Lexicala API:
A new era in dictionary data

Ilan Kernerman and Dorielle Lonke

The Lexicala API is a REST API providing access to cross-lingual lexical data of K Dictionaries (KD) across 50 languages, including monolingual linguistic resources, over 150 language pairs, and numerous multilingual combinations. It enables flexible search options and returns **lexicalaAPI** ▶ JSON responses, as well as JSON-LD encoding RDF representation of KD data for Linked Data purposes and full integration with Semantic Web technologies. The formal release is on July 11, 2019.

KDICTIONARIES

▶

The Lexicala API Team

Maayan Orner, Manager

Vova Dzhuranyuk, Developer

Roi Sadika, Developer

Dorielle Lonke, Coordinator

1. Resources

The Lexicala API offers data from three different KD resources: *Global*, *Password*, and *Random House*.

The Global series. A network of multi-layered and inter-linked lexicographic datasets for 25 European and Asian languages. Each language has at the basic layer its own monolingual core, featuring detailed and varied semantic and syntactic information, including alternative spellings and scripts; phonetic transcription; grammatical categorization, gender and number; sense disambiguation and attributes such as synonyms, antonyms, subject domain, register, etc.; examples of usage and different types of multiword units. Most of these cores (22 languages) have translation equivalents for each sense, example and expression to at least one other language (e.g., Korean to Japanese) and up to 18 languages (in the case of French). When several bilingual versions are available, they are juxtaposed and form a multilingual set, i.e. each item includes translations in several languages.

The Password series. A semi-bilingual English learner's dictionary with translations in 45 languages. The English entries include a definition and example(s) of usage for each sense, as well as a brief translation equivalent of that sense of the headword, including for each multiword unit and sub-entry. Most of the language versions are complemented by a bilingual index to the specific sense(s) of the English entry. The index is then expanded into a multilingual glossary by automatically adding the other language translations of *Password*, thus generating translations indirectly from the core language to any of the other languages via the English intermediary.

Random House Webster's College Dictionary (RHWCD). A comprehensive monolingual dictionary of American English. The last edition of this legacy dictionary was published by Random House in 2005, and in 2009 KD acquired full rights for its use. It includes over 133,000 entries with 191,000 senses, and offers a deep and extensive description of contemporary English language, including etymological, geographical and biographical information. KD has continued to reformat the data and update the contents in making RHWCD its flagship English dictionary, serving both native speakers of English and upper-level non-native users.

These resources combine various forms of human-curated and machine-generated content, and are supplemented by rich morphological lists of inflected word forms.

2. Infrastructure and Functionalities

The Lexicala API uses ElasticSearch as a back end, utilizing language specific functionalities for a flexible, fast search, that deals with natural language challenges for multiple languages. It is hosted on Amazon Web Services (AWS), prioritizing service reliability and scalability.

A basic API search is performed by looking up a headword, which returns all corresponding entries. This type of query returns a JSON document containing partial lexical information on entries that match the search criteria, including their unique entry ID. It is also possible to search for inflected forms, as well as by grammatical gender, number, part of speech and subcategorization, to obtain more specific results.

The inflected forms are provided either by morphological word form lists or by an automated *stemmer* functionality, allowing for greater flexibility when searching for a specific lemma. The purpose of the stemmer is to create a stem form from the analyzed word. The stem does not have to be a valid word, for example, a stemming algorithm can reduce *fishing*, *fished* or *fishes* to the stem *fish*, or the words *argue*, *argued* or *arguing* to the stem *argu*.

It is also possible to query the entire collection of dictionary entries (or senses) by a unique entry (or sense) ID. This type of response consists of a full dictionary entry, including translations, syntactic and semantic information, compositional phrases and usage examples. This type of result contains elaborate information about each headword, and offers all available translations at once, emphasizing the uniqueness of KD resources in multi- and cross-linguality and linking between languages and datasets.

3. RDF and Linked Data

Besides relying on unique multilingual resources, a prominent feature setting the Lexicala API apart from other dictionary APIs is the option of obtaining JSON-LD formatted RDF representation of lexical data, designed for Linked Data (LD) interoperability.

The RDF data is modelled according to the state-of-the-art *Lexicog* module of the OntoLex-*lemon* model – which was designed with linguistic and lexicographic data in mind and constitutes the *de facto* standard of lexicographic data representation in RDF – and is serialized in JSON-LD, a popular format in the Semantic Web (cf. W3C Ontology-Lexica Community Group at https://www.w3.org/community/ontolex/ and *Lexicog* module specification at http://www.w3.org/ns/lemon/lexicog#).

LD methods are at the forefront of the current generation of powerful language

technology solutions, and are at the heart of human-machine interaction. Providing quality cross-lingual lexical data, with the LD-driven option of linking to other sources, substantially widens the offering of data resources to Lexicala API users and for various integration, research and development purposes. The possibility of linking KD data to other enriched or annotated resources can be of great value for NLP and machine-learning tasks, which places the Lexicala API in leverage to other Dictionary APIs, providing added value in computational aspects to traditional lexicography and language related content.

4. Market and Users

The Lexicala API targets a broad range of users, starting with individual developers of a wide variety of applications who are looking for quality lexical data with rich multilingual extensions, through NLP researchers and computer scientists in need of large lexical corpora for processing, parsing or training a machine, and on to all types and sizes of online and offline translation, localization, learning and other language services.

The possibilities for using KD data are diverse, as the varying focal points of each resource and the wide selection of languages and information offer solutions to many different issues, and the flexible search and accessibility allow easy processing and seamless integration with other applications.

The new Lexicala API has an important role in two ongoing projects funded by the European Union's Horizon 2020 research and innovation programme, in which KD is participating: *Lynx* – Building the Legal Knowledge Graph for Smart Compliance Services in Multilingual Europe (No. 780602, http://lynx-project.eu/); and, *Elexis* – European Lexicographic Infrastructure (No. 731015, https://elex.is/).

