Web Application “Latvian Language of Science” for Academic and Research Text Types and Phrases

Raitis Ralfs VECMANIS¹, Raita ROLLANDE¹, Agnese DUBOVA¹, Diāna LAIVENIECE²

¹ Ventspils University of Applied Sciences, Ventspils, Latvia
² Liepaja University, Liepaja, Latvia
raichsh@gmail.com, raitab@venta.lv, agnesed@venta.lv, diantra@gmail.com

Abstract. The paper is based on corpus linguistics research: scientific articles (published from 2008 to 2018) focusing on macrostructure and microstructure analysis to determine the phrases used in Latvian scientific language, argumentation and other aspects. This type of research is needed for all languages, so that academic and research text writers in different levels of education and research can receive support and at the same time improve the quality of the scientific language use.

To ensure that research results are available to a wide target audience in Latvian – school students, university students and researchers – the web application “Latvian Language of Science” (zinatnesvaloda.lv) was developed. It is an easy-to-use digital tool, that users can access in their smart devices. The web app contains three main parts: (1) descriptions of 40 types of scientific texts, separating them into texts that are written in school, university and research; (2) sample phrases that are used in scientific language, which phrase classification based on the structure of a scientific text structure – introduction, main body and conclusions, as well as linking words and reference creation; (3) a glossary that defines 56 terms used in scientific text type descriptions, such as hypothesis, quantitative research, theoretical usage, etc.

The system was developed using an Agile project development framework Dynamic systems development method. The system’s interface is developed using Angular 8 framework. Interface compatibility with and responsiveness to mobile devices is provided by the MDBootstrap framework. The system API is developed using the Node.js environment and the Express.js library. Data storage and maintenance in the system is provided by the PostgreSQL database engine.

Web application “Latvian Language of Science”, which was published on November 30, 2020 has also been added to the CLARINE-LV repository. In the 3 months time since being published, the web application has been used by more than 5400 users with 98.8% engaged sessions per user.

Keywords: information system, web application, Latvian scientific language, academic and research text types, phrases, glossary, corpus linguistic methods
1 Introduction

There is currently no unified system in Latvia where someone can obtain information on how to create academic and research texts and build phrases (linguistic constructions) for scientific texts. Websites have a variety of information resources with conflicting information that creates confusion. As a result, the quality of scientific research texts is at stake generally. In order for students and scientists to produce high-quality scientific research texts and use suitable phrases, there is a need for a unified system accessible to everyone. To solve this problem specialists from the field of linguistics and information technology came together to initially create a theoretical basis for the creation of academic and research texts and linguistic constructions and then to develop an information system that would make the result accessible to any interested person.

For English and German, there already exist websites that have been developed and published about academic text types and phrases used in academic writing. One example of this is the website developed by University Duisburg Essen under the lead of Ulrike Pospiech in 2006, called Der Schreibtrainer (The Writing Couch (German), 2006). This website has short text type descriptions in studies and professional work, given sources for references about grammatical forms and examples, including a short German dictionary with grammar and spelling aid. The second example is Manchester University’s Academic Phrasebank (WEB[a]), which was developed under the lead of John Morley. This phrasebank is based on the selection and classification of phrases used in academic writing, based on authentic sources, that is, 100 dissertations defended at Manchester University in various fields. The German website focuses on text types and their structures, whereas the English language focuses on phrases used in academic writing.

The web app “Latvian Language of Science” (original: “Zinātnes valoda”) offers (1) descriptions of 40 types of scientific texts ranged for school, university and research levels; (2) samples of phrases used in scientific texts; (3) glossary with 56 terms used in the description of scientific texts.

The aim of the article is to describe the web app “Latvian Language of Science”, delineating the app’s structure and it’s usability for a wide target audience – high school and university students and researchers.

2 Academic and scientific text types and linguistic constructions

As the intended application is a manual-type reference material, it is important that the text types included in it are described in texts with similar style and these could be used as separate entries, thus being clear and easy to use. Based on the feasibility study and experience and knowledge about most types of scientific and academic texts, it was decided to describe each text type with the following structural elements: (1) the name of the text type; (2) the name of the text type in foreign languages – English, Russian, German; (3) synonym(s) of the text type name (if any exist); (4) definition and description of the text type; (5) text type structure; (6) references in Latvian where more information on a particular text type may be sought, if the user needs it.
To make the academic and scientific writing application easy for anyone to use (taking into account users with different academic and scientific writing experience levels), three major groups of academic and scientific texts were established:

1. academic texts in school,
2. academic and scientific texts at university,
3. scientific texts in professional environments (research institutions).

