



UNIVERSITAT DE
BARCELONA

*Accesibilidad comunicativa en el entorno clínico:
una propuesta inclusiva desde el lenguaje claro*

Trabajo Final de Máster de Humanidades Digitales

Junio 2025

Autora: Chen Ma

Tutora: Mariona Taulé Delor, Montserrat Nofre

Curso 2024 – 2025



Agradecimientos

El tiempo ha pasado volando y, con ello, el máster llega a su fin. En realidad, mi interés por el estudio del corpus y del discurso surgió mucho antes de iniciar mis estudios en Filología Hispánica, cuando apenas conocía el alcance real de estas áreas. Nunca imaginé que un día podría escribir una tesis centrada precisamente en estos temas, así que este trabajo representa también la realización de un sueño académico personal. Por ello, quiero aprovechar estas líneas para expresar mi más sincero agradecimiento.

Quiero agradecer profundamente a Mariona Taulé, mi directora y también profesora de la asignatura de corpus, por haber influido decisivamente en mi elección de tema. Su manera de enseñar y su pasión por la lingüística computacional despertaron en mí un gran interés por el estudio de corpus y del discurso. Además, le estoy muy agradecida por su acompañamiento, disponibilidad y orientación constante durante todo el proceso de elaboración de este trabajo.

Mi más sincero agradecimiento también a Montse Nofre, mi tutora, sin cuya ayuda esta tesis no habría sido posible. Su rigor académico, su generosidad en compartir conocimientos sobre lenguaje médico y su implicación en cada detalle del trabajo me han enseñado muchísimo. Gracias por su paciencia, sus comentarios precisos y su constante apoyo.

Gracias a Alex, mi pareja, por tu apoyo incondicional a lo largo de este proceso. Gracias por estar siempre disponible para ayudarme, animarme y darme la fuerza necesaria en los momentos difíciles.

Por último, gracias a mi familia y a mis amigos, cuya compañía y apoyo emocional han sido fundamentales para completar esta etapa. Me siento muy afortunada de tenerlos en mi vida.

A todos, mi mayor agradecimiento.

Resumen

La comprensión efectiva de los textos médicos por parte de los pacientes es fundamental para garantizar una atención sanitaria segura, equitativa y participativa. Sin embargo, muchos documentos clínicos presentan un nivel de complejidad que dificulta su comprensión. Este Trabajo Fin de Máster aborda esta problemática mediante la aplicación de un proceso de simplificación textual a documentos médicos reales, con el fin de mejorar su legibilidad y adecuación al público general.

Se trabajó con un corpus formado por planes de seguimiento y planes terapéuticos, sobre los cuales se aplicaron dos fases de simplificación manual basadas en criterios lingüísticos definidos y apoyadas por herramientas como Legible y Voyant Tools. La evaluación incluyó índices de legibilidad, como el índice de Fernández Huerta (1959) e INFLESZ (2008), métricas complementarias (edad mínima de comprensión, nivel educativo requerido y tiempo de lectura), así como análisis léxico.

Los resultados muestran una mejora significativa en la accesibilidad de los textos: se redujeron tanto la edad estimada para su comprensión como el nivel de escolarización necesario, sin pérdida de contenido médico. A partir de estos hallazgos, se propone un conjunto de mejoras aplicables al entorno sanitario, como la creación de guías de redacción clara, el desarrollo de herramientas automáticas de apoyo a la escritura y la validación de los textos con pacientes reales.

Abstract

Effective comprehension of medical texts by patients is essential for safe, equitable, and participatory healthcare. However, many clinical documents are written in language that exceeds the reading capabilities of the general population. This Master's Thesis addresses this issue through a process of textual simplification applied to real medical documents, aiming to improve their readability and patient accessibility.

The study focused on a corpus composed of follow-up plans and therapeutic plans. Two phases of manual simplification were applied, based on defined linguistic criteria and supported by tools such as Legible and Voyant Tools. The evaluation included readability indices (Fernández Huerta and INFLESZ), complementary metrics (minimum comprehension age, required education level, and estimated reading time), and lexical analysis.

The results show a significant improvement in text accessibility: both the estimated comprehension age and the required schooling level decreased notably, while the medical content remained intact. Based on these findings, the study proposes a set of practical improvements for healthcare settings, including the creation of plain language writing guidelines, the development of automated support tools, and validation of simplified texts with real patients.

Índice de contenido

1. Introducción	7
2. Marco teórico	8
2.1. Terminología médica	8
2.2. Lenguaje Claro	9
2.3. Simplificación lingüística	9
2.4. Legibilidad en textos médicos	10
2.5. Herramientas digitales para el análisis textual	10
3. Objetivo	12
4. Estado del arte	12
5. Metodología	15
5.1. Recopilación del corpus	15
5.2. Herramientas de análisis	16
5.2.1. Legible	17
5.2.2. Voyant Tools	19
5.3. Criterios de simplificación	20
5.3.1 Simplificación sintáctica	20
5.3.2 Simplificación léxica	22
5.3.3 Simplificación de los números	23
5.4. Procedimiento	24
6. Análisis y resultados	26
6.1. Análisis en subgrupo “Plan de seguimiento”	26
6.1.1 Análisis de índices	27
6.1.2 Análisis de léxico	30
6.1.3 Análisis de métricas	32
6.2. Análisis en subgrupo “Plan terapéutico”	35
6.2.1. Análisis de índices	36
6.2.2. Análisis de léxico	38
6.2.3. Análisis de métricas	41
7. Propuesta	43
8. Conclusiones	46
9. Bibliografía y referencias	48
10. Anexos	50
10.1. Anexo 1. Documento de trabajo	50

Índice de figuras

Figura 1. Fórmula para obtener la legibilidad mediante el índice Fernández-Huerta	17
Figura 2. Fórmula para obtener la legibilidad mediante el índice INFLESZ.....	18
Figura 3. Fórmula para obtener la edad mínima de comprensión	19
Figura 4. Fórmula para obtener la legibilidad mediante el índice INFLESZ.....	19
Figura 5. Matriz de transición de niveles de dificultad sobre el índice Fernández Huerta, primera simplificación en subgrupo “Plan de seguimiento”	27
Figura 6. Matriz de transición de niveles de dificultad sobre el índice Fernández Huerta, segunda simplificación en subgrupo “Plan de seguimiento”.	28
Figura 7. Matriz de transición de niveles de dificultad sobre el índice INFLESZ, primera simplificación en subgrupo “Plan de seguimiento”	29
Figura 8. Matriz de transición de niveles de dificultad sobre el índice INFLESZ, segunda simplificación en subgrupo “Plan de seguimiento”	30
Figura 9. Frecuencia de palabras más utilizadas en subgrupo original “Plan de seguimiento” junto a su nube de palabras.	31
Figura 10. Frecuencia de palabras más utilizadas en subgrupo primera y segunda simplificación “Plan de seguimiento” junto a sus nubes de palabras respectivamente	31
Figura 11. Evolución de la proporción de palabras por tipos tras las simplificaciones	32
Figura 12. Evolución de la edad mínima de comprensión tras las simplificaciones.....	33
Figura 13. Evolución de la edad mínima de comprensión tras las simplificaciones en cada texto	33
Figura 14. Evolución del nivel de escolarización necesario tras las simplificaciones	34
Figura 15. Evolución del nivel de escolarización necesaria tras las simplificaciones en cada texto.....	34
Figura 16. Evolución del tiempo estimado de lectura tras las simplificaciones.....	35
Figura 17. Evolución del tiempo estimado de lectura tras las simplificaciones en cada texto.	35
Figura 18. Matriz de transición de niveles de dificultad sobre el índice Fernández Huerta, primera simplificación en subgrupo “Plan terapéutico”	36
Figura 19. Matriz de transición de niveles de dificultad sobre el índice Fernández Huerta, segunda simplificación en subgrupo “Plan terapéutico”	37
Figura 20. Matriz de transición de niveles de dificultad sobre el índice INFLESZ, primera simplificación en subgrupo “Plan terapéutico”	37
Figura 21. Matriz de transición de niveles de dificultad sobre el índice INFLESZ, segunda simplificación en subgrupo “Plan terapéutico”	38

Figura 22. Frecuencia de palabras más utilizadas en subgrupo original “Plan de seguimiento” junto a su nube de palabras	39
Figura 23. Frecuencia de palabras más utilizadas en subgrupo primera simplificación “Plan de seguimiento” junto a su nube de palabras	39
Figura 24. Frecuencia de palabras más utilizadas en subgrupo segunda simplificación “Plan de seguimiento” junto a su nube de palabras	40
Figura 25. Evolución de la proporción de palabras por tipos tras las simplificaciones	40
Figura 26. Evolución de la edad mínima de comprensión tras las simplificaciones.....	41
Figura 27. Evolución de la edad mínima de comprensión tras las simplificaciones en cada texto	41
Figura 28. Evolución del nivel de escolarización necesario tras las simplificaciones	42
Figura 29. Evolución del nivel de escolarización necesaria tras las simplificaciones en cada texto.....	42
Figura 30. Evolución del tiempo estimado de lectura tras las simplificaciones.....	43
Figura 31. Evolución del tiempo estimado de lectura tras las simplificaciones en cada texto.	43
Figura 32. Tabla recopilatoria de textos originales y simplificados, subconjunto de seguimiento médico	50
Figura 33. Tabla recopilatoria de textos originales y simplificados, subconjunto de plan terapéutico	51
Figura 34. Tabla recopilatoria de resultados índices y métricas, subconjunto de seguimiento médico	51
Figura 35. Tabla recopilatoria de resultados índices y métricas, subconjunto de plan terapéutico	52
Figura 36. Tabla comparativa de resultados, subconjunto de seguimiento médico	52
Figura 37. Tabla comparativa de resultados, subconjunto de plan terapéutico	52

Índice de tablas

Tabla 1. Interpretación del índice Fernández Huerta.....	17
Tabla 2. Interpretación del índice INFLESZ	18

1. Introducción

En los entornos clínicos la comunicación escrita entre los profesionales de la salud y los pacientes desempeña un papel fundamental: transmite información precisa sobre diagnósticos, procedimientos terapéuticos, nombres de medicamentos y consentimiento informado. Sin embargo, numerosos estudios han demostrado que estos textos suelen tener un alto grado de tecnicismo, estructuras sintácticas complejas y terminología especializada, que los hace difíciles de comprender, especialmente para personas sin formación médica o pacientes con discapacidades cognitivas, sensoriales o intelectuales.

Esta dificultad de comprensión impacta directamente en la comunicación eficiente, la calidad del cuidado y el derecho a la información del paciente. Cuando una persona no entiende la información que se le transmite, no puede tomar decisiones informadas sobre su salud ni participar activamente en el proceso de atención. En este sentido, la accesibilidad comunicativa es esencial para garantizar un cuidado equitativo e inclusivo.

Ante esta realidad, la importancia del enfoque de lenguaje claro se vuelve cada vez más evidente. A nivel internacional, el movimiento por el lenguaje claro ha promovido una serie de estrategias lingüísticas y discursivas que buscan redactar textos comprensibles, fáciles de entender y útiles para el lector. Estas estrategias incluyen recomendaciones como el uso de oraciones cortas, evitar estructuras pasivas, reducir la terminología técnica innecesaria y emplear términos comunes o explicativos. Aunque en ámbitos jurídicos y administrativos ya se han desarrollado métodos sistemáticos de simplificación (como los de Pistola y da Cunha (2024), Montolío y Tascón (2020), entre otros), su aplicación en contextos clínicos sigue siendo limitada, y aún existen pocas propuestas centradas en la simplificación terminológica para personas vulnerables como los mayores de la tercera edad o personas con discapacidad.

En este sentido, el presente trabajo de fin de máster tiene como objetivo promover la accesibilidad comunicativa en el ámbito sanitario mediante la simplificación de la terminología presente en los textos clínicos, con el fin de mejorar la comprensión de dichos textos por parte de aquellos colectivos en situación de vulnerabilidad.

2. Marco teórico

2.1. Terminología médica

La terminología médica es una rama altamente especializada dentro del lenguaje técnico-científico, cuyo propósito fundamental es facilitar una comunicación precisa y eficaz en el ámbito de la atención sanitaria. Tal como señala Jéssica López Hernández (2022), “El lenguaje médico trata de un tipo de discurso que principalmente busca informar y transmitir conceptos.” (p. 39).

El lenguaje médico abarca, en su mayoría, un extenso conjunto de términos especializados, caracterizados por una densa presencia de acrónimos, abreviaturas y unidades fraseológicas complejas, muchos de los cuales tienen su origen en las lenguas clásicas, como el griego y el latín. Esta elevada densidad terminológica resulta funcional en contextos profesionales especializados donde el público objetivo cuenta con el conocimiento y comprende sus significados; sin embargo, representa una barrera significativa para pacientes, familiares y otros agentes no expertos involucrados en el proceso asistencial.

Uno de los principales problemas asociados al lenguaje médico en contextos clínicos es su escasa transparencia semántica. Numerosos términos poseen significados técnicos alejados del uso común o, incluso, inexistentes en la lengua general. En cuanto a las estructuras lingüísticas utilizadas, los informes médicos tienden a adoptar construcciones gramaticales no estándar, como la nominalización (“realización de una intervención”) o la voz pasiva impersonal (“se procederá a realizar”), así como la omisión del sujeto o la ausencia de puntuación. Asimismo, por la necesidad de registrar información con rapidez en un tiempo limitado, la redacción de los informes suele tener una alta incidencia de errores lingüísticos, dado que los profesionales redactan los textos de forma apresurada y sin un proceso de revisión posterior. También es habitual el uso de términos abreviados, siglas y vocablos de origen extranjero, adoptados por su utilidad práctica en contextos clínicos, aunque su interpretación puede resultar confusa para los lectores. Todo ello contribuye a dificultar la comprensión del mensaje por parte del receptor y plantea desafíos de comunicación en el ámbito de la salud.

Esta barrera lingüística afecta especialmente a personas vulnerables, quienes pueden encontrar serias dificultades para entender informes médicos, consentimientos informados, prospectos farmacológicos u otros documentos clínicos. En estos casos, la terminología se convierte en una forma de exclusión y puede incluso limitar el ejercicio del derecho a la salud de la ciudadanía. Por ello, es necesario repensar el uso de la terminología médica desde una perspectiva inclusiva, mejorando la comprensibilidad del lenguaje; en este contexto, surge la corriente denominada *Lenguaje Claro* (o *Plain Language*, en inglés), cuyas primeras referencias en español datan en los inicios del 2000 como Daniel Cassany (2005) interviniendo en la Conferencia Internacional de Plain Language de 2005.

2.2. Lenguaje Claro

El lenguaje claro es un enfoque de redacción que busca garantizar la comprensión efectiva de los textos por parte de sus destinatarios convirtiéndolos en más comprensibles, estructurados y accesibles, independientemente de su nivel educativo o experiencia. Según la definición propuesta por la Organización Internacional de Normalización (2023), un texto se considera claro cuando su redacción, estructura y diseño son tan transparentes que los lectores pueden encontrar lo que necesitan, entender lo que encuentran y usar esa información.

Desde un punto de vista práctico, el lenguaje claro aboga por el uso de un estilo directo y la eliminación de tecnicismos innecesarios. Se caracteriza por el uso preferente de oraciones cortas, la voz activa y una organización lógica del contenido. Su principal objetivo es garantizar que los lectores puedan encontrar la información que necesitan, comprenderla con facilidad y utilizarla de manera eficaz.

En el contexto hispano, este enfoque está siendo adoptado cada vez más por instituciones públicas en países hispanohablantes, como lo demuestra la creación de la Red Panhispánica de Lenguaje Claro (RAE, 2022), así como la publicación de guías por parte de gobiernos regionales y locales (Ayuntamiento de Madrid, 2017). Asimismo, este enfoque no se limita a una cuestión de estilo, sino que se articula como un derecho fundamental a la comprensión, especialmente cuando la información afecta a derechos o servicios esenciales, como sucede en el ámbito jurídico-administrativo o sanitario. Como subrayan Montolío y Tascón (2020) en su libro sobre El derecho a entender, en España existen multitud de normas legales en las que se hace referencia a la necesidad de que la información que se dirige a la ciudadanía resulte comprensible y accesible.

Desde un punto de vista técnico, las estrategias de lenguaje claro pueden clasificarse, como proponen Pistola y da Cunha (2024), en tres niveles:

- **Discursivo:** Implementando una organización textual, uso de listas y estructura jerárquica.
- **Morfosintáctico:** Evitando el uso de oraciones pasivas, gerundios, nominalizaciones y favoreciendo el uso de frases breves.
- **Léxico:** Fomentando el reemplazo de tecnicismos, eliminación de arcaísmos y un uso de sinónimos más comunes.

2.3. Simplificación lingüística

Diversos autores han descrito la simplificación lingüística como una de las principales vías operativas del lenguaje claro, centrada en eliminar obstáculos terminológicos, sintácticos o discursivos que dificultan la comprensión. Como las dichas estrategias de lenguaje claro, se podría realizar la simplificación desde distintos aspectos, sin pérdida de su contenido esencial. En cuanto al ámbito médico, como bien comenta Jérica López Hernández (2022), la simplificación de informes médicos puede resultar especialmente útil para aquellos pacientes

que tienen acceso a estos desde aplicaciones, como el Portal del Paciente, y que puede ayudar a una mayor comprensión de estos.

Diversas estrategias han sido desarrolladas con este fin, como las guías de redacción en Plain Language, el proyecto Clara-MeD (Campillos-Llanos et al., 2024) para textos médicos en español, o los modelos automáticos propuestos en tareas como CLEF-Simplext (SanJuan et al., 2024). La mayoría de estas estrategias abarcan los tres niveles de estrategias de lenguaje claro, recomendando evitar el uso de oraciones largas, usar voz activa y evitar pasivas innecesarias, y preferir términos concretos frente a abstractos, siendo estos de mayor conocimiento común entre la población. En el plano terminológico, se recomienda expandir siglas, usar sinónimos más frecuentes o proporcionar definiciones breves.

2.4. Legibilidad en textos médicos

El concepto de legibilidad hace alusión al grado en que un texto puede ser leído y comprendido con facilidad. En el ámbito de la documentación sanitaria, este factor resulta especialmente relevante dado el impacto directo que tiene sobre la comprensión y el seguimiento de las indicaciones médicas por parte del paciente.

En español, se han desarrollado diversos índices de legibilidad adaptados lingüísticamente, entre los que destacan el índice Fernández-Huerta (Fernández Huerta, 1959), el cual es una adaptación del Flesch Reading Ease (Flesch, 1948), y el índice Szigriszt-Pazos, especialmente diseñado para textos administrativos y sanitarios (Szigriszt Pazos, 1993). Otros como INFLESZ (Barrio-Cantalejo et al., 2008) han ampliado este enfoque al integrar múltiples métricas y estimar niveles de comprensión en función de la edad o el grado educativo necesario.

No obstante, estas métricas presentan limitaciones importantes. Por ejemplo, muchas se basan exclusivamente en variables superficiales como la longitud de las palabras o de las oraciones, sin considerar aspectos semánticos, pragmáticos o discursivos como bien remarcan Barrio-Cantalejo et al. (2008) creadores del índice INFLESZ. Asimismo, su aplicación automática a textos especializados como los médicos puede generar puntuaciones engañosas si no se acompaña de un análisis cualitativo del contenido.

