Leveraging Domain Corpora for Enhanced Terminology: The Case of Estonian-English Remote Sensing Termbase

Liisi Jakobson, Jelena Kallas, Erko Jakobson
Tartu Observatory of the University of Tartu, Institute of the Estonian Language
liisi.jakobson@ut.ee, jelena.kallas@eki.ee, erko.jakobson@ut.ee

Abstract

This article addresses methodological issues related to developing domain corpora and a terminological database from scratch. We present an ongoing project focused on creating an Estonian-English Remote Sensing Termbase. First, we describe the compilation process of the Estonian Remote Sensing Corpus 2022[1], which served as the primary data source for the termbase. The corpus was compiled by crawling the web and adding files using the Corpus Query System Sketch Engine (Kilgarriff et al., 2004). In the next step, we employed the Term Extraction module (Kilgarriff et al., 2014; Fišer et al., 2016; Blahuš et al., 2023) to identify terms, which were subsequently registered in the Estonian Remote Sensing Termbase[2] using the Dictionary Writing System Ekilex (Tavast et al., 2018). For each term, we provided definitions, variants, and usage contexts. In the final stage, remote sensing experts reviewed and edited the terms, their variants, and usage contexts. Finally, we provide insights and outline directions for future work in this area.

Keywords: termbase, domain corpora, remote sensing, bilingual, Estonian

1. Introduction

Remote sensing is the process of detecting and monitoring the physical characteristics of an area by measuring its reflected and emitted radiation at a distance (typically from satellite or aircraft)[3]. It allows for the collection of valuable information, such as monitoring forest fires, floodings, heatwaves, snow, and various environmental phenomena. Over the past few decades, Estonia has seen a substantial increase in the adoption of remote sensing technology (Noorma et al., 2020a). With a wealth of high-quality data made freely accessible through the European Union's Copernicus program, the Estonian public sector has started to use this data to improve public services (Noorma et al., 2020b), and has embraced remote sensing data for the development of various products (e.g. Voormansik et al., 2020; Domnich et al., 2021; Khoshkhah et al., 2022; Komisarenko et al., 2022). What was once a topic primarily discussed in English and primarily by scientists and experts has now become a subject of widespread exploration among Estonian-speaking non-experts.

Despite the growing interest in this field, the remote sensing terminology for Estonian has not been comprehensively analysed, and standardisation is needed. However, it has been analysed for other languages, as evidenced by resources such as the English-Bulgarian dictionary of remote sensing terms (Kancheva, 2013), the multilingual Polish-English-German dictionary covering various subfields, including remote sensing (Kwiatek, 2017), and the Romanian-English glossary for terms related to the domains of remote sensing, geodesy, topography, and cadaster (Stan, 2023). Some of remote sensing terms have also been registered in the IATE database4 (Zorrilla-Agut & Fontenelle, 2019).

In Estonia, terminological work is curated by the state. Several documents regulate the development and maintenance of terminology: the national program 'Principles for Supporting the Management of Estonian Special Language and Terminology Work 2019–2027'[5] and the action program 'Estonian Terminology Work Action Plan for the Years 2023–2025'[6]. According to these programs, the Institute of the Estonian Language[7] annually organises contests for terminological commissions, providing financial support. These commissions typically comprised four to ten domain experts and one or two language specialists well-versed in terminology development (Erelt, 2007). Traditionally, termbases in Estonia were compiled solely through expert knowledge. Nowadays, however, the process of termbase creation has become more automated, encompassing both the creation and management of databases as well as the compilation and analysis of domain corpora.

For compiling general language dictionaries and termbases, the Institute of the Estonian Language has developed an in-house Dictionary Writing System named Ekilex (Tavast et al., 2018). The Lenoch classification system[8] is used to cover subject fields within Ekilex. For corpus data analysis, the Institute uses the Corpus Query Systems Sketch Engine (Kilgarriff et al., 2004), and KORP[9].

