Neology in Practice:Lexicographic and Terminological Approaches to Lexical Innovation

Jelena Kallas[1], Kristina Koppel[1], Kris Heylen[2], Ilan Kernerman[3], Ana Ostroški Anić[4], Federica Vezzani[5], Špela Arhar Holdt[6]

[1] Institute of the Estonian Language, Tallinn, Estonia
[2] Instituut voor de Nederlandse Taal, Leiden, the Netherlands
[3] Lexicala by K Dictionaries, Tel Aviv, Israel
[4] Institute for the Croatian Language, Zagreb, Croatia
[5] Università degli Studi di Padova, Padua, Italy
[6] University of Ljubljana, Ljubljana, Slovenia
E-mail: jelena.kallas@eki.ee, kristina.koppel@eki.ee, kris.heylen@ivdnt.org, ilan@lexicala.com, aostrosk@ihjj.hr, federica.vezzani@unipd.it, spela.arharholdt@ff.uni-lj.si

Abstract

The COST Action 'European Network on Lexical Innovation' (ENEOLI) has conducted a comprehensive survey in October-November 2024 regarding the methods, practices, tools, and resources used in the study and documentation of lexical innovations, including neologisms and novel senses. The 249 respondents from 50 countries represented linguists, lexicographers, terminologists, translators, software developers, and educators. Respondents could indicate more than one field of expertise, and 169 noted theirs as linguistics (70%), 107 lexicography (44%) and 105 terminology (43%). In this paper, we focus on the responses of those indicating their field of expertise as lexicography and/or terminology, and we analyzed their approaches to the identification and documentation of neologisms, the composition of project teams and the use of corpora and digital tools. Special attention is given to training pathways and professional needs, offering insights into the evolving skills required in the field of lexical innovation.

Keywords: neologisms; neology; lexical innovation; lexicography; terminology; survey

1. Introduction

The main objective of the COST Action 'European Network on Lexical Innovation' (ENEOLI, CA22126[1]) is to connect researchers, educators, translators, terminologists, and other stakeholders involved in the field of neology. ENEOLI aims to explore the core terminology of this discipline, showcase advanced methodologies, conduct comparative studies of lexical innovation across languages, and provide targeted training for various

[1] https://eneoli.eu (7 October 2025).

professional groups. The Action began in 2023 and will continue until 2027, addressing several fundamental challenges in the evolving field of neology.

One of the central aims of ENEOLI is to define and standardize the core terminology of neology as an emerging discipline. This will be achieved through the creation of a born-digital, open-access multilingual glossary to facilitate research and collaboration on an international scale. In parallel, the Action promotes the adaptation and advancement of digital tools and methodologies for the identification and analysis of lexical innovations. By engaging institutions, experts, and the wider public through crowdsourcing, the network promotes awareness of neological activities and their broader societal implications, encourages linguistic creativity in different languages, and fosters clarity in scientific communication.

Another important area of focus is the comparative study of lexical innovation across European and other languages, particularly in relation to lexical borrowings and their native equivalents. In the context of ENEOLI, lexical innovation is understood as a concept broader than neology, although it is often used interchangeably. *Lexical innovation* is a linguistic phenomenon that not only refers to the creation of new words, but also to semantic extensions and metaphorical usage that indicate changes in meaning. *Neology* is the term that denotes both the process and the study of how new words and phrases enter a language.[2] The studies of lexical innovation within ENEOLI aim to shed light on shared trends and language-specific patterns in lexical development. At the same time, the Action places strong emphasis on capacity building by developing specific training materials in neology. These are designed for translators, editors, journalists, and educators, and are intended to be adaptable to any European language context.

As part of establishing a comprehensive understanding of existing practices, tools, and resources in the field, Working Group 2 of ENEOLI (Methods, Digital Resources, and Tools for Neology) has conducted a large-scale survey as a foundational step toward achieving its broader goals. To the best of our knowledge, no major studies have previously investigated how neologism work is conducted within lexicographic and terminological projects in Europe, which suggests that this aspect merits more explicit attention. Hence, in this paper, we focus on presenting the lexicography- and terminology-related findings of this survey.

The remainder of this paper is structured as follows. Section 2 describes the methodology and practical aspects of the survey design and implementation. Section 3 presents the overall survey results, mainly emphasizing its diverse lexicographic and terminographic aspects and implications, and Section 4 draws the conclusions and perspectives of this work.

[2] See NeoVoc, a multilingual vocabulary of Neology, for a definition of neology: https://eneoli.wikibase.cloud/wiki/NeoVoc (7 October 2025).

2. Methodology and Practical Aspects of Survey Design and Implementation

2.1 General Principles and Aims

The main aim of the survey was to get an overview of the methods, practices, tools, and resources employed in the study and documentation of lexical innovation. The survey targeted research groups and individual researchers or language specialists involved in this field, and it was organized around specific projects or research topics. Projects may have been linked to resource creation in lexicography and terminology, software development for neology tracking, or academic research in lexicology, linguistics, or translation studies. Additionally, a dedicated section focused on teaching and training practices related to neology.

2.2 Structure and Implementation

The survey questionnaire was developed over several months in 2024, with contributions from experts in lexicography, terminology, corpus linguistics, NLP, translation, and neology teaching. A draft version of the survey was internally tested and iteratively revised to enhance clarity, technical functionality, and content relevance. The final version of the questionnaire was implemented on the 1KA platform[3] and consisted of 294 questions, generating a total of 850 variables.

The questionnaire was modular and dynamically structured. Respondents began with a general metadata section, providing information about their background and the specific project or context on which they were reporting. Based on their self-identified roles and expertise, participants were then guided to (one or more) tailored sections pertaining to: *Lexicography*, *Terminology*, *Linguistics* and *Language Research*, *Translation*, *Software* and *Corpus Development*, and *Teaching and Training*. The survey included both closed-ended and open-ended questions. Although participants were encouraged (and reminded) to answer all items, it was possible to proceed through the questionnaire without responding to every question.

Dissemination of the survey followed a broad, yet targeted strategy designed to reach relevant academic and professional audiences across disciplines. The promotion was through ENEOLI's internal mailing lists and website, LinkedIn, Facebook, as well as via relevant institutional newsletters, professional associations, discussion lists, and at academic events. To maintain visibility and encourage participation, several reminders were issued during the data collection period. Referral data indicate that most

[3] https://1ka.arnes.si/ (7 October 2025).

responses were submitted via a direct link – highlighting the continued effectiveness of email distribution – while additional responses came from the ENEOLI website, social media platforms, and established academic channels such as *linguistlist.org*.

The survey was open from 24 September to 19 November 2024, with some additional responses collected in December 2024. During this period, 1,824 unique users (human or bots) accessed the introductory page, 524 proceeded to the first page of questions, and 307 began responding. After excluding fully empty or idle sessions, 249 responses were retained for analysis–139 completed in full and 110 completed partially. The median completion time for valid responses was 15 minutes and 19 seconds.

After the survey closed, responses were exported from the platform for quantitative and qualitative analysis.