Sample API extracts:
1. excerpt from *Abbau* in Global German
2. *chair* (noun) in the Password multilingual set
3. *smile* in RHWCD
4. results for *azul* in Global Spanish, when searching by parameters

1.
```
        },
        "tr": {
            "text": "atlıkarıncanın sökülmesi"
        }
      }
    }
  ]
},
{
    "id": "DE_SE00000042",
    "semantic_subcategory": "von Bodenschätzen",
    "translations": {
      "ar": [
        {
            "text": "إِسْتِخْرَاج",
            "pronunciation": {
              "value": "ʔistixˈraːdʒun"
            },
            "gender": "masculine",
            "number": "singular"
        },
        {
            "text": "إِسْتِغْلالْ",
            "pronunciation": {
              "value": "ʔistiɣˈlaːlun"
```

2.
```
},
"ko": {
  "text":
},
"lt": {
  "text":
},
"lv": {
  "text":
},
"ml": {
  "text":
},
"nl": {
  "text":
},
"no": {
  "text":
},
"pl": {
  "text": "krzesło"
},
"prs": {
  "text": "صندلی"
},
"ps": {
  "text": "چوکۍ: کرسی: رئیس (دغونډی"
},
"pt": {
  "text": "cadeira"
},
"ro": {
  "text": "scaun"
},
"ru": {
  "text": "стул"
},
"sk": {
  "text": "stolička"
},
"sl": {
  "text": "stol"
},
"sr": {
  "text": "stolica"
},
"sv": {
  "text": "stol"
},
"th": {
  "text": "เก้าอี้"
```

3.
```
{
  "id": "RDE00064769_0",
  "source": "random",
  "language": "en",
  "headword": {
    "text": "smile",
    "pronunciation": {
      "value": "smaɪl"
    },
    "pos": [
      "verb",
      "noun"
    ]
  },
  "senses": [
    {
```

4.
```
{
  "n_results": 2,
  "page_number": 1,
  "results_per_page": 10,
  "n_pages": 1,
  "available_n_pages": 1,
  "results": [
    {
      "id": "ES_DE00006683",
      "language": "es",
      "headword": {
        "text": "azul",
        "pos": "noun"
      },
      "senses": [
        {
          "id": "ES_SE00009139",
          "definition": "color parecido al cielo"
        }
      ]
```

K Dictionaries & Lexicala Workshops

Arleta Adamska-Sałaciak at the K Dictionaries & Lexicala Seminar at Euralex 2018

Philippe Climent at the K Dictionaries & Lexicala Seminar at Euralex 2018

Margit Langemets at the K Dictionaries & Lexicala Seminar at Euralex 2018

Dorielle Lonke at the K Dictionaries & Lexicala Seminar at Euralex 2018

Mehmet Gürlek at the K Dictionaries & Lexicala Workshop at Asialex 2019

Shigeru Yamada at the K Dictionaries & Lexicala Workshop at Asialex 2019

Roi Sadika at the K Dictionaries & Lexicala Workshop at Asialex 2019

KD & Lexicala Seminar at Euralex 2018

The K Dictionaries & Lexicala Seminar was co-located with the 18th Euralex International Congress and held at the University of Ljubljana on July 17, 2018. In addition to an introduction of KD by Ilan Kernerman and overviews of its editorial policy by Yifat Ben Moshe and Lexicala by Dorielle Lonke, the seminar featured presentations by four guests: Arleta Adamska-Sałaciak from Adam Mickiewicz University in Poznań, on Bilingual Learner's Dictionaries from a Practical and Theoretical Perspective; Margit Langemets from the Estonian Language Institute in Tallinn, on Translating Password Dictionary and Editing its Index; Philippe Climent from IDM in Paris, on Data Development and Online Publication with DPS; and John McCrae from the National University of Ireland, Galway, on Linked Data Lexicographical Modelling in OntoLex. https://euralex2018.cjvt.si/programme/co-located-events/

KD & Lexicala Workshops at Asialex 2019

K Dictionaries held a pre-conference workshop for over 50 students of the Turkish Department at Istanbul University on June 18, and another workshop for participants of the Asialex 2019 conference on June 21. The workshops featured guest editors Mehmet Gürlek (Turkish) and Shigeru Yamada (Japanese), alongside KD's Raya Abu Achmad (Arabic), Roi Sadika (online and API), and an overview by Ilan Kernerman. http://asialex2019.istanbul.edu.tr/en

KD & Lexicala Data Workshop at eLex 2019

K Dictionaries will hold a workshop focusing on lexical data issues at the closure of the 6th biennial conference on electronic lexicography in Sintra on October 3, 2019. Topics include the automatic generation and manual creation of data; formats, standards, and the Ontolex *Lexicog* module; data dissemination, usage, and interoperability; and a live API demonstration. In addition to KD's Yifat Ben Moshe, Ilan Kernerman and Dorielle Lonke, confirmed guests include Jorge Gracia (Universidad de Zaragoza) and Julia Bosque-Gil (Universidad Politécnica de Madrid). https://elex.link/elex2019/programme/k-dictionaries-lexicala-data-workshop/

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.

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.

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.

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)

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.

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

globaLex

dic·tion·ar·y
SOCIETY of NORTH AMERICA

Lexicography in higher education institutions: European Master in Lexicography with an Erasmus Mundus joint degree

Stefan J. Schierholz

Stefan Schierholz is a full professor of German Linguistics and Lexicography at Friedrich-Alexander-Universität Erlangen-Nürnberg in Germany, and the coordinator of the Erasmus Mundus Joint Master Degree Programme 'European Master in Lexicography (EMLex)'. He is the series editor of WSK (*Wörterbücher zur Sprach- und Kommunikationswissenschaft / Dictionaries of Linguistics and Communication Science*), a series of 25 volumes in print that will include (over) 40,000 dictionary articles of which 12,000 are already available online, and the volume editor of WSK 1 (Grammar, with Pál Uzonyi). His metalexicographic works deal primarily with Dictionary Research, Methods in (Meta)Lexicography, Dictionary Criticism and Lexicography for Special Purposes.
stefan.schierholz@fau.de

The European Master in Lexicography (EMLex) is an international master degree, which was developed in Erlangen in 2007 and is supported by the Interdisciplinary Center of Valency, Lexicography and Collocations, located at Erlangen University. The history, development and progress of this master degree is documented in the international yearbook *Lexicographica*, starting from vol. 26 (Schierholz 2010) and continued year by year in the column Lexicography in Higher Education (cf. Jesenšek / Schierholz 2011, Buchi 2012, Schierholz 2013, Domínguez Vázquez / Schierholz 2014, Gouws / Iriarte 2015, Hollós / Schierholz 2017, Bielinska / Schierholz 2017, Buchi / Blanck 2019). The EMLex has been an Erasmus Mundus Joint Master Degree programme funded by the EACEA (Education, Audiovisual and Culture Executive Agency) since 2015 and through 2024.

