

INTELIGENCIA ARTIFICIAL Y LENGUAJE CLARO EN ESPAÑOL

Por Iria Da Cunha (*)

I. EL LENGUAJE CLARO EN EL CONTEXTO DE LA INTELIGENCIA ARTIFICIAL

La International Plain Language Federation (IPLF) define el término *lenguaje claro* de la siguiente manera: “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”⁽¹⁾. Esta definición, que es una de las más consensuadas en el ámbito, recoge tres aspectos fundamentales que deben tenerse en cuenta: la redacción, la estructura y el diseño. Estos mismos aspectos se señalan en la norma ISO titulada *Plain language — Part 1: Governing principles and guidelines*, que se publicó en junio del 2023⁽²⁾.

A su vez, el lenguaje claro se enmarca en el ámbito de la comunicación clara, que la consultora Prodigioso Volcán⁽³⁾ define de la siguiente manera: “comunicación clara significa transmitir de forma fácil, directa, transparente, simple y eficaz información relevante para la ciudadanía por cualquiera de los diferentes canales actuales (papel, teléfonos móviles, ordenadores y otros dispositivos) y adaptada a sus particularidades”. En relación

(*) Es doctora en Ciencias del Lenguaje y Lingüística Aplicada por la Universitat Pompeu Fabra. Actualmente es profesora contratada doctora en la Facultad de Filología de la UNED, donde previamente fue investigadora Ramón y Cajal. Es académica de número de la Academia Joven de España, vocal de la Junta Directiva de la Asociación Española de Terminología y vocal del Comité del Órgano Técnico AEN/CTN 191 de la Asociación Española de Normalización. Es autora de 120 publicaciones científicas entre las que destacan los libros *El discurso del ámbito de la Administración. Una perspectiva lingüística*, y *Lenguaje claro y tecnología en la Administración*. Ha participado en 23 proyectos de investigación y ha dirigido de cuatro de ellos, uno financiado por la Fundación BBVA a través de una Beca Leonardo y tres por el Ministerio de Ciencia e Innovación de España. Es la creadora de arText claro, un *software* de asistencia a la redacción en lenguaje claro.

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

(2) <https://www.iso.org/standard/78907.html>

(3) <https://comunicacionclara.com/>

con este concepto, Montolío y Tascón (2020)⁽⁴⁾ establecen cuatro ejes para la comunicación clara: el lenguaje claro, el lenguaje visual, los lenguajes de especialidad, y el neurolenguaje y las ciencias del comportamiento.

Uno de los nuevos ejes que debe añadirse en el ámbito del lenguaje claro y la comunicación clara es la inteligencia artificial (IA), especialmente una de sus ramas, el procesamiento del lenguaje natural (PLN). En este sentido, precisamente, según Montolío y Tascón (2020), la ingeniería de *software* es una de las áreas que más novedades vienen aportando últimamente en este ámbito. Lo mismo opina María Lorena Tula del Moral, jueza y directora del Observatorio de Lenguaje Claro de la Facultad de Derecho de la Universidad de Buenos Aires, quien afirma que “hoy la tecnología cruza todos los trabajos, todas las investigaciones, todas las aplicaciones; entonces creemos que la tecnología aplicada al lenguaje claro puede traer grandes ventajas” [afirmación recogida en Cervellini y Rebella (2022)⁽⁵⁾]. También la publicación del libro *Lenguaje claro y tecnología en la Administración* (Da Cunha, 2022)⁽⁶⁾, que presenta la investigación más reciente en el ámbito, va en esta línea.

Evidencias de este nuevo enfoque son las recientes iniciativas de divulgación al respecto, como jornadas, cursos y *workshops*. Una de estas iniciativas de divulgación a nivel internacional es la *Clear Writing for Europe Conference*, que la Dirección General de Traducción de la Comisión Europea organiza cada dos años en Bruselas. Este año 2023 se tituló “Supporting European democracy and transparency through clear language” y por primera vez contó con un *workshop* titulado “Drafting and editing tools and resources”⁽⁷⁾. En este *workshop*, se presentaron dos sistemas tecnológicos para la redacción en lenguaje claro, uno para el francés (Lisible)⁽⁸⁾ y otro para el español (arText claro)⁽⁹⁾, al que me referiré en el apartado II.1).

En España, en los últimos años también han tenido lugar eventos centrados específicamente en este tema, como en 2021 la jornada “La tecnología al servicio del lenguaje claro en la Administración”⁽¹⁰⁾, organizada por la

(4) MONTOLÍO, E.; TASCÓN, M., “El derecho a entender: la comunicación clara, la mejor defensa de la ciudadanía”, Prodigioso Volcán y Catarata, 2020.

(5) CERVELLINI, M. C.; REBELLA, M., “El software como instrumento de mejora para la comunicación e información jurídica en lenguaje claro”, III Congreso Nacional de Estudios de Administración Pública “Las Transformaciones en la Administración Pública ante la nueva complejidad”, 2022, en línea. https://aaeap.org.ar/wp-content/uploads/2023/02/Cervellini_Rebella.pdf

(6) DA CUNHA, I. (Ed.), “Lenguaje claro y tecnología en la Administración”, Ed. Comares, Granada, 2022.

(7) https://commission.europa.eu/events/clear-writing-europe-2023-2023-05-23_en

(8) <https://lisible.com/>

(9) <http://sistema-artext.com/>

(10) <https://www.youtube.com/watch?v=FFeKv2X1Qkw&t=5598s>

Región de Murcia con la colaboración de la Fundación NovaGob. Además, se están organizando multitud de eventos de divulgación en los que se incluyen mesas redondas y conferencias al respecto, como, entre otros:

- Las jornadas EnClaro⁽¹¹⁾, especialmente en su edición de 2021⁽¹²⁾.
- La jornada “Comunicación clara y lenguaje administrativo. Por una administración al servicio de las personas”⁽¹³⁾, organizada por la Generalitat Valenciana en 2022.
- La mesa “Comunicación clara” en el curso “Fundamentos del Gobierno Abierto”⁽¹⁴⁾, organizado por el Instituto Nacional de Administración Pública en 2022.

En Latinoamérica también se están llevando a cabo este tipo de iniciativas de divulgación. Por ejemplo:

- El IV Seminario internacional de lenguaje claro, organizado por la Red de Lenguaje Claro de Colombia en 2021, que contó con una mesa redonda titulada “Midiendo la claridad: testeo de contenidos y diálogo de saberes”⁽¹⁵⁾.
- Las I Jornadas de Lenguaje Claro en Español. Transparencia, integridad y la visibilización del sector, organizadas por la Universidad del Salvador (Argentina) en 2023, con un panel titulado “El derecho a entender”⁽¹⁶⁾, en el que se habló de inteligencia artificial y lenguaje claro.
- Las II Jornadas de Corrección de Textos en Español, organizadas también por la Universidad del Salvador en 2023, en las que se habló de herramientas para el lenguaje claro en dos paneles: “Tecnologías aplicadas para la intervención de textos I”⁽¹⁷⁾ y “Tecnologías aplicadas para la intervención de textos II”⁽¹⁸⁾.
- La Conferencia Internacional Plain Language 2023⁽¹⁹⁾, organizada por el Observatorio de Lenguaje Claro de la Facultad de Derecho de la Universidad de Buenos Aires, con un panel titulado “La inteligencia artificial como instrumento para clarificar el lenguaje”.

(11) <https://enclaro.es/>

(12) <https://vimeo.com/623470170/4fa8b6b667>

(13) <https://www.youtube.com/watch?v=65uhpO-HwS0>

(14) <https://www.youtube.com/watch?v=ROxV0sSdz7s>

(15) <https://www.youtube.com/watch?v=7CtH16-m1vE&t=8279s>

(16) <https://www.youtube.com/watch?v=-xl2MG3Qiz0>

(17) https://www.youtube.com/watch?v=_dD9HGJrMp4&t=1s

(18) <https://www.youtube.com/watch?v=oFautPYb8SQ>

(19) <http://www.derecho.uba.ar/institucional/conferencia-internacional-plain-language/>

Otra evidencia del interés que despierta actualmente la sinergia entre la inteligencia artificial y el lenguaje claro en español es el desarrollo de proyectos de investigación sobre este tema. Estos proyectos se están llevando a cabo tanto en universidades y centros de investigación, como en administraciones y empresas. Los proyectos que hoy en día cuentan con productos tecnológicos desarrollados se abordarán con detalle en el apartado II, por ejemplo, arText claro (UNED), Clara⁽²⁰⁾ (Prodigioso Volcán) y CLAPI⁽²¹⁾ (Gobierno de la Ciudad de Buenos Aires).

Además de estos proyectos pioneros, hay otras iniciativas que se están gestando en la actualidad. En primer lugar, una de estas iniciativas se plantea en el marco de la Red Panhispánica de Lenguaje Claro⁽²²⁾, impulsada por la Real Academia Española (RAE). Esta red se constituyó oficialmente el 9 de junio de 2022 en Santiago de Chile mediante un acuerdo suscrito por el presidente de la Corte Suprema de Justicia de la República de Chile, el director de la RAE y la Secretaría de la Red de Lenguaje Claro Chile en representación de las instituciones que conforman la red. Su objetivo es fomentar el lenguaje claro y accesible como fundamento de los valores democráticos, y promover el compromiso de las autoridades para asegurarlo en todos los ámbitos de la vida pública. En el anexo I⁽²³⁾ del sitio web de la red se ofrece un completo resumen de las principales iniciativas de lenguaje claro desarrolladas a nivel nacional y supranacional en diferentes lenguas.