3. Results

3.1 General Data

A total of 249 respondents from 50 countries participated in the survey, forming a highly international group. The ten most frequently represented countries were Italy (10%), Montenegro (9%), Spain and France (8% each), Croatia and Germany (4% each), followed by Estonia, Belgium, Serbia, and Sweden (3% each). Although the majority of respondents were based in Europe, the survey also reached participants from South and North America (Argentina, Brazil, Chile, United States, Canada), the Middle East and Asia (Lebanon, Israel, Afghanistan, South Korea, India, Georgia, Turkey), and Africa (Côte d'Ivoire, Senegal, Algeria).

In terms of disciplinary expertise, respondents (n=243) could select multiple options (see Figure 1). Most respondents had backgrounds in linguistics, with substantial representation also in lexicography, terminology, and lexicology. Many respondents identified their main domains as academic teaching, translation studies, corpus linguistics, and NLP / Computational linguistics. A smaller group reported working in artificial intelligence or software development, and some respondents identified additional areas of expertise, including sign language, phonology, etymology, language planning, and English literature.

The respondents' professional experience varied in length, but most had long-standing experience in their field. The largest group (33%) reported 16–25 years of experience, followed by 25% with over 25 years of experience. Additionally, 44% indicated membership in the ENEOLI network, suggesting that the survey had a broader reach than its original scope.

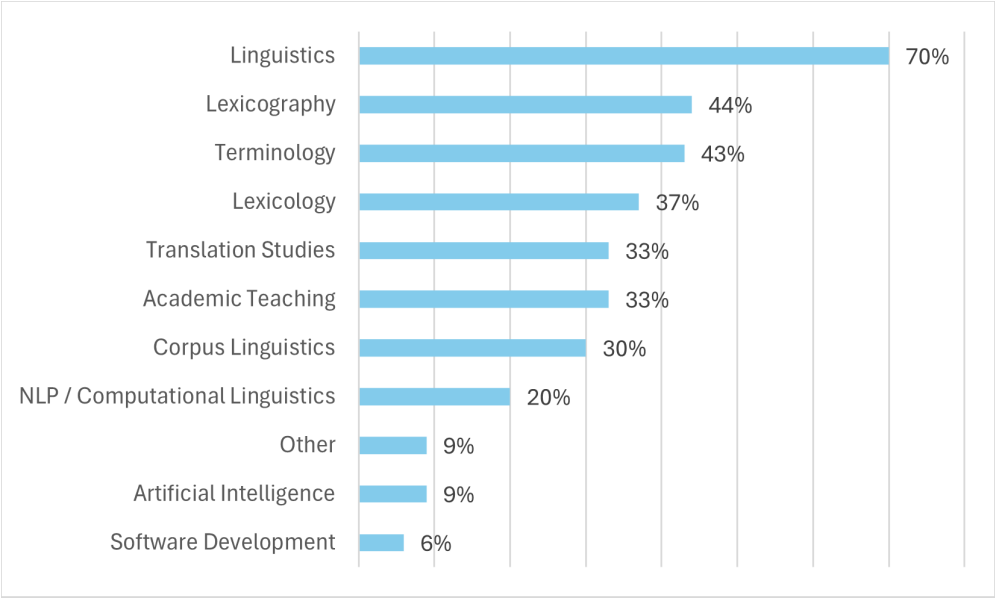

Figure 1: Disciplinary expertise of survey respondents (n=243)

Engagement in neology-related activities was reported by 244 respondents, with the possibility of selecting multiple activities. These activities were diverse, with roles ranging from research and academic teaching to lexicographic and terminological work, as well as computational linguistics and translation (see Figure 2).

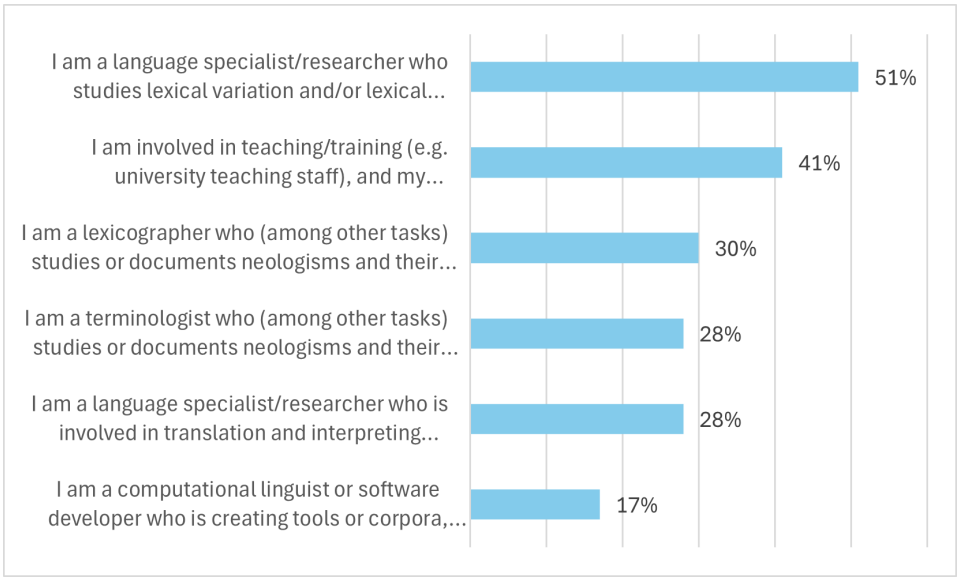

Figure 2. Engagement in neology-related activities (n=244), with most respondents active in research, teaching, and lexicographic work

For this paper, the focus is mainly on sections regarding responses of lexicographers and terminologists, whose majority is affiliated with European institutions. In addition, some lexicographers and terminologists were from other parts of the world, including

South America (Brazil, Argentina), the USA, the Middle East and Africa (e.g., Turkey, Algeria, Senegal), and Asia (e.g., Korea). Most respondents work within the educational, academic, public, or governmental sectors.

Although lexicography and terminology differ in methodology and scope, their contrastive analysis is regarded as complementary and enriching for the study of methods, practices, and tools for lexical innovation.

3.2 Lexicographic and Terminology Resources

Lexicographers participating in the survey reported a total of 27 distinct dictionary projects that address lexical innovations, including, e.g., Neologismenwörterbuch[4] (Klosa-Kückelhaus et al., 2023), Woordenboek van Nieuwe Woorden[5] (Waszink, 2020), The Danish Dictionary[6] (Trap-Jensen, 2016; Nimb et al., 2020), Dicionário da Língua Portuguesa[7] (Salgado et al., 2023), and EKI Combined Dictionary[8] (Koppel et al., 2019). Terminologists reported 19 distinct projects that address lexical innovations, including, e.g., bistro - Information System for Legal Terminology[9] (Kranebitter & Ralli, 2022), TriMED - Medical Termbase[10] (Vezzani et al., 2018), and NeoTriTerm-M[11] (François, 2024). Close to half (42%) of the lexicographic resources described possess an official status within their respective countries and are primarily general-purpose dictionaries or lexicons. Only 33% of the reported resources are specifically dedicated to neologisms. Among these, the vast majority (92%) address neologisms in general, while only a single respondent indicated that their resource targets a specific domain or subject area.