2.5. Herramientas digitales para el análisis textual

En los estudios de análisis textual aplicados a la lingüística y a la comunicación especializada, las herramientas digitales han cobrado un papel relevante. Estas aplicaciones permiten realizar mediciones cuantitativas sobre corpus extensos, facilitando la evaluación objetiva de aspectos como la frecuencia léxica, la estructura sintáctica, la densidad terminológica o la legibilidad general. El uso de herramientas computacionales en lingüística aplicada, particularmente en proyectos de simplificación textual, se ha consolidado como una metodología eficaz para

garantizar la reproducibilidad, el análisis a gran escala y la validación empírica de intervenciones lingüísticas.

Entre las herramientas empleadas en este tipo de investigaciones destacan programas de análisis de corpus como AntConc¹, desarrollada por Laurence Anthony, que permite analizar frecuencias de palabras, listas de concordancias, patrones de colocación y n-gram. Su funcionalidad es especialmente útil para observar cómo se repiten estructuras léxicas o fórmulas métricas en textos originales frente a textos simplificados.

Otra vía para el análisis automatizado es el uso de instrucciones de Unix a través de la línea de comandos. Esta metodología se basa en comandos como `grep`, `wc`, `cut`, `sort` o `uniq`, que permiten realizar tareas rápidas de conteo, filtrado, segmentación y ordenación de datos textuales. Aunque requiere conocimientos básicos de programación en entorno terminal, estas instrucciones permiten un tratamiento flexible y potente de grandes volúmenes de texto. Su uso es habitual en proyectos de análisis de corpus masivos y constituye una estrategia eficaz para investigadores que buscan precisión en la extracción de patrones lingüísticos sin necesidad de interfaces gráficas reduciendo el tiempo necesario para la obtención de los resultados.

En este contexto amplio de tecnologías de análisis textual, esta investigación ha optado por emplear dos herramientas específicas: Voyant Tools y Legible. Ambas han sido seleccionadas por su accesibilidad, su compatibilidad con el español y su utilidad específica en los objetivos del trabajo para analizar la evolución del contenido léxico y evaluar la legibilidad tras la simplificación.

Voyant Tools² es una plataforma de análisis textual en línea desarrollada por Sinclair y Rockwell en 2016. Se trata de una de las herramientas disponibles en el repositorio europeo Social Sciences and Humanities Open Marketplace³, lo que garantiza su reconocimiento en el ámbito académico y su compatibilidad con estándares de investigación en humanidades digitales. Su interfaz permite cargar textos o corpus completos y visualizar información como frecuencia de palabras, nubes léxicas, distribución temática o relaciones contextuales entre términos. En el caso de este proyecto, Voyant Tools se ha utilizado para analizar el vocabulario predominante en los textos originales y simplificados, identificar tecnicismos frecuentes y observar la evolución léxica a lo largo del corpus.

Por su parte, Legible⁴ es una aplicación web diseñada específicamente para evaluar la legibilidad de textos en español. Ha sido desarrollada por Alejandro Muñoz Fernández mediante lenguaje Python y es de libre acceso. Esta herramienta calcula diversos índices clásicos de legibilidad adaptados al contexto hispanohablante, como Fernández-Huerta (1959);

¹ <https://www.laurenceanthony.net/software/antconc/>

² <http://voyant-tools.org/>

³ <https://marketplace.sshopencloud.eu/>

⁴ <https://legible.com>

Gutiérrez de Polini (1972), Szigriszt-Pazos (1993) o INFLESZ (2008), además de otros indicadores funcionales como el nivel educativo requerido, el tiempo estimado de lectura o la edad mínima para comprender el texto. También identifica problemas de puntuación, palabras no reconocidas y frases potencialmente complejas. Emplearemos esta herramienta para valorar, desde una perspectiva empírica, el grado de accesibilidad alcanzado tras la simplificación de los textos médicos.

3. Objetivo

El presente trabajo de fin de máster tiene como objetivo promover la accesibilidad comunicativa en el ámbito sanitario mediante la simplificación de la terminología presente en textos clínicos, con el fin de mejorar la comprensión de dichos textos por parte de aquellos colectivos en situación de vulnerabilidad. Para ello, se utilizarán textos procedentes de informes clínicos reales, los cuales a diferencia de los casos clínicos que son redactados con fines docentes, científicos o divulgativos y ya han tenido revisiones previas a sus publicaciones, estos son creados específicamente durante la atención real de un paciente, cuyo destinatario suele ser este o personas cercanas de su entorno por lo que no cuentan con un proceso de revisión que certifique una fácil legibilidad de los textos. Este estudio lo realizaremos con el propósito de identificar los principales obstáculos sintácticos y léxicos, buscando simplificarlos y reformularlos según recomendaciones lingüísticas específicas. El proceso se realizará aplicando una simplificación lingüística siguiendo reglas de simplificación identificadas, siguiendo los estándares anteriormente descritos de Lenguaje Claro y comunicación efectiva. Para analizar el impacto de las transformaciones se compararán los textos originales y los textos simplificados mediante herramientas de análisis lingüístico, con el fin de cuantificar y visualizar los resultados obtenidos tras la aplicación de estrategias de lenguaje claro. Además, se propondrá una estrategia de simplificación destinada a ofrecer orientaciones inclusivas a los profesionales de la salud o a los responsables de la elaboración de materiales sanitarios para futuras iniciativas de comunicación en salud.

4. Estado del arte

Con el fin de contextualizar el estado actual de la investigación sobre simplificación de textos médicos en lengua española, se llevó a cabo una revisión exploratoria (scoping review) de la bibliografía. Esta metodología permite mapear de forma estructurada las publicaciones relevantes, sin limitarse a una pregunta de investigación cerrada.

La búsqueda se realizó en bases de datos como Scopus, Google Scholar y PubMed, utilizando palabras clave como “text simplification”, “legibilidad médica”, “lenguaje médico comprensible”. Se incluyeron artículos publicados entre 2010 y 2024, en español, que abordan la simplificación lingüística, la legibilidad de textos médicos o la adaptación de lenguaje para pacientes.

Se seleccionaron 6 artículos, los estudios abarcaban desde propuestas de modelos automáticos de simplificación (Saggion, 2024), simplificación sistemática de lenguaje administrativo (Pistola y da Cunha, 2024), hasta recomendaciones terminológicas para textos dirigidos a pacientes (Bartolomé et al., 2023). La mayoría de los trabajos coinciden en señalar que el lenguaje técnico y la complejidad sintáctica representan barreras significativas para la comprensión de los textos médicos por parte de pacientes con baja alfabetización en salud.

Aunque los esfuerzos por clarificar textos administrativos han sido más abundantes que en el ámbito médico, ofrecen una base metodológica útil para su transferencia disciplinar. El estudio de Pistola y da Cunha (2024) representa un hito en esta línea, al aplicar 19 recomendaciones lingüísticas explícitas para reformular 384 unidades fraseológicas extraídas de documentos administrativos reales. Estas recomendaciones abarcan niveles discursivos, morfosintácticos y léxicos, y su eficacia fue evaluada mediante análisis cuantitativo de las intervenciones. El corpus resultante fue integrado en la herramienta arText⁵, demostrando que la simplificación no solo mejora la comprensión, sino que también puede implementarse de forma automatizada en contextos institucionales.

Dentro del entorno clínico, los textos presentan obstáculos adicionales. Como muestra la tesis de Jélica López Hernández (2022), los informes médicos en español contienen errores sistemáticos debidos a la presión asistencial, el uso de tecnicismos, abreviaturas, y estructuras elípticas o agramaticales. Su estudio, basado en un corpus de más de dos millones de palabras de informes reales (urgencias, UCI, cirugía y psiquiatría), revela que los errores más frecuentes están vinculados a omisiones, ambigüedades y usos no normativos del lenguaje médico. Esta investigación resalta la necesidad no solo de corregir errores, sino de repensar la estructura comunicativa de los textos médicos para que resulten legibles para públicos no expertos.

Por otro lado, vemos cómo Victoria Rosales (2021) pone énfasis en la importancia de una comunicación eficaz en la relación médico-paciente, considerando esta como un eje central de la práctica clínica. A través de una revisión conceptual, señala que una interacción basada en el respeto, la claridad y la empatía puede mejorar significativamente la adherencia terapéutica y el bienestar emocional del paciente. Destaca la obligación ética de informar de forma comprensible sobre diagnósticos, tratamientos y pronósticos, lo cual implica necesariamente el uso de un lenguaje adaptado a las capacidades del paciente.

También encontramos diferentes trabajos como el de Araujo (2017) quien analiza críticamente el uso generalizado de epónimos y abreviaciones en el lenguaje médico, señalando su origen histórico y su función como estrategias de economía comunicativa. Sin embargo, destaca que estos recursos, aunque útiles dentro de contextos especializados, pueden dificultar la comprensión entre profesionales de diferentes niveles asistenciales o disciplinas, y mucho más para pacientes con bajo nivel de alfabetización en salud. El autor alerta sobre la ambigüedad

⁵ <https://sistema-artext.com/es/>

y falta de equivalencias internacionales en estos términos, lo cual puede derivar en interpretaciones erróneas con consecuencias clínicas graves.

El proyecto CLARA-MeD (Bartolomé et al., 2023) representa uno de los desarrollos más relevantes en simplificación médica en español. Construido a través del análisis de más de 1.000 textos de ensayos clínicos del repositorio EudraCT⁶, con versiones originales y versiones simplificadas, tanto a nivel sintáctico como léxico. Este trabajo sistematiza 13 operaciones de simplificación sintáctica (como evitar la voz pasiva, dividir frases largas, sustituir nominalizaciones por verbos) y 7 operaciones léxicas (como la expansión de siglas, el uso de sinónimos comunes, o la glosa de tecnicismos). Además, propone una taxonomía operativa para evaluar la dificultad de cada intervención y liberar conjuntos de datos útiles para el entrenamiento de sistemas automáticos de simplificación.

Adentrándonos en proyectos más técnicos en cuanto a tecnología aplicada, Horacio Saggion (2024) propone un modelo automatizado de simplificación textual que combina inteligencia artificial y procesamiento de lenguaje natural (PLN) con el objetivo de mejorar la legibilidad de textos públicos y médicos. El modelo se basa principalmente en recursos lingüísticos — como diccionarios de sustituciones léxicas— y técnicas de aprendizaje automático, permitiendo realizar simplificaciones léxicas y sintácticas sin alterar el significado original. Saggion enfatiza el principio de “priorizar la comprensibilidad” y defiende la incorporación de módulos que adapten la simplificación a las capacidades del lector, como niveles ajustables para distintos perfiles (niños, personas mayores o con dificultades cognitivas). Además, se presentan resultados experimentales de la herramienta sobre diversos corpus, mostrando su potencial para mejorar tanto la velocidad como la comprensión lectora, abogando por una mayor implantación de este tipo de tecnologías en ámbitos médicos, administrativos y educativos.

En conjunto, tras la revisión hemos identificado una creciente preocupación por la accesibilidad lingüística de los textos especializados en español, especialmente en los sectores médico y administrativo. La bibliografía revisada señala que las barreras terminológicas, la complejidad sintáctica y la escasa adaptación al lector general siguen siendo obstáculos significativos para la comprensión. Al mismo tiempo, se constata un desarrollo progresivo de herramientas automáticas, directrices de estilo y métricas de legibilidad que apuntan hacia una mayor sistematización de la simplificación textual como estrategia de salud pública y de inclusión comunicativa. Esta revisión establece así un marco teórico y metodológico robusto sobre el cual se justifica la presente investigación.

⁶ <https://eudract.ema.europa.eu/>

5. Metodología

5.1. Recopilación del corpus

El presente trabajo tiene como objetivo principal el análisis de la simplificación lingüística en textos médicos, con especial atención a su impacto en la legibilidad. Por ello, se ha empleado el corpus textual construido a partir de fuentes documentales reales pertenecientes al ámbito clínico a través del corpus CARMEN-I (Corpus of Anonymized Records for Medical information Extraction) (Farre Maduell et al., 2023; 2024).

CARMEN-I se ha desarrollado en el contexto del Plan de Impulso de las Tecnologías del Lenguaje (Plan TL) promovido por la Secretaría de Estado de Digitalización e Inteligencia Artificial (SEDIA) y del convenio entre el Barcelona Supercomputing Center-Centro Nacional de Supercomputación (BSC-CNS) y el Hospital Clínic de Barcelona. Está compuesto por 2.000 registros clínicos anonimizados recopilados entre marzo de 2020 y marzo de 2022 pertenecientes a este centro clínico. Así, la mayoría de los textos corresponden a pacientes con COVID-19 y diversas comorbilidades, como insuficiencia renal, enfermedades cardiovasculares, cáncer e inmunosupresión.

Además, el conjunto de datos ha sido tratado con un proceso de anonimización exhaustivo, siguiendo protocolos validados por expertos clínicos y lingüistas para así reemplazar los datos sensibles por equivalentes sintácticos con el fin de asegurar la confidencialidad de los datos sin perder su mensaje original.

Dado que el corpus CARMEN-I es extenso y abarca múltiples subgéneros dentro de la documentación médica, se ha optado por una selección parcial de sus contenidos. En concreto, se han elegido fragmentos clasificados bajo las categorías de seguimiento clínico y plan terapéutico. Esta decisión responde a varios criterios metodológicos, siendo en primer lugar que ambos tipos textuales presentan estructuras representativas del lenguaje médico en su dimensión práctica, ya que están destinados a orientar la conducta del paciente tras la hospitalización o durante un tratamiento, también porque suelen incluir instrucciones, prescripciones y recomendaciones dirigidas al paciente, lo que los convierte en una muestra relevante para el análisis de inteligibilidad y adecuación al receptor y por último se tuvo en cuenta la extensión y formato ya que permiten una simplificación viable sin perder el contenido informativo esencial, facilitando así la comparación entre versiones originales y adaptadas.

El proceso de extracción de los textos se ha realizado de manera manual, identificado los fragmentos pertinentes dentro de la base de datos del corpus CARMEN-I, y se han transcrito a un archivo editable. Se han excluido los documentos demasiado técnicos, dirigidos exclusivamente a profesionales sanitarios, o aquellos cuyo contenido no presentaba potencial para ser reescrito en un lenguaje más accesible como por ejemplo aquellos textos con longitud muy reducida, además de aquellos textos en catalán que representan el 2 % del corpus. En

total, se han seleccionado 65 textos, 37 de los cuales pertenecen al subgrupo de seguimiento clínico y 28 al subgrupo de plan terapéutico. Cada uno de ellos representa un caso diferente y redactado por profesionales médicos con distintos estilos. Esto ha permitido asegurar cierta heterogeneidad textual y tipológica dentro de un mismo subgénero, garantizando un enriquecimiento necesario para el análisis posterior.

Una vez seleccionados los textos, se procedió a la elaboración de sus respectivas versiones simplificadas. Este proceso fue realizado de forma manual, con base en principios de adecuación textual, reducción de tecnicismos, eliminación de estructuras sintácticas complejas, y sustitución de construcciones impersonales por formas activas o más cercanas al lenguaje común. No se recurrió a modelos automáticos de simplificación, con el fin de mantener un mayor control sobre los cambios lingüísticos aplicados. Para cada texto, se generó una ficha que contiene: el texto original, su versión simplificada y una breve descripción de las estrategias aplicadas.

Los textos originales y sus versiones adaptadas fueron organizados en una hoja de cálculo de Excel con el objetivo de sistematizar el análisis, así contamos con dos bloques diferenciados uno para cada subgrupo que analizaremos. Dentro de estos bloques recogemos información como el nombre del documento que implícitamente define características de este ya que sigue un formato homogéneo incluyendo el tipo de reporte. Por ejemplo, uno de los nombres de archivos del documento es "CARMEN-I_IA_ANTECEDENTES_2.txt", del cual podemos conocer que pertenece a la clasificación de informes de alta (IA), y pertenece a la tipología de antecedentes, al igual que su numeración dentro del subgrupo. También recogemos en esta base de datos los textos originales y simplificados además de un listado de transformaciones que han sido realizadas a cada uno de ellos, siguiendo los criterios de simplificación que se describen en el apartado 5.3. Además, este documento nos sirve para recopilar los resultados de los índices y las métricas obtenidas en cada simplificación, pudiendo tener una visión global y detallada de los resultados.

Cabe señalar que, si bien el corpus no pretende ser exhaustivo ni estadísticamente representativo de todo el lenguaje médico escrito en español, sí permite observar patrones recurrentes en la producción textual médica dirigida al paciente. En particular, se observan fenómenos como la sobrecarga informativa, el uso excesivo de tecnicismos, la falta de organización discursiva y la escasa personalización del mensaje. Todos estos factores influyen negativamente en la legibilidad y comprensión por parte de los destinatarios no expertos.

5.2. Herramientas de análisis

Con el fin de evaluar de forma cuantitativa el impacto de la simplificación en los textos médicos seleccionados, se han utilizado dos herramientas digitales complementarias: Legible.es y Voyant Tools. Ambas permiten realizar un análisis textual automatizado, aunque con enfoques diferentes. Mientras que Legible.es se centra en la legibilidad de los textos en

español, Voyant Tools está orientado al análisis de frecuencia léxica y visualización de patrones lingüísticos en corpus textuales amplios. A continuación, se detallan las características, funcionamiento y criterios de selección de los indicadores utilizados en cada herramienta.

5.2.1. Legible

Legible es una herramienta en línea de análisis automático de textos en español, desarrollada para calcular distintos índices de legibilidad clásicos y adaptados a la lengua castellana. Estos índices cuantifican el grado de dificultad de un texto para el lector promedio, tomando en cuenta factores como la longitud de las palabras, la complejidad de las oraciones y la densidad léxica.

