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 Unpublished presentation, Congreso Internacional El Diccionario: neología, lenguaje de especialidad, computación, Ciudad de México, 28-30 October 2013.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 Unpublished presentation, VII Congreso Internacional de Lexicografía Hispánica, Valencia, 28 June 2018.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.

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].

---

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)

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

---

6   https://ekilex.eki.ee (accessed March 30, 2019)