3.2.1 Practices in the Compilation of Lexicographic and Terminological Resources

Respondents were asked how they compile their resources, with five options provided: semi-automatically with manual post-editing, manually, automatic data extraction with manual post-editing, automatic data extraction without manual post-editing, and other. Among the 44 lexicographers who responded, two thirds (66%) reported using some form of automation in their workflow. Of these, 47% indicated the use of semi-automatic methods followed by manual post-editing, making it the most common approach. 19% relied on fully automated data extraction, also followed by manual post-

[4] https://www.owid.de/docs/neo/start.jsp (7 October 2025).

[5] https://neologismen.ivdnt.org/search (7 October 2025).

[6] https://ordnet.dk/ddo (7 October 2025).

[7] https://dicionario.acad-ciencias.pt (7 October 2025).

[8] https://sonaveeb.ee/?uilang=en (7 October 2025).

[9] https://bistro.eurac.edu/ (7 October 2025).

[10] https://shiny.dei.unipd.it/TriMED/index.html (7 October 2025).

[11] Not available yet online.

editing. In contrast, among terminologists (n=20), only 50% reported using automated methods: 30% reported using semi-automatic compilation followed by manual post-editing, while 20% utilized fully automated data extraction, also followed by manual post-editing. Despite these differences, both groups highlighted complex workflows that combine several tools and procedures. For instance, one terminologist described using Sketch Engine to compile a corpus of medical texts, automatically extract candidate terms, manually identify relevant terms, and compile entries using the FAIRterm web application (Di Nunzio & Vezzani, 2025; Vezzani, 2021) before transferring validated entries into the TriMED system. Another terminologist referenced their local language web-based portal, which comprises over 70 domain-specific terminology collections, etc. The analysis revealed slight differences between lexicographers and terminologists in their approaches to compiling lexical resources, particularly in terms of the extent to which automated methods are employed. These responses illustrate that while lexicographers adopt automation more broadly, terminologists tend to favor tailored, domain-specific solutions that integrate expert validation and editorial oversight. Notably, no respondents reported using fully automatic extraction without post-editing.

Regarding team composition, 36 lexicographers provided information about the contributors involved in their projects. In addition to the involvement of lexicographers as core contributors, which was reported by all respondents (100%), 69.4% indicated participation by computational linguists, and 63% listed other contributors such as software developers, computational advisors, consulting editors, UX designers, technical advisors, proofreaders, publishers, language experts, etymological consultants, and members of advisory or editorial boards. Some projects also involve students, interns, and interdisciplinary collaborators, such as members of the Association for Computational Linguistics (ACL) or specialists in neology. In comparison, from terminologists (n=18) who reported on team composition and their roles within terminology projects, 50% identified their main role as terminologists, while 39% served as project leaders, and a smaller share (11%) worked as editors. Several terminologists reported holding multiple or hybrid roles, such as terminologist and translator, or as project leader and terminologist, or as language expert and project leader. Compared to lexicographic teams, which show a high degree of technical and interdisciplinary integration, terminological projects appear to involve smaller teams with more clearly defined and often overlapping roles.

Regarding the data formats used for lexicographic projects (n=21), relational databases and XML structures were the most frequently used (49%). Less structured formats, such as word-processor files or table-based formats (e.g., CSV, TSV, XLS), were mentioned approximately half as often. No specific information about data formats is available for terminologists, as this question was not included in their section of the survey.

3.2.2 Data Sources for Identifying and Researching Lexical Innovations and Terms

The survey indicates that corpora serve as the primary data source for neology tracking among both lexicographers and terminologists. Among lexicographers (n=41), who responded to the question whether they use corpora, 88% reported using them, while only 12% do not. The most commonly used are Web corpora (78%), followed by dynamic corpora (67%), i.e., corpora that capture ongoing language usage and are continuously updated to reflect changes over time. Static corpora, i.e., corpora representing language at a specific point in time, are used by 53%. However, spoken corpora are used less frequently (14%), indicating that their potential as a source of neology remains largely unexploited in current lexicographic practice. Terminologists, in contrast, rely heavily on structured, domain-specific corpora tailored to their area of specialization. The source texts they currently use or plan to use reveal a wide array of both general and discipline-focused resources. These include e.g., stratified bilingual corpora in engineering (e.g., RAL Leitfaden, UNI 11673), astronomy corpora such as DLR and TAFOC, legal corpora stratified by national legal systems (Italy, Austria, Germany, Switzerland, EU), as well as corpora in translation and interpreting, materials engineering, and cybersecurity (e.g., the English-Lithuanian parallel and comparable corpora). Additionally, projects use national resources (e.g., Estonian National Corpus 2023), diachronic corpora (e.g., the Bulgarian Diachronic Mathematical Resource), and specialized terminology databases (e.g., 6,500 concepts and 32,000 terms in materials engineering).

The survey results indicate that public engagement also plays a notable role in neology tracking. Of the 40 lexicographers, 63% reported relying on public suggestions—typically submitted via a form on a dictionary website—for identifying new words or senses, though opinions on the value of these contributions vary: over half (52%) of lexicographers rated public suggestions as "maybe" important to their projects, while 36% considered them important or very important. These findings suggested that while public contributions are commonly welcomed, they are often viewed as supplementary to lexicographic work. Lexicographers gather new words through various channels, including dictionary websites, social media platforms such as Instagram, Facebook, Twitter, and TikTok, social media corpora, log files, and, in one case, a Pybossa[12] installation was also mentioned as a tool for input. Terminologists make significant use of social media corpora (e.g., Twitter, Instagram), as well as press corpora (e.g., Le Monde, Libération, Le Figaro), literary texts, academic texts, and oral resources, e.g., transcripts from the XAIDA Summer School. These diverse corpora are often managed through specialized software like Sketch Engine and complemented by terminology banks (e.g., TriMED, Icelandic TermPortal), demonstrating a hybrid methodology that combines automatic extraction with manual validation by domain experts.

[12] https://pybossa.com/ (7 October 2025).

Overall, while lexicographers emphasize real-time language tracking and public engagement, terminologists apply more domain-specific and systematically curated approaches. Lexicographic methods favor broad linguistic surveillance across media landscapes, whereas terminological work hinges on targeted corpora and structured validation to support precision in specialized fields.

3.2.3 Software and Tools for Neology Tracking and Management

A dedicated section of the survey focused on software and tools for neology tracking and management. Respondents were asked whether they use specific tools for identifying and managing neologisms and new word senses, what the tools are used for, what types of data they retrieve, and what additional functionalities they would like to see implemented.

