

Towards Standardized Annotation and Parsing for Korean FrameNet

<https://github.com/jungyeul/k-framenet>

k-FrameNet

Yige Chen[†] Jae Ihn^{‡*} KyungTae Lim[¶] Jungyeul Park[‡]

[†]The Chinese University of Hong Kong, Hong Kong, [‡]The University of British Columbia, Canada,

[¶]SeoulTech, South Korea. *Currently at Rocketbrew Inc., Canada.

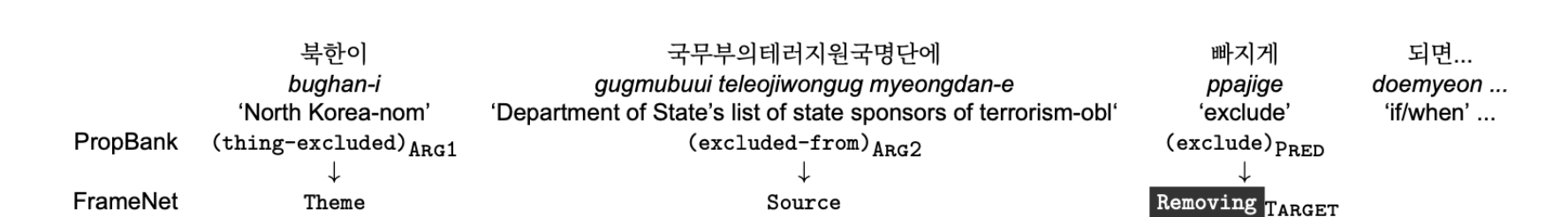


Figure 1: Comparisons between annotations on the same instance in Korean PropBank and the Korean FrameNet dataset. The meaning of the above instance is 'If North Korea were excluded from the Department of State's list of state sponsors of terrorism...', which is part of a sentence in the Korean PropBank.

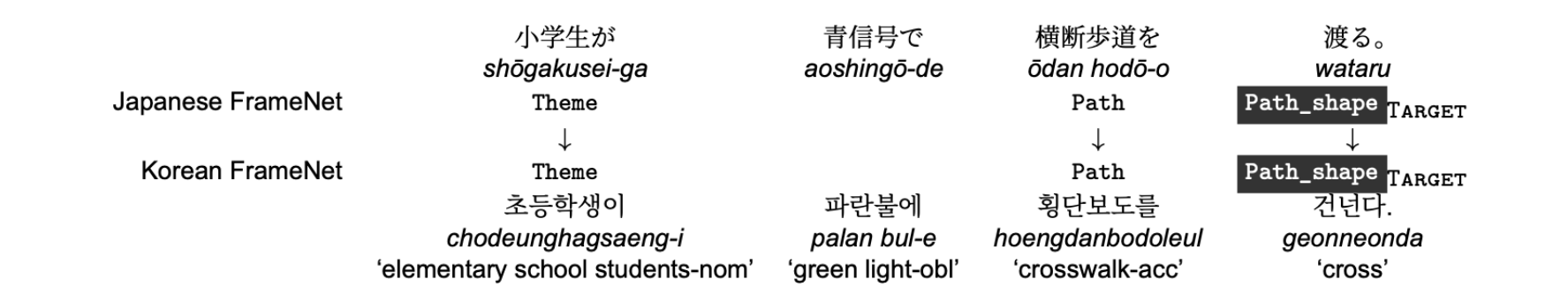


Figure 2: Comparisons between annotations on the same instance in the Japanese FrameNet dataset and the Korean FrameNet dataset. The meaning of the above instance is 'elementary school students cross a crosswalk on the green light'.

