ten years ago are still in use. These tasks are carried out using a web-based neologism extractor and a web crawler respectively. The extraction of new words is performed semi-automatically, since the automatic web-based neologism extractor is combined to manual identification. Since 2012, all the neologisms collected for task 1 have been added to the database of the online dictionary *Urimalsaem*, which became accessible to the public in 2016. *Urimalsaem* and the *Standard Korean Language Dictionary* (SKLD) are the main dictionaries of the Korean language. Both are state-run dictionaries, but have nonetheless distinct identities. *Urimalsaem* is a partly crowdsourced dictionary that enables contribution of dictionary users, while SKLD is a prescriptive dictionary for the use of standard language and grammar. As a result of task 2, the neologisms that are still in continuous use after ten years can be considered as headword candidates for SKLD.

At the outset in 1994, the methodology adopted for the project consisted of reading texts and searching for new words with the naked eye. Crucial methodological changes have been introduced since then, including the construction of a large-scale corpus (2005) and the use of the web crawler and web-based neologism extractor (2012). In 2015, a ten-year usage investigation for the neologisms extracted in 2005 and 2006 began. The following year, a pattern-based methodology of neologism extraction was introduced, and the minimum threshold of frequency occurrence for neologism candidates was increased to three. Despite these adjustments, the precision and recall levels of automatic neologism detection are still not satisfactory. Moreover, there are a number of other issues for improvement that are addressed in this paper, such as the difficulty of conducting a consistent frequency survey due to the dynamic nature of the web as corpus, the identification of semantic neologisms that are not morphological neologisms, and the dependency on manual processes. Some of these issues can be approached in terms of Korean natural language processing or from a typological perspective of Korean as an agglutinative language. In their ten-year cycle investigation of neologism usage, Nam et al. (2016) have found that only 75% of the neologisms survived after ten years. Whether this result constitutes a suitable criterion for lexicographic inclusion is also re-examined in the current study.

Keywords: Korean neologisms, neologism extraction, neologism usage investigation, headword candidates, *Urimalsem*, *Standard Korean Language Dictionary*

References

Barnhart, D.K. 2007. A Calculus for New Words. *Dictionaries: Journal of the Dictionary Society of North America* 28, 132-138.

Nam, K., Lee, S., Jung, H.-Y. and Choi, J. 2016. The Life and Death of Neologisms: On What Basis Shall We Include Neologisms in the Dictionary? In *Proceedings of the XVIII EURALEX International Congress,* 389-393.

***Standard Korean Language Dictionary* [SKLD].** http://stdweb2.korean.go.kr/main.jsp.

Urimalsaem. https://opendict.korean.go.kr/main.

New words in Japanese and the design of *UniDic* electronic dictionary

Teruaki Oka

Teruaki Oka graduated from Toyohashi University of Technology in 2010, and received his masters and PhD degrees in Engineering from Nara Institute of Science and Technology (NAIST), Ikoma, in 2012 and 2015, respectively. From 2015 to 2016 he was a Program-Specific Researcher at Kyoto University, and in 2016 he joined the National Institute for Japanese Language and Linguistics (NINJAL), where he currently serves as a Project Assistant Professor. His research interests are computational and corpus linguistics.
teruaki-oka@ninjal.ac.jp

The National Institute for Japanese Language and Linguistics (NINJAL) is involved in developing Japanese language corpora, including the Balanced Corpus of Contemporary Written Japanese, Corpus of Spontaneous Japanese, Corpus of Historical Japanese, and NINJAL Web Japanese Corpus. In the development processes we often encounter new words that are formed by using different character types (e.g., Hiragana, Katakana, Kanji) and their heterographs, with their combinations, even for writing a single word (e.g., *big*: おおきい, 大きい, オオキイ, ぉぉきぃ, 大キィ), which could be ‘literal’ (e.g., *as it was expected*: 矢張り), ‘somewhat colloquial’ (やっぱり), ‘colloquial’ (やっぱし), ‘abbreviated’ (やぱ), and so on. Thus, new words can appear as orthographic variants (おおきい vs. 大キィ), form variants (矢張り vs. やぱ) and new lemmas (such as エモい *emotional*), and be classified at these three levels (orthographic, form, lemma).

