TeresIA. Spanish Access Portal to Terminologies and Artificial Intelligence Services

Nava Maroto*1,2*

1 Universidad Politécnica de Madrid, Avda. Complutense, 30, Madrid, E-28040, Spain
2 AETER (Spanish Terminology Association), Spain

Abstract

The TeresIA project aims to improve the creation and management of terminologies in Spanish and Latin American contexts using artificial intelligence. The portal features a metasearch engine for unified access to high-quality terminologies from various projects. Specific functionalities include term extraction tools, expert validation, and user management. A significant aspect involves developing a human-in-the-loop validation service for collaborative terminology management. The subsequent phase involves real-world applications in legal, biomedical, and engineering domains, highlighting the project's impact on information retrieval and scientific communication in Spanish.

Keywords

Terminology management, AI-driven metasearches, human-in-the-loop validation, Spanish scientific communication

1. Introduction

Artificial Intelligence has long been an ally for terminology management. Ever since Nedobity [1] issued his article "Terminology and Artificial Intelligence", these two disciplines have influenced each other.

This contribution presents the latest action carried out in Spain regarding the creation of a common platform for the consultation and utilization of terminology, especially oriented towards its use in artificial intelligence applications. TeresIA, funded by the Secretary of State for Digitalization and Artificial Intelligence (SEDIA) of the Ministry of Digital Transformation of Spain within the framework of the Strategic Project of the New Language Economy and the Recovery, Transformation and Resilience Plan (the so-called *PERTE de la Lengua*), will be developed during the period 2023-2025.

The paper is structured as follows. First, we provide an overview of the background of TeresIA, both at a Spanish and at an international level (section 2). Then, the different modules of TeresIA will be briefly described (section 3). Finally, the focus will be set on the activities in which the Spanish Association for Terminology is particularly involved, namely, terminology validation, from a technical as well as a linguistic perspective (Section 4). Section 5 reflects on future work and the main challenges envisaged.

2. Background of TeresIA

TeresIA is a highly ambitious project to harmonize the terminology of Spanish and the languages of Spain, with a special focus on the development of terminologies to be used in AI applications.

In this section the background of the TeresIA project will be described in detail. First, the efforts carried out at a national level within the Terminesp project are presented (section 2.1). Then, an overview of similar initiatives at an international level that have been inspiring in developing our proposal are outlined (section 2.2).

🆔 0000-0002-0349-7716 (N. Maroto)

2.1. The Terminesp initiative

It is only fair to acknowledge that, long before TeresIA finally saw the light of day at the end of 2023, the efforts to harmonize the terminology of Spanish and the languages of Spain promoted by the Spanish terminology association (AETER) had been numerous [2] [3] [4] [5] [6] [7] [8]. This section summarizes the Terminesp project, the precursor of the current TeresIA project.

Terminesp was an AETER initiative, launched in 2005 from the project designed by M. Teresa Cabré [2]. Its initial objectives were to organize Spanish terminology in Spain; to articulate the organization of Spanish terminology with the terminology of the different autonomous regions with a language other than Spanish (namely Catalan, Basque and Galician); to promote the organization of terminology management in the Spanish-speaking countries, more specifically the countries of Latin America; and, finally, to organize a network that combines the Latin American and peninsular Spanish nodes in a single organization.

To achieve these objectives, three phases and three modules were envisaged. As for the phases, the first phase consisted in the organization of Spanish terminology in Spain. The second stage would articulate Spanish terminology with the terminology organizations of the non-Spanish-speaking autonomous communities. Finally, during the third stage, the terminology of peninsular Spanish would be articulated with the Spanish terminology in Latin America.

Regarding the three modules envisaged for Terminesp, Module 1 would consist of the creation of a terminology access platform and its organization for consultation. The second module would encompass the design and implementation of a terminology sanctioning system through expert committees called Valiter (VALIdación TERminológica, terminological validation). Finally, Module 3 would consist of a linguistic commission for Spanish terminology called COLTE. COLTE stands for Comisión Lingüística para la Terminlogía del Español (Linguistic Commission for Spanish Terminology), and it was convened by the Spanish Academy (RAE) and the Spanish Terminology Association (AETER), with the participation of the Instituto Cervantes, the Fundación del Español Urgente (Fundéu), the European Commission and experts from the universities of Salamanca and Alcalá de Henares in 2006.

