

Korean Quantification in Abstract Meaning Representation

Kiyong Lee, Harry Bunt,
Byong-Rae Ryu, Chongwon Park,
Younggyun Hahm

Korea University, Tilburg University,
Chungnam National University, University of Minnesota Duluth,
Teddysum AI
ikiyong@gmail.com, Harry.Bunt@tilburguniversity.edu,
ryu@cnu.ac.kr, cpark2@d.umn.edu,
hahmyg@teddysum.ai

Abstract

This paper explores the meaning of quantification in Korean and how it is encoded in Abstract Meaning Representation (Banarescu et al., 2019) and an enriched version AMR⁺ accommodating Uniform Meaning Representation (UMR, 2022) and some of the contextual constraints proposed by Bos (2020). The extension makes five special references: Bunt et al. (2018), Bunt and Lee (2025), Pustejovsky et al. (2019), Bos (2020), and ISO (2025), the main reference. The aim of this paper is threefold. First, it focuses on implementing Korean AMR with the rich specification of QuantML (ISO, 2025) and its partially DRT-based semantics (Kamp and Reyle, 1993). Second, it supports the AMR multilingual development project by exploring methods for constructing a large-scale Korean AMR-annotated corpus. This line of research is necessary because Korean AMR resources remain severely underdeveloped. To date, only one study has examined a Korean AMR corpus (Choe et al., 2020). In addition, Korean’s agglutinative morphology and head-final syntax challenge AMR frameworks that are largely based on the analytic inflectional language English. Third, it advances the current state of the UMR 2026 multilingual shared task by contributing more fine-grained annotations of quantification specified by ISO QuantML (ISO, 2025) for resource domain, individuation, distributivity, and determinacy, as well as by treating coreference and lexical or scope ambiguities in Korean. Nonetheless, this paper postpones a review of the general semantics of Korean quantification until particular issues are addressed in future work.

Keywords: quantification, AMR, UMR, QuantML, contextual constraint

1. Aim and Motivation

This paper explores the meaning of quantified or polarity-sensitive quantifier-related sentences in Korean, not at a full scale, but at an experimental level, to examine how they are encoded in Abstract Meaning Representation (Banarescu et al., 2019) and (Knight et al., 2020), enriched by the ISO international standard QuantML (ISO, 2025). The enriched AMR⁺ is also extended by accommodating Bos (2020) and Uniform Meaning Representation (UMR, 2022) with some modifications. Our extension draws on Bunt et al. (2018), Pustejovsky et al. (2019), Bunt (2021, 2024), Bunt and Lee (2025), and Bos (2020) to accommodate scope constraints within UMR and also to formulate the syntax of UMR in the manner of AMR.

Furthermore, unlike categorial grammar-based Montague semantics or phrase-structure dependent Discourse Representation Theory (DRT) (Kamp and Reyle, 1993), AMR-based semantics abstracts away from syntactic variations, having no syntactic pre-processing as required. On the other hand, Korean AMR is expected to be enriched with its rich morpho-syntactic information. For example, different case markings, such as the nominative marker -가 -ga or -이 i, in contrast to the topic marker -는 neun or -은 -eun, differentiate semantic interpretations.

The aim of this paper is threefold. First, it focuses on implementing Korean AMR with the rich specification of quantification in natural language by various efforts, such as Bunt (2021, 2024), Bunt et al. (2018, 2022), and Bunt and Lee (2025) to formulate ISO 24617-12:2025 on quantification as an ISO international standard. Second, this paper aims to support the AMR multilingual development project by exploring methods for constructing a large-scale Korean AMR-annotated corpus. This line of research is necessary because Korean AMR resources remain severely underdeveloped. To date, only one study has examined a Korean AMR corpus (Choe et al., 2020). In addition, Korean’s agglutinative morphology and head-final syntax challenge AMR frameworks that are largely based on English. Third, it advances the current state of the UMR 2026 multilingual shared task¹ by contributing Korean data with more fine-grained annotations of quantification, including event distributivity and lexical or scope ambiguities in Korean.

The adoption of AMR for annotating quantification in Korean is motivated by both practical and theoretical considerations. From a practical perspective, AMR has gained wide acceptance in the

¹See the homepage of the shared task (<https://ufal.mff.cuni.cz/umr-parsing/submission>) for instructions on how and where to submit a system output (as well as where to download the blind input).

computational linguistics community and scales well to large annotated corpora (Bos, 2016).

From a theoretical perspective, recent work, such as Bos (2020) and Pustejovsky et al. (2019), has shown the potential of AMR and UMR to encode quantificational phenomena in a principled yet compact way (Lee et al., 2025; Bunt and Lee, 2025). Focusing on events and their argument or adjunct structures, AMR performs most of the linguistic process of annotating data, while Bos (2020) and UMR (UMR, 2022) introduce discourse constraints on coreference, temporal or modal relations, and relative scopes of quantifiers and events to interpret AMR-based linguistic annotation structures.

