

A comparison between Korean and Mongolian eomi and josa for Korean to Mongolian machine translation system

Enkhsaruul A., Chang Geun Song, and Yu-Seop Kim
Department of Computer Engineering, Hallym University
{saruul, cgsong, yskim01}@hallym.ac.kr

Abstract

In this paper we propose comparison of both verb and noun endings between Korean and Mongolian languages. It is based on the similarity between two languages which have the same sentence structures and their eomi and josa structure. Korean verb and noun endings match into those of corresponding Mongolian endings. Josa and eomi are classified as a one-to-one, a one-to-many, and a many-to-many cases as well as some abnormal cases. In order to encourage development of Korean to Mongolian machine translation system, this paper would introduce one of the significant units in grammar.

1. Introduction

New translated major books are important factor in Mongolian society. Especially there is very big demand of major books which are necessary in Mongolian education. The lack of professional translators and finance limitation make it difficult to translate and publish major books regularly. Likewise we can mention many reasons. Therefore, we try to introduce comparison between Korean and Mongolian josa and eomi in order to support machine translation system.

Like Japanese and Korean machine translation systems [1,2] it would be able to be implemented. Korean and Mongolian languages have much similarities which would be advantage for developing machine translation system. Both languages use almost same word order and have josa and eomi. In other word, it is possible to translate every word in a Korean sentence into a Mongolian sentence almost same word order as Korean have.

For example, we have English sentence here 'Let's walk home from here'. Now let us look at Korean and Mongolian forms. A '여기(에)서부터 집까지 걸어 갑시다' is same meaning as 'Эндээс гэр хүртэл алахаад явцгаая'. All vocabularies in the sentence would match as following.

Here = 여기 = Энд

From = '부터' = 'эс'

Home = 집 = гэр

To = '까지' = хүртэл

walk = 걸어 = алахаад

Let's go = 갑시다 = явцгаая

Therefore our paper focuses on matching Korean josa and eomi into Mongolian josa and eomi based on same word order. Solving both languages josa and eomi correspondence would be great contribution for grammar part for machine translation system.

The josa and eomi make both languages more complicated to be applied in machine translation system. However, as shown in previous example we can apply it. Therefore, in here we built tables¹ including those josa and eomi of both Korean and

Mongolian languages. We also use Korean and English examples from grammar book for foreigners by Yonsei University [3,4]. Mongolian examples were translated by us with [5].

Following parts would be a josa, an eomi, frequency information of those josa and eomi, and a conclusion.

2. Josa

The Korean josa is not only translated into exactly one meaning, also it can be translated into two or more in Mongolian. The most of josas can find accurate corresponding target words. However, some of them cannot be translated accurately. In other words, those kinds of josas include more than one meaning in Mongolian language. Therefore, we classified josas into a one-to-one, a one-to-many, an abnormal, and no corresponding cases according to translated meanings in Mongolian. We built four tables for josa and attached explanations for necessary part. Each table has examples of some josas with explanations. A structure of following examples would be first Korean, Mongolian, and English sentences, respectively.

Table 1 is for a one-to-one case and it contains 16 josas. The table divides josas into two parts according to their dependency on their gender of translated word.

Table 1. One-to-one case of josa

<i>Followings are independent on gender</i>	
Korean	Mongolian
-같이	<i>xamm</i>

- The powers after some josas and additional words convey types according to gender of translated words. If power would be 4 it means josa or additional word has four different types.
- A dash before josa, eomi, and additional words convey that those will continuously connected after words.
- Without dash conveys connection after word with space.
- Italic words in the Mongolian part convey not josa and eomi instead those are additional word or substitute of them.

¹ In the table:

-도	ч
-조차	ч
-까지	хүртэл
-만	зөвхөн
-마저	ч гэсэн
-마다	бүр
-처럼	шиг
-하고	Ба/бөгөөд

Followings are dependent on gender

-의	-ийн,-ын,-ний,-ны
-에게 / -한테	-тай ³
-께	-тай ³
-보다	-аас ⁴
-부터	-аас ⁴
-부터 ~ -까지	-аас ⁴ ~хүртэл
-에게서 / -한테서	-аас ⁴ +аа
-을/를	-ийг,-ыг,-г

Here are examples of one-to-one case.

