

Abstract Meaning Representation

(AMR)

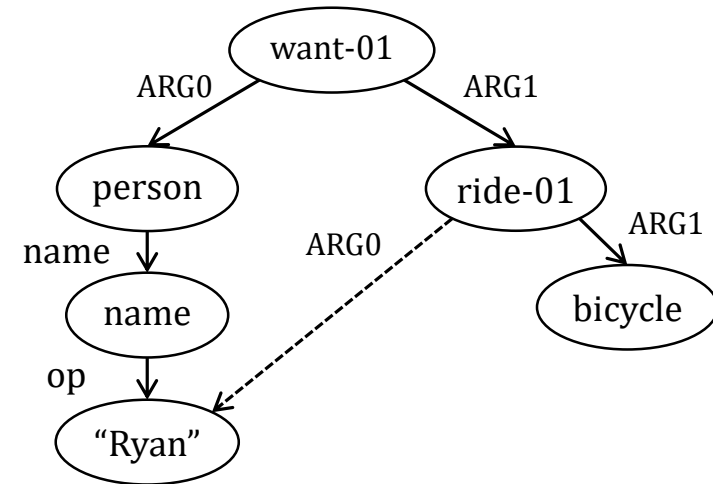
1. 개요

- 문장의 의미 구조를 그래프로 표현하기 위해 제안된 개념

Abstract Meaning Representation (AMR)은 Banarescu, Laura, et al .*가 제시한 개념

전체 문장의 의미 구조를 나타내는 sembank (semantic treebank)를 생성하는 것을 목표로 함

Ryan wants to ride bicycle.



Origin Text

AMR Graph

2. AMR

- Limitations of AMR

- 보편 양화사(universal quantifier)를 고려하지 못함

 - “all”, “for all”, “given any”과 같은 보편 양화사는 주요 개념(Head concept)을 변경할 수 있음

- 시제 및 수 표현에 따른 접사 및 모음의 변화를 포함하지 않음

 - 그래프의 주석 (annotation) 표기 속도를 증가시킬 수 있음

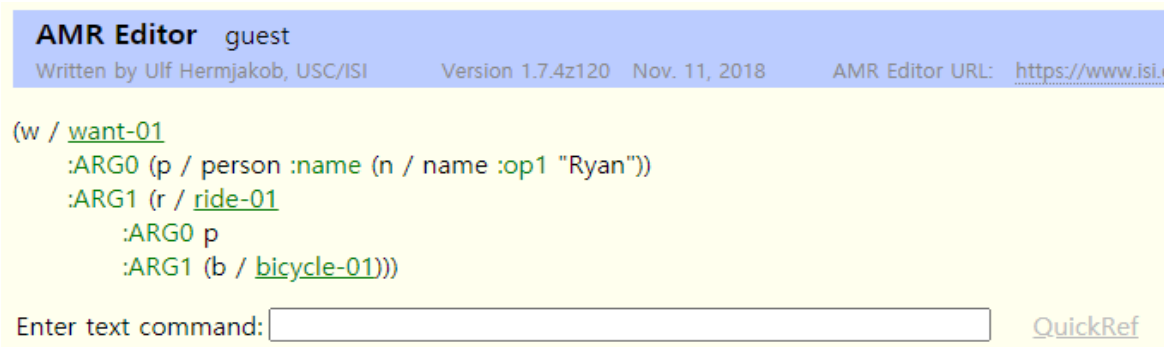
 - 예시 : *-ed, -ing, -s*

- 실제로 일어난 일과 미래, 가정, 소망을 구별하지 못함

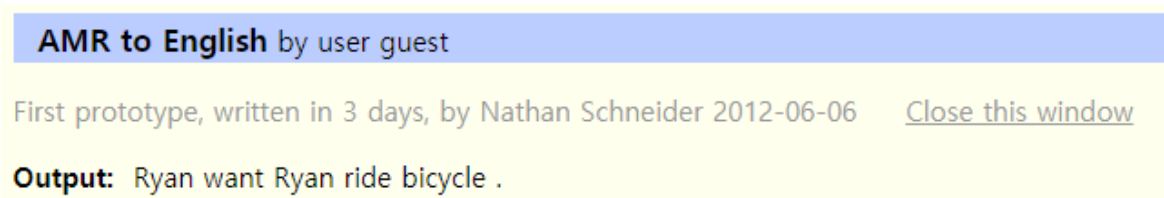
2. AMR

- AMR Editor*

Banarescu, Laura, et al.은 AMR Editor 서비스를 배포하였으며,
AMR 그래프 생성 및 생성된 그래프를 텍스트로 변환하는 기능을 제공함



The screenshot shows the AMR Editor interface. At the top, it says "AMR Editor guest" and "Written by Ulf Hermjakob, USC/ISI". Below that, it shows the AMR graph for the sentence "Ryan want Ryan ride bicycle". The graph is represented as a nested list of triples: (w / want-01, :ARG0 (p / person, :name (n / name, :op1 "Ryan")), :ARG1 (r / ride-01, :ARG0 p, :ARG1 (b / bicycle-01))). Below the graph is a text input field with the placeholder "Enter text command:" and a "QuickRef" link.



The screenshot shows the "AMR to English" output. It says "AMR to English by user guest" and "First prototype, written in 3 days, by Nathan Schneider 2012-06-06". Below that, it shows the output: "Output: Ryan want Ryan ride bicycle .".

AMR Annotation Release 3.0 (20/01/15)

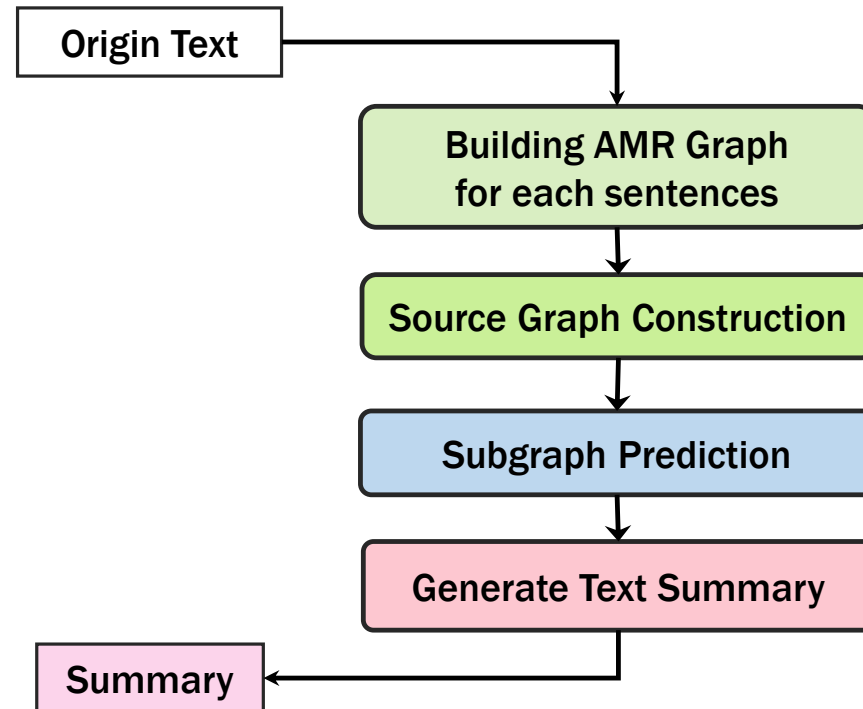
소설, 웹 텍스트, 웹 포럼 등으로부터 수집한
59,255개의 영어 문장 데이터를 기반으로 구축됨

* AMR Editor : amr.isi.edu/editor.html

3. 관련 연구 – 요약

- Summarization using AMR (English)*

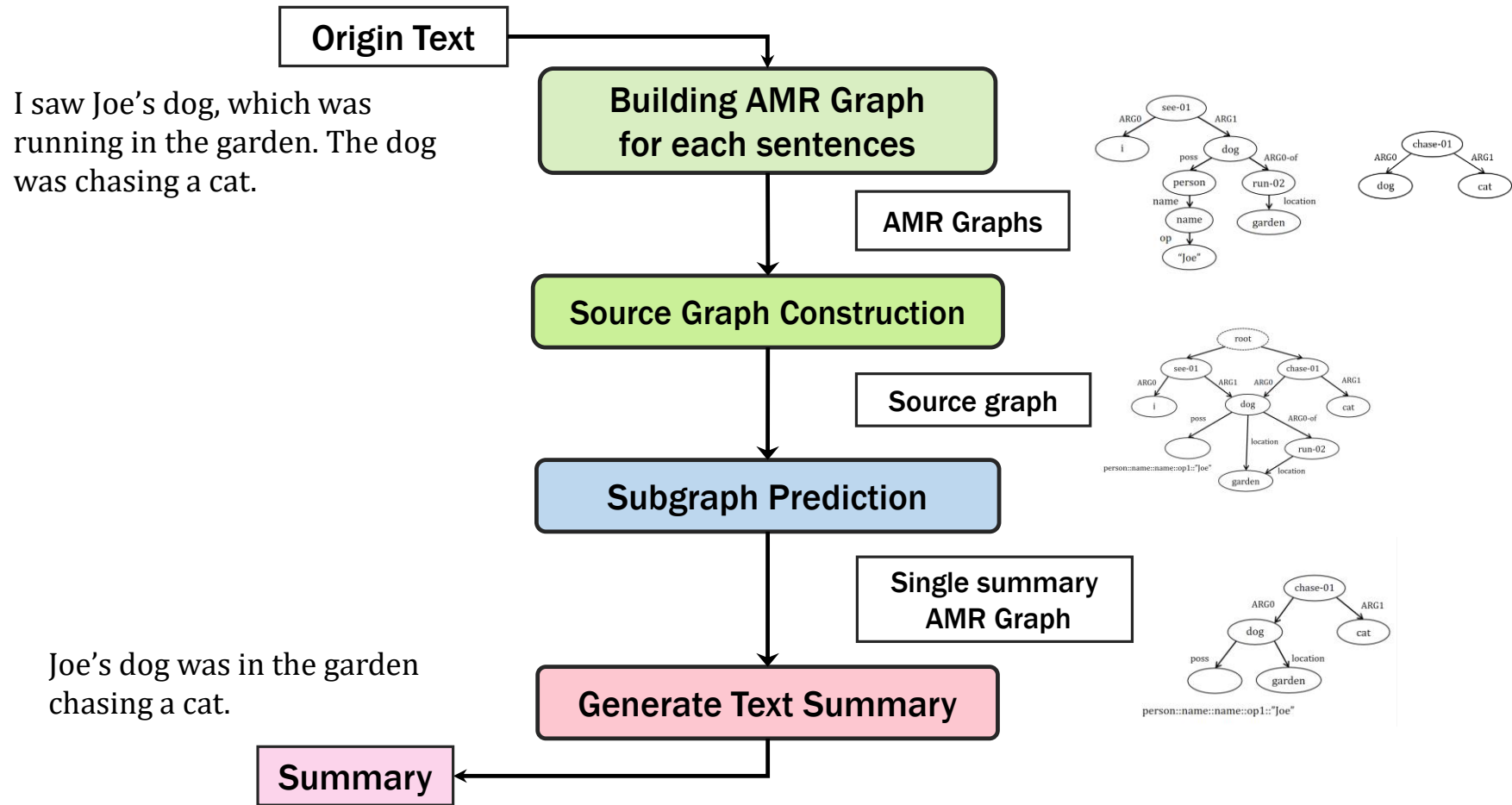
AMR 그래프를 사용하여 추상 요약 (Abstractive Summarization) 수행



* Liu, Fei, et al. "Toward abstractive summarization using semantic representations." Proceedings of the 2015 conference of the north american chapter of the association for computational linguistics: human language technologies. 2015.

