

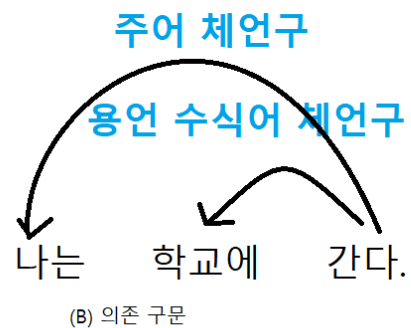
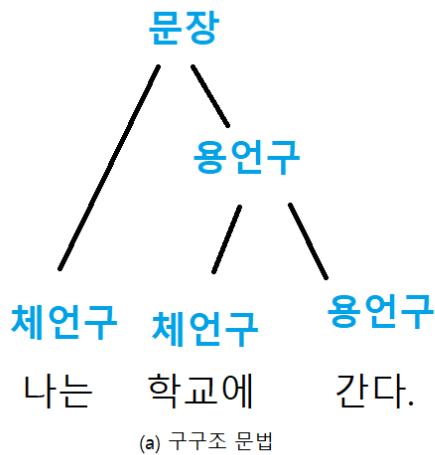
Ch 6 구문 분석

#1 구문 분석

- 자연어 문장에서 구성 요소들의 문법적 구조를 분석하는 기술
- 문법적 구조 정보를 자동으로 추출함으로써, 기계 번역, 정보 검색 전문가 시스템 등의 자연어 처리 기술에서 문장 의미의 분석을 돕는 세부 기술로 활용 가능
- 규칙 기반 구문 분석: 인간이 직접 정의한 문법 규칙을 적용하여 구문 분석을 수행
- 통계 기반 구문 분석: 확률적인 문법 규칙을 통계적으로 계산하여 구문 분석을 수행
- 딥러닝 기반 구문 분석: 인간이 구축한 구문 분석 데이터 셋으로부터 딥러닝 모델을 학습하여 구문 분석을 수행

#1.2 구문 문법

- 문법적 구성 요소 들로부터 문장을 생성하고, 반대로 문장을 구성 요소들로 분석하는 문법
- 구구조 문법과 의존 문법이 대표적임



#1.2.1 구구조 문법

- 구성소 관계에 기반하여 문장 구조를 분석
- 단어들이 모여 절을 구성하며 이러한 절과 단어들의 계층적 관계에 따라 문장을 분석
- 문장 구성 요소의 구조가 비교적 고정적인 언어에 적합(영어)

#1.2.1-1 규칙 기반 구구조 구문 분석

- 인간이 가지고 있는 언어학적 지식을 컴퓨터가 이해할 수 있는 형태의 문법 규칙으로 미리 정의한 뒤 이를 자연어 문장에 정의해 구문 분석을 수행
- 구구조 문법을 규칙으로 활용
- 대표적인 알고리즘으로 CYK알고리즘이 있다.

문법 규칙 예시

문법 규칙
$S(\text{문장}) \rightarrow NP(\text{명사구}) + VP(\text{동사구})$
$NP(\text{명사구}) \rightarrow NN(\text{명사}) + XSN(\text{접사}) + JK(\text{조사}) / NN(\text{명사}) + JK(\text{조사})$
$VP \rightarrow NP(\text{명사구}) + VP(\text{동사구}) / VV(\text{동사의 어간}) + EP(\text{선어말 어미}) + EF(\text{어말 어미})$

#1.2.1-2 통계 기반 구구조 구문 분석

- 통계적으로 확률적 구구조 문법을 계산하여 이를 바탕으로 구문 분석을 수행
- 각 규칙에 대한 조건부 확률이 정의된다.

#1.2.1-3 딥러닝 기반 구구조 구문 분석

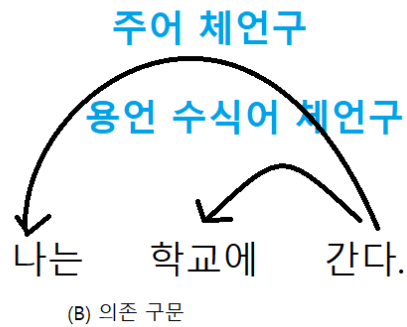
- 인간이 구축한 구구조 구문 분석 데이터셋으로부터 딥러닝 모델을 학습하여 구문 분석을 수행
- 자연어 문장의 표층 정보와 의미적 정보를 입력으로 한다.
- 가장 대표적인 것으로 전이 기반 파싱이 있다.

#1.2.2 의존 문법

- 의존 관계에 기반하여 문장 구조를 분석
- 문장을 구성하는 단어들 간의 계층적인 의존 관계에 따라 문장을 분석
- 각 단어를 노드(node)로 하고, 단어간의 의존 관계를 에지(edge)로 나타내는 트리 구조

#1.2.2-1 규칙 및 통계 기반 의존 구문 분석

- 지배소(Head): 의존 관계 표현에서 절의 중심이 되는 단어
- 의존소(Modifier): 절 안에서 지배소에 의존하는 단어



'간다'는 절의 중심이 되는 어절이므로 전체 문장에서 지배소이다.

'나는', '학교에'는 지배소 '간다'의 의존소이다.

- CG(Constraint Grammar): 이런 규칙 기반 의존 구문 분석을 위한 문법.

#1.3 구문 중의성

- 하나의 문장에서 가능한 구문 구조가 여러 개일 때 발생
- 해소를 위해선 의미, 문맥 등의 추가적인 정보가 필요