All termbases created using Ekilex are accessible to the public through the dictionary portal Sõnaveeb[10], with each termbase having its dedicated homepage, such as the Estonian-English Remote Sensing Termbase. Currently, there are 130 databases spanning diverse fields, and this collection continues to grow.

The paper reports on an ongoing project aimed at creating an Estonian-English Remote Sensing Termbase to standardise terminology and reduce ambiguity. The paper is organised as follows: Chapter 1 describes the process of compiling the Estonian Remote Sensing Corpus 2022 ('Kaugseire korpus 2022'), along with the term extraction process (Fišer et al., 2016; Blahuš et al., 2023), and the evaluation of

[1] The corpus is accessible at sketchengine.eu
[2] The termbase is accessible at sõnaveeb.ee
[3] https://www.usgs.gov/faqs/what-remote-sensing-and-what-it-used
[4] iate.europa.eu
[5] http://tinyurl.com/2tx3969f
[6] http://tinyurl.com/yc53jdv5
[7] https://portaal.eki.ee/
[8] https://www.uibk.ac.at/translation/termlogy/lenoch.html
[9] korp.keeleressursid.ee
[10] sõnaveeb.ee

LREC-COLING 2024, pages 10347–10351
20-25 May, 2024.

extracted term candidates. In Chapter 2, we describe the database structure and the process of the compilation of Remote Sensing Termbase. Finally, we provide insights and outline directions for future work in this area.

2. Estonian Remote Sensing Corpus 2022 compilation and Term Extraction using Sketch Engine

2.1 Corpus compilation from files and the web

For corpora creation from scratch, we used a Corpus Query System Sketch Engine (Kilgarriff et al., 2004), which supports approximately 100 languages (Kostka, 2022) and provides tools for users to create their corpus either by uploading their texts[11] or by building the corpus semi-automatically from the web[12] according to the keywords given by the user. Sketch Engine also supports the compilation of parallel corpora[13].

We compiled four corpora in the remote sensing field: the monolingual English Remote Sensing Corpus from the web, two parallel English-Estonian corpora from files, and the monolingual Estonian Remote Sensing Corpus 2022 from files and from the web. All corpora can be accessed through the Sketch Engine.

The monolingual English Remote Sensing Corpus was primarily crawled from the web using seed words. We managed to reach a size of 5 million words. This corpus was used as a primary source for definitions and examples of English terms.

Texts for two parallel English-Estonian corpora were mainly collected from EUR-Lex[14]. We managed to reach a size of 0,5 million words. Parallel corpora were employed to identify corresponding English and Estonian terms.

The monolingual Estonian Remote Sensing Corpus 2022 was compiled from both files and the web. Our initial goal was to reach a size of 5 million words in each corpus, as suggested in Kallas et al. (2017). However, despite including nearly all available texts in Estonian related to remote sensing, we managed to gather only approximately 3 million words. In total, 347 documents were uploaded, constituting 58% of the corpus size. Documents were sourced from the University of Tartu Library database, encompassing BA and MA theses, handbooks, projects, and study materials. To obtain permission to add files to corpora, scientists/researchers were directly requested to provide their materials.

Texts from the web were crawled by providing seed words such as here: 'kaugseire' (remote sensing), 'spektraalne lahutusvõime' (spectral resolution), 'spektraalmõõtmine' (spectral measurement), 'multispektraalne seade' (multispectral instrument), 'keskkonnasatelliit' (Earth observation satellite, environmental satellite), 'andmetele ja teabele juurdepääsu teenus' (DIAS), and 'ülelennu sagedus' (revisit time, revisit period). Most of the texts were from leading popular science journals, legal documents, materials related to remote sensing enterprises, the Estonian Agricultural Registers and Information Board, materials from the Estonian remote sensing community homepage, and materials from remote sensing scientific projects. Notably, there were no materials from social media sources such as user forums, blogs, comments, etc., as this topic has not been extensively covered in Estonian social media.

