



La comunicación en contextos de salud: generación de recursos tecnológicos multilingües para la mejora de la eficacia comunicativa del Consentimiento Informado

Communication in healthcare contexts: Multilingual technological resources to improve the communicative effectiveness of the Informed Consent

Isabel García-Izquierdo

Universitat Jaume I
Castellón, Comunidad Valenciana, España
igarcia@uji.es
<https://orcid.org/0000-0001-8208-4479> 

Anabel Borja Albi

Universitat Jaume I
Castellón, Comunidad Valenciana, España
borja@uji.es
<https://orcid.org/0000-0001-7053-0152> 

Resumen: Los avances científicos y la complejidad del contexto sociológico en el cual se practica la medicina en un mundo cada día más globalizado e interconectado plantean nuevos interrogantes éticos y legales sobre los derechos y obligaciones de los pacientes, los profesionales de la salud y los servicios públicos de atención sanitaria. A pesar del avance significativo e innegable que supone el cambio de paradigma en la relación médico-paciente y el gran desarrollo que ha conocido el derecho médico y la bioética en los últimos años, la población sigue encontrando serios obstáculos lingüísticos y culturales para ejercer su derecho a la información y a la toma de decisiones en relación con su salud. En este trabajo presentamos parte de los resultados del proyecto de investigación HIPÓCRATES del grupo español GENTT, dirigido a humanizar la atención sanitaria mediante la mejora de la comunicación médico-paciente en entornos multilingües. En concreto, nos centraremos en los avances realizados en la investigación de los aspectos textuales del Consentimiento Informado (CI), un género médico cuya redacción y traducción tiene un gran impacto en contextos clínicos debido a sus implicaciones éticas y legales. La complejidad léxico-sintáctica de este género y la falta de patrones estandarizados en su redacción dificultan la comprensibilidad, así como la identificación de las secuencias retóricas en las que se plasman la funciones comunicativa y legal, especialmente cuando emisor y receptor tienen idiomas y culturas médicas diferentes. Presentaremos la herramienta de análisis textual ProText GENTT, desarrollada en el contexto del proyecto, que utiliza técnicas tradicionales de explotación de corpus combinadas con aprendizaje automático e

inteligencia artificial. Con la ayuda de esta herramienta hemos realizado un análisis del corpus multilingüe (español, catalán e inglés) de textos de CI compilado por nuestro equipo (GENTT_Corpus), que nos ha permitido identificar los principales elementos lingüístico-textuales y retóricos que dificultan su comprensión. A partir de estos resultados, nuestro grupo trabaja actualmente en la propuesta de modelos optimizados con distintos niveles de complejidad tanto en formato tradicional como digital (*e-consent*).

Palabras clave: comunicación médica; redacción y traducción en contextos clínicos; tecnologías del lenguaje; consentimiento informado; explotación de corpus multilingües.

Abstract: Scientific advances and the complexity of the sociological context in which medicine is practised in an increasingly globalised and interconnected world raise new ethical and legal questions about the rights and obligations of patients, health professionals and public health care services. Despite the significant and undeniable progress brought about by the paradigm shift in the doctor-patient relationship and the great development of medical law and bioethics in recent years, patients continue to encounter serious linguistic and cultural obstacles to exercising their right to information and decision-making in relation to their health. In this paper we present part of the results of the HIPOCRATES research project of the Spanish GENTT group, aimed at humanising health care by improving doctor-patient communication in multilingual environments. Specifically, we will focus on the advances made in the textual analysis of the Informed Consent (IC), a medical genre of great relevance in clinical translation and writing due to its ethical and legal implications. Its lexical-syntactic complexity and the lack of standardised patterns in its writing hinder its comprehensibility and makes it difficult to identify the rhetorical sequences in which the communicative and legal functions of this genre are embodied. We present a textual analysis tool (ProText GENTT), which uses machine learning to exploit corpora with traditional techniques and artificial intelligence and with which we have analysed a multilingual (Spanish, Catalan and English) corpus of IC texts compiled by our research group (GENTT_Corpus). This analysis has allowed us to identify the linguistic-textual and rhetorical elements that hinder comprehension. From these results, our team is currently working on proposing optimized models with different levels of complexity both in printed and digital formats (*e-consent*).