The survey results showed that the use of specialized tools – such as corpus query systems, neology tracking tools, or in-house software for tracking and managing neologisms and new word senses – is not yet widespread. Of the 42 lexicographers, 55% reported not using any specific software or tools, while 45% indicated that they do. Terminologists (n=19) reported even lower use of tools, with only 26% indicating that they use any software for neology tracking and management. This suggests that, despite the availability of automated tools, a significant number of respondents across both communities still rely on more traditional or manual methods for tracking lexical innovation, and the overall adoption of neology tracking tools remains limited.

Altogether, around twenty tools were mentioned in the survey. The most commonly cited corpus query tool among them was Sketch Engine[13] (Kilgarriff et al., 2004), a finding that aligns with previous surveys on lexicographic practices conducted within the Horizon 2020 project ELEXIS[14] (Tiberius et al., 2024). Other tools referenced by respondents included AntConc[15] (Anthony, 2013), KORP[16] (Borin et al., 2012), DURel[17] (Schlechtweg et al., 2023), Neoveille (Cartier, 2017), NeoCrawler (Kerremans et al., 2012), Word2Dic[18] (Sørensen & Nimb, 2018), COSMAS II[19], and CoREST[20]. Terminologists reported using only Sketch Engine and Neoveille in their current projects.

Tools were widely used for creating corpora and supporting the study of regional

[13] https://www.sketchengine.eu (7 October 2025).

[14] https://elex.is (7 October 2025).

[15] https://www.laurenceanthony.net/software/antconc (7 October 2025).

[16] https://spraakbanken.gu.se/korp/ (7 October 2025).

[17] https://durel.ims.uni-stuttgart.de (7 October 2025).

[18] https://korpus.dsl.dk/resources/details/word2vec.html#english (7 October 2025).

[19] https://cosmas2.ids-mannheim.de/cosmas2-web (7 October 2025).

[20] https://korpus.dsl.dk/corest (7 October 2025).

language varieties (e.g., German vs. Swiss or Austrian German), as well as for automatic data retrieval (see Figure 3).

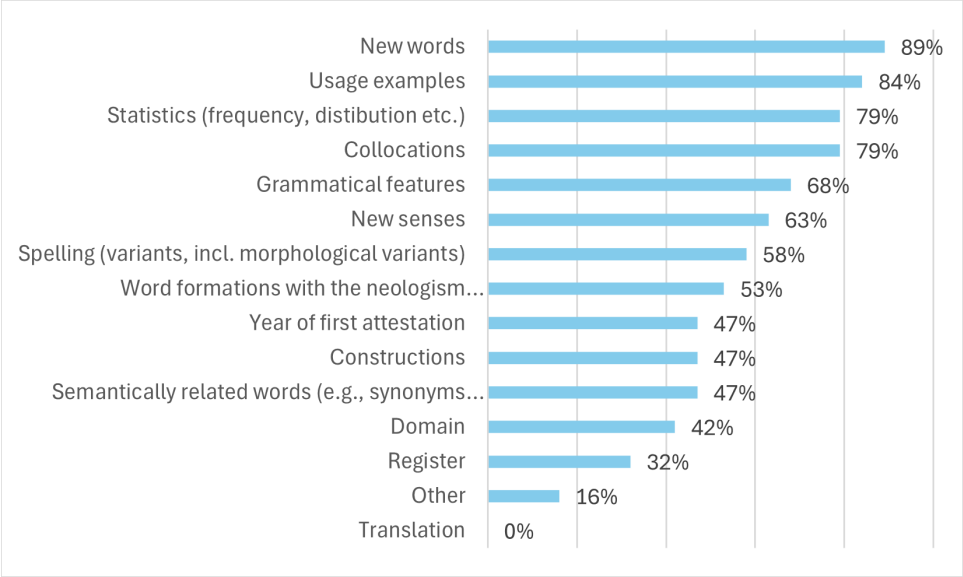

Figure 3: Types of lexicographic data retrieved with tools (n=19)

The most retrieved lexicographic data (n=19) include new words, usage examples, statistics, and collocations. Less frequently retrieved types of data include semantic relations, register and domain information.

Terminologists (n=4) reported retrieving term candidates, translation equivalents, statistics (e.g., frequency, termhood, unithood) and contexts, but also monitoring usage trends and regional variations of neologisms. The specific methods used to extract each type of information were not explored.

Twelve lexicographers responded to the question about the functionalities they would like to have in their tools. They mentioned the need for better visualization of trending words and collocates, enhanced GDEX[21] functionality for selecting examples (Kilgarriff et al. 2008), advanced semantic analysis, a more intuitive user interface, automated reporting features, and access to external databases. One comment included the wish "to have a specially developed tool for detecting neologisms in Sketch Engine". Some lexicographers have also expressed a wish for better integration with Large Language Models (LLMs), such as AI-driven detection of new meanings and expressions. Terminologists expressed a wish to add functionalities that, while not directly related to terminology, could significantly support terminology research.

One suggestion was to integrate named entity recognition (NER) and topic modelling tools into platforms like Sketch Engine: "NER would help identify and classify entities such as organisations, locations, and product names, which are important for analysing domain-specific vocabulary. Topic modelling would allow for the categorisation of texts

[21] https://www.sketchengine.eu/documentation/manual-for-gdex (7 October 2025).

into topics. This would help identify the main subjects around which terms are clustered, providing insights into the conceptual structure of domains and assisting in contextualising terms within their thematic frameworks".

Seven terminologists responded to the question about problems that could potentially be addressed with appropriate technical support. Their answers revealed a range of challenges, including the lack of resources in certain languages (e.g., Arabic), which limits access to term extraction technologies. Other issues included the representation of terminological gaps in knowledge structures, inefficient multi-word term extraction, the extraction of hapax legomena, which necessitated continued manual work to cover new terms in domains that had already been explored from a terminological perspective, and the need for support in tracking and discovering new meanings and new terms together with the necessary information, such as definitions, contexts, subject fields, sources. Additional concerns involved the lack of IT experts who have the time and resources to cooperate with terminologists, restrictions on access to web data, and an interest in applying Linked Linguistic Open Data (LLOD) technologies to link the data with other resources, enhancing interoperability, improving data accessibility, and enabling dynamic connections to external datasets and ontologies.

At the time the survey was planned, the use of LLMs in lexicography and terminology started emerging. Since then, developments in this area have evolved rapidly, such that the use of LLMs now merits a dedicated survey of its own. Among those respondents answering about LLM use (n=38), only 16% indicated that using them for identifying and tracking neologisms. ChatGPT was the most frequently mentioned model, followed by ClaudeAI. The respondents mentioned using LLMs for suggestions on word meanings, definitions, and new senses of words. One project (Sköldberg et al., 2024) reported using the DURel software with XL-LEXEME model, which was optimized for Lexical Semantic Change Detection (Cassotti et al., 2023). Terminologists did not indicate the use of LLMs in their everyday practice.

The results of the survey indicated that the potential of corpora for detecting new words and senses had not yet been fully realized, and existing tools do not entirely meet the needs of lexicographers, suggesting there is clear room for improvement and future development.

