Propuesta recibida en febrero 2026 y aceptada para publicación en agosto 2026.

El lenguaje claro como herramienta para la preedición en traducción automática

Plain language as a pre-editing tool in machine translation

José Manuel Bustos Gisbert
Universidad de Salamanca

Andrea Fernández Vivanco
Universidad de Salamanca

Resumen

El presente trabajo plantea la adopción del lenguaje claro como estrategia de preedición orientada a mejorar la calidad de las pretraducciones generadas por sistemas de traducción automática neuronal. Se parte de la premisa de que la claridad textual no constituye una propiedad absoluta, sino contextual, determinada por el género discursivo y por el perfil del destinatario. A partir de una revisión bibliográfica reciente se identifican recomendaciones de lenguaje claro en cuatro niveles de intervención (discursivo, ortotipográfico, morfosintáctico y léxico-semántico), que se evalúan atendiendo a su pertinencia para la preedición con fines traductológicos y a su compatibilidad con la preservación de la naturalidad lingüística y de las convenciones de género del texto origen. Como resultado, se formula una propuesta de 20 normas, ordenadas según su secuencia de aplicación y definidas mediante parámetros que permiten ajustar su alcance al contexto comunicativo. Se sostiene que la aplicación de estas normas podrá incrementar la interpretabilidad del texto por parte del sistema de traducción y reducir el esfuerzo de posedición, al desplazar parte de la carga cognitiva hacia una revisión intralingüística en la lengua de partida.

Palabras clave

lenguaje claro, traducción automática, preedición, traducibilidad, lecturabilidad, lenguajes controlados

Abstract

This paper examines plain language as a framework for pre-editing intended to enhance the quality of draft translations generated by neural machine translation systems. It is based on the premise that textual clarity is not an absolute property but a contextual one, determined by discourse genre and the profile of the intended audience. Based on a recent literature review, plain-language recommendations are identified across four levels of intervention (discursive, orthotypographic, morphosyntactic, and lexico-semantic). Guidelines are assessed in terms of relevance for translation-oriented pre-editing, and compatibility with the preservation of the source text's fluency and genre conventions. The study results in a proposal comprising 20 rules, ordered according to their sequence of application and defined through parameters that allow their scope to be adjusted to the communicative context. It is argued that the implementation of these rules can enhance the interpretability of the text for the translation system and reduce post-editing effort by shifting part of the cognitive load to intralingual revision in the source language.

Keywords

plain language, machine translation, pre-editing, translatability, readability, controlled languages

1. Introducción

La investigación que se presenta se inserta en un proyecto más amplio que persigue valorar la aplicación de los principios del denominado lenguaje claro (LCL) a las labores de preedición propias de la traducción automática. Entendemos el LCL como el uso de una formulación lingüística que facilite el acceso del lector a la comprensión del texto. Continúa el trabajo iniciado en Fernández Vivanco & Bustos Gisbert (2025), cuyo objetivo fundamental era evaluar la validez actual de la aplicación de los lenguajes controlados como estrategia de preedición para mejorar la calidad de una traducción automática neuronal (TAN). Tras analizar el rendimiento de las normas aplicadas a un texto sometido a un motor de traducción, concluimos que tal vez era posible mejorar el resultado proponiendo reglas asociadas no tanto al lenguaje controlado como al lenguaje claro. Dedicaremos las páginas siguientes a sentar las bases conceptuales sobre las que llevar a cabo ese salto cualitativo. Sin embargo, antes parece oportuno adelantar algunas reflexiones acerca del concepto mismo del LCL, así como sobre su alcance en el ámbito de la traducción.

En primer lugar, no se ha de perder de vista que el LCL nace como una exigencia global en la comunicación entre la administración pública y los ciudadanos, reivindicación que ha alcanza-

do su máximo exponente en el ámbito jurídico. Como bien apuntan Montolío Durán & Tascón (2020), los ciudadanos tienen el derecho a entender aquello que la administración del estado tiene a bien hacerles saber, sea cual sea el contexto en el que esa interacción se produce. No obstante, se debe subrayar que se observa cierta renuencia a la aplicación de un lenguaje claro y, en última instancia, democratizador, que se explica desde una percepción de la comunicación como una forma de dominio. Según Van Dijk (2000), el poder organiza y condiciona muchas de las relaciones que se establecen entre el discurso y la sociedad. Dominar el primero se convierte en una vía de ejercer el poder y también de preservarlo.

En segundo término, el LCL no es un concepto absoluto, pues se ha de valorar considerando las coordenadas en las que se aplica, entendidas desde al menos dos variables fundamentales, como son el género textual empleado y el perfil del destinatario previsto. En relación con los géneros, suscribimos la percepción de García Izquierdo (2012), quien los concibe como constructos de interacción dinámicos y solo relativamente estables[1]. Se definen en términos de su propósito social: cada género maneja el lenguaje de una determinada forma según su cometido. Se construye, pues, una convención textual y la responsabilidad de su diseño recae en la comunidad discursiva que lo emplea; así, son los usuarios del género quienes acaban por darle forma. Sin embargo, es importante subrayar que las comunidades discursivas no son homogéneas en el sentido de que sus miembros no siempre desempeñan funciones idénticas: habrá géneros en los que todos pueden acceder a ellos no solo como destinatarios, sino también como generadores de textos, pues su domino del ámbito temático y de las convenciones expresivas es similar y suficiente. Pero también habrá otros en los que una parte de la comunidad solo actúa como destinataria y con un dominio limitado del ámbito del conocimiento y de las convenciones. En este caso, estamos ante *comunidades en desequilibrio discursivo*. Desde la posición argumentada por los defensores del lenguaje claro es en estos casos en los que se hace necesaria una adaptación textual.

En tercer lugar, no cabe hablar sin más de normas de lecturabilidad cuya aplicación tiene como resultado que el texto producido sea necesariamente más claro, pues esta cualidad no se puede interpretar de manera ajena al acto de comunicación y se mide, como no puede ser de otro modo, en términos de la recepción, desde la cual se valora tanto la inteligibilidad de los conceptos como la intención que lleva al acto de escritura. Entendemos, pues, que no existen reglas absolutas conducentes al LCL, como tampoco hay modelos universales para calibrarlo. Tanto la nómina de reglas de LCL que se quiera proponer como su peso y su importancia cualitativa han de asociarse necesariamente a un contexto de comunicación y deben tener en cuenta el género textual sobre el que se apliquen. Del mismo modo, se ha de partir del convencimiento de que su implementación difícilmente será mecánica. Fundamentalmente porque el concepto de género es esencialmente dinámico y no se define en términos canónicos, sino de uso.

Argüíamos más arriba que la aplicación de normas LCL será tanto más pertinente cuanto mayor sea el desequilibrio en el interior de la comunidad discursiva que se enfrenta a un texto determinado. Este se consigna, al menos, en tres clases de géneros mediante los cuales se establece una relación entre agente(s) especializado(s) y destinatario(s) lego(s) o de conocimiento limitado tanto de la materia de la que se trate como de las convenciones discursivas que el primero ha creado de forma normalmente unilateral. Ente ellos destacan aquellos con los que se establece algún tipo de relación legal, administrativa, sanitaria o contractual con efectos directos e inmediatos. Asimismo, los que reúnen textos de naturaleza divulgativa destinados a informar a destinatarios no especializados acerca de cualquier avance en el conocimiento. Y, finalmente, aquellos que persiguen informar de la actualidad nacional e internacional de manera fiel, veraz y accesible a destinatarios que pueden conocerla solo parcialmente. Esta investigación se vincula a un género correspondiente a los del tercer tipo, como es el caso de la noticia periodística.

Sin embargo, nuestro propósito no es mejorar la claridad de los textos, sino la traducción en bruto y sin poseditar ni revisar que obtenemos de un motor de traducción automática o una herramienta de IA a partir de un texto original que podemos encuadrar en uno de los tres grupos indicados. Para ello, partimos de la hipótesis confirmada en Fernández Vivanco & Bustos Gisbert (2025) de que la preedición mejora globalmente la TA. Sin embargo, es necesario perfilar las normas de preedición empleadas hasta ahora, que tienen su origen en los llamados *lenguajes controlados*. Estas muestran dos insuficiencias de calado. La primera es que su rendimiento es relativo y solo algunas tienen impacto real en la

[1] Para la definición del género textual y de sus características se siguen en lo esencial las propuestas del grupo GENTT, de la Universidad Jaume I de Castellón, especialmente las recogidas en García Izquierdo (2011, 2012) y Ezpeleta Piorno & Borja Albi (2017).

calidad de las traducciones automáticas. La segunda es que su aplicación afecta estilísticamente al texto origen en lo relativo a las convenciones de género. Esta cuestión no es menor, puesto que puede tener efectos no deseados en la funcionalidad de los motores al condicionar su entrenamiento con muestras de lengua de dudosa validez idiomática. Consideramos que la preedición puede mejorar cualitativamente si se aplican normas propias del LCL por dos motivos fundamentales: primero porque los textos intervenidos con ellas no pierden su naturalidad lingüística y también discursiva en la lengua original; después porque, siendo su objetivo general aumentar la claridad y la lecturabilidad de los textos, esa mejora también facilitará, como sostiene Bowker (2024), su interpretación por parte de un motor de TA, especialmente si, como debería ser, está entrenado con textos de calidad y de lecturabilidad alta en un contexto de comunicación dado.

2. Hipótesis y metodología aplicada

Esta investigación se asienta en la convicción de que incrementar el trabajo de preedición del texto expuesto a TAN se traducirá en un descenso en el esfuerzo de posedición. Esa inversión en el foco de la actividad tendrá como resultado una reducción en el esfuerzo cognitivo, dado que este será fundamentalmente intralingüístico y se llevará a cabo exclusivamente en la lengua origen. Desgraciadamente, la extensión de este trabajo impide dar respuesta completa a toda la hipótesis fijada, por lo que estableceremos solo dos objetivos: decidir las reglas de lectura clara que aplicar en el proceso de preedición y dar una definición consistente de las mismas y de su modo de aplicación.

El trabajo se ha organizado en dos fases. La primera de ellas está vinculada a las fuentes primarias de las que parte la investigación. En ella se lleva a cabo una justificación del modo en el que se han seleccionado, se analizan las propuestas de cada uno de ellas y se propone una síntesis. Nos ocuparemos de extraer las reglas sobre las que la investigación reciente parece coincidir a la hora de fijar cómo ha de ser el LCL en lengua española. Es importante subrayar que una norma de LCL se diseña para intervenir sobre textos que, siendo correctos en su expresión, pueden mejorar en términos de lecturabilidad; por ese motivo, excluiremos del análisis cualquier norma que no asuma ese presupuesto. A continuación, se procederá a una evaluación crítica de la lista de normas con el objeto de calibrar su grado de adecuación a la hipótesis enunciada más arriba. A tal efecto, será necesario definirlas con precisión y describir los parámetros con los que se fijará su peso en cada contexto comunicativo y discursivo. Toda la labor se realizará para un contexto de traducción que involucra a las lenguas inglesa y española en una direccionalidad ES > EN. A ellos dedicaremos el epígrafe 3 del trabajo.

En la segunda fase se propone una definición de las reglas y se articula un procedimiento destinado a ordenar su uso. La propuesta se asienta en una experiencia piloto que ha supuesto la aplicación de las normas a cuatro textos periodísticos completos, si bien en el caso del primero de ellos se emplearon dos veces de manera independiente. Son los siguientes:

- Manuel Ansede: "Revolución en la medicina: una molécula producida por bacterias del intestino causa la aterosclerosis, culpable de millones de muertes". El País, 16 de julio de 2025[2].
- Karelia Vázquez: "He estado en Mallorca en agosto y no he visto (casi) a nadie". El País, 24 de agosto de 2025[3].
- EFE Avilés: "La llegada de un tiburón foca a una playa de Asturias genera una gran expectación científica". El Mundo, 16 de enero de 2026[4].
- Ricardo F. Colmenero: "La captura de un tiburón blanco en aguas de Alicante confirma su presencia esporádica en España". El Mundo, 10 de febrero de 2026[5].

