

Brain neuro  
Language  
Processing

# Introduction to Computational Korean Language Processing

고려대학교 컴퓨터학과

Brain-neuro Language Processing Lab

임희석

[limhseok@korea.ac.kr](mailto:limhseok@korea.ac.kr)

<http://blp.korea.ac.kr>

# Index

---

2

1

Human Language  
Processing and  
Computer Language  
Processing

2

Basic Techniques  
for NLP

3

Application  
Areas of NLP

# NLP

---

3

- Natural Language vs. Artificial Language

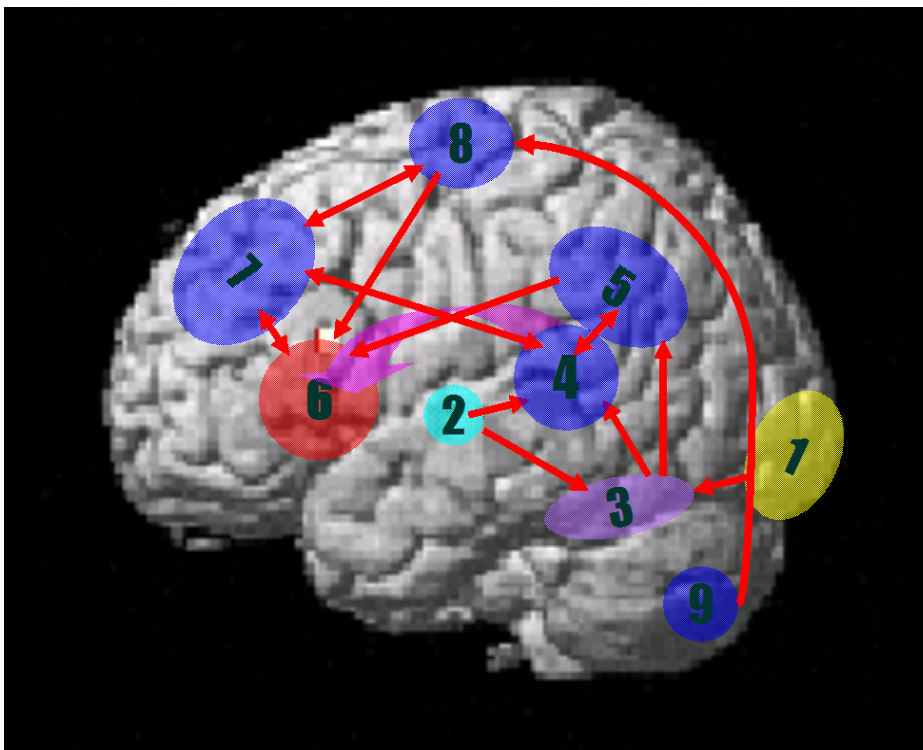
# Two Language Processors

---

- Human Brain
- Computer

# Human Neural Language Processing in Brain

5



1:Primary visual cortex : 초기 시각정보처리

2:Primary auditory cortex : 초기 청각정보처리

3:Middle temporal cortex : 의미정보처리

**4:Wernike's area : 언어이해중추**

5:SMG, angular gyrus : 음운표상 저장 및 처리

**6:Broca's area : 언어산출중추**

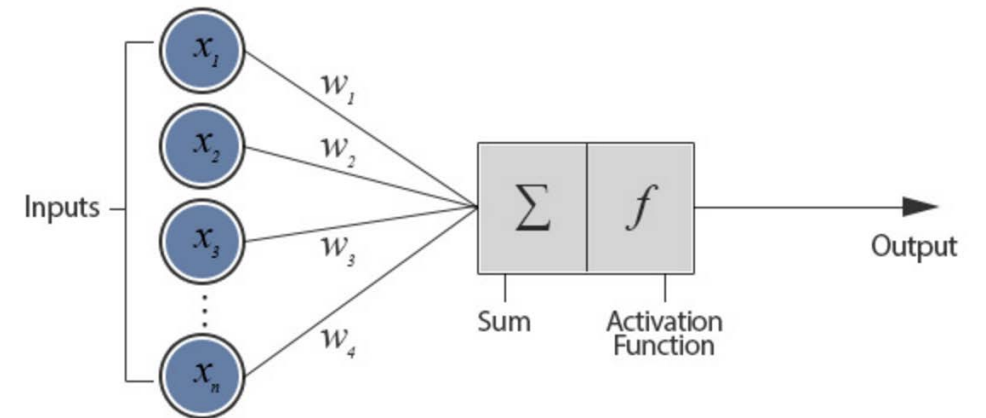
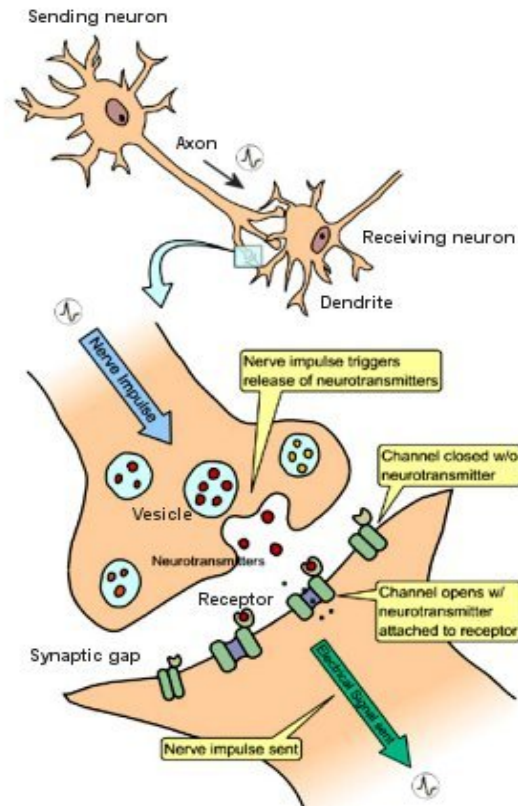
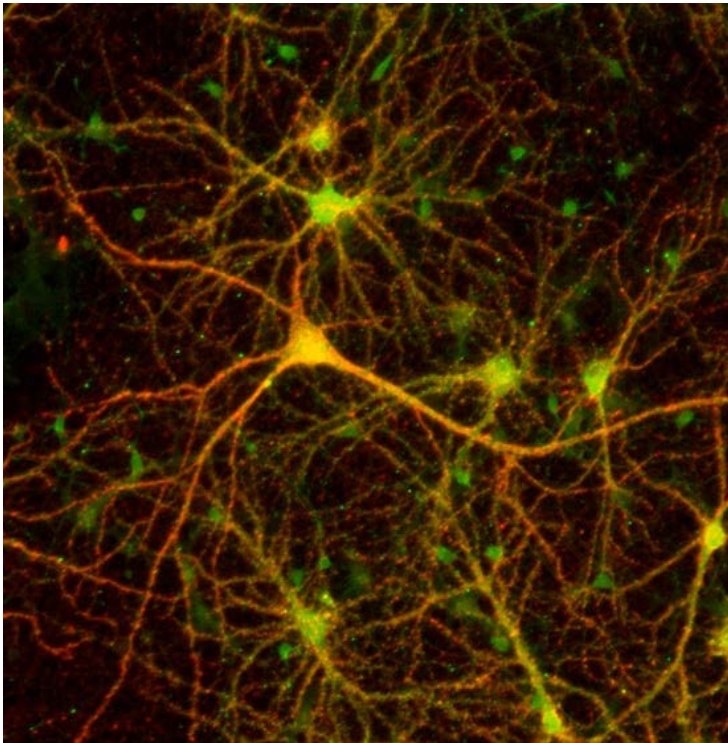
7:Prefrontal area : 주의, 정보선택