3.2.4 Practices in Lexicographic and Terminological Description

This section reports on questions related to the selection, description, and follow-up of neologisms in lexicographic and terminological projects, and provides an overview of the types of language awareness and policy activities that projects are involved in. The questions for lexicographers and terminologists were prepared by different teams and, unlike in the previous section, are not fully parallel. More specifically, the questions for

lexicographers benefited from the work being done within ENEOLI task 2.5[22] on neology from a lexicographic perspective. At the time of survey preparation, there was not yet a similarly dedicated task group on terminology and neology. This section, therefore, focuses on the more elaborate questions posed to lexicographers and only contrasts these with the more limited information about terminological projects where possible.

3.2.4.1 Selection of Lexical Innovations

First, lexicographers were asked about the systematicity with which they document neologisms. Of the respondents (n=41), 39% reported documenting neologisms systemically on a permanent basis, 17% on a project basis, and 32% only added neologisms occasionally. Of those responding, 53% reported having explicit policies or guidelines for lexical innovation, while 47% did not.

Next, lexicographers were asked about the selection criteria they use for inclusion into the database. They were offered a multiple-choice list of inclusion and exclusion criteria that were mainly drawn from the extensive work by Freixa (2022a, 2022b) for Catalan (see Figure 4). Commonly used inclusion criteria for lexicographic projects include minimum frequency thresholds in a project's main monitor corpus (59%), attestation in multiple publication types (61%), an anticipated need among language users for guidance on spelling or potential confusion (34%) and regional diversity (27%). A majority also enlist language specialists for additional validation (56%). Additional inclusion factors were user suggestions (37%), completion of thematic, semantic, or derivative series (27–39%), the neologism being already included in other lexical resources (29%), like an official spelling dictionary or general language dictionary, or occurring in other textual sources than the project's main monitor corpus (27%), e.g., Google searches or other corpora.

Lexicographers' responses (n=34) on exclusion criteria show that 62% routinely exclude proper nouns, 50% filter out grammatically incorrect formations, and around one-third reject orthographic variants, metalinguistically marked forms, or LLM-generated forms. Only 21% answered that they followed the official language policy advising against the use of specific types of neologisms, such as anglicisms and gallicisms.

When asked how long it takes between the selection for inclusion of a neologism and its actual publication in a dictionary, the most common answer among lexicographers (n=40) was 1 year (33%), but time lags vary dramatically from 1 day (18%) up to "many years" (n=1).

In their free-text answers, terminologists (n=18) reported few structured selection policies. Although 65% terminologists indicate the new term usage status in their database, the majority has no specific inclusion criteria for new terms but relies instead

[22] https://eneoli.eu/working-groups (7 October 2025).

on detecting the need to fill a terminological gap. One-third (33%) have explicit workflows for identifying new terms and, in that case, they emphasize inclusion criteria like domain needs, conceptual gaps, translation needs, or user feedback from translators and experts. As can be expected from terminologists, whose work on defining specialized knowledge often includes collaborating with specialists, 95% terminologists include domain experts in the validation process of new terms.

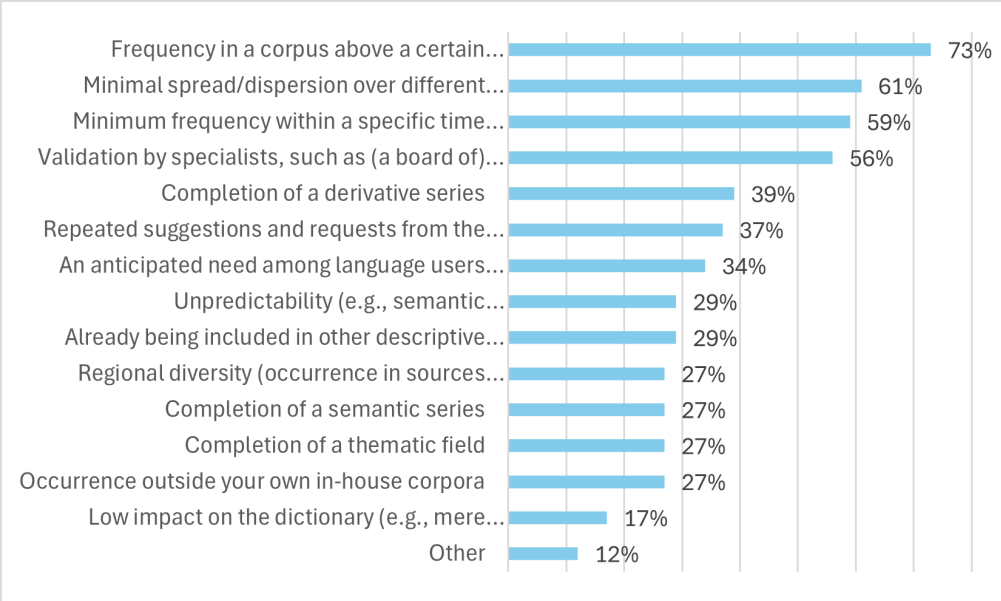

Figure 4: Inclusion criteria for lexical innovations in lexicographic projects (n=41), showing frequency and attestation as the most widely applied criteria

3.2.4.2 Description of Lexical Innovations in the Database

Lexicographers were offered a multiple-choice list of possible microstructure components and asked to indicate which of those their dictionary projects use in the description of neologisms. The list was mainly based on the ELEXIS surveys on general lexicography practices (Tiberius et al., 2024), supplemented with specific microstructure components for neologisms that are used in dedicated neologism dictionaries, like the Woordenboek van Nieuwe Woorden for Dutch[23] (Waszink, 2020) or the IDS-Neo2020+ for German (Storjohann, 2024). In general, lexicographers (n=40) use the same standard microstructure components to describe neologisms as they do for other entries (see Figure 5): lemma form, parts of speech, usage examples, meaning descriptions, and spelling variants. Other common elements include additional grammatical features, word formation type, collocations, semantic relations, origin/etymology, and usage labels, such as register, style, or domain. Specifically, neology-related microstructure components were relatively less used: just over half of projects record the date of first attestation, but more elaborate explanations on the socio-cultural background of a neologism's origins are rare, as are explicit neologism labelling, advice on use or frequency and diffusion labels.

[23] https://ivdnt.org/woordenboeken/nieuwe-woorden/ (7 October 2025).