Text type descriptions are equipped with a glossary of terms used therein, where the definition of the term can be seen as an “active word” separately or all together in the Glossary section.

2.1 Types of academic texts at school

Educational standards and programs, textbooks and teaching aids, student handbooks, and information available on school websites were studied to select types of texts for school academic texts. As a result, eleven types of text were defined and described: abstract (e.g., abstract of a scientific article); description (e.g., experiment description) argumentative essay; notes (for example, a popular science article notes); a summary (for example, a summary of the project work); protocol (such as field work protocol); review (e.g., school textbook review); presentation; student research work; theses; reports.

The choice of these text types for inclusion in the application was complicated by the fact that there are synonyms for names of text genres and some of the texts described therein as a result of academic writing may, in general, not be written in the style of the scientific language. For example, a review of a non-scientific work may be written in a journalistic style, meeting protocols are usually written in a business style, but descriptions can also be a work of fiction, not non-fiction. This might be a confusing aspect for the user of this group of texts, the pupil, given his or her limited scientific writing experience. The second aspect that raised doubts about whether the text could be defined and distinguished as a separate text type is the question about the relationship between primary and secondary texts. For example, a student’s research work as a primary text may include an abstract, theses, and summary as secondary texts. In other learning situations, however, these secondary texts may qualify for the status of primary text, such as project theses, which are based on oral text reports on the progress of the project.

2.2 Types of academic and scientific texts at university

To select the types of academic and scientific texts most used in higher education, the websites of universities and higher education institutions were researched, and the focus was set on the process and outcome of students’ scientific writing. Information on the requirements for student research paper writing can be found on the websites of almost every university. Several monographs (including translations) have been published in the Latvian language, which deal with scientific works and their writing during the study process (Torgāns, 2011; Klāviņš, 2005; Eko, 2006; Kristapsone, 2008; Martīnsone and Pipere, 2011; Laivenece, 2014; Martīnsone et al., 2016; Martīnsone and Pipere, 2018).
In both Latvian monographs and on the web pages of higher education institutions, methodological guidelines are focused on the research process, workflow and technical characteristics of text formatting, as well as the macro text structure, but not the writing process itself: text, terminology, text micro structure elements, skill to choose and organize a scientific text. Similarly, it has been found that in the student work guidelines available on university sites there are descriptions of a text type, but no definitions, for example, what is a semester work or diploma project. This has been taken into account when developing the text type descriptions for the student target audience.

Initially, two groups of text types were distinguished: (1) academic texts that are written during the study process, and (2) scientific texts that are written for the purpose of obtaining a degree or qualification. The following eight types of texts are distinguished and described as academic papers: term paper, course project, laboratory work description, internship report, report, semester work, seminar report and study project. There are six types of scientific texts: bachelor’s thesis, diploma thesis, diploma project, qualification paper, master’s thesis and master’s diploma paper. In the final version of the web app, this division does not exist as the text types have been arranged alphabetically.

There are a number of factors that can be considered problematic in creating this compilation:

1. the diversity of study programs and specific differences in the research process, which also lead to major differences in the writing of single-title text types, for example, term papers in the humanities and term papers in natural sciences are two quite different representations of the research process in textual form;
2. a different understanding of academic/scientific writing and its outcome – text – in the methodological guidelines of study papers developed not only by each university, but often also by separate faculties of the same university.
3. difficulties in the selection and structuring of information, as the end product of academic/scientific writing – text – is difficult to separate from the research process itself. In turn, this application focuses on the writing process of a work – writing a text based on the language of science.

In both Latvian monographs and on the web pages of higher education institutions, methodological guidelines are focused on the research process, workflow and technical characteristics of text formatting, as well as the macro text structure, but not the writing process itself: text, terminology, text micro structure elements, skill to choose and organise a scientific text. While extensive English literature on academic writing is available in Latvia (Biggam 2020; Godfrey 2018; Hopkins and Reid 2018; Samuels and Garbati 2018; Weaver-Hightower 2019), but it reflects another tradition of academic writing, and therefore the use of this literature directly in teaching materials in Latvian schools and in universities is limited, all the more so because pupils’ and students’ knowledge of the Latvian language of science, their skills in scientific writing and textual writing are still in the process of development.

