

# 문장 의미의 그래프 구조 표상을 위한 한국어 Abstract Meaning Representation 가이드라인

최현수<sup>0</sup>, 한지윤, 박혜진<sup>1</sup>, 오태환<sup>1</sup>, 박석원, 김한샘<sup>†</sup>

연세대학교 언어정보학협동과정, 연세대학교 국어국문학과<sup>1</sup>  
{choehyonsu, clinamen35, hjp1010, ghksl0604, pswon27, khss}@yonsei.ac.kr

## Korean Abstract Meaning Representation (AMR) Guidelines for Graph-structured Representations of Sentence Meaning

Hyonsu Choe<sup>0</sup>, Jiyeon Han, Hyejin Park, Taehwan Oh, Seokwon Park, Hansaem Kim<sup>†</sup>  
Institute of Language and Information Studies, Yonsei University

### 요약

이 논문은 한국어 Abstract Meaning Representation (AMR; 추상 의미 표상) 가이드라인 1.0\*을 소개한다. AMR은 통합적인 의미 표상 체계로, 의미 분석(semantic parsing)의 주요 Task 중 하나로 자리매김하고 있다. 한국어 AMR 가이드라인은 현행 AMR 1.2.6을 심도 있게 분석하고 이를 한국어 상황에 맞게 로컬라이징한 것이다. 해당 가이드라인은 추후 한국어 AMR 말뭉치 구축(semanticbanking)에 대비하여 일관된 주석 세부 지침을 제공하기 위해 작성되었다.

**주제어:** Abstract Meaning Representation, 한국어 AMR, 추상 의미 표상, 의미 분석

### 1. 서론

그간의 한국어 의미 주석 체계 및 분석(parsing) 연구는 개체명, 의미역, 상호 참조, 담화 연결사, 시간 및 공간 정보 등 언어 표현이 가진 여러 의미론적 측면 중 일부를 주석하는데 초점을 맞추어 왔다. 그러나 언어 표현과 그 의미 사이의 복잡성을 고려할 때 개별 층위로 파편화된 의미 분석을 넘어서는 통합된 의미 표상 체계 및 언어 자원의 필요성이 대두되고 있다.

이 논문에서는 영어 Abstract Meaning Representation (AMR) 1.2.6 Specification[1,2]과 Korean Propbank (LDC2006T03)[3]를 기반으로 한국어의 특성과 상황에 맞게 로컬라이징한 한국어 AMR 가이드라인 1.0을 소개한다. 한국어 AMR 가이드라인은 영어 및 중국어 AMR[4]을 비롯하여 관련 연구가 활발하게 진행되고 있는 스페인어[5,6], 브라질 포르투갈어[7,8], 베트남어[9] 등 다양한 언어권의 연구들을 폭넓게 참고하여 작성되었다. AMR은 사건 의미를 표상하기 위해 기존의 의미 주석 자원을 필요로 하는데, 한국어 AMR에서는 의미역 의존 언어 자원으로 Korean Propbank를 채택한다. 이는 국내외 공개 언어 자원 및 다국어 AMR 데이터와의 호환성에 중점을 둔 것으로, 앞서 언급한 각 언어에서의 AMR 연구가 vnPropbank, Verbo-Brasil 등 Propbank 계열 언어 자원을 바탕으로 수행된 점, 한국전자통신연구원(ETRI)이 공개한 엑소브레인(Exobrain) 언어처리 학습데이터의 의미역 인식 말뭉치가 Propbank 방식으로 구축된 점[10] 등을 감안한 것이다.

한국어 AMR 가이드라인은 영어 AMR의 주석 지침 전반

을 면밀히 분석하고 이를 한국어에 적용하는 파일럿 연구 단계를 거쳐 세부 지침들을 확정하는 과정을 거쳤다. 본고에서는 통합적 의미 표상 체계로써 AMR의 가능성을 논의하고자 다양한 한국어 예시를 풍부하게 제시하고자 하였다.