We apply a design policy called “hierarchical definition of word indexes” to register new words in *UniDic*, our electronic Japanese word dictionary, for annotating plain texts with morphological information. Using the hierarchical definition of word indexes, a single lemma (e.g., 矢張り) has its various word forms written in Katakana characters (e.g., 矢張り←ヤハリ, ヤッパリ, ヤッパシ, ヤパ) as its children, with each form having its orthographic variants as its children (e.g., ヤハリ←矢張り, やはり, ヤハリ). *UniDic* contains about 200 thousand lemmas and one million of their form and orthographic variants with rich morphological information (e.g., part of speech, lemmatized form, pronunciation, accent). To annotate morphological information in plain unsegmented texts, we select optimal records for character strings in the texts from UniDicDB, a word database system. The records and their morphological information are manually registered to UniDicDB when new words are detected during the annotation phase. We also employ UniDicExplorer, an annotator-friendly user interface capable of searching and registering words. Another feature is UniDicMA, a dictionary software for the morphological analyzer, which is derived from UniDicDB and can attach the hierarchical structure of *UniDic* to each word in an input plain unsegmented text automatically (https://unidic.ninjal.ac.jp/). Only UniDicMA is open to the public, whereas all other UniDics are not accessible outside NINJAL.

In this paper, we discuss what is a ‘new word’ in Japanese, our hierarchical definition of word indexes, and how to register new words in UniDicDB using UniDicExplorer.

Keywords: electronic dictionary, Japanese, corpus, annotation, database system, morphological analyzer, neologisms

Adding neologisms to the Hebrew online dictionary *Rav-Milim*

Noga Porath

Noga Porath has studied at the Department of Hebrew Language in Tel-Aviv University, and received a PhD for her dissertation examining metaphors in the language of developmental cognitive psychology and special education in 2017. She is a lexicographer at Melingo Ltd, which publishes online the Hebrew dictionary *Rav-Milim* and the English/Hebrew dictionary *Morfix*.
nogap@melingo.com

This paper describes the process of finding Hebrew neologisms and adding them to the online dictionary *Rav-Milim*. The editorial board of the dictionary uses different methods to find such neologisms, including crowdsourcing (suggestions from users), and tracking new terms in the media and in official announcements by the Academy of the Hebrew Language. We discuss the criteria and methodology for adding new words to the dictionary, with emphasis on the decision-making process of labelling foreign words (mainly from English) as neologisms in Hebrew. Various kinds of neologisms have been added to the dictionary in recent years: new technological terms, including terms for new tools and appliances (רחפן, *rachfan*, ‘drone’); internet and social media slang; terms that have emerged in recent years in media coverage of news events; terms that have arisen in recent general discourse regarding new concepts (מזון-על, *mezon-al*, ‘superfood’); new military terms; neologisms added by the Academy of the Hebrew Language, some of which are the equivalents of existing loanwords. Most of these types of neologisms include loanwords, that are mainly borrowed from English.

Our dictionary is a practical, descriptive tool rather than an etymological documentation project. Therefore, new words in the dictionary are, in general, not indicated as such, though we do note whether a neologism has been formally suggested by the Academy of the Hebrew Language. These neologisms are linked to earlier loanwords with the same meaning.

Rav-Milim has also added new meanings to existing entries. New technological meanings have emerged in words like ענן (*anan*, ‘cloud’). In other cases, existing terms have been replaced with new ones due to considerations of political correctness in contexts such as gender and disability.

GWLN 2019. **Globalex Workshop on Lexicography and Neologism** https://globalex.link/events/workshops/gwln2019/

DSNA. **Dictionary Society of North America** https://dictionarysociety.com/

DSNA 22. **22nd biennial meeting of the Dictionary Society of North America** https://indiana.edu/~iucweb/dsna/

DICTIONARIES. **Journal of the Dictionary Society of North America** https://dictionarysociety.com/journal/

GLOBALEX. **Global Association for Lexicography** http://globalex.link/

Keywords: neologisms, Hebrew, foreign words, internet slang

The formation of neologisms in a lesser used language: The case of Frisian

Hindrik Sijens and Hans Van de Velde

Hindrik Sijens studied Frisian language and literature and lexicography at the University of Amsterdam, and has written on neologisms, spelling and lexicography. He is a lexicographer at the Fryske Akademy at Leeuwarden/Ljouwert, and currently serves as editor of the *Online Dutch-Frisian Dictionary* and of *Taalweb*, a website with Frisian language tools such as online dictionaries, spelling tools and automatic translation.
hsijens@fryske-akademy.nl

Hans Van de Velde is chair of sociolinguistics at Utrecht University, and specializes in language variation and change and in standardization processes. He is a senior researcher at the Fryske Akademy, focusing on Frisian, Dutch and the mixed varieties spoken in Friesland, and is responsible for the development of Frisian language tools such as online dictionaries, spelling tools, automatic translation and speech recognition.
hvdvelde@fryske-akademy.nl