

II Encuentro: "El español en el entorno digital: accesibilidad, representación y desarrollo de tecnologías del lenguaje"



OBSERVATORIO
NEBRIJA DEL
ESPAÑOL

**Lenguaje
accesible y
textos reglados:
del enfoque
correctivo a la
reutilización
estructural**

María Xesús Bello Rivas



Junio 2026



UNIVERSIDAD
NEBRIJA

Lenguaje accesible y
textos reglados:
del enfoque correctivo a
la reutilización estructural

María Xesús Bello Rivas



© Universidad Nebrija

DOI: 10.5281/zenodo.21493980



Cualquier forma de reproducción, distribución, comunicación pública o transformación de esta obra solo puede ser realizada con la autorización de sus titulares, salvo excepción prevista por la ley. Diríjase a CEDRO (Centro Español de Derechos Reprográficos) si necesita fotocopiar o escanear algún fragmento de esta obra ([www. conlicencia. com](http://www.conlicencia.com); 91 702 19 70 / 93 272 04 45).

Índice



GLOSARIO	6
1. EL OBJETO: textos reglados, la comprensión como obligación	8
2. EL DIAGNÓSTICO: de los principios al método	8
2.1 Lo que la norma establece	8
2.2 El enfoque correctivo y su límite	9
3. LA EVIDENCIA: lo que la investigación empírica revela	9
3.1 La inteligencia artificial y la simplificación automática	9
3.2 Evidencia arquitectónica desde el corpus AIP	10
3.3 La distinción que la intervención superficial no ve	10
4. LA PROPUESTA: BELA™, de los principios a las estructuras	11
4.1 Las macrofunciones discursivas	12
4.2 Lo que los LLM no pueden hacer sin arquitectura previa	13
4.3 El proceso: de la arquitectura al banco	14
4.4 La metodología de anotación	15
5. LA PRUEBA: Madrid Comprensible, diagnóstico sobre corpus AIP (Acceso a la Información Pública)	16
5.1 El corpus piloto	16
5.2 Criterios operativos de la evaluación	16
5.3 Decisiones de diseño del diagnóstico automático	17
6. EL ALCANCE: formalizar para transferir	18
6.1 De la corrección a la infraestructura	18
6.2 Transferibilidad del marco	18
6.3 Del banco al despliegue: la herramienta de generación asistida	19
Bibliografía	20

Resumen¹

El ordenamiento jurídico impone, de manera creciente, que determinados textos —resoluciones administrativas, contratos, prospectos, comunicaciones de servicios regulados— resulten comprensibles para sus destinatarios. Los estándares formulan los principios del lenguaje claro (UNE-ISO 24495-1:2024) y de la lectura fácil (UNE 153101:2018 EX); lo que no existe es un método para convertirlos en criterios concretos y reutilizables en géneros cuya estructura está condicionada por norma jurídica o convención institucional.

Este trabajo propone ese método: analiza cómo están organizados los textos reglados, formaliza la función de cada uno de sus bloques y construye un banco piloto de estructuras que permite producir versiones accesibles de forma sistemática, verificable y trazable. La propuesta se desarrolla y valida sobre un piloto con resoluciones de acceso a la información pública del Ayuntamiento de Madrid.

GLOSARIO

banco. Colección de microframes atestiguados en un corpus real, con trazabilidad a su documento de origen y a su diagnóstico. No es una biblioteca de plantillas: cada entrada existe porque un diagnóstico la ha justificado.

diagnóstico. Evaluación del estado de cada macrofunción en un texto concreto mediante criterios verificables, previa a cualquier intervención de reescritura.

¹ Agradezco las aportaciones surgidas del debate en el II Encuentro de la línea «El español en el entorno digital» (Observatorio Nebrija del Español, Universidad Nebrija, Madrid, 22 de junio de 2026), y en particular las de Cristina Carretero González (Universidad Pontificia Comillas), Óscar García Muñoz (Dilofácil), Ana Medina Reguera (Universidad Pablo de Olavide), Lourdes Moreno López (Universidad Carlos III de Madrid), Claudia Poblete Olmedo (Pontificia Universidad Católica de Valparaíso) y Mari Carmen Suárez-Figueroa (Universidad Politécnica de Madrid), que han corregido y precisado varios puntos.

enfoque correctivo. Denominación empleada en este trabajo para la práctica dominante de intervención en lenguaje claro: la que actúa sobre la formulación lingüística de un texto ya redactado — léxico, sintaxis, longitud de los párrafos, tratamiento de las referencias normativas— y que, cuando aborda la organización del contenido, lo hace mediante recomendaciones generales no derivadas del análisis del género concreto. Mejora la legibilidad, pero no verifica que cada bloque cumpla la función que el destinatario necesita, porque carece de criterios explícitos por género para comprobarlo. No se opone al lenguaje claro ni a su normativa: opera en un nivel distinto del análisis arquitectónico previo que aquí se propone.

estructura base. Formulación abstracta de un fragmento textual, despojada de los datos del caso y conservando la secuencia y el encadenamiento propios de la macrofunción a la que sirve.

etiquetas del diagnóstico (OK, PARCIAL, FALLO, AUSENTE). OK: la función está presente y formulada en lenguaje claro. PARCIAL: está presente, pero no en lenguaje claro. FALLO: está presente y gravemente deficiente (campos sin rellenar, datos incoherentes, remisiones vacías). AUSENTE: no aparece en el texto. PARCIAL es, por tanto, un juicio sobre la formulación, no sobre la presencia.

macrofunción. Unidad de la arquitectura global del texto que responde a una pregunta del destinatario y agrupa los fragmentos que cumplen esa misma finalidad discursiva. Se sitúa en el nivel de la macroestructura (van Dijk y Kintsch, 1983), no en el de las metafunciones de la lingüística funcional sistémica (Halliday y Matthiessen, 2014), que operan de forma simultánea en cada enunciado. Las macrofunciones no forman un catálogo fijo: se inducen del corpus de cada género. Las cuatro de este trabajo —EXPONER, DECIDIR, FUNDAMENTAR, ORIENTAR— son propias de las resoluciones AIP.

microframe. Conjunto de estructuras base con sus slots, asociado a una combinación concreta de rasgos del documento. En el corpus AIP, esa combinación es tipo de resolución, causa jurídica y alcance. Es la unidad recuperable del banco.