### 2. 관련 연구

2013년에 AMR이 처음으로 발표된[1] 이후 관련 연구들이 지속적으로 이어지고 있다. 영어 AMR 말뭉치가 확장되는 과정에서 Propbank의 동사 프레임셋만으로는 해결될 수 없는 사례들에 대응하기 위해 특수 프레임과 개체 타입을 추가하거나[11] 단일 문장 단위 주석을 넘어서는 다중 문장 AMR(MS-AMR) 주석 지침이 추가되는 등[12] 세부적인 주석 지침이 계속해서 개정되어 왔다. 또한 시제와 상을 누락시키고 보편 양화사와 작용역을 적절하게 처리할 수 없는 등의 한계를 극복하여[13,14] 더 나은 의미 표상 체계를 개발하려는 시도들이 활발하게 이어지고 있다.

특히 AMR 파서(parser) 및 응용 연구는 가장 활발하게 수행된 연구 주제다. 2019년 현재 가장 뛰어난 성능을 보이는 모델은 그래프 기반의 모델로[15] LDC2016E25 데이터셋 대상 SMATCH score 76.3의 성능을 보이고 있다. 그 밖에 AMR을 기반으로 문장을 생성하거나[16] 환언 탐지[17], 추출 요약[18,19], 대화 시스템[20]에 응용하는 등 다양한 시도들이 있다.

AMR을 특정 도메인에 적용한 연구로는 의생명 분야에서 구축한 BioAMR이 있다. 이는 의료문서에서 약물간 반

\* <https://github.com/klr1-yonsei/korean-amr-guidelines/blob/master/korean-amr-guidelines.md>

응[21]이나 생체 분자 상호작용을 추출[22]하는데 AMR을 활용한 것이다.

### 3. 한국어 AMR의 체계와 구축

AMR은 문장 단위의 의미 표상(Meaning representation) 프레임워크로 단일 또는 여러 문장의 논리적 의미를 그래프 구조로 표상한다. AMR 그래프는 단일 루트(single-rooted)를 기준으로 하위에 여러 개념(concept)들을 포괄하는 유향 비순환 그래프(DAG)로 산출되는데 이때 루트는 표상의 초점(focus)으로 기능한다. 일단 초점이 정해지면 다른 개념들은 의미론적 관계에 따라 순차적으로 결합된다.

AMR은 형태통사적 특질을 드러내기보다 논리적이고 추상적인 표상을 하는 체계이다. AMR 그래프 내에서 노드(node)로 놓이는 각각의 개념들은 명사나 동사로 취급되기보다 사건(event)이나 개체(entity)로 취급된다. 때문에 AMR에 나타난 개념들과 자연어 문장에서 나타난 어휘는 서로 엄밀하게 정렬되지 않는 것이 일반적이다.

#### 3.1 AMR의 의미 표상

아래는 한국어 문장 “소년은 소녀가 자신을 좋아한다고 생각한다.”의 표상 예시를 세 가지 형식(format)으로 보인 것이다. 아래 (2)는 (3)의 텍스트 친화적인 형태로, Korean Propbank(이하 KPB)의 용언 프레임(frame)을 활용하여 주석한 것이다.

##### (1) LOGIC 형식

```

    ㄴ 생, 소, 좋, 소2:
    instance(생, 생각-01) ^ instance(소, 소년) ^ instance(좋, 좋-02) ^
    instance(소2, 소녀) ^ arg0(생, 소) ^ arg1(생, 좋) ^ arg0(좋, 소2) ^
    arg1(좋, 소)
    
```

