Building Korean Abstract Meaning Representation Corpus

Hyonsu Choe, Jiyoon Han[†], Hyejin Park[◇], Teahwan Oh[◇], Hansaem Kim[†]

Language AI Lab, NCSOFT Corp. / Seongnam, South Korea

[†]Institute of Language and Information Studies, Yonsei University / Seoul, South Korea

[◇]Department of Korean Language and Literature, Yonsei University / Seoul, South Korea

choehyonsu@ncsoft.com, {clinamen35, hjp1010, ghksl0604, khss}@yonsei.ac.kr

Abstract

To explore the potential sembanking in Korean and ways to represent the meaning of Korean sentences, this paper reports on the process of applying Abstract Meaning Representation to Korean, a semantic representation framework that has been studied in a wide range of languages, and its output: the Korean AMR corpus. The corpus which is constructed so far is a size of 1,253 sentences and its raw texts are from ExoBrain Corpus, a state-led R&D project on language AI. This paper also analyzes the result in both qualitative and quantitative manners, proposing discussions for further development.

1 Introduction

This paper aims to prepare the annotation guideline for applying Abstract Meaning Representation (AMR)—the annotation framework to represent the meaning of the sentence in a graph structure—to Korean and construct a corpus based on it as well as to explore the possibility of new Korean semantic annotation methodologies and language resources.

Until now, the field of Korean natural language processing attempts to construct a corpus targeted to specific phenomena such as polysemy disambiguation, Named entities, temporal and spatial information and semantic roles. However, given the complex relationship between the expression and its meaning, integrated meaning representation system and language resources that far exceed the fragmented analysis are necessary.

In response, this study creates annotation guidelines covering sentence structure and grammatical phenomenon in Korean and describes the following result, a Korean AMR corpus with a size of 1,253 sentences. Korean AMR uses predicate frames from Korean PropBank (Palmer et al., 2006) to represent the relation between events and concepts. Parts of the procedures are automated by using feature annotation form preconstructed parallel corpus. As a result, the search cost has been reduced and the annotation process has been simplified. Section 4 will analyze the corpus in qualitative and quantitative perspectives and provide insight for further research.

2 Related Works

Ever since AMR was first proposed in Banarescu et al. (2013), efforts to construct corpus have continued in English-speaking nations. *The Little Prince Corpus*, containing 1,562 sentences, has been continuously providing the foundation for multilingual AMR research; *Bio AMR Corpus*, comprising 6,952 sentences, is known for proving its applicability in the biomedical domain.[1] In particular, the release of *Abstract Meaning Representation (AMR) Annotation Release 3.0* (Knight et al., 2020) demonstrates how the English AMR managed to enter a stable phase.

AMR research in non-English-speaking nations has begun its expansion; recent years have seen concrete developments in corpus construction in various parts of the world. In 2014, 100-sentence-sized Chinese and Czech AMR corpus was first built for basic research on multilingual AMR annotation.

---

[1]https://amr.isi.edu/download.html

(Xue et al., 2014) In 2016, Annotation Specification for Chinese AMR (CAMR) and the Chinese version of *The Little Prince Corpus* was released. (Li et al., 2016) According to Wang et al. (2018) and Li et al. (2019), the currently known size of Chinese AMR Bank is 10,149 sentences; such a size can be said to approach the level that allows a significant parsing ability. Follow-up efforts have since continued, including Song et al. (2020), which attempted to expand predicate lexicon in order to improve the Chinese AMR.

Research efforts based on *The Little Prince Corpus* act as a 'calling water' in the multilingual AMR studies. In Spanish, a basic research was performed on how to represent grammatical phenomena that are characteristic of individual languages based on AMR (Migueles-Abraira et al., 2018); In Brazilian Portuguese, *The Little Prince Corpus* (AMR-BR) was constructed (Anchiêta and Pardo, 2018), laying the foundations for the construction of the general-purpose corpus. (Sobrevilla Cabezudo and Pardo, 2019) AMR-based studies that aim to develop semantic representation resources have been done in Vietnamese (Linh and Nguyen, 2019) and Turkish (Azin and Eryiğit, 2019) as well.

In Korean, preliminary studies have been done to illustrate some grammatical phenomena that are unique to Korean, such as Copula construction, its negation, and case-stacking, and address how they can be represented within the grammatical system of AMR. (Choe et al., 2019a) Further studies have developed a Korean Guideline v1.0, thereby laying the foundations for corpus construction. (Choe et al., 2019b) This study expands upon the aforementioned studies by discussing the process and outcomes of the construction of Korean AMR corpus.

