

Treball de fi de màster de Recerca

*Evaluación automática de legibilidad en sentencias
judiciales argentinas:
lenguaje claro y acceso a la justicia*

Nom i Cognoms: *Dalila Natalia Milio*

Màster: *Lingüística Teòrica i Aplicada*

Edició: **2021-2022**

Directora: *María Núria Bel Rafecas*

Any de Defensa: **2022**

Col.lecció: **Treballs de fi de màster**

Departament de Traducció i Ciències del Llenguatge

“J’ai compris que tout le malheur des hommes venait de ce qu’ils ne tenaient pas un langage clair. J’ai pris le parti alors de parler et d’agir clairement.”

Albert Camus - La Peste (1947)

RESUMEN

¿Son las sentencias judiciales claras para la ciudadanía? Con el objetivo de contribuir a las iniciativas del lenguaje claro, este estudio se centra en el desarrollo de una maqueta de evaluador automático de legibilidad en el ámbito judicial para la República Argentina que pueda utilizarse como indicador tanto para informar sobre el nivel de claridad de la producción de sentencias publicadas en un determinado período, como para el uso personal de los jueces responsables de su redacción. Para ello, se analizó manualmente un corpus inicial compuesto por sentencias de la Cámara Nacional de Apelaciones en lo Civil de la Ciudad Autónoma de Buenos Aires y, posteriormente, se utilizó un segundo corpus de características similares para evaluar el funcionamiento de la maqueta. Esta demostró resultados eficaces en la evaluación y permitió desarrollar la propuesta del indicador general del nivel de claridad.

Palabras clave: *lenguaje claro, sentencia judicial, acceso a la justicia, República Argentina, evaluación de legibilidad, evaluador automático, indicador.*

Keywords: *plain language, final judgment, access to justice, Argentina, evaluation of legibility, automatic evaluation, index.*

ÍNDICE

1. INTRODUCCIÓN	4
2. ESTADO DEL ARTE	5
2.1 A nivel nacional (República Argentina)	5
2.2 A nivel internacional (fuera de la República Argentina)	8
2.3 Otras referencias relevantes	10
3. OBJETIVOS DEL PRESENTE ESTUDIO	11
4. METODOLOGÍA	12
4.1 Corpus lingüístico	12
4.2 Características observadas y su análisis	13
4.3 Creación de la maqueta de evaluador automático de legibilidad	25
4.4 Propuesta para el indicador general del nivel de claridad	26
5. RESULTADOS	29
6. CONCLUSIÓN	36
7. BIBLIOGRAFÍA	37
8. ANEXO	40

1.INTRODUCCIÓN

La sentencia judicial es un género textual de especialidad. Se trata de un documento predominantemente argumentativo cuyo objeto es la justificación de la decisión del juez sobre un caso concreto. La producción narrativa que hace a la práctica judicial tiene características propias que deben ser consideradas a la hora de analizar una sentencia. Por su parte, la sentencia judicial tiene especial trascendencia para el ciudadano, ya que las consecuencias de lo allí enunciado recaerán directamente sobre él. Sin embargo, si bien las sentencias son accesibles a la ciudadanía, muchas veces, se trata de textos cuyo lenguaje es excesivamente formal, hermético y críptico. Y esta redacción poco clara impide la comprensión por parte de personas que no tienen conocimientos de Derecho. Este derecho a comprender está relacionado con el acceso a la justicia real y efectivo, ya que de poco sirve que un documento esté disponible si la falta de claridad hace que su contenido sea ininteligible.

Por su parte, el movimiento del lenguaje claro tuvo sus orígenes a mediados del siglo XX con el objeto de mejorar la comunicación entre los representantes de la Administración Pública y la ciudadanía. Actualmente, instituciones de todo el mundo aplican guías de lenguaje claro en sus escritos y el movimiento ha crecido en pos de una mejora y simplificación de la redacción de documentos jurídicos y administrativos para garantizar a la ciudadanía y, especialmente, a personas de bajos recursos económicos y educativos, el derecho a la información en lenguaje claro.

Es por ello que el presente trabajo se motivó en el interés de servir de apoyo a la iniciativa del lenguaje claro en el ámbito jurídico argentino y de poder facilitar y automatizar tareas con el uso de técnicas de procesamiento del lenguaje natural. Todo ello, en un intento de combinar algunos de los intereses que motivaron esta investigación: los idiomas, el Derecho, el lenguaje claro y la lingüística, tanto teórica como computacional. Además, con

este trabajo se busca aplicar los conocimientos adquiridos en el Máster en Lingüística Teórica y Aplicada a una problemática local (República Argentina), con el objeto de contribuir de algún modo con el país que vio nacer a quien suscribe el presente trabajo.

2. ESTADO DEL ARTE

2.1 A nivel nacional (República Argentina):

En la actualidad, en la República Argentina existe la Red de Lenguaje Claro que, en el año 2019, publicó la “Guía SAIJ de Lenguaje Claro” de la Dirección Nacional del Sistema Argentino de Información Jurídica. Se trata de un manual que se complementa con un sitio web (www.derechofacil.gob.ar), ambos destinados a orientar la redacción de textos claros. Este manual enumera cuestiones que hacen a la claridad de los textos e incluye recomendaciones sobre diseño, estilo, puntuación, gramática y hasta lenguaje inclusivo.

Asimismo, el Centro de Información Judicial (<https://www.cij.gov.ar/sentencias.html>) publica sentencias, resoluciones y acordadas que emanan de los tribunales que las emiten en el marco de la acordada 24/13 y de las normas que regulan el Sistema Informático de Gestión Judicial.

Con anterioridad a estos hechos, el Dr. Guillermo González Zurro, actual Juez de la sala M de la Cámara Nacional de Apelaciones en lo Civil e integrante de la Red Nacional de Lenguaje Claro, escribió el artículo “Sentencias en lenguaje claro” (2018). Dicho artículo fue la fuente de inspiración inicial del presente trabajo. El autor, como consecuencia de años de experiencia en el dictado de sentencias, describe una serie de características propias de este género textual y propone diferentes recursos para otorgar claridad al mensaje. Es destacable su intento de posicionar al justiciable, como usuario del servicio de justicia, desde un lugar en el que pueda ejercer su derecho a comprender el mensaje emitido en virtud de una sentencia judicial. Para ello, el magistrado se remite al artículo 42 de la Constitución de la Nación

argentina que garantiza, en lo que respecta al Derecho del Consumidor, el derecho a obtener “*información adecuada y veraz*”. Por su parte, la misma Carta Magna hace referencia, en su artículo 18 referido al “debido proceso” que la defensa en juicio de la persona y de sus derechos es inviolable. Por lo tanto, el Dr. González Zurro afirma que “*el acceso a la justicia comprende el derecho a la información en lenguaje claro*” y que el individuo debe obtener información clara durante toda la tramitación del proceso y, especialmente, en la instancia final de la sentencia. Además, el autor resalta que al tratarse de un documento emanado de un juez competente que contiene información sobre la culpabilidad o la inocencia de un sujeto justiciable es de suma importancia que la o las personas involucradas puedan entender el contenido de dicho documento; esto hace al control democrático de los actos emanados de la autoridad pública que le otorgan legitimidad a la sentencia misma. Por último, el magistrado sostiene que la sentencia no debe ser ambigua y que es decisión y responsabilidad del juez expresarse de manera tal que las personas involucradas y, como consecuencia, la sociedad toda, puedan entender el resultado de su escrito.

Por su parte, en el artículo “Una propuesta de lenguaje jurídico claro para la Argentina” (2019), escrito por la Dra. Carmen de Cucco Alconada, la Dra. Mónica Graiewski y otros, se hace referencia a la clara necesidad de aunar esfuerzos para lograr un cambio de paradigma y de la importancia de abrir el camino hacia un lenguaje jurídico claro. La complejidad del lenguaje jurídico, su “ritualismo” y el hecho de que muchos términos jurídicos tienen otro significado en lenguaje común hacen de este tipo de documentos un enigma difícil de descifrar. Lo más destacable de este artículo es la idea de que, en las universidades de Derecho argentinas, los estudiantes se encuentran expuestos constantemente a textos mal escritos y complicados. El resultado da lugar a la adopción de vicios de redacción y a la creencia de que, cuanto más rebuscado sea el texto, más jurídico o académico será. Por lo

tanto, estas personas que se gradúan como abogados y abogadas escriben con menos claridad que cuando ingresaron a la universidad.