En este estudio, se han seleccionado los siguientes indicadores ofrecidos por Legible, por su validez empírica en contextos hispanohablantes:

Índice de Fernández-Huerta

Fue creada por José Fernández-Huerta (1958), adaptando el índice Flesch (1956) al español. Este índice mide la legibilidad o “lecturabilidad” de un texto, siendo cuán de fácil o difícil es de entender considerando el número de sílabas por palabra y de palabras por oración. Se rige mediante la siguiente fórmula:

$$\text{Lecturabilidad} = 206.84 - 0.60P - 1.02F$$

P = promedio de sílabas por palabra

F = promedio de palabras por frase

Figura 1. Fórmula para obtener la legibilidad mediante el índice Fernández-Huerta

Así, cuanto más alto sea el valor mayor legibilidad contendrá el texto analizado, los resultados del índice son interpretados siguiendo los siguientes rangos:

Lecturabilidad	Nivel
90 - 100	Muy fácil
80 - 90	Fácil
70 - 80	Algo fácil
60 - 70	Normal
50 - 60	Algo difícil
30 - 50	Difícil
0 - 30	Muy difícil

Tabla 1. Interpretación del índice Fernández Huerta

Índice de Szigriszt-Pazos (INFLESZ)

Esta fórmula es ampliamente utilizada en textos sanitarios en español permitiendo calcular la legibilidad considerando también la extensión silábica y la complejidad estructural. Fue creada por Inés Marçia Barrio-Cantalejo (2008) a partir de la fórmula de Perspicuidad de Szigriszt (1993), que a su vez es una adaptación de la fórmula Flesch original. Su fórmula es:

$$INFLESZ = 206.835 - 62.3 \left(\frac{S}{P} - \frac{P}{F} \right)$$

S = número total de sílabas

P = número total de palabras

F = número total de frases

Figura 2. Fórmula para obtener la legibilidad mediante el índice INFLESZ

Al igual que el anterior índice, cuando mayor sea su valor mayor es la legibilidad del texto, los resultados del índice son interpretados siguiendo los siguientes rangos:

Lectorabilidad	Nivel
80 - 100	Muy fácil
65 - 80	Bastante fácil
55 - 65	Normal
40 - 55	Algo difícil
0 - 40	Muy difícil

Tabla 2. Interpretación del índice INFLESZ

Además de los índices anteriores, Legible proporciona también una serie de *parámetros auxiliares* que permiten contextualizar mejor los resultados cuantitativos:

- **Palabras no reconocidas:** La herramienta permite obtener un listado con las palabras no registradas en la base de datos SUBTLEX-ESP, revisando 94.337 palabras únicas del alfabeto castellano, así se recogen la presencia de tecnicismos, neologismos o errores tipográficos.
- **Signos de puntuación:** Listado de signos de puntuación empleados junto a su frecuencia de uso en el texto, siendo muy útil para valorar la densidad estructural del texto.
- **Edad mínima de comprensión:** Legible también nos proporciona una estimación basada en la fórmula propuesta por José Antonio García López (1999) que indica a partir de qué edad un lector promedio podría entender el texto. Se rige bajo la siguiente fórmula, expresada en años:

$$\text{Edad mín. comprensión} = 0,2495 P + 6,4763 S - 7,1395$$

P = número de palabras por frase

S = número de sílabas por palabra

Figura 3. Fórmula para obtener la edad mínima de comprensión

- **Nivel educativo mínimo (años):** Estimación según la fórmula de Crawford (Alan Crawford, 1989), la cual nos permite estimar cuántos años de escolaridad serían necesarios para comprender el texto, facilitando el análisis en relación con el público destinatario. La fórmula empleada es la siguiente:

$$\text{Años escolarización} = -0,205 OP + 0,049 SP - 3,407$$

OP = número de oraciones por cien palabras

SP = número de sílabas por cien palabras

Figura 4. Fórmula para obtener la legibilidad mediante el índice INFLESZ

- **Tiempo estimado de lectura (minutos):** Estimación en base a una velocidad promedio de lectura. Sirve para valorar la carga informativa del texto de forma temporal.

5.2.2. Voyant Tools

Voyant Tools es una plataforma de análisis textual desarrollada por Stéfán Sinclair y Geoffrey Rockwell en 2016. Su funcionalidad está orientada al procesamiento de corpus extensos mediante análisis estadístico de texto y visualización interactiva.

En este trabajo, se ha utilizado Voyant para analizar el corpus completo, tanto en su versión original como simplificada, para obtener las siguientes mediciones:

- **Frecuencia léxica:** La herramienta nos permite identificar las palabras más frecuentes de cada conjunto de textos. De este modo, si identificamos un descenso en la frecuencia de tecnicismos o palabras poco comunes tras la simplificación puede indicar una mejora en la accesibilidad léxica.
- **Nube de palabras (WordCloud):** Visualización gráfica del vocabulario predominante. Se emplea para detectar visualmente si los textos simplificados presentan un vocabulario más básico y centrado en verbos o sustantivos comunes.
- **Longitud media de palabra:** Ayuda a determinar si las versiones simplificadas tienden a reducir el uso de palabras largas o complejas, lo que suele correlacionarse con una mayor comprensibilidad.

- **Distribución de términos clave:** Esta función permite observar cómo se distribuyen las palabras más frecuentes a lo largo del corpus. Sirve para identificar la coherencia temática y consistencia estilística entre documentos.

Ambas herramientas, aunque distintas en enfoque, ofrecen métricas complementarias que permiten valorar de manera integrada la transformación lingüística efectuada sobre los textos médicos. Mientras Legible se centra en la experiencia de comprensión desde una perspectiva cognitiva y a nivel específico de cada texto, Voyant Tools permite observar cómo la reescritura afecta la densidad, el registro y la estructura semántica de todo el conjunto.

5.3. Criterios de simplificación

Con el objetivo de transformar los textos médicos originales en versiones más comprensibles para un público general no especializado, en este trabajo se han aplicado criterios sistemáticos de simplificación lingüística. Estas reglas se basan principalmente en las recomendaciones a nivel morfosintáctico y léxico del manual de estilo Clara-MeD para comunicación médica, así como en adaptaciones propias para el manejo de datos numéricos, extraídas de guías de redacción accesible y experiencias previas de proyectos similares. Las reglas se agrupan en tres niveles principales: simplificación sintáctica, léxica y numérica.

5.3.1 Simplificación sintáctica

La estructura gramatical de los textos originales fue modificada con el fin de facilitar la lectura lineal, mejorar la cohesión y reducir la carga cognitiva del lector. Las principales operaciones fueron:

- I. **Uso de verbos en lugar de nominalizaciones:** Se transformaron estructuras sustantivas en verbales. Esta modificación permite una construcción más directa y menos densa, favoreciendo la claridad. Por ejemplo,

“la vacunación” → “ponerse la vacuna”

- II. **Conversión de voz pasiva a activa:** La voz activa resulta más intuitiva para el lector medio y mejora la comprensión inmediata.

“Deberá ser controlado por...” → “... controlará”

- III. **Supresión de aposiciones no restrictivas:** Las frases apositivas con función explicativa no esencial se eliminaron, a menos que aportaran información clínica necesaria y relevante para el mensaje.

“Trinomia (AAS/Ramipril/Atorvastatina)” → “Seguir tomando Trinomia”

(Se omiten los componentes detallados (AAS/Ramipril/Atorvastatina) porque Trinomia ya los incluye, así que no se especifica más)

IV. **Control del uso de conjunciones:** Se evitó la coordinación excesiva de ideas en una misma oración. Se prefirieron estructuras más breves, con una sola idea central por enunciado.

“Seguir en aislamiento domiciliario hasta nuevas indicaciones por parte de Salud Laboral” → “Seguir aislado en casa hasta que Salud Laboral diga lo contrario”

(Se sustituyó una conjunción subordinada formal por una más natural y directa)

V. **Eliminación de redundancias:** Se suprimieron repeticiones innecesarias que no aportaban nuevo contenido informativo.

“Medicamentos habituales: [...] Omeprazol 20 mg: un comprimido al día por la mañana” → “Omeprazol 20 mg: tomar un comprimido por la mañana”

(Se elimina la repetición de “al día” cuando ya se da contexto de horario)

VI. **Reformulación de negaciones ambiguas:** Las expresiones negativas que podían generar confusión se transformaron en afirmaciones claras.

“No reincorporarse a trabajar hasta el 27 de julio” → “No debe volver a trabajar hasta el 27 de julio”

(Se transforma una negación ambigua en una frase afirmativa y directa con un sujeto claro lo que facilita su comprensión por parte del paciente)

VII. **Clarificación de correferencias:** Se resolvieron ambigüedades anafóricas (como “ellos”, “lo”, “que”), repitiendo el sustantivo cuando fue necesario para asegurar la referencia correcta.

“Debe consultar con Salud Laboral en caso de presentar de nuevo sintomatología compatible con COVID-19” → “Si vuelves a tener síntomas parecidos al COVID-19, contacta con Salud Laboral”

(Se aclara que el paciente es el que debe realizar la acción)

VIII. **Evitar la reducción excesiva:** Aunque se aplicó un criterio de brevedad, se cuidó que las modificaciones no alteraran el contenido clínico esencial.

“Oxígeno a demanda para mantener SatO₂ >90%” → “Oxígeno: a demanda, mantener saturación mayor a 90%”

(Se evita reducir todo a términos técnicos sin contexto (se añade “Oxígeno:” como encabezado explicativo))

IX. **Corrección de errores gramaticales y puntuación:** Se revisaron posibles fallos en el uso de preposiciones, concordancia, orden de palabras y uso excesivo de paréntesis.

“Hoy Dolquine (Hidroxicloroquina) 200mg un comprimido en la cena, y suspender”
→ “Hoy tomar un comprimido de Dolquine 200 mg en la cena y dejar de tomar”

(Se corrige puntuación y se transforma la frase inconclusa a una oración clara)

Además, algunos textos con estructuras excesivamente esquemáticas (por ejemplo, listas técnicas de más de 10 elementos sin conectores) no fueron incluidos en el corpus final por falta de contexto interpretativo.

5.3.2 Simplificación léxica

El vocabulario técnico especializado fue adaptado mediante distintas estrategias para hacerlo más accesible, sin perder precisión conceptual. Se aplicaron los siguientes procedimientos:

- I. **Expansión de abreviaturas y acrónimos:** Debe buscarse una explicitación de términos que hayan sido reducidos.

“MAP” → “Médico de Atención Primaria” o “CAP” → “Centro de Atención Primaria”

(Ambos términos se escriben directamente en su forma desarrollada)

- II. **Añadición de lexemas aclaratorios:** Se incorporaron palabras que ayudan a completar el sentido de los términos.

“EPO” → “Medicación para la anemia (EPO)”

(Se añade “medicación para la anemia” como lexema aclaratorio para contextualizar EPO)

- III. **Eliminación de palabras innecesarias:** Se retiraron elementos léxicos que no afectaban la comprensión del mensaje

“Metformina kern pharma 850 mg comprimidos recubiertos con película efg, 50 comprimidos. 850 MG. Vía oral. Una vez al día, en la cena” → “Metformina 850 mg: un comprimido por la noche”

(Eliminando el término “kern pharma” y “recubiertos con película efg”, ya que son demasiado descriptivos del medicamento, ya se entiende a cuál se hace referencia)

- IV. **Sustitución por hiperónimos:** Se reemplazaron palabras específicas por categorías más generales

“Control por su médico habitual. Seguimiento de especialistas habituales” → “Le harán seguimiento sus médicos habituales”

(Se engloban ambos elementos dentro de un concepto más amplio que puede abarcar ambos)

- V. **Paráfrasis o definición:** Algunos términos se reescribieron usando una breve definición o reformulación más sencilla.

“Control con médico de cabecera” → “Revisión con su médico de siempre”

(Se explica en un lenguaje más coloquial, resultando más fácil de entender)

- VI. **Uso de sinónimos más comunes:** Se buscaron equivalentes léxicos de mayor frecuencia y más conocidos por la población general

“cefalea” → “dolor de cabeza” o “torácico” → “pecho”

(Los términos se reemplazan por sus equivalentes coloquiales)

Estas modificaciones se aplicaron teniendo en cuenta la fidelidad semántica del contenido y evitando ambigüedades.

5.3.3 Simplificación de los números

Dado que la información cuantitativa es frecuente en los textos médicos y puede representar una barrera de comprensión, se aplicaron normas específicas para su presentación:

- I. **Números del 1 al 10 se escriben con letras:** Excepto si están acompañados por unidades de medida.

“1 comprimido cada 8 horas” → “Un comprimido cada ocho horas”

(Sin embargo “5 mg” permanece como está)

- II. **Evitar símbolos técnicos innecesarios:**

“25 mEq 1-0-0” → “una por la mañana”

(Se reemplazó la pauta 1-0-0 a su forma verbal.)

- III. **Reducción de precisión excesiva:** Se resumieron cifras muy exactas si no alteraban la comprensión clínica.

“Tomar 1,5 comprimidos de Enalapril al día” → “Tomar un comprimido y medio de Enalapril al día”

(Se transforma el número decimal en una expresión más natural y entendible)

- IV. **Homogeneidad de formatos:** Se evitó mezclar cifras y letras en una misma frase.

“/24h” → “una vez al día”

(Se unifica el formato “24h”, “/24h”, “24 horas” para todas las pautas en “una vez al día”)

V. **Evitar abreviaciones no comunes:**

“CCEE” → “consultas externas”

(Se traduce completamente la abreviatura técnica sin conservar la sigla, priorizando comprensión)

VI. **Expresar fechas y duraciones en lenguaje natural:**

“hasta el 30/08/2012” → “hasta el 30 de agosto de 2012”

(Se escribe la fecha en formato natural)

VII. **Redondeo razonable:**

“Cita para el 17 de julio a las 08:27 h” → “Tiene cita el 17 de julio a las 8:30 h”

(Se redondea a la media hora más cercana, siendo más fácil de recordar por el paciente)

VIII. **Incluir la unidad siempre que se mencionen cantidades:**

“Hidroferol 266” → “Hidroferol: una cápsula (266 microgramos)”

(Se aclara la dosis en microgramos, no dejando independiente ninguna cantidad)

Estas reglas numéricas facilitan la lectura y evitan errores de interpretación en pacientes con bajo nivel de alfabetización numérica.

En conjunto, la aplicación de estas reglas permitió una transformación sistemática, consistente y controlada de los textos médicos originales. Su implementación se realizó de manera manual, caso por caso, garantizando la comprensión sin comprometer el contenido clínico.

5.4. Procedimiento

Tras la recopilación y simplificación del corpus, el análisis se articuló mediante una estrategia doble, basada en la evaluación automática de la legibilidad y el estudio de patrones lingüísticos, con el objetivo de comprobar de forma sistemática los efectos de las transformaciones aplicadas. El corpus, compuesto por textos médicos reales en español extraídos de los subgrupos “Seguimiento médico” y “Plan terapéutico” de CARMEN-I, fue procesado en su versión original y en su versión simplificada donde se aplicarán todas las reglas anteriormente descritas. Además, se realizará una segunda simplificación para abarcar aquellos textos que contengan características más específicas y requieran de una atención especial. El procedimiento se estructuró en dos fases complementarias que integran tanto análisis cuantitativo como exploración textual: por un lado, la evaluación del grado de comprensibilidad a través de métricas formales con la herramienta Legible; por otro, el estudio de la distribución léxica y la transformación del contenido terminológico mediante la plataforma Voyant Tools.

Para la primera fase, se utilizó Legible como plataforma en línea especializada en la medición de legibilidad para textos en español. Cada fragmento del corpus (en versión original y en versión simplificada) fue introducido de forma individual en el sistema, y se extrajeron los valores generados para todos los índices disponibles. De este modo, cada par de textos se analizó bajo las mismas condiciones, lo que permitió una comparación directa de las puntuaciones obtenidas en los indicadores Fernández-Huerta e INFLESZ, así como otros datos derivados como la edad mínima de comprensión estimada o el nivel educativo requerido. Estos datos se registraron en un archivo de Excel, como se muestra en Anexo 1, estructurado por variables, lo que permitió la elaboración posterior de tablas comparativas y gráficas que reflejan el efecto cuantificable de la simplificación textual.

El valor añadido de esta fase no residió únicamente en la lectura aislada de los índices, sino en su comparación sistemática. Se calculó la variación porcentual de cada índice entre la versión original y la simplificada, observando tendencias comunes y desviaciones individuales. Por ejemplo, se detectó un aumento homogéneo en la puntuación INFLESZ en más del 90 % de los textos, lo que implica una mejora general en la facilidad de lectura. También se compararon métricas menos técnicas, como el tiempo estimado de lectura o el número de palabras no reconocidas, que resultaron útiles para evaluar la presencia de tecnicismos y la economía discursiva.

La segunda fase del procedimiento consistió en el análisis textual de carácter exploratorio, llevado a cabo con la herramienta Voyant Tools. A diferencia de Legible, esta plataforma permite el procesamiento simultáneo de múltiples documentos y genera visualizaciones dinámicas que representan la frecuencia, recurrencia y distribución de los elementos lingüísticos del corpus. Para ello, se crearon dos conjuntos de documentos: uno con todos los textos originales y otro con todas las versiones simplificadas. Cada conjunto fue tratado como un corpus único, permitiendo así la identificación de tendencias globales.

La carga del corpus en la interfaz de Voyant generó una serie de visualizaciones automáticas propias de la herramienta. En primer lugar, se prestó atención al módulo de términos más frecuentes, que permitió observar qué palabras eran centrales en el discurso original y cómo estas se transformaban tras la simplificación. En el corpus original, la presencia de tecnicismos médicos como “ceftriaxona”, “intramuscular”, “posología” o “mg/12h” era notoria; en contraste, el corpus simplificado mostraba un aumento relativo de verbos comunes y expresiones cotidianas como “tomar”, “pastilla”, “cada día” o “por la mañana”. Esta evolución sugiere una transformación del registro especializado hacia un léxico de mayor accesibilidad.

Del mismo modo, se analizó la nube de palabras generada para ambos corpus, que sirvió como representación visual de los términos más frecuentes ponderados por su aparición. Se identificaron patrones que confirmaban la reducción de complejidad léxica, así como la presencia de nuevos elementos introducidos en la simplificación para clarificar o parafrasear los conceptos originales. Complementariamente, el módulo de “contextos” permitió examinar cómo se usaban ciertas palabras en su entorno sintáctico, y si las sustituciones léxicas

mantenían la coherencia con el sentido clínico original. Este paso fue especialmente útil para validar que la simplificación no había afectado negativamente la precisión informativa de los textos.

La integración de los resultados obtenidos a partir de Legible y Voyant permitió triangular la información desde dos niveles: cuantitativo y discursivo. Mientras que los índices de legibilidad ofrecían una medida objetiva del cambio formal, las visualizaciones y frecuencias léxicas brindaban una comprensión cualitativa del tipo de transformaciones efectuadas. Esta combinación metodológica aseguró no solo que los textos eran más fáciles de leer, sino también que el contenido se mantenía funcional y clínicamente relevante.

Finalmente, todos los datos fueron organizados en una tabla recopilatoria que permitía analizar de forma paralela, para cada texto, los cambios métricos (longitud media de oración, número de sílabas por palabra, edad mínima de comprensión) y los cambios léxicos (sustitución de tecnicismos, expansión de siglas, aparición de sinónimos de mayor frecuencia). Este enfoque comparativo riguroso constituye la base sobre la cual se desarrollan los resultados y discusiones presentadas en los capítulos posteriores.

6. Análisis y resultados

A continuación, realizaremos el análisis de los resultados obtenidos, buscando correlaciones con tratamientos realizados sobre el corpus original. Para ello, se realizan dos capas de simplificación, una primera donde se emplean los criterios de simplificación comentados anteriormente y, tras ella, analizaremos también los datos sobre una segunda transformación, enfocada a estudiar casos aislados que no hayan sido mejorados por tener una naturaleza característica diferente al resto de textos.

6.1. Análisis en subgrupo “Plan de seguimiento”

El subgrupo del corpus relacionado perteneciente a planes de seguimiento médico contiene textos de naturaleza clínica cuyo objetivo principal es dar instrucciones tras el alta para garantizar la continuidad del cuidado del paciente. Estos textos forman parte del proceso asistencial hospitalario, específicamente del momento de transición entre el entorno hospitalario y el seguimiento en atención primaria o especializada.