##### (2) AMR 형식 (PENMAN notation)

```

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

##### (3) 그래프 형식

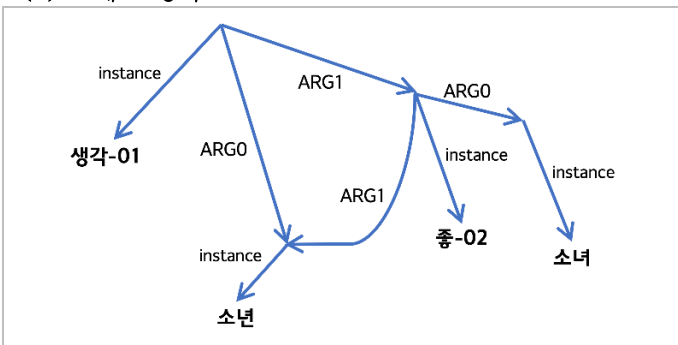


그림 1 소년은 소녀가 자신을 좋아한다고 생각한다.

(2)의 AMR에서 “생”은 “생각-01(think)”의 인스턴트

스(instance)이며, “소(소년)”, “소2(소녀)”, “좋(좋-02; like)” 등과 함께 그래프의 노드들을 구성한다. 이때 “소(소년)”는 사건 “생(생각-01)” 뿐만 아니라 “좋(좋-02)”의 프레임 논항(frame argument)으로도 쓰였는데, AMR에서는 인스턴트의 재진입(reentrancy)을 표상하기 위해 단순히 “:ARG1 소”와 같이 표기한다. 즉, “소”는 다른 위치에서 변항(variable)으로 기능한다.

AMR은 문장의 통사 구조를 분석한 것이 아닌 명제적 의미에 대한 표상이다. 따라서 “소년은 ‘소녀가 나를 좋아하네’라고 생각했다.”와 같이 문장의 구조에 차이를 보이는 경우에도 그림 1과 같이 표상될 수 있다. AMR에서는 여러 문장들의 의미가 동일하다면 그 문장들은 하나의 AMR로 표상될 수 있으며, 비가역적인 표상 체계인 만큼 어떤 하나의 AMR로부터 유일한 자연어 문장을 결정할 수는 없다.

#### 3.2 AMR의 기본 요소

AMR은 Propbank의 프레임셋(frameset)과 일반 어휘들을 활용하여 개념과 개념 간의 관계를 나타낸다. 둘 이상의 개념 또는 논항 노드들의 사이를 잇는 간선(edge)에 해당하는 것이 바로 역할(role)이다. AMR에서는 숫자가 매겨진 필수역(core roles; numbered roles)과 수의역(non-core roles) 등 일반적인 의미역(semantic roles)을 비롯하여 부정, 접속, 명령, 소유, 양보, 수량, 시간, 위키(wikification) 등 다양한 관계를 나타내기 위해 약 100여 개의 역할들을 폭넓게 활용한다.

구분	역할 표지	비고
필수역	:ARGA, :ARG0, :ARG1, :ARG2, :ARG3	개별 프레임에 의존
수의역	:accompanier, :age, :beneficiary, :concession, :condition, :consist-of, :degree, :destination, :direction, :domain, :duration, :example, :extent, :frequency, :instrument, :location, :manner, :medium, :mod, :mode, :name, :ord, :part, :path, :polarity, :polite, :poss, :purpose, :quant, :range, :scale, :source, :subevent, :time, :topic, :unit, :value, :wiki 등	용언에 따라 필수역에 포함된 경우 :ARGn으로 주석
date-entity	:calendar, :century, :day, :dayperiod, :decade, :era, :month, :quarter, :season, :timezone, :weekday, :year, :year2 등	
opX 및 opN 관계	:op1, :op2, :op3, :op4 등	
특수조사	:postp-는커녕, :postp-으로-의 등	제한적 사용
접속	:conj-뿐-아니라 등	제한적 사용
역관계	:ARG0-of, :ARG1-of, :polarity-of :location-of, :consist-of-of 등	관계화

표 1 AMR에서 활용되는 관계 (한국어 가이드라인 기준)

필수역은 KPB에서 개별 용언 프레임별로 정의된 룰셋(roleset)을 따른다. 예를 들어 KPB의 두 표제항인 “주-01(give)”과 “걱정-01(worry)”의 룰셋을 살펴보면

문장의 주어가 언제나 “:ARG0”로 주석되는 것은 아님을 짐작할 수 있다.

수준의 처리 지침을 제공하고 있다. 이러한 문법 현상들 역시 AMR에서는 개념과 개념 간의 관계로 다룬다.

frame	argnum	argrole
주.01 (cu.kor.xml)	:ARG0	“agent”
	:ARG1	“thing given”
	:ARG2	“given to”
걱정.01 (keok-ceong.kor.xml)	:ARG0	“topic of worry”
	:ARG1	“worrier”

