

# Teaching Administrative Law through its Textual Genres: Compilation, Annotation and Corpus Analysis

Sara Pistola Grille<sup>1</sup>

<sup>1</sup> Universidad Nacional de Educación a Distancia (UNED), Senda del Rey, 7, 28040, Madrid, Spain

## Abstract

Public administration communicates with citizens through a series of textual genres with specific linguistic characteristics that can be difficult to write. In this context, this PhD thesis proposes to create, annotate and analyse a corpus of textual genres in the field of public administration in order to obtain a model structure with information on the prototypical sections, titles and contents of each textual genre. It will also integrate representative phraseology that can be used to express these contents adapted to plain language. These model structures will be integrated into the arText system, an assisted editor based on Natural Language Processing (NLP) strategies that helps to write administrative texts in plain language. Finally, the automatic system will be used in a didactic proposal that integrates the teaching of Administrative Law, academic writing and the use of Information and Communication Technologies (ICT) in the university context.

## Keywords

Corpus linguistics, administrative discourse, higher education, textual genres

## 1. Justification of the research

As a specialised discourse, the discourse of the administration domain has specific linguistic characteristics such as: the use of archaic terms and structures, the utilisation of complex syntactic constructions, the excessive length of sentences and paragraphs, etc. [1], rendering this discourse difficult for citizens to understand [2]. In this sense, two recent reports [3, 4] analyse the clarity with which the Spanish public administration communicates with citizens and, in particular, with vulnerable groups, and determine that those public institutions tend to use unclear language. As a result, an international trend has emerged to promote the use of plain language by public institutions.

Furthermore, administrative texts can be difficult to write for public administration employees, as they have to take into account specific linguistic characteristics that vary according to the genre and the receiver [5]. Moreover, these professionals are often not trained in linguistics or in academic and professional writing. In fact, a recent publication [6] asked a group of students at a Spanish university whether they receive training in academic writing. The results show that the students perception is that the training received on this subject is insufficient and inadequate.

In this context, the research project *Tecnologías de la Información y la Comunicación para la e-Administración: hacia la mejora de la comunicación entre Administración y ciudadanía a través del lenguaje claro* (TIC-eADMIN) is being developed. This project is carried out in the *Departamento de Filologías Extranjeras y sus Lingüísticas* of the *Facultad de Filología* of the UNED and it is funded by the Spanish Ministry of Science and Innovation. Its main objective is to develop, in the context of e-Administration, an online technological tool to assist public employees in drafting administrative texts in Spanish, in order to contribute to ensuring that the texts received by citizens are written in plain language.

In order to achieve the main objective of this project, three specific objectives have been established:

---

Doctoral Symposium on Natural Language Processing from the PLN.net network 2022 (RED2018-102418-T), 21-23 September 2022, A Coruña, Spain.

EMAIL: [spistola@flog.uned.es](mailto:spistola@flog.uned.es)

ORCID: 0000-0002-6011-8639



© 2022 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).  
CEUR Workshop Proceedings (CEUR-WS.org)

1. Investigate the possibilities of NLP to automatically detect linguistic phenomena in administrative texts that interfere with plain language writing.
2. Analyse and systematise the prototypical linguistic features of some specific textual genres in the field of public administration.
3. Design and implement an ICT tool that helps public employees to write texts in plain language addressed to the citizens, in order to integrate it into the technological infrastructure of the arText system<sup>2</sup>, an editor based on NLP strategies that assists in the writing of texts in specialised fields and texts in plain language.

More information on the project can be found on its website<sup>3</sup>. Within the framework of this project, I have been awarded a predoctoral training contract *Formación de Personal Investigador* (FPI) to collaborate in the achievement of specific objective 2. Thus, my PhD thesis intends to analyse a corpus of textual genres in the field of public administration. The linguistic information extracted from this analysis will be integrated into the arText system. In turn, this technological tool will be used in the design of a didactic proposal that integrates the teaching of Administrative Law, academic writing and the use of ICT in the university context.

In section 2, work related to this research is reviewed. In section 3, the thesis proposal is described in detail, with its corresponding objectives and hypotheses. In section 4 the methodology is explained. In section 5, some elements of the research that could be discussed are specified.