Del mismo modo, la Dra. María Paula Pontoriero en su artículo “Acceso a la información jurídica como política pública: servicios innovadores para contribuir a la construcción de ciudadanía” plantea la idea de acceso a la justicia como política pública y la necesidad de desarrollar e implementar una política integral que permita al ciudadano, a través de diferentes herramientas, ejercer sus derechos. La autora afirma que *“los sistemas democráticos dependen de la participación ciudadana, pero para el ciudadano no es posible ejercer sus derechos y participar de manera activa si no comprende la terminología utilizada por los agentes productores del derecho”*. El ciudadano lego queda entonces en una posición de vulnerabilidad ante una ley que debe cumplir y que se asume que conoce, pero que en verdad no comprende.

En otro orden de ideas, durante los últimos años en la República Argentina se ha llevado a cabo un proceso de adaptación, digitalización y sistematización de las comunicaciones oficiales entre la Administración Pública y la ciudadanía. El sistema GDE (Gestión Documental Electrónica) se encuentra operativo en el ámbito estatal y la plataforma TAD (trámites a distancia) le permite al ciudadano realizar numerosas gestiones desde su hogar. Además, el sistema de consultas web del Poder Judicial ha flexibilizado las gestiones judiciales y, hoy en día, permite la presentación de documentos con firma ológrafa escaneada o firma digital.

En lo que respecta al campo de la inteligencia artificial y del Procesamiento del Lenguaje Natural en la República Argentina, aunque todavía quede un largo camino por recorrer, mucho se ha avanzado en los últimos años. Por ejemplo, el Gobierno de la Ciudad Autónoma de Buenos Aires cuenta con un sistema de Chatbox llamado “BOTI” que facilita algunas gestiones y solicitudes de citas. En lo relativo al ámbito jurídico, existe un estudio sobre

detección automática de entidades en textos legales desarrollado por investigadores de Argentina para el programa de investigación e innovación Horizonte 2020 de la Unión Europea, en el marco del proyecto MIREL. El estudio se llama “Mejora del acceso a Infoleg mediante técnicas de procesamiento automático del lenguaje” y plantea un análisis de los diferentes tipos de entidades que se encuentran en los textos legales, sus realizaciones lingüísticas y sus dificultades para el análisis automático.

Finalmente, la Facultad de Derecho de la Universidad de Buenos Aires cuenta con un Posgrado Internacional de Inteligencia Artificial y Derecho y con el IALab, un laboratorio de inteligencia artificial especializado en la investigación académica para impulsar la creación de un ecosistema que combine la detección inteligente, la automatización, la asistencia inteligente y la predicción. En ese contexto, el IALab se asoció con el Grupo de Procesamiento de Lenguaje Natural de la Facultad de Matemática, Astronomía, Física y Computación de la Universidad Nacional de Córdoba para la elaboración de una síntesis de conceptos sobre el Procesamiento del Lenguaje Natural.

2.2 A nivel internacional (fuera de la República Argentina):

Una de las referencias más recientes del movimiento del lenguaje claro en el idioma español es la creación de la Red Panhispánica de Lenguaje Claro. Se trata de un proyecto impulsado por la Real Academia Española (RAE) que nace con el objeto de integrar diferentes iniciativas sobre el lenguaje claro y accesible que se están llevando a cabo en el mundo hispanohablante con el fin de mejorar la comunicación entre el poder público y la ciudadanía. El trabajo de esta red está supervisado por la Asociación de Academias de la Lengua Española (ASALE) y las instituciones integrantes asumen el compromiso de trabajar de forma conjunta para implementar proyectos y medidas que promuevan el uso del lenguaje claro y así lograr una comunicación más eficiente con la ciudadanía.

Por otra parte, España adoptó en el año 2002 la “Carta de Derechos de los Ciudadanos ante la Justicia” gracias al trabajo de la Comisión de Modernización del Lenguaje Jurídico. Dicha Carta enuncia que se debe trabajar en pos de una justicia moderna y abierta a los ciudadanos, que la justicia debe ser comprensible y que *“El ciudadano tiene derecho a que las sentencias y demás resoluciones judiciales se redacten de tal forma que sean comprensibles por sus destinatarios, empleando una sintaxis y estructura sencillas, sin perjuicio de su rigor técnico”*.

Finalmente, Iria da Cunha y M. Ángeles Escobar en su artículo titulado “Recomendaciones sobre el lenguaje claro en español en el ámbito jurídico-administrativo: análisis y clasificación” (2021) destacan la importancia que tiene para la administración pública el hecho de comunicarse con la ciudadanía en lenguaje claro y realizan un relevamiento y análisis de las recomendaciones dadas por diferentes fuentes y organismos. Este artículo sirvió de marco teórico para la implementación de la herramienta informática “ArText claro” (<http://sistema-artext.com/>). Se trata de una herramienta que contiene un editor de texto en el que el usuario debe introducir un documento y, como resultado, el programa le brinda recomendaciones de redacción en lenguaje claro. Este programa fue creado como una herramienta de ayuda a la redacción para la Administración Pública española y ha sentado un precedente en el ámbito del lenguaje claro en español.

A continuación, se detallarán las diferencias entre “ArText claro”, que ha servido de inspiración para el presente trabajo, y el programa que se propondrá a continuación. En primer lugar, “ArText claro” es una herramienta de ayuda en la redacción, mientras que la maqueta de evaluador automático tiene como objetivo específico mostrarle al magistrado qué es lo que medirá el indicador general del nivel de claridad en sentencias judiciales. En segundo lugar, ambos programas se diferencian en el público meta: en el caso de “ArText claro” se dirige al personal de la Administración pública española, mientras que la maqueta

de evaluador automático se dirige al magistrado argentino y también a los profesionales que, en un futuro, evaluarán los resultados del indicador general del nivel de claridad. Finalmente, otra diferencia importante tiene que ver con el género textual, ya que “ArText claro” analiza textos administrativos generalistas y la maqueta de evaluador automático analiza sentencias judiciales argentinas, con sus especificidades.

2.3 Otras referencias relevantes:

La elección de las características frecuentes en sentencias judiciales que tienen un impacto negativo sobre la claridad del lenguaje y que se detallarán a continuación derivó de las recomendaciones de diferentes fuentes de referencia nacionales e internacionales. A nivel nacional, se tuvieron en cuenta los siguientes documentos: el artículo “Sentencias en lenguaje claro” del Dr. Guillermo González Zurro (2018); la “Guía SAIJ de Lenguaje Claro” del Ministerio de Justicia y Derechos Humanos (2019); el “Manual de Lenguaje Claro y Estilo” del Consejo de la Magistratura del Poder Judicial de la Ciudad de Buenos Aires (2021); el “Protocolo para el uso de lenguaje claro” de la Cámara Segunda de Apelación en lo Civil y Departamental, el Centro de Modernización en Litigación y Gestión Judicial no Penal de la Facultad de Ciencias Jurídicas y Sociales de la UNLP (2020); la “Guía de lenguaje claro de la Revista de Interés Legislativo” de la Cámara de Diputados de la Provincia de Buenos Aires (2019); y la “Guía de Lenguaje Claro” del Poder Judicial de Formosa (2020). A nivel internacional, se tuvieron en cuenta los siguientes documentos de referencia: la guía “Lenguaje Claro. Comprender y hacernos entender” (2017) de Clarity y el Instituto de Lectura Fácil y la guía “Cómo escribir con claridad” de la Comisión Europea (2015). Cabe destacar, en este caso, que las dos fuentes internacionales que se utilizaron como referencia fueron inspiradas en el trabajo de da Cunha et al. (2021).

3.OBJETIVOS DEL PRESENTE ESTUDIO

El presente estudio tiene como objetivo general contribuir a las iniciativas relacionadas con el lenguaje claro a través del desarrollo de una maqueta de evaluador automático de legibilidad en el ámbito de la justicia argentina para que pueda utilizarse como indicador tanto para informar sobre el nivel de claridad de la producción de sentencias publicadas en un determinado período, como para el uso personal de los jueces responsables de la redacción de sentencias.

Para conseguir este objetivo general, se han definido los siguientes objetivos específicos:

- 1) Se creará un corpus de sentencias judiciales argentinas correspondientes a un fuero, instancia y período determinados que servirá como objeto de análisis para este trabajo.
- 2) Se recopilará y analizará bibliografía nacional e internacional relacionada con el lenguaje claro en el ámbito jurídico y se seleccionará un conjunto de textos de referencia que servirán para la fundamentación teórica del presente trabajo.
- 3) Se utilizará el corpus mencionado anteriormente para estudiar y analizar manualmente características comunes a las sentencias judiciales argentinas, tales como elementos lingüísticos, sintácticos y de estilo.
- 4) De acuerdo con los criterios y recomendaciones sobre lenguaje claro enunciados en la bibliografía de referencia elegida, se seleccionarán algunas de las características frecuentes en las sentencias judiciales que tienen un impacto negativo sobre la claridad del lenguaje y que pueden ser susceptibles de identificación automática.
- 5) Se procederá a la implementación de una maqueta de evaluador automático de legibilidad, con el objeto de identificar en cada sentencia las características elegidas anteriormente. Para la evaluación de la herramienta, se utilizará un segundo corpus con características similares a las del corpus inicial.