표 2 KPB 용언 프레임별 논항 역할 (Predicate-sensitivity)

(주 / 주-01 :ARG0 (소 / 소년) :ARG1 (책 / 책) :ARG2 (소2 / 소녀))
--

예 1 소년은 소녀에게 책을 주었다.

(걱 / 걱정-01 :ARG1 (그 / 그녀) :ARG0 (고 / 고양이) :ARG0-of (나 / 나가-01 :ARG2 (집 / 집)))
---

예 2 그녀는 집을 나간 고양이를 걱정했다.

이와 같이 :ARGx 관계는 각 용언의 격들에 의해 결정되는 상대역이다. AMR은 Propbank 계열 자원에 강하게 의존하므로 사건 또는 개념의 필수 논항을 주석할 때에는 KPB의 프레임을 기준으로 주석해야 한다. 또 일반적으로는 수의역에 해당하지만 경우에 따라 착점(destination) 또는 수혜자(beneficiary)가 특정 용언의 필수역에 해당할 경우, 해당 용언의 롤셋에 따라 :ARGx로 주석한다.

역관계는 복문에서 관계관형절을 표상할 때 주로 사용된다. AMR에서 x :relation y의 관계로 이루어진 트리플(triple)은 y :relation-of x의 관계로 치환될 수 있는데, 이를 이용하면 표상의 초점이 달라도 의미는 동일한 여러 AMR 그래프를 작성할 수 있다.

(모 / 모으-01 :ARG1 (사 / 사람 :quant (수 / 수 :ARG1-of ( 많 / 많-01)))
--

예 3 많은 수의 사람이 모였다.

(많 / 많-01 :ARG1 (수 / 수 :quant-of (사 / 사람 :ARG1-of (모 / 모으-01)))
--

예 4 모인 사람의 수가 많다.

### 3.3 문법 현상의 표상

AMR 가이드라인은 양태(modality), 부정(negation), 의문(question), 소유(possession), 접속(conjunction), 명령(imperative) 등 여러 문법 현상에 대해 기초적인

(가 / 가능-01 :polarity - :ARG1 (걸 / 걸-01 :ARG0 (로 / 로봇 :poss (박 / 박사)))
---

예 5 박사의 로봇은 걸지를 못한다.

(찾 / 찾-01 :ARG1 (소2 / 소년) :location (a / amr_unknown))
--

예 6 소년을 어디서 찾은 거야?

한국어 AMR 가이드라인은 기존의 영어 AMR의 주석 지침을 최대한 수용하는 한편, 한국어에서 특징적으로 나타나는 현상들을 처리하는 방법을 추가로 제시하였다. 일례로 이중주어문과 이중목적어문을 들 수 있다. 이는 동일한 문장 내에 주격 또는 목적격 조사가 두 번 이상 할당된 경우를 말한다. 대부분의 경우 필수역 또는 수의역을 통해 격중출 구문에서의 중의성을 해소할 수 있지만 논항 간의 관계를 뚜렷하게 결정할 수 없는 경우에는 “:domain”을 활용하여 주석할 수 있다.[23]

(좋 / 좋-01 :ARG1 (향 / 향) :domain (와 / 와인))
---

예 7 와인이 향이 좋다.

(빽 / 빽-01 :ARG1 (짱 / 짱) :domain (제 / 제비))
---

예 8 제비를 짱을 빽았다.

### 3.4 개체 및 특수 프레임 활용 표상

AMR은 다양한 타입의 개체를 활용하여 시간이나 위치, 거리와 같은 추상적인 개체에 대해 더 쉽고 직관적인 주석이 가능하다. 아래 예 (9)에서는 date-entity를 통해 시간을, relative-position을 통해 상대적 위치를, distance-quantity를 통해 거리를 표상하였다.

(추 / 추락-01 :ARG1 (비 / 비행기) :location (r / relative-position :op1 (c / city :wiki "모스크바" :name (이 / 이름 :op1 "모스크바")) :quant (d / distance-quantity :quant 50 :unit (마 / 마일)) :direction (동 / 동쪽)) :time (d / date-entity :month 3 :day 24 :time 15:36))
--

예 9 비행기는 3월 24일 오후 3시 36분 모스크바로부터 동쪽으로 50마일 지점에 추락했다.

AMR에서 개체명(named entity)은 ‘:name’ 역할을 통해 표상된다. 그러나 텍스트에서 나타나는 개체명은 규범 표기가 아닌 경우가 많다. AMR은 이러한 문체를 위키피케이션(wikification)을 통해 보완한다. 한국어 AMR에서는 한국어 위키백과의 항목명을 대표형으로 하여 ‘:wiki’ 역할을 주석할 수 있다.