P-criterio. Pregunta de verificación de respuesta binaria que se aplica en orden dentro de una macrofunción y de cuyo conjunto emerge la etiqueta de diagnóstico, que no puede asignarse por apreciación global.

slot. Posición marcada dentro de una estructura base que solo admite un tipo determinado de contenido. El término procede de la semántica de marcos (Fillmore) y se conserva sin traducir por ser el habitual en procesamiento del lenguaje natural; no equivale a campo de formulario, que carece de esa restricción semántica.

texto reglado. Documento cuya estructura y cuyo contenido están condicionados por una norma jurídica o una convención institucional que, además, exige su comprensibilidad para el destinatario. Resoluciones administrativas, contratos de consumo, prospectos de medicamentos o notificaciones tributarias son textos reglados: parte de su contenido y de su disposición es exigible, de modo que no admiten reescritura libre.

1. EL OBJETO: textos reglados, la comprensión como obligación

Una resolución administrativa que el destinatario no puede entender no cumple su función comunicativa aunque sea jurídicamente correcta; un contrato cuyas cláusulas resultan opacas para el consumidor no es propiamente un acto de comunicación. En ambos casos, y en otros muchos que comparten la misma lógica —prospectos de medicamentos, notificaciones tributarias, escrituras de propiedad—, la norma jurídica no solo determina el contenido y la forma del texto, sino que exige su comprensibilidad para el destinatario. Esta categoría de textos, que llamaremos textos reglados, constituye el objeto de este trabajo.

El problema que plantean estos textos no es de principios, ya que está bien establecido qué debe cumplir un texto para ser accesible a su destinatario; el problema es más bien de método: cómo aplicarlos de forma sistemática a géneros cuya arquitectura está condicionada por obligaciones jurídicas que no pueden ignorarse. Entre la norma que establece el qué y la práctica que determina el cómo hay un vacío de naturaleza arquitectónica.

2. EL DIAGNÓSTICO: de los principios al método

2.1 Lo que la norma establece

Para que un texto reglado sea accesible a su destinatario, la norma UNE-ISO 24495-1:2024 fija cuatro principios: que sea relevante (que contenga la información que esa persona necesita), localizable (que la organice de forma que pueda encontrarla), comprensible (que la exprese de forma que pueda entenderla) y utilizable (que le permita actuar a partir de ella). No son principios de estilo, sino el estándar internacional de referencia en la materia.

Estos cuatro principios no derivan de una teoría unificada de la lectura, sino de un consenso profesional: la definición que la International Plain Language Federation fijó en 2014 y que la ISO formalizó en 2023. Su sustrato cognitivo es parcial —la teoría de la carga cognitiva y la fluidez de procesamiento respaldan sobre todo la comprensibilidad, es decir, el trabajo sobre el léxico y la sintaxis—, mientras que los otros tres —relevante, localizable y utilizable— pertenecen más al diseño de la información y al modelo de tareas del destinatario. De ahí una paradoja que la propia Federación reconoce: plain language es un nombre engañoso, porque solo en torno al 7 % de las recomendaciones de su estándar se ocupa de la elección de palabras; el resto atañe a la estructura y al diseño. Sin embargo, cuando se implementa el lenguaje claro —a través de guías, formación y práctica correctiva— se interviene casi siempre sobre la frase y la palabra. El nivel que el estándar señala como más determinante es, precisamente, el que menos se sabe operacionalizar.

En el contexto de los textos reglados, estos principios se traducen en preguntas que el texto debe responder de forma accesible. Las preguntas concretas varían según el género —una resolución administrativa, un contrato o una receta médica plantean necesidades distintas a su destinatario—, pero responden siempre a la misma lógica: qué se me está diciendo, qué se decide o establece, por qué, y qué puedo o debo hacer.

2.2 El enfoque correctivo y su límite

La aplicación de esos principios se realiza, de forma predominante, sobre la forma lingüística del texto: se sustituye léxico especializado, se acortan párrafos, se reordenan elementos sintácticos, se simplifican referencias normativas. Las guías de estilo y los sistemas de asistencia automática disponibles operan mayoritariamente en ese nivel y, cuando abordan la organización del contenido, lo hacen mediante recomendaciones generales — situar lo importante al principio, agrupar la información relacionada, encabezar las secciones— que no derivan del análisis del género concreto. El resultado es un texto más legible; el límite es que ninguna de estas intervenciones dispone de criterios verificables, propios del género, para comprobar qué función cumple cada bloque del documento para el destinatario —qué se le comunica, qué efecto jurídico se produce, por qué razón, qué puede hacer al respecto—, y sin ese análisis funcional previo la intervención sobre la superficie no puede garantizar que el texto sea comprensible para quien lo recibe.