These approaches demonstrate that quantificational meanings can be captured without relying on heavy meta-theoretic machinery or extensive inventories of categorial features, making them particularly suitable for large-scale annotation. These observations are relevant at the level of human annotators and programmers, and may only be insignificant at the level of computational operations. Nevertheless, the coverage of AMR at the current stage is not rich enough to accommodate the complex mechanism of quantification in natural language, as treated in the current version of quantML (ISO, 2025).

It should be noted at the outset that the primary aim of this paper is to lay a preliminary basis for inviting shared work for developing the AMR-based construction of a Korean semantic database that allows the direct use of Hangul script in annotation, as is done extensively for Chinese with its complex but simplified Han-script. It thus postpones the task of formally introducing the general semantics of Korean quantification, even though more than a dozen excellent works on the subject exist; to cite one, see Lee (2018). We refer to them in depth as we treat particular relevant topics on Korean quantification in future projects.

The paper develops as follows: Section 2 outlines some issues in Korean quantification, chosen somewhat arbitrarily for this preliminary work; Section 3 illustrates how Korean data are annotated in AMR; Section 4 shows how AMR-annotated data is converted to logical form for interpretation. Here, we discuss how to treat various discourse constraints on scope ambiguity and other semantic phenomena. Section 5 sketches how to construct an AMR-based Korean semBank, and Section 6 concludes the paper.

2. Issues in Korean Quantification

2.1. Preamble

AMR 1.2.6 Specification (Banarescu et al., 2019) states: “AMR does not have a deep representation

for quantifiers, but only *canonicalizes* their position.” This paper, however, treats scope and other contextual constraints on variables related to quantification by extending the predicate-core level to another level, called *document level*, introduced by UMR (UMR, 2022). Following Lee et al. (2025), we call this level *discourse level*, treating scope, coreference, and other discourse constraints that have been discussed by Bos (2020) on variables, introduced at the predicate level of AMR.

As in English, Korean exhibits several challenges related to quantifiers, which may belong to nominal (e.g., “everyone”), adjectival (e.g., “all men”), or adverbial (e.g., “all done”) categories. This paper addresses three such challenges. First, we propose a semantic classification of Korean quantifiers, as illustrated in 2.2. Second, we examine the lexical ambiguity of the universal quantifier \forall *all*, as shown in 2.2.3. Finally, we analyze issues of scope and distributivity, focusing on quantifiers such as 씩 *ssik* ‘each’. The semantics of Korean quantification, discussed here, follows QuantML (ISO, 2025), which makes use of Discourse Representations (DRSS) as defined in DRT (Kamp and Reyle, 1993) to represent logical forms in Neo-Davidsonian semantics (Davidson, 1967; Parsons, 1990), apart from the frequent use of second-order representations as in QuantML. The use of DRSS is preferred over the use of predicate-logic formulas, which are semantically equivalent, since they are more easily derived from the annotation structures of the QuantML abstract syntax.

2.2. Taxonomy of Korean Quantifiers

2.2.1. Overview

The semantic classification of Korean quantifiers is not much different from that of other languages. Notable differences, however, are shown in the morpho-syntactic variations of Korean quantifiers. These differences thus do not affect their classification, except for how they may be implemented computationally or how they have developed historically. Nevertheless, we briefly introduce Korean quantifiers for references in the remaining part of this paper.

Based on the official eGovernment-supported open media Korean dictionary (kDictionary)², 우리말샘 (Urimalsaem, 2020), we classify quantifiers in Korean roughly into three types: universal, polarity-sensitive, and existential quantifiers.³

²<https://opendict.korean.go.kr>

³In this paper, we adopt the Korean standard romanization system, adopted for computational processing, instead of the Yale romanization of the Korean characters Hangul 한글, more often used in linguistic journals abroad.

2.2.2. Classification of Quantifiers (tentative)

1. Universal Quantifiers:

다_{003adv,007noun} da, "all"

다_{004adv} da,

"reaching the end of an action or a state"

Other items: 모두 modu, 모든 modeun.

(a) Distributive universal (Choe, 1987):

-씩_{003sx} ssik, "all each"

"suffixed to a quantitative word in a distributive or divisive sense",

Other items: 각각 gakgak, 마다 mada, 매 mae.

(b) Partitive Universal:

온_{001adv} "whole", "entire"

E.g.: 온 나라 "the whole (all parts of the country)"

(c) Free choice universal:

누구(나)_{003,pro}, nugu(na), "any"

아무(나)_{003,det}, amu(na), "any".