3. 관련 연구 - 요약

- Framework

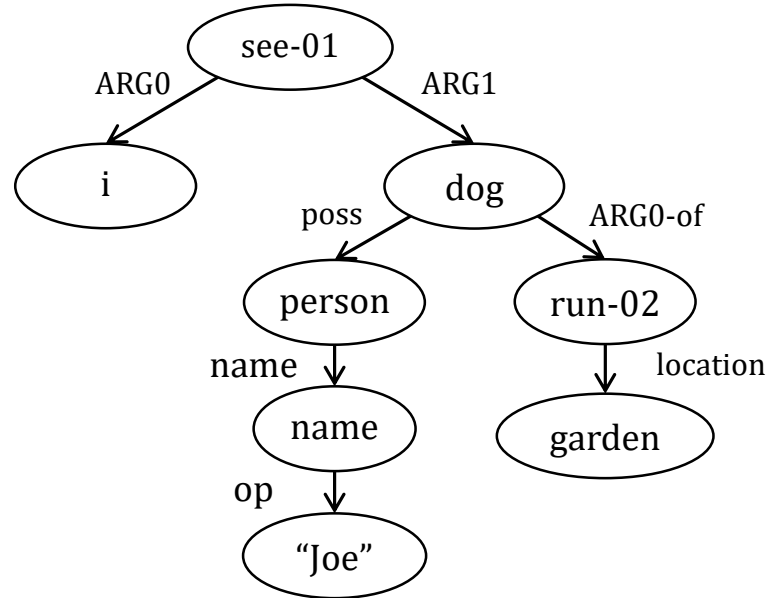


3. 관련 연구 – 요약

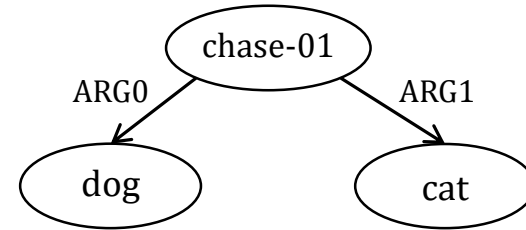
- Summarization using AMR (English)

1. 각 문장으로부터 AMR 그래프 생성

I saw Joe's dog, which was running in the garden.



The dog was chasing a cat.



3. 관련 연구 - 요약

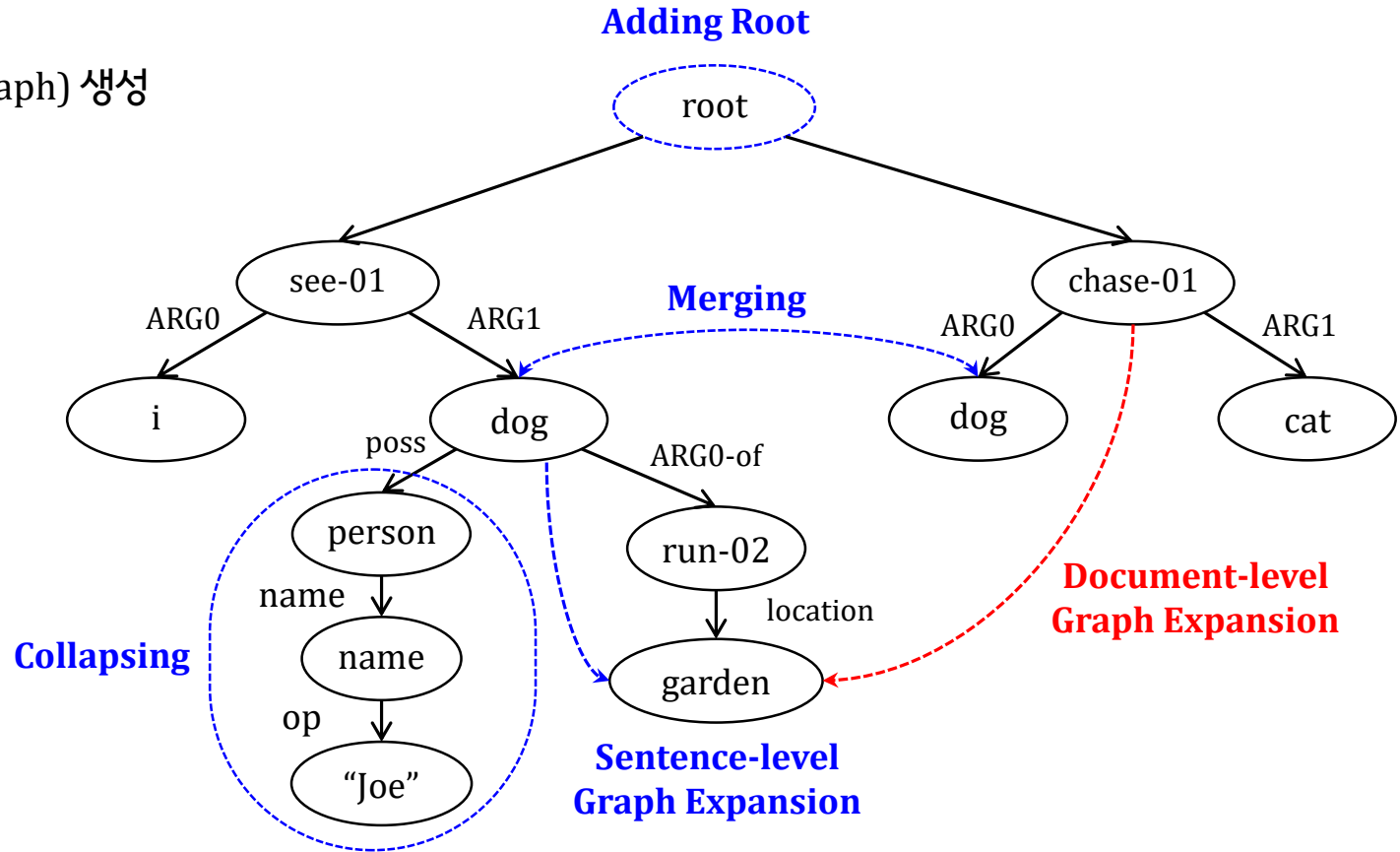
- Summarization using AMR (English)

2. 요약 그래프 (Single Summary Graph) 생성

1) 각 AMR 그래프로부터 하나의 소스 그래프(Source Graph) 생성

Collapsing

date-entity 또는 name-entity가 이끄는 부분 그래프의 구조가 평평한 경우, 해당 구조를 붕괴한 후 새로운 라벨을 부여함



3. 관련 연구 - 요약

- Summarization using AMR (English)

2. 요약 그래프 (Single Summary Graph) 생성

1) 각 AMR 그래프로부터 하나의 소스 그래프(Source Graph) 생성

Graph Expansion

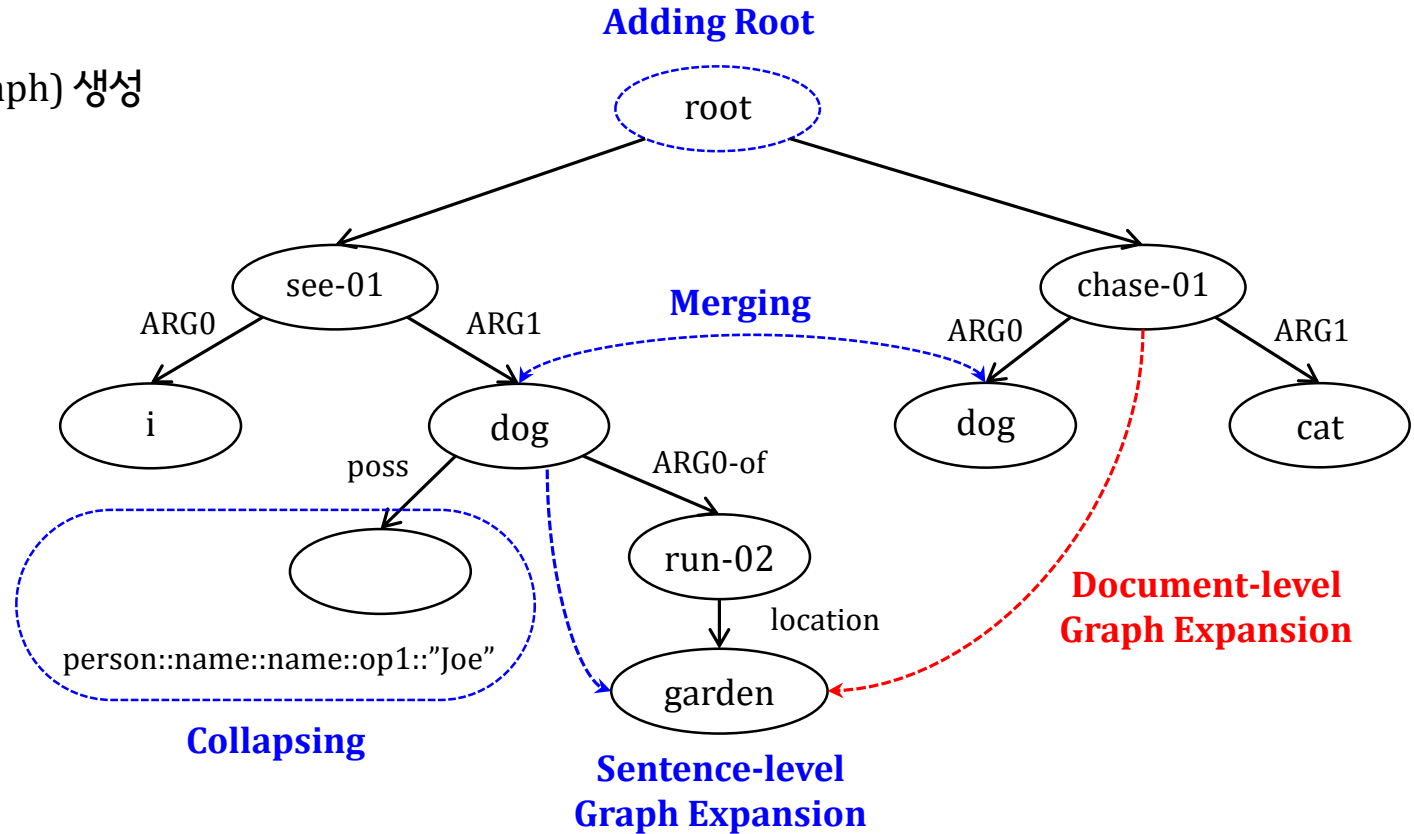
의미적으로 완전 밀집 그래프 (fully dense graph)를 만들기 위해,
노드 간의 간선을 추가함

