

Un redactor asistido para adaptar textos administrativos a lenguaje claro

A writing assistant to adapt administrative texts into plain language

Iria da Cunha

Universidad Nacional de Educación a Distancia

iriad@flog.uned.es

Resumen: El lenguaje claro aboga por que los textos dirigidos a los ciudadanos estén redactados en un lenguaje más sencillo y transparente, para que estos puedan entender fácilmente el mensaje que se les quiere transmitir. En este contexto, nuestro objetivo es desarrollar un redactor asistido para el español que ayude al personal de la Administración pública a escribir en lenguaje claro los textos que dirige a la ciudadanía. El sistema, gratuito y en línea, integra diferentes herramientas de Procesamiento de Lenguaje Natural (PLN) para detectar en los textos escritos por los usuarios los rasgos lingüísticos que interfieren con las recomendaciones sobre lenguaje claro. Asimismo, ofrece al usuario información para hacer más sencillo su texto. Para evaluar los algoritmos se empleó un corpus anotado manualmente, y las medidas de precisión y cobertura. Los resultados son muy positivos, aunque también reflejan algunos aspectos que se pueden mejorar en el futuro.

Palabras clave: Lenguaje claro, redacción asistida, Procesamiento de Lenguaje Natural (PLN), Administración pública.

Abstract: Plain language advocates that texts addressed to citizens should be written in simpler and more transparent language, so that they can easily understand the message to be conveyed. In this context, our aim is to develop an assisted writing tool for Spanish to help Public Administration staff to write texts addressed to citizens in plain language. The system is free and online. It integrates different Natural Language Processing (NLP) tools to detect the linguistic features that interfere with plain language recommendations in the texts written by users. It also provides users with information to make their text clearer. A manually annotated corpus, and precision and recall measures were used to evaluate the algorithms. The results are very positive, although they also highlight some aspects that could be improved in the future.

Keywords: Plain Language, Assisted Writing, Natural Language Processing (NLP), Public Administration.

1 Introducción

El lenguaje claro es una corriente que aboga por que los textos que se dirijan a los ciudadanos estén redactados en un lenguaje más sencillo y transparente, para que estos puedan entender el mensaje que se les quiere transmitir y, así, contribuir a que ejerzan sus derechos y a que cumplan con sus obligaciones. Según el sitio web de la International Plain Language Federation,¹ la comunicación en lenguaje claro tiene unos rasgos concretos: “A communication

is in plain language if its wording, structure, and design are so clear that the intended readers can easily find what they need, understand what they find, and use that information”.

Esta corriente adquiere una relevancia especial en el marco de la modernización del lenguaje jurídico y administrativo, uno de los grandes retos de la Administración española (Cassany, 2005; Montolío, 2012; Carretero y Fuentes, 2019; Montolío y Tascón, 2020; Bayés, 2021). El reto es cambiar una tradición en la Administración que hace que los textos que recibe la ciudadanía (notificaciones, requerimientos, resoluciones, multas, etc.)

¹ <https://www.iplfederation.org/plain-language/>

tengan unas características lingüísticas que “oscurecen” el texto, como párrafos larguísimos, oraciones con múltiples subordinadas e incisos, verbos en voz pasiva, en gerundio y en participio, formas verbales arcaicas, terminología y fraseología muy compleja, multitud de negaciones y formas subjetivas, entre otros rasgos (De Miguel, 2000; Alcaraz, Hugues y Gómez, 2014).

Por esta razón, en los últimos años se han elaborado guías y manuales para ayudar a redactar en lenguaje claro los textos que se dirigen a los ciudadanos. Algunos ejemplos son las publicaciones de la Comisión Europea (2015), Carretero et al. (2017), y Montolío y Tascón (2017). Sin embargo, a pesar de estas meritorias aportaciones, son escasos los trabajos que establecen sinergias entre el lenguaje claro en español y la tecnología.

En este contexto, el objetivo de la presente investigación es desarrollar el primer redactor asistido para el español que ayude al personal de la Administración pública a escribir en lenguaje claro los textos que dirige a la ciudadanía. Este redactor asistido se llama “arText claro”, puede utilizarse gratuitamente y está disponible en línea desde la dirección: <http://sistema-artext.com/>. El sistema integra diferentes herramientas de Procesamiento de Lenguaje Natural (PLN), gracias a las cuales logra detectar en los textos escritos por los usuarios los rasgos lingüísticos que interfieren con las recomendaciones sobre lenguaje claro más habituales. Asimismo, ofrece al usuario información para hacer más claro y sencillo su texto. Las recomendaciones que integra el sistema están relacionadas con tres niveles de la lengua: discursivo, morfosintáctico y léxico.

En el apartado 2 se incluye un breve estado la cuestión sobre herramientas que tienen que ver con la revisión de textos, especialmente, en relación con el lenguaje claro. En el apartado 3 se muestran las fases de la metodología de la investigación. En el apartado 4 se detallan las funcionalidades y la implementación del redactor asistido. En el apartado 5 se explica la evaluación del sistema. Finalmente, en el apartado 6, se exponen las conclusiones y se plantean algunas líneas de trabajo futuro.