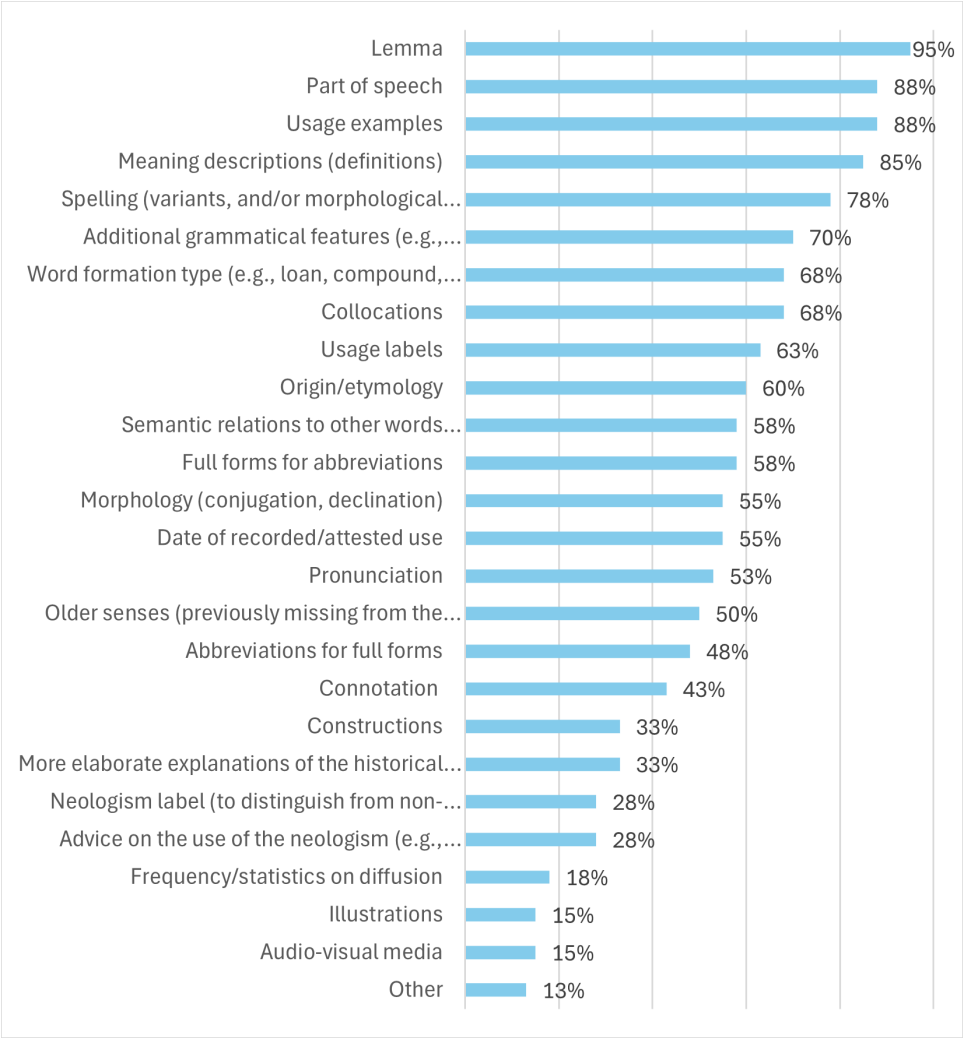

Figure 5. Microstructure components in neologism descriptions (n=40), with lemma form, POS, usage examples, and meaning descriptions used most consistently

In compiling entries, a large majority (32 of 40) follow standard templates, with about half (48%) adapting templates by lemma type, mostly part-of-speech. As for lay-out considerations, most lexicographic projects (36 of 40) include corpus examples, and present them either as a block (56%) or spread throughout entries (50%). Direct linking from dictionaries to corpora is still relatively uncommon (23%).

Terminologists (n=17) generally indicate usage status (65%) in terms of preferred or deprecated, with some additional anecdotal reporting of frequency labels, contextual notes, or acceptability ratings. Terminologists also often (n=14) include term variants that are identified using corpus searches, manual synonym checks, expert validation and keyword tools.

3.2.4.3 Follow-up of Lexical Innovations after Inclusion in the Database

Among lexicographers (n=40), 30% report having special policies for removing or marking obsolete neologisms. These include having information about the circulation of the words (old, out-of-use, etc.), putting older words into distinct categories and

removing words that are no longer used from the dictionary, as well as explicitly deciding on maintaining words in the dictionary that have become part of the standard language. 30% say they update neologisms differently than other entries, mainly adding a date and or explicit labeling as a neologism. Finally, 28% collect user feedback on neologism entries, for example by using social media to collect feedback from users, using e-mail or a separate feedback form.

Terminologists (n=15) report that they maintain accuracy through regular updates, periodic checks, domain expert reviews, corpus analyses, and user feedback loops. This continuous review is intended to ensure that definitions remain accurate and relevant over time.

3.2.4.4 Language Promotion Activities

Lexicographers (n=36) actively engage in language promotion and awareness activities: 35% report creating new words for lexical gaps or as alternatives for borrowings, 20% promote indigenous words as alternatives, and others mention awareness-raising activities such as publications, talks, and engagement with the media.

Among terminologists (n=19), 79% actively participate in creating new terms, often collaborating with national bodies, ministries, or expert councils. This work is reported as being intended to help maintain lexical vitality and align with language planning objectives, especially for specialized or minority language contexts.

3.2.5 Policies, Recommendations and Guidelines Related to Lexical Innovation

Lexicographers and terminologists approach lexical innovation through notably different policy frameworks, reflecting the distinct nature of their disciplines. In lexicographic projects, explicit policies related to lexical innovation are relatively common. Among 36 respondents, 53% reported having formal or semi-formal guidelines in place. These range from highly structured editorial rules–such as requiring the documented presence of a neologism in at least two usage registers or its sustained use over a period of three years–to more flexible, internal criteria focusing on frequency, dissemination, durability, and relevance to general language. Several projects prioritize the inclusion of well-established neologisms found in newspaper texts or across multiple digital sources. Others emphasize interdisciplinary collaboration, ethical considerations in language innovation, and user-centered validation processes involving corpus analysis and expert review. Some teams have institutionalized their principles through published documents or online presentations, while others rely on internal style guides and entry development instructions.

Terminological projects, in contrast, report less frequent adoption of explicit innovation policies–only 44% of 18 respondents indicated such guidelines exist within their projects. Nevertheless, among those who do follow policies, the standards are often

closely tied to formalized national or institutional frameworks. For example, terminologists cited the use of ISO standards (e.g., ISO 704 on terminology work), national guidelines for terminology work (e.g., the Institute of the Estonian Language), and well-established references like Guide to Terminology (Suonuuti, 1997). In some projects, internal commissions (e.g., the South Tyrolean Terminology Commission) or digital tools (e.g., Ekilex[24] (Tavast et al., 2018)) structure the creation and validation of new terms. Methodologically, terminological policies often emphasize semantic processes like derivation and semantic extension over borrowing and prioritize adherence to standardized terminology creation practices rather than spontaneous lexical innovation.

In sum, while both communities recognize the need to manage lexical innovation, lexicographers typically frame it as part of a broader language documentation effort, often shaped by corpus evidence and public engagement. Terminologists, on the other hand, adopt a more prescriptive and structured approach aligned with formal standards and domain-specific precision. This divergence reflects their differing goals: capturing evolving general language vs. ensuring consistency and clarity in specialized domains.

3.3 Needs and Training in Neology Across Profiles

Both lexicographers and terminologists demonstrate a clear interest in improving their competencies in neology, although their training backgrounds and expressed needs reveal some nuanced differences. A common trend among both groups is the predominance of informal or self-directed learning. Among lexicographers, 69% reported acquiring their knowledge through learning-by-doing, with only 28% having taken university courses and 18% having participated in Continuing Professional Development (CPD) seminars. Similarly, 84% of terminologists reported self-directed learning, with slightly lower figures for formal education (21% for university courses, 26% for CPD workshops). Both groups included respondents who had doctoral-level qualifications, but few had received systematic instruction specifically focused on neology.