## 2. Background and related work

The public administration communicates continuously with citizens in the context of several situations, such as the publication of open calls, and the use of a service or the consultation of an institutional web portal by a citizen. In this context of communication, a series of prototypical textual genres of the specialised field of the Administration are used, constituting the main communication tools between citizens and the public administration [7].

Over the last few decades, a number of studies have been published regarding texts in the field of administration. These publications are mostly of a theoretical-descriptive nature, that is to say, they list, classify, describe and characterise the prototypical texts in the field of public administration, and highlight the need to modernise the language they use. To this end, some studies advocate the use of corpus analysis [8].

The corpus analysis carried out in these works usually focuses on the lexical and grammatical characteristics of the texts. As an example, the main characteristics of administrative discourse are established through the analysis of a series of texts extracted from the *Boletín Oficial del Estado* (BOE) [9]. More recently, some reports have also focused on citizens' perception of the clarity of texts. For example, two recent reports [3, 4] analyse a corpus of texts issued by different public institutions and determine their level of clarity. However, the organisational scheme of the text or superstructure [10, 11], the contents (also known as moves) [12], and representative phraseology are not usually examined in depth.

In recent years, there has also been a growing number of studies on the writing of textual genres in the legal-administrative field at university, especially in Latin America [13, 14, 15]. However, in practice, these proposals have limited application in Spain, as it is difficult to integrate them into the curriculum of Administrative Law courses. Moreover, they do not usually integrate the use of ICT.

Finally, at the professional level, public employees have various style manuals [1, 7, 16] and guides on plain communication [17, 18, 19] published by their institutions. The aim of these publications is to provide public employees with a tool that helps them to improve the texts addressed to citizens, by making them clearer, correcting grammatical and spelling mistakes, and unifying criteria and structural models. However, consulting these manuals is a great investment of time and effort for public employees, who, moreover, as already mentioned, do not usually have previous training in either linguistics or specialised text writing.

Hence, the development of tools and resources based on NLP regarding the drafting of administrative texts in plain language could contribute to facilitate the writing task of public employees,

---

<sup>2</sup> <http://sistema-artext.com/>.

<sup>3</sup> <http://iriadacunha.com/FundacionBBVA2015/es/ministerio2019-2022/>.

and, ultimately, to improve written communication between the Administration and the citizens. However, this topic is still little explored. In Spanish, there are currently two technological tools for writing administrative texts in plain language. Firstly, there is the arText claro system<sup>4</sup>, an assisted editor that detects issues in the text written by the user that interfere with the most common plain language recommendations, and offers information to make the text clearer and simpler [20]. Secondly, the beta version of the Clara tool<sup>5</sup> is a textual clarity measurement system that allows the user to test the clarity of texts according to nine evaluation metrics [21]. Both tools are available online free of charge.

### 3. Description of the proposed research

The general objectives of this PhD thesis are: firstly, to analyse the prototypical linguistic characteristics of a corpus that includes five administrative textual genres written by the Spanish public administration and addressed to citizens, and secondly, to design and evaluate a didactic proposal for teaching Administrative Law through its textual genres that integrates the use of ICT and can be used in a university context of self-learning and online teaching.

To achieve these general objectives, eight specific objectives are set out. The specific objectives related to the first general objective are:

- To list all the textual genres in the administrative field and classify them according to the sender-receiver parameter into: written by the Administration and addressed to the citizens; written by the Administration and addressed to the Administration itself; and written by the citizens and addressed to the Administration.
- To select five textual genres written by the public administration and addressed to the citizens based on the following criteria: frequency and difficulty of writing by public employees; frequency and difficulty of writing by university students of Administrative Law; and amount of free text that the sender needs to write.
- To compile a corpus of texts of the Spanish public administration that belong to the five administrative textual genres previously selected.
- To establish the structure and prototypical contents of those textual genres based on a corpus analysis (textual level).
- To extract prototypical phraseology that can be used to write the contents of those textual genres through a corpus analysis and to adapt it to plain language (lexical level).
- To elaborate a model structure that contains the linguistic information extracted from the corpus analysis at textual and lexical level in order to integrate it in arText and to use it in the design of a didactic proposal for teaching Administrative Law through its textual genres.