En este marco, la RAE llevará a cabo diversas acciones en tres líneas de trabajo complementarias: I) Lenguaje claro y Administraciones públicas, II) Lenguaje accesible y III) Acción institucional. En relación con la línea de trabajo I), en la web de la RAE se señala que tienen intención de abordar la sinergia entre lenguaje claro e inteligencia artificial a través de las acciones 3 y 4⁽²⁴⁾:

3. Desarrollo, con técnicas de inteligencia artificial, de una herramienta que permita «traducir» automáticamente los tecnicismos propios del lenguaje institucional y administrativo a un lenguaje sencillo, propio de la vida cotidiana de los ciudadanos. El traductor automático se pondrá a disposición de las instituciones como contribución a la generalización del uso del lenguaje claro.

4. Prestación de asistencia técnica para el contraste y mejora del uso de la lengua a través del verificador automático de la RAE que, a modo de corrector, aplicará el canon lingüístico de las academias para informar a

(20) <https://clara.comunicacionclara.com/>

(21) <https://www.youtube.com/watch?v=-xl2MG3Qiz0>

(22) <https://www.rae.es/la-institucion/red-panhispanica-de-lenguaje-claro>

(23) <https://www.rae.es/la-institucion/red-panhispanica-de-lenguaje-claro/desarrollo-historico-y-situacion-actual-del/anexo>

(24) <https://www.rae.es/la-institucion/red-panhispanica-de-lenguaje-claro/programa-de-actuacion-y-acciones-de-la-rae-en>

los usuarios sobre la adecuación de sus textos, no solo en lo que respecta a su corrección ortográfica y gramatical, sino también en cuanto al estilo y a su claridad e inteligibilidad. Permitirá garantizar la utilización continua y sistemática de un lenguaje con la calidad debida por parte de instituciones relevantes para el funcionamiento del Estado de derecho.

De hecho, durante el IX Congreso de la Lengua Española, que organizó la RAE en Cádiz en 2023, tuvo lugar el panel “Lenguaje claro y accesible en contextos multiculturales: un derecho fundamental”⁽²⁵⁾, en el que el académico Salvador Gutiérrez Ordóñez puso de relieve la necesidad de desarrollar este tipo de sistemas para el español. Asimismo, destacó el trabajo pionero del equipo arText.

En segundo lugar, en España está en marcha el Proyecto estratégico para la recuperación y transformación económica (PERTE) de la Nueva Economía de la Lengua⁽²⁶⁾. Este proyecto tiene como objetivo desarrollar las oportunidades que presenta el español como un activo para impulsar la economía y cuenta con un presupuesto de 1.100 millones de euros de inversión pública. El proyecto tiene cinco ejes estratégicos: a) conocimiento en español y lenguas cooficiales, b) inteligencia artificial en español, c) ciencia en español, d) aprendizaje del español y e) industrias culturales. El eje relativo a la inteligencia artificial tiene asignada una partida presupuestaria muy elevada: 330 millones de euros. Durante la tercera reunión del Consejo Consultivo de este PERTE, el 10 de julio de 2023, se acordó de forma unánime aprovechar la existencia de este proyecto para potenciar todas las iniciativas en marcha que fomentan el lenguaje claro, inclusivo, accesible y libre de barreras por razón de edad o formación⁽²⁷⁾. Asimismo, en dicha reunión, la vicepresidenta de Asuntos Económicos y Transformación Digital del Gobierno de España, Nadia Calviño, destacó el importante desarrollo de la inteligencia artificial.

En tercer lugar, desde el Gobierno de España también se está llevando a cabo un proyecto en el que se establece una sinergia entre lenguaje claro y tecnología. María Ramos, directora de Análisis y Estudios del Gabinete de la Presidencia del Gobierno, fue la encargada de presentar el proyecto en la *Clear writing for Europe Conference 2023*, concretamente en el panel “*Making plain language the default setting to reach the widest audience*”⁽²⁸⁾. Este proyecto tiene como objetivo clarificar las respuestas que se ofrecen desde la Presidencia del Gobierno a las sugerencias, peticiones o quejas (entre otros géneros textuales) enviadas por la ciudadanía. El proyecto estuvo

(25) <https://www.youtube.com/watch?v=eO-B5kCbJz8&t=4694s>

(26) <https://planderecuperacion.gob.es/como-acceder-a-los-fondos/pertes/perte-nueva-economia-de-la-lengua>

(27) <https://portal.mineco.gob.es/es-es/comunicacion/Paginas/Consejo-Consultivo-PERTE-de-la-Lengua-.aspx>

(28) https://commission.europa.eu/events/clear-writing-europe-2023-2023-05-23_en

nominado a los Premios Archiletras de la Lengua en la edición 2023 en la categoría Innovación⁽²⁹⁾, con el título “El proceso modernizador en la comunicación con la ciudadanía de la Presidencia del Gobierno”.

Finalmente, desde las administraciones autonómicas de España también existe mucho interés en el tema. Prueba de ello es que en 2023 la Consejería de Inclusión Social, Juventud, Familias e Igualdad de la Junta de Andalucía sacó a licitación la contratación de servicios para la realización de un sistema conversor de documentos a lenguaje claro⁽³⁰⁾. El importe de la licitación es de más de dos millones de euros y está financiado con los fondos europeos Next Generation, vinculados al Mecanismo de Recuperación y Resiliencia de la Unión Europea. A día de hoy, no ha trascendido más información sobre este proyecto.

En paralelo a estas iniciativas, también hay otras relacionadas con la lectura fácil. Por ejemplo, en la Universidad Politécnica de Madrid, el grupo OEG (*Ontology Engineering Group*) está investigando desde hace unos años sobre el desarrollo de un sistema informático para adaptar textos a lectura fácil (Figuerola et al. 2022)⁽³¹⁾. Por su parte, el grupo GPLSI (Procesamiento del lenguaje y sistemas de información) de la Universidad de Alicante ha obtenido un proyecto del Ministerio de Ciencia e Innovación de España con fondos Next Generation, gracias al cual desde 2023 están comenzando a investigar sobre cómo adaptar a lectura fácil el contenido digital de los organismos del sector público mediante PLN, para que puedan beneficiarse las personas con discapacidades cognitivas⁽³²⁾. Finalmente, en julio de 2023, se anunció en los medios de comunicación que el Ministerio de Justicia de España está desarrollando una herramienta de IA en línea que traduce los documentos jurídicos a un formato de lectura fácil⁽³³⁾. De momento no ha trascendido más información sobre el producto informático derivado de este proyecto.

En definitiva, en los últimos años estamos viviendo un resurgimiento de las iniciativas relacionadas con la comunicación clara y el lenguaje claro, especialmente desde la perspectiva del desarrollo de herramientas y recursos tecnológicos basados en el PLN y la IA. En el apartado II, se abordará la

(29) <https://www.archiletras.com/actualidad/36-finalistas-ii-premios-archiletras-de-la-lengua/>

(30) https://www.juntadeandalucia.es/haciendayadministracionpublica/apl/pdc_sirec/perfiles-licitaciones/detalle-licitacion.jsf?idExpediente=465008

(31) SUÁREZ-FIGUEROA, M. C.; DIAB, I.; RUCKHAUS, E.; CANO, I., “First steps in the development of a support application for easy-to-read adaptation”, *Universal Access in the Information Society*, 2022, en línea. <https://link.springer.com/article/10.1007/s10209-022-00946-z>

(32) <https://cleartext.gplsi.es/>

(33) <https://confilegal.com/20230703-el-ministerio-de-justicia-desarrolla-una-herramienta-de-ia-online-que-traduce-los-documentos-juridicos-a-un-formato-de-lectura-facil/>

evolución de los paradigmas del PLN y su aplicación al lenguaje claro, con ejemplos reales. Finalmente, en el apartado III, se reflexionará sobre la situación actual en el ámbito y se plantearán perspectivas de trabajo futuro.

II. EVOLUCIÓN DE LOS PARADIGMAS DEL PLN Y APLICACIÓN AL LENGUAJE CLARO

El concepto de IA, que, como ya se ha indicado, incluye entre una de sus tecnologías el PLN, ha cobrado una gran importancia en la última década. Especialmente, se ha hecho conocida para el público general a raíz de que la empresa OpenAI diese acceso libre a su (ya famoso) ChatGPT⁽³⁴⁾ en noviembre de 2022. Sin embargo, tal como resaltan González-Blanco, Ros y Fresno (2022)⁽³⁵⁾, hace muchos años que se viene trabajando sobre este tipo de tecnologías. Estos autores hacen hincapié, además, en la evolución de los paradigmas del PLN (González-Blanco, Ros y Fresno (2022: 189):

Por otra parte, y como ya hemos comentado, el camino del PLN ha sido largo y no exento de luces y sombras. Pero este camino sí tiene una clara evolución que ha marcado su historia y que ha sido caracterizado por los diferentes paradigmas que han ido configurando lo que conocemos ahora como técnicas del PLN (ilustración 1). De esta manera podemos observar cómo ha ido pasando, como hemos dicho, por diferentes paradigmas computacionales que surgieron a la luz de los avances en *hardware* y de diseño de algoritmos. Así se pasa desde los sistemas basados en reglas hasta los modelos de lenguaje neuronales, siendo estos últimos posibles gracias los avances en *hardware* actuales.

