
Jun Choi/Hae-Yun Jung

ON LOANS IN KOREAN NEW WORD FORMATION AND IN LEXICOGRAPHY

Abstract This study examines a list of 3,413 neologisms containing one or more borrowed item, which was compiled using the databases built by the Korean Neologism Investigation Project. Etymological aspects and morphological aspects are taken into consideration to show that, besides the overwhelming prevalence of English-based neologisms, particular loans from particular languages play a significant role in the prolific formation of Korean neologisms. Aspects of the lexicographic inclusion of loan-based neologisms demonstrate the need for Korean neologism and lexicography research to broaden its scopes in terms of methodology and attitudes, while also providing a glimpse of changes.

Keywords Neologisms; lexicography; loans; clipping; blending; word formation

1. Introduction

The objective of this paper is to understand what role loans play in Korean neologism formation and whether they weigh in the inclusion of loan-based neologisms in Korean dictionaries by analysing the neologisms that include at least one borrowed element. These neologisms were retrieved from the lists of all neologisms that were extracted from 2006 to 2019 within the framework of the Korean Neologism Investigation Project (a project funded and supervised by the National Institute of Korean Language) and in 2020 by the Centre for Korean Language Information Studies (Kyungpook National University). In section 2.1, an overview of the government-affiliated project is presented, along with an explanation of the methodology used to extract Korean neologisms.

By ‘borrowed element’ is meant any type of loans – whether a loanword, a loan-morpheme, or a clipped loanword to form blend neologisms. In this paper, loans are analysed according to three of their characteristics. First, they are examined from an etymological perspective (section 2.2). The language origins of borrowed items may range from European languages to Asian languages, but one language type that is not considered for loans is Traditional Chinese, or Hanja, as opposed to Simplified Chinese. Hanja-based words (also called Sino-Korean words) are regarded as fully Korean, even though they are distinguished from native Korean words (sometimes referred to as pure Korean words). Second, loan-based neologisms as well as Korean neologisms are analysed from a morphological perspective (section 3), that is, based on Korean word formation processes (3.1) and with regard to their potential productivity (3.2). Lastly, they are discussed from a lexicographic perspective (section 4), not only in terms of statistics but also in terms of the attitudes of the Korean academia towards loanwords and neologisms.

2. Methodology

2.1 The Korean Neologism Investigation Project

Affiliated to the South Korean Ministry of Culture and Tourism since 1991, the National Institute of Korean Language (NIKL) undertook the task of collecting and analysing new words in Korean language from 1994 to 2019 under the project name ‘Korean Neologism Investigation Project’ (KNIP). KNIP was carried out on a yearly basis by the research team from a research centre or a university, who produced annual reports, for the most part available on the NIKL website¹. The reports present the neologisms in order of high frequency as well as in alphabetical order, and provide crucial information for neologism and dictionary research. Indeed, they describe each neologism in lexicographic style (i.e., in terms of pronunciation, etymology, part-of-speech, domain when applicable, definition, and examples) and indicate not only the date of first occurrence but also frequencies of the neologisms

The project was temporarily discontinued in 2011 and resumed in 2012 with the Centre for Korean Language Information Studies at Kyungpook National University (KNU) carrying out the project under the supervision of NIKL until it came to an end in 2019. In 2020, the investigation into Korean neologisms was conducted independently by the abovementioned KNU Centre, using the same framework and criteria set by NIKL for the consistency of the data² (Nam et al. 2021).

In the early stages of KNIP, new words³ were manually retrieved from printed newspaper articles and news broadcasting scripts. The development of the Internet and computing tools allowed a number of methodological improvements from the early 2000s onwards, including the distinction between neologisms proper and words that are simply not included in the *Standard Korean Language Dictionary* (SKLD), the expansion of source texts with ever-increasing online media⁴ and, from 2005, the construction and use of a Web-based corpus to extract the neologisms automatically (in addition to the manual extraction). From 2012 onwards, neologism candidates have been retrieved using a Web-based neologism extractor based on whether or not a candidate word is represented in the online dictionary *Urimalsaem*^{5,6}. The list of neologism candidates is then manually checked by researchers to narrow down the candidates to neologism headword candidates.

¹ https://www.korean.go.kr/front/reportData/reportDataList.do?mn_id=207.