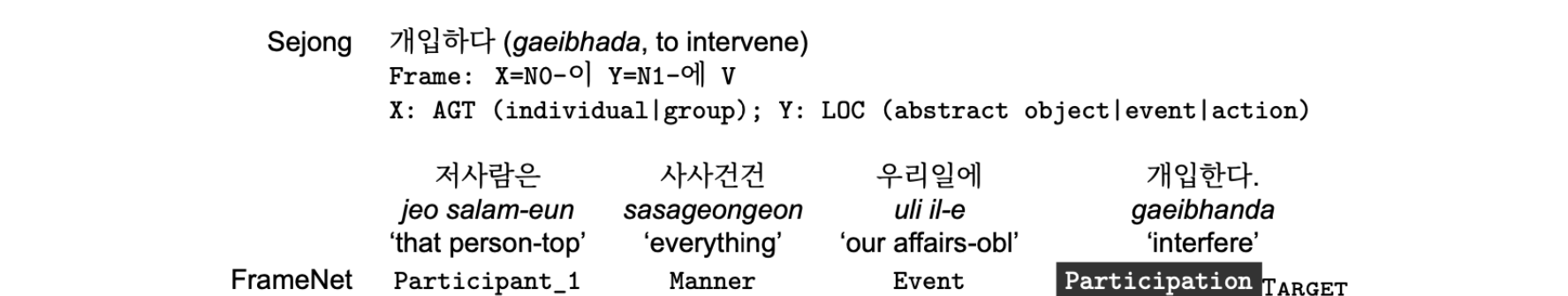


Figure 3: Comparisons between the corresponding information in Sejong Dictionary and the annotation in the Korean FrameNet dataset with regard to a single instance. The meaning of the above instance is "that person interferes in our affairs constantly and meddles in everything".

index	word	lexeme	target	frame	annotation
...					
16	30	30	-	-	B-Time
17-19	여년간	-	-	-	I-Time
17	여	-	-	-	I-Time
18	년	-	-	-	I-Time
19	간	-	-	-	I-Time
20-21	오스트리아를	-	-	-	B-Dependent_entity
20	를	-	-	-	I-Dependent_entity
21	오스트리아	-	-	-	I-Dependent_entity
22-24	통치한	-	-	-	B-FrameTarget
22	통치	-	-	-	I-FrameTarget
23	하	-	-	-	I-FrameTarget
24	은	-	-	-	I-FrameTarget
25-26	좌익이	-	-	-	B-Controlling_entity
25	좌익	-	-	-	I-Controlling_entity
26	이	-	-	-	I-Controlling_entity
...					

Figure 4: Example of the morphologically enhanced FrameNet data: 30yeonyeongan oseuteulialeul jibaehan jwaigi... ('The left wing that ruled Austria for over 30 years...')

Korean FrameNet Dataset

# of LUs	PKFN	JKFN	SKFN
Noun	0	755	0
Verb	644	500	2,252
Adjective	6	155	0
Others	0	14	0
Total	650	1,424	2,252

Distributions of the lexical units (LUs) of the targets

# of frames per LU	PKFN	JKFN	SKFN
Noun	0	1.109	0
Verb	1.183	1.276	1.274
Adjective	1.167	1.290	0
Others	0	1.286	0
Overall	1.183	1.189	1.274

The number of frames per lexical unit

	PKFN	JKFN	SKFN
# of sent	1,767	1,357	5,703
# of inst	2,350	2,919	5,703
# of frames per sent	1.330	2.151	1.000

Numbers of sentences and instances

Conclusion

- We propose a morphologically enhanced scheme to annotate Korean FrameNet datasets, which is motivated by the linguistic features of the Korean language.
- We convert existing Korean FrameNet data into our proposed format through an alignment algorithm, and further train parsers on the standardized morpheme-based data as well as the original word-based data for the comparison purpose.
- The results show that the Korean FrameNet data, once enhanced morphologically, improves the parsing outcomes when using datasets in which annotations are securely inherited from their sources.
- We consider the proposed morpheme-based scheme a standardized way to annotate Korean FrameNet datasets for parsing.

		KoELECTRA-Base			KR-BERT-char16424		
		PKFN	JKFN	SKFN	PKFN	JKFN	SKFN
exact	eojeol	0.2523 ± 0.0215	0.3968 ± 0.0445	0.8091 ± 0.0003	0.2964 ± 0.0229	0.3493 ± 0.0281	0.8041 ± 0.0009
	morph	0.3319 ± 0.0807	0.6528 ± 0.0135	0.6054 ± 0.0056	0.3070 ± 0.0868	0.6256 ± 0.0127	0.5343 ± 0.0042
partial	eojeol	0.3051 ± 0.0224	0.4438 ± 0.0444	0.8279 ± 0.0003	0.3475 ± 0.0226	0.4010 ± 0.0267	0.8241 ± 0.0008
	morph	0.4091 ± 0.0694	0.7152 ± 0.0096	0.7373 ± 0.0047	0.4094 ± 0.0677	0.6929 ± 0.0083	0.6627 ± 0.0036

The cross validation mean \pm standard deviation of exact and partial F_1 scores on *eojeol*- and morpheme-based variants of PKFN, JKFN and SKFN datasets.

Acknowledgement: This work was supported by National Research Foundation of Korea grants funded by the Korean government (Ministry of Science and ICT) (2021R1F1A1063474) for KyungTae Lim.



Download the paper →

