

CHAPTER 7

의미 분석

목차

■ 7.1 단어와 단어 의미 중의성

- 중의성이란

■ 7.2 단어 의미 중의성 해소 기법

- 단어 중의성 해소
- 지식기반 방법
- 지도학습기반 방법

■ 7.3 의미역 분석

- 의미역이란
- 의미역의 종류
- 지도학습기반 방법

■ 7.4 의미표현

- 의미표현이란
- 추상 의미 표현

7.1 단어와 단어 의미 중의성

▪ 중의성이란

- 둘 이상의 의미를 가지는 표현
- 해석의 혼동으로 인해 원하는 결과에 방해를 주기때문에 적절한 문맥 정보와 함께 표현되어야 함
 - 예) 어휘적 중의성, 구조적 중의성

7.1 단어와 단어 의미 중의성

■ 어휘적 중의성

- 다의어에 의한 중의성
 - 예) 손 좀 보다
 - "신체 일부"를 의미
 - "수리 하다."의 의미
 - "혼을 내다."의 의미
- 동음어에 의한 중의성
 - 예) 밤이 좋다
 - "시간"를 의미
 - "음식"을 의미

7.1 단어와 단어 의미 중의성

- 구조적 중의성

- 수식어에 의한 중의성

- 예) 부유한 철수와 영희가 명품 매장을 갔다.
 - “부유한”의 수식어가 철수와 영희를 수식할 경우

7.2 단어 의미 중의성 해소 기법

▪ 단어 의미 중의성 해소

- 문장 내 중의성을 가지는 어휘를 사전에 정의된 의미와 매칭하여 어휘적 중의성을 해결하는 문제