La teoría de la macroestructura de van Dijk y Kintsch (1983) establece que la comprensión de un texto requiere la construcción de una representación mental de su estructura global, no solo el procesamiento de sus oraciones individuales. Un texto puede tener oraciones cortas y léxico accesible y, sin embargo, ser difícil de comprender si su macroestructura no responde a las expectativas informativas del lector. Masson y Waldron (1994) muestran que esto se aplica específicamente a los textos legales: la comprensibilidad depende de la estructura global del documento —qué información aparece primero, cómo se organiza la progresión temática, qué relaciones lógicas se hacen explícitas—, no solo de sus rasgos lingüísticos locales. Zódi (2019), desde el análisis del derecho continental europeo, concluye que la opacidad de los textos jurídicos es en gran medida estructural y que las intervenciones de lenguaje claro que no abordan esa dimensión producen mejoras limitadas. Fernández-Silva y Núñez Cortés (2025) lo confirman empíricamente en español administrativo, aunque su experimento se limita al nivel terminológico.

3. LA EVIDENCIA: lo que la investigación empírica revela

3.1 La inteligencia artificial y la simplificación automática

En los últimos años, la disponibilidad de modelos de lenguaje de gran escala ha generado expectativas sobre la posibilidad de automatizar la producción de textos en lenguaje claro y lectura fácil. La idea es atractiva: si un modelo puede resumir, parafrasear y adaptar el registro de un texto en segundos, ¿por qué no habría de producir directamente una versión accesible? La lectura fácil permite poner a prueba esa expectativa con especial claridad, porque su norma —la UNE 153101:2018 EX— es explícita, detallada y verificable: establece sin ambigüedad qué debe cumplir un texto, lo que hace posible comprobar con criterios objetivos si un modelo lo respeta o no.

La evidencia disponible muestra que los modelos no resuelven por sí solos esta tarea de forma fiable, y que una parte decisiva del límite no es técnica sino conceptual. Madina, González-Dios y Siegel (2024) hicieron esa comprobación para el español y concluyeron que ChatGPT-4 simplifica el léxico y la sintaxis, pero no sigue las reglas estructurales ni

adapta el contenido a las necesidades cognitivas del destinatario. El texto resultante parece más simple, pero no funciona para quien lo tiene que leer.

Ese límite no es una particularidad de la lectura fácil ni de un modelo concreto. Los trabajos reunidos en el Joint Workshop on Readability and Text Simplification (READixTSAR, LREC 2026) lo encuentran en distintas lenguas y dominios: los modelos resuelven lo visible —léxico, sintaxis— y dejan intacto lo que determina si el texto cumple su función, que es cómo está construido como documento.

La tabla de la sección 4.2 detalla tres estudios que convergen en este punto desde ángulos distintos.

3.2 Evidencia arquitectónica desde el corpus AIP

En el género analizado, la resolución administrativa articula cuatro macrofunciones: expone de qué trata el asunto, fundamenta la decisión, decide, y orienta al destinatario sobre qué puede hacer. El examen de las resoluciones de acceso a la información pública del Ayuntamiento de Madrid —datos abiertos del Portal de Transparencia— revela que el contenido de las cuatro funciones casi siempre está presente, pero su disposición y su formulación lo vuelven inaccesible para el ciudadano:

- En la exposición, el objeto de la solicitud es casi siempre identificable, pero la presentación es deficiente: campos de plantilla sin rellenar, fechas incoherentes, maquetación que mezcla bloques.
- En la fundamentación, la base jurídica es correcta, pero el motivo real de la denegación solo se formula en lenguaje técnico, y a menudo la resolución remite la justificación detallada a un informe adjunto.
- La decisión aparece con verbo explícito, pero diluida entre información procesal y formulada en lenguaje técnico, sin una expresión clara de qué se resuelve para el destinatario.
- La orientación —las vías de recurso— está siempre presente, con plazos y órganos, pero redactada en tercera persona y con referencias normativas que el destinatario no puede seguir.

El patrón es siempre el mismo: el contenido exigible está, no así su accesibilidad.

3.3 La distinción que la intervención superficial no ve

La distinción entre legibilidad y comprensibilidad es aquí fundamental. La legibilidad es una propiedad del texto medible mediante índices cuantitativos; la comprensibilidad es una propiedad de la relación entre el texto y su destinatario, y depende en gran medida de que la información esté organizada según las expectativas y necesidades cognitivas de ese destinatario.

Esta distinción es más operativa que meramente teórica. Señala exactamente dónde falla la intervención correctiva y qué tipo de herramienta hace falta para ir más allá. Lo que se necesita no es un criterio mejor para corregir la superficie, sino un método para analizar, formalizar y transformar la arquitectura del texto.

4. LA PROPUESTA: BELA™, de los principios a las estructuras

BELA™ —Banco de Estructuras de Lenguaje Accesible— es un método de formalización de la arquitectura discursiva de textos reglados. Analiza cómo está organizado un documento, formaliza la función de cada uno de sus bloques y construye con ellos un banco de estructuras reutilizables, clasificadas por género y causa jurídica, que permite producir versiones accesibles de forma sistemática, verificable y trazable. El banco es su producto: lo que se transfiere, se audita y se reutiliza. El método se desarrolla en el marco del proyecto "Accesibilidad web en español a través de la lectura fácil" (<https://accesofacil.es>), del que procede también nuestra obra: *Lectura fácil y web comprensible. Un libro blanco en el marco del nuevo reglamento*.

En ese proceso BELA™ emplea un modelo de lenguaje en dos funciones distintas que conviene no confundir. La primera es analítica: el modelo diagnostica la arquitectura del texto y extrae sus estructuras, sin redactar nada nuevo; la segunda es generativa y opera en dos momentos distintos del proceso. Durante la construcción del banco, el modelo produce a partir del texto original la versión en lenguaje claro (UNE-ISO 24495-1:2024) y el borrador de lectura fácil (UNE 153101:2018 EX), guiado por el diagnóstico y por la estructura previamente abstraída; de esas versiones se extraen las estructuras accesibles que ingresan en el banco. Una vez el banco está anotado y validado, el modelo genera los documentos nuevos operando sobre esas estructuras formalizadas y no sobre la superficie de un texto original. En ambos momentos la generación va precedida de análisis y seguida de validación humana. La diferencia con el enfoque correctivo no está en usar o no inteligencia artificial generativa, sino en que BELA™ la somete a una estructura previa en lugar de aplicarla directamente al texto sin análisis previo.