1. -만 (зөвхөн)

그는 공부는 안 하고 잠 만 자요.

Тэр хичээлээ хийхгүй зөвхөн унтдаг.

He only sleeps and doesn't study. (Without studying he only sleeps.)

In this example, '잠 자요' is translated to 'унтдаг'. '잠' cannot be translated separately because it express same meaning as '자요' in mongolian translation, and therefore we only translate '자요' to 'унтдаг'.

2. -보다 (-аас⁴)

비행기가 기차 보다 빨라요.

Онгоц галт тэрэгнээс хурдан.

An airplane is faster than a train.

Table 2 is for a one-to-many case and it contains four josas. Each josa has more than one corresponding josa or additional word. Three of them have special cases which have no role for some translations. In that case, we can just ignore the josa and translate the word.

Table 2. One-to-many case of josa

Korean	Mongolian
-(으)로	-руу,-рүү -аар ⁴ -no role
-(이)나	эсвэл орчим ч
-에서	-аас ⁴ -д,-г -no role
-에	-д,-г -no role

1. -(으)로 is matched into three different cases according to its own meaning. This Korean josa acts like three different ways.

First, it conveys the direction of the movement. Second, it describes a manner or a turning point of an action. Finally, it describes the cause of an event and also shows meaning like by something. For example,

1) -(으)로: -руу,-рүү

내일 유럽 으로 떠납니다.

Маргааш европруу явна.

I will leave for Europe tomorrow.

2) -(으)로: no role

물이 얼어서 얼음으로 변했어요.

Ус хөлдөөд мөс болсон.

Water froze and turned to ice.

In this example no rule means just translating the word '얼음' to 'мөс'. Underlined word is the word without any josa.

3) -(으)로: -аар⁴

한국에 비행기 로 왔어요.

Солонгост онгоцоор ирсэн.

I came to Korea by an airplane.

2. -에서 shows three different meanings to the corresponding josas. One is indicating the place of occurrence of an action given in the predicate clause. Sometimes it is used to show a temporal or spatial point of departure. Another case is with '에', '에서' can be used with nouns associated with a place. However, they may not be used interchangeably. They are selectively used depending upon the meaning communicated by the predicate.

For example,

1) -에서: no role (underlined word in mongolian translation is the word without any josa)

어디 에서 일하십니까?

Хаана ажилладаг вэ?

Where do you work?

2) -에서: -аас⁴

어니 나라 에서 오셨어요?

Аль улсаас ирсэн бэ?

Which country do you come from?

3) -에서: -д, -г

학생들이 교실 에 있습니다.

Оюутнууд ангид байна.

There are students in the classroom.

학생들이 교실 에서 공부합니다.

Оюутнууд ангид хичээлээ хийж байна.

The students are studying in the classroom.

Table 3 is for an abnormal case which contains special one josa.

Table 3. Abnormal case of josa

Some cases we can use corresponding josa in sentences but there is case which we do not use any of them	
Korean	Mongolian
-와/과	ба/бөгөөд -тай ³

-와/과 is matched into two corresponding josas in Mongolian

grammar. But these two corresponding cases can not replace each other. Both perform different role in different sentences according to sentences' meaning.

One is 'ба' and 'бөгөөд' that usually both can replace each other in same sentence. But for some case those may have different usage. It will depend on meaning of the sentence. However, corresponding josa of -와/과 is special that if this josa connects just two words then it can be used but if those connected words are used in sentence then usually it is not used practically in Mongolian written and spoken language. Though, it does not mean that it has different meaning, it really contains same meaning for either of translations. For example,

1. -와/과: ба/бөгөөд

For following example it is not translated corresponding josa instead just two words are written one after another.

이 방에 의자 와 책상이 많아요.

Энэ өрөөнд сандаг ширээ их.

There are many chairs and desks in this room.

Another match is '-тай³' that it is used for rest of cases. It will be connected to word.

2. -와/과: -тай³

In following sentence '-тай³' is added on the pronoun and connects it with adjective.

