

Una aplicación tecnológica que ayuda a la ciudadanía a escribir textos a la Administración pública

A technological application that helps citizens write texts to the Public Administration

Iria da Cunha 

Universidad Nacional de Educación a Distancia (UNED)

iriad@flog.uned.es

Resumen

En este trabajo se presenta una aplicación tecnológica gratuita y en línea que ayuda a la ciudadanía a escribir textos dirigidos a la Administración pública. Concretamente, ayuda a redactar cinco géneros textuales: alegación, carta de presentación, queja, reclamación y solicitud. La aplicación tiene forma de editor de textos e incluye tres módulos para: I) estructurar y añadir contenidos en el texto, II) corregirlo ortográficamente y darle formato, y III) obtener sugerencias de mejora sobre aspectos léxicos y discursivos. Integra diferentes herramientas de Procesamiento del Lenguaje Natural (PLN), como un analizador morfosintáctico y un segmentador discursivo. Las evaluaciones *data-driven* y *user-driven* realizadas ofrecen resultados positivos.

Palabras clave

administración pública, ciudadanía, géneros textuales, asistente a la redacción en español

Abstract

This article presents a free and online technological application that helps citizens write texts addressed to the Public Administration. Specifically, it helps to draft five textual genres: allegation, cover letter, letter of complaint, claim and application. The technological application is a text editor that includes three modules: I) structure and contents of the text, II) spelling and format correction, and III) suggestions on vocabulary and discourse. It integrates different Natural Language Processing (NLP) tools, such as a morphosyntactic tagger and a discourse segmenter. The data-driven and user-driven evaluations performed show positive results.

Keywords

public administration, citizenship, textual genres, Spanish writing assistant

1. Introducción

Escribir textos de ámbitos especializados no es fácil, ya que han de poseer unas características muy concretas que deben tenerse en cuenta para que sean adecuados (Cabré, 1999; Gotti, 2008). Estas características dependerán del ámbito de los textos y del género textual que se quiera producir (Van Dijk, 1989; Bhatia, 1993; Parodi, 2010). Como indica Swales (1990), un género textual es una estructura convencionalizada a través de la cual se organizan los intercambios comunicativos de una determinada comunidad discursiva. Por lo general, cuando se habla de géneros textuales producidos en ámbitos especializados, suele partirse de la idea de que los autores serán especialistas de dichos ámbitos (Cabré, 1999), como por ejemplo un abogado en el ámbito legal. Sin embargo, hay determinados ámbitos especializados en los que son los ciudadanos quienes deben enfrentarse a la tarea de la redacción de textos. Un ejemplo muy evidente es el ámbito de la Administración pública, a quien los ciudadanos deben dirigirse por escrito en muchas ocasiones, para presentar por ejemplo una reclamación, una alegación o una queja. La ciudadanía no suele estar familiarizada con este tipo de géneros textuales administrativos, y su redacción resulta difícil y frustrante, lo cual se agrava al tratarse de documentos que tienen una repercusión directa en su vida y en su bienestar.

En el contexto de la lengua española, en los últimos años ha habido algunos esfuerzos por listar los géneros textuales prototípicos del ámbito administrativo y caracterizarlos lingüísticamente (Ayala et al., 2000; Castellón, 2001; Sánchez Alonso, 2014). A partir de estos trabajos, en da Cunha & Montané (2019) se seleccionaron los géneros textuales que tienen como emisores a los ciudadanos y, posteriormente, se analizaron empíricamente cuáles de estos géneros deben redactar con más frecuencia y cuáles les generan una mayor dificultad de redacción.



DOI: 10.21814/lm.12.1.316

This work is Licensed under a

Creative Commons Attribution 4.0 License

En las conclusiones de ese estudio se destacaron cinco géneros: alegación, carta de presentación, queja, reclamación y solicitud. Asimismo, en ese trabajo, se analizaron los problemas que plantean los ciudadanos a la hora de escribir textos destinados a la Administración, y se concluyó que son principalmente cuatro: 1) estructurar el texto, 2) utilizar el vocabulario adecuado a la situación comunicativa, 3) decidir el contenido del texto y 4) utilizar el grado de formalidad adecuado. Recientemente, en da Cunha & Montané (2020), se realizó un análisis lingüístico (textual, léxico y discursivo) basado en corpus de los cinco géneros mencionados, para obtener una perspectiva global de las características de cada uno de ellos.

Tomando estos estudios como base, el objetivo del presente trabajo ha sido desarrollar una aplicación tecnológica que ayuda a la ciudadanía a escribir textos dirigidos a la Administración. Esta aplicación se ha denominado sistema arText e integra diferentes herramientas de Procesamiento del Lenguaje Natural (PLN).

En el Apartado 2 se presenta un estado de la cuestión sobre los recursos de ayuda a la redacción en el ámbito de la e-Administración. En el Apartado 3 se detalla la metodología del trabajo, haciendo hincapié en el marco teórico, en las funcionalidades de arText y en su implementación. En el Apartado 4 se exponen los resultados y la evaluación realizada, tanto *data-driven* como *user-driven*. En el Apartado 5 se incluyen las conclusiones y las líneas de trabajo futuro.

2. Estado de la cuestión

En España, la comunicación electrónica entre el ciudadano y la Administración es el procedimiento habitual en la actualidad, ya que se está tendiendo a la e-Administración. El 2 de octubre del 2015 se publicó en el Boletín Oficial del Estado (BOE) la Ley 39/2015, de 1 de octubre, del Procedimiento Administrativo Común de las Administraciones Públicas¹. Esta Ley establece una regulación completa de las relaciones entre las Administraciones y los administrados, en diversos aspectos. Uno de ellos está relacionado con las Tecnologías de la Información y la Comunicación (TIC), cuyo desarrollo en los últimos años ha repercutido en la forma y contenido de las relaciones entre la Administración, y los ciudadanos y las empresas. En la introducción de la Ley 39/2015 se indica:

Si bien la Ley 30/1992, de 26 de noviembre, ya fue consciente del impacto de las nuevas tecnologías en las relaciones administrativas, fue la Ley 11/2007, de 22 de junio, de acceso electrónico de los ciudadanos a los Servicios Públicos, la que les dio carta de naturaleza legal, al establecer el derecho de los ciudadanos a relacionarse electrónicamente con las Administraciones Públicas, así como la obligación de éstas de dotarse de los medios y sistemas necesarios para que ese derecho pudiera ejercerse. [...] Porque una Administración sin papel basada en un funcionamiento íntegramente electrónico no sólo sirve mejor a los principios de eficacia y eficiencia, al ahorrar costes a ciudadanos y empresas, sino que también refuerza las garantías de los interesados.

Sin embargo, hasta ahora son pocos los esfuerzos que se han hecho para desarrollar herramientas TIC que tengan como objetivo mejorar la comunicación escrita entre la Administración y la ciudadanía, a pesar de que ya en 2010 en el Informe de la Comisión de modernización del lenguaje jurídico del Ministerio de Justicia² se especificaba que:

Las tecnologías de la información pueden prestar soporte para el análisis de la claridad de los textos, facilitando una redacción más comprensible, sin suponer por ello una ralentización en los tiempos de escritura.

En este sentido, son de especial utilidad los analizadores gramaticales y estadísticos de textos (longitud de las frases, número de oraciones, longitud media de las palabras, etc.), así como los programas de escritura semiautomática empleados por los profesionales de la interpretación y traducción. Estos últimos, por ejemplo, almacenan repositorios de frases o párrafos predeterminados que, una vez validados, se proponen a los redactores de los textos.

En definitiva, se recomienda a las instituciones del sector promover el uso de este tipo de programas entre los profesionales del derecho e invertir en el desarrollo de nuevas aplicaciones.

Es cierto que desde hace años y hasta la actualidad se han escrito diversas publicaciones interesantes que han tratado sobre las características del lenguaje jurídico (Rodríguez-Aguilera, 1969; Duarte & Martínez, 1995; Alcaraz & Hughes,

¹<https://boe.es/boe/dias/2015/10/02/pdfs/BOE-A-2015-10565.pdf>

²<https://lenguajeadministrativo.com/wp-content/uploads/2013/05/cmlj-recomendaciones.pdf>

2002; Samaniego, 2005; González Salgado, 2009; Montolío Durán, 2012; Jiménez Yañez, 2016, entre otros). Sin embargo, los trabajos centrados en el lenguaje administrativo son más escasos, y principalmente han abordado el tema desde un punto de vista descriptivo y, por lo general, teniendo en cuenta como potencial emisor de los textos a los profesionales del ámbito y no a la ciudadanía (Ayala et al., 2000; de Miguel Aparicio, 2000; Castellón, 2001; Sánchez Alonso, 2014, entre otros).

También existen algunos sistemas automáticos en el ámbito del PLN que pueden ser de utilidad a la hora de escribir textos, ya que detectan diferentes tipos errores, principalmente ortográficos, gramaticales y de estilo, como por ejemplo:³

- Stilus,⁴ textos generales en español;
- LanguageTool,⁵ textos generales en español e inglés, entre otras lenguas;
- Estilector,⁶ textos académicos universitarios en español;
- Grammarly,⁷ textos generales y algunos géneros textuales en inglés;
- ACROLINX,⁸ textos generales en inglés.
- GrammarChecker,⁹ textos generales en inglés.
- SWANfootnote,¹⁰ textos científicos en inglés;
- CALeSE,¹¹ textos científicos en inglés;
- eWriting Pal,¹² textos académicos en inglés.

Otro caso interesante dentro de la Península Ibérica es el de LinguaKit¹³, un corrector que analiza el texto buscando errores ortográficos, léxicos, gramaticales o de estilo. Actualmente únicamente funciona para el gallego, pero, según los autores, su arquitectura se podría adaptar a cualquier lengua, como el español, en caso de tener los recursos necesarios para ello, como diccionarios y analizadores morfosintácticos (Gama-llo Otero et al., 2015).

³Aunque este trabajo se centra en la lengua española, se incluyen aquí también referencias de sistemas automáticos para el inglés para dejar patente la escasa investigación previa en el ámbito incluso en esta lengua, que es la lengua para la cual, por lo general y como es sabido, se suelen desarrollar más aplicaciones y herramientas informáticas.

⁴<http://www.mystilus.com/>

⁵<https://www.language-tool.org/>

⁶<http://www.estilector.com/>

⁷<http://www.grammarly.com/>

⁸<http://www.acrolinx.com/>

⁹https://www.e-uned.es/correctme/cm_english/

¹⁰<https://cs.joensuu.fi/swan/>

¹¹<http://www.nilc.icmc.usp.br/calese/>

¹²<http://www.ewritingpal.com/>

¹³<https://linguakit.com/es/supercorrector>

No obstante, ninguno de estos sistemas se centra en textos del ámbito de la Administración y, además, la mayor parte de ellos son sistemas comerciales. Es de destacar también que, en el ámbito del PLN, normalmente se emplea información lingüística de varios tipos, pero en el campo de la redacción asistida es difícil encontrar investigaciones basadas en conocimiento extraído del análisis del discurso especializado para desarrollar herramientas informáticas relacionadas con la lengua, aunque en los últimos años ha habido autores que han remarcado la necesidad de usar información discursiva como entrada para dichas herramientas (Zhou et al., 2014).