2.1.1. Phases and landmarks

In the development of the attempts to fully deploy the Terminesp project, three distinct stages can be identified. During the first actions (2005-2014), the promoting commission, composed of entities such as AETER, the Directorate-General for Translation of the European Commission, the Virtual Center of the Instituto Cervantes, the Foundation "El Español, lengua de traducción", the Iberoamerican Network for Terminology (RITERM) and the Union Latine, carried out key actions. This included the search for institutional support, the creation of the COLTE, the transfer of Spanish standards issued by AENOR (UNE standards) to become a terminology database, the testing of the Valiter module and the publication of a preliminary terminological database containing the terminology in UNE standards on the Wikilengua platform [9].

In the second stage (2014-2018), efforts were made to revitalize the project. A collaboration agreement was established between AETER and the Instituto Cervantes. In 2016, the Instituto Cervantes took over the project, planning the technical and financial design of the terminological platform and Valiter, the creation of a portal with terminology resources, the preparation of the White Paper on Spanish Terminology, the conversion of UNE standards to a database and the revitalization of COLTE, which were actions that eventually could not be fully deployed.

Finally, the third stage (2019-2023) involved ceding the Terminesp database to the Spanish Academy (RAE) and the Spanish Foundation for Science and Technology (FECYT) for its implementation in the Platform to Support Scientific and Technological Communication in Spanish called *Enclave de Ciencia* [10]. The Terminesp project was taken up again to be integrated into a platform for unified access to Spanish terminology. An interest group was formed with the participation of the Instituto Cervantes, the Directorate General for Translation of the European Commission, the Spanish National Research Council (CSIC) and AETER. Work was carried out on a White Paper on Spanish Terminology, a terminology validation system through expert committees and the creation of a new linguistic commission to establish criteria for term formation and loan adaptation in Spanish.

These initiatives, interrupted by the COVID-19 pandemic, were resumed in 2021, with the four partners (Instituto Cervantes, Directorate General for Translation of the European Commission, CSIC and AETER) collaborating to draft a new project. They were strengthened thanks to the Alliance for Spanish in Science and Technology (ALESCYT) promoted by the Ministry of Science and Innovation, FECYT and Instituto Cervantes [11]. The name TeresIA was proposed, reflecting Terminology in Spanish and Artificial Intelligence and paying tribute to Mª Teresa Cabré i Castellví, the first promoter of the project. The project was adapted to the new technological approaches related to Artificial Intelligence and the requirements of the Spanish Secretary of State for Digitalization and Artificial Intelligence (SEDIA). The incorporation of two technological partners, the Barcelona Supercomputing Center (BSC) and the Ontology Engineering Group of the Universidad Politécnica de Madrid (OEG-UPM), was proposed.

At the end of 2023, an agreement was signed between the Spanish National Research Council (CSIC) -acting as project leader- and the Secretary of State for Digitalization and Artificial Intelligence (SEDIA) for the creation of TeresIA, a portal for access to terminologies in artificial intelligence services within the framework of the Strategic Project for Economic Recovery and Transformation on the new economy of language, the so-called PERTE de la Lengua.

2.2. Similar projects that inspire TeresIA

In this section we will refer to different initiatives which might be considered to some extent similar to TeresIA. On the one hand, we will enumerate several terminology portals that have been inspiring (section 2.2.1). On the other hand, we will set our spotlight on the criteria defined for terminology development and adaptation by different institutions for the conceptual validation and linguistic sanction module, which will be described further in section 4 (section 2.2.2).

2.2.1. Terminology portals

Over the years, several projects aimed at harmonizing the access to terminologies have been developed in different European countries. At European level, the first project that comes to mind is EuroTermBank [12] which ended in 2022 after several European-funded projects that were chained together. EuroTermBank is defined as a centralized online termbank of EU and Icelandic languages, interlinked to other terminology banks and resources. It enables exchange of terminology data with existing European terminology databases. EuroTermBank focuses on the harmonisation and consolidation of terminology work in new EU member states, transferring experience from other European Union terminology networks and accumulating competencies and efforts of the accessed countries [12].

Although Eurotermbank aggregates the results of all the thesauri/terminologies included, they are not presented as proper linguistic linked open data (LLOD), as different senses of the same term may appear without being previously disambiguated. At TeresIA the plan is to show previously disambiguated terms.

Other comparable terminology access portal is the one provided by TermCoord [13] where pointers to many other portals and terminological resources are placed together. TermCoord is the terminology coordination unit within the European Parliament's translation department. Again, this resource offers links to other portals, both internal (stemming from European Union institutions) and external, but there is no single access point to the terminology contained in these resources.

Within Spain, the term bases compiled at TERMCAT (the Catalan Center for Terminology) and the guidelines provided would be an example of what TeresIA plans to do for terminological resources in Spanish [14]. However, the glossaries compiled within the Cercaterm search engine available to the public do not support linked data formats, either.