이 물건은 내 것 과 비슷합니다.

Энэ юм минийхтэй адилхан.

This item is similar to mine.

Table 4 is for no corresponding case and it contains three josas. For this case it is not translated in Mongolian sentence therefore it just ignore the josas and just translate the words themselves. And there would be no trouble to complete the whole sentence. Translation meaning would be match correctly as Korean has.

Table 4. No corresponding case of josa

Followings express infinitive form of translated word	
Korean	Mongolian
-이/가	no corresponding case
-께서	no corresponding case
-는/은	no corresponding case

Here is examples of no corresponding case and underlined word is the word without josa.

1. -이/가 (No corresponding ending)

눈이 옵니다.

Цас орж байна.

It is snowing.

2. -는/은 (No corresponding ending)

이 연필 은 참 좋아요.

Энэ харандаа маш сайн.

This pencil is very nice.

3. Eomi

As same as josa, Korean eomi is not only translated into one exact meaning of Mongolian eomi, but also it can be translated into two or more. The most of eomis can match accurate meaning. However, some of them cannot be translated accurately. In other words, those kinds of eomis include more than one meaning in

Mongolian language. Therefore, we are classified eomis into a one-to-one, a one-to-many, and an abnormal cases according to translated meanings in Mongolian. We built three tables for eomi and attached explanations. Each table has examples of some eomis with explanations. A structure of following examples would be first Korean, Mongolian, and English sentences, respectively.

Table 5 is for a one-to-one case and it contains 21 eomis. In addition, some eomi has additional letters or word after them. It conveys whole meaning of Korean eomi. But those are not real eomi in Mongolian grammar so we called it additional words or letters.

Table 5. One-to-one case of eomi

Korean	Mongolian
<i>1. Followings are dependent from words' gender that those would be four types and some of them have additional parts after the josa</i>	
-습니다, -습시다, -읍니다	-на ⁴
-십시오, -으십시오	-аарай ⁴
-(으)려고	-аар ⁴
-(으)러	-аар ⁴
-(으)면	-вал ⁴
-(으) 면서	-нгаа ⁴
-는	-даг ⁴
-(으)ㄴ	-сан ⁴
-았(었/였) 던	-сан ⁴
-비시다, -읍시다	-цгаа ⁴ + я/e/ë
-자	-нгууг ² + аа ²
-기 전에	-аас ⁴ + -аа ⁴ өмнө
<i>2. Followings are independent from words' gender and some of them have additional part after the josa</i>	
-(으)며	ч
-(으)르 수록	тусмаа/тусам
-던	-ж байсан
-기	нь
-기 위해서(위하여)	гэж
-기 때문에	-ж байгаа учраас
<i>3. -If word is adjective '-май⁴' will added then 'ч'</i> <i>-If word is verb '-can⁴' will added then 'ч'</i>	
-아(어/여)도	ч
<i>4. -Independent from words' gender</i> <i>-Can use either of two</i> <i>-Words written after josa '-ж' are additional</i>	
-지만	-ж байгаа ч -ж байгаа боловч
<i>5. '-нгаа⁴' is dependent from words' gender and it would be four types</i> <i>- '-ж байгаад' is independent from word's gender</i> <i>-Can use either of two</i>	
-다가	-нгаа ⁴ / -ж байгаад

Examples of one-to-one case

1. -아(어, 여)도 (ч)

음식이 매워도 괜찮아요.

Хоол халуун ногоо ихтэй ч, зүгээр үү?

It is okay that the food is (spicy) hot?

2. -다가 (-нгаа⁴, -ж/- ч байгаад)

a. 학교에 가다(가) 친구를 만났어요.

Сургуульруугаа явж байгаад найзтайгаа уулзсан.

Сургуульруугаа явангаа найзтайгаа уулзсан.

On the way to school, I met a friend.

b. 시장에 갔다가 은행에도 들렀습니다.

Зах гарангаа банк орсон.

Зах гарч байгаад банк орсон.

I went to the market, then went to the bank.