Esta fase será el objeto del epígrafe 4 y constituye la sección nuclear de nuestra investigación.

3. Análisis de las fuentes primarias

3.1. Selección de fuentes

Para el presente trabajo, se han considerado 9 fuentes, que presentamos en el cuadro 1.

A la hora de seleccionarlas, se ha partido de dos criterios fundamentales: su actualidad y su representatividad. En relación con la primera, se

[2] https://elpais.com/salud-y-bienestar/2025-07-16/revolucion-en-la-medicina-una-molecula-producida-por-bacterias-del-intestino-causa-la-aterosclerosis-culpable-de-millones-de-muertes.html

[3] https://elpais.com/elviajero/escapadas/espana/2025-08-24/he-estado-en-mallorca-en-agosto-y-no-he-visto-casi-a-nadie.html

[4] https://www.elmundo.es/ciencia-y-salud/ciencia/2026/01/16/6969f70ce85ece6e668b4574.html

[5] https://www.elmundo.es/ciencia-y-salud/ciencia/2026/02/10/698b1ecde85ecef2578b45b8.html

Nº	Autoría y título
F1	Ministerio de Justicia (2011). *Informe de la Comisión de Modernización del Lenguaje Jurídico*. Madrid: Ministerio de Justicia.
F2	Montolío Durán, E. y Tascón, M. (2020). *El derecho a entender*. Madrid: Prodigioso Volcán/Catarata.
F3	Da Cunha, I. y Escobar, M. Á. (2021). Recomendaciones sobre lenguaje claro en español en el ámbito jurídico-administrativo: análisis y clasificación. *Pragmalingüística*, (29), 129–148.
F4	Romanik, W.; Piekot, T.; Tymiński, R. (2022). Adaptation of informed consent for anaesthesia to plain language standard. *NOWOTWORY J Oncol*, 72, 265–273.
F5	Estopà, R. (2021). Culturalización en salud: aportaciones desde la Lingüística. *Cadernos de Linguística*, 2(1), 01–22.
F6	Da Cunha, I. (2024). Inteligencia artificial y lenguaje claro en español. En A. Retegui (ed.), *Lenguaje claro en Iberoamérica. Principios y prácticas* (pp. 417–444). Thomson Reuters.
F7	Briva-Iglesias, V. y Peñuelas Gil, I. (2025). Simplifying healthcare communication: Evaluating AI-driven plain language editing of informed consent forms. En *Proceedings of the 1st Workshop on Artificial Intelligence and Easy and Plain Language in Institutional Contexts*, 55–65.
F8	Fernández Vivanco, A. y Sanfilippo, V. (2025). La lecturabilidad de textos escritos en lengua española: revisión crítica de métricas y propuesta de nuevas variables. *ELUA*, 44, 29–56.
F9	Bowker, L. (2024). Plain Language in the Age of Neural Machine Translation: An Opportunity for Translators. En Y. Peng, H. Huang y D. Li (eds.), *New Advances in Translation Technology*. Springer, Singapore.

Cuadro 1: Fuentes analizadas para la extracción de reglas de LCL.

ha optado por propuestas realizadas desde el año 2020; no obstante, se ha incluido una anterior por su máxima relevancia en lo relativo al estudio del lenguaje claro en lengua española en el ámbito de la comunicación jurídica y administrativa (F1), hasta el punto de que aparece recogida en prácticamente todos los trabajos posteriores dedicados al lenguaje claro en lengua española. Junto a F2, suma dos hitos esenciales en el estudio del LCL en España[6].

[6]F3 no plantea estrictamente recomendaciones originales, sino que recopila de diversas fuentes. No obstante, dado que el criterio de frecuencia no se ha tenido en cuenta para la selección de las reglas, no parece que su elección pueda generar sesgos en la lista final.

En lo que tiene que ver con la segunda, dos fuentes (F4 y F5) se han incluido por evaluar el LCL en un contexto menos frecuente, como es el sociosanitario, en concreto en el relativo a los consentimientos informados; por su parte dos textos (F6 y F7) evalúan las posibilidades del aprovechamiento de la IA generativa en la adaptación de los textos al lenguaje claro. En ambos casos, al ser ámbitos de aplicación más novedosos y con menos referencias disponibles se ha seleccionado una fuente referida al inglés y otra al español[7].

Por su parte, F8 ha sido elegida en tanto que realiza una revisión exhaustiva y muy actualizada referida a propuestas de lecturabilidad en los ámbitos del lenguaje controlado y el lenguaje claro. Finalmente, se ha incluido F9 porque es el único caso en el que nos consta que se evalúan las posibilidades de la aplicación del LCL a la traducción automática. Esta aportación, de Bowker (2024), será tomada en cuenta dos veces, pues informa de propuestas de lecturabilidad que extrae de dos fuentes independientes: la primera incluye pautas para preparar resúmenes de artículos científicos en lenguaje claro para revistas académicas propuestas por dos editoriales (Taylor & Francis y SAGE) y la segunda alude a indicaciones para escribir textos que se van a traducir mediante traducción automática neuronal propuestas por la Dirección General de Traducción (Unión Europea). Contaremos, por tanto, con nueve referencias bibliográficas y diez formulaciones de reglas.

3.2. Cuantificación de las normas extraídas

Pese a la diversidad de objetivos, los trabajos seleccionados establecen de manera más o menos explícita una serie de normas cuya aplicación garantizaría un aumento en el grado de lecturabilidad de los textos. Hemos identificado 58, que hemos clasificado en cuatro grandes grupos: las referidas al nivel discursivo de la construcción de los textos (cuadro 2), las relativas a cuestiones ortográficas (cuadro 3), las de contenido morfosintáctico (cuadro 4) y las de orientación léxica (cuadro 5). En dichas tablas se propone para cada regla una redacción provisional y de compromiso que quiere sintetizar las formulaciones aportadas por los distintos autores. Para cada caso se indica también en cuántas fuentes aparece cada regla. Todas las normas identificadas han sido incluidas en la evaluación realizada en la sección siguiente de este trabajo.

[7]Para valorar la cuestión referida a las similitudes y las diferencias detectables entre el LCL en español e inglés, véase Escobar & Samaniego (2022).

	Reglas	F1	F2	F3	F4	F5	F6	F7	F8	F9a	F9b	Total
R.01	Evitar elipsis y anáforas que puedan se ambiguas.	•		•	•	•			•		•	6
R.02	Usar conectores.	•	•	•		•	•				•	6
R.03	Evitar enunciados largos.	•	•	•	•				•	•	•	7
R.04	Evitar incisos extensos.	•							•			2
R.05	Medir la longitud de frase en palabras.						•	•	•		•	4
R.06	Medir la longitud del texto en frases.								•			1
R.07	Evitar enunciados muy largos.	•	•	•	•				•	•	•	7
R.08	Usar párrafo breve.	•	•	•								3
R.09	Usar párrafo monotemático.	•	•	•							•	4
R.10	Proponer un orden lógico de párrafos.	•										1
R.11	Evitar párrafos con un solo enunciado.	•										1
R.12	Cuidar la estructura interna del párrafo.		•									1

Cuadro 2: Reglas de LCL extraídas de las nueve fuentes analizadas: nivel discursivo.

	Reglas	F1	F2	F3	F4	F5	F6	F7	F8	F9a	F9b	Total
R.13	Vigilar los errores ortográficos.	•				•			•	•	•	5
R.14	Vigilar los errores de puntuación.	•				•	•			•	•	5
R.15	Evitar siglas, acrónimos y abreviaturas.		•			•			•	•	•	5
R.16	Evitar paréntesis.								•			1
R.17	Evitar corchetes y signos poco habituales.								•	•	•	3
R.18	Comprobar la expresión de cifras.								•	•		2
R.19	Destacar tipográficamente las enumeraciones largas.	•	•		•							3
R.20	Evitar líneas en blanco innecesarias.								•			1
R.21	Editar el texto de tal manera que el paciente pueda pasar por él rápidamente.				•							1

Cuadro 3: Reglas de LCL extraídas de las nueve fuentes analizadas: nivel ortográfico.

	Reglas	F1	F2	F3	F4	F5	F6	F7	F8	F9a	F9b	Total
R.22	Usar un orden natural de los elementos sintácticos.		•	•		•			•		•	5
R.23	Reducir la distancia entre sujeto y verbo.					•			•			2
R.24	Reducir el número de subordinadas por frase.	•		•			•		•		•	5
R.25	Evitar perífrasis verbales.								•			1
R.26	Evitar tiempos verbales compuestos.								•			1
R.27	Evitar en una oración dos SPrep con la misma preposición.								•			1
R.28	Evitar el uso del gerundio.	•	•	•								3
R.29	Vigilar la falta de sistematicidad en el uso de la persona y el tiempo verbal.	•							•			2
R.30	Escapar de formas negativas, giros negativos, dobles negaciones, etc.		•	•					•			3
R.31	Evitar el uso del subjuntivo y el condicional.	•							•			2
R.32	Evitar pasivas sin agente y reflejas.	•	•	•		•	•		•	•		7
R.33	Evitar oraciones impersonales.				•	•			•	•		4
R.34	Evitar los adverbios en -mente.								•			1
R.35	Evitar nominalizaciones.								•			1
R.36	Evitar los relativos poco usuales.	•										1
R.37	Evitar superlativos.									•		1
R.38	Evitar las cadenas de dos o más sustantivos.				•					•	•	3
R.39	Respeta las normas de la lengua.		•									1

Cuadro 4: Reglas de LCL extraídas de las nueve fuentes analizadas: nivel sintáctico.

	Reglas	F1	F2	F3	F4	F5	F6	F7	F8	F9a	F9b	Total
R.40	Medir la longitud de palabra en letras.								•			1
R.41	Medir la longitud de palabra en sílabas.							•	•			2
R.42	Medir la densidad de palabras raras o no comunes.						•	•	•			3
R.43	Calcular las palabras de más de tres sílabas.							•	•			2
R.44	No usar diferentes términos para un mismo referente.								•		•	2
R.45	Evitar palabras de significado vacío.							•	•		•	3
R.46	Evitar palabras o conceptos abstractos.			•					•			2
R.47	Evitar palabras polisémicas.								•			1
R.48	Evitar el uso de tecnicismos, jergas y xenismos.	•	•	•	•	•	•		•	•		8
R.49	Evitar extranjerismos.								•			1
R.50	Evitar falsos amigos.								•			1
R.51	Evitar modismos, expresiones locales y proverbios.								•	•	•	3
R.52	Incluir definiciones e información contextual.								•			1
R.53	Evitar el lenguaje figurado y las metáforas.								•	•		2
R.54	Explicar el significado de los términos si es necesario utilizarlos.	•		•								2
R.55	Eliminar palabras innecesarias.			•								1
R.56	Evitar arcaísmos.			•								1
R.57	No usar palabras fuera del diccionario.						•					1
R.58	Usar las palabras más adecuadas al sentido y el tono.		•									1

Cuadro 5: Reglas de LCL extraídas de las nueve fuentes analizadas: nivel léxico.

Parece oportuno recordar que todas las reglas de lenguaje claro que en Fernández Vivanco & Bustos Gisbert (2025) demostraron rendimiento elevado como herramientas de preedición para textos tratados por un motor de TAN están incluidas en la nómina de reglas aplicadas para LCL.

