lexicala Review

Multilingual Data & Knowledge

Lexicala's multilingual lexical data
solutions for language service providers

Lexicala by K Dictionaries is launching a new range of services for the language technology industry and the academia. The focus is on multilingual lexical data solutions that merge human created and curated language resources with automatically generated content. The solutions feature expert parallel corpora, domain classification, morphology, text annotation, and other premium resources for enhancing natural language processing (NLP) tasks and training machine learning (ML) models.

Lexicala is the trade name used by K Dictionaries (KD) to offer cross-lingual resources to language service providers (LSPs) and for collaboration in research and innovation projects. It relies on and expands datasets and methodologies developed by KD over the last three decades, which eventually enable infinite ways of extracting refined monolingual, bilingual and multilingual components and applying them for translation, spellchecking, semantic technologies, speech recognition, knowledge management, language learning, as well as for online dictionaries.

KD has been an international dictionary leader since the 1990s. It began with the breakthrough semi-bilingual dictionary for learners of English, published by prominent local publishing partners in dozens of language editions for millions of students worldwide, and reaching many more users in recent years in digital formats, mainly through collaboration with Cambridge University Press & Assessment on Cambridge Dictionary online, the world's no. 1 website for English learner's dictionaries. At the turn of the century, KD became involved in innovative multilingual dictionaries, starting in cooperation with Finland's leading electronic dictionary publisher, Kielikone, and currently made available on Naver's Open Dictionary PRO platform. In the early 2000s, it pioneered a new approach to creating multi-layer lexicographic resources as part of the Global series, built on an extensive core for each language. This monolingual core serves as a base for developing bilingual pairs and multilingual networks, which are all inter-connected within an overall framework sharing a common technical infrastructure and interoperable with exterior resources. In the last decade, KD was proactive in pioneering the incorporation of lexicographic data and linguistic linked (open) data technologies, primarily in connection with the W3C OntoLex community group, as well as in EU projects, particularly with Semantic Web Company.

The first major step in the new Lexicala direction has been accomplished this year with the delivery of expert parallel corpora to one of Asia's top information technology conglomerates and completing the integration of the first language pair into its Neural Machine Translation systems. As part of this venture, Lexicala provided over a quarter-of-a-million bilingual segments to train ML models and improve the translation engines' performance. The segments consisted of usage examples composed of full sentences and short phrases that were extracted from dictionary entries between the required Asian language and European ones. The corpora were developed by converging manually created content with smart data generation methods, then followed by thorough curation of each sentence pair by local language experts. Contracting with such a leading global player demonstrates the ability to offer highest quality cross-lingual lexical data supported by automated processes and perfected by professional linguists and translators.

Lexicala's expert parallel corpora resources contain millions of bilingual and multilingual segments in more than 20 languages, including low-resourced combinations. Moreover, many of these segments can apply to languages for specific purposes, as they stem from selected senses of dictionary entries that include a subject field label.

Besides the existing collection of KD resources that encompass 50 languages, the new Lexicala services aim to cover all human languages. They include bilingual and multilingual translation, domain classification, morphological forms, part-of-speech tagging, sense alignment, name entity annotation, IPA and audio pronunciation, and other cross-lingual NLP applications. This underlines Lexicala's expansion to the world of AI and illustrates its potential to offer outstanding solutions to the language technology community.

Ilan Kernerman
CEO, Lexicala by K Dictionaries

lexicala Solutions

Lexical Data

Parallel Corpora

API

Dictionaries

Lexicala
Multilingual Lexical Data Solutions

COLING 2022, Gyeongju

COLING, the International Conference on Computational Linguistics, is one of the premier events for natural language processing and computational linguistics. First established in 1965, the biennial COLING conference is held in diverse parts of the globe with participants ranging from top-ranked research centers to emerging countries. Today, the most important developments in our field are taking place not only in universities and academic research institutes, but also in industrial research departments, including technology startups. COLING provides opportunities for all these communities to showcase their exciting discoveries.

COLING 2022 will be held in Gyeongju, the Republic of Korea, from October 12 to 17, with the special theme of **Tackling the grand challenges of the world by promoting mutual understanding through language**. The conference has received a record-breaking number of papers, which feature substantial, original research in all aspects of natural language and computation. The accepted papers will be presented at the main conference on October 12-15, along with three invited talks from Martha Palmer (University of Colorado, USA), Marco Baroni (ICREA and Pompeu Fabra University, Spain), and Kentaro Inui (Tohoku University, Japan). COLING 2022 will host also 20 workshops and 7 tutorials, from October 16 to 17.

website ▶

The 7 Societal Grand Challenges

- Health, demographic change, and wellbeing
- Food security, sustainable agriculture and forestry, marine and maritime and inland water research, and the bioeconomy
- Secure, clean and efficient energy
- Smart, green and integrated transport
- Climate action, environment, resource efficiency, and raw materials
- Changing world – inclusive, innovative, and reflective societies
- Secure societies – protecting freedom and security of the world

COLING 2022

THE 29TH
INTERNATIONAL CONFERENCE ON
COMPUTATIONAL LINGUISTICS

October 12-17, 2022
Gyeongju, Republic of Korea

workshops

The Fifth Workshop on Technologies for Machine Translation of Low-Resource Languages (LoResMT 2022)

The 6th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (LaTeCH-CLfL 2022)

8th Workshop on Noisy User-generated Text (W-NUT 2022)

Pattern-based Approaches to NLP in the Age of Deep Learning (Pan-DL)

Automatic Summarization for Creative Writing

Fifth Workshop on Computational Models of Reference, Anaphora and Coreference (CRAC 2022)

3rd Workshop on Computational Approaches to Discourse (CODI)

Third Workshop on Scholarly Document Processing (SDP 2022)

Fifth Workshop on NLP for Internet Freedom (NLP4IF)

When Creative AI Meets Conversational AI

First workshop on Transcript Understanding (TU)

Third Workshop on Threat, Aggression and Cyberbullying (TRAC-2022)

The First workshop on Customized Chat Grounding Persona and Knowledge

9th Workshop on Argument Mining

The 9th Workshop on Asian Translation (WAT2022)

Graph-based Methods for Natural Language Processing (TextGraphs-16)

Field matters: The first workshop on NLP applications to field linguistics

Social Media Mining for Health Research and Applications – Workshop & Shared Task (7th SMM4H)

Ninth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial)

Performance and Interpretability Evaluations of Multimodal, Multipurpose, Massive-Scale Models

tutorials

- Cognitive and Linguistic BERTology: An Idiomatic Multiword Expression Perspective. Harish Tayyar Madabushi, Carlos Ramisch, Marco Idiart, Aline Villavicencio

- Information Theory in Linguistics: Methods and Applications. Ryan Cotterell, Richard Futrell, Kyle Mahowald, Clara Meister, Tiago Pimentel, Adina Williams

- Uncertainty Estimation for Natural Language Processing. Adam Fisch, Robin Jia, Tal Schuster

- Knowledge Graph Embeddings for NLP: From Theory to Practice. Luca Costabello, Adrianna Janik, Eda Bayram, Sumit Pai

- NS4NLP: Neuro-Symbolic Methods for Natural Language Processing. Maria Leonor Pacheco, Sean Welleck, Dan Goldwasser, Vivek Srikumar, Yejin Choi, Dan Roth

- Analyzing Human Communication: Sociopragmatic and Pragmalinguistic Models of Impoliteness and Verbal Aggression. Ritesh Kumar, Daniel Kadar and Juliane House

- Spatial Language Understanding: Representation, Reasoning, and Grounding. Parisa Kordjamshidi, James Pustejovsky, Marie-Francine Moen

SEMANTiCS2022 –
In the Era of Knowledge Graphs

Martin Kaltenböck

The SEMANTiCS conference series brings together the semantic and language technology communities in Europe and beyond! The 18th edition takes place this year from 13 to 15 September in Vienna, Austria.