Table 6 is for a one-to-many case and it contains seven eomis. Each eomi match into corresponding two or more cases. Corresponding cases express different meaning, it cannot replace one another.

Table 6. One-to-many case of eomi

Korean	Mongolian
-고	ба, бөгөөд / -аад ⁴ / -ж
-아(어, 여)서	-аас ⁴ / -аад ⁴
-(으)니까	-аас ⁴ / -д
-거든	-вал ⁴ / байна
-는(은)데	-ж байна / -ж байгаа ч
-아(어, 여)야	-х хэрэгтэй / -сан ⁴ ч
-도록	-нгаа ⁴ / -гүйгээр

1. -고 is matched into 3 different general cases and one of them has two types according to its usage in the sentence that connects two word in same level. Second eomi is used according to its gender. Last one is added on word that is independent from gender. -고 lists two or more sentences simply and it has several different usages. First, two events take place concurrently. Second, two events are in a temporal relationship. Third, it is depending upon the verbs used, the first action still continues while the second takes places. For example,

a. -고: ба, бөгөөд

하늘은 높고, 바다는 깊어요.

Тэнгэр уудам далай гүн./Тэнгэр уудам **ба** далай гүн.

Sky is high, and sea is deep.

In this example, we can translate Korean sentence into two options. The reason is 'ба, бөгөөд' can be omitted from sentence. However result would be same.

b. -고: -аад⁴

형이 새 옷을 입고 외출했습니다.

Ах шинэ хувцас өмсөөд гарсан.

My brother went out wearing new clothes.

c. -고: -ж

나는 피아노를 치고, 동생은 노래를 불렀습니다.

Би төгөлдөр хуур тоглож дүү дуу дуулсан.

I was playing the piano, while my sister sang.

5. -는(은)데 has two corresponding cases. But actually it is one josa with two different additional words. -는(은)데 has several usages that first two sentences are connected in such way that the action taking place in the first sentence still continues in the second sentence. Sometimes it is used as a terminative ending, and conveys the speaker's willingness to listen or to gather the

hearer's response. But this case Mongolian josa '-ж' will be omitted from sentence so only addition word 'байна' is used. Second, two sentences are contrasted in a time concept, in meaning, or stand in a cause / effect relationship.

For example,

1a. -는(은)데: -ж байна

비가 오는데, 우산이 있습니까?

Бороо орж байна шүхэр байгаа юу?

It is raining, do you have an umbrella?

1b. -는(은)데: байна (josa is omitted)

오늘은 날씨가 꽤 좋는데.

Өнөөдөр цаг агаар үнэхээр сайхан байна.

Today is a fine day.

2. -는(은)데: -ж байгаа ч

계속 치료를 하는데, 감기가 낫지 않아요.

Эмчилгээгээ үргэлжлүүлж байгаа ч гэсэн ханиад эдгэхгүй байна.

Even if I continue with treatment, I am not recovering from my cold.

Table 7 is for an abnormal case and it contains four speciality. Some of them not real eomi in Mongolian and they just express the exact meanings.

Table 7. Abnormal case of eomi

Korean	Mongolian
-ㅁ니까, -습니까, -옵니까	yy ² ?
<i>'yy²' depends on words' gender and it has two types It is addition letter and it is used after verb</i>	
<u>갑니까?</u>	
Явах <u>yy</u> ?	
(Will you) go?	
-(으)ㅁ	no regular case
<i>There is a regulation to convert verb to noun but it is not for all. There is an irregular conversion for most word. For example,</i>	
1. 걷다 - 걸음	
Алхах - алхаа	
to walk - step	
2. 웃다 - 웃음	
Инээх - инээд	
to laugh - laughter	
3. 아프다 - 아픔	
Өвдөх - өвчин	
to hurt - pain	
4. 자다 - 잠	
Унтах - нойр	
to sleep - sleep	
5. 울다 - 울음	
Уйлах - уйлах	
to cry - crying	
-(으)ㄹ	-x
<i>'-x' is corresponding josa and it is same as verbs' last alphabet so without adding this josa we can use infinitive verb itself in the sentenc. For example,</i>	
1. 비가 오니까 쓰고 <u>갈</u> 우산을 준비해야 합니다.	
Бороо орж байгаа учраас хэрэглээд явах шүхрээ бэлдэх хэрэгтэй.	