3.3. Evaluación de las normas

La nómina de las reglas recogidas en los cuadros 2-5 no es consistente en términos taxonómicos, pero da una idea cabal de la amplitud de ámbitos que se han visitado en los estudios de lecturabilidad de los textos escritos. En la construcción de la propuesta final de normas, renunciaremos a todas aquellas que proponen métricas objetivas de lecturabilidad, dado que no pueden indicar por sí mismas la complejidad de lectura y comprensión de un texto, como han argumentado Bailin & Grafstein (2016) o Porras Garzón & Estopà Bagot (2020), entre otros. Asimismo, descartaremos algunas que carecen de rendimiento en nuestro contexto de aplicación. Dado que la formulación de las reglas exhibe en algunos casos variaciones según las fuentes que las dictan, propondremos una síntesis que permita aplicarlas de forma más eficiente.

3.3.1. Nivel discursivo

El análisis ha permitido extraer 12 reglas: dos tienen que ver con la cohesión, cinco con la construcción de los enunciados y otras cinco con la de los párrafos; aparecen recogidas en el cuadro 6[8]. En primer lugar, desechamos las reglas de índole cuantitativa (R.05 y R.06) y descartamos R.08 y R.10 por no ser de aplicación en el contexto de la traducción automática neuronal, toda vez que esta usa como unidad de traducción el enunciado.

Asimismo, también renunciamos a R.09 y R.12, dado que es difícil contar con argumentos que permitan valorar de manera objetiva lo referido a la organización interna y a la unidad temática de los párrafos. En ello seguimos a Bustos Gisbert (2012, p. 84), que ha comprobado que el párrafo no se construye conforme a criterios fijos y estables, por lo que cualquier caracterización suele chocar con la realidad del uso.

Finalmente, fundimos en una sola regla R.03 y R.07, puesto que hacen referencia a la misma cuestión, como es la extensión máxima aconsejable para los enunciados textuales.

[8] Indicamos la codificación alfanumérica de la regla, una formulación de compromiso y el número de fuentes en las que se consigna.

Nº	Reglas	Total
R.01	Evitar elipsis y anáforas que puedan ser ambiguas.	6
R.02	Usar conectores.	6
R.03	Evitar enunciados largos.	7
R.04	Evitar incisos extensos.	2
R.05	Medir la longitud de frase en palabras.	3
R.06	Medir la longitud del texto en frases.	1
R.07	Evitar enunciados muy largos.	7
R.08	Usar párrafo breve.	3
R.09	Usar párrafo monotemático.	4
R.10	Proponer un orden lógico de párrafos.	1
R.11	Evitar párrafos con un solo enunciado.	1
R.12	Cuidar la estructura interna del párrafo.	1

Cuadro 6: Reglas del nivel discursivo y representación en las fuentes analizadas.

3.3.2. Nivel ortotipográfico

Hemos identificado en este nivel nueve indicaciones, de las que seis son de naturaleza ortográfica y tres tienen que ver con la presentación visual del texto. Están resumidas en el cuadro 7.

Nº	Reglas	Total
R.13	Vigilar los errores ortográficos.	5
R.14	Vigilar los errores de puntuación.	5
R.15	Evitar siglas, acrónimos y abreviaturas.	5
R.16	Evitar paréntesis.	1
R.17	Evitar corchetes/signos poco habituales.	3
R.18	Comprobar la expresión de cifras.	2
R.19	Destacar tipográficamente las enumeraciones largas.	3
R.20	Evitar líneas en blanco innecesarias.	1
R.21	Editar el texto de tal manera que el paciente pueda pasar por él rápidamente.	1

Cuadro 7: Reglas del nivel ortotipográfico y representación en las fuentes analizadas.

Proponemos fundir R.13 y R.14 en una sola, pues la puntuación es parte de la ortografía de la lengua. Asimismo, agrupamos R.15, R.16 y R.17 y R.18 por entender que en todos los casos se trata de usos específicos y especialmente complejos de la ortografía[9]. Por último, renunciamos a R.19, R.20 y R.21 porque son irrelevantes para su aplicación en el contexto de la traducción automática.

3.3.3. Nivel morfosintáctico

19 son las reglas propuestas en este nivel de análisis, si bien su grado de concreción no es homogéneo en absoluto, tal y como se muestra en el cuadro 8.

[9]Están tratados en los capítulos III.4, V y VIII de la Ortografía de la lengua española (2010).

A la vista de la diversidad de reglas, propondremos una simplificación de la nómina. Para ello, suprimimos 8 de las normas identificadas e integramos dos de ellas en otras de las propuestas. Estas son las eliminadas:

- R.26 es de pertinencia dudosa si se quiere que los textos adaptados a LCL gocen de naturalidad expresiva. Esta cuestión es de especial importancia en el caso de determinados géneros textuales: así sucede con la noticia periodística, que es con el que trabajamos en este proyecto.

Nº	Reglas	Total
R.22	Usar un orden natural de los elementos sintácticos.	5
R.23	Reducir la distancia entre sujeto y verbo.	2
R.24	Reducir el número de subordinadas por frase.	5
R.25	Evitar perífrasis verbales.	1
R.26	Evitar tiempos verbales compuestos.	1
R.27	Evitar en una oración dos sintagmas encabezados con la misma preposición.	1
R.28	Evitar el uso del gerundio.	3
R.29	Vigilar la falta de sistematicidad en el uso de la persona y el tiempo verbal.	2
R.30	Escapar de formas negativas, giros negativos, dobles negaciones, etc.	3
R.31	Evitar uso de subjuntivo y condicional.	2
R.32	Preferir la construcción activa a las pasivas sin agente y a las pasivas reflejas.	7
R.33	Evitar oraciones impersonales.	4
R.34	Evitar los adverbios en -mente.	1
R.35	Evitar nominalizaciones.	1
R.36	Evitar relativos poco usuales.	1
R.37	Evitar superlativos.	1
R.38	Separar las cadenas de dos o más sustantivos.	3
R.39	Respetar las normas de la lengua.	1

Cuadro 8: Reglas del nivel sintáctico y representación en las fuentes analizadas.

- R.29 se incluye en dos fuentes, que la proponen en las secciones correspondientes a la sintaxis. No obstante, parece que hacen más bien referencia a errores relativos a la coherencia de tiempo y de persona en la construcción discursiva, por lo que no tiene especial encaje en el nivel sintáctico. Por otro lado los tipos de error subyacentes a la incoherencia en construcción del tiempo y la persona no parecen estar tanto en el nivel del lenguaje claro como en el de la coherencia textual.
- R.31 alude a modos verbales cuya aparición está vinculada a determinados géneros textuales. Por otro lado, no parece que el uso del condicional y del subjuntivo sea un problema para la gestión de la sintaxis que lleva a cabo la TAN. Téngase, además, en cuenta que

la aplicación de R.24, asociada a la reducción de construcciones subordinadas, ya puede llevar aparejada una disminución en el uso del subjuntivo.

- R.34 se propone exclusivamente desde un contexto en el que el objetivo no es la lectura clara, sino la lectura fácil[10]. Por otro lado, las construcciones terminadas en *-mente* no suponen un problema a la hora de gestionar un texto con TAN.
- R.36 es excesivamente imprecisa: no se designa en ningún momento cuáles son esos relativos de poca frecuencia. Además, la única fuente que la propone circunscribe su aplicación a la lengua oral.
- R. 37 solo se recoge en Bowker (2024), quien sostiene que su uso se ha de evitar solamente en el caso de un único género textual, como son los *abstracts* de artículos científicos.
- R.38 hace referencia a un uso sintáctico propio de la lengua inglesa pero inexistente en el caso de español, como es el encadenamiento de dos o más sustantivos en posición contigua y con función de adjetivo: "e.g. instead of *sensitive protein function monitoring system*, use *a sensitive system that monitors protein function*"(Bowker, 2024, p. 175).
- R.39 tiene una naturaleza excesivamente genérica y no se fija en ningún concepto sintáctico concreto.

Por su parte estas son las integradas:

- R.22 se suma a R.23, pues la distancia entre sujeto y verbo en español está muy vinculada al orden natural de los constituyentes oracionales en español.
- R.32 se funde con R.33. En el caso de la segunda, la propuesta más repetida se orienta a evitar el uso de se impersonal. En ese contexto se ha de tener en cuenta que hay casos en los cuales no es tan evidente trazar la diferencia entre una construcción impersonal y una pasiva refleja, hasta el punto de que, en el *Glosario de términos gramaticales*, (2020), la Academia denomina impersonales reflejas a las impersonales de SE. De hecho, según el texto citado, unas y otras coinciden en que ocultan un argumento del verbo.

[10]Es una regla recogida en Fernández Vivanco & Sanfilippo (2025) a partir de García Muñoz (2012).

3.3.4. Nivel léxico y semántico

El análisis nos ha permitido identificar 19 reglas de contenido léxico o semántico, que resumimos en el cuadro 9.

Eliminamos R.40, R.41, R.42 y R.43 por estar asociadas a métricas matemáticas. Asimismo, renunciamos a R.58 porque su formulación va más allá de los objetivos propios del lenguaje claro: emplear palabras inadecuadas al sentido que se quiere transmitir no dificulta la comprensión; más bien la imposibilita.

Nº	**Reglas**	**Total**
R.40	Medir la longitud de palabra en letras.	1
R.41	Medir la longitud de palabra en sílabas.	2
R.42	Medir la densidad de palabras raras o no comunes.	3
R.43	Calcular las palabras de más de tres sílabas.	2
R.44	No usar diferentes términos para un mismo referente.	2
R.45	Evitar palabras de significado vacío.	3
R.46	Evitar palabras abstractas.	2
R.47	Evitar palabras polisémicas.	1
R.48	Evitar el uso de tecnicismos, jergas y xenismos.	8
R.49	Evitar extranjerismos.	1
R.50	Evitar falsos amigos.	1
R.51	Evitar modismos, expresiones locales y proverbios.	3
R.52	Incluir definiciones e información contextual.	1
R.53	Evitar el lenguaje figurado y las metáforas.	2
R.54	Explicar el significado de los términos si es necesario utilizarlos.	2
R.55	Eliminar palabras innecesarias.	1
R.56	Evitar arcaísmos.	1
R.57	No usar palabras fuera del diccionario.	1
R.58	Usar las palabras más adecuadas al sentido y el tono.	1

Cuadro 9: Reglas del nivel léxico-semántico y representación en las fuentes analizadas.

Por su parte se simplifica la lista de reglas combinando algunas de ellas:

- R.44, R.48, R.52 y R.54 se integran en una regla relativa a la manera en la que se debe evaluar el uso de la terminología en los textos.
- R.45, R.47 y R.55 se combinan en una norma destinada a evitar el léxico de significado vacío, polisémico y redundante.
- R.49 y R.50 se unen para dar cuenta de todo lo relativo a los préstamos léxicos y semánticos. Se incluye la referencia a xenismos, que aparece en R.49.
- R.46, R.56 y R.57 se funden en una norma que evalúe el uso del léxico infrecuente que no se

incluya en la terminología del ámbito temático al que corresponde el texto. Engloba también el uso de léxico innecesariamente abstracto.

- R.51 y R.53 se suman para dar cuenta de todo lo vinculado al lenguaje figurado y con alta carga connotativa, así como al léxico excesivamente local.

3.4. Recapitulación y propuesta

El resultado del análisis conduce a una nómina de 20 normas distribuidas en cuatro niveles: discursivo (5), ortográfico (2), morfosintáctico (8) y léxico-semántico (5). El proceso aparece resumido en el cuadro 10: en la columna de la izquierda aparecen las normas extraídas de los documentos de referencia y en la derecha, la propuesta realizada.