1. Basic idea and general objectives
The EMLex programme aims to:

- promote the international and interdisciplinary training of lexicographers
- provide an attractive international course of studies at different universities in Europe and elsewhere
- bring together students from all over the world
- teach lexicographic theory at an international level
- encourage active research on behalf of the students
- emphasize applicability in the dictionary making practice

Graduates will be able to work in a publishing house, technology firm or academic institute in all fields of lexicography, from the conception of printed and electronic dictionaries to specific tasks and technical realization of lexicographic products. Besides, they are thoroughly trained in the analytic-theoretical aspects of international lexicography in order to find a field of activity in scientific research or to become advising-conceptional lexicographers in the fields of publishing or further education.

In addition to their subject-specific skills, graduates acquire practical knowledge through an internship that enables them to take on responsibility for demanding tasks in various areas of industry, culture, media, school

and adult education after having graduated. Graduates will be able to teach the theory and practice of international lexicography and dictionary research, and complete practical tasks as professional lexicographers by using advanced academic methods.

The four-semester programme commences each winter term, and includes the second semester abroad. The overall number of ECTS points of the programme is 120. The total number of students, who can be registered for EMLex at the participating universities is limited to 30 per year, due to the particular design of the summer-term block seminars in which the international component of the course manifests itself.

The contribution of independent student work can be estimated to be very high, since a high level of motivation is expected. The modules are either e-learning-modules, specific lexicography modules, or import modules of other courses of study. The post-graduates become international specialists all over Europe since EMLex is unique in Germany, in Europe and worldwide. On the basis of the contract of cooperation and the accreditation process undergone by all participating universities, it is possible to provide the students with a high-ranking certificate. All students get a Joint Master Degree signed by the presidents or rectors of all partner universities.

2. Conditions of admission
The entry into the master programme is subject to special conditions. A first university degree with an outstanding grade and a special entrance examination are necessary (cf. examination regulations). Basic knowledge of English and German are essential for participation as the teaching is done exclusively in these two languages. Besides other languages (French, Hungarian, Italian, Polish, Portuguese, Spanish), these two can be deepened in the first semester if necessary.

3. Participating universities
The following universities and institutes participate in the EMLex programme (responsible teachers are listed):

- Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany: Prof. Dr. Stefan Schierholz, Prof. Dr. Thomas Herbst, Prof. Dr. Stefan Evert, Dr. Michael Klotz

- Université de Lorraine (Nancy), France: Prof. Dr. Éva Buchi, Prof. Dr. Alain Polguère, Prof. Dr. Yvon Keromnes
- Universidade de Santiago de Compostela, Spain: Prof. Dr. Maria José Domínguez Vázquez, Prof. Dr. Carlos Valcárcel Riveiro, Prof. Dr. Ernesto González Seoane
- Universidade do Minho (Braga), Portugal: Prof. Dr. Álvaro Iriarte Sanromán, Prof. Dr. Idalete Maria da Silva Dias
- Uniwersytet Śląski w Katowicach, Poland: Dr. habil. Monika Bielińska, Dr. Ewa Myrczek-Kadłubicka
- Università degli Studi Roma Tre, Italy: Prof. Dr. Martina Nied Curcio, Prof. Dr. Stefania Nuccorini, Prof. Dr. Paolo Vaciago
- Károli Gáspár Református Egyetem (Budapest), Hungary: Dr. habil. Zita Dötschné Hollós, Dr. habil. Dóra Pödör, Dr. Petra Szatmári
- Universität Hildesheim, Germany: Prof. Dr. Ulrich Heid, Prof. Dr. Friedrich Lenz, Prof. Dr. Laura Giacomini
- University of Stellenbosch, South Africa: Prof. Dr. Rufus H. Gouws, Prof. Dr. Ilse Feinauer.

Associated partners in the EMLex, who also provide staff as guest lecturers, include:

- Leibniz-Institut für Deutsche Sprache (German Language Institute): Prof. Dr. Stefan Engelberg, Prof. Dr. Carolin Müller-Spitzer, Dr. Annette Klosa, Dr. Dominik Brückner, Dr. Sascha Wolfer
- Ivané Javakhishvili Tbilisi State University, Georgia: Prof. Dr. Tinatin Margalitadze
- Universidade Federal do Rio Grande do Sul, Brazil: Prof. Dr. Maria José Bocorny Finatto
- University of Texas (Austin)
- Lomonossow University Moscow
- Université de Montréal

4. Curriculum

The curriculum is set up by compiling the modules and laying down the concrete aims, contents and purposes of the various seminars. A programme and additional details have been agreed upon. The following list contains the essential topics:

- Lexicographic entries, typology of dictionaries
- Dictionary research, metalexicography
- History of lexicography
- Research into dictionary usage
- Users' needs and preconditions of usage
- Critical research into dictionaries

- Functions of dictionaries
- Work in the lexicographic workshop
- Corpus lexicography, lexicographic tools
- Bilingual lexicography
- Lexicography for special purposes
- Knowledge of terminology and terminography
- Computational lexicography
- Electronic dictionaries, encyclopaedic dictionaries
- Corpus work
- Lexicology
- Etymology

Soft skills contain language modules and a computer science module. Aims, contents and special regulations for the courses are listed in the schedule in detail below and can be found on the EMLex website.

5. Schedule

The framework of the programme is shown in Figure 1.

- Notes on the 2nd semester (block seminars)

In the second semester, one university serves as the common study place for all the students. In consequence, the number of participants is limited; their travel is financed by the Erasmus Mundus programme, Erasmus+ or other programmes. The teachers of foreign universities stay for 4 to 7 days; their travel is financed by the EMLex consortium and the Erasmus Mundus programme. One block seminar lasts from Monday to Friday. Information about the exams is given each year by the responsible teachers who develop an organisation plan for the students.

- Notes on the 3rd semester

The in-depth module 1 is done via the EMLex Moodle platform, which is hosted by the University of Minho (Braga).

- Notes on the 4th semester

The ECTS points granted for the master thesis according to the individual university regulations allow for further opportunities (taking accompanying seminars, etc).