The specific objectives related to the second general objective are:

- To investigate on how to integrate in a cohesive manner the use of ICT, theoretical contents of Administrative Law, and a proposal on writing textual genres that takes into account the linguistic analysis carried out and the clear writing guidelines, in order to elaborate the didactic proposal mentioned above.
- To evaluate the didactic proposal that has been developed through its implementation in the university context of the UNED's online studies of Administrative Law.

This work is based on the hypothesis that the textual genres of the Spanish public administration have a particular structure and contents that can be used as a basis to articulate NLP tools that help professionals and students in the administration field to structure and write texts in a simpler and clearer way. In addition, administrative textual genres can be used to design didactic proposals for teaching Administrative Law at the university level.

### 4. Methodology

The methodology of this PhD thesis includes the following stages:

---

<sup>4</sup> <http://sistema-artext.com/lenguaje-claro>.

<sup>5</sup> <https://clara.comunicacionclara.com/>.

- Stage 1: Classification of the administrative textual genres according to the sender and receiver.
- Stage 2: Selection of five administrative textual genres written by the public administration and addressed to the citizens.
- Stage 3: Compilation of the corpus of texts that will be analysed.
- Stage 4: Selection of linguistic features for corpus analysis.
- Stage 5: Annotation of the corpus with the ATLAS.ti software.
- Stage 6: Linguistic characterisation of each textual genre based on the results of the corpus annotation.
- Stage 7: Elaboration of a didactic proposal for teaching Administrative Law through its textual genres.
- Stage 8: Evaluation of the didactic proposal.

The stages of the methodology are summarised in Figure 1:

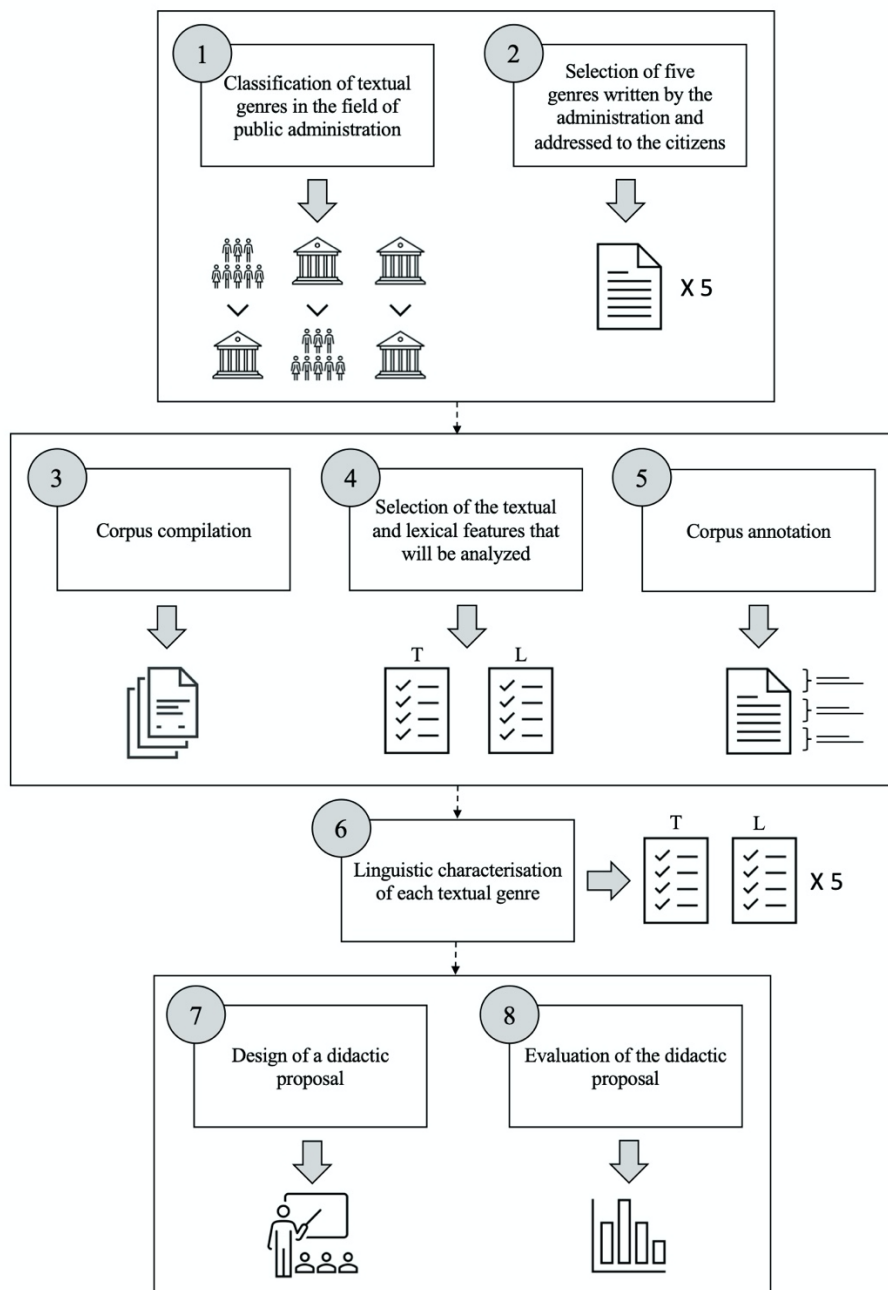


Figure 1: Stages of the methodology

Stages 1-6 have already been completed. As for stage 1, a proposal has been developed to enumerate and designate prototypical texts in the field of administration. This proposal has been adapted to the concept of textual genre, classified according to their sender and receiver, systematic, updated, developed taking into account the current legislation, and developed in an interdisciplinary linguistic-legal context [22].

In relation to stage 2, two surveys were carried out among the employees of the Madrid City Council and the students of the degrees *Grado en Derecho* and *Grado en Ciencias Jurídicas de las Administraciones Públicas* of the UNED. The results were analysed in relation to the following parameters: frequency and difficulty of writing administrative textual genres by public employees; frequency and difficulty of writing by university students of Administrative Law; and amount of free text that needs to be written by the issuer. Based on the results of the survey, five administrative textual genres were selected to be the focus of the research: contract, inspection report, resolution published on institutional portals, administrative procedure resolution sent to individual citizens, and requirement.

For stage 3, a corpus of 100 administrative texts (20 for each selected textual genre) has been compiled. The texts corresponding to the genres that are made publicly available on the Internet were compiled through the search engines of institutional portals. On the other hand, the corpus of texts addressed to private citizens was compiled with the collaboration of the Madrid City Council. The corpus is not made publicly available for confidentiality reasons.

With regard to stage 4, the textual (sections, titles and contents) and lexical (phraseology) features that will be used in the analysis of the corpus have been selected on the basis of the methodology used in previous work related to the development of arText [23, 24].

As regards stage 5, the corpus was manually annotated using the ATLAS.ti tool. Thus, an annotated corpus of administrative texts was obtained with the indication of the sections, titles, contents and representative phraseology to express the contents. Figure 2 shows an example of the annotation of one of the corpus texts using ATLAS.ti.