4. Definición de las reglas y procedimiento de aplicación

Una vez fijadas las reglas correspondientes a cada nivel, se aportará un descriptor que las justifica y las precisa conceptualmente, de manera que facilite su aplicación a textos concretos. El grado de concreción será diferente según la naturaleza de la norma de la que se trate.

D.01. Evitar párrafos construidos con un solo enunciado

Esta norma se asienta en la concepción de que el párrafo es una unidad significativa que integra varios enunciados relacionados entre sí por el contenido. Como ya hemos comprobado en trabajos anteriores (Bustos Gisbert, 2012), se trata de una unidad informativa compleja compuesta normalmente de dos o más enunciados. En cualquier caso esta caracterización estará siempre condicionada por la naturaleza del género textual al que se le apliquen las normas de lecturabilidad. En su uso como regla de preedición para TAN nos vamos a fijar de manera exclusiva en los párrafos compuestos por un solo enunciado normalmente muy extenso, por lo que la norma se solapa con D.02. En el ejemplo 1-a, hemos recogido un párrafo unioracional de 56 palabras, que en 1-b se distribuyen en dos enunciados de 24 y 30 respectivamente[11]:

1-a Sancho destaca que el trabajo se ha podido hacer gracias a la colaboración de los miles de empleados voluntarios del Banco Santander en Madrid, pero también por las ayudas de un millón de euros de la Fundación "la Caixa", 150.000 euros del Consejo Europeo de Investigación y 100.000 euros de la Agencia Estatal de Investigación.

1-b Sancho destaca que el trabajo se ha podido hacer gracias a la colaboración de los miles de empleados voluntarios del Banco Santander en Madrid. También ha recibido varias ayudas: un millón de euros de la Fundación "la Caixa", 150.000 euros del Consejo Europeo de Investigación y 100.000 euros de la Agencia Estatal de Investigación.

D.02. Evitar enunciados de más de 30 palabras

D.02 se expresa de forma normalmente vaga, aunque se repite en casi todas las fuentes. Por ello, hemos optado por unificar la formulación sobre un máximo de 30 palabras a partir de Bustos Gisbert (2017), donde acreditamos que es el límite de los enunciados más frecuentes en textos escritos en español correspondientes a contextos no especializados. Será, no obstante, una extensión máxima aproximada. Así, el enunciado recogido en el ejemplo 2-a, de 62 palabras, se divide en 2-b en dos enunciados de 24 y 37 palabras respectivamente:

2-a La investigación, realizada con técnicas caras y complejas como la tomografía axial computarizada (TAC) y la tomografía por emisión de positrones (PET), ha posibilitado detectar que en fases muy tempranas de la aterosclerosis ya hay niveles altos de propionato de imidazol, lo que podría facilitar enormemente el diagnóstico en esa etapa invisible en la que la persona está en riesgo sin saberlo.

2-b La investigación se ha realizado con técnicas caras y complejas como la tomografía axial computarizada (TAC) y la tomografía por emisión de positrones (PET). Ha posibilitado detectar niveles altos de propionato de imidazol en fases muy tempranas de la aterosclerosis, lo que podría facilitar enormemente el diagnóstico en esa etapa invisible en la que la persona está en riesgo sin saberlo.

Se podrán mantener enunciados de mayor extensión siempre que la complejidad sintáctica de los mismos no sea elevada. Remitimos a las consideraciones incluidas en S.09 para lo referido a esa cuestión.

[11] Incluimos en cada caso el texto original (a) y el texto preeditado (b). Todas las negritas incluidas son nuestras.

		Nivel discursivo		
R.11	Evitar párrafos con un solo enunciado.	>	D.01	Evitar párrafos construidos con un solo enunciado.
R.03	Evitar enunciados largos.	>	D.02	Evitar enunciados de más de 30 palabras.
R.07	Evitar enunciados muy largos.			
R.04	Evitar incisos extensos.	>	D.03	Evitar incisos intermedios.
R.01	Evitar elipsis y anáforas ambiguas.	>	D.04	Evitar las elipsis. Evitar las anáforas que puedan se ambiguas.
R.02	Usar conectores.	>	D.05	Incorporar conectores textuales.
R.05	Medir la longitud de frase en palabras.			
R.06	Medir la longitud del texto en frases.			
R.08	Usar párrafo breve.			
R.09	Usar párrafo monotemático.			
R.10	Proponer un orden lógico de párrafos.			
R.12	Cuidar la estructura interna del párrafo.			
		Nivel ortotipográfico		
R.13	Vigilar errores ortográficos	>	O.06	Vigilar los errores ortográficos.
R.14	Vigilar errores de puntuación.			
R.15	Evitar siglas, acrónimos y abreviaturas.	>	O.07	Evitar siglas, acrónimos, abreviaturas, símbolos y signos de puntuación poco habituales, y sistematizar la expresión de cifras.
R.16	Evitar paréntesis.			
R.17	Evitar signos poco habituales.			
R.18	Comprobar la expresión de cifras.			
R.19	Destacar tipográficamente las enumeraciones largas.			
R.20	Evitar líneas en blanco innecesarias.			
R.21	Editar el texto de tal manera que el paciente pueda pasar por él rápidamente.			
		Nivel morfosintáctico		
R.22	Usar un orden natural de los elementos sintácticos.	>	S.08	Evaluar la distancia entre los núcleos oracionales y entre estos y sus complementos sintagmáticos.
R.23	Reducir la distancia entre sujeto y verbo.			
R.24	Reducir el número de subordinadas por frase.	>	S.09	Reducir la complejidad sintáctica en los enunciados.
R.25	Evitar perífrasis verbales.	>	S.10	Limitar construcciones con infinitivo.
R.27	Evitar repetición de preposiciones.	>	S.11	Evitar en una oración la coincidencia contigua de sintagmas encabezados con la misma preposición.
R.28	Evitar gerundio.	>	S.12	Evitar el uso del gerundio más allá de las construcciones de contenido temporal o modal
R.30	Evitar negaciones múltiples.	>	S.13	Evitar la acumulación de polaridad negativa.
R.32	Preferir la construcción activa a las pasivas sin agente y a las pasivas reflejas.	>	S.14	Preferir activa y evitar construcciones con SE
R.33	Evitar oraciones impersonales.			
R.35	Evitar nominalizaciones.	>	S.15	Evitar la nominalización sintáctica.
R.29	Vigilar la falta de sistematicidad en el uso de la persona y el tiempo verbal.			
R.26	Evitar tiempos verbales compuestos.			
R.31	Evitar el subjuntivo y el condicional.			
R.34	Evitar los adverbios en -mente.			
R.36	Evitar relativos poco usuales.			
R.37	Evitar superlativos.			
R.38	Separar las cadenas de 2 o más sustantivos			
R.39	Respetar las normas de la lengua.			

		Nivel léxico		
R.45	Evitar palabras de significado vacío.	>	L.16	Evitar palabras vacías, polisémicas e innecesarias.
R.47	Evitar palabras polisémicas.			
R.55	Eliminar palabras innecesarias.			
R.46	Evitar palabras o conceptos abstractos.	>	L.17	Evitar palabras de baja frecuencia.
R.56	Evitar arcaísmos.			
R.57	No usar palabras fuera del diccionario.			
R.44	No usar diferentes términos para un mismo referente.	>	L.18	Evaluar el uso de terminología.
R.48	Evitar el uso de tecnicismos, jergas y xenismos.			
R.52	Incluir definiciones e información contextual.			
R.54	Explicar el significado de los términos.			
R.49	Evitar extranjerismos.	>	L.19	Evitar extranjerismos y falsos amigos.
R.50	Evitar falsos amigos.			
R.51	Evitar modismos, expresiones locales y proverbios.	>	L.20	Evitar localismos y lenguaje figurado.
R.53	Evitar el lenguaje figurado y las metáforas.			
R40	Medir la longitud de palabra en letras.			
R41	Medir la longitud de palabra en sílabas.			
R42	Medir la densidad de palabras raras o no			
R43	comunes.			
R58	Calcular las palabras de más de tres sílabas.			
	Usar las palabras más adecuadas al sentido y el tono.			

Cuadro 10: Conjunto de reglas de LCL tras su evaluación.

D.03. Evitar incisos intermedios

Aun cuando los documentos referidos al LCL suelen incluir con frecuencia esta recomendación, normalmente, las indicaciones suelen ser poco precisas. Efectivamente, la mayor dificultad de esta regla estriba en decidir la naturaleza de los incisos que se deben evitar. Por ello, proponemos, en primer lugar, fijar la atención especialmente en aquellos en posición intermedia que distancian el verbo de cualquiera de sus argumentos en cinco o más palabras. En segundo lugar, se evitarán siempre que tengan una extensión superior al 50 % del total del enunciado en el que se integra. Veamos lo que ocurre en el ejemplo 3. El inciso separa el sujeto del verbo y tiene una extensión de 13 palabras dentro de un enunciado de 27:

3-a El descubrimiento, **gracias a un experimento con cientos de trabajadores del Banco Santander en Madrid**, abre la puerta a nuevos tratamientos más allá de reducir el colesterol.

3-b El descubrimiento ha sido posible gracias a un experimento con cientos de trabajadores del Banco Santander en Madrid y abre la puerta a nuevos tratamientos más allá de reducir el colesterol.

Por último, es aconsejable eliminar incisos que dividen construcciones subordinadas, como sucede en el ejemplo 4, donde una construcción concesiva funciona como inciso que fragmenta otra de naturaleza causal:

4-a Esto es muy relevante porque, **aunque reducir el colesterol mediante los medicamentos denominados estatinas puede disminuir eficazmente el riesgo de enfermedad cardiovascular,** no siempre es suficiente.

4-b Esto es muy relevante porque la reducción del colesterol mediante los medicamentos denominados estatinas puede disminuir eficazmente el riesgo de enfermedad cardiovascular, pero no siempre es suficiente.

D.04. Evitar las elipsis. Evitar las anáforas que puedan se ambiguas

Se ha buscado una redacción de compromiso para D.04, ya que los seis documentos que la proponen presentan diferencias de matiz. El efecto de la anáfora y la elipsis en la comprensión de los textos ha sido ampliamente estudiado desde el análisis del discurso y la psicolingüística. En ese contexto, apoyamos nuestra posición en la defendida por Martin & McElree (2008, p. 903), quienes sostienen que el nivel de lecturabilidad

desciende en el caso en el que los elementos correferenciales están a cierta distancia[12].

En nuestra percepción, esta regla implica, en primer lugar, evitar las elipsis verbales; los enunciados deben tener, preferiblemente, forma sintáctica de oración con verbo explícito. Es la intervención realizada en el ejemplo 5:

5-a Sancho destaca que el trabajo se ha podido hacer gracias a la colaboración de los miles de empleados voluntarios del Banco Santander en Madrid. **También por las ayudas** de un millón de euros de la Fundación "la Caixa", 150.000 euros del Consejo Europeo de Investigación y 100.000 euros de la Agencia Estatal de Investigación.

5-b Sancho destaca que el trabajo se ha podido hacer gracias a la colaboración de los miles de empleados voluntarios del Banco Santander en Madrid. **También contaron con las ayudas** de un millón de euros de la Fundación "la Caixa", 150.000 euros del Consejo Europeo de Investigación y 100.000 euros de la Agencia Estatal de Investigación.