Están redactados en un estilo directo, estructurado y orientado a la acción, con un lenguaje que puede ir desde técnico hasta adaptado al lenguaje sencillo, buscando la comprensión de cuáles son los pasos siguientes para realizar por el paciente.

6.1.1 Análisis de índices

Con el objetivo de evaluar la efectividad del proceso de simplificación se realizó una comparación de los niveles de dificultad antes y después de la intervención, utilizando como referencia primeramente el índice de legibilidad de Fernández Huerta. Para ello, se aplicaron las dos transformaciones consecutivas de simplificación textual, obteniéndose los índices Fernández-Huerta e INFLESZ y los resultados se representaron mediante matrices de transición. Así, mediante esta visualización podemos ver claramente los cambios entre los niveles de legibilidad de los textos, consecuencia de aplicar las simplificaciones a los textos originales. En cada casilla de nivel se indicará el número de textos que cumplen las transformaciones, siendo el eje vertical el nivel que tenía el registro en el texto original y el eje horizontal el nuevo registro que está obteniendo el texto tras ser simplificado.

En la primera simplificación (véase Figura 5), se observa una clara tendencia hacia la reducción del nivel de dificultad en 30 de los 37 casos. Dentro de estas mejoras observamos que 13 de textos fueron originalmente clasificados como “Normal” y han pasado a categorías más accesibles siendo los niveles “Algo fácil” en 10 casos y “Fácil” en 3 de ellos. Los textos considerados “Algo difíciles” se simplifican más todavía llegando hasta el nivel “Muy fácil” en un caso y “Fácil” en otro caso. Estos valores sugieren que la transformación inicial tuvo un efecto positivo en términos de accesibilidad, logrando suavizar la complejidad lingüística sin introducir grandes alteraciones que incrementen la dificultad de lectura. Solo se registraron 2 casos donde las simplificaciones no surgieron efecto, siendo un caso en el que un texto originalmente considerado “Muy fácil” donde pasó a clasificarse como “Fácil” y otro que pasa de “Muy fácil” a “Fácil”. Estos dos textos se consideran el foco del análisis de mejora en la segunda simplificación que comentaremos a continuación.

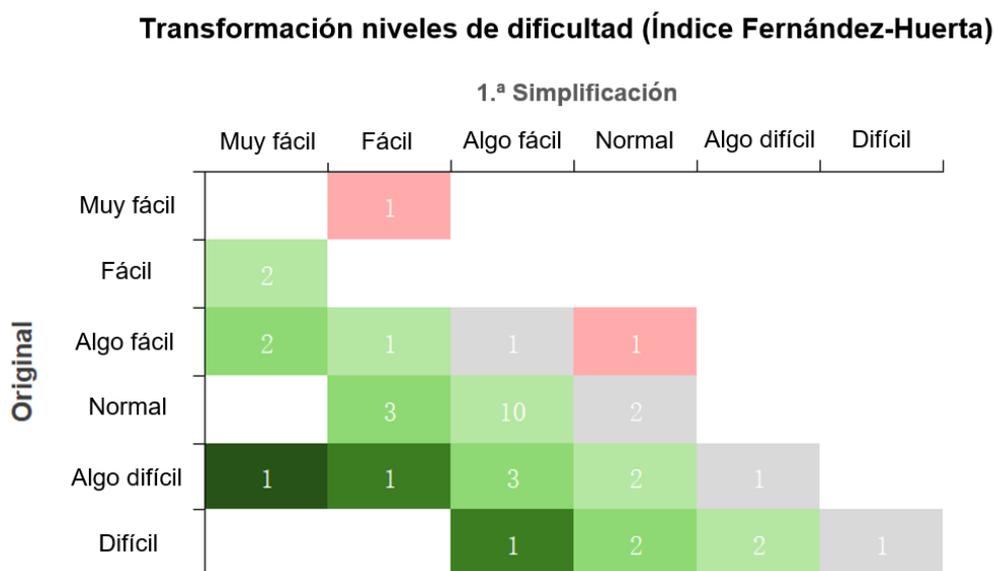


Figura 5. Matriz de transición de niveles de dificultad sobre el índice Fernández Huerta, primera simplificación en subgrupo “Plan de seguimiento”

La segunda simplificación (véase Figura 6) tuvo como objetivo simplificar en mayor medida aquellos casos recogidos tras la primera simplificación que no surgieron efecto de mejora, realizando un análisis más exhaustivo en la aplicación de las reglas. De este modo, encontramos como el texto cuyo original anteriormente era categorizado como “Algo fácil” y los resultados de la primera simplificación lo convertían a “Normal” ha mejorado posicionándose en un nivel “Fácil”.

El texto que en la primera simplificación pasó de “Muy fácil” a “Fácil” sigue sin mejorar. Si profundizamos en el contenido de ese texto, podemos considerarlo anómalo, porque su longitud es inferior a 5 palabras y, además, se trata de siglas (MAP = “médico de atención primaria”), con lo que su comprensión resulta difícil. Por ello, se obtiene un peor resultado.

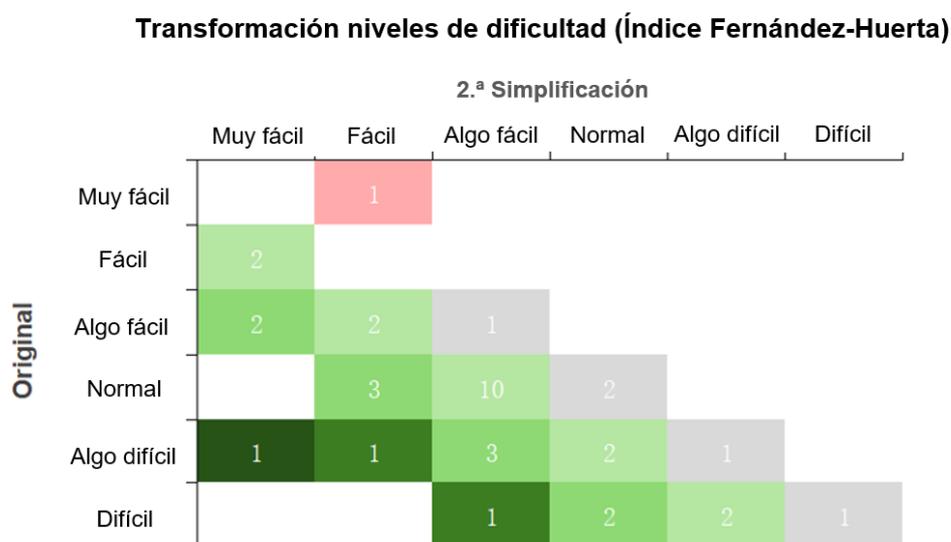


Figura 6. Matriz de transición de niveles de dificultad sobre el índice Fernández Huerta, segunda simplificación en subgrupo “Plan de seguimiento”.

Para completar el análisis se empleó también el índice INFLESZ con el fin de contrastar y reforzar los resultados obtenidos en el proceso de simplificación. Este índice nos permite realizar una medición más detallada de la legibilidad, siendo particularmente adecuado para textos en español orientados a pacientes, gracias a su mayor sensibilidad a estructuras léxicas y morfosintácticas comunes en contextos clínicos. La evaluación se realizó sobre el mismo corpus textual de planes de seguimiento médico, aplicando igualmente una doble simplificación progresiva.

En la primera simplificación (véase Figura 7), los resultados evidencian una mejora significativa de la legibilidad donde 28 de los 37 casos, aunque en menor medida que la recogida por el índice anterior. Se observa que los textos originalmente clasificados como de dificultad “Normal” la mayoría se convierten a “Bastante fácil”, siendo 12 de los casos y transformándose al nivel “Muy fácil” en uno de los casos. De los textos marcados originalmente como “Algo difícil”, algunos se reubicaron como “Bastante fáciles” (5 casos), “Normales” (2 casos) e incluso hubo un texto el cual se catalogó como “Muy fácil” siendo el

nivel máximo de legibilidad posible lo que evidencia una reducción general en la carga cognitiva de los mismos.

No obstante, al igual que en el análisis con Fernández Huerta, se detectaron los mismos casos en los que la simplificación pareció haber generado un pequeño aumento en la dificultad percibida los cuales corresponden a los mismos textos que no resultaban en mejora en el estudio del índice Fernández-Huerta, consolidando los resultados obtenidos mediante el anterior índice.

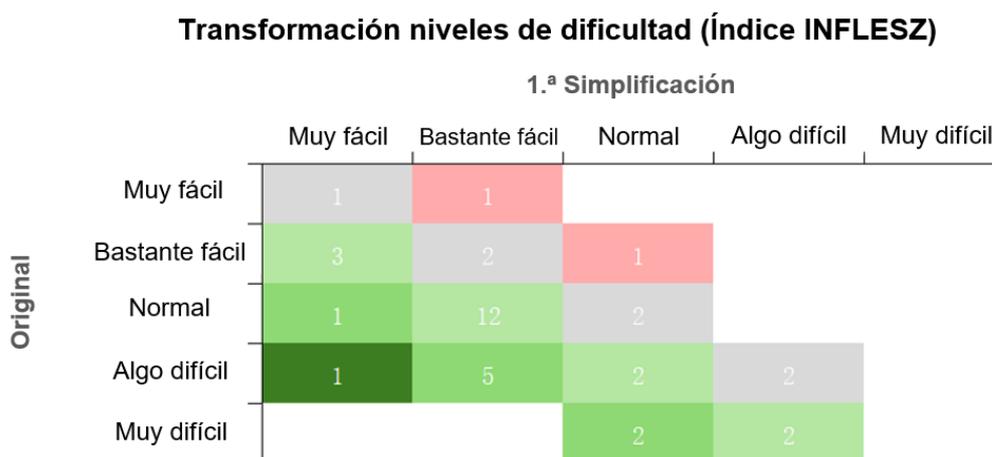


Figura 7. Matriz de transición de niveles de dificultad sobre el índice INFLESZ, primera simplificación en subgrupo “Plan de seguimiento”

En la segunda simplificación (véase Figura 8), del mismo modo que fue realizado con el índice Fernández-Huerta se incide en aquellos textos que no han podido arrojar valores positivos tras su primera transformación.

De este modo, encontramos resultados similares al anterior índice, donde el texto cuyo original anteriormente era categorizado como “Bastante fácil” tras la primera simplificación se convertía a “Normal” ha mejorado posicionándose en un nivel “Bastante fácil”, siendo el mismo nivel que el original.

El texto que en la primera simplificación pasó de “Bastante fácil” a “Normal” sigue sin mejorar. Recordamos que este podemos considerarlo anómalo, ya que su longitud es inferior a 5 palabras y, además, se trata de siglas (MAP = “médico de atención primaria”), con lo que su comprensión resulta difícil.

Transformación niveles de dificultad (Índice INFLESZ)

2.ª Simplificación

	Muy fácil	Bastante fácil	Normal	Algo difícil	Muy difícil
Muy fácil	1	1			
Bastante fácil	3	3			
Normal	1	12	2		
Algo difícil	1	5	2	2	
Muy difícil			2	2	

Figura 8. Matriz de transición de niveles de dificultad sobre el índice INFLESZ, segunda simplificación en subgrupo “Plan de seguimiento”

La ausencia de retrocesos significativos encontrada en las simplificaciones de ambos índices indica que el método de simplificación es consistente y controlado, y que las adaptaciones aplicadas no alteran negativamente la estructura de comprensión del texto. Este comportamiento es muy importante en el ámbito sanitario, donde la claridad y accesibilidad de la información escrita tienen un impacto directo en la comprensión y adherencia del paciente.

Además, los resultados arrojados en las segundas iteraciones nos muestran que el margen de mejora adicional se ve reducido, pero según la naturaleza de los textos en algunos casos aún es posible ajustar ciertos aspectos del texto para ganar en claridad sin comprometer el contenido.

6.1.2 Análisis de léxico

Tras medir la evaluación de la legibilidad mediante los índices Fernández Huerta e INFLESZ, se ha realizado un análisis léxico del subgrupo “Plan de seguimiento” del corpus textual mediante la herramienta Voyant Tools, con el objetivo de identificar los términos más frecuentes, detectar patrones de repetición y observar cambios en la densidad y variedad del vocabulario tras las simplificaciones aplicadas.

En la versión original del corpus, las palabras más frecuentes fueron “control” (22 apariciones), “visita” (20), “paciente” (17), “alta” (15) y “aislamiento” (14), seguidas de otros términos recurrentes como “acudir” (12), “seguimiento” (9), “urgencias” (9) o “médico”(9). El vocabulario refleja con claridad el contexto clínico del contenido, con un predominio de sustantivos técnicos relacionados con la atención, el seguimiento y las instrucciones al paciente.

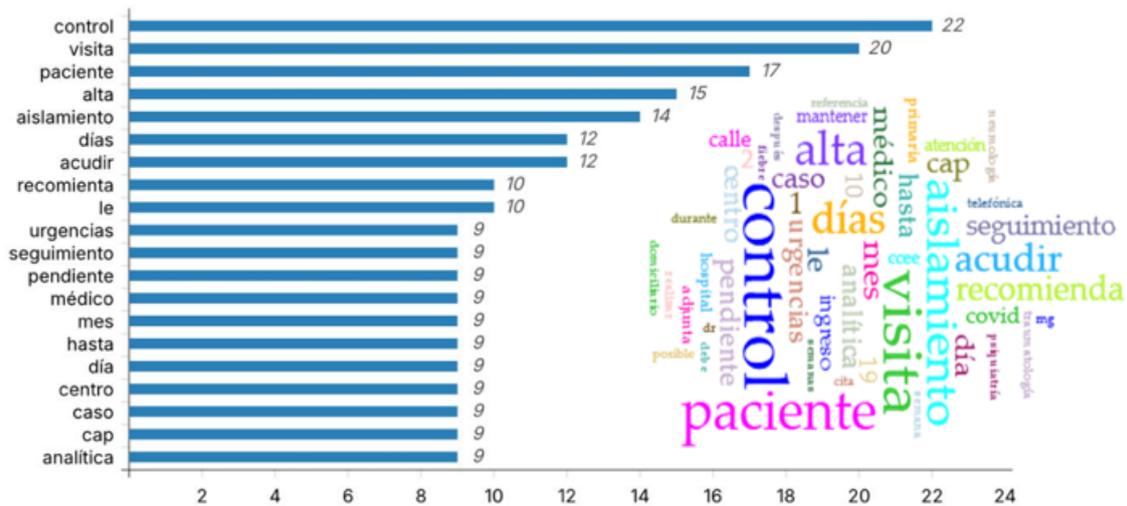


Figura 9. Frecuencia de palabras más utilizadas en subgrupo original “Plan de seguimiento” junto a su nube de palabras.

Tras la primera y segunda simplificación, se aprecia un cambio notable tanto en la frecuencia como en la naturaleza de los términos más utilizados. En las versiones simplificadas, palabras como “debe” (20 apariciones), “cita” (16), “centro” (15), “médico” (13), “alta” (13) y “seguir” (12) ganan relevancia. Este cambio léxico indica una transformación hacia un lenguaje más directo, orientado a la acción, y con un estilo más instructivo. Además, la presencia destacada del verbo “debe” sugiere una estructura más normativa y clara en las recomendaciones, lo cual es coherente con un enfoque de simplificación centrado en la claridad comunicativa.

Otro apunte interesante es la desaparición de términos menos comunes o más técnicos, como “control” o “aislamiento” lo que refuerza la hipótesis de que el proceso de simplificación ha implicado una reducción de tecnicismos y estructuras complejas, en favor de un vocabulario más cotidiano y fácilmente comprensible, como puede ser representado por los términos “revisión” y “casa”

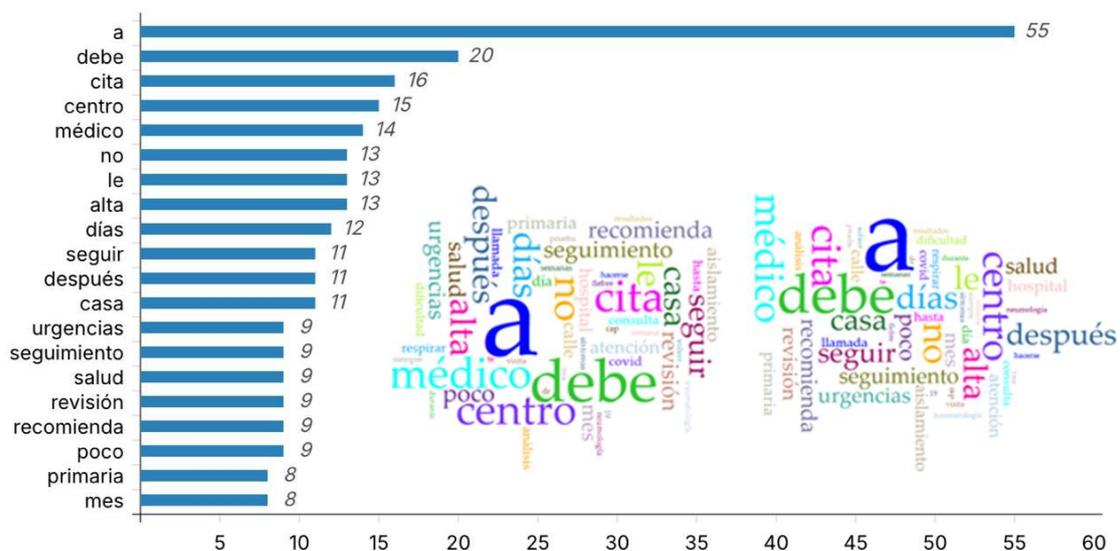


Figura 10. Frecuencia de palabras más utilizadas en subgrupo primera y segunda simplificación “Plan de seguimiento” junto a sus nubes de palabras respectivamente

También se evaluó la proporción de palabras por tipo (sustantivos, verbos, artículos, etc.), con el objetivo de detectar posibles desplazamientos estructurales en el subgrupo. El porcentaje de palabras por tipo se mantuvo prácticamente estable, aumentando ligeramente del 81 % en el texto original al 82 % tanto en la primera como en la segunda simplificación. Este dato indica que, si bien se ha modificado el contenido léxico, la estructura gramatical general de los textos no ha sido alterada de forma radical, lo que contribuye a preservar la coherencia sintáctica y semántica del discurso médico.



Figura 11. Evolución de la proporción de palabras por tipos tras las simplificaciones

Esta estabilidad sugiere que el proceso de simplificación ha sido cuidadoso en cuanto a su aplicación, no ha consistido en eliminar frases o reducir drásticamente el contenido, sino en reformular de forma más accesible los mensajes, manteniendo una proporción similar de tipos de palabra. El leve aumento del 1 % se asocia a la incorporación de expresiones más directas o a la sustitución de estructuras nominales complejas por construcciones verbales más simples.