2 Estado de la cuestión

Actualmente, existen herramientas tecnológicas que ayudan a revisar y corregir textos en español. Estas herramientas incluyen diferentes

funcionalidades relacionadas con diversos niveles de la lengua. Estos niveles pueden ir desde el más simple, como el ortográfico, hasta el más complejo, como el discursivo o pragmático, pasando por el léxico y el sintáctico. Además del clásico corrector de los procesadores de textos como Word y OpenOffice, hay otros sistemas destacables que se han desarrollado en los últimos años, como, por ejemplo, LanguageTool,² OutWrite,³ Stilus⁴ y Estilector.⁵ Sin embargo, ninguno de estos correctores está diseñado específicamente para ayudar a redactar en lenguaje claro.

Para la lengua inglesa, sí existen algunos ejemplos muy recientes. Por ejemplo, los sistemas comerciales con demos gratuitas Hemingway Editor⁶ y VisibleThread.⁷ Estos sistemas otorgan una puntuación a la legibilidad o comprensibilidad del texto. Para asignar estas puntuaciones, tienen en cuenta diferentes cuestiones gramaticales y de estilo que hacen que los textos resulten más claros. Por ejemplo, recomiendan limitar el uso de los adverbios y de la voz pasiva; proponen alternativas más simples para algunas expresiones utilizadas en el texto; y señalan las oraciones que resultan difíciles de leer para que el usuario las acorte o divida. Cada una de estas cuestiones se marca en el texto con un color diferente.

En el ámbito de las tecnologías de la lengua en español, hay aún mucho camino por recorrer en relación con el lenguaje claro. Por ejemplo, es de destacar la herramienta en versión beta Clara,⁸ gratuita y en línea, que permite realizar un test de claridad a textos en español, principalmente documentos administrativos y contratos de servicios. A partir de un texto escrito por el usuario, de un mínimo de 40 palabras y un máximo de 120, la herramienta ofrece un porcentaje global de claridad. Asimismo, indica un porcentaje específico de claridad para cada una de las métricas que incluye y ofrece propuestas de mejora. La herramienta incluye nueve métricas de evaluación (Torrijos y Oquendo, 2021: 123-124):

² <https://www.languagetool.org>

³ <https://es.outwrite.com/>

⁴ <http://www.mystilus.com/>

⁵ <http://www.estilector.com/>

⁶ <https://hemingwayapp.com/>

⁷ <https://www.visiblethread.com/>

⁸ <https://clara.comunicacionclara.com/>

- *Métrica 1: uso de palabras fuera del diccionario.*
- *Métrica 2: uso de conectores discursivos.*
- *Métrica 3: uso de la puntuación.*
- *Métrica 4: citas y referencias a leyes.*
- *Métrica 5: uso de la voz pasiva.*
- *Métrica 6: uso de nexos subordinados.*
- *Métrica 7: uso de tecnicismos financieros y administrativos.*
- *Métrica 8: uso de palabras del ranking de las 1000 más comunes en español.*
- *Métrica 9: número medio de palabras por frase.*

No obstante, Clara no marca en el texto escrito por el usuario las cuestiones específicas que interfieren con el lenguaje claro, ya que no es un redactor asistido ni un corrector, sino un sistema de medición de la claridad textual.

3 Metodología

La metodología de esta investigación incluye tres fases, que se detallan a continuación:

Fase 1. Búsqueda de fuentes bibliográficas sobre lenguaje claro en el ámbito administrativo en español. Para seleccionar las recomendaciones sobre lenguaje claro que integra el redactor asistido, partimos del trabajo de Da Cunha y Escobar (2021). En este estudio se analizan las principales fuentes sobre lenguaje claro en el ámbito jurídico-administrativo en español peninsular (Vilches y Sarmiento, 2010; Ministerio de Justicia, 2011; Comisión Europea, 2015; Jiménez Yáñez, 2016; Carretero *et al.*, 2017; Montolío y Tascón, 2017; Carretero, 2019) y se cuantifican las recomendaciones que recogen, para seleccionar las más frecuentes. A continuación, se dividen las recomendaciones seleccionadas en función de los tres niveles de la lengua mencionados en el apartado 1:

- Nivel discursivo. Ejemplo de recomendación: “Se recomienda redactar oraciones cortas”.
- Nivel morfosintáctico. Ejemplo de recomendación: “Se recomienda utilizar la voz activa en vez de la voz pasiva”.
- Nivel léxico. Ejemplo de recomendación: “Se recomienda evitar los arcaísmos”.

Fase 2. Diseño e implementación del sistema.

Fase 2a. Selección de herramientas de PLN en abierto para la lengua española que permitan un procesamiento lingüístico del texto escrito por el usuario, concretamente:

- Lematización.
- Análisis *Part of Speech* (POS).
- Segmentación oracional.
- Segmentación discursiva intraoracional.

Fase 2b. Diseño de algoritmos que detecten en el texto escrito por el usuario los rasgos lingüísticos que interfieren con el lenguaje claro. Para ello se tienen en cuenta las recomendaciones seleccionadas en la Fase 1 y el resultado del procesamiento lingüístico del texto de la Fase 2a. Por ejemplo, en el caso de la recomendación “Se recomienda redactar oraciones cortas”, el algoritmo detecta todas las oraciones del texto y contabiliza las palabras que incluye cada una. Si una oración supera las 25 palabras, la subraya en amarillo en el texto escrito por el usuario y le ofrece la recomendación correspondiente para adaptarla a lenguaje claro. En el caso de la recomendación “Se recomienda utilizar la voz activa en vez de la voz pasiva”, el algoritmo marca en el texto del usuario los verbos en voz pasiva obtenidos gracias al analizador POS y le ofrece la recomendación correspondiente.