In terms of training needs, lexicographers emphasized the importance of access to tools and corpora – 50% rated this as "absolutely necessary" – along with training in editorial policies, semantics, and word-formation. Terminologists, on the other hand, pointed to the need for a more structured framework: suggestions included university-level courses, dedicated training schools, and clear distinctions between terminology and neology. Some respondents also called for better understanding of sociolinguistic dynamics, particularly for those working in minority languages, and highlighted the importance of ethical and inclusive terminology creation.

Suggestions for improving training among lexicographers focused on hands-on, tool-

[24] https://ekilex.ee (7 October 2025).

oriented approaches. These included greater familiarity with corpus tools (e.g., Sketch Engine, Neoveille), disambiguation techniques, the selection of neologisms for dictionary inclusion, and collaboration with domain experts. Respondents also proposed interdisciplinary training that bridges lexicography with fields such as sociology, psychology, and computational linguistics. Terminologists similarly expressed a need for practical, digitally supported training – such as programming courses covering Python, NLP, and data analysis – to reduce dependency on external technical tools and increase autonomy in term extraction and validation.

Both groups emphasized the value of professional development and networking opportunities. Lexicographers suggested more conferences and panel discussions, while terminologists stressed the importance of community-building, follow-up webinars, and sharing best practices. In sum, while the two profiles differ slightly in their methodological emphasis–terminologists leaning more toward systematized approaches and lexicographers toward dynamic, media-driven methods–both converge on the need for accessible, practical, and collaborative training opportunities to better support their work in the evolving field of neology.

4. Conclusion

This paper presents findings from the ENEOLI survey, conducted in October–November 2024, focusing on the methods, practices, tools, and methodologies adopted by lexicographers and terminologists in researching and documenting lexical innovation. These largely reflect approach differences in defining knowledge and language in lexicography and terminology. The findings highlight variations in workflows, data resources and design, and the integration of digital tools, as well as shared challenges related to training and the evolving role of public engagement in promoting neologisms.

The findings suggest several directions for future research. First, further systematic investigation is needed into how the varying degrees of automation and corpus usage influence the quality and sustainability of neologism tracking in both fields. Second, the contrasting ways in which lexicographers and terminologists handle community input and public engagement could be further explored, especially in relation to how compilation practices are communicated. Third, the potential role of emerging technologies, including LLMs, warrants closer examination, particularly given the expressed needs for more intuitive tools and domain-specific corpora. Moreover, future work should also address how data is structured, organized, and linked in both lexicographic and terminology resources, with particular attention to questions of interoperability, data management, and cross-disciplinary data exchange.

Finally, a deeper comparative study could help clarify how the two disciplines might better learn from each other's strengths. For example, combining the broader, real-time surveillance typical of lexicography with the rigorous domain-specific validation of terminology work can lead to a better understanding of how specialized terms become

general language neologisms in a process of determinologization. Such insights could inform more targeted training pathways, collaborative resource development, and policy recommendations to support the dynamic field of neology in both general and specialized contexts.

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT-4 for grammar and spelling checks. The authors have subsequently reviewed and edited the content and take full responsibility for the publication's final version.

5. Acknowledgments

This paper is based on work from COST Action ENEOLI, CA22126, supported by the European Cooperation in Science and Technology (COST), https://www.cost.eu/actions/CA22126/. This work was financially supported by the Estonian Research Council grant (PRG 1978), by the EKKD-III1 project *"Applying Large Language Models to Lexicography: New Opportunities and Challenges"*, and by the Slovenian Research and Innovation Agency (ARIS) through the core programme *Language Resources and Technologies for Slovene* (P6-0411). We would like to thank Judit Freixa, Petar Božović, and Hiwa Asadpour for their valuable assistance in compilation and conducting the survey, and Andreína Adelstein, Anna Anastassiadis-Symeonidis, Hiwa Asadpour, Patrizio Bellan, Silvia Bonavero, Milica Božić Sinčuk, Petar Božović, Esther Odilia Breuer, Vesna Bulatovic, Marc Van Campenhoudt, Pierluigi Cassotti, Anila Çepani, Dragana Čarapić, Mauro Le Donne, Germana Foscale, Océane Foubert, Marius François, Ana Frankenberg, Greta Franzini, Judit Freixa, Rosa Maria Frejaville, Radovan Garabík, Marta Garcia-Casado, Anca Gata, Voula Giouli, Sanel Hadziahmetovic Jurida, Linnea Hanell, Andrew Hawke, Margrethe Heidemann Andersen, Tarja Riitta Heinonen, Louise Holmer, Michaela Hrotekova, John Humbley, Riaz Hussain, Balša Ivanović, Susanna Ivaldi, Burcu Koc, Iztok Kosem, Nuran Kır, Ganna Krapivnyk, Jelena Krkljes, Irena Krumes, Priit Kulu, Katarzyna Kłosińska, Ivana Lazić Konjik, Paul Lerner, Veronika Lipp, Michaela Lišková, Henrik Lorentzen, Sarah MacDougall, Ajay Malik, Tinatin Margalitadze, Bruo Maroneze, Christine Michaux, Nikolche Mickoski, Lina Miloshevska, Anela Mulahmetović Ibrišimović, Rahela Nisavic, Eugenia Esperanza Núñez Nogueroles, Atul Ojha, Radmila Otovic, Christos Papanagiotou, Elbaz Pascale, Elizabeth Peterson, Aurélie Picton, Sigita Rackevičienė, Natascia Ralli, Marianne Rathje, Marie-Alice Rebours, Rossella Resi, Nuria Rodriguez Pedreira, Michael Rosner, Benjamin Rouxel, Xhevdet Rusinovci, Jan Oliver Rüdiger, Signe Saar, Marta Sabater, Lina Sader Feghali, Dominik Schlechtweg, Ineke Schuurman, Taner Sezer, Manuel Singala, Eva Skafte Jensen, Emma Sköldberg, Réka Sólyom, Lena Ruth Stefanovic, Petra Storjohann, Velislava Stoykova, Giovanni Tallarico, Elena Isabelle Tamba, Lars Trap-Jensen, Giedre Valunaite Oeskeviciene, Mari Vaus, Andrejs Veisbergs, Corina Veleanu, Federica Vezzani, Andrea Zaninello,

Ágústa Þorbergsdóttir for contributing to the survey.

6. References

Anthony, L. (2013). Developing AntConc for a new generation of corpus linguists. *Proceedings of the Corpus Linguistics Conference (CL 2013)*, pp. 14–16. Available at: https://laurenceanthony.net/research/20130722_26_cl_2013/cl_2013_paper_final.pdf.

