

LEXICC: The Design and Development of an Online Dictionary Writing System

Udiluz del Carmen Monsalve Muñoz, *Instituto Caro y Cuervo,*
Bogotá, Colombia (udiluz.monsalve@caroycuervo.gov.co)
(<https://orcid.org/0009-0003-5225-3181>)

Johnatan E. Bonilla, *Instituto Caro y Cuervo, Colombia;*
and Humboldt-Universität zu Berlin, Berlin, Germany
(johnatan.bonilla@caroycuervo.gov.co)
(<https://orcid.org/0000-0002-8166-3548>)

Ruth Yanira Rubio López, *Instituto Caro y Cuervo,*
Bogotá, Colombia (ruth.rubio@caroycuervo.gov.co)
(<https://orcid.org/0000-0002-1396-9238>)

and

Andrés Steban Luna Cortés, *Instituto Caro y Cuervo,*
Bogotá, Colombia (andres.luna@caroycuervo.gov.co)
(<https://orcid.org/0009-0002-6021-1785>)

Abstract: The Instituto Caro y Cuervo (Caro and Cuervo Institute, ICC) was initially founded to complete Rufino José Cuervo's *Diccionario de Construcción y Régimen* (Dictionary of Construction and Usage) (Cuervo and ICC 1998) and has since expanded its mission to include the research and promotion of Colombia's linguistic heritage. Following this lexicographic tradition, the Institute developed the *Diccionario de Colombianismos* (Dictionary of Colombianisms, DiCol) (ICC 2018) using the proprietary software TshwaneLex, which facilitated the production of its print version but created a dependency on third-party resources, the need for a more flexible and independent solution became apparent. In response, this report introduces *LEXICC – Diccionarios y Lenguajes* (Dictionaries and Languages, LEXICC), a new, tailored online Dictionary Writing System (DWS) developed from scratch as an open-source solution. LEXICC empowers researchers, linguists, lexicographers, and anyone interested in dictionaries to create and manage their lexicographic resources separately. This paper details the design and development process of LEXICC, highlights its main functionalities, and discusses the electronic adaptation of the DiCol, now accessible online through LEXICC.

Keywords: ELECTRONIC DICTIONARIES, DICTIONARY WRITING SYSTEM, COLOMBIAN SPANISH, CARO AND CUERVO INSTITUTE, DICTIONARY OF COLOMBIANISMS, NON-

FUNCTIONAL REQUIREMENTS, FUNCTIONAL REQUIREMENTS, DEMO DICTIONARY, LEXICOGRAPHER DIRECTOR

Opsomming: LEXICC: Die ontwerp en ontwikkeling van 'n aanlyn woordeboekskryfprogram. Die Instituto Caro y Cuervo (Caro- en Cuervo-Instituut, ICC) is aanvanklik gestig om Rufino José Cuervo se *Diccionario de Construcción y Régimen* (Woordeboek van samestelling en gebruik) te voltooi (Cuervo en ICC 1998). Sedertdien is die Instituut se missie uitgebrei om die navorsing en bevordering van Colombia se linguïstiese erfenis in te sluit. In navolging van hierdie leksikografiese tradisie, het die Instituut die *Diccionario de Colombianismos* (Woordeboek van Colombianismes, DiCol) ontwikkel (ICC 2018) met behulp van die kopieregsgateware TshwaneLex, wat die vervaardiging van die gedrukte weergawe daarvan vergemaklik het, maar wat terselfdertyd 'n afhanklikheid van derdeparty-hulpbronne geskep het. Dit het die behoefte aan 'n buigsamer en meer onafhanklike oplossing duidelik laat word. In reaksie hierop stel hierdie verslag *LEXICC – Dictionaries and Languages* (Woordeboeke en Tale, LEXICC) 'n nuwe, pasgemaakte aanlyn woordeboekskryfprogram (WSP) bekend wat reg van die begin af as 'n oopbron-oplossing ontwikkel is. LEXICC bemagtig navorsers, linguïste, leksikograwe, en enigiemand wat in woordeboeke belangstel om hul leksikografiese hulpbronne afsonderlik te skep en te bestuur. In hierdie artikel word die ontwerp en ontwikkelingsproses van LEXICC uitvoerig beskryf, die hoof funksies daarvan word beklemtoon, en die elektroniese aanpassing van die DiCol, wat nou aanlyn via LEXICC toeganklik is, word bespreek.

Slutelwoorde: ELEKTRONIESE WOORDEBOEKE, WOORDEBOEKSKRYFPROGRAM, COLOMBIAANSE SPAANS, CARO- EN CUERVO-INSTITUUT, WOORDEBOEK VAN COLOMBIANISMES, NIE-FUNKSIONELE VEREISTES, FUNKSIONELE VEREISTES, DEMO-WOORDEBOEK, LEKSIKOGRAAF-BESTUURDER

1. The lexicographical tradition of the Instituto Caro y Cuervo

Founded in 1942 to complete Rufino José Cuervo's *Diccionario de Construcción y Régimen* (DCR, Dictionary of Construction and Usage) (Cuervo and ICC 1998), the Instituto Caro y Cuervo (ICC) has significantly contributed to the development of dictionaries for both Spanish and endangered languages in Colombia. Cuervo began this monumental work in 1872, and the ICC's Lexicography Department published the final volumes in 1994. Concurrently, the ICC enhanced Colombian lexicography by creating a lexicographic glossary (Montes et al. 1986) for the *Atlas Lingüístico-Etnográfico de Colombia* (ALEC, Linguistic and Ethnographic Colombian Atlas) (ICC 1983) and supporting the creation of the *Nuevo Diccionario de Colombianismos* (NDC, New Dictionary of Colombianisms) (Haensch and Werner 1993).

Building on these efforts, ICC researchers have been part of various lexicographic projects. One example is the *Diccionario Básico de la Lengua de Señas Colombiana* (DBLSC, Basic Dictionary of Colombian Sign Language) (INSOR and ICC 2011), the first Colombian sign language dictionary representing dialectal variation as it was developed from corpora collected in two major cities: Cali and

Bogotá. In 2014, ICC researchers were involved in two online publications implementing MediaWiki for dictionary management. The first was the *Diccionario Bilingüe Sáliba–Español*¹ (DBSE, Sáliba–Spanish Bilingual Dictionary) (ICC 2014), developed after extensive fieldwork and documentation in collaboration with indigenous knowledge keepers from the Asociación de Autoridades Indígenas de Orocué (Association of Indigenous Authorities of Orocué, ASAISOC) (Dueñas and Gómez 2015). The second was the *Diccionario Académico de Medicina*² (DIACME, Academic Dictionary of Medicine) (Academia Nacional de Medicina de Colombia 2023), primarily advised by the ICC (Bernal et al. 2020). The ICC's lexicographic work also extends into pedagogical and sociocultural projects and research. Notable contributions include the *Glosario de aprendizaje del Español de Colombia* (Glossary for Learning Colombian Spanish) (Nieto 2017) and the *Léxico de la Violencia en Colombia 1948–1970* (Lexicon of Violence in Colombia 1948–1970), based on the literature of the violence canon (Rozo et al. 2020).

However, the most significant work in recent times is the *Diccionario de Colombianismos* (DiCol, Dictionary of Colombianisms) (ICC 2018). This dictionary was created after reviewing various dictionaries of Americanisms and Colombianisms, such as the NDC (Haensch and Werner 1993), the *Diccionario de Americanismos* (Dictionary of Americanisms) (Asociación de Academias de la Lengua Española 2010), and the *Breve Diccionario de Colombianismos* (Brief Dictionary of Colombianisms) (Academia Colombiana de la Lengua 2012). It was further refined through collaboration with researchers from various Colombian regions, resulting in 6,000 entries, 8,000 definitions, and 4,500 examples (ICC 2018: 16).

The DiCol was developed with the commercial Dictionary Writing System (DWS), TshwaneLex (TLex). However, despite TLex's robust capabilities, its proprietary nature poses significant challenges. In the first place, the procurement of a paid license may present a notable obstacle for budget-constrained researchers, small institutions and native communities. Additionally, in the case of DiCol, the team of lexicographers needed the assistance of a developer to build the structure of the entries, which affected the autonomy in the initial phase of the project. Lastly, for the online publication of DiCol it was necessary to devise another system that allowed both its online viewing and constant updating of the entries by the team of lexicographers.