Fase 2c. Redacción de las recomendaciones sobre lenguaje claro ofrecidas al usuario. En cada recomendación se incluye la siguiente información:

- Título de la recomendación.
- Sugerencia para la adaptación a lenguaje claro.
- Ejemplo (en caso de que se considere necesario para que el usuario entienda la recomendación).

En el Anexo 1 se detallan todas las recomendaciones que ofrece el sistema.

Fase 2d. Implementación del redactor asistido (los detalles técnicos se incluyen en el apartado 4).

Fase 3. Evaluación del sistema. Para evaluar el sistema, se compila un corpus de textos del ámbito de la Administración, concretamente, resoluciones del BOAM (Boletín Oficial del Ayuntamiento de Madrid). Una persona con formación en lingüística y experiencia en

anotación de corpus anota manualmente en cada texto los diferentes rasgos lingüísticos correspondientes a las recomendaciones que se quieren evaluar (detallados en el apartado 5). A continuación, se aplica el sistema sobre los textos del corpus para obtener automáticamente las recomendaciones sobre lenguaje claro. Finalmente, se calcula la precisión y cobertura de los resultados del sistema en contraposición con la anotación manual.

4 *Funcionalidades e implementación del redactor asistido*

El sistema se desarrolló en un entorno Linux con un servidor Apache. También se usaron distintos recursos en el *back-end* (Bash, Perl y PHP, con un entorno de trabajo Laravel) y en el *front-end* (HTML, CSS, JavaScript, con AJAX y jQuery). El sistema está optimizado para utilizarse con el navegador Google Chrome.

El redactor integra dos herramientas de PLN existentes para el español que permiten procesar lingüísticamente el texto escrito por el usuario:

- 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. Este analizador permite detectar rasgos lingüísticos que son utilizados por los algoritmos del sistema en las recomendaciones que tienen que ver principalmente con el nivel morfosintáctico, como, por ejemplo, los verbos en voz pasiva, los gerundios, los participios, los verbos en futuro de subjuntivo, o los verbos en 1.^a persona del singular y del plural.
- Un segmentador discursivo (Da Cunha et al., 2010, 2012b), que permite dividir el texto en oraciones y, además, en segmentos discursivos intraoracionales, a partir de la definición de 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”. Este segmentador permite detectar los rasgos utilizados por los algoritmos en las recomendaciones relativas al nivel discursivo, como la segmentación

discursiva de las oraciones largas y la sugerencia de conectores alternativos.

Además de integrar estas dos herramientas, el sistema incluye diversos algoritmos desarrollados en el marco de nuestra investigación. Estos algoritmos toman como entrada el texto procesado lingüísticamente por las dos herramientas de PLN mencionadas y detectan en el texto del usuario los rasgos lingüísticos necesarios para poder ofrecer las recomendaciones asociadas a cada uno de ellos. Estos rasgos son:

- Párrafos-oración.
- Párrafos largos, con un umbral de 135 palabras (teniendo en cuenta las fuentes recopiladas en la Fase 1).
- Oraciones largas, con un umbral de 25 palabras (teniendo en cuenta las fuentes recopiladas en la Fase 1).
- Conectores discursivos interoracionales e interoracionales que evidencian ocho relaciones discursivas: antítesis, causa, concesión, condición, contraste, propósito, reformulación y resumen (Da Cunha et al., 2012a).
- Nominalizaciones verbales. Concretamente, el algoritmo detecta los sustantivos acabados en *-ción* (en singular y plural) que comienzan por minúscula, excepto los incluidos en una lista de exclusión predefinida que integra términos del ámbito de la Administración, como “licitación” y “notificación” (Da Cunha, 2022).
- Unidades que expresan negación de una lista predefinida, que incluye unidades como “no”, “ni” y “ninguno”.
- Unidades léxicas que indican subjetividad, como ciertos adjetivos (ej. “bueno”), adverbios (ej. “evidentemente”) y frases (ej. “sin ninguna duda”), extraídas de Otaola (1988).
- Siglas propias (Giraldo, 2008) y sus correspondientes términos desplegados. 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. Ej. “Plan de Emergencias Invernales

- del Ayuntamiento de Madrid” > “PEIAM”.
- Términos difíciles de entender que tienen una variante sinonímica más sencilla de una lista predefinida (Da Cunha, 2022). Por ejemplo, la variante más sencilla del verbo “adverar” es “certificar” y la variante más sencilla del sustantivo “aquiescencia” es “consentimiento”.
 - Expresiones difíciles de entender que tienen una variante sinonímica más sencilla de una lista predefinida (Da Cunha, 2022). Por ejemplo, el latinismo “ad valorem” tiene como variante en español “según en valor” y la expresión arcaica “a tenor de” tiene como variante más clara “según”.
 - Palabras poco precisas de una lista predefinida (Da Cunha, 2022), como “cosa”, “varios”, “alguno”, “muy” y “poco”.
 - Expresiones redundantes de una lista predefinida (Da Cunha, 2022), como “está claro que”, “mi opinión personal” y “como es bien sabido”.

- Palabras largas que tienen una variante sinonímica más breve de una lista predefinida (Da Cunha, 2022), como “gratuitamente/gratis” y “encomendar/encargar”.