3 Towards Korean AMR Corpus

3.1 Guidelines for Korean AMR Annotation

In order to construct Korean AMR corpus, above all, a guideline that is applicable to the Korean language must be set up. Choe et al. (2019a) has already proposed specific but partial annotation guidelines, such as Copula Construction, its negation, and Case Stacking, that must be supplemented and reinforced to better apply AMR in Korean. After that, based on those guidelines, Korean AMR Guideline v1.0 has been developed and released.[2] (Choe et al., 2019b)

Korean AMR guideline was prepared by taking into account a wide array of elements, including morphological and syntactic characteristics of Korean. While its basic structure is derived from that of English AMR, specific representation guidelines regarding some grammatical phenomena such as negation, modality, named-entity representation, and their examples were significantly reinforced. Below demonstrates some of the features of Korean AMR.

Using Korean PropBank – Korean AMR annotation is based on the verb frames of Korean PropBank. (Palmer et al., 2006) This is the result of taking into account that in several languages such as Chinese, Brazilian-Portuguese and Vietnamese AMR is constructed based on PropBank-related language resources and that criteria for determining semantic roles in Exobrain wiseNLU[3], a state-led R&D project, are also based on PropBank style. (Lim et al., 2015) Therefore, such an annotation system may be advantageous in the context of compatibility between different language resources and multilingual research.

Adopting special frames and entities of English AMR as metalanguage – Usage of special frames and entities defined in English AMR will be maximized in Korean AMR corpus. Special frames and entities that can both allow more intuitive annotation (e.g., `rate-entity-91` and `have-org-role-91`) and normalization of meaning will be widely accepted. Reifications such as `cause-01` and `exemplify-01` that correspond to certain relations will also be accepted as the grammar of Korean AMR. These decisions are especially useful in circumstances in which exact correspondence between verb frames cannot be provided due to differences between the two languages. For instance, deontic modality in Korean does not have a lexicalized unit such as the English verb "obligate" but is rather expressed via periphrastic construction such as "-야 하/되-"(*have to (be)*). In cases like this, English verb frame such as `obligate-01` would be used. Such a system would be applied to

---

22

the counting particles (`:unit`) of `X-quantity` and `X-entity` in the same way, thereby normalizing types of meaning.

Coverage of `:polarity` – Syntactic negation in Korean can be categorized into 1) short-form negation that utilizes adverbial negators '안'(*not*) and '못'(*cannot*) and 2) long-form negation that utilizes auxiliary verb '아니하-'(*do not*), '않-'(*don't*) and '못하-'(*cannot do*). In addition, negation of copula '-이-'(copula*be*') uses a lexicalized '아니-'(*be not*). In imperative, '말-'(*desist*), a deontic modal negator, is used to realize the negation. In Korean AMR, `:polarity` is annotated to these basic negators as well as negative prefix such as '비-', '불-', '미-', '무-'(corresponding to *non-, un-, im-, ir-, il-, dis-*) and negative verbs such as '없-'(*not exist*; antonym of adjective '있-', which means *exist*), '모르-'(*not know*; antonym of verb '알-', which means *know*). (e.g. "*X를 모른다.*" → 알-01 `:polarity` - `:ARG1 X`)

Expanded usage of `:domain` for case-stacking – Korean is a topic-subject prominent language; there are Multiple Nominative Construction, where the nominative marker '-이/-가/-은/-는' is licensed to two or more components, and Multiple Accusative Construction, where the accusative marker '-을/-를' is also licensed to two or more components. Although it is desirable in AMR that relational structure is annotated based on the roleset of an appropriate verb frame and not on the syntactic structure, components that are of topical focus are annotated as `:domain` when it is difficult to determine the semantic role between the constituents. (The detailed discussion on case-stacking in Korean and its representational choice can be found in Choe et al. (2019a).)

Restricted set of NE Types – Current criteria for NE annotation allows for the selection of arbitrary NE Type rather than canonical type based on context. In Korean, the list of basic named-entity types is limited to that presented in English AMR; English vocabulary is adopted when there is a need for expansion. This effectively introduces meta-language in NE annotation and can be helpful for regularized annotation. Furthermore, difficulties in determining canonical form were resolved by annotating wikification for all NE representations within corpus.