2. Polarity sensitive:

누구-도₀₀₁ nugu-do, "neg ... any"

아무(도)_{004,det} amu(do) "neg ... any"

Note: 누구 nugu and 아무 have other senses, "who" and "someone", respectively, which are free of polarity.

3. Existential:

• Logical existential: Lee (2002)

적어도 하나 이상 jeogeodo hana isang "at least more than one"

어떤 oddeon, 모 mo "a certain", "some"

• Scalar existential:

많다 manh-da, 많은 manh-eun "many"

적다 jeog-da, 적은 jeog-un "a few" or "a little"

대부분 daebubun "most"

• Numeric existential:

하나 hana, 한 han "one"; 둘 "dul", 두 "du" "two"; 셋 ses 세 se "three"; etc.

The universal type has three subtypes: distributive -씩 (ssik), partitive 온 (on), and free choice 누구-나 (nugu-na). The existential type has logical, scalar, and numerical types. These quantifiers also have subtypes, depending on whether they are associated with countables or uncountable mass.

2.2.3. Senses of the Quantifier 다 da

The quantifier 다 da has at least three senses:

- (1) • 다₀₀₃ Adnominal (floating): "all"
- 다₀₀₄ Adverbial: meaning an action or state reaching its limit.⁴

⁴This sense is controversial, as to whether it should be treated as a quantifier.

- 다₀₀₇ Noun: "All came"

Quantifiers in general may quantify over either individual entities (e.g., 손님들 guests) or events (e.g., 왔다 came, 떠났습니다 leave). Here are examples:

- (2) a. 모든₀₀₃ 손님들이 경주에 왔다.
modeun sonnim-deul-i gyeongju-e wa-ss-da
all guest-PI-Nom city-Loc come-Pst-Decl
"All the guests came to Gyeongju."
 $\forall x[guest(x) \rightarrow \exists e[arrive(e), past(e), agent(e, x)]]$
- b. 다(들)₀₀₇/모두가 떠났습니다.
da(deul)/modu-ga tteona-ss-seumni-da
all-PI-Nom leave-PI-Honorific-End
"All left."
 $\forall x\exists e[person(x), arrive(e), past(e), agent(e, x)]$
- c. 기차가 경주역에 (거의) 다₀₀₄/*모두 왔어.
train-Nom Gyeongju-yeog-e da/*modu wa-ss-eo
"The train (almost) arrived at the Gyeongju Station."
 $\exists\{x, y, e\}$
 $[train(x), gyeongju(y), reach(e), past(e), goal(e, y)]$
- d. 회의가 다₀₀₄ 끝났어.
hwe.eui-ga da ggunna-ss-eo
meeting-Nom all end-Pst-Ending
The meeting all ended.
 $\exists e[meeting(e), end(e), past(e)]$

Examples (2a,b) illustrate how the individual entities are quantified by the universal quantifier. In Example (2c,d), the two occurrences of 다₀₀₄ are treated as quantifying over the events, as quantification is extended to events.

2.2.4. Scope

- (3) 동네 총각들이 여자한테 다들 미쳤다.
dongne chongak-deul-i yeoja-hanthe da-deul michyeo-ss-da
village bachelor-PI-Nom woman-Dat all-PI get-mad-Pst-Decl
"All the bachelors in the village got mad at a woman/women."

This example is interpreted either as indicating that there's a specific woman that all the bachelors got mad at, or that each of them got mad at some woman, not necessarily at the same woman.

2.2.5. Case-dependent Meaning Differences

Case markers such as the topic marker "-는 -neun" or /"-은 -eun" and the nominative marker "-가 -ka" or "이 -i" differentiate meanings. Consider:

- (4) a. 개가 짖는다. gae-ga jin-neun-da
dog-Nom bark-Pres-Decl
"A dog is barking (progressive)".
b. 개는 짖는다. gae-neun jin-neun-da
dog-Topic "Dogs bark."

Unlike (a), Sentence (b) with the subject marked with a topic marker has a *generic* reading. The verb in (a) is interpreted as progressive, whereas the verb in (b) is interpreted as present.

3. Annotating Korean Data in AMR

We annotate some of the examples given in Section 2 in the graph-based PENMAN format in AMR according to the guidelines in [Banarescu et al. \(2019\)](#). We, however, make some necessary modifications such as treating plurals or quantification. Note that AMR annotates raw data without any morphological or syntactic analysis. Note also that, as has been applied to Chinese large language data, AMR processors as well as the Python programming language are designed to accommodate Unicode-defined characters, including Chinese Han characters or Korean Hangul characters.