En total, el redactor incluye 22 recomendaciones sobre lenguaje claro. Como se ha indicado, en el Anexo 1 se detallan todas ellas, divididas en función de los tres niveles de la lengua mencionados en el apartado 3.

En la Figura 1 se ofrece una captura de pantalla de “arText claro” en donde se muestra la recomendación sobre párrafos largos en un texto del corpus de evaluación.

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. Para ello, existen varias opciones de exportación de documentos en local: .doc, .pdf, .txt, .html, etc. Para poder importar un texto posteriormente en el redactor, debe utilizarse un formato creado específicamente para este sistema: .artext.

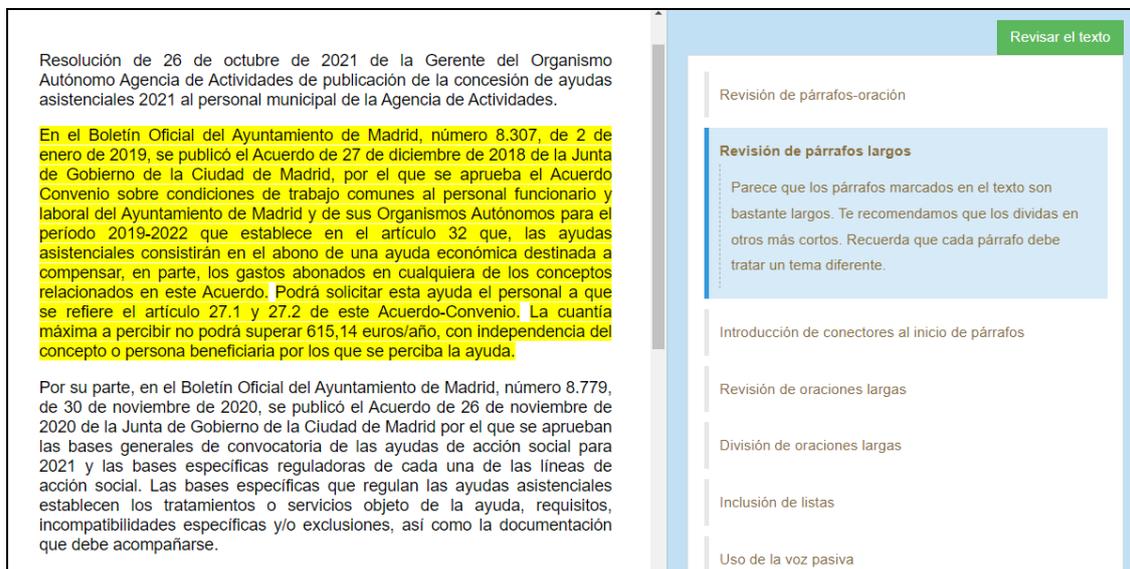


Figura 1: Captura de pantalla de “arText claro” en donde se muestra la recomendación sobre párrafos largos en un texto del corpus de evaluación.

5 Evaluación

Como se avanzaba en el apartado 3, una vez implementado el sistema, se llevó a cabo una evaluación *data-driven* utilizando un corpus formado por 10 resoluciones del BOAM publicadas en el año 2021, que en total suman

8.052 palabras. El texto más corto tiene 436 palabras y el texto más largo, 1398. En el Anexo 2 se recogen los títulos de las resoluciones empleadas.

Para comparar los resultados del sistema con los resultados de la anotación manual, se calculó la precisión y cobertura de los rasgos

lingüísticos anotados en los textos del corpus. Los rasgos anotados se incluyen en la Tabla 1.

Nivel de	Rasgos anotados
Discursivo	a1. Párrafos-oración
	a2. Párrafos largos
	a3. Párrafos que no comienzan por un conector discursivo
	a4. Oraciones largas
	a5. Oraciones largas que pueden dividirse en segmentos discursivos
	a6. Conectores discursivos que aparecen 3 veces o más
	a7. Listas
Morfo-sintáctico	b1. Verbos en voz pasiva
	b2. Gerundios
	b3. Participios
	b4. Verbos en futuro de subjuntivo
	b5. Verbos en 1.ª persona del plural y del singular
	b6. Nominalizaciones verbales
	b7. Unidades que expresan negación
Léxico	c1. Unidades que expresan subjetividad
	c2. Siglas que no aparecen con su correspondiente término desplegado la 1.ª vez que aparecen en el texto
	c3. Términos desplegados para los que se introdujo su sigla previamente en el texto
	c4. Términos difíciles de entender que tienen una variante sinónima más sencilla
	c5. Expresiones difíciles de entender que tienen una variante más sencilla
	c6. Palabras poco precisas
	c7. Expresiones redundantes
	c8. Palabras largas que tienen una variante sinónima más breve

Tabla 1: Rasgos lingüísticos anotados manualmente en el corpus de evaluación.

Los resultados de la evaluación de cada recomendación se muestran en la Tabla 2 (“ID” se refiere al identificador de la recomendación).