From 2010 to 2016, the EMLex programme and the journeys were financed by Erasmus+. The study programme started in winter term 2010 with five students from Erlangen and Nancy, and in the following years the number of students has gradually increased. In the first Erasmus Mundus cohort, however, we had 15 students, then 24 the following year, and 21 students started the EMLex in 2018. The summer terms were held in Maribor (2011), Barcelona (2012),

The main homepage of EMJMD-EMLex provides links to the participating partners and displays all the activities:
https://www.emlex.phil.fau.eu/

	ECTS	Aims and contents
Winter term 1st semester		During the first semester the students of the various universities should reach a comparable level of knowledge to enable international cooperation and joint studying as soon as possible.
	10	<u>Basic module B1</u> (e-learning, approx. 4 semester hours): Basics of lexicography: dictionary usage, lexicographical entries, typology of dictionaries, lexicography and culture, theory and practice, computational lexicography.
	10	<u>Basic module B2</u> (home module): A lexicographically-oriented module (approx. 4 semester hours) with university teachers participating in the master programme.
	5 5 5 5	<u>Soft skills module</u> (home module): <u>B3a course</u>: English (4-6 semester hours), if needed, see [1] <u>B3b course</u>: German (4-6 semester hours), if needed, see [1] <u>B3c optional module</u>: language course (Portuguese, Hungarian, Polish, Spanish, French, Italian) B3d <u>optional module</u>: computer science [1] From section B3a to B3d: 5 ECTS are obligatory, 15 ECTS are possible, 10 ECTS can be credited.
Sum: 25-30	**During the first semester all students should acquire 25-30 ECTS points.** Those who acquired only 25 ECTS points (i.e. students with a BA in German and English) have to acquire 35 ECTS in the second semester.

Erlangen (2013), Santiago de Compostela (2014), Braga (2015), Budapest (2016), Katowice (2017), Nancy (2018), and Rome (2019).

6. What is Erasmus Mundus?

The goals of the Erasmus Mundus programme are to improve the competitiveness of European universities, to become global players in higher education, to intensify the inner connection of European countries and people, to bring together students from different countries and to foster excellence and jointness in the educational sector. The objectives of the EACEA in Brussels are to support students with scholarships, to facilitate the organisation of joint programmes and to foster excellence in Europe. The demands of EACEA are to develop a stable Master programme, to guarantee joint teaching, to foster multilingualism, to ensure coherence of the consortium members and the permanent networking in the consortium (including with other Erasmus Mundus consortia), to provide additional value by the study programme, to execute perfect management in all areas, to enable students' participation, to develop worldwide promotion of the programme, to undertake quality assurance and monitoring, to assure sustainability, to establish networks with the private sector, and to award the students a Joint Degree, e.g. one diploma for all graduates that is equally accepted by universities and employers all over Europe.

The EMLex consortium fullfilled these criteria during the first period of EMJMD funding (2015-2021) excellently. The application for a second period of EMJMD funding (2018-2024) was thus evaluated as the best among 121 applications in 2018. This presents a great challenge for the consortium to continue its work every day with the same enthusiasm and activity as we did in the past. The coordinator is only a coordinator, but the work is done by everyone in our group and was supported by the secretaries in Erlangen.

References

Bielińska, M. / Schierholz, S.J. 2017. Der Europäische Master für Lexikographie 2017 im Erasmus Mundus Joint Master Degree Programm. *Lexicographica* 33: 467-471.

Buchi, É. 2012. First European Master in Lexicography (EMLex) diplomas awarded. *Lexicographica* 28: 441-442.

Buchi, É. / Blanck, W. 2019. "Lexicographers of all countries, unite!". About the common semester of the European Master in Lexicography (EMLex) in Nancy. *Lexicographica* 34: 367-372.

Domínguez Vázquez, M.J. / Schierholz, S.J. 2014. Der Europäische Master für Lexikographie in 2014. *Lexicographica* 30: 613-616.

Summer term 2nd semester		The second semester provides an overview of the whole range of lexicography. Each university offers special knowledge. The local organisers change every year.
	5 each	Advanced modules 1 to 10: 10 modules as block seminars: All students have to travel. The teachers of the partner universities hold the block seminars, the local teacher holds his/her seminar throughout the semester. The students have to acquire 30 (35) ECTS, but may, however, acquire more; the organisation plan is set up according to special announcement. A1: Metalexicography A2: History of lexicography A3: Data modeling and representing data in digital lexicography A4: Research into dictionary use A5: Special field lexicography and terminology A6: Computational lexicography A7: Dictionaries and translation A8: Lexicography and lexicology A9: Planning and making a dictionary A10: Learner lexicography
	Sum: 30	**All students have to have 60 ECTS points after the first year.**

Winter term 3rd semester		The third semester is dedicated to specialisation and application of the knowledge acquired in the second semester; at the same time, the master thesis is prepared. The contents depend on the respective universities.
	10 10 10 Sum: 30	In-depth module 1: Academic writing in lexicography: (a) Series of three lectures by teachers of different universities via video (online: e-learning). Oral exam with or written report by the students; (b) Writing a review on a dictionary. In-depth module 2 (home module): Special topic (depending on the seminars offered at the home university). Practical module (home module): A four-to-twelve-week internship at a publisher or a company or a research institute (during the semester break or during the semester).
Summer term 4th semester	0-5 25-30 Sum: 30	Accompanying seminar to the Master thesis (home module). Master thesis
	Sum: 120	Total number of points

Figure 1: Study programme of the EMLex

Gouws, R.H. / Iriarte Sanromán, Á. 2015. The European Master in Lexicography in 2015: 2nd EMLex Colloquium on Lexicography and the international Summer Semester. *Lexicographica* 31: 358-362.

Hollós, Z. / Schierholz, S.J. 2017. Der Europäische Master für Lexikographie im Erasmus Mundus Joint Master Degree Programm. *Lexicographica* 32: 219-223.

Jesenšek, V. / Schierholz, S.J. 2011. Der Europäische Master für Lexikographie. *Lexicographica* 27: 331-333.

Schierholz, S.J. 2010. EMLex: Europäischer Master für Lexikographie – European Master in Lexicography. *Lexicographica* 26: 343-350.

Schierholz, S.J. 2013. Der Europäische Master für Lexikographie in 2013. *Lexicographica* 29: 339–341.

