On loans in Korean new word formation and in lexicography

Jun CHOI c-juni@hanmail.net

Hae-Yun JUNG haeyun.jung.22@gmail.com
Objective

- Examining the role of loans in Korean neologism formation
- Examining the relationships between borrowing, word formation and the status of loan-based neologisms in Korean dictionaries
**Standard Korean Language Dictionary (SKLD)**
- National Institute of Korean Language
- Print dictionary (1999)
- Digitalized since 2002
- 422,894 headwords
- stdict.korean.go.kr/main/main.do

**Urimalsaem (lit. ‘our language pond’)**
- National Institute of Korean Language
- Online dictionary (2016)
- Macrostructure based on the structure of SKLD
- Semi-crowdsourced dictionary
- 1,144,607 headwords
- opendict.korean.go.kr/main
Contents

1. Loans in Korean language
2. Word formation in Korean language
3. Data & methodology
4. Analysis & results
5. Significance & implications for Korean lexicography
1. Loans in Korean language

- **Loan in Korean**
  - Loanword
    - 마스크 *masukhu* EN mask
  - Loan morpheme
    - 언 *en* EN un- 언택트 *entaykthu* EN un- + EN [con]tact ‘no-contact’
  - Clipped loanword
    - 템 *theym* EN [i]tem 에코템 *eykhotheym* EN eco + EN [i]tem ‘eco-friendly product’

- **Hanja (Traditional Chinese) ≠ loan**
2. Word formation in Korean language

Korean words

Simplex words
- Single Native Korean or Hanja word
- Loanword
- Derivatives (root + affix)

Complex words
- Compounds (stem + stem)
- Blends (1+ clipped word(s))

Examples:
- 커엽다 kheyepta
  KO ‘cute’
- 코비디엇 khopities
  EN covidiot
- 메이비족 meyipicok
  EN maybe + HA cok ‘tribe’
- 프티숍 phuthisyop
  FR petit + EN shop ‘shop-in-shop’
- 에코템 eykhotheym
  EN eco + EN [i]tem
2. Word formation in Korean language

Loans in Korean language

Complex words in original language → simplex loanwords in Korean

Compounding, clipping of borrowed words/morphemes regardless of the morphological rules in the original language

Cryptojacking (EN crypto[currency] + EN [hi]jacking) → 크립토재킹 khulipthocaykhing

에코템 eykhotheyym EN eco + EN [i]tem

(ten Hacken & Panacová 2020)
3. Data & methodology

- 6,554 neologisms
- 3,413 loan-based neologisms

Korean Neologism Investigation Project (KNIP), National Institute of Korean language
Online news articles
2006~2020 (-2011)
### 3. Data & methodology

**Korean Neologism Investigation Project (KNIP) supervised by the National Institute of Korean Language (NIKL) from 1994-2019**

Investigation carried out by the Centre for Korean Language Information Studies at Kyungpook National University

<table>
<thead>
<tr>
<th>Year</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>Project temporarily discontinued</td>
</tr>
<tr>
<td>2011</td>
<td>Automatic extraction from a Web-based corpus of online news</td>
</tr>
<tr>
<td>2012</td>
<td>Automatic extraction from online news using a Web-based neologism extractor</td>
</tr>
<tr>
<td>2019</td>
<td>Neologism investigation carried out according to the same criteria as those of KNIP</td>
</tr>
<tr>
<td>2020</td>
<td>Project discontinued</td>
</tr>
</tbody>
</table>
4. Analysis & results

6,554 neologisms

3,413 loan-based neologisms

Predominance of loan-based blends (1,247) and compounds (1,526)

Predominance of English loans (95%)

52% of all neologisms
4. Analysis & results

Number of loan-based neologisms with at least 1 element from the borrowed language

<table>
<thead>
<tr>
<th>Language (ISO Code)</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>English (EN)</td>
<td>3235</td>
</tr>
<tr>
<td>French (FR)</td>
<td>97</td>
</tr>
<tr>
<td>Japanese (JA)</td>
<td>70</td>
</tr>
<tr>
<td>Italian (IT)</td>
<td>52</td>
</tr>
<tr>
<td>Chinese (ZH)</td>
<td>22</td>
</tr>
<tr>
<td>German (DE)</td>
<td>22</td>
</tr>
<tr>
<td>Latin (LA)</td>
<td>14</td>
</tr>
<tr>
<td>Greek (EL)</td>
<td>10</td>
</tr>
<tr>
<td>Spanish (ES)</td>
<td>5</td>
</tr>
<tr>
<td>Danish (DA)</td>
<td>4</td>
</tr>
<tr>
<td>Russian (RU)</td>
<td>3</td>
</tr>
<tr>
<td>Hindi (HI)</td>
<td>2</td>
</tr>
<tr>
<td>Pashto (PS)</td>
<td>2</td>
</tr>
<tr>
<td>Indonesian (ID)</td>
<td></td>
</tr>
<tr>
<td>Portuguese (PT)</td>
<td></td>
</tr>
<tr>
<td>Sanskrit (SA)</td>
<td></td>
</tr>
<tr>
<td>Turkish (TR)</td>
<td></td>
</tr>
</tbody>
</table>
4. Analysis & results