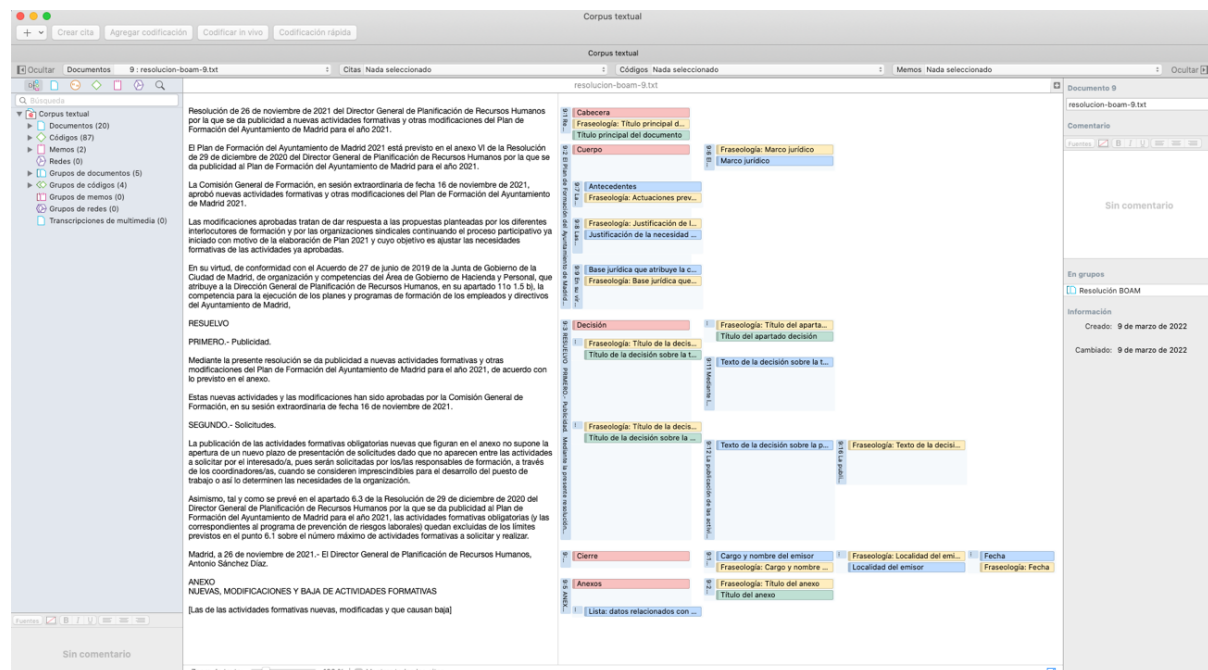


Figure 2: Example of the annotation

In relation to stage 6, the frequency with which textual features appear in the corpus has been counted and those which appear with a frequency > 50 % have been established as prototypical of each genre, following the methodology proposed in previous studies on arText development [23, 24]. Optional textual features (between 40% and 50%) are also included. In addition, phrases that are used in the corpus to express the prototypical contents of the genre and that may be useful for the writer have been selected and adapted to the most frequent plain language recommendations for Spanish [25]. As a result, a model structure with the prototypical sections, contents and phraseology of each textual genre has

been obtained [26, 27]. Table 1 contains an extract of the model structure of the resolution published on institutional portals.

**Table 1**

Fragment of the model structure of the resolution published on institutional portals: section *Cuerpo*

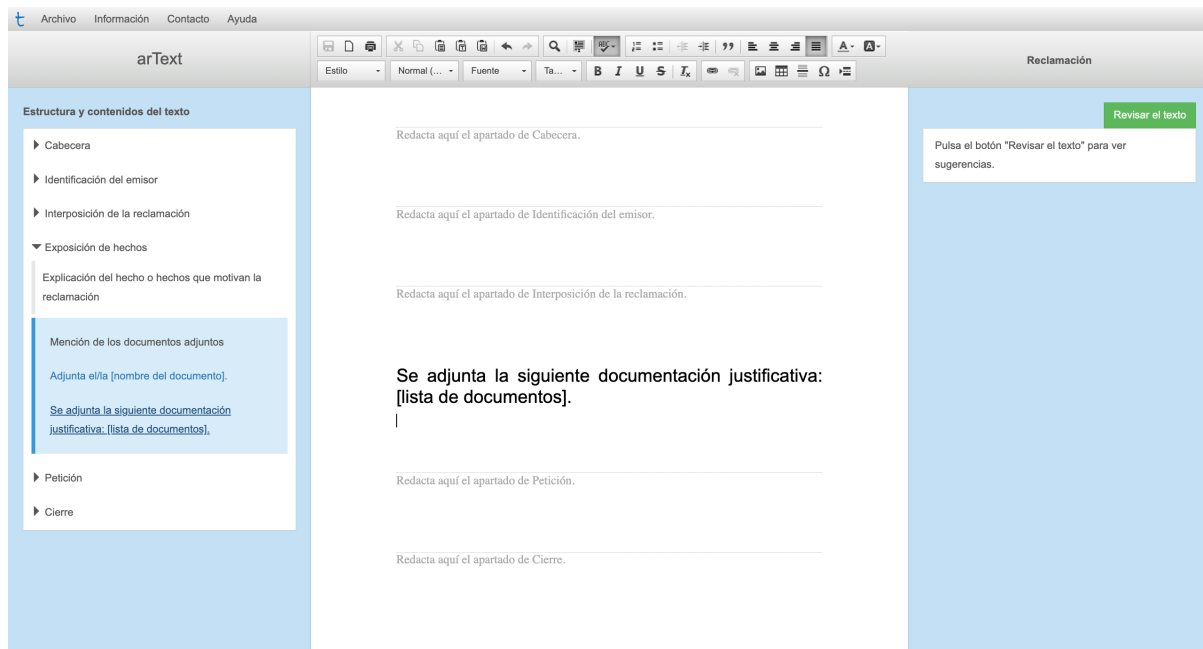
Section	Contents	Phraseology
Cuerpo	Marco jurídico	<ul style="list-style-type: none"> <li>- Según los artículos [nº] y siguientes del/de la [normativa aplicable: ley/reglamento/etc.], de [fecha], en adelante, [sigla por la que se conoce la normativa aplicable],</li> <li>- El/La [normativa aplicable: ley/reglamento/etc.] (en adelante, [sigla por la que se conoce la normativa aplicable]) señala en su artículo [nº] que "[cita textual]".</li> <li>- El/La [artículo o parte de la norma] de la [normativa aplicable: ley/reglamento/etc.], que entró en vigor el [fecha], establece</li> </ul>
	Antecedentes	<ul style="list-style-type: none"> <li>- Se ha concluido el procedimiento y la Comisión de Valoración ha dictado la propuesta de resolución del concurso.</li> <li>- La Comisión de Valoración se nombró conforme a la base [nº] de las Bases Generales de la convocatoria de [temática de la convocatoria].</li> <li>- Se confirma que se han llevado a cabo los procesos que se prevén en las bases y en las disposiciones supletorias.</li> </ul>