6.1.3 Análisis de métricas

Además del estudio mediante índices de legibilidad clásicos y análisis léxico, se llevó a cabo una evaluación de tres métricas clave asociadas al grado de comprensión de los textos: edad mínima de comprensión, nivel de grado educativo necesario y tiempo estimado de lectura. Estas variables permiten valorar el impacto real de las simplificaciones desde una perspectiva práctica, es decir, cómo cambia la accesibilidad del texto para los lectores tras las intervenciones aplicadas, especialmente relevante para aquellos que encuentran dificultades en la comprensión de los textos.

La edad mínima de comprensión, expresada en años, representa la edad estimada a partir de la cual un lector medio podría comprender adecuadamente el texto. En el corpus original para el subgrupo “Plan de seguimiento”, la media fue de 9,4 años, lo que indica que algunos textos podrían ser entendidos un lector infantil de entre tercer y cuarto curso de educación primaria. Tras la primera simplificación, implementando las reglas de simplificación como la sustitución de palabras complejas por sinónimos de conocimiento más general, esta edad descendió a 8,0

años, reflejando una mejora significativa en la comprensión potencial de los textos (-1,4 años). La segunda simplificación mantuvo esta mejora, conservando el nivel de 8,0 años.



Figura 12. Evolución de la edad mínima de comprensión tras las simplificaciones

En el siguiente gráfico de línea, que detalla esta métrica por cada texto, podemos observar una mejora sostenida en la mayoría de los textos individuales. En algunos casos, las reducciones fueron especialmente marcadas (por ejemplo, textos que pasaron de más de 11 años a 8 años o menos). Esta homogeneidad en la mejora sugiere que el proceso de simplificación fue eficaz en reducir la complejidad lingüística sin depender del contenido concreto del texto, lo cual valida la robustez del método empleado.

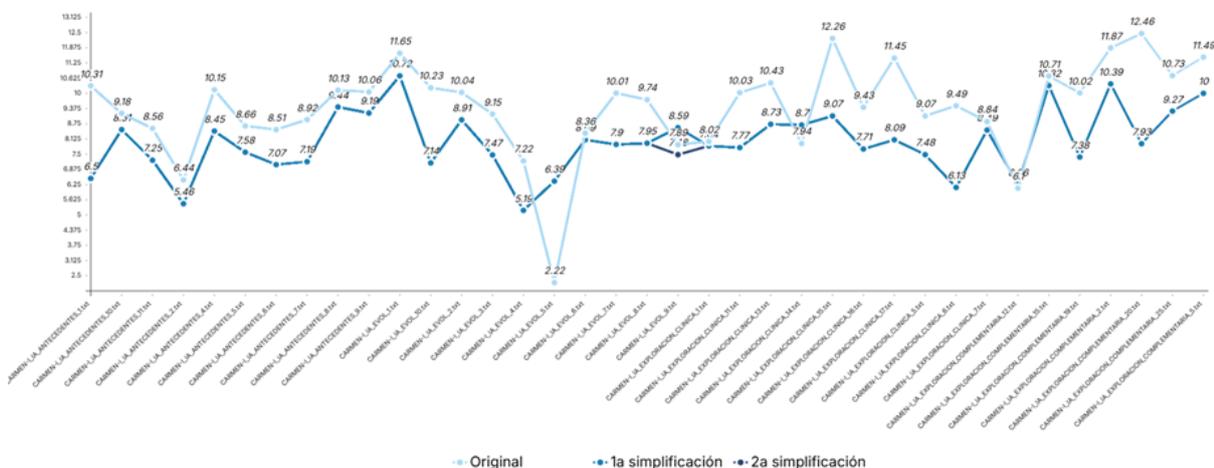


Figura 13. Evolución de la edad mínima de comprensión tras las simplificaciones en cada texto

El nivel educativo necesario, también expresado en años, estima la escolarización formal mínima requerida para comprender el texto, esta métrica es clave para comprender la situación a la que los pacientes con bajos niveles educativos o con dificultades de comprensión lectora deben enfrentarse. En los textos originales, la media se situaba en 4,1 años de escolarización, mientras que tras la primera simplificación descendió a 3,2 años, y se mantuvo en este valor tras la segunda intervención. La reducción de casi un año completo (-0,9 años) es un indicador clave del impacto de la simplificación, ya que implica que el texto se ha vuelto más accesible incluso para lectores con niveles de formación más bajos.

Nivel educativo necesario (en años)



Figura 14. Evolución del nivel de escolarización necesario tras las simplificaciones

Este cambio es coherente con la reducción observada en la edad mínima de comprensión, y reafirma el éxito de la simplificación para adaptar los textos a audiencias con escasa formación o dificultades de comprensión lectora, como pueden ser algunos perfiles de pacientes. En el gráfico de línea (véase Figura 15), se aprecia una caída generalizada a lo largo de los registros, con un comportamiento más plano tras la segunda simplificación, sugiriendo un umbral de simplificación a partir del cual no se pueden reducir más los requerimientos sin comprometer la claridad o precisión del contenido médico.

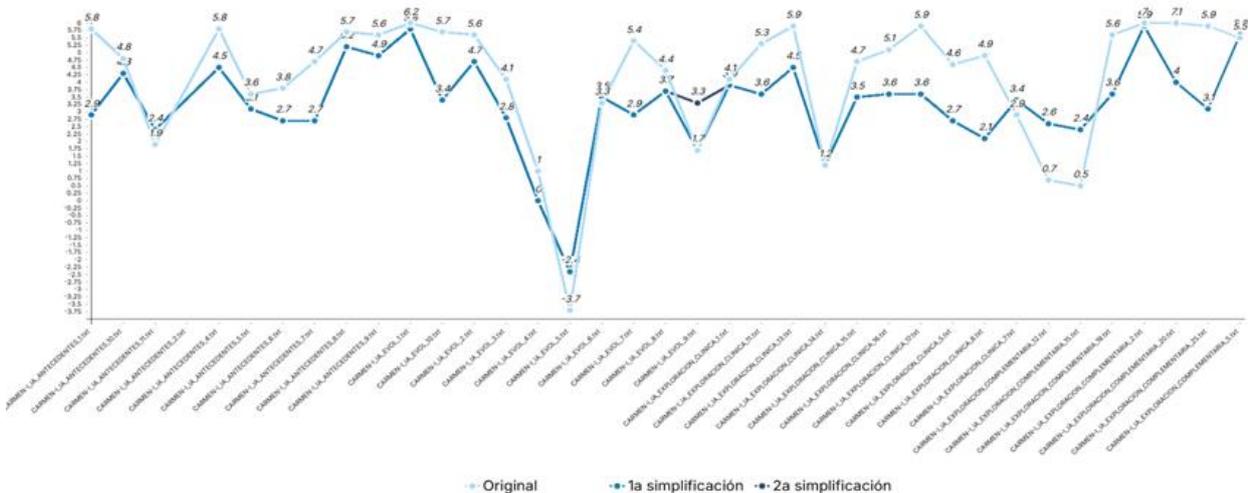


Figura 15. Evolución del nivel de escolarización necesaria tras las simplificaciones en cada texto

El tiempo estimado de lectura, medido en minutos, se mantuvo constante en 0,3 minutos (aproximadamente 18 segundos) a lo largo de las tres versiones del texto (véase Figura 16). Aunque no se evidencian diferencias significativas en esta métrica, el análisis por registro muestra pequeñas oscilaciones entre los textos, algunas de las cuales reflejan mínimas reducciones tras las simplificaciones.

6.2.1. Análisis de índices

A continuación, realizaremos el estudio de los textos de forma análoga como al anterior subgrupo. Enfocándonos primero en el índice Fernández Huerta, se observa una clara tendencia descendente en la dificultad (véase Figura 18) donde 20 de los 28 textos han obtenido una mejora en los niveles de dificultad originales.

Encontramos textos originalmente clasificados como “Muy difíciles” que se desplazaron hacia categorías más accesibles como “Algo difícil” en dos ocasiones, “Normal” en otras dos ocasiones y transformaciones hasta “Algo fácil” en un caso. Del mismo modo, destacan los textos originales marcados como “Difíciles” que se desplazaron hacia categorías más accesibles como “Normal” en una ocasión, o “Algo fácil” en dos ocasiones y hasta un caso donde tras la simplificación el texto se considera como “Fácil”.

A pesar de los buenos resultados, también se han identificado dos textos en los cuales no ha podido obtenerse un mejor nivel de legibilidad tras la primera simplificación. En la siguiente simplificación se buscará poder mejorar estos dos registros tras aplicar nuevas reglas analizando más en profundidad sus naturalezas.

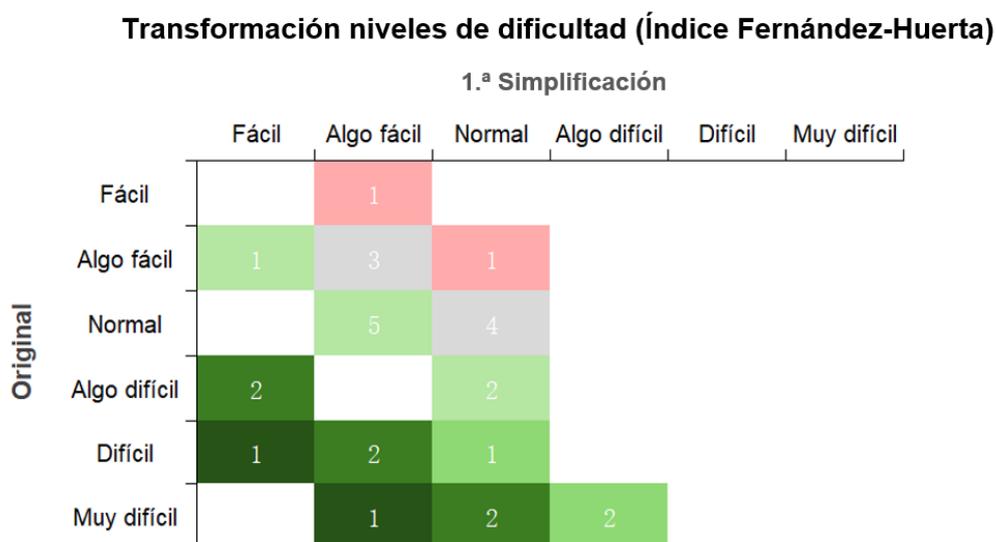


Figura 18. Matriz de transición de niveles de dificultad sobre el índice Fernández Huerta, primera simplificación en subgrupo “Plan terapéutico”

En la segunda interacción (véase Figura 19) podemos observar la mejora de estos dos textos anteriormente comentados, los cuales han dejado de transformarse a niveles más complejos. Haciendo un enfoque más personalizado sobre ellos, estos contenían textos de menos de 15 palabras donde, tras un cambio en la transformación siendo más incisivos en las acciones a realizar por el paciente se consiguió la mejora.

Además, también fueron mejorados el resto de textos los cuales ya contaban con mejoras en la primera simplificación, esto muestra que en este subgrupo dedicado a planes terapéuticos el margen de mejora es muy amplio, debido a la naturaleza instructiva de estos hacia los pacientes.

Transformación niveles de dificultad (Índice Fernández-Huerta)

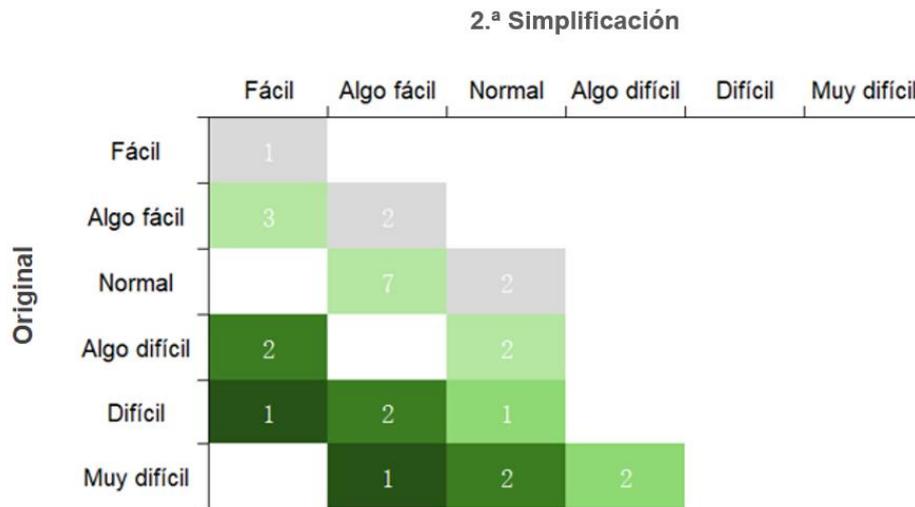


Figura 19. Matriz de transición de niveles de dificultad sobre el índice Fernández Huerta, segunda simplificación en subgrupo “Plan terapéutico”

Los resultados del índice INFLESZ (véase Figura 20) refuerzan esta misma conclusión. De forma similar, se evidencian mejoras con desplazamientos desde niveles como “Algo difícil” hacia “Normal” en dos ocasiones, hacia “Bastante fácil” en una ocasión e incluso hasta “Muy fácil” en otra ocasión, también destacan los cambios del nivel “Muy difícil”, el peor de todos, donde tras las simplificaciones todos ellos mejoran llegando a haber un registro que ha sido transformado hasta el mejor nivel de legibilidad de “Muy fácil”.

Transformación niveles de dificultad (Índice INFLESZ)



Figura 20. Matriz de transición de niveles de dificultad sobre el índice INFLESZ, primera simplificación en subgrupo “Plan terapéutico”

En cuanto a la segunda iteración (véase Figura 21), vemos como se repite el comportamiento descrito en el índice anterior, donde aquellos dos textos donde se hizo especialmente énfasis por su naturaleza, que empeoraban su nivel de dificultad mejoran nuevamente.

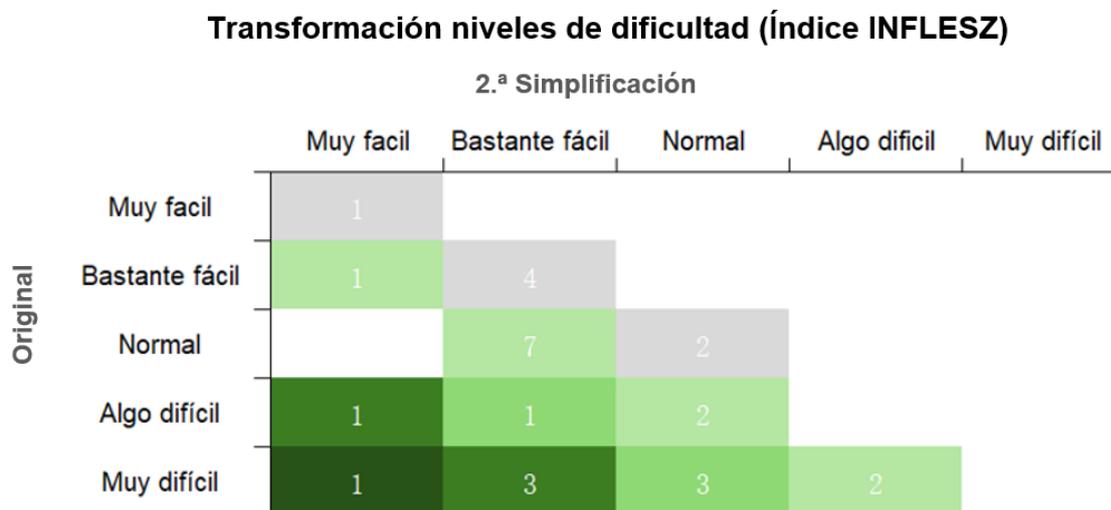


Figura 21. Matriz de transición de niveles de dificultad sobre el índice INFLESZ, segunda simplificación en subgrupo “Plan terapéutico”

6.2.2. Análisis de léxico

El análisis de frecuencia léxica mediante nube de palabras y gráficas de barras permitió identificar los términos más comunes en este subgrupo, tanto en la versión original como en las simplificadas.

En el texto original, como avanzábamos por la naturaleza de este subconjunto predominan términos técnicos relacionados con posología y forma farmacéutica, como “mg” (149 apariciones), “vía” (86), “oral” (77), “comprimidos” (76), “desayuno” (43), “efg” (34), o “recubiertos” (15). Esto es consistente con el contenido terapéutico, pero plantea barreras para pacientes con bajo nivel de alfabetización en salud.

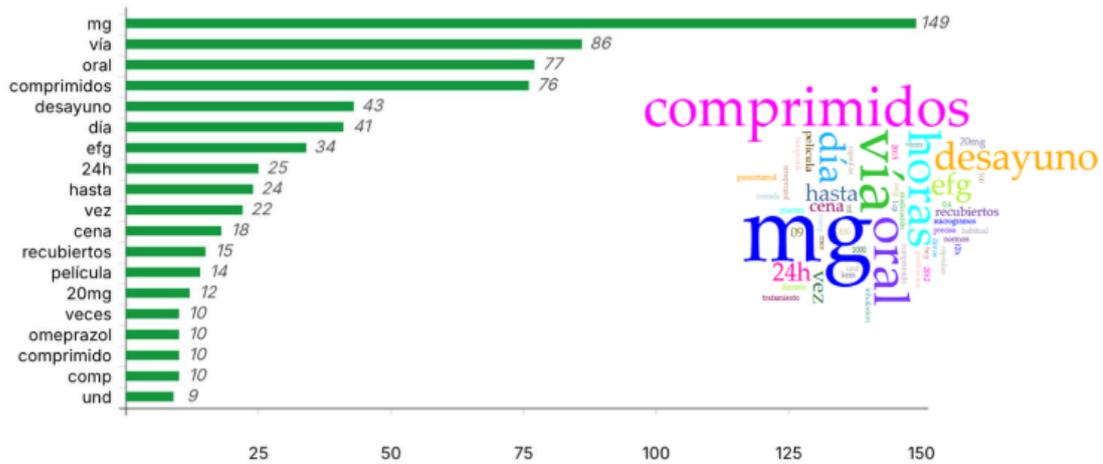


Figura 22. Frecuencia de palabras más utilizadas en subgrupo original “Plan de seguimiento” junto a su nube de palabras

Tras la primera simplificación, se mantiene el núcleo temático mediante los términos “mg” (130), “comprimido” (61), “día” (91), “pastilla” (14), “tomar” (44), pero se sustituye o reduce el uso de terminología técnica como “efg”, “recubiertos”, “película” o abreviaturas farmacológicas. También vemos cómo palabras como “tomar”, “vez”, “pastilla” o “mañana” ganan peso en las versiones simplificadas, favoreciendo un lenguaje más accesible y orientado a la acción sugiriendo un esfuerzo por reformular el contenido en estructuras más comprensibles, sin pérdida de información terapéutica.