- Sentence-level expansion

같은 문장 내의 노드 간에 가능한 간선을 모두 추가

- Document-level expansion

전체 Source graph 내의 노드 간에 가능한 간선을 모두 추가
(Unlabeled Edge가 급증하기 때문에 비효율적임)

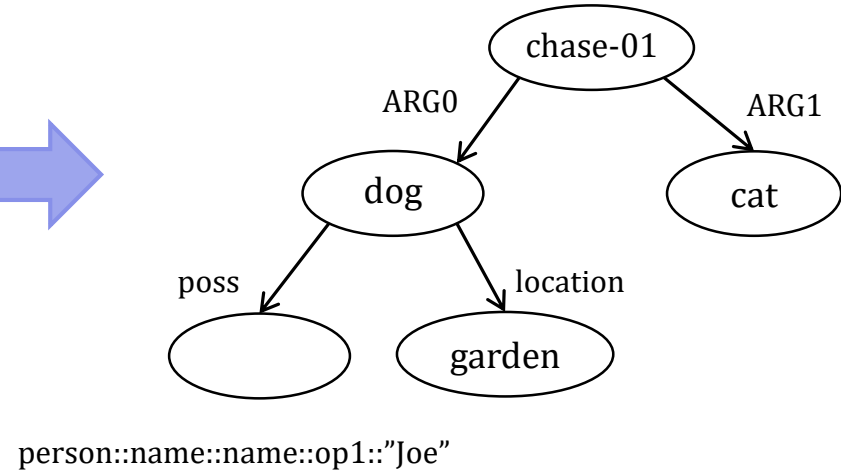
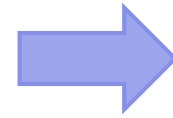
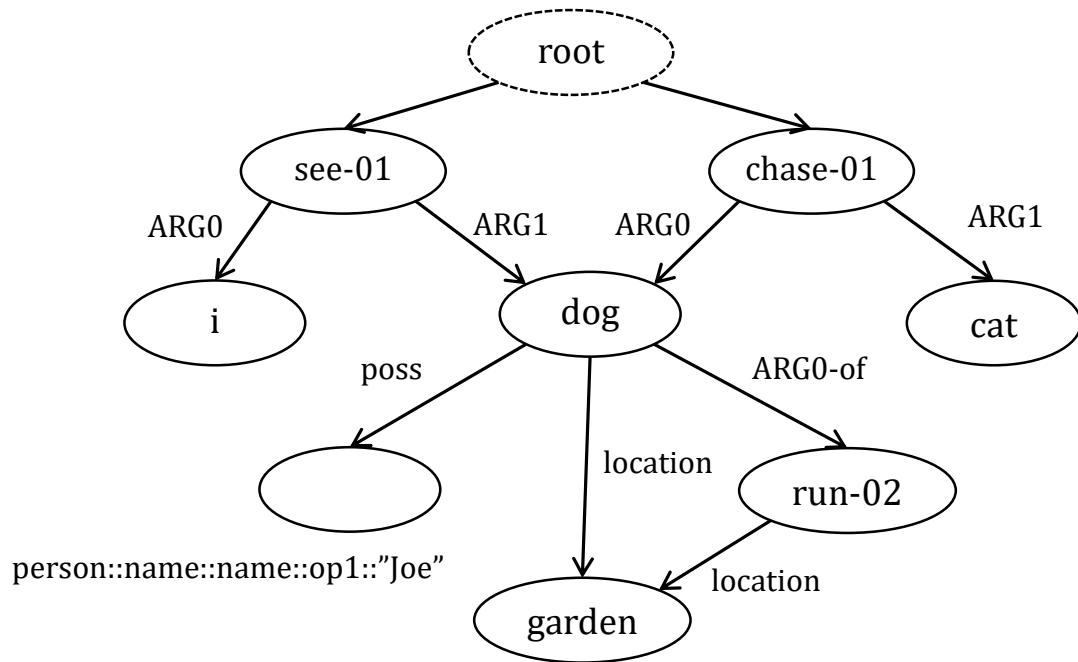


3. 관련 연구 - 요약

- Summarization using AMR (English)

2. 요약 그래프 (Single Summary Graph) 생성

2) 소스 그래프로부터 요약 그래프 예측



3. 관련 연구 – 요약

- Summarization using AMR (English)

2. 요약 그래프 (Single Summary Graph) 생성

2) 소스 그래프로부터 요약 그래프 예측

병합된 전체 그래프를 $G = (V, E)$ 라 할 때, score가 가장 높은 부분 그래프 (V', E') 를 추출

$$score(V', E'; \theta, \psi) = \sum_{i=1}^N v_i \underbrace{\theta^\top \mathbf{f}(i)}_{\text{node score}} + \sum_{(i,j) \in E} e_{i,j} \underbrace{\psi^\top \mathbf{g}(i,j)}_{\text{edge score}}$$

* θ, ψ 는 학습 모델의 계수(coefficient)이며, 선형 모델을 사용하여 학습함

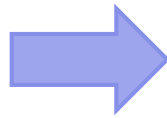
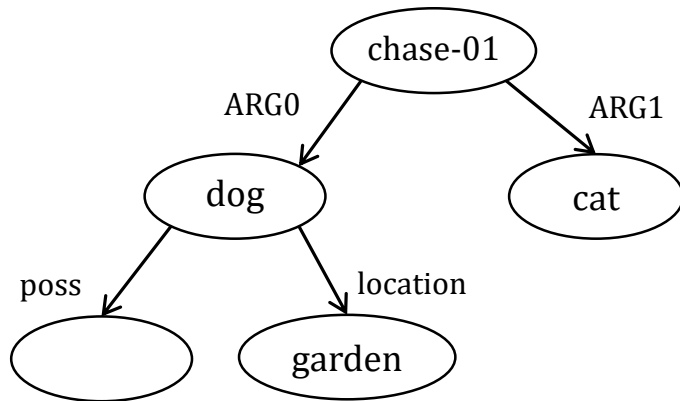
* $\mathbf{f}(v)$ 와 $\mathbf{g}(e)$ 는 각각 노드 v 와 간선 e 의 특징 표현(feature representation)을 나타냄

3. 관련 연구 – 요약

- Summarization using AMR (English)

3. 요약 그래프를 바탕으로 요약문 생성

요약문 생성 시, 노드에 대한 **특정 순서가 정해지지 않았기 때문에** 평가 방법으로 ROUGE-1을 사용함
(완전한 문장 생성이 아닌 단어 나열의 형식을 띠)



Joe's dog was in the garden chasing a cat.
(Ideal)

person::name::name::op1::"Joe"

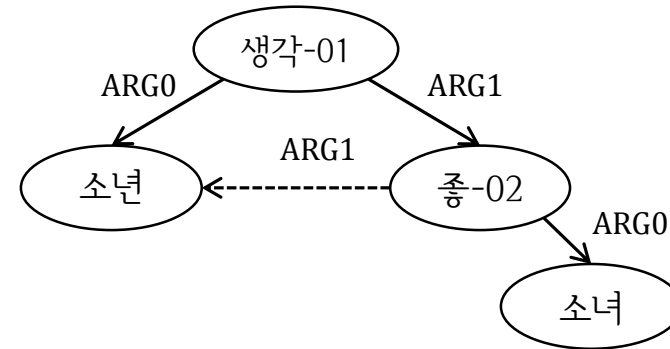
3. 관련 연구 – 한국어 AMR

- Korean AMR*

- 한국어의 특성을 반영한 한국어 AMR 가이드라인 소개

- 영어 및 중국어 AMR을 비롯한 스페인어, 브라질 포르투갈어, 베트남어 등 다양한 언어권의 연구들을 참고함

소년은 소녀가 자신을 좋아한다고 생각한다.



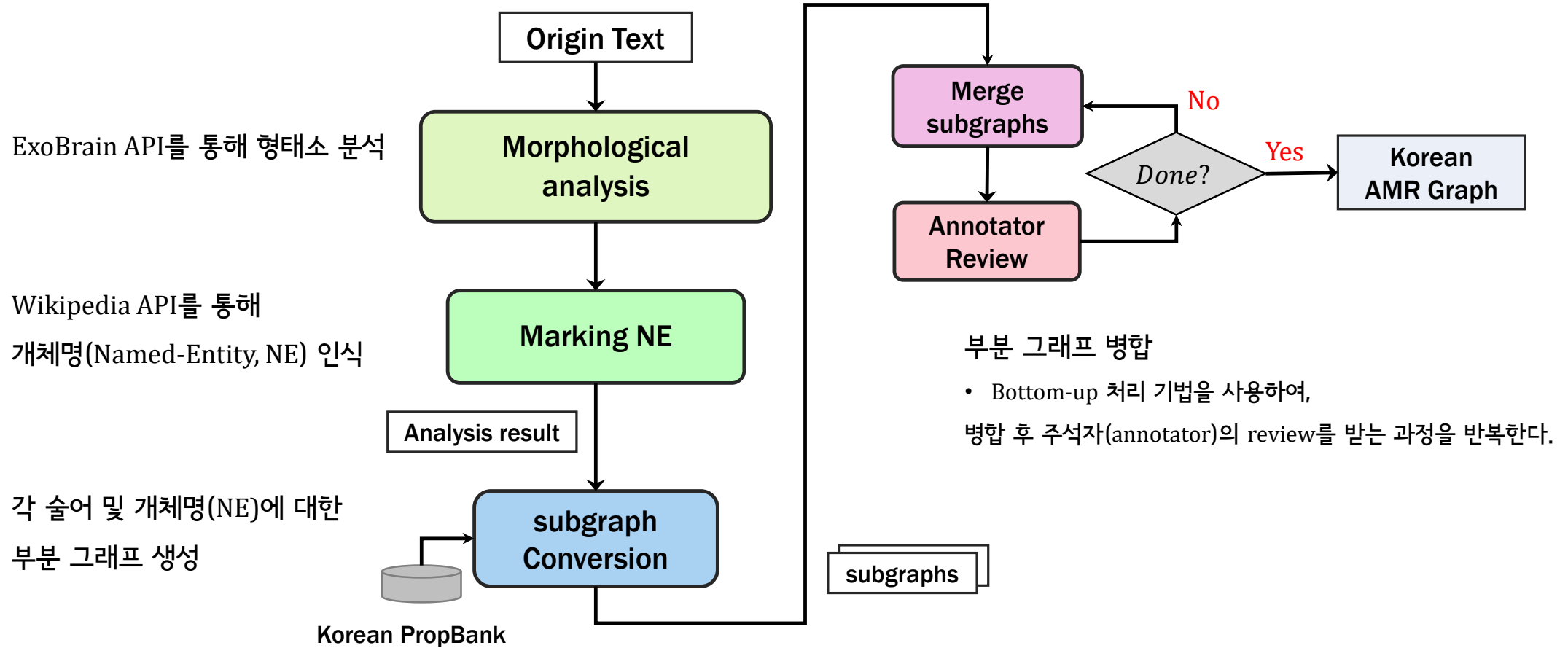
Origin Text

AMR Graph

* Choe, Hyonsu, et al. "Building Korean Abstract Meaning Representation Corpus." *Proceedings of the Second International Workshop on Designing Meaning Representations*. 2020.