² The report on the 2020 neologisms is not available in the NIKL website as the investigation was not carried out as part of the government-funded KNIP but has been published by Hankwukmwunhwasa.

³ We specifically use the term ‘new word’ and not ‘neologism’ here as the project aimed to retrieve any words that were not included in the *Standard Korean Language Dictionary* published by NIKL, whether neologisms or not.

⁴ The online news articles used for the project are those provided by the Naver News portal (<https://news.naver.com/>).

⁵ While the macrostructure of *Urimalsaem* is based on *SKLD* and administered by NIKL, it is nonetheless much bigger than *SKLD*. As an online dictionary, *Urimalsaem* has no printing limitations and also allows users to suggest headword candidates.

⁶ For more details on the methodological changes that were brought to the Korean Neologism Investigation Project, see Nam/Lee/Jung (2020, pp. 107–110); Choi (2020, p. 153).

2.2 The neologisms under study

The present study focuses on the years 2006 to 2020⁷ – that is, after the project started to use the corpus methodology and the retrieval of neologisms thus became more systematized. More specifically, the study targets the neologisms that are either full loanwords or partly borrowed. During this time frame, there were in total 6,554 neologisms collected, from which we extracted a list of 3,413 neologisms with at least one borrowed element by excluding all neologisms that are composed of solely native Korean and Hanja characters. Loan-based neologisms represent a little over 52% of the Korean neologisms collected in the past fifteen years or so. Table 1 shows the ratio of such neologisms for each year of the time frame and Table 2 the ratio of neologisms comprising at least one borrowed element from a given language, presented in order of high frequency.

Year	Number of neologisms	Number of loan-based neologisms	Percentage
2006	530	254	47.92
2007	702	369	52.66
2008	475	167	35.16
2009	588	219	37.24
2010	368	170	46.20
2012	511	298	58.32
2013	488	282	57.79
2014	339	212	62.54
2015	285	166	58.25
2016	649	349	53.78
2017	396	237	59.85
2018	460	291	63.26
2019	358	212	59.22
2020	405	187	46.17
Total	6,554	3413	52.07

Table 1: Ratio of loan-based neologisms per year

Borrowed language	Number of loan-based neologisms with at least 1 element from the borrowed language	Percentage
English (EN)	3,235	94.78
French (FR)	97	2.84
Japanese (JA)	70	2.05
Italian (IT)	52	1.52
Chinese (ZH)	22	0.64
German (DE)	22	0.64

⁷ Except for 2011, where there is no data.

Borrowed language	Number of loan-based neologisms with at least 1 element from the borrowed language	Percentage
Latin (LA)	14	0.41
Greek (EL)	10	0.29
Spanish (ES)	5	0.14
Danish (DA)	4	0.11
Russian (RU)	3	0.08
Hindi (HI)	2	0.05
Pashto (PS)	2	0.05
Indonesian (ID)	1	0.02
Portuguese (PT)	1	0.02
Sanskrit (SA)	1	0.02
Turkish (TR)	1	0.02

Table 2: Ratio of loan-based neologisms per borrowed language

Although English is clearly and by far the predominant language as regards Korean neologisms, neologisms may borrow from a variety of languages. More importantly, they may combine elements from several languages. Thus, they can be divided into neologisms borrowing from a single language (1a), hybrid neologisms composed of borrowed elements from multiple foreign origins (i. e., other than native Korean and Traditional Chinese) (1b), and hybrid neologisms composed of a borrowed element and a native Korean and/or Hanja element (1c).

- (1) a. *kheyleynsia* (ES *querencia*); *teykacicum* (FR *dégagisme*); *koltu misu* (EN *gold miss*)
 b. *hwuykeylaiphu* (DA *hygge* + EN *life*); *lamulliey* (JA *râ[men]* + FR [*som*]*melier*); *weyting alpa* (EN *wedding* + DE *Arbe[it]*)
 c. *lattey appa* (IT *latte* + Korean (KO) ‘daddy’), *takkwuin* (EN *di[ary]* + KO *kkwu[minun]* ‘decorating’ + Hanja (HA) ‘person’) ⁸

Hybrid neologisms are the most common neologisms with borrowed elements, counting 2,225 items and constituting almost two thirds of all loan-based neologisms. Among those, only 106 neologisms fall into the (1b) type of hybrids. In other words, 95.2% of hybrid neologisms include at least one element in native Korean or Sino-Korean. Neologisms consisting of a single borrowed language are not necessarily loanwords. As seen in (1a), *koltu misu* (EN *gold* + EN *miss*), which is composed of English words, falls into the category of Konglish neologisms, whereby English morphemes are borrowed to form words in English which do not exist in the English language. Instead, Konglish neologisms follow the semantic and cognitive patterns of Korean. The following section explains the processes and trends in Korean word formation.