In summary, the experience in lexicographic projects and the approach to electronic lexicography (such as the creation of DiCol) have awakened multiple demands in the Institute: firstly, there is the need for efficient and user-friendly tools for the development of dictionaries to facilitate the work of the Institute's lexicographers and researchers. For example, the *Glosario de Aprendizaje del Español de Colombia* (Glossary for Learning Colombian Spanish) (Nieto 2017) was edited through a word processor which implies difficulties in the organisation and efficient search of entries, limitations in the handling of large volumes of data, and problems integrating and updating information in a collaborative and automated manner.

Secondly, there is the need to ensure the publication and consultation of these resources in an accessible and efficient way for different audiences. Several of the Institute's dictionaries are in printed format or in spaces that do not guarantee the safeguarding and easy consultation of the data. Thirdly, it is important to integrate the functionalities that optimize the processes of creating, updating, and handling lexicographic data. For instance, the migration of the DBSE (ICC 2014) from its MediaWiki format will be vital to ensure the safeguarding of information and the constant updating of the dictionary by the community. Finally, it is vital to highlight the lexicographic projects that are currently under development and that reflect the need for a platform for their publication, access, and consultation: the *Diccionario Español–Andoque* (Spanish–Andoque Dictionary) (Andoque and Landaburu 2023), created with members of the Aduche Reserve in the Amazon (Colombia), and the *Corpus Léxico del Español de Colombia* (CorlexCo, Lexical Corpus of Colombian Spanish) (Nieto 2020), a linguistic corpus composed of lexical combinations used in Colombia and other Spanish-speaking countries, which could also become a dictionary in the future.

The ICC Corpus and Computational Linguistics (LICC) research line initiated the project *LEXICC – Diccionarios y Lenguajes* (Dictionaries and Languages, LEXICC)³. This open-source DWS allows the creation, management, publishing, and consultation of the lexicographic products of the Instituto Caro y Cuervo (whether new or existing), and any dictionary created by platform users, specialists, or not in lexicography. This report outlines the construction process of LEXICC, its requirements, its main features (users and interfaces), future improvements, and the digitisation process of the DiCol, the first printed dictionary published on the platform.

2. Building LEXICC

To design and develop the LEXICC platform, a multidisciplinary team was formed, consisting of a full-stack web developer, a graphic designer, and four linguists. The developer handled the technological implementation of the platform and supporting technical requirements; the graphic designer worked on the layout and visualization of the system; and the linguists conducted preliminary research for platform development, proposed requirements, advised on development and design tasks, and tested the platform's usability.

Methodologically, a cascade-hybrid system development was adopted. This approach, based on the six phases of the system development life cycle — feasibility study, analysis, design, coding, testing, and maintenance (Leau et al. 2012), permitted the execution of each phase systematically and interdependently. The construction of LEXICC began in 2019 with a bibliographical inquiry into electronic lexicography, systems or tools for dictionary creation and publication, and Corpus Query Systems (CQS) and how they contribute to lexicographic work. The characteristics of the Institute's dictionaries and various types of electronic dictionaries were also reviewed. Based on this analysis, the platform

requirements outlined in Section 3 were established. In the first semester of 2020, the design and development phases commenced with the definition of the technical specifications. During this semester, the database was built, and the back-end services were implemented ensuring traceability through GitLab⁴ (these technical specifications are detailed in Section 4).

In 2021, the focus shifted to developing user interfaces presented in Section 5. During this period, interfaces for different user types and specific services for dictionary structures were created. These developments have continually improved to enhance the user experience up to 2023. By then, the dictionary creation functionality was refined, the coordinator lexicographer interface (see Section 5.3) was developed for efficient dictionary management, and the formatting of both print and digital versions of the articles was improved. Finally, as presented in Section 6, the 2018 DiCol was migrated to the new database, so it is now accessible and managed through the platform.

3. Requirements analysis

To identify and fulfil the specific needs for LEXICC, a comprehensive three-part requirements analysis was conducted. Firstly, key aspects of electronic lexicography and features from renowned dictionary writing systems (DWS) were analysed. This included systems such as TLex (TshwaneDJe HLT 2023), EELex (Institute of the Estonian Language 2023), Dictionary Production System or DPS (IDM France n.d.), and Lexonomy (Měchura 2017). Secondly, existing ICC dictionaries (mentioned in Section 1) were reviewed in detail, and consultations were held with their authors to gain deeper insights into the need for a customised platform for the institute. Finally, the experience of the LICC group in developing linguistic information management systems was leveraged. This includes projects like the *Corpus Lingüísticos del Instituto Caro y Cuervo (CLICC)*⁵ (CLICC, Linguistic Corpora of the ICC) (Rubio et al. 2023; Rubio and Bernal 2019) and the *Sistema de Información Geográfica del ALEC*⁶ (SIGALEC, ALEC's Geographic Information System) (Bonilla et al. 2020; Bonilla and Bernal 2020), particularly their collaborative methodologies in interdisciplinary processes and project management.

Based on this analysis, three fundamental components of a DWS were identified: a database, a dictionary entry editing interface, and administrative tools for user registration and dictionary tracking. However, this structure was refined for current development with: (1) a database (DB) (Section 4) to store and manage all dictionary data; (2) a General Interface (GI) (Section 5.1) for the presentation of the project to anonymous users, and options for registering or logging in; (3) a Visualization Interface (VI) (Section 5.3.2) for the query of public dictionaries; (4) an Administrative Interface (AdI) (Section 5.2) for user management: roles, projects, and editable aspects of website interfaces; (5) a Demo Interface (DI) (Section 5.3.1) that allows users to create and consult demo dictionaries for educational purposes; and (6) a Lexicographer Interface (LI) (Section 5.3) for expert

researchers and their collaborators to manage their dictionary projects comprehensively. This refined structure ensures a customised interface to the specific needs of its users, from public interaction to specialised research and project management.

To achieve this, a list of requirements was gathered. These requirements were divided into three categories: non-functional (Table 1), functional (Table 2), and users and use cases. According to Sommerville (2011), functional requirements refer to what the system can do; therefore, they describe the functions or services that the system should offer. In contrast, non-functional requirements relate to the properties of the system that enable it to adequately provide those functions or services, such as performance, response time, and storage capacity, among others. Finally, use cases identify the different scenarios in which users can interact with the system.

Regarding LEXICC, non-functional requirements ensure the platform's reliability and user experience, while functional requirements and use cases specify the tasks each user role can perform within the system. Each requirement was organised in tables with the name and description of the requirement, the requirement source, and the priority level (High/Essential, Medium/Desired, Low/Optional).

Table 1: Non-functional requirements of LEXICC

No.	Requirement Name	Description	Priority
RNF01	Flexibility	Store various types of information and create new dictionaries.	High/Essential
RNF03	Performance	Queries responded to in less than 5 seconds.	Medium/Desired
RNF04	Information Update	Constant updates and immediate saving.	High/Essential
RNF05	Security	Protection against unauthorised access.	High/Essential
RNF07	Backup Copies	Weekly backups of data and developments.	High/Essential
RNF08	Support and Maintenance	Documentation for future adjustments.	High/Essential
RNF09	Graphical Interface Policies	Compliance with design policies.	High/Essential
RNF10	Query Browsers	Accessible from the most used browsers.	High/Essential
RNF11	Applications	Offline use as a downloadable application.	Medium/Desired

The non-functional requirements of LEXICC (Table 1) focused on key aspects such as flexibility in storing diverse information types, quick response times for

queries, and constant updates with immediate saving. Security measures protect against unauthorised access, with weekly data backups and comprehensive support documentation ensuring system integrity and maintainability. It was required that the graphical interface comply with design policies and be accessible from major browsers, with offline use available as an optional feature.