```
(g / government-organization
 :wiki "대한민국_국가정보원"
 :name (이 / 이름
       :op1 "국정원")
 :mod (c / country
       :wiki "대한민국"
       :name (이2 / 이름 :op1 "한국"))))
```

예 10 한국의 국정원

원칙적으로 AMR에서 개체로 취급되는 모든 노드는 “:name” 관계를 가질 수 있다. 그러나 개체 타입을 특정하기 어려운 경우를 위해 일반적인 개체명 타입들의 목록을 제공한다. 즉 아래 개체 타입 중 ‘company’는 한국어 AMR에서 ‘회사’나 ‘기업’의 alias로 취급될 수 있다.

인물	person, family, animal, language, nationality, ethnic-group, regional-group, religious-group, political-movement
기관	organization, company, government-organization, military, criminal-organization, political-party, market-sector, school, university, research-institute, team, league
장소	location, city, city-district, county, state, province, territory, country, local-region, country-region, world-region, continent; ocean, sea, lake, river, gulf, bay, strait, canal; peninsula, mountain, volcano, valley, canyon, island, desert, forest, moon, planet, star, constellation
시설	facility, airport, station, port, tunnel, bridge, road, railway-line, canal, building, theater, museum, palace, hotel, worship-place, market, sports-facility, park, zoo, amusement-park
사건	event, incident, natural-disaster, earthquake, war, conference, game, festival
인공물	product, vehicle, ship, aircraft, aircraft-type, spaceship, car-make, work-of-art, picture, music, show, broadcast-program
저작물	publication, book, newspaper, magazine, journal
자연물	natural-object
기타	award, law, court-decision, treaty, music-key, musical-note, food-dish, writing-script, variable, program
의생명 도메인 특화 NE타입 (BioAMR)	molecular-physical-entity, small-molecule, protein, protein-family, protein-segment, amino-acid, macro-molecular-complex, enzyme, nucleic-acid
	pathway, gene, dna-sequence, cell, cell-line, species, taxon, disease, medical-condition

표 3 AMR에서 쓰이는 일반적인 개체 타입

이 밖에도 AMR에서는 역할을 일급 개념(first-class concept)으로 구체화(reification)하여 활용하거나 have-org-role-91 또는 have-rel-role-91 등의 특수 프

레이미를 통해 집단 내에서의 직위나 사회적 상호관계를 표상할 수 있다.

