社団法人 電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report

A Contrastive Study on the Adnominal Constructions in Japanese and Korean
－Relative Frequency of ‘-no’ vs. ‘-ui’ －

Jae-Woong CHOE [1] Sachiko SHUDO [2] and Yasunari HARADA [2]

1 Department of Linguistics, Korea University, Anam-dong, Seongbuk-gu, Seoul, 136-701 Korea
2 Faculty of Law, Waseda University 1-6-1 Nishi-Waseda, Shinjuku-ku, Tokyo, 169-8050 Japan
E-mail: 1 jchoe@korea.ac.kr, 2 {shudo, harada}@waseda.jp

Abstract Syntax in Japanese and Korean shows “striking similarities” especially when we consider that the two languages are not directly connected historically the way Spanish and Italian are, for instance. Such similarities also pertain in the way two nouns (phrases) are combined to form a larger nominal expression. However, when we take a careful look into the distribution and relative frequency of adnominal constructions with particles ‘-no’ and ‘-ui’ against noun-noun compounds, we notice an interesting difference between the two languages. Here we report a preliminary result of our survey using a small parallel corpus, corroborated with data from larger monolingual corpora.

Keyword Adnominal, frequency, parallel corpora, Japanese, Korean

日本語と韓国語における名詞句連接の対照研究
― 「の」と「의」の相対的頻度の比較―

崔 在雄 [1] 首藤 佐智子 [2] 原田 康也 [2]

1 高麗大学校言語学科 〒136-701 ソウル市城北区安岩洞
2 早稲田大学法学学術院 〒169-8050 東京都新宿区西早稲田 1-6-1
E-mail: 1 jchoe@korea.ac.kr, 2 {shudo, harada}@waseda.jp

あらまし 日本語と韓国語の統語法を比べてみると、この二言語にヨーロッパ諸言語のような系統的関連性がないと考えられていることからするとある意味で「驚くべき並行性」がいろいろな側面で観察できる。こうした並行性の一つに、名詞（句）と名詞（句）を組み合わせてより大きな名詞句を構成する点も挙げることができる。しかし、日本語の「の」や韓国語の「-ui」を用いた名詞句連接と名詞（句）と名詞（句）の複合表現を比べると、両言語の間に興味深い相違があることに気が付く。本稿では、比較的小規模なタグ付き二言語並行コーパスの分析結果と、それぞれの言語のより大規模なコーパスの調査結果を合わせて報告する。

キーワード 名詞句連接、頻度、並行コーパス、日本語、韓国語

1. Introduction

Syntax in Japanese and Korean shows “striking similarities” especially when we consider that the two languages are not directly connected historically the way Spanish and Italian are, for instance. Such similarities also pertain in the way two nouns (phrases) are combined to form a larger nominal expression. However, when we take a careful look into the distribution and relative frequency of adnominal constructions with particles ‘-no’ and ‘-ui’ against noun-noun compounds, we notice an interesting difference between the two languages. Here we report a preliminary result of our survey using a small parallel corpus, corroborated with data from larger monolingual corpora.

2. Background

In this study, we focus on the adnominal construction in which a noun followed by the adnominal particle, ‘-no’ in Japanese and ‘-ui’ in Korean, forms an adnominal phrase which modifies a noun that immediately follows. While ‘-no’ and ‘-ui’ generally correspond, the constraints do not seem to be the same. In both Japanese and Korean, noun-noun (N-N) compounds, in which two nouns are simply combined, with the first modifying the second, are also possible. Since they behave basically the same as English N-N compounds, they are alternative forms for the adnominal constructions. Chung, el al. (1999) indicates that translation from Korean to Japanese often requires the insertion of ‘-no’ between nouns for Korean N-N compounds. This seems to suggest that the

Japanese N-N compound construction is more constrained than its Korean equivalent.

However, there seems to have been no quantitative study in which the distributions between the two constructions in the two languages are compared in a systematic way. This study, using Sejong Parallel Corpus, parallel corpora of Korean and Japanese, produces quantitative data, examines the distribution of the adnominal construction in each language, and compares the distribution patterns with each other. We corroborate our findings with data from monolingual corpora.

