Mikyung Baek/Jinsan An/Yelin Go

A DISTRIBUTIONAL APPROACH TO KOREAN SEMANTIC NEOLOGISMS
Identifying their first occurrences and investigating their spread

Diachronic corpora have enabled extensive research on semantic change (Gulordava/Baroni 2011; Kim/ et al. 2014; Hamilton/Leskovec/Jurafsky 2016), from which both theoretical linguistics and applied linguistics can benefit. In lexicography in particular, the description of new meanings proves crucial not only for communication, but also language teaching and translation, including automatic translation. In neologism research, words that display a shift in meaning are referred to as semantic neologisms and may encompass various types of semantic changes. Broadly speaking, these may include metaphorical and metonymic uses, as well as widening and narrowing of meanings (Ullman 1957; Boussidian 2013; Nam/Lee/ Choi 2018). While Nam/Lee/Choi (2018) also add the time of first occurrence as an important criterion for determining semantic neologisms, a satisfying systematic methodology for identifying semantic neologisms is yet to be developed. Thus, this research seeks to overcome this shortcoming as an empirical study proposing a methodology to identify the time of first occurrence of four semantic neologisms (Examples 1–4) as well as the period when they settled in language.

(1)   인생 *insayng* 'lifetime' [S1], 'best' [S2]

(2)   폭풍 *phokphwung* 'storm' [S1], 'extreme' [S2]

(3)   공화국 *konghwakwuk* 'republic (as a political system)' [S1], 'country (of a particular type)' [S2]

(4)   영접 *yengcep* 'formal reception (of an official person)' [S1], 'welcome (of object/food)' [S2]

In order to achieve this, we first built a set of Web-crawled corpora consisting of news articles dating from January 1990 to December 2019, each subset corresponding to one semantic neologism in (1–4). We then extracted all noun phrases that comprise the above semantic neologisms either as a modifier or as a modificand. As this included both noun phrases where the word is used in its original meaning and noun phrases where it is used with a new meaning, we sorted out the noun phrases regarded as semantic neologisms by all three authors for analysis. In the next step, we applied the main two methodologies used in semantic change research, that is, collocation analysis (within a left/right window of four words) and word similarity analysis. The results yielded by these two methodologies only proved meaningful in the case of *phokphwung* 'storm' [S1], 'extreme' [S2], but were rather inconclusive for the others. Thus, we proposed another methodology here, which consisted of analyzing the collocate types (vs tokens) and the semantic categories of the collocates.

The collocate type analysis showed a clear pattern in all four semantic neologisms, which allowed us to determine the time of first occurrence and the establishment period, as shown in pictures 1–4.
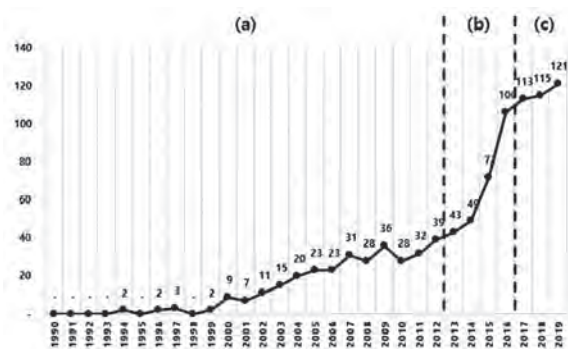
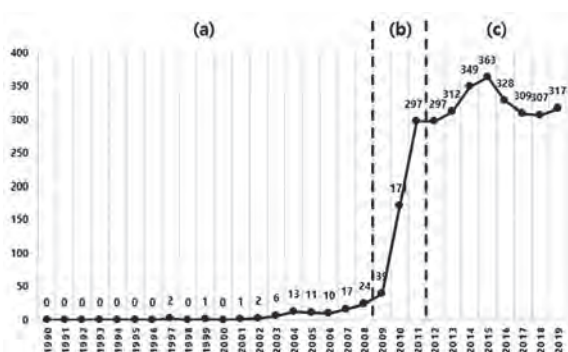**Fig. 1:** Collocate type pattern for '*insayng* [S2] + N'

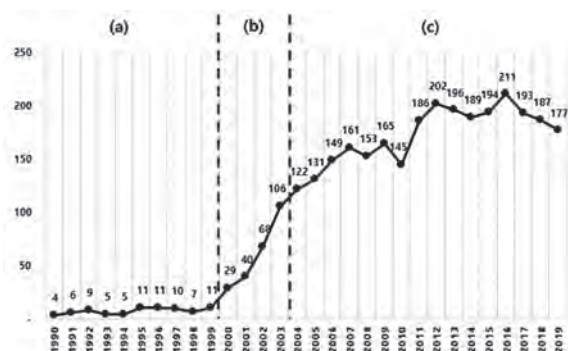**Fig. 2:** Collocate type pattern for '*phokphwung* [S2] + N'