The annual conference – that started in Vienna with 80 participants in 2005 – has turned into the biggest European industry event for semantic technologies and is returning to Vienna in 2022 as a hybrid event including 250 participants on-site and another 150 participants online.

The SEMANTiCS series stands for:

- Transfer – exchange ideas between academia and industry

- Engineering – dive deep into semantic web and linked data technologies

- Community – discuss trends and initiate projects

This year the event is organised together with the annual LT-Innovate Industry Summit 2022: Language Intelligence @ Work: Putting the Customer in the Centre, taking place on September 14-15 as an integrated track in SEMANTiCS2022.

This co-organization brings together two vibrant communities that have been working together for many years, but still operate separately most of the time: the Language Technology community and the Semantic Web / Technology community.

Although, on the one hand, Language Technology (LT) is making use of Semantic Web, Linked Data and Knowledge Graph principles and software tools in several areas, as for instance in data modeling or storing data as RDF triples, on the other hand, the Semantic Web crowd has been making use of computational linguistics, language resources and other related data for some time, these two groups collaborate very little. While both disciplines deal with natural language, meaning and context, the overlap of the use of each other's methodologies and technologies is rather small.

SEMANTiCS2022 bridges this gap and brings together researchers, practitioners, software developers and service vendors, as well as end users and customers of both worlds, to discuss best-in-breed approaches, practices, solutions and technologies.

Martin Kaltenböck is co-founder and managing partner of Semantic Web Company, and as CFO responsible for financial, legal and organizational issues. He studied communication science, psychology and marketing at the University of Vienna, is tutor and publishes on semantic data, information and knowledge management, Linked (Open) Data as well as Open (Government) Data and Knowledge Graphs, leads and works in national and international research, industry and government projects, and is on the permanent advisory board of the SEMANTiCS conference series. martin.kaltenboeck@ semantic-web.com

website ▶

The conference starts on September 13 with a workshop and tutorial day, including 4 workshops, 7 tutorials, and a full DBpedia Day, hosted by the DBpedia Association.

In the evening, the Semantic AI MeetUp takes place, bringing together the conference participants with the local AI and semantic web professionals in the form of a free networking event.

On September 14 and 15, the main conference features 6 keynote talks and dozens of scientific and industry presentations spread over the two days, including a conference dinner in the evening of September 14 for all on-site participants.

SEMANTiCS2022 features the following keynote talks

- **Taxonomies at the heart of a corporate brain for contact centers.** Jans Aasman, Franz Inc.

- **Multilingual virtual assistants: Research, applications, and scaling to 1,000 languages.** Jack Fitzgerald, Amazon Alexa

- **UX design and Knowledge Graphs: The perfect match.** Adam Keresztes, IKEA Knowledge Graph

- **(Machine) Translation with a vision: Shaping a translation ecosystem in a biotech company.** Chantal Moucouveia, Roche, and Kerstin Berns, Berns Language Consulting

- **Taming Turing's legacy.** Steve Moyle, Amplify Intelligence and Freeman Clarke

- **Detection and analysis of narratives and propaganda of today's big players: The example of Russian narratives with conclusions from the Ukraine "Special Military Operation".** Christoph Prinz, Anne Dudenhoefer and Gerhard Backfried, Hensoldt Analytics

SEMANTiCS2023 is already in the planning phase, and the date and location will be announced at the closing of SEMANTiCS2022 by the next conference chair, followed by an announcement. Join in for 2022 and stay tuned for 2023!

In the Era of Knowledge Graphs

VIENNA & ONLINE, SEPTEMBER 13-15, 2022

SEMANTiCS & LTI, Vienna 2022

LTI22 — Language Intelligence @ Work: Putting the Customer in the Centre

Philippe Wacker

On 14-15 September 2022 in Vienna (Austria), LT-Innovate organizes its 9th yearly Language Technology Summit, entitled **Language Intelligence @ Work: Putting the Customer in the Centre**.

Language Intelligence @ Work is Europe's major event showcasing the latest developments in multilingual Artificial Intelligence: speech interaction, deep meaning processing, and multilingual communication and cognition. The Summit is an ideal meeting point between technology providers, users and integrators, business leaders, leading-edge developers and researchers, investors and analysts, serving to learn, exchange, network, forge new partnerships, find new clients, and get tuned to the multilingual intelligent future.

After two years of remission due to the COVID-19 pandemic, LT-Innovate decided to team up with five other organizations and to co-organize and co-locate LTI22 with SEMANTiCS2022.

The extensive program has five parallel strands, including these themed highlights of LTI2022:

- Multilingual AI by Jochen Hummel (Coreon), Jan Oevermann (Plusmeta) and Thomas Lindemann (Fyrfeed)

- Towards One-stop-shop Language Intelligence Partner Spheres in the Cloud by Erik de Vrieze (RWS), Klaus Fleischmann (Kaleidoscope), Raniero Romagnoli and Vincenzo Sciacca (Almawave)

- LangOps – A New Industry Paradigm by Britta Aagaard (Semantix), Vasco Pedro (Unbabel), Jochen Hummel (ESTeam) and Bruno Hermann (LT-Innovate Board of Advisors)

Philippe Wacker is Executive Director of LT-Innovate, the association of the Language Technology industry. He has spent the last 30 years in the high tech sector as a manager of leading digital technology networks and his skills encompass technology evangelization and advocacy, strategic project planning, development and management, international project finance, multicultural networking and communication, and digital transformation.

phw@lt-innovate.org

LANGUAGE INTELLIGENCE @ WORK
PUTTING THE CUSTOMER
IN THE CENTRE
Vienna & Online
14-15 September 2022
@LTINNOVATE #LTI22

LT-INNOVATE.ORG
The language technology industry association

http://lt-innovate.org/summit

How can semantic technologies help to
revolutionize the way we do business?

Victoria Penker and Angela DaSilva

If we have learned anything during the Covid-19 pandemic, it is that business as usual is a concept of the past. The way we manage and think about businesses has fundamentally changed as a result of the global crisis and political instability.

In a world of ever-evolving innovation, new technologies and business models are not only disrupting the way corporations used to operate but also transforming the way we regard product, supply, and customer relationships.

Learning how to scale up operations during these times is a necessary skill that both employees and managers will have to adapt to. Digital technologies are a crucial component of this on-going process and hold the potential to build a resilient organization that thrives despite uncertainty.
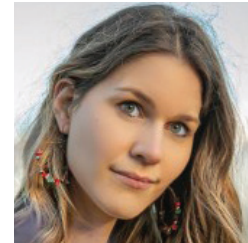
The workplace revolution

Companies have had to quickly adopt new ways to maintain business operations, and they began to recognize the value of a digital workspace that reflects modern work styles, maturing technologies, and new communication approaches.

Following the success of remote work during the pandemic, many businesses are already in the process of transitioning from traditional to flexible workplaces.