3. Metodología

En este apartado se detalla el marco teórico del trabajo, junto con las funcionalidades de ar-Text y los detalles de su implementación.

3.1. Marco teórico

En este trabajo se emplean tres marcos teóricos complementarios. En primer lugar, con respecto al nivel textual, se parte, por un lado, de los trabajos de Van Dijk (1977, 1989), quien afirma que los géneros textuales siguen un patrón claramente codificado y ampliamente aceptado. Por ejemplo, un artículo de investigación suele incluir ciertos apartados prototípicos: Introducción, Estado de la cuestión, Metodología, Resultados y Conclusiones. Este autor, además, define la superestructura como la estructura organizativa textual, que varía dependiendo del tipo de texto y que, en general, se muestra mediante distintos apartados, que suelen incluir títulos y diferentes contenidos. Por otro lado, se utiliza también el concepto *moves* (“movidas retóricas” o “movimientos”) de Swales (1990) para caracterizar la superestructura de los cinco géneros textuales que se incluyen en este trabajo, en la línea del análisis basado en corpus propuesto por Biber et al. (2007). Concretamente, en este marco, un *move* representa un fragmento de texto que sirve para una función comunicativa y semántica particular (Upton & Cohen, 2009), y que suele insertarse en alguno de los apartados del documento.

En segundo lugar, en cuanto al nivel léxico, se sigue la Teoría Comunicativa de la Terminología (TCT) de Cabré (1999), que es una teoría de la terminología y del discurso especializado que pone de relieve la dimensión comunicativa de los textos de especialidad. Según la TCT, los géneros textuales producidos en ámbitos especializados presentan algunas características globales, como

la precisión, la concisión, la sistematicidad, la impersonalidad y la objetividad. Estas características se hacen evidentes en los textos por medio que diferentes rasgos lingüísticos. Por ejemplo, la concisión se puede lograr a través del uso de siglas y la objetividad se puede alcanzar evitando marcadores de subjetividad, entre otras estrategias (Cabré et al., 2010).

En tercer lugar, en relación con el nivel discursivo, se emplea la *Rhetorical Structure Theory* (RST) de Mann & Thompson (1981), que es una teoría de organización textual que permite describir un documento caracterizando su estructura mediante las relaciones retóricas que mantienen sus segmentos discursivos. Estas relaciones pueden ser núcleo-satélite o multinucleares. En las relaciones núcleo-satélite hay dos elementos: uno de ellos es más relevante de cara a los propósitos del emisor (el núcleo), mientras que el otro (el satélite) aporta una información adicional sobre el núcleo. En las relaciones multinucleares, en cambio, puede haber más de dos elementos, todos ellos núcleos, que se relacionan al mismo nivel, es decir, todos tienen la misma importancia de cara a los propósitos del autor. Estos elementos se llaman segmentos discursivos o *elementary discourse units* (EDUs), y pueden definirse como indican (Tofiloski et al., 2009, p.77):

Discourse segmentation is the process of decomposing discourse into elementary discourse units (EDUs), which may be simple sentences or clauses in a complex sentence, and from which discourse trees are constructed.

Los criterios de segmentación discursiva específicos para el español utilizados en este trabajo se extrajeron de da Cunha & Iruskieta (2010). En el ejemplo (1) se incluye una oración (que podría ser parte de una alegación) donde se han marcado los tres segmentos discursivos que contiene (indicados entre corchetes, junto con el número de segmento):

- (1) [Ese día el Ayuntamiento no tuvo en cuenta la seguridad ciudadana,]SEGMENTO1 [es decir, no adoptó las medidas necesarias que garantizaran un correcto funcionamiento de los servicios públicos municipales.]SEGMENTO2 [pero sí es cierto que supo reconocer su error públicamente.]SEGMENTO3

da Cunha & Montané (2020) seleccionaron ocho relaciones discursivas para el análisis junto con sus correspondientes conectores discursivos: siete relaciones núcleo-satélite (antítesis, causa,

concesión, condición, propósito, reformulación y resumen) y una relación multinuclear (contraste). Esta selección se realizó a través de la búsqueda de relaciones discursivas que aparecen de manera frecuente en el corpus de referencia anotado con relaciones discursivas en español, el *RST Spanish Treebank*¹⁴, y que a la vez se evidencian habitualmente por medio de conectores. En el marco del presente trabajo, y de cara a integrarlos en arText, se hace una diferencia entre conectores discursivos intraoracionales (conectores que enlazan segmentos dentro de una misma oración) y conectores interoracionales (conectores que enlazan diferentes oraciones). Por ejemplo, la oración mostrada en el ejemplo (1) incluye dos conectores intraoracionales que encabezan los segmentos 2 y 3: “es decir”, que refleja una relación de reformulación, y “pero”, que marca una relación de antítesis. Si esta oración se dividiese en tres oraciones diferentes, más breves, esos dos conectores intraoracionales deberían sustituirse por conectores interoracionales que expresasen la misma relación discursiva, como por ejemplo “en otras palabras” y “sin embargo”, respectivamente, tal como se recoge en el ejemplo (2), donde ambos conectores se han marcado en cursiva:

- (2) [Ese día el Ayuntamiento no tuvo en cuenta la seguridad ciudadana.]SEGMENTO1 [*En otras palabras*, no adoptó las medidas necesarias que garantizaran un correcto funcionamiento de los servicios públicos municipales.]SEGMENTO2 [*Sin embargo*, sí es cierto que supo reconocer su error públicamente.]SEGMENTO3

Partiendo de estos tres marcos teóricos complementarios, uno de los resultados del trabajo de da Cunha & Montané (2020) fue una estructura modelo de cada uno de los cinco géneros textuales analizados. La estructura modelo de cada género incluye sus apartados prototípicos, la indicación de si estos suelen incluir un título y cuál es en caso de existir, y los *moves* (Swales, 1990) habituales en cada apartado. Además, se ofrece fraseología habitual que puede utilizarse para expresar dichos *moves*. Por ejemplo, en una alegación los apartados prototípicos detectados son los siguientes: “Cabecera”, “Identificación del emisor”, “Exposición de hechos”, “Presentación de alegaciones”, “Petición” y “Cierre”. En cuanto a los *moves*, en el caso del apartado “Presentación de alegaciones”, por ejemplo, se identificaron cuatro habituales: “Fórmula de introducción de alegaciones”, “Título introductorio al listado

¹⁴http://www.corpus.unam.mx/rst/index_es.html

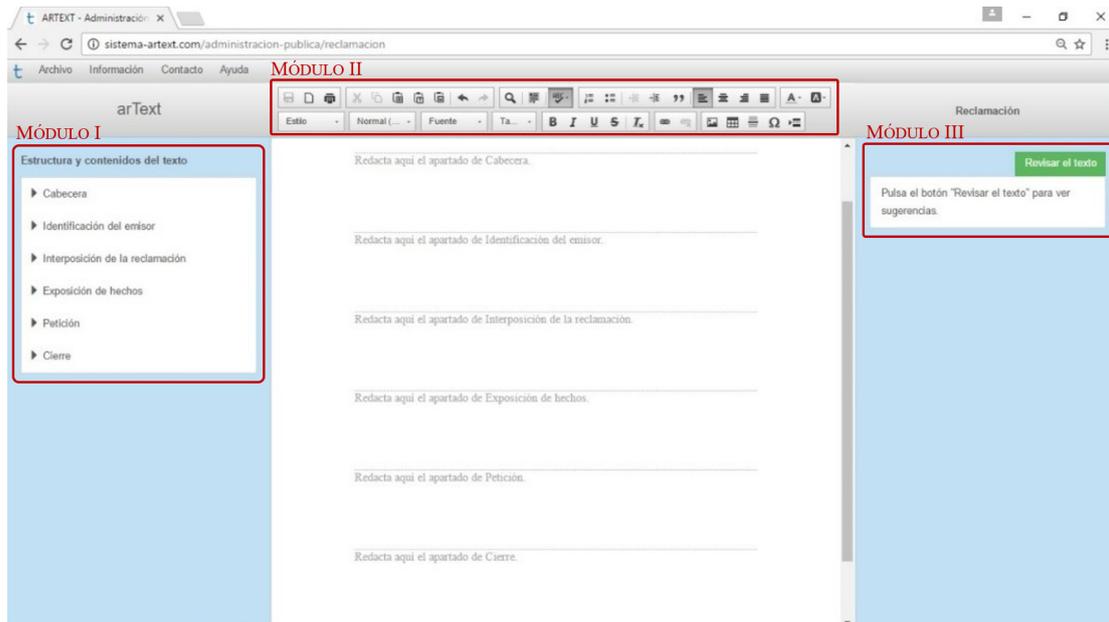


Figura 1: Captura de pantalla del editor de textos en línea de arText, con indicación de sus tres módulos en color rojo.

de alegaciones”, “Listado de las alegaciones presentadas por el emisor” y “Mención de los documentos adjuntos”. En el caso del primero de estos cuatro *moves*, en la estructura modelo se incluyen algunas unidades fraseológicas (que en este caso se corresponden con oraciones completas, aunque no siempre es así, ya que depende del género textual del que se trate) que pueden ayudar a expresarlo en un texto, como por ejemplo:¹⁵

Que mediante este escrito, y sin perjuicio de lo que pueda manifestar en el trámite de audiencia, formula las siguientes alegaciones.

Que no estando conforme con los hechos denunciados, interpone el presente escrito en base a las siguientes alegaciones.

Que dentro del plazo que le ha sido concedido y al amparo de lo previsto en [documentación legal aplicable: artículos de leyes, leyes, reglamentos, etc.], formula alegaciones y presenta los documentos y justificaciones necesarios.

Otro resultado de dicho trabajo fue una caracterización lingüística de cada género textual, que incluye tanto ciertos aspectos léxicos como discursivos. En relación con los aspectos léxicos, se incluye, por ejemplo, información sobre el uso de siglas, de unidades subjetivas, y de la primera persona del singular y del plural. En cuanto a aspectos discursivos, se aporta información sobre la cantidad de oraciones, de segmentos discursivos y de conectores discursivos que suelen contener.

¹⁵Los corchetes indican información variable en cada texto.

El diseño de arText, cuyo desarrollo es el objetivo del presente artículo, parte, como se verá en el apartado 3.2, por un lado, de las estructuras modelo mencionadas para cada uno de los cinco géneros textuales, y, por otro, de su caracterización lingüística.