El análisis de género en textos jurídicos y administrativos tiene una tradición consolidada. Bhatia (1993, 2004) establece que los géneros profesionales y jurídicos poseen una integridad genérica —una constelación estable de correlaciones forma-función— que determina cómo se organiza el texto y qué función cumple cada bloque en relación con su propósito comunicativo. Swales (1990) sistematiza el concepto de *move* como unidad retórica con función discursiva específica. Ambas tradiciones describen la arquitectura de los textos jurídicos con precisión analítica, pero ninguna da el paso de la descripción a la formalización operativa. El análisis de género identifica qué hace cada parte de un texto; BELA™ convierte ese conocimiento en estructuras reutilizables, clasificadas por género y causa jurídica, que permiten producir versiones accesibles de manera sistemática. Es la diferencia entre cartografiar un territorio y construir infraestructura sobre él.

La misma distinción separa a BELA™ de las ontologías jurídicas conceptuales —LKIF, LRI-Core²—, que modelan entidades y relaciones: normas, obligaciones, procedimientos, sujetos. Son ontologías del *qué*. BELA™ opera en un nivel distinto: modela la estructura y función del discurso, cómo se construye el texto jurídico como género, qué bloques lo componen y en qué posición deben aparecer para que el destinatario pueda comprenderlo. Opera en el nivel del *cómo*. Ese nivel —la formalización discursiva aplicada al derecho— está escasamente desarrollado frente a las ontologías conceptuales y no hemos localizado, en la bibliografía examinada, una formalización equivalente para el español administrativo con orientación a la accesibilidad.

² LKIF (Legal Knowledge Interchange Format): formato para representar conocimiento jurídico de forma computacional, desarrollado en el proyecto europeo ESTRELLA. LRI-Core: ontología jurídica conceptual desarrollada en la Universidad de Leiden que clasifica entidades del derecho (actos, agentes, estados, procesos). Ambas modelan conceptos jurídicos, no arquitectura discursiva.

BELA™ se propone como una primera instancia empírica en ese espacio: un marco construido desde corpus real y con criterios operativos explícitos.

4.1 Las macrofunciones discursivas

BELA™ trabaja con criterios funcionales de diagnóstico específicos de cada género textual, inducidos de su corpus. En el piloto AIP-Madrid, el análisis funcional de las resoluciones de acceso a la información pública ha dado lugar a cuatro macrofunciones discursivas, que responden cada una a una pregunta del destinatario. La tabla siguiente resume esa correspondencia:

Macrofunción	Pregunta del destinatario
EXPONER	¿Qué se me está diciendo?
DECIDIR	¿Qué se ha decidido?
FUNDAMENTAR	¿Por qué se ha decidido así?
ORIENTAR	¿Qué puedo hacer?

Sobre esa identificación de bloques, BELA™ operacionaliza en criterios verificables los cuatro principios de la UNE-ISO 24495-1:2024 (figura 1): relevante (incluir solo la información que el destinatario necesita para entender la decisión y actuar), localizable (situar lo esencial al inicio y agrupar la información relacionada en un orden de bloques estable), comprensible (léxico preciso, frases en voz activa con estructura sujeto-verbo-objeto, un tema por párrafo) y utilizable (que el destinatario pueda actuar: vías de recurso en segunda persona, con verbo de acción, plazo y órgano). Las macrofunciones definen qué bloque es cada parte del texto; los principios determinan si ese bloque está formulado de forma accesible. El cruce de ambos ejes constituye la matriz ISO × macrofunciones, que convierte cada principio general de la norma en un criterio verificable sobre un bloque textual determinado y sobre la que opera el diagnóstico (sección 5.2).

Operacionalización de UNE-ISO 24495-1:2024 en BELA™				
BELA — AIP - resoluciones de acceso a la información				
	EXPONER	DECIDIR	FUNDAMENTAR	ORIENTAR
Relevante La información necesaria está	¿Consta la fecha de entrada? ¿Está identificado el objeto? ¿Aparece el canal de respuesta? ¿Hay campos sin rellenar?	¿Aparece la decisión con verbo explícito? ¿Identifica con claridad qué se decide?	—	—
Localizable Se puede encontrar	—	—	¿Aparece razón concreta en el cuerpo principal? ¿Aparece razón concreta en el informe adjunto?	—
Comprensible Se puede entender	¿El objeto es comprensible sin conocimiento del procedimiento?	¿La decisión es comprensible sin conocimiento jurídico?	¿La razón está en lenguaje claro, sin jerga jurídica?	¿Plazos y órganos están en lenguaje claro?
Utilizable Se puede actuar	—	—	—	¿Hay al menos un plazo concreto? ¿Hay al menos un órgano identificado?

Las celdas vacías indican que ese principio no requiere criterio de verificación explícito en esa macrofunción.

El orden propuesto: EXPONER → DECIDIR → FUNDAMENTAR → ORIENTAR responde al esquema cognitivo del destinatario: primero que reconozca el contexto, luego que sepa qué se ha decidido, luego que comprenda por qué, finalmente que sepa qué hacer. Este orden difiere del convencional heredado de la práctica judicial (antecedentes → fundamentos → dispositiva): el orden judicial sí está legalmente prescrito —artículo 248.4 de la LOPJ—; el administrativo no: el artículo 88 de la Ley 39/2015 únicamente exige que la resolución contenga la decisión motivada y la indicación de recursos. BELA™ no sustituye ni reordena la resolución administrativa: propone incorporar, antes del desarrollo jurídico completo, una capa funcional de orientación inicial que permite al destinatario identificar de forma inmediata el contexto, la decisión, la razón principal y las vías de actuación, manteniendo íntegro el contenido jurídico exigible.