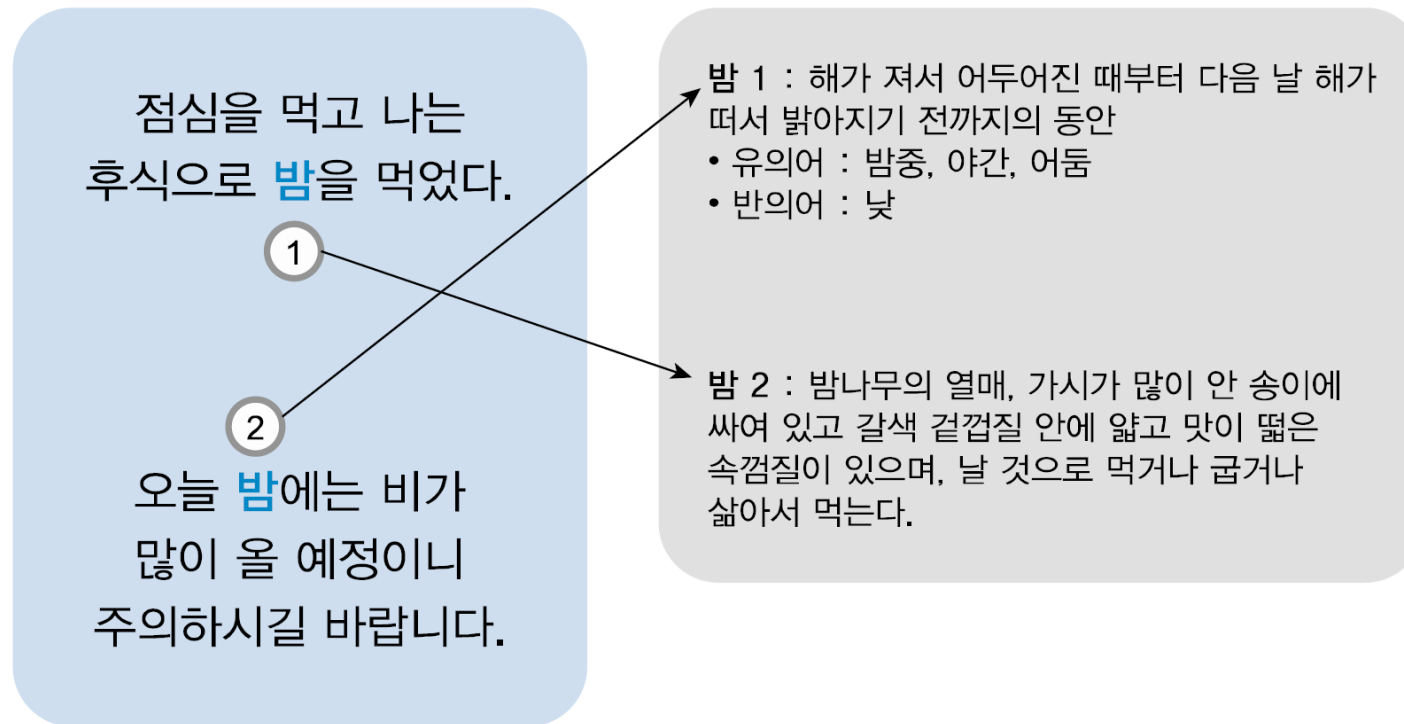


그림 7-1

단어 의미 중의성 해소

단어 의미 중의성 해소 기법

7.2 지식기반 방법

- 문장에 등장한 단어들을 사전에 정의된 어휘 지식을 활용하여 예측하는 방법
 - 예) WordNet, ConceptNet, FreeBase, BabelNet 등
- 사전 정의 기반 방법
 - 사전에 정의된 문장의 단어들을 기반으로 의미 추론
- 그래프 기반 방법
 - 의미 관계를 가지는 사전들의 관계성을 보고 의미 추론

7.2 지식기반 방법

■ 사전 정의 기반 방법 - Lesk 알고리즘

- 중의성 단어의 사전 뜻풀이에 쓰인 단어들과 중의성 단어의 주변 문맥에 나타난 단어의 사전 뜻풀이에 쓰인 단어들 사이에 중복되는 단어가 가장 많은 의미를 중의성 단어의 의미로 선택

그 **사람**은 **수술**을 통해 불편한 **다리**를 고쳤다.

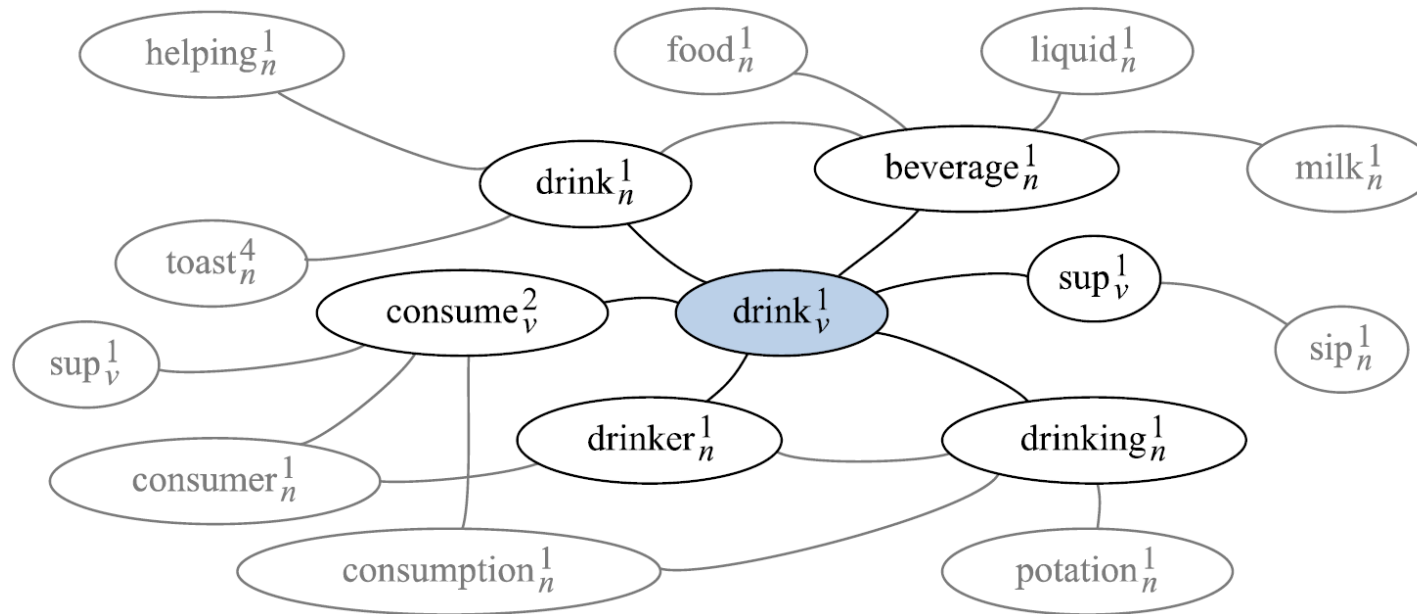
단어		사전 뜻풀이에 쓰인 단어
함께 나타난 단어	사람	생각, 언어, 만들다, 쓰다, 사회, 살다, 동물 , ...
	수술	피부, 점막, 조직, 기계, 병, 고치다, ...
		...
중의성 단어	다리 01	사람, 동물 , 몸통, 신체, ...
	다리 02	물, 건너다, 시설물, ...

