05.09 CL 08

#17

센텐스는 명사구와 동사구로 이루어짐. S-> NP VP, NP-> DET N, VP -> V NP. 이런식으로 specify할 수 있음.

PSG의 문제점 : 영어 문장의 구조를 이런식으로 스페시파이한다면 over generation할 수 있다.

예를 들어, The dog died the cat. 비문법적인 문장인데 우리가 만든 문법에 따르면 맞는 문장이 된다. 이런 문제를 해결하기 위해서는 카테고리를 pos를 세분화한다. 동사는 타동사, 자동사.(v\_iv)이런식으로.

#19

nodes. 그 노드를 constitudent라고 부르기도 한다. 문장 구조를 수용도처럼 표현하는 방법중 또 다른 것은 bracketing(괄호치기).

구성구조문법(PSG)의 표현 방법이 rewriting rule이라고 했는데, 그것의 장점 중 하나가 자연언어에 나타나는 무한한 문장들을 만들어 낼 수 있는 능력, 무한히 긴 문장들을 유한개의 규칙들로 설명할 수 있다는 것이다. 언어학에서 던지는 질문은 어떻게 한정된 지적 능력과 지능으로 무한히 많은 문장을 만들어 낼 수 있겠는가? 하는 것이다. 아이들이 어떻게 언어를 습득하는 가? 한가지 이론은 아이들은 엄마 아빠가 주변에서 하는 말을 듣고 copy하는 것이다. 라는 가설이 있는데 이는 어른들이 말을 안 할법한 말을 아이들이 한다는 점도 있다는 것이다.

* Recursivity (회기성)

다시 출현한다는 것이다.

NP->NP PP 하나의 규칙인데

PP -> P NP 이런 식으로 반복될 수 있음. 이는 자연언어의 특징이고, 그 특징을 파악해주는 포멀 랭귀지의 장점이기도 한다. 그런데 생각해보면 자연언어에는 이런 회기성이라는 특징이 있다. 전치사구 뿐만 아니라 관계대명사절도 가능. NP-> NP S.

#20

문장을 보다보면 대부분 단어들이 자기 주변 근처에 있는 것들과 관련이 있음. I do not love the subject. 에서 not이 영향을 미치는 단어는 love이다. 그런데 떨어져 있는 경우도 있음.

The woman, who read the bookd, succeeded. 주어와 동사가 떨어져 있음. 이런 경우를 non local dependency라고 한다.

S-> NP VP

NP -> DET N

VP -> V NP

The man liked the music. 이 문장을 구조 분석하려면 첫번째. 토크나이즈, pos 태깅을 한다. 그 이후에 문장의 구조를 알아낸다. 이런 것을 Parsing 이라고 한다. 숨어있는 구조를 다시 만들어 내는 것을 Parsing이라고 한다. shallow parsing 이라고 하는 것, 겉핥기식 구조 분석이라고 한다.

\*\*기말 과제 : 영어의 명사구를 분석하는 그래머를 쓰고, 그 그래머를 활용해서 영어 문장에서 NP를 추출할 수 있는 코드

영어 문법에서 명사구라는게 더 많은 형태가 있을 것임. 영어에 존재하는 모든 명사구를 커버하진 못해도 된다. The man whom I met in the building that was located in northern Seoul. 이런 명사구 안에도 여런 명사구가 있음. 가장 작은 단위의 명사구가 또 다른 명사구 안에 포함되고 그런걸 얘기하는게 아니라 큰 명사구이긴 하지만 최소의 단위를 알려는 것. 이걸 하려면 샘플 사전이 있어야 함. 내가 만든 문법과 파이썬 코드로 다음과 같은 문장에서 이런 명사구를 추출할 수 있다는게 필요할 것임.

그래서 숙제.

1. fileread

2. tokenize

3. pos\_tagging

4. grammer, 명사구 NP -> ~~ 규칙들을 만들기

5. parsing if pos = DET : 이런식으로 ,,, 안 배운 테크닉을 활용해야 할 수도 있음. 구글 검색하기.

이런 파이썬 코드와 그래머와 예문들을 테스트 해봤더니 어떻게 됐더라 이런걸 하기.

예문 길이는 자유다. 기말고사 시작 전까지 제출. 이메일로 제출.

명사구의 구조는 정말 많음.

파이썬 파일 하나, 그래머를 정리하고 설명한 워드 하나, 돌린 결과의 스크린샷 정도로 내기.

#21

모든 유럽어에 있는 현상인데 영어로 하면 which book should hans buy? wh는 앞으로 나오는 것. 어떤 학자들은 이동했다고 함.

#23

Dependency Grammer 의존 문법, 독일과 프랑스에서 나온 이론.

여태 배운 것과 완전 새로운 얘기.

문장의 구조를 단어와 단어 간의 의존관계가 핵심이다. 문장에서 항상 원자들이 결합하듯이 핵심이 되는 원자가 있는데 그 핵심은 동사. 모든 단어들은 동사와 어떻게 해서든 관련이 되어 있다고 가정.