4.2 Lo que los LLM no pueden hacer sin arquitectura previa

La propia experiencia de diseño del diagnóstico de BELA™ confirma lo que señala la investigación reciente: sin criterios arquitectónicos explícitos, los modelos no distinguen entre la presencia de un contenido y su accesibilidad (la sección 5.3 documenta este patrón de error y la solución adoptada). Los tres estudios siguientes convergen en ese punto desde ángulos distintos.

Estudio	Objeto	Límite identificado
Madina, González-Dios y Siegel (2024)	Adaptación automática al español en Lectura Fácil (ChatGPT-4, norma UNE 153101:2018 EX)	El modelo simplifica léxico y sintaxis pero no sigue las restricciones estructurales ni adapta el contenido a las necesidades cognitivas del destinatario. El texto parece más simple; no funciona para quien lo tiene que leer
Mathas, Bakker y Kamps (2026)	Simplificación guiada por plan explícito a nivel de oración (corpus Newsela)	La planificación frase a frase no es equivalente al análisis de la arquitectura del documento: la precisión en el seguimiento del plan cae de 70,8 % a 44,7 % al ampliar el contexto de uno a cinco párrafos; el modelo sabe qué hacer con cada oración y pierde la coherencia funcional del conjunto
Cardon y Doğruöz (2026)	Efecto del género textual y del perfil del destinatario sobre cinco modelos de lenguaje en simplificación médica en francés	Los modelos reproducen la arquitectura del texto que reciben sin analizarla explícitamente: el género del input impacta el output más que las instrucciones sobre el destinatario incluidas en el prompt

En el espacio iberoamericano, el proyecto Human_AI (Moreno López y Martínez Fernández, UC3M, en curso; humanai.uc3m.es) desarrolla un sistema multiagente con supervisión humana para la generación de texto en lenguaje claro y lectura fácil. Su enfoque —IA como proceso controlado, evaluable y gobernado, no como caja negra— es convergente con el principio de formalización previa que BELA™ operacionaliza en el nivel arquitectónico: el modelo opera sobre estructura formalizada externamente, no sobre texto sin estructura previa.

En el mismo ámbito, Suárez-Figueroa, Diab, Ruckhaus y Cano (2024) desarrollan una aplicación de apoyo a la adaptación a Lectura Fácil en español que asiste al adaptador humano mediante pautas y criterios verificables. Su enfoque confirma que la asistencia automática solo opera con eficacia cuando los criterios de adaptación están formalizados externamente al modelo.

Lo que BELA™ añade es el esquema que estos trabajos señalan como ausente: una formalización de las macrofunciones discursivas que permite determinar qué va primero, por qué y con qué consecuencias para el destinatario. Una vez formalizado, la IA puede operar de manera controlada. La secuencia es: análisis arquitectónico → formalización en estructuras discursivas³ → generación asistida → validación humana. La inteligencia artificial interviene en dos momentos —analítico en el diagnóstico y la extracción, generativo en la producción de las versiones accesibles—, pero nunca antes de que la arquitectura esté formalizada: lo que el enfoque correctivo hace primero, BELA™ lo hace después.

4.3 El proceso: de la arquitectura al banco

La figura siguiente resume el flujo completo, que se ordena en cinco fases. El diagnóstico funcional determina qué macrofunción cumple cada bloque y en qué estado se encuentra. La abstracción estructural extrae de cada bloque su estructura base, despojada de los datos del caso. La generación produce la versión en lenguaje claro y el borrador de lectura fácil a partir del texto original, guiada por el diagnóstico y por la estructura abstraída, y de ambas versiones se extraen las estructuras accesibles que ingresan en el banco. La anotación y la validación las revisan. La redacción, por último, instancia las estructuras ya validadas con los datos de cada caso nuevo. El modelo interviene en el diagnóstico, en la abstracción y en la generación; la anotación, la validación y la decisión de incorporar una estructura al banco son operaciones humanas.



³ *Estructura discursiva*: unidad mínima del banco BELA™. Registra un patrón del texto original, su versión en lenguaje claro y las restricciones que garantizan su accesibilidad, clasificados por macrofunción y causa jurídica.



4.4 La metodología de anotación

El marco de anotación se organiza en tres capas que representan un continuo de accesibilidad creciente: el texto original, la versión en lenguaje claro y el borrador de lectura fácil. La norma de referencia para la capa de lenguaje claro es la UNE-ISO 24495-1:2024; para la capa de lectura fácil, la UNE 153101:2018 EX.

La arquitectura de estructuras discursivas permite la generación controlada de versiones accesibles sin improvisación por parte de los sistemas de IA. Cada estructura contiene los referentes léxicos controlados para asegurar la coherencia terminológica, las restricciones explícitas de la versión en lenguaje claro, y los criterios de validación para la lectura fácil.