ID	Precisión	Cobertura
a1	0,99	0,74
a2	0,7	0,8
a3	0,84	0,71
a4	1	0,7
a5	1	0,89
a6	1	1
a7	1	0,83
b1	1	0,67
b2	1	1
b3	0,86	1
b4	1	1

b5	0,17	1
b6	1	1
b7	1	1
c1	Rasgo sin ocurrencias en el corpus	
c2	0,6	1
c3	1	1
c4	1	1
c5	1	1
c6	1	1
c7	Rasgo sin ocurrencias en el corpus	
c8	1	1

Tabla 2: Resultados de la evaluación del sistema.

Como puede apreciarse, los resultados obtenidos son, en general, positivos para la mayor parte de las recomendaciones evaluadas. Destaca especialmente que el sistema obtiene la máxima precisión y cobertura en relación con la detección de 10 rasgos: los conectores que aparecen más de tres veces en el texto, los gerundios, los verbos en futuro de subjuntivo, las nominalizaciones verbales, las unidades que expresan negación, los términos desplegados para los que se introdujo su sigla previamente en el texto, los términos difíciles de entender que tienen una variante sinónima más sencilla, las expresiones difíciles de entender que tienen una variante más sencilla y las palabras poco precisas.

En cuanto a las recomendaciones relacionadas con los párrafos (a1, a2, a3), también se obtienen resultados positivos, aunque bajan ligeramente. Esto se debe, principalmente, a signos de puntuación que interfieren en la segmentación oracional, lo cual tiene consecuencias en la detección correcta de párrafos:

- Citas a artículos y leyes. Ej. “los artículos 8.1 y 46.1 de la Ley”.
- Elementos numerados con puntuación. Ej. “1.”, “2.”.
- Enlaces web. Ej. “en la intranet/extranet municipal ayre (<https://ayre.madrid.es>) y en la web <https://jubilacion.madrid.es>”.

En cuanto a las recomendaciones relacionadas con la detección de oraciones largas y segmentación discursiva (a4, a5), la precisión es muy alta, pero en cambio baja ligeramente la cobertura. El motivo es que en este tipo de textos administrativos en ocasiones hay oraciones que acaban con una coma o directamente sin ningún signo de puntuación.

Por tanto, el sistema no logra recuperarlas. Por ejemplo:

- (1) “En su virtud, de conformidad con el Acuerdo de 27 de junio de 2019 [...] la competencia para la ejecución de los planes y programas de formación de los empleados y directivos del Ayuntamiento de Madrid,”
- (2) “En virtud de las facultades que me han sido conferidas por [...], esta Gerencia”

En cuanto a la detección de rasgos morfosintácticos, los que no alcanzan la máxima precisión y cobertura son los siguientes:

- Verbos en voz pasiva (b1). En este caso, los errores en la cobertura (0,67) provienen del procesamiento lingüístico con Freeling. Por ejemplo, no se detectan las formas “hubiera sido cesado” o “han sido conferidas”.
- Participios (b3). En esta ocasión el problema es de precisión (0,86) y se debe a la desambiguación de Freeling. Por ejemplo, se detecta “propuesta” y “puesto” como participios cuando en el texto tienen función de sustantivo (“la propuesta”, “el puesto de trabajo”).
- Verbos en 1.^a persona del plural y del singular (b5). Este ha sido el rasgo que ha obtenido peores resultados en la evaluación, con un 0,17 de precisión, debido también a problemas en la desambiguación. Por ejemplo, se detectan como 1.^a persona del singular las formas “sea” y “haga”, cuando en realidad en el texto son formas de 3.^a persona singular (“sea esta accidental o intencionada”, “cuando la previsión meteorológica haga previsible”).

En relación con el nivel léxico, como se ha visto, se obtienen muy buenos resultados. Únicamente baja la precisión (0,6) en la detección de siglas que no aparecen con su correspondiente término desplegado la primera vez que aparecen en el texto (c2). El motivo principal es que en estos documentos suele haber números romanos (ej. “III”, “IV”) y el sistema los marca erróneamente como siglas. También señala otras secuencias de letras en mayúscula (como los acrónimos) que sí aparecen con su término desplegado y que, por tanto, no sería pertinente marcar en el texto del

usuario. Ej. “Plan Territorial Superior de la Comunidad de Madrid (PLATERCAM)”.

Finalmente, como se observa en la Tabla 2, hay dos recomendaciones del nivel léxico que no se han podido evaluar (c1 y c7), porque no se han detectado en el corpus unidades que expresan subjetividad ni expresiones redundantes.

6 Conclusiones y líneas de trabajo futuro

Como se ha visto en este trabajo, el lenguaje claro es un tema de investigación en el que aún queda mucho por explorar, especialmente desde el punto de vista del PLN. El objetivo de este trabajo ha sido desarrollar el primer redactor asistido para el español para escribir textos administrativos en lenguaje claro. La herramienta desarrollada tiene un gran potencial de aplicación en las diferentes dependencias de la Administración pública, como ayuntamientos, diputaciones, consejerías, ministerios, etc. Los empleados públicos dispondrán, así, de una herramienta tecnológica que les ayudará a revisar sus textos y adaptarlos al lenguaje claro.

La evaluación de “arText claro” ofrece buenos resultados, aunque aún quedan cuestiones que se pueden mejorar, como la precisión en la detección de siglas, o las interferencias con los signos de puntuación que provocan errores en la detección de párrafos y oraciones largas. También sería interesante ampliar el corpus de evaluación con más textos anotados manualmente, de otros géneros textuales del ámbito administrativo, con el objetivo de validar los resultados obtenidos en esta investigación.