Artificial intelligence (AI) has the potential to enhance service operating journeys, streamline activities, and eliminate vulnerabilities, enabling decision makers to facilitate internal and external procedures. As part of this process, usage of semantic technologies is growing in popularity within the context of digital transformation and has become a multifarious powerhouse of corporate operations. Business leaders now have the task of turning data into knowledge and building a solid basis for their company.

The Semantic Web Company

Founded in 2004, the Vienna-based Semantic Web Company is the vendor of the PoolParty Semantic Suite, a complete and advanced semantic middleware on the global market. PoolParty Semantic Suite

Victoria Penker
joined Semantic Web Company in 2022 in the role of Partner Success and Marketing Manager. She studied Comparative Literature, Iberian Studies and German Philology at the University of Vienna and has a diverse background in education, the startup industry and content creation.
victoria.penker@
semantic-web.com

leverages machine learning, natural language processing, and graph technologies to enrich and interlink enterprise data. It uses innovative means to help organizations build and manage enterprise knowledge systems as a basis for their AI strategy.

Content and digital asset management are of particular interest to companies these days, where semantic technologies are being sought out to add an intelligent layer to workflows that have been broken down by hybrid workspaces. To combat these issues, the PoolParty approach to content management is making content more accessible to both internal and external stakeholders. The approach begins with a taxonomy in the PoolParty Thesaurus Server, and encompasses the following:
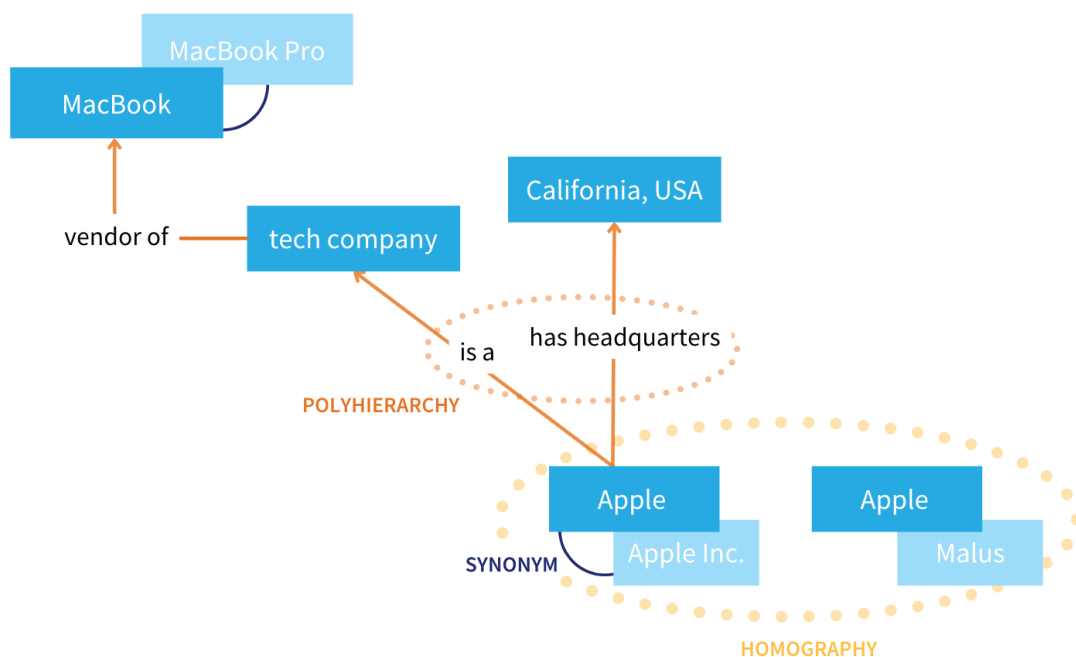
Text analytics and concept tagging

Using natural language processing, PoolParty enables the extraction of semantic concepts to mitigate linguistic challenges commonly found in unstructured text. If we take language ambiguity, for example, *apple* is a fruit, but it is also the name of a technology company; while humans understand the difference, a machine cannot process it unless it is trained.

Angela DaSilva is Content Strategist and Associate Director of Marketing at Semantic Web Company. Previously, she worked as a technical writer and a communications specialist in the software and government fields. She holds a BA in English Language and Literature and in Public Relations from Rhode Island College.
angela.dasilva@semantic-web.com

taxonomy

Semantic text mining can distinguish such occurrences based on a taxonomy and the local context of the ambiguous concepts. Representing a specific knowledge domain, the taxonomy helps to define parameters for which entities can be extracted and synced back to the taxonomy before they are sorted into their corresponding concept schemes and classes. The benefit to maintaining semantic tagging in a taxonomy is the consistency it provides through its hierarchical structure and controlled vocabularies.

Furthermore, semantic concepts in themselves are a uniquely powerful alternative to keyword-based tagging. Semantic concepts ensure that documents are tagged with more than just a singular term, but rather its synonyms and multilingual/alternative labels. In this way, a product like the Apple MacBook can be bundled with words like *laptop*, *notebook*, and *computadora* in a Spanish subsidiary, or by its versioning like the MacBook Pro or MacBook Air.

Semantic search and recommender systems

When content is tagged with concepts, the chief benefit is a vastly enhanced search engine. Semantic search is a solution that leverages the work already done with advanced text mining and concept tagging to take on all the linguistic challenges from a search perspective.

Using the MacBook example from above, if a technical writer enters *MacBook* in the search field, the semantic search engine would find all the articles tagged within that concept. The technical writer who wishes to update all the relevant Help articles with crucial product information can be assured that no articles are missed in the process.

Similarly, the semantic search engine can interpret and gather results based on the user's natural language. Consider the MacBook example from the customer side, where the customer wants to upgrade to a better internal processor but first needs to understand the size of their current one; they get a technician on the phone to help with the upgrade process. The user can search with terms they understand and use daily (e.g. *specs of an Apple notebook*), but their search will also automatically expand to include terms that are synonymous with other parties (i.e. *support teams*, *MacBook Pro specifications*). As the semantic search engine draws upon natural language and the semantics concepts, users can easily find the information they need regardless of their input.

This is especially helpful for recommender systems, which give users' results that are precisely tailored to their search query and can suggest "further reading" content to get a deeper dive into a product explanation. The semantic recommender system does not simply

recommend "more of the same," but enriches the experience with additional information.

A semantic recommender system is unique for its ability to make complex pairings between a user request and the content at hand. The following example is pulled from the PoolParty Wine + Cheese Sample Recommender Demo, which illustrates how a semantic knowledge model in the background helps to establish connections between two different, products (cheese and wine) that complement each other based on their relation to each other in a knowledge graph.

recommender

The facets used in the demo represent the different descriptors that are then weighed against each other to make the flavour pairings.

Increasing findability across digital workspaces

All this is to point towards having better findability and reusability of content, which companies often lack today. Many companies use traditional content management systems that have weak metadata and even poorer user search.

PoolParty Semantic Suite provides a solution with the Intelligent Content Hub, which goes beyond simple curation and maintenance of business assets and adds a much-needed layer of intelligent search and recommendations. As the most complete Semantic AI middleware on the market, one of PoolParty's chief principles is being highly integrable with existing systems to provide semantic search and recommendations.

Within the Intelligent Content Hub, a company can connect its content infrastructure to PoolParty to enable the semantic layer. The semantic layer spans an umbrella over existing content and data management systems to link them together (Knowledge graphs), and essentially "push out" consistency and findability (concept tags and semantic search) back into a hub.