```
(p / person
 :wiki "도널드_트럼프"
 :name (이 / 이름 :op1 "트럼프")
 :ARG0-of (h / have-org-role-91
           :ARG1 (c / country :wiki "미국"
                  :name (이2 / 이름 :op1 "미국")))
           :ARG2 (대 / 대통령)))
```

예 11 미국의 트럼프 대통령

```
(h / have-rel-role-91
 :ARG0 (그 / 그)
 :ARG1 (나 / 나)
 :ARG2 (처 / 처남))
```

예 12 그는 내 처남이다.

이 밖에도 주소 객체를 위한 street-address-91이나, 확인 의문문(tag question)의 처리를 위한 request-confirmation-91, '~르(을)수록 더 ~하다' 구문(The X-er, The Y-er construction)을 처리하기 위한 correlate-91 등 특수 목적에 맞게 확장된 특수 프레임들이 다방면으로 활용될 수 있다. 이는 데이터 주석의 난이도를 경감시키고 더 효율적인 표상을 가능케 하며 개별 언어의 상황에 맞게 확장도 가능하다.[11]

### 3.5 상호 참조를 위한 무형 대용어 표상

자연어 문장에서 외현적 대명사(overt pronouns)와 영 대명사(zero pronouns)는 흔히 상호 참조(coreference)를 실현하는 데 쓰인다. AMR에서는 인스턴스를 변항(variable)으로 취급하므로 암시된 역할을 명시적으로 표상할 수 있는 방법을 제공한다.