⁸ Romanization of Korean follows the Yale romanization system and literal translation is provided in single quotation marks where needed. Round brackets show the original word in case of loans; square brackets show elements that have been dropped in the word creation process.

3. Loans and word formation in Korean

3.1 Word formation processes in Korean neologisms

From a morphological perspective, Korean words can be divided into two main categories, that is, simplex words and complex words. The formation of simplex neologisms encompasses the rather rare generation of purely native Korean or Sino-Korean forms⁹ as well as the borrowing of a wordform from a foreign language. On the other hand, complex words are formed by combining either a root and an affix (derivatives), or two or more word stems (compounds), or two or more clipped words (blends). In practice, the lines delimitating these categories are not clear-cut. For some native Korean/Sino-Korean neologisms for example, it is unclear whether they are simplex or complex words. This is the case of the 2018 neologism *pposilaeki* ‘little one’, used to designate young or small and cute people or animals. It could be argued that the form *ppo-* is short for *ppoccak*, which is a dialectal form for *paccak* ‘close(ly)’ and has been widely used by netizens with the meaning of ‘cute’ when describing animals or talking of Korean pop idols online. However, the word has been categorized as a simplex form, as a semantic neologism of *pposilaeki* ‘crumb’ in a southwestern dialect by analogy with the netizens’ use of *ppoccak*.

Regarding loanwords, their categorization is not always obvious depending on whether they are considered from the point of view of word formation or from the perspective of the resulting word form. As ten Hacken/Panocová put it, even “if the borrowed word is the result of a word formation rule in the original language, the word formation origin is lost in the receiving language”, because “[w]ord formation rules are not borrowed” (ibid. 2020, p. 4), only the final product is. The authors illustrated this with the English simplex word but originally German compound *kindergarten*. There are such examples in Korean neologisms, such as the 2018 loanword *khuliphocaykhing*, from the English blend ‘cryptojacking’ (*crypto[currency] + [hi]jacking*) which is perceived as a simplex word in the formation process of Korean neologisms. Even if there are about twice as many loanwords as neologisms generated from native Korean or Sino-Korean characters, they still only constitute about 6.7% of the total loan-based neologisms, most of which being blends or compounds.

Before getting into complex neologisms, a new (minor) process of word formation has developed following the ever-increasing user-generated content platforms, which can be somewhat considered as a morphological ‘anomaly’. This process consists of forming neologisms by replacing Korean characters¹⁰ with other characters of similar shape, regardless of their semantic or phonetic similarities, as the two 2018 neologisms illustrate in (2).

- (2) 땡땡미 *taengtayngmi* ‘someone as cute as a puppy’ for KR 멍멍이 *mengmengi* ‘doggy’;
네주얼 *neycwuel* ‘visual’ instead of the English loanword 비주얼 *picwuel* ‘visual’

These neologisms can be categorized as graphic neologisms or ‘pictorial representations’ (Kim 2016). Example (2) also shows that this graphic word formation process can be applied to both native Korean or Sino-Korean words and loanwords.

⁹ Within the scope of our study, there are only 94 such neologisms out of the total 6,554 neologisms formed from 2006 to 2020.

¹⁰ The Korean script can be classified as a syllabic alphabet in that it does not consist of ideograms as in Chinese, but of alphabetic letters that are combined in square clusters to form a character (i. e. a syllable).

3.2 Compounding and blending in Korean neologisms: the case of French-based neologisms

As just mentioned above, most Korean neologisms, including loan-based neologisms, are formed through compounding and blending. A little more than a third of the neologisms under study are blends (1247 items) and nearly 45% of them are compounds (1526 items). When examined according to the language origin, we can observe clear patterns emerging between both categories. To illustrate these patterns, the following analysis focuses on French-based compound and blend neologisms.