3.2. Módulos y funcionalidades de arText

El sistema arText es una aplicación tecnológica que tiene forma de editor de textos en línea, y que puede utilizarse gratuitamente y sin necesidad de registro¹⁶. Hay disponible un tutorial en línea sobre sus funcionalidades en Canal UNED¹⁷. Una vez el usuario entra en la página principal del sistema, debe seleccionar el botón “EMPIEZA A USAR ARTEXT” y, a continuación, indicar el género textual que desea redactar. El usuario puede elegir entre los cinco géneros textuales del ámbito de la Administración mencionados en el apartado 1: alegación, carta de presentación, queja, reclamación y solicitud. Una vez seleccionado el género textual, el usuario entrará automáticamente en el editor en línea (véase Figura 1, en donde se ha seleccionado el género reclamación). Desde este editor, puede comenzar a redactar su escrito en la hoja de texto central, utilizando los tres módulos de ayuda con los que cuenta el sistema, que se detallan a continuación:

¹⁶<http://sistema-artext.com/>

¹⁷<https://canal.uned.es/mmobj/index/id/54433>

I. Módulo de estructura y contenidos del texto

El primer módulo ayuda a estructurar el documento y a comenzar a redactarlo, utilizando las estructuras modelo de cada género. Este módulo se encuentra en la columna izquierda de la pantalla y permite:

- Insertar los apartados prototípicos del documento.
- Añadir contenidos habitualmente presentes en cada apartado (los mencionados *moves* de Swales (1990)).
- Incorporar fraseología relacionada con los contenidos.

Tal como muestra la Figura 1, lo primero que visualiza el usuario en la columna izquierda es la lista de apartados sugeridos por arText para el género textual seleccionado, la reclamación. Asimismo, como se aprecia también en la Figura 1, el usuario verá estos mismos apartados en la hoja de texto, siempre precedidos por la siguiente indicación en color gris: “Redacta aquí el apartado de...”. Una vez el usuario comience a escribir el texto en el apartado correspondiente, cada una de estas indicaciones se eliminará automáticamente.

Al hacer clic en alguno de los apartados sugeridos en la columna izquierda, se despliega debajo una lista que incluye los contenidos que se recomienda insertar en él, en el orden propuesto. El título del apartado, si procede incorporarlo, se muestra como el primer contenido de la lista. Al situar el cursor sobre un contenido, este se evidencia mediante un recuadro con fondo azul. En la Figura 2 se ofrece un ejemplo, en que el apartado “Exposición de hechos” incluye dos contenidos: “Explicación del hecho o hechos que motivan la reclamación” y “Mención de los documentos adjuntos”.

A su vez, al hacer clic sobre cada contenido, se muestra una lista de frases prototípicas que pueden utilizarse para expresarlo, que no aparecen en un orden determinado. El usuario debe elegir cuáles de estas frases quiere incorporar en su texto, teniendo en cuenta el objetivo del documento que está redactando. Estas frases están escritas en color azul y, al poner el cursor encima de una de ellas, aparecerá subrayada. Para incorporar las frases deseadas en el documento, basta con hacer clic sobre cada una de ellas en la columna de la izquierda y automáticamente se cargarán en el texto, dentro del apartado correspondiente. Las frases propuestas por arText pueden incluir varios componentes:

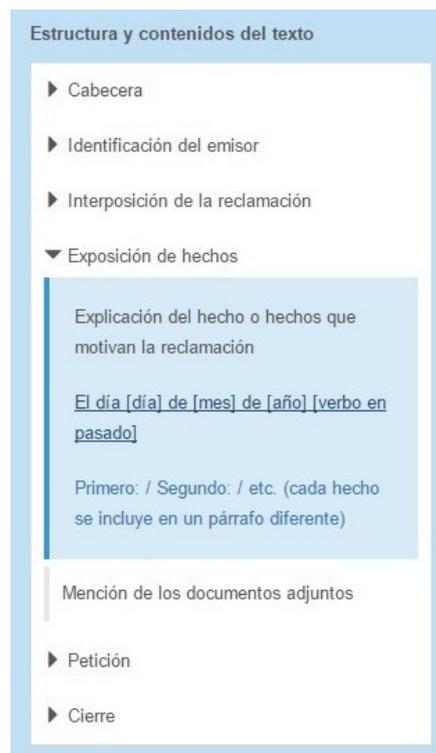


Figura 2: Detalle de una captura de pantalla de arText donde se muestran los apartados y contenidos incluidos en el módulo 1 para el género reclamación.

- Elementos sin marca alguna: son fragmentos literales que el usuario podría incorporar en su texto directamente, como por ejemplo: “Mediante el presente escrito presenta reclamación de responsabilidad patrimonial de la administración en base a los siguientes hechos:”.
- Elementos entre corchetes: se trata de componentes variables que el usuario debe modificar para adaptar a su texto. Por ejemplo, si arText ofrece la siguiente secuencia “El día [día] de [mes] de [año] [verbo en pasado]” el usuario debería transformarla sustituyendo la información entre corchetes, como por ejemplo de la siguiente manera: “El día 5 de enero de 2015 solicitó”.
- Elementos entre paréntesis: se trata de instrucciones sobre la estructura textual sugeridas al usuario. Ej. “(cada hecho se incluye en un párrafo diferente)” quiere decir que cada uno de los hechos redactados por el usuario debe ocupar un párrafo distinto en el texto.

En la Figura 2 se incluye la fraseología asociada al contenido “Explicación del hecho o hechos que motivan la reclamación”, ubicado en el apartado “Exposición de hechos”.

II. Módulo de corrección ortográfica y formato

El segundo módulo incorpora una barra superior que incluye diferentes opciones de formato mediante las cuales el usuario tendrá acceso a las funcionalidades más habituales de los editores de texto, como seleccionar el tamaño de letra y la fuente, asignar estilos, insertar tablas, hacer listas, copiar, pegar, buscar, etc. (véase Figura 1). Asimismo, en esta barra de formato se ha integrado un corrector ortográfico.

III. Módulo de sugerencias sobre léxico y discurso

El sistema arText cuenta con un tercer módulo de sugerencias que permite al usuario procesar lingüísticamente su texto, y visualizar recomendaciones sobre cuestiones relacionadas con aspectos lingüísticos (léxicos y discursivos, principalmente) que se podrían mejorar. Estas recomendaciones están basadas en la caracterización lingüística de los cinco géneros textuales analizados por da Cunha & Montané (2020).

Para visualizar las sugerencias, el usuario debe tener escrito su texto en el editor. En el ejemplo (3) se incluye un texto ficticio que se corresponde con el género textual de la reclamación (donde se ha eliminado la información sobre el emisor y el receptor):

- (3) Mediante el presente escrito presenta reclamación de responsabilidad patrimonial de la administración en base a los hechos que se describen a continuación.

El día 2 de mayo de 2019 el personal del Ayuntamiento instaló cuatro pivotes de gran tamaño en el número 46 de la calle Galileo. Creemos que esta decisión no se debió de tomar teniendo en cuenta las normas urbanísticas del Plan General de Ordenación Urbana de Madrid (PGOUM). En este Plan General de Ordenación Urbana de Madrid se determinaron las situaciones en las cuales es posible instalar este tipo de elementos urbanos, es decir, se definieron supuestos en los que dicha instalación se debería efectuar. Evidentemente, la instalación de estos cuatro bolardos supone una negligencia, porque imposibilitan la salida de vehículos por la puerta del garaje de la vivienda que se encuentra delante de los mismos, pero considero que es posible solucionar la situación mediante una opción alternativa. Esta opción sería eliminar dos de los cuatro pivotes, es decir, mantener únicamente dos, separados a una distancia de 3 metros. Creo que debemos ser partícipes de las decisiones tomadas

por la administración que nos afecten, es decir, debe tenerse en cuenta la opinión de la comunidad de vecinos de la zona porque es su derecho poder decidir sobre estas cuestiones. En otras comunidades, como la AVIT, hace tiempo que negocian sobre cómo alcanzar un pacto para la regeneración de ideas en relación con esta materia. Deberíamos fijarnos también en la CVBM para observar maneras distintas de gestionar nuestros procesos de toma de decisiones, porque no es deseable caer en los errores de siempre.

Se adjunta la siguiente documentación justificativa:

Ley 30/1992, de 26 de noviembre, de Régimen Jurídico de las Administraciones Públicas y del Procedimiento Administrativo Común.

Real Decreto 429/1993, de 26 de marzo, por el que se aprueba el Reglamento de las Administraciones Públicas en materia de responsabilidad patrimonial.

Solicita al Ayuntamiento que admita esta reclamación patrimonial, junto con los documentos que se acompañan.

Una vez escrito el texto, el usuario debe hacer clic sobre el botón “Revisar el texto” de la columna derecha de la pantalla. Entonces, el sistema le ofrecerá en esa misma columna diferentes recomendaciones de mejora del texto, como se recoge en la Figura 3.

En función de la recomendación sobre la que el usuario haga clic, aparecerán marcadas en el texto que ha escrito diferentes cuestiones, como oraciones largas, siglas, palabras repetidas, etc. A continuación se detallan las recomendaciones de mejora ofrecidas al usuario al procesar el texto, junto con un ejemplo representativo de cada una de ellas:

- a) “*División de oraciones largas.* Parece que las oraciones marcadas podrían dividirse en otras más cortas. Te recomendamos que lo hagas. Haz clic en cada oración para ver dónde podrías segmentarla.”

En este caso, el sistema marca en amarillo en el texto las oraciones consideradas demasiado largas para el género textual seleccionado (véase Figura 4)¹⁸.

¹⁸Se toman como referencia los umbrales de palabras para cada género detectados por da Cunha & Montané (2020).

Mediante el presente escrito presenta reclamación de responsabilidad patrimonial de la administración en base a los hechos que se describen a continuación.

El día 2 de mayo de 2019 el personal del Ayuntamiento instaló cuatro pivotes de gran tamaño en el número 46 de la calle Galileo. Creemos que esta decisión no se debió de tomar teniendo en cuenta las normas urbanísticas del Plan General de Ordenación Urbana de Madrid (PGOUM). En este Plan General de Ordenación Urbana de Madrid se determinaron las situaciones en las cuales es posible instalar este tipo de elementos urbanos, es decir, se definieron supuestos en los que dicha instalación se debería efectuar. Evidentemente, la instalación de estos cuatro bolardos supone una negligencia, porque imposibilitan la salida de vehículos por la puerta del garaje de la vivienda que se encuentra delante de los mismos, pero considero que es posible solucionar la situación mediante una opción alternativa. Esta opción sería eliminar dos de los cuatro pivotes, es decir, mantener únicamente dos, separados a

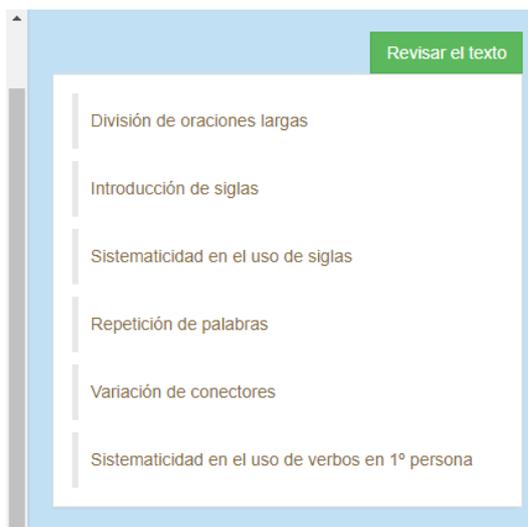


Figura 3: Detalle de una captura de pantalla de arText que muestra las recomendaciones de mejora obtenidas al procesar un texto.

