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