Dictionaries for the future – the future of dictionaries: Challenges for lexicography in digital society

From November 26 to 28, 2018, 18 experts from the fields of Dictionary Research, Practical Lexicography, German as a Foreign Language, Italian Studies, Translation Science and Empirical Linguistics congregated at the Centro Italo-Tedesco per l'Eccellenza Europea / Deutsch-Italienisches Zentrum für Europäische Exzellenz, *Villa Vigoni* on Lake Como, for a workshop on the topic of future challenges for lexicography. Supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG), the workshop was initiated and organised by Stefan Schierholz, Martina Nied Curcio, Laura Balbiani and Anne-Kathrin Gärtig-Bressan. Held within this framework of German-Italian cooperation in humanities, the discussion dealt with aspects of the digital society of the future. The dramatic transformation from print to online dictionaries during the last two decades has changed the academic and commercial worlds of dictionaries, lexicography and relevant research. Users' preferences for costless online data, waiving proof of the data reliability, has drawn many publishing houses into a deep crisis, which is being slowly overcome these days. In Germany, this change is felt in the last couple of years as a few publishers survived and are successful again on the market today. In addition, some academic dictionary projects have been reinforced with the support of public funding. No fundamental progress is observable at the theoretical basis and as with regard to dictionary research at universities since, (a) the issue is still not focused and (b) the near future perspectives look grim. Research on online information systems, digital dictionaries, user's behaviour, and dictionary critics and criticism are no top priorities at universities in Germany. In this context, the main aims of the workshop were to develop ideas and concepts for the future, to initiate a critical use of the data offered on the internet, and to set new impulses for the state of affairs in dictionary research. At the closure of the workshop the participants adopted *The 15 Villa Vigoni Theses on Lexicography*[1], which are listed below.

Stefan Schierholz

1. **Dictionaries of the future** consist of lexical or linguistic information systems in which existing lexicographic data are conflated, entrenching multilingualism and linguistic variety, and which provide people who are confronted with gaps in their knowledge with answers as well as support in the writing and formulation processes of texts.
2. Lexical information systems must become a significant topic of **public discourse**. **Awareness** of the fact that the respective online data available should provide the **requisite high quality** must be publicly nurtured.
3. Practical lexicography must be constantly aware of its **social responsibility** and must strive for a comprehensive, pluralistic **description of linguistic and factual realities**. In the process, the demarcation between the subject area and the selective lexicographic prioritization must be rendered perceptible.
4. As independent social institutions, universities and public research facilities must actively participate in **critical discussions and evaluations** regarding lexical information systems.
5. Lexicographic **amendments** in online information systems must be **chronicled and preserved** so that they remain permanently available as well-documented evidence of academic processes.
6. Lexicography requires **partners and allies**: the solutions and challenges for the lexicography of the future

The workshop participants included the following persons:
Dr. Andrea Abel (Eurac Research, Bozen), Prof. Dr. Laura Balbiani (Università della Valle d'Aosta), Wiebke Blanck (Universität Erlangen-Nürnberg), Prof. Dr. Gualtiero Boaglio (Institut für Romanistik-Universität Wien), Prof. Dr. Stefan Engelberg (Institut für Deutsche Sprache, Mannheim), Dr. Anne-Kathrin Gärtig-Bressan (Università degli Studi di Trieste), Prof. Dr. Luisa Giacoma (Università della Valle d'Aosta), Dr. Laura Giacomini (Universität Heidelberg), Prof. Dr. Christine Konecny (Universität Innsbruck), Dr. Kathrin Kunkel-Razum (Bibliographisches Institut GmbH, Berlin), Dr. Fabio Mollica (Università degli Studi di Milano), Dr. Carolin Müller-Spitzer (Institut für Deutsche Sprache, Mannheim), Prof. Dr. Martina Nied Curcio (Università degli Studi Roma Tre), Prof. Lorenza Rega (Università degli Studi di Trieste), Prof. Dr. Elmar Schafroth (Universität Düsseldorf), Dr. Rüdiger Scherpe (Klett Verlag – ELI), Prof. Dr. Stefan Schierholz (Universität Erlangen-Nürnberg), Dr. Francesco Urzi (European Parliament – Université de Genève).

---

1 Cf. https://www.emlex.phil.fau.de/2019/01/

demand, with a view to European perspectives, an **interdisciplinary exchange** between research institutes, academies, publishing houses and other representatives of the private sector.

7. One significant **task for the lexicography** of the digital future is the orderly conflation of data which has been generated automatically by text corpora and specifically processed, as well as a user-orientated presentation. The social relevance of such information systems will be consolidated once the underlying corpora mirror the entire linguistic diasystem and are freely available to researchers.

8. Dictionary research must be considered a **cultural science** which, through interdisciplinary projects, conflates practical lexicography, linguistics, computer science, book science and documentation science.

9. In a modern information society, we require academic studies to advance a **standardization process** for metalexicographic **core terminology**, as a solid theory induces multifaceted improvements in practical lexicography.

10. **Academic lexicography** should be increasingly visually creative, and with regard to digital formats it should venture into **experiments**, thereby availing itself of the interest of people in linguistic questions. **State funding** must concentrate on lexicographic innovations.

11. Lexicographic projects should be oriented towards the **specific needs of the users** (towards the first language and the foreign language, towards translating, etc.) as well as towards the users' linguistic acts or communicative intentions, as language is the subject matter of lexicography, and learning and understanding languages is a central competence in a globalised world.

12. Academic findings regarding the use of lexicographic information systems as well as **teaching practice and translation practice** should be increasingly incorporated into the **lexicographic process**.

13. Lexicography is called upon to develop concepts for productive **user participation** in lexicographic information systems.

14. The digital supply of data in the information systems of the future must be regarded as a significant means for 'lifelong learning' so that the **critical use of resources** can be established as a **strategic key competence**. This must also be firmly entrenched in the **training and continued education of teachers**.

15. Lexicography requires **pedagogical concepts** in order to be able to accomplish the **didactic implementation** of lexicographic information systems. In the process, this should integrate the media competence of the users.

Jacek Fisiak (1936-2019)

Professor Jacek Fisiak of Adam Mickiewicz University in Poznań, Poland, passed away on June 3, 2019. He was a renowned linguist, an indefatigable organizer of academic life, and a capable dictionary editor.

Jacek Fisiak first came to Poznań in the late 1960s, with the ambitious plan of reviving English Studies there. He not only succeeded spectacularly, but soon came to be regarded as the chief figure among Polish Anglicists; it was not long before he received similar recognition from historical linguists elsewhere in the world. He was the go-to person for anyone in need of assistance with elusive sources pertaining to the history of Germanic languages, his private library being a frequent port of call for scholars from all over Poland.

