง

| | get | see | use | hear | eat | kill |
|---|---|---|---|---|---|---|
| knife | 51 | 20 | 84 | 0 | 3 | 0 |
| cat | 52 | 58 | 4 | 4 | 6 | 26 |
| dog | 115 | 83 | 10 | 42 | 33 | 17 |
| boat | 59 | 39 | 23 | 4 | 0 | 0 |
| cup | 98 | 14 | 6 | 2 | 1 | 0 |
| pig | 12 | 17 | 3 | 2 | 9 | 27 |
| banana | 11 | 2 | 2 | 0 | 18 | 0 |