Con respecto a la metodología, una posible línea de trabajo futuro sería la aplicación de estrategias de aprendizaje automático. Este enfoque ya está siendo utilizado en algunas investigaciones, como es el caso del sistema de medición de la claridad textual Clara (Torrijos y Oquendo, 2021), mencionado en el apartado 2. La principal dificultad de este enfoque, sin embargo, es la necesidad de contar con corpus muy extensos de textos originales y sus correspondientes textos clarificados.

Otra de nuestras líneas de investigación será llevar a cabo estudios de evaluación de la percepción de la claridad y de la comprensión de los textos escritos con la herramienta por parte de los destinatarios. Finalmente, nuestro

objetivo es implementar el sistema en las dependencias de la Administración pública española.

Agradecimientos

Este trabajo se deriva del proyecto de investigación titulado “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 Estatal de Generación de Conocimiento (referencia PGC2018-099694-A-I00), y desarrollado en el Departamento de Filologías Extranjeras y sus Lingüísticas de la Facultad de Filología de la Universidad Nacional de Educación a Distancia (UNED), en el marco del grupo de investigación ACTUALing y en colaboración con el grupo IULATERM (IULA-UPF).

Bibliografía

- Alcaraz, E., B. Hugues, y A. Gómez. 2014. *El español jurídico*. 3.^a edición. Ariel, Barcelona.
- Atserias, J., B. Casas, E. Comelles, M. González, Ll. Padró, y M. Padró. 2006. FreeLing 1.3. Syntactic and semantic services in an open-source NLP library. En *LREC 2006 Proceedings. 5th Edition of the International Conference on Language Resources and Evaluation*, páginas 48-55, European Language Resources Association (París).
- Da Cunha, I. (Ed.). 2022. *Lenguaje claro y tecnología en la Administración*. Granada, Comares. En prensa.
- Da Cunha, I. y M. Á. Escobar. 2021. Recomendaciones sobre lenguaje claro en español en el ámbito jurídico-administrativo: análisis y clasificación. *Pragmalingüística*, 29:129-148.
- Da Cunha, I., E. SanJuan, J-M. Torres-Moreno, M. T. Cabré, y G. Sierra. 2012a. A Symbolic Approach for Automatic Detection of Nuclearity and Rhetorical Relations among Intra-sentence Discourse Segments in Spanish. *Lecture Notes in Computer Science*, 7181:462-474.
- Da Cunha, I., E. SanJuan, J-M. Torres-Moreno, M. Lloberes, e I. Castellón. 2012b. DiSeg 1.0: The First System for Spanish Discourse Segmentation. *Expert Systems with Applications*, 39(2):1671-1678.
- Da Cunha, I., E. SanJuan, J-M. Torres-Moreno, M. Lloberes, e I. Castellón. 2010. DiSeg: Un segmentador discursivo automático para el español. *Procesamiento del Lenguaje Natural*, 45:145-152.
- Bayés, M. 2021. Análisis del impacto de una selección de (meta)indicaciones de redacción clara en la percepción de claridad de un documento administrativo: estudio de caso. Tesis doctoral, Universitat de Barcelona.
- Carretero, C. 2019. *Comunicación para juristas*. Tirant lo Blanch, Valencia.
- Carretero, C., y J. C. Fuentes. 2019. La claridad del lenguaje jurídico. *Revista del Ministerio Fiscal*, 8:7-40.
- Carretero, C., J. M. Pérez, L. Lanne-Lenne, y G. de los Reyes. 2017. *Lenguaje Claro. Comprender y hacernos entender*. Instituto de Lectura Fácil y Clarity, Sevilla, Madrid.
- Cassany, D. 2005. Plain Language in Spain. *Clarity*, 53:41-44.
- Comisión Europea. 2015. *Cómo escribir con claridad*. Oficina de Publicaciones de la Unión Europea, Luxemburgo.
- De Miguel, E. 2000. El texto jurídico-administrativo: análisis de una orden ministerial. *Círculo de lingüística aplicada al a comunicación*, 4: en línea. Disponible en: <https://webs.ucm.es/info/circulo/no4/demiguel.htm>
- Giraldo, J. J. 2008. Análisis y descripción de las siglas en el discurso especializado de genoma humano y medio ambiente. Tesis doctoral. Universitat Pompeu Fabra, Institut de Lingüística Aplicada (IULA), Barcelona.
- Jiménez Yáñez, R. M. 2016. *Escribir bien es de justicia*. Aranzadi, Cizur Menor.
- Ministerio de Justicia. 2011. *Informe de la Comisión de modernización del lenguaje jurídico*. Ministerio de Justicia, Madrid. En línea.

- Montolío, E. 2012. La modernización del discurso jurídico español impulsada por el Ministerio de Justicia. Presentación y principales aportaciones del Informe sobre el lenguaje escrito. *Revista de Llengua i Dret*, 57:95-121.
- Montolío, E. y M. Tascón. 2020. *El derecho a entender: la comunicación clara, la mejor defensa de la ciudadanía*. La Catarata, Madrid.
- Montolío, E. y M. Tascón. 2017. *Comunicación Clara. Guía Práctica*. Ayuntamiento de Madrid, Madrid.
- Otaola, C. 1988. La modalidad (con especial referencia a la lengua española). *Revista de Filología Española*, 68(1):97-117.
- Tofiloski, M., J. Brooke, y M. Taboada. 2009. A syntactic and lexical-based discourse segmenter. En *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, páginas 77-80, Association for Computational Linguistics (Singapur).
- Torrijos, C. y S. Oquendo. 2021. ¡Hola! Soy clara y mido la claridad de tu texto” *Archiletras Científica*, Vol. VI:119-133.
- Vilches, F. y R. Sarmiento. 2010. *Manual de lenguaje Jurídico-Administrativo*. Dykinson, Madrid.