Keywords: health communication; writing and translation in clinical contexts; language technologies; informed consent; exploitation of multilingual corpora.

I. Introducción

Los avances científicos y la complejidad del contexto sociológico en el cual se practica la medicina en un mundo cada día más globalizado e interconectado plantean nuevos interrogantes éticos y legales sobre los derechos y obligaciones de los pacientes, los profesionales de la salud y los servicios públicos de atención sanitaria. A pesar del avance significativo e innegable que supone el cambio de paradigma en la relación médico-paciente (Epstein *et al.*, 2005; Montalt & García-Izquierdo, 2016) y el gran desarrollo que ha conocido el derecho médico y la bioética en los últimos años, la población sigue encontrando serios obstáculos lingüísticos y culturales para ejercer su derecho a la información y a la toma de decisiones en relación con su salud.



Por este motivo, el grupo de investigación GENTT (www.gentt.uji.es) trabaja, desde hace una década, en proyectos de investigación cuyo eje conductor ha sido la mejora y humanización de la comunicación médico-paciente en contextos multilingües, mediante el análisis y las propuestas de mejora de la redacción y la traducción de los géneros médicos. Uno de nuestros últimos proyectos de investigación, HIPÓCRATES, ha abordado esta cuestión centrándose en el Consentimiento informado (CI), un género médico-jurídico de gran relevancia en la comunicación clínica. La complejidad léxico-sintáctica del género y la inexistencia de patrones estandarizados en su redacción en español, sumados a su naturaleza multimodal (acto comunicativo oral y escrito), dificultan la comprensibilidad del género (Borja Albi, 2012; García-Izquierdo & Montalt, 2017), por lo que se hace necesario investigar para proponer mejoras en la redacción (que puedan contribuir también a una mejor traducción).

2. El proyecto HIPÓCRATES

El proyecto HIPÓCRATES nace con la finalidad de mejorar la comunicación multilingüe (tanto escrita como oral) en el contexto médico-sanitario, mediante la generación de materiales y recursos para diferentes colectivos, desde el personal sanitario a los especialistas en redacción, traducción e interpretación médica. Para lograr este objetivo, consideramos necesario definir, en primer lugar, las normas textuales, cognitivas y sociales que regulan los géneros propios de la comunicación entre profesionales de la medicina y pacientes, con el fin de responder adecuadamente a las necesidades de ambos grupos. Como ya se ha indicado, en el proyecto pusimos el foco de análisis en el género CI, abordándolo como acto multimodal de comunicación oral y escrita.

Como objetivo último del proyecto nos planteamos crear una plataforma de recursos en línea para mejorar la comunicación médico-paciente y favorecer la autonomía del paciente, con especial atención a los aspectos lingüísticos, terminológicos, conceptuales, socioculturales, institucionales y legales del CI. Para ello, ha sido necesario abordar la investigación desde diferentes perspectivas, con el fin de analizar de manera pormenorizada todos los aspectos citados, aplicando un proceso de análisis multidimensional, y triangulando a posteriori los resultados. Los bloques de la investigación se organizaron alrededor de las cuatro dimensiones del género CI que describimos a continuación:

- *Marco legal*: desde el punto de vista jurídico recopilamos y analizamos la legislación aplicable al CI en España, tanto en el ámbito autonómico como estatal y en el plano internacional. En la web del proyecto se recogen todas las disposiciones legales que regulan la autonomía y los derechos y obligaciones de paciente en materia de información y documentación clínica en el Estado español, así como la legislación internacional al respecto utilizando mapas interactivos. Es importante señalar que la normativa regula tanto cuestiones de redacción como aspectos relativos al acto de comunicación oral en el que se enmarca la administración del CI. Toda esta normativa se puede consultar en el apartado *Marco legal* de la web del proyecto HIPÓCRATES¹;

¹ <https://hipocratesgentt.uji.es/>.