6) Finalmente, y de acuerdo con los resultados obtenidos tanto del análisis manual como automático, se propondrá un sistema de ponderación numérica para cada característica observada para que, a través de la suma de los puntajes de cada una, se obtenga un resultado que sirva como indicador general del nivel de claridad (entre el 0 y el 20, tal como se explicará en la sección “Metodología”). La intención es que este indicador pueda implementarse para el análisis de grandes cantidades de datos, lo que permitirá conocer el estado de situación de un momento dado y plantear objetivos de mejora a futuro.

En definitiva, el evaluador automático de legibilidad cumplirá una doble función: no solo se utilizará como métrica para la creación del indicador general del nivel de claridad, sino que también servirá como recurso para que los jueces puedan conocer e identificar cuáles son los parámetros que mide el indicador y colaborar con el cumplimiento de los objetivos generales de mejora propuestos.

4. METODOLOGÍA

4.1 Corpus lingüístico

Para crear el corpus inicial, se compilaron 65 sentencias judiciales de 13 salas (de la A a la M) de la Cámara Nacional de Apelaciones en lo Civil de la Ciudad Autónoma de Buenos Aires (Argentina). Se seleccionaron de forma aleatoria 5 sentencias de cada sala dictadas durante el primer trimestre del año (entre el 01/01/2022 y el 30/04/2022) con una extensión promedio de entre 500 y 1200 palabras.

Cada una de las sentencias se anonimizó de forma manual, tanto en lo que respecta a los jueces que la redactaron como a las partes involucradas y sus datos personales, con el objetivo de poner el corpus a disposición de futuras investigaciones relacionadas con el tema. El número de sentencias y la extensión promedio se debió, principalmente, a que el análisis preliminar fue manual, a través de la lectura de cada sentencia.

A seguir, teniendo en cuenta algunas de las cuestiones enunciadas en los documentos de referencia que dan cuenta de una falta de claridad en el lenguaje, se identificó manualmente la aparición y repetición de dichas características en las sentencias que conforman el corpus con el fin de seleccionar las que serían más propicias y adecuadas para la detección automática.

Finalmente, se seleccionaron de forma aleatoria otras 13 sentencias (una de cada sala) de la Cámara Nacional de Apelaciones en lo Civil de la Ciudad Autónoma de Buenos Aires, correspondientes al mismo período que se había delimitado para el corpus inicial. Estas sentencias conformaron un nuevo corpus que se utilizó para evaluar la funcionalidad de la maqueta de evaluador automático.

4.2 Características observadas y su análisis

Cantidad de palabras, cantidad de oraciones y promedio de la cantidad de palabras por oración:

De acuerdo con los lineamientos de la “Guía SAIJ de Lenguaje Claro” (2019), la extensión de las oraciones no debe ser de más de 30 palabras para garantizar la claridad del mensaje y que se desarrolle solo una idea por oración. Por ese motivo, se desaconseja el uso de contenido extenso entre comas, paréntesis o corchetes y se aconseja utilizar el punto para separar las ideas. González Zurro (2018) adhiere a la cantidad de palabras indicadas anteriormente. Según lo enunciado en el “Manual de Lenguaje Claro y Estilo” (2021) del Consejo de la Magistratura del Poder Judicial de la Ciudad de Buenos Aires, *“las oraciones deben ser cortas, concretas y expresar una idea. Se recomienda evitar el uso de oraciones con excesiva cantidad de subordinadas”*. A nivel internacional, en la guía “Lenguaje Claro. Comprender y hacernos entender” (2017) de Clarity y el Instituto de Lectura Fácil se recomienda *“redactar oraciones y párrafos de extensión breve”* y *“evitar la profusión de*

oraciones subordinadas”. Por último, en la guía “Cómo escribir con claridad” (2015) de la Comisión Europea se recomienda un promedio de 20 palabras por oración [lineamiento seguido por da Cunha et al. (2021)].

Cabe destacar que establecer fórmulas estructuradas sobre la extensión de una oración puede llevar a clasificar como no claras oraciones que en verdad lo sean. A veces, es preciso conservar ciertos niveles retóricos esenciales para la comunicación, tales como las oraciones subordinadas, ya que ayudan a construir el pensamiento. Es decir, que una oración tenga más de 30 palabras no significa, necesariamente, que deba ser reformulada, pero la advertencia funciona, principalmente, para identificar si el documento presenta una tendencia a las oraciones extensas.

Cantidad de comas por oración:

De acuerdo con A. M. Zorrilla en “El uso de la puntuación en español” (2012), la coma se utiliza para destacar ciertos elementos en la oración, tales como el vocativo, la aposición o inciso explicativo, el adjetivo explicativo, la cláusula absoluta, la interjección, la oración compuesta y subordinada y los conectores.

Por lo tanto, la presencia de muchas comas en una oración puede dar cuenta de que esta contiene aposiciones o incisos, es decir, información adicional. Este contenido entre comas, por lo general, podría eliminarse sin que la oración perdiera el sentido. Con ese criterio, es preciso analizar, a la hora de redactar un texto, si el contenido entre comas es relevante o si puede formar parte de una nota al pie de página.

También las comas pueden indicar la presencia de oraciones subordinadas que tienden a evitarse en exceso en lenguaje claro. Respecto de este asunto, en el “Manual de Lenguaje Claro y Estilo” (2021) se recomienda evitar las “oraciones-párrafo” en las que solo hay comas y usar el punto seguido o el punto y coma para organizar mejor los párrafos.

En otros casos, la presencia de muchas comas en una oración puede deberse a la llamada (informalmente) “coma criminal”, es decir, la coma ubicada erróneamente entre el sujeto y el verbo, el verbo y el objeto, o el verbo y el atributo. Tal como se indica en la guía “Lenguaje Claro. Comprender y hacernos entender” (2017), es preciso “*cuidar los signos de puntuación, los cuales son esenciales para comprender*”. Por su parte, en la guía “Cómo escribir con claridad” (2015) se recomienda establecer un orden correcto de las oraciones para evitar ambigüedades.

En consecuencia, es posible concluir que la presencia de más de 4 comas en una oración (independientemente de la cantidad de palabras que esta tenga) es una característica a tener en cuenta para mejorar la claridad del mensaje.

Cantidad de pares de paréntesis por oración:

En la misma línea de lo explicado en el punto anterior, los paréntesis se utilizan para enmarcar aclaraciones o datos complementarios, e indican que el contenido se aísla de lo que expresa la oración principal (es decir, que no tienen relación directa con el resto de la oración). Otro uso habitual de los paréntesis es para encerrar citas textuales dentro del cuerpo del texto (A. M. Zorrilla en “El uso de la puntuación en español”).

Tal como se indica en la “Guía SAIJ de Lenguaje Claro” (2019), el uso de los paréntesis debe limitarse a aclarar brevemente algún concepto o término técnico o a introducir un ejemplo. Por su parte, el “Protocolo para el uso de lenguaje claro” (2020) de la Cámara Segunda de Apelación en lo Civil y Comercial de La Plata y el Centro de Modernización en Litigación y Gestión Judicial no penal de la Facultad de Ciencias Jurídicas y Sociales de la UNLP recomienda “*Buscar ser concreto sin alargar los párrafos innecesariamente con información, citas o detalles sobreabundantes*”. Por su parte, la “Guía de Lenguaje Claro” (2020) del Poder Judicial de Formosa recomienda el uso de paréntesis

únicamente para “*aclarar brevemente algún concepto o término técnico o para introducir un ejemplo*”.

Asimismo, de acuerdo con el Diccionario panhispánico de dudas de la Real Academia Española, el paréntesis es un “*signo ortográfico (...) que se usa para insertar en un enunciado una información complementaria o aclaratoria*”. En consecuencia, en líneas generales, el contenido entre paréntesis podría reemplazarse por una nota al pie de página.

A nivel de la oración, es posible utilizar como parámetro de referencia lo mismo que para las comas. Si una oración tiene más de dos pares de paréntesis significa, en la mayoría de los casos, que abunda el contenido accesorio o aclaratorio, lo que podría generar problemas de comprensión por parte del lector.

Por último, del análisis manual de las sentencias judiciales, se identificó el siguiente patrón: en una gran cantidad de casos, el contenido entre paréntesis es extenso (más de 15 palabras) y contiene en su interior citas textuales o referencias a jurisprudencia, otra sentencia, ley, decreto o resolución. Por lo tanto, identificar si en una misma oración hay dos o más pares de paréntesis es relevante para el presente estudio, ya que sirve como evidencia de una falta de claridad.