Además, si el sistema logra dividir automáticamente la oración en segmentos discursivos más breves, también ofrece esta información al usuario en la columna derecha. Veamos el ejemplo (4), que incluye una oración extraída del texto del ejemplo (3):

- (4) Evidentemente, la instalación de estos cuatro bolardos supone una negligencia, porque imposibilitan la salida de vehículos por la puerta del garaje de la vivienda que se encuentra delante de los mismos, pero considero que es posible solucionar la situación mediante una opción alternativa.

En este caso, arText propone al usuario dividir esta oración larga en tres oraciones más breves, teniendo en cuenta los tres segmentos siguientes (véase Figura 4):

- I. *Evidentemente, la instalación de estos cuatro bolardos supone una negligencia,*
- II. *porque imposibilitan la salida de vehículos por la puerta del garaje de la vivienda que se encuentra delante de los mismos,*
- III. *pero considero que es posible solucionar la situación mediante una opción alternativa.*

Asimismo, en caso de que los segmentos propuestos comiencen por conectores discursivos intraoracionales, el sistema propone al usuario opciones de conectores interoracionales alternativos que expresen la misma relación discursiva. Por ejemplo, siguiendo con la misma oración, en el texto se marcarían en naranja dos conectores: “porque” (que refleja una relación de causa) y

“pero” (que refleja una relación de antítesis). Si el usuario hace clic en “pero”, verá en la columna de la derecha las siguientes propuestas de conectores interoracionales alternativos, que expresan la misma relación de antítesis: “Con todo”, “De todas formas”, “De todas maneras”, “De todos modos”, “No obstante” y “Sin embargo” (véase Figura 4). De esta manera, el usuario obtendrá información útil que le ayudará a decidir si desea conservar en su texto la oración original o si, por el contrario, prefiere dividirla en otras oraciones más breves y mantener explícita la relación discursiva que existe entre ellas a través de conectores. Es el usuario quien decide en todo momento qué recomendaciones desea incorporar en su texto. Para hacerlo, simplemente debe realizar los cambios directamente en la hoja del editor de textos.

- b) “*Introducción de siglas.* Las unidades marcadas parecen siglas. Si es así, ten en cuenta que la primera vez que se utiliza una sigla en un texto suele ir acompañada del término desplegado.”

En este caso, el sistema marca en amarillo en el texto las siglas que no se acompañan de su correspondiente término desplegado la primera vez que aparecen en el texto, lo cual no es recomendable. El sistema solo detecta siglas propias (Giraldo, 2008), es decir, aquellas que están formadas únicamente por las unidades léxicas incluidas en la estructura sintagmática del término desplegado, como por ejemplo “UNED”, que se corresponde con “Universidad Nacional de Educación a Distancia”. En la Figura 5 puede verse que el sistema ha detectado dos de estos casos, como son “AVIT” y “CVBM”.

Mediante el presente escrito presenta reclamación de responsabilidad patrimonial de la administración en base a los hechos que se describen a continuación.

El día 2 de mayo de 2019 el personal del Ayuntamiento instaló cuatro pivotes de gran tamaño en el número 46 de la calle Galileo. Creemos que esta decisión no se debió de tomar teniendo en cuenta las normas urbanísticas del Plan General de Ordenación Urbana de Madrid (PGOUM). En este Plan General de Ordenación Urbana de Madrid se determinaron las situaciones en las cuales es posible instalar este tipo de elementos urbanos, es decir, se definieron supuestos en los que dicha instalación se debería efectuar. Evidentemente, la instalación de estos cuatro bolardos supone una negligencia, porque imposibilitan la salida de vehículos por la puerta del garaje de la vivienda que se encuentra delante de los mismos, pero considero que es posible solucionar la situación mediante una opción alternativa. Esta opción sería eliminar dos de los cuatro pivotes, es decir, mantener únicamente dos, separados a una distancia de 3 metros. Creo que debemos ser participes de las decisiones tomadas por la administración que nos afecten, es decir, debe tenerse en cuenta la opinión de la comunidad de vecinos de la zona porque es su derecho poder decidir sobre estas cuestiones. En otras comunidades, como la AVIT, hace tiempo que negocian sobre cómo alcanzar un pacto para la regeneración de ideas en relación con esta materia. Deberíamos fijarnos también en la CVBM para observar maneras distintas de gestionar nuestros procesos de toma de decisiones, porque no es deseable caer en los errores de siempre.

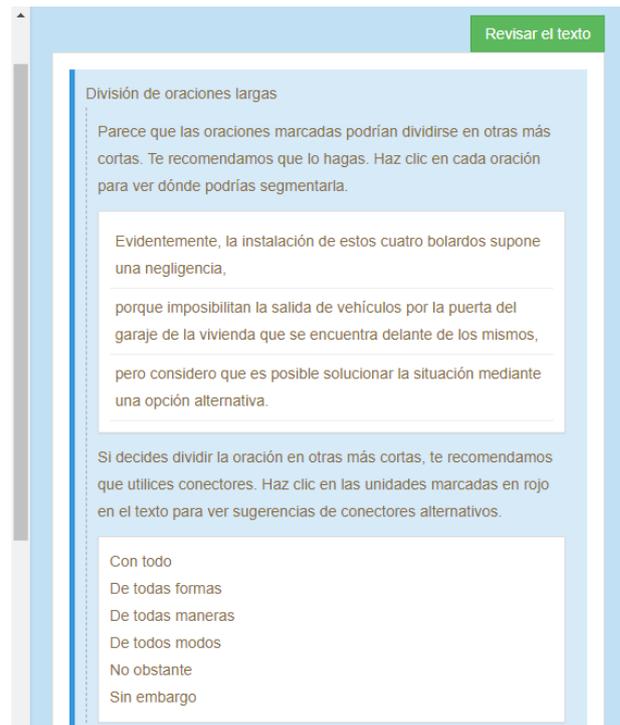


Figura 4: Detalle de una captura de pantalla de arText que refleja la recomendación *División de oraciones largas*.

Creo que debemos ser participes de las decisiones tomadas por la administración que nos afecten, es decir, debe tenerse en cuenta la opinión de la comunidad de vecinos de la zona porque es su derecho poder decidir sobre estas cuestiones. En otras comunidades, como la AVIT, hace tiempo que negocian sobre cómo alcanzar un pacto para la regeneración de ideas en relación con esta materia. Deberíamos fijarnos también en la CVBM para observar maneras distintas de gestionar nuestros procesos de toma de decisiones, porque no es deseable caer en los errores de siempre.

Se adjunta la siguiente documentación justificativa:

Ley 30/1992, de 26 de noviembre, de Régimen Jurídico de las Administraciones Públicas y del Procedimiento Administrativo Común.

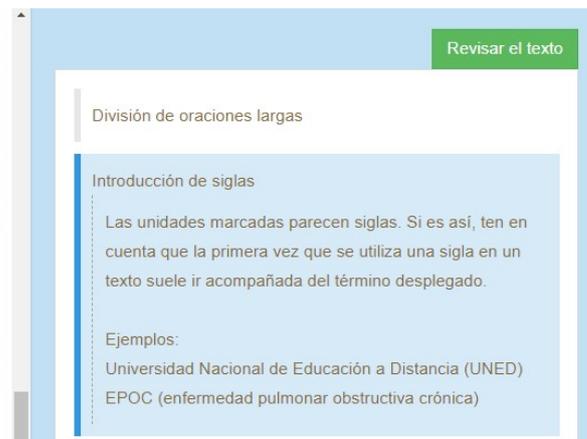


Figura 5: Detalle de una captura de pantalla de arText que refleja la recomendación *Introducción de siglas*.

- c) “*Sistematicidad en el uso de siglas.* Las unidades marcadas parecen el término desplegado de siglas que utilizas en el texto. Si es así, ten en cuenta que, una vez se introduce una sigla en un texto, se suele seguir utilizando la sigla y no el término desplegado.”

Esta recomendación también está relacionada con las siglas. Sin embargo, en esta ocasión el sistema marca en el texto las ocurrencias de términos desplegados que ya habían sido introducidos en el texto previamente junto con su correspondiente sigla, lo cual es desaconsejable. La Figura 6 muestra un ejemplo, en que la unidad “Plan General de Ordenación Urbana de Ma-

drid” está marcada dos veces. La primera vez aparece en una oración junto con su sigla entre paréntesis, correctamente. Sin embargo, también aparece marcada en la siguiente oración, ya que se quiere hacer notar al usuario que en este caso sería más conveniente sustituir esta unidad por su sigla.

- d) “*Repetición de palabras.* Las unidades de la lista siguiente se repiten varias veces en el texto. Ten en cuenta que en este tipo de textos puedes utilizar variantes como sinónimos, explicaciones, paráfrasis, etc. Haz clic en cada unidad para visualizar sus ocurrencias en el texto.”

El día 2 de mayo de 2019 el personal del Ayuntamiento instaló cuatro pivotes de gran tamaño en el número 46 de la calle Galileo. Creemos que esta decisión no se debió de tomar teniendo en cuenta las normas urbanísticas del **Plan General de Ordenación Urbana de Madrid** (PGOUM). En este **Plan General de Ordenación Urbana de Madrid** se determinaron las situaciones en las cuales es posible instalar este tipo de elementos urbanos, es decir, se definieron supuestos en los que dicha instalación se debería efectuar. Evidentemente, la instalación de estos cuatro bolardos supone una negligencia, porque imposibilitan la salida de vehículos por la puerta del garaje de la vivienda que se encuentra delante de los mismos, pero considero que es posible solucionar la situación mediante una opción alternativa. Esta opción sería eliminar dos de los cuatro pivotes, es decir, mantener únicamente dos, separados a una distancia de 3 metros.

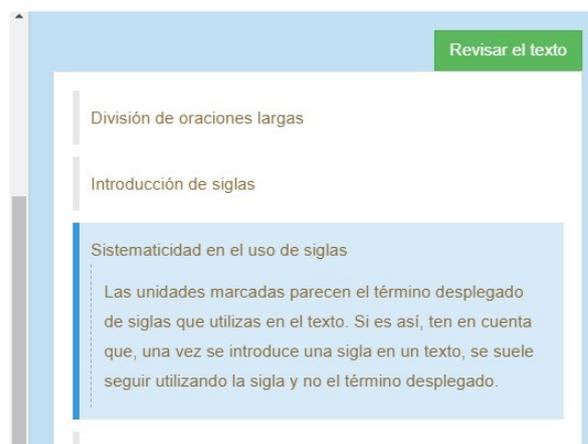


Figura 6: Detalle de una captura de pantalla de arText que refleja la recomendación *Sistematicidad en el uso de siglas*.