3. Distributional properties of adnominal constructions in Japanese and Korean

3.1. Sejong KR-JP Parallel Corpus

Sejong Korean-Japanese Parallel Corpus consists of four types of data, originally Korean data (21.6%), Japanese data translated from the originally Korean data (19.2%), Japanese originally data (27.5%), and Korean data translated from the originally Japanese data (31.6%).

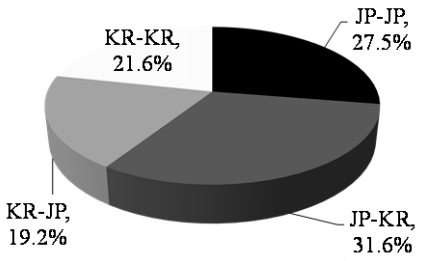

Chart 1: Source language-Target language

The best merit of using a parallel corpus for our purpose here is that we would not have to be concerned with the possible factors that may play roles in selecting one form over the other, such as genre, topic, content, etc.

1.		
2.	<s id=1.1.p1.s1>	
3.	ー	ー/NNR
4.	時間	時間/NADP
5.	**の**	**の/PCS**
6.	奇跡	奇跡/NG
7.	、	、/SYP
8.	海	海/NG
9.	を	を/PJKG
10.	歩こ	歩く/VIN
11.	う	う/AU
12.	！	！/SYG
13.	！	！/SYG
14.	</s>	
15.		

1.		
2.	<s id=1.1.p1.s1>	
3.	한	한/MM
4.	시간의	시간/NNG+
5.		**의/JKG**
6.	기적,	기적/NNG+
7.		,/SP
8.	바다를	바다/NNG+
9.		를/JKO
10.	걷는다!!	걷/VV+
11.		는다/EF+
12.		!/SF+
13.		!/SF
14.	</s>	
15.		

Samples from Sejong Parallel Corpus

The Japanese and Korean data show little difference in size in terms of sentences; Japanese data have 4,045 sentences while the Korean data have 4,038 sentences. On the other hand, the data show some difference in size in terms of morphemes; The Japanese data include 89,083 morphemes, though the Korean data have 101,548. As sample data show, sentences and morphemes often correspond to their counterparts in the other language. The differences in numbers of morphemes and sentences, however, may have resulted from their tagging techniques.

3.2. Frequency of each category

Tables 2 and 3 show the frequency of each category in Japanese and Korean data, respectively. The frequencies of nouns are compatible; nouns account for 33.6% of morphemes in Japanese while those in Korean account for 29.0%.

cat	freq	
Nouns	29903	33.6%
Particles	24090	27.0%
S	11539	13.0%
V	11009	12.4%
A	10869	12.2%
C	1402	1.6%
F	170	0.2%
E	100	0.1%
U	1	0.0%
sum	89083	100.0%

Table 2: Japanese

cat	freq	
Nouns	29484	29.0%
E	19993	19.7%
J(particles)	17109	16.8%
V	13057	12.9%
S	10351	10.2%
X	7147	7.0%
M	4307	4.2%
I	100	0.1%
sum	101548	100.0%

Table 3: Korean

3.3. Frequency of ‘no’ and ‘ui’

Tables 4 and 5 show the frequency of each morpheme in Japanese and Korean, respectively.

string	cat	freq
、	SYP	4636
の	**PCS**	**3841**
。	SYF	3364
た	AU	3078
は	PRE	2796
を	PJKG	2678
に	PJKG	2429
だ	AU	2184
て	PJC	2136
する	VIN	2009
が	PJKG	1946
私	NNPG	898
「	SYPO	877
」	SYPC	875
も	PRE	857

Table 4: Japanese

string	cat	freq
.	SF	3757
다	EF	2584
ㄴ	ETM	2174
하	XSV	2016
을	JKO	1931
에	JKB	1866
의	JKG	1794
는	JX	1495
이	VCP	1443
는	ETM	1382
이	JKS	1339
,	SP	1319
은	JX	1181
를	JKO	1164
고	EC	1156

Table 5: Korean

In the Japanese data, the adnominal particle '-no' is found 3841 times. With the overall 89,083 morphemes of the data, '-no' is found at the rate of one in 23.2 morphemes (Table 6), making it the second most frequently occurring morpheme next to the clause-marking comma (Table 4). For the Korean adnominal '-ui', which is found 1794 times, the ratio is one in 56.6 morphemes (Table 6). The Korean adnominal construction obviously appears far less than its Japanese counterpart.