The Latvian scientific language historically has been developing based upon German, and later Russian science language samples. These are strong traditions that have impacted the specific creation of Latvian scientific text, and Galtung described this as the Teutonic style. This style has the so-called pyramid model – with the least number...
of premises achieving as many conclusions as possible, there is little variety in opinions in the discussions part, the research is centred around theory, that is illustrated by data, not vice versa. Whereas the Anglo-Saxon style, that is prevalent in English and American English scientific language is much freer, more democratic and one could possibly say – more tolerant and positive to other research. (Galtung, 1985)

2.3 Types of texts in science

Both spoken and written forms are used for scientific communication. Although scientists write and publish different types of texts in Latvian in their professional activities, the classification of the types of texts used in the Latvian scientific discourse has not been developed yet. Consequently, written scientific types of texts in Latvian have been compiled and described in the project web app.

Different types of sources are used to describe types of texts: Classification of Latvian Council of Science publications 2012 (LCS [2012]), laws and regulations (MKN [2005]), general and specialized dictionaries and encyclopaedias, publications about scientific publication types (Pipere et al. [2018]) and about separate scientific text types (scientific articles, theses). When looked at sources about scientific texts in Latvian, it could be concluded that very little research in this area has been done, and existing research is fragmentary. Taking into account the aforementioned, the web app includes 14 text types, that are used in Latvian written scientific communication: monograph, scientific article, dissertation, patents, technical reports, computer programs, abstract, theses, summary of a doctoral dissertation, review, introductory article, conference or congress scientific review, literature review and editorial note.

2.4 Phrases in academic writing

The web app not only describes the types of research texts at school, university and scientific professional communication, but also the textual components – phrases used in Latvian scientific language. When writing scientific texts, school students, university students and even experienced researchers need support for text structural element creation in Latvian. Therefore, the web app offers the opportunity to find the phrases used in academic writing, that is nominal and verbal wording of scientific texts, as well as will help to choose the appropriate linguistic wording for different types of scientific text structural elements in Latvian.

This resource includes constructions or vocabulary used in general scientific language, which are found in scientific texts among words or terms of special use, and which can be used in both discipline specific communication and everyday communication (Ehlich [1999]). Nominal and verbal constructions related to the formulas of structural elements are selected from the corpus of authentic texts, Latvian scientific articles with concordance software AntConc via keyword search, concordance, and file viewer sync where also synonyms of text element names at both word and phrase level when capturing a phrase are taken into account. AntConc was chosen for this task because it “has a freeware license, it includes an easy-to-use, intuitive graphical user interface and offers a powerful concordance, word and keyword frequency generators, tools for cluster and lexical bundle analysis, and a word distribution plot” (Anthony [2004]).
Web Application “Latvian Language of Science” 171

The software gets regular updates and is usable for various OS (WEB). The selected wording is then evaluated and described in relation to the structural element of the text. Sample phrases are arranged in five groups. Three groups represent the three main parts of an academic or scientific works: introduction, main body and conclusions. The fourth group includes phrases that are needed to form references and the fifth group offers phrases to link text parts. So, for example, the web app includes phrases used in introductions, which are used to express purpose (pētījuma mērķis ir analizēt .. /the goal of the research is to analyze ..), tasks (pētījumā tiek izvirzīts darba uzdevums /the research task is set as ..), research object (Kā pētījuma objekts tika izvēlēts .. /X was chosen as the research object ..), question (X ir atsevišķa .. pētījuma jautājums /X is the question for another study ..), method or methodology (X metode lāj noteikt Y /X method allows to find Y) hypothesis (Sākotnēji tika pieņemta hipotēze .. /The initial hypothesis was ..).

3 System development

The system was developed using an Agile project delivery framework Dynamic systems development method (DSDM). First created in 1994 by a consortium of United Kingdom companies, DSDM is an Agile method that focuses on the full project lifecycle (WEB; DSDM Consortium, 2014). The origins of DSDM are in rapid application development (Abrahamsson et al., 2003).

The eight principles of DSDM (DSDM Consortium, 2014):
– Focus on the business need
– Deliver on time
– Collaborate
– Never compromise quality
– Build incrementally from firm foundations
– Develop iteratively
– Communicate continuously and clearly
– Demonstrate control