2.2. The Terminology Extraction Module and the compilation of the term list

To extract term candidates for the Estonian Remote Sensing Termbase, we primarily relied on the monolingual Estonian Remote Sensing Corpus 2022.

Sketch Engine incorporates a keyword and terminology extraction module designed to identify term candidates. This task involves two arbitrary corpora: a focus corpus from which keywords should be extracted and a reference corpus against which term candidates from the focus corpus are compared (Jakubíček & Šmerk, 2016). In our case, we utilised the Estonian Remote Sensing Corpus 2022 as the focus corpus and the Estonian National Corpus 2021 (Koppel & Kallas, 2022) as the reference corpus.

To enhance the accuracy of terminology extraction, Sketch Engine exclusively selects grammatically valid phrases. This two-step process, as described in (Fišer et al., 2016), consists of:

1. unithood: The first step is rule-based and language-dependent. It assesses the grammatical validity of a phrase (unithood) using the term grammar[15] . The term grammar defines grammatically plausible terms using regular expressions based on available annotations in the corpus, such as morphosyntactic tags and lemmas;
2. termhood: Candidate phrases generated in the first step are then compared with the reference corpus using the 'simplemath' statistic (Kilgarriff, 2001). This statistic contrasts their normalised frequencies, focusing on either less frequent or more frequent phrases.

The currently employed Estonian term grammar, version 2.0, is rooted in evidence-based term grammar design principles. It was developed by Marek Blahuš in close collaboration with Eleri Aedmaa and Merily Plado (Koppel & Kallas, 2022). The term grammar primarily relies on term patterns that utilise various combinations of the defined default attributes to identify and represent the extracted term candidates. The following parts of speech were considered as possible elements of term patterns: nouns, adjectives, indeclinable adjectives, adverbs, verbs, proper nouns, conjunctions, ordinal numbers, and acronyms. Version 2.0 of the Estonian term grammar enables term extraction for single and multi-word terms, with a maximum length of up to 5 words.

In total, the term grammar encompasses 37 distinct term patterns:
5-grams: 2 patterns
4-grams: 10 patterns
3-grams: 16 patterns
bigrams: 7 patterns
unigrams: 2 patterns

This approach enabled the identification of the most distinctive term candidates within the focus corpus, which in our case is the Estonian Remote Sensing Corpus 2022, compared to the reference corpus, represented by the Estonian National Corpus 2021 (Koppel & Kallas, 2022).

To employ the evaluation, extracted term candidates were organised according to their frequency within the

[11] Sketch Engine: create a corpus from files
[12] Sketch Engine: create a corpus from web
[13] Sketch Engine: parallel corpora
[14] https://eur-lex.europa.eu/homepage.html
[15] Sketch Engine: *term grammar*

focus corpus. The evaluation was performed on the 500 top-ranking single-word term candidates and the 500 top-ranking multi-word terms. A large majority of the multi-word term candidates were bigrams, with only a few 3- and 4-grams:

- bigrams: 445 (89%) term candidates;
- 3-grams: 43 (8.6%) term candidates;
- 4-grams: 12 (2.4%) term candidates.

All candidates were manually examined, and out of 1,000 terms, approximately 250 were identified as potential candidates for inclusion in the database (60% single-word terms and 40% multi-word terms).

The main problems we encountered during the manual examination included the following:

1) the appearance of English terms in the list, which was a result of the corpus having some parts of texts in English, such as BA and MA theses with English summaries;
2) terms from interconnected domains, primarily physics, biology, forestry, agriculture, metrology, meteorology, and climatology (e.g., 'fütoplankton' (phytoplankton), 'heljum' (suspension), and 'neeldumine' (absorption);
3) general language items (e.g. 'majandus' (economy), and 'võrdlus' (comparison);
4) mistakes in lemmatisation and morphological analysis.

Another task was the detection of variability in terminology. For example, we identified seven potential variants for the English term 'reflectance' (see Table 1).