Borin, L., Forsberg, M., & Roxendal, J. (2012). Korp – the corpus infrastructure of Språkbanken. *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, *Istanbul, Turkey*. European Language Resources Association (ELRA), pp. 474–478.

Cartier, E. (2017). Neoveille, a Web Platform for Neologism Tracking. *Proceedings of the Software Demonstrations of the 15th Conference of the European Chapter of the Association for Computational Linguistics, Valencia, Spain*. Association for Computational Linguistics, pp. 95–98.

Cassotti, P., Siciliani, L., de Gemmis, M., Semeraro, G., & Basile, P. (2023). Xl-lexeme: Wic pretrained model for cross-lingual lexical semantic change. *Proceedings of the 61th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Association for Computational Linguistics, pp. 1577–1583.

Di Nunzio, G. M., & Vezzani, F. (2025). FAIRterm 2.0: Towards FAIR terminologies resources for EOSC. *IEEE International Conference on Cyber Humanities (IEEE-CH), IEEE Explore*. IEEE.

François, M. (2024). Exploitation d'un corpus oral en anglais dans le domaine de la modélisation climatique: contribution à la description d'une néologie «première». *Approches épistémologiques de la terminologie - Enjeux actuels*, Réseau LTT - Lexicologie, Terminologie, Traduction, Oct 2024, Paris, France.

Freixa, J. (2022a). The Dictionarisation of Neologisms: The NADIC as a Model. In J. Freixa, M. I. Guardiola, J. Martines, M. A. Montané (Eds.), *Dictionarization of Catalan Neologisms*, Peter Lang, pp. 15–37.

Freixa, J. (2022b). 'Garbell: L'avaluador Automàtic de Neologismes Catalans (Garbell: The Automatic Catalans Neologisms Analyzer)'. *Terminàlia*, 26, pp. 7–16.

Kerremans, D., Stegmayr, S., & Schmid, H.-J. (2012). The NeoCrawler: Identifying and retrieving neologisms from the Internet and monitoring ongoing change. In K. Allan & J. A. Robinson (Eds.), *Current methods in historical semantics*. De Gruyter Mouton, pp. 59–96.

Kilgarriff, A., Rychlý, P., Smrž, P., & Tugwell, D. (2004). The Sketch Engine. In G. Williams & S. Vessier (Eds.) *Proceedings of the 11th EURALEX International Congress*. Université de Bretagne-Sud, pp. 105–115.

Kilgarriff, A., Husák, M., McAdam, K., Rundell, M., and Rychlý, P. (2008). GDEX: Automatically finding good dictionary examples in a corpus. *Proceedings of the 13th EURALEX International Congress. Spain, July 2008*, pp. 425–432.

Klosa-Kückelhaus, A., & Rüdiger, J. O. (2023). Introducing NeoRate. *Lexicography*, 10(2), pp. 117–137.

Koppel, K., Tavast, A., Langemets, M., & Kallas, J. (2019). Aggregating dictionaries into the language portal Sõnaveeb: issues with and without a solution. In I. Kosem, T. Zingano Kuhn, M. Correia, J. P. Ferreria, M. Jansen, I. Pereira, J. Kallas, M. Jakubíček, S. Krek & C. Tiberius (Eds.) *Proceedings of the eLex 2019 conference. 1-3 October 2019, Sintra, Portugal. Brno: Lexical Computing CZ, s.r.o.*, pp. 434−452.

Kranebitter, K., & Ralli, N. (2022). Quanto può influire l'utente nello sviluppo di uno strumento terminologico? L'esperienza di bistro. *Risorse e strumenti per l'elaborazione e la diffusione della terminologia in Italia. Eurac Research*, pp. 102−116.

Nimb, S., Sørensen, N. H., & Lorentzen, H. (2020). Updating the dictionary: Semantic change identification based on change in bigrams over time. *Slovenščina 2.0*: *empirične, aplikativne in interdisciplinarne raziskave*, 8(2), pp. 112–138.

Salgado, A., Simões, A., Iriarte, A., Vieira, R., Ferreira, M., Carmo, R., Pinheiro, C. (2023). Dicionário da Língua Portuguesa: a new lexicographic resource of Academia das Ciências de Lisboa. *Electronic lexicography in the 21st century (eLex 2023): Invisible Lexicography. Proceedings of the eLex 2023 conference. Brno, 27–29 June 2023. Brno: Lexical Computing CZ s.r.o.*, pp. 72-75.

Schlechtweg, D., Virk, S. M., Sander, P., Sköldberg, E., Linke, L. T., Zhang, T., Tahmasebi, N., Kuhn, J., & Walde, S. S. I. (2023). The DURel Annotation Tool: Human and Computational Measurement of Semantic Proximity, Sense Clusters and Semantic Change. *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations, St. Julians, Malta. Association for Computational Linguistics*, pp. 137–149.

Sköldberg, E., Virk, S. M., Sander, P., Hengchen, S., & Schlechtweg, D. (2024). Semantic Variation in Swedish Using Computational Models of Semantic Proximity – Results From Lexicographical Experiments. In K. Š. Despot, A. O. Anić, I. Brač (Eds.) *Proceedings of the XXI EURALEX International Congress: Lexicography and Semantics. Cavtat (Croatia): Institute for the Croatian Language*, 2024. pp. 169–182.

Sørensen, N. H., & Nimb, S. (2018). Word2Dict – Lemma Selection and Dictionary Editing Assisted by Word Embeddings. In J. Čibej, V. Gorjanc, I. Kosem, & S. Krek (Eds.) *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts. Ljubljana University Press, Faculty of Arts*, pp. 819–826.

Storjohann, P. (2024). IDS-Neo 2020+: A Novel Resource for New German Words in Use. *International Journal of Lexicography, 37(4),* pp. 389–403.

Suonuuti, H. (1997). Guide to terminology. Helsinki: Tekniikan Sanastokeskus.

Tavast, A., Langemets, M., Kallas, J., & Koppel, K. (2018, July). Unified data modelling for presenting lexical data: The case of EKILEX. In Ed. J. Čibej, V.

Gorjanc, I. Kosem & Simon Krek (Eds.) *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts, Ljubljana*, pp. 749–761.

Tiberius, C., Kallas, J., Koeva, S., & Langemets, M. (2024). A lexicographic practice map of Europe. *International Journal of Lexicography*, 37(1), pp. 1–28.

Trap-Jensen, L. (2016). Lexicography at the Society for Danish Language and Literature. *Kernerman Dictionary News*, 16.

Vezzani, F., Di Nunzio, G. M., & Henrot, G. (2018). Trimed: A multilingual terminological database. *LREC 2018-11th International Conference on Language Resources and Evaluation*, pp. 4367–4371.

Vezzani, F. (2021). La ressource FAIRterm: entre pratique pédagogique et professionnalisation en traduction spécialisée. *Synergies Italie*, 17, pp. 51–64.

Waszink, V. (2020). Neologisms in an online portal: The Dutch Neologismenwoordenboek (NW). *Dictionaries: Journal of the Dictionary Society of North America 41, no. 1 (2020)*, pp. 27–44.