	JP	KR
morphemes	89,083	101,548
nouns	29,903	29,484
adnConsts with –no, -ui	3,842	1,795
morphs/adnConst	**23.2**	**56.6**
nouns/admConst	**7.8**	**16.4**

Table 6: Frequencies of '–no' and '-ui'

The Sejong Parallel Corpus consists of 50 files for each language. When we examine the occurrences of theadnominal construction in each file, the distribution is somewhat stable in Japanese across its files. On the other hand, the Korean files show greater fluctuation (see Tables 7 and 8). The standard deviation for the occurrence of the adnominal construction '-no' (average 22.0 morphemes per '-no') is 4.9. In the case of Korean, however, the standard deviation for the occurrence of the adnominal construction '-ui' (average 50.4 morphemes per '-ui') is 31.8. This great variation among files suggests that the usages of the Korean adnominal '-ui' may be influenced by the type of discourse or the content.

file	morphs	nn	gen(no)	nn/no	mor/no	mor/nn
6JT_12kj	1090	451	74	6.1	14.7	2.4
6JT_13kj	1068	472	64	7.4	16.7	2.3
6JT_14kj	942	400	58	6.9	16.2	2.4
6JT_15kj	661	270	43	6.3	15.4	2.4
6JT_24kj	1008	424	47	9.0	21.4	2.4
6JT_25kj	1292	651	61	10.7	21.2	2.0
6JT_26kj	383	176	16	11.0	23.9	2.2
6JT_27kj	1138	466	50	9.3	22.8	2.4
50 files	…	…	…	…	…	…
sum	89083	29903	3842			
average	1781.7	598.1	76.8	8.5	**22.0**	2.6
stdev	4601.7	1226.3	177.4	1.9	**4.9**	0.4

Table 7: Japanese files

file	morphs	nn	gen(의)	nn/gen	mor/gen	mor/nn
6jt_12kk	1072	373	39	9.6	27.5	2.9
6jt_13kk	1062	375	36	10.4	29.5	2.8
6jt_14kk	1089	399	37	10.8	29.4	2.7
6jt_15kk	672	243	25	9.7	26.9	2.8
6jt_24kk	906	274	13	21.1	69.7	3.3
6jt_25kk	1405	500	20	25.0	70.3	2.8
6jt_26kk	357	144	2	72.0	178.5	2.5
6jt_27kk	1141	390	7	55.7	163.0	2.9
50 files						
sum	101548	29484	1795			
average	2031.0	589.7	35.9	16.4	**50.4**	3.1
stdev	5493.6	1341.2	61.2	11.3	**31.8**	0.4

Table 8: Korean files

4. Corroborating data

As we wish to use parallel corpora, we realize the weakness of the current study is that the corpora are small. In order to validate the ratio of occurrences of the adnominal constructions to the total morphemes, we compare the figures with those in monolingual corpora of Korean and Japanese, respectively. For Korean, we use Sejong Corpus which is sense-tagged. For Japanese, Kyoto University Text Corpus (Version 4.0) is used. Since the original sources for the Kyoto U Corpus are newspaper articles, we only use the files originating from newspaper articles in Sejong Corpus. The size of the selected Korean data was 573,447 sentences (cf. 4,038 sentences in the Korean data of the parallel corpora), which consists of 15,435,299 morphemes (cf. 101,548 morphemes in the parallel corpora). The size of Kyoto U Corpus is 38,400 sentences (cf. 4,045 sentences), which consists of 972,894 morphemes (cf. 89,083 morphemes).

