Abstracts from the Globalex Workshop on Lexicography and Neologism 2019

The Globalex Workshop on Lexicography and Neologism (GWLN 2019) was held in conjunction with DSNA22, the 22nd biennial meeting of the Dictionary Society of North America, at Bloomington, Indiana, on May 8, 2019. It brought together 13 papers on 12 languages from Africa, Asia, Europe and North America (two on English), highlighting issues related to the detection of neologisms – including new words, new meanings of existing words, and new multiword units – and their representation in lexicography and dictionaries, such as:

- How to find neologisms (corpus analysis and editorial means of identification; evaluation of data, e.g. blogs and chats)
- How to interoperate lexicographic datasets with online resources and incorporate neologisms into the digital dictionary (the media, formatting, labeling, etc.)
- How to deal with grammatical/orthographic/ pronunciation variation (description vs. prescription)
- How to explain meaning with/without encyclopedic information, and how to use illustrations and audio-visual media
- How differently, if at all, should neologisms be treated in different dictionary types (e.g. in historical comprehensive ones as opposed to those focusing on current usage; in monolingual vs. bilingual dictionaries; in special domain dictionaries)
- How to deal with neologisms that are no longer *new* and those no longer used
- How can dictionary users help with finding and informing about neologisms

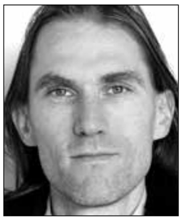
The proceedings of GWLN 2019 are undergoing peer-review for publication in 2020 as a special issue of *Dictionaries: Journal of the Dictionary Society of North America*, guest edited by the workshop organizers Annette Klosa-Kückelhaus and Ilan Kernerman. The presentation slides are available from the GWLN 2019 website.

Annette Klosa-Kückelhaus holds an MA and PhD in German linguistics from the universities of Munich and Bamberg. She has been a lexicographer for *Duden* and has (co-)authored extensively on lexicography. Currently she heads the Lexicography and Language Documentation area and is chief editor of an online dictionary of neologisms at the Leibniz-Institute for the German Language (IDS) at Mannheim.
klosa@ids-mannheim.de

Ilan Kernerman heads K Dictionaries, leading the development of resources, collaboration with industrial, academic and other professional partners, and interoperability with other domains. He is a former President of Asialex and currently chairs the Globalex management committee.
ilan@kdictionaries.com

Linguistics terminology and neologisms in Swahili: Rules vs. practice

Gilles-Maurice de Schryver and Jutta De Nul



Gilles-Maurice de Schryver is the President of Euralex (2018-2020), and a two-term past President of Afrilex (2009-2013). He holds an MSc in microelectronic engineering, as well as an MA and PhD in African languages and cultures from Ghent University. Currently a research professor at the Ghent University Centre for Bantu Studies, he has (co-)authored about 300 books, book chapters, journal articles and conference papers on lexicography.
gillesmaurice.deschryver@ugent.be

Jutta De Nul obtained an MA in African languages and cultures at Ghent University, and is currently working on her PhD entitled ‘Theoretical underpinnings for a user-friendly, corpus-driven, semi-bilingual, digital dictionary of Swahili’ under the supervision of Gilles-Maurice de Schryver and Koen Bostoen.
jutta.denul@ugent.be

In this paper we discuss the use of Swahili terminology in the field of linguistics. In particular, we are interested in finding out whether the rules laid out by scholars in the scientific literature for the creation of terminological neologisms in Swahili correspond with actual practice. In order to do this, three steps are taken. In Step 1 we undertake the semi-automatic extraction of linguistics terminology, by comparing occurrence frequencies in a special-purpose corpus consisting of ten Swahili language/linguistics textbooks, with their corresponding frequencies in a 22-million-token general-language reference corpus. In Step 2 we study the source languages and actual word formation processes of the terms and neologisms with the highest keyness values obtained during the previous step. This discussion is divided into several sections, one section per source language. In Step 3, the terms and neologisms that have been found are compared with their treatment (or absence thereof) in two existing reference works, a general dictionary and a linguistics terminology list. These three steps are preceded by brief introductions to (i) the Swahili language; (ii) its dictionaries and terminology lists; (iii) its metalexicographical, terminological and neologism studies; and (iv) our use of the term ‘neologism’. The three steps are followed by a discussion of our findings and a conclusion.