La Figura 1 recoge un esquema de la evolución de los paradigmas del PLN ofrecido por González-Blanco, Ros y Fresno (2022: 189).

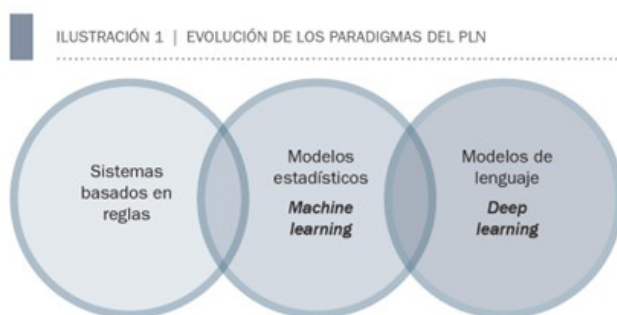


Figura 1. Esquema de la evolución del PLN (González-Blanco, Ros y Fresno, 2022: 189).

(34) <https://chat.openai.com/>

(35) GONZÁLEZ-BLANCO, E.; ROS, S.; FRESNO, V., “Cómo hablan las máquinas español: logros, retos y oportunidades para la inteligencia artificial aplicada al lenguaje”, El español en el mundo 2022. Anuario del Instituto Cervantes, 2022, ps. 183-199.

Los sistemas basados en reglas, que son los que se comenzaron a utilizar en el ámbito del PLN, logran llevar a cabo una tarea determinada basándose en un grupo de reglas lingüísticas elaboradas a mano. Estas reglas pueden estar basadas en información relativa a diferentes niveles de la lengua en función de la tarea que se necesite resolver, como el nivel fonético-fonológico, el morfológico, el léxico, el sintáctico, el semántico, el discursivo y el pragmático.

Por su parte, los modelos estadísticos basados en aprendizaje automático comenzaron a adquirir fuerza en los años 80. Estos modelos pueden procesar y analizar cantidades de datos enormes, aprender de ellos y, ante datos nuevos, predecir una respuesta. Tal como resaltan Torrijos y Sánchez (2023: 77)⁽³⁶⁾ en su reciente y recomendable libro *La primavera de la inteligencia artificial*:

Este es el gran cambio, y no es cualquier cosa: es el hecho de que los sistemas pueden inferir patrones sobre datos conocidos y generalizar, extrapolar ese conocimiento a datos nuevos. No necesitan reglas predeterminadas. Si observan una cantidad suficiente de textos clasificados en una serie de categorías, son capaces de predecir la categoría que va a corresponder a un texto que no han visto nunca.

Estos modelos estadísticos evolucionaron de tal manera que, en la última década, han explotado los modelos de lenguaje basados en redes neuronales. De hecho, desde 2017, las grandes tecnológicas han dedicado presupuestos altísimos para crear sus propios modelos. Estos modelos pueden predecir con una gran precisión la palabra que tiene más probabilidades de ser la siguiente de un fragmento textual, considerando el contexto que precede a dicha palabra. Para obtener una respuesta, el usuario debe escribir un *prompt*, que puede ser un fragmento para que el sistema pueda continuarlo o un fragmento en forma de petición (“resume”, “traduce”, “haz una lista”, etc.). También se puede solicitar al sistema que imite un determinado estilo, registro o variedad de la lengua, entre otras cuestiones.

Aunque ha habido muchos modelos pioneros en el ámbito (sobre todo para el inglés, pero también para el español)⁽³⁷⁾, la revolución llegó en noviembre del 2022, cuando la empresa OpenAI lanzó su chat conversacional ChatGPT (que utiliza el modelo GPT-3.5), para su uso en línea y gratuitamente⁽³⁸⁾. El lanzamiento de este chat ha supuesto una verdadera revolución

(36) TORRIJOS, C.; SÁNCHEZ, J. C., “La primavera de la inteligencia artificial”, Prodigioso Volcán y Catarata, 2023.

(37) Para aquellas personas interesadas en profundizar en el tema, se recomienda leer el capítulo 4 de Torrijos y Sánchez (2023).

(38) Posteriormente, en marzo de 2023, OpenAI lanzó una nueva versión del chat, llamada ChatGPT Plus (que utiliza el modelo GPT-4, más avanzado y con más capacidades), en este caso de pago. No obstante, sigue manteniendo activo y en abierto ChatGPT.

a nivel científico y social, por la gran repercusión que está teniendo. Tanto los expertos en el ámbito del PLN como la ciudadanía en general están muy sorprendidos por la calidad de los textos que genera. Aunque el uso de esta tecnología está despertando sentimientos encontrados entre sus defensores y detractores, lo cierto es que ya se está aplicando a múltiples tareas con resultados impresionantes (aunque mejorables en muchos aspectos)⁽³⁹⁾.

El PLN, sea cual sea el paradigma que utilice, se ha venido aplicando a diferentes tareas a lo largo de las últimas décadas, como la traducción automática, el resumen automático, la detección de opinión y sentimientos, la detección de noticias falsas, y la ayuda a la redacción, entre otras. Como he adelantado en la introducción, en los últimos años ha comenzado a aplicarse a la tarea de adaptación de textos a lenguaje claro. En los siguientes apartados detallaré las tecnologías existentes más relevantes para el español, divididas en los tres paradigmas del PLN explicados hasta ahora: sistemas basados en reglas (apartado II.1), modelos estadísticos (apartado II.2), y modelos de lenguaje (apartado II.3).

II.1. Sistemas basados en reglas

ArText claro es una herramienta computacional que se ha integrado en la infraestructura tecnológica del sistema arText (Da Cunha, Montané e Hysa, 2017), desarrollado previamente en la UNED gracias a una Beca Leonardo de la Fundación BBVA concedida en 2015. Tanto arText como arText claro pueden utilizarse gratuitamente en línea y sin necesidad de registro. Puede verse un videotutorial que explica sus funcionalidades en: <https://www.youtube.com/watch?v=yr99-FXCJao>. También puede verse el vídeo oficial de lanzamiento de arText claro, con la participación del rector de la UNED en: <https://canal.uned.es/video/60acbc7bb609230ea20ba3b4>.

Concretamente, arText claro es un redactor asistido para el español que ayuda a escribir textos administrativos en lenguaje claro (Da Cunha, 2022)⁽⁴⁰⁾. Se ha desarrollado en el Departamento de Filologías Extranjeras y sus Lingüísticas de la Facultad de Filología de la UNED, en el marco del grupo de investigación ACTUALing y con la colaboración del grupo IULATERM (Institut de Lingüística Aplicada, Universitat Pompeu Fabra). Concretamente, es fruto de dos proyectos de investigación del Ministerio de Ciencia e Innovación de España⁽⁴¹⁾. El objetivo de estos proyectos es, en el marco

(39) Sobre este tema, recomiendo la charla de Julio Gonzalo, director del grupo de investigación Procesamiento del Lenguaje Natural y Recuperación de Información de la UNED, impartida en la IV edición de la Jornada “Un Mundo de Fusión Digital”: <https://canal.uned.es/video/646740e26ff5a71a2d294b42>.

(40) DA CUNHA, I., “Un redactor asistido para adaptar textos administrativos a lenguaje claro”, *Procesamiento del Lenguaje Natural*, 69, 2022, págs. 39-49.

(41) Los proyectos son: I) “Tecnologías de la Información y la Comunicación (TIC) para la e-Administración: hacia la mejora de la comunicación entre Administración y ciu-

de la e-Administración, ayudar a la redacción de textos administrativos en español por parte del personal público, para contribuir a que los textos que reciba la ciudadanía estén redactados en lenguaje claro, utilizando recursos y herramientas de PLN. En este contexto, en 2021 se firmó un protocolo de colaboración en investigación científica en materia de comunicación clara con el Ayuntamiento de Madrid, a través de la Subdirección General de Calidad y Evaluación⁽⁴²⁾, que seguirá vigente hasta 2024.

ArText claro tiene dos vías de entrada, tal como puede verse en la Figura 2: a) puede ayudar a estructurar y a redactar en lenguaje claro ciertos géneros textuales del ámbito de la Administración, y b) puede ayudar a revisar la adecuación de textos administrativos (sin tener en cuenta un género textual específico) a las recomendaciones de lenguaje claro.



Figura 2. Captura de pantalla inicial del sistema arText.