- *Perspectiva socioprofesional*: la implicación de los profesionales sanitarios es fundamental para entender las condiciones reales de la comunicación y mejorarla. Desde esta perspectiva, analizamos las percepciones del personal de medicina y enfermería en cuanto a la administración del CI. Utilizando métodos cuantitativos y cualitativos, obtuvimos datos sobre el grado de cumplimiento de las disposiciones legales sobre el acto de comunicación verbal en el que el profesional debe informar al paciente sobre la finalidad de la firma del CI y solucionar posibles dudas; sobre los problemas de tiempo que genera esta obligación; y sobre las percepciones y propuestas de mejora que plantea el colectivo;
- *Orientaciones pedagógicas*: la formación sobre CI es clave en el entorno actual de las *Humanidades médicas* y, por tanto, se hizo hincapié en la mejora de la formación de los profesionales de la salud para asegurar que la administración del género CI se realiza con todas las garantías éticas y legales;
- *Aspectos textuales*: el CI, en tanto que documento escrito, presenta importantes problemas de comprensibilidad. Para poder analizar el género y proponer mejoras en su redacción (y guías para su traducción), utilizamos las teorías de análisis del registro (Biber, 2019; Biber & Egbert, 2016) y desarrollamos una herramienta de análisis textual: ProText GENTT. Utilizando esta herramienta, como veremos a continuación, compilamos y analizamos un corpus multilingüe de textos de CI, lo que nos permitió: 1) identificar los elementos lingüístico-textuales que dificultan la comprensión; 2) identificar las secuencias retóricas prototípicas de este género; y 3) proponer la generación de modelos optimizados con distintos niveles de complejidad, tanto en formato tradicional como digital (*e-consent*).

En este artículo, por cuestiones de espacio, nos centraremos únicamente en describir el diseño de este último bloque de investigación y avanzaremos los primeros resultados del análisis textual obtenidos con la herramienta ProText GENTT.

3. El Consentimiento Informado y la comprensibilidad

En el artículo 3 de la Ley 41/2002 (española), de 14 de noviembre, básica reguladora de la autonomía del paciente y de derechos y obligaciones en materia de información y documentación clínica, encontramos la siguiente definición del CI: “[...] la conformidad libre, voluntaria y consciente de un paciente, manifestada en el pleno uso de sus facultades después de recibir la información adecuada, para que tenga lugar una actuación que afecta a su salud” (Gobierno de España, 2002, s. p.). Tal y como está previsto en la legislación, su finalidad es proteger al paciente y asistirlo para tomar una decisión informada. Resulta paradójico, sin embargo, que, según los datos y percepciones aportados por los profesionales en los cuestionarios y *focus group* realizados a lo largo de nuestro proyecto, el CI funciona en la práctica como una herramienta de protección para los profesionales sanitarios y las instituciones de salud frente a posibles reclamaciones legales (García-Izquierdo & Bellés, en prensa).



Siguiendo a Borja Albi (2012), el CI es un género híbrido, situado entre la Medicina y el Derecho. Los géneros médico-jurídicos son textos “en los que aparecen combinados conceptos propios de la medicina y conceptos legales [...], o bien textos médicos que pueden surtir efectos legales” (Borja Albi, 2012, p. 167). Casado (2008, p. 19), por su parte, señala que los documentos médico-jurídicos son “[...] todas aquellas actuaciones escritas que utiliza el médico en sus relaciones profesionales con las autoridades, los organismos, las instituciones o con cualquier persona”. Según la taxonomía de textos médico-jurídicos propuesta por Borja Albi (2012), el CI se sitúa en el bloque de *Declaraciones unilaterales de voluntad*, ya que el paciente da su consentimiento sin obtener a cambio una promesa o compromiso por parte del personal sanitario.

Tabla 1: Taxonomía de textos médico-jurídicos

1	Textos normativos de temática médica: leyes, reglamentos y normativas relativas a cuestiones de salud.
2	Textos judiciales de temática médica: demandas por negligencia médica o sentencias de incapacitación, entre otros.
3	Acuerdos de voluntades: contratos de suministro médico, contratos de seguros de salud.
4	Declaraciones unilaterales de voluntad: consentimientos informados, testamentos vitales, donaciones de órganos, etc.
5	Textos administrativos que requieren la intervención de un médico para tener eficacia jurídica o resoluciones administrativas sobre aspectos de salud: informes forenses, certificados de defunción, certificados médicos para una solicitud de reconocimiento de invalidez o autorizaciones, entre otros.