En segundo lugar, se debe solucionar cualquier elipsis o anáfora cuando incluye términos de significado poco preciso o cuando la correferencia exhibe diferencias de género o de número (primera intervención en el ejemplo 6). Por último, se recomienda aplicar la norma siempre que el correferente no se encuentre en el enunciado inmediatamente anterior (segunda intervención en el ejemplo 6):

6-a "Esto podría ayudar a explicar por qué algunas personas desarrollan enfermedades cardiacas a pesar de tener pocos o ningún factor de riesgo tradicional. Entre **ellos** se encuentran el colesterol alto o la hipertensión", **subraya**.

6-b "Esto podría ayudar a explicar por qué algunas personas desarrollan enfermedades cardiacas a pesar de tener pocos o ningún factor de riesgo tradicional. Entre **dichos riesgos** se encuentran el colesterol alto o la hipertensión", **subraya Haghikia**.

Por último, debemos solucionar cualquier elipsis que induzca a la ambigüedad, especialmente cuando involucra términos de significado poco preciso. Así sucede en el ejemplo 7:

7-a "El hecho de que **dos grupos distintos** hayan llegado a la misma conclusión refuerza la confianza en que se trata de un descubrimiento significativo y relevante", argumenta Haghikia, de la Universidad del Ruhr de Bochum (Alemania).

7-b "El hecho de que **dos equipos de investigación** distintos hayan llegado a la misma conclusión refuerza la confianza en que se trata de un descubrimiento significativo y relevante", argumenta Haghikia, de la Universidad del Ruhr de Bochum (Alemania).

D.05. Incorporar conectores textuales

Con D.05 se persigue explicitar las relaciones entre enunciados que puedan resultar más complejas de interpretar, pero manteniendo la idiomaticidad de los textos. Nuestro punto de partida será Portolés (1993, p. 146-147), quien sostiene que tan poco recomendable es construir el texto sin conectores como sobrecargarlo innecesariamente de ellos, especialmente cuando las relaciones discursivas son evidentes. En ese sentido, parece oportuno tener en cuenta el porcentaje promedio en español de explicitación de la conexión discursiva (Bustos Gisbert, 2014), que se sitúa en torno al 25 % en los textos de base tipológica dominante expositiva y en casi el 17 % cuando son de base narrativa.

Por todo lo dicho, la recomendación será aumentar en aproximadamente un 10 % esos porcentajes. El ejemplo 8 está compuesto por dos enunciados; proponemos introducir un conector en el segundo de ellos:

8-a Ese contexto revela el papel desconocido de los microbios intestinales en algunas enfermedades humanas. El biotecnólogo español Cayetano Pleguezuelos y sus colegas del Instituto Hubrecht (Países Bajos) demostraron en febrero de 2020 que una cepa de Escherichia coli produce una molécula tóxica, llamada colibactina.

8-b Ese contexto revela el papel desconocido de los microbios intestinales en algunas enfermedades humanas. **Así**, el biotecnólogo español Cayetano Pleguezuelos y sus colegas del Instituto Hubrecht (Países Bajos) demostraron en febrero de 2020 que una cepa de Escherichia coli produce una molécula tóxica, llamada colibactina.

D.05 se solapa con D.02, pues la fragmentación de enunciados puede tener como resultado la incorporación de conectores textuales. Por ello, es oportuno que se apliquen en el orden propuesto.

[12] En el contexto de la TA, se ha aconsejado con mucha frecuencia vigilar las elipsis y las anáforas en las propuestas de aplicación de lenguajes controlados. Es el caso de O'Brien (2003, 2004) y Ramírez Polo (2012).

Es lo que sucede en el ejemplo 9, donde un solo enunciado se convierte en dos independientes conectados por un enlace contraargumentativo:

9-a Esto es muy relevante porque, aunque reducir el colesterol —mediante los medicamentos denominados estatinas, por ejemplo— puede disminuir eficazmente el riesgo de enfermedad cardiovascular, una proporción considerable de personas sigue experimentando eventos cardiovasculares adversos, como infartos de miocardio o accidentes cerebrovasculares", alertan.

9-b Esto es muy relevante porque reducir el colesterol —mediante los medicamentos denominados estatinas, por ejemplo— puede disminuir eficazmente el riesgo de enfermedad cardiovascular. **Sin embargo**, una proporción considerable de personas sigue experimentando eventos cardiovasculares adversos, como infartos de miocardio o accidentes cerebrovasculares", alertan.

O.06. Vigilar los errores ortográficos

Aun cuando siempre se debe cuidar con esmero la ortografía, esta no compete realmente al LCL, sino a la corrección lingüística. Por ello, O.06 se interesará por cuestiones donde la norma dictada por la *Ortografía de la lengua española* (2010) plantea dudas incluso para los usuarios más competentes: se fijará específicamente en lo que tiene que ver con el uso de mayúsculas y minúsculas, y con la ortografía de los nombres propios y de las expresiones procedentes de otras lenguas. En el ejemplo 10 se observa la aplicación de la regla en lo referido a nombres propios de instituciones:

10-a También contaron con las ayudas de un millón de euros de la **Fundación "la Caixa"**, 150.000 euros del Consejo Europeo de Investigación y 100.000 euros de la Agencia Estatal de Investigación.

10-b También contaron con las ayudas de un millón de euros de la **Fundación La Caixa**, 150.000 euros del Consejo Europeo de Investigación y 100.000 euros de la Agencia Estatal de Investigación.

O.07 Vigilar el uso de siglas, acrónimos, abreviaturas y símbolos, sistematizar la expresión de cifras, y evitar los signos de puntuación poco habituales.

La aplicación de O.07 se asocia a las intervenciones en siglas, acrónimos, abreviaturas y símbolos; estará determinada por el género textual y por el dominio del tema que se suponga al destinatario. Así, en el ejemplo 11 se elimina la fórmula química final, pues el texto se ha publicado en un periódico cuyo lector meta no está previsto que tenga conocimientos avanzados de química:

11-a Los nuevos resultados muestran que algunas bacterias del intestino producen propionato de imidazol en determinados estados. Se trata de una molécula sencilla con seis átomos de carbono, ocho de hidrógeno, dos de nitrógeno y dos de oxígeno **($C_6H_8N_2O_2$)**.

11-b Los nuevos resultados muestran que algunas bacterias del intestino producen propionato de imidazol en determinados estados. Se trata de una molécula sencilla con seis átomos de carbono, ocho de hidrógeno, dos de nitrógeno y dos de oxígeno.

La expresión de las cifras es uno de los errores prototípicos de la TAN. En términos de preedición, se abren dos posibles estrategias: seguir la norma ortográfica o adaptar las cifras, medidas y magnitudes a las convenciones de la lengua meta. Optaremos por seguir las indicaciones de la *Ortografía de la lengua española* (2010), en su capítulo VIII. Así hemos hecho en el ejemplo 12:

12-a El proyecto también ha recibido varias ayudas: un millón de euros de la Fundación La Caixa, **150.000** euros del Consejo Europeo de Investigación y **100.000** euros de la Agencia Estatal de Investigación.

12-b El proyecto también ha recibido varias ayudas: un millón de euros de la Fundación La Caixa, **150 000** euros del Consejo Europeo de Investigación y **100 000** euros de la Agencia Estatal de Investigación.

Por otro lado, se recomienda evitar en lo posible los denominados *signos auxiliares* en el capítulo III.4 de la *Ortografía de la lengua española* (2010). Nos referimos al guion, la barra, la antilambda, la llave, el apóstrofo, el asterisco, la flecha el calderón y el signo de párrafo. Su uso solo se aceptará cuando sea resultado de una exigencia derivada de las convenciones de género.

S.08. Evaluar la distancia entre los núcleos oracionales y entre estos y sus complementos sintagmáticos

El orden de las palabras es una cuestión largamente debatida por la crítica en los estudios de lingüística (véase, por ejemplo, Sologuren Insúa & Castillo Fadić 2018). En lo referido específicamente a la lengua española es un lugar común

afirmar que esta tiene, en general, un grado de flexibilidad elevado en lo relativo al orden de los constituyentes oracionales, que está a menudo condicionado por la estructura informativa de los enunciados. En ese sentido, se sostiene que la ordenación lógica se registra sobre todo en su modalidad escrita y lleva a ubicar al principio la información valorada como temática o conocida, mientras que la información novedosa o remática suele aparecer en la segunda parte de los enunciados (Hernando Cuadrado, 2005).

Según hemos podido constatar, el respeto de esas posiciones no suele generar dificultades en la aplicación de TAN. Sin embargo, algunos procedimientos sintácticos (muy frecuentes, por otra parte, en la lengua oral, aunque quizá menos en la variante escrita) pueden resultar problemáticos, por lo que conviene preeditarlos. Hablamos de estructuras como las siguientes:

- Las construcciones topicalizadas, que sirven para fijar al principio de los enunciados determinadas informaciones temáticas.
- Las construcciones focalizadas, en las que se antepone la información novedosa o remática para destacarla.
- Las copulativas enfáticas, construidas mediante el verbo ser, una oración de relativo sin antecedente expreso y un segmento focal. El ejemplo 13-a incluye la intervención en una construcción de este tipo, que se ha eliminado en 13-b:

13-a Lo que resulta especialmente llamativo es que el propionato de imidazol parece promover la aterosclerosis incluso cuando los niveles de colesterol son normales.

13-b Es especialmente llamativo que el propionato de imidazol parece promover la aterosclerosis incluso cuando los niveles de colesterol son normales.

En segundo lugar, es importante evaluar la distancia que se pueda observar entre los núcleos verbales y sus complementos argumentales: se ha de prestar especial atención a la establecida con los sujetos oracionales. No obstante, se ha de tener en cuenta que S.08 se ha de aplicar después de D.04 (referida a la reducción de los incisos), con la que se solapa y que, de haberse empleado adecuadamente, reducirá ahora la intervención en el texto.

S.09. Reducir la complejidad sintáctica en los enunciados

Nuestra propuesta tiene su origen en un estudio previo (Bustos Gisbert, 2017) en el que propusimos un análisis de los enunciados en textos no especializados que nos permitió identificar tres conjuntos de datos relevantes en relación con S.09. El primero tiene que ver con la naturaleza sintáctica de los enunciados: el 67 % se construye como oraciones compuestas, mientras que solo el 31,5 % son oraciones simples. Además, se prefiere la subordinación a la coordinación, que suma apenas el 13 % de las oraciones con más de un predicado. Por otro lado, los grupos sintácticos no oracionales apenas dan cuenta del 1,5 % de las unidades. El segundo conjunto de datos se fija en el número de predicaciones subordinadas: el límite de complejidad se establece en tres predicaciones de esta naturaleza. Por último, en lo referido a los árboles de dependencia, el trabajo realizado desveló que se evita ir más allá de dos niveles de incrustación oracional: solo un 5,72 % de las predicaciones verbales identificadas quedan fuera de tal tendencia.

Así las cosas, para la aplicación de S.09 proponemos, en primer lugar, intervenir en todos los enunciados que sean grupos sintácticos no oracionales. En segundo lugar, se ha de limitar el uso de coordinaciones que no sean copulativas o adversativas. En tercer término, se reducirá a un máximo de tres las predicaciones subordinadas incluidas en un mismo enunciado; podrán ser cuatro cuando dos de ellas aparezcan coordinadas entre sí. Por último, se evitarán niveles de incrustación superiores a 2 en todos los enunciados. En el ejemplo 14-a se ha intervenido en un enunciado con cuatro predicaciones verbales subordinadas con un nivel 3 de incrustación. El resultado en 14-b muestra tres verbos subordinados y un nivel 2 de incrustación:

14-a Este tipo de aterosclerosis se caracteriza porque las placas de grasa **tienen** mayor tendencia **a romperse y dar lugar** a los trombos que **provocan** infartos y accidentes cerebrovasculares.