- (3) a. Compound neologisms: *hompakhangsucok* (EN *home* + FR *vacances* + HA ‘tribe’)
 b. Blend neologisms: *nuckhangsucok* (KO ‘late’ + FR [*va*]cances + HA ‘tribe’); *kolkhangsucok* (EN *golf* + FR [*va*]cances + HA ‘tribe’); *molkhangsu* (EN *ma[ll]* + FR [*va*]cances); *holkhangsucok* (KO ‘alone’ + FR [*va*]cances + HA *cok* ‘tribe’); *khakhangsu* (FR *ca[fɛ]* + FR [*va*]cances); *phwulkhangsu* (EN *pool* + FR [*va*]cances)

The most striking feature of the French-based neologisms is the salience of the loanword *pakhangsu* (*vacances*), especially used as the clipped loan *-khangsu* ([*va*]cances) to form blend neologisms. While (3.a) shows the only example of a compound formed based on the loanword, (3.b) features only a handful of blends formed with the clipped loan. Table 3 presents the number and percentage of such blends per year.

Year	Number of neologisms containing <i>khangsu</i> ([<i>va</i>]cances)	Percentage
2010	2	3.7
2011	0	0
2012	1	1.9
2013	4	7.4
2014	0	0
2015	1	1.9
2016	0	0
2017	1	1.9
2018	3	5.5
2019	14	25.9
2020	2	3.7
Total	28	51.9

Table 3: Ratio of blend neologisms containing the clipped French loan *-khangsu* ([*va*]cances) to French-based blend neologisms per year

28 out of the 54 French-based blend neologisms include the clipped loan *-khangsu*. While the *-khangsu* neologisms seemed to have been particularly trendy in 2019, distribution across the remaining years is rather balanced, which leads us to think that this particular loan will most likely continue to be used in the future. In fact, the loanword *pakhangsu* (*vacances*) and its clipped version have long been used to create Korean neologisms, many of which were included in *Urimalaem*: twelve *pakhangsu* neologisms and fifteen *-khangsu* neologisms, including four from our list. The case of the French loan *-khangsu* can be ex-

tended to a few other donning languages, especially where the number of neologisms is higher¹¹, as illustrated in Table 4.

Language	Number of loan-based neologisms	Most productive loan	Number of neologisms with most productive loan	Examples
Japanese	70	<i>otaku</i> ‘geek’	2 compounds; 24 blends	<i>otekcil</i> (JA <i>otak[u]</i> + KO ‘attitude’); <i>sengtek</i> (KO ‘successful’ + JA [<i>o</i>] <i>tak[u]</i>)
Italian	52	<i>paparazzi</i>	3 compounds; 24 blends	<i>phaynphalachi</i> (EN <i>fan</i> + IT [<i>pa</i>] <i>parazzi</i>); <i>kyenphalachi</i> (HA ‘muzzle’ + IT [<i>pa</i>] <i>parazzi</i>)
Chinese	22	<i>mala</i> (spicy seasoning)	1 compound; 3 blends	<i>malamama</i> (ZH <i>mala</i> + ZH ‘mummy’)
German	22	<i>Arbeit</i>	1 compound; 7 blends	<i>alpaleylla</i> (DE <i>Arbe[it]</i> + EN [<i>Cinde[rel]la</i>])
Latin	14	<i>homo</i> + attribute ¹²	3 compounds; 5 blends	<i>homo cheyexhwusu</i> (LA <i>homo</i> + EN <i>chair</i> + LA –[Australopithe]cus)

Table 4: Most productive loans per language with higher number of neologisms containing the most productive loan

Table 4 shows that some loans yield many neologisms. It also confirms that their productivity is more prominent in blending when used in their clipped forms. This means that these particular loans are well established in the mental lexicon of Korean language speakers.

4. Lexicographic representation of and attitudes towards Korean loan-based neologisms

Just as *vacances*, such productive loanwords as *otaku*, *paparazzi*, *Arbeit*, and *homo*, as well as a number of the compound and blend neologisms they produced, have been included in *Urimalsaem*. However, they are not equally represented, as shown in Table 5.