Esta recomendación da al usuario una lista de las unidades léxicas repetidas en el texto, concretamente de cuatro categorías gramaticales: nombres, verbos, adjetivos y adverbios. Cuando el usuario hace clic en alguna de estas unidades, en el texto se destacan en amarillo sus diferentes ocurrencias. El sistema lematiza todas las palabras del documento, por lo que se marcarán en el texto todas las formas detectadas de los nombres y de los adjetivos (singular, plural, masculino y femenino) y de los verbos (independientemente de su tiempo, modo, número y persona). Por ejemplo, en la Figura 7, se observa en la columna derecha un fragmento de la lista de palabras repetidas, que incluye “administración”, “deber”, “decir”, “decisión”, “patrimonial” y “urbano”. En el caso del verbo “deber”, se marcan en el texto las formas “debería”, “debemos”, “debe” y “deberíamos”. Es importante destacar que el hecho de que se marquen estas unidades en el texto no quiere decir que sean incorrectas. Lo que se pretende con esta recomendación es ofrecer al usuario información para decidir si desea eliminar o modificar alguna de ellas.

- e) “*Variación de conectores*. Los conectores de la lista siguiente se repiten varias veces en el texto. Haz clic en cada conector para ver sugerencias de conectores alternativos.”

En este caso, el sistema ofrece en la columna derecha un listado de los conectores discursivos que se repiten en el texto tres o más veces. Si el usuario hace clic en alguno de ellos, verá en el texto marcadas sus ocurrencias y además obtendrá en la columna derecha un listado de conectores alternativos (tanto intraoracionales como interoracionales) que expresan la misma relación discursiva. Por ejemplo, en la Figura 8 se observan marcadas en el texto tres ocurrencias del conec-

tor de reformulación “es decir” y en la columna derecha una lista con cuatro propuestas de conectores alternativos, como son “dicho de otro modo”, “en otras palabras”, “esto es” y “o sea”.

- f) “*Sistematicidad en el uso de verbos en 1º persona*. Las unidades marcadas en verde parecen verbos en 1º persona del singular y las marcadas en azul parecen verbos en 1ª persona del plural. Te recomendamos que optes por el singular o el plural para que el texto sea sistemático.”

Esta recomendación tiene que ver con la falta de sistematicidad en el uso de la primera persona del singular y del plural, especialmente en relación con el autor del escrito. Hay ocasiones en que en un texto es necesario utilizar ambas opciones, pero, en los textos administrativos, es importante que el autor utilice el singular o el plural de manera sistemática, en función de quién o quiénes emitan la alegación, queja, reclamación, etc. Para ello, el sistema marca en el texto las formas verbales en singular en verde y las formas en plural en azul, dando así información al autor para tomar la decisión de hacer modificaciones en este sentido. Por ejemplo, en la Figura 9, se marcan en verde las formas en singular “considero” y “creo”, mientras que se destacan en azul las formas “creemos”, “debemos” y “deberíamos”.

- g) “*Uso de indicadores de subjetividad*. Las unidades marcadas podrían ser indicadores de subjetividad. Ten en cuenta que este tipo de textos suelen ser objetivos. Te recomendamos que revises estas unidades para confirmar que son adecuadas en tu texto.”

Las recomendaciones anteriores se aplican a los cinco géneros textuales incluidos en arText.

El día 2 de mayo de 2019 el personal del Ayuntamiento instaló cuatro pivotes de gran tamaño en el número 46 de la calle Galileo. Creemos que esta decisión no se debió de tomar teniendo en cuenta las normas urbanísticas del Plan General de Ordenación Urbana de Madrid (PGOUM). En este Plan General de Ordenación Urbana de Madrid se determinaron las situaciones en las cuales es posible instalar este tipo de elementos urbanos, es decir, se definieron supuestos en los que dicha instalación se debería efectuar. Evidentemente, la instalación de estos cuatro bolardos supone una negligencia, porque imposibilitan la salida de vehículos por la puerta del garaje de la vivienda que se encuentra delante de los mismos, pero considero que es posible solucionar la situación mediante una opción alternativa. Esta opción sería eliminar dos de los cuatro pivotes, es decir, mantener únicamente dos, separados a una distancia de 3 metros.

Creo que **debemos** ser partícipes de las decisiones tomadas por la administración que nos afecten, es decir, **debe** tenerse en cuenta la opinión de la comunidad de vecinos de la zona porque es su derecho poder decidir sobre estas cuestiones. En otras comunidades, como la AVIT, hace tiempo que negocian sobre cómo alcanzar un pacto para la regeneración de ideas en relación con esta materia. **Deberíamos** fijarnos también en la CVBM para observar maneras distintas de gestionar nuestros procesos de toma de decisiones, porque no es deseable caer en los errores de siempre.

Se adjunta la siguiente documentación justificativa:

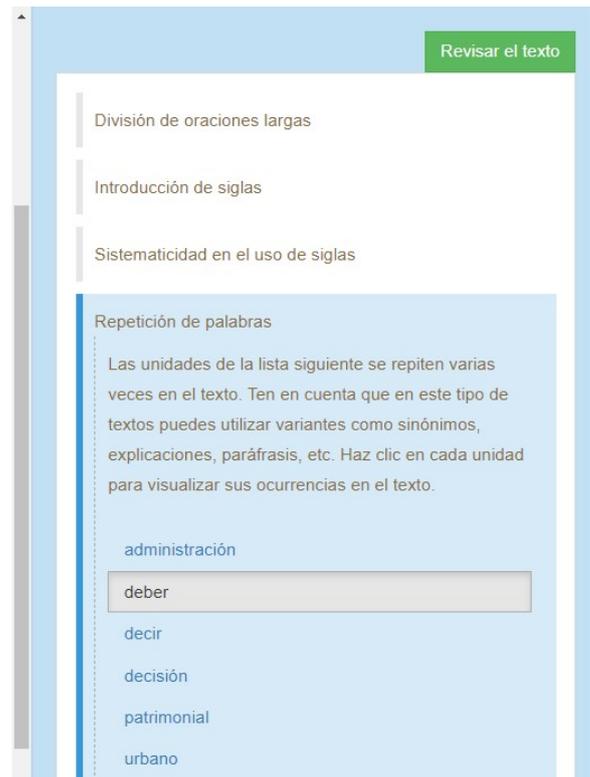


Figura 7: Detalle de una captura de pantalla de arText que refleja la recomendación *Repetición de palabras*.

Madrid (PGOUM). En este Plan General de Ordenación Urbana de Madrid se determinaron las situaciones en las cuales es posible instalar este tipo de elementos urbanos, **es decir**, se definieron supuestos en los que dicha instalación se debería efectuar. Evidentemente, la instalación de estos cuatro bolardos supone una negligencia, porque imposibilitan la salida de vehículos por la puerta del garaje de la vivienda que se encuentra delante de los mismos, pero considero que es posible solucionar la situación mediante una opción alternativa. Esta opción sería eliminar dos de los cuatro pivotes, **es decir**, mantener únicamente dos, separados a una distancia de 3 metros.

Creo que debemos ser partícipes de las decisiones tomadas por la administración que nos afecten, **es decir**, debe tenerse en cuenta la opinión de la comunidad de vecinos de la zona porque es su derecho

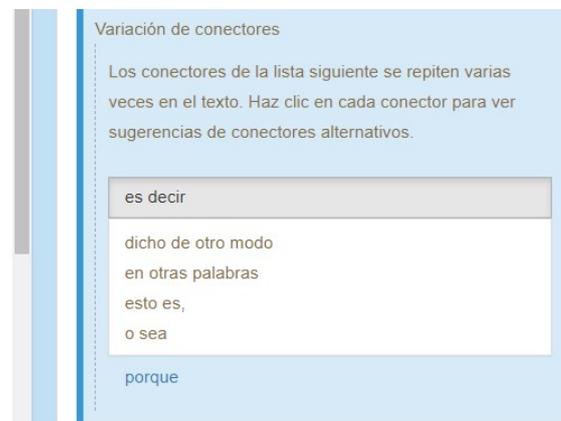


Figura 8: Detalle de una captura de pantalla de arText que refleja la recomendación *Variación de conectores*.

Sin embargo, en este caso, se trata de una recomendación específica para el género textual solicitud, ya que en [da Cunha & Montané \(2020\)](#) se observó que en este tipo de género se tiende a evitar la subjetividad (en contraposición con una queja o una carta de presentación, por ejemplo). Así, como puede observarse en la Figura 10, si se selecciona este género, se marcan en el texto las unidades indicadoras de subjetividad, como “evidentemente”, para que el usuario decida si quiere eliminarlas de su escrito.

3.3. Implementación del sistema

El sistema arText se desarrolló en un entorno Linux usando un servidor Apache. Se utilizaron también diferentes recursos tanto en el *back-end* (Bash, Perl y PHP, con un entorno de trabajo Laravel) como en el *front-end* (HTML, CSS, JavaScript, con AJAX y jQuery). Algunos de los recursos principales se detallan en este apartado. El sistema está optimizado para su utilización con el navegador Google Chrome.

El día 2 de mayo de 2019 el personal del Ayuntamiento instaló cuatro pivotes de gran tamaño en el número 46 de la calle Galileo. **Creemos** que esta decisión no se debió de tomar teniendo en cuenta las normas urbanísticas del Plan General de Ordenación Urbana de Madrid (PGOUM). En este Plan General de Ordenación Urbana de Madrid se determinaron las situaciones en las cuales es posible instalar este tipo de elementos urbanos, es decir, se definieron supuestos en los que dicha instalación se debería efectuar. Evidentemente, la instalación de estos cuatro bolardos supone una negligencia, porque imposibilitan la salida de vehículos por la puerta del garaje de la vivienda que se encuentra delante de los mismos, pero **considero** que es posible solucionar la situación mediante una opción alternativa. Esta opción sería eliminar dos de los cuatro pivotes, es decir, mantener únicamente dos, separados a una distancia de 3 metros.