Cantidad de pares de comillas por oración:

Si bien las guías de estilo y de recomendaciones sobre el lenguaje claro en Argentina no brindan sugerencias específicas sobre el uso de comillas, sabemos que se trata de un signo ortográfico cuya principal función es la de enmarcar la reproducción de citas textuales. En una sentencia judicial, el juez hará uso de numerosas citas de autoridad de leyes, decretos, códigos o manuales para fundar su decisión. También es muy frecuente que se haga referencia a instancias previas del proceso, ya sea que se trate de sentencias previas, fallos de la Corte Suprema de Justicia de la Nación que hayan sentado un precedente sobre el tema en cuestión o, incluso, declaraciones testimoniales y/o de las partes del proceso. Estas citas

textuales, en algunos casos, pueden resultar de utilidad. Sin embargo, en otros, se trata de información que no hace al fondo de la cuestión.

Para evitar el exceso de citas, el proceso de digitalización de la justicia facilitó un nuevo recurso que los jueces pueden utilizar. Se trata de la “hipervinculación”, una herramienta que sirve para lograr la integración de documentos de diferente tipo y la remisión a documentos de referencia. La hipervinculación no solo sirve para reducir el uso excesivo de comas, paréntesis y comillas, sino que también sirve para evitar el exceso de notas al pie de página y para permitir al lector acceder a la referencia completa que, muchas veces, le brindará más contexto para la comprensión.

Una vez más y, tal como se lo describió en el punto anterior, a nivel de la oración es posible aplicar en este caso el mismo criterio que para los paréntesis y las comas. Si una oración tiene más de dos pares de comillas, es probable que su contenido pueda reemplazarse por una nota al pie de página o por la hipervinculación al documento de referencia.

En lo que respecta al uso de notas al pie de página, el “Manual de Lenguaje Claro y Estilo” (2021) aconseja su uso para *“evitar distraer la atención del lector del argumento que se desarrolla”*. Con respecto al uso excesivo de citas textuales, indica que *“La idea o argumento tiene que estar desarrollado y explicado en el texto en forma sólida y debe sustentarse sin la cita. Se recomienda acudir a las citas para ejemplificar si hay otros tribunales o doctrina que se expresó en el mismo sentido, pero no las usamos para hablar a través de ellas”*.

Uso de latinismos

Tal como lo indica el Centro de Información Judicial argentino en su página web, *“el lenguaje jurídico se caracteriza por el empleo de latinismos (palabras o frases originales del latín) y extranjerismos (términos procedentes de lenguas vivas extranjeras)”*. Lo que nos interesa analizar en este trabajo es el caso de los latinismos no adaptados, es decir, aquellos

que conservan la grafía y pronunciación originarias y presentan rasgos ortográficos distintos de los del español. Existen dos tipos de latinismos no adaptados: los necesarios o muy extendidos, que son aquellos que no tienen un equivalente exacto en español o cuyo empleo está muy arraigado; por otro lado, los superfluos o innecesarios son aquellos que se usan aun cuando existen equivalentes en español de uso común.

En el caso del discurso jurídico y tal como lo enuncia María Elena Azofra Sierra en su artículo “Criterio normativo y uso general culto en el Diccionario panhispánico de dudas” (2018), *“el carácter de superestrato de la lengua latina respecto al español exige que estos préstamos tengan un tratamiento diferente al de otros extranjerismos. En el uso culto, la voluntad del hablante se inclina por dar a su mensaje un eco de prestigio que solo se consigue si estas voces no se hispanizan demasiado”*. El uso de expresiones en latín, entonces, es considerado como una representación de la lengua culta y evoca prestigio y excelencia. Es decir que, la mayoría de las veces, el hablante emplea un latinismo no solo en función de su significado denotativo, sino también del valor connotativo que trae aparejado.

Todo esto se ve reflejado, por ejemplo, en el “Protocolo para el uso de lenguaje claro” (2020), donde se recomienda *“prescindir completamente la utilización de palabras o frases en latín o castellano antiguo”*. A la misma idea adhiere el Poder Judicial de Formosa en la “Guía de Lenguaje Claro” (2020) y González Zurro (2018). Finalmente, en el “Manual de Lenguaje Claro y Estilo” (2021) se sugiere reemplazar los latinismos por su expresión equivalente en español o explicar su significado, si su uso es inevitable.

Dicho todo lo anterior y con el objeto de garantizar la claridad del mensaje en una sentencia judicial dirigida a ciudadanos no especializados en Derecho, se prefiere el uso de las variantes hispanizadas de los latinismos o su explicación correspondiente.

Expresión de cantidades

Según la Real Academia Española, se escribirán preferentemente con letras los números que pueden expresarse en una sola palabra. Asimismo, la “Guía SAIJ de Lenguaje Claro” (2019), entre sus recomendaciones, contiene la de escribir los números en cifras y evitar el uso de números romanos. A la misma idea adhiere González Zurro (2018), que recomienda expresar las cantidades únicamente en cifras. También, tanto en el “Protocolo para el uso de lenguaje claro” (2020) como en la “Guía de Lenguaje Claro” (2020) se recomienda escribir las cantidades en cifras, a menos que deban escribirse en letras, indefectiblemente, por disposición legal.

Por su parte, la Real Academia Española también indica que *“no es recomendable mezclar en un mismo enunciado números escritos con cifras y números escritos con letra”*. Por lo tanto, la expresión de cantidades duplicadas en letras y cifras, tal como puede verse en muchos escritos judiciales, no es una técnica que mejore la claridad del lenguaje, sino que torna el texto innecesariamente extenso y denso.

Uso del futuro del subjuntivo

El futuro del subjuntivo es un tiempo verbal que tiene un carácter hipotético y se utiliza para expresar eventualidad, posibilidad y contingencia. En general, expresa acciones anteriores a otras acciones que se realizarán en el futuro. Puede ser simple (tuviere) o compuesto (hubiere tenido).

Se trata de un tiempo verbal que ha caído en desuso en el lenguaje común pero que aún se utiliza en el lenguaje jurídico y administrativo; sin embargo, su frecuencia de uso y sus límites son difusos. De acuerdo con lo enunciado en la guía “Lenguaje Claro. Comprender y hacernos entender” (2017), se recomienda evitar los arcaísmos (y se brinda un ejemplo de una oración escrita en futuro del subjuntivo).

En el lenguaje jurídico argentino aún quedan resabios del futuro del subjuntivo, especialmente en textos de doctrina y jurisprudencia más antiguos, aunque su uso se reduce en textos más actualizados tales como el nuevo Código Civil y Comercial (2015). Por ejemplo, el Código Penal de la Nación Argentina (1921) aún conserva este tiempo verbal, principalmente para oraciones condicionales, aunque se trata de una variación meramente estilística que no sigue patrones específicos, ya que se utiliza de forma indistinta e intercambiable con el presente del subjuntivo y con el pretérito imperfecto del subjuntivo. Un ejemplo de ello se ve reflejado en el artículo 94 que reza: “*Si las lesiones fueran de las descritas en los artículos 90 o 91 y fueren más de una las víctimas lesionadas, el mínimo de la pena prevista en el primer párrafo, será de seis (6) meses o multa de tres mil (3.000) pesos e inhabilitación especial por dieciocho (18) meses*”.

Por su parte, la Real Academia Española se pronuncia al respecto al decir que “*El futuro de subjuntivo ha desaparecido prácticamente de la lengua ordinaria y solo pervive en textos normativos y otros de índole oficial que se caracterizan por su lenguaje arcaizante*”. En esa misma línea, Gerhard Edelman, en el artículo “El futuro de subjuntivo en el lenguaje jurídico español: ¿una forma verbal destinada a desaparecer?”, sostiene que esta forma verbal conlleva un alejamiento y distanciamiento del legislador que se sirve del futuro del subjuntivo con un objetivo solemne para enfatizar la distancia entre él y los justiciables. Esto no hace más que alimentar la idea de una autoridad omnipotente que debe ser obedecida. En el mundo moderno, donde se intenta otorgar claridad al lenguaje jurídico para que sea inteligible y asequible al ciudadano común, este tiempo verbal se torna obsoleto, incompatible y hasta forzado.