3.2 Annotation Overview

3.2.1 Source Texts

The Korean AMR Corpus is composed of texts that are released to the public for research purposes, including a Korean-version text of ⟨*The Little Prince*⟩,[4] example sentences of verb entries in the Basic Korean Dictionary[5], and subset of ExoBrain Corpus v4.0[6]. Most of the texts are from ExoBrain corpus, which is comprised of 19.5M sentences of news texts, Korean Wikipedia documents and web texts. Corpus is highly useful in that it contains various layers of annotation information such as, dependency structure, word senses, NE and semantic roles. Each layer of annotation guideline prepared for the construction of ExoBrain corpus is gradually being accepted as a standard method for automatic analysis of Korean; the analysis engine whose learning is based on a vast amount of data is provided in the form of open API so that Korean NLP researchers and technologists can use it.

3.2.2 Korean PropBank

Korean PropBank (Palmer et al., 2006) comprises 33,295 predicates tokens from Virginia Corpus (54.5K words), Newswire Corpus (131.8K words) and 2,749 Verb Frames. Through the ExoBrain project, Korean PropBank became a customary standard for Korean semantic role labeling, and further research endeavors including an expansion of verb frames followed. (Bae and Lee, 2015) Taking into account that multilingual PropBank resources are already established in diverse languages and that AMR annotation based on this is in trial, Korean AMR corpus was also annotated based on Verb Frames of Korean PropBank. For the Verb Frames of Korean PropBank are limited in size (2,749), additional guidelines are needed when annotating unlisted predicates; therefore, while synonymous predicates are selected when annotating for the unlisted predicates, annotations are made based on the roleset of the verb frame that has similar case-frame information and valency required by the predicates.

---

[4]Text source: *The Little Prince Collection* (`https://phasis68.blogspot.com/2017/03/korean.html`)

[5]Provided by National Institute of Korean Language. (`https://krdict.korean.go.kr/`)

[6]Provided by ETRI. (Electronics and Telecommunications Research Institute) (`http://aiopen.etri.re.kr/service_dataset.php`)

Sentence to be converted:

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

ExoBrain API
morph, NE features

Sub-graph
Conversion

Korean PropBank   Wikipedia API (ko-kr)

Lemma	Subgraph snippets for predicates
앓.V (suffer)	(앓 / 앓-01 :ARG1 thing_suffering :ARG1 suffer_from)
전하.V (report)	(전 / 전하-01 :ARG0 agent :ARG1 thing_reported :ARG2 report_to)

Span	Subgraph snippets for NE
"뉴욕타임스" OGG_MEDIA	(c / company|organization :wiki "뉴욕_타임스" :name (이 / 이름 :op1 "뉴욕타임스"))
"그레이엄" PS_NAME	(p / person :wiki "그레이엄?" :name (이 / 이름 :op1 "그레이엄"))
"목사" CV_OCCUPATION	:ARG0-of (h / have-org-role-91 :ARG1 :ARG2 (목 / 목사))
"전립선암" TMM_DISEASE	(d / desease :wiki "전립선암" :name (이 / 이름 :op1 "전립선암"))
"파킨슨병" TMM_DISEASE	(d / desease :wiki "파킨슨병" :name (이 / 이름 :op1 "파킨슨병"))

Korean AMR Guidelines & Wikipedia

Human examination
⇩
Merge subgraphs manually

Result:

(전 / 전하-01
    :ARG0 (c / company
        :wiki "뉴욕_타임스"
        :name (이 / 이름 :op1 "뉴욕타임스"))
    :ARG1 (앓 / 앓-01
        :ARG0 (p / person
            :wiki -
            :name (이2 / 이름 :op1 "그레이엄")
                :ARG0-of (h / have-org-role-91
                        :ARG2 (목 / 목사)))
        :ARG1 (등 / 등
            :op1 (d / desease
                :wiki "전립선암"
                :name (이3 / 이름 :op1 "전립선암"))
            :op2 (d2 / desease
                :wiki "파킨슨병"
                :name (이4 / 이름 :op1 "파킨슨병")))))
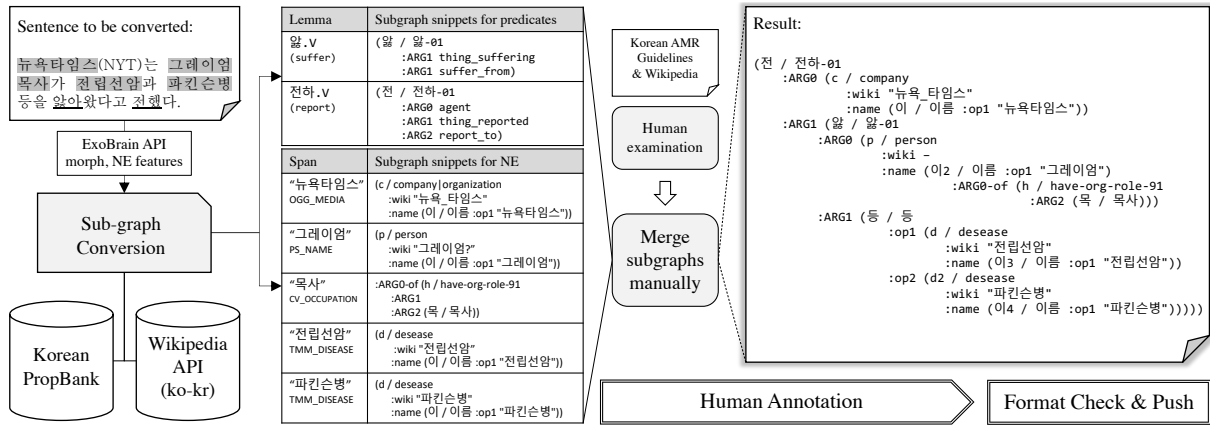
Human Annotation   Format Check & Push