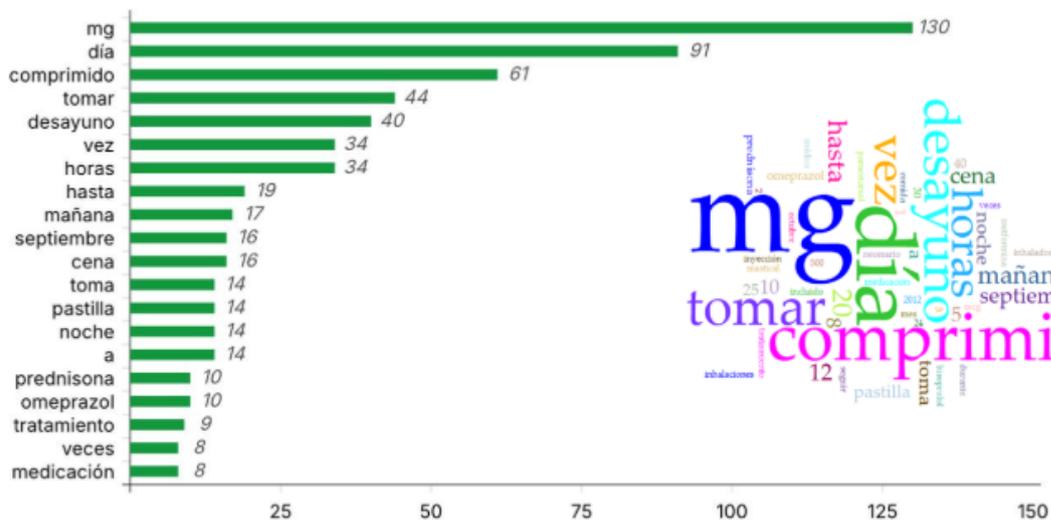


Figura 23. Frecuencia de palabras más utilizadas en subgrupo primera simplificación “Plan de seguimiento” junto a su nube de palabras

Vemos cómo en la segunda iteración aparecen nuevas palabras entre las más usadas como “pastilla”, esto es debido a que la segunda iteración dentro de este subgrupo de “Plan terapéutico” se ha enfocado en estandarizar los discursos entre los textos, donde, la

mayoría de las ocasiones se realizan listas de tareas que el paciente debe realizar para su llevar a cabo su tratamiento.

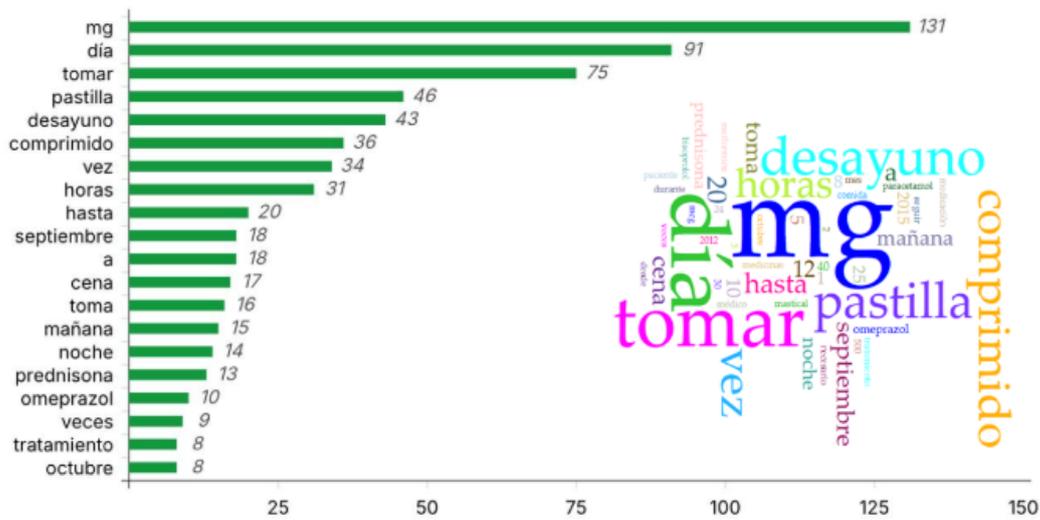


Figura 24. Frecuencia de palabras más utilizadas en subgrupo segunda simplificación “Plan de seguimiento” junto a su nube de palabras

En cuanto al porcentaje de palabras por tipos, se detecta un descenso progresivo: del 68 % en el original al 64 % en la primera simplificación, y 63 % en la segunda. Este descenso se explica por la sustitución de formas técnicas complejas por construcciones verbales simples, lo que incrementa la presencia de conectores, pronombres y verbos auxiliares, típicos de un lenguaje más conversacional.



Figura 25. Evolución de la proporción de palabras por tipos tras las simplificaciones

6.2.3. Análisis de métricas

La edad mínima de comprensión pasó de 10,8 años en el texto original a 8,1 años tras la primera simplificación, y 7,8 años en la segunda. Esta reducción de más de tres años representa una mejora sustancial en términos de accesibilidad, permitiendo que lectores con menor desarrollo cognitivo comprendan instrucciones críticas sobre su medicación.

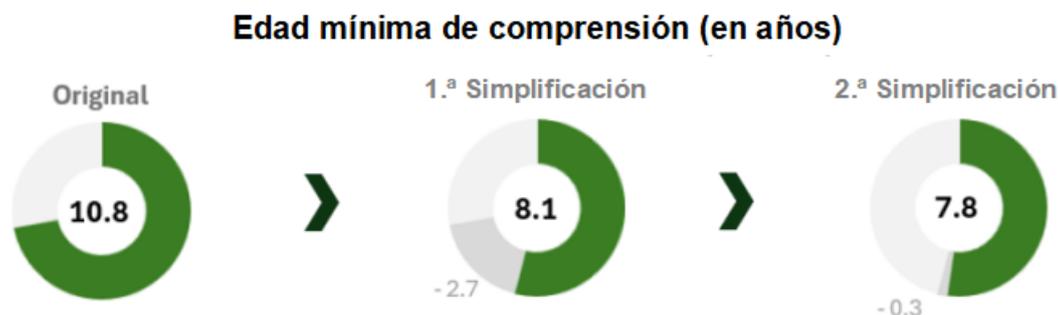


Figura 26. Evolución de la edad mínima de comprensión tras las simplificaciones

El gráfico de línea por documento muestra que esta mejora se reproduce en la mayoría de los registros, destacando especialmente casos que pasan de valores por encima de los 13 años a cifras por debajo de los 9, lo cual representa un avance notable en equidad comunicativa en salud.

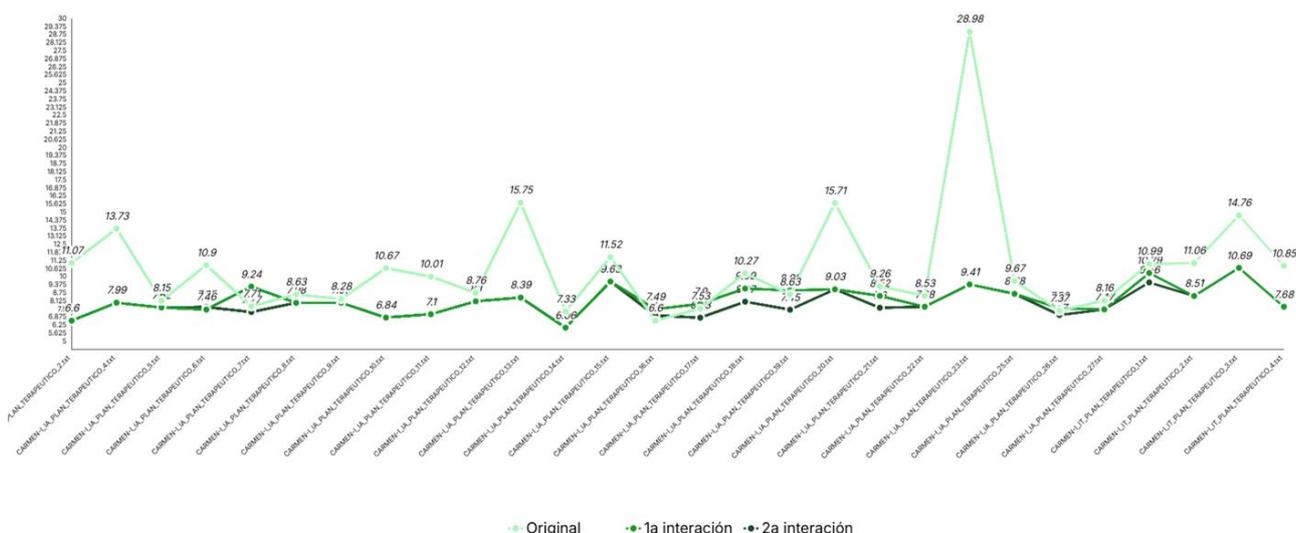


Figura 27. Evolución de la edad mínima de comprensión tras las simplificaciones en cada texto

Se registró una disminución desde los 4,4 años de escolarización promedio hasta 3,0 y finalmente 2,5 años tras la segunda simplificación. Esto indica que, en su versión final, el texto puede ser comprendido por personas con niveles muy bajos de escolarización formal, lo que es particularmente relevante en poblaciones vulnerables o con bajo acceso educativo.

Nivel educativo necesario (en años)



Figura 28. Evolución del nivel de escolarización necesario tras las simplificaciones

El gráfico de transformación del nivel educativo necesario refleja de forma más evidente el efecto de la simplificación textual. En la versión original, algunos textos requerían más de 8 o incluso 12 años de escolarización, lo cual suponía una barrera significativa para pacientes con bajo nivel educativo. Tras la primera simplificación, se aprecia una caída generalizada a lo largo de los registros, reduciendo el nivel necesario en muchos casos a rangos más accesibles, entre 2 y 4 años. Con la segunda simplificación, el comportamiento de la curva se vuelve más plano y homogéneo, sin nuevos descensos pronunciados.

Transformación nivel de grado educativo (en años)

Módulo Plan Terapéutico

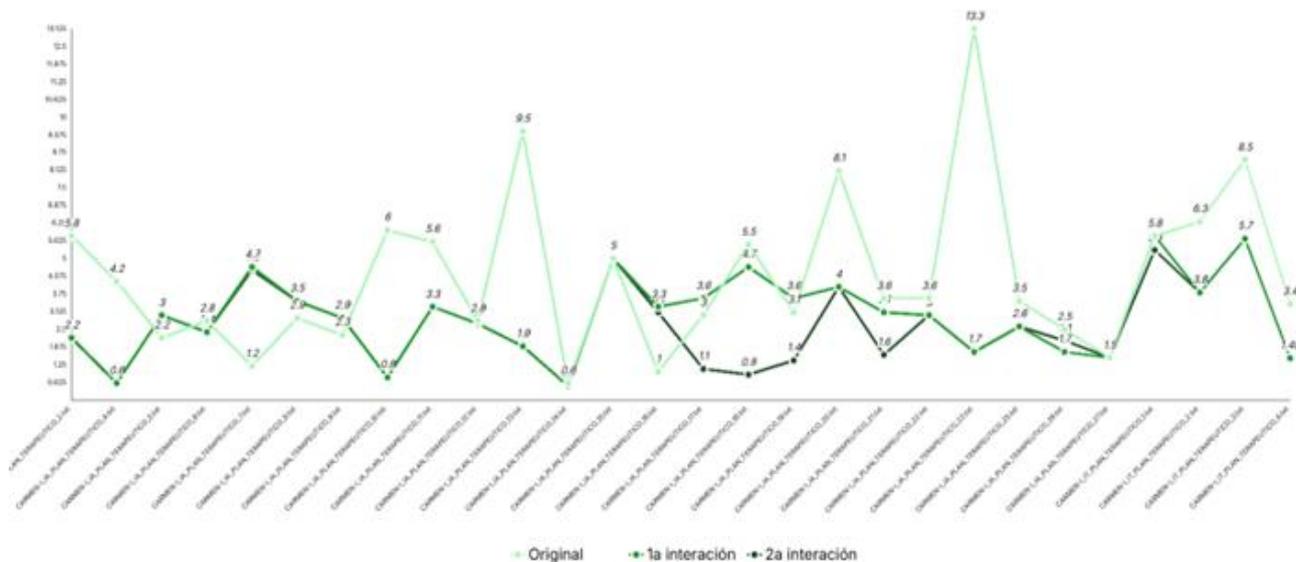


Figura 29. Evolución del nivel de escolarización necesaria tras las simplificaciones en cada texto

El tiempo medio de lectura se mantuvo en 0,4 minutos tras la primera simplificación y subió ligeramente a 0,6 minutos tras la segunda. Este pequeño está relacionado con la incorporación de aclaraciones o reformulaciones más explícitas en las listas de tareas a realizar por el paciente para lograr una mayor claridad en el mensaje y objetivos. Además, se valora positivamente que la accesibilidad se haya incrementado sin aumentar excesivamente la extensión del texto.



Figura 30. Evolución del tiempo estimado de lectura tras las simplificaciones

El análisis del tiempo estimado de lectura para los textos del subgrupo “Plan terapéutico” confirma una tendencia de estabilización tras las simplificaciones. Aunque algunos textos originales presentaban tiempos de lectura relativamente elevados, incluso superiores al minuto y medio, tras la primera simplificación se observa una ligera reducción en ciertos casos.

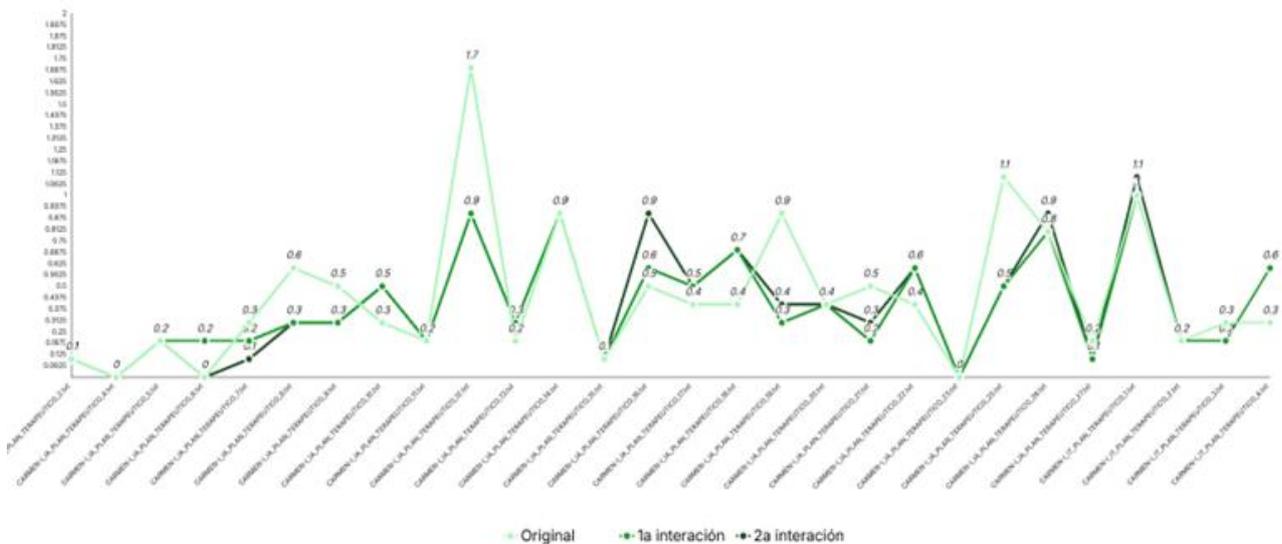


Figura 31. Evolución del tiempo estimado de lectura tras las simplificaciones en cada texto

7. Propuesta

A partir de los resultados obtenidos mediante el análisis cuantitativo y cualitativo del corpus, se identifican una serie de líneas de mejora que pueden orientar futuras iniciativas de simplificación textual en el ámbito clínico. Estas propuestas se fundamentan en la efectividad demostrada de las intervenciones realizadas y buscan ampliar su aplicabilidad, sistematización y sostenibilidad dentro del entorno sanitario real.

Uno de los principales logros del presente trabajo ha sido la aplicación sistemática de un conjunto de criterios de simplificación lingüística en textos médicos reales. A partir de esta

experiencia, se considera necesario avanzar hacia la sistematización de estos criterios en forma de una guía de estilo clínica, que pueda servir como referencia práctica para profesionales sanitarios a la hora de redactar documentos destinados al paciente. Esta guía no debería limitarse a recomendaciones generales, sino que conviene que contemple secciones específicas adaptadas a los distintos bloques que suelen aparecer en los informes clínicos. Por ejemplo, en la sección de instrucciones terapéuticas, se sugiere priorizar el uso de imperativos positivos (“tome”, “acuda”, “espere”) frente a fórmulas impersonales o negativas ambiguas; en los bloques más relacionados con la posología y dosificación, debería promoverse la expresión natural de pautas horarias (“por la mañana”, “una vez al día”) en lugar de formatos abreviados más complejos de entender (“1-0-0”, “/24h”). En cuanto a la presentación de fechas o resultados numéricos, se recomienda evitar símbolos técnicos y expresar los datos de forma completa, con unidades claramente mencionadas y fechas escritas en formato natural. También deberían recogerse ejemplos de paráfrasis frecuentes para tecnicismos recurrentes (como “cefalea” → “dolor de cabeza”, “recubiertos con película” → “pastillas normales”), así como un listado de expresiones a evitar por su complejidad o ambigüedad. En definitiva, la guía debería funcionar como un recurso vivo, útil no solo para redactar de forma clara, sino también como herramienta de revisión y adaptación lingüística en la producción de materiales clínicos.

Asimismo, el análisis realizado ha evidenciado que la necesidad de simplificación no es uniforme en todas las tipologías textuales clínicas, sino que cada subgénero presenta desafíos particulares que requieren estrategias diferenciadas. En el caso del subgrupo de planes de seguimiento, el discurso presenta una narrativa más continua, con frases de carácter explicativo que requieren una reformulación centrada en reducir la complejidad sintáctica y reorganizar la información de manera más jerarquizada. La prioridad aquí es mejorar la claridad global del mensaje y evitar sobrecarga informativa. En cambio, en los planes terapéuticos, el contenido se organiza de forma esquemática, con predominio de listados de medicación, pautas y horarios. Estos textos requieren una atención especial a los formatos numéricos, unidades de medida y abreviaturas, que muchas veces resultan incomprensibles para los pacientes. Este hallazgo sugiere que una futura estrategia de simplificación debería contemplar protocolos diferenciados según el tipo de documento médico, de modo que se ajusten no solo a la estructura del texto, sino también al nivel de riesgo asociado a su correcta interpretación por parte del paciente.

En segundo lugar, y como prolongación natural de lo anterior, se propone avanzar hacia la integración de herramientas automáticas de apoyo a la escritura accesible en el entorno clínico. A lo largo del presente trabajo, se ha evidenciado el potencial de plataformas como Legible o Voyant Tools para evaluar la legibilidad y el perfil léxico de los textos. No obstante, su aplicación ha sido posterior a la redacción del texto, lo que limita su impacto inmediato. Por ello, una vía de mejora clave sería explorar el desarrollo de funcionalidades que permitan operar durante el propio proceso de escritura. En este sentido, una de las limitaciones detectadas en este estudio ha sido la falta de herramientas específicas que actúen de forma proactiva —como sugerencia de sinónimos, reformulación de estructuras complejas, o alertas sobre tecnicismos—, algo que sin duda enriquecería la práctica clínica diaria. Esta limitación

metodológica se reconoce como una de las fronteras del presente trabajo, pero también como una oportunidad de desarrollo futura. La integración de sistemas de apoyo lingüístico, similares a correctores gramaticales, pero orientados a legibilidad clínica, podría implantarse en plataformas de historia médica electrónica, ayudando al personal sanitario a producir mensajes más comprensibles sin necesidad de conocimientos técnicos en lingüística o edición.