Capa	Descripción	Norma
ORIGINAL	Fragmento literal del corpus. No se modifica.	Datos abiertos institucionales
LENGUAJE CLARO	Versión con intervención mínima. El contenido jurídico permanece intacto.	UNE-ISO 24495-1:2024
LECTURA FÁCIL (borrador)	Generada mediante IA a partir del texto original, aplicando la norma UNE 153101:2018 EX de forma independiente a la versión en lenguaje claro. Requiere validación con usuarios reales.	UNE 153101:2018 EX

5. LA PRUEBA:

Madrid Comprensible, diagnóstico sobre corpus AIP (Acceso a la Información Pública)

5.1 El corpus piloto

El piloto de referencia se aplica a resoluciones de acceso a la información pública del Ayuntamiento de Madrid disponibles como datos abiertos en el Portal de Transparencia. El corpus piloto comprende 49 resoluciones (2018-2026) correspondientes a las cuatro causas jurídicas con mayor representación en el registro AIP: protección de datos personales, prevención de ilícitos penales, intereses económicos y comerciales, e igualdad de partes en proceso judicial. La selección se estratificó por causa jurídica y año, y se cerró por saturación estructural: se examinaron alrededor de 200 resoluciones hasta que las nuevas dejaron de aportar patrones discursivos no observados, y el corpus de referencia quedó fijado en esas 49. Del análisis de ese corpus se derivaron las macrofunciones y los criterios operativos del diagnóstico.

La tabla siguiente desglosa la composición del corpus por causa jurídica, con su base legal y el número de resoluciones correspondiente:

Causa jurídica	Base legal	Resoluciones
Protección de datos personales	Art. 15 LTAIBG · RGPD 2016/679 · LO 3/2018	17
Prevención de ilícitos penales	Art. 14.1.e LTAIBG	14
Intereses económicos y comerciales	Art. 14.1.h LTAIBG	9
Igualdad de partes en proceso judicial	Art. 14.1.f LTAIBG	9

El diagnóstico se ha implementado en una herramienta interactiva que opera sobre este corpus, presentada en el Encuentro del 22 de junio. Puede consultarse una demostración en <https://accesofacil.es/bela-presentacion/#demo-video>.

5.2 Criterios operativos de la evaluación

El sistema de diagnóstico evalúa cada resolución en las cuatro macrofunciones con criterios operativos derivados del análisis del corpus real. Para cada función, el diagnóstico distingue cuatro estados: OK (la función está presente y formulada en lenguaje claro), PARCIAL (presente pero no en lenguaje claro), FALLO (presente pero gravemente deficiente: campos sin rellenar, datos incoherentes, remisiones vacías) y AUSENTE (la función no aparece en el texto).

Dos criterios críticos derivados del análisis del corpus:

- **FUNDAMENTAR:** el diagnóstico toma en consideración el texto completo, incluido el informe adjunto de la unidad informante cuando lo haya: la delegación de la fundamentación al informe no equivale a su ausencia.

- **ORIENTAR:** FALLO o AUSENTE solo se aplican cuando faltan el plazo o el órgano; la formulación técnica con plazos y órganos presentes es PARCIAL.

El examen del corpus muestra un patrón constante: ninguna de las resoluciones examinadas resuelve de forma accesible las cuatro macrofunciones. El contenido jurídicamente exigible está presente en prácticamente todos los casos; lo que falla es su formulación y su disposición. Esto confirma sobre el corpus lo observado en la sección 3.2: el problema no es de omisión, sino de arquitectura.

5.3 Decisiones de diseño del diagnóstico automático

El sistema de diagnóstico opera mediante un modelo de lenguaje grande (*large language model*, LLM)⁴ al que se le instruye para aplicar los criterios operativos BELA™ sobre cada resolución. La experiencia de aplicación sobre el corpus AIP ha revelado un patrón de error sistemático que tiene consecuencias metodológicas directas.

Sin un protocolo de razonamiento explícito, el modelo tiende a colapsar dos propiedades distintas del texto: la *presencia* de un contenido y su *accesibilidad*. Un objeto de solicitud formulado en lenguaje complejo pero perfectamente identificable es diagnosticado como ausente porque el modelo procesa dificultad como equivalente a ausencia. El resultado es un diagnóstico de FALLO donde el criterio operativo prescribe PARCIAL.

Este error no es aleatorio: es estructural y predecible. Emerge cada vez que el modelo infiere el estado directamente, sin descomponer la evaluación en preguntas binarias previas. La solución adoptada es un *protocolo de decisión secuencial basado en P-criterios*: antes de asignar cualquier estado, el modelo debe responder en orden una secuencia de preguntas de verificación binaria para cada macrofunción. La etiqueta emerge del conjunto de respuestas y no puede contradecirlas: no se asigna por apreciación global, sino que se deriva de comprobaciones auditables una a una. Así, si el modelo ha confirmado que el objeto de la solicitud está presente, no puede concluir FALLO en EXPONER independientemente del grado de complejidad lingüística del texto.

El hallazgo es específico y verificable: sin criterios formalizados externamente al modelo, el diagnóstico automático no distingue de forma fiable entre *inaccesible* y *ausente*. La formalización previa es, por tanto, condición de posibilidad del diagnóstico automático, no un complemento opcional.⁵

Una segunda mejora introduce *ejemplos de diagnóstico correcto (few-shot examples)*⁶ directamente en el protocolo de instrucción al modelo. Estos ejemplos, extraídos del corpus AIP, anclan el razonamiento del modelo en casos reales con diagnóstico conocido, reduciendo la variación entre ejecuciones sobre textos similares.

⁴ *Large language model* (LLM): sistema de inteligencia artificial entrenado sobre grandes volúmenes de texto que genera respuestas en lenguaje natural a partir de instrucciones. En BELA™ se utiliza como motor de diagnóstico y generación, siempre bajo instrucciones formalizadas externamente.

⁵ Este punto está documentado empíricamente en Madina, González-Dios y Siegel (2024) para la lectura fácil en español, y en Humm et al. (2025) para los límites de la autonomía de los sistemas de IA en tareas de adaptación lingüística. Las referencias completas figuran en la bibliografía del documento.