8:SMA, Premotor cortex : 주의, 정보의 선택

9:Cerebellum : 운동/발화 계획, 조음기관 통제

# Neural Network

6



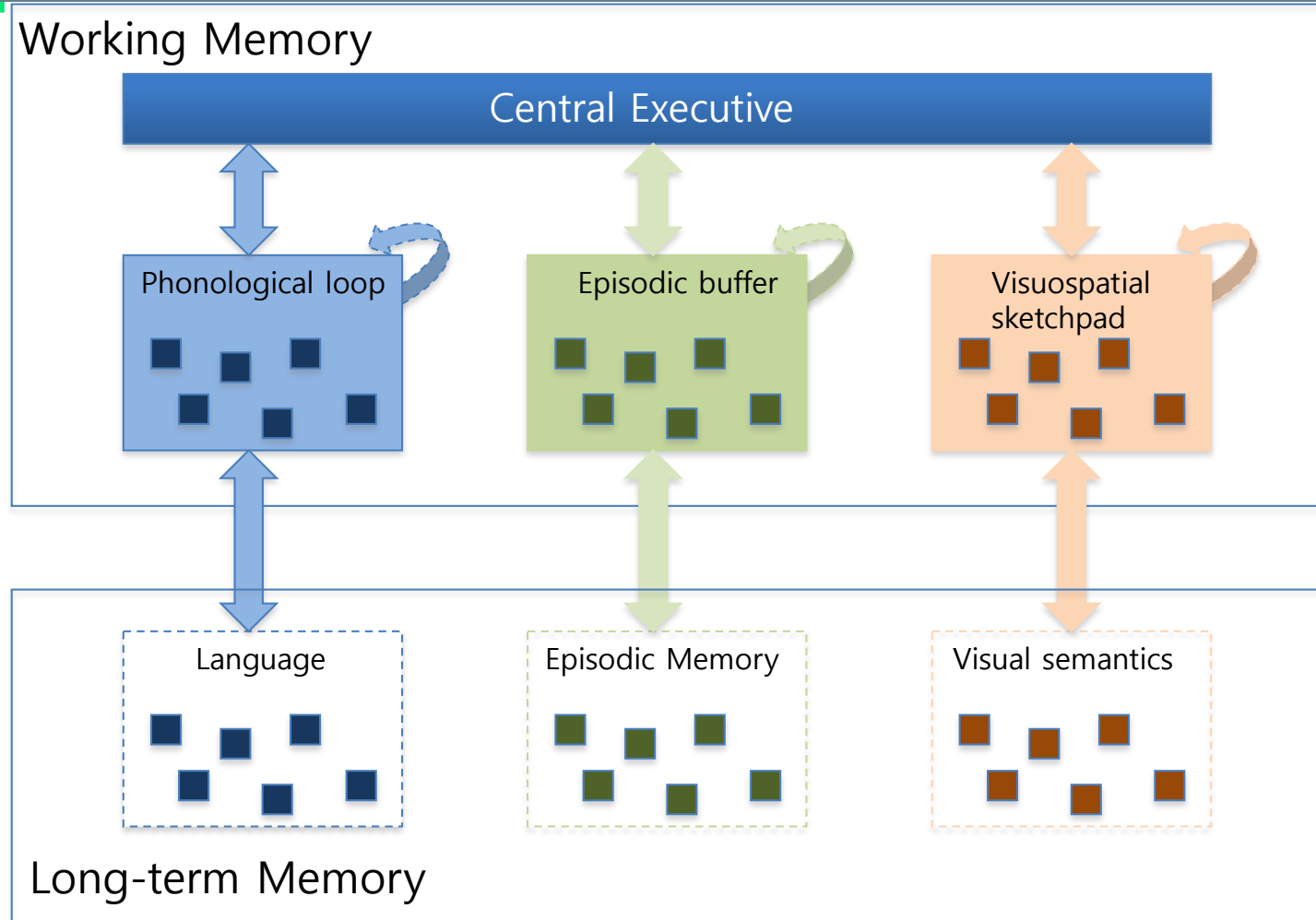
Neural Network in Brain

A Neuron and A Synapse

An Artificial Neuron

# Working and Long-term Memory

7



Permanent Learning is transition of information in WM to LTM?

# Computer Language Processing? (1/2)

8

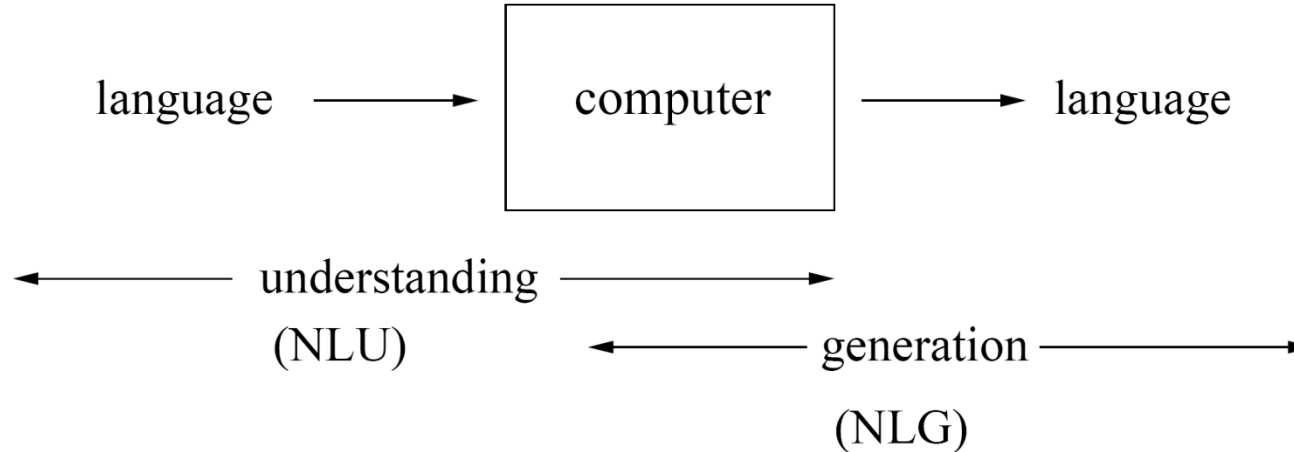
- Definition of Natural Language Processing
  - *A subfield of artificial intelligence and linguistics for making computers “understand” statements written in human languages*
- Two motivations for NLP (Allen, 1994)
  - The scientific or linguistic motivation is to understand the nature of language through the tools provided by computer science
  - The technological motivation is to improve communication between humans and machines



# Computer Language Processing? (2/2)

9

- One simple (but practical) answer
  - Computer using natural language as input and/or output



# Why do we research NLP?

---

10

- Two possible answers
  - Because natural language is the most effortless and effective communicative way of humans
  - Great majority of valuable information is still described and recorded by natural language

# Speech Recognition Applications(1/2)

11



Siri.  
말하는 대로  
이루어집니다.

Siri를 이용해 목소리로 메시지를 보내고, 회의 일정을 잡아 보세요. 전화를 걸어 보는 건 어떨까요? Siri는 다양하고 많은 일을 할 수 있습니다. 평소에 말하듯 자연스럽게 Siri에게 말해 보세요. Siri는 당신이 하는 말을 알아듣고 그 뜻을 파악해 목소리로 답을 준답니다. 정말 사용하기 쉽고 정말 많은 일을 하는 Siri. 쓰면 쓸수록 점점 활용 방법이 다양해 집니다.



# Speech Recognition Applications(2/2)

12

Dragon Speech Recognition Software

[+ See all Dragon products](#)

## Let your voice do all the typing

Dragon, the world's best-selling speech recognition software, lets you dictate documents, search the web, email and more on your computer — quickly and accurately — just by using your voice. You don't even have to lift a finger.