The stages of the methodology that remain to be completed are the 7th and 8th. With reference to stage 7, it has been planned the design of a didactic proposal for teaching Administrative Law through its textual genres that integrates academic writing and the arText system as an ICT tool. This proposal will be carried out within the framework of the *Trabajo de Fin de Grado* (TFG) course for students of the degrees *Grado en Derecho* and *Grado en Ciencias Jurídicas de las Administraciones Públicas* of the UNED. Finally, in stage 8, it is planned to evaluate this didactic proposal through a survey addressed to the aforementioned students.

## 5. Specific issues of research to be discussed

In the last few months, the compilation, annotation and analysis of a corpus of texts in the field of public administration has been completed. Thanks to this analysis, five model structures were obtained with linguistic information on the following genres: inspection report, contract, resolution addressed to private citizens, resolution published on institutional portals and requirement.

In the framework of the TIC-eADMIN project, the linguistic information extracted from the corpus analysis I have carried out in my PhD thesis will be used to extend the functionalities of the arText system. Thus, in addition to helping users to make their administrative texts clearer through linguistic recommendations, the system will also provide information on the sections, titles and contents of five text genres in the field. It will also allow users to incorporate into their texts phrases in plain language that are used to express those contents. Figure 3 includes a preview of what this functionality of the arText system will allow the user to do.



**Figure 3:** Preview of the arText functionality about the structure and contents of the text

Within the framework of my PhD thesis, I am going to develop a didactic proposal for teaching Administrative Law through its textual genres in which university students use the arText system. Subsequently, the didactic proposal will be evaluated. Hence, it is open to debate whether it is possible to teach Administrative Law contents through a didactic proposal on academic writing of specialised texts in which, in addition, ICTs are used. It is also worth discussing how to effectively integrate an NLP tool within the framework of teaching Law at university level. Furthermore, suggestions for evaluating the didactic proposal could be discussed. These suggestions would take into account statistical assessment and feedback from the students involved in the study.

## 6. Acknowledgments

This work stems from a *contrato de formación predoctoral FPI*, funded by the Ministry of Science and Innovation in the 2019 call of the *Subprograma Estatal de Formación* of the *Programa Estatal de Promoción del Talento y su Empleabilidad en I+D+i*, in the framework of the *Plan Estatal de Investigación Científica y Técnica y de Innovación 2017-2020* (predoctoral reference PRE2019-088419). The contract was awarded to Sara Pistola Grille, who is carrying out her PhD thesis as a trainee researcher at the *Escuela Internacional de Doctorado* of the *Universidad Nacional de Educación a Distancia (UNED)*, "*Programa de Doctorado en Filología. Estudios lingüísticos y literarios*", *Departamento de Filologías Extranjeras y sus Lingüísticas, Facultad de Filología, UNED*. The PhD thesis is co-directed by Iria da Cunha and Susana Viñuales. The contract is associated with the research project entitled "*Tecnologías de la Información y la Comunicación para la e-Administración: Hacia la mejora de la comunicación entre Administración y ciudadanía a través del lenguaje claro (TIC-eADMIN)*", funded by the Ministry of Science and Innovation in the 2018 call of *Proyectos I+D del Subprograma Estatal de Generación de Conocimiento* (reference PGC2018-099694-A-I00), and developed in the *Departamento de Filologías Extranjeras y sus Lingüísticas* of the *Facultad de Filología* of the UNED, in the framework of the research group ACTUALing, with the collaboration of the group IULATERM (IULA-UPF). Within the framework of this project, a "*Protocolo de colaboración en investigación científica en materia de comunicación clara*" has been signed with the Madrid City Council.