	Korean	Japanese
source	SJ_semTag*	Kyoto4.0
content	newspaper, mostly *Chosun*	newspaper, *Mainichi*
files	255	28
sentences	573,447	38,400
morphemes	15,425,299	972,894
morphemes/sentence	26.9	25.3
(STDEV)	(17.2)	(14.0)
adnConst with –ui, -no	342,650	53,241
adnConsts/sentence	0.6	1.4
(STDEV)	(0.9)	(1.3)
morphemes/adnConst	45.0	18.3
NNG&NNP	3,709,032	356,629
NNs/adnConst	10.8	6.7

Table 9: Data from monolingual corpora

The result from monolingual corpora confirms that the Japanese adnominal construction with '-no' occurs far more frequently than the Korean adnominal construction with '-ui' (see Table 9). In Kyoto U Corpus, the ratio of the adnominal '-no' construction to morphemes is one in 18.3 morphemes. The occurrences are slightly higher than what we find in the Japanese data from Sejong Parallel Corpus. In Sejong Corpus, the ratio of the adnominal '-ui' construction to morphemes is one in 45.0. Again, the occurrences are slightly more frequent than the Korean data from the parallel corpus. The gap between the ratio of '-no' in Japanese data and that of '-ui' in Korean data is even greater than what we see in Sejong Parallel Corpus.

5. Discussion

As mentioned in the beginning, why the Japanese '-no' occurs more frequently than its Korean counterpart '-ui' is probably largely related to the difference between the constraints on forming N-N compounds in Japanese and Korean. In order to explain the difference, let us first introduce part of a table from Nishiguchi (2009), where the spectrum of various semantic relationships between the two nouns in the Japanese adnominal construction is provided.

Table 10 succinctly shows the so-called 'ambiguity' of the Japanese 'N-no N' construction and how various semantic relationships may hold between the two nouns. Can the Japanese expressions in Table 10 be expressed in N-N compounds? What is interesting from our point of view is that the adnominal particle 'no' is obligatory in most of the cases, as shown in the third column (J-compound) in Table 11, where '-no' is deleted. This contrasts with the optionality of the Korean adnominal particle '-ui', as shown in the third and fourth columns where the Korean counterpart adnominal and compound expressions are given respectively.

Relation	Japanese Possessive	English Possessive	English Noun Compound
I possession	Naomi-no kaban	Naomi's bag	*Naomi bag
II part-whole	Naomi-no kao	Naomi's face	*Naomi face
III location	Tokyo-no	*Tokyo's	*Tokyo relative
IV time	yugata-no koen	*evening's park	evening park
	natsu-no kyuka	*summer's vacation	summer vacation
	7-ji-no nyusu	*7 oclock's	7 oclock news
V accomp-animent	kaban-no hito	*bag's man	the bag man
	boshi-no fujin	*hat's lady	hat lady
VI trade	Kaban-no Kochi	*Bags' Coach	Bags Coach
	Ningyo-no Morishige	*Dolls' Morishige	Dolls Morishige
VII activity	Maaruboro-no kuni	*Marlboro's country	Marlboro country
	biiru-no machi	*the beer's city	*the beer city
VIII property	chisee-no hito	*intelligence's man	*intelligence man
	osu-no tora	*male's tiger	male tiger
	aoi-me-no ningyo	*blue eyes' doll	blue eyes doll
	tsutsuji-no koen	*azaleas' park	azalea park
IX quantity	1-kiro-no pasokon	*1 kg's computer	1 kg computer
	3-bon-no pen	*three's pen	three pens
X intensional property	nise-no fukahire	*fake's shark fin	fake shark fin
	nise-no keisatsukan	*personator's impersonator	a police police officer
XI kind	osu-no tora	*male's tiger	a male tiger

Table 10: from Nishiguchi (2009)