3. 관련 연구 – 한국어 AMR

- Annotation process of Korean AMR



3. 관련 연구 – 한국어 AMR

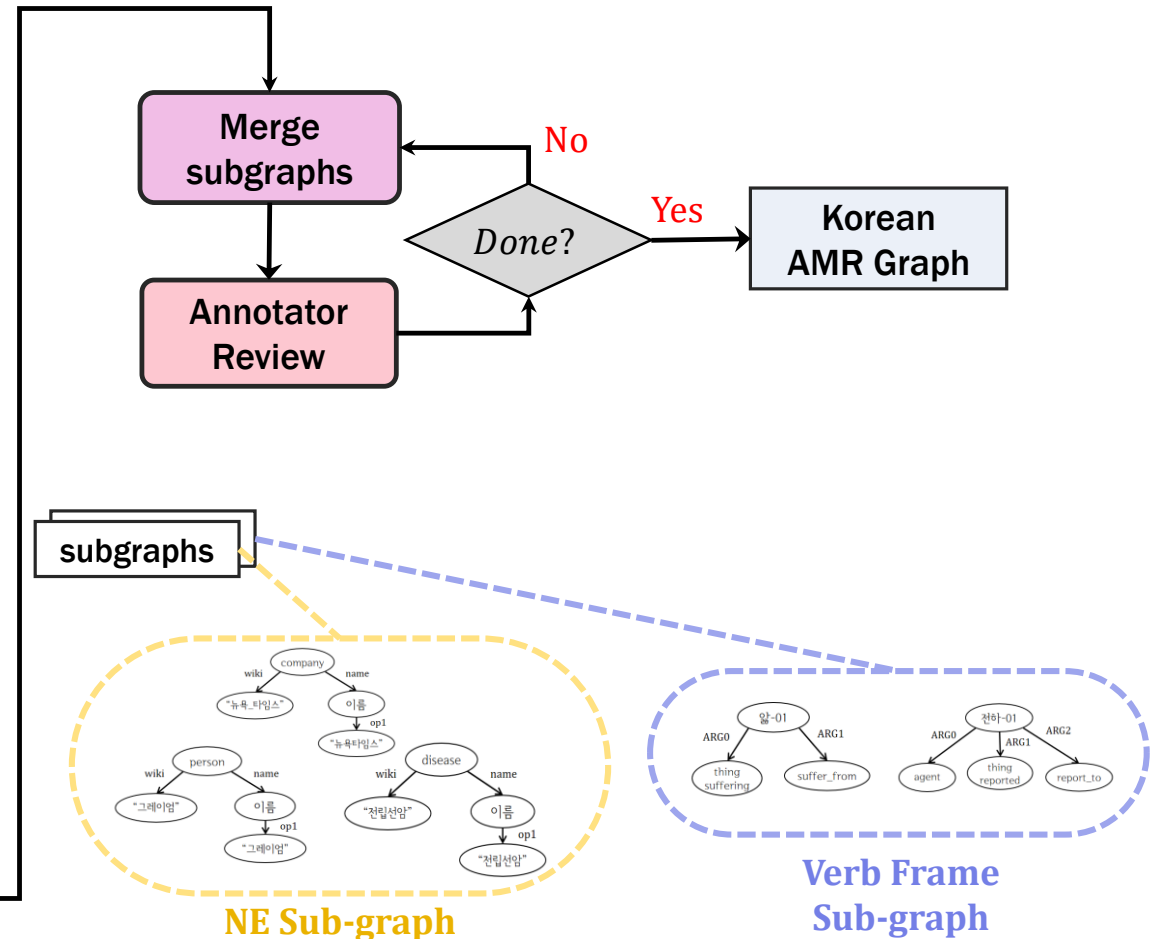
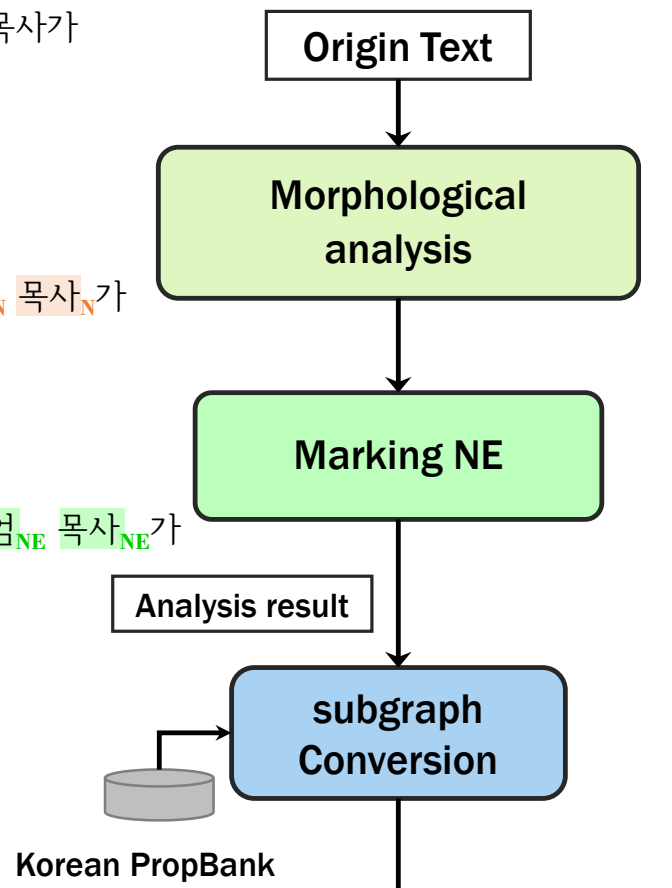
- Annotation process of Korean AMR

뉴욕타임스(NYT)는 그레이엄 목사가
전립선암과 파킨슨병 등을
앓아왔다고 전했다.

뉴욕타임스_N(NYT_N)는 그레이엄_N 목사_N가
전립선암_N과 파킨슨병_N 등_N을
앓_V아왔다고 전했다.

뉴욕타임스_{NE}(NYT_{NE})는 그레이엄_{NE} 목사_{NE}가
전립선암_{NE}과 파킨슨병_{NE} 등_N을
앓_V아왔다고 전했다.

NE : Named-Entity
N : Noun
V : Verb



3. 관련 연구 - 한국어 AMR

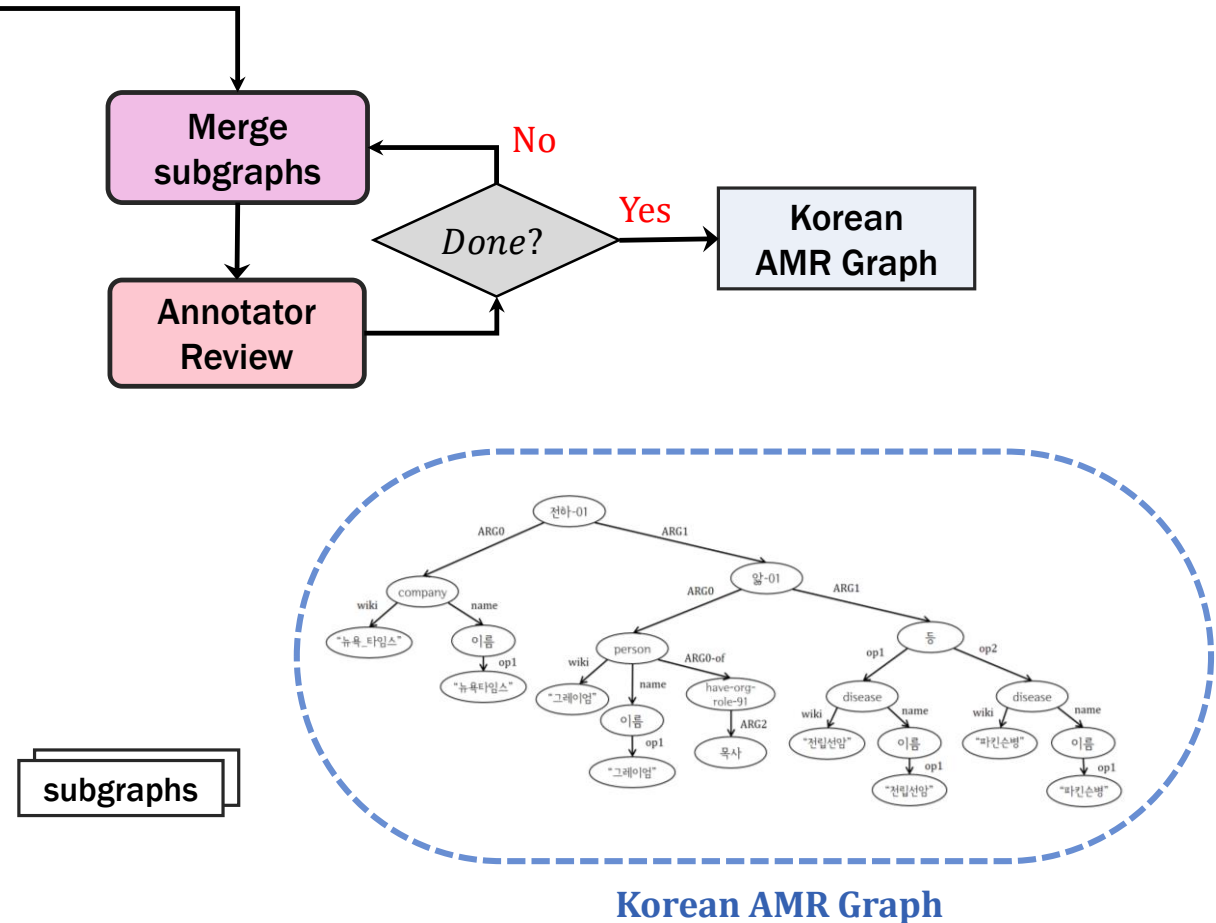
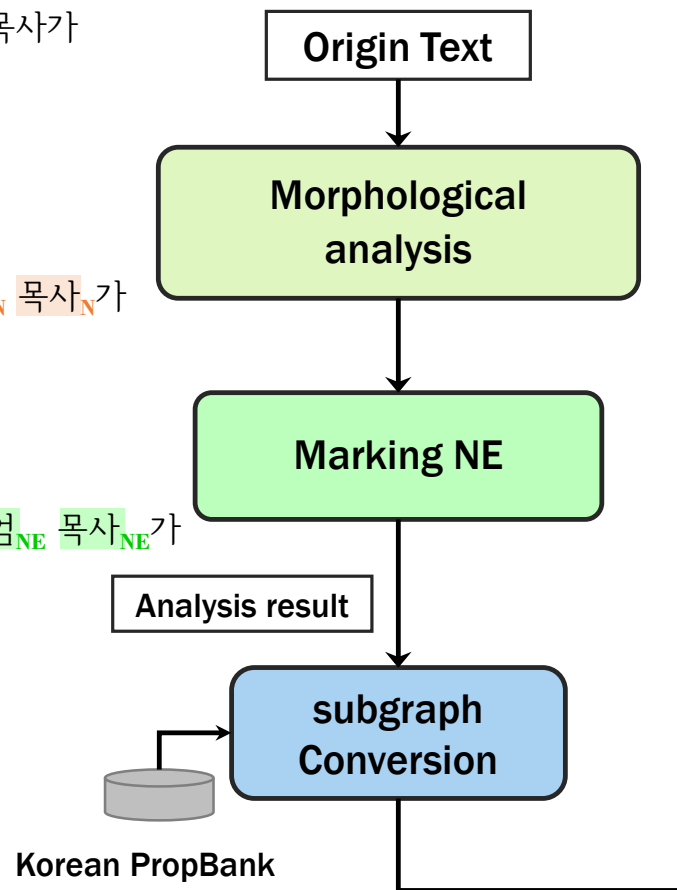
- Annotation process of Korean AMR