Other similar projects worth mentioning because they have "collected" resources and can be a good starting point for identifying domain terminologies are the following:
- the Linguistic Linked Open Data cloud [15], which is a collaborative effort pursued by several members of the Open Linguistics Working Group (OWLG) to develop a Linked

Open Data (sub-)cloud of linguistic resources. However, most of the resources that appear classified as terminologies are rather thesauri.
- the European CLARIN infrastructure [16], which provides access to digital language datasets.
- the ELRA Catalogue of Language Resources [17], which offers a repository of language resources in the various fields of Human Language Technology (HLT).
- the European Language Grid (ELG) [18], which develops and deploys a scalable cloud platform, providing access to hundreds of commercial and non-commercial Language Technologies for all European languages, including running tools and services as well as data sets and resources.

As we can see, the idea of "aggregating" terminological resources is by no means new. However, none of these very useful repositories or catalogs fully shares objectives with TeresIA.

2.2.2. Terminology validation criteria

One of the issues that TeresIA is most concerned about is terminology validation, both from a conceptual and a linguistic point of view. In order to carry out this validation process, which will be described in more detail in section 4, have been inspiring in previous efforts made by several terminology institutions across different countries.

It is well known that linguistic standardization efforts in those territories with minority languages such as French in Canada or Catalan have been the main breeding ground for the development of criteria for the adoption of new terms, and in particular for the adaptation of borrowings from other languages.

The Office Québequois de la Langue Française (OQLF) issued a set of criteria for the adoption of loan words as early as 1981 [19]. After the experience acquired over the years, this policy was revised in 2007 [20]. In the same vein, TERMCAT has developed a series of guides that help in the adoption of new terms for Catalan [21]. These guides encompass linguistic and methodological criteria. Within the linguistic criteria, detailed guides for the formation of terms with elements derived from Latin and Greek, as well as for the management of loanwords and calques, or the use of acronyms have been issued. All of these will serve as a very good starting point for setting the criteria that will be used by Spanish terminologists during the linguistic sanction process.

We also follow closely the case of France, where a French language enrichment program has been in place for more than fifty years [22]. The current system, instituted by the decree of July 3, 1996 (amended on March 25, 2015), has the primary mission of filling gaps in the French scientific and technical vocabulary, in particular by identifying new concepts that generally appear under foreign names, most often in English, and then creating equivalent terms in French. The project includes a commission for the enrichment of the French language, which is coordinated by the Delegation Générale à la Langue Française et aux Langues de France (DGLFLF). Experts in the scientific and technical fields, as well as representatives from the French Administration, the Académie Française, the Académie des Sciences and standardization bodies (AFNOR), correspondents from linguistic institutions in French-speaking countries and academics specializing in language collaborate in the proposal of new terms. Experts from nineteen professional associations are responsible for proposing the necessary terms to the Enrichment Commission, along with their definitions. Once approved by the Académie française, the terms adopted by the Commission are published in the Official Journal. They are compulsory for use in government departments and institutions, and can serve as a reference for translators and technical writers, and more generally for anyone who wants to be understood by as many people as possible.

Although our approach is more descriptive and less prescriptive than the French government's, we value all these initiatives in order to adopt a protocol that enables the linguistic sanction and the conceptual endorsement of the terminology identified within the corpus compiled for the TeresIA project. The terminology extracted needs to be as reliable and coherent as possible and should comply with the rules governing the formation of words in Spanish.

In the next section we will describe in full the different modules of the TeresIA project.

3. General overview of TeresIA

The Spanish language, with nearly 500 million native speakers, holds international significance and is poised for growth, given the substantial number of people studying it annually [23]. The digitization process and the knowledge economy's advancement create an environment for developing systems that harness information stored in text repositories to enhance public administration, thus improving citizens' quality of life. Working on a language with natural language processing techniques presents opportunities for strengthening the language in science, promoting multilingualism in scientific communication, and retrieving scientific content generated in Spanish, crucial within established European scientific information infrastructures. The strategic position of Spanish globally provides an advantage for fostering the growth of the Spanish data industry and positioning it as a leader in language technologies.

Digital transformation has generated vast textual data in sectors like R&D, healthcare, and law, necessitating systems for efficient data access and diagnostic assistance. Much information is encapsulated in textual data, making it essential to develop systems for classifying, structuring, and retrieving information to utilize institutional and organizational resources effectively.