7.2 지식기반 방법

- 사전 정의 기반 방법 - Lesk 알고리즘의 한계점
 - 단어 간의 정확한 일치가 기반
 - 사전 정의의 굉장히 의존적
- 실습: Lesk 알고리즘
 - https://github.com/nlpai-lab/nlp-bible-code/tree/master/07%EC%9E%A5_%EC%9D%98%EB%AF%B8%EB%B6%84%EC%84%9D

7.2 지식기반 방법

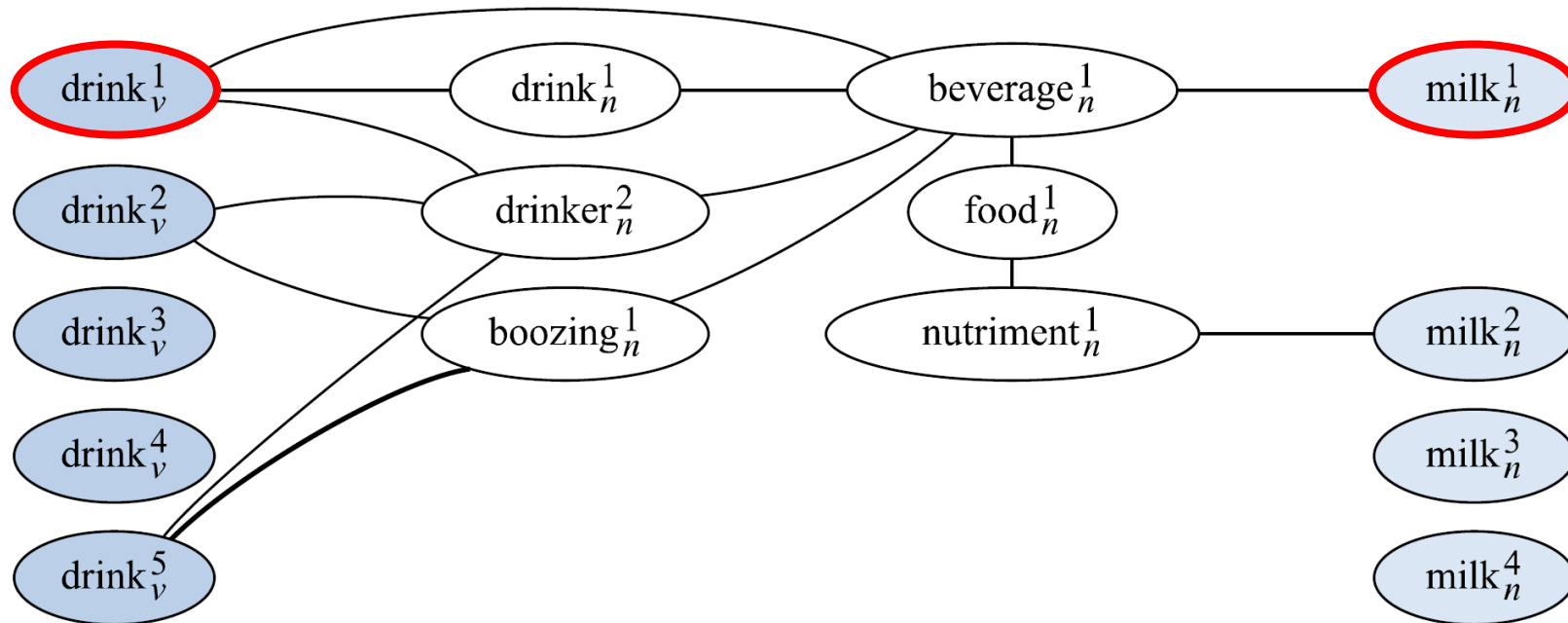
- 그래프 기반 방법 - 단순 그래프 기반 방법
 - 예) "She drank($l_1 = drink$) some milk($l_2 = milk$)"
 - $l_1 = drink: drink(v, 1), drink(v, 2) \dots l_2 = milk: milk(n, 1), milk(n, 2) \dots$
 - 중의성 단어(lemma)인 drink와 milk의 synset들을 WordNet에서 추출