¹¹ We leave alone the case of English, which is overwhelmingly higher and thereby would present many cases of productive (clipped) loanwords.

¹² That is, neologisms that imitate forms such as ‘homo Australopithecus’ or ‘homo sapiens’ to designate people living a certain lifestyle. For example, *homo cheyexhwusu* ‘homo chaircus’ refers to those who spend their day sitting, such as office workers or academics.

Loanword	Number of related neologisms	Number of related neologisms included in <i>Urimalsaem</i>	Number of other related neologisms included in <i>Urimalsaem</i> which are not from our list
<i>pakhangsu</i> (FR <i>vacances</i>)	28	4	23
<i>othakhwu</i> (JA <i>otaku</i> ‘geek’)	26	15	–
<i>phaphalachi</i> (IT <i>paparazzi</i>)	27	5	29
<i>alupaithu</i> , <i>alpa</i> ¹³ (DE <i>Arbeit</i>)	8	4	29
<i>homo</i> + attribute (LA <i>homo</i>)	8	7	18

Table 5: Representation of highly productive loans in *Urimalsaem*

While very few neologisms from our list made it to the dictionary, quite surprisingly, many other neologisms formed with the loanwords (compounded or blended) from Table 5 have been added to *Urimalsaem* but are not from the list of neologisms collected within the scheme of the KNIP. On the one hand, this highlights the limits of neologism extraction from the sole genre of ‘news media’. On the other hand, it implies that these ‘other’ neologisms that eluded the project have been probably included following the dictionary users’ suggestions. Indeed, *Urimalsaem* allows users to participate in the making of the dictionary and suggest new headwords with their definitions under a separate tab. Users’ suggestions are then reviewed by language experts and potentially added if they have lexicographic value.

In addition to the inconsistent representation of these neological ‘families’, only FR *vacances* and DE *Arbeit* are represented in *SKLD*, moreover only in their full, unaltered forms. Although both dictionaries are managed by the NIKL and considered as language authorities, they also have different characteristics. Unlike *SKLD* that started off as a print dictionary that has been digitalized, *Urimalsaem* is solely an online dictionary which has some characteristic features of online content. For example, as mentioned earlier, it allows user-generated content, although still supervised by experts. Thus, *Urimalsaem* is more inclusive – at least on paper – and its macrostructure, albeit based on *SKLD*, grows at a faster rate than *SKLD* which is more of a traditional dictionary, that is, prescriptive and more conservative towards neologisms, and particularly loan-based neologisms.

In the mid-1970s, the government undertook the task to ‘purify’ the Korean language. The main objective of the task was to refine Korean by replacing improper or dialectal words by correct, standard words, complicated words or expressions by simpler ones, and words of foreign origins by native Korean words (Kim 2019; Seo 2019). *SKLD* and *Urimalsaem* are government-affiliated dictionaries and may reflect some aspects of the language purification policy. In the case of *SKLD*, it is shown from its macrostructure itself. In the latest statistical report on the dictionary content, which is accessible on the *SKLD* website¹⁴, it ap-

¹³ *Alpa* is the shortened form of *alupaithu*.

¹⁴ https://stdict.korean.go.kr/statistic/dicStat.do#static_menu3_3.

pears that words of foreign origin constitute a mere 5.6% of the headwords and hybrids account for 20.5%; Korean words, including native Korean (20.9%) and Sino-Korean (53%), make up the vast majority of the macrostructure. The low ratio of foreign words together with hybrids (26.1%), may be explained by the dictionary's passivity before neologisms in general, and loan-based neologisms in particular. As a matter of fact, one of the rare neologisms that were included in *SKLD* in the 2000s is *colipep* 'recipe', which was actually presented as a Korean replacement to the English loanword *leysiphi* 'recipe'.

As for *Urimalsaem*, it does accept more neologisms and loans; however, it has not neglected its role of language prescription. Example 4 shows a couple of cases of 'normative information' in loanword entries.