R	J-Poss.	J-Compound	K-Poss.	K-Compound
I	Naomi-no kaban	*Naomi kaban	Naomi-ui gabang	Naomi gabang
II	Naomi-no kao	*Naomi kao	Naomi-ui eolgul	Naomi eolgul
III	Tokyo-no shinseki	*Tokyo shinseki	Tokyo-ui chincheog	Tokyo chincheog
IV	yugata-no koen	*yugata koen	? jeonyeog-ui gong-won	jeonyeog gong-won
	natsu-no kyuka	*natsu kyuka	? yeoleum-ui banghag	Yeoleum banghag
	7-ji-no nyusu	*7-ji nyusu	(?) 7si-ui nyuseu	7si nyuseu
V	kaban-no hito	??kaban hito	*gabang-ui salam	?? gabang salam
	boshi-no fujin	??boshi fujin	?? moja-ui yeoin	? moja yeoin
VI	Kaban-no Kochi	*Kaban Kochi	*gabang-ui kochi	*gabang kochi
	Ningyo-no Morishige	*Ningyo Morishige	*inhyeong-ui molisige	! Inhyeong molisige
VII	Maaruboro-no kuni	*Maaruboro kuni	Malbolo-ui nala	Malbolo nala
	biiru-no machi	??biiru machi	maegju-ui dosi	maegju dosi
VIII	chisee-no hito	*chisee hito	*jiseong-ui salam	*jiseong salam (cf. jiseongin)
	osu-no tora	??osu tora	? susnom-ui holangi	susnom holangi
	aoi-me-no ningyo	??aoi-me ningyo	puleunnun-ui inhyeong	Puleunnun inhyeong
	tsutsuji-no koen	tsutsuji koen	? jindallae-ui gong-won	jindallae gong-won
IX	1-kiro-no pasokon	1-kiro pasokon	1 kg-ui computer	1 kg computer
	3-bon-no pen	*3-bon pen	sejalu-ui pen	sejalu pen
X	nise-no fukahire	nise fukahire	? gajja-ui sang-eojineuleomi	gajja sang-eojineuleomi
	nise-no keisatsukan	nise keisatsukan	?? bunjang-ui gyeongchalgwan	! bunjang gyeongchalgwan
XI	osu-no tora	osu tora	? susnom-ui holangi	susnom holangi

Table 11: Productivity in Japanese and Korean compounding

As for examples in relational categories Possession (I), Part-whole (II), Location (III), Time (IV), Activity (VII), and Quantity (IX), Japanese N-N compounds are not acceptable, whereas their Korean N-N compounds are acceptable. In some cases, such as those in the Time (IV) category, the examples with the '–ui' construction are not as acceptable as those in the N-N compound form. Though the distribution of the grammaticality needs to be expanded to further related examples, those given in Table 11 seem to suggest some initial interesting contrasts between the Japanese and Korean adnominal constructions on the one hand, and between the adnominal and compound constructions for each language on the other.

Another explanation for less occurrence of the '-ui' construction in Korean can be found in the existence of a particular particle *sai-sio*s or 'intervening character sios (ㅅ)[s]. There is a long list of studies concerning whether the particle in question is an adnominal partical (the so-called 'short-form' adnominal particle) or it simply functions as a kind of 'facilitator' for N-N compounding (Cheon 1999). Either way, it partially shows why the admoninal '-ui' in Korean is less frequent than the corresponding '-no' in Japanese; there is another option for Korean to relate the two nouns.

Idiomaticity might also be involved in the less frequent use of '-ui'. Chung, et al. (1999) suggested that some Korean N-N compounds have become idiomatic expressions so that they would not allow an intervening adnominal particle at all. For example, '-ui' is typically absent when nouns are followed by a set of space-temporal expressions in Korean, such as *ap/jeon* ('front'), *yeop* ('side'), *dwi/hu* ('back'), *wi* ('above'), *alae* ('below'), *sai* ('inbetween'), *sog/an/jung* ('inside'), *achim* ('morning'), *jeonyeog* ('evening'). In Japanese, while there are similar space-temporal expressions, originating from Chinese expressions, it seems only a limited number of nouns can precede them without '-no,' such as *yushoku-go* ('after dinner') and *sotsugyo-go* ('after graduation').