He did all the usual things that academics do, but tended to do them exceptionally fast, efficiently, and with deep passion. He organized dozens of international conferences on historical and contrastive linguistics, gave over 60 papers at international congresses and over 150 guest lectures, supervised 61 doctoral dissertations and countless MA theses. He served as Head of the Department of English (1969-2005), Rector of Adam Mickiewicz University (1985-1988), and Minister of National Education (1988-1989). In recognition of his achievements, he received honours and awards too numerous to mention, including two honorary doctorates (from the Universities of Jyväskylä in Finland and Opole in Poland), as well as, unusually for a non-British citizen, an Order of the British Empire.

While his primary interest had always been historical linguistics, especially medieval English dialectology, in the post-1989 period he also tried his hand at dictionary making. He was instrumental in breathing new life into bilingual English-Polish lexicography, acting as editor-in-chief in a number of projects with HarperCollins, Longman, Kernerman Publishing, and others. Among those, *The New Kościuszko Foundation Dictionary* (2003; 2nd, digital edition 2015) was, and still is, the largest English-Polish, Polish-English dictionary in terms of coverage, as well as the only one featuring the American variety of English, while *Longman Słownik Współczesny* (2004, 2nd edition 2011) was the first – and so far the only – truly active bilingual dictionary designed specifically for Polish learners of English.

As a person, Professor Fisiak was generous, outgoing, and likeable. He will be remembered not only for his wide-ranging knowledge and professional achievements, but also for his irresistible charisma and his lust for life.

Arleta Adamska-Sałaciak
Faculty of English, Adam Mickiewicz University, Poznań

Deny Arnos Kwary
1 January 1975 – 4 June 2019

A speech made at the opening of the Asialex 2019 conference in Istanbul on June 19, 2019, by Dora Amalia, of Badan Bahasa, Jakarta.

Good morning participants, colleaques, and distinguished guests.

I am standing here today with a grieving heart, for I am going to give a eulogy for a man who was very dear to many of us: Dr. Deny Arnos Kwary. Barely a fortnight ago, at the end of Ramadan, the holy month for Moslems, one day before Eid al-Fitr, we were shocked by the sad news that Deny Arnos Kwary has just passed away due to a sudden heart attack. It was with great sadness that we received this news of Deny's passing. He was, but now is no more, with us. *Inna lillahi wa inna ilaihi rojiun (Surely we belong to Allah and to Him we shall return).*

I have known Deny personally since 2009, when he was still a PhD student at Aarhus University in Denmark. We originally met through Facebook thanks to our mutual interest in lexicography, which is quite uncommon, when he took the initiative to greet me virtually. It took only a short while before we became friends, partly because of his friendly and gentle approach.

Since then, Deny has been an unusually good and helpful friend of mine. He was bright and humble, a dedicated young scholar in his prime, and has personally helped me on inumerable occasions.

Once I was struggling to finish my studies in lexicography at the University of Indonesia. I called Deny for help, and he was willing to spend his precious time to help me in the matter I found difficulties in, enabling me to graduate successfully. Knowing his capacity, I asked him to be my co-promotor right after he obtained his PhD. At first he declined, saying it would be too challenging for him, but after I persisted, being certain that he was the right person for the position, he finally accepted my request. I will always remember how Deny has helped and guided me by generously sharing his time, expertise and experience in lexicography. He also encouraged me to write journal papers and to join many regional and international workshops and seminars on lexicography, including this AsiaLex conference.

These are but a few recollections that I can share in this limited time, where he helped me personally as a friend. His sudden death has left me with a grief which is too difficult to describe with words, although I am a lexicographer with experience in dictionary writing – his contributions were simply too great.

Having shared my personal experience with Deny, I am also sure that many of you knew him personally through his work in lexicography, both internationally and locally. He was the Treasurer and an Executive Board member of AsiaLex, as well as the Review Editor of *Lexicography – Journal of AsiaLex*. Besides, he was an associate professor and head of the English Department in the Faculty of Humanities at Airlangga University in Surabaya.

Deny finished his undergraduate studies at Airlangga University in 1996. In 2002 he completed his masters program at University of Indonesia, and in 2010 he earned a doctor of philosophy (PhD) from Aarhus University. It was during his doctoral study that he began to have an interest in the "dictionary" as he often found his thesis advisor reading dictionaries. Deny finally decided to dive deep into lexicography.

He was a very productive researcher and had conducted many studies since 2006, with at least 39 of his studies being published in national and international journals. He was also a frequent speaker in international conferences, an active member in a number of associations such as

Directory of Linguists, TEFLIN (Teachers of English as a Foreign Language in Indonesia), Global Electronic Member, TESOL (Teachers of English to Speakers of Other Languages), LSA (Linguistic Society of America), and Global CALL (Computer Assisted Language Learning). During his career Deny had received two prestigious awards: as an outstanding lecturer in 2013 and as a brand ambassador of Casio Electronic Dictionaries in 2014.

There are only a handful of persons like Deny in Indonesia, especially with such remarkable expertise. Lexicography began to develop locally only recently, and this was inseparable from Deny's major role. Badan Bahasa began to hold the Indonesian Lexicography Seminar annually since 2016, and in September this year we will be holding the fourth seminar. At some point in our journey, more people became interested in lexicography and in 2017 we established Perkamusi (Association for Indonesian Lexicographers) with Deny as Chairman. This is the first official association in Indonesia specialized in lexicography. All these local advances in lexicography are, to a great extent, driven by none other than Deny, as without him it would have been impossible for us to reach this state. Surely, Deny's passing is a great loss for the lexicography community, not only in Indonesia but also worldwide.

In the midst of all his activities, Deny had also conducted community services, such as 'The right corpus design for Indonesia', a self-funded service, in 2015. Based on his studies, Badan Bahasa began building Korpus Indonesia (Indonesian Corpus) in 2016, with Deny as the developer as well as consultant. Korpus Indonesia was released last year, and this was the last project we shared with him.

There is a saying by Goethe that "a useless life is an early death." This is not true of Deny's early death, on the contrary, I would say: For a useful life such as Deny's, his works will outlast his death, and will be remembered by the generations after him.

To conclude, there are numerous things Deny has done for the lexicography community in Indonesia and for Badan Bahasa. May his works and dedication to advance people's knowledge earn him the reward promised by Allah. May Allah record all his goodness and give him the most beautiful place in heaven. Let us

also pray that Deny's family will be given strength and fortitude to face this sadness and continue their lives.