To take full advantage of the hub format, more and more tools can be added to the platform such as an autosuggestion writing tool, a translation engine, a chatbot, and so on. All these platforms and capabilities can come together to ensure that stakeholders have access to data and documents.

The enterprise knowledge graph sits at the center of all these siloed systems and tools, enabling the interoperability of them. When this is in place, companies can use the content hub to its full capabilities, leveraging the relations defined in the knowledge graph and the advanced search performance to access data and content more efficiently.

PoolParty is a tool that does not solely appeal to technicians and knowledge engineers but can also benefit content creators, marketing, or legal experts. The Semantic Suite links business objects and content assets automatically, extracting value from masses of data to make it actionable. It is easy to use, secure, and as simple to integrate as can be. Its user-friendly interface allows users to seamless transition to the software, and its low-learning curve is supported by the PoolParty Academy, which will get you up and running with the software in no time.

Lexicala by K Dictionaries is a technology partner of Semantic Web Company.

Links and Further Reading

Semantic Web Company

PoolParty Semantic Suite

The Knowledge Graph Cookbook

How Text Mining and NLP Drives Your Semantic AI

SALLD-2 @ LREC 2022:
Reinforcing synergies between Sentiment Analysis and Linguistic Linked Data

The second workshop on Sentiment Analysis and Linguistic Linked Data (SALLD-2) was held in conjunction with LREC 2022 in Marseille, France, on June 24, 2022, aimed to foster research and discussion about the interoperability of Linguistic Linked Data (LLD) principles and Sentiment Analysis (SA). More specifically, the half-day workshop provided a platform to explore relevant methodologies, resources, tools, and applications underpinning this connection, as well as to understand the primary approaches – with their respective advantages and limitations – and case studies.

Following on SALLD-1, co-located with LDK 2021 – 3rd Conference on Language, Data and Knowledge, held in Zaragoza, Spain, on September 1, 2021 – the SALLD series was initiated in the framework of NexusLinguarum COST Action (European network for Web-centred linguistic data science, CA 18209), and with its support.

SALLD-2 featured an invited talk by Anna Fensel, Associate Professor at Wageningen University and Research and at the University of Innsbruck, entitled 'From Data to Meaning in Representation of Emotions'. The talk focused on the research by Fensel and her team regarding emotion analysis based on knowledge graphs.

Five peer-reviewed papers were accepted to the workshop: three related to sentiment lexicons, one to the harmonization of language resources, and one to the exploitation of information available in the Linked Open Data cloud for sentiment analysis.

website ▶

publication ▶

video recordings ▶

In the first category, the paper 'Movie Rating Prediction using Sentiment Features', by Apóstolo et al., describes a new sentiment lexicon, Expanded OntoSenticNet (EON), which combines OntoSenticNet with SentiWordNet and is used to predict movie ratings. Likewise, 'Sentiment Analysis of Serbian Old Novels', by Stanković et al., describes a polarity lexicon in Serbian based on three existing lexicons (NRC, Affin, and Bing), the Ontolex-lemon model, and the sentiment vocabulary Marl. Finally, 'Evaluating a New Danish Sentiment Resource: the Danish Sentiment Lexicon, DSL', by Schneidermann and Pedersen, describes an evaluation of Danish sentiment lexicons. We are pleased to highlight the importance of publishing non-English sentiment lexicons for resource-scarce languages.

The second category has been addressed in the paper 'O-Dang! The Ontology of Dangerous Speech Messages', by Stranisci et al., which defines an interoperable knowledge graph to link linguistic resources related to dangerous speech. The graph was populated with eight Italian language resources on dangerous speech, demonstrating the considerable potential of linked data technology to interlink language resources.

Finally, the use of open linked data for sentiment analysis is exploited in 'Correlating Facts and Social Media Trends on Environmental Quantities Leveraging Commonsense Reasoning and Human Sentiments', by McNamee et al. This paper analysed how external information (e.g. temperature or pollution) can be combined with the insights of sentiment analysis.

While SALLD is gradually evolving into a platform where different communities (computer science, linguistics, information science, knowledge engineering, language data science) interact and cooperate within the scope of SA and LLD, we note that the interoperability of these two factors is still rare in most of the submissions we receive. It is hoped that future editions of this workshop will contribute to further strengthening the relevant collaboration between these two fields.

SALLD-2 Organizers

Ilan Kernerman
Lexicala by
K Dictionaries

Sara Carvalho
Universidade de Aveiro

Rachele Sprugnoli
Università degli Studi
di Parma

Carlos A. Iglesias
Universidad Politécnica
de Madrid

Nexus Linguarum

Introducing ODP: Naver's Open Dictionary PRO platform

Jonghwan Kim

Open Dictionary PRO (ODP) is a free online dictionary platform of Naver Corporation. It was launched in 2018 as part of the Naver Dictionary platform, which features over 60 different dictionary services. ODP has been growing exponentially to include over 3 million entries and 4,400 dictionaries, of which 450 turned into Best Dictionary. There are 300 popular languages, such as English, Korean, Chinese, Japanese, French, Spanish, German, Vietnamese, Russian, Arabic, Hindi, Thai, Italian and Turkish, as well as some rare or low-resourced languages such as Belarusian, Daur, Sanskrit and Urdu.

The rapid growth and great success of ODP can be attributed to its smart compilation tools, including video and audio uploaders, and promoting user feedback and proactive participation, as follows:

1. Open Dictionary PRO offers a user-participating dictionary service, consisting of an open-type free dictionary compilation tool that anyone can use to develop a new dictionary easily.

2. The platform allows the inclusion of various forms of dictionaries, glossaries, and encyclopedias, from those having simple structures that are easily manageable by common users to complex ones that are suitable for language experts.

3. Furthermore, the platform is particularly flexible to accommodate the characteristics of diverse languages and the data structures can be freely customized according to different dictionary types.

4. Multimedia editors, as well as image, video, and audio uploaders are available for free.

5. Users may conduct massive compilation work, using Excel uploading and downloading functions, which makes data sharing more friendly.

6. Dictionary makers, owners, and publishers can communicate with their users directly and receive various types of feedback through comments, bookmarks, and cheering messages.