14-b Este tipo de aterosclerosis se caracteriza porque las placas de grasa **se rompen** más a menudo y dan lugar a los trombos que **provocan** infartos y accidentes cerebrovasculares.

S.10. Evitar el uso del gerundio más allá de las construcciones de contenido temporal o modal

Como veremos en S.13, los estudios de LCL consideran que las formas no personales del verbo tienden a ser menos transparentes que las conjugadas. En el caso de los gerundios se suma el problema de que hay en su uso frecuentes errores derivados de la influencia de otras lenguas, muy especialmente del inglés. Además, la norma académica al respecto es ciertamente poco clara. Por todo ello, se sugiere reducir el empleo del gerundio a tres contextos: en usos perifrásticos con ESTAR, LLEVAR, SEGUIR y CONTINUAR; cuando introduce una acción que de forma evidente se pueden interpretar como simultánea a las designada por el verbo del que depende; y cuando presenta una acción que explica el modo en que se lleva a cabo lo significado por el verbo principal. Así, en el ejemplo 15-b se mantiene el primer gerundio, pero no el segundo:

15-a Los casos de cáncer colorrectal se **están disparando** en las personas menores de 50 años por causas desconocidas, **duplicándose** en muchos países en las últimas dos décadas.

15-b Los casos de cáncer colorrectal se **están disparando** en las personas menores de 50 años por causas desconocidas **y se han duplicado** en muchos países en las últimas dos décadas.

S.11. Evitar la acumulación de términos de polaridad negativa

En relación con los efectos de la acumulación de términos de polaridad negativa, determinados estudios en el ámbito de la psicolingüística (véase, por ejemplo, Haase et al., 2019) han concluido que, en el proceso de comprensión, los lectores tienden a construir representaciones positivas antes de negarlas o transformarlas con el contexto negativo. Por ello, la presencia de un término de polaridad negativa parece requerir recursos de procesamiento adicionales. En consecuencia, aumentan el tiempo y el esfuerzo de comprensión. Así pues, en la aplicación de S.11 se evitará en lo posible el uso de las formas de polaridad negativa (NADA, NADIE, NINGUNO/A, NUNCA, JAMÁS, TAMPOCO, SINO, NI, NO y SIN), especialmente cuando se acumulen dos o más en un mismo enunciado. Véase el ejemplo 16:

16-a Los nuevos resultados demuestran que la aterosclerosis **no es solo una enfermedad** causada por las grasas, **sino que también presenta** un componente inflamatorio y autoinmune, en palabras de Sancho, nacido en Cartagena (Murcia) hace 53 años.

16-b Los nuevos resultados demuestran que la aterosclerosis **es una enfermedad** causada por las grasas **y además presenta** un componente inflamatorio y autoinmune, en palabras de Sancho, nacido en Cartagena (Murcia) hace 53 años.

S.12. Preferir la construcción activa a la pasiva. Evitar las construcciones de SE.

En esta regla se acometen dos cuestiones interrelacionadas. Con la primera, se prioriza la voz activa a la pasiva. De ese modo, se respeta la tendencia natural de la lengua española, en la que, como reconoce la Academia en su *Glosario de términos gramaticales*(2020), es especialmente limitado tanto el uso de las pasivas perifrásticas, formadas con SER + PARTICIPIO de verbo transitivo, como el de las construcciones llamadas inacusativas del tipo "*cuando sea llegado el momento...*".

La segunda cuestión tiene que ver con el uso de las construcciones de SE, que son comparativamente mucho más frecuentes en español y entre las que debemos incluir las construcciones impersonales, las pronominales, las reflexivas y recíprocas, y las pasivas reflejas: el hecho mismo de que existan varias construcciones con significados distintos es fuente tanto de problemas formales como de adscripción. Las dificultades de interpretación por parte de los motores de TAN afectan especialmente a las pronominales y las impersonales. Por ello, se recomienda evitar las construcciones de SE siempre que sea posible. Es la norma que hemos aplicado a 17-a, donde se acumulan dos construcciones de SE:

17-a Los casos de cáncer colorrectal **se están disparando** en las personas menores de 50 años por causas desconocidas y **se han duplicado** en muchos países en las últimas dos décadas.

17-b Los casos de cáncer colorrectal **aumentan** rápidamente en las personas menores de 50 años por causas desconocidas y **han duplicado su número** en muchos países en las últimas dos décadas.

S.13. Limitar las construcciones con infinitivos

La sobreutilización de infinitivos tiene efectos en la claridad de los enunciados, pues se trata de una forma verbal invariable, por lo que no incluye información personal, temporal y modal que

el lector debe reconstruir a partir del contexto. Es por esa razón por la que las construcciones no personales son, comparativamente, menos frecuentes que las de verbo conjugado. Así lo confirmamos en Bustos Gisbert (2017): las formas no personales de los verbos rozan el 20 % de frecuencia y las perífrasis exhiben un uso medio de casi el 14 %. El infinitivo es la forma no personal más habitual y soporta las perífrasis más frecuentes.

Por todo ello, S.13 va a tener como objetivo limitar el infinitivo a las oraciones subordinadas sustantivas en función argumental cuando no se disponga de alternativas expresivas aptas. Asimismo, se restringirá (en las mismas condiciones de idiomaticidad) el uso de las perífrasis de infinitivo, especialmente las de valor modal. Es lo que sucede en el ejemplo 18:

18-a Sancho destaca que el trabajo **se ha podido hacer** gracias a la colaboración de los miles de empleados voluntarios del Banco Santander en Madrid.

18-b Sancho destaca que el trabajo **ha sido posible** gracias a la colaboración de los miles de empleados voluntarios del Banco Santander en Madrid.

S.14. Evitar la nominalización sintáctica

La relación entre la nominalización sintáctica y el LCL ha sido analizada con mucha precisión por Azpiazu (2025, p. 105-106), quien ha demostrado que la predicación nominal o la asentada en formas verbales no personales condensan la expresión y la hacen más opaca al elevar el nivel de abstracción conceptual.

Así las cosas, la aplicación de S.14 supondrá intervenir en los usos propios del estilo nominal, como, entre otros, la tendencia a apoyar el discurso en unidades nominales en perjuicio de las verbales, la preferencia por las construcciones copulativas frente a las predicativas o el alargamiento sintáctico de base preposicional y no conjuntivo. En el ejemplo 19 se observa un cambio de base conceptual nominal a verbal:

19-a Sin embargo, **señalan "cambios** en la ecología microbiana intestinal" tras modificaciones en la comida, **con un enriquecimiento** de géneros bacterianos como Escherichia, Shigella y Eubacterium.

19-b Sin embargo, **observan que la dieta "modifica** la ecología microbiana intestinal"**y aumenta** la presencia de géneros bacetrianos como Escherichia, Shigella y Eubacterium.

En 20, además de un cambio de base conceptual, se evita un verbo copulativo en beneficio de otro predicativo:

20-a El nuevo trabajo muestra que **los niveles** en sangre de propionato de imidazol **son más bajos** en **las personas con una dieta rica en** verduras, frutas y cereales integrales.

20-b El estudio revela que **las personas que comen** más verduras, frutas y cereales integrales **muestran niveles más bajos** de propionato de imidazol en sangre.

Y en 21 se sustituye una locución prepositiva por una preposición y se corrige una expresión nominal por otra verbal:

21-a El objetivo era estudiarlos a fondo durante décadas **con el fin de entender el nacimiento** de las enfermedades cardiovasculares en las personas sanas.

21-b El objetivo era estudiarlos a fondo durante décadas **para comprender cómo se originan** las enfermedades cardiovasculares en personas sanas.

Los usos nominales considerados en S.14 se solapan con las cuestiones que interesan a S.10, S.12 y S.13, por lo es oportuno que se aplique después de ellas, pues se reducirá la intervención en el texto.

S.15. Evitar en una oración la coincidencia contigua de sintagmas encabezados con la misma preposición

El efecto en la comprensión textual que tiene la acumulación contigua de unidades sintácticamente idénticas ha sido ampliamente estudiado en el ámbito de la psiocolíngüística (por ejemplo, Pinker, 1999).

En nuestra percepción, estas construcciones pueden darse en cualquier entorno sintáctico, aunque son especialmente frecuentes cuando se trata de sintagmas preposicionales. Por ello, se propone aplicar la norma cuando encadenen tres o más sintagmas preposicionales encabezados por la preposición DE. Reduciremos el número a dos cuando se trate de otra preposición. Es lo que sucede en 22:

22-a Este descubrimiento facilitaría enormemente el diagnóstico **en esa** etapa invisible **en la que** la persona desconoce que está en riesgo.

22-b Este descubrimiento facilitaría enormemente el diagnóstico **durante** esa etapa invisible **en la que** la persona desconoce que está en riesgo.

L.16. Evitar palabras vacías, polisémicas e innecesarias

El efecto en la traducción automática del léxico de esta naturaleza ha sido evaluado fundamentalmente dos direcciones. Una tiene que ver con la descontextualización en la interpretación de esas unidades (Naveen & Trojovský, 2024). Otra observa que su uso supone una reducción de la variedad de equivalentes léxicos: el sistema tiende a escoger la opción más frecuente, lo que implica empobrecimiento y pérdida de matices semánticos (Espinosa Giménez & Forcada, 2024).

A partir de las consideraciones enunciadas, L.16 persigue que el texto mejore en precisión léxica y semántica. Desde la primera perspectiva, su aplicación estará destinada a eliminar las palabras innecesarias; desde la segunda, se limitará la presencia de vocablos de contenido léxico impreciso. Así, se pondrá especial atención en reducir la frecuencia de aparición de sustantivos comodín de significado muy general. Es la intervención aplicada en el ejemplo 23:

23-a La buena noticia es que, si el propionato de imidazol **provoca el problema** en un porcentaje relevante de los pacientes, es posible intervenir para evitarlo.

23-b La buena noticia es que, si el propionato de imidazol **provoca la enfermedad** en un porcentaje relevante de los pacientes, es posible intervenir para evitarlo.

También se evitarán los verbos de contenido vago (DAR, PONER, HACER, DECIR, HABER, TENER, etc.), como sucede en 24, y las expresiones verbales complejas, en las que se combina uno de ellos con un sustantivo que concentra el significado:

24-a Igualmente, destacó que es sencillo prevenirlos con medidas como **hacer ejercicio**, llevar una alimentación sana y evitar el tabaco.

24-b Igualmente, destacó que es sencillo prevenirlos con medidas como **practicar ejercicio**, llevar una alimentación sana y evitar el tabaco

Se vigilarán también las palabras y locuciones de moda, como la que se emplea en 25-a:

25-a Al estudiar a fondo a los empleados del Santander con máquinas de imagen médica avanzada, los científicos descubrieron pronto otro **dato impactante**.

25-b Al estudiar a fondo a los empleados del Santander con máquinas de imagen médica avanzada, los científicos descubrieron pronto otro **hallazgo relevante**.

Por último, se intervendrá también en aquellos usos en los que a determinadas unidades léxicas se les atribuyen significados que realmente no les corresponden. Es lo que sucede en el ejemplo 26:

26-a Este inesperado vínculo entre los microbios y las enfermedades cardiovasculares —primera causa de muerte en la humanidad— **es un cambio de paradigma**.

26-b Este inesperado vínculo entre los microbios y las enfermedades cardiovasculares —primera causa de muerte en la humanidad— **supone una transformación profunda en la investigación médica**.