Terminologies are vital for communication among experts, transcending scientific communication to society through dissemination and translation. The efficiency of scientific content access and reuse in Spanish depends on terminology work and integration into multilingual retrieval systems. This work has commercial implications, especially in linguistic technologies for translation, as well as in the field of specialized translation.

Spain has a rich tradition in terminology research and practice, with notable institutions and groups contributing to the discipline. Efforts by organizations like TERMCAT and academic initiatives highlight the importance of terminology work. The TeresIA project is seen as a meeting point for terminologies in Spain and Latin America, offering developed technologies to various organizations. The project aims to accelerate terminology generation through agile tools supported by language technologies and artificial intelligence, benefiting strategic sectors in the Spanish economy.

3.1. Functionalities of TeresIA

The functionalities of the TeresIA portal are designed to address existing challenges in the creation, reuse, and harmonization of terminologies. The project aims to develop digital tools based on artificial intelligence, language technologies, and data interoperability to establish a common access point for terminologies in Spain and Latin America. The portal aims to enhance the efficiency of creating, expanding, reusing, and applying terminological resources. The specific functionalities include:

1. **Unified Access Portal**: Development of a portal implementing a metasearch engine for unified access to high-quality Spanish and co-official language terminologies. In this first version Spanish will be the starting point, whereas co-official languages will be dealt with in future extensions of the project. This access portal will provide unified access to a vast array of terminological resources, irrespective of whether they support LLOD technology or not.

2. **Term Metasearch Engine**: Retrieval of terms and associated information from various terminological projects, specialized dictionaries, thesauri, etc., previously converted to Linked Data formats and interconnected.
    The main difference between the Unified Access Portal and the Term Metasearch Engine is that, whereas the first relies on already existing terminologies as such, the envisaged Term Metasearch Engine will serve to retrieve terms from existing terminologies interconnected following the guidelines of LLOD.

3. **Terminology Extraction Tools**: Tools for extracting terminologies, adapting them to Linked Data formats, and incorporating them into the metasearch engine based on input corpora.

4. **Validation and linguistic Sanctioning System**: Implementation of a system for expert validation and linguistic sanctioning of terminologies. The portal includes a module for collaborative review and editing by domain experts and linguists. This module will be further described in the next section (section 4).

5. **Application Scenarios for Terminology Generation:** This work package focuses on real-world applications of services generated in major project technical packages. Three terminology application scenarios with distinct goals and uses are outlined: terminology generation in the legal domain, enrichment of existing terminologies in the biomedical domain, and engineering.

4. The expert validation and linguistic sanctioning module of TeresIA

The validation module of TeresIA is conceived as a terminology service that calls for the involvement of experts, both in linguistics and the different domains following the human-in-the-loop model [24]. The service will consist of a user-friendly interface that allows the collaborative management of terminologies created from documentary sources resulting from the work in previous phases of the TeresIA project, and of the links established between created and existing terminologies as a result of the extraction of terms and relationships using AI. This service will aim to ensure the proper management of terminology noise (overlaps, unjustified duplications, etc.), and the treatment of terminological silences, for example, in the absence of a response to a query made in the metasearch engine or to fill gaps in the systematicity and coverage of the terminology detected at the level of ontological relations of a given resource available on the platform. For this purpose, a restricted access management will be defined for specialists in the different areas. Collaborative management will be implemented as a workflow with different levels of interaction that will be defined during the development of the project.

With a similar approach, a service will be enabled so that expert terminologists can carry out the linguistic sanctioning of the terminology, detecting those units that are not well formed in Spanish (from a pan-Hispanic perspective), or that have been poorly adapted, in the case of borrowings from another language. It is essential that the results of this double validation be communicated to society as immediately and as widely as possible, using the channels of the participating institutions and those of collaborating entities, such as the Fundéu.

It is important to emphasize that one of the main objectives pursued by TeresIA is to achieve a high-quality terminology in Spanish. That is why the linguistic resources (textual corpora) incorporated must meet high standards of quality in terms of content and form. Hence, the terminology obtained from the textual corpora must be validated by both field and linguistic experts (sanctioned by experts in terminology) in order to obtain the endorsement that would suggest their inclusion in the terminologies of the metasearch engine.

Moreover, it is worth clarifying that when we talk about generating this double validation cycle, our aim is to incorporate expert human knowledge in the whole process, following the human-in-the-loop approach [24]. This concept has long been used in machine learning as an umbrella term that encompasses different ways in which human expertise can be introduced within the process of using AI for activities such as machine learning or machine teaching. From our perspective, this human interaction is beneficial because it incorporates human knowledge in the process as a way to validate the results of the terminology extraction activities.