Find your Dragon:

[For PC](#)

[For Mac](#)

[What's new](#) [Why choose Dragon](#) [Dragon business](#) [Support](#)

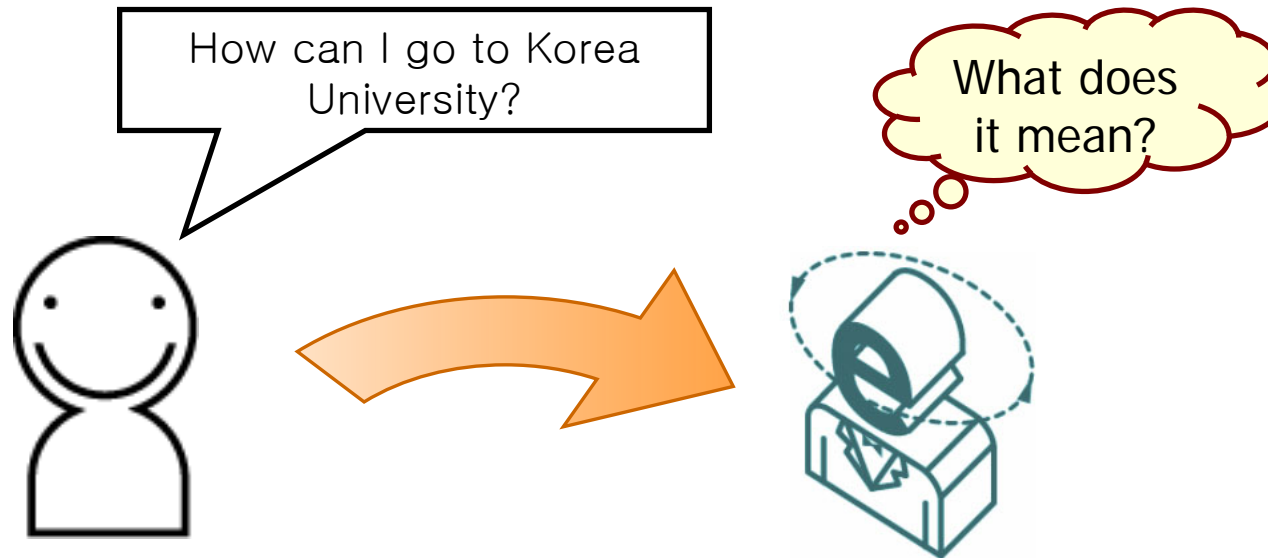


- <https://todaysmeet.com/ailecture>

# NLP techniques for Speech Recognition Applications(1/2)

14

- At least two areas, we need natural language processing techniques for building such applications
  - To map a sentence into a computer-understandable form (컴퓨터가 이해할 수 있는 형태로 변환)

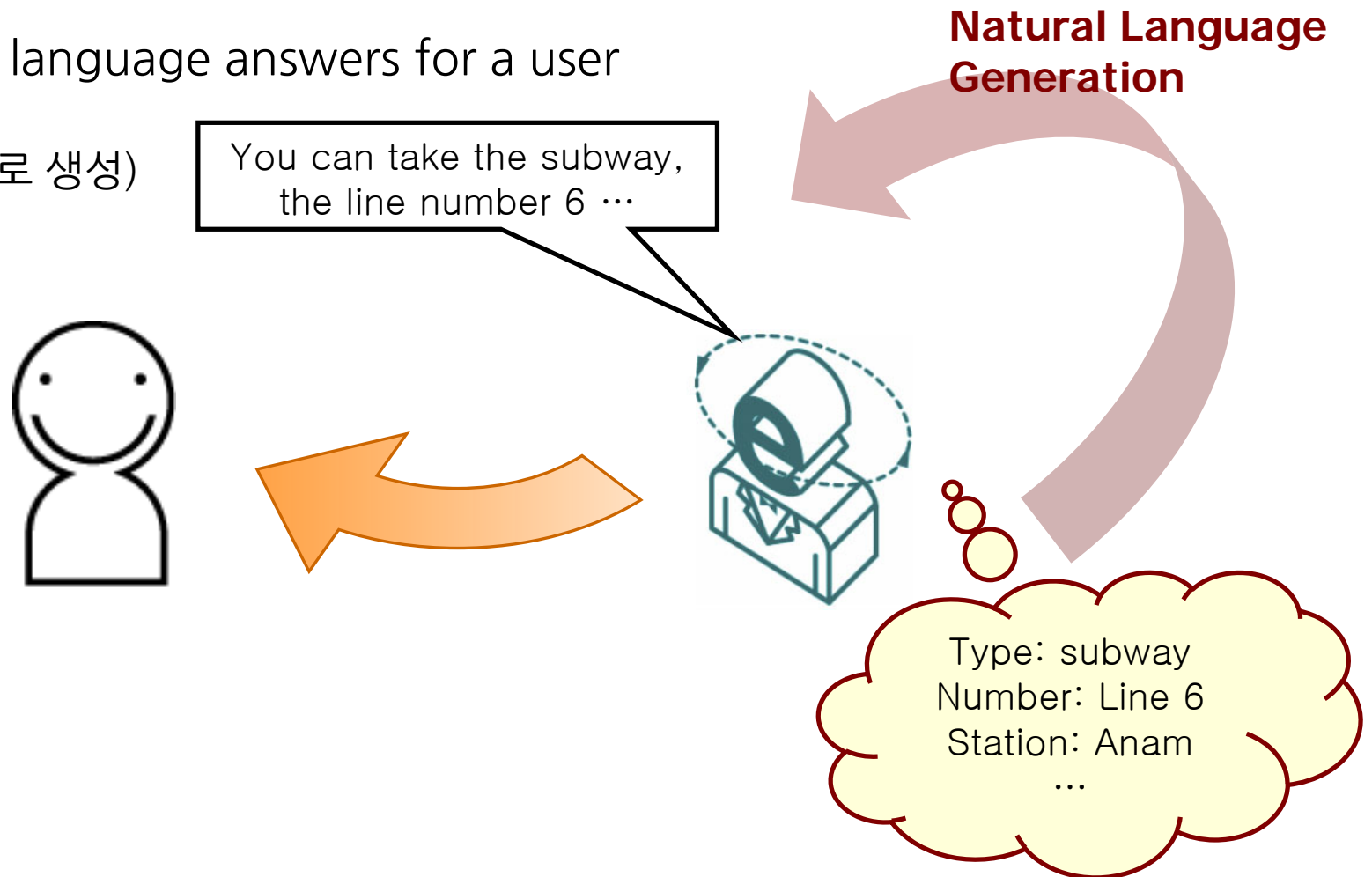


# NLP techniques for Speech Recognition Applications(2/2)

15

- To generate natural language answers for a user

(사용자를 위하여 자연어로 생성)



# Why NLP is hard?

---

16

Because of **ambiguity!**

“At last, a computer that understands you like your mother”



# Ambiguity (중의성)

17

“At last, a computer that understands you like your mother”

- Possible interpretations
  - A computer understands you as your mother understands you  
(어머니가 당신을 이해하는 것처럼 당신을 이해하는 컴퓨터)
  - It understands (that) you like your mother  
(당신이 당신의 어머니를 좋아한다는 사실을 컴퓨터가 이해한다.)
  - It understands you as it understands your mother  
(컴퓨터가 당신의 어머니를 이해하는 것처럼 당신도 이해한다.)
- What is the correct interpretation?