Figure 1: Annotation process: Human annotators take automatically converted subgraph snippets including Korean PropBank frame and roleset, head NE types, name strings and wikification.

3.2.3 Assisted Annotation: Reviewing & Merging Preprocessed Subgraph Snippets

In this study, the corpus was constructed not with AMR Editor (Hermjakob, 2013), a specialized tool for AMR annotation, but rather using text editor with advanced features such as code completion. Such a decision was made to overcome the limitation that AMR Editor was not designed in a way suitable for multilingual annotation as well as reduce the search cost by taking the full advantage of the feature annotation that existing resources have. In the whole annotation process of this study, annotators have completed an AMR graph in a 'assisted bottom-up' process, in which annotator carefully reviewed all the auto-suggested subgraph snippets and merged them manually to the entire AMR graph.

Generating NE subgraphs with API response: When many NEs are included in the raw text, search cost for making AMR increases. To address this problem, substring span of potential NE is used as query of Wikipedia API (ko-kr) to automatically convert NE subgraph, including wikification, thereby reducing the search cost. Without any particular command input, an annotator can examine wikification as well as the type of subgraph that is already created; when required, some can be modified and reflected in the entire graph.

Generating verb frame subgraphs with morphological annotation: Same method of automatic conversion can be applied to verb frame selection as well. Verb Frame of Korean PropBank can be fetched based on verb, adjective, roots of predicative nominals from the morphological analysis result of ExoBrain API response; this verb frame is suggested to the annotator after converted into a PENMAN format subgraph. By doing so, the search becomes simplified and proposing a list of subgraph snippets in advance in such a way can prevent the annotator from omitting certain verbs in long, complicated sentences.

4 Evaluation

4.1 Corpus Specification

The construction process of the Korean AMR corpus can be categorized into trial phase and actual construction phase. In the trial phase, annotators, who were trained based on sentences that are relatively easy to annotate, annotated some parts of the Korean ⟨*The Little Prince*⟩ text and example sentences of verb entries in the Basic Korean Dictionary to verify the feasibility of Korean AMR guidelines. Actual construction was done on a portion of ExoBrain corpus v4.0; The corpus consists of 1,253 sentences was constructed from September 2019 to April 2020. Specific construction sizes of the subgroups are shown below. Constructed corpus was released to the public via online repository.[7]

The constructed Korean AMR Corpus is a size of 1,253 sentences with 20,050 nodes and 18,797 edges. About 6K nodes (30%) represent general concepts, and approximately 3.9K Korean Propbank frames

---

[7]https://github.com/choe-hyonsu-gabrielle/korean-amr-corpus

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%)

Table 1: Organization of the Korean AMR Corpus.