On the one hand, experts in different fields validate that the conceptual content is appropriate, and, on the other hand, linguists confirm that the new terms are well-formed from a linguistic point of view. Hence, the aim of this module is not to adopt a prescriptive perspective, to standardize the terminology in Spanish, but rather to ensure that the terms have the best chance of being adopted by the experts, while at the same time complying with the rules of the Spanish language.

The imperative for action transcends the mere accumulation of terms. The focus must be set on assembling a repository of high-quality terms that meet the needs of both human users and machine systems, enabling learning and effective operation. Quality in terminological content is multi-faceted, influenced by factors such as the source (terminological databases, textual corpora) and the endorsement process for newly proposed terms, ensuring their terminological

adequacy from a linguistic and conceptual point of view, as well as from the specialized domain's conventions.

The critical question arises: who, whether an individual or an institution, stands behind this endorsement? This validation process is not just about a stamp of approval; it is about laying the foundation for subsequent dissemination and utilization. Therefore, it must be representative, supported by a diverse array of stakeholders, and descriptive, capturing the nuances of usage and context. Moreover, it should carry either explicit or implicit approval from end-users, reinforcing its operational and functional value.

Creating a robust and effective collaborative validation system demands careful planning and calibration of steps. This involves not only the expert conceptual validation itself but also the subsequent linguistic sanctioning process. Asking the right questions consistently and seeking answers that align with the goals and objectives of the project are paramount. A successful validation system should exhibit certain characteristics, including centralization to ensure consistency and coherence, coordination among stakeholders to streamline the process, sequencing to prioritize tasks and manage resources efficiently, and an effective interplay between technical tools and human expertise to leverage the strengths of both.

Figure 1 shows how the work unfolds in several distinct phases, beginning with the identification of relevant terms, which in TeresIA will involve mostly automated processes depending on the complexity and scope of the domain. This is followed by the collection of terms, often requiring close collaboration with domain experts to ensure comprehensiveness and accuracy. The heart of the process lies in the validation phase, where domain experts scrutinize each term to ensure its accuracy, relevance, and adherence to established criteria in the knowledge field. Subsequently, linguists (terminologists) check that the forms validated by the experts conform the rules of the Spanish language (linguistic sanction), and finally, the sanctioned terms are disseminated to various stakeholders, including experts, the general public, and translators, facilitating their uptake and integration into practice.
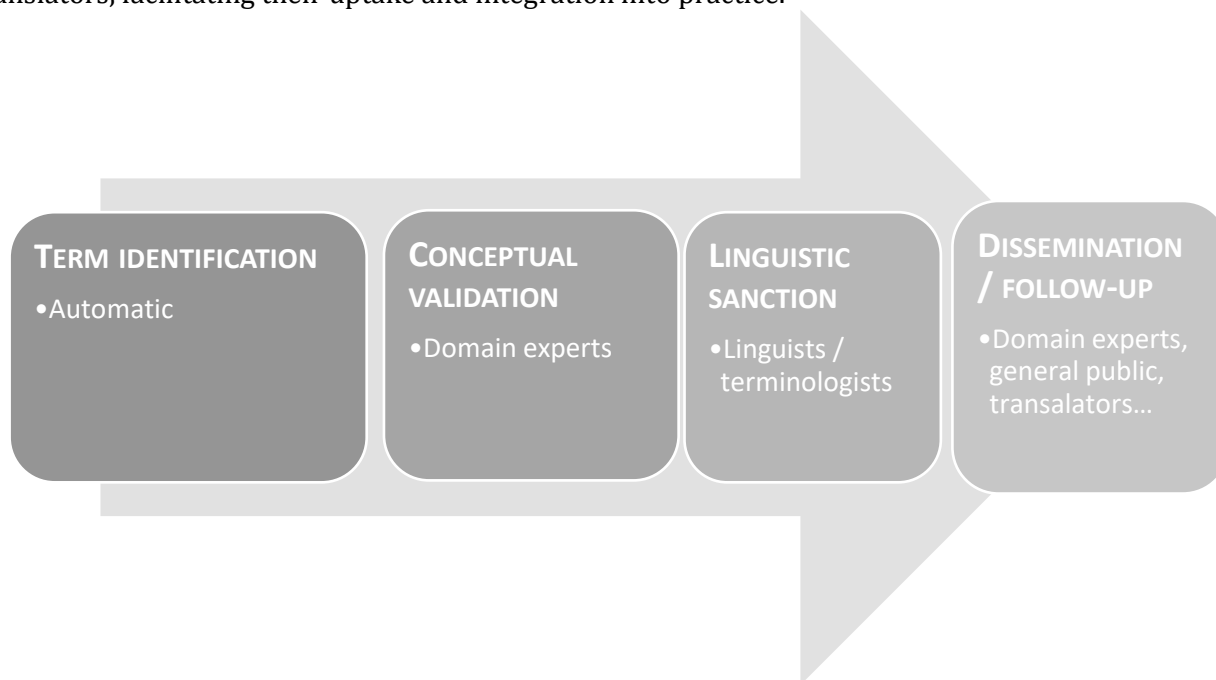

Figure 1: Basic workflow in the double endorsement of terms.