## 7. References

- [1] Ministerio de Administraciones Públicas (MAP), Manual de estilo del lenguaje administrativo, 6th. reprint, Ministerio de Administraciones Públicas, Madrid, Spain, 1997.
- [2] Ministerio de Justicia, Informe de la Comisión de modernización del lenguaje jurídico, Ministerio de Justicia, Madrid, Spain, 2011. URL: [https://www.administraciondejusticia.gob.es/paj/PA\\_PAGAJGenerico/download/Informe%20de%20la%20Comisi%C3%B3n%20de%20Modernizaci%C3%B3n%20del%20Lenguaje%20Jur%C3%ADdico.pdf?idFile=41ae8714-a3a7-462e-b496-6da3ffcb7421](https://www.administraciondejusticia.gob.es/paj/PA_PAGAJGenerico/download/Informe%20de%20la%20Comisi%C3%B3n%20de%20Modernizaci%C3%B3n%20del%20Lenguaje%20Jur%C3%ADdico.pdf?idFile=41ae8714-a3a7-462e-b496-6da3ffcb7421).
- [3] Prodigioso Volcán, ¿Habla claro la Administración pública? Prodigioso Volcán, Madrid, Spain, 2020. URL: <https://comunicacionclara.com/lenguaje-claro/prodigioso-volcan-habla-claro-la-administracion.pdf>.
- [4] Prodigioso Volcán, ¿Habla claro la Administración a los públicos vulnerables? Prodigioso Volcán, Madrid, Spain, 2021. URL: <https://comunicacionclara.com/habla-claro-la-administracion/>.
- [5] M. T. Cabré, La terminología: Representación y comunicación. Elementos para una teoría de base comunicativa y otros artículos, 1st. ed., Institut de Lingüística Aplicada (IULA), Universitat Pompeu Fabra, Barcelona, Spain, 1999.
- [6] A. N. Romero, M. N. Álvarez, Representaciones sociales de los estudiantes universitarios de grado sobre la escritura académica, Íkala, Revista de Lenguaje y Cultura 24(1) (2019) 103–118. <https://doi.org/10.17533/udea.ikala.v24n01a05>.
- [7] Ministerio para las Administraciones Públicas (MAP), Manual de documentos administrativos, 3rd. ed., Editorial Tecnos, Madrid, Spain, 2003.
- [8] M. Lorente, La lingüística de corpus para el estudio del discurso especializado de los ámbitos jurídico y administrativo, in: M. Eurrutia (Coord.), El lenguaje jurídico y administrativo en el ámbito de la extranjería. Estudio multilingüe e implicaciones culturales, Peter Lang, Berna, Switzerland, 2016, pp. 167-193.
- [9] L. Calvo, Introducción al estudio del lenguaje administrativo. Gramática y textos, Editorial Gredos, Madrid, Spain, 1980.
- [10] T. A. Van Dijk, Text and context: Explorations in the semantics and pragmatics of discourse, 1st. ed., Longman, London, United Kingdom, 1977.
- [11] T. A. Van Dijk, La ciencia del texto: un enfoque interdisciplinario, 1st. ed., Paidós Comunicación, Barcelona, Spain, 1989.
- [12] J. M. Swales, Genre analysis: English in academic and research, 1st., ed., Cambridge University Press, Cambridge, United Kingdom, 1990.
- [13] C. Poblete, L. Arenas, A. Córdova, Alfabetización académica en derecho: Experiencia y resultados de un programa de escritura, Revista Pedagogía Universitaria y Didáctica del Derecho 7(2) (2020) 23–50. URL: <https://pedagogiaderecho.uchile.cl/index.php/RPUD/article/view/60562>.
- [14] P. Meza, F. González-Catalán, A. Pastén, M. Barahona, Clases textuales de la formación en derecho: Descripción y orientaciones para su enseñanza, Revista Pedagogía Universitaria y Didáctica del Derecho 7(1) (2020) 63–90. URL: <https://pedagogiaderecho.uchile.cl/index.php/RPUD/article/view/56926>.
- [15] C. De Cucco, El lenguaje jurídico como materia obligatoria, Academia. Revista sobre enseñanza del Derecho 18(35) (2020) 69–93. URL: [http://www.derecho.uba.ar/publicaciones/rev\\_academia/revistas/35/el-lenguaje-juridico-como-materia-obligatoria.pdf](http://www.derecho.uba.ar/publicaciones/rev_academia/revistas/35/el-lenguaje-juridico-como-materia-obligatoria.pdf).
- [16] R. Sarmiento, Manual del Lenguaje Administrativo del Ayuntamiento de Madrid, Ayuntamiento de Madrid, Madrid, Spain, 2007. URL: <https://www.madrid.es/UnidadesDescentralizadas/Calidad/Publicaciones/Ficheros/ManualLenguajeAdministrativo1.pdf>.
- [17] E. Montolío, M. Tascón (Dirs.), Guía Práctica Comunicación Clara, Ayuntamiento de Madrid, Madrid, Spain, 2017. URL: <https://www.madrid.es/UnidadesDescentralizadas/Calidad/LenguajeClaro/ComunicacionClara/Documentos/GuiaPracticaCCLara.pdf>.