# Ambiguities at Multi Levels (1/3)

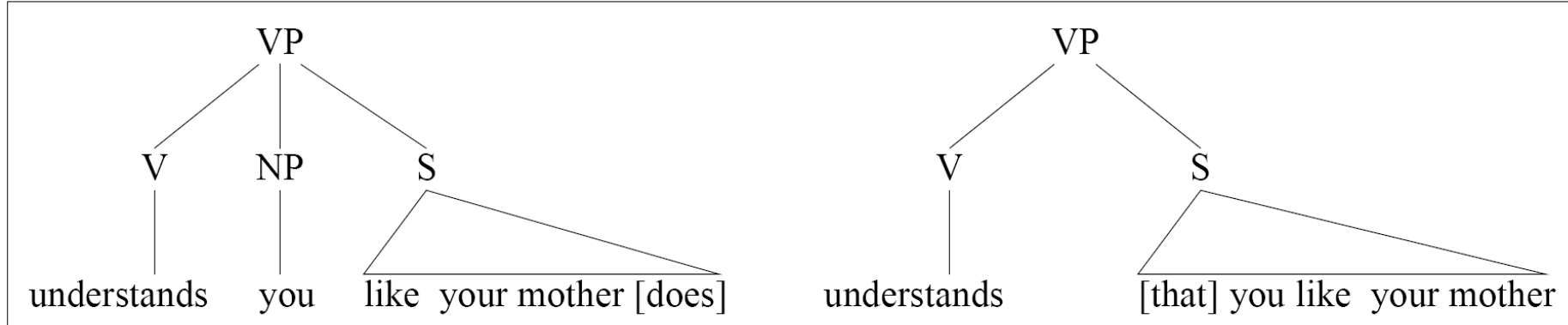
18

- At the word level (word ambiguity)
  - “At last, computer understand you *like* your mother”
  - Part of speech tag of the word “*Like*”
    - can be [Verb]
    - Or can be [Preposition]

# Ambiguities at Multi Levels (2/3)

19

- At the syntactic level (structural ambiguity)
  - “At last, a computer that understands you like your mother”



- Different structures lead to different semantic interpretations

# Ambiguities at Multi Levels (3/3)

20

- At the semantic level (word sense ambiguity)
  - “At last, a computer understand you like your *mother*”
  - The word ‘Mother’ has two different meanings in English

#1. 엄마

#2. 효모 (a stringy slimy substance consisting of yeast cells and bacteria)

# How to deal with ambiguities?

21

- We need
  - Knowledge about language (언어에 관한 지식)
    - For example, grammar, dictionary, ...
  - Knowledge about the world (세상에 관한 지식)
    - Language acts are the product of agents in the world, either human or computer
    - For example, knowledge about the preference of a human, situation, ...

# Major Issues for NLP

---

22

- Thus, major issues for NLP would be
  - How to acquire necessary knowledge?
  - How to use such knowledge for resolving ambiguity?

# Two Major Approaches for NLP

23

- Symbolic approach (or rule-based approach)
  - Built all necessary information by human's hand
  - Require excessive manual works
- Statistical approach
  - Try to automatically infer knowledge from samples (manually annotated corpus)
- Neural approach - Deep Learning
  - Build a neural language model through Deep Learning
  - Does not need handcrafted knowledge or probabilities from samples

# 기반 기술

- Preprocessing (전처리)
- Morphological Level (형태소 단계)
- Syntactic Level (구문단계)
- Semantic Level (의미 단계)
- Discourse Level (담화단계)



# Preprocessing (Error Correction)

25

- Problem domain
  - Correct all typographical errors in text before processing the given text
  - Previous research of error correction has focused on two types of errors:
    - Spacing error (띄어쓰기 오류)
      - “아버지가방에 들어 가셨다”
        - “아버지가 방에 들어 가셨다.” vs. “아버지 가방에 들어 가셨다.”
    - Spelling error (철자오류)
      - “나는 하교에 갔다”
        - “나는 학교에 갔다”
    - Both errors will significantly deteriorate performance of whole NLP system.
    - Complex errors (복합 오류) (e.g. “lemme C” => “Let me see.” )

# Morphological Analysis

26

- 형태소 분석기 역할
  - 어절의 가능한 모든 형태소 분석열을 생성
  - 한국어 정보처리를 위한 필수적 역할

나는	수영이	나는	이유를	물었다
나/N+는/J 나/V+는/E 날/V+ㄴ/E	수영/N+이/J	나/N+는/J 나/V+는/E 날/V+ㄴ/E	이유/N+를/J	물/V+었/PE+다/E 묻/V+었/PE+다/E

- 필요성
  - 문장 이해를 위해서 어절에 대한 정보가 필요함
  - 생성 가능한 어절의 수가 무한하여 어절 단위로 전자 사전을 구성하는 것이 불가능

# POS Tagging

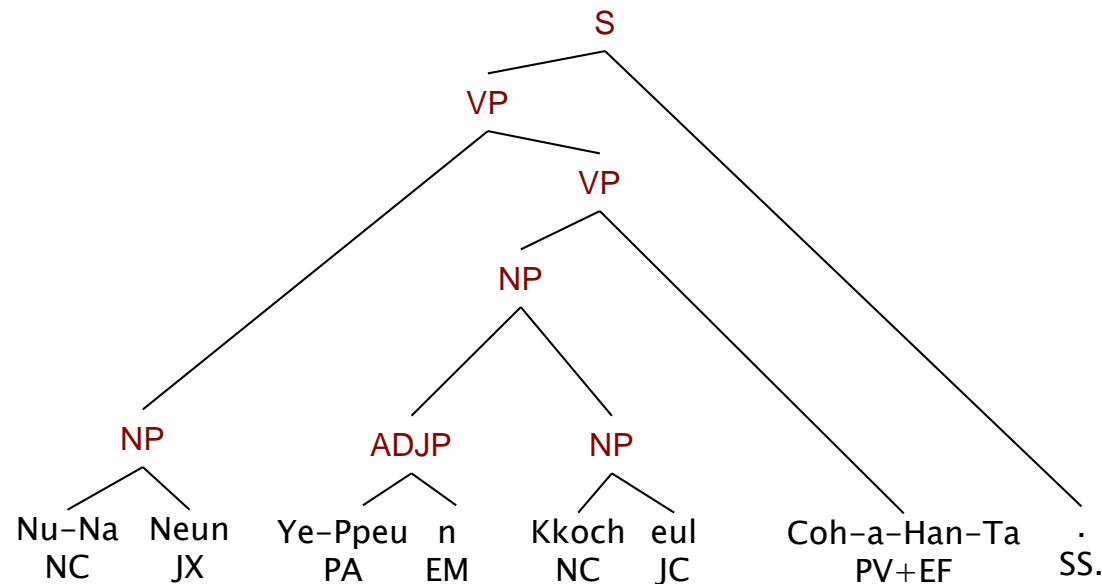
- 품사 태깅이란?
  - 단어(어절)의 형태론적 중의성을 해소하여 올바른 품사를 할당하는 작업
  - 단어(어절) 중의성 해소 작업

나는	수염이	나는	이유를	물었다
나/N+는/J 나/V+는/E 날/V+ㄴ/E	수염/N+이/J	나/N+는/J 나/V+는/E 날/V+ㄴ/E	이유/N+를/J	물/V+았/PE+다/E 묻/V+았/PE+다/E

# Syntactic Parsing 구문분석 (1/2)

28

- Goal
  - Find out the syntactic structure with a specific grammar for a given sentence
    - Example of parsed sentence with the phrasal structure grammar
      - “누나는 예쁜 꽃을 좋아한다.”

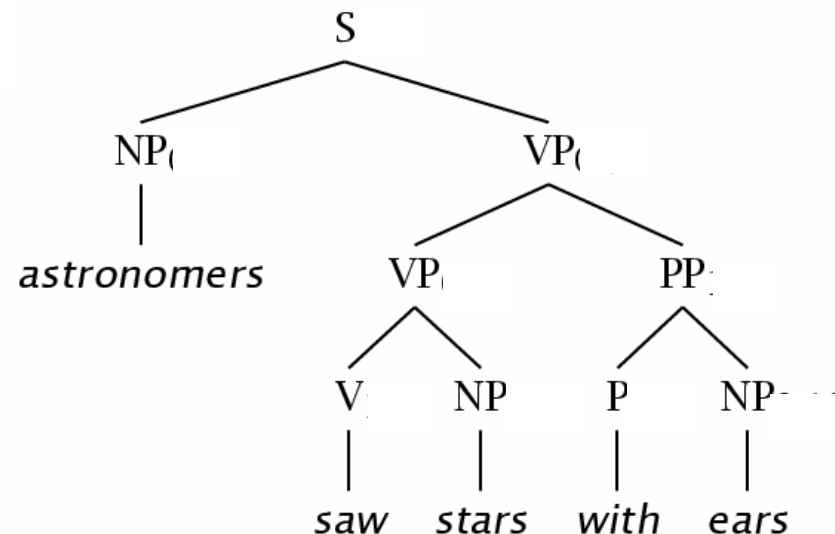
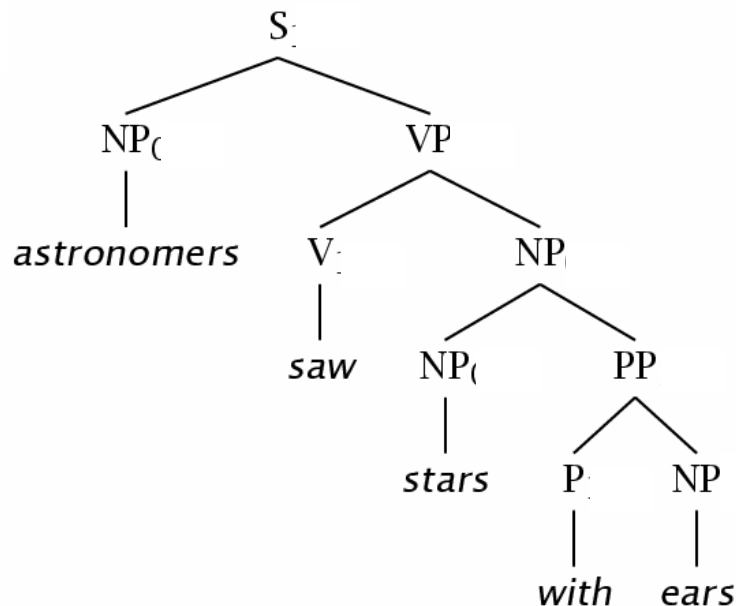