Concepts	Node freq.(%)	Relations	Edge freq.(%)
General Concepts	6,026 (30.1%)	Core roles (`:ARGx/-of`)	6,119 (32.6%)
NE related	3,408 (17.0%)	`:opN` & `:opX`	3,826 (20.4%)
Name span & valid wikification	3,257 (16.2%)	`:name` & `:wiki`	3,032 (16.1%)
Korean PropBank Frames	3,184 (15.8%)	`:mod`	1,469 (7.82%)
Unlisted frames (*-00)	809 (4.1%)	`:time`	595 (3.3%)
Numerics & Scalar	772 (3.8%)	`:location`	395 (2.1%)
Conjunctions	635 (3.2%)	`:manner`	391 (2.1%)
Date-entity & Temporal-quantity	558 (2.8%)	`:quant`	301 (1.6%)
Special frames (*-91/*-01)	455 (2.3%)	`:topic`	251 (1.3%)
Polarity & truth-value	434 (2.2%)	`:poss`	224 (1.2%)

Table 2: 10 most frequent elements in the Korean AMR Corpus.

are used to represent events. Out of the verb frame instances in corpus, 800 instances (554 types) are unlisted entries to Korean Propbank, ending with -00 suffix. About 3.4K (17%) nodes are related to NE representation, including newly added NE Types specific to Korean AMR such as `brand`, `service`, `cultural-asset`, `hospital`, `weapon`.

Out of all relation markers, the core-roles including inverse role accounts for 32.6% in order of `:ARG1/-of`, `:ARG0/-of`, `:ARG2/-of` and the others. As the raw texts contain much information, `:opN` and `:opX`–often used to represent conjunction, enumeration, exemplification, and stretch—are frequently used (20.4%) while `:wiki` and `:name` for NE representation recorded 16.1%. Temporal (`:time`), quantitative (`:quant`, `:unit`) and locative (`:location`) representation are relatively used often. `:ARGA` - `:ARG4`, `:source`, `:location`, `:destination`, `:beneficiary`, `:instrument` for general semantic roles (Saeed, 1997) only accounts for 34.8%.
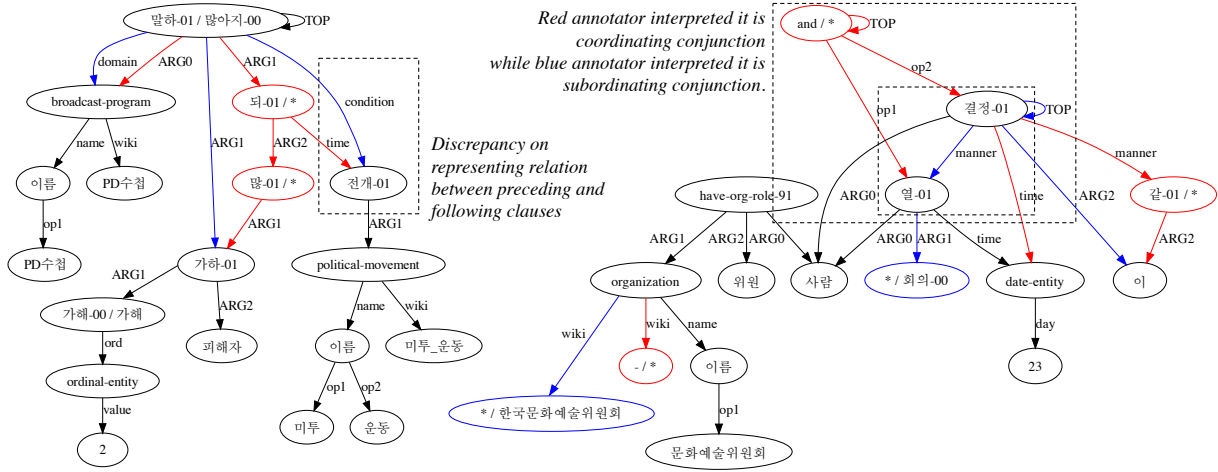
The Inter-annotator agreement of four annotators who participated in building and editing the corpus, based on 50 sentences, reached Smatch 0.79 comparable to the previously reported 0.79 to 0.83 from English or 0.72 from Brazil-Portuguese.

4.2 Disagreement Analysis

4.2.1 Adverbial Clause and Conjunction

Conjunction is a connection between two or more syntactic units (phrases, clauses, sentences and etc) that can be subdivided depending on whether the relationship is subordinate or not, and if so, what the specific semantic relationship is. However, representing the conjunction is not an easy task because distinguishing between the embedded adverbial clauses and (subordinate) conjunction is difficult, and specifying the semantic relation between two clauses is controversial among annotators, resulting in a lack of consistency among the outputs.

A following sentence (on the left part of Figure 2) demonstrates this issue: *"PD 수첩은, "미투 운동이 전개되면서, 피해자들에게 2차 가해를 가하는 것도 많아지고 있다."* (*The PD Note (said), "As Me-Too movement expands, more and more victim-blamings are inflicted to the victims."*). The annotators had

Figure 2: Annotators showed conflicting views on representing conjunction markers in Korean.

conflicting views when representing this sentence: whether the latter clause (led by '-면서,'; *after*) has temporal relation(:time) or logical relation (:condition or else). The controversy is due to the fact that a single conjunction marker can represent multiple meanings.