⁶ *Few-shot examples* (ejemplos de referencia): casos con respuesta correcta conocida que se incluyen en las instrucciones al modelo para que aprenda el patrón esperado por imitación, no solo por descripción abstracta. Es una técnica estándar de ingeniería de instrucciones (*prompt engineering*).

6. EL ALCANCE: formalizar para transferir

6.1 De la corrección a la infraestructura

Una vez formalizada la arquitectura discursiva de un género regulado para una causa jurídica determinada, las versiones accesibles de los documentos correspondientes pueden generarse de manera controlada, con criterios verificables y con trazabilidad al corpus real. La validación humana —que sigue siendo necesaria e irremplazable— se apoya en las estructuras del banco como criterio estable, sin partir de cero en cada documento.

El proyecto de norma ISO/DIS 24495-4 desplaza el lenguaje claro desde el plano de la recomendación estilística hacia el de la gobernanza comunicativa. Esta norma no fija solo criterios de redacción, sino que exige que las organizaciones dispongan de procesos, recursos, responsabilidades y evidencias que permitan aplicar de forma sistemática los principios de la UNE-ISO 24495-1:2024. En esta lógica, BELA™ se presenta como una herramienta metodológica alineada con los requisitos que la norma en desarrollo establece: un sistema de estructuras lingüísticas reutilizables, funciones comunicativas operativas, criterios de evaluación y evidencias de intervención sobre textos administrativos.

6.2 Transferibilidad del marco

El marco está diseñado desde su origen como metodología transferible a otros corpus institucionales, otros géneros documentales y otras lenguas. La condición para la transferencia es la disponibilidad de un corpus de datos abiertos institucionales en la lengua de destino y la revisión de los criterios de diagnóstico para incorporar las especificidades del género y la lengua correspondientes.

El banco, una vez construido para una tipología, no sustituye el método que lo generó. La normativa cambia y genera tipos textuales nuevos que requieren análisis arquitectónico antes de poder ser formalizados. El banco exige auditoría: verificar que un texto producido a partir de sus estructuras cumple efectivamente la arquitectura no es una operación automática sino un juicio lingüístico experto. Y el perímetro sectorial no está cerrado: el banco de resoluciones AIP no es transferible sin adaptación a contratos de consumo, a comunicaciones de telecomunicaciones o a prospectos farmacéuticos. Cada nuevo sector requiere el mismo proceso de análisis, formalización y validación.

En géneros cuya estructura está legalmente fijada —el prospecto de medicamento, la escritura notarial, el contrato de crédito normalizado— BELA™ formaliza cada bloque funcional y verifica que resuelve su cometido comunicativo en la posición que la norma le asigna. La formalización es lo que BELA™ transfiere, con independencia de si el resultado puede proponer un orden funcional o solo optimizar la resolución interna de cada bloque.

6.3 Del banco al despliegue: la herramienta de generación asistida

Una vez construido el banco de estructuras para un género y una causa jurídica determinados, el marco BELA™ habilita un segundo nivel de operación: la generación de nuevos documentos accesibles directamente desde las estructuras formalizadas, sin necesidad de partir del texto institucional original.

El flujo operativo tiene dos modos de entrada. En el primero, la institución dispone de un texto existente: el sistema detecta automáticamente la causa jurídica y el tipo de documento, recupera del banco el microframe correspondiente a esa combinación y mapea sobre sus slots los datos del caso —número de expediente, objeto de la solicitud, motivo, órgano, plazos—. En el segundo, la institución introduce directamente los datos del caso: el sistema selecciona el microframe pertinente, rellena sus slots con los datos proporcionados y genera el documento desde la arquitectura formalizada. Como cada slot solo admite un tipo determinado de contenido, el relleno queda acotado y es verificable.

En los dos casos, el resultado es el mismo: una versión en lenguaje claro estructurada según el orden funcional derivado del análisis del corpus correspondiente, generada a partir de estructuras validadas sobre corpus real, con trazabilidad explícita a los criterios operativos del banco. La versión en lectura fácil se genera de forma independiente, aplicando los criterios de la norma UNE 153101:2018 EX sobre las estructuras del banco y no como derivación de la versión en lenguaje claro, y requiere validación con usuarios reales antes de cualquier uso.

Esta capacidad de generación está implementada y en fase de validación sobre el corpus AIP-Madrid. La afinación de las estructuras del banco para cada causa jurídica es el trabajo en curso.

Los textos reglados son el lugar donde la norma alcanza directamente al ciudadano: en una resolución administrativa, en un contrato, en un prospecto. Que ese ciudadano no entienda lo que se le dice no es un problema de redacción ni de principios: los estándares internacionales de lenguaje claro establecen con precisión qué debe cumplir un texto accesible. El problema es operativo: cómo aplicar esos principios a géneros cuya arquitectura está condicionada por norma jurídica. BELA™ resuelve ese paso. Formaliza la arquitectura discursiva de los géneros reglados en estructuras reutilizables, construidas desde corpus real, que permiten producir versiones accesibles con criterios verificables y trazabilidad explícita. Lo que queda en una institución después de un piloto BELA™ no es un conjunto de textos corregidos. Es la capacidad de operar sobre la arquitectura de sus propios documentos de manera autónoma y duradera. Esa es la diferencia entre corrección e infraestructura.