There is no word to describe how deep is the loss we experience over Deny's passing. But I am confident that what Deny has done will always be remembered as precious legacy to our nation, and will always live in our hearts as long as we draw breath. The Greek quote by Menander says, ὃν οἱ θεοὶ φιλοῦσιν ἀποθνήσκει νέος (*hon hoi theoi pilusin apothniskei neos*), meaning 'He whom the gods love, dies young'. God loves you, Deny. Goodbye. Your toil is over now. May you rest in peace.

Dora Amalia
Badan Pengembangan Bahasa dan Perbukuan

In-Memoriam

It came as a shock to all of us to hear about the sudden and untimely death of our friend and fellow lexicographer Deny Arnos Kwary.

I first met Deny in 2007 when he started as a PhD fellow working in the field of lexicography at the Aarhus School of Business, Denmark. I was Deny's supervisor from 2007 to 2011, when he submitted his thesis with the title *Theoretical Considerations in Creating an Internet English Dictionary of Finance for Indonesian Students*, for which he was awarded his PhD. His project focused on the lexicographic function of reception by Indonesian university students reading English business texts and the needs for help provided by dictionaries.

As a researcher, Deny was a conscientious and dedicated person who was always eager to learn and to try new things in order to constantly improve his knowledge of lexicography and other matters. During discussions, he displayed an ability to make thoughtful and perceptive analyses of both concepts and practices in practical and theoretical lexicography.

Deny had an engaging personality and an outgoing manner, which enabled him to interact well with students, faculty and researchers in various contexts, such as Denmark and Danish culture, which were new to him. For instance, Deny was very disappointed to find that the streets were not covered in snow when he first arrived in Aarhus one day in late November 2007. His work was consistently of good quality and he soon published in international peer-reviewed journals as well as presented papers at international conferences on lexicography. Through his dedication and winning personality, Deny developed a large network within the lexicographic community, which led to, among other things, a three-month stay at the University of Birmingham, as well as becoming editor of the dictionary *Password. English Learner's Dictionary for Speakers of Bahasa Indonesia*. Furthermore, Deny quickly became a household name in lexicography both as a presenter at and an organizer of international conferences. It was a pleasure and privilege to know, supervise and work with Deny Arnos Kwary; he will be sadly missed. May he rest in peace.

Sandro Nielsen
Centre for Lexicography
Aarhus University

AsiaLex 2020
Lexicography and Language Documentation

ASIALEX
The Asian Association for Lexicography

The 14th AsiaLex Conference, AsiaLex 2020, will be held in Yogyakarta in June 2020, and is hosted by Badan Pengembangan Bahasa dan Perbukuan (Agency for Language and Book Development), Ministry of Education and Culture, Republic of Indonesia.

Inspired by the hundreds of local languages spoken in Indonesia, the theme of the conference will highlight language documentation in lexicographical contexts, including the following main topics:

- Lexicography of Indonesian languages
- Lexicography of endangered languages
- Lexicography of specialized languages
- Language data for lexicographic work
- Multimodal data for lexicography
- Corpus lexicography
- Electronic lexicography
- Pedagogical lexicography

META≡ FORUM 2019

Introducing the European Language Grid

8-9 October 2019, Brussels

META-FORUM is the international conference series on powerful and innovative language technologies for the multilingual information society.

The program of the eighth edition of the conference – META-FORUM 2019 – will provide a comprehensive picture of the most recent developments in the European Language Technology industry and research, including language-centric AI.

In addition to presentations and panel discussions, the event will feature an exposition showcasing the latest research and innovation projects, hosting experts from the academia and industry.

META-FORUM 2019 will also offer an opportunity to acquaint with the emerging European Language Grid platform, whose system will be demonstrated for the very first time. The European Language Grid is funded by the European Union H2020 research and innovation programme (https://www.european-language-grid. eu/). It aims to promote technologies tailored to all European languages and cultures and adapt to their social and economic needs. The target groups are users from science, industry and society alike. The goal is to become the central data and information portal and market place for Language Technology developed in Europe for Europe.

Key topics include:

- Overview and demo of the European Language Grid (ELG)
- ELG approach, plan and technical architecture
- Information on the ELG open calls, to be published in 2020 (financial support for third parties)
- Current EU projects in the Language Technology area
- Language Technology and Artificial Intelligence: building bridges in the EU
- European Language Technology community consolidation
- News from the communities
- Reports from industry
- Expo with booths from industry and research
- https://www.european-language-grid.eu/meta-forum-2019/

The invited keynote speakers include Lan Li, of the Chinese University of Hong Kong; Sandro Nielsen, of Aarhus University, Denmark; Dadang Sunendar, of Badan Pengembangan Bahasa dan Perbukuan, Indonesia; and René van den Berg, of SIL International.

Yogyakarta is the capital city of Indonesia's Special Region of Yogyakarta (Sultanate). One of its south-eastern districts, Kotagede, was the capital of the Mataram Sultanate between 1587 and 1613, and the city has served as the capital of Indonesia from 1946 to 1948 during the Indonesian National Revolution. As the only Indonesian royal city still ruled by a monarchy, it is a centre for classical Javanese fine arts and culture. Also renowned as a centre of Indonesian education, Yogyakarta is home to a large student population and to dozens of schools and universities, including Gadjah Mada University, the country's largest institute of higher education and one of its most prestigious.

Partially a result of its proximity to the Borobudur and Prambanan temples, and because of the Javanese court Kraton culture of Kraton Yogyakarta, Yogyakarta hosts a sizable tourist industry. Borobudur temple is a huge Buddhist temple located in a gorgeous setting amidst beautiful mountains and landscape, which houses the largest Buddhist structure in the world. It is merely an hour's drive from the town center and is easily accessible by public transport. Prambanan temple is a massive and impressive complex that comprises several smaller Hindu temples set around the main cluster. Each of the temples is slightly different from the other and is beautiful in its own way, setting a great example of Hindu architecture.

We look forward to welcoming in Yogyakarta for AsiaLex 2020!

Dora Amalia (Convener) and **Luh Anik Mayani** (Co-Convener)
Badan Pengembangan Bahasa dan Perbukuan

Badan Bahasa is a government agency that handles linguistics and literary issues in Indonesia. It is divided into four centers, namely Center for Language Development and Protection, Center for Language Cultivation, Center for Language Development and Diplomacy, and Center for Book Development.