3.1. Existentials

- (5) ::Snt-1 개가 짖는다.
::Interpretation: A dog is barking.
::id=dog1-2026-03-29
- ```
(b / 짖다 bark
 :agent (d / 개 dog)
 :aspect Progressive)
```
- (6) ::Snt-2 개 두 마리가 짖고 있다.  
::Interpretation: Two dogs are barking  
::id=dog2-2026-03-29-KL
- ```
(b / 짖다 bark
  :agent (d / 개 dog
          :quant 2))
```
- (7) ::Snt-3 여자 셋이서 맥주 한 캔씩 마셨다.
::Interpretation: There were three women.
They each drank a can of beer.
::id=woman3-2026-03-29
- ```
(d / 마시다 drink
 :agent (w / 여자 woman
 :quant 3)
 :theme (b / 맥주 beer
 :quant 1
 :unit can))
```

#### 3.2. Universal Quantifiers

There are three types of universal quantifiers: nominal, adnominal (adjectival), and adverbial. However, AMR does not differentiate them.

- (8) ::snt: 세 사람 모두가 왔다.  
:: interpretation: All of (the) three men came.  
:: id:8, 2025-12-28

```
(c / 오다-01 come-01
 :: interpretation: \\
 All of three men came.nt 3
 :quant A))
```

The universal quantifier is annotated as a logical constant  $A$ , like the negative polarity  $-$ : e.g.,  $:quant A$ , and numeric quantities like 3 are treated as constants and annotated as  $:quant 3$ .

#### 3.3. Negation and Polarity-sensitive Quantifiers

Korean has two types of negation: a short form and a long form. The short form 아니 "ani" or 안 "an" is adverbial like the English *not*, modifying a verb. The long form *않-다* "anha" or *아니하-다* "anihada" is an auxiliary verb requiring the stem of a main verb ending in a suffix *-지* "-ji". This distinction is ignored in AMR-based annotation.

- (9) a. 친구가 안 왔다.  
chingu-ga an wa-ss-da  
friend-NOM NEG come-Past-Declarative  
"A friend did not come."  
b. 친구가 오지 않았다.  
chingu-ga o-ji anhae-ss-da  
friend-Nom come-COMP NEG-Past-Decl  
"A friend did not come."
- (10) :: Snt: 친구가 안 왔다/ 오지 않았다.  
:: Interpretation: A friend didn't come.  
:: id=neg1-2026-03-29
- ```
(c / 오다-01 come
  :agent (f / 친구 friend)
  :polarity -)
```

The polarity $-$ above negates the event 오다 "come". The universal quantifier 아무도 amudo is polarity sensitive, as in Example (11).

- (11) :: snt: 아무도 나를 돕지 않았다.
:: interpretation: No one helped me.
:: id: k0011-2025-12-28
- ```
(h / 돕다-01 help
 :agent (p / 사람 person
 :quant
 (ex / E
 :polarity -))
 :theme (m / 나 me))
```

ex / E stands for the existential quantifier, just as A stands for the universal quantifier.

(12) :: snt: 개 한 마리도 나한테 짖지 않았다.  
 :: interpretation: No dog barked at me.  
 :: id: k0012, 2025-12-28

```
(b / 짖다-01 bark
 :agent (d / 개 dog
 :quant
 (q / quantity
 :polarity -))
 :theme (m / 나 me))
```

(13) :: 손님들이 아무도/한 분도 오지 않았다.  
 :: interpretation: None of the guests came.  
 :: id: k0013, 2025-12-28

```
(c / 오다-01 come
 :agent (g / 손님 guest
 :quant (p / Pl
 :polarity -)))
```

*Pl* stands for plurality. As shown, many syntactic details are ignored in the annotation to keep AMR annotations simple and straightforward.

### 3.4. Senses of Adverbial 다 da

The adverbial form 다, which is generally treated as a universal in Korean, may also carry the sense of expressing the state of reaching the completion of an action or motion.

(14) :: Sentence: 기차가 서울에 다 왔습니다.  
 :: Interpretation: The train reached/all arrived in Seoul.  
 ::: id: k0014, 21026-02-25

```
(c / 오다-01 come
 :agent (t / 기차 train)
 :goal (s / Seoul
 :reached yes))
```

## 4. Interpreting AMR-annotated Korean Data

### 4.1. Conjunctive Logical Forms

As it is programmed, AMR is designed to automatically produce the graph-based conjunctive logical forms corresponding to the PENMAN format in annotating data. Consider Annotation (5) in the PENMAN format, copied here:<sup>5</sup>

(15) ::Snt-1 개가 짖는다.  
 ::Interpretation: A dog is barking.  
 ::id=dog1-2026-03-29

```
(b / 짖다 bark
 :agent (d / 개 dog))
```

<sup>5</sup>aspect is ignored here because the current version (Banarescu et al., 2019) does not annotate the progressive aspect.

This just says there is a dog that barks, but nothing else. AMR is rich enough to specify various meaningful attributes of a dog mentioned, as in QunatML.