A Anexo 1: Recomendaciones que ofrece “arText claro”

a) Recomendaciones del nivel discursivo

- a1. Título de la recomendación: *Revisión de párrafos-oración*.
Texto de la recomendación: *Parece que los párrafos marcados en el texto solo incluyen una oración. Ten en cuenta que suele recomendarse que cada párrafo incluya al menos dos oraciones.*
- a2. Título de la recomendación: *Revisión de párrafos largos*.
Texto de la recomendación: *Parece que los párrafos marcados en el texto son bastante largos. Te recomendamos que los dividas en otros más cortos. Recuerda que cada párrafo debe tratar un tema diferente.*
- a3. Título de la recomendación: *Introducción de conectores al inicio de párrafos*.
Texto de la recomendación: *Parece que los párrafos marcados en el texto no comienzan con una marca explícita que los relacione con su*

párrafo anterior. Te recomendamos que enlaces los diferentes párrafos a través de conectores discursivos. Haz clic en cada acción para ver sugerencias de conectores que pueden servirte de ayuda para introducir párrafos:

*Introducir un tema nuevo
Marcar un orden
Distinguir
Seguir el mismo tema
Enfatizar y reformular
Detallar
Resumir
Terminar
Indicar causa
Indicar consecuencia
Indicar oposición
Indicar objeción
Mostrar elementos en forma de lista*

- a4. Título de la recomendación: *Revisión de oraciones largas*.
Texto de la recomendación: *Las oraciones marcadas son muy largas. Te recomendamos que las revises. Por ejemplo, podrías dividir la oración en otras más cortas o eliminar la información poco relevante.*
- a5. Título de la recomendación: *División de oraciones largas*.
Texto de la recomendación: *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. Si decides dividir la oración en otras más cortas, te recomendamos que utilices conectores. Haz clic en las unidades marcadas en rojo en el texto para ver sugerencias de conectores alternativos.*
- a6. Título de la recomendación: *Variación de conectores*.
Texto de la recomendación: *Los conectores de la lista siguiente se repiten varias veces en el texto. Haz clic en cada conector para ver sugerencias de conectores alternativos.*
- a7. Título de la recomendación: *Inclusión de listas*.
Texto de la recomendación: *Parece que no has incluido ninguna lista en tu texto utilizando las opciones disponibles en la barra superior de herramientas (numeración o viñetas). Recuerda que las listas, si están bien construidas, son muy eficaces para transmitir información de manera clara. Te recomendamos que añadas alguna lista a tu texto.*
Ejemplo:
Para presentar su solicitud, debe enviar cuatro documentos:
 - 1. El formulario de solicitud cumplimentado.*
 - 2. Una fotocopia de su DNI.*

3. *Su acta de nacimiento.*
4. *Su certificado de empadronamiento.*

b) Recomendaciones del nivel morfosintáctico

b1. Título de la recomendación: *Uso de la voz pasiva.*

Texto de la recomendación: *Las unidades marcadas parecen verbos en voz pasiva. Ten en cuenta que en este tipo de textos es más habitual la voz activa.*

Ejemplo:

La oración "El suministro que se contrata fue aprobado por este organismo" podría sustituirse por "Este organismo aprobó el suministro que se contrata".

b2. Título de la recomendación: *Revisión de gerundios.*

Texto de la recomendación: *Las unidades marcadas en el texto parecen verbos en gerundio. Te recomendamos que evites estas formas verbales, ya que pueden causar ambigüedades, alargar demasiado la oración y hacer difícil la comprensión.*

Ejemplo:

En la siguiente oración, no queda claro quién es el sujeto del gerundio: "El aspirante a la plaza dio su documentación al empleado del registro solicitando una copia."

Si el sujeto es el aspirante, sería más conveniente decir: "El aspirante a la plaza dio su documentación al empleado del registro, a quien solicitó una copia."

En cambio, si el sujeto es el empleado del registro, sería más adecuado decir, por ejemplo: "El aspirante a la plaza dio su documentación al empleado del registro y solicitó una copia."

b3. Título de la recomendación: *Revisión de participios.*

Texto de la recomendación: *Las unidades marcadas en el texto parecen verbos en participio. A no ser que sea imprescindible utilizarlos, te recomendamos que evites estas formas verbales, ya que pueden causar ambigüedades, alargar demasiado la oración y hacer difícil la comprensión.*

Ejemplos:

En las siguientes oraciones, no queda claro quién es el sujeto del participio:

"Enviada la resolución, el aspirante tiene diez días para reclamar."

"Presentado el documento, se cerró el expediente."

Podría especificarse de la siguiente manera, por ejemplo:

"Una vez el comité evaluador envió la resolución, el aspirante tiene diez días para reclamar."