Term in Estonian	**Raw Frequency**	**Frequency per million**
heleduskordaja	242	61
heleduskoefitsient	158	40
heledustegur	12	3
peegeldustegur	246	62
peegeldumistegur	69	17
peegelduskoefitsient	15	4
peegeldumiskoefitsient	29	7

Table 1: An example of how one English term, 'reflectance' has been translated into Estonian in seven different ways within the remote sensing field.

As the project is ongoing, we initially chose 150 term candidates for consultation with experts in the first stage. Ten remote sensing experts from UT Tartu Observatory participated by completing a questionnaire addressing terms with unclear meanings, variants, and usage inconsistencies. The questionnaire comprised two parts: first, experts evaluated terms identified through corpus analysis, and second, they had the opportunity to propose additional terms for consideration, especially for the concepts that lacked Estonian equivalents. For instance, the English term 'handheld lidar' was suggested to be translated as '*käsilidar*', and 'backpack lidar' as '*märsilidar*'. The results of this questionnaire were deliberated upon in a seminar held on June 2, 2022, at Tartu Observatory. The final list of 100 terms served as a headword list for compiling the Estonian Remote Sensing Termbase using the Dictionary Writing System Ekilex (Tavast et al., 2018). The work will continue.

3. The Compilation of Remote Sensing Termbase using Dictionary Writing System Ekilex

Ekilex (Tavast et al., 2018) is the in-house Dictionary Writing System developed by the Institute of the Estonian Language. Data is stored in Ekilex's PostgreSQL database and is accessible through an API[16]. Ekilex is hosted in the Estonian Scientific Computing Infrastructure (ETAIS) cloud[17]. A term entry in Ekilex is concept-based[18], structured as a unit containing term variants, definitions, contexts, source references, notes, and domain information. Figure 1 illustrates an entry for the concept *'artificially generated color image in which blue, green, and red colors are assigned to the wavelength regions to which they do not belong in nature*'.

Definitions in Estonian and English are presented in the upper box, along with an English explanation. It's worth noting that an entry can have multiple explanations in both languages.

The term variants with associated contexts are displayed in the lower boxes. Variants are categorised as preferred, admitted, former, deprecated, or new, following the classification outlined by Vaus (2022). Given that the Remote Sensing topic is relatively new, only a very limited number of terms have been marked as 'former' (e.g., the term 'peegeldustegur' (reflectance) is being replaced by the term 'heledustegur'). Currently, we do not have any deprecated terms.

Definitions were an area where expert assistance was particularly crucial. While some definitions were available in handbooks, many were developed based on English definitions. Following the principles outlined in Vezzani et al. (2018), our goal was to create comprehensible definitions for remote sensing experts, translators, and everyday end-users.

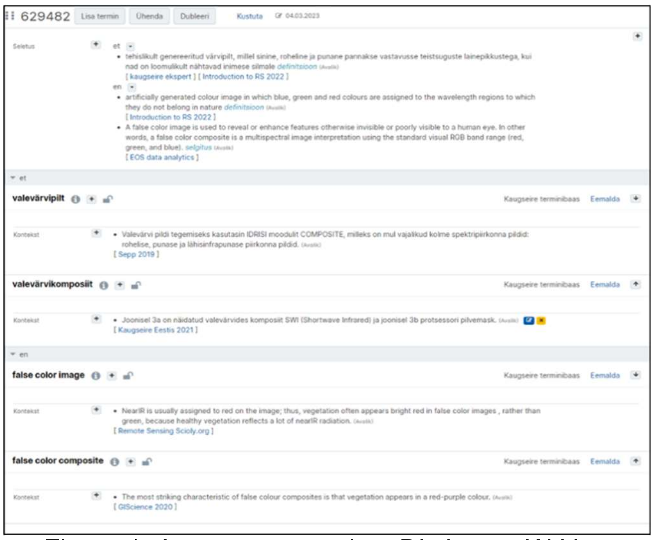

Figure 1: A concept entry in a Dictionary Writing System Ekilex.