This PENMAN annotation can be converted to a conjunctive logical form, as below:

(16)  $\{b, d\}$   
 $[i(b, \text{짖다 } bark), i(d, \text{개 } dog), agent(b, d)]$

Here, there are two variables  $\{e, d\}$  which may be bound by an existential quantifier  $\exists$ . The comma (,) stands for the logical connective  $\wedge$ , thus forming a sequence of conjunctive logical forms.  $i$  stands for a logical relation that instantiates a semantic concept, such as an event or a property, while relating a variable to it.  $i(b, \text{짖다 } bark)$ , for instance, stands for  $instance(b, \text{짖다 } bark)$ , where the relation *instance* instantiates the event 짖다 of barking and  $b$  is its variable.

Consider another example, copied from (17), dealing with a negative polarity.

(17) :: Snt: 손님이 안 왔다/ 오지 않았다.  
 :: Interpretation: The guest didn't come.  
 :: id=neg1-2026-03-29

```
(c / 오다-01 come
 :agent (g / 손님 guest)
 :polarity -)
```

From this annotation, we can derive a sequence of conjunctive logical forms with negation.

(18)  $\{e, x\}$   
 $[i(c, \text{오다 } come), i(g, \text{손님 } guest), agent(c, g), polarity(c, -)]$

Here, the polarity negates the event of a friend's coming.

AMR cannot, however, treat plurality or quantification with its logical conjunction. We thus propose different ways to treat them.

### 4.2. Quantity and Plurality

(19) ::Snt-2 개 두 마리가 짖고 있다.  
 ::Interpretation: Two dogs are barking  
 ::id=dog2-2026-03-29-KL

```
(b / 짖다 bark
 :agent (d / 개 dog
 :quant 2))
```

Here is a sequence of conjunctive logical forms:

(20)  $\{b, d\}$   
 $[i(b, \text{짖다 } bark), i(d, \text{개 } dog), agent(b, d), quant(d, 2)]$

Here,  $quant(x, 2)$  may be defined in set-theoretic terms.

- (21) Definition of Quantity  
Given a countable object  $x$  and a positive numeral  $n$ ,  
 $quant(x, n)$  is defined as:  $[x \in X, |X|=n]$ .

With this definition,  $quant(x, 2)$  can be replaced by (22), as in QuantML(ISO, 2025) below:

- (22)  $[X | |X|=2, x \in X \rightarrow dog(x)]$

Likewise, plurality (PI) can be defined:

- (23) Definition of Plurals  
Given a countable object  $x$  and a positive numeral  $n$ ,  
 $Pl(x, n)$  is defined as:  $[x \in X, |X| \geq 2]$ .

### 4.3. Universal Quantification

Consider:

- (24) :: snt: 개들은 다 짖는다.  
gae-deul-eun da jin-nun-da  
dog-PL-topic ALL bark-Pres-Decl  
"Dogs all bark."  
:: interpretation:  
All dogs bark.  
:: id: k0024, 2026-03-28-KL
- (b / 짖다-01 bark  
:agent (d / 개 dog  
:num Pl  
:quant A))
- (25) Conjunctive Logical Form:  
 $\{b, d\}$   
 $[i(e, \text{짖다 } bark), i(b, \text{개 } dog),$   
 $agent(e, d), quant(d, Pl), quant(d, A)]$

Here,  $b$  and  $d$  are variables in AMR, and treated as discourse referents in DRT. The relation *instance*, which we abbreviate as  $i$ , instantiates each event or object. The comma “,” is a conjunctive logical connective, usually represented by  $\wedge$ , making the entire list the sequence of conjunctive formulas.

The universal quantification  $quant(d, A)$  may also be defined in set-theoretic terms, as in DRS.

- (26) Definition of Universal Quantification  
Given a set  $X$ , such that  $|X| \geq 2$ , of discourse referents, and a property  $P$  or its instantiation  $i(x, P)$  in AMR:  
 $quant(x, A) =_{df} [X | x \in X \rightarrow i(x, P)]$ .

With these definitions, Logical Form (25) is modified as below:

- (27) Conjunctive Logical Form with Universal Quantification (for Individual distributivity):  
 $\{b, d\}$   
 $[i(b, \text{짖다 } bark), i(d, \text{개 } dog), agent(e, x)]$   
-----  
 $\{X\}$

$$[[x=d, x \in X, |X| \geq 2] \rightarrow \{E\} \\ [[e=b, e \in E] \rightarrow agent(e, x)]]$$

Note here that  $x=d$  and  $e=b$  links the blower box of implicational logical forms for quantification to the first block with a sequence of conjunctive forms, created from AMR’s PENMAN format.

In addition to the first block of AMR’s conjunctive logical forms, quantification creates another block of logical forms, as shown in (27). This section mainly concerns how to relate this block to the basic block of AMR’s conjunctive logical forms. At the same time, it treats scope and other semantic phenomena that depend on contextual constraints in discourse.