L.17. Evitar palabras de baja frecuencia

Los estudios que nos preceden parecen confirmar que el uso de palabras de baja frecuencia afecta a la comprensión y al recuerdo (Ludewig et al. 2023; Horno Chéliz 2024). La percepción de la frecuencia y la familiaridad de las palabras es subjetiva y susceptible a sesgos cognitivos. Por lo tanto, se recomienda emplear herramientas objetivas para su identificación, como es el caso de la frecuencia relativa en corpus generales o estimaciones de disponibilidad léxica.

La aplicación de L17 busca evitar palabras de baja frecuencia de uso y que no pertenezcan a la terminología propia del ámbito temático del texto, como sucede en el ejemplo 27:

27-a La aterosclerosis **era omnipresente**.

27-b La aterosclerosis **estaba muy extendida**.

Téngase en cuenta, además, que no es infrecuente que la incorporación de unidades léxicas inhabituales suponga, además, que se les asignen significados que no les son propios, por lo que L.17 se solapa con L.16. Tal hecho justifica la intervención realizada en 28:

28-a La revista *Nature*, considerada **la vitrina** de la mejor ciencia mundial, publicará el estudio este miércoles.

28-b La revista *Nature*, considerada **un referente** de la mejor ciencia mundial, publicará el estudio este miércoles

L.18 Evitar extranjerismos y falsos amigos

Varios estudios de LCL coinciden en que las palabras tomadas de lenguas extranjeras revierten en una mayor dificultad de lectura (Montiel et al., 2024). Asimismo, si seguimos el modelo dual de Coltheart podemos afirmar que, *a priori*, las ortografías extranjeras dificultan la vía subléxica del procedimiento de lectura (Cuetos Vega et al., 2018). Por otro lado, la dirección más frecuente de intercambio lingüístico en las últimas décadas se produce desde el inglés al resto de las lenguas (García Palacios, 2022).

Ahora bien, aun siendo una tendencia respaldada por evidencia empírica, no puede asumirse que todas las lenguas recurran a los mismos préstamos. Por lo tanto, la aplicación de la regla dependerá en todo caso de la combinación lingüística en la que se aplique la preedición. Algo similar sucede con los cognados y los heterosemánticos o falsos amigos, que también requieren de un análisis interlingüístico y un conocimiento profundo de la lengua original y la lengua meta (Vandevoorde & Lefever, 2023). En ese sentido, los vocablos con similitudes morfológicas y divergencias semánticas constituyen una de las interferencias más problemáticas en el ejercicio traductológico (Hernández Palma et al., 2025). Si bien es cierto que la TAN ha mejorado considerablemente en este sentido, perduran los errores ligados a los falsos amigos terminológicos (Lerner & Yvon, 2025).

Por ende, L.18 se orientará a eliminar préstamos crudos y falsos amigos cuando generen una interferencia semántica en la combinación lingüística en la que se va a realizar la traducción. Es lo que se observa en 29:

29-a El microbiólogo argentino Federico Rey y el patólogo indio Vaibhav Vemuganti aplauden "las **excitantes oportunidades**" que abre el nuevo estudio para la prevención y el tratamiento de las enfermedades cardiovasculares.

29-b El microbiólogo argentino Federico Rey y el patólogo indio Vaibhav Vemuganti aplauden "las **estimulantes oportunidades**" que abre el nuevo estudio para la prevención y el tratamiento de las enfermedades cardiovasculares.

L.19 Evitar los localismos y el lenguaje figurado

La eliminación de localismos garantizará la comprensión por parte de destinatarios menos marcados en términos dialectales. Por otro, la reducción del lenguaje figurado facilitará globalmente la comprensión, especialmente en aquellos casos en los que las dimensiones poética y argumentativa son secundarias en el acto de comunicación establecido a través del texto. En ello coincidimos con Naveen & Trojovský (2024), quienes han estudiado detalladamente la manera en la que los motores de traducción automática gestionan las expresiones idiomáticas y el lenguaje figurado. En su percepción, la sustitución de las metafóricas por alternativas expresivas ancladas a la literalidad del significado facilita globalmente la traducción. Es lo que sucede en el ejemplo 30:

30-a El descubrimiento ha sido posible gracias a un experimento con cientos de trabajadores del Banco Santand.er en Madrid y **abre la puerta a** nuevos tratamientos más allá de reducir el colesterol

30-b El descubrimiento ha sido posible gracias a un experimento con cientos de trabajadores del Banco Santander en Madrid y **permite el desarrollo de** nuevos tratamientos más allá de reducir el colesterol

Por todo lo dicho, la aplicación de L.19 estará en correlación directa con el género textual empleado y con la relevancia de las dimensiones textuales en su construcción.

L.20. Evaluar el uso de la terminología

La limitación en el uso de la terminología es una de las recomendaciones que se recoge con frecuencia en los tratados referidos al LCL. De hecho, es la que se consigna en mayor número de ocasiones en los textos de referencia empleados para esta investigación. No obstante, parece necesario efectuar algunas consideraciones al respecto. La primera y más importante tiene que ver con que los términos son unidades de forma y contenido que activan conocimiento especializado en un contexto discursivo y en cuyo empleo es esencial considerar la naturaleza del emisor y del destinatario previsto. Por lo tanto, la terminología es un recurso indispensable para representar y comunicar los conocimientos especializados. Puesto que muchas veces el propósito de los géneros LCL es precisamente trasmitir ese conocimiento experto, descartamos cualquier regla orientada a *eliminar* la terminología de forma indiscriminada y, en su lugar, proponemos *evaluarla*.

Por todo ello, L.20 se va a fijar en dos variables fundamentales. La primera tiene que ver con la consistencia terminológica, orientada a garantizar que no se emplean diferentes denominaciones para designar un mismo referente especializado. La segunda se centra en valorar la adecuación de los usos terminológicos en relación con el género textual y con el destinatario previsto. Así, en un género dirigido a un destinatario en desequilibrio discursivo, tiene sentido que se intervenga en dos direcciones. Por un lado, se evitarán, como en 31, aquellos términos cuyo empleo no sea imprescindible a la hora de garantizar tanto el contenido como la intención general del escrito:

31-a David Sancho y sus colegas reconocen que hará falta más investigación para identificar las **cepas específicas** de bacterias con capacidad para producir la molécula.

31-b David Sancho y sus colegas reconocen que hará falta más investigación para identificar qué **tipos concretos** de bacterias tienen capacidad para producir la molécula.

Y por otro, se parafrasearán, como en 32, los términos cuyo significado se considere opaco.

32-a También observan que las modificaciones en la dieta "**alteran la ecología microbiana intestinal"** y aumentan la presencia de géneros bacterianos como Escherichia, Shigella y Eubacterium.

32-b También observan que las modificaciones en la dieta "**alteran el equilibrio de bacterias en el intestino"** y hacen que aumenten algunas como Escherichia, Shigella y Eubacterium.

Si tenemos en cuenta que nuestro objetivo final es reducir el esfuerzo de posedición, hemos de contemplar que la equivalencia terminológica en traducción no se limita a la búsqueda de un equivalente semántico en lengua meta (Kerremans & Temmerman, 2016), sino que contempla un amplio abanico de procedimientos traductológicos, como hiperónimos, paráfrasis, perífrasis, omisiones, modulaciones o transposiciones. El ejemplo 33 es especialmente representativo porque reproduce el titular de una noticia periodística publicada en un medio generalista. Originalmente, incluía un término, que se reformula en un lenguaje con menor grado de especialización:

33-a Revolución en la medicina: una molécula producida por bacterias del intestino causa **la aterosclerosis**, culpable de millones de muertes

33-b Revolución en la medicina: una molécula producida por bacterias del intestino causa **una enfermedad de las arterias**, culpable de millones de muertes

5. Conclusión

La investigación llevada a cabo ha cumplido dos objetivos fundamentales. En primer lugar, ha permitido extraer un número viable de normas de LCL aplicables a la preedición de textos en la dirección ES > EN que van a ser tratados con un motor TA para su traducción y posterior revisión. Este proceso de preedición tendrá más rendimiento cuando se aplique a textos correspondientes a géneros textuales utilizados por comunidades en situación de desequilibrio discursivo.

Se propone la aplicación de 20 normas distribuidas en cuatro niveles: discursivo (5), ortográfico (2), morfosintáctico (8) y léxico-semántico (5). El orden de administración de las normas será el propuesto en el cuadro 11.

La ordenación propuesta está fijada a partir de la relación que se establece entre las normas. Se basa en una evaluación del texto sobre el que se interviene orientada de lo general a lo particular. Así, se parte de las normas de naturaleza discursiva porque valoran la lecturabilidad en términos más globales y su aplicación facilita la de las normas sintácticas y las ortográficas.

Estas son las relaciones más evidentes:

- D.01 (Evitar las elipsis. Evitar las anáforas que puedan se ambiguas) > S.10 (Limitar las construcciones con infinitivos).
- D.02 (Evitar enunciados de más de 30 palabras) > S.09 (Reducir la complejidad sintáctica en los enunciados).
- D.04 (Evitar incisos intermedios) > S.08 (Evaluar la distancia entre los núcleos oracionales y entre estos y sus complementos sintagmáticos).
- D.02 (Evitar enunciados de más de 30 palabras) y D.04 (Evitar incisos intermedios) > O.06 (Vigilar los errores ortográficos).

Dentro de cada grupo, el orden es razonablemente más libre. No obstante, se han de tener en cuenta algunas cuestiones asociadas a un principio de exhaustividad en la construcción de un texto más comprensible. Así, en el plano discursivo, la aplicación de D.02 (Evitar enunciados de más de 30 palabras) puede suponer que D.05

Nº	Reglas
D.01	Evitar párrafos construidos con un solo enunciado.
D.02	Evitar enunciados de más de 30 palabras.
D.03	Evitar incisos intermedios.
D.04	Evitar las elipsis. Evitar las anáforas que puedan ser ambiguas.
D.05	Incorporar conectores textuales.
O.06	Vigilar los errores ortográficos.
O.07	Evitar siglas, acrónimos, abreviaturas, símbolos y signos de puntuación poco habituales, y sistematizar la expresión de cifras.
S.08	Evaluar la distancia entre los núcleos oracionales y entre estos y sus complementos sintagmáticos.
S.09	Reducir la complejidad sintáctica en los enunciados.
S.10	Evitar el uso del gerundio más allá de las construcciones de contenido temporal o modal.
S.11	Evitar la acumulación de términos de polaridad negativa.
S.12	Preferir la construcción activa a las pasivas. Evitar las construcciones de SE.
S.13	Limitar las construcciones con infinitivos.
S.14	Evitar la nominalización sintáctica.
S.15	Evitar en una oración la coincidencia contigua de sintagmas encabezados con la misma preposición.
L.16	Evitar palabras vacías, polisémicas e innecesarias.
L.17	Evitar palabras de baja frecuencia.
L.18	Evitar extranjerismos y falsos amigos.
L.19	Evitar los localismos y el lenguaje figurado.
L.20	Evaluar el uso de la terminología.

Cuadro 11: Propuesta de reglas de LCL.

(Evitar párrafos construidos con un solo enunciado) no se aplique. Por su parte, entre las reglas sintácticas la aplicación de S14 (Preferir la construcción activa a la pasivas. Evitar las construcciones de SE) tiene normalmente como resultado una mayor nominalización del texto; por ello, la aplicación de S15 (Evitar la nominalización sintáctica) debe ser posterior para, en su caso, evitar esa tendencia.