[16] github.com/keeleinstituut/ekilex/wiki/Ekilex-API
[17] etais.ee/
[18] Ekilex guidelines

As a standard practice, definitions are supplemented with contexts. The guidelines for Ekilex state that context should clarify or complement the definition when necessary).To find appropriate contexts, we used the GDEX tool in Sketch Engine (Kilgarriff et al., 2008), which ranks corpus sentences according to predefined criteria, assigning a numerical score to each sentence. GDEX configurations have been developed for several languages (see, e.g. Kosem et al., 2019; for Estonian module Koppel et al., 2019). This mechanism can be seen as a filter, as it helps to find more relevant citations even though they have not been manually annotated. Most of the contexts in the Remote Sensing Termbase in Ekilex are from the Estonian Remote Sensing Corpus 2022.

4. Discussion

The practical advantages that corpora offered during the creation of the Estonian Remote Sensing Termbase include the ability to:

1) choose relevant terms based on frequency through term extraction;
2) detect possible variants of the same term;
3) analyse the usage context of variants, helping to understand in which domains the term is used. For instance, the English term 'reflectance' might be used as 'heleduskordaja' in the water remote sensing domain, while in vegetation remote sensing domain, the term '*peegeldustegur*' is preferred;
4) identify preferred variants based on statistical data;
5) distinguish old variants based on data from different time spans (Kilgarriff et al., 2015). By dividing the data into two periods: 1993-2010 and 2011-2022, we found that Estonian equivalent for English term 'reflectance' '*peegeldustegur*' is significantly more used after 2010, while '*heleduskordaja*' is significantly less used after 2010. This analysis indicates that '*peegeldustegur*' is substituting the term '*heleduskordaja*';
6) clarify the meaning based on context analyses;
7) find contexts and definitions from trustworthy sources.

It's worth noting that while our uploaded documents and most web crawling results primarily consisted of official or journalistic documents, we can reasonably conclude that the noise level was low, even though noisiness is a common characteristic of web texts (Baldwin et al., 2013). Our web crawling results also didn't encounter a significant problem often associated with web corpora (Jakubíček et al., 2020) – machine-translated texts. However, we did face challenges related to the minimal set of metadata, which is typical of web corpora (Jakubíček et al., 2020). Additionally, the metadata of the uploaded files also required additional details. We can assert that the texts in our corpus are generally reliable with minimal noise, but the corpus lacks metadata. Also, it is important to keep in mind that the results reflect the current state of the included texts, which may change slightly if new texts are added.

5. Conclusions and future perspectives

Creating a corpus-based termbase for languages with a small number of native speakers is an important topic to be addressed. Our work demonstrates that adopting a corpus-based approach is viable even when dealing with relatively new topics. The Estonian Remote Sensing Corpus 2022 was compiled using the Corpus Query System Sketch Engine. The term extraction module was used to find significant remote sensing terms and their variants. One of the primary advantages of corpus data lies in its ability to uncover various term variants along with their frequencies of occurrence, a feat that is not achievable with solely expert-intuited data. However, it is crucial to underscore that corpora do not replace expert knowledge in the termbase creation process. Instead, a corpus-based approach should complement an expert-based approach, as most terms still require expert consultation.

We used the dictionary writing system Ekilex to compile the Estonian Remote Sensing Termbase, which currently contains 100 terms. This termbase includes definitions in Estonian and English, usage contexts, variants, and occasionally indicates the status of a term variants.

Our future plans involve expanding the Remote Sensing Termbase in Ekilex by adding new terms and revising existing ones based on user feedback. We aim to make the Estonian Remote Sensing Corpus 2022 publicly accessible through the Corpus Query System KORP, with future plans to provide end-users access through the KORP API also in Sõnaveeb.

6. Acknowledgements

This work was supported by the Estonian Research Council grant (PRG 1978).