Creo que **debemos** ser partícipes de las decisiones tomadas por la administración que nos afecten, es decir, debe tenerse en cuenta la opinión de la comunidad de vecinos de la zona porque es su derecho poder decidir sobre estas cuestiones. En otras comunidades, como la AVIT, hace tiempo que negocian sobre cómo alcanzar un pacto para la regeneración de ideas en relación con esta materia. **Deberíamos** fijarnos también en la CVBM para observar maneras distintas de

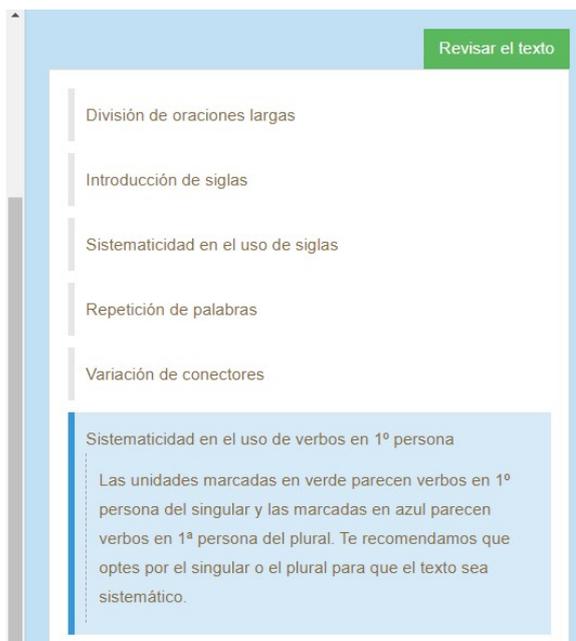


Figura 9: Detalle de una captura de pantalla de arText que refleja la recomendación *Sistematicidad en el uso de verbos en 1º persona*.

El día 2 de mayo de 2019 el personal del Ayuntamiento instaló cuatro pivotes de gran tamaño en el número 46 de la calle Galileo. Creemos que esta decisión no se debió de tomar teniendo en cuenta las normas urbanísticas del Plan General de Ordenación Urbana de Madrid (PGOUM). En este Plan General de Ordenación Urbana de Madrid se determinaron las situaciones en las cuales es posible instalar este tipo de elementos urbanos, es decir, se definieron supuestos en los que dicha instalación se debería efectuar. **Evidentemente**, la instalación de estos cuatro bolardos supone una negligencia, porque imposibilitan la salida de vehículos por la puerta del garaje de la vivienda que se encuentra delante de los mismos, pero **considero** que es posible solucionar la situación mediante una opción alternativa. Esta opción sería eliminar dos de los cuatro pivotes, es decir, mantener únicamente dos, separados a una distancia de 3 metros. **Creo** que **debemos** ser partícipes de las decisiones tomadas por la administración que nos afecten, es decir, debe tenerse en cuenta la opinión de la comunidad de vecinos de la zona porque es su derecho poder decidir sobre estas cuestiones. En otras comunidades, como la AVIT, hace tiempo que negocian sobre

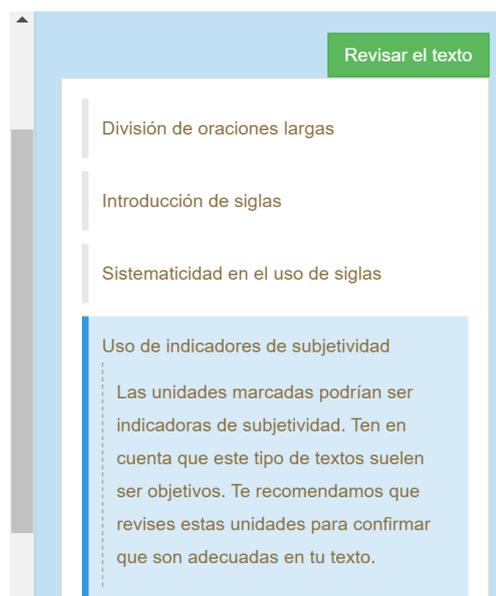


Figura 10: Detalle de una captura de pantalla de arText que refleja la recomendación *Uso de indicadores de subjetividad* en el género textual solicitud.

El principal recurso empleado para la implementación del Módulo I fue una base de datos MySQL, donde se guarda la información relativa a la estructura textual de cada uno de los cinco géneros que incluye arText, es decir, los apartados, los contenidos (incluyendo aquí los títulos) y la fraseología. MySQL es un sistema de gestión de bases de datos relacionales de código abierto con un modelo cliente-servidor que nos pareció adecuado por la facilidad en la gestión de los datos y por los requerimientos del sistema, ya que se necesitaba flexibilidad para crear o editar la es-

tructura (ya que se tiene prevista la creación de nuevos ámbitos y géneros, con diferentes apartados, contenidos y fraseología).

Por su parte, el Módulo II, además de la barra de formato, incorpora un corrector ortográfico *opensource* (WebSpellChecker Ltd). El acceso a WebSpellChecker se hace por medio de una API desde la nube. Se trata de un *plugin* del editor de texto de código abierto CKEditor, utilizado también en la implementación del sistema.

Finalmente, el Módulo III incluye dos herramientas de PLN existentes actualmente para el español que permiten realizar un procesamiento lingüístico del texto escrito por el usuario. En primer lugar, incorpora el analizador morfosintáctico de Freeling (Atserias et al., 2006), mediante el cual se lematizan todas las unidades léxicas del texto y se asigna una categoría gramatical a cada una de ellas. En segundo lugar, incorpora el segmentador discursivo DiSeg (da Cunha et al., 2012), que permite dividir el texto en oraciones y, además, en segmentos discursivos intraoracionales, siguiendo, el concepto de segmento discursivo de Tofiloski et al. (2009) y los criterios de segmentación para el español de da Cunha & Irukieta (2010), mencionados en el apartado 3.1.

Asimismo, se implementaron en el Módulo III diferentes algoritmos que toman como entrada el texto procesado lingüísticamente por las dos herramientas de PLN mencionadas. Estos algoritmos permiten detectar en el texto escrito por el usuario los elementos lingüísticos necesarios para poder ofrecer las recomendaciones correspondientes asociadas a cada uno de ellos. Estos elementos son:

- Oraciones largas, con un umbral diferente de palabras para cada género textual.
- Conectores discursivos interoracionales e interoracionales que evidencian las ocho relaciones discursivas utilizadas en la investigación, extraídos del trabajo de da Cunha & Montané (2020).
- Conectores discursivos que se repiten tres o más veces en el texto, extraídos del trabajo de da Cunha & Montané (2020).
- Unidades léxicas que indican subjetividad, como marcas de superlativos (ej. “-ísimo”), y ciertos adjetivos (ej. “bueno”), adverbios (ej. “evidentemente”) y frases (ej. “sin ninguna duda”), extraídas del trabajo de Otaola Olano (1988).
- Siglas propias y sus correspondientes términos desplegados. En este caso, para hacer la correlación entre la sigla y su término desplegado, se tiene en cuenta que la letra inicial de las unidades léxicas incluidas en el término (excepto las *stopwords*) se correspondan, en el mismo orden, con las mismas letras que incluye la sigla.

En el Módulo III se gestiona el lado cliente por medio de clases/objetos JavaScript que se encuentran en un fichero PHP (*Hypertext Pre-Processor*). Los algoritmos implementados procesan los resultados de las clases lingüísticas que

se encuentran en este fichero PHP. La clase principal es “app” y esta procesa varias clases, como oraciones siglas, oraciones, unidades subjetivas, conectores discursivos, etc. El motivo de esta elección es que JavaScript es un lenguaje de programación que permite realizar actividades complejas en una página web, y PHP es un lenguaje de código abierto muy popular especialmente adecuado para el desarrollo web y que puede ser incrustado en HTML. Por tanto, al ser arText un sistema para ser empleado en la web, nos pareció una opción acertada.

La arquitectura general del Módulo III de arText puede verse en la Figura 11.

En cuanto al formato de persistencia de los documentos de arText, se usó HTML5, que es la última versión de HTML, con nuevos elementos, atributos y comportamientos, y que contiene un conjunto más amplio de tecnologías que permite a los sitios web y a las aplicaciones ser más diversas y de gran alcance. La elección de este formato se debe a que su uso es requerido por el editor CKEditor.

En relación con la exportación e importación de documentos, por una cuestión de protección de datos, se decidió que el sistema no guardase en su servidor los textos escritos por los usuarios. Por tanto, si el usuario desea guardar un texto escrito en línea en arText, debe hacerlo en local. Para ello, existen varias opciones de exportación de documentos: .pdf, .txt, .html, y .arText. Para poder importar un texto posteriormente en arText debe utilizarse el formato creado específicamente para esta aplicación, el formato .arText.

El sistema permite incluir en el documento imágenes, siempre que estas tengan asignada una URL, es decir, una dirección web. No se permite subir imágenes desde local porque, como actualmente no es necesario registrarse para utilizar el sistema, nuestro equipo no puede hacerse responsable de las imágenes subidas por los usuarios desde sus ordenadores personales. En la opción “Cómo subir imágenes desde Google Drive” (ubicada en la pestaña de Ayuda en la parte superior izquierda del editor; véase Figura 1) se explica cómo asignar una URL a una imagen.¹⁹ Todas estas cuestiones se explican con detalle en el manual de uso del sistema.²⁰

¹⁹http://sistema-artext.com/doc/como_subir_imagenes_desde_Drive.pdf

²⁰<http://sistema-artext.com/doc/manual.pdf>

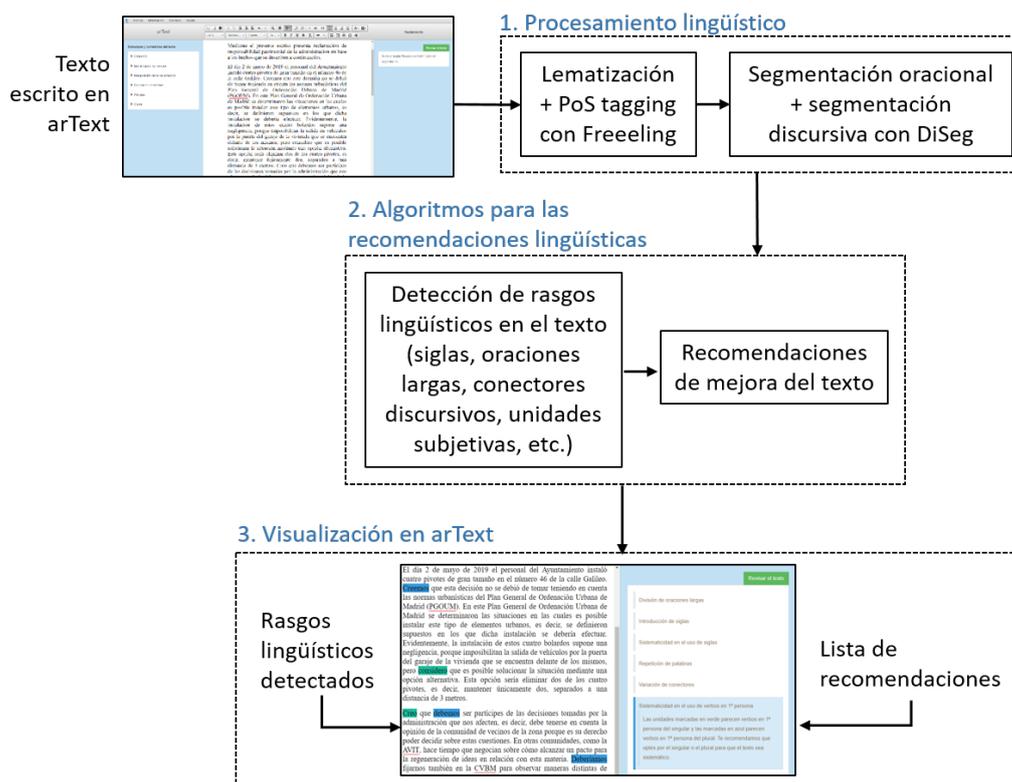


Figura 11: Arquitectura general del Módulo III de arText.