En tercer lugar, considerando que la redacción de documentos clínicos comprensibles no suele formar parte explícita de la formación sanitaria, se propone incorporar programas de formación específicos en lenguaje claro y accesible dirigidos a profesionales de la salud. Esta formación debería integrarse tanto en la etapa universitaria como en la formación continua, especialmente para profesionales que redactan instrucciones clínicas, informes de alta o materiales informativos para pacientes. Lejos de ser una competencia accesorio, la habilidad para comunicar con claridad forma parte esencial de una atención sanitaria segura y equitativa. La propuesta no se limitaría a talleres puntuales, sino que implicaría la creación de módulos permanentes dentro de los planes docentes, con contenidos que aborden desde la identificación de barreras lingüísticas hasta la práctica de reescritura de textos reales, pasando por el análisis de casos de malentendidos clínicos relacionados con la ambigüedad textual. Este tipo de formación no solo dotaría a los profesionales de recursos para escribir de forma más clara, sino que también permitiría concienciar sobre el impacto que el lenguaje tiene en la comprensión, adherencia terapéutica y autonomía del paciente.

Por último, se plantea la necesidad de incorporar mecanismos de validación directa con usuarios reales como parte del proceso de evaluación de legibilidad. Aunque los índices de legibilidad y métricas lingüísticas utilizadas en este estudio han demostrado ser útiles para medir mejoras objetivas en los textos, no sustituyen la percepción y experiencia reales del lector final. Por tanto, una mejora sustancial consistiría en realizar sesiones de validación participativa con pacientes, especialmente aquellos con bajo nivel educativo o con antecedentes de dificultades de comprensión médica. Estas sesiones podrían adoptar la forma de entrevistas semiestructuradas, pruebas de lectura con recogida de impresiones o tareas de comprensión lectora guiada. La información obtenida permitiría identificar puntos críticos que los algoritmos de legibilidad no detectan, como ambigüedades contextuales, interpretaciones erróneas o sensaciones de inseguridad. Además, este enfoque participativo aportaría una dimensión ética y humanizadora al proceso de simplificación, permitiendo ajustar las estrategias no solo desde parámetros técnicos, sino también desde las necesidades y expectativas reales de los pacientes. La validación social del texto, en definitiva, garantizaría que la claridad alcanzada sea efectiva, funcional y alineada con la diversidad del público receptor.

En resumen, los puntos más relevantes de nuestra propuesta son:

1. Seguimiento de una guía de estilo clínica para la redacción de textos dirigidos al paciente, basada en criterios de simplificación como los descritos en el apartado 5.3 de este proyecto. Pudiendo así sistematizar de una manera más completa la función de escritura

de estos textos y por ende mejorando su accesibilidad. También debe considerarse la tipología del documento clínico ya que, como bien hemos visto, existen diferencias en sus formas de comunicación y objetivo.

2. Contemplar la integración de herramientas automáticas de apoyo durante la redacción. Estos programas facilitan la redacción accesible de forma cómoda y transparente al usuario a la vez que se ven potenciados los niveles de accesibilidad.
3. Incorporación de formación específica en lenguaje claro y accesible en el ámbito sanitario. De este modo, los profesionales dedicados a la redacción de los textos dirigidos a pacientes conocerán las mejores prácticas para comunicar sus textos de la manera más accesible y conocerán los posibles puntos donde pueden aparecer barreras que impidan a los pacientes comprender sus informes.
4. Fomento de la validación de los escritos con usuarios reales, poniendo a prueba los textos redactados en sesiones participativas para entender si estos informes realizados realmente son accesibles de manera directa a todos los pacientes. De este modo, el plan de formación comentado en el punto anterior quedaría completado, no solo atendiendo de manera teórica sino también desde un punto de vista práctico, que reafirme los conocimientos aprendidos.

8. Conclusiones

Este trabajo se ha centrado en el desarrollo y aplicación de una metodología de simplificación textual orientada a mejorar la legibilidad de documentos clínicos dirigidos a pacientes, concretamente en los subgrupos de “Plan de seguimiento” y “Plan terapéutico”. A partir de una perspectiva mixta (lingüística, técnica y asistencial), el estudio ha demostrado que es posible transformar textos médicos con un alto grado de tecnicismo y complejidad estructural en versiones más accesibles, sin que ello suponga pérdida de contenido ni ambigüedad comunicativa.

El punto de partida del trabajo ha sido el reconocimiento de una problemática real en el ámbito de la comunicación médico-paciente: muchos documentos clínicos, a pesar de estar dirigidos a personas sin formación sanitaria, presentan niveles de legibilidad que exigen edades y niveles educativos superiores a los que puede asumir una parte significativa de la población. Este desajuste no solo representa un obstáculo para la comprensión de la información médica, sino que también compromete la adherencia a tratamientos, la toma de decisiones informadas y, en última instancia, la seguridad del paciente.

Ante esta situación, se ha diseñado una metodología de simplificación en dos fases, mediante una primera intervención donde se reformularon manualmente textos auténticos procedentes de un corpus clínico real, respetando su contenido médico, pero modificando su estructura sintáctica, elección léxica y presentación general y posteriormente un análisis de las

simplificaciones realizadas junto al uso de herramientas de análisis lingüístico (como Legible y Voyant Tools).

Los resultados obtenidos a través de los índices de legibilidad Fernández Huerta e INFLESZ mostraron reducciones significativas en la dificultad textual, con desplazamientos evidentes desde categorías como “Difícil” o “Muy difícil” hacia niveles como “Normal”, “Algo fácil” o “Fácil”. Estas mejoras se observaron de forma consistente en ambas fases de simplificación, aunque la primera produjo el cambio más marcado y la segunda actuó como ajuste fino para casos más específicos. En ningún caso se registraron retrocesos generalizados, lo cual respalda la estabilidad del método.

Complementariamente, el análisis léxico evidenció un cambio en los términos dominantes, con una disminución del uso de tecnicismos, abreviaturas médicas y formas pasivas, en favor de construcciones verbales más comunes y directas. Este cambio semántico se produjo sin alterar el contenido funcional del mensaje, pero incrementando su comprensibilidad potencial.

Las métricas complementarias (edad mínima de comprensión, nivel educativo requerido y tiempo de lectura) aportaron una dimensión más práctica a la evaluación. La edad media de comprensión descendió en ambos subgrupos hasta situarse por debajo de los 8 años, y el nivel de escolarización necesario pasó de valores cercanos a 4,5 años en los textos originales a 2,5 años tras la segunda simplificación. Estos resultados sitúan los textos adaptados dentro de rangos asumibles para un gran número de pacientes, incluyendo aquellos con bajo nivel educativo o con dificultades lectoras. Además, el tiempo de lectura se mantuvo estable, confirmando que la simplificación no se logró a costa de eliminar contenido relevante.

El análisis diferenciado entre los dos subgrupos permitió identificar características específicas en cada tipo de documento. En el caso del plan de seguimiento, el discurso era más narrativo y explicativo, lo que requería una estrategia centrada en la reorganización del contenido, la simplificación de estructuras subordinadas y la eliminación de ambigüedades. En cambio, en los planes terapéuticos, el reto consistía en interpretar y reescribir listados complejos de medicación, dosificación y horarios, lo cual exigía una estandarización léxica y la transformación de códigos médicos en expresiones comprensibles.

Una de las aportaciones más valiosas del trabajo ha sido la formulación de una propuesta de mejora. A partir de los hallazgos, se ha planteado la necesidad de desarrollar una guía de estilo para la redacción médica accesible, incorporando criterios concretos según el tipo de contenido (dosis, fechas, instrucciones, tecnicismos). Asimismo, se ha identificado como línea de futuro la exploración de herramientas automatizadas de asistencia a la redacción clara, capaces de ofrecer alertas, sugerencias y reformulaciones durante el proceso de escritura. Aunque estas herramientas aún presentan limitaciones y su integración en los sistemas sanitarios sería compleja, su desarrollo representaría un avance significativo hacia la normalización del lenguaje claro en el entorno clínico.

En conjunto, este trabajo ha demostrado que la simplificación textual médica no solo es viable, sino necesaria. A través de un proceso metodológico riguroso, y con herramientas accesibles y replicables, es posible mejorar de forma significativa la comprensión de los documentos clínicos dirigidos a pacientes. Esta transformación tiene implicaciones éticas, clínicas y sociales, ya que promueve la equidad en salud, empodera al paciente y reduce riesgos derivados de la mala comprensión.

Como línea de continuidad, se propone extender esta metodología a otros géneros textuales médicos como pueden ser consentimientos informados, informes de alta hospitalaria, documentos de consentimiento terapéutico, entre otros para descubrir aquellas pautas que puedan convertirlos en más accesibles.

9. Bibliografía y referencias

Araujo, J. C. (2017). Reflexiones en torno al lenguaje médico actual, los epónimos y abreviaciones: las razones de su existencia y los principales problemas que plantea su uso. *Biosalud*, 16(1), 93-104. <http://www.scielo.org.co/pdf/biosa/v16n1/v16n1a10.pdf>

Ayuntamiento de Madrid. (2017). Comunicación Clara. *Guía práctica*. <https://www.madrid.es/UnidadesDescentralizadas/Calidad/LenguajeClaro/ComunicacionClara/Documentos/GuiaPracticaCClara.pdf>

Barrio-Cantalejo, I. M., Simón-Lorda, P., Melguizo, M., Escalona, I., Marijuán, M. I., y Hernando, P. (2008). Validación de la Escala INFLESZ para evaluar la legibilidad de los textos dirigidos a pacientes. *Atención Primaria*, 40(7), 371–376. http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1137-66272008000300004&lng=es&tyng=es.

Bartolomé Rodríguez, R., Campillos Llanos, L., y Terroba Reinales, A. R. (2023). Guideline for sentence simplification of medical texts in Spanish. *Repositorio Digital CSIC*.

Campillos-Llanos, L., Ortega-Riba, F., Terroba, A. R., Valverde-Mateos, A., y Capllonch-Carrión, A. (2024). CLARA-MeD Tool - A System to Help Patients Understand Clinical Trial Announcements and Consent Forms in Spanish. *Studies in health technology and informatics*, 316, 95–99. <https://doi.org/10.3233/SHTI240354>

Crawford, A. N. (1989). Fórmula y gráfico para determinar la comprensibilidad de textos de nivel primario en castellano. *Lectura y Vida*, 10(4). http://www.lecturayvida.fahce.unlp.edu.ar/numeros/a6n4/06_04_Crawford.pdf

de la Victoria Rosales, M. (2021). Comunicación en la relación médico–paciente. *Vida y ética*, 22(2), 9-26. <https://erevistas.uca.edu.ar/index.php/VyE/article/view/3967>

Farre Maduell, E., Lima-Lopez, S., Frid, S. A., Conesa, A., Asensio, E., Lopez-Rueda, A., Arino, H., Calvo, E., Bertran, M. J., Marcos, M. A., Nofre Maiz, M., Tañá Velasco, L., Martí, A., Farreres, R., Pastor, X., Borrat Frigola, X., & Krallinger, M. (2023). CARMEN-I: A resource of anonymized electronic health records in Spanish and Catalan for training and testing NLP tools (version 1.0). *PhysioNet*. <https://doi.org/10.13026/bxrx-y344>.

Farre Maduell, E., Lima-Lopez, S., Frid, S. A., Conesa, A., Asensio, E., Lopez-Rueda, A., Arino, H., Calvo, E., Bertran, M. J., Marcos, M. A., Nofre Maiz, M., Tañá Velasco, L., Martí, A., Farreres, R., Pastor, X., Borrat Frigola, X., y Krallinger, M. (2024). CARMEN-I: A resource of anonymized electronic health records in Spanish and Catalan for training and testing NLP tools (version 1.0.1). *PhysioNet*. <https://doi.org/10.13026/x7ed-9r91>

Fernández Huerta, J. (1959). Medidas sencillas de lecturabilidad. *Consigna: Revista pedagógica de la Sección Femenina de Falange ET y de las JONS*, (214), 29–32.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32(3), 221–233. <https://doi.org/10.1037/h0057532>

García López JA y Arcos A. Medida de la legibilidad del material escrito. *Pharm Care Esp* 1999; 1(6): 412-9.

López Hernández, J. (2022). Análisis y tipificación de errores lingüísticos para una propuesta de mejora de informes médicos en español (*Tesis doctoral, Universidad de Murcia, Escuela Internacional de Doctorado*).

Montolío, E., y Tascón, M. (2020). El derecho a entender. *La comunicación clara, la mejor defensa de la ciudadanía. Catarata*.

Organización Internacional de Normalización. (2023). ISO 24495-1:2023 Plain language — Part 1: Governing principles and guidelines. <https://www.iso.org/standard/78907.html>

Pistola, S., y da Cunha, I. (2024). Análisis y clarificación de un corpus de fraseología administrativa: Una propuesta orientada al uso del lenguaje claro en la administración española. *Logos: Revista de Lingüística, Filosofía y Literatura*, 34(2), 494–522. <https://doi.org/10.15443/RL3421>

Real Academia Española. (2022). Red Panhispánica de Lenguaje Claro y Accesible. <https://www.rae.es/la-institucion/red-panhispanica-de-lenguaje-claro-y-accesible>

Saggion, H. (2024). Artificial intelligence and natural language processing for easy-toread texts. *Revista de Llengua i Dret: Journal of Language and Law*, 82, 84–103. <https://doi.org/10.58992/rld.i82.2024.4362>

SanJuan, É., Huet, S., Kamps, J., y Ermakova, L. (2024). Overview of the CLEF 2024 SimpleText Task 1. *CEUR Workshop Proceedings (Vol. 3740)*. <https://ceur-ws.org/Vol-3740/paper-305.pdf>

Szigriszt Pazos, F. (1993). Sistemas predictivos de legibilidad del mensaje escrito: Fórmula de Perspicuidad (*Tesis doctoral, Universidad Complutense de Madrid*). <https://webs.ucm.es/BUCM/tesis/19911996/S/3/S3019601.pdf>

10. Anexos

10.1. Anexo 1. Documento de trabajo.

El archivo Excel donde son almacenados los resultados contiene una tabla estructurada que presenta los textos originales de cada subgrupo junto a sus correspondientes versiones adaptadas bajo criterios de lenguaje claro.

En ambos subgrupos encontramos las mismas columnas, pertenecientes a un nombre descriptivo del texto dentro del corpus, el texto original como tal y las versiones simplificadas, al igual que el tratamiento que se han aplicado a cada uno de ellos, siguiendo las reglas propuestas en el proyecto (apartado 5.3).

Nombre	Texto original	Versión simplificada	Uso de solución
CARMEN-IJA_ANTECEDENTES_1.txt	<ul style="list-style-type: none"> - Acudir a cita en consulta externa según volante adjunto. - Acudir a urgencias en caso de dolor incontrolable con analgesia pautada, fiebre elevada, náuseas o vómitos, empeoramiento del estado general, etc. - Retirar grapas en CAP de zona después de 7-10 días de la intervención. 	<p>Ir a la consulta externa según la cita que le han dado.</p> <p>Ir a urgencias si tiene dolor que no mejora con los calmantes indicados, fiebre alta, náuseas, vómitos o si se siente peor en general.</p> <p>Quitar las grapas en su centro de atención primaria (CAP) entre siete y diez días después de la operación.</p>	<p>5.3.1.I. Uso de verbos en lugar de nominalizaciones</p> <p>5.3.1.II. Conversión de voz pasiva a activa</p> <p>5.3.1.IV. Control del uso de conjunciones</p> <p>5.3.1.V. Eliminación de redundancias</p> <p>5.3.2.I. Expansión de abreviaturas y acrónimos</p> <p>5.3.2.VI. Uso de sinónimos más comunes</p> <p>5.3.3.I - Números del 1 al 10 escritos con letras (cuando no van con unidades técnicas)</p> <p>5.3.3.IV. Homogeneidad de formatos</p>
CARMEN-IJA_ANTECEDENTES_4.txt	<ul style="list-style-type: none"> - Seguimiento de especialistas habituales - Se recomienda la vacunación anual de la gripe. - Se recomienda la vacunación antineumocócica si no ha sido administrada con anterioridad. - Deberá mantener el aislamiento domiciliario hasta transcurridos un mínimo de 10 días de inicio de síntomas, siempre y cuando haya permanecido los últimos 3 días sin síntomas. - Se está estudiando la administración de plasma de paciente convalesciente como tratamiento de la infección por el coronavirus. Dado que usted ha pasado la enfermedad podría ser donante, si está interesado póngase en contacto con el Banc de sang i teixits a través del siguiente enlace: https://www.donarsang.gencat.cat/covid19 o contactando con el siguiente número de teléfono +50 88 078 68 49. 	<p>Le harán seguimiento sus médicos habituales.</p> <p>Se recomienda ponerse la vacuna de la gripe cada año.</p> <p>Si no la tiene puesta, también se aconseja vacunarse contra el neumococo.</p> <p>Debe seguir en casa sin salir al menos diez días desde que comenzaron los síntomas, y solo podrá dejar el aislamiento si lleva los últimos tres días sin síntomas.</p> <p>Se está estudiando usar plasma de personas que ya han pasado la infección como tratamiento para otros pacientes. Como usted ya ha pasado la enfermedad, podría ser donante. Si le interesa, puede contactar con el Banc de Sang i Teixits en donarsang.gencat.cat/covid19 o llamando al +50 88 078 68 49.</p>	<p>5.3.1.I. Uso de verbos en lugar de nominalizaciones</p> <p>5.3.1.II. Conversión de voz pasiva a activa</p> <p>5.3.1.VII. Clarificación de correferencias</p> <p>5.3.2.V. Paráfrasis o definición</p> <p>5.3.2.VI. Uso de sinónimos más comunes</p> <p>5.3.3.I. Números del 1 al 10 se escriben con letras</p>
CARMEN-IJA_ANTECEDENTES_6.txt	<p>Acudirá a CCEE de Traumatología (Paseo del Duque de Mandas 17) en 2 semanas en Sala de Curas, para control evolutivo (Dr. Callejas).</p> <p>Se cita en CCEE de Traumatología Calle de Pedro Manuel Collado 10, con Dr. Callejas en 1 mes, se adjuntan volantes.</p> <p>Se recomienda seguimiento con sus especialistas habituales</p>	<p>Tiene una cita en la consulta de Traumatología (Paseo del Duque de Mandas 17) dentro de dos semanas, en la Sala de Curas, con el Dr. Callejas, para revisión.</p> <p>También tiene otra cita con el Dr. Callejas en la consulta de Traumatología (Calle Pedro Manuel Collado 10) dentro de un mes. Se entregan los documentos necesarios.</p> <p>Se recomienda continuar el seguimiento con sus médicos de siempre.</p>	<p>5.3.1.I. Uso de verbos en lugar de nominalizaciones</p> <p>5.3.1.IV. Control del uso de conjunciones</p> <p>5.3.2.VI. Uso de sinónimos más comunes</p> <p>5.3.3.I. Números del 1 al 10 se escriben con letras</p> <p>5.3.3.V. Evitar abreviaciones no comunes</p>
CARMEN-IJA_EVOL_10.txt	<p>Se adjuntan medidas de aislamiento. El paciente podrá realizar el aislamiento donde considere necesario, sea en su vivienda habitual o en su segunda residencia. Por tanto se le autoriza a desplazarse hasta su segunda residencia en caso de que realice allí la cuarentena estipulada.</p> <p>Control por su CAP de referencia. Si dentro de un mes aproximadamente persiste con días de esfuerzo se puede valorar la necesidad de realizar un TC de tórax de forma ambulatoria.</p>	<p>Se le dan las medidas para hacer el aislamiento. Puede quedarse en su casa o en su segunda residencia.</p> <p>Debe seguir en contacto con su centro de salud.</p> <p>Si en un mes sigue con dificultad para respirar al hacer esfuerzo, podrían hacerle una prueba del pecho.</p>	<p>5.3.1.I. Uso de verbos en lugar de nominalizaciones</p> <p>5.3.1.II. Conversión de voz pasiva a activa</p> <p>5.3.1.VII. Clarificación de correferencias</p> <p>5.3.2.VI. Uso de sinónimos más comunes</p> <p>5.3.2.V. Paráfrasis o definición</p>
CARMEN-IJA_EXPLORACION_CLINICA_15.txt	<p>Ante cualquier sintomatología contactar con Salud Laboral.</p>	<p>Si tienes algún síntoma, deberías avisar a Salud Laboral.</p>	<p>5.3.1.I. Uso de verbos en lugar de nominalizaciones</p> <p>5.3.1.VII. Clarificación de correferencias</p> <p>5.3.2.VI. Uso de sinónimos más comunes</p>