Table 2: Functional requirements of LEXICC

Interface	General Description of Requirements	Reference Platforms or Dictionaries
General Interface	Project data, menu with resources and dictionaries, options to login and sign up, it allows to visit available dictionaries with Visual Interface and to access the Demo Interface.	All
Administrative Interface	The administrator can assign roles, consult, and export query statistics, and edit news and content. The system will log changes and the users who made them.	CLICC, TLEX
Lexicographer Interface	Tools for configuring dictionaries, editing, and modifying information, viewing entries in three formats, adding lexical relations, and managing editing windows for bilingual dictionaries. Integration with CQS.	Sáliba–Spanish Dictionary, CLICC, TLEX, Lexonomy, TERMSTAR, Golden Dict, Termania
Demo Interface	Menu to customise searches, create DEMO dictionaries/glossaries, download in Word, PDF, XML, save changes in the system, and participate in consultation forums.	Wordreference, DEBWrite
Visualization Interface	Search box with various search types: approximate, exact, alphabetical, begins with, contains, ends with, anagrams, random, manual configuration, region, semantic field, marks, visual dictionary.	RAE, INSOR Dictionary

On the other hand, functional requirements (Table 2) defined the specific capabilities needed for each interface within LEXICC. For instance, The AdI allows administrators to manage user registrations, assign roles, export query statistics, and edit content, with all changes logged in. The LI provides tools for configuring and editing dictionaries, viewing entries in different formats, adding lexical relations, and managing bilingual dictionaries, integrated with Corpus Query Systems (CQS). The DI enables users to customise searches, create demo dictionaries, download content, save changes, and participate in forums. Finally,

VI offers various search options within the dictionaries.

In terms of user experience, the requirements imply dedicated access and features for distinct types of users. The requirements defined per user experience were as follows:

- **Administrator User (AU)**
 - Manage the entire system, including user and dictionary management.
 - Register, modify, search, or delete users.
 - Activate or deactivate roles.
 - Add, delete, and manage the display of dictionaries.
 - Edit content viewed by users.
 - Obtain statistical reports on user activity and general queries.
- **Guest or Anonymous Users (GU)**
 - Navigate the general webpage.
 - Explore available dictionaries through the VI.
 - Create dictionaries using the DI but cannot save them.
 - Access sign-up and login interfaces.
- **Registered User (RU)**
 - All Anonymous Users capabilities
 - Save and download information from created and consulted dictionaries.
 - Participate in collaborative tools.
 - Choose query preferences.
 - Review recent searches.
 - Select specific dictionaries to consult.
 - Save consulted words.
 - Forums participation.
- **Lexicographer Director (LD)**
 - Define the type of dictionary and its general structure.
 - Define styles and fonts of dictionary.
 - Grant Coordinator and Editor roles to users.
 - Supervise the team and indicate progress states of an entry (unassigned, in editing, for review, approved).
 - Grant consultation and download permissions for their dictionary.
 - View statistics of the dictionary's development.
 - Create, edit, and delete entries.
 - Manage forums, comments, and dictionary suggestions.
 - Include notes in entries.
- **Lexicographer Coordinator (LC)**
 - Create, edit, and delete entries.
 - Manage forums, comments, and dictionary suggestions.
 - Include notes in entries.

- **Lexicographer Editor (LE)**
 - Create, edit, and delete entries.
 - Include notes in entries.

To summarise, the user experience requirements are tailored to provide dedicated access and features for distinct types of users, ensuring each role can efficiently perform its designated tasks. AUs manage the entire system and oversee user activities, while GUs have limited navigation and dictionary exploration capabilities. RUs gain additional privileges, such as saving and downloading information and participating in collaborative tools. LDs and LCs have more specialised functions related to the creation and management of dictionaries, with the LD also overseeing team roles and progress. LEs focus on entry creation and editing. Each role's specific capabilities ensure a structured and efficient workflow, promoting effective dictionary management and user interaction.

4. Technical implementation and development

The technical implementation for the LEXICC project aims to address challenges identified from previous developments by the LICC group and to remain at the forefront of current software development. While the CLICC and SIGICC platforms (Rubio et al. 2023; Rubio and Bernal 2019; Bonilla et al. 2020; Bonilla and Bernal 2020) utilised relational databases constructed in MySQL and PostgreSQL respectively, and PHP, LEXICC has adopted the MERN stack (MongoDB, ExpressJS, React, and NodeJS) (Aggarwal and Verma 2018). These technologies are widely used, constantly evolving, and provide dynamic and efficient interfaces, significantly enhancing user experience and data visualization fluency.

Firstly, MongoDB is a NoSQL database designed for scalability, flexibility, and storing large volumes of data. Unlike traditional table-based relational databases, MongoDB uses collections and documents, facilitating the configuration of a dynamic and adaptable structure (Győrödi et al. 2022). This allows the registration of dictionaries without a predefined structure. Users can create their own configurations for their dictionaries without requiring specialist intervention, enabling the storage of attributes not defined during the requirements phase based on a relational model.

As described in Section 3, LEXICC's database allows for the storage of information regarding dictionaries, users, and queries. It is essential for registering and systematising lemmas, definitions, parts of speech, and other entry components, as well as for administrative actions such as user activation and deactivation, recording entry progress states, organising lexicographer teams and task assignments, and generating statistics, among others. Additionally, the database facilitates the import of dictionaries created in other systems, which is crucial for populating the platform. The LEXICC database is flexible for storing any type of information (documents, videos, audios, texts, images, etc.), and its per-

formance supports the continuous growth of system data. This solution effectively resolves a challenge encountered during the development of CLICC. Specifically, the integration of new corpora and metadata with differing attributes necessitated normalisation to the sixth normal form. This process involved identifying and eliminating all multivariable relationships within the database, allowing for the registration of a table to store attributes. Subsequently, information was recorded in another table where responses to these attributes were stored and linked to the newly registered corpora entity (Rubio et al. 2023).

Secondly, ExpressJS is a web framework for NodeJS used to create APIs simply and efficiently. This tool was used for data persistence and to meet the group's need to share information between applications and even entities that required it. Thirdly, React is a JavaScript library widely used in modern web development due to its reactive interfaces, allowing users to have a smooth and fast experience when querying or configuring dictionaries. This implementation was based on an atomic design methodology, which is grounded in the creation of modular and scalable systems, making each created component reusable. Lastly, Node.js, the server-side JavaScript runtime environment, integrates seamlessly with Express.js and React to provide a comprehensive and efficient solution for web application development (Saundariya et al. 2021).

Finally, the development process was supported by traceability management in GitLab, a tool that allows for managing the LEXICC repository and controlling all platform versions. GitLab has also been very useful for tracking the development process, as there is a dedicated space available to help plan, organise, review, and approve workflow tasks until the development project is released to production.

5. LEXICC interfaces

Following rigorous research and requirements gathering, five key interfaces that power LEXICC were developed: General Interface (GI), Administrative Interface (AdI), Demo Interface (DI), Lexicographer Interface (LI), and Visualization Interface (VI). In the following sections, their main characteristics will be explained.

5.1 General Interface (GI)

The GI, depicted in Figure 1, serves as the homepage to navigate dictionaries and familiarise oneself with LEXICC. The numbered modules in this figure are described below.