뉴욕타임스(NYT)는 그레이엄 목사가
전립선암과 파킨슨병 등을
앓아왔다고 전했다.

뉴욕타임스_N(NYT_N)는 그레이엄_N 목사_N가
전립선암_N과 파킨슨병_N 등_N을
앓_V아왔다고 전했다.

뉴욕타임스_{NE}(NYT_{NE})는 그레이엄_{NE} 목사_{NE}가
전립선암_{NE}과 파킨슨병_{NE} 등_N을
앓_V아왔다고 전했다.

NE : Named-Entity
N : Noun
V : Verb



Korean AMR Graph

3. 관련 연구 – 한국어 AMR

- Korean PropBank (KPB)

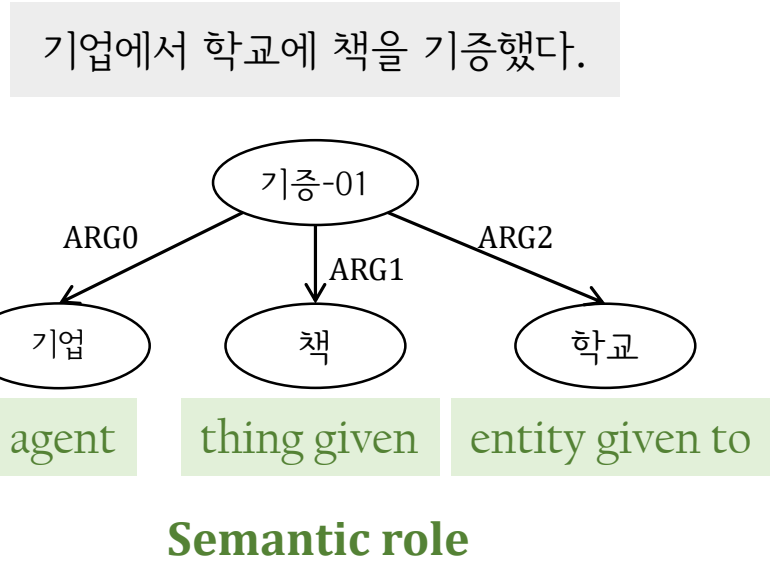
사건(event)의 의미 표상을 위해 기존의 의미 주석 자원(semantic annotation resource)을 사용

한국어 AMR에서는 의미역(semantic role) 의존 언어 자원으로 Korean PropBank를 채택함

KPB는 총 2,749개의 동사 프레임(Verb frame)을 가짐

Frame	argnum	argrole
기증.01 (ki-ceung.kor.xml)	:ARG0	“agent”
	:ARG1	“thing given”
	:ARG2	“entity given to”

* Korean PropBank 내 ‘기증’의 동사 프레임 예시



3. 관련 연구 – 한국어 AMR

- Korean AMR Corpus

Korean AMR을 위한 Korean AMR 말뭉치 구축

총 1,253개의 주석된 문장으로 구성되었으며,

1,253개의 AMR 그래프는 총 20,050의 절점(node)와 18,797개의 간선(edge)을 포함한다.

Source	Subcategory	Snts. (%)
ExoBrain Corpus v4.0	Wikipedia QA Corpus	356 (28.4%)
	Newswire Corpus	256 (20.4%)
	Paraphrase Dataset	253 (20.1%)
	Wikipedia Corpus	234 (18.6%)
Basic Korean Dictionary	Sentence examples of verb entries	120 (9.5%)
The Little Prince (Korean Ed.)	Chapter I (parallel)	34 (2.7%)
		1,253 (100.0%)

* Korean AMR 말뭉치 구성

3. 관련 연구 – 한국어 AMR

- Challenges on Korean Sembanking

1. 정확한 의미 표상

1) 주석자 지원 도구 개발

주석 예제 사전, 개념 및 개체 타입 사전 등을 통해 주석의 일관성 확보

2) AMR 주석 가이드라인의 모호한 표현 수정

3) Korean PropBank의 프레임셋 확장

현재 2,749개의 동사 프레임은 규모나 다양성면에서 부족함

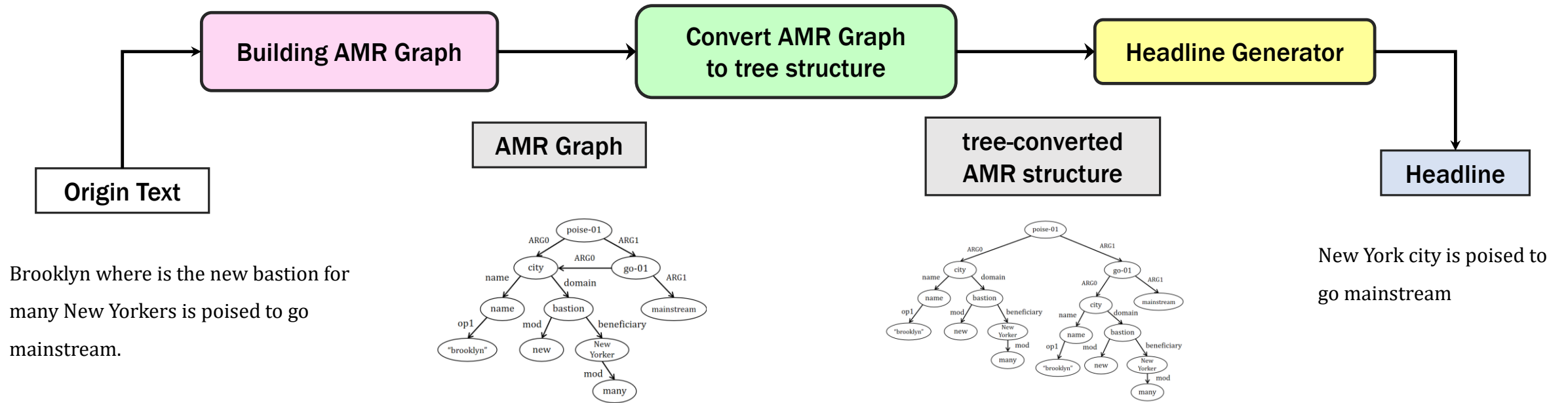
2. AMR 말뭉치의 양적 확장

1) 주석 과정의 부분 자동화

3. 관련 연구 – 제목 생성

- Headline Generation using AMR*

AMR 그래프를 사용하여 제목 생성 (Headline Generation) 수행

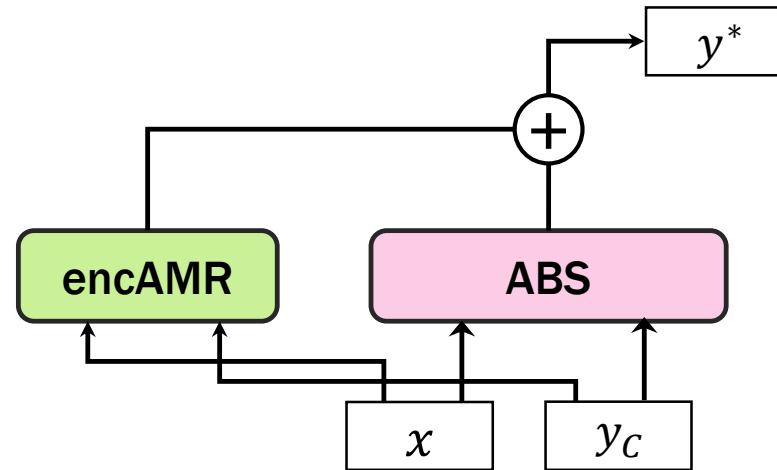


* Takase, Sho, et al. "Neural headline generation on abstract meaning representation." Proceedings of the 2016 conference on empirical methods in natural language processing. 2016.

3. 관련 연구 – 제목 생성

- Headline Generator

주의집중 기반 요약 (Attention-Based Summarization, ABS) 모델¹에 AMR 그래프 특성에 따라 수정된 Tree-LSTM²을 결합한 모델을 사용



¹Rush, Alexander M., et al. "A neural attention model for sentence summarization." ACLWeb. Proceedings of the 2015 conference on empirical methods in natural language processing. 2017.

²Tai, Kai Sheng, Richard Socher, and Christopher D. Manning. "Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks." Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). 2015.