```
(기 / 기소-01
 :ARG1 (그 / 그)
 :ARG2 (a / and
       :op1 (주 / 주폭)
       :op2 (저 / 저항-01
             :ARG0 그
             :ARG1 (체 / 체포-01
                   :ARG1 그))))
```

예 13 그는 주폭과 체포 저항으로 기소되었다.

위 예시 (13)에서는 주어진 자연어 문장과 다르게 변항 ‘그’가 2군데 더 나타나는 것을 볼 수 있다. AMR에서 표면형에서 해당 표현이 실현되었는가는 크게 중요하지 않다. 다른 사람이 아닌 ‘그 자신’의 체포에 “그”가 저항한 것이 자명하므로 “저항-01(resist)”의 “:ARG0”과 “체포-01(arrest)”의 “:ARG1”에 변항 “그”를 놓을 수 있다.

다중 문장 주석을 위한 프레임인 multi-sentence를 활용하면 문장 경계를 넘어서는 무형 대용어와 상호 참조도 표상할 수 있다.[12] 아래 예시 (14)는 두 문장 “김침지는 종로를 따라 광장시장으로 걸었다.”와 “그는 정오에서야 도착했다.”를 하나의 AMR로 표상한 것이다.

```
(m / multi-sentence
:snt1 (길 / 길-01
:ARG0 (p / person
:wiki -
:name (이 / 이름 :op1 "김철지"))
:ARG1 (r / road
:wiki "종로_(서울)"
:name (이2 / 이름 :op1 "종로"))
:ARG3 (m2 / market
:wiki "광장시장"
:name (이3 / 이름 :op1 "광장시장")))
:snt2 (도 / 도착-01
:ARG1 p
:ARG3 m2
:time (정 / 정오)))
```

예 14 김철지는 종로를 따라 광장시장으로 걸었다. 그는 정오에서야 도착했다.

위 예시에서 두 번째 문장의 표상에서는 앞선 문장에서 표상된 인스턴스 변항을 그대로 활용하고 있다. 다중 문장 AMR은 기존의 단일 문장 AMR들을 바탕으로 간단하게 생성할 수 있다.

#### 4. 한국어 AMR 가이드라인의 특징

AMR은 영어를 중심으로 개발된 의미 표상 체계이다. 따라서 현행 AMR 정책은 영어에 맞춰져 있다. 비록 AMR이 개별 언어의 형태-통사적 특질보다는 추상적인 의미에 더 가까운 표상을 지향하지만 유사한 뜻을 나타내고 간주되는 여러 문장들이 단 하나의 AMR로 완벽히 표상될 수는 없다. 자연어 문장의 의미는 그 언어 고유의 어휘, 품사 범주, 어순 등의 여러 형태-통사적 특질들로부터 완전히 자유로울 수는 없기 때문이다. 그러나 AMR이 언어 표현의 표층보다는 기저의 의미에 집중하는 만큼 서로 다른 언어 의미 그래프 간의 정렬은 비교적 수월한 것으로 알려졌다.[24]

한국어 AMR 가이드라인은 영어 AMR을 전면적으로 분석하고 무리없이 들여올 수 있는 주석 지침을 가급적 그대로 수용하고자 하였다. 영어 및 타 언어 Propbank 계열 자원간의 호환성을 고려하여 한국어 용언을 주석함에 있어서는 Korean Propbank를 채택하였고, '-91' 류 특수 프레임 및 AMR 고유의 개체(entity) 타입은 별도의 번역이나 수정 없이 한국어 AMR 체계로 수용하고자 하였다. 비록 한국어의 특성을 엄밀하게 반영하는 데에는 부족함이 있겠으나 한국어 AMR의 활용성 측면에서는 더 바람직한 접근이 될 것이다.

또한 기존 영어 AMR에서 제시한 기준들만으로는 한국어를 직접 처리하기에는 부족함이 있는 것이 사실이다. 이에 한국어 AMR 가이드라인에서는 지정사 '-이다'와 어휘화된 부정형용사 '아니다'의 취급 방침을 명확히 제시하고, 한국어의 특징적인 현상 중 하나인 격중출 구문의 처리 방침 등을 보장하는 등 한국어 상황에 맞는 확장 지침들을 제공할 필요가 있었다. 특히 영어의 품사 범주인 Adjective와 한국어의 형용사 사이의 간극을 다각적으로 검토하고, 한국어 상황에 맞는 지침을 제

시하고자 하였다. 아래는 영어 AMR의 내용을 한국어 상황에 맞게 고친 부분을 나열한 것이다.

영어 AMR에서 제시된 내용	한국어 AMR에서 수정 보완한 내용	비고
Negation	부정	'아니다', '모르다', '없다' 등 대표적인 부정 용언 및 명관류의 부정접두사에 대한 기준 지침 제시
Articles, plurals, tense, aspect, quotes, hyphens	지시사, 복수형, 시제, 상, 인용, 하이픈	지시관형사, 수관형사, 수사의 기준 지침 제시
Main verb "be"	지정사 '-이다'와 계사적 쓰임을 갖는 '있다'	한국어 계사 및 존재형용사 '있다'의 처리 지침 제시, 부정형용사 '아니다'와 분열문 기준 지침 제시
Nouns that invoke predicates	동사 파생이 될 수 있는 서술성 명사	동사 파생 접미사, 전성 어미, '화(化)', '자(者)', '인(人)', '사(社)' 등의 한자 형태소에 대한 처리 기준 제시
Adjectives that invoke predicates	형용사 파생 및 보조 형용사	형용사 파생 접미사, '감(感)' 등의 한자 형태소, '직하다', '만하다' 등의 보조 형용사에 대한 처리 기준 제시
Phrasal verbs	동사구	본 용언-보조 용언 구성의 처리 기준 제시
Prepositions	한국어의 조사와 시간 및 공간 관련 표현	영어의 전치사의 의미적 기능에 대응하여 특수 조사, 명사 등의 처리 기준 제시
-	높임과 공손성	높임 및 존칭어, 공손성 처리 기준 제시
-	이중주어문과 이중 목적어문	주격 조사, 보조사, 목적격 조사 중출에 대한 처리 기준 제시
-	압축적 서술과 명사 상당 어구	뉴스 텍스트에서 자주 나타나는 압축적 서술 및 명사 상당 어구에 대한 심층적 처리 기준 제시
-	우연적 구성	공개 예정
-	특수 의문문	공개 예정

표 4 영어 AMR 대비 한국어 AMR 가이드라인 로컬라이징 중점 사항

#### 5. 결론

본 논문은 층위별로 파편화된 한국어 의미 분석을 보완할 수 있는 대안으로 통합적 의미 표상 체계인 한국어 AMR 구축을 위한 가이드라인을 소개하였다.

한국어 AMR의 활용도를 높이기 위해서는 기준이 되는 Korean Propbank의 프레임셋에 대한 확장이 필요하다. 현재 표제항 기준 2,749개의 용언 프레임은 규모나 다양성 면에서 부족함이 있기에 한국어 용언에 최적화된 프레임셋의 대대적인 보완이 선결 과제라 할 수 있다.

본고의 연구진들은 2020년 하반기 공개를 목표로 한국어 AMR 말뭉치를 구축 중에 있다. 이는 한국전자통신연구원(ETRI)이 제공하는 엑소브레인(Exobrain) 언어분석 말뭉치의 뉴스 텍스트, 위키백과, 위키QA 데이터 일부와 국립국어원 한국어 기초 사전에 수록된 예문 일부를 대상으로 한다. 2020년 하반기까지 약 1,500 문장 분량을 우선 구축 및 공개할 계획이며, 구축 과정 전반을 통해 한국어 AMR 가이드라인 또한 활발히 개정될 것이다.

## 참고문헌

- [1] Banarescu, Laura, et al., “Abstract meaning representation for sembanking”, *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, 2013.
- [2] Banarescu, Laura, et al., “Abstract Meaning Representation (AMR) 1.2.6 Specification”, 2019. <https://github.com/amrisi/amrguidelines/blob/master/amr.md>