En el caso de la entrada a), los géneros textuales que ayuda a redactar el sistema son: requerimientos, actas de inspección, contratos, resoluciones publicadas en portales institucionales y resoluciones dirigidas a un particular. La selección de estos cinco géneros textuales se llevó a cabo gracias al trabajo de Pistola y Viñuales-Ferreiro (2021)⁽⁴³⁾, quienes compilaron un lis-

dadanía a través del lenguaje claro (TIC-eADMIN)” (PGC2018-099694-A-I00), financiado en la convocatoria del 2018 de Proyectos I+D del Subprograma Estatal de Generación de Conocimiento, y II) “Un prototipo tecnológico para la redacción en lenguaje claro: incorporación en la Administración pública y análisis de su impacto en la ciudadanía (ar-Text_IMPACT)” (PDC2022-133935-I00), financiado en la convocatoria 2022 de Ayudas a Proyectos de Prueba de Concepto (MCIN/ AEI/10.13039/501100011033/ y la Unión Europea NextGenerationEU/PRTR).

(42) <https://www.madrid.es/portales/munimadrid/es/Inicio/Actualidad/Noticias/El-Ayuntamiento-y-la-UNED-firman-un-protocolo-de-comunicacion-clara/?vgnextfmt=default&vgnextoid=0d90f1feb7d9c710VgnVCM2000001f4a900aRCRD&vgnextchannel=a12149fa40ec9410VgnVCM100000171f5a0aRCRD>

(43) PISTOLA, S.; VIÑUALES-FERREIRO, S., “Una clasificación actualizada de los géneros textuales de la Administración pública española”, Revista de Llingua i Dret, 75,

tado actualizado de los géneros textuales de la Administración pública española, y de Pistola y Da Cunha (2023)⁽⁴⁴⁾, quienes hicieron un estudio para detectar cuáles de esos géneros suponen más dificultades de escritura para el personal público de la Administración, y para estudiantes de derecho y ciencias jurídicas.

Para comenzar a redactar, debe escogerse un género textual. De esta manera, se entra directamente en el editor de texto, donde se verán tres módulos:

- Módulo I: Estructura y contenidos del texto (columna izquierda).
- Módulo II: Formato y corrección ortográfica (barra superior).
- Módulo III: Revisión del texto y obtención de recomendaciones sobre lenguaje claro (columna derecha).

Estos módulos se diseñaron en consonancia con la ya mencionada definición de “lenguaje claro” de la IPLF, que tiene en cuenta: la redacción (módulo III), la estructura (módulo I) y el diseño (módulo II). En la Figura 3 se muestra una captura de pantalla del editor, después de haber seleccionado el género textual “resolución publicada en portales institucionales”.

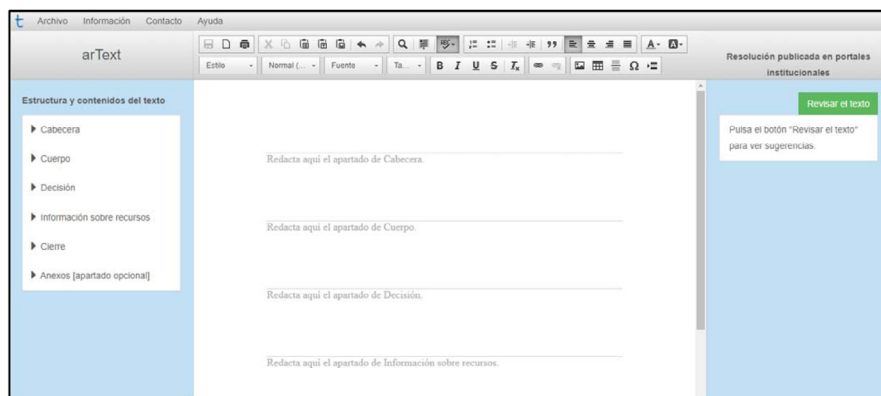


Figura 3. Captura de pantalla del editor de arText con la selección del género textual “resolución publicada en portales institucionales”.

A través del módulo I, se puede acceder a la estructura recomendada para el género escogido, es decir, a sus apartados prototípicos. En el caso

2021, págs. 181-203.

(44) PISTOLA, S.; DA CUNHA, I., “La escritura de los géneros textuales administrativos en el contexto académico y profesional”, Íkala, Revista de Lenguaje y Cultura, 28(3), 2023.

de la Figura 3, pueden verse los siguientes apartados: Cabecera, Cuerpo, Decisión, Información sobre recursos, Cierre y Anexos [apartado opcional]. Asimismo, si se clicla encima de cada uno de los apartados, se obtendrán los contenidos que suelen incorporarse en cada uno de ellos, así como frases prototípicas adaptadas a lenguaje claro que pueden ayudar a expresar dichos contenidos. Si se clicla en cualquiera de dichas frases, estas se cargan automáticamente en el texto. La selección de estos apartados, contenidos y frases se llevó a cabo mediante un estudio empírico basado en corpus (Pistola, 2022⁽⁴⁵⁾; Pistola y Da Cunha, 2022⁽⁴⁶⁾). En la Figura 4 se muestran los contenidos del apartado “Cuerpo” y en la Figura 5 se reflejan las frases asociadas al contenido “Marco jurídico”.

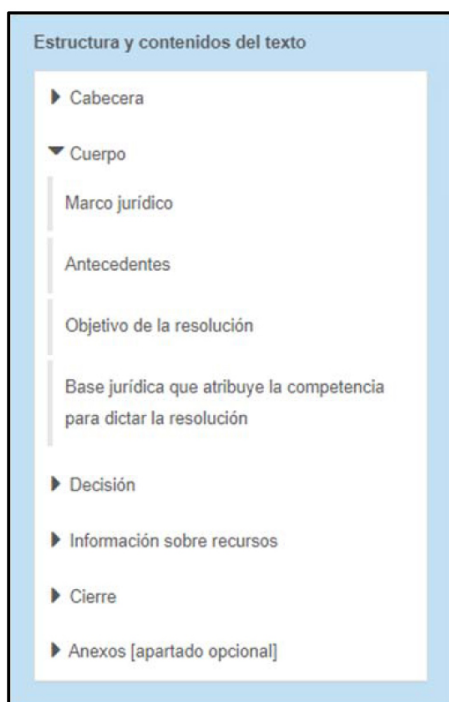


Figura 4. Contenidos del apartado “Cuerpo”.

(45) PISTOLA, S., “Análisis de la estructura y contenidos de un corpus de géneros textuales del ámbito de la Administración”, *Lenguaje claro y tecnología en la Administración*, Comares, 2022, págs. 77-116.

(46) PISTOLA, S.; DA CUNHA, I. “La resolución administrativa publicada en portales web institucionales: un análisis basado en corpus”, *Estudios de Lingüística*. Universidad de Alicante, 38, 2022, págs. 17-45.

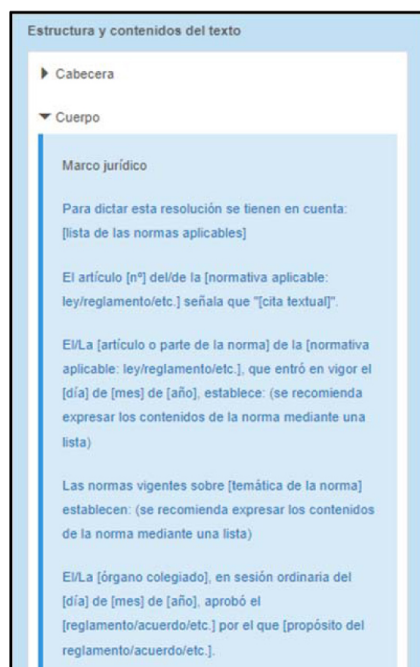


Figura 5. Frases asociadas al contenido "Marco jurídico".

Por su parte, mediante el módulo II, se puede hacer una revisión ortográfica del documento mediante el corrector de código abierto que tiene integrado el sistema (WebSpellChecker LLC.). Asimismo, se puede acceder a las opciones de formato más habituales para maquetar el documento.

Finalmente, el módulo III permite revisar el texto y detectar cuestiones lingüísticas que interfieren con las principales pautas de lenguaje claro en español, así como obtener recomendaciones de mejora. Estas recomendaciones, concretamente 22, se seleccionaron a partir del estudio de Da Cunha y Escobar (2021)⁽⁴⁷⁾. Tienen que ver con rasgos discursivos (p. ej. párrafos, oraciones, conectores discursivos), morfosintácticos (p. ej. voz pasiva, formas en gerundio, negaciones) y léxicos (p. ej. siglas, terminología, fraseología).

ArText claro es un sistema que sigue el paradigma del PLN basado en reglas. Para poder aplicar dichas reglas, el sistema hace un preprocesamiento del texto mediante dos herramientas de PLN que tiene integradas: el analizador morfosintáctico de FreeLing (Atserias et al., 2006)⁽⁴⁸⁾ y el segmentador discursi-

(47) DA CUNHA, I.; ESCOBAR, M. Á. (2021), "Recomendaciones sobre lenguaje claro en español en el ámbito jurídico-administrativo: análisis y clasificación", *Pragmalingüística*, 29, 2021, págs. 129-148.