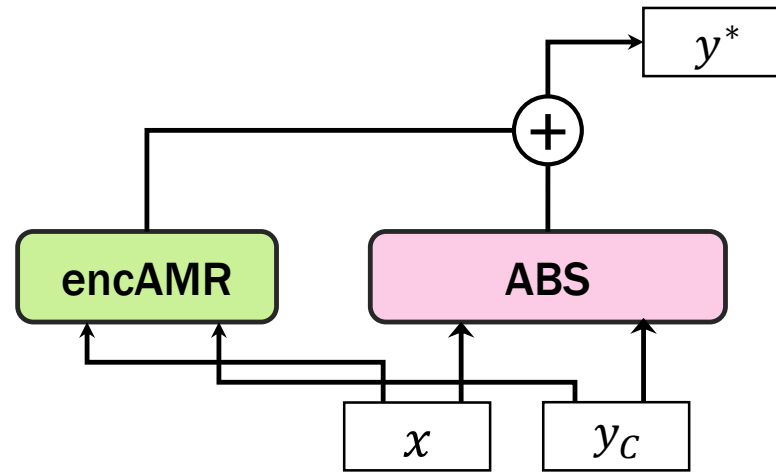
3. 관련 연구 – 제목 생성

- **Headline Generator**

주의집중 기반 요약 (Attention-Based Summarization, ABS) 모델¹에 AMR 그래프 특성에 따라 수정된 Tree-LSTM²을 결합한 모델을 사용

- x M 개의 단어 x_1, \dots, x_M 로 구성된 input sequence
- y N 개의 단어 y_1, \dots, y_N 로 구성된 output sequence ($N < M$)
- y_C window size C 내의 부분 output sequence $[y_{i-C+1}, \dots, y_i]$
- y^* predicted output sequence

- encAMR** Tree-LSTM을 활용한 Attention-based AMR Encoder
- ABS** Attention-based Summarization model
Rush et al.에서의 제목 생성 모델 그대로 사용

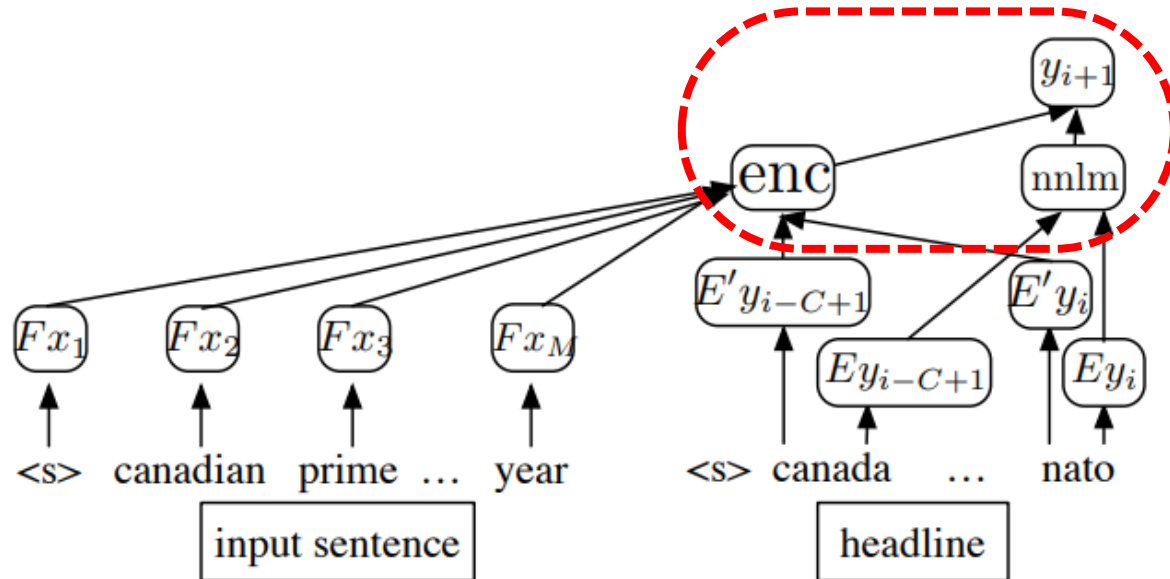


3. 관련 연구 – 제목 생성

- **Headline Generator**

1. 주의집중 기반 요약 (Attention-Based Summarization, ABS) 모델

Attention-based Encoder와 **NNLM**(feed-forward Neural Network Language Model) **Decoder**를 이용하여 다음 출력 단어에 대한 확률 분포를 계산한다.