# Syntactic Parsing 구문분석 (2/3)

29

- Example of statistical parsing method based on PCFG
  - The sentence “*Astronomers saw stars with ears*” can be analyzed in two ways;

How to decide which one is correct?

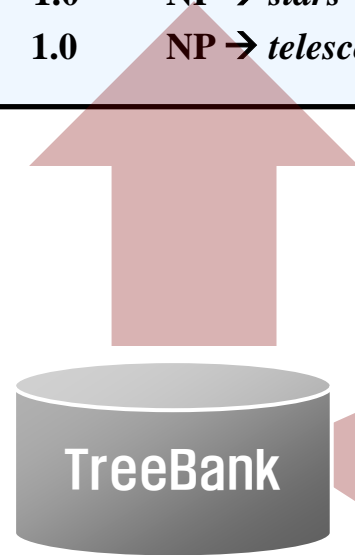


# Syntactic Parsing 구문분석 (3/3)

30

- Infer PCFG grammar from Tree bank

$S \rightarrow NP VP$	1.0	$NP \rightarrow NP PP$	0.4
$VP \rightarrow V NP$	0.7	$NP \rightarrow \textit{astronomers}$	0.1
$VP \rightarrow VP PP$	0.3	$NP \rightarrow \textit{ears}$	0.18
$PP \rightarrow P NP$	1.0	$NP \rightarrow \textit{saw}$	0.04
$P \rightarrow \textit{with}$	1.0	$NP \rightarrow \textit{stars}$	0.18
$V \rightarrow \textit{saw}$	1.0	$NP \rightarrow \textit{telescope}$	0.1



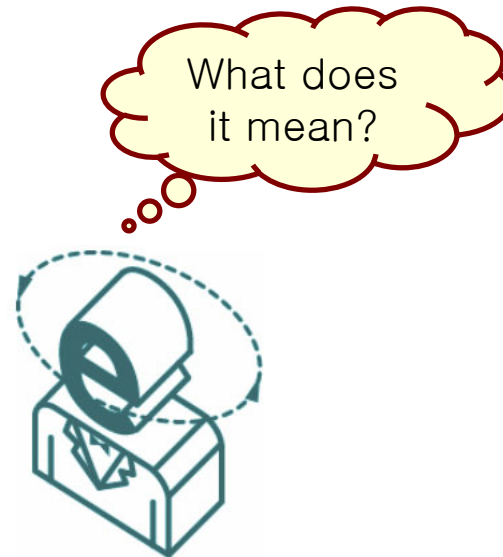
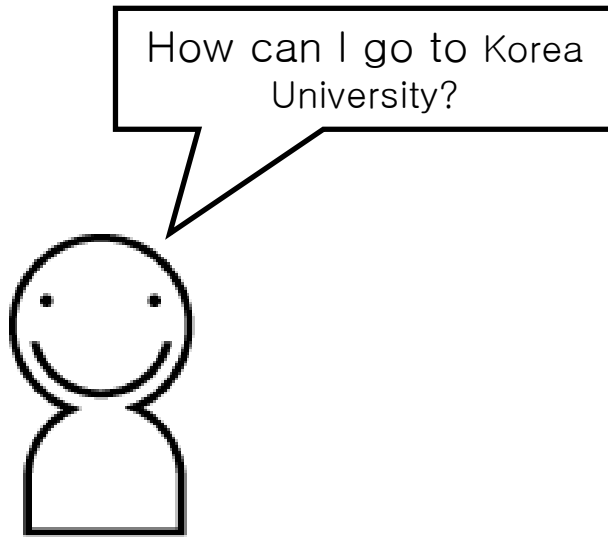
Example of treebank

; 사람이 스스로 만물의 영장이라 하고 우짚대는 까닭이 여기에 있다.  
(S  
(ADJP  
(NP  
(VP (NP 사람 )+이  
(VP  
(VP (ADVP 스스로 )  
(VP  
(VP  
(NP (NP 만물 )+의  
영장 )+이 )+라  
하 )+고  
우짚대 )+는 까닭 )+이  
(ADJP (NP 여기 )+에 있 )+다 +. )

# Semantic Analysis (1/2)

31

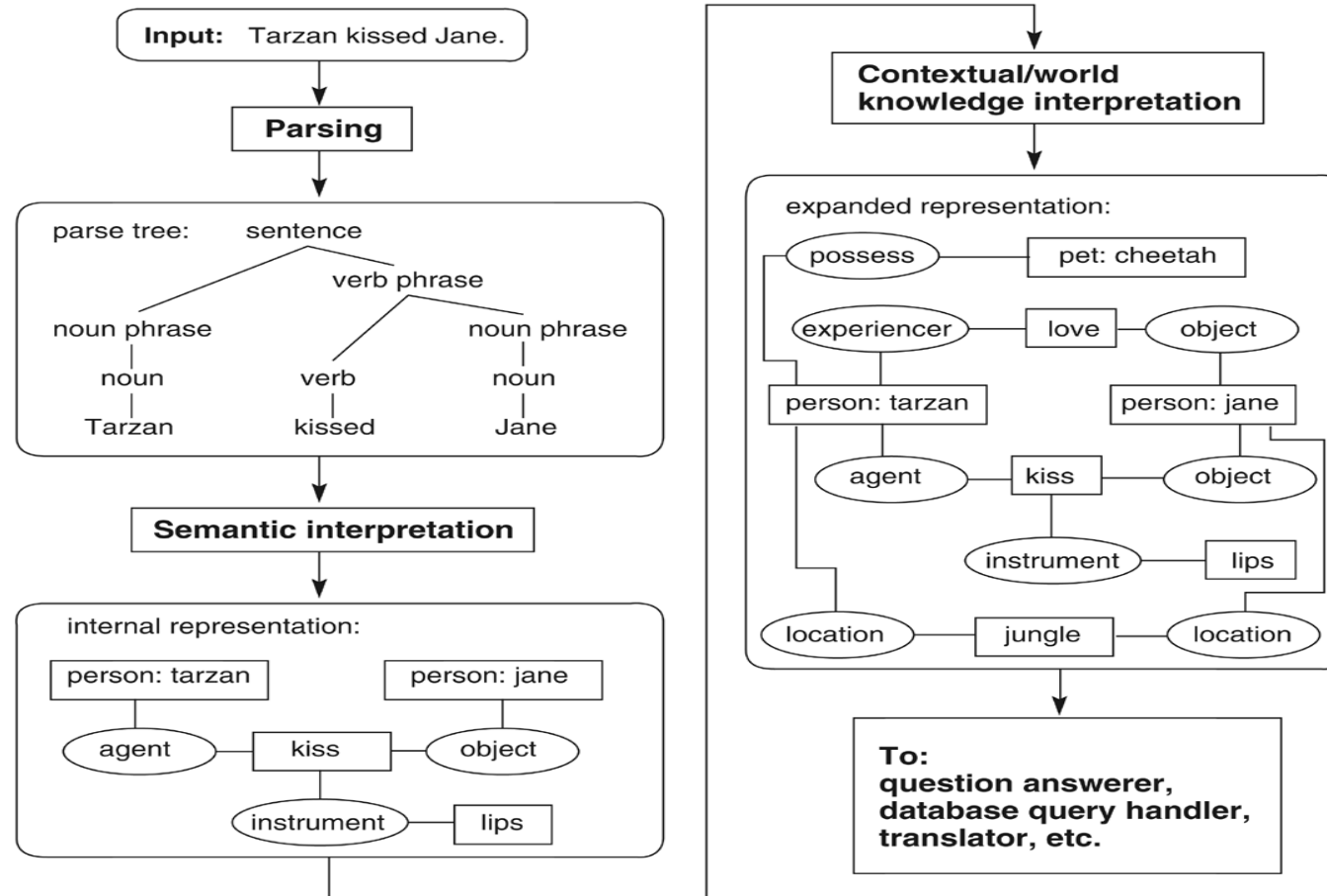
- Semantic analysis is
  - The process whereby **meaning representations are composed** and assigned to natural language text