- [18] Gobierno de Aragón, Manual de estilo de lenguaje claro. Criterios para la redacción de textos claros en internet, Gobierno de Aragón, Aragón, Spain, 2020a. URL: <https://www.aragon.es/documents/20127/6649155/Manual+de+estilo+de+lenguaje+claro.PDF>.
- [19] Gobierno de Aragón, Guía de buenas prácticas de Comunicación Clara. Aplicación a los contenidos web del Gobierno de Aragón, Gobierno de Aragón, Aragón, Spain, 2020b. URL: <https://www.aragon.es/documents/20127/6649155/Guia+de+buenas+pr%C3%A1cticas+de+comunicaci%C3%B3n+clara.PDF>.
- [20] I. Da Cunha, Un redactor asistido para adaptar textos administrativos a lenguaje claro, *Procesamiento del Lenguaje Natural* 64 (2022). To appear.
- [21] C. Torrijos, S. Oquendo, ¡Hola! Soy clara y mido la claridad de tu texto, *Archiletras Científica* 6(1) (2021) 119–133.
- [22] S. Pistola, S. Viñuales-Ferreiro, Una clasificación actualizada de los géneros textuales de la Administración pública española, *Revista de Lengua i Dret* 75 (2021) 181–203. <https://doi.org/10.2436/rld.i75.2021.3587>.
- [23] I. Da Cunha, M. A. Montané, A corpus-based analysis of textual genres in the administration domain, *Discourse Studies* 22(1) (2020) 3–31. <https://doi.org/10.1177/1461445619887538>.
- [24] I. Da Cunha, M. A. Montané, B. Fisas, M. Á. Escobar, Comparing textual genres in Spanish: the case of the tourism domain, *Ibérica* 42 (2021) 163–190. URL: <https://revistaiberica.org/index.php/iberica/article/view/603/193>.
- [25] I. Da Cunha, M. Á. Escobar, Recomendaciones sobre lenguaje claro en español en el ámbito jurídico-administrativo: análisis y clasificación, *Pragmalingüística* 29 (2021) 129–148. <https://doi.org/10.25267/Pragmalinguistica.2021.i29.07>.
- [26] S. Pistola, I. Da Cunha, La resolución administrativa publicada en portales web institucionales: un análisis basado en corpus, *Revista de Estudios de Lingüística. Universidad de Alicante (ELUA)* 38 (2022). To appear.
- [27] S. Pistola, Análisis de un corpus de géneros textuales prototípicos de la Administración, in: I. Da Cunha (Ed.), *Lenguaje claro y tecnología en la Administración*, Comares, Granada, Spain. To appear.