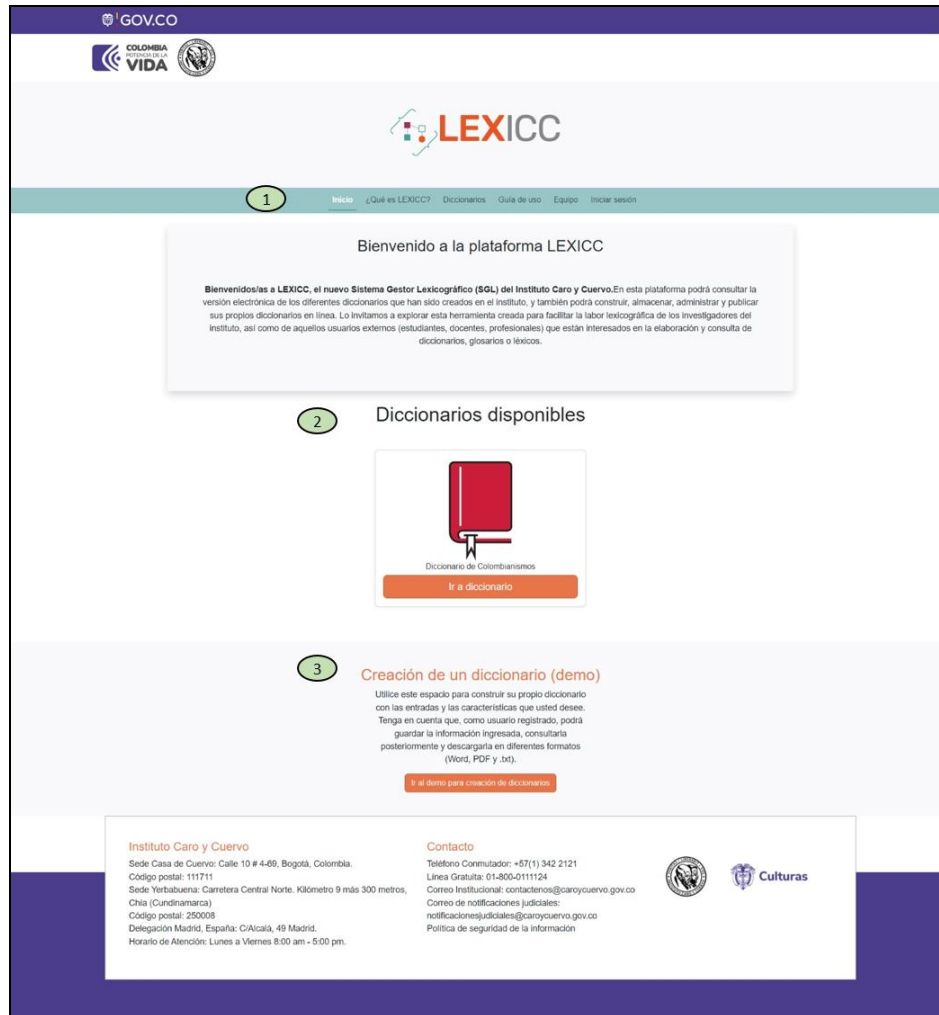


Figure 1: Homepage in the GI

The main menu (Module 1) consists of 6 hyperlinks: (1) *Inicio* (Home), which appears by default when entering the platform to welcome the users; (2) *¿Qué es LEXICC?* (What is LEXICC?) presents a description of the objectives and characteristics of the platform; (3) *Diccionarios* (Dictionaries) allows access to the available dictionaries; (4) *Guía de uso* (User Guide) is a dropdown menu containing the platform's usage manuals, including the RU manual and the LD manual, which provide detailed information about the system tools; (5) *Equipo* (Team) displays the names and contact information of the research group responsible for creating LEXICC, including linguists, designers, and developers; finally,

(6) *Iniciar sesión* (Log in) enables users to register in LEXICC and access the platform as a RU. If they do not have an account in the system, they fill out the registration form that appears at the link *Crea una cuenta ahora* (Create an account now) (Figure 2). However, if they already have an account, they can log in by entering their email and password.



Figure 2: Interface to log in

Referring back to Figure 1, users can also access the dictionaries of the system in Module 2 via the *Inicio* (Home) hyperlink. By clicking on *Ir al diccionario* (Go to Dictionary) users are directed to the VI of the DiCol, which is currently the only available dictionary (see Section 5.3.2). Notably, registration on the platform is not required to access the dictionaries, if they have been enabled by the AU. Additionally, in Module 3, users can access the DI without registration. They simply need to click on the *Ir al demo para creación de diccionarios* (Go to dictionary creation demo) option and they will immediately be directed to the DI (presented in Section 5.3).

5.2 Administrative Interface (AdI)

To access the AdI, users must log in and select the AU role. This selection is made by clicking the blue arrows next to the role, which will display a window where the AU role can be chosen (Figure 3). The AdI enables the management users and dictionaries within the system.

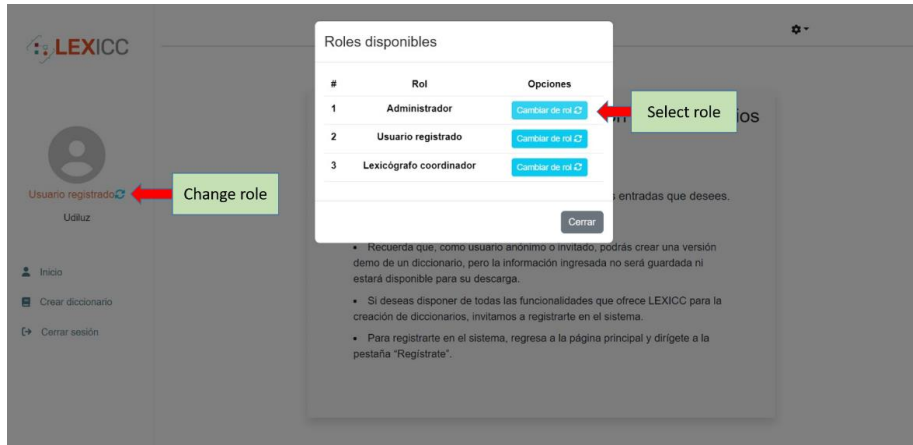


Figure 3: Selection of the AU role by a RU

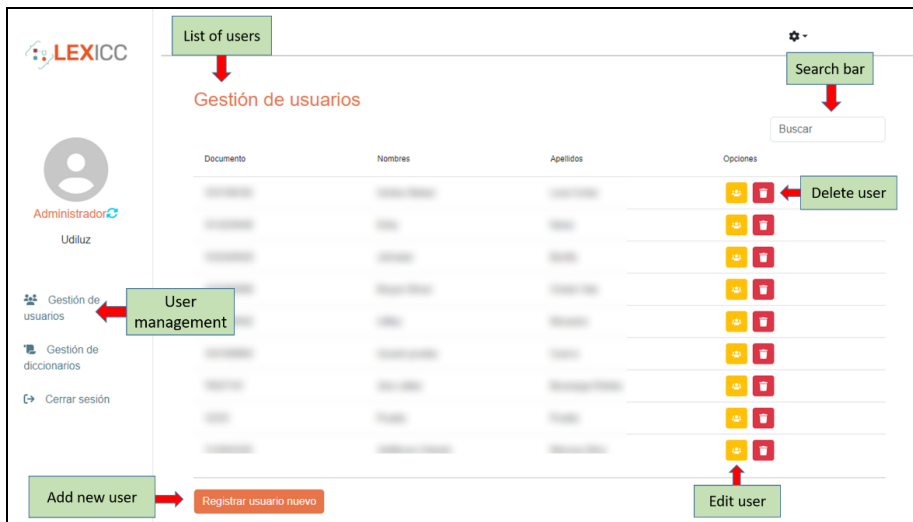


Figure 4: User management in the AdI

For user management, the AU must select the *Gestión de usuarios* (User management) option in the left menu of the interface (see Figure 4). This will display a list of registered users within the system. Users can be deleted or edited using the options on the right-hand margin. Additionally, specific users can be searched using the search bar in the top-left corner. New users can be added by clicking the *Registrar usuario nuevo* (Register new user) button at the bottom of the interface.

Editing a user involves modifying their personal data or assigning permissions within the system. To edit a user, the AU must click on the yellow symbol in the right margin of the interface, which will display an editing window (see Figure 5). In this window, the AU can change the roles of the RU and enable the dictionaries in which they can participate as directors or editors. The image indicates that the user has the DL role deactivated and two assigned dictionaries. The AU can modify these permissions at any time.



Figure 5: User editing window in the in the AdI

5.3 Lexicographer Interface (LI)

The LI is used by the three types of lexicographer users described in Section 3 (LD, LC, and LE). To access this interface, users must log in with their e-mail and password (Figure 2, Section 5.1), and then select their assigned role from the role change window (Figure 3, Section 5.2). The LI currently includes a user management menu, an option to create new dictionaries, a menu to select the dictionary where lexicographers wish to work, and the DI for creating and editing dictionaries. Figure 6 shows the LI used from the perspective of an LD. On the left side, DiCol is shown as the selected dictionary for editing, and a new dictionary can be created using the option below. After this, the LD can manage the users of the dictionaries under their responsibility, similar to the AU's user management (Figure 4, Section 5.2). To access the DI, the LD clicks on *Mi diccionario* (My dictionary), where the structure and entries configuration of the chosen dictionary are managed. Figure 6 also displays the General module of the DI, where the general description of the chosen dictionary (DiCol in this case)

is entered. For further details, Section 5.3.1 will explain what the DI consists of and the process of creating a dictionary.