7. New dictionaries can be compiled around a specific topic and developed together in a community with invited members. Thus, users may create a dictionary based on their needs and choose to compile it alone or in a group.
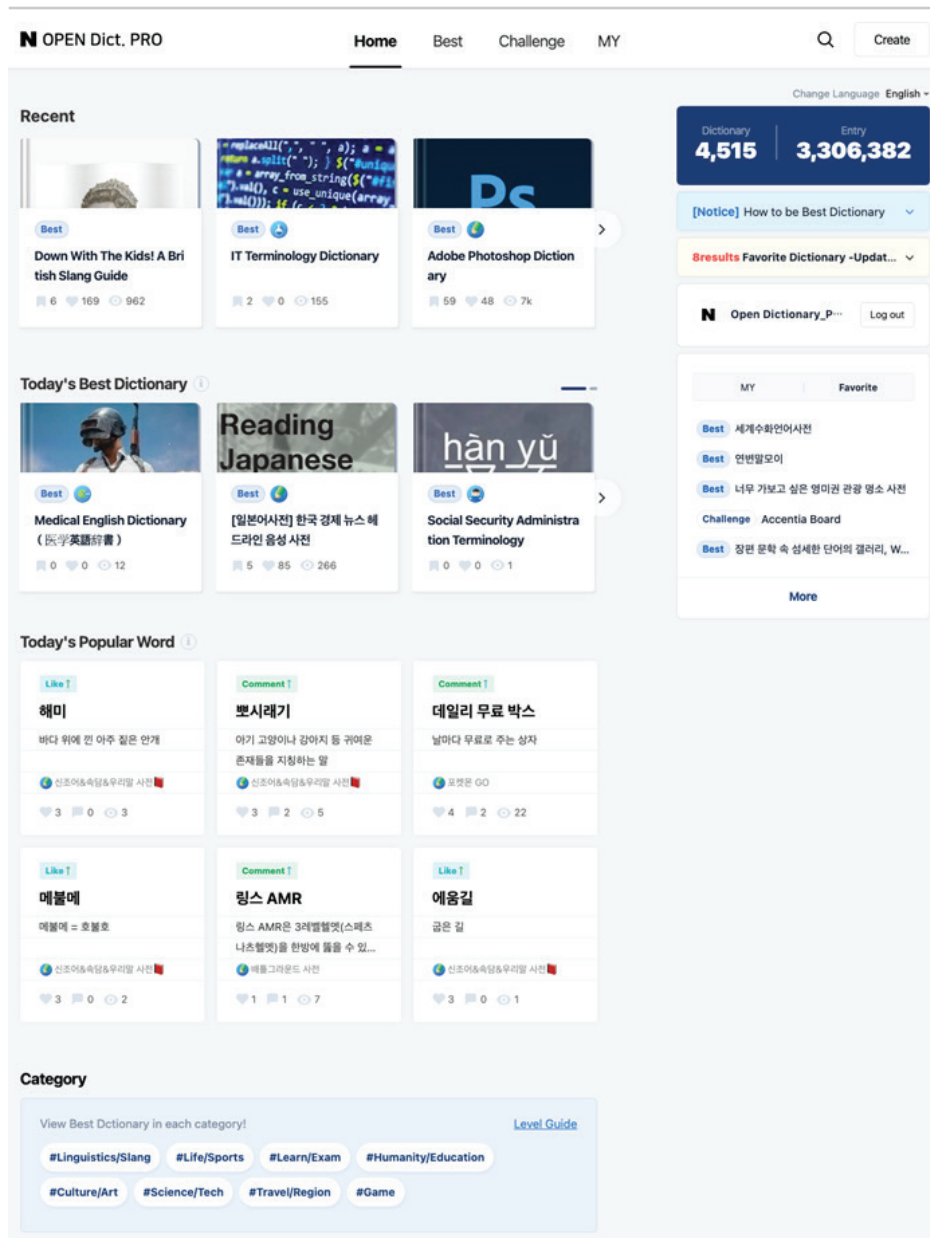
Jonghwan Kim has been leading the dictionary department of Naver Corporation since 2008. He studied Korean History at Seoul National University and has rich working experience at the IT industry in multimedia content planning and managing search content teams. tim.kim@navercorp.com

OPEN Dict. PRO

The wide range of dictionaries on ODP are gaining attention for their diverse topics, such as slang, language learning, everyday words, game glossary, etc., and are becoming popular among writers who need a private space for lexicographic work. More dictionaries of various specialized topics are expected to be constantly created in the future.

NAVER

NAVER Corp. is a global ICT company, providing South Korea's number one search portal NAVER, and its subsidiaries and affiliates offer a wide range of services, including LINE messenger, SNOW camera app, digital comics platform NAVER WEBTOON, group social media platform NAVER BAND, and metaverse platform ZEPETO. The sustained research and development of AI, robotics, mobility, and other future technology trends are propelling NAVER forward, in pursuit of the transformation and innovation of technology platforms, while remaining devoted to a paradigm of shared growth, joining hands with vast numbers of users and partners worldwide. In 2018, NAVER was ranked as the 9th most innovative company by Forbes and top 6th Future 50 company by Fortune magazine.

Lexicala by K Dictionaries collaborates with NAVER Corporation and has provided Naver Dictionary with trilingual Korean dictionaries as well as with parallel corpora including over a quarter of a million bilingual segments to enrich its lexicographic resources, and is about to release dozens of titles including various multilingual dictionaries on the **Open Dictionary PRO** platform.

The 25th Anniversary of ASIALEX:
Past, Present, Future

Hai Xu

This year marks the 25th anniversary of the Association for Asian Lexicography (ASIALEX). Over the past two-and-a-half decades, ASIALEX has grown and exerted a profound impact on the lexicography community in Asia and beyond. In the following, I give a brief overview of the formation, development, and prospects of our association.

1. Birth

ASIALEX was established in 1997 in Hong Kong, at the conference **Dictionaries in Asia: Research and Pedagogical Implications**, which was held at Hong Kong University of Science and Technology on March 27-29. The founders included Gregory James, R.R.K. Hartmann, Tom McArthur, and Amy Chi. Jianhua Huang, from Guangdong University of Foreign Studies, was elected as the first president of the association.

The aim of ASIALEX is "to foster scholarly and professional activities in the field of lexicography by facilitating the exchange of information and ideas through meetings, publications and other means" (cf. https://asialex.org/#about). Before its foundation, four sister lexicography associations were established, namely the Dictionary Society of North America (DSNA, 1975), the European Association for Lexicography (EURALEX, 1983), the Australasian Association for Lexicography (AUSTRALEX, 1990), and the African Association for Lexicography (AFRILEX, 1995). Compared with the diversity of languages and prosperity of dictionary-making in this region, metalexicographic studies in Asia were relatively moderate. An association was therefore called for. Asia boasts a diversity of languages: of the world's 7,151 known living languages, 2,314 can be found in Asia (cf. https://ethnologue.com). There is also a time-honoured tradition of lexicography in Asia. The first wordlists in China were compiled around 2,500 years ago, and Sanskrit lexicography can be dated back to c. 1,500 BC (Considine 2019).

2. Development

ASIALEX held its international conferences biennially until 2015. Since 2016, the conferences have been held annually. Over the past

Hai Xu is a Professor of Applied Linguistics in the Centre for Linguistics and Applied Linguistics at Guangdong University of Foreign Studies. His research interests include lexicography, vocabulary acquisition, and corpus linguistics. He is currently President of the Asian Association for Lexicography (2021-2023), Co-Editor-in-Chief of *Lexicography. Journal of ASIALEX*, and on the editorial board of *International Journal of Lexicography*. xuhai1101@gdufs.edu.cn

website ▶

25 years ASIALEX conferences were hosted in 9 countries: China, India, Indonesia, Japan, Korea, the Philippines, Singapore, Thailand, and Turkey, ranging from Eastern Asia and Southern Asia to Western Asia. Focusing on various themes, these conferences highlighted lexicography of Asian indigenous languages, bilingual lexicography, pedagogical lexicography, corpus lexicography, digital lexicography, dictionary use and education, language documentation, etc. In addition to participants from Asia, we are delighted to meet in each conference delegates from all over the world.

Besides international conferences and pre-conference workshops, ASIALEX has organized other events. Most recently, in December 2021, the First International Symposium on Cross-Cultural Lexicography was successfully held online by Sichuan International Studies University.