# Semantic Analysis (2/2)

32

- Example of semantic analysis





# Semantic Role Labeling

33

- In a sentence, linguistic components (i.e., word, phrase) take some kind of semantic roles
  - e.g., Taehee kissed Bee
    - Taehee is an *Agent* which did certain action (kiss)
    - Bee is a *Patient* which got an effect from *Agent*
  - By analyzing the predicate and its semantic arguments, we can know the *shallow semantic structure* of the sentence
  - It gives us an identical semantic point of view in syntactic variations

**Taehee kissed Bee**

**Bee was kissed by Taehee**



Predicate: kiss  
Agent : Taehee  
Patient: Bee

# Word Sense Disambiguation

34

- Word Sense Disambiguation
  - Problem of **determining the proper sense** of a word used in a given sentence (단어의 적합한 의미를 결정하는 문제)

“A mother *crane* soon laid an egg.”

- crane 1 : a bird that has a long neck, long legs, and a long bill
- crane 2 : a device that lifts and moves heavy objects

- WSD is required :
  - For human language understanding
    - Word is the smallest meaningful unit
    - Sentences or documents consist of one or more words
  - For machine translation or information retrieval

# Discourse Analysis

35

- Reference resolution (대용어 해결)

- The omitted words (or phrases) and the pronominal **references** are complemented by the use of common sense and discourse information

U: I would like to open a **fixed deposit account**.

S: For what amount?

U: Make **it** for 8000 dollars.

- Speech Act Identification (화행분석)

- Speech Act: The **communicative intention** represented by each utterance (진술, 주장, 추측, 명령, 요청, 언약, 정표 등) **발화의 의도**

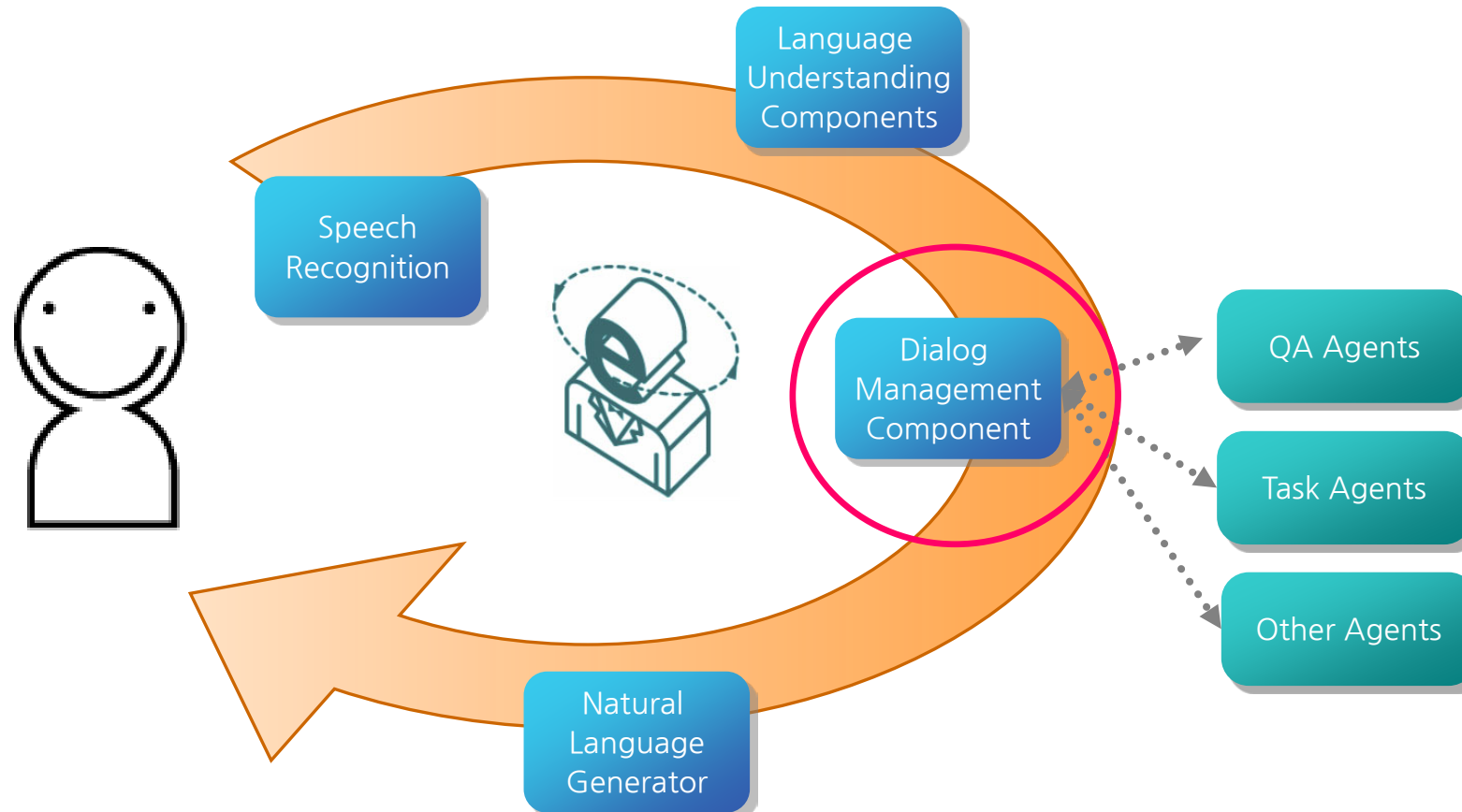
# 응용 기술

- Dialogue System
- Question and Answering System
- Clustering and Classification System
- Information Extraction
- Machine Translation

# Dialogue System (1/2)

37

- Typical Architecture of Dialog System



# Dialogue System (2/2)

38

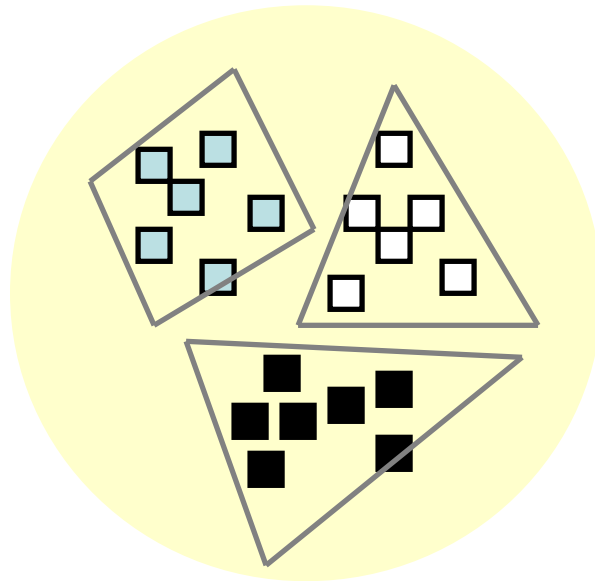
- Speech Techniques
- Text Understanding Techniques