(48) ATSERIAS, J.; CASAS, B.; COMELLES, E.; GONZÁLEZ, M.; PADRÓ, LL; PADRÓ, M., "FreeLing 1.3. Syntactic and semantic services in an open-source NLP library", *LREC*

vo DiSeg (Da Cunha et al., 2010)⁽⁴⁹⁾. Además de integrar estas dos herramientas, el sistema incluye diversos algoritmos desarrollados por el equipo del proyecto. Estos algoritmos toman como entrada el texto procesado lingüísticamente por las dos herramientas de PLN mencionadas y detectan en el texto escrito los rasgos lingüísticos necesarios para poder ofrecer las recomendaciones asociadas a cada uno de ellos. El sistema detecta, por ejemplo, párrafos-oración, párrafos largos, oraciones largas (y, cuando es posible, las segmenta en fragmentos más breves), formas verbales obsoletas (como el futuro de subjuntivo), formas verbales que suelen utilizarse incorrectamente o que generan ambigüedad (como gerundios y pasivas), términos difíciles de entender (y ofrece alternativas más claras), expresiones en desuso, expresiones redundantes y siglas sin desplegar (véase la lista completa en Da Cunha, 2022).

En el caso de la entrada b) en el sistema arText claro, se accede al mismo editor. La diferencia principal reside en el módulo I, ya que, en este caso, al tratarse de textos administrativos en general, se ofrece únicamente la estructura genérica de los textos administrativos (encabezamiento, cuerpo y pie) con una breve explicación del contenido de cada apartado. A través de esta entrada, los módulos más relevantes son el II y el III.

En la Figura 6 se ofrece una captura de pantalla en la que aparecen las recomendaciones ofrecidas por arText claro utilizando el texto de prueba que está colgado en la web del sistema⁽⁵⁰⁾, que es una reclamación ficticia.

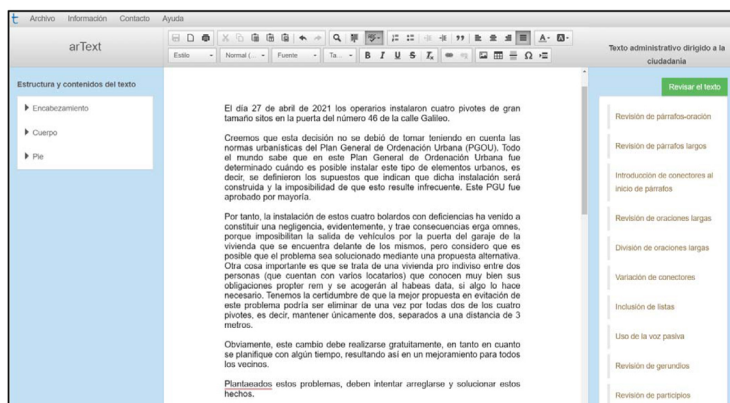


Figura 6. Captura de pantalla con las recomendaciones ofrecidas por arText claro.

2006 Proceedings. 5th Edition of the International Conference on Language Resources and Evaluation, European Language Resources Association, 2006, págs. 48-55.

(49) DA CUNHA, I.; SANJUAN, E.; TORRES-MORENO, J.-M.; LLOBERES, M.; CASTELLÓN, I., "DiSeg: Un segmentador discursivo automático para el español", *Procesamiento del Lenguaje Natural*, 45, 2010, págs. 145-152.

(50) http://sistema-artext.com/texto_de_pruebas_artext.html

En cuanto a la evaluación del sistema, por un lado, se llevó a cabo una evaluación *data-driven* de las recomendaciones de lenguaje claro obtenidas automáticamente (Da Cunha, 2022). Para ello, se utilizó un corpus de evaluación formado por 10 resoluciones del Boletín Oficial del Ayuntamiento de Madrid (BOAM) publicadas en el año 2021, que en total suman 8052 palabras. Los resultados obtenidos son, en general, positivos para la mayor parte de las recomendaciones evaluadas: en promedio se obtiene un 0,908 de precisión y un 0,917 de cobertura. Por otro lado, en el trabajo de Núñez y Da Cunha (2022)⁽⁵¹⁾ se llevó a cabo una evaluación *user-driven* de arText claro. Concretamente, se evaluó el impacto del uso del sistema en la comprensión lectora y en la percepción de claridad de textos administrativos por parte del alumnado universitario. Este experimento constató mediante técnicas estadísticas que el uso de un lenguaje claro tiene efectos positivos en las personas destinatarias de los textos y también que el *software* arText claro es útil para clarificar textos administrativos.

Actualmente, el sistema cuenta con más de 220.000 usuarios en la mayor parte de países del mundo y está siendo utilizado en diversas administraciones. El proyecto estuvo nominado a los Premios Archiletras de la Lengua en la edición 2022 en la categoría Investigación⁽⁵²⁾.

II.2. Modelos estadísticos

II.2.a. Clara

Clara es un sistema que sigue el paradigma de PLN de aprendizaje automático. Su versión beta, gratuita y en línea, permite realizar un *test* de claridad a textos en español, principalmente a fragmentos de 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. Para la implementación de la herramienta se utilizaron técnicas de aprendizaje y las siguientes nueve métricas de evaluación (Torrijos y Oquendo, 2021: 123-124)⁽⁵³⁾:

- Métrica 1: uso de palabras fuera del diccionario.
- Métrica 2: uso de conectores discursivos.

(51) NÚÑEZ, J. A.; DA CUNHA, I., "Impacto del uso arText claro en la comprensión lectora y en la percepción de claridad de textos administrativos por parte de estudiantes universitarios", *Lenguaje claro y tecnología en la Administración*, Comares, 2022, págs. 189-212.

(52) <https://www.archiletras.com/actualidad/el-jurado-selecciona-36-finalistas-en-los-premios-archiletras-de-la-lengua/>

(53) Torrijos, C.; Oquendo, S., "¡Hola! Soy Clara y mido la claridad de tu texto", *Archiletras Científica*, VI, 2021, págs. 119-133.

- 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.

Para ello, se compiló un corpus de textos administrativos y financieros, que se clarificaron manualmente por parte de lingüistas. Finalmente, se obtuvo un corpus de 1000 textos, de los cuales aproximadamente la mitad se etiquetaron como claros y la otra mitad como no claros. De este corpus, el 75% de los textos se utilizó para entrenar el sistema, mientras que el 25% restante se utilizó para evaluarlo. En la evaluación con el corpus de prueba, se obtuvo un *accuracy* (acierto) del 72%, lo cual es un resultado muy prometedor.

En la Figura 7, se incluye una captura de pantalla de la página inicial de Clara, donde se ha insertado el primer párrafo de texto de prueba de la web de arText (véase nota 51), y en la Figura 8 se ve el resultado de Clara una vez procesado dicho fragmento.

¡Hola! Soy Clara. Te ayudo a analizar la claridad de los textos en español.

Test de claridad (Beta)

Creemos que esta decisión no se debió de tomar teniendo en cuenta las normas urbanísticas del Plan General de Ordenación Urbana [PGOU]. Todo el mundo sabe que en este Plan General de Ordenación Urbana fue determinado cuándo es posible instalar este tipo de elementos urbanos, es decir, se definieron los supuestos que indican que dicha instalación será construida y la imposibilidad de que esto resulte infrecuente. Este PGU fue aprobado por mayoría.

[Términos de uso](#)

Número de palabras: 71

Analizar

Figura 7. Captura de pantalla de la página principal de Clara.

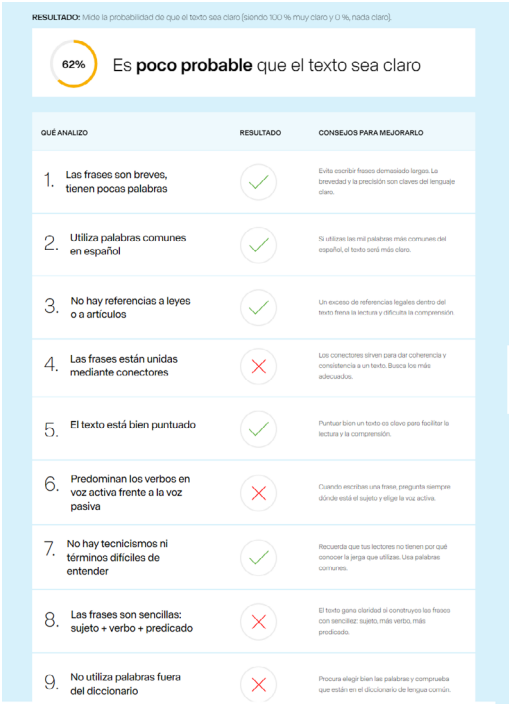


Figura 8. Captura de pantalla del resultado de Clara

II.2.b. Clasificador de sentencias en lenguaje claro del Observatorio de Lenguaje Claro de la Universidad de Buenos Aires

Este tipo de proyectos también despierta interés en países de Latinoamérica, especialmente en Argentina. Un caso destacable es el trabajo que llevan a cabo en el Observatorio de Lenguaje Claro de la Facultad de Derecho de la Universidad de Buenos Aires. Este observatorio está integrado por diferentes grupos, uno de los cuales es el Grupo de Tecnología, que explora la intersección entre lenguaje claro e IA. En el marco de este grupo, Cerveellini y Rebella (2022) presentaron un estudio con dos objetivos. El primer objetivo fue hacer una búsqueda de herramientas disponibles relacionadas con el lenguaje claro, para después analizarlas y evaluarlas a partir de una matriz de criterios que permitiese ponderarlas.