In fostering collaboration between linguists and domain experts, it is essential to adopt a non-directive approach that encourages open dialogue and mutual respect. This involves creating opportunities for scientists to share their expertise, such as by identifying terms and providing input on their linguistic and conceptual validity. By facilitating maximal collaboration and integration of scientific expertise into the project, the collective wisdom of both domain experts and terminologists will result in the advance of terminological research and practice.

The issue about whether the inclusion of domain experts and terminologists within the project is going to be paid or voluntary work is still under discussion. One of the ideas that we are considering is niche sourcing, which has already been used as a valid methodology in the context of terminology [25] [26].

The details of the collaborative platform that will be implemented to carry out this process are still under development. Therefore we can only provide our desiderata, rather than presenting a fully deployed platform. However, we already have the experience of the Valiter commission that worked between 2006 and the mid-2010s [27].

Valiter was a collaborative term validation service open to translators, editors and professionals from all sectors and was based on the constitution of terminology committees by expert field, whose task was to validate the terms received by means of a form. This form was available at a website, which also hosted a wiki, where different mailing lists were managed by fields of expertise, thus facilitating terminology discussion and validation.

Content editing was reserved for users with editor rights (terminologists, translators or specialists who take on this role), but all the edited and archived information (conclusions and previous discussions) were made available to the public.

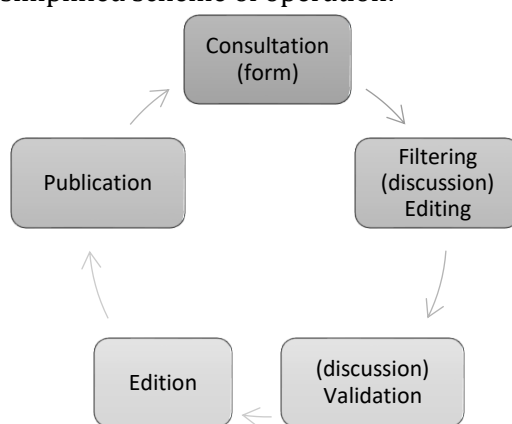Figure 2 represents the simplified scheme of operation:

Figure 2: Scheme of operation of the Valiter collaborative validation service. Adapted from [27].

This experience and the lessons learned from it serve as a valuable starting point to develop a validation platform where the sanctioning protocols can be incorporated seamlessly so that both field experts and linguists can perform their linguistic and conceptual endorsement tasks.

Following are the main steps that will allow us to complete the validation module of TeresIA:

1. Definition of the validation and linguistic sanctioning protocols from a pan-Hispanic perspective that considers the different geographical variants of this language and addresses both the formation of new terms using Spanish rules and the adaptation of terms from other languages.
2. Development of computer support for the implementation of a workflow with several levels of interaction that allows sequential, parallel, synchronous and asynchronous editing, as well as communication and discussion functionalities by experts, providing a friendly collaborative editing environment.
3. Technical and non-technical evaluation (usability, acceptability and impact) of the validation and sanctioning service.
4. Generation of domain terminologies.
5. Creation of terminology resources in those areas of interest for the Spanish Justice and Public Administration, as well as the growing field of national and international mediation and arbitration, and their linking with national and international terminology resources.
6. Enrichment of biomedical terminologies with units extracted and linked semi-automatically from data sets to improve their coverage and the representativeness of terms in peninsular Spanish.

7. Development of structured terminologies in the field of multilingual scientific information retrieval, in the format of the Web of Data, which will allow the systematic indexing of these documents.

As for the data models to be used for data transformation and linking, there is a proposal based on Ontolex-lemon that is currently under consideration [28] that would allow to interlink the resulting terminologies with available resources that are implemented according to the same representation formalism. It is called Termlex and it is a machine-readable format of the Semantic Web improves interoperability between terminological resources in order to capture the information included in authoritative terminological resources such as the various sources of term descriptions and the quality indicators related to terms. Termlex is based on the OntoLex-lemon model that combines the conceptual structure of the SKOS model with the lexical information as modelled in OntoLex-lemon. New classes and properties are defined to cover the specific needs of terminological resources coming from a variety of approaches.