Sin embargo, existen opiniones diversas, más y menos extremas sobre el uso de este tiempo verbal en el discurso de especialidad jurídico. Así como existen autores que se oponen a su uso por los motivos enunciados anteriormente, otros de posición más intermedia se

inclinan por no eliminarlo del todo, siempre que este sea correctamente empleado. Por su parte, los defensores de continuar utilizando el futuro del subjuntivo consideran que, tal como su uso aún persiste en el idioma portugués, la lengua española es una lengua rica en tiempos verbales y perderlo sería perder la posibilidad de enriquecer el lenguaje jurídico y de resaltar matices de elegancia que con otros tiempos verbales no sería posible resaltar. En esa línea de ideas, Daniel Fajardo (1997) en el “Boletín de los traductores españoles de las instituciones de la Unión Europea” enuncia que el futuro del subjuntivo añade un matiz de improbabilidad e incertidumbre que lo hace característico y lo diferencia del resto de los tiempos verbales.

Uso de siglas y abreviaturas

Según María Marta García Negroni en el libro “Escribir en español. Claves para una corrección de estilo” (2011), *“las abreviaturas y las siglas constituyen distintas formas de abreviar, mediante la supresión de determinadas letras o sílabas, la representación gráfica de una palabra o de una expresión”*. Por un lado, las abreviaturas consisten en eliminar algunas de las letras o sílabas de la escritura completa de la palabra. Existen las abreviaturas por truncamiento, que suprimen las letras o sílabas finales (artículo y art.), y las abreviaturas por contracción, que eliminan las letras centrales de las palabras y conservan las más representativas (señor y sr.). Con respecto al uso de mayúscula o minúscula en las abreviaturas, se siguen las reglas de la palabra de origen. Por su parte, las siglas son yuxtaposiciones de las iniciales de las palabras que integran un enunciado o entidad y se usan para referirse de modo abreviado a organismos, instituciones, empresas, entre otros. A diferencia de las abreviaturas, las siglas se escriben sin puntos y sin espacios de separación entre las letras que las constituyen.

En el lenguaje jurídico, los límites entre abreviaturas y siglas son muy difusos, ya que se observan tanto abreviaturas sin punto al final como siglas con cada una de sus letras separadas por un punto, lo cual genera aún más confusión para la lectura e interpretación.

Asimismo, si bien algunas siglas y abreviaturas son de uso extendido en el ámbito del Derecho y forman parte de la jerga jurídica, el magistrado no puede asumir que el ciudadano lego conoce a qué hace referencia esa abreviatura o sigla.

Por su parte, en la guía “Cómo escribir con claridad” (2015) se ofrecen tres opciones interesantes para el tratamiento de siglas y abreviaturas: *“escribir los términos con todas sus letras, en caso de que solo aparezcan una o dos veces en el documento, o, si aparecen en más ocasiones, escribir los términos con todas sus letras la primera vez, seguidos de su sigla o abreviatura entre paréntesis, y utilizar esta después en el resto del documento, y/o adjuntar una lista de siglas/abreviaturas con su significado o un hipervínculo que explique ese significado”*. Asimismo, la “Guía de Lenguaje Claro” (2020) recomienda *“evitar el uso de abreviaturas sin aclarar cuál es su significado la primera vez que son utilizadas (ej.: Código Civil, en adelante C.C.)”*. Finalmente, en el “Manual de Lenguaje Claro y Estilo” (2021) se recomienda también evitar las abreviaturas salvo que, en la primera mención se aclare su significado.

Uso del gerundio

El gerundio es una forma no finita del verbo que tiene un significado durativo y, en líneas generales, expresa una acción simultánea o anterior a la del verbo principal. El gerundio funciona, desde un punto de vista sintáctico, como un adverbio de modo que modifica al verbo y, por ello, es invariable en género y número.

Tal como lo describe A. M. Zorrilla en su libro “El uso del verbo y del gerundio en español” (2004), *“el gerundio simple expresa acción durativa e inacabada, imperfecta, simultánea o inmediatamente anterior al verbo personal de la oración en que aparece”*. Por su parte, *“el gerundio compuesto expresa acción perfecta, acabada, anterior a la del verbo principal de la oración en que se encuentra”*.

Sin embargo, debido a la influencia de idiomas como el inglés y el francés y a siglos de traducciones poco precisas y de calcos, el uso del gerundio en español constituye un punto delicado y de discusión recurrente entre gramáticos de la lengua. Es por ello que, si bien existen reglas específicas, los límites en el uso del gerundio se vuelven difusos y derivan en muchos casos de ambigüedad. De acuerdo con el “Manual de Lenguaje Claro y Estilo” (2021), se deben evitar los gerundios que no permiten identificar con facilidad quién realiza la acción o cuál es la relación lógica entre las acciones (cuál es la causa y cuál la consecuencia, el antecedente y el consecuente, etc.).

Asimismo, María Marta García Negroni (2011) describe que los tres casos más comunes de uso incorrecto del gerundio se dan en el caso del “gerundio de posterioridad”, “gerundio de consecuencia” y “gerundio especificativo”. El gerundio de posterioridad es aquel en el que el gerundio expresa una acción posterior a la del verbo principal (“*Terminó el trabajo entregándolo al día siguiente.*” bien podría reemplazarse por “*Terminó el trabajo y lo entregó al día siguiente.*”). Este uso del gerundio no se recomienda, aunque tanto Manuel Seco (1961) como A. M. Zorrilla (2004) coinciden en que no resulta incorrecto usarlo cuando la acción posterior a la del verbo principal denota una posterioridad inmediata. Por su parte, el gerundio de consecuencia hace referencia al hecho de que el verbo en gerundio sea la consecuencia directa del verbo principal (“*Estudió durante el verano pudiendo recibirse a principios de marzo.*” podría reemplazarse por “*Estudió durante el verano, por lo que pudo recibirse a principios de marzo.*”). Finalmente, el gerundio especificativo ocurre cuando el gerundio pierde su carácter adverbial como complemento del verbo y se convierte en un complemento de un sustantivo, por lo que pasa a cumplir la función de adjetivo (“*La ley regulando esta problemática.*” que puede reemplazarse por una proposición relativa como “*La ley que regula esa problemática.*”).

En consecuencia y tal como lo indica Manuel Seco en su “Diccionario de Dudas de la lengua española” (1961), *“la abundancia de construcciones de gerundio en una oración, aun en el caso de que sean gramaticalmente correctas, denota pobre dominio del idioma; sobre todo cuando la única finalidad de su uso es estirar la oración, convirtiendo en una sola oración larguísima lo que lógicamente habría de ser una serie de oraciones independientes más cortas”*.

En el discurso jurídico, abundan los gerundios mal utilizados que no denotan simultaneidad y que bien podrían reformularse para ofrecer una versión más clara de la misma oración y evitar la ambigüedad. En el material de referencia y, más particularmente, en la “Guía de lenguaje claro de la Revista de Interés Legislativo” de la Cámara de Diputados de la Provincia de Buenos Aires (2019), se indica que el uso del gerundio puede afectar la claridad del texto por lo que la revista ofrece ejemplos de construcciones incorrectas y sus posibles cambios.

Uso de la voz pasiva perifrástica

La voz pasiva es una particularidad sintáctica de los verbos transitivos. La pasiva perifrástica se construye con el verbo “ser” o “estar” conjugado en el tiempo y modo que corresponden, seguido del participio pasado concordado del verbo de la oración activa. El objeto directo de la voz activa se transforma en el sujeto de la voz pasiva, el verbo se convierte en una perífrasis verbal y el sujeto de la voz activa se convierte en el complemento agente (si el sujeto de la oración principal es expreso), precedido por la preposición “por”.

El tema del uso y abuso de la voz pasiva en español es un debate que data de largo tiempo entre gramáticos y lingüistas. Sin embargo, nada indica la Real Academia Española sobre el impedimento de su uso, sino que aconseja utilizarla con precaución.

En el ámbito jurídico, el uso excesivo de la voz pasiva proviene, en general, de un intento por parte del legislador o magistrado de atenuar su discurso, de tomar distancia y

ocultar, en cierto modo, su identidad (la del agente) para generar despersonalización o exención de responsabilidad sobre lo que se expresa.

Asimismo, de acuerdo con la “Guía SAIJ de Lenguaje Claro” (2019), el uso de la voz pasiva puede afectar a la claridad y comprensión del mensaje ya que no se sigue el orden canónico de la oración (sujeto + verbo + complemento). Además, la voz pasiva torna, indefectiblemente, la oración más extensa. Las mismas ideas enuncian la guía “Lenguaje Claro. Comprender y hacernos entender” (2017) y la guía “Cómo escribir con claridad” (2015), en las que se recomienda utilizar más construcciones activas que pasivas, siempre que sea posible.

4.3 Creación de la maqueta de evaluador automático de legibilidad

La maqueta de evaluador automático de legibilidad creada para el presente trabajo está basada en reglas manuales. Estas reglas representan algunas de las características distintivas del lenguaje poco claro en sentencias judiciales argentinas, que se observaron a partir del análisis manual del corpus inicial y que son susceptibles de identificación automática. Además, se tomó como fuente de inspiración el artículo da Cunha et al. (2021) y la herramienta “ArText claro”.