System development process using DSDM includes (DSDM Consortium, 2014):
1. Pre-project: ensures that only the right projects are started, and that they are set up correctly, based on a clearly defined business objective. At this stage, the goals, tasks and development deadlines of the new system were agreed.
2. Project, during which the four main phases of DSDM are applied:
   (a) Sequential Phases: studying the business domain and performing a preliminary analysis of the system:
      i. Feasibility. In this phase, the new system justification was developed.
      ii. Foundations. At this stage, 2nd year students of the bachelor’s study program “Computer Science” were involved in the project to define the solution foundation, business foundations and management foundation. At this stage, students actively worked together with the system customer to identify the new system requirements and display them in the Priority Requirements List and to make the Outline Plan refined into the Delivery Plan.
(b) Iterative Phases (The Development Cycle):

i. Evolutionary Development - iterative and incremental analysis, design, coding. In this phase, a programmer was involved in the development of the new system, who clarified the requirements with the customer and started the iterative development of the system based on the Priority Requirements List. In order for the customer to better understand the system’s capabilities, a system prototype was developed. As a result, the customer understood the functionality of the system much better and thus a better list of requirements for the new system was obtained. The developed prototype was based on CRUD (London 2017), CRUDQ (Plumley 2017) and SCRUD (Diepstraten 2015).

ii. Deployment. For each Increment of the project the solution is made available. In this phase, the programmer gradually delivers the developed system.

3. Post-project: After the final Deployment for a project, the Post-Project phase checks how well the expected business benefits have been met. This phase will be implemented once the system is fully implemented.

The developed web app includes 3 basic parts:

1. descriptions of different text types at school, university, and research level.
2. a glossary explaining the terms and concepts used in the descriptions of the types of scientific texts.
3. sample phrases explaining the phrases used in the scientific texts.

The developed web app is intended for a wide range of users - pupils, teachers, students, university lecturers, and scientists. Each user group can find relevant information on the web app, for example- text types are grouped into 3 categories for schools (report, summary, etc.), universities (bachelor’s thesis, internship report, etc.), and science (scientific article, review, etc.). Each category contains the text types that correspond to it so that each of the user groups can quickly retrieve the necessary information about the text types. In addition to the text types, each of the user groups can get information about frequently used phrases in the Latvian language of science, which could be useful when creating scientific texts.

4 System architecture

The created website is a progressive web application, meaning that visitors can access it from most modern web browsers on both desktop computers and mobile devices. Figure 1 shows the web application’s technology stack and the interaction between those technologies. Furthermore, they can be grouped into three groups: the client and Angular framework as the front-end, Node.js and express as the back-end and PostgreSQL as the database.
4.1 Front-end

The front-end of the web application is built on the Angular 2+ framework. Its design is based on a mobile-first approach and most of the design elements come from the MD-Bootstrap library, which is a combination of Twitter Bootstrap 4 and Material Design. Websites created with modern frameworks and libraries like Angular, React and Vue.js are known as single-page applications (SPA), which means that they contain a single page that is dynamically rewritten to serve the desired content in a faster, more fluid and responsive manner than traditional multi-page websites.

However, single-page applications cause a problem for search engines, which makes them unable to correctly crawl the website and index pages. This happens because the crawlers usually do not execute JavaScript code, therefore all the crawler sees is an incompletely rendered page. This problem can be solved using search engine optimization (SEO). For this web app SEO was implemented via server-side rendering, which detects search engine crawler requests, renders the website and executes JavaScript code on the server side, and responds to the search engine’s crawler with a rendered website.

The created website is also a Progressive Web App (PWA), which allows the user to install the app on their device and run it as if it was a native application. This makes the application easier to use and it also allows caching data, therefore letting the user browse the web app when offline.

The web app is compiled and served using Docker. The Dockerfile contains a list of commands to install dependencies and build the application. The application build stage is implemented using multi-stage builds, which ensures that the source code is not present in the container that will serve the app over the Internet. In the last build stage the compiled code is copied to a Docker container, which contains an nginx web server that will serve the app to the network.

4.2 Back-end

The system’s back-end is based on Node.js, which is a runtime environment that allows running JavaScript code outside of a web browser. The API is based on Express.js, which is a very popular framework for Node.js with 12.6 million weekly downloads."WEB [3].
Interfacing with the database is provided by the pg-promise library, user password hashing is handled by the ‘bcrypt’ library and logging is provided by Morgan.

The Node.js application can also be built into a Docker image using the Dockerfile.

### 4.3 Database

The web application’s database is built using the PostgreSQL database engine. This database engine was chosen because initially it was thought that the data will be more complex and would require relations. However, in the end the research data was condensed to fairly simple entities, therefore a NoSQL database, such as MongoDB would have been sufficient.

The database is served using a custom Docker image.

### 4.4 Deployment

The system can be deployed using Docker Compose, which allows running and managing it with little to no additional interaction from the user.