7.2 지식기반 방법

■ 그래프 기반 방법 - 단순 그래프 기반 방법

- DFS(Depth First Search), BFS(Breath First Search) 알고리즘을 이용하여 검색되는 Edge를 아래와 같이 추출하여 Subgraph 생성
- 의미 간 연결성 측정을 통해 가장 많이 연결된 의미를 선택
- 같은 점수는 의미 순서가 높은 걸로 선택



7.2 지도학습 기반 방법

- 각종 기계 학습 알고리즘을 통해 단어 의미를 분석
- 기계학습 분류기 모델은 사용자가 정의한 규칙에 맞춰 선택된 자질에 따라 성능을 높여옴
 - 예) 공기어, 의존 관계 정보, 형태소 정보 등
- 성능을 높이기 위해서는 대규모의 의미 태깅된 말뭉치가 필요
- 학습한 특정 중의성 단어에 대해서만 해결 가능

7.2 지도학습 기반 방법

▪ Naïve Bayes Classifiers

- Naïve Bayes Classifier를 이용한 단어 중의성 해소 방법

$$\begin{aligned}c &= \arg \max_{c_k} P(c_k | \vec{x}) \\&= \arg \max_{c_k} \frac{P(\vec{x} | c_k)}{P(\vec{x})} P(c_k) \\&= \arg \max_{c_k} [\log P(\vec{x} | c_k) + \log P(c_k)] \\&= \arg \max_{c_k} \left[\sum_{v_j \in x} \log P(v_j | c_k) + \log P(c_k) \right]\end{aligned}$$

7.2 지도학습 기반 방법

▪ k-Nearest Neighbor Classifiers

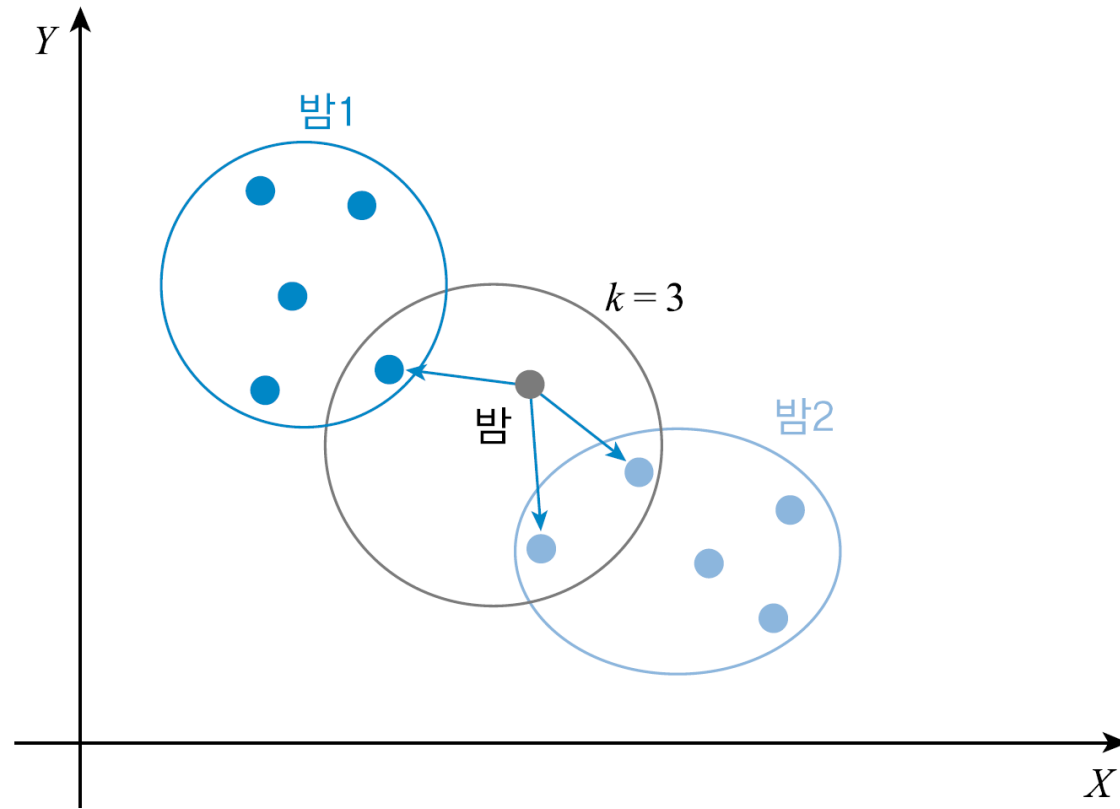
- 벡터 공간에 표현된 자질들을 정해진 k값에 따라서 가장 많이 묶이는 자질들의 의미클래스로 선택
- 유클리드 거리(Euclidean Distance)와 코사인 유사도(Cosine Similarity)로 벡터 공간에서의 자질을 묶음