El segundo objetivo del estudio fue llevar a cabo una prueba de concepto para explorar la utilidad de aplicar aprendizaje automático, y detectar si una sentencia está escrita o no en lenguaje claro. Para ello, compilaron un corpus de sentencias judiciales, que clasificaron en dos grupos: sentencias claras y no claras. Mediante el uso de la técnica de bolsa de palabras (*bag of words*), representaron el contenido de las sentencias en vectores numéricos interpretables por algoritmos de aprendizaje automático. Estos vectores se

utilizaron para entrenar el clasificador. Los resultados mostraron un 74 % de acierto. Como trabajo futuro, se plantean definir más métricas que permitan entrenar otro modelo de aprendizaje automático, así como ampliar el corpus de sentencias. Por el momento, el sistema no está disponible para el público general, por lo que no puedo aportar capturas de pantalla o pruebas específicas con el sistema.

II.3. Modelos de lenguaje

II.3.a. ChatGPT

El ya mencionado ChatGPT, basado en modelos de lenguaje, no está especialmente diseñado para adaptar textos a lenguaje claro. Sin embargo, en función del *prompt* que se le indique, puede resultar de utilidad para ello, tal como mostraré con algunas pruebas que he hecho al respecto. Para ello, he utilizado su versión gratuita disponible en línea. En primer lugar, le he formulado la siguiente pregunta: “¿Qué es ‘lenguaje claro’?”. La respuesta, adecuada, está recogida en la Figura 9, que sirve asimismo para mostrar la interfaz de este chat conversacional.

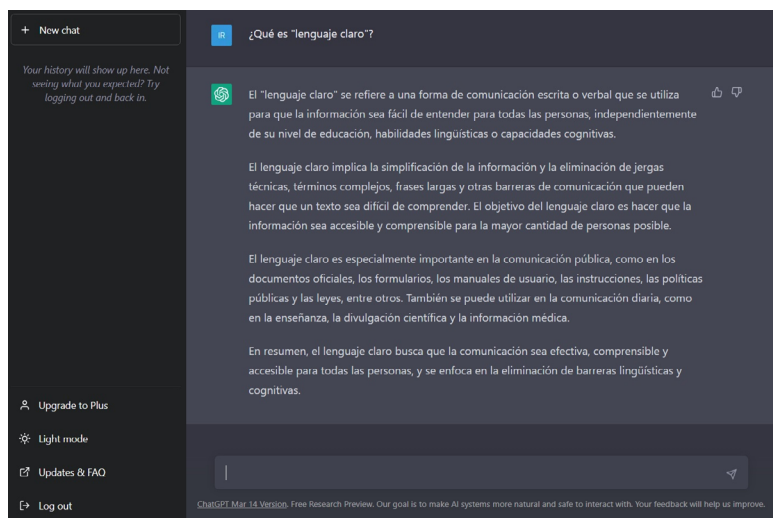


Figura 9. Resultado obtenido por ChatGPT con el *prompt* “¿Qué es ‘lenguaje claro’?”

En segundo lugar, le he pedido algo más complicado: que adapte a lenguaje claro el texto de prueba de la web de arText (véase nota 51). Para ello, he usado el siguiente *prompt*: “Adapta el siguiente texto a lenguaje claro:” [Texto de prueba de la web de arText]. El sistema devolvió un texto con diversas modificaciones, algunas de las cuales están en consonancia con las recomendaciones de lenguaje claro en español, mientras que otras son cuestionables. El resultado se muestra en la Figura 10.

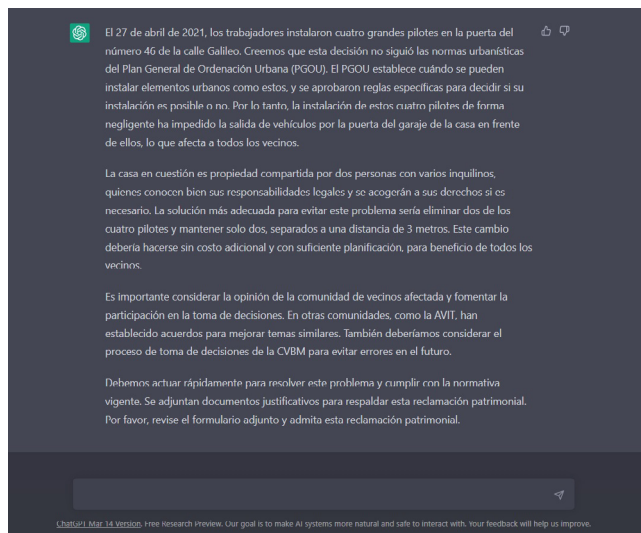


Figura 10. Resultado de ChatGPT con el prompt “Adapta el siguiente texto a lenguaje claro:” [Texto de prueba de la web de arText].

Para ilustrar estos resultados, a continuación, se ofrece un análisis lingüístico del texto generado. Concretamente, se analizan los tres niveles lingüísticos que afectan principalmente al lenguaje claro: el discursivo, el morfosintáctico y el léxico (Da Cunha y Escobar, 2021).

II.3.a.i. Nivel discursivo

Algunas cuestiones que modifica el sistema adecuadamente tienen que ver con los párrafos. Por ejemplo, en el texto original hay un párrafo-oración que debería evitarse:

El día 27 de abril de 2021 los operarios instalaron cuatro pivotes de gran tamaño sitios en la puerta del número 46 de la calle Galileo.

En este caso, el sistema lo fusiona con la siguiente oración para evitar dicho párrafo-oración. Otro caso tiene que ver con la detección y división de párrafos largos. Por ejemplo, el texto original incluye este párrafo largo:

Por tanto, la instalación de estos cuatro bolardos con deficiencias ha venido a constituir una negligencia, evidentemente, y trae consecuencias erga omnes, 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 que el problema sea solucionado mediante una propuesta alternativa. Otra cosa importante es que se trata de una vivienda pro indiviso entre dos personas (que cuentan con varios locatarios) que conocen muy bien sus obligaciones propter rem y se acogerán al habeas data, si algo lo hace necesario. Tenemos la certidumbre de que la mejor propuesta en evitación de este problema po-

dría ser eliminar de una vez por todas dos de los cuatro pivotes, es decir, mantener únicamente dos, separados a una distancia de 3 metros.

El sistema divide este párrafo en dos más cortos (inserta un salto de línea antes de la segunda oración, que, asimismo, reformula) y, además, logra mantener en cada uno una unidad temática: la explicación y la justificación de la negligencia que se indica en la reclamación, y la mención a los habitantes de la vivienda:

El 27 de abril de 2021, los trabajadores instalaron cuatro grandes pilotes en la puerta del número 46 de la calle Galileo. Creemos que esta decisión no siguió las normas urbanísticas del Plan General de Ordenación Urbana (PGOU). El PGOU establece cuándo se pueden instalar elementos urbanos como estos, y se aprobaron reglas específicas para decidir si su instalación es posible o no. Por lo tanto, la instalación de estos cuatro pilotes de forma negligente ha impedido la salida de vehículos por la puerta del garaje de la casa en frente de ellos, lo que afecta a todos los vecinos.

La casa en cuestión es propiedad compartida por dos personas con varios inquilinos, quienes conocen bien sus responsabilidades legales y se acogerán a sus derechos si es necesario. La solución más adecuada para evitar este problema sería eliminar dos de los cuatro pilotes y mantener solo dos, separados a una distancia de 3 metros. Este cambio debería hacerse sin costo adicional y con suficiente planificación, para beneficio de todos los vecinos.

En cuanto a la longitud oracional, el sistema adapta las oraciones para hacerlas más breves. Por ejemplo, en el texto original, se encuentra la siguiente oración larga:

Por tanto, la instalación de estos cuatro bolardos con deficiencias ha venido a constituir una negligencia, evidentemente, y trae consecuencias erga omnes, 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 que el problema sea solucionado mediante una propuesta alternativa.

El sistema ofrece la siguiente adaptación:

Por tanto, la instalación de estos cuatro pilotes de forma negligente ha impedido la salida de vehículos por la puerta del garaje de la casa en frente de ellos, lo que afecta a los vecinos.

Es cierto que esta oración es más breve, ya que cuenta con 35 palabras, pero, según indica la bibliografía, el umbral en el caso de lenguaje claro en español serían 25. En este caso, además, lo que es cuestionable es el tipo de reducción que hace el sistema: no hace una segmentación de la oración en otras más cortas, sino que directamente elimina información de la oración.

Concretamente, elimina la información siguiente (que incluye otras 35 palabras):

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 que el problema sea solucionado mediante una propuesta alternativa.

Ocurre algo similar en muchos otros fragmentos del texto. Por ejemplo, se elimina la siguiente información: “Este PGU fue aprobado por mayoría”, que, a priori, podría resultar una información relevante. De hecho, de 605 palabras que tiene el texto original, el sistema genera un texto de 247 palabras⁽⁵⁴⁾. En el ámbito del lenguaje claro no se tiende a eliminar información, sino a clarificarla. El hecho de que el sistema elimine fragmentos de contenido supone un riesgo, puesto que quizás esa información puede ser importante para las personas destinatarias o debe incluirse en el texto por algún motivo legal, por ejemplo.