Figure 6: General module of the DI viewed from the LI

5.3.1 Demo Interface (DI)

The DI comprises four modules:

- (1) *General* (General module): Used to enter general information about the dictionary, such as its name, description, abbreviation, team involved, information about the lexicographical project, and copyright information (Figure 6, Section 5.3).
- (2) *Anexos* (Annexes module): Allows including additional information such as the dictionary's usage guide, abbreviations, symbols, and any other relevant documents (Figure 7).
- (3) *Estructura* (Structure module): Allows to organise the structure of the dictionary entries, including the headword, definition(s), example(s), etc. (Figure 9).
- (4) *Mi diccionario* (My Dictionary module): Facilitates the creation of dictionary entries following the predefined structure (Figure 16).
- (5) *Estilos* (Styles module): Allows users to edit the style of the entry structure and the style of the VI (Figures 17 and 18).

For GU and RU, only modules (1), (3) and (4) are available, while lexicographer users (LD, LC, and LE) can use all of them. This section of the article showcases the modules of the DI using the interface of an LD, as this user has the most comprehensive editing functionalities. To explain the entry structure, also known as

microstructure (Atkins and Rundell 2008), the term *entry components* will often be used, referring to "the separate pieces which go to make up the dictionary entry" (Atkins and Rundell 2008: 202). These pieces or elements vary depending on the type of dictionary (monolingual or bilingual), its purposes, and the intended audience. Generally, entries in a monolingual dictionary will consist of a headword, a definition, grammatical information, and other linguistic labels regarding usage (dialect, register, attitudes).

To create a dictionary in LEXICC, users first select the *Nuevo diccionario* (New dictionary) option from the left menu of the DI (Figure 6, Section 5.3), which will open the form that needs to be filled out. The first part of the form is the General module, presented in Figure 6 (Section 5.3), where users complete the dictionary description and click the *Crear diccionario* (Create dictionary) option at the end of the module (Figure 7).

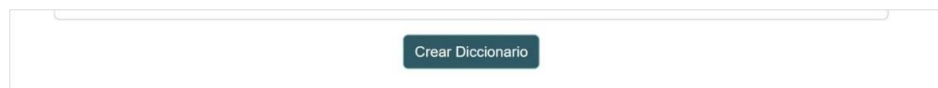


Figure 7: Button to create a dictionary

After this, they should proceed to the Annexes module to enter additional information about the dictionary (Figure 8). Both modules have WYSIWYG editors (What You See Is What You Get) for inserting and editing information, and users can decide which texts or documents appear published in the VI using a toggle button. The general description of the dictionary will appear at the top of the VI, and its annexes will appear on the left side (Figure 19, Section 5.3.2).

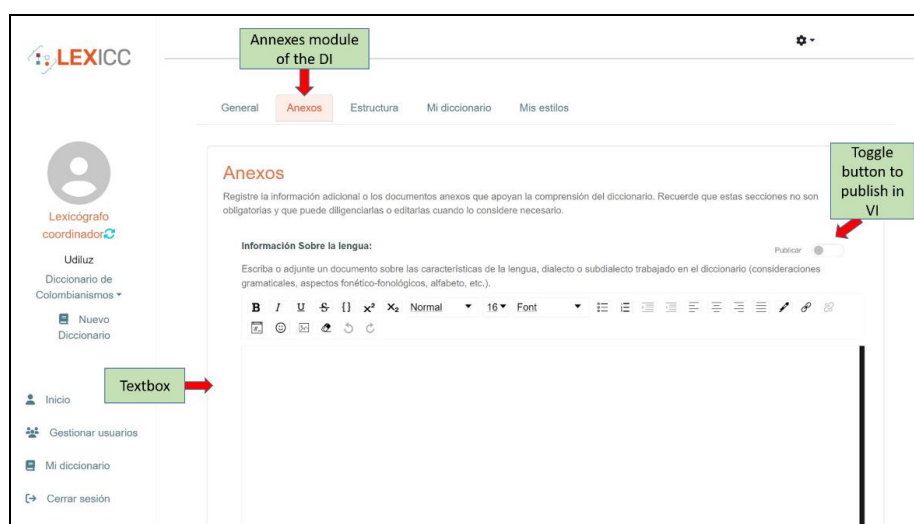


Figure 8: Annexes module in the DI

Next, users establish the entry structure in the Structure module illustrated in Figure 9. To add entry components, the first step is to click on the blue button with the plus sign (+), indicated in the same figure. This action will display the window presented in Figure 10, where users must choose the component type and the symbols accompanying the component.

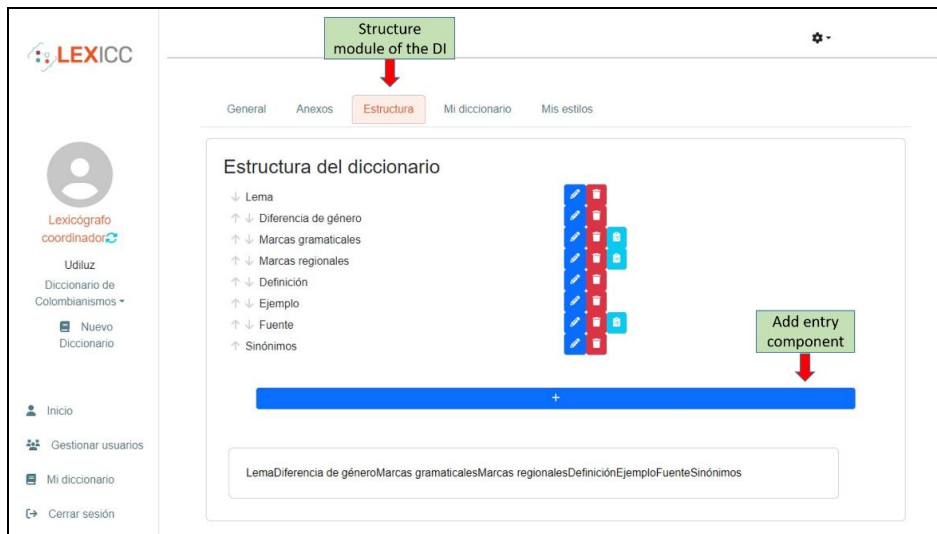


Figure 9: Structure module in the DI: Button to add entry component

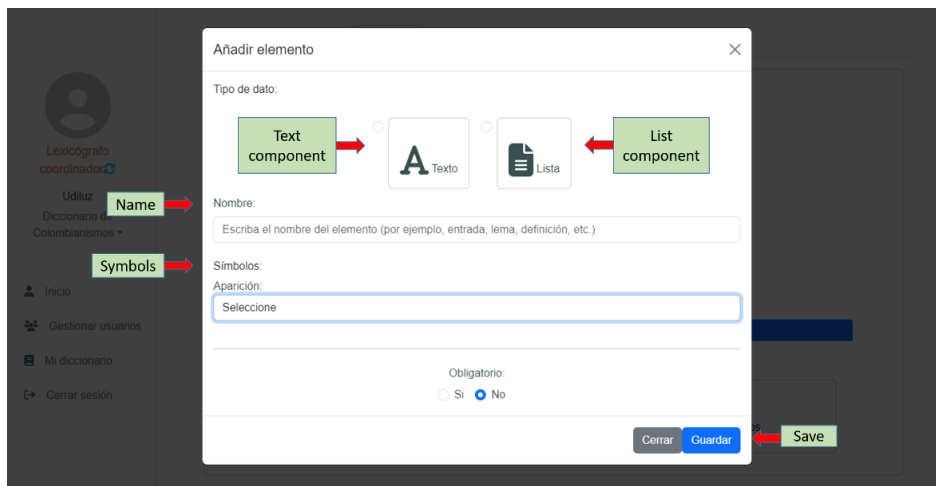


Figure 10: Menu to add entry components in the DI

An entry in LEXICC can include text components (e.g., definitions) and list components where a specific option is selected (e.g. parts of speech). If users want to create a text component named *Lema* (headword), they should click on the circle corresponding to text data (Figure 11, step 1) and specify the name of the entry component (Figure 11, step 2). Afterward, the user designates the placement of the symbols accompanying the component (left, right, or both), and the type of symbol (e.g., period, comma, colon, bars, etc.) (Figure 11, steps 3 and 4).