Figura 32. Tabla recopilatoria de textos originales y simplificados, subconjunto de seguimiento médico

Nombre archivo	Texto original	Versión simplificada	Uso de solución	Versión simplificada iteración 2	Uso de solución iteración 2
CARMEN_LI_LA_PLAN_TERAPEUTICO_2.txt	Utilizar mascarilla quirúrgica en puesto de trabajo y en domicilio durante 1 semana.	Usar mascarilla en el trabajo y en casa por una semana.	5.3.1.I. Uso de verbos en lugar de nominalizaciones 5.3.2.VI. Uso de sinónimos más comunes 5.3.2.I. Números del 1 al 10 se escriben con letras		
CARMEN_LI_LA_PLAN_TERAPEUTICO_5.txt	- Masticar el sabor naranja 500 mg/400 comprimidos. 1 UND. Via oral. Una vez al día, en el desayuno. Mantener hasta visita con Neurología. - Prednisona 30 MG. Via oral. Una vez al día, en el desayuno. Mantener hasta visita con Neurología. RESTO VISITA.	Tomar Masticar O (500 mg/400 UI, sabor naranja), un comprimido al desayuno hasta la cita con Neurología. Tomar Prednisona 30 mg, una vez al día en el desayuno hasta la cita con Neurología. Lo demás, igual.	5.3.1.I. Uso de verbos en lugar de nominalizaciones 5.3.2.VI. Uso de sinónimos más comunes 5.3.2.II. Eliminación de palabras innecesarias 5.3.2.III. Números del 1 al 10 se escriben con letras 5.3.2.V. Parafrasas o definiciones		
CARMEN_LI_LA_PLAN_TERAPEUTICO_11.txt	Utilizar mascarilla quirúrgica en puesto de trabajo y en domicilio durante 1 semana. Seguir en aislamiento domiciliario hasta nuevas indicaciones por parte de Salud Laboral. Revisión de cita para PCR por parte de Salud Laboral.	Usar mascarilla en el trabajo y en casa por una semana. Seguir aislado en casa hasta que Salud Laboral diga lo contrario. Espere la cita para la PCR de Salud Laboral.	5.3.1.I. Uso de verbos en lugar de nominalizaciones 5.3.2.VI. Uso de sinónimos más comunes 5.3.1.VII. - Reformulación de regiones ambiguas 5.3.1.VIII. - Clarificación de concordancias 5.3.3.I. - Números del 1 al 10 se escriben con letras		
CARMEN_LI_LA_PLAN_TERAPEUTICO_13.txt	Origen: a demanda para mantener SatO2>90% Furosemida 40 mg/6B (V) Nitroglicerina 50 mg/10 Bisoprolol 5 mg 1-0-0.5 Etoricoxib 60 mg/6 Piracetam 20 mg/20 mgEq 1-0-0 Metformina 850 mg/6	Origen: a demanda, mantener saturación mayor a 90%. Furosemida 40 mg, via endovenosa: cada seis horas. Nitroglicerina 50 mg, via oral: una vez al día. Bisoprolol 5 mg: una por la mañana y una mitad por la noche. Etoricoxib 60 mg una vez al día. Piracetam 20 mg: una por la mañana. Metformina 850 mg: una vez al día.	5.3.2.II. - Expansión de abreviaturas y acrónimos 5.3.2.VI. - Parafrasas o definiciones 5.3.3.II. - Incluir ambos textos innecesarios 5.3.3.IV. - Homogeneidad de formatos		
CARMEN_LI_LA_PLAN_TERAPEUTICO_21.txt	Se adjunta medicación de base: - Metformina kern pharma 850 mg comprimidos recubiertos con película atg, 50 comprimidos, 100 MG. Via oral. Una vez al día, en la casa. - Co-dovon forte 160 mg/25 mg comprimidos recubiertos con película ef, 28 comprimidos, 1.25 UND. Via oral. Cada 24 horas. - Simvastatina kern pharma 20 mg comprimidos recubiertos con película ef, 28 comprimidos, 20 MG. Via oral. Cada 24 horas. - Fexipex (40 mg/14ml) solución oral, 20 viales bebibles de 15 ml, 40 MG. Via oral. Cada 24 horas. - Acido acetilsalicílico kern pharma 100 mg comprimidos gastroresistentes atg, 30 comprimidos, 100 MG. Via oral. Cada 24 horas.	Lista de medicamentos habituales: Metformina 850 mg: un comprimido por la noche. Co-Dovon forte 160/25 mg: un comprimido y medio al día. Simvastatina 20 mg: un comprimido al día. Fexipex (40 mg): en viales bebibles al día. Acido acetilsalicílico (aspirina) 100 mg: un comprimido al día.	5.3.2.III. - Eliminación de palabras innecesarias 5.3.2.VI. - Uso de sinónimos más comunes 5.3.2.V. - Parafrasas o definiciones 5.3.3.VIII. - Incluir la unidad siempre que se mencionen cantidades	Medicines que toma todos los días: Metformina 850 mg: tomar una pastilla por la noche. Co-Dovon forte 160/25 mg: tomar una pastilla y media al día. Simvastatina 20 mg: tomar una pastilla al día. Fexipex (40 mg): tomar un vial bebible al día. Acido acetilsalicílico 100 mg (aspirina): tomar una pastilla al día.	5.3.2.III. Eliminación de palabras innecesarias 5.3.2.VI. Uso de sinónimos más comunes 5.3.2.V. Parafrasas o definiciones 5.3.3.VIII. Incluir la unidad siempre que se mencionen cantidades 5.3.1.I. Uso de verbos en lugar de nominalizaciones 5.3.3.I. Números del 1 al 10 con letras 5.3.3.IV. Sustitución por hiperónimos

Figura 33. Tabla recopilatoria de textos originales y simplificados, subconjunto de plan terapéutico

Además de estas primeras columnas encontramos los resultados de la herramienta Legible, correspondiente a los índices Fernández-Huerta e INFLESZ y métricas como la edad mínima de comprensión, nivel educativo y tiempo estimado de lectura. Además, también de notificar descripciones adicionales de los textos como el listado de palabras no reconocidas por la herramienta, signos de puntuación usados y frecuencia de cada uno de ellos. Todos estos resultados los anotamos tanto para el texto original como los textos simplificados.

Nombre	Índices legibilidad				Métricas				Índices legibilidad				Métricas					
	Fernández Huerta	Nivel Fernández Huerta	INFLESZ	Nivel INFLESZ	Palabras no reconocidas	Signos de puntuación	edad mínima entendimiento (años)	Nivel de grado educativo (años)	Tiempo estimado de lectura (minutos)	Fernández Huerta spif.	Nivel Fernández Huerta spif.	INFLESZ spif.	Nivel INFLESZ spif.	Palabras no reconocidas spif.	Signos de puntuación spif.	edad mínima entendimiento (años) spif.	Nivel de grado educativo (años) spif.	Tiempo estimado de lectura (minutos) spif.
CARMEN_LI_LA_ANTECEDENTES_1.txt	56.73	Algo difícil	53.94	Algo difícil	1enalegra 2empesamiento 3cap	1Comas (J4) 2Oviones (J4) 3Puntos (J3)	10.31	5.80	0.20	91.89	Muy fácil	87.79	Muy fácil	cap	1Comas (J3) 2Puntos (J3) 3Parentesis (y J2)	6.50	2.90	0.30
CARMEN_LI_LA_ANTECEDENTES_4.txt	60.39	Normal	55.09	Normal	1antineumococica 2aislamiento 3sintomas 4conalescienta 5coronavirus 6pongase 7banc 8sang 9ovitas 10tups 11tense 12donarsang 13gnicat 14cat 15covid	1Puntos (J8) 2Oviones (J5) 3Barras diagonales (J5) 4Comas (J2) 5Dos puntos (J2)	10.15	5.80	0.60	76.72	Algo fácil	72.26	Bastante fácil	1neumococo 2aislamiento 3banc 4sang 5ovitas 6donarsang 7gnicat 8cat 9covid	1Puntos (J8) 2Comas (J4) 3Barras diagonales (J1)	8.45	4.50	0.60
CARMEN_LI_LA_ANTECEDENTES_6.txt	69.44	Normal	64.43	Normal	1ccoe 2callajes 3colleido 4edjstan	1Puntos (J4) 2Parentesis (J4) 3Comas (J3)	8.51	3.80	0.20	82.70	Fácil	78.36	Bastante fácil	1callajes 2colleido	1Puntos (J6) 2Parentesis (y J4) 3Comas (J3)	7.07	2.70	0.30
CARMEN_LI_LA_EVOL_10.txt	62.09	Normal	57.29	Normal	1aidjstan 2aislamiento 3acuerntena 4cap 5diana 6tc	1Puntos (J5) 2Comas (J1)	10.23	5.70	0.40	66.46	Fácil	62.52	Muy fácil	1aislamiento	1Puntos (J4) 2Comas (J1)	7.14	3.40	0.20
CARMEN_LI_LA_EXPLORACION_CLINICA_15.txt	31.67	Difícil	25.34	Muy difícil		1Puntos (J1)	12.26	4.70	0.00	62.45	Normal	56.97	Normal		1Comas (J1) 2Puntos (J1)	9.07	3.50	0.00

Figura 34. Tabla recopilatoria de resultados índices y métricas, subconjunto de seguimiento médico

Nombre archivo	TEXTO ORIGINAL								TEXTO SIMPLIFICADO (Iteración 1)								TEXTO SIMPLIFICADO (Iteración 2)										
	Índices legibilidad				Métricas				Índices legibilidad				Métricas				Índices legibilidad				Métricas						
	Nivel Fernández Huerta	Nivel INFLESEZ	Nivel INFLESEZ	Palabras no reconocidas	Signos de puntuación	edad mínima	Nivel de comprensión (Lefk)	Tiempo estimado de lectura	Fernández Huerta diff1	Nivel Fernández Huerta diff1	Nivel INFLESEZ diff1	Nivel INFLESEZ diff1	Palabras no reconocidas	Signos de puntuación	edad mínima	Nivel de comprensión	Tiempo estimado de lectura	Fernández Huerta diff2	Nivel Fernández Huerta diff2	Nivel INFLESEZ diff2	Nivel INFLESEZ diff2	Palabras no reconocidas	Signos de puntuación	edad mínima	Nivel de comprensión	Tiempo estimado de lectura	
CARMEN-IA_PLAN_TERAPEUTICO_2.txt	45.92	Difícil	39.89	Muy difícil	1Puntos (1)	11.07	5.8	0.1	86.63	Facil	82.40	Muy facil	1Puntos (1)	6.6	2.2	0.1	86.63	Facil	82.40	Muy facil	0.00	1Puntos (1)	6.60	2.20	0.10		
CARMEN-IA_PLAN_TERAPEUTICO_3.txt	70.25	Algo facil	65.29	Bastante facil	1maestical 2d 3mg 4c 5und 6neurologia 7prediccion	1Puntos (10) 2Comas (2) 3Circunflexo (2) 4Barras diagonales (1)	8.15	2.20	0.20	79.09	Algo facil	74.62	Bastante facil	1maestical 2d 3mg 4c 5und 6neurologia 7prediccion	1Comas (4) 2Puntos (3) 3Parentesis (1) 4Barras diagonales (1)	7.52	3	0.2	79.09	Algo facil	74.62	Bastante facil	1maestical 2d 3mg 4Barras diagonales (1) 5neurologia 6prediccion	1Comas (4) 2Puntos (3) 3Parentesis (1) 4Barras diagonales (1)	7.62	3	0.2
CARMEN-IA_PLAN_TERAPEUTICO_11.txt	58.97	Algo difícil	53.58	Algo difícil	1abastamiento 2por	1Puntos (3)	10.81	5.60	0.20	84.94	Facil	80.51	Muy facil	1por	1Puntos (3)	7.1	3.3	0.2	84.94	Facil	80.51	Muy facil	1por	1Puntos (3)	7.10	3.30	0.20
CARMEN-IA_PLAN_TERAPEUTICO_13.txt	16.7	Muy difícil	10.26	Muy difícil	1talo 22 3ruosemida 4c 5h 6ev 7hidroclorotiazida 8vo 9neurologia 10metformina 11am 12pharma 13mg 14g 15g 16g 17co 18dolan 19forte	10Juncos (4) 2Punto y coma (1) 3Comas (2) 4Barras diagonales (1)	13.75	9.50	0.20	68.43	Normal	63.52	Normal	1g 2ruosemida 3mg 4hidroclorotiazida 5pila 6Sotoprolol 7edotaban 8Tasapirico 9man	1Puntos (17) 2Dos puntos (17) 3Comas (3) 4Punto y coma (1)	8.39	1.9	0.3	68.43	Normal	63.52	Normal	1g 2ruosemida 3mg 4hidroclorotiazida 5pila 6Sotoprolol 7edotaban 8Tasapirico 9man	1Puntos (17) 2Dos puntos (17) 3Comas (3) 4Punto y coma (1)	8.39	1.9	0.3
CARMEN-IA_PLAN_TERAPEUTICO_21.txt	60.69	Normal	55.07	Normal	1Comprimidos 2und 3metformina 414 514 614 714 814 914 1014 1114 1214 1314 1414 1514 1614 1714 1814 1914 2014 2114	1Puntos (20) 2Comas (6) 3Circunflexo (6) 4Parentesis (1) 5Barras diagonales (2) 6Dos puntos (1)	9.26	3.60	0.50	67.75	Normal	62.74	Normal	1metformina 2mg 3co 4dolan 5forte 6y4 7metformina 8Barras diagonales 9Barras diagonales 10Barras diagonales 11Barras diagonales	1Dos puntos (1) 2Comas (3) 3Comas (3) 4Punto y coma (1)	8.52	3.1	0.2	75.17	Algo facil	70.22	Bastante facil	1metformina 2mg 3co 4dolan 5forte 6y4 7metformina 8Barras diagonales 9Barras diagonales 10Barras diagonales 11Barras diagonales	1Dos puntos (1) 2Comas (3) 3Comas (3) 4Punto y coma (1)	7.60	1.60	0.30

Figura 35. Tabla recopilatoria de resultados índices y métricas, subconjunto de plan terapéutico

Al final de la tabla, se reservan columnas para anotar las diferencias entre los índices recogidos, siendo comparativas entre los resultados de los textos simplificados contra los textos originales, de este modo podemos observar el impacto que han tenido las simplificaciones y los cambios realizados en cada texto.

Name	DIFERENCIA							
	Iteración 1				Iteración 2			
	Fernandez Huerta diff.	Nivel Fernández Huerta diff.	INFLESEZ diff.	Nivel INFLESEZ diff.	Fernandez Huerta diff.2	Nivel Fernández Huerta diff2	INFLESEZ diff2	Nivel INFLESEZ diff2
CARMEN-IA_ANTECEDENTES_1.txt	33.16	Algo difícil -> Muy fácil	33.85	Algo difícil -> Muy fácil	33.16	Algo difícil -> Muy fácil	33.85	Algo difícil -> Muy fácil
CARMEN-IA_ANTECEDENTES_4.txt	16.33	Normal -> Algo fácil	17.17	Normal -> Bastante fácil	16.33	Normal -> Algo fácil	17.17	Normal -> Bastante fácil
CARMEN-IA_ANTECEDENTES_6.txt	13.26	Normal -> Fácil	13.93	Normal -> Bastante fácil	13.26	Normal -> Fácil	13.93	Normal -> Bastante fácil
CARMEN-IA_EVOL_10.txt	24.37	Normal -> Fácil	25.24	Normal -> Muy fácil	24.37	Normal -> Fácil	25.24	Normal -> Muy fácil
CARMEN-IA_EXPLORACION_CLINICA_15.txt	30.78	Difícil -> Normal	31.63	Muy difícil -> Normal	30.78	Difícil -> Normal	31.63	Muy difícil -> Normal

Figura 36. Tabla comparativa de resultados, subconjunto de seguimiento médico

Nombre archivo	DIFERENCIA							
	Iteración 1				Iteración 2			
	Fernández Huerta diff.	Nivel Fernández Huerta diff	INFLESEZ diff.	Nivel INFLESEZ diff	Fernández Huerta diff2	Nivel Fernández Huerta diff2	INFLESEZ diff2	Nivel INFLESEZ diff2
CARMEN-IA_PLAN_TERAPEUTICO_2.txt	40.71	Difícil -> Fácil	42.51	Muy difícil -> Muy fácil	40.71	Difícil -> Fácil	42.51	Muy difícil -> Muy fácil
CARMEN-IA_PLAN_TERAPEUTICO_5.txt	8.84	Algo fácil -> Algo fácil	9.33	Bastante fácil -> Bastante fácil	8.84	Algo fácil -> Algo fácil	9.33	Bastante fácil -> Bastante fácil
CARMEN-IA_PLAN_TERAPEUTICO_11.txt	25.97	Algo difícil -> Fácil	26.93	Algo difícil -> Muy fácil	25.97	Algo difícil -> Fácil	26.93	Algo difícil -> Muy fácil
CARMEN-IA_PLAN_TERAPEUTICO_13.txt	51.73	Muy difícil -> Normal	53.26	Muy difícil -> Normal	51.73	Muy difícil -> Normal	53.26	Muy difícil -> Normal
CARMEN-IA_PLAN_TERAPEUTICO_21.txt	7.06	Normal -> Normal	7.67	Normal -> Normal	14.48	Normal -> Algo fácil	15.15	Normal -> Bastante fácil

Figura 37. Tabla comparativa de resultados, subconjunto de plan terapéutico