$$Euclidean\ Distance = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

$$Cosine\ similarity = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

7.2 지도학습 기반 방법

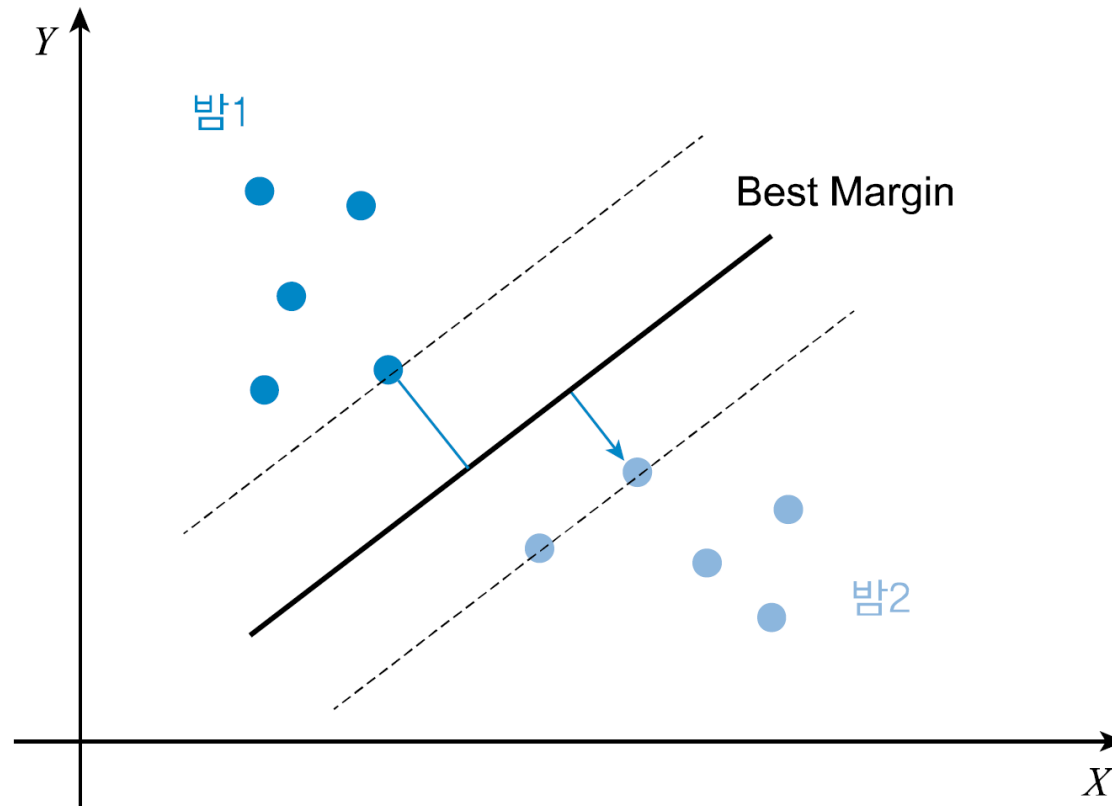
- k-Nearest Neighbor Classifiers
 - kNN을 이용한 '밤' 중의성 해소 방법



7.2 지도학습 기반 방법

▪ Support Vector Machine

- 벡터 공간에 표현된 자질로부터 의미 클래스를 분류하기 위해 의미 클래스 간에 가장 넓은 거리를 사용하는 방향으로 선을 그어 의미를 분류



의미역 분석

7.3 의미역의 구분

■ 의미역은 두가지로 구분

- 서술어의 의미를 구성하는데 필수적으로 요구되는 **필수적 의미역**
 - 행동주, 도구, 피동주, 경험자, 수혜자, 출처, 도달점 등으로 분류
- 서술어의 의미를 보충하는 **수의적 의미역**
 - 장소/위치, 이유, 목적, 경로, 시간, 방법 등으로 분류
- 통사적인 분석에서 의미역 기준에 따라 하나의 명사구는 하나의 의미역만을 가짐