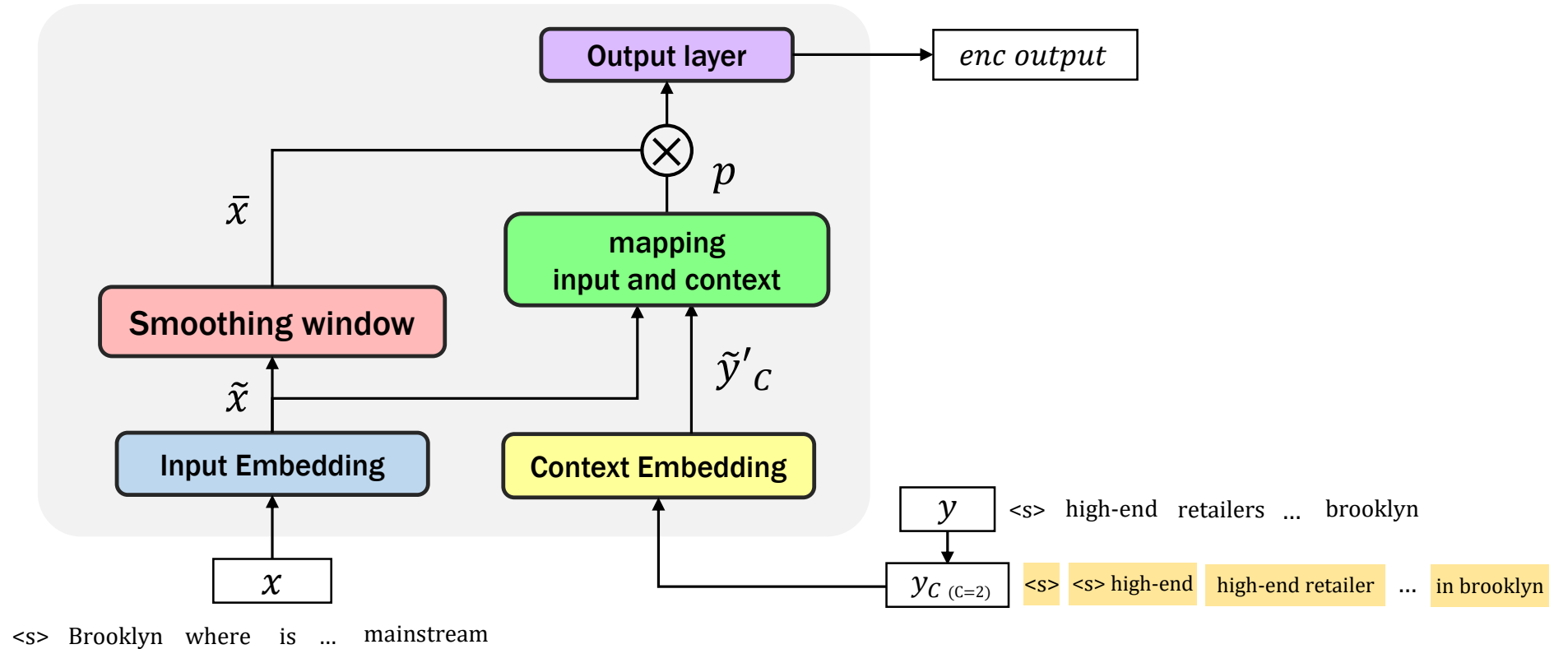


3. 관련 연구 – 제목 생성

- Headline Generator

1. 주의집중 기반 요약 (Attention-Based Summarization, ABS) 모델

1) Encoder



3. 관련 연구 – 제목 생성

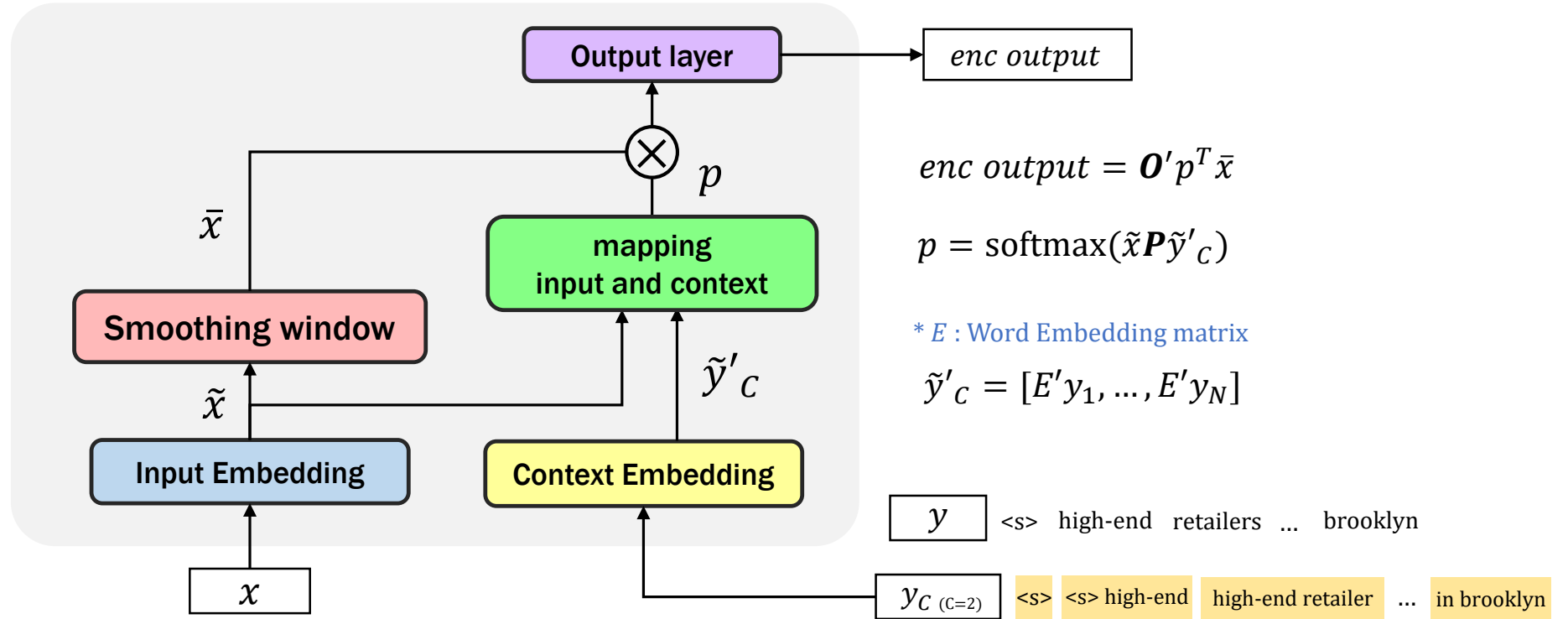
- Headline Generator

1. 주의집중 기반 요약 (Attention-Based Summarization, ABS) 모델

1) Encoder

$$\bar{x} = \frac{1}{Q} \sum_{q=i-Q}^{i+Q} \tilde{x}_i$$

$$\tilde{x} = [Fx_1, \dots, Fx_M]$$



$$enc\ output = O' p^T \bar{x}$$

$$p = \text{softmax}(\tilde{x} P \tilde{y}'_C)$$

* E : Word Embedding matrix

$$\tilde{y}'_C = [E'y_1, \dots, E'y_N]$$

y <s> high-end retailers ... brooklyn

y_C (C=2) <s> <s> high-end high-end retailer ... in brooklyn

<s> Brooklyn where is ... mainstream

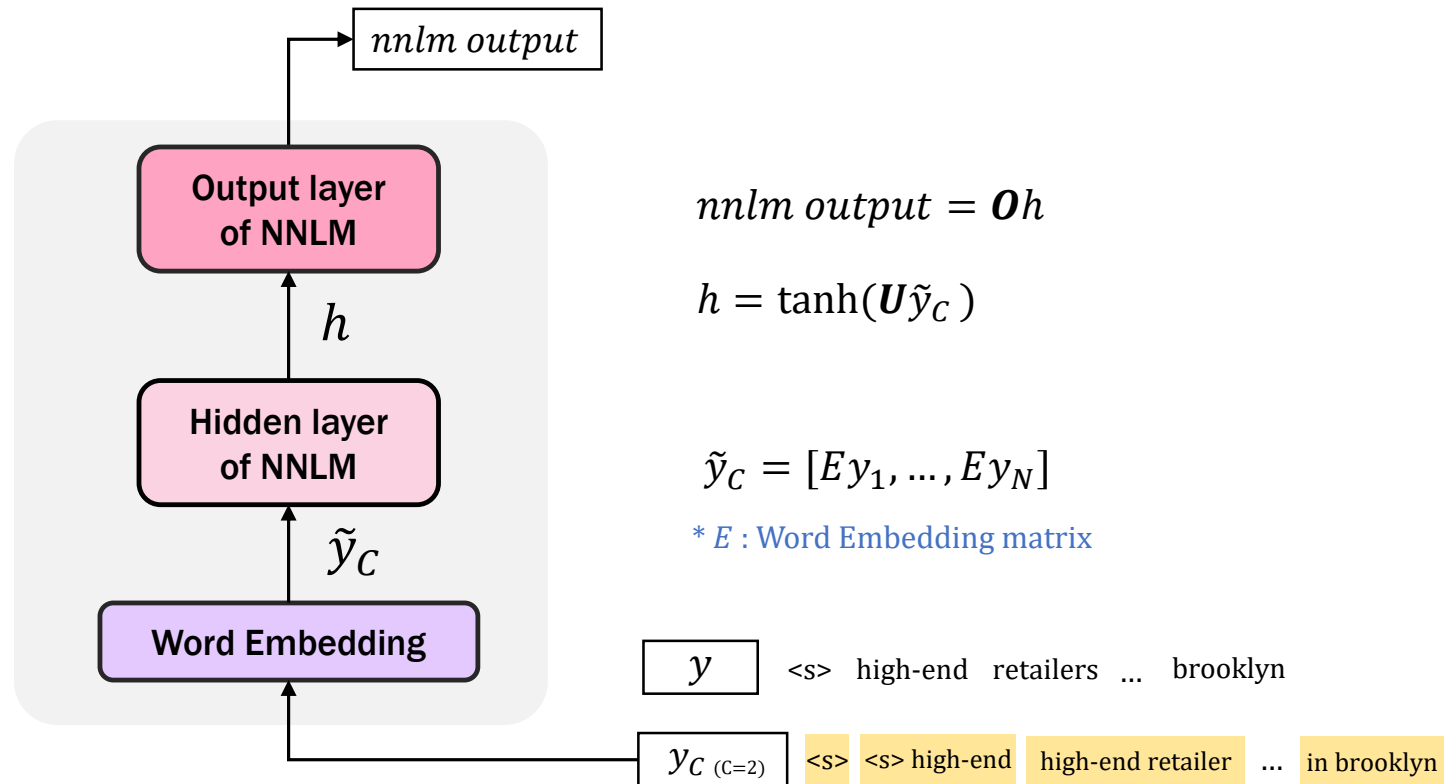
* Q와 C는 각각 window size

3. 관련 연구 – 제목 생성

- Headline Generator

1. 주의집중 기반 요약 (Attention-Based Summarization, ABS) 모델

2) feed-forward Neural Network Language Model (NNLM) Decoder

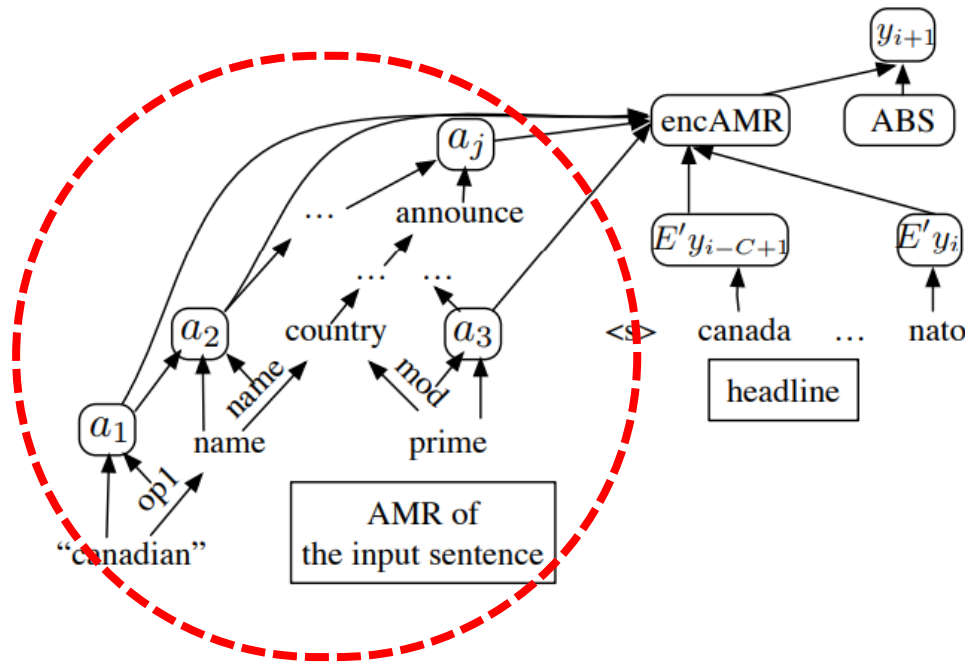


3. 관련 연구 – 제목 생성

- Headline Generator

2. AMR 그래프 특성에 따라 수정된 Tree-LSTM (encAMR)

기존의 순차적 구조가 아닌 트리 구조의 LSTM

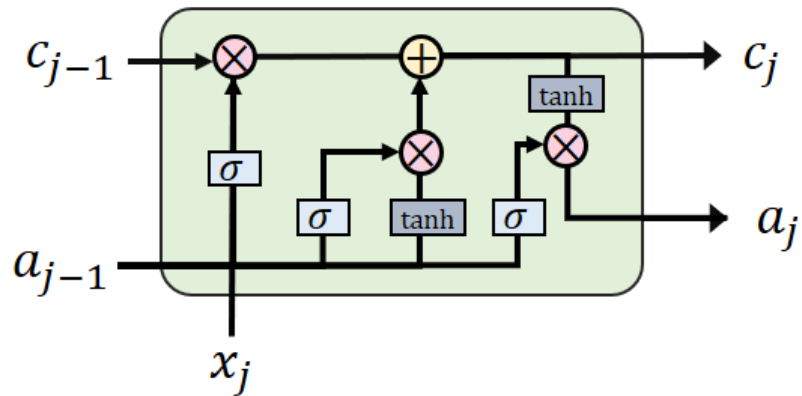


3. 관련 연구 – 제목 생성

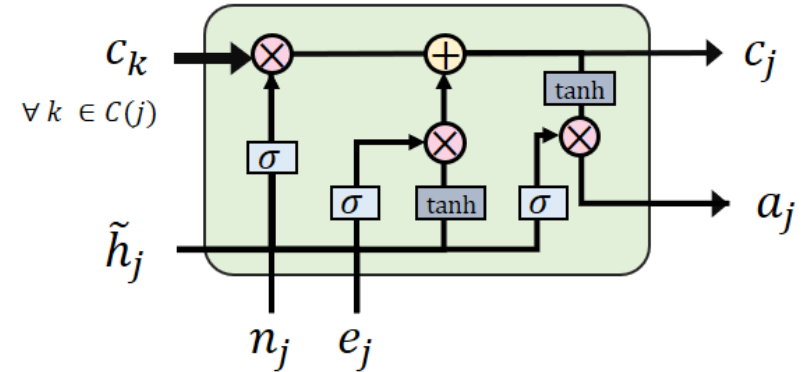
- Headline Generator

2. AMR 그래프 특성에 따라 수정된 Tree-LSTM (encAMR)

기존의 순차적 구조가 아닌 트리 구조의 LSTM



기존 LSTM의 cell structure



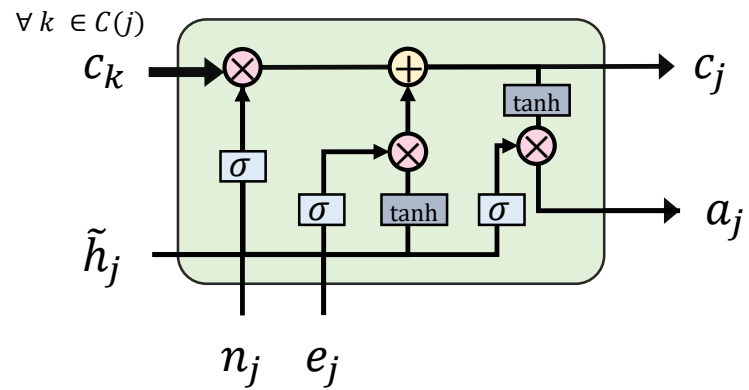
child-Sum Tree-LSTM의 cell structure

3. 관련 연구 – 제목 생성

- Headline Generator

2. AMR 그래프 특성에 따라 수정된 Tree-LSTM (encAMR)

child-Sum Tree-LSTM의 셀 구조



child-Sum Tree-LSTM의 cell structure

- $C(j)$ 노드 j와 연결된 노드 집합
- c_j j번째 셀의 셀 상태
- a_j j번째 셀의 은닉 상태

- n_j j번째 노드 임베딩 벡터 (dim=N)
- e_j j번째 노드의 부모 방향 간선 임베딩 벡터 (dim=E)
- \tilde{h}_j $C(j)$ 내 노드들의 은닉 상태 합