Since it is raining, (You) must prepare an umbrella to use.	
-(으)시	Ta
Korean honorific form is included in the verb but Mongolian does not have this form. However, beginning of the sentence we can use 'Ta' which means honorific calling for elders or strangers. Just like we call our parent 'mom and dad' before saying something. 'Ta' locates beginning of the sentence which means calling her or him instead of their names. For example, 1. 가시겠습니까? Ta, явах уу? Would you go?	

4. Frequency Information

This information includes josa and eomi which are used in our paper. Table 8 shows the 20 most frequently used Korean Josa selected by [6]. The table is composed of four columns. The first column shows the name of the josa itself and the second shows the frequency number counted from Corpus data. The third means the proportion of the josa and the final column means the corresponding class classified as in chapter 2: one-to-one, one-to-many, abnormal, and no corresponding cases. The number 1 to 4 shown in the fourth column corresponds the classes respectively.

Table 8. The frequency data of josas

Josa	Frequency	Proportion (%)	Class
의	72,437	11.4	1
을	71,370	11.2	1
에	57,750	9.1	2
이	50,894	8.0	4
이다	49,507	7.8	2
는	45,348	7.1	4
를	41,429	6.5	1
은	35,606	5.5	4
가	32,530	5.1	4
도	23,812	3.8	1
으로	21,559	3.3	2
에서	19,693	3.1	2
로	15,511	2.5	2
과	9,914	1.6	3
고	9,732	1.5	2
와	6,471	1.0	3
만	6,057	1.0	1
이	4,949	0.8	4
에게	4,769	0.7	1
과	4,435	0.7	3

The case 1 and the case 4 does not need an additional operation to translate the josas. The Korean josas only have to find out the target josa in Mongolian. About 67.5% of josas are included in the case 1 and case 4.

We didn't make a table for eomi, because the names of eomis shown in [3] and [6] are different. However, about over 60% of all eomis are inferred to be translated into Mogolian eomi without any specific operations.

5. Conclusion

This paper concentrates on Korean and Mongolian josa and eomi using two languages' similarity. The almost same word order gives big advantage to implement machine translation system. Therefore, we introduced how we should handle Korean and Mongolian josa and eomi for machine translation system.

The tables classify josa and eomi into a one-to-one, a one-to-many, an abnormal, and no corresponding cases. From the tables we specially focus on one-to-many, abnormal, and no corresponding cases. That's why here would be more explanations and examples involved in those tables. This paper did not include any solution for choosing which corresponding cases would be correct for current sentence. However, we plan to research the solutions for the future. For example, if we can analyze the characteristics of roots of the words, we can find the corresponding translation more accurately.

We hope that this paper will give encouragement to the people who have willing to implement Korean to Mongolian machine translation system.

Reference

- [1] Sanghwa Yuh, Hanmin Jung, Won Chang, Taewan Kim, Dong-In Park, Jungyun Seo, "FromTo/JK: A Japanese-Korean Maching Translation System," Procs. of NLPRS'97, pp.613-616, 1997.
- [2] Yu-Seop Kim, Sung Dong Kim, Seong Bae Park, Jong Woo Lee, Jeong Ho Chang, Jyu Baek Whang, Jang Min O, and Yung Taek Kim, "Machine Translation Systems: E-K, K-E, J-K, K-J," Procs. of 4th Conference of the Association for Machine Translation in the Americas, pp.405-406, 2000.
- [3] Tae-Sung Kim, Ph.D, *한국어 1*. Yonsei University Press., 1992.
- [4] Ho-bin Ihm, Kyung-pyo Hong, Suk-in Chang. *Korean Grammar for International Learners*. Yonsei University Press., 2001.
- [5] Sainzaya. L, Yanjin. T, Enhke. D, Ganbold. D., *Mongolian Grammar Handout*, 2001.
- [6] The National Institute of the Korean Language, "Modern Korean Language Usage Frequency Research Results," <http://www.korean.go.kr>, 2003.