7. Bibliographical References

Baldwin, T., Cook, P., Lui, M., MacKinlay, A., Wang, L. (2013). How Noisy Social Media Text, How Diffrnt Social Media Sources? In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, Nagoya, Japan. Asian Federation of Natural Language Processing, pp. 356–364.

Blahuš, M., Jakubíček, M. Cukr, M., Kovář, V. Suchomel, V (2023). Development of Evidence-Based Grammars for Terminology Extraction in OneClick Terms. In *Electronic lexicography in the 21st century. Proceedings of the eLex 2023 conference*, pp. 650-662.

Domnich, M., Sünter, I., Trofimov, H., Wold, O., Harun, F., Kostiukhin, A., Järveoja, M., Veske, M., Tamm, T., Voormansik, K., Olesk, A., Boccia, V., Longepe, N., Cadau, E.G. (2021). KappaMask: AI-Based Cloudmask Processor for Sentinel-2. In *Remote Sens*. 13 (20), 4100. doi.org/10.3390/rs13204100.

Erelt, T. (2007). *Terminiõpetus*. Tartu:Tartu Ülikooli Kirjastus.

Fišer, D., Suchomel, V., Jakubíček, M. (2016). Terminology Extraction for Academic Slovene Using Sketch Engine. In *Proceedings of Recent Advances in Slavonic Natural Language Processing (RASLAN)*. Karlova Studánka, Czech Republic, pp.135–141.

Jakubíček, M., Šmerk, P. (2016). Large Scale Keyword Extraction using a Finite State Backend. In *Proceedings of Recent Advances in Slavonic Natural Language Processing* (RASLAN). Karlova Studánka, Czech Republic, pp. 143–146.

Jakubíček, M., Kovář, V., Rychlý, P., Suchomel, V. (2020). Current Challenges in Web Corpus Building. In *Proceedings of the 12th Web as Corpus Workshop*. Marseille, European Language Resources Association (ELRA), pp. 1–4.

Kallas, J., Suchomel, V., Khokhlova, M. (2017). Automated Identification of Domain Preferences of

Collocations. In *Proceedings of eLex 2017 conference.* Leiden, Netherlands, pp. 309–321.

Kancheva, R. (2013). Remote sensing terminology: past experience and recent needs, In *Proceedings of SPIE,* 8887, Remote Sensing for Agriculture, Ecosystems, and Hydrology XV, 88871Z. Dresden, Germany. doi.org/10.1117/12.2029229,

Khoshkhah, K., Medianovskyi, K., Kolesnykov, D., Hadachi, A., Voormansik, K. (2022). A Hidden Markov Model Method for Non-Stationary Noise Reduction, Case study on Sentinel Data for Mowing Detection. In *Signal, Image and Video Processing,* 17, pp. 3477–3483. doi.org/10.1007/s11760-023-02571-6.

Kilgarriff, A. (2001). Comparing corpora. In *International journal of corpus linguistics* 6(1), pp.97–133, doi.org/10.1075/ijcl.6.1.05kil.

Kilgarriff, A., Marcowitz, F., Smith, S., Thomas, J. (2015). Corpora and Language Learning with the Sketch Engine and SKELL. In *Revue française de linguistique appliquée*, XX, *pp*. 61–80.

Kilgarriff, A.; Rychly, P.; Smrž, P. and Tugwell, D. (2004). The Sketch Engine. In *Proceedings of the XI Euralex International Congress*. Lorient: Université de Bretagne Sud, pp.105–116.

Kilgarriff, A., Jakubíček, M., Kovář, V., Rychlý, P., Suchomel, V. (2014). Finding terms in corpora for many languages with the Sketch Engine. In *Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics*. Gothenburg, Sweden, pp. 53–56.

Kilgarriff, A., Husák, M., McAdam, K., Rundell, M., Rychlý, P. (2008). GDEX: Automatically Finding Good Dictionary Examples in a Corpus. In *Proceedings of the XIII EURALEX International Congress*. Barcelona, pp. 425–432.