Finalmente, con respecto al nivel discursivo, el sistema no introduce conectores al inicio de los párrafos. Esta es una recomendación habitual del lenguaje claro, ya que estas unidades ayudan a hilar las diferentes ideas de los párrafos del texto.

II.3.a.ii. Nivel morfosintáctico

En cuanto al nivel morfosintáctico, el sistema logra eliminar elementos que suelen causar ambigüedad, como las voces pasivas y los gerundios. Por ejemplo, transforma el fragmento “Todo el mundo sabe que en este Plan General de Ordenación Urbana fue determinado cuándo es posible instalar este tipo de elementos urbanos” por “El PGOU establece cuándo se pueden instalar elementos urbanos como estos”, cambiando la forma en voz pasiva “fue determinado” por el indicativo “establece”. Asimismo, sustituye el fragmento “resultando así en un mejoramiento para todos los vecinos” por “para beneficio de todos los vecinos”, sustituyendo de esta manera el gerundio “resultando” por la frase preposicional “para beneficio”.

En el caso de las formas en futuro de subjuntivo “estuvieren” y “hubieren considerado”, que se encuentran en el texto original y están obsoletas, no se incluyen en el texto generado. Sin embargo, no es posible saber por qué formas las habría sustituido el sistema, puesto que directamente ha eliminado todo el párrafo (77 palabras) en el texto resultante.

(54) Posteriormente, realicé más pruebas con el *prompt* “Adapta el texto siguiente a ‘lenguaje claro’ sin omitir información incluida en el texto original:”, pero en todas ellas el texto generado omitía gran parte de la información del texto original.

En cuanto a las negaciones, abundantes en el texto original, el sistema elimina varias de ellas, pero otras las mantiene, como puede verse en siguiente fragmento, donde cada oración contiene la unidad léxica “no”:

Creemos que esta decisión no siguió las normas urbanísticas del Plan General de Ordenación Urbana (PGOU). El PGOU establece cuándo se pueden instalar elementos urbanos como estos, y se aprobaron reglas específicas para decidir si su instalación es posible o no.

También elimina unidades léxicas que indican negación como “jamás”, “nadie”, “ni” y “tampoco”. Sin embargo, en estos casos el texto generado omite por completo las oraciones en las que están incluidas estas unidades, por lo que es complicado valorar este aspecto.

II.3.a.iii. Nivel léxico

En relación con el nivel léxico, pueden comentarse diversas cuestiones que son adecuadas en el texto generado. En primer lugar, se incluye variación terminológica de diferentes tipos. Por ejemplo, se pasa del sustantivo “operarios” del texto original a su hiperónimo “trabajadores”; se cambia el sustantivo “locatarios” por un sinónimo más claro “inquilinos”, y se sustituye el verbo “dirimirse” por uno más claro también, como es “resolver”.

En segundo lugar, en el texto original se repite dos veces el término desplegado “Plan General de Ordenación Urbana”. En cambio, en el texto generado por el sistema se sustituye la segunda ocurrencia por su sigla, “PGOU”, que ya había sido introducida a su vez con anterioridad.

En tercer lugar, el sistema logra sustituir los latinismos incluidos en el texto original. Por ejemplo:

trae consecuencias *erga omnes* > lo que afecta a todos los vecinos
se acogerán al *habeas data* > se acogerán a sus derechos
conocen muy bien sus obligaciones *propter rem* > conocen bien sus obligaciones legales

En la Tabla 1 puede verse que la sustitución es correcta en los tres casos, teniendo en cuenta el significado de dichos latinismos ofrecido en el Diccionario Panhispánico del Español Jurídico (DPEJ) de la RAE. El caso de *propter rem* es el que difiere ligeramente en cuanto a su significado, pero, en general, la sustitución, especialmente de *erga omnes* y de *habeas data*, es correcta.

Latinismo y definición en el DPEJ
<i>erga omnes</i> 1. Gral. Para o frente a todos. Tendrá efectos <i>erga omnes</i> , y no solo entre las partes del litigio. Se utiliza para expresar los efectos generales de una norma o sentencia.

<i>habeas data</i> 1. Adm. Acción constitucional que puede ejercer cualquier persona incluida en un registro de datos para acceder al mismo y recabar la información que le afecte, así como para solicitar su eliminación o corrección si tal información fuera falsa o estuviera desactualizada. <i>Constitución Política de Nicaragua, art. 190.1. Constitución de la República del Paraguay, art. 135.</i>
<i>obligación propter rem</i> Sublema de obligación ¹ Adm. Obligación que sigue a la propiedad del bien, independientemente de quien sea su titular. TRLSRU, art. 27.

Tabla 1. Definición de “erga omnes”, “habeas data” y “obligación propter rem” en el DPEJ.

Todos estos cambios léxicos, relacionados con la variación terminológica, las siglas y los latinismos, son coherentes con las pautas de lenguaje claro y coinciden, además, con las recomendaciones ofrecidas por arText claro sobre el mismo texto de prueba.

No obstante, también se detectan algunos casos que no son adecuados en el texto generado. Por ejemplo, en el texto original se habla de “pivotes” o “bolardos”, mientras que en el texto generado se cambia por “pilotes”. Aunque estas unidades puedan pertenecer al mismo campo semántico, tal como se indica en el Diccionario de la Lengua Española (DEL) de la RAE, su significado es diferente. En la Tabla 2 se ofrece una tabla que recoge la definición de “bolardo” y “pilote”, junto con una imagen que ilustra cada término.

Término	Definición del DLE	Imagen
bolardo	1. m. Obstáculo de hierro, piedra u otra materia colocado en el suelo de una vía pública y destinado principalmente a impedir el paso o aparcamiento de vehículos.	

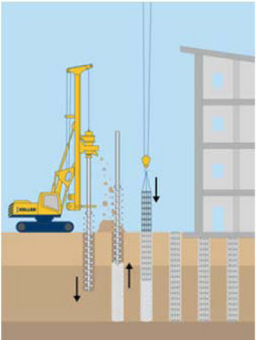
<p>pilote</p>	<p>1. m. Pieza larga a modo de estaca, de madera, hierro u hormigón armado, que se hince en el terreno para consolidar los cimientos de una construcción.</p>	 El diagrama ilustra el uso de un pilote en la construcción. A la izquierda, una excavadora amarilla está sobre el terreno. En el centro, un pilote largo y cilíndrico se está hincando en el suelo, guiado por una estructura vertical. A la derecha, se muestra una sección transversal del terreno con varios pilotes ya instalados, representados como columnas grises. Arriba a la derecha, se ve una parte de un edificio gris. Flechas indican la dirección de la hincada y la consolidación del terreno.
---------------	---	--

Tabla 2. Definición e imagen de los términos “bolardo” y “pilote”.

Lo mismo ocurre en otros casos, como en la sustitución de “protocolos” por “normativa”. Si bien su campo semántico puede ser el mismo, en este contexto su significado varía y sustituir dicha unidad léxica puede tener consecuencias legales.

II.3.b. CLAPPI

Otro sistema que se está desarrollando en la actualidad siguiendo el paradigma de los modelos de lenguaje es CLAPPI. Este proyecto se está llevando a cabo en el seno del Gobierno de la Ciudad de Buenos Aires en Argentina por parte de un equipo interdisciplinar formado por empleados públicos de dicho gobierno, especialistas en lenguaje claro de la Universidad de El Salvador (Argentina) y una empresa de tecnología externa contratada por medio de una licitación pública. El sistema usa la tecnología de ChatGPT y, asimismo, está siendo entrenado con miles de documentos legales de diferentes géneros textuales, como decretos, ordenanzas, resoluciones, edictos, requerimientos, reglamentos, avisos y anexos de normas, disponibles en la base de datos del Servicio de Información Normativa de la Secretaría Legal y Técnica de dicho gobierno.

Las recomendaciones que ofrecerá el sistema están en la línea con las que se vienen utilizando hasta ahora en el ámbito del lenguaje claro en español, como la segmentación de oraciones largas y la eliminación de voces pasivas. Asimismo, se tiene previsto trabajar con la estructura de los diferentes géneros textuales para que el sistema la tenga en cuenta en el texto generado.

Actualmente, CLAPPI está en su fase final de desarrollo y necesita ser evaluado antes de lanzarse oficialmente en 2024. La idea de sus creadores es que sirva de apoyo al personal público del Gobierno de la Ciudad de Buenos Aires, pero que en un futuro pueda abrirse a empleados públicos de otras administraciones.

III. DISCUSIÓN Y PERSPECTIVAS FUTURAS

Como se ha visto a lo largo de este capítulo, la sinergia entre la IA y el lenguaje claro ha cobrado una gran relevancia recientemente, y el español es una de las lenguas pioneras en este ámbito. Una evidencia es que se está investigando sobre ello desde universidades, empresas y administraciones públicas, tanto de España como de Latinoamérica (principalmente desde Argentina), y que se están desarrollando productos tecnológicos con diversas finalidades y que utilizan diferentes paradigmas de PLN.