So far, work has begun on the tasks necessary to establish the criteria for the linguistic sanction of terms, which are related to:
1. Lessons learned from the history of the Spanish language regarding the adaptation of foreign terms.
2. Lessons learned from the work done in other nearby languages in which these processes have already been addressed (French in Québec and France, Catalan and Basque).
3. Lessons learned from the work done in languages farther away from Spanish in which these processes have already been addressed such as Dutch or the Nordic languages.
4. Also, the first steps have been taken towards defining with the technical partners of TeresIA the features of the double-factor validation platform, in order to define the terms and accompanying information to be extracted from the textual corpora.

Although there is still a long way to go, we would like to emphasize that TeresIA does not intend to adopt a prescriptive or purely standardizing position, but to apply common sense and bear in mind the knowledge of the facts and the history of the Spanish language. In order to achieve this, the diachronic dimension of the terminology, that is, the historical evolution of terms over the years, will guide our decisions when recommending a term, as well as the actual terms that specialists use on a daily basis. See, in this regard, for example, the approach to the study of chemistry terminology proposed by [29]. After careful consideration, a set of criteria for linguistic sanction and expert validation will be approved and implemented by all TeresIA members with the rest of the implemented tool.

5. Future work

At the present time, the different work teams of the institutions involved in TeresIA are beginning to develop the different functionalities envisaged. Currently, AETER is collaborating with the OEG and the Directorate-General for Translation of the European Commission on the requirements that the validation tool should have in order to introduce this double factor of human endorsement (by experts in the field and by linguists). Once this validation tool is ready, it will be put into practice considering the case studies proposed for the legal, biomedical and engineering fields.

TeresIA seeks to harmonize access to Spanish terminology that is usable by both machines and humans, and that considers the proper use of Spanish. Nevertheless, the challenges to be addressed remain numerous. Among them we could mention the harmonization of terminological information coming from different sources, the difficulty to maintain the project beyond the initial funding, the challenge of ensuring that the expert and linguist groups do not limit their contribution to specific moments but commit themselves in the long term to maintaining collaboration networks, to mention just a few. Technical questions such as automatic detection of terminological neologisms and the use of generative artificial intelligence and large language models for assisting in the definition-writing process remain some of the most challenging issues confronted by TeresIA.

Finally, we are convinced that the generalization of artificial intelligence as an ever-present tool calls for us terminologists to take action in order to establish international networks of

terminologists, domain experts and AI technologists to ensure the quality and validation of terminological resources, which are deemed essential. Projects such as TeresIA project expect to contribute towards establishing accessible quality terminology in the context of digital transformation.

We believe that at last, after almost 20 years of strenuous efforts by different people and institutions, the long-awaited unified portal to Spanish terminology will finally see the light of day.

Acknowledgements

References

[1]   W. Nedobity, "Terminology and artificial intelligence", KNOWL. ORG., vol. 12, n.º 1, 17-19, 1985.

[2]   M.T. Cabré, La terminología del español: organización, normalización y perspectivas, in: Consuelo Gonzalo y Pollux Hernúñez (Eds.), CORCILLUM: estudios de traducción, lingüística y filología dedicados a Valentín García Yebra, Arco/libros, Madrid, 2005, 721-733. ISBN 84-7635-648-X.

[3]   M.T. Cabré, Organizar la terminología del español en su conjunto: ¿realidad o utopía?, in IV Congreso Internacional de la Lengua Española: Cartagena, 2007. Cartagena de Indias. ISBN 978-84-691-5709-1. https://congresosdelalengua.es/cartagena/paneles-ponencias/ciencia-tecnica-diplomacia/cabre-mt.htm.

[4]   M.T. Cabré, Una propuesta de organización de la terminología del español: el proyecto TERMINESP, Donde dice... Boletín de la Fundación del Español Urgente 9 (2007) 4-6. http://www.fundeu.es/files/revistas/DondeDiceN09.pdf.

[5]   M.T. Cabré, La Plataforma TERMINESP, in L. González & P. Hernúñez (Eds.), Traducción: contacto y contagio. Actas del III Congreso "El español, lengua de traducción", 12-14 July 2006, Puebla (México). 2008, ESLEtRA: Bruselas, 255-261. http://cvc.cervantes.es/lengua/esletra/pdf/03/020_cabre.pdf.