Like English and other languages, Korean data show how various discourse constraints, for instance, involving coreference and scope, can be represented at the discourse level. Each scopal constraint involving quantification can be represented by incorporating contextual constraints, such as those proposed by Bos (2020), into the document level, renamed *discourse level* (Lee et al., 2025), in UMR. Here are Bos’s contextual constraints, some of which are directly incorporated into UMR’s discourse level:

- (28) 1. identical (=)  
2. negated ( $\neg$ )  
3. implication ( $\Rightarrow$ )  
4. presuppositional ( $<$ )  
5. inclusion (:)

The identity relation (=) represents the identity of a pronoun with its antecedent. The negation ( $\neg$ ) negates the truth-value of a proposition, either by negating events of some properties, both of which are treated as *concepts* in AMR. The conditional ( $\Rightarrow$ ) is used to represent universal quantification. The presupposition ( $<$ ) involves the existential presupposition of names or definite descriptions with determinacy. The inclusion (:) relates contextual constraints. This paper illustrates some of these notions.

### 4.4. Logical Forms

Graph-based AMR annotation structures at the predicate level represent their logical forms as a list of conjuncts, where concepts are instantiated while relations relate the variable of a mother node to its daughters. Based on these conjunctive logical forms, the discourse-level constraints control the derivation of logical forms in first or second-order logic. Here is an example:

AMR produces a logical form corresponding to the PENMAN representation. Below, we produce a

minimal logical form that eliminates the representation of PL and A, which stand for plurality and the universal quantification.

- (29) Minimal Logical Form  
 $i(b, \text{짖다})$   
 $i(d, \text{개})$   
 $agent(b, d)$

Here,  $i$  instantiates semantic concepts like the event 짖다 bark or the property 개 dog. By a contextual constrain, Bos (2020)'s implicational relation ( $\Rightarrow$ ) holds between the property of being a dog and the event of barking in Figure 1.

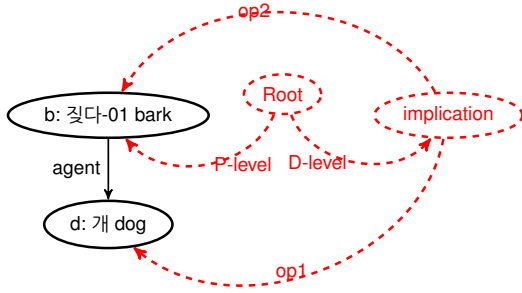


Figure 1:  
Adding the Discourse Level to the AMR Graph

In Figure 1, the `Root` node is introduced as the top node dominating the AMR-based predicate level and the discourse level, marked in red. The *implication* ( $\rightarrow$ ) relation dominates the AMR predicate structure under the root node. The `op1` of the implicational constraint has a wider scope over the `op2`, as represented in the second block of Logical Form (30).

- (30)  $\{b, d\}$   
 $[i(b, \text{bark}), i(d, \text{dog}), agent(b, d)]$   
 -----  
 $\{X, E\}$   
 $[ [|X| \geq 2, x = d, x \in X = ] \rightarrow$   
 $[ |E| \geq 2, e = b, e \in E ] \rightarrow agent(e, x) ] ]$

This logical form is interpreted as asserting that there are dogs, and each of them bark.

#### 4.5. Negative Quantification

Consider:

- (31) :: snt: 개들이 다 안 짖고 있다.  
 :: two interpretations:  
 a. Not all of those dogs are barking.  
 b. All of those dogs are not barking.  
 :: id: k0031, 2026-01-13-KL

Interpretation (a) is annotated as in (32):

- (32) (e / 짖다-01 bark  
 :agent (x / 개 dog  
 :quant (a / A  
 :polarity -)))  
 )))  
 :discourse-level  
 (c / constraints  
 :negation (a / A  
 :polarity -))

Here, the universally quantified 개 “dog” is negated, deriving the following logical form:

- (33) Not all the dogs bark:  
 $[x \neg i(x, \text{dog}) \rightarrow [e | i(e, \text{bark}), agent(e, x)]]$

Interpretation (b) is represented as below:

- (34) (e / 짖다-01 bark  
 :agent (x / 개 dog  
 :quant (a / 다 all))  
 :polarity -))

From this, we derive the following logical form in first-order logic::

- (35) All the dogs do not bark:  
 $[x | i(x, \text{dog}) \rightarrow \neg [e | i(e, \text{bark}), agent(e, x)]]$

## 5. Towards Constructing an AMR-based Korean SemBank

Designing a Korean SemBank (kAMR-SemBank) requires a sophisticated pipeline that can bridge the gap between Korean’s rich morphosyntactic features and the abstract nature of AMR. Unlike English, Korean’s agglutinative morphology, where particles (Josa) and endings (Eomi) carry significant semantic weight, must be pre-processed to ensure accurate predicate-argument structures. Some of them also have morphologically complex derivational structures. Applying AMR-based annotation to data such as this requires time-consuming morpho-syntactic preprocessing to facilitate each annotation step.