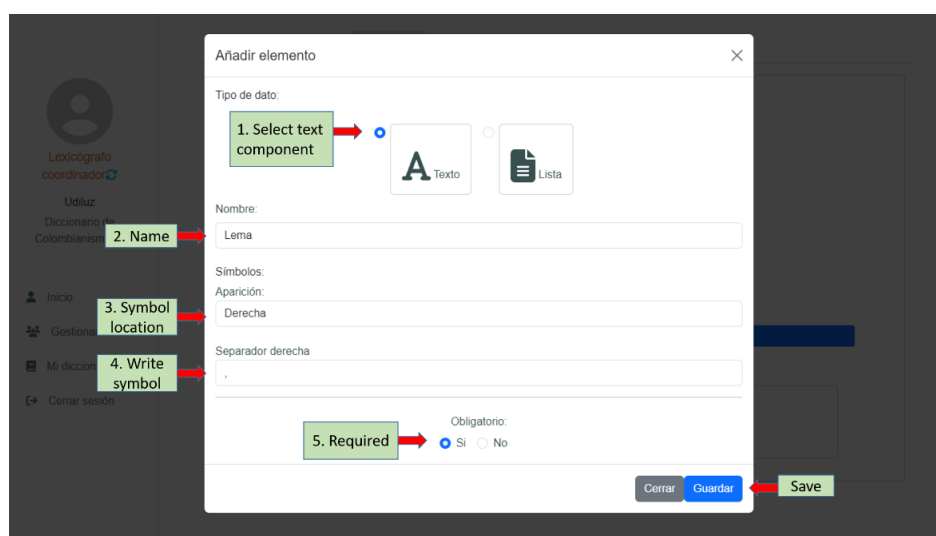


Figure 11: Menu to add entry components: Example of text component

In the example shown in Figure 11, the headword will be accompanied by a comma on the right side, and it will be a required component in the dictionary (step 5), meaning it will appear in all registered entries. If users wish to include a list component in the entry structure, the procedure described is repeated, but users must click on the circle corresponding to list data. Figure 12 illustrates all the steps to add the list component named *Categoría gramatical* (part of speech), which will appear in parentheses in the dictionary.

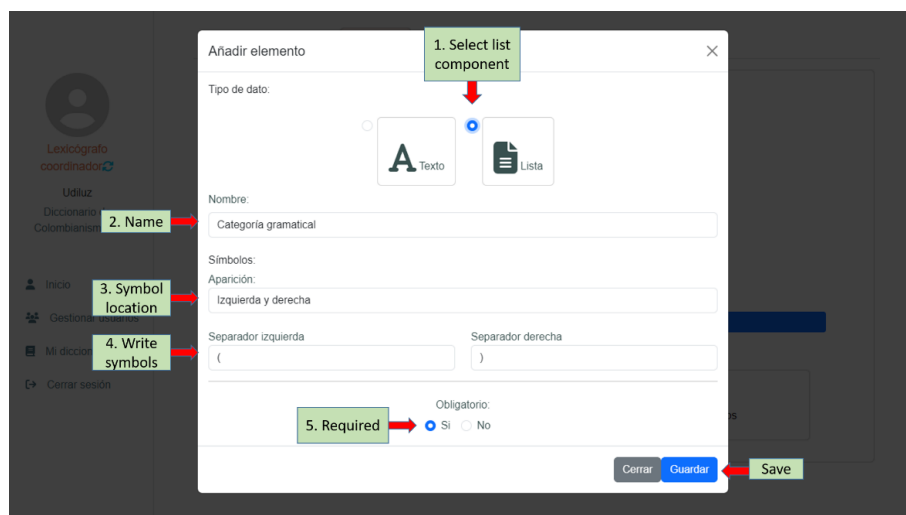


Figure 12: Menu to add entry components: Example of list component

Considering that a list component consists of different attributes, the corresponding menu must be filled out. To access this menu, users must return to the main interface of the Structure module and click on the blue icon located next to the list component that was previously added (Figure 13). This action will display the menu shown in Figure 14, where users can enter the attribute information that makes up the component.

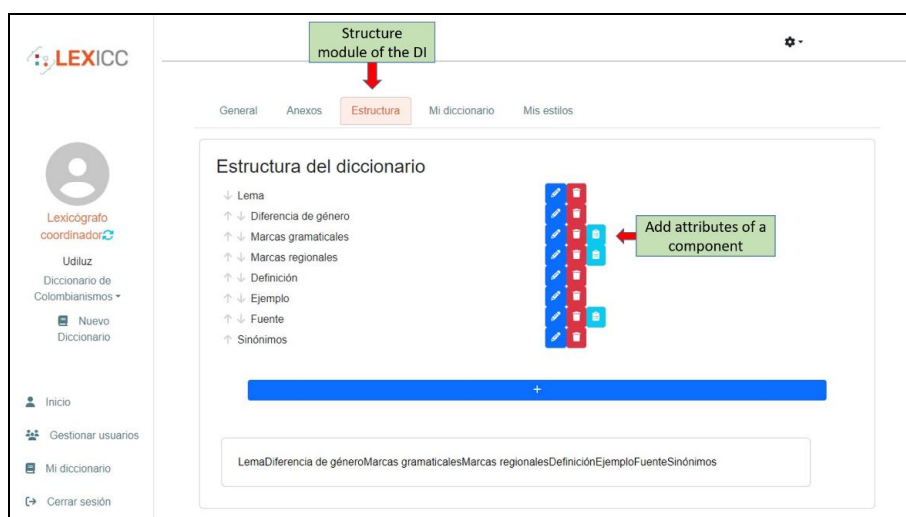


Figure 13: Structure module in the DI: Button to add attributes of a component

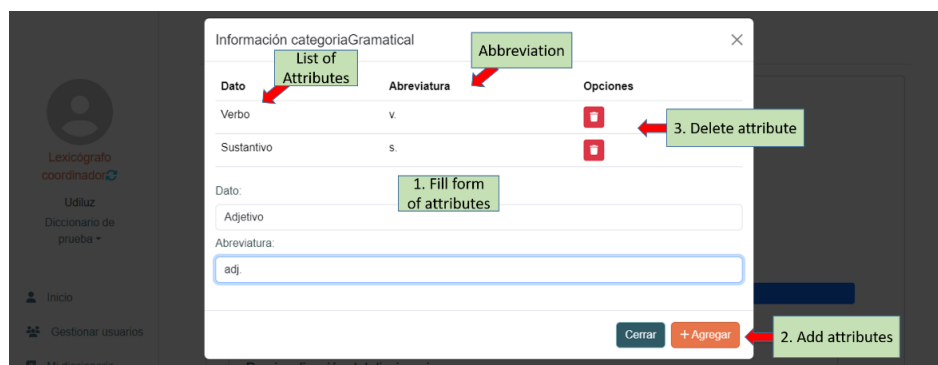


Figure 14: Menu to add attributes of an entry component in the DI

The first step is to include the name of the attribute and its abbreviation. For example, the *Adjetivo* (Adjective) is an attribute for a parts of speech component, and the form *adj.* is its abbreviation (Figure 14, Step 1). The second step is to click the *Agregar* (Add) button (Figure 14, Step 2), and the new attribute will appear at the top of the window. If the attribute is no longer needed, it can be deleted by clicking on the trash can icon (Figure 14, Step 3).

The main interface of the Structure module also allows for other actions as indicated in Figure 15: delete components that are no longer needed by clicking on the trash can icon, edit components as required by selecting the pencil icon, and reorder components by clicking on the gray arrows on the left side.