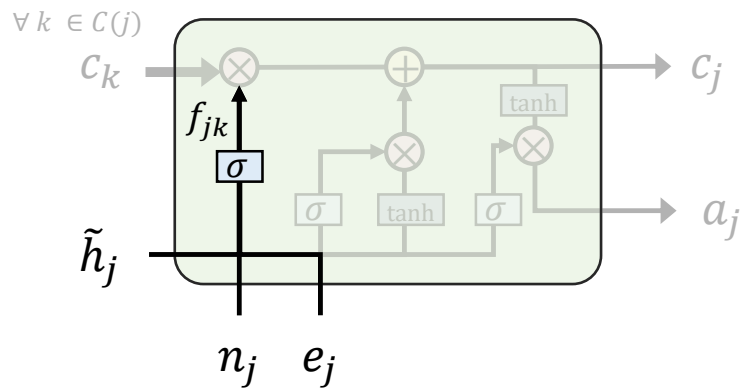
$$\tilde{h}_j = \sum_{k \in C(j)} a_k$$

3. 관련 연구 – 제목 생성

- Headline Generator

2. AMR 그래프 특성에 따라 수정된 Tree-LSTM (encAMR)

child-Sum Tree-LSMT의 셀 연산 과정



$$\forall k \in C(j)$$

$$f_{jk} = \sigma(W_{fn}n_j + W_{fe}e_j + W_{fh}a_k)$$

forget gate

자식 노드들의 은닉 상태와 셀의 입력 정보를 바탕으로

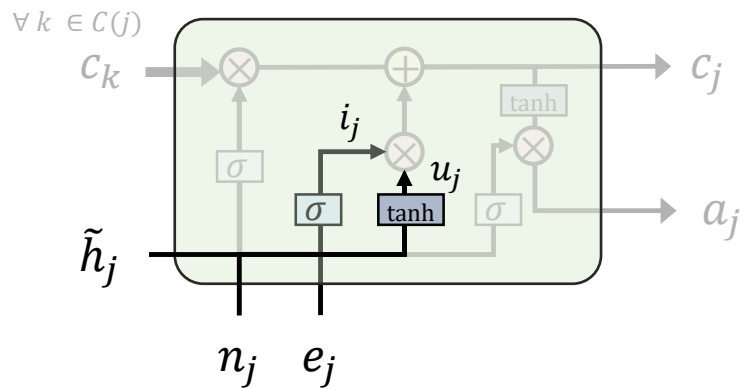
모든 자식 노드에 대하여, 각 노드의 정보를 현재 셀 상태에 유지할 것인지 판단

3. 관련 연구 – 제목 생성

- Headline Generator

2. AMR 그래프 특성에 따라 수정된 Tree-LSTM (encAMR)

child-Sum Tree-LSMT의 셀 연산 과정



$$i_j = \sigma(W_{in}n_j + W_{ie}e_j + W_{ih}\tilde{h}_j)$$

input gate

$$u_j = \tanh(W_{un}n_j + W_{ue}e_j + W_{uh}\tilde{h}_j)$$

New candidate values

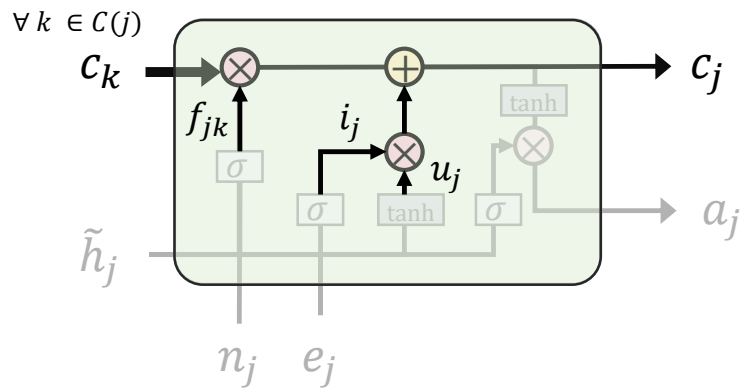
입력된 정보 중 셀 상태에 추가할 새로운 값을 결정

3. 관련 연구 – 제목 생성

- Headline Generator

2. AMR 그래프 특성에 따라 수정된 Tree-LSTM (encAMR)

child-Sum Tree-LSMT의 셀 연산 과정



$$c_j = i_j \cdot u_j + \sum_{k \in C(j)} f_{jk} \cdot c_k$$

cell state

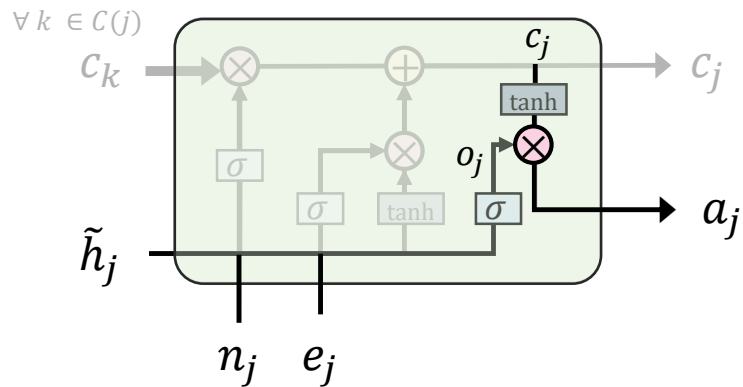
자식 노드들로부터 유지할 정보와 셀의 입력 정보로부터 새롭게 추가할 정보를 통해
현재 셀 상태 갱신

3. 관련 연구 – 제목 생성

- Headline Generator

2. AMR 그래프 특성에 따라 수정된 Tree-LSTM (encAMR)

child-Sum Tree-LSMT의 셀 연산 과정



$$o_j = \sigma(W_{on}n_j + W_{oe}e_j + W_{oh}\tilde{h}_j)$$

$$a_j = o_j \cdot \tanh(c_j)$$

hidden state

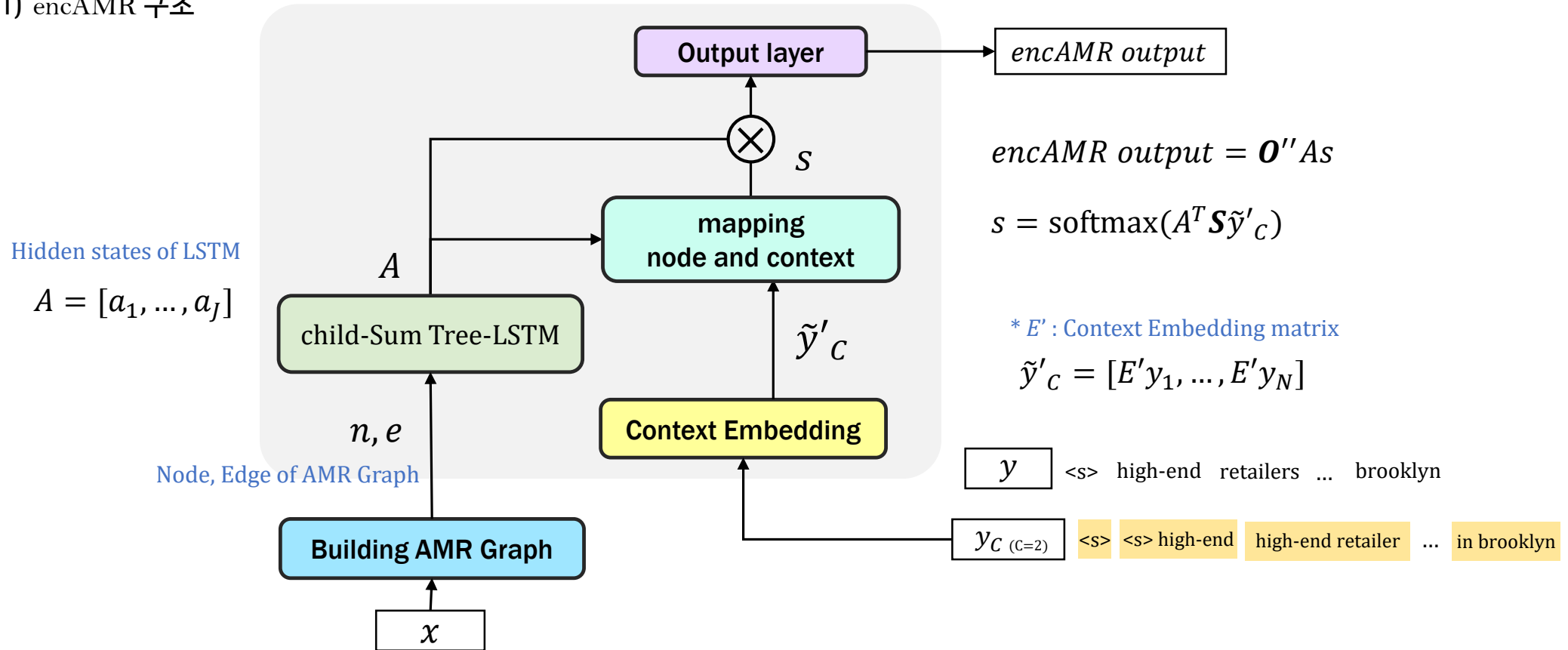
갱신된 셀 상태와 입력 정보들을 바탕으로 은닉 상태 (출력 값) 갱신

3. 관련 연구 – 제목 생성

- Headline Generator

2. AMR 그래프 특성에 따라 수정된 Tree-LSTM (encAMR)

1) encAMR 구조



3. 관련 연구 – 제목 생성

- Headline Generator

3. ABS Model과 AMR Encoder를 결합한 최종 모델

$$y^* = \arg \max_Y s(x, y)$$

$$s(x, y) = \log p(y|x; \theta)$$

Score function

출력 단어에 대한 확률 분포를 계산

$$\log p(y|x; \theta) \approx \sum_{i=0}^{L-1} \log p(y_{i+1}|x, y_C; \theta)$$

$$\log p(y_{i+1}|x, y_C; \theta)$$

$$\propto \exp(\text{nnlm}(y_C) + \text{enc}(x, y_C) + \text{encAMR}(x, y_C))$$

NNLM Decoder of ABS

Encoder of ABS

AMR Encoder