In 2014 another important event occurred: the inauguration of *Lexicography. Journal of ASIALEX*, the first academic journal devoted to lexicography in Asia. The journal was first published by Springer, and since 2021 it is published by Equinox. "Given this diversified yet also unique nature of Asian languages, it is extremely urgent and important to conduct more research to keep a record of and to preserve languages in this part of the world. Promoting the collection of words through fieldwork, the compilation of dictionaries or even small glossaries for indigenous languages, among other things, would be much needed to achieve this goal." (Tono 2014: 2) Together with *International Journal of Lexicography*, *Lexikos*, *Dictionaries. Journal the Dictionary Society of North America*, and *Lexicographica*, our journal promotes research all over the world through publishing the latest lexicographic studies, sharing experience in dictionary making, and disseminating state-of-the-art knowledge of lexicography. In April 2020, the journal reached a milestone by being indexed in Scopus, and was ranked Class A (the top rank) by the Italian Ministry for Education and Scientific Research (ANVUR).

3. 25 years ahead

In the coming years it is expected that ASIALEX will hold conferences in a wider range of venues, attract more participants, and upgrade its journal. Currently, the venues of ASIALEX international conferences are mainly confined to Eastern Asian countries, which may partly be due to the prosperity of lexicography in this region. It is good to see that in recent years there has been a growing interest in lexicography in more countries, like Malaysia and Vietnam. In the post-pandemic days when the economy is gradually revitalizing, we

---

ASIALEX international conferences

1997 Hong Kong
Dictionaries in Asia: Research and Pedagogical Implications

ASIALEX 1999 Guangzhou, China
National Experience in Lexicography or Dictionary Compilation and Bilingual Lexicography

ASIALEX 2001 Seoul, Korea
Asian Bilingualism and the Dictionary

ASIALEX 2003 Tokyo, Japan
Dictionaries and Language Learning: How can Dictionaries Help Human and Machine Learning?

ASIALEX 2005 Singapore
Words in Asian Cultural Context

ASIALEX 2007 Chennai, India
Asian Lexicography: Retrospect and Prospect

ASIALEX 2009 Bangkok, Thailand
Dictionary in Education

ASIALEX 2011 Kyoto, Japan
Lexicography: Theoretical and Practical Perspectives

ASIALEX 2013 Bali, Indonesia
Lexicography and Dictionaries in the Information Age

>

believe that more Asian countries will bid to host of an ASIALEX international conference.

We are also trying our best to attract more participants, particularly the younger generation of lexicographers and researchers. In the digital era, there is more room for fascinating research in lexicography, such as the building and application of digital lexicographic databases, semi-automatic dictionary making, and digital literacy in lexicography. Through conferences, workshops, symposiums, etc., we will encourage more young people to pursue this line of research.

Last but not least, we hope that *Lexicography. Journal of ASIALEX* will be indexed in SSCI and A&HCI in the near future. Instead of being many, there are very few journals devoted to lexicography. One more SSCI/A&HCI journal will certainly extend the influence of lexicography in academia.

Journals

Dictionaries: Journal of the Dictionary Society of North America. https://muse.jhu.edu/journal/540

International Journal of Lexicography. https://academic.oup.com/ijl

Lexicographica. https://degruyter.com/journal/key/lexi/html

Lexicography. Journal of ASIALEX. https://journal.equinoxpub.com/lexi

Lexikos. https://lexikos.journals.ac.za/pub

References

Considine, J. (ed.) 2019. *The Cambridge World History of Lexicography*. Cambridge: Cambridge University Press.

Ethnologue: Languages of the World. Accessed on August 9, 2022. https://ethnologue.com

Tono, Y. 2014. Lexicography in Asia: Its future and challenges. *Lexicography*. *Journal of ASIALEX* 1.1, 1-5.

>

ASIALEX 2015
Hong Kong, China
Words, Corpora and Dictionaries: Innovations in Reference Science

ASIALEX 2016
Manila, the Philippines
Advancing Language Teaching with Lexicography and Corpus-Building

ASIALEX 2017
Guangzhou, China
Lexicography in Asia: Challenges, Innovations, and Prospects.

ASIALEX 2018
Krabi, Thailand
Lexicography in the Digital World

ASIALEX 2019
Istanbul, Turkey
Lexicography: Past, Present and Future

ASIALEX 2021
Online (host: Badanbahasa, Indonesia)
Lexicography and Language Documentation

ASIALEX 2022
Online (host: Guangxi Minzu University, China)
Asian Lexicography in the Digital Age: Challenges and Solutions

ASIALEX
The Asian Association for Lexicography

AISALEX 2023, Seoul

Kilim Nam and Yelee An

The 16th International Conference of the Asian Association for Lexicography (ASIALEX) will be organized by the Korean Association for Lexicography (Korealex) and held on June 22-24, 2023, at Yonsei University in Seoul, South Korea. Founded in 2003, Korealex organizes two conferences a year and publishes the *Journal of Korealex* annually. Yonsei University was established in 1885 and today figures among the oldest and most prestigious Korean universities. It will be host to ASIALEX for the second time, 22 years after the second biennial conference held successfully in 2001.

ASIALEX aims to promote scholarly and professional activities in the field of lexicography. Founded in 1997, it strives to deepen lexicographic research activities and broaden networks among lexicographers and other stakeholders all over the world, by organizing events and publishing the peer-reviewed journal, *Lexicography. Journal of Asialex*.

The theme of ASIALEX 2023 is **Lexicography, Artificial Intelligence, and Dictionary Users**. While proposals on other topics related to the study and use of dictionaries are also welcome, ASIALEX 2023 aims to provide opportunities to discuss the changes and challenges that go beyond the realms of traditional lexicography and seek new directions and perspectives for lexicography and dictionaries to cope with social problems and changes. These issues will also be broached by the four keynote speakers, namely, Jonghwan Kim (Naver Corporation), Erin McKean (Wordnik.com), Michael Rundell (Lexical Computing), and Yukio Tono (Tokyo University of Foreign Studies), primarily in relation to innovative achievements in lexicography, AI integration, and dictionary user research.

Dictionaries, including their accompanying resources and tools, methodologies and technologies, platforms and publication formats, have been continuously evolving according to changes in the trends and cultural contexts of the times. As lexicography undergoes periods of transition, researchers have questioned the future of dictionaries and dictionary makers (e.g. Grefenstette's 'The Future of Linguistics and Lexicographers: Will there be Lexicographers in the year 3000?', as early as 1998; Esposito 2002; or the EURALEX 2010 roundtable discussion on the theme 'Will there be people who make dictionaries in 2020?', cf. Bae 2016). Now that we have passed 2020, fortunately,

Kilim Nam is a professor at the Department of Korean Language and Literature at Kyungpook National University. She has a PhD in Korean linguistics (on the copula ida structures in contemporary Korean, 2004) from Yonsei University and was Visiting Scholar at SOAS University of London and UCLA Center for Korean Studies. She is the vice-president of Asialex. nki@knu.ac.kr

Yelee An is Associate Professor of Korean linguistics at the Academy of Korean Studies. She has a PhD in Korean linguistics from Yonsei University (2013), and was Visiting Fellow at Harvard Yenching Institute and Visiting Scholar at the University of Hawaii at Manoa. She is the convener of ASIALEX 2023. anyelee@aks.ac.kr

website ▶

the activities of many associations and researchers in the field of lexicography remain strong and ongoing. Although commercial models based on profit structures of print dictionaries no longer exist, the demand for refined language resources and the power of language information seem to have become more pertinent. The questions we are faced with are thus related to what opportunities as well as crises dictionaries and lexicography face. With this in mind, we look forward to discussing the cultural roles of lexicography and lexicographers, the value of language information in the AI era, and dictionary users themselves as major topics. The following points detail our intention to propose the theme of Lexicography, Artificial Intelligence, and Dictionary Users for ASIALEX 2023.