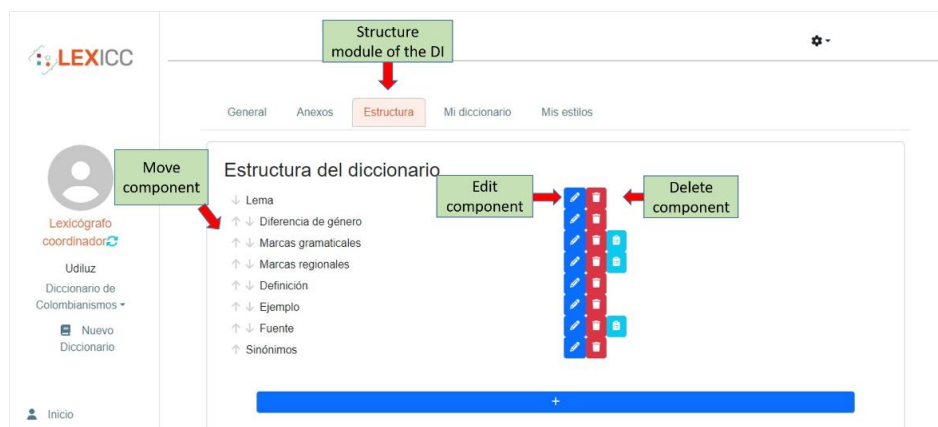


Figure 15: Structure module in the DI: Buttons to move, edit and delete a component

Once the dictionary entries are structured, the user should navigate to *Mi diccionario* (My dictionary module) to fill out the proposed form (Figure 16). In this same window, entries can be searched, previsualized, and downloaded in their current state, whether draft or finalised.

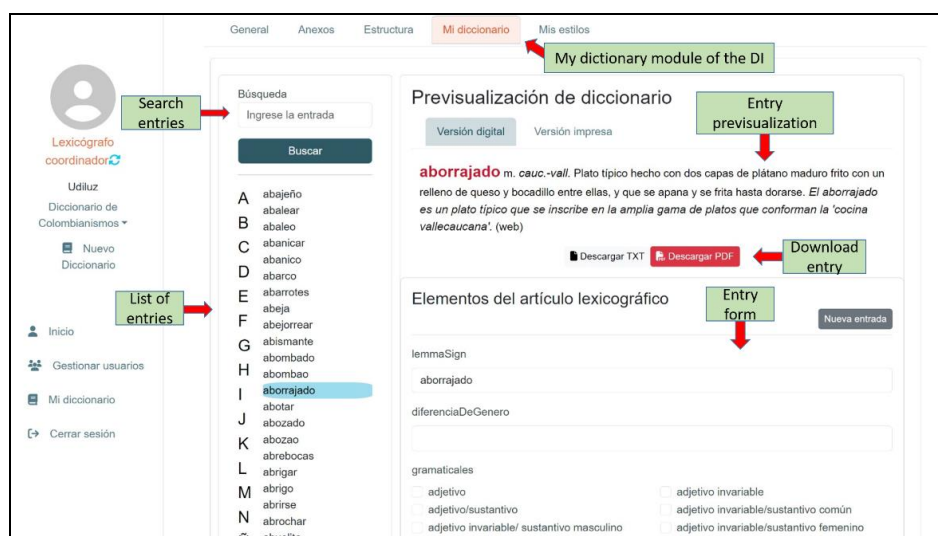


Figure 16: My dictionary module of the DI viewed from the LI. This figure shows the form for an entry in the DiCol

Each of the entry components can have the style assigned by the LD. Modifications can be made in the *Estilos* module (Styles), which is only available in the LI. Figure 17 shows the entry styles interface, where the LD can choose font colour, types of fonts, and font size, among other options. On the other hand, Figure 18 depicts the dictionary styles interface, where the LD can select the main and secondary colours that users will see in the VI. Figure 19 in Section 6 shows the VI of the DiCol, characterised by red as the main colour and black as the secondary colour. However, for other dictionaries, these colours may vary according to the preference of the lexicography team in charge.



Figure 17: Styles module of the DI: Entry styles interface

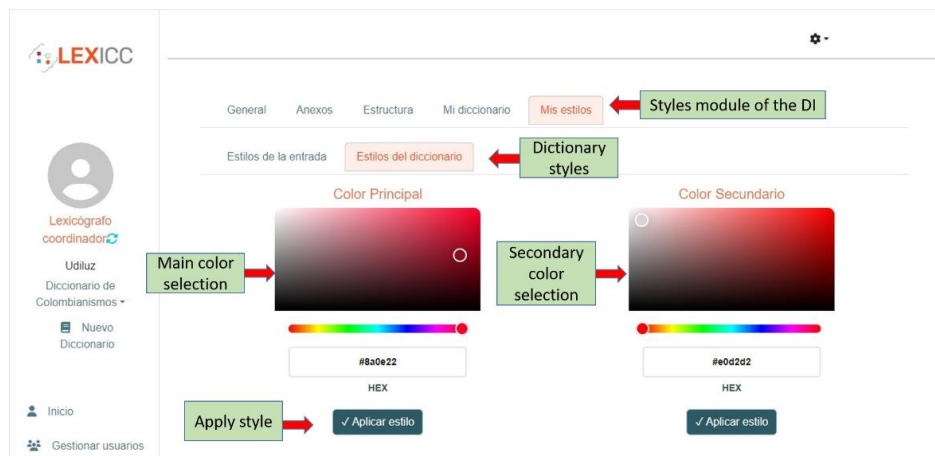


Figure 18: Styles module of the DI: Dictionary styles interface

5.3.2 Dictionary Visualization Interface (VI)

This space is where any user can consult available dictionaries, their features, project information, and contact details of the involved team. Currently, only the DiCol is available, as shown in Figure 19.



Figure 19: VI of the DiCol

The online version of DiCol in LEXICC consists of 5,915 entries. To achieve this transition to LEXICC, the team developer exported the DiCol data from TLex as a MySQL extension file, which was then transformed into JSON format using a script for better data handling. This JSON file loaded successfully into the LEXICC database in MongoDB, where data structure configuration was fundamental to ensure correct visualization. During this configuration, the ID naming the entry components in TLex was identified and collected in MongoDB, with other tags related to position, symbols, and styles (font, colour, text-alignment). These characteristics were implemented in the VI, giving the electronic version of the DiCol its resemblance to the printed version. Also, at the top of this interface, is an alphabetical search and the search box for word lookup, as depicted in Figure 19. Based on Almind's suggestions (2005), electronic dictionaries should display a visible search box, preferably at the top, and legible text, such as in DiCol electronic version. Furthermore, this new format aims to empower our lexicography team to draft entries separately without the constant assistance of a systems engineer. The DiCol editing interface allows lexicographer users to modify the entries' structure, edit existing entries, and create new ones, as shown in Figure 16.

6. Conclusions

The research work of the Instituto Caro y Cuervo has enabled the electronic and printed publication of various lexicographical projects on the languages and varieties of Spanish spoken in Colombia. Although they are meaningful to Colombian linguistic diversity recognition, their dispersion in different formats hampers their access and visibility for interested users, such as students, teachers,

and researchers from the same institution. Therefore, LEXICC emerges as an innovative platform whose primary purpose is to host and manage the lexicographic projects of the ICC. One such venture is the DiCol, a project initially published in print in 2018, and its electronic version hosted on LEXICC promises to become a more accessible resource for different kinds of users, offering the opportunity to delve into the diverse regional varieties of Spanish-spoken in Colombia.

The creation of LEXICC adhered to the software development lifecycle methodology, which includes the planning and organisation of requirements, the design and development of the platform, functionality testing, and eventual deployment. This rigorous process culminated in an alpha version of LEXICC, incorporating the DiCol's database migrated from TLex. LEXICC, with its varied user interfaces and types, promises a user-friendly navigation experience and easy handling of its functionalities. For non-specialist users, LEXICC provides a demo dictionary, where can be created the structure of a basic dictionary. For specialist users, LEXICC offers the Lexicographer Interface, which allows the creation of a dictionary with different styles.