Fuente: Borja Albi (2012, p. 168)

El paciente consiente y exonera de responsabilidad. De ahí, la importancia de que el consentimiento sea prestado de manera *informada* y con todos los requisitos de procedimiento que dispone la ley y que se pueden observar gráficamente en la Figura 1. Con esto queremos resaltar la importancia que tiene el acto comunicativo de carácter oral, previo a la firma, en el que el profesional debe informar y asegurarse de que la firma del documento cumple los requisitos legales. En este sentido la Ley 41/2002 establece que su uso clínico real requiere de la interacción oral en el caso de determinados procesos e intervenciones (Gobierno de España, 2002). En su artículo 8, la ley dispone que “Toda actuación en el ámbito de la salud de un paciente necesita el consentimiento libre y voluntario del afectado, una vez que, recibida la información prevista en el artículo 4, haya valorado las opciones propias del caso” (Gobierno de España, 2002, s. p.). En el párrafo 2 de este mismo artículo, se establece que: “El consentimiento será verbal por regla general. Sin embargo, se prestará por escrito en los casos siguientes: intervención quirúrgica, procedimientos diagnósticos y terapéuticos invasores y, en general, aplicación de procedimientos que suponen riesgos o inconvenientes de notoria y previsible repercusión negativa sobre la salud del paciente” (Gobierno de España, 2002, s. p.).

Figura 1: El Consentimiento Informado como acto multimodal



Fuente: Proyecto Hipócrates. Gentt (2022)

Son muchas las voces que, desde el ámbito de la medicina y la investigación clínica – la Organización Médica Colegial (2016) española, en su decálogo sobre el CI de 2016; la Organización Mundial de la Salud y el Comité de Evaluación Ética de la Investigación español, entre otros – reivindican la necesidad de prestar mayor atención a este género multimodal, que, en la mayoría de casos, no se ajusta a los estándares básicos de comprensibilidad exigibles para que la comunicación funcione eficazmente. Además, es necesario mejorar la formación específica en comunicación de los profesionales sanitarios, prestar atención a la calidad y mejorar la adecuación de la comunicación que se mantiene en la interacción médico-paciente (García-Izquierdo & Montalt, 2013, 2017), particularmente en su dimensión oral, pero siempre con el apoyo de documentos escritos con un buen nivel de comprensibilidad.

El CI ha sido ampliamente estudiado tanto desde el derecho como desde la medicina, pero centrándose casi exclusivamente en el análisis de su legibilidad en modo escrito, a partir de fórmulas lingüísticas preestablecidas (Simón Lorda, 2000; Dubay, 2004), sin profundizar en aspectos textuales no recogidos en las fórmulas de legibilidad y sin tener en cuenta aspectos cualitativos fundamentales. Así, se ha dedicado poca atención a tener una visión integral de la realidad de su funcionamiento en la práctica clínica y a las consecuencias que realmente se derivan de su utilización. Sin embargo, como hemos dicho, en muchos casos encontramos deficiencias en la redacción de los textos escritos que, unida a la citada falta de intermediación oral, puede tener repercusiones negativas en la autonomía de los pacientes e incidir significativamente tanto en su falta de adhesión a los procesos terapéuticos, como en la eficacia del sistema de salud (Montalt & García-Izquierdo, 2016).

Como conclusión de este apartado podemos afirmar que el estudio socioprofesional del contexto de situación en el que se administra el CI en los hospitales y centros de salud de la Comunidad Valenciana realizado en el marco del proyecto HIPÓCRATES puso de manifiesto que hay dificultades para que los pacientes entiendan el texto y que, por tanto, no se cumple con el objetivo comunicativo del género, que es fundamentalmente un acto performativo de aceptación. Según las percepciones de los profesionales médicos entrevistados, estas dificultades están determinadas, entre otras cosas, por:

1. La complejidad de la información que se incluye (que pertenece a dos ámbitos de especialidad —jurídico y médico— ya de por sí complejos), unida a las deficiencias (de tipo gramatical y léxico) en la redacción;
2. La multifuncionalidad del texto, que incluye secuencias de diferentes tipos (expositivas, argumentativas, performativas);
3. La estructura variable del documento en función de los hospitales, a pesar de la prototipicidad y la recurrencia que debería exhibir el género y la existencia de algunos modelos normalizados;
4. La falta de tiempo para la lectura comprensiva del documento (como no hay tiempo, el paciente no lo lee y, por tanto, no lo comprende).