# Document Clustering

39

- 문서 군집화 정의
  - 문서집합을 유사성이 높은 부분집합으로 분해하는 작업

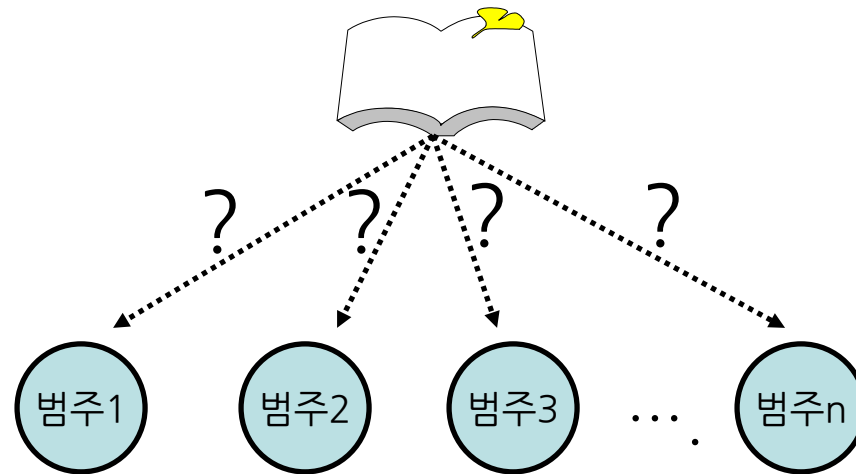


클러스터링

# Document Classification

40

- 문서 분류의 정의
  - 문서를 입력받아 입력문서가 포함될 수 있는 범주(분류)를 할당하는 작업



문서 분류

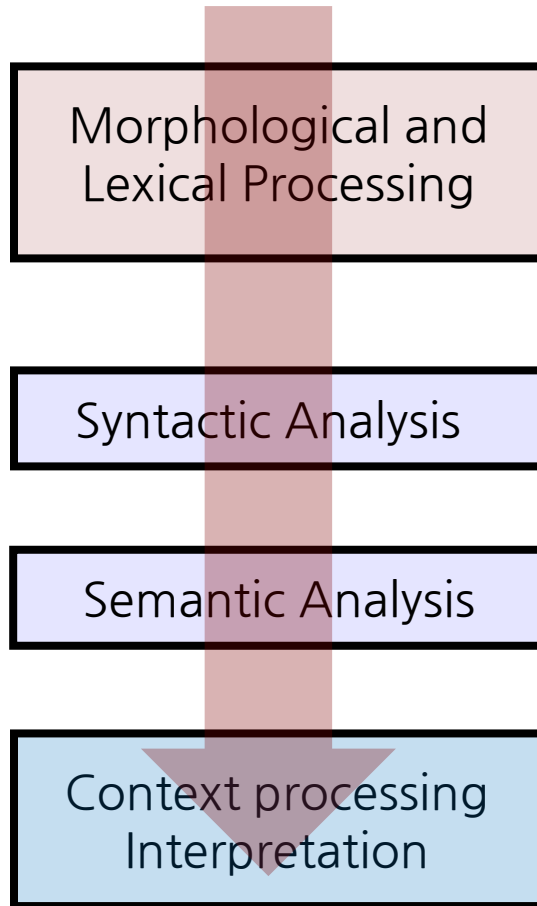


# Information Extraction

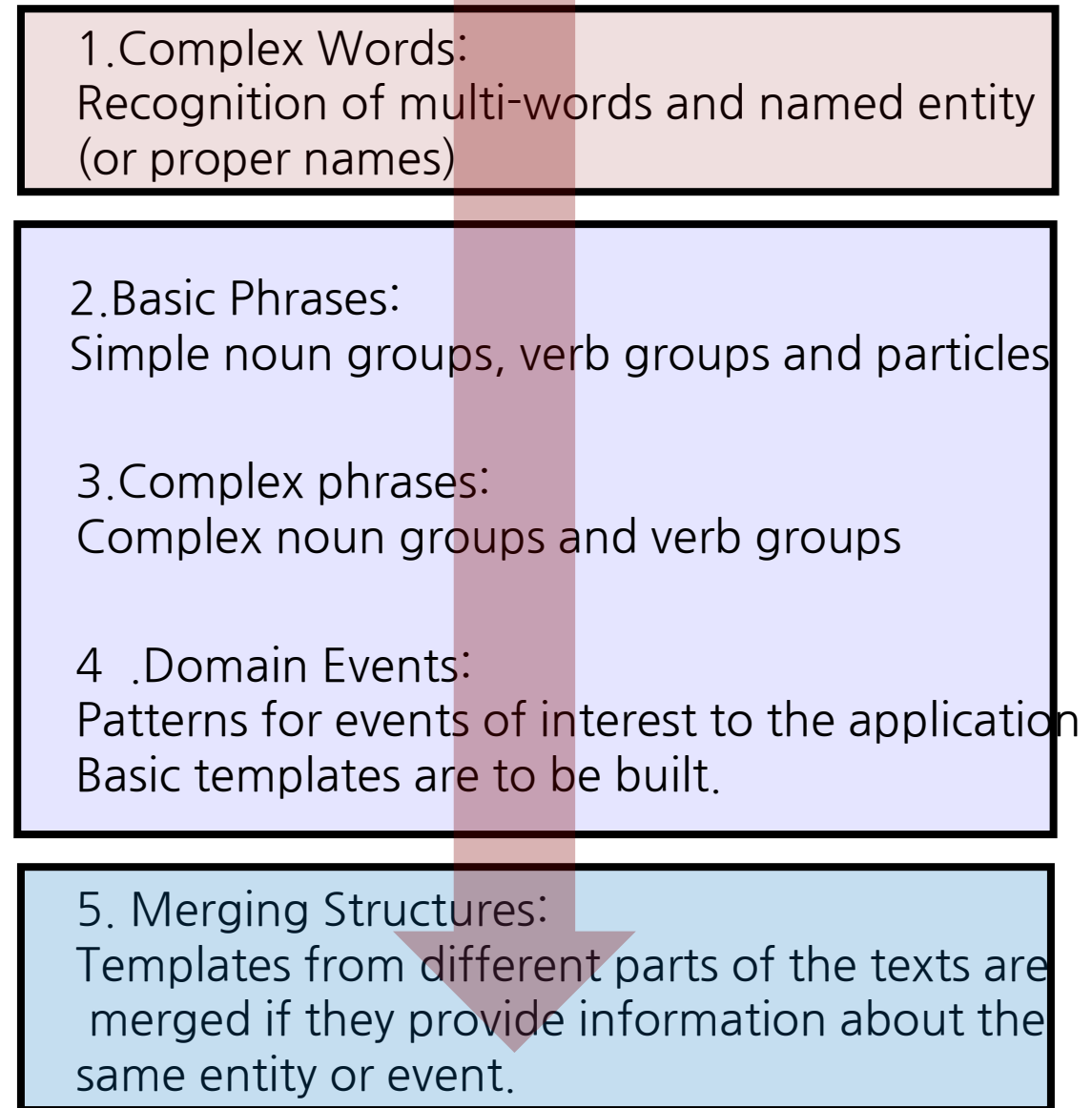
41

- Takes input a unrestricted text and summarizes it with respect to a pre-specified domain or topic of interest
  - Target of opportunity for Information Extraction (IE):
    - Text whose function is the communication of factual information
    - Intended audience for texts is broad
    - No requirement for modeling special character of either author or audience
    - Prime example: News stories
- In contrast to an in-depth NL understanding system, IE systems **skim a text to find relevant sections** and then **focus only on processing these sections**

## General Framework of NLP



## Framework of typical IE system



# Machine Translation approaches

43

- Machine translation (MT) is automated translation. It is the process by which computer software is used to translate a text from one natural language (such as English) to another (such as Korean).