### 5 Results

As a result of this research, data on types of scientific research texts and phrases used in academic writing has been obtained and the analysis of the obtained data has been performed. Macrostructure and microstructure of different types of scientific texts were studied using corpus linguistics methods to analyze their vocabulary and textual elements, thus promoting the use of high-quality Latvian science language in the national academic environment and preparing the preconditions for further research of scientific discourse. Based on the research results the structure of the web application has been created [Figure 2].

![Fig. 2. Navigation structure of the web application](image)

In [Figure 3], [Figure 4], [Figure 5], [Figure 6] and [Figure 7] the system can be seen running on a mobile device.
Figure 3 is a screenshot of the landing page running in mobile view. This is what the visitor will see when navigating to www.zinatnesvaloda.lv from their smartphone. On the bottom of Figure 3, you can see a purple bar, which is used as the main navigation for mobile devices. The bottom navigation provides easy access to the most important parts of the web application - home, text types, phrases and glossary.

Furthermore, if the visitor wants to browse other pages, such as the list of used literature, the information about the project, the team or contacts, they can use the hamburger menu seen in Figure 4.

Figure 4. Hamburger navigation menu

Figure 5 is a screenshot of the text types list. By using the blue links, the user can navigate to each individual text type.

Figure 6 is a screenshot of one of the text types. Each text type has translations in foreign languages, an explanation, the description and structure. They can also have synonyms and used literature but those are optional fields. In Figure 6 some of the words are highlighted in blue and purple colors. This indicates that they are clickable links. The words highlighted in blue represent a link to a text type, which, when clicked or tapped, navigates the user to a corresponding text type such as seen in Figure 6. However, the words highlighted in purple, when clicked or tapped, navigate the user to the corresponding explanation in the glossary, which can be seen in Figure 7.

Figure 7 is a screenshot of the glossary. The glossary contains a list of explanations for terms used in the Latvian language. The list is a group of collapsible elements.
meaning that, when the term is clicked, the element expands and the description becomes visible.

**Figure 8** is a screenshot of the phrases page. This page contains a summary of commonly used wordings or phrases in the Latvian language.

In the 3 months time since being published, the web application has been used by more than 5400 users with 98.8% engaged sessions per user. The average engagement time of sessions from visitors in Latvia is 1 minute and 6 seconds. 66.4% of users used a desktop computer to browse the website and 33.6% did it with a mobile device. The users have generated more than 45 thousand page views. The most popular text type is research at school with over 500 users visiting the page. On average, each of those users has viewed the text type 2.82 times.

### 6 Conclusion

In the project “Intra-lingual Aspects of the Latvian Scientific Language”, linguists and IT researchers and specialists collaborated by creating and publishing a modern web application that can be used as a source of reference on the Latvian language of science for a wide range of stakeholders.

In the current development stage of the web app – after two years of research – the users are offered writing support in three categories: (1) descriptions; (2) phrases used in various parts of scientific text in Latvian; (3) a glossary, that explains scientific language terms used in the web app. Such research should be continued, as the collaboration between linguists and IT specialists shows that even a complex topic like scientific
writing can be attractive to a large audience, especially easy to use and appealing to students.

The developed system provides a single entry point for finding and displaying high-quality Latvian science language in the form of a Web App. The front-end framework together with the added libraries and responsive design make browsing the website easy on both desktop and mobile platforms. The Node.js back-end allows for fast and secure retrieval of stored information, which is stored using the PostgreSQL database engine. Due to the simplicity of the entities currently stored in the database, other database engines, such as MongoDB could be considered.

In the future, further research into scientific text types is needed, as this would create updates and additions for the web app, such as descriptions of argumentation and its expression in the Latvian scientific language. To promote multilingual scientific communication, the web app could also be supplemented with comparative language material in English, Russian and German, to help researchers who work internationally, in several languages.

Acknowledgements

Research funded by the Latvian Council of Science, project Intra-lingual Aspects of the Latvian Scientific Language, project No. Izp-2018/2-0131.

The publication and presentation is made with the financial support of the project “Promotion of research, innovation and international cooperation in science at Liepaja University”, Project No. 1.1.1.5/18/I/018.
References


Hopkins, D., Reid, T. (2018). The academic skills handbook: Your guide to success in writing, thinking and communicating at university, SAGE.


Mārtinsone, K., Pipere, A. (2018). Scientific writing and dissemination of research findings (Latvian), RSU.


The Writing Couch (German) (2006). [1]


WEB (a). The University of Manchester, [2]
WEB (b). AntConc Homepage, [3]
WEB (c). What is DSDM?, [4]
WEB (d). Express, [5]

Received October 31, 2020, revised February 27, 2021, accepted April 15, 2021