La maqueta se desarrolló en lenguaje de programación Python dentro de la plataforma Google Colaboratory, para que cualquier persona que cuente con el enlace de acceso pueda probarla. En ella se representa una forma de identificar las características explicadas en los apartados anteriores, aunque es posible que también existan otras formas de obtener resultados semejantes. Es por ello que se espera que esta maqueta pueda servir para desarrollos futuros y queda abierta para propuestas de mejora.

El programa intercala instrucciones y líneas de código para que el usuario pueda guiarse sin mayores inconvenientes. En una primera instancia, se importan diferentes librerías

de Python que se utilizan para tareas relacionadas con el Procesamiento del Lenguaje Natural, tales como Pandas, Nltk, Re, Spacy y Word cloud. Pandas es una librería que se utiliza para la manipulación y el tratamiento de datos. Nltk es una librería que permite realizar tareas de tokenización, clasificación, etiquetado, y análisis, entre otras. La tokenización es un método que separa un texto en unidades más pequeñas, llamadas tokens, que se utilizan para diferentes tipos de análisis. Asimismo, el módulo “re” (de regular expressions o regex) permite recurrir al uso de expresiones regulares, es decir, secuencias de caracteres que conforman un patrón de búsqueda aplicable a una cadena de texto. Por su parte, Spacy es una biblioteca que ofrece herramientas para el etiquetado de categorías gramaticales (PoS). En este tipo de etiquetado gramatical, se le asigna a cada token o palabra una categoría gramatical. Por último, WordCloud es un complemento que permite crear nubes de palabras para visualizar textos de forma gráfica.

Todas estas herramientas sirvieron para la creación de la maqueta de evaluador automático de legibilidad, cuyos resultados se describirán en la siguiente sección, y a la que es posible acceder (junto con las 13 sentencias anonimizadas que conforman el corpus de evaluación) a través del enlace disponible en el “Anexo”.

4.4 Propuesta para el indicador general del nivel de claridad

Tal como se describió en la sección “Objetivos”, la maqueta de evaluador automático de legibilidad se creó con el fin último de asignar una ponderación a cada característica observada para crear un indicador general del nivel de claridad que pueda aplicarse a grandes cantidades de datos. Para dicha ponderación, se tuvieron en cuenta no solo el análisis manual de las 65 sentencias del corpus inicial, sino también la evaluación realizada posteriormente sobre 13 sentencias con la maqueta de evaluador automático y la frecuencia de aparición promedio de cada característica (ver sección de “Resultados”).

A partir de ese análisis, se creó un sistema de puntuación del 0 al 20. El objetivo es que, a futuro, el evaluador automático pueda contar automáticamente el número de apariciones de cada característica y, según la regla preestablecida, devolver como resultado un valor de 0, 1 o 2. La suma de los resultados de todas las características dará como resultado final un número del 0 al 20 que servirá de indicador general del nivel de claridad, cualquiera sea la longitud del documento o la cantidad de datos a medir. Un indicador más cercano a 20 representará una tendencia hacia un lenguaje poco claro y un indicador más cercano a 0 representará una tendencia hacia un lenguaje más claro. Por lo tanto, se espera que con la observación del puntaje o índice general obtenido en un momento dado, se puedan establecer objetivos de mejora en un plazo determinado para, al finalizar ese plazo, volver a efectuar el análisis y comparar los resultados con la medición anterior.

La propuesta para la ponderación de cada una de las características estudiadas se definió a nivel de la oración (frecuencia observada cada 10 oraciones) y a nivel de la sentencia (frecuencia observada cada 1000 palabras). En la página siguiente, se presenta una tabla explicativa de la propuesta:

Propuesta para el indicador general del nivel de claridad

Cada 10 oraciones:	Cant. de oraciones	Puntaje	Cant. de oraciones	Puntaje	Cant. de oraciones	Puntaje
Más de 30 palabras por oración	0	0	1-3	1	4 o +	2
Más de 4 comas por oración	0		1 o 2		3 o +	
Más de 2 pares de paréntesis por oración	0		1		2 o +	
Más de 2 pares de comillas por oración	0		1		2 o +	
Cada 1000 palabras:	Cant. de apariciones	Puntaje	Cant. de apariciones	Puntaje	Cant. de apariciones	Puntaje
Uso de latinismos	0	0	1	1	2 o +	2
Expresión de cantidades	0		1		2 o +	
Uso del futuro del subjuntivo	0		1		2 o +	
Uso de siglas y abreviaturas	0		1-4		5 o +	
Uso del gerundio	0		1 o 2		3 o +	
Uso de la voz pasiva perifrástica	0		1 o 2		3 o +	

5. RESULTADOS

Separación de la sentencia en oraciones:

Con la descarga del módulo “punkt” de la librería NLTK, se obtienen modelos para la tokenización de textos en diferentes idiomas (incluido el español). En una primera instancia, el proceso de tokenización se realizó a nivel de la oración, con el objetivo de identificar todas las oraciones que conforman la sentencia.

Sin embargo, dada la complejidad de las sentencias y la gran variedad de siglas y abreviaturas que, además, no se expresan de forma consistente en los diferentes documentos, la tarea de separación en oraciones resultó dificultosa. Es preciso destacar la importancia de que este primer paso de separación en oraciones funcione correctamente, ya que una posible falla afectaría el resultado de los siguientes pasos que analizan el documento a nivel de la oración (cantidad de oraciones de la sentencia, cantidad promedio de palabras por oración y advertencia en caso de que la oración tenga más de 30 palabras, cantidad de comas, paréntesis y comillas de cada oración).

Por ese motivo, se buscó una librería en la que se pudieran añadir manualmente abreviaturas comunes a este tipo de documento. Asimismo, fue preciso agregar un paso previo a la tokenización para realizar algunos reemplazos tipográficos, tales como la sustitución de oraciones que terminan en “.-” por “.” (típico de la redacción de escritos judiciales).

Finalmente, el texto original (“sentencia”), se tokenizó y quedó separado en oraciones (“oraciones”).

En líneas generales, los resultados de la separación en oraciones han sido satisfactorios para las 13 sentencias que integraron el corpus de evaluación. Sin embargo, es posible que el sistema falle cuando la sentencia original contenga errores de puntuación (puntos en el lugar

incorrecto) o siglas y abreviaturas poco comunes separadas por puntos. Es por ello que una de las primeras instrucciones de la maqueta de evaluador automático es que el usuario pase el corrector ortográfico de Word y se asegure de que todas las oraciones se encuentran separadas por un punto.

Gráfico de frecuencia de uso de palabras:

Si bien se trata de un parámetro adicional que, en principio, no será tenido en cuenta para el indicador general del nivel de claridad, el gráfico se agregó para identificar las palabras o términos más utilizados y/o relevantes del documento importado, lo que permite visualizar, a través de las palabras clave, si el objetivo comunicacional se cumple.

Luego de realizar pruebas con diferentes sentencias, se tomó la decisión de transformar el texto original (“sentencia”) a minúscula para evitar repeticiones de palabras y de incluir una lista de reemplazos para eliminar “StopWords” o palabras vacías, es decir, aquellas palabras que no forman parte del léxico de especialidad y que solo son funcionales en el contexto de la oración y en su interacción con otras palabras. En español, la lista de palabras vacías está formada por artículos, preposiciones, conjunciones y pronombres. Esto permitió mejorar los resultados de la nube de palabras.

Cantidad de palabras, cantidad de oraciones y promedio de la cantidad de palabras por oración:

Se aplicó la función “len” para obtener el resultado de la cantidad de palabras y oraciones de la sentencia. A continuación, se calculó el número promedio de palabras por oración.

También el programa ofrece la posibilidad de visualizar cuáles son las oraciones que tienen más de 30 palabras (y que, según lo explicado anteriormente, podrían ser susceptibles de adaptación) y, además, cuántas palabras tiene cada una de esas oraciones.

Dado que la separación en oraciones funcionó con éxito en las 13 sentencias del corpus de evaluación, todos estos parámetros también funcionaron correctamente. De este modo, se pudo observar que el promedio general de palabras por oración fue de 40,26 y que, también en promedio, más de la mitad de las oraciones de cada sentencia tenían más de 30 palabras.

Cantidad de comas, pares de paréntesis y pares de comillas por oración:

Para el caso del cálculo de la cantidad de comas por oración, se tomó como base el texto tokenizado (“oraciones”) y se programó un contador que devuelve como resultado únicamente aquellas oraciones que tienen más de 4 comas.