In the following sentence, (on the right part of Figure 2) *"문화예술위원회 위원들은 23일 회의를 열고 이같이 결정했다."* (*The members of the Culture and Arts Council held a meeting on the 23rd and made the decision.*), there was also a difference between annotators as to whether to represent the clause lead by '-고' with coordinate conjunction (and) or adverbial clause (:manner). In terms of meaning, it is similar to :manner; however, '-고' as well as 'and' can represent both coordinate conjunction (enumerate) and temporal relationship (sequential, simultaneous). Therefore, this disagreement among annotators depends on one's perspective: whether to focus on the form ('-고'; *and*) or on the meaning. Further guidelines with detailed examples of specific conjunctive markers are required to solve this issue.

4.2.2 Collocations

The institutionalised collocation in Korean refers to a case in which different morphemes gather to build a semantic unit. The Korean AMR Corpus annotated these cases through 1) using similar Propbank frames 2) using elements from English AMR. For instance, '-ㄹ 수 있-'(*can / be able to*) that represent modality of possibility can be annotated by using 가능-01 similar to the English way of using possible-01. In contrast, cases with '-아/어야 하-'(*should be*) that represent deontic modality or '-X을 수록 Y하-' that correspond to 'The X-er, The Y-er' constructions can cite obligate-01 or correlate-91 for Korean AMR annotation.

However, the crux of this discussion is that collocations which consist of morphemes with various POS serve diverse semantic functions. For instance, auxiliary verb '하-' from grammatical collocations '-아/어야 하-' function as a part of a unit representing deontic modality, while '하-' in -려(고) 하-'(*in order to / be intended to do*) serve as a part of a unit representing the purpose or intention. Furthermore, '-고는 하-' has a similar meaning to English 'used to', and '-게 하-' functions as English 'make' or 'cause'. Finally, the auxiliary verb '하-' in '-기도 하-' work as a light verb.

Lexical collocations share this complexity. In Korean, a verb '영향받-'(*be influenced*) which is derived by attaching denominal suffix '-받'(*take*) to the noun '영향'(*influence*). However, its relational antonymous expression '영향을 주-' (*give influence*) counts as a collocation. Unlike the antonymous relationship between verb '주-'(*give*) and '받-'(*take*), denominal suffix '-주'(*give*) does not exist in Korean; therefore, annotators did not reach a consensus for representing a sentence such as *"바그너의 음악에 영향을 주었다."* (*It influenced Wagner's music.*) Some annotator used Korean PropBank frame '영향-01' while another used 주-01, considering '주-'(*give*) as a main verb. (See left part of Figure 3.)
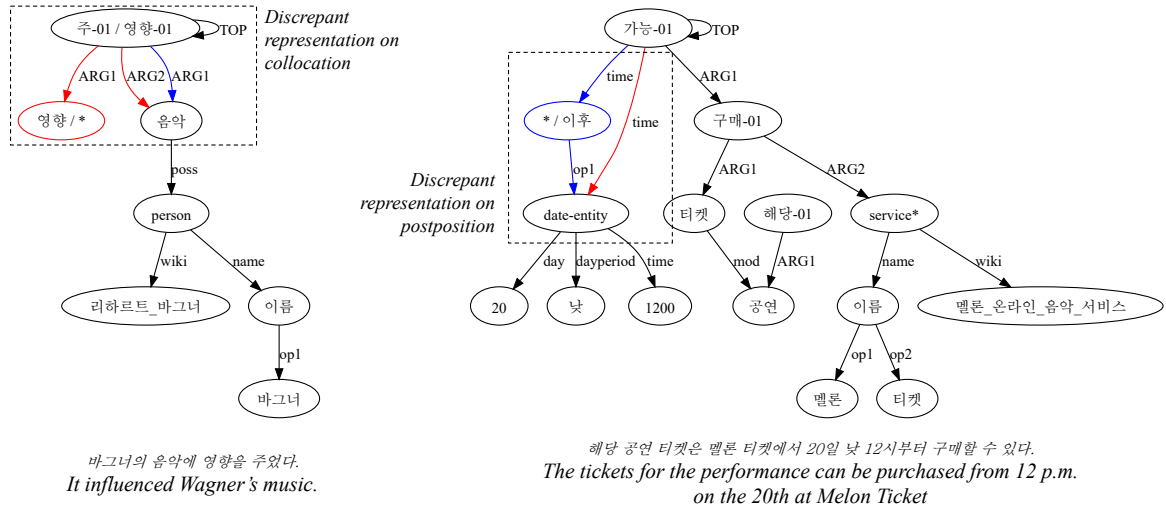
26

Figure 3: Annotators showed conflicting views on representing collocation and postposition.