Keywords: Bantu, Swahili, corpora, semi-automatic term extraction, linguistics terminology, terminological neologisms, terminology, lexicography, digital dictionaries

References

Baker, M. 1992. *In Other Words: A coursebook on translation*. London: Routledge.

de Schryver, G.-M. 2008. Why does Africa need Sinclair? *International Journal of Lexicography* 21.3: 267-291.

de Schryver, G.-M., Joffe, D., Joffe, P. and Hillewaert, S. 2006. Do Dictionary Users Really Look Up Frequent Words? – On the Overestimation of the Value of Corpus-based Lexicography. *Lexikos* 16: 67-83.

Gibbe, A.G. 2008. Maendeleo ya istilahi za Kiswahili [Development of Swahili terminology]. In Kiango, J.G. (ed.), *Ukuzaji wa Istilahi za Kiswahili,* 79-99. Dar es Salaam: TUKI.

Hillewaert, S. and de Schryver, G.-M. 2004. Online Kiswahili (Swahili) – English Dictionary. http://africanlanguages.com/swahili/ (accessed May 1, 2009).

Massamba, D.P.B. 2004. *Kamusi ya Isimu na Falsafa ya Lugha [Dictionary of Linguistics and the Philosophy of Language]*. Dar es Salaam: TUKI.

Mtintsilana, P.N. and Morris, R. 1988. Terminography in African languages in South Africa. *South African Journal of African Languages* 8.4: 109-113.

Mwansoko, H.J.M. 1990. *The Modernization of Swahili Technical Terminologies: An investigation of the linguistics and literature terminologies*. unpublished PhD dissertation. University of York, York.

Mwansoko, H.J.M. 2001. Uboreshaji wa mfumo wa uingizaji wa istilahi za kimataifa katika Kiswahili [Improvement of the system for inserting international terminology into Swahili]. In Mdee, J.S. and Mwansoko, H.J.M. (eds.), *Makala ya kongamanola kimataifa Kiswahili 2000. Proceedings,* 318-331. Dar es Salaam: TUKI.

Scott, M. 1996-2019. WordSmith Tools. http://www.lexically.net/wordsmith/ (accessed May 1, 2019).

Taljard, E. and de Schryver, G.-M. 2002. Semi-automatic term extraction for the African languages, with special reference to Northern Sotho. *Lexikos* 12: 44-74.

Tumbo-Masabo, Z.N.Z. 1990. *The Development of Neologisms in Kiswahili: A diachronic and synchronic approach with special reference to mathematical terms*. unpublished PhD dissertation. Teachers College, Columbia University, New York, NY.

Tumbo-Masabo, Z.N.Z. 1992. Uundaji wa istilahi za Kiswahili [Formation of Swahili terminology]. In Tumbo-Masabo, Z.N.Z. and Mwansoko, H.J.M. (eds.), *Kiongozi cha Uundaji wa Istilahi za Kiswahili [Guidelines for the Formation of Swahili Terminology]*, 21-42. Dar es Salaam: TUKI.

Zgusta, L. 1971. *Manual of Lexicography*. (Janua Linguarum Series Maior 39). Prague/The Hague: Academia/Mouton.