Los datos empíricos, cuantitativos y cualitativos, obtenidos en el estudio socioprofesional —*focus group*, entrevistas, etc.— (Muñoz Miquel & García-Izquierdo 2020; Martínez-Carrasco & Ordóñez-López, 2023; García-Izquierdo & Bellés, en prensa), demuestran que no se cumple la legislación aplicable de carácter internacional, comunitaria y nacional respecto a la obligación de proporcionar información y explicaciones verbalmente antes de la firma. En este artículo, como ya hemos señalado, nos centraremos en los resultados del análisis formal (textual) de los documentos del corpus de CI, para intentar detectar las deficiencias en la redacción y proponer mejoras a la Administración pública que puedan implementarse en la documentación para la práctica clínica.

4. Análisis textual para mejorar la eficacia comunicativa del Consentimiento Informado

A pesar de la generalización, desde principios de los 2000, de las llamadas *Humanidades médicas* y la comunicación centrada en el paciente (Epstein *et al.*, 2005; Epstein & Street, 2011; Suojanen *et al.*, 2014), continúa siendo necesaria la formación de los redactores y traductores médicos para apropiarse adecuadamente de las normas que regulan el funcionamiento textual, cognitivo y social de los géneros para pacientes y poder responder adecuadamente a las expectativas y necesidades de estos últimos. Esto es especialmente relevante cuando hablamos de géneros, como el CI, objeto de la investigación de GENTT en el proyecto HIPÓCRATES, que, como hemos apuntado, puede tener consecuencias jurídicas importantes y cuya administración está regulada por la ley.

En el bloque de investigación textual, como hemos visto en líneas anteriores, se consideró necesario realizar el análisis del documento escrito de CI para detectar deficiencias y proponer mejoras. Y para ello se plantearon los siguientes *objetivos específicos*:

1. Compilar un corpus multilingüe (español, catalán e inglés¹), representativo de consentimientos informados en modo escrito, utilizados cuando el paciente debe someterse a una intervención quirúrgica o a procedimientos diagnósticos o terapéuticos invasores;

¹ El español y el catalán son las lenguas propias de la Comunitat Valenciana, donde se lleva a cabo la investigación. Y el inglés se seleccionó por ser la *lingua franca* de la comunicación internacional.

2. Analizar formalmente los documentos del corpus aplicando técnicas de análisis cualitativo y cuantitativo, a fin de valorar y mejorar su comprensibilidad (tipográfica, lingüística y cognitiva, etc.) desde la perspectiva del paciente, aplicando técnicas de observación manual, así como diversas métricas del análisis tradicional de corpus y de aprendizaje automático para determinar cómo influyen los aspectos macro y microlingüísticos en la eficacia comunicativa del género. Partiendo de los postulados del análisis del registro, destacamos la necesidad de distinguir entre el uso del contexto lingüístico y del contexto situacional desde una perspectiva retórica. Por tanto, decidimos poner nuestro foco en la gramática y el léxico del CI, así como en el funcionamiento de las secuencias retóricas que utiliza para responder a los objetivos concretos de la situación comunicativa en la que se enmarca el CI (Carrillo Guerrero, 2005);
3. Analizar, mediante el asesoramiento de expertos, el ajuste textual de los documentos a la legislación vigente y a las recomendaciones de diversas comisiones multidisciplinares y comités de bioética en España;
4. Generar bases de datos de documentos tipo en los idiomas de trabajo y glosarios especializados útiles para el traductor o redactor de este género, desde una perspectiva multilingüe.

En la actualidad, nuestro equipo de investigación está diseñando sistemas de acceso en línea a dichas bases de datos para que puedan pasar a formar parte de los protocolos de actuación de los centros sanitarios participantes en el proyecto. Para ello trabajamos en la creación de una plataforma virtual con toda la información resultante de las acciones anteriores para dar acceso a todos los grupos de interés en condiciones óptimas de usabilidad a los resultados y productos del proyecto: bases de datos de gestión documental multilingüe; protocolos; recomendaciones y análisis de resultados del estudio de campo y de la reflexión teórica que sustenta la investigación.