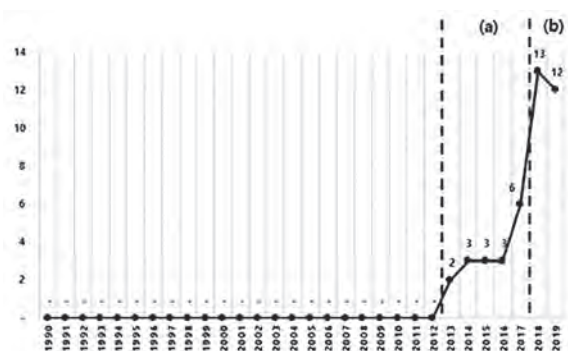**Fig. 3:** Collocate type pattern for 'N + *konghwakwuk* [S2]'

Fig. 4: Collocate type pattern for 'N + *yengcep* [S2]'

Frequency trends can be divided in three phases (*a*, *b*, *c*). Phase *a* corresponds to a low number of collocate types, which consists of nonce expressions. Phase *b* shows a sudden and rapid diversification of collocate types, corresponding to Schmid's 'consolidation' stage (2008). Finally, the diversification of types slows down in phase *c*, which indicates that the new meaning is being established. While this methodology seems to provide systematicity, it also presents some limitations. For example, the word *seytay* 'generation' is also typically used in such constructions as 'N + *seytay*'. When analyzing the collocate types of the *seytay* noun phrases, it appears to show a similar pattern. However, these noun phrases are not indicative of a semantic shift in the word *seytay*. One explanation lies in the semantic category the collocate falls into: in the case of 'N + *seytay*', N would most likely fall in the category of 'society'. This would be in the same line with the example presented by Gulodarva/Baroni (2011) whereby the distribution of the collocates for *sleep* may have changed (e.g., deep sleep (1960s); REM sleep (1990s)) but without impacting the meaning of *sleep*. However, as the semantic category analysis shows, the diversification of collocate types is concomitant with a clear semantic shift in the semantic category of the collocates, as shown in Table 1.

	[S1]	[S2]
insayng 'lifetime'	*insayng mokphyo* 'life goal'; *insayng keyhoyk* 'life plan'	*insayng yenghwa* 'best movie'; *insayng meynyu* 'best menu'
phokphwung 'storm'	*phokphwung kyengpo* 'storm warning'; *phokphwung yeypo* 'storm forecast'	*phokphwung sengcang* 'extreme growth'; *phokphwung cilcwu* 'extreme speeding'
konghwakwuk 'republic'	*mincwu konghwakwuk* 'democratic republic'; *cheykho konghwakwuk* 'Czech Republic'	*senghyeng konghwakwuk* 'the country of plastic surgery'; *aphathu konghwakwuk* 'the country of apartments'
yengcep 'formal reception'	*taythonglyeng yengcep* 'formal reception of the President'; *chongli yengcep* 'formal reception of the Prime Minister'	*chikhin yengcep* 'welcome of the chicken'; *sutheyikhu yengcep* 'welcome of the steak'

Table 1:   Comparative table of the collocates for the semantic neologisms under study

The semantic category of the collocates for the original meaning of the words in question is rather restricted: 'life/lifestyle' for *insayng*; 'nature/weather forecast' for *phokphwung*; 'political system' for *konghwakwuk*; 'political figure' for *yengcep*. However, the collocate types have greatly diversified in terms of semantic categories in the case of semantic neologisms. Only *yengcep* shows a more restricted semantic category pattern; nonetheless, the categories of the original and new meanings are semantically different enough to consider *yengcep* as a semantic neologism.

The methodology we presented in this study showed that the distribution of the collocate types somewhat evidences semantic change. A sole quantitative analysis proved insufficient and instead, a qualitative analysis allowed us to refine our findings. Nonetheless, our study presents a number of limitations, including the scale of our data as our corpora only covers a time span of thirty years and the selection criterion for the semantic neologisms under study, which is ultimately based on our intuition of native speakers.

References

Boussidan, A. (2013): Dynamics of semantic change. PhD Dissertation. Lyon.

Gulordava, K./Baroni, M. (2011): A distributional similarity approach to the detection of semantic change in the Google Books Ngram corpus. In: Proceedings of the GEMS 2011 Workshop on Geometrical Models of Natural Language Semantics. Edinburgh, pp. 67–71.

Hamilton, W. L./Leskovec, J./Jurafsky, D. (2016): Diachronic word embeddings reveal statistical laws of semantic change. arXiv preprint arXiv:1605.09096.

Kim, Y./Chiu, Y. I./Hanaki, K./Hegde, D./Petrov, S. (2014): Temporal analysis of language through neural language models. arXiv preprint arXiv:1405.3515.

Nam, N. Y./Lee, S. J./Choi, J. (2018): Research trends and issues on semantic neology using web corpus. In: Korean Lexicography 31, pp. 55–84.

Schmid, H. (2008): New words in the mind: concept-formation and entrenchment of neologisms, In: Anglia 126 (1), pp. 1–36.

Ullmann, S. (1957): The principles of semantics. Oxford.

Contact information

Mikyung Baek
Kyungpook National University
bmg0128@hanmail.net

Jinsan An
Kyungpook National University
siveking@naver.com

Yelin Go
Kyungpook National University
goyelin08@naver.com