7.3 의미역 설명

▪ 필수적 의미역

- 행동주
 - 어떤 행위를 하는 주체로 대개 감정이나 의지를 가지는 사람이나 동물
 - 고의성이나 의도성을 가짐
- 도구
 - 행위, 이동의 의미를 표현하는 동사의 수단
- 피동주/수동자
 - 동사가 행위를 표현하는 경우며, 행위에 영향을 받거나 상태 변화를 겪음
- 경험자
 - 인지, 지각, 감정을 나타내는 용언의 경우 그 현상의 경험 주체가 되는 논항에 의해 내적인 상태에 영향을 받는 것
 - 행동주와 달리 의도성을 나타내는 부사(구)와 결합할 수 없음
- 수혜자
 - 술어가 나타내는 사건에 의해 어떤 대상이 그로부터 이동하게 된 원래의 장소
 - 행동의 동기나 이유가 출처가 되기도 함
- 도달점/목표
 - 행위가 목적하는 곳으로 출처와 대칭적인 개념
 - 공간적 목적지뿐만 아니라 심정적, 주관적 목적지까지 포함

7.3 필수적 의미역 예시

표 7-1 필수적 의미역

의미역	예시
행동주(agent)	철수(agent)가 돈을 낸다.
도구(instrument)	철수(agent)가 망치(instrument)로 못을 박는다.
피동주/수동자(patient)	철수가 민희(patient)를 사랑했다.
경험자(experiencer)	영희(experiencer)가 사랑에 빠졌다.
수혜자(benefactive)	내가 철수(benefactive)를 위해 밥을 사줬다.
출처/근원(source)	나는 식당(source)에서 밥을 주문했다.
도달점/목표(goal)	나는 책을 서랍(goal) 안에 보관했다.

7.3 지도학습 기반 의미역 분석

- 단어 의미 중의성 해소의 지도학습기반 방법과 마찬가지로 기계학습을 사용
- 형태소, 구문정보와 같은 문법적 자질을 사용할 때
 - '타다' 라는 서술어가 같은 행동주와 피동주/수동자를 갖는다면 유사한 의미인 '승차하다' 도 같은 의미역을 갖는다고 유추하여 의미역을 파악하는데 도움을 줌
 - 반대로 '타다' 의 동형이의어가 존재할 경우 '탈 것' 을 의미하는 것이 아니라 '불씨가 번지거나 불꽃이 일어난다.'로 해석될 수 있기 때문에 단어 그대로가 아닌 단어가 갖는 의미도 고려해야함
- 개체명과 같은 의미적 자질을 사용할 때
 - '눈'이라는 단어가 '얼음 결정체'인지 '감각기관' 인지도 파악해서 의미역을 파악하는데 도움을 줌
- 의미역 성능을 높이기 위해 이러한 문법적, 의미적 자질을 사용하면 높은 성능을 보이지만, 자질을 추출하기위해 대량의 학습데이터가 학습된 분류기를 따로 구축해야한다는 단점이 존재