5. Metodología: compilación y análisis del corpus GENTT de Consentimientos Informados

Para obtener datos empíricos sobre el comportamiento textual del CI, decidimos compilar un corpus de este género procedente de los servicios de Salud de la Comunidad Valenciana en castellano, catalán e inglés. En total recogimos 1343 textos y aplicamos metodología cuantitativa de explotación de corpus, a fin de detectar los problemas de comprensión fundamentales y proponer modelos normalizados, tanto desde el punto de vista de la macroestructura, como desde el punto de vista microestructural.

Todos los documentos del corpus aparecen documentados según distintos atributos en una base de datos compartida con todos los miembros del grupo. Los documentos se organizaron en una carpeta general y se crearon subcarpetas atendiendo a los descriptores que se definen más abajo, con el objetivo de poder alimentarlo y estudiarlo de manera sistemática.

El *diseño del corpus* responde a las necesidades de análisis planteadas en el proyecto y cada texto se caracterizó con 11 descriptores:



Descriptores	
1. Título descriptivo y código:	<p>El criterio para nombrar los documentos del corpus pretende ser lo más transparente y explícito posible, con el objetivo de poder localizar rápidamente un texto dentro del corpus, sin necesidad de abrirlo. La nomenclatura de los documentos del corpus se compone de una serie de elementos separados por una barra baja (sin espacios):</p> <p><i>CI_número_tipo de procedimiento_estatus_idioma</i></p>
2. Número	Número consecutivo y único asignado a cada texto del corpus a medida que estos se han añadido a la base de datos.
3. Estatus	El estatus marca la naturaleza del texto: original (OR), traducido (TR) o bitexto (BT; es decir, un texto a dos columnas, con el original a la izquierda y su correspondiente traducción a la derecha).
4. Idioma o combinación de lenguas:	<p>En el caso de los textos cuyo estatus fuera original, aparece el código ISO /639-I correspondiente. En el caso de nuestro corpus, se utilizaron los códigos correspondientes para el español (ES), el inglés (EN) y el catalán (CA). En el caso de que fuera una traducción o un bitexto, se combinaron los códigos en función del par de lenguas.</p> <p><i>Ejemplos:</i> <i>CI_0001_I_OR_ES: es un CI original en español que se utilizará para una intervención</i> <i>CI_0012_D_TR_EN_ES: es la traducción al español de CI original en inglés que se utilizará para el diagnóstico</i> <i>CI_0087_I_TR_EN_ES: es el bitexto con el original en inglés (columna de la izquierda) y la traducción al español (columna de la derecha) de un CI que se utilizará para una intervención</i></p>
5. Modelo/Auténtico:	Los CI recogidos en el corpus pueden ser modelos o plantillas, a disposición del público, generalmente a través de internet (portales de los hospitales, consejerías de sanidad de las Comunidades Autónomas, fundaciones, colegios, etc.) o textos auténticos, utilizados en la práctica

	real y diaria de los profesionales sanitarios y proporcionados al grupo [GENTT] directamente por parte de estos.
6. Género textual:	En este corpus, todos los documentos se marcaron como «consentimiento informado» (CI). Esta columna se añadió porque puede resultar de utilidad si se decide ampliar el corpus con otros géneros textuales, en investigaciones paralelas o futuras.
7. Diagnóstico/Intervención:	Generalmente, los CI suelen utilizarse para dar el consentimiento ante un procedimiento diagnóstico o una intervención, de modo que se decidió añadir esta distinción en el corpus, como complemento al resto de etiquetas
8. Especialidad médica:	<p>Los CI se clasificaron según el tipo de CI basándose en la especialidad médica. se decidió añadir a qué especialidad o rama de la medicina pertenece el CI.</p> <p>Se etiquetaron, de forma inductiva, los CI según las siguientes especialidades: alergología, anestesia y reanimación, angiología y cirugía vascular, aparato digestivo, artroscopia y cirugía articular, biopsia, cardiología, cirugía abierta del estómago, cirugía de la columna, cirugía de la mano, cirugía del pie, cirugía general, cirugía general y del aparato digestivo, cirugía oral y maxilofacial, de cabeza y del cuello, cirugía ortopédica y traumatología, cirugía protésica, dermatología, digestivo, endocrinología y nutrición, ginecología, hematología y hemoterapia, implantología, medicina interna, medicina nuclear, neumología, obstetricia, oftalmología, otorrinolaringología, otros, radiología, reumatología, traumatología y urología.</p>
9. Autor