- (4) a. *pakhangsu* (FR *vacances*):
Purification (notice for the correction of daily life terms (Ministry of Culture and Sports Notice No. 1996-13, March 23, 1996))
Instead of 'vacances', use the refined terms of *yelum hyuka* (KO 'summer' + HA 'vacation') or *hyuka* (HA 'vacation') if possible.
- b. *alupaithu* (DE *Arbeit*):
Purification (notice for the correction of daily life terms (Ministry of Culture and Sports Notice No. 1996-13, March 23, 1996))
'pwuep' (HA 'part-time job') can be used along 'Arbeit'

It is nonetheless safe to say that many normative forms of the Ministry of Culture and Sports fall into oblivion.

5. Conclusive remark

As for a conclusion, we propose to have a last look into our list of loan-based neologisms and check how, overall, they are represented in *Urimalsaem* and how they compare with native Korean and Sino-Korean neologisms.

Year	Total neologisms	Total neologisms included in <i>Urimalsaem</i>	Number of native Korean and Hanja neologisms represented	%	Number of loan-based neologisms represented	%
2006	530	91	58	63.7	33	36.3
2007	702	148	70	47.3	78	52.7
2008	475	85	56	65.9	29	34.1
2009	588	97	70	72.2	27	27.8
2010	368	17	8	47.1	9	52.9
2012	511	178	73	41.0	105	59.0
2013	488	152	56	36.8	96	63.2
2014	339	196	85	43.4	111	56.6
2015	285	206	85	41.3	121	58.7
2016	649	302	143	47.4	159	52.6
2017	396	211	76	36.0	135	64.0

Year	Total neologisms	Total neologisms included in <i>Urimalsaem</i>	Number of native Korean and Hanja neologisms represented	%	Number of loan-based neologisms represented	%
2018	460	201	72	35.8	129	64.2
2019	358	20	8	40.0	12	60.0
2020	405	8	3	37.5	5	62.5
Total	6,554	1,912	863	45.1	1,049	54.9

Table 6: Lexicographic representation of native Korean and Hanja neologisms against loan-based neologisms per year

Table 6 allows us to take a look at the big picture. The inclusion of loan-based neologisms can be divided into three main stages. Until 2009, there were generally fewer loan-based neologisms included in the dictionary than native Korean and Hanja neologisms, regardless of whether their ratio was higher. Then, in the first half of the 2010s, they seemed to gain ground in the lexicographic race. Finally, since 2017, even though fewer neologisms have been added to the dictionary, there tend to be twice as much inclusion of loan-based neologisms as native Korean and Hanja neologisms. Despite the normative attitude of Korean dictionaries and the efforts of language policies to minimize the impact of loanwords on the Korean language, it is the language speaker who ultimately shapes the language by creating new words and choosing the words to use and giving momentum to loan-based neologisms.

References

Choi, J. (2020): On the methodology for the detection of Korean neologisms. In: *Language Facts and Perspectives* 51, pp. 151–172.

ten Hacken, P./Koliopoulou, M. (2020): Dictionaries, neologisms, and linguistic purism. In: *International Journal of Lexicography* 33 (2), pp. 127–134.

ten Hacken, P./Panacová, R. (eds.) (2020): *The interaction of borrowing and word formation*. Edinburgh.

Kim, E. J. (2016): Anglicized Korean neologisms of the new millennium: an overview. In: *English Today* 32 (3), pp. 52–60.

Kim, J. (2019): The flows and changes of Korean language purification policy. In: *Culture and Convergence* 41 (4), pp. 1279–1303.

Nam, K./Lee, S./Jung, H. (2020): The Korean Neologism Investigation Project: current status and key issues. In: *Dictionaries: Journal of the Dictionary Society of North America* 41 (1), pp. 105–129.

Nam, K./Lee, S./Choi, J./Seo, E./Kang, H./Baek, M./Jeong, H./Kim, H./An, J. (2021): *Neologisms of 2020. A new language of the COVID-19 pandemic*. Seoul.

Seo, H. (2019): On the meaning correspondence of refining words. In: *Korean Semantics* 64, pp. 131–153.

Standard Korean Language Dictionary (2002).

https://stdict.korean.go.kr/main/main.do#main_logo_id (last access: 27-03-2022).

Urimalsaem (2016). <https://opendict.korean.go.kr/main> (last access: 27-03-2022).

Contact information

Jun Choi

Kyungpook National University
c-juni@hanmail.net

Hae-Yun Jung

Kyungpook National University
haeyun.jung.22@gmail.com