In order to achieve this, our proposed design utilizes a multi-source data collection strategy, as depicted in Figure 2.

For correctly and efficiently obtaining a balanced output of the AMR-based Korean SemBank (kAMR-based SemBank) enriched with QuantML for very large data, the proposed pipeline goes through the following three stages:

1. **Structured Resource Integration:** We prioritize the collection of data from the *Sejong Corpus* and the National Institute of Korean Language’s *Modu Corpus* that already contain dependency annotations. Since Korean follows head-final syntax, the use of pre-structured syntactic data significantly reduces

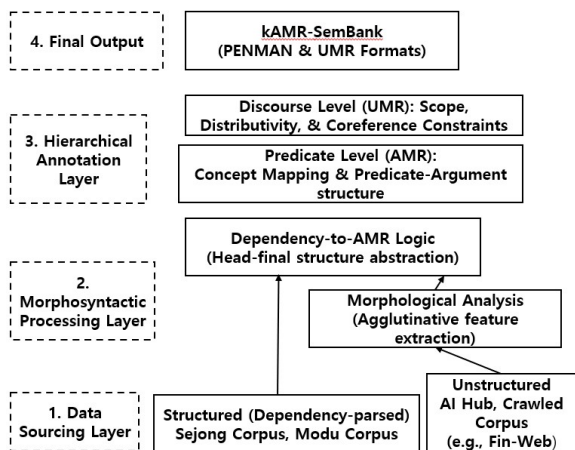


Figure 2: Proposed Pipeline for kAMR-SemBank Construction: Integrating Large-scale Raw Corpora with Morphosyntactic Enrichment

the time required to extract core argument structures for AMR.

2. **Large-scale Raw Corpus Acquisition:** To ensure the diversity and scalability of the SemBank, we incorporate extensive raw corpora from *AI Hub* (covering news, books, and web data) and crawled web-scale datasets such as *Fin-Web*. These resources provide a wealth of contemporary Korean quantification examples in various discourse contexts.
3. **Dual-layered Annotation Process:** The pipeline incorporates a hierarchical approach: a predicate-level for core semantic relations and a discourse-level (as defined in UMR) to handle Korean-specific quantification, distributivity features of events, and scope constraints.

## 6. Concluding Remarks

This paper argues that Korean quantification can be modeled in AMR while benefiting from UMR and Bos (2020)’s discourse-level constraints. We proposed a practical taxonomy of Korean quantifiers and used it to guide consistent annotation. We also treated the multiple senses of  $\square$ , including completion readings that quantify over events. At the predicate level, we normalized Korean surface variation, including nominal, adnominal, and adverbial forms, as well as short and long negation. We marked universals and existentials with logical constants and explicitly represented polarity, which keeps the graphs compact and readable. At the discourse level, we added constraints for presupposition, scope, and coreference, enabling first-order logical forms to be derived in a principled way.

This separation clarifies contrasts, such as *none* versus *not all*, and supports distinctions between *individual* and *collective distributivity* features, although we did not provide an adequate treatment of these features here. Finally, we sketched a corpus-building pipeline that combines structured resources with large-scale raw data and a dual-level annotation workflow. Future work will refine the guidelines and tools, expand coverage across genres, and validate the scheme through annotation studies and shared-task evaluation.

To initiate the project, TeddySum AI will first secure large-scale government funding to organize a shared task for constructing an AMR-based Korean semBank for quantification. The content of the shared work will be enriched by the ISO international standard 246217-12:2025 (ISO, 2025).

## 7. Acknowledgments

Kiyong Lee, the first author, is deeply grateful to the four reviewers for their detailed comments, which substantially improved this paper. He also sincerely thanks his coauthors, Chongwon Park, Harry Bunt, Byonrae Ryu, and Younggyun Hahm, for their patience, generosity, and sustained engagement with difficult issues, including many that extend beyond the Korean data itself. Their criticism, suggestions, and encouragement have been invaluable.

## 8. Limitations

The first author also regrets that he was not able to accommodate all of the reviewers’ suggestions in the present version, especially those concerning the taxonomy of Korean quantifiers. He nevertheless remains deeply appreciative of their thoughtful and constructive comments.

In particular, the first author wishes to express special gratitude to Harry Bunt for his valuable comments on several of the paper’s central theoretical issues, especially the inadequacy of AMR in the treatment of plurals, determinacy, individuation, and distributivity; in the use of DRSS to represent logical form; and in possible ways of relating AMR’s predicate-structure level to UMR’s discourse-structure level. Regrettably, not all of these issues could be covered in the present version of the paper due to limitations of time and space.

