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)