Las finalidades de estos sistemas son principalmente tres: la medición automática de la claridad textual, la redacción asistida en lenguaje claro y la generación automática de textos más claros. Las tres tareas son relevantes en varios contextos y para diferentes destinatarios. En primer lugar, la medición automática de la claridad textual puede servir, por ejemplo, para evaluar los textos que las instituciones dirigen a la ciudadanía y mostrar evidencias de su falta de claridad. También puede servir para evaluar las iniciativas que se estén llevando a cabo en las instituciones y comprobar si las medidas instauradas (como el uso de una guía o un manual de claridad, la impartición de cursos de formación, etc.) realmente son efectivas y tienen repercusión en los textos que se generen en el marco de dichas instituciones. Por ejemplo, Prodigioso Volcán (2020⁽⁵⁵⁾, 2021⁽⁵⁶⁾) ha utilizado Clara para la elaboración de sus informes sobre la claridad (o falta de) de los documentos de la Administración pública española.

En segundo lugar, la redacción asistida en lenguaje claro puede ser útil principalmente para dos cuestiones. Por un lado, para la adaptación de textos a lenguaje claro. En los últimos años, en diversas dependencias de la Administración (en España y en Latinoamérica) se han publicado guías con recomendaciones sobre comunicación clara y lenguaje claro, tanto impresas como en línea (en este caso, principalmente en formato pdf o como página web). Estas recomendaciones en ocasiones son difíciles de aplicar para el personal público, puesto que les resulta complicado ir leyéndolas una por una y detectar en sus textos los fragmentos concretos que deben clarificar. Además, estas guías contienen terminología lingüística que no tienen por qué conocer (p. ej. “subordinadas”, “conectores discursivos”, “voz pasiva”). De hecho, en el marco del ya mencionado protocolo de colaboración del equipo arText con el Ayuntamiento de Madrid, se envió una encuesta a su personal público mediante la que se constató que casi el 90 % consideraba que le sería de gran utilidad disponer de una herramienta tecnológica que

(55) Prodigioso Volcán, “¿Habla claro la Administración pública?”, 2020, en línea. <https://comunicacionclara.com/lenguaje-claro/prodigioso-volcan-habla-claro-la-administracion.pdf>.

(56) Prodigioso Volcán, “¿Habla claro la Administración a los públicos vulnerables?”, 2021, en línea, <https://comunicacionclara.com/habla-claro-la-administracion/>.

le ayudase a adaptar a lenguaje claro los textos dirigidos a la ciudadanía (Pistola y Da Cunha, 2023).

Por otro lado, los redactores asistidos pueden ser útiles como herramienta didáctica, puesto que ponen en el centro a la persona que escribe el texto. Estos sistemas no generan automáticamente un texto nuevo, sino que guían en el proceso de redacción de dos maneras: primero, ofreciendo al usuario la estructura prototípica del género textual que necesita escribir y, segundo, marcando en el texto las cuestiones que interfieren con las recomendaciones sobre lenguaje claro. Así, a medida que la persona va redactando, va interiorizando tanto la estructura como las recomendaciones, lo cual es parte de un proceso didáctico. Precisamente, el Ayuntamiento de Madrid ha incorporado en su curso “Introducción a la comunicación clara”, con cinco ediciones anuales, una práctica con el sistema arText claro, para que su personal público conozca sus funcionalidades y lo utilice en su día a día profesional.

En tercer lugar, la generación automática de textos más claros tiene múltiples aplicaciones. La Administración puede desarrollar sistemas específicos para sus diferentes dependencias con el objetivo, por ejemplo, de redactar modelos y formularios que se envían masivamente a la ciudadanía, como resoluciones o notificaciones (cambiando los datos personales, y algunos datos específicos como cuantías y fechas). También pueden servir para ayudar a clarificar automáticamente textos ya escritos con anterioridad por el personal público, siempre, eso sí, teniendo en cuenta que la efectividad de estos sistemas nunca será total y, que, necesariamente, el texto generado deberá ser revisado por alguna persona especialista en el ámbito. Esta revisión es fundamental para evitar que en el texto generado “se cuelen” las ya conocidas “alucinaciones” de los modelos de lenguaje, es decir, información inventada que no es fiable y que, al estar redactada de forma gramaticalmente correcta, pueda parecerlo. Por ejemplo, el mencionado sistema CLAPI, sobre el que se está trabajando actualmente, va en esta dirección.

En cuanto a los paradigmas de PLN, como he explicado, se han desarrollado sistemas basados en reglas, en aprendizaje automático y en modelos de lenguaje. Uno de los retos futuros será combinar estos enfoques de manera eficiente. Por ejemplo, en el caso de la redacción asistida, fusionar el enfoque basado en reglas con los modelos de lenguaje es algo que ya se está comenzando a hacer en otras lenguas, como, por ejemplo, en el ya mencionado sistema Lisible, para el francés. Este sistema utiliza un paradigma basado en reglas (como arText para el español), pero recientemente ha incorporado ciertas funcionalidades (por ejemplo, la segmentación oracional o la sugerencia de variantes sinónimas) gracias a la incorporación de ChatGPT. También, en el caso de la generación de texto, si se utilizan modelos de lenguaje como base, será necesario aplicar alguna capa de *fine-tuning* que utilice reglas lingüísticas de cara a refinar los resultados y evitar errores

relacionados con diferentes niveles de lengua. Como se ha visto en el apartado II.3.a., si se lleva a cabo un análisis lingüístico exhaustivo de los resultados de los modelos generativos, es posible detectar los casos más comunes en los que fallan (en los niveles discursivo, morfosintáctico y léxico) y aplicar reglas que puedan mejorarlos. Asimismo, tanto en la redacción asistida como en la generación de texto, es fundamental tener en cuenta ciertas reglas lingüísticas relativas a la estructura de los diferentes géneros textuales.

Uno de los mayores retos de este tipo de herramientas es que lleguen a utilizarse de verdad por parte de los colectivos profesionales a quienes van dirigidas. Y parece que poco a poco se está consiguiendo. Además de los ejemplos citados hasta ahora, hay otros. Por ejemplo, recientemente apareció en los medios de comunicación que un juez del Superior Tribunal de Justicia de Corrientes (Argentina) está ya utilizando arText claro y ChatGPT para clarificar sus escritos dirigidos a la ciudadanía⁽⁵⁷⁾.

Otro de los retos es adaptar estas herramientas a otras lenguas. Para el inglés existen dos herramientas muy conocidas, Hemingway Editor⁽⁵⁸⁾ y VisibleThread⁽⁵⁹⁾, que están disponibles para textos generales no especializados. Para el francés, ya he mencionado la herramienta Lisible. Sin embargo, para las lenguas cooficiales del estado español aún no se ha desarrollado ninguna. En este sentido, destaco el proyecto de adaptación de arText claro al catalán, que se está desarrollando en la UNED con la financiación obtenida por el Govern de les Illes Balears y la Generalitat de Catalunya⁽⁶⁰⁾.

En conclusión, la intersección entre la IA y el lenguaje claro en español es un tema puntero que, a pesar de los importantes avances que se han hecho, tiene un amplio margen de recorrido. Se están abriendo multitud de posibilidades inspiradoras y, con seguridad, quedan muchas otras por descubrir, tal como resalta Camacho (2023: 18)⁽⁶¹⁾.

La inteligencia artificial vive un momento de apogeo, especialmente por la popularidad de herramientas como ChatGPT y otras que están en desarrollo. Estas herramientas abren multitud de posibilidades que aún están por descubrir. Entre las que ya son realizables está la de pedirle que expliquen conceptos de manera sencilla, que desarro-

(57) <https://www.diarioepoca.com/1287991-utilizaron-inteligencia-artificial-para-dictar-una-sentencia-en-beron-de-astrada->

(58) <https://hemingwayapp.com/>

(59) <https://www.visiblethread.com/>

(60) https://intranet.caib.es/pidip2front/ficha_noticia.xhtml?lang=es&urlSemantica=strongspan-stylecolorNavyCONSELL-DE-GOVERN-spanstrongSE-FACILITA-LA-ADAPTACIOacuteN-DEL-PROYECTO-ARTEXT-AL-CATALAacuteN--0

(61) CAMACHO, R., "El lenguaje claro, una llave para abrir la transparencia", Transparència i Govern Obert. Sección: Sala de ponents, 2023, en línea. https://governobert.diba.cat/sites/governobert.diba.cat/files/el_lenguaje_claro_llave_para_abrir_la_transparenciacc.pdf.

llen las ideas básicas o presentaciones sobre algún tema y las que queremos imaginar para la automatización en nuestro ámbito de trabajo.

Finalizo este capítulo con una cita muy pertinente de Taboada (2022: 11)⁽⁶²⁾, quien resalta la utilidad de estos sistemas para mejorar las relaciones entre la ciudadanía y las administraciones y, así, contribuir al bienestar social:

Los sistemas de PLN para el lenguaje claro constituyen un ejemplo de aplicaciones con fines sociales. Dentro del marco de la ciencia aplicada y del desarrollo del software, las aplicaciones con fines sociales persiguen desarrollar sistemas que contribuyan a mejorar la sociedad y apoyar el desarrollo individual.

(62) TABOADA, M., "Lingüística computacional y discurso: Hacia la solución de problemas reales", *Lenguaje claro y tecnología en la Administración*, Comares, 2022, págs. 1-13.