Adam Kilgarriff Prize, 2019

As Trustees of the Adam Kilgarriff Prize – which was set up in 2016 in honour of our brilliant friend and colleague – we approached this year's round of applications with some nervousness. Could the success of the first iteration of the Prize be repeated, when we received eight submissions, including several high-quality ones, from a wide spectrum of researchers, and found a truly worthy winner in Dr. Paweł Rutkowski (see KDN 25, 2017)? This set a high bar, but I'm pleased to report that the 2019 crop of applications, nine in all, again included many excellent proposals and products.

These covered a wide range, from corpus linguistics, to translation studies, to software development. Submissions included projects as diverse as corpus-building for less-well-resourced languages, an online reference resource for a specialized domain, and a software tool for tracking the phonological development of young children. Two applications deserve particular mention. We were impressed by a corpus-driven study of academic English, which demonstrated the central role of phraseology in writers' vocabulary choices. Another especially strong submission looked at ways of analysing textual variation in register-diverse corpora, and proposed new approaches for identifying keywords and accounting for lexical dispersion. What made this research especially appealing was that the whole approach – in questioning widely-accepted assumptions – was so reminiscent of the way Adam himself was always ready to challenge received wisdom

Once again, we found ourselves contemplating an embarrassment of riches, and the judging process involved several in-depth discussions among the Trustees. In the end, we agreed to award this year's Adam Kilgarriff Prize to Matt Kohl for his hip-hop dictionary, *The Right Rhymes*. Firmly based on empirical data, *The Right Rhymes* is a model of corpus-driven lexicography. At the same time, it incorporates many features (such as data visualization and links to external data) which are frequently talked about in the lexicographic literature, but not so often implemented. The result is a serious interactive dictionary which is also a lot of fun to use. Matt will be talking about his project at the forthcoming eLex conference in Sintra, later this year, and once again Adam's spouse Gill has agreed to come and award the Prize.

We were impressed by all the applications, and we thank the other applicants very warmly, not least for giving us the opportunity to learn more about the interesting work they are doing. Finally, I would like to thank my fellow Trustees for the time and energy they devoted to ensuring that every application was rigorously and thoughtfully evaluated.

The Trustees of the Adam Kilgarriff Prize are Miloš Jakubíček, Ilan Kernerman, Iztok Kosem, Michael Rundell (Chair), Pavel Rychlý, and Carole Tiberius.

Michael Rundell

K Dictionaries & Lexicala News

Interns at KD in 2018-2019

KD has offered the following student internship programs during the 2018-2019 academic year: Leyre Abril Calvo and María Castelló Fabregat from the Department of English at Universitat Jaume I in Castello (Spain) have taken part in post editing automatically-generated translations for an English-Spanish learner's dictionary; eleven students from the Department of Turkish at Istanbul University have been involved in a project of reviewing frequency lists for selecting new Turkish headwords: Cansu Akkaya, Züleyha Altunkaynak, Gressa Ipek Balci, Ahmet Mert Dagli, Ceren Görücü, Berkay Havuk, Yiğithan Kocaoğlu, Tahir Meylani, Kaan Tuglaci, Kaan Turan, Necatican Yildirim; and, EMLex students have participated in various lexicographic projects in conjunction with their main language skills: Yuliia Bolbit from Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) worked on a Ukrainian-English glossary; Mochamad Wahyu Hidayat from Universidade do Minho worked on editing translations for a German-Indonesian learner's dictionary; Gulzhazira Yesmachanova from FAU worked on a Kazakh headword list; Zarina Liurova from FAU worked on editing translations for an English-Russian learner's dictionary; and, Chien-An Wang from Uniwersytet Śląski w Katowicach worked on a Traditional Chinese dictionary.

Henrik Køhler Simonsen research visit at KD

Henrik Køhler Simonsen from Copenhagen Business School and SmartLearning visited KD on a grant from the ELEXIS transnational visit program from December 5 to 13, 2018. The visit was held as part of the *Business Models & Startegies for Lexicography* research project, with the aim of exploring new methods, examining new ideas and generating new strategies, including reflections on technologies, platforms, user groups, applications, eco systems, revenue streams, and interoperability with other domains.

Dr Køhler Simonsen carried out a series of seminars on business model innovation, including presentations, interviews and work on business models with the KD team and experts from different domains, such as lexicography, computational linguistics, Artificial Intelligence, machine learning, data and knowledge sciences, digital publishing and business development, and concluded the visit by conducting a business model innovation webinar including presentations by representatives of IBM, Lemma.com, Dictus, Invixio, and KD.

Ilan Kernerman at EMLex Spring Meeting 2019

Ilan Kernerman presented KD and Lexicala to EMLex staff and students at their Spring Meeting held at FAU in Erlangen on February 18, 2019 as part of a workshop on synergy between EMLex and the private sector, including representatives of Duden, PONS, Zanichelli, and Inbenta. https://www.emlex.phil.fau.eu/2019/02/25/workshop-synergy-effects-between-erasmus-mundus-emlex-higher-education-institutions-and-the-private-sector/

PONS Standardwörterbuch ARABISCH / Lexicala Arabic German Dictionary Online

KD's new Arabic-German / German-Arabic dictionary has been published simultaneously on February 18, 2019 in print by PONS and online by KD. The dictionary has been developed by KD on the basis of the Arabic and German

lexicographic cores in its Global series, and the current edition is targeted primarily at speakers of German who are learning Arabic. https://pons.lexicala.com/

PONS is one of Germany's leading language publishers, with a portfolio covering 32 languages and serving 125 million users worldwide. The company is based in Stuttgart and is part of the Klett Group. https://pons.com/

KD wins LTI Award 2019

KD is among the winners of the LT-Innovate Award 2019 of the Association of Language Technology Industry. The award was presented at the closing session of the LTI Summit in Brussels on June 25, 2019. https://www.lt-summit.com/the-sound-of-intelligence

Ilan Kernerman at Korealex 2019

The Korean Association for Lexicography (Korealex) has invited Ilan Kernerman from KD to speak at its 35th conference. The conference will be held at Ewha Womans University in Seoul on August 31, 2019 with the topic of Social Changes and Lexicography. Korealex holds regular conferences twice a year, and this will be the first time it will have an invited keynote speaker from outside Korea.