"Cuando el solicitante presentó el documento, se cerró el expediente".

b4. Título de la recomendación: *Eliminación de formas verbales arcaicas.*

Texto de la recomendación: *Las formas verbales marcadas en el texto están en desuso y son innecesarias. Te recomendamos que las evites y que utilices en su lugar otras formas más sencillas y actuales.*

Ejemplos:

En vez de "si resultare necesario", podría decirse "si resulta necesario", "si resultara necesario" o "si resultase necesario".

En vez de "si hubiere sido necesario", podría decirse "si hubiera sido necesario" o "si hubiese sido necesario".

b5. Título de la recomendación: *Sistematicidad en el uso de verbos en 1.ª persona.*

Texto de la recomendación: *Las unidades marcadas en verde parecen verbos en 1.ª persona del singular y las marcadas en azul parecen verbos en 1.ª persona del plural. Si estas formas verbales se refieren al emisor del texto, te recomendamos que optes por el singular o el plural para que el texto sea sistemático.*

b6. Título de la recomendación: *Revisión de nominalizaciones verbales.*

Texto de la recomendación: *Parece que algunas de las palabras marcadas en el texto son nombres derivados de verbos y, por tanto, pueden resultar difíciles de entender. A no ser que sean términos de este ámbito que no se puedan cambiar, te recomendamos que las sustituyas por sus correspondientes verbos para hacer el texto más claro y dinámico.*

Ejemplos:

La oración "Se llevará a cabo una evaluación de los riesgos" podría sustituirse por "Se evaluarán los riesgos".

La oración "Se efectuó la instalación de los programas informáticos" podría sustituirse por "Se instalaron los programas informáticos".

b7. Título de la recomendación: *Reformulación de ideas expresadas en negativo.*

Texto de la recomendación: *Parece que las palabras marcadas en el texto indican negación. Te recomendamos que optes por la formulación afirmativa de tus ideas en caso de ser posible, puesto que así se favorece la legibilidad y la interpretación del mensaje.*

Ejemplos:

La oración "No es infrecuente que se acepten nuevos proyectos" podría formularse en positivo de la siguiente manera: "Es habitual que se acepten nuevos proyectos".

c) Recomendaciones del nivel léxico

c1. Título de la recomendación: *Uso de indicadores de subjetividad.*

Texto de la recomendación: *Las unidades marcadas podrían indicar 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.*

c2. Título de la recomendación: *Introducción de siglas.*

Texto de la recomendación: *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.*

Ejemplos:

Universidad Nacional de Educación a Distancia (UNED)

EPOC (enfermedad pulmonar obstructiva crónica)

c3. Título de la recomendación: *Sistematicidad en el uso de siglas.*

Texto de la recomendación: *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.*

c4. Título de la recomendación: *Utilización de términos más transparentes.*

Texto de la recomendación: *Los términos de la lista siguiente aparecen en tu texto y pueden resultar difíciles de entender. Te recomendamos que los sustituyas por otros más transparentes o que los aclares entre paréntesis la primera vez que aparecen. Haz clic en cada uno de ellos para ver sugerencias de términos alternativos más claros.*

c5. Título de la recomendación: *Sustitución de expresiones difíciles de entender.*

Texto de la recomendación: *Las expresiones de la lista siguiente aparecen en tu texto y pueden resultar difíciles de entender, por ser formas arcaicas, en desuso o latinismos. Te recomendamos que las sustituyas por otras más claras o, si no es posible, que las expliques en el texto la primera vez que aparecen (entre paréntesis o en una nota, por ejemplo). Haz clic en cada una de ellas para ver sugerencias de variantes alternativas más comprensibles o una explicación de su significado.*

c6. Título de la recomendación: *Sustitución de palabras poco precisas.*

Texto de la recomendación: *Parece que las palabras marcadas en el texto son poco precisas. Te recomendamos que las elimines o las sustituyas por otras palabras más precisas.*

Ejemplos:

En vez de “Se trataron algunos temas en la reunión”, sería más preciso decir “Se trataron cuatro temas en la reunión”.

En vez de “Dijeron que había que hacer una reunión”, sería más preciso decir “Dijeron que había que convocar una reunión”.

c7. Título de la recomendación: *Eliminación de expresiones redundantes.*

Texto de la recomendación: *Las expresiones marcadas en el texto incluyen información redundante. Te recomendamos que las elimines para hacer el texto más breve.*

c8. Título de la recomendación: *Revisión de palabras largas.*

Texto de la recomendación: *Las palabras de la lista siguiente aparecen en tu texto y tienen variantes alternativas más cortas. Te recomendamos que las utilices para favorecer la legibilidad. Haz clic en cada palabra para ver cuál es su alternativa más corta.*

B Anexo 2: Resoluciones del BOAM incluidas en el corpus de evaluación

1. BOAM 8933 (19/07/2021). Resolución 1927.
2. BOAM 8980 (22/09/2021). Resolución 2393.
3. BOAM 9001 (22/10/2021). Resolución 2773.
4. BOAM 9030 (07/12/2021). Resolución 3277.
5. BOAM 8919 (29/06/2021). Resolución 1691.
6. BOAM 9027 (01/12/2021). Resolución 3241.
7. BOAM 9023 (25/11/2021). Resolución 3167.
8. BOAM 9005 (28/10/2021). Resolución 2831.
9. BOAM 8997 (18/10/2021). Resolución 2728.
10. BOAM 9021 (23/11/2021). Resolución 3120.