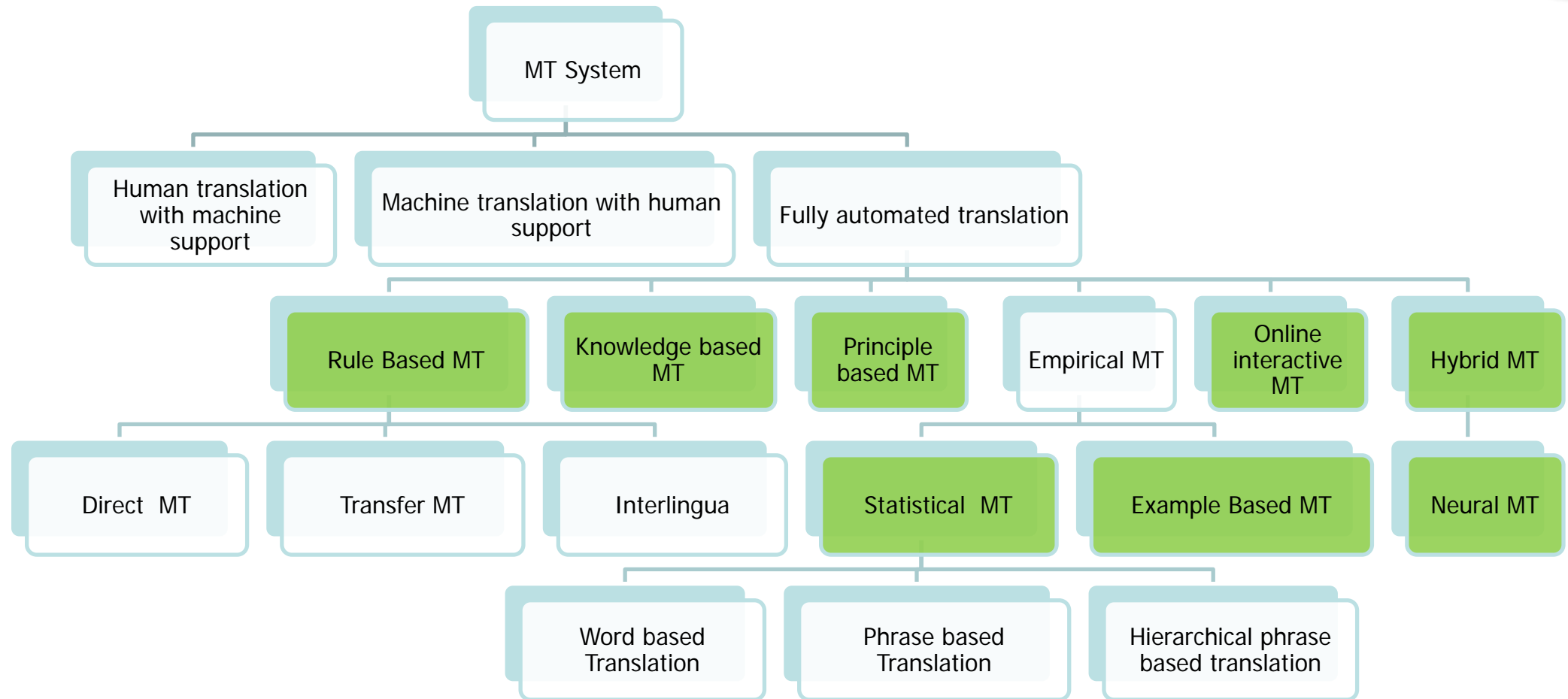


- Ex:-



# Machine Translation approaches

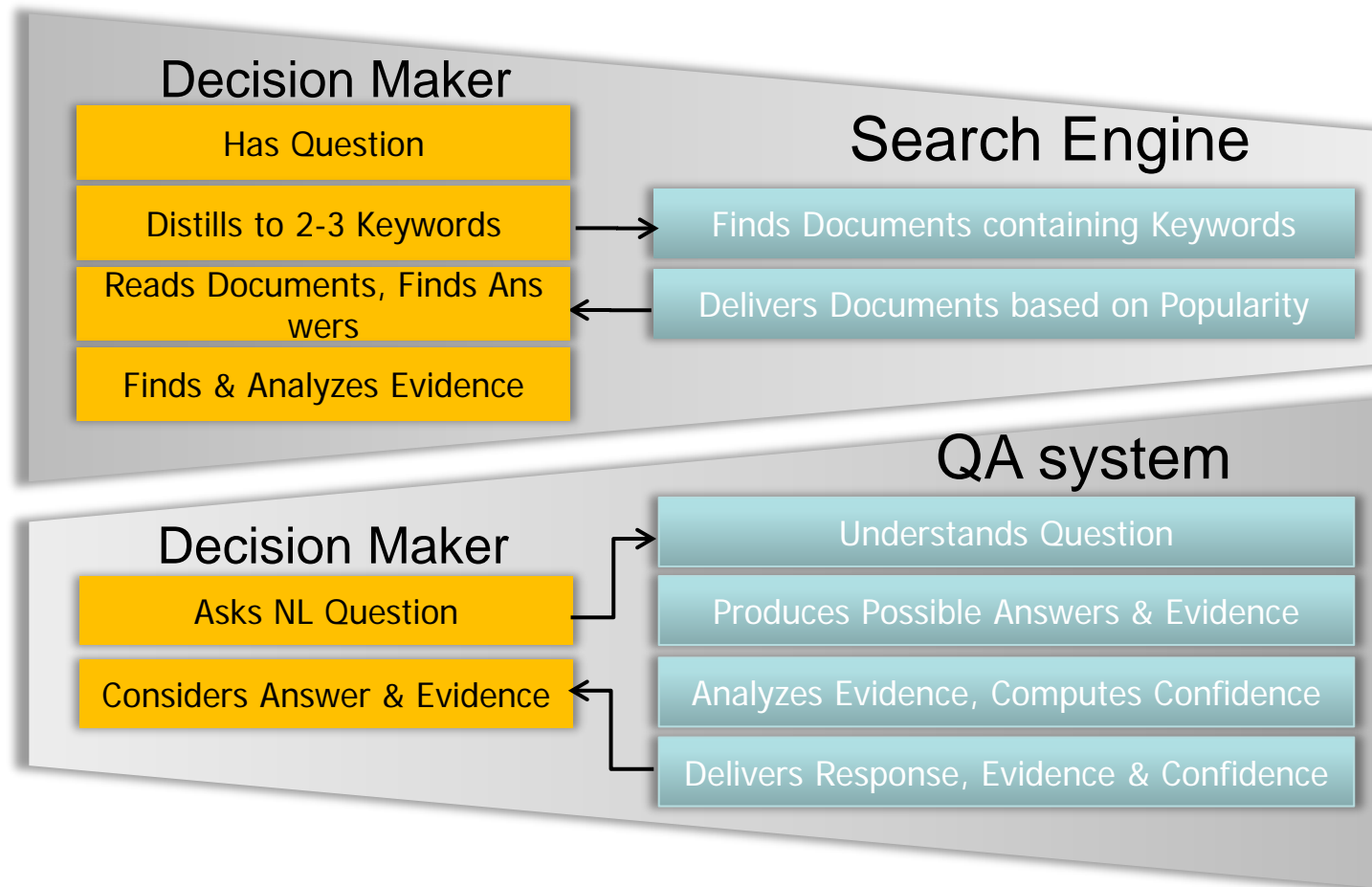
44



# Question & Answering

45

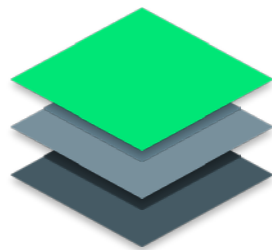
- 사용자의 질의에 대한 정답을 생성



# Question & Answering

46





감사합니다.

**Heuseok Lim**

Brain-neuro Language Processing Lab

Korea University, Seoul

limhseok@korea.ac.kr