Bibliografía

- Bello Rivas, M. X. (2026). BELA™: *Banco de Estructuras de Lenguaje Accesible. Marco metodológico y esquema de anotación* (Versión 11). Zenodo. <https://doi.org/10.5281/zenodo.21166817>
- Bello Rivas, M. X. y García Muñoz, Ó. (2026). Lectura fácil y web comprensible. Un libro blanco en el marco del nuevo reglamento. ARCO/libros.
- Bhatia, V. K. (1993). *Analysing Genre: Language Use in Professional Settings*. Longman.
- Bhatia, V. K. (2004). *Worlds of Written Discourse: A Genre-Based View*. Continuum.
- Cardon, R., y Doğruöz, A. S. (2026). Automatic text simplification for French medical documents with LLMs: The role of target audience and genre. En *Proceedings of the Joint Workshop on Readability and Text Simplification (READIXTSAR) @ LREC 2026* (pp. 164-180). ELRA.
- Carretero González, C. (2015). La claridad y el orden en la narración del discurso jurídico. *Revista de Llengua i Dret*, 64, 63–85.
- Carretero González, C. (2019). *Comunicación para juristas*. Tirant lo Blanch.
- Fernández-Silva, S., y Núñez Cortés, J. A. (2025). Exploring the effect of plain terminology on processing and comprehension of administrative texts in Spanish: A self-paced reading experiment. *International Journal of Applied Linguistics*, 35, 659-671. <https://doi.org/10.1111/ijal.12650>
- Guía panhispánica de lenguaje claro y accesible* (2024). RAE/ASALE, Espasa.
- Halliday, M. A. K., y Matthiessen, C. M. I. M. (2014). *An Introduction to Functional Grammar* (4.ª ed.). Routledge.
- Humm, B. G. et al. (2025). New trends in applied machine intelligence. *Informatik Spektrum*, 48, 21-37. <https://doi.org/10.1007/s00287-025-01589-z>
- Indecopi (2026). *¿Cómo simplificar una resolución? Guía práctica RÁPIDA*. Instituto Nacional de Defensa de la Competencia y de la Protección de la Propiedad Intelectual. Depósito Legal BNP N.º 2026-05140. <https://barreras.indecopi.gob.pe/>
- ISO 24495-1:2023. Plain language — Part 1: Governing principles and guidelines. ISO.
- UNE-ISO 24495-1:2024. Lenguaje claro — Parte 1: Principios rectores y directrices. UNE.
- ISO 24495-2:2025. Plain language — Part 2: Legal communication. ISO.
- ISO/DIS 24495-4. Plain language — Part 4: Requirements for implementing the plain language principles in organizations. ISO. (Borrador internacional en fase DIS, bajo desarrollo)
- Jiménez Hurtado, C., y Medina Reguera, A. (2022). Metodología de la traducción a lectura fácil: retos de investigación. En M. P. Castillo Bernal y M. Estévez Grossi (eds.), *Translation, Mediation and Accessibility for Linguistic Minorities* (pp. 205–222). Frank & Timme.
- Ley 39/2015, de 1 de octubre, del Procedimiento Administrativo Común de las Administraciones Públicas. BOE n.º 236, 2 de octubre de 2015.
- Madina, M., González-Dios, I., y Siegel, M. (2024). A preliminary study of ChatGPT for Spanish E2R text adaptation. *LREC-COLING 2024*, 1422-1434. <https://aclanthology.org/2024.lrec-main.126/>

- Masson, M. E. J., y Waldron, M. A. (1994). Comprehension of legal contracts by non-experts: Effectiveness of plain language redrafting. *Applied Cognitive Psychology*, 8(1), 67-85.
- Mathas, P., Bakker, J., y Kamps, J. (2026). Plan-guided text simplification with extended contexts. En *Proceedings of the Joint Workshop on Readability and Text Simplification (READixTSAR) @ LREC 2026* (pp. 121-129). ELRA.
- Moreno López, L., y Martínez Fernández, P. (en curso). Proyecto Human_AI: sistema multiagente con supervisión humana para la generación de texto en lenguaje claro y lectura fácil. Universidad Carlos III de Madrid. <https://humanai.uc3m.es/>
- Pistola, S., y Da Cunha, I. (2022). La resolución administrativa publicada en portales web institucionales: Un análisis basado en corpus. *ELUA: Estudios de Lingüística*. Universidad de Alicante, (38), 17-45. <https://doi.org/10.14198/ELUA.22442>
- Pistola, S., y Da Cunha, I. (2024). Análisis y clarificación de un corpus de fraseología administrativa: una propuesta orientada al uso del lenguaje claro en la Administración española. *Logos: Revista de Lingüística, Filosofía y Literatura*, 34(2), 494-522. <https://doi.org/10.15443/RL3421>
- Poblete Olmedo, C., y Yepes Villegas, P. (2023). Propuesta de un instrumento para medir la claridad de los textos jurídico-administrativos. En *Crónica de la lengua española 2022-2023*. RAE. <https://www.rae.es/recursos-academicos-de-apoyo-para-el-lenguaje-claro/cronica-de-la-lengua-espanola-2022-2023-2>
- Suárez-Figueroa, M. C., Diab, I., Ruckhaus, E., y Cano, I. (2024). First steps in the development of a support application for easy-to-read adaptation. *Universal Access in the Information Society*, 23(1), 365-377. <https://doi.org/10.1007/s10209-022-00946-z>
- Swales, J. M. (1990). *Genre Analysis: English in Academic and Research Settings*. Cambridge University Press.
- UNE 153101:2018 EX. Lectura Fácil. Pautas y recomendaciones para la elaboración de documentos. AENOR.
- van Dijk, T. A., y Kintsch, W. (1983). *Strategies of Discourse Comprehension*. Academic Press.
- Zödi, Zs. (2019). The limits of plain legal language: understanding the comprehensible style in law. *International Journal of Law in Context*, 15(3). <https://doi.org/10.1017/S1744552319000260>



OBSERVATORIO
NEBRIJA DEL
ESPAÑOL

observatorionebrijadelespañol.com
ONE@nebrija.es

www.nebrija.com



UNIVERSIDAD
NEBRIJA