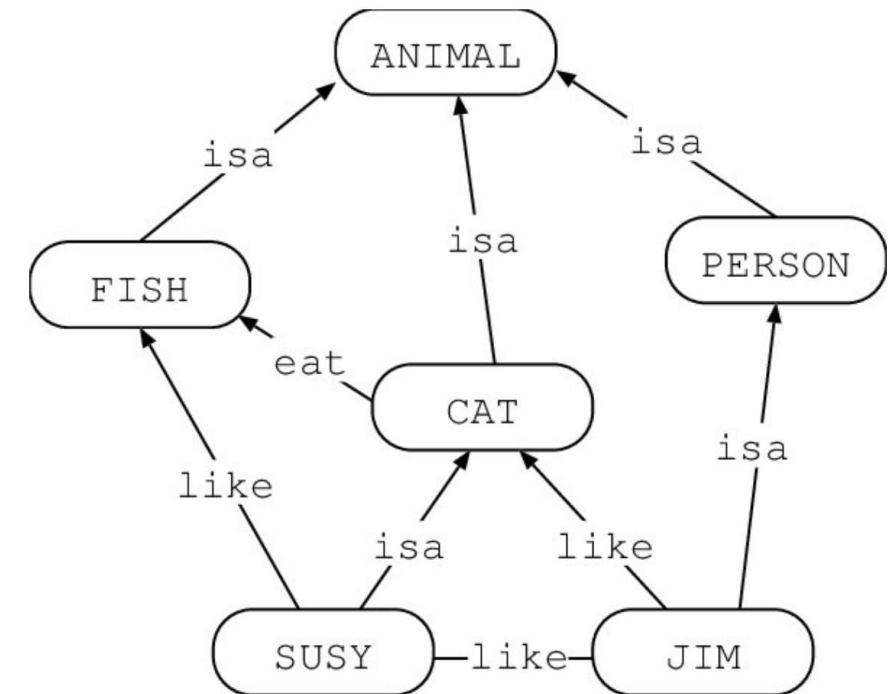
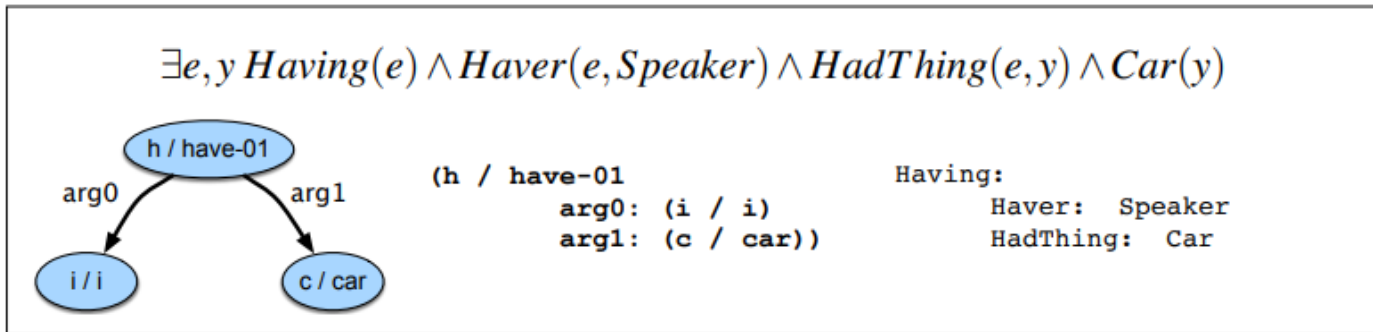
의미 표현

7.4 의미 표현

- 화자들은 의미 구조에 대한 지식으로 모국어의 문장을 의미와 명제 내용에 따라 논항 구조를 파악하고, 문장 사이의 모순관계, 함의관계, 중의성 등의 관계를 정확하게 파악
 - “나는 사과를 먹는다.”와 “나는 사과를 했다.”라는 문장을 구분할 때 ‘사과’가 ‘먹는다.’, ‘했다.’와의 관계를 통해 다른 의미로 쓰이는 것을 정확하게 파악
 - “오늘 날씨가 너무 춥다.”라는 말을 했을 때 상황에 따라 창문을 닫거나 선생님을 위로하듯이 은유적, 관용적 구성을 사용하고, 지시, 가리킴말, 맥량, 화행, 및 전제, 함축을 상황에 맞춰서 유연하게 사용
- 문장의 의미표현은 이러한 화자들의 다양한 언어적 표현을 파악하는 단계
- 의미 분석 연구는 개체명, 의미역, 상호참조, 담화 연결사, 시간 및 공간 정보 등 언어 표현이 가진 여러 의미론적 측면 중 일부에 초점을 맞추어 진행되어 옴

7.4 의미표현

- FOL (First-Order Logic)
- Conceptual Dependency Diagram
- Semantic Network
- Frame-based Representation
- Abstract Meaning Representation (AMR)



7.4 의미표현

■ AMR 예시

- '사건(생각-01)', 주체(:ARG0)'는 '소년'이 되고, '대상(:ARG1)'은 '좋-02'의 사건이 됨
- "소년이 ~ 좋아한다고 생각한다."
- '좋-02' 사건의 '주체(:ARG0)'는 '소녀'가 되고 '대상(:ARG1)'은 앞서 상위에 언급되었던 '소년' 이 됨
- "소년은 소녀가 자신을 좋아한다고 생각한다."

(생 / 생각-01
: ARG0 (소 / 소년)
: ARG1 (좋 / 좋-02
: ARG0 (소2 / 소녀)
: ARG1 소))

“소년은 소녀가 자신을 좋아한다고 생각한다.” AMR 예시