In Korean, multiple morphemes with various POS often gather to create collocations or periphrastic construction. Without providing detailed guidelines for each of them, it will be troublesome to maintain annotation consistency. The ideal methods would be to obtain a rich collection of examples, such as AMR Dictionary, as well as expanded multi-word predicate frames and rolesets which facilitate intuitive annotations of various constructions in Korean.

4.2.3 Special Postpositions

Postposition of Korean are consists of case postposition, which function as case marker, and special postposition which highlights or adds meaning. Both types of postposition are functional words realized after contents words, establishing postposition cluster. They can be omitted if enough context is provided to interpret the meaning. In such case, case postposition is more frequently omitted than special postposition.

Words that occur independently in other languages including English, (e.g. lexical items such as *'and'* or *'only'*) correspond to special postpositions in Korean. In Korean AMR, these special postpositions are characterized as function words that do not affect the proposition of the sentence. Often, they are omitted in AMR or represented with relations, such as ':postp-는커녕 and :postp-으로-의 because it is not easy to generalize their various uses.

Despite reservations regarding the representation of special postposition, the discrepancy of special postposition interpretation play crucial role in IAA. Special postposition refers to the starting point of the temporal or spatial interval, '-부터'(*from*) illustrates this issue well enough. (See right part of Figure 3.) In "해당 공연 티켓은 멜론 티켓에서 20일 낮 12시부터 구매할 수 있다." (*The tickets for the performance can be purchased from 12 p.m. on the 20th at Melon Ticket*), special postposition '-부터' which combined with temporal expression corresponds to English preposition 'from'. Some annotators considered *'12시 부터'*(*from 12 p.m.*) as a starting point of the specific time interval and following times; therefore, they added '이후' (*after*) during the annotation. However, some simply omitted '-부터' as *'from 12 p.m.'* meaning 'the start of the sale'.

Special postpositions of Korean vary its meaning and importance, thereby creating different representations depending on the intuition of annotators and types of special postposition. For example, annotators reached to consensus when special postposition '-까지' (*til, until / by*) that marks the end point of the time interval is used alone. (e.g. *"9월 25일까지"*(*by 25 September*) → date-interval :op2 (date-entity :month 9 :day 25) while *"내일까지"*(*until tomorrow*) → :time (이전 :op1 내일)). The special postposition words are roughly 30 types in Korean. Even though they are closed set, as multiple special postpositions stacked to add meaning and make syntactic relationships with other grammatical phenomena such as scopal polarity, the generalization of

27

their semantic function is not easy. Therefore, the specific guidelines are needed to capture different usages of special postpositions.

5 Challenges on Korean Sembanking

For Korean AMR to be actually applied to systems that understand and process the natural language, beyond specific tasks such as semantic parsing, there are many challenges to overcome. The Korean AMR corpus presented in this paper is in small size with less specific guidelines. Therefore, it is necessary to set specific goals for both the quality and quantity aspects of data.

The first challenge that Korean AMR is facing is developing several supporting resources for the annotation while simultaneously revising guidelines. To provide more precise meaning representation, ambiguous descriptions in the guideline must be changed to concrete wordings and extensive examples. Moreover, The Korean PropBank should expand its predicates lexicon with enlarged multi-word expression, such as collocation and idioms.

The second challenge is quantitative expansion. To overcome lack of training data, integrated annotation environment that reduces search cost and fosters more intuitive annotation is required, and part of the annotation procedures must be automated to simplify the entire process.

6 Conclusion

This paper introduced the procedure for building a Korean AMR corpus, discussing several issues for further development. The Korean AMR corpus for this study is a size of 1,253 sentences, serving as the first result of foundational studies. Even with small size, the Korean AMR corpus will be the source of empirical clues for developing Korean AMR. It will contribute to laying the foundation for future applications of natural language understanding and generation.

References

Rafael Anchiêta and Thiago Pardo. 2018. Towards AMR-BR: A SemBank for Brazilian Portuguese language. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May. European Language Resources Association (ELRA).