Komisarenko, V., Voormansik, K., Elshawi, R., Sakr, S. (2022). Exploiting time series of Sentinel-1 and Sentinel-2 to detect grassland mowing events using deep learning with reject region. In *Scientific Reports,* 12, pp. 1–15. doi.org/10.1038/s41598-022-04932-6

Koppel, K., Kallas, J. (2022). Eesti keele ühendkorpuste sari 2013-2021: mahukaim eestikeelsete digitekstide kogu. In *Eesti Rakenduslingvistika Ühingu Aastaraamat* 18, pp. 207–228. doi.org/doi:10.5128/ERYa18.12

Koppel, K., Tavast, A., Langemets, M., Kallas, J. (2019). Aggregating Dictionaries into the Language Portal Sõnaveeb: Issues With and Without Solutions. In *Proceedings of the eLex 2019 conference.* Sintra, Portugal, pp.1–3.

Kosem, I., Koppel, K., Zingano Kuhn, T., Michelfeit, J., Tiberius, C. (2019). Identification and automatic extraction of good dictionary examples: the case(s) of GDEX. In *International Journal of Lexicography* 32, pp. 119–137. doi.org/10.1093/ijl/ecy014.

Kostka, M. (2022). Pipeline Effectiveness in the Sketch Engine. In *Proceedings of the Sixteenth Workshop on Recent Advances in Slavonic Natural Languages Processing (RASLAN).* Brno, pp. 123–130.

Kwiatek, E. (2017). Evaluation of Multilingual Land Surveying Dictionaries – Part II. In *Geomatics and Environmental Engineering*, 11, 1, pp. 67–78. dx.doi.org/10.7494/geom.2017.11.1.67.

Noorma, A., Jakobson, L., Lang, M., Kutser, T., Oja, T., Uiboupin, R., Voormansik, K., Puust, R., Post, P., Sepp, K., Liibusk, A. (2020a). Kaugseire andmete kasutuselevõtt avalike teenuste väljatöötamisel ja arendamisel (report from project *RITA1 KAUGSEIRE*). doi.org/10.23673/RE-255.

Noorma, A., Lang, M., Alikas, K., Jakobson, L., Olesk, A., Kutser, T., Post, P., Toll, V., Uiboupin, R., Sipelgas, L., Raudsepp, U., Rikka, S., Oja, T., Sagris, V., Puust, R., Liibusk, A., Sepp, K., Järveoja, M., Voormansik, K. (2020b). Kaugseire jätkusuutliku rakendamise kava. doi.org/10.23673/RE-250.

Stan, A. (2023). Compiling a specialised bilingual glossary: challenges and outcomes. In *Language and Communication in the Digital Age* (2023), pp. 218–234.

Tavast, A., Langemets, M., Kallas, J., Koppel, K. (2018). Unified Data Modelling for Presenting Lexical Data: The Case of EKILEX. In *Proceedings of the XVIII EURALEX International Congress: EURALEX*: *Lexicography in Global Contexts*. Ljubljana, Slovenia, pp. 749–761.

Vaus, M. (2022). Workshop about Ekilex, May 2022. terminoloogia.ee/wp-content/uploads/2022/05/Terminitoo-alused_Mari-Vaus_EKI_25.05.2022.pdf

Vezzani, F., Nunzio, G.M.D., Henrot, G. (2018). TriMED: A Multilingual Terminological Database. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018).* Miyazaki, Japan, pp. 4367–4371.

Voormansik, K., Zalite, K., Sünter, I., Tamm, T., Koppel, K., Verro, T., Brauns, A., Jakovels, D., Praks, J. (2020). Separability of Mowing and Ploughing Events on Short Temporal Baseline Sentinel-1 Coherence Time Series. In *Remote Sensing.* 12, 3784. doi.org/10.3390/rs12223784.

Zorrilla-Agut, P. and Fontenelle, T. (2019). IATE 2. In *Terminology* 25, pp. 146–174. https://doi.org/10.1075/term.00034.zor.