4. Resultados y evaluación

Se realizaron dos tipos de evaluaciones de arText: *data-driven* (basada en datos) y *user-driven* (basada en usuarios). Por un lado, la evaluación *data-driven* se llevó a cabo principalmente para comprobar que los algoritmos desarrollados en el Módulo III funcionaban correctamente. Para ello, en primer lugar, se compiló un corpus específico para la evaluación, que incluyó un total de 24 textos. De estos, 8 fueron del ámbito administrativo (concretamente, solicitudes), pero también se decidió añadir textos de diferentes ámbitos y géneros textuales, para confirmar que arText puede funcionar con otro tipo de textos. Así, se añadieron al corpus de evaluación 8 resúmenes de artículos de investigación del ámbito médico y 8 artículos de divulgación de revistas especializadas en turismo. En segundo lugar, un anotador con formación en lingüística y experiencia en anotación de corpus anotó manualmente los diferentes rasgos lingüísticos que se pretendían evaluar, incluidos en la Tabla 1.

En tercer lugar, se revisaron los textos del corpus automáticamente con arText. Finalmente, se compararon los resultados provenientes del análisis manual y de la revisión automática realizada con el sistema. Para ello se calculó la precisión y cobertura de los resultados de arText en contraposición con la anotación manual. Los resul-

Rasgos anotados

- a) División de oraciones largas.
 - a1) Detección de oraciones largas.
 - a2) Segmentación discursiva.
 - a3) Detección de conectores discursivos interoracionales e intraoracionales que expresan la misma relación discursiva.
- b) Introducción de siglas.
- c) Sistemática en el uso de siglas.
- e) Variación de conectores.
- f) Sistemática en el uso de verbos en 1^o persona.
 - f1) Sistemática en el uso de verbos en 1^a persona singular.
 - f2) Sistemática en el uso de verbos en 1^a persona singular.

Tabla 1: Rasgos anotados para la evaluación *data-driven* de arText.

tados se muestran en la Tabla 2. Como puede observarse, los resultados obtenidos son en general muy positivos para todos los rasgos lingüísticos analizados.

Destaca especialmente que el sistema detectó correctamente todas las oraciones largas del corpus, todos los conectores discursivos incluidos en la Base de datos 1 y todas las unidades subjetivas incluidas en la Base de datos 2.

Recomendación	Precisión	Cobertura
a1)	1	1
a2)	0,74	0,87
a3)	1	1
b)	0,76	0,68
c)	0,75	0,94
e)	1	1
f1)	0,87	0,97
f2)	1	0,97

Tabla 2: Resultados de la evaluación *data-driven* de arText.

En el caso de la recomendación d) *Repetición de palabras*, la evaluación fue ligeramente diferente: no se anotaron manualmente en el corpus todas las palabras repetidas, sino que se comprobó si las palabras listadas por arText estaban marcadas en el texto y eran realmente diferentes formas del mismo lema. Los resultados indicaron que el 93,2 % de las unidades detectadas sí lo eran. Las principales dificultades se observaron en la detección de abreviaturas como “Ilmo.” o “Sr.”.

Con respecto a las causas de los errores detectados, pueden dividirse en errores derivados de las dos herramientas de PLN integradas en arText y errores de los algoritmos propios del sistema. Con respecto a los primeros, los errores relacionados con las recomendaciones sobre palabras repetidas y verbos en primera persona vienen derivados del analizador morfosintáctico, mientras que los errores en relación con la recomendación sobre segmentación discursiva son generados por el segmentador discursivo. Al tratarse de herramientas externas, en estos casos será difícil realizar mejoras en el sistema para evitar este tipo de errores. En cambio, sí podremos trabajar en el futuro en la búsqueda de soluciones a los errores producidos por los algoritmos propios de arText. Estos tienen que ver principalmente con la cobertura en la detección de siglas (como se ha mencionado en el apartado 3.2, actualmente únicamente se detectan siglas propias) y sus correspondientes términos desplegados.

Por otro lado, la evaluación *user-driven* permitió conocer la percepción de la utilidad de la aplicación por parte de usuarios reales. Para ello, en primer lugar, se diseñó un cuestionario de valoración con la herramienta Google Forms en que se pedía a los usuarios que probasen arText y, a continuación, se les preguntaba su opinión sobre diferentes aspectos. El cuestionario contiene 3 bloques, que incluyen distintas preguntas, detalladas en el Apéndice A:

1. Breve cabecera de contextualización del proyecto.
2. Pasos a seguir para la utilización de arText.
3. Preguntas de respuesta cerrada sobre:
 - Accesibilidad.
 - Estructuración y contenidos del texto.
 - Corrección ortográfica y formato.
 - Revisión del texto.
 - Valoración final.

En segundo lugar, se distribuyó el cuestionario a 25 ciudadanos, con estudios universitarios, entre 30 y 50 años, y con manejo de internet. En general, los participantes hicieron una valoración muy positiva de la aplicación. La valoración general indica que el 84 % de los usuarios considera que arText es muy útil y el 100 % lo recomendaría a otras personas. Las opciones ofrecidas por el editor resultaron muy claras o bastante claras al 100 % de los usuarios. El aspecto mejor valorado fue el Módulo I de arText, que incluye información sobre la estructura textual, ya que un 80 % de los usuarios considera que le ha parecido de mucha utilidad. El segundo aspecto mejor valorado tiene que ver con las sugerencias de revisión lingüística del texto ofrecidas en el Módulo III, puesto que al 76 % de los usuarios le han parecido muy útiles o bastante útiles, y, además, al 80 % le ha parecido que están redactadas de manera muy clara o bastante clara. La barra de formato del Módulo II parece que también ha sido de bastante utilidad a los usuarios para maquetar el documento.

El manual de uso, la forma de exportar e importar documentos, y la manera de insertar imágenes tuvieron asimismo una valoración positiva, aunque hay un porcentaje de usuarios que reconoce que no usó el manual (24 %), que no exportó o importó ningún documento (12 % y 20 %, respectivamente), o que no insertó ninguna imagen (24 %).

Uno de los aspectos en donde hay menos consenso y que parece ser de los peor valorados es el uso de las frases sugeridas en el Módulo I, ya que, aunque un 36 % de los usuarios dice haber insertado muchas de las frases y un 28 % ha insertado bastantes, un 16 % indica que no ha insertado ninguna y un 20 % ha insertado pocas. Las repuestas sobre la utilidad del corrector ortográfico también van en esta línea.

En el Apéndice A se recogen los resultados del cuestionario de valoración de arText.

5. Conclusiones y trabajo futuro

Como se ha visto, aunque la comunicación electrónica entre el ciudadano y la Administración es el procedimiento habitual en España en el contexto de la e-Administración, son escasas las iniciativas para desarrollar herramientas TIC que tengan como objetivo mejorar la comunicación escrita entre ambos. Este trabajo busca contribuir en este sentido. Así, el objetivo de este artículo ha sido mostrar el diseño e implementación de arText, una aplicación tecnológica que ayuda a la ciudadanía a escribir textos dirigidos a la Administración pública, y que integra diferentes herramientas y recursos de PLN, como un analizador morfosintáctico y un segmentador discursivo.

La aplicación tiene forma de editor de textos en línea e incluye tres módulos, que ayudan al usuario a: I) estructurar y redactar el documento, II) darle formato y corregirlo ortográficamente, y III) revisar la adecuación del texto, mediante recomendaciones lingüísticas relacionadas con oraciones largas, conectores discursivos, siglas, palabras repetidas, verbos en primera persona del singular y plural, y unidades léxicas subjetivas. El Módulo I es especialmente relevante porque no existía hasta la fecha ninguna aplicación tecnológica que ayudase a la ciudadanía a estructurar géneros textuales del ámbito administrativo, incorporando sus apartados prototípicos, y añadiendo en cada uno de ellos los títulos, contenidos y fraseología habituales. El Módulo III es útil principalmente para revisar la adecuación del texto escrito por la ciudadanía, atendiendo a las características globales que debe cumplir este tipo de documentos, como son la adecuación gramatical y estilística, la precisión, la concisión, la objetividad y la sistematicidad.

Las evaluaciones *data-driven* y *user-driven* realizadas a arText ofrecen resultados positivos. La evaluación *data-driven* se realizó para evaluar el funcionamiento de los algoritmos utilizados en el Módulo III. En este caso, se observa que los resultados obtenidos de precisión y cobertura son en general buenos para todos los rasgos lingüísticos analizados. La evaluación *user-driven* permitió conocer la percepción de la utilidad del sistema por parte de usuarios reales, quienes hicieron una valoración muy favorable, destacando la utilidad y claridad de la aplicación. El aspecto mejor valorado fue el Módulo I, que incluye información sobre la estructura textual, y el segundo fue el Módulo III, que ofrece sugerencias de revisión lingüística del texto.

El aspecto en el que hubo menos consenso en la evaluación fue el uso de las frases sugeridas en el Módulo I. Por tanto, la revisión, modificación y ampliación de estas unidades será una de las líneas prioritarias de trabajo futuro. Asimismo, se refinarán las recomendaciones relacionadas con las siglas en el Módulo III, sobre todo para que logren una mayor cobertura, es decir, para que se detecten un mayor número de casos. También se investigará sobre la posibilidad de incluir nuevas recomendaciones no contempladas por el momento. En cuanto a la evaluación, sería interesante ampliar los textos del corpus, realizar la anotación por parte de diferentes anotadores y medir el acuerdo entre ellos (*interannotator agreement*). Asimismo, se podría realizar una evaluación *user-driven* adicional enfocada a otros colectivos, con distintos perfiles (en cuanto a edad y formación) y capacidades (en cuanto a manejo de internet). Finalmente, se investigará sobre la posibilidad de adaptar arText a otros géneros textuales, ámbitos especializados y lenguas. Actualmente el sistema ya está disponible para la redacción de géneros textuales en español de otros dos ámbitos: medicina y turismo. En el caso de la medicina, permite redactar un artículo de investigación, un artículo de revisión, una historia clínica, un resumen de artículo de investigación y un Trabajo de Fin de Grado (TFG). En el caso del turismo, permite redactar un artículo de divulgación, una entrada de blog de viajero, un informe, una normativa y un plan de negocio.