In the long term, the optimization of the platform will allow the Lexicographer Director to work collaboratively with a team by assigning tasks and drafting notes in entries to assess the progress of the lexicographic project. LEXICC users also can make suggestions to the available dictionaries through forums that will be available soon. Moreover, search capabilities will significantly improve within DiCol's electronic version with searches by linguistic labels: part of speech, dialect, register, domain, or attitude. Additionally, the system will offer integrated geolocation services in partnership with ICC's Geographic Information System platform. This integration aims to transform LEXICC into a multifunctional platform (cf. Kruyt 2003) capable of interacting with other electronic linguistic resources. Finally, a comprehensive methodology will be implemented to test the platform with external stakeholders, including ICC staff, DiCol coordinators, and other institutional researchers. The main goal is to ensure that LEXICC will provide a valuable, dynamic resource for Colombian Spanish exploration and comprehension.

Acknowledgments

This work is financed by the Instituto Caro y Cuervo (ICC), an institution adjunct to the Colombian Ministry of Cultures, under the project "ICC 2.0: Technological Integration for Linguistic Preservation and Diffusion Phase 2024". This project is being developed by the line of research in corpus and computational linguistics.

Endnotes

1. <https://saliba.caroycuervo.gov.co/>
2. <https://www.idiomamedico.net/>

3. <https://lexicc.caroycuervo.gov.co/>
4. GitLab is a web-based tool that helps developers manage their code, track issues, and automate the process of testing and deploying applications. It integrates various aspects of software development and operations (commonly referred to as DevOps) into a single platform. For more information, visit <https://about.gitlab.com>.
5. <https://clicc.caroycuervo.gov.co/>
6. <https://alec.caroycuervo.gov.co/sig-alec.php>

References

- Academia Colombiana de la Lengua.** 2012. *Breve diccionario de colombianismos*. Fourth Edition. Bogotá: Academia Colombiana de la Lengua.
- Academia Nacional de Medicina de Colombia.** 2023. *Diccionario Académico de Medicina*. <https://www.idiomamedico.net/>
- Aggarwal, S. and J. Verma.** 2018. Comparative Analysis of MEAN Stack and MERN Stack. *International Journal of Recent Research Aspects* 5(1): 133-137.
- Almind, R.** 2005. Designing Internet Dictionaries. *Hermes* 18(34): 37-54.
- Andoque, H. and J. Landaburu.** 2023. *Diccionario preliminar de la lengua del pueblo gente de Hacha-Andoque del Amazonas*. Bogotá.
- Asociación de Academias de la Lengua Española (ASALE).** 2010. *Diccionario de americanismos*. Madrid: Santillana.
- Atkins, B.T.S. and M. Rundell.** 2008. *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Bernal, J., W. López, D. Moreno and F. Mendieta.** 2020. *Lexicografía electrónica especializada: el caso del diccionario académico de medicina — DIACME*. Bogotá: Instituto Caro y Cuervo.
- Bonilla, J.E. and J. Bernal.** 2020. Modelamiento de una base de datos espacial para el Atlas Lingüístico-Etnográfico de Colombia. *Revista signos* 53(103): 346-368.
- Bonilla, J.E., R. Rubio, A. Llanos, D. Bejarano and J. Bernal.** 2020. Proyecto de digitalización y nuevas perspectivas tecnológicas del "Atlas Lingüístico-Etnográfico de Colombia". Gallego, Á. and F. Roca (Coords.). *Dialectología digital del español*: 13-28. Santiago de Compostela: Universidad Santiago de Compostela, Servicio de Publicaciones = Servizo de Publicacións.
- Cuervo, R.J. and Instituto Caro y Cuervo.** 1998. *Diccionario de Construcción y Régimen de la Lengua Castellana*. Barcelona: Herder.
- Dueñas, G. and D. Gómez.** 2015. Diccionario electrónico sáliba-español: una herramienta interactiva para la documentación de la lengua y de la cultura sáliba. *Forma y Función* 28(2): 49-61.
- Győrödi, C.A., D.V. Dumșe-Burescu, D.R. Zmaranda and R.Ş. Győrödi.** 2022. A Comparative Study of MongoDB and Document-based MySQL for Big Data Application Data Management. *Big Data and Cognitive Computing* 6(2). <https://doi.org/10.3390/bdcc6020049>
- Haensch, G. and R. Werner.** 1993. *Nuevo diccionario de Americanismos. Nuevo diccionario de colombianismos. Vol. 5*. Bogotá: Instituto Caro y Cuervo.
- IDM France.** n.d. *DPS User Manual*. <https://dps.cw.idm.fr/>

- Instituto Caro y Cuervo.** 1983. *Atlas Lingüístico-Etnográfico de Colombia (ALEC)*. Bogotá: Instituto Caro y Cuervo.
- Instituto Caro y Cuervo.** 2014. *Diccionario bilingüe sáliba-español*.
<https://saliba.caroycuervo.gov.co/>
- Instituto Caro y Cuervo.** 2018. *Diccionario de colombianismos*. Bogotá: Instituto Caro y Cuervo.
- Instituto Nacional para Sordos (INSOR) and Instituto Caro y Cuervo.** 2011. *Diccionario básico de la lengua de señas colombiana*. Bogotá: Imprenta Nacional de Colombia.
http://www.insor.gov.co/descargar/diccionario_basico_completo.pdf
- Institute of the Estonian Language.** 2023. *EElex* (version 3.4) [software].
<https://eelex.eki.ee/>
- Kruyt, T.** 2003. Multifunctional Linguistic Databases: Their Multiple Use. Van Sterkenburg, P. (Ed.). 2003. *A Practical Guide to Lexicography*: 194-203. Amsterdam: John Benjamins.
- Leau, Y.B., W.K. Loo, W.Y. Tham and S.F. Tan.** 2012. Software Development Life Cycle AGILE vs Traditional Approaches. *International Conference on Information and Network Technology* 37(1): 162-167.
- Měchura, M.** 2017. Introducing Lexonomy: An Open-Source Dictionary Writing and Publishing System. Kosem, I., C. Tiberius, M. Jakubíček, J. Kallas, S. Krek and V. Baisa (Eds.). 2017. *Electronic Lexicography in the 21st Century: Proceedings of eLex 2017 Conference, Leiden, The Netherlands, 19–21 September 2017*: 662-679. Brno, Czech Republic: Lexical Computing CZ s.r.o.
- Montes, J.J., J. Figueroa, S. Mora and M. Lozano.** 1986. *Glosario Lexicográfico del Atlas Lingüístico-Etnográfico de Colombia (ALEC)*. Bogotá: Instituto Caro y Cuervo.
- Nieto, G.** 2017. *Glosario de Aprendizaje del Español de Colombia*. Bogotá: Instituto Caro y Cuervo.
<https://spanishincolombia.caroycuervo.gov.co/documentos/imagenes/Hecho%20en%20Colombia-glosario.pdf>
- Nieto, G.** 2020. *Corpus léxico del español de Colombia (CorlexCo)*.
<https://clicc.caroycuervo.gov.co/corpus/CorlexCo> [29 may 2024]
- Rozo, N., M.B. Espejo, G. Duarte, D. Guevara and S. Lamprea.** 2020. *Léxico de la Violencia en Colombia 1948–1970*. Bogotá: Instituto Caro y Cuervo.
- Rubio, R. and J. Bernal.** 2019. Corpus Oral del Instituto Caro y Cuervo: reestructuración, diseño y construcción. *Lexis* 43(1): 195-219.
- Rubio, R., A. Luna and N. Solano.** 2023. Corpus Lingüísticos del Instituto Caro y Cuervo (CLICC): una plataforma en línea para el almacenamiento, sistematización y consulta de corpus. *Linguística* 15(2): 89-96.
<https://doi.org/10.21814/lm.15.2.407>
- Saundariya, K., M. Abirami, K.R. Senthil, D. Prabakaran, B. Srimathi and G. Nagarajan.** 2021. Webapp Service for Booking Handyman Using MongoDB, express JS, React JS, node JS. 2021 *3rd International Conference on Signal Processing and Communication (ICPSC), Coimbatore, India, 13–14 May 2021*: 180-183. Danvers, MA: Institute of Electrical and Electronics Engineers (IEEE).
- Sommerville, I.** 2011. *Ingeniería de Software*. Ninth edition. Mexico City: Pearson Education.
- TshwaneDJe HLT.** 2023. *TshwaneLex Suite: Dictionary Compilation Software* (version 15).
<http://tshwanedje.com/tshwanelex/>