## 9. Bibliographical References

Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and

- Nathan Schneider. 2019. *Abstract Meaning Representation (AMR) 1.2.6 Specification*. The DMR Working Group.
- Johan Bos. 2016. *Squib: Expressive power of abstract meaning representations*. *Computational Linguistics*, 42(3):527–535.
- Johan Bos. 2020. *Separating argument structure from logical structure in AMR*. In *Proceedings of the 2nd International Conference on Designing Meaning Representations*, pages 13–20, Barcelona, Spain (online), December 13, 2020.
- Harry Bunt. 2021. *The ISA-17 quantification challenge: Background and introduction*. In *Proceedings of the 17th Joint ACL - ISO Workshop on Interoperable Semantic Annotation*, pages 33–40, Groningen, The Netherlands (online). Association for Computational Linguistics.
- Harry Bunt. 2024. *ISO 24617-12: A new standard for semantic annotation*. In *Proceedings of LREC-COLING 2024*, pages 9361–9371, Turin, Italy. ELRA Language Resource Association: CC BY-NC 4.0.
- Harry Bunt, Maxime Amblard, Johan Bos, Karën Fort, Bruno Guillaume, Philippe de Groote, Chuyuan Li, Pierre Ludmann, Michel Musiol, Siyana Pavlova, Guy Perrier, and Sylvain Pogdalla. 2022. *Quantification annotation in ISO 24617-12, second draft*. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 3407–3416, Marseille, France. European Language Resources Association.
- Harry Bunt and Kiyong Lee. 2025. *The representation of QuantML annotations in UMR – an exploration*. In *Proceedings of the 21th Joint ACL–ISO Workshop on Interoperable Semantic Annotation (ISA-21)*, Heinrich Heine University Düsseldorf, Germany. IWCS 2025.
- Harry Bunt, James Pustejovsky, and Kiyong Lee. 2018. *Towards an ISO standard for the annotation of quantification*. In *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018)*, pages 1787–17949, Miyazaki, Japan. ELRA.
- Hyonsu Choe, Jiyeon Han, Hyejin Park, Tae Hwan Oh, and Hansaem Kim. 2020. *Building Korean Abstract Meaning Representation corpus*. *Proceedings of the Second International Workshop on Designing Meaning Representation*, pages 21–29.
- Jae-Woong Choe. 1987. *Anti-quantifiers and a Theory of Distributivity*. Ph.D. dissertation, University of Massachusetts, Amherst.
- Donald Davidson. 1967. *The logical form of action sentences*. In *The Logic of Decision and Action*, pages 181–120, Pittsburgh. University of Pittsburgh.
- ISO. 2025. *ISO 24617-12:2025 Language resource management – Semantic annotation framework (SemAF) – Part 12: Quantification*. The International Organization for Standardization, Geneva. Project leader: Harry Bunt.
- Hans Kamp and Uwe Reyle. 1993. *From Discourse to Logic: An Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and DRT*. Kluwer, Dordrecht.
- Kevin Knight, Bianca Badarau, Laura Baranescu, Claire Bonial, Madalina Bardocz, Kira Griffitt, Ulf Hermjakob, Daniel Marcu, Martha Palmer, Tim O’Gorman, and Nathan Schneider. 2020. *Abstract Meaning Representation Release 3.0*. Linguistic Data Consortium, Philadelphia, PA. Technical Report LDC2020T02.
- Chungmin Lee. 2002. *(in)definites, case markers, classifiers and quantifiers in Korean*. *Harvard Studies in Korean Linguistics*, 3:469–487.
- Eunhee Lee. 2018. *Korean Syntax and Semantics*. Cambridge University of Press, Cambridge, UK.
- Kiyong Lee, Harry Bunt, James Pustejovsky, Alex C. Fang, and Chongwon Park. 2025. *Representing ISO-annotated dynamic information in UMR*. In *Proceedings of the Sixth International Workshop on Designing Meaning Representations*, pages 49–58, Prague, Czechia. Association for Computational Linguistics.
- Terence Parsons. 1990. *Events in the Semantics of English: A Subatomic Semantics*. The MIT Press, Cambridge, MA.
- James Pustejovsky, Nianwen Xue, and Kenneth Lai. 2019. *Modeling quantification and scope in abstract meaning representations*. In *Proceedings of the First International Workshop on Designing Meaning Representation*, pages 28–33. Association for Computational Linguistics. Florence, Italy, August 1, 2019.
- Working Group UMR. 2022. *Uniform Meaning Representation (UMR) 0.9 Specification*. UMR Working Group for Guidelines.