En segundo lugar, se ha justificado la pertinencia de las normas y se ha especificado su procedimiento de aplicación. Todo ello se ha realizado desde una perspectiva multidisciplinar que bebe de aportaciones de los estudios de lenguaje claro y lecturabilidad, y de la gramática descriptiva y la normativa, pero también del análisis del discurso, la psicolingüística, la psicología cognitiva y los estudios de traducción automática y procesamiento del lenguaje natural. En lo relativo a la forma en la que deben aplicarse, se ha querido aportar propuestas viables y realistas en relación con su objetivo final, que es aumentar la eficiencia de los procesos de preedición de textos tratados con motores de traducción automática. Es esencial no perder de vista que la guía de aplicación está necesariamente asociada al par de lenguas definido (ES > EN) y a la gestión de un género textual en el que interactúan comunidades discursivas en desequilibrio (en este caso, la noticia periodística). Carece, por tanto, de validez universal.

Quedan pendientes dos cuestiones que no podemos afrontar en este trabajo por su complejidad y diversidad. La primera tiene que ver con la valoración del rendimiento real de las normas propuestas una vez aplicadas a textos concretos. Para ello ya está en marcha un análisis que sigue la metodología empleada en Fernández Vivanco & Bustos Gisbert (2025) y que se ha aplicado a dos de los textos empleados para esta investigación.

La segunda está asociada a la implementación de mecanismos de aplicación semiautomatizada de aquellas que demuestren relevancia. Esta línea está también en pleno desarrollo y supone tres retos fundamentales que ahora solo enunciamos. El primero tiene que ver con la selección de varios modelos LLM de inteligencia artificial generativa para valorar el rendimiento de cada uno de ellos; por el momento estamos trabajando con ChatGPT (Open AI), Gemini (Google) y Claude (Anthropic). El segundo persigue valorar el efecto de la llamada temperatura en inteligencia artificial en las respuestas aportadas por los modelos. Por último, es esencial validar los *prompts* creados con el objeto de optimizarlos y, en su caso, adaptarlos a las características propias de cada IA.

Es nuestro objetivo tratar de dar respuesta a los dos retos que ahora solo hemos enunciado en el marco del el proyecto de investigación en el que se incardina nuestra labor.

Agradecimientos

Este artículo se ha realizado en el marco delproyecto COMPLETEXT: Complejidad textual y lecturabilidad. Estudio aplicado a la IA generativa y la didáctica de las lenguas (Proyecto financiado por la JCyL FEDER-UE, SA067P24) y de un contrato predoctoral cofinanciado por la Junta de Castilla y León y el Fondo Social Europeo, al amparo de la Orden de 1 de diciembre de 2023, de la Consejería de Educación (BOCYL 233/2023).

Referencias

Azpiazu, Susana. 2025. El estilo nominal en el contexto de la lecturabilidad. *Revista Española De Lingüística* 55(2). 71–114. 10.31810/rsel.55.2.2

Bailin, Alan & Anne Grafstein. 2016. *Readability: Text and context*. Londres: Palgrave Macmillan. 10.1057/9781137388773

Bowker, Lynne. 2024. Plain language in the age of neural machine translation: An opportunity for translators. Em *New Advances in Translation Technology: Applications and Pedagogy*, 163–180. Singapore: Springer Nature. 10.1007/978-981-97-2958-6_9

Briva-Iglesias, Vincent & Isabel Peñuelas Gil. 2025. Simplifying healthcare communication: Evaluating AI-driven plain language editing of informed consent forms. Em *1st Workshop on Artificial Intelligence and Easy and Plain Language in Institutional Contexts (AI & EL/PL)*, 55–65.

Bustos Gisbert, José Manuel. 2012. Interpretación multifuncional del párrafo. *ELUA: Estudios de Lingüística. Universidad de Alicante* 26. 55–88. 10.14198/ELUA2012.26.03

Bustos Gisbert, José Manuel. 2014. El uso de conectores en hablantes nativos de español. Em *Procedimientos de conexión discursiva en español: adquisición y aprendizaje*, 147–195. Berna: Peter Lang

Bustos Gisbert, José Manuel. 2017. Naturaleza sintáctica de los enunciados textuales en el discurso escrito. *ELUA: Estudios de Lingüística. Universidad de Alicante* 31. 67–95. 10.14198/ELUA2017.31.04

Cuetos Vega, Fernando, Julio González Álvarez & Manuel de Vega Rodríguez. 2018. *Psicología del lenguaje*. Madrid: Editorial Médica Panamericana.

Da Cunha Fanego, Iria. 2024. Inteligencia artificial y lenguaje claro en español. Em *Lenguaje claro en Iberoamérica: Principios y prácticas*, 417–441. Thomson Reuters

Da Cunha Fanego, Iria & María Ángeles Escobar. 2021. Recomendaciones sobre lenguaje claro en español en el ámbito jurídico-administrativo: análisis y clasificación. *Pragmalingüística* 29. 129–148. 10.25267/Pragmalinguistica.2021.i29.07

Escobar, M. Ángeles & Eva Samaniego. 2022. Análisis contrastivo inglés-español de las recomendaciones sobre lenguaje claro en el ámbito jurídico-administrativo en los últimos 10 años dentro de la unión europea. Em *Lenguaje claro y tecnología en la Administración*, 15–40. Comares.

Espinosa Giménez, Miriam & Mikel L. Forcada. 2024. Lexical impoverishment in neural machine translation. *Tradumàtica* 22. 162–180. 10.5565/rev/tradumatica.428

Estopà Bagot, Rosa. 2021. Culturalización en salud: aportaciones desde la lingüística. *Cadernos de Linguística* 2(1). 1–22. 10.25189/2675-4916.2021.v2.n1.id244

Ezpeleta Piorno, Pilar & Anabel Borja Albi. 2017. The multilingual university website (MUW) genre ecology: content analysis and translation processes. *Revista española de lingüística aplicada* 30(2). 636–661. 10.1075/resla.00010.ezp

Fernández Vivanco, Andrea & José Manuel Bustos Gisbert. 2025. Los lenguajes controlados y su aplicación a la traducción automática: 20 años después. *Tradumàtica* 23. 36–64. 10.5565/rev/tradumatica.501

Fernández Vivanco, Andrea & Vincenzo Sanfilippo. 2025. La lecturabilidad de textos escritos en lengua española. *ELUA: Estudios de Lingüística. Universidad de Alicante* 44. 29–56. 10.14198/ELUA.29266

García Izquierdo, Isabel. 2011. New perspectives for the analysis and formalisation of specialised genres: The GENTT proposal. *Terminàlia* 3. 13–21. 10.2436/20.2503.01.21

García Izquierdo, Isabel. 2012. *Competencia textual para la traducción*. Valencia: Tirant lo Blanch

García Muñoz, Óscar. 2012. *Lectura fácil: métodos de redacción y evaluación*. Madrid: Real Patronato sobre Discapacidad. Ministerio de Sanidad, Servicios Sociales e Igualdad

García Palacios, Joaquín. 2022. El préstamo y los préstamos en español: hacia una teoría completa e integradora sobre los préstamos. Em *El préstamo como procedimiento de actualización léxica*, 1–24. Granada: Editorial Comares

Haase, Viviana, Maria Spychalska & Markus Werning. 2019. Investigating the comprehension of negated sentences employing world knowledge: An event-related potential study. *Frontiers in Psychology* 10. Article 2184. 10.3389/fpsyg.2019.02184

Hernando Cuadrado, Luis Alberto. 2005. La elipsis en el análisis e interpretación de textos.

CAUCE, Revista Internacional de Filología y su Didáctica 28. 169–181.

Hernández Palma, Jonathan, Eleazar Morales Vázquez & Andrés Arias de la Cruz. 2025. Entre el error y la estrategia: análisis contrastivo de falsos cognados frecuentes en inglés y español desde una perspectiva traductológica. *Estudios y Perspectivas Revista Científica y Académica* 5(2). 2665–2677. 10.61384/r.c.a..v5i2.1304

Horno Chéliz, María del Carmen. 2024. *Un cerebro lleno de palabras: descubre cómo influye tu diccionario mental en lo que sientes y piensas*. Barcelona: Plataforma Editorial

Kerremans, Koen & Rita Temmerman. 2016. Finding the (un)expected? quantitative and qualitative comparisons of term variants and their translations in a parallel corpus of EU texts. Em *Corpus-based Approaches to Translation and Interpreting: from Theory to Applications*, 43–63. Frankfurt am Main: Peter Lang. 10.3726/b10354

Lerner, Paul & François Yvon. 2025. Towards the machine translation of scientific neologisms. Em *31st International Conference on Computational Linguistics*, 947–963.

Ludewig, Ulrich, Nicolas Hübner & Sascha Schroeder. 2023. Vocabulary, text coverage, word frequency and the lexical threshold in elementary school reading comprehension. *Reading and Writing* 36. 2409–2431. 10.1007/s11145-022-10385-0

Martin, Andrea E. & Brian McElree. 2008. A content-addressable pointer mechanism underlies comprehension of verb-phrase ellipsis. *Journal of Memory and Language* 58(3). 879–906. 10.1016/j.jml.2007.06.010

Ministerio de Justicia. 2011. *Informe de la comisión de modernización del lenguaje jurídico*. Madrid: Ministerio de Justicia, Centro de Publicaciones.

Montiel, María Leticia, Valeria Magalí Colella & Nuria Gómez Belart. 2024. *Manual de lenguaje claro*. Buenos Aires: Secretaría Legal y Técnica. Gobierno de la Ciudad Autónoma de Buenos Aires

Montolío Durán, Estrella & Mario Tascón. 2020. *El derecho a entender. la comunicación clara, la mejor defensa de la ciudadanía*. Madrid: Prodigioso volcán / Catarata

Naveen, Palanichamy & Pavel Trojovský. 2024. Overview and challenges of machine translation for contextually appropriate translations. *iScience* 27(10). 110878. 10.1016/j.isci.2024.110878

O'Brien, Sharon. 2003. Controlling controlled English. Em *EAMT Workshop: Improving MT through other language technology tools: resources and tools for building MT*,

O'Brien, Sharon. 2004. Machine translatability and post-editing effort: How do they relate. Em *Translating and the Computer 26*,

Pinker, Steven. 1999. *El instinto del lenguaje: cómo crea el lenguaje la mente*. Madrid: Alianza Editorial. Traducción de José Manuel Igoa González

Porras Garzón, Jorge & Rosa Estopà Bagot. 2020. Escalas de legibilidad aplicadas a informes médicos. *Círculo de Lingüística Aplicada a la Comunicación* 83. 205–216. 10.5209/clac.70574

Portolés, José. 1993. La distinción entre los conectores y otros marcadores del discurso en español. *Verba* 20. 141–170.

RAE & ASALE. 2020. *Glosario de términos gramaticales*. Salamanca: Ediciones Universidad de Salamanca

RAI & ASALE. 2010. *Ortografía de la lengua española*. Madrid: Espasa.

Ramírez Polo, Laura. 2012. Los lenguajes controlados y la documentación técnica: mejorando la traducibilidad. *Tradumàtica* 10. 192–204. 10.5565/rev/tradumatica.25

Romanik, Wojciech, Tomasz Piekot & Radosław Tymiński. 2022. Adaptation of informed consent for anaesthesia to plain language standard. *Nowotwory Journal of Oncology* 72(4). 265–273. 10.5603/NJO.a2022.0034

Sologuren Insúa, Enrique & María Natalia Castillo Fadić. 2018. El orden de palabras en los estudios gramaticales: un recorrido a través de diferentes perspectivas lingüísticas. *Nueva revista del Pacífico* 68. 123–146. 10.4067/S0719-51762018000100123

Van Dijk, Teun. 2000. El discurso como interacción en la sociedad. Em *El discurso como interacción social*, 19–66. Barcelona: Gedisa

Vandevoorde, Lore & Els Lefever. 2023. Who's afraid of false friends? Cognate ratios in translated and non-translated Dutch. *Across Languages and Cultures* 24(1). 73–84. 10.1556/084.2022.00204