Dictionaries in the Age of Artificial Intelligence

In the current era of AI, dictionaries exist not just for human beings, but also for machines, and this shift urges us to deepen the discussion of theoretical lexicography and to expand the scope of dictionaries more flexibly. While the word has long been considered the basic unit of dictionary entries, it is now necessary to consider how to better adopt typically unregistered categories, such as neologisms, non-standard forms, loanwords, hate speech, slang, and pragmatic or nonverbal information, which have often been neglected in traditional lexicography. As Sinclair et al. (2004) referred to an ideal dictionary as containing all semantic units, it is time to consider the useful extensions and forms of a dictionary containing all such semantic units used in everyday communication.

Implication and Significance for and of Dictionary Users

Not only have the boundaries of what is considered a dictionary expanded, but the definitions of dictionary users have extended as well. As the term 'machine readable' indicates, nowadays dictionary users include machines as well as humans. Nonetheless, even dictionaries designed for machines ultimately aim to represent human language and to serve human beings. For a dictionary to properly function as a medium connecting human intuition and machines, it is necessary to think about how to represent knowledge of the world more precisely.

Popularization of Lexicography and the Role of Professionals

Finally, we hope that this conference will lead to discussions on popularizing dictionaries and fostering subsequent generations of lexicographers. Dictionaries are found all around us, and they are used everywhere in our daily lives, although we may not be aware of their presence. Despite the achievements of lexicography throughout

KEYNOTES

Jonghwan Kim
Leader, Global Dictionary Production Center, Naver Corporation
Development of User Participation at the Naver Dictionary: A case study of Open Dictionary PRO and Accentia

Erin McKean
Founder, Wordnik.com
Investigating the ROI of AI for Lexicography

Michael Rundell
Chief Lexicographic Officer, Lexical Computing
Automating the creation of dictionaries: Are we nearly there?

Yukio Tono
Professor in Corpus Linguistics, Tokyo University of Foreign Studies
Research into dictionary use in an era of e-lexicography

human history and the relatively recent corpus revolution (Rundell and Stock 1992, Rundell 2008, Hanks 2012), the study of dictionaries does not seem to be widely appreciated by the public. For the public in general, dictionaries are still difficult to use, and lexicography is an unknown area. Scholars and professionals in lexicography thus need to seek out the desired identity of dictionaries as required in modern times by approaching and interacting with the public. We hope that ASIALEX 2023 will present opportunities to diagnosing modern social communication problems by gaining a better understanding of the public use of language, and listening to the needs of a new, modern era with a more flexible attitude toward the structures, forms, and boundaries of lexicography and dictionaries.

References

Bae, Y. 2016. The evolution and future of a dictionary. *Journal of New Korean Language Life*, 26.4, 43-64. https://www.korean.go.kr/nkview/nklife/2016_4/26_0403.pdf

Esposito, J.J. 2002. Dictionaries, another Netscape? *Kernerman Dictionary News*, 10, 1-3. https://lexicala.com/wp-content/uploads/kdn10_2002.pdf

Grefenstette, G. 1998. The future of linguistics and lexicographers: will there be lexicographers in the year 3000? In *Proceedings of the 8th Euralex Congress*, Liege, 25-41. http://euralex.org/wp-content/themes/euralex/proceedings/Euralex%201998%20Part%201/Gregory%20GREFENSTETTE%20-The%20Future%20of%20Linguistics%20and%20Lexicographers_%20Will%20there%20be%20Lexicographers%20.pdf

Hanks, P. 2012. The Corpus Revolution in Lexicography. *International Journal of Lexicography*, 25.4, 398-436. https://academic.oup.com/ijl/article-abstract/25/4/398/923874?redirectedFrom=fulltext

Rundell, M., 2008. The corpus revolution revisited. *English Today*, 24.1, 23-27. https://www.cambridge.org/core/journals/english-today/article/abs/corpus-revolution-revisited/ABEFA306EE99B1980D954D8A00ECBC6E

Rundell, M. and Stock, P. 1992. The corpus revolution. *English Today*, 8.4, 45-51. https://www.cambridge.org/core/journals/english-today/article/abs/corpus-revolution/20B4172D6CF9505E7E35B4CC99E3227E

Sinclair, J., Jones, S. and Daley, R. 2004. *English Collocation Studies: The OSTI Report*. London: Continuum. https://www.bloomsbury.com/us/english-collocation-studies-9780826474889/

ASIALEX
The Asian Association for Lexicography

GWLN — Globalex Workshop on Lexicography and Neology.
An overview

Annette Klosa-Kückelhaus and Ilan Kernerman

Lexicography has been around for thousands of years and has always had to adapt to developments in society and language, apparently more than ever in the past generation with its notably fast and fundamental changes. Neology has been there even before, ever since human language itself was born, driving it from the start and so-to-speak inciting lexicography. In the recent decades it is also drawing more attention in research communities and practical applications, such as at university observatories or in the language technology industry, as well as with the public. The speed of novelty in daily life accelerates and the volume of innovations grows exponentially – all defined by language and both affected by and affecting language. Overall, there is now greater interest in neologisms and in the role of lexicographic methodologies and resources to capture and communicate them to the world.

The vast field of research of neologisms (referring here to new words, new multiword units, new elements of word formation, and new meanings of either of them) pertains to their origin (stemming from a given language, as in new word formation, or loan words from other languages, today mainly from English, as well as combining both), distribution (in general language or domain-specific, that is terminology), identification (applying corpus linguistics methods, editorial methods, user generated candidates, and comparison of different methods), evaluation (such as in blogs and chats), and more.

The Globalex Workshop on Lexicography and Neology (GWLN) series was launched with the support of the Global Alliance for Lexicography (Globalex) at the 22nd biennial meeting of the Dictionary Society of North America (DSNA 2019). Since then, GWLN has been held annually, alternating in conjunction with the international conferences of the continental associations for lexicography, with the workshop papers published in different channels. Our overall aim is to explore this intersection of neology and lexicography worldwide, uncover common factors and highlight individual features, expose and share the findings with each other, and enhance relevant mutual understanding, professional competence

Annette Klosa-Kückelhaus heads the Lexicography and Language Documentation program area in the Department of Lexicology at the Leibniz-Institute for the German Language (IDS) in Mannheim. She has hosted the XX EURALEX International Congress (2022) and currently acts as EURALEX president.

klosa@ids-mannheim.de

and cooperation, and user satisfaction. The broad issues begin with the detection and identification of neologisms and continue with their analysis and classification and lexicographic treatment and representation. The main GWLN topics concern the following lexicography-driven or -oriented aspects of neology and neologisms:

- How to interoperate lexicographic datasets with online resources and incorporate neologisms into dictionaries (the media, formatting, labelling, etc.)