문장은 동사가 핵심이다. Hans saw a man at the table. 이 문장의 핵심은 saw이다. 그리고 여기에 의존하고 있는 애들은 주어와 목적어인 Hans, man 이다. 문장을 동사 중심으로 분석, 우리나라 기업 뿐만 아니라 외국 기업에서 많이 사용한다. 그 이유는 PSG가 굉장히 정교하게 문법을 기술하는데 그거를 실제 산업분야에서 적용하려면 분석이 안 되는게 너무 많음. 보다라는 동사는 문장을 만들기 위해 몇 개의 아규먼트를 취할까? 두개 : 주어, 목적어. 이렇게 문장 핵심을 이루는 동사들 마다 결합가가 있다. 동사들이 결합가가 무엇인지를 알아야 한다.

#24

See라는 동사를 중심으로 주어와 목적어를 취하는데 얘네를 arguments라고 한다. 논리학에서 온 것.

1. semantic roles 의미적 역할. 아규먼트가 그 문장 내에서 어떤 역할을 하느냐. 역할이라는 건데 그 역할은 일반적으로 언어학에서 아규먼트가 어떤 의미적 역할을 하냐고 묻는다면 그 역할을 말하기 위핸 set가 있음. Jane hit john. jane은 행위자 john은 theme(행위의 대상)이라고 한다.

2. grammatical relations. 문법적 기능을 의미. 혹은 grammatical function이라고 한다.

그러면 영어에서 주어의 semantic role은 뭘까?

agent, theme, recipient, instrument, goal, instrument, location. 여섯 개로 영어 문장에서 명사가 하는 역할을 기술할 것임. 영어 문장의 주어들이 갖는 의미가 뭘까? 상황에 따라 다름. 수동태도 있으니까. 수동태일 때는 주어는 theme 이다.

The train arrived at time. 이 문장에서 train은 agent가 아니라 theme 이다.

I got a present from Jane. 이 문장에서 I 는 recipient이다. 그래서 어떤 언어학자들은 영어나 독일어나 서양언어에 나오는 수동현상들에서 수동이 뭐냐, 왜 필요하냐?

어떤 정보를 추측하려고 할 때 누가 무엇을 했고, 영향을 주었다는 정보를 알려면 문장 구조 분석 이후 어떤 행위를 나타내는 문장인지 행위의 주체가 누구인지 대상이 누구인지를 알 필요가 있음. 이를 알기 위해서 필요한 정보는 주어, 목적어, 간접목적어 등.. 뿐만 아니라 모든 주어가 경우에 따라서 주어가 agent, recipient, theme 인 경우가 있음. 단순히 주어와 목적어만을 안다고 해서 되는게 아니라.. 의미적 역할까지 파악해야 함.

A contain B 라는 문장. contain이라는 동사는 argument를 일반적으로 2개 취한다.

The box contained a coke and two sprites. 기존에는 pos 태깅만 했다면 이 문장에는 명사 세개 동사 하나밖에 없음. contain 이라는 의미가 있고 무엇이 어디에 담겨있는지를 알려면 더 분석을 해야함.The box는 location, a coke and two sprites는 theme. 이런 정보가 필요함

A contain B

A : Subj, location

B : Obj, theme << <이런 정보를 알아야 함.

문장의 의존 구조를 분석하면서 의미적인 이야기까지 왔음. 우리가 만약에 의미 분석을 하면서 추가적으로 given information에다가 새로운 정보를 더할 뿐만 아니라 추론도 할 수 있다.

#25

Grammatical Relations in English. 영어는 어순이 고정되어 있는 편이다.

영어에서는 어순으로 그 펑션을 추측할 수 있다.

pos 태깅 후 동사 찾기. 사전을 보면 사전에는 contain은 두개의 아규먼트를 취한다는 정보가 있음. 동사를 기준으로 왼쪽에 NP가 있으면 주어, 오른쪽에 있으면 목적어. 이렇게 하면 대략적 분석이 될 것임.

#26

문장을 의존 구조의 관점에서 바라보면 지금 마치 문장은 동사가 있고 그 동사에 의존하는 모든 아규먼트들만 있는 것처럼 느껴졌는데 꼭 그렇지는 않음.

I had lunch with my friend. 에서 with my friend는 문장의 필수 성분은 아님. 부가적 설명. 그런 애들을 Adjuncts. 라고 한다. 문장은 그래서 크게 argument와 adjuncts로 볼 수 있다.

yesterday, I went to the restaurants. yesterday – adjuncts.

#27

She put the book on the table.

on the table 이 argument냐 adjuncts냐? 교수 개인적으로 argument가 맞다고 봄. 의미적으로 put의 행위의 완결점을 알려주기 때문에. 사실은 아규먼트와 어정트의 클리어 컷은 어렵다. 애매함.