Zahra Azin and Gülşen Eryiğit. 2019. Towards Turkish Abstract Meaning Representation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pages 43–47, Florence, Italy, July. Association for Computational Linguistics.

Jangseong Bae and Changki Lee. 2015. Extending korean propbank for korean semantic role labeling and applying domain adaptation technique. *Korean Journal of Cognitive Science*, 26(4):377–392.

Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract Meaning Representation for sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 178–186, Sofia, Bulgaria, August. Association for Computational Linguistics.

Hyonsu Choe, Jiyoon Han, Hyejin Park, and Hansaem Kim. 2019a. Copula and case-stacking annotations for Korean AMR. In *Proceedings of the First International Workshop on Designing Meaning Representations*, pages 128–135, Florence, Italy, August. Association for Computational Linguistics.

Hyonsu Choe, Jiyoon Han, Hyejin Park, Taehwan Oh, Seokwon Park, and Hansaem Kim. 2019b. Korean Abstract Meaning Representation (AMR) Guidelines for Graph-structured Representations of Sentence Meaning. In *Proceedings of the 31th Annual Conference on Human and Cognitive Language Technology*, pages 252–257, Seoul, Republic of Korea, October. SIGHCLT, Korean Institute of Information Scientists and Engineers. *In Korean*.

Ulf Hermjakob. 2013. Amr editor: A tool to build abstract meaning representations. Technical report, USC Information Sciences Institute.

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 (amr) annotation release 3.0. *LDC Catalog No.: LDC2020T02, ISBN 1-58563-915-X*.

Bin Li, Yuan Wen, Weiguang Qu, Lijun Bu, and Nianwen Xue. 2016. Annotating the little prince with Chinese AMRs. In *Proceedings of the 10th Linguistic Annotation Workshop held in conjunction with ACL 2016 (LAW-X 2016)*, pages 7–15, Berlin, Germany, August. Association for Computational Linguistics.

Bin Li, Yuan Wen, Li Song, Weiguang Qu, and Nianwen Xue. 2019. Building a Chinese AMR bank with concept and relation alignments. In *Linguistic Issues in Language Technology, Volume 18, 2019 - Exploiting Parsed Corpora: Applications in Research, Pedagogy, and Processing*. CSLI Publications, July.

Soojong Lim, Minjung Kwon, Junsu Kim, and Hyunki Kim. 2015. Korean Proposition Bank Guidelines for Exobrain. In *Proceedings of the 27th Annual Conference on Human and Cognitive Language Technology*, pages 250–254, Seoul, Republic of Korea, October. SIGHCLT, Korean Institute of Information Scientists and Engineers. *In Korean*.

Ha Linh and Huyen Nguyen. 2019. A case study on meaning representation for Vietnamese. In *Proceedings of the First International Workshop on Designing Meaning Representations*, pages 148–153, Florence, Italy, August. Association for Computational Linguistics.

Noelia Migueles-Abraira, Rodrigo Agerri, and Arantza Diaz de Ilarraza. 2018. Annotating Abstract Meaning Representations for Spanish. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May. European Language Resources Association (ELRA).

Martha Palmer, Shijong Ryu, Jinyoung Choi, Sinwon Yoon, and Yeongmi Jeon. 2006. Korean propbank. *LDC Catalog No.: LDC2006T03, ISBN 1-58563-374-7*, pages 1–58563.

John Saeed. 1997. I. 2003. Semantics. *GB: Blackwell Publishing*.

Marco Antonio Sobrevilla Cabezudo and Thiago Pardo. 2019. Towards a general Abstract Meaning Representation corpus for Brazilian Portuguese. In *Proceedings of the 13th Linguistic Annotation Workshop*, pages 236–244, Florence, Italy, August. Association for Computational Linguistics.

Li Song, Yuling Dai, Yihuan Liu, Bin Li, and Weiguang Qu. 2020. Construct a sense-frame aligned predicate lexicon for Chinese AMR corpus. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 2962–2969, Marseille, France, May. European Language Resources Association.

Chuan Wang, Bin Li, and Nianwen Xue. 2018. Transition-based Chinese AMR parsing. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 247–252, New Orleans, Louisiana, June. Association for Computational Linguistics.

Nianwen Xue, Ondřej Bojar, Jan Hajič, Martha Palmer, Zdeňka Urešová, and Xiuhong Zhang. 2014. Not an interlingua, but close: Comparison of English AMRs to Chinese and Czech. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 1765–1772, Reykjavik, Iceland, May. European Language Resources Association (ELRA).