6. Further questions

We have seen that the Japanese '-no' construction is used almost twice as often as the Korean '-ui' construction. We have seen cases in which both construction are possible and cases in which Japanese noun phrases with the '-no' construction correspond to Korean N-N compounds. Are there cases in which Korean noun phrases with the '-ui' construction correspond to Japanese N-N compounds? We have not seen solid cases. This is not surprising as we assume that forming Japanese N-N compounds is more constrained than forming Korean N-N compounds.

An interesting discrepancy in Table 10 between Japanese and Korean is found in the Japanese noun phrases in the relational category Nishiguchi calls Trade, such as Kaban-no Kochi (which is roughly translated to "Coach, the bag maker") and Ningyo-no Morishige ("Morishige, the doll maker") do not have Korean counterparts. Such Japanese expressions are productive in the manner in which the name of a brand name is modified by the adnominal phrase with a noun expressing its products or services.

While Nishiguchi (2009) has attempted to categorize the relation between nouns, whether the relations are limited to certain types is an unanswered question. Harada and Shudo (2009), for example, claim that the relation between the nouns in the adnominal construction with '-no' is pragmatically constrained. For the Korean language, Im (1981/1998) indicates that the Korean adnominal construction with '-ui' allows various meanings because the particle requires the existence of the particular relation between referents. It seems that some relations described in Nishiguchi (2009) provide 'the prototypical relations' between nouns. The presence of the prototypical relations seems to blur the constraint on the relation between nouns. While we often address '-no' and '-ui' as a genitive or possessive marker, it is not clear that the labeling is accurate. The above issues give us a larger question; what do the adnominal constructions with '-no' and '-ui' really do? That will be beyond our scope for this paper. More thorough comparative analyses between the languages using parallel corpora may be at least a start.

7. Acknowledgements

The authors would like to acknowledge that their joint research reported here was supported in part by JSPS Grant-in-Aid for Challenging Exploratory Research 21652041, titled Toward Construction of Linguistics of *BA*: Semantics and Pragmatics of *BA*, and Grants-in-Aid for Scientific Research (C) 23520475, titled A New Approach to Presupposition: Analysis of Marginal Cases of Presupposition Manipulation. The authors would like to thank Sanghoun Song for providing them with a refined version of the Sejong Korean-Japanese Parallel Corpus. They would also like to express their appreciation for the Academic Exchange Program between Korea University and Waseda University, with

which one of the authors was able to spend an extended period of time with the other two authors in a productive environment. Yasunari Harada would like to reiterate his academic indebtedness to late Megumi Kameyama in his researches pertaining to adnominal constructions in Japanese.

Language Resources

[1] Sejong Corpus (with sense tagging) http://www.sejong.or.kr/

[2] Sejong Parallel Corpus (Korean-Japanese) http://www.sejong.or.kr/

[3] Kyoto Corpus 4.0 http://nlp.kuee.kyoto-u.ac.jp/nl-resource/corpus-e.html

References

[1] Cheon, Cheol-Woong, “Sai-sios,” In: Where Does the Study of Korean Language Stand, Korea, pp. 186-194, Dong-A Publishers, Seoul, Korea, 1999.

[2] Cheon, Jeong-rye, “The study on the historical change in the genitive and enumerative cases in Korean and Japanese,” Kukyeol Yeongwu 6, pp. 229-254, 2000.

[3] Chung, Yujin, Namwon Heo, Jong-Hyeok Lee, Tai-Suk Kim, “Restoration of Japanese particle ‘-no’ in Korean-to-Japanese MT system,” Proceedings of the 11th Human & Cognitive Language Technology, pp. 115-120, Cheonbuk Univ, Korea, October 1999.

[4] Harada Yasunari and Sachiko Shudo, “Contextual effects of the Japanese adnominal particle no,” In: Proceedings of the 26th Annual Meeting of the Japanese Cognitive Science Society, pp. 218-219, 2009.

[5] Im, Hong Pin, “The existential presupposition and the possessive marker '-ui',” In: The Deep Structure of the Korean Grammar II, Taehaksa, Seoul, Korea, 1981/1998.

[6] Nishiguchi Sumiyo, Possessive Disambiguation, PhD Dissertation, Osaka University, 2009.