- How to deal with grammatical/orthographic/pronunciation variation (descriptive vs. prescriptive approaches)

- How to explain meaning with/without encyclopedic information, and how to use illustrations and audio-visual media

- How well are neologisms that are integrated in dictionaries accepted by the community (rejection of new words and language purism)

- 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 dictionaries of neologisms; in special domain dictionaries)

- How to deal with neologisms that are no longer *new* and with those no longer used

- How can dictionary users help with finding and informing about neologisms

GWLN-1 @ DSNA 2019

The first workshop took place at the 22nd Biennial Meeting of the Dictionary Society of North America at Indiana University, Bloomington on 8 May 2019. It comprised thirteen invited papers from around the world on twelve languages (including two on English), with a few of the papers presented online. Eight of these form a special issue of the DSNA's journal *Dictionaries*, published the following year.

(webpage ▶) (program ▶) (abstracts ▶) (publication ▶)

Ilan Kernerman is CEO of Lexicala by K Dictionaries. His main interests include the interoperability of language and technology through multilingual data and knowledge management. He is involved in various research and innovation projects, workshop organization, journal editing and reviewing, is founder of Globalex and was the first chair of its management committee, and past president of Asialex.
ilan@lexicala.com

GWLN-2 @ EURALEX 2020/2021

GWLN 2020 was due to be held in Alexandroupolis, Greece as part of the 19th biennial Euralex Congress, but was postponed because of the COVID-19 pandemic to 2021 and then had to be changed to virtual mode. Five of the papers were given as part of a workshop we held independently online on 4 November 2020, attracting 112 participants worldwide, and five others were presented on 8 September 2021 as part of the Euralex conference online (including nearly 30 participants). Altogether, six of these papers formed a special issue of *International Journal of Lexicography* entitled Lexicographic Neology and Neological Lexicography (three of which concern Spanish, along with Czech, Slovak, and Brazilian Portuguese), and another one was published in the EURALEX Proceedings.

webpage ▶   program ▶   video recordings ▶   publication ▶

GWLN-3 @ Australex 2021

The third iteration of GWLN was held in conjunction with Australex 2021, online too, and was devoted entirely to Coronavirus-related neologisms in lexicographic context. Thirteen papers presented at this workshop are currently being prepared for publication in a volume of *Lexicographica. Series Maior* (Issue 163, due 16 December 2022). They focus on ten languages: one Altaic (Korean), one Finno-Ugric (Hungarian), one Slavic (Croatian), two Germanic (English and German) and four Romance (French, Italian, [Brazilian and European] Portuguese, and [Pan-American and European] Spanish), as well as the Sign Language of New Zealand. 125 participants from all over the world joined the workshop.

webpage ▶   publication ▶

GWLN-4 @ EURALEX 2022

In 2022, the fourth iteration of GWLN was held as an in-conference workshop in Mannheim, Germany as part of EURALEX 2022, on July 15. Six papers on African, Asian, European, and South American languages were presented, of which four have been published as part of the conference proceedings.

webpage ▶   publication ▶

Dictionaries: Journal of the Dictionary Society of North America, 41.1
Special Issue: Global Viewpoints on Lexicography and Neologisms
Guest Editors: Annette Klosa-Kückelhaus and Ilan Kernerman
2020. DSNA

International Journal of Lexicography, 34.3
Special Issue: Lexicographic Neology and Neological Lexicography
Guest Editors: Ilan Kernerman and Annette Klosa-Kückelhaus
September 2021.
Oxford University Press

>

GWLN-5 @ ASIALEX 2023

The fifth edition of GWLN will be held in conjunction with the 16th International Conference of the Asian Association for Lexicography, Asialex 2023, at Yonsei University, Seoul, Korea (June 22-24), and selected papers will be published as a special issue of *Lexicography. Journal of Asialex* (Issue 10.2, December 2023). The main topic of GWLN-5 will be unregistered words, i.e., lexical items not registered in dictionaries, including – but not limited to – words and expressions found in speech corpora, social media, hate speech, poetry, etc., or simply non-standard words or newly coined terminology. The first call for papers was issued in July 2022.

> webpage ▶

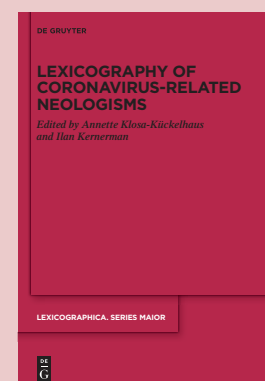**Annette Klosa-Kückelhaus** and **Ilan Kernerman** are the GWLN organizers and publication editors.

Kilim Nam is advisor for GWLN-5.

>

Lexicographica. Series Maior, 163
Lexicography of Coronavirus-related Neologisms
Edited by Annette Klosa-Kückelhaus and Ilan Kernerman
December 2022. Berlin: De Gruyter

globaLex

The Adam Kilgarriff Prize: Fourth iteration

Once again, we are seeking applications for the Adam Kilgarriff Prize. If you're interested in submitting an application, you will find all the information you need at https://kilgarriff.co.uk/prize/.

The deadline for submissions is 30 September 2022, and a winner will be announced on or before 31 December 2022. As always, the Prize itself will be awarded at the forthcoming eLex conference. The 2023 edition of eLex will be held next June in Brno, Czech Republic.

The Adam Kilgarriff Prize was established to honour our friend and colleague Adam, who died far too young in 2015. Adam was a rare talent, who was equally at home in the very different (but complementary) worlds of computational linguistics and lexicography. While contributing hugely to developments in natural language processing (NLP), Adam worked closely with dictionary developers, applying his NLP expertise to finding smart solutions to practical lexicographic problems. In the process, numerous innovative technologies came into being which are now routinely used across the lexicographic world in particular and computational linguistics in general.

In the mid-1980s, Adam embarked on a PhD at the University of Sussex, with the simple title 'Polysemy'. This was a tremendously ambitious investigation, but, as his co-supervisor Roger Evans (2015) said: "Pulling off the one-word title is a tall order, requiring a unique combination of focus and coverage, breadth and depth, and, most of all, authority. Adam completely nailed it".

Adam continued his research at Brighton University where, along with David Tugwell, he ran the WASPS project investigating the links between collocation and word senses. This in turn fed into the development of word sketches, a primitive version of which was used in the late 1990s, during the compilation of the *Macmillan English Dictionary* (Rundell and Kilgarriff 2011, 267-269).

Together with Sue Atkins and myself, Adam set up Lexicom in 2001 — an intensive 5-day workshop in lexicography and computational linguistics — and the Lexicom workshops have continued to run annually ever since. One of the participants at the first Lexicom, Pavel Rychlý, demonstrated an impressively fast concordancing program he had developed. Adam immediately teamed up with Pavel to create Sketch Engine, which brought together a corpus (initially the British National Corpus), a concordancer, and Adam's word sketch

website ▶

technology. Sketch Engine has been in continuous development since then, and is now an industry-standard package used in language projects worldwide.

Adam became ill in late 2014, and died from colon cancer the following spring. His unique talents, remarkable achievements, and enormous impact on NLP, corpus linguistics and lexicography inspired some of his friends to set up a prize in his honour. The biennial Adam Kilgarriff Prize has already had three exceptional winners from very diverse areas of the field, and we are now looking for a fourth. Please get your applications in by 30th September!

Michael Rundell

Chair of Trustees, Adam Kilgarriff Prize

Evans, R. 2015. Obituary for Adam Kilgarriff. *Computational Linguistics*, 41.4, 719-721.

Rundell, M. and Kilgarriff, A. 2011. Automating the creation of dictionaries: where will it all end? In Meunier F., De Cock S., Gilquin G. and Paquot M. (eds.), *A Taste for Corpora. A tribute to Professor Sylviane Granger*. Benjamins: 257-281.