Si bien, en líneas generales, las pruebas han resultado satisfactorias, es importante mencionar que el resultado puede devolver algunos falsos positivos. Esto ocurrió, por ejemplo, en casos en los que la coma se utiliza para separar elementos cortos de una enumeración, tales como sustantivos, adjetivos, verbos o adverbios, por lo que la cantidad de comas de la oración no implica, necesariamente, que esta sea particularmente extensa.

El mismo tipo de contador mencionado anteriormente se aplicó para devolver como resultado aquellas oraciones que tienen más de 2 pares de paréntesis y de comillas.

Por último, cabe destacar una particularidad con respecto a las comillas. De acuerdo con lo enunciado en “El uso de la puntuación en español” (2012), las comillas utilizadas en español pueden ser españolas (« »), inglesas (“ ”) y simples (‘ ’). Esta clasificación no tiene en cuenta las comillas rectas utilizadas en inglés (" ") que, actualmente, se observan con bastante frecuencia en textos en otros idiomas. En virtud de esta variedad de opciones y con el objeto de lograr la identificación de todas las comillas, se decidió aplicar el reemplazo en el texto sin tokenizar (“sentencia”) de todas las comillas de apertura españolas, simples y rectas a las comillas inglesas (“”) para unificar el criterio de búsqueda y lograr resultados más precisos.

Luego de evaluar el programa en las 13 sentencias del corpus de evaluación, se comprobó que, con un promedio general de 15,84 oraciones por sentencia, el promedio de oraciones con 4 comas o más fue de 4,38, mientras que el de más de dos pares de paréntesis y de comillas por oración fue mucho menor: 0,61 y 0,38, respectivamente.

Uso de latinismos:

Para la identificación de latinismos en la sentencia, se procedió a crear una lista de latinismos propios del discurso jurídico y se estableció una expresión regular para identificar los latinismos tanto en letra minúscula como mayúscula. Como esta identificación es a nivel de la sentencia, se utilizó el texto sin tokenizar (“sentencia”).

Cabe destacar también que, en un primer momento, se intentó identificar latinismos luego de tokenizar el texto por palabras, pero el sistema presentaba errores, ya que identificaba como dos latinismos diferentes a “ipso” y “facto” o a “prima” y “facie”.

Los resultados del programa fueron precisos con respecto a la identificación de latinismos en las 13 sentencias que se utilizaron para la evaluación y se observó que, teniendo en cuenta que el promedio general de palabras por sentencia fue de 629,38, la cantidad de latinismos fue, en promedio, de 0,84 por sentencia.

Expresión de cantidades:

El mismo intento de tokenizar el texto por palabras falló en el caso de la expresión de cantidades, porque el objetivo del programa no era identificar solo números cardinales, sino cantidades de dinero (que son las que habitualmente aparecen en los textos jurídicos en cifras y letras) y que, en Argentina, vienen acompañadas de la palabra “pesos” o “dólares”.

Asimismo, se intentó utilizar el etiquetador de categorías gramaticales (PoS) para identificar los números cardinales, pero no se pudo establecer un patrón uniforme que los identificara a todos. Por otro lado, el etiquetador no identificaba palabras en mayúscula y, en

la mayoría de los casos, las cantidades de dinero expresadas en letras aparecen en mayúscula en las sentencias.

Por ese motivo, se llevó a cabo el mismo proceso que en el caso de los latinismos, con una lista de expresiones regulares que se aplicara tanto para minúsculas como para mayúsculas.

El programa demostró ser efectivo en la identificación de las expresiones de cantidad en las 13 sentencias que conformaron el corpus de evaluación. Una vez más, tomando como base el promedio general de cantidad de palabras por sentencia (629,38), el promedio de la cantidad de latinismos por sentencia fue de 0,46.

Uso del futuro del subjuntivo:

En lo que respecta a la identificación de los casos de uso del futuro del subjuntivo, se intentó utilizar el etiquetador de categorías gramaticales, pero dadas las limitaciones del sistema para el idioma español y la poca frecuencia de uso del futuro del subjuntivo, el etiquetador no pudo identificar dicho tiempo verbal.

En consecuencia, se procedió a utilizar el texto sin tokenizar (“sentencia”) con una expresión regular que identificara todas las palabras terminadas en “are”/”ares” y “ere”/”eres”, sin distinción de mayúsculas o minúsculas.

Los resultados de la evaluación demostraron que el programa funciona correctamente, es decir, que identifica todos los casos de futuro del subjuntivo (el promedio de casos de futuro de subjuntivo cada 629,38 palabras fue de 0,38). Sin embargo, este valor no es preciso porque el programa devuelve algunos falsos positivos, como es el caso de los verbos terminados en “ere” en segunda y tercera persona del singular del presente del indicativo (requiere/refiere) y los verbos terminado en “are” y “ere” en segunda y tercera persona del singular del presente del subjuntivo (declare/supere). Por lo tanto, se espera poder mejorar este aspecto del programa a futuro.

Uso de siglas y abreviaturas:

Tal como se explicó anteriormente, la presencia de una gran cantidad y variedad de siglas y abreviaturas poco consistentes en el corpus dificultó no solo la tarea de separación del texto en oraciones, sino también la de identificación de siglas y abreviaturas propiamente dichas. Asimismo, muchas abreviaturas no terminan con un punto como es debido y muchas siglas contienen puntos entre las letras que las conforman, lo que hizo que la lista de expresiones regulares para abreviaturas tuviera que adaptarse por completo al tipo de documento analizado. Es por ello que se observan los signos (¿?), que indican que la búsqueda se realiza sin discriminar minúsculas o mayúsculas, y luego se presentan ejemplos de cada sigla y abreviatura observada en el corpus inicial separada por puntos y sin puntos.

En lo que respecta a los resultados y, teniendo en cuenta el promedio de 629,38 palabras por sentencia, se observaron, en promedio, 6,46 siglas y abreviaturas por sentencia. Además, en 3 de las 13 sentencias se pudo observar que la forma de expresar una misma sigla o abreviatura no era consistente (por ejemplo, “CSJN” y “C.S.J.N” o “IGJ” e “I.G.J.”). Es importante destacar, con respecto a esta última observación, el esfuerzo realizado para simplificar la expresión regular y poder abarcar la mayor cantidad de casos posibles. Sin embargo, el sistema está limitado a la lista de siglas y abreviaturas observadas en el corpus inicial, por lo que es posible que, al probarlo en otras sentencias, existan falsos negativos. En resumen, en líneas generales se observa una alta frecuencia de uso de siglas y abreviaturas en sentencias judiciales argentinas.

Uso del gerundio:

En el caso del gerundio, se estableció una regla para identificar la aparición de gerundios dentro del texto tokenizado (“oraciones”) a través del uso del etiquetador de categorías gramaticales (PoS). Por lo tanto, luego de haber identificado la etiqueta correspondiente al gerundio como forma verbal “Ger”, el programa logró devolver como

resultado todos los casos de gerundio que aparecen en las 13 sentencias de evaluación con muy buenos resultados, excepto aquellos casos de gerundios que se encuentran al principio de la oración (contienen mayúscula inicial) o todas sus letras en mayúscula. Esto se debe a la limitación del etiquetador de categorías gramaticales en la identificación de palabras con mayúscula.

En lo que respecta a la frecuencia en el corpus de evaluación, se observó una frecuencia promedio de 1,38 gerundios cada 629,38 palabras.

Uso de la voz pasiva perifrástica:

Finalmente, el mismo procedimiento que para los gerundios se utilizó para identificar las construcciones pasivas perifrásticas con buenos resultados. El programa se configuró para identificar de forma precisa y correcta todas las construcciones pasivas perifrásticas que contienen el verbo “ser” o “estar” como auxiliar (etiqueta “AUX”), seguido de un participio (forma verbal “Part”).

Las 13 sentencias que conforman el corpus de evaluación demostraron una frecuencia promedio de 0,84 expresiones en voz pasiva perifrástica cada 629,38 palabras. Sin embargo, este resultado no es preciso porque el etiquetador de categorías gramaticales no identificó (falsos negativos) los casos en los que el verbo “ser” se conjugó en futuro del indicativo (esto ocurrió en 4 de las 13 sentencias con los siguientes ejemplos: será remitida, serán desestimados, serán detalladas y serán distribuidas). Es posible que esto se deba al uso de la tilde en el verbo “ser” en futuro. Por lo tanto, se espera poder mejorar este aspecto del programa a futuro.

6. CONCLUSIÓN

El análisis efectuado en el contexto del presente trabajo demostró la necesidad de impulsar la iniciativa del lenguaje claro en la redacción de sentencias judiciales argentinas para promover un cambio de paradigma en la forma de redactar y poder garantizar así un mayor grado de entendimiento por parte de la ciudadanía y un acceso a la justicia más igualitario.