[6]   G. Aguado de Cea, AETER y Terminesp. in L. González, P. Hernúñez (Eds.), El español, lengua de traducción para la cooperación y el diálogo. Actas del IV Congreso "El Español, Lengua de Traducción", 8-19 May 2008, Toledo (España). 2010, ESLEtRA, Brussels, pp. 261-265.

[7]   J. García Palacios, Terminología y colaboración. Puntoycoma, nº 170 (abril/mayo/junio/2021), 32-36. https://ec.europa.eu/translation/spanish/magazine/documents/pyc_170_es.pdf.

[8]   N. Maroto & G. Aguado de Cea, Les possibilités des données linguistiques liées ouvertes pour la terminologie et la traduction, in R. Agost Canós & D. Rouz (Eds.), Traductologie, terminologie et traduction, Classiques Garnier; Paris, pp. 63-76.

[9]   Wikilengua, Wikilengua: Terminesp, 2024. URL: https://www.wikilengua.org/index.php/Wikilengua:Terminesp.

[10] Real Academia Española, Enclave de Ciencia, 2024. URL: https://enclavedeciencia.rae.es/contenidos/inicio.

[11] ALESCYT. Alianza por el español en la ciencia y la tecnología, 2023. URL: https://aeter.org/2023/02/28/alescyt-alianza-por-el-espanol-en-la-ciencia-y-la-tecnologia/.

[12] Eurotermbank, Accessible terminology management – for everyone, 2024. URL: https://www.eurotermbank.com/about/.

[13] Termcoord, Terminology Coordination, European Parliament, 2024. URL: https://termcoord.eu/terminology-websites/.

[14] TERMCAT, TERMCAT, centre de terminologia, 2024. URL: https://www.termcat.cat/ca.

[15] Insight Centre for Data Analytics at National University of Ireland Galway, Linguistic Linked Open Data cloud, 2018. URL: https://linguistic-lod.org/llod-cloud.

[16] European Research Infrastructure Consortium (ERIC), Clarin Virtual Language Observatory, 2024. URL: https://www.clarin.eu/content/data.

[17] European Language Resources Association (ELRA), ELRA catalogue, 2018. URL: https://catalogue.elra.info/en-us/.

[18] European Language Grid Consortium, European Language Grid (ELG), Release 3, 2024. URL: https://live.european-language-grid.eu/catalogue/?page=1.

[19] Office de la Langue Française (OLF), Enoncé d'une politique rélative à l'emprunt de formes linguistiques étrangères, Terminogramme, 7/8 (1981), 2-5.

[20] Office Québequois de la Langue Française (OQLF), Politique de l'emprunt linguistique, Québec, Office Québecois de la Langue Française, 2007.

[21] TERMCAT, Criteris terminologics, 2024. URL: https://www.termcat.cat/ca/recursos/criteris

[22] Ministère de la Culture de France, Franceterme. Le dispositif d'enrichissement de la langue française, 2024. URL: https://www.culture.fr/FranceTerme/Le-dispositif-d-enrichissement-de-la-langue-francaise.

[23] Instituto Cervantes, El español en el mundo. Anuario del Instituto Cervantes, 2023. URL: https://cvc.cervantes.es/lengua/anuario/anuario_23/default.htm.

[24] E. Mosqueira-Rey, E. Hernández-Pereira, D. Alonso-Ríos et al, Human-in-the-loop machine learning: a state of the art. Artif Intell Rev 56, 3005–3054 (2023). doi: https://doi.org/10.1007/s10462-022-10246-w.

[25] A. Cox, K. Kerremans, R. Temmerman, Niche sourcing and transexplanations for the enhancement of doctor-patient comprehension in multilingual hospital settings, in: G. Aguado de Cea, N. Aussenac-Giles (Eds.), Proceedings of the 10th International Conference on Terminology and Artificial Intelligence TIA, 2013, pp. 33-36.

[26] J. Enqvist, T. Onikki-Rantajääskö, K. Pitkänen-Heikkilä, Terminology work as open, communal and collaborative crowdsourcing practice of academic communities. Terminology, 27(1), 2021: 56-79. doi: https://doi.org/10.1075/term.00058.enq.

[27] L. González, La red de validación terminológica Valiter, puntoycoma 121 (2011): 13-16.

[28] P. Martín-Chozas, T. Declerck, E. Montiel-Ponsoda, V. Rodríguez-Doncel, Representing terminological data in the Semantic Web. A proposal based on OntoLex-lemon. Terminology. doi: https://doi.org/10.1075/term.22037.mar.

[29] C. Garriga Escribano, The Language of Chemistry in the Romance Languages, Oxford Research Encyclopedia of Linguistics, https://doi.org/10.1093/acrefore/9780199384655.013.475.