Agradecimientos

Este trabajo se ha llevado a cabo en el marco de un contrato Ramón y Cajal (RYC-2014-16935) financiado por el Ministerio de Economía, Industria y Competitividad, vinculado al Departamento de Filologías Extranjeras y sus Lingüísticas de la Facultad de Filología de la Universidad Nacional de Educación a Distancia (UNED). Los resultados se derivan de dos proyectos de investigación. Por un lado, del proyecto “Un sistema automático de ayuda a la redacción de textos especializados de ámbitos relevantes en la sociedad española actual”, financiado en la Convocatoria 2015 de Ayudas Fundación BBVA a Investigadores y Creadores Culturales. Por otro lado, del proyecto “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)”, financiado por el Ministerio de Ciencia, Innovación e Universidades en la convocatoria 2018 de Proyectos I+D del Subprograma Es-

tatal de Generación de Conocimiento (PGC2018-099694-A-I00). Ambos proyectos se han desarrollado en el marco del grupo de investigación AC-TUALing de la UNED, en colaboración con el grupo de investigación IULATERM del Institut de Lingüística Aplicada (IULA) de la Universitat Pompeu Fabra (UPF). Quiero agradecer su participación en la investigación a todos los miembros del equipo de trabajo de ambos proyectos, especialmente a M. Amor Montané y a Luis Hysa.

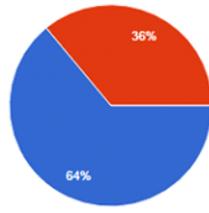
Referencias

- Alcaraz, Enrique & Brian Hughes. 2002. *El español jurídico* Ariel Derecho. Grupo Planeta.
- Atserias, Jordi, Bernardino Casas, Elisabet Comelles, Meritxell González, Lluís Padró & Muntsa Padró. 2006. Freeling 1.3. syntactic and semantic service in an open-source NLP library. En *5th International Conference on Language Resources and Evaluation (LREC)*, 48–55.
- Ayala, Pilar, Elena Domínguez Ortega, Francisca Martel Ruiz, Jacqueline Montelongo Sánchez, Dolores Morales Sosa, Orlando J. Socorro Lorenzo & Elena Suárez Manrique de Lara. 2000. Manual de normalización de documentos administrativos. Informe técnico. Universidad de Las Palmas de Gran Canaria.
- Bhatia, Vijay Kumar. 1993. *Analyzing genre: language use in professional settings*. Longman.
- Biber, Douglas, Ulla Connor & Thomas A. Upton. 2007. *Discourse on the move: Using corpus analysis to describe discourse structure*. John Benjamins. doi 10.1075/sc1.28.
- Cabré, M. Teresa, Carme Bach, Iria da Cunha, Albert Morales & Jorge Vivaldi. 2010. Comparación de algunas características lingüísticas del discurso especializado frente al discurso general: el caso del discurso económico. *Modos y formas de la comunicación humana, Ways and Modes of Human Communication* 453–460.
- Cabré, Maria Teresa. 1999. *La terminología: Representación y comunicación. Elementos para una teoría de base comunicativa y otros artículos*. IULA, Universitat Pompeu Fabra.
- Castellón, Heraclia. 2001. *El lenguaje administrativo. Formas y uso*. Editorial La Vela.
- da Cunha, Eric SanJuan, Iria, Juan M. Torres-Moreno, Marina Lloberes & Irene Castellón. 2012. DiSeg 1.0: The first system for Spanish discourse segmentation. *Expert Systems with Applications* 39(2). 1671–1678. doi 10.1016/j.eswa.2011.06.058.
- da Cunha, Iria & Mikel Iruskietia. 2010. Comparing rhetorical structures in different languages: The influence of translation strategies. *Discourse Studies* 12(5). 563–598. doi 10.1177/1461445610371054.
- da Cunha, Iria & M. Amor Montané. 2019. Textual genres and writing difficulties in specialized domains. *Revista Signos. Estudios de Lingüística* 52(99). 4–30. doi 10.4067/S0718-09342019000100004.
- da Cunha, Iria & M. Amor Montané. 2020. A corpus-based analysis of textual genres in the administration domain. *Discourse Studies* 22(1). 3–31. doi 10.1177/1461445619887538.
- Duarte, Carles & Anna Martínez. 1995. *El lenguaje jurídico*. AZ Editora.
- Gamallo Otero, Pablo, Marcos García, Iria del Río & Isaac González López. 2015. Avalingua: Natural language processing for automatic error detection. En *Learner Corpora in Language Testing and Assessment*, 35–58. John Benjamins. doi 10.1075/sc1.70.02gam.
- Giraldo, John J. 2008. *Análisis y descripción de las siglas en el discurso especializado de genoma humano y medio ambiente*: IULA, Universitat Pompeu Fabra. Tesis Doctoral.
- González Salgado, J. Antonio. 2009. El lenguaje jurídico del siglo XXI. *THEMIS: Revista de Derecho* 57. 235–245.
- Gotti, Maurizio. 2008. *Investigating specialized discourse*. Peter Lang.
- Jiménez Yañez, Ricardo-Maria. 2016. *Escribir bien es de justicia*. Pamplona, Navarra: Editorial Aranzadi.
- Mann, William & Sandra A. Thompson. 1981. Rhetorical structure theory: Toward a functional theory of text organization. *Text & Talk* 8(3). 243–281. doi 10.1515/text.1.1988.8.3.243.
- de Miguel Aparicio, Elena. 2000. El texto jurídico-administrativo: Análisis de una orden ministerial. *Círculo de Lingüística Aplicada a la Comunicación* 4.
- Montolío Durán, Estrella. 2012. *Hacia la modernización del discurso jurídico*. Publicacions i Edicions de la Universitat de Barcelona.

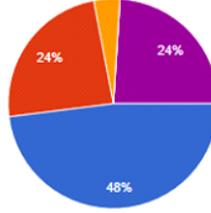
- Otaola Olano, Concepción. 1988. La modalidad (con especial referencia a la lengua española). *Revista de Filología Española* 68(1/2). 97–117.
- Parodi, Giovanni. 2010. *Academic and professional discourse genres in Spanish*. John Benjamins. doi: 10.1075/scl.40.
- Rodríguez-Aguilera, Cesáreo. 1969. *El lenguaje jurídico*. Barcelona Bosch.
- Samaniego, Eva. 2005. El lenguaje jurídico: Peculiaridades del español jurídico. En *Lengua y Sociedad: Aportaciones recientes en lingüística cognitiva, lenguas de contacto, lenguajes de especialidad y lingüística del corpus*, 273–310. Universidad de Valladolid, Centro Buendia.
- Swales, John M. 1990. *Genre analysis: English in academic and research settings*. Cambridge University Press.
- Sánchez Alonso, Fernando. 2014. *Lenguaje y estilo administrativo. Redacción de documentos*. Escuela de Formación e Innovación. Administración Pública.
- Tofiloski, Milan, Julian Brooke & Maite Taboada. 2009. A syntactic and lexical-based discourse segmenter. En *47th Annual Meeting of the Association for Computational Linguistics*, 77–80.
- Upton, Thomas A. & Mary A. Cohen. 2009. An approach to corpus-based discourse analysis: The move analysis as example. *Discourse Studies* 11(5). 585–605. doi: 10.1177/1461445609341006.
- Van Dijk, Teun A. 1977. *Text and context: Explorations in the semantics and pragmatics of discourse*. Longman.
- Van Dijk, Teun A. 1989. *La ciencia del texto: un enfoque interdisciplinario*. Paidós Comunicación.
- Zhou, Lanjun, Li Binyang, Wei Zhongyu & Wong Kam-Fai. 2014. The CUHK Discourse Treebank for Chinese: Annotating Explicit Discourse Connectives for the Chinese Treebank. *International Conference on Language Resources and Evaluation (LREC)* 942–949.

A. Resultados de las preguntas del cuestionario de valoración de arText

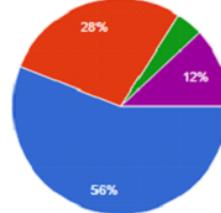
¿Le han parecido claras las opciones ofrecidas por el editor? ¿Le ha parecido útil el manual de uso? ¿Le ha parecido clara la manera de exportar documentos?



● Muy claras
● Bastante claras
● Poco claras
● Nada claras

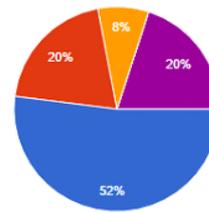


● Mucho
● Bastante
● Poco
● Nada
● No lo he usado

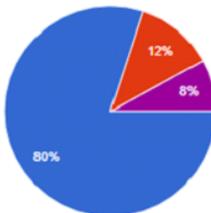


● Muy clara
● Bastante clara
● Poco clara
● Nada clara
● No he exportado ningún documento

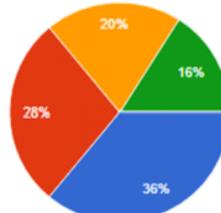
¿Le ha parecido clara la manera de subir al sistema documentos guardados previamente? ¿Le ha parecido útil la información incluida en la opción "Estructuración y contenidos de arText"? ¿Ha insertado en el texto alguna de las frases sugeridas?



● Muy clara
● Bastante clara
● Poco clara
● Nada clara
● No he subido ningún documento

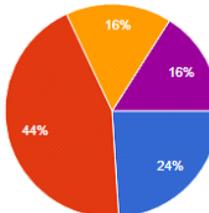


● Mucho
● Bastante
● Poco
● Nada
● No la he usado

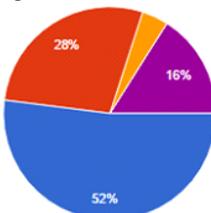


● Muchas
● Bastantes
● Pocas
● Ninguna

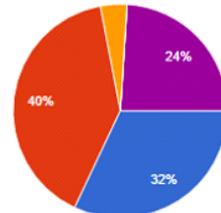
¿Le ha parecido útil el corrector ortográfico, que subraya las palabras desconocidas en rojo? ¿Le han parecido útiles las opciones de formato que incluye la barra superior de la página web? ¿Le ha parecido clara la manera de insertar imágenes en el texto?



● Mucho
● Bastante
● Poco
● Nada
● No lo he usado

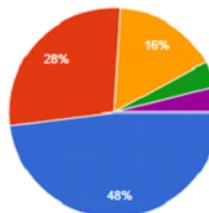


● Mucho
● Bastante
● Poco
● Nada
● No las he usado

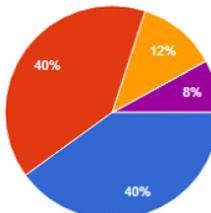


● Muy clara
● Bastante clara
● Poco clara
● Nada clara
● No he insertado imágenes

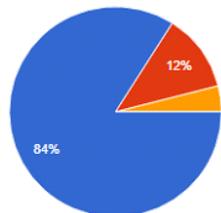
¿Le han parecido útiles las sugerencias ofrecidas al seleccionar la opción "Revisar el texto"? ¿Le ha parecido clara la redacción de las sugerencias ofrecidas? En general, ¿considera útil este sistema?



● Mucho
● Bastante
● Poco
● Nada
● No he usado esta opción



● Mucho
● Bastante
● Poco
● Nada
● No he usado esta opción



● Mucho
● Bastante
● Poco
● Nada