Asimismo, si bien las observaciones se realizaron sobre el género “sentencia judicial”, es probable que el análisis de otros textos del ámbito jurídico proporcione resultados semejantes, dados los vicios de redacción que caracterizan a los textos jurídicos en general.

En lo que respecta a la elección de las características utilizadas para la evaluación de legibilidad, es posible reconocer que algunas de las más generales coinciden con las identificadas por “ArText claro”, tales como la advertencia de oraciones extensas (sobre 20 palabras en lugar de 30), la identificación del futuro del subjuntivo (observado dentro de la categoría “arcaísmos”), la identificación de los gerundios y la voz pasiva. Sin embargo, la originalidad de la maqueta de evaluador automático es que permite identificar otras características propias, tales como el número de comas, paréntesis y comillas a nivel de la oración, los latinismos, la expresión de cantidades en letras y las siglas y abreviaturas características de la sentencia judicial argentina.

Por su parte, la maqueta de evaluador automático demostró buenos resultados en la identificación de las características enunciadas, lo que permitió calcular el número de apariciones de dichas características en las 13 sentencias del corpus de evaluación para crear la ponderación aplicable al modelo de indicador general del nivel de claridad propuesto. Se espera que este modelo sirva, a futuro, para poder aplicar el indicador general del nivel de claridad en el ámbito jurídico argentino y poder así identificar tendencias y generar propuestas de mejora hacia un lenguaje jurídico cada vez más claro.

En conclusión, el desarrollo del presente estudio resultó productivo y permitió, en líneas generales, lograr los objetivos propuestos. Además, se espera que este trabajo sirva de base para el desarrollo de investigaciones futuras y más exhaustivas sobre el tema en cuestión.

7. BIBLIOGRAFÍA

Libros

García Negroni, M. M. (2011). *Escribir en español: claves para una corrección de estilo*. - 2a ed. - Buenos Aires: Santiago Arcos editor.

Tula del Moral, M. L.; Núñez Gelvez, P. (2021). *Manual de lenguaje claro y estilo*. - 1a ed. - Ciudad. Autónoma de Buenos Aires: Editorial Jusbaire.

Zorrilla, A. M. (2012). *El uso de la puntuación en Español*. 3a ed. - Buenos Aires: Fundación Litterae.

Zorrilla, A. M. (2004). *Uso del verbo y del gerundio en español*. Buenos Aires: Fundación Litterae.

Artículos

Azofra Sierra, M. E. (2018). *Criterio normativo y uso general culto en el Diccionario panhispánico de dudas*. Revista de Lexicografía, 13, 7-16.
<https://doi.org/10.17979/rlex.2007.13.0.4782>

Cardellino, F.; Cardellino, C.; Haag, K.; Soto, A.; Teruel, M., Alonso i Alemany, L; Villata, S. (2019). Mejora del acceso a Infoleg mediante técnicas de procesamiento automático del lenguaje. XVIII Simposio Argentino de Informática y Derecho (SID)-JAIIO 47 (CABA, 2018).

da Cunha, I., & Escobar, M. A. (2021). Recomendaciones sobre lenguaje claro en español en el ámbito jurídico-administrativo: análisis y clasificación. *Pragmalingüística*, (29), 129-148. <https://doi.org/10.25267/Pragmalinguistica.2021.i29.07>

De Cucco Alconada, C. y Graiewski, M. (2019). Una propuesta de lenguaje jurídico claro para la Argentina. Publicado en: *Temas de Derecho Penal y Procesal Penal*, Erreius. Cita digital: IUSDC286619A

Edelmann, G. (2010). El futuro de subjuntivo en el lenguaje jurídico español: ¿una forma verbal destinada a desaparecer?

Fajardo, D. (1997), *puntoycoma*, Boletín de los traductores españoles de las instituciones de la Unión Europea, [en línea]. Disponible en: <http://ec.europa.eu/translation/bulletins/puntoycoma/47/pyc475.htm> [Fecha de la consulta: 26/06/2022].

Fernandez Tucci, T.; Carro, M. V.; Almarcha C., M. A (2021). Informe de Inteligencia Artificial y Procesamiento del Lenguaje Natural. UBA IALAB – UNC. Disponible en: <https://ialab.com.ar/wp-content/uploads/2021/09/Informe-UBA-IALAB-UNC.-Inteligencia-Artificial-y-Procesamiento-de-Lenguaje-Natural.pdf>

González Zurro, G. (2018). *Sentencias en lenguaje claro*. Thomson Reuters La Ley.

Pontoriero, M. P. (2015). Acceso a la información jurídica como política pública: servicios innovadores para contribuir a la construcción de ciudadanía. XX Congreso Internacional del CLAD sobre la Reforma del Estado y de la Administración Pública, Lima, Perú.

Diccionarios

Real Academia Española: Diccionario panhispánico del español jurídico (DPEJ) [en línea]. Disponible en: <https://dpej.rae.es/> [Fecha de la consulta: 26/06/2022].

Real Academia Española: Diccionario panhispánico de dudas (DPD) [en línea].
Disponible en: <https://www.rae.es/dpd/> [Fecha de la consulta: 26/06/2022].

Seco, M. (1961). Diccionario de dudas de la lengua española. Madrid: Aguilar.

Recursos digitales

Carta de Derechos de los ciudadanos ante la Justicia (2002). Ministerio de Justicia. España. Disponible en: <https://sedejudicial.justicia.es/carta-de-derechos-de-los-ciudadanos> [Fecha de la consulta: 26/06/2022].

Código Civil y Comercial de la Nación. (2015). Buenos Aires: Honorable Cámara de Diputados de la Nación. Disponible en: <http://servicios.infoleg.gob.ar/infolegInternet/anexos/235000-239999/235975/norma.htm> [Fecha de la consulta: 26/06/2022].

Código Penal de la Nación Argentina. (1921). Buenos Aires: Honorable Congreso de la Nación Argentina. Disponible en: <http://servicios.infoleg.gob.ar/infolegInternet/anexos/15000-19999/16546/texact.htm> [Fecha de la consulta: 26/06/2022].

Guía SAIJ de Lenguaje Claro. Sistema Argentino de Informática Jurídica. Publicado en: lenguajeclaro.gob.ar (2019). Disponible en: <http://lenguajeclaroargentina.gob.ar/wp-content/uploads/2019/08/Gu%C3%ADa-SAIJ-de-lenguaje-claro.pdf> [Fecha de la consulta: 26/06/2022].

Guía de lenguaje claro. Revista de Interés Legislativo (2019). Buenos Aires. Cámara de Diputados. Disponible en: <https://institucional.hcdiputadosba.gov.ar/ppi/camaraaccesible/publicaciones/Version%20WEB%20Lenguaje%20Claro.pdf> [Fecha de la consulta: 26/06/2022]

Guía de Lenguaje Claro. Poder Judicial de Formosa. (2020). Disponible en: <http://lenguajeclaroargentina.gob.ar/guia-de-lenguaje-claro-en-el-poder-judicial-de-formosa/>

[Fecha de la consulta: 26/06/2022].

Guía Lenguaje Claro. Comprender y hacernos entender, publicada por el Instituto de Lectura Fácil y Clarity (C. Carretero *et al.*, 2017). Disponible en: <http://lenguajeclaroargentina.gob.ar/lenguaje-claro-comprender-y-hacernos-entender/> [Fecha de la consulta: 26/06/2022].

Guía Cómo escribir con claridad de la Comisión Europea (2015). Disponible en: <http://lenguajeclaroargentina.gob.ar/como-escribir-con-claridad-comision-europea/> [Fecha de la consulta: 26/06/2022].

Página web del Centro de Información Judicial de la Corte Suprema de Justicia de la Nación Argentina para la descarga de sentencias que conforman el corpus lingüístico: <https://www.cij.gov.ar/sentencias.html> [Fecha de la consulta: 26/06/2022].

Protocolo para el uso de lenguaje claro. Cámara Segunda de Apelación en lo Civil y Comercial de La Plata y Centro de Modernización en Litigación y Gestión Judicial no penal de la Facultad de Ciencias Jurídicas y Sociales de la UNLP (2020). Disponible en <http://lenguajeclaroargentina.gob.ar/protocolo-de-lenguaje-claro-pba/> [Fecha de la consulta: 26/06/2022].

8. ANEXO

A continuación, se comparte el enlace para acceder a la maqueta de evaluador automático de legibilidad de sentencias judiciales argentinas y a las 13 sentencias anonimizadas que conforman el corpus de evaluación: https://drive.google.com/drive/folders/1AINgalGQCrb1wqh14rxdV0_WJA9dz3qE?usp=sharing