**French (97)**

- **vacances**
- 28 blends, 2 compounds, 1 phrase
- 32%

**바캉스 얼리 버드족 pakhangsu elli petucok**
FR vacances + EN early bird + HA cok ‘tribe’

**홀캉스족 holkhangsucok**
KO hol ‘solo’ + FR [va]cances + HA cok ‘tribe’

**몰캉스 molkhangsu**
EN mall + FR [va]cances
### 4. Analysis & results

<table>
<thead>
<tr>
<th>Language</th>
<th>Count</th>
<th>Example Words and Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japanese</td>
<td>70</td>
<td>otaku ‘geek’</td>
</tr>
<tr>
<td>German</td>
<td>22</td>
<td>Arbeit/Arbe[it]</td>
</tr>
</tbody>
</table>

오덕질 otekcil
JA otak[u] + KO jil ‘attitude’
‘comics/anime geek’

알바렐라 alpaleylla
DE Arbe[it] + EN [Cinde]rella
part-time working student
### 4. Analysis & results

<table>
<thead>
<tr>
<th>Language</th>
<th>Count</th>
<th>Description</th>
<th>Blends/Compounds</th>
<th>Blends/Compounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Italian</td>
<td>52</td>
<td>paparazzi</td>
<td>24 blends and 3 compounds</td>
<td>52%</td>
</tr>
<tr>
<td>Latin</td>
<td>14</td>
<td>homo + attribute</td>
<td>5 blends or 3 compounds</td>
<td>57%</td>
</tr>
</tbody>
</table>

**Italian**

- **paparazzi**: 24 blends and 3 compounds, 52%

**Latin**

- **homo + attribute**: 5 blends or 3 compounds, 57%

**Korean**

- **영파라치** (yengphalachi): HA yeng[hwa] ‘movie’ + IT [pa]parazzi ‘a person who reports movie pirates’
- **호모 체어쿠스** (homo cheyekhwusu): LA homo + EN chair + LA -[Australopithe]cus ‘long hours office worker’
5. Significance & implications for Korean lexicography

Representation of highly productive loans in *Urimalsaem*

<table>
<thead>
<tr>
<th>Loan</th>
<th>Number of related neologisms</th>
<th>Number of related neologisms included in Urimalsaem</th>
<th>Number of other related neologisms included in Urimalsaem which are not from our list</th>
</tr>
</thead>
<tbody>
<tr>
<td>pakhangsu (FR vacances)</td>
<td>4</td>
<td>23</td>
<td>31</td>
</tr>
<tr>
<td>othakhwu (JA otaku ‘geek’)</td>
<td>0</td>
<td>15</td>
<td>26</td>
</tr>
<tr>
<td>phaphalachi (IT paparazzi)</td>
<td>5</td>
<td>27</td>
<td>29</td>
</tr>
<tr>
<td>alupaithu, alpa (DE Arbeit)</td>
<td>4</td>
<td>18</td>
<td>29</td>
</tr>
<tr>
<td>homo + attribute (LA homo)</td>
<td>7</td>
<td>18</td>
<td>29</td>
</tr>
</tbody>
</table>
5. Significance & implications for Korean lexicography

Limitation of extracting neologisms from a single genre (news media)

User participation by suggesting new headwords (reviewed by language experts and potentially added if lexicographic value)
5. Significance & implications for Korean lexicography

- **Print dictionary**
- **More prescriptive**
- **More conservative towards neologisms and foreign words**

- **Online content features**
- **More inclusive**
- **Grows faster**

**Standard Korean Language Dictionary**

바캉스 *pakhangsu*  
FR *vacances*

아르바이트 *alupaithu*  
DE *Arbeit*
5. Significance & implications for Korean lexicography

Language purification policy by the Korean government (1970s)

- improper or dialectal words
  - correct, standard words
- complicated expressions
  - simpler expressions
- words of foreign origins
  - native Korean or Hanja words
5. Significance & implications for Korean lexicography

- **SKLD macrostructure**
  
  - **Native Korean words**
    - 21%
  
  - **Hanja**
    - 53%
  
  - **Foreign words**
    - 6%
  
  - **Hybrid words**
    - 20%

- **SKLD’s passivity before neologisms and especially loan-based neologisms**

  - **Urimalsaem**
    - More neologisms and loans
    - ‘Normative information’ in some loanword entries
5. Significance & implications for Korean lexicography

바캉스 *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 여름휴가 *yelum hyuka* (KO ‘summer’ + HA ‘vacation’) or 휴가 *hyuka* (HA ‘vacation’) if possible.
5. Significance & implications for Korean lexicography

부업 pwuep (HA ‘part-time job’) can be used along ‘Arbeit’
5. Significance & implications for Korean lexicography

Native Korean and Hanja neologism representation in *Urimalsaem* 45.1%

Loan-based neologism representation in *Urimalsaem* 54.9%
References


Thank you for your attention!

Q & A
4. Analysis & results

Number of blend neologisms containing the clipped French loan -$khangsu$ ([va]cances) to French-based blend neologisms per year
4. Analysis & results

vacances

바캉스

va- -can- -ces
4. Analysis & results

자부심 *capwusim*

-심 ‘a sense of’

a sense of pride

맵부심 *mayppwusim*

맵- ‘spicy’

[자]부심 ‘pride’

pride in eating (and handling) spicy food
4. Analysis & results

Greek (10)

alpha

5 compounds (50%)

3 compounds with beta and omega (80%)

알파맘 alpha mam
GR alpha + EN mum
‘a mother who pushes her children to success’

오메가족 omeykacok
GR omega + HA cok ‘tribe’
‘mother(s) of ‘alpha girl(s)’ (successful women)’