- [3] Palmer, Martha, et al., “Korean Propbank LDC2006T03”, Linguistic Data Consortium, 2006.
- [4] LI, Bin, et al. “Annotating the little prince with chinese amrs”, *Proceedings of the 10th linguistic annotation workshop held in conjunction with acl 2016 (law-x 2016)*, pp. 7-15, 2016.
- [5] Migueles Abairra, Noelia. “A Study Towards Spanish Abstract Meaning Representation”, MSc Thesis, 2017.
- [6] Migueles Abairra, Noelia, et al., “Annotating Abstract Meaning Representations for Spanish”, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018)*, 2018.
- [7] Anchiêta, Rafael; Pardo, Thiago. “Towards AMR-BR: A SemBank for Brazilian Portuguese Language”, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018)*. 2018.
- [8] Cabezudo, Marco Antonio Sobrevilla; Pardo, Thiago. “Towards a General Abstract Meaning Representation Corpus for Brazilian Portuguese”, *Proceedings of the 13th Linguistic Annotation Workshop*, pp. 236-244, 2019.
- [9] Linh, Ha; Nguyen, Huyen. “A Case Study on Meaning Representation for Vietnamese”, *Proceedings of the First International Workshop on Designing Meaning Representations*. pp. 148-153, 2019.
- [10] 임수중, 권민정, 김준수, 김현기, “ExoBrain을 위한 한국어 의미역 가이드라인 및 말뭉치 구축”, *제 27회 한글 및 한국어 정보처리 학술대회 논문집*, pp. 250-254, 2015.
- [11] Bonial, Claire, et al., “Abstract Meaning Representation of constructions: The more we include, the better the representation”, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018)*. 2018.
- [12] O’Gorman, Tim, et al., “AMR beyond the sentence: the Multi-sentence AMR corpus”, *Proceedings of the 27th International Conference on Computational Linguistics*, pp. 3693-3702, 2018.
- [13] Donatelli, Lucia, et al. “Annotation of tense and aspect semantics for sentential AMR”, *Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)*, pp. 96-108, 2018.
- [14] Pustejovsky, James; Lai, Ken; Xue, Nianwen. “Modeling Quantification and Scope in Abstract Meaning Representations”, *Proceedings of the First International Workshop on Designing Meaning Representations*. pp. 28-33. 2019.
- [15] Zhang, Sheng, et al. AMR Parsing as Sequence-to-Graph Transduction. *arXiv preprint arXiv:1905.08704*, 2019.
- [16] Damonte, Marco; Cohen, Shay B. “Structural Neural Encoders for AMR-to-text Generation”, *arXiv preprint arXiv:1903.11410*, 2019.
- [17] Liao, Kexin; Lebanoff, Logan; Liu, Fei. “Abstract meaning representation for multi-document summarization”, *arXiv preprint arXiv:1806.05655*, 2018.
- [18] Vlachos, Andreas, et al. “Guided neural language generation for abstractive summarization using Abstract Meaning Representation”, *arXiv preprint arXiv:1808.09160*, 2018.
- [19] Issa, Fuad, et al. “Abstract meaning representation for paraphrase detection”, *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pp. 442-452, 2018.
- [20] Bonial, Claire, et al. “Augmenting Abstract Meaning Representation for Human-Robot Dialogue”, *Proceedings of the First International Workshop on Designing Meaning Representations*, pp. 199-210, 2019.
- [21] Wang, Yanshan, et al. “Dependency and AMR embeddings for drug-drug interaction extraction from biomedical literature”, *Proceedings of the 8th acm international conference on bioinformatics, computational biology, and health informatics*, ACM, pp. 36-43, 2017.
- [22] Garg, Sahil, et al. “Extracting biomolecular interactions using semantic parsing of biomedical text”, *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.
- [23] Choe, Hyonsu, et al., “Copula and Case-Stacking Annotations for Korean AMR”, *Proceedings of the First International Workshop on Designing Meaning Representations*. pp. 128-135, 2019.
- [24] Xue, Nianwen, et al., “Not an Interlingua, But Close: Comparison of English AMRs to Chinese and Czech”, *LREC*, pp. 1765-1772, 2014.