Beyond frequency: On the dictionarisation of new words in Spanish

Judit Freixa and Sergi Torner

Judit Freixa is a tenured lecturer at the Department of Translation and Language Sciences in the Universitat Pompeu Fabra, Barcelona. She is the director of Observatori de Neologia (in IULATERM group) and her research focuses on the lexicon, specifically on neology and terminology, with particular attention to sociolinguistic aspects.
judit.freixa@upf.edu

Sergi Torner is a tenured lecturer of Spanish language at the Department of Translation and Language Sciences in the Universitat Pompeu Fabra, Barcelona. He is the principal investigator of the InfoLex research group, whose members do research on lexicography in Spanish. His research focuses on Spanish lexicon in the major areas of lexical semantics and lexicography, with particular attention to learners' lexicography.
sergi.torner@upf.edu

The most recent literature on neology has discussed the criteria that must be taken into account in order to include new words in dictionaries (Metcalf 2002, Barnhart 1985, Cook 2010, Ishikawa 2006, O'Donovan and O'Neill 2008, Freixa 2016, Sanmartín 2016, among many others). Although there are other factors that must be considered, such as morphologic features or semantic transparency (Adelstein and Freixa 2013, Bernal et al. 2018), authors broadly agree that frequency plays a central role, given that high frequency in a corpus may be taken as evidence of the institutionalization of a lexical unit. However, it has also been pointed out that frequency is a complex criterion in itself, and, therefore, aspects such as stabilization in use (Cook 2010) or a possible longitudinal change in frequency (Metcalf 2002, Ishikawa 2006) must also be taken into account when measuring frequency in corpora.

In this presentation, we approach lexical frequency as a criterion to evaluate whether neologisms must be included in Spanish dictionaries from a new point of view. Specifically, we compare data concerning change in frequency of neologisms through time with the speakers' perception about their newness, known as 'neological feeling' in the specialized literature (Gardin et al. 1974, Salayrolles 2003). Data about speakers' perception are obtained from online questionnaires carried out within the framework of the Neómetro project[1] (Bernal et al.

[1] 'NEÓMETRO: La medición de la neologicidad y la diccionariabilidad de los neologismos del español' project - supported by the Ministerio de Economía y Competitividad (ref. FFI2016- 79129-P), and

in press). A set of questionnaires was launched in which 100 subjects evaluated their perception of about 130 neologisms in Spanish according to four different criteria (correct formation, frequency, novelty and necessity of inclusion in dictionaries). On the other hand, frequency data are taken from an extensive corpus of texts from the press, FACTIVA, which provides histograms of frequency through time.

For this study, we analyze 40 neologisms that were perceived as the most and the least frequent in the questionnaires. We analyze their frequency curve through time in FACTIVA to find correlations between stabilization in time and speakers' perception about their institutionalization. The data allow us to improve the predictive capacity of frequency as a measure to decide which neologisms must be included in dictionaries, as it introduces factors (formal, semantic, or of use) that favor or hinder institutionalization in the equation.

Keywords: Spanish, neologism, frequency, histogram, institutionalization

References

Adelstein, A. and Freixa, J. 2013. Criterios para la actualización lexicográfica a partir de datos de observatorios de neología. Unpublished presentation, Congreso Internacional El Diccionario: neología, lenguaje de especialidad, computación, Ciudad de México, 28-30 October 2013. https://repositori.upf.edu/handle/10230/34891

Barnhart, D.K. 1985. Prizes and pitfalls of computerized searching for new words for dictionaries. *Dictionaries* 7, 253-260.

Bernal, E., Freixa, J. and Torner, S. 2018. Frecuencia, estabilidad... ¿y después qué? Criterios para la diccionarización de neologismos. Unpublished presentation, VII Congreso Internacional de Lexicografía Hispánica, Valencia, 28 June 2018.

Bernal, E., Freixa, J. and Torner, S. in press. Néologicité et dictionnairiabilité: deux conditions inverses?. In Proceedings of *4ème Congrès international de néologie des langues romanes, Lyon, 4 July 2018.*

Cook, C.P. 2010. *Exploiting Linguistic Knowledge to Infer Properties of Neologisms*. PhD dissertation. Toronto: University of Toronto.

[Factiva] Dow Jones. 1989–. *Factiva.* New York: Dow Jones & Company. https://global.factiva.com.

Freixa, J. 2016. Reagrupamiento lexicográfico de neologismos frecuentes. In Bernal, E. and Torner, S. (eds.), *Relaciones morfológicas y diccionario.* A Coruña: Universidade da Coruña, 95-113.

Gardin, B., Lefevre, G., Marcellesi, C. and Mortureux, M.F. 1974. A propos du sentiment néologique. *Langages 36,* 45-52.

Ishikawa, S. 2006. When a word enters the dictionary: A data-based analysis of neologism. In JACET Society of English Lexicography, *English Lexicography in Japan.* Bunkyo-ku: Taishukan, 39-52.

Metcalf, A. 2002. *Predicting New Words*. Boston, MA: Houghton Mifflin Company.

O'Donovan, R. and O'Neill, M. 2008. A systematic approach to the selection of neologisms for inclusion in a large monolingual dictionary. In Bernal, E. and DeCesaris, J. (eds.), *Proceedings of the XIII EURALEX International Congress (Barcelona, 15-19 July 2008).* Barcelona: IULA-UPF, 571-579.

Sablayrolles, J.-F. 2003. Le sentiment néologique. In Sablayrolles, J.-F. (ed.), *L'innovation lexicale*. Paris: Honoré Champion, 279-295.

Sanmartín, J. 2016. Sobre neología y contextos de uso: Análisis pragmalingüístico de lo ecológico y de lo sostenible en normativas y páginas web de promoción turística. *Ibérica* 31, 175-198.

founded by the Agencia Estatal de Investigación (AEI) and the Fondo Europeo de Desarrollo Regional (FEDER) - is led by Judit Freixa and Sergi Torner.

New words for the *Duden*

Kathrin Kunkel-Razum

Kathrin Kunkel-Razum studied German and History at Leipzig University, where she received her PhD in Linguistics (phraseology of the German language) in 1986. She was editor at *Zeitschrift für Germanistik* (1986-1991), and teacher of German as a foreign language at Universidad Complutense de Madrid. In 1992 she became an editor at Duden, where she is editor-in-chief since 2016. kathrin.kunkel-razum@duden.de

Every three or four years there is a new printed edition of the *Rechtschreibduden* [*Duden's Orthographic Dictionary*], the most well-known dictionary of the German language. The past five or six editions boasted approximately 5,000 newly added lemmata each (currently available: the 27th edition with 145,000 lemmata), and since 1996, the year of the "Rechtschreibreform" [national reform of orthography], public response to each new edition has focused primarily on these new additions. When a word is included in the *Duden* it is considered to have become officialized. There are people who wonder whether words not included in the *Rechtschreibduden* exist, although even its online version offers an additional 100,000 lemmata.

So, what are the criteria applied by the *Duden*'s editorial staff when deciding which new words to include? Which sources are used? What is the editors' position in the ongoing discussion about the – arguably excessive – use of Anglicisms in the German language and the addition of terms and grammatical adaptions related to or dictated by political correctness? How about the ratio of new entries in the printed edition of *Rechtschreibduden* as opposed to its online version, and what are the procedures for inclusion? On what grounds, finally, are words deleted from the dictionary?

In this paper I refer to these issues and, with regard to future editions of *Rechtschreibduden*, I also talk about which new sources the *Duden* will have to consider and work with to remain the predominant dictionary of the German (standard) language.

Keywords: German, orthographic dictionary, Anglicisms, print vs. online dictionary

New Estonian words and senses: Detection and description

Margit Langemets, Jelena Kallas, Kaisa Norak and Indrek Hein

Margit Langemets (PhD) is a senior lexicographer and the chief editor of dictionaries at the Institute of the Estonian Language. Her research interests include e-lexicography, corpus linguistics and lexical semantics. She has been involved in several bilingual and monolingual dictionary projects, as well as in the development of the in-house Ekilex dictionary writing system.
margit.langemets@eki.ee

Jelena Kallas (PhD) is a computational lexicographer at the Institute of the Estonian Language. Her research interests include corpus lexicography, automated lexicography, scholarly lexicography, dictionary use and innovative ways for presenting lexicographic data. She has been a member of the Euralex Executive Board since 2014 and on the organizing and scientific committee of the eLex conferences since 2013.
jelena.kallas@eki.ee

The web era has brought about the urgent need for the automatic monitoring of language, including the extraction of new words and senses. In order to monitor language, especially lexical changes, the Institute of the Estonian Language, in cooperation with Lexical Computing Ltd., crawls the web every two years. Corpora are used through the corpus query system Sketch Engine (Kilgarriff et al. 2004)[2] and CQS KORP[3]. The most recent corpus is the Estonian Reference Corpus 2017 (1.1 billion words); the next corpus will be crawled in 2019. We also implement crowdsourcing techniques for neologism registration by offering our users the opportunity to propose new words or senses. They can do this by using the feedback forms on our dictionary portals Sõnaveeb (‘Wordweb’)[4] and e-keelenõu (‘e-Language advice’)[5].

In this paper, we present the results of an experimental study on neologism detection on the basis of text collection, which was compiled at the Institute from 2016 to 2018. We describe the method for neologism detection and evaluate the results. This is the first study for Estonian aimed at the development of a tool to supply lexicographers with neologism candidates for inclusion in a dictionary.

In addition, we discuss the practice of providing both prescriptive and descriptive information about new words.

The prescriptive data concern mostly orthography and inflection and should point out what belongs to standard Estonian and what does not. However, it is not a trivial task dealing with neologisms. Within the unified single database Ekilex[6], we will present both descriptive and prescriptive data.

Keywords: neologisms, corpus lexicography, dictionary portal, Estonian

References

Kilgarriff, A., Rychly, P., Smrž, P. and Tugwell, D. 2004. The Sketch Engine. In *Proceedings of the XI Euralex International Congress, (eds.),* Williams G. and Vessier, S. Lorient: Université de Bretagne Sud, 105–116.

A system for evaluating multiple data inputs to prioritize neologisms for inclusion in dictionaries

Katherine Connor Martin

Katherine Connor Martin holds degrees in history from Yale University and the University of Iceland. Her career in lexicography began in 2003 as an editor for the *Oxford English Dictionary*, and currently she is Head of Lexical Content Strategy at Oxford University Press, New York.
katherine.martin@oup.com

With today’s massive web-based corpus resources, the key challenge facing lexicographers of new words in languages with a major digital presence is no longer *identification* of neologisms, but rather *prioritization* for inclusion in the dictionary. There are many possible data points that can be leveraged to prioritize the most editorially significant from among tens of thousands of candidates, including frequency in corpora, evidence of reader interest via web searches, prior registers of the word’s existence, and salience of the item in particular regions, registers, or domains of editorial interest. The most effective way to use these data inputs is to take a holistic approach, considering multiple factors simultaneously. This paper will discuss the use of a

[2] https://sketchengine.eu/ (accessed March 30, 2019)
[3] https://korp.keeleressursid.ee/ (accessed March 30, 2019)
[4] https://sonaveeb.ee (accessed March 30, 2019)
[5] http://keeleabi.eki.ee/ (accessed March 30, 2019)
[6] https://ekilex.eki.ee (accessed March 30, 2019)

new system, Oxford's New Words Prioritization Engine (NWPE), developed by Oxford Dictionaries to facilitate prioritization of large sets of candidate words by combining multiple sources of data in a single interface for analysis and by capturing human judgments about particular words so that they can be leveraged to improve future results.

Keywords: corpora, neologisms, prioritization

Using the Hypothes.is web annotation tool for neologism collection

Erin McKean

Erin McKean is the founder of Wordnik.com and works on open source strategy for Google. She was the editor-in-chief for American Dictionaries at Oxford University Press, and the editor of the *New Oxford American Dictionary, 2E.* She has written books on words and on dresses, was a regular columnist for *The Boston Globe* and the *Wall Street Journal*, and has served as an advisor to the American National Corpus, *American Speech,* and the Wikimedia Foundation.
erin@wordnik.com

Dictionary citation collection programs (sometimes called 'reading programs') involving both dedicated amateurs and paid professionals are not new, but have often required either cumbersome marking of print materials or creation of paper slips or access to private computer systems specific to individual projects. However, given the development and adoption of open standards for web annotation, citation collection by readers in and outside of dictionary programs can now be done easily without expensive proprietary tools or resorting to paper slips.

In this paper, we give an overview of Wordnik's reading program (currently in beta), which uses the free and open-source Hypothes.is web annotation tool to select, tag, and share citations from the open web directly for use on Wordnik.com. Using the Hypothes.is API, it is possible to import user-generated citations and their accompanying metadata directly into editorial workflows, including importing into KWIC corpora or other databases.

Since Wordnik is a radically inclusive dictionary (all words are eligible for inclusion), we discuss how this approach influences readers' marking of terms, and whether terms selected by readers are more likely to be typical neologisms (newly-coined words) or words overlooked by traditional dictionaries (e.g. jargon, slang, nonce, or other low-frequency words).

Keywords: dictionary users, web annotation, neologisms, hypothes.is, free-range definition

The Korean Neologism Investigation Project: Current status and key issues

Kilim Nam, Soojin Lee and Hae-Yun Jung

Kilim Nam has a PhD in Korean linguistics (on the copula *ida* structures in contemporary Korean, 2004) from Yonsei University (Seoul). She is a professor at the Department of Korean Language and Literature in Kyungpook National University (Daegu), has been the principal investigator of the Korean Neologisms Investigation Project since 2012, and is currently a board member of Korealex. Her research focuses on corpus linguistics and language performance.
nki@knu.ac.kr

Soojin Lee is a lecturer at the Department of Korean Language and Literature in Kyungpook National University (Daegu), where she obtained her MA (on academic keywords) and is doing her PhD. She has been a member of the research group for the Korean Neologisms Investigation Project since 2012. Her research interests include lexicography, lexicology as well as neology.
sjmano27@naver.com

Hae-Yun Jung received her MA in Korean Studies from SOAS (London) and is currently doing her PhD at Kyungpook National University (Daegu) under the supervision of Kilim Nam. Her PhD thesis is concerned with the treatment of pragmatic information in bilingual French-Korean lexicography, with particular attention to politeness. Her research interests include lexicography and cross-cultural pragmatics.
haeyun.jung.22@gmail.com

This paper reports on the Korean Neologism Investigation Project and discusses a number of unresolved issues related to neologism research. Since 1994, when the Korean government initiated the project, the use of the Internet and mobile phones has increased exponentially and the methods and scope of the investigation into Korean neologisms have been modified accordingly. The two major tasks carried out within the scheme of the project consist of (1) collecting all the neologisms that appear each year in news articles on the Naver portal, and (2) investigating the usage development of neologisms within the past decade in order to determine whether those collected

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.

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/

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.

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

Frisian is the language spoken in the Dutch Province of Friesland. Its approximately 440,000 speakers use it mainly for informal and oral communication. Dutch is the official language in the Netherlands, also in Friesland. With approximately 24 million speakers worldwide, Dutch is used in almost all areas of society. It is a widely supported standard language with a large written production.

Frisian has a limited tradition as a written language and consequently has a large number of lexical gaps. For many Dutch or international concepts, there are simply no Frisian equivalents. When it comes to new words, Frisian does not keep pace with Dutch either. Because of the limited use of Frisian and the omnipresence of Dutch, there are almost no spontaneously formed Frisian neologisms. Dutch neologisms often have a Frisian equivalent that is based on Dutch or no equivalent at all. Sometimes Dutch words are adopted literally, sometimes they are adapted in the pronunciation or replaced by a loan translation. Because Frisians live in a dominant Dutch context and have an excellent command of this language (as opposed to [written] Frisian), they easily adopt Dutch neologisms.

However, there is an unmistakable, partly ideologically-driven, effort towards a certain standardization in written language, which creates a need for Frisian variants of neologisms. This endeavour to purify Frisian has an impact on the treatment of neologisms in dictionaries. The a-symmetrical bilingual situation outlined above also has its impact on the spontaneous creation of Frisian neologisms and their subsequent incorporation in dictionaries of Frisian.

De Fryske Akademy is working on an extensive bilingual online Dutch-Frisian production dictionary (ONFW). That dictionary has a large, standardized, autonomous language, as its source language, whereas the target language is small, dependent, and far less standardized. The macrostructure of the contemporary *Algemeen Nederlands Woordenboek* (ANW) is the basis for that of ONFW, which means that the ONFW mainly incorporates neologisms identified by ANW. The Fryske Akademy also has at its disposal a corpus of bilingual news items (Dutch and Frisian). This is an interesting source, because the news editors constantly have to think of Frisian equivalents for neologisms from mostly Dutch-language news.

In this paper we discuss the possibilities there are for forming Frisian neologisms, as well as the ideological responsibility of the lexicographer to form neologisms that have the greatest potential to be accepted by the language user, as only widely accepted neologisms contribute to the vitality of Frisian.

Keywords: Frisian, Dutch, lesser used language, dominant language, language ideology, purification, standardization, bilingual dictionary

References

van der Kuip, F. and Visser, W. 2018. Introduction. *International Journal of Lexicography*, 31.2.1. 127–131. https://doi.org/10.1093/ijl/ecy005

Sijens, H. 2004. Neologismen yn it Frysk, 'Wat wy net hawwe, dat liene wy'. *It Beaken*, 66.3-4, 256-298.

Anglicisms and language-internal neologisms: Dealing with new words and expressions in *The Danish Dictionary*

Lars Trap-Jensen

Lars Trap-Jensen has a background in general linguistics, Greenlandic, and social studies. Since 1994 he has been working as a practical lexicographer at the Society for Danish Language and Literature, Copenhagen, since 2003 as the managing editor of *The Danish Dictionary* and the dictionary site ordnet.dk. He is a former president of Euralex and currently serves as its representative on the Globalex management committee.
ltj@dsl.dk

The corpus-based online *The Danish Dictionary* contains just over 100,000 entries. The dictionary is updated on a regular basis, with batches published two or three times a year. Whenever a new batch is released, it almost certainly becomes the object of public attention. The media love new words and usually assume that a new word in the dictionary is also a new word in the language – a neologism. Of course, popular belief is far from the truth: many newly published words have been in the language for a long time, but were perhaps too infrequent to be included previously.

Given their popularity, neologisms are obviously interesting for the dictionary staff, and in this paper I analyse the ones that have been included recently, and consider whether special selection criteria should apply. The editors do not use a specific method to detect neologisms in particular, but we have, on the one hand, various tools to assist us in finding lemma candidates in general, and on the other, we can analyse the batches that have already been published in recent years. I pursue both these approaches, addressing questions such as the following:

- What broad types of neologisms exist and what are their characteristics?
- How does the pressure from English affect the vocabulary of the dictionary?
- Are Anglicisms dominant or used increasingly over time as compared with language-internal neologisms? Does globalisation promote the import of words from other languages, too?
- Do dictionary users suggest and look up neologisms, and in particular Anglicisms, more often than other words?

Although the notion of 'neologism' pertains to a range of linguistic phenomena,in this context I confine myself to words and multiword expressions as (potential) entries.

Keywords: corpus-based lexicography, lemma selection criteria, Anglicisms, dictionary use, neologisms

Exploring criteria for the inclusion of trademarks in general language dictionaries of Modern Greek

Anna Vacalopoulou

Anna Vacalopoulou holds an MA in Lexicography from Exeter University. She has designed and (co-)written more than 25 dictionaries and other reference works, including paper, electronic, monolingual, bilingual, multilingual, and multimodal (sign language) works, and also contributed in the design and implementation of the first online language corpus of Modern Greek. She is currently a scientific associate at ILSP, Athens.
avacalop@ilsp.gr

This paper explores the inclusion of genericized trademarks in Greek dictionaries. Genericized trademarks constitute a special type of neologism, balancing between the non-lexical and the lexical, ‘proper’ and ‘common’. Although the goal of creating a brand name is to make a specific product easily distinguishable from the rest of its kind, the trademark might become so well-known and widely used that it starts denoting all similar products, becomes part of the general vocabulary and gains lemma status in dictionaries. Given the fact that very little, if any, documentation exists on the subject, be it publicized lexicographic policies or style guides, dictionary notes, or any other reference in the relevant literature, the main aim of the article is to explore some of the criteria by which such proprietary eponyms make their way into dictionaries of Modern Greek. First, a historical account of genericized brand names in dictionaries is given, demonstrating how this type of neologism has been gaining ground in recent years. Then, a list of genericized trademarks found in current dictionaries is compared to similar lemmas in contemporary English dictionaries to investigate which of them also constitute imported neologisms. In this respect, the paper investigates how many genericized trademarks are borrowed by other languages compared to Greek, which languages these are, and which fields constitute neologism pools for eponyms in Greek. Finally, the list of the proprietary eponyms that are included in dictionaries of Modern Greek is crosschecked against the Hellenic National Corpus to compare the frequency of lexical use to that of their non-lexical use. Traditionally, the main criteria used to differentiate the two forms of use include the existence of capitalization, the inclusion of the article, and the formation of words belonging to different parts of speech. The paper attempts to test whether these measures can help to determine the source and status of such neologisms in Modern Greek or whether other/more criteria are necessary.

Keywords: Modern Greek lexicography, genericized trademarks, lemma selection, neologisms

Neologisms in a Dutch online portal

Vivien Waszink

Vivien Waszink is a researcher at the Instituut voor de Nederlandse Taal (Dutch Language Institute) in the Netherlands. She works as a lexicographer for the *Algemeen Nederlands Woordenboek* (an online dictionary of present-day Dutch) and *Neologismenwoordenboek* (dictionary of neologisms), and is the author of books about youth language, hip-hop language and new words.
vivien.waszink@ivdnt.org

Every year, thousands of neologisms, or new words, are coined. Most neologisms are compounds or derivations. Already existing words used in a new meaning (for example, Dutch *slim* ‘smart’, often used attributively before a machine or device), new multiword units (*urban gym*) and new loanwords (*frosecco*, *thighbrow*, et cetera) are treated as neologisms as well.

Not every neologism is widely used and the majority of new words will disappear. The more widely adopted or firmly rooted neologisms are often described in dictionaries, for example in the *Algemeen Nederlands Woordenboek* (ANW), an online dictionary of present-day Dutch. Why are some new words adopted, while others are ignored? Is it necessary to register and describe neologisms that are likely to disappear, for example in a dictionary of neologisms? And what should such a dictionary of neologisms look like?

In this paper I present a pilot version of a new dictionary of Dutch neologisms. Firstly, I will explain how Dutch neologisms are created. Secondly, I demonstrate why it is necessary to register and describe neologisms (also those that are not adopted in present-day Dutch) in an online dictionary portal. Then I show how potential neologisms in Dutch can be detected with the aid of the computer tool Neoloog and through corpus analysis. Finally, I will go into the lemma structure of this special-domain dictionary of neologisms and discuss how it differs from the *ANW* in the way it describes neologisms.

Keywords: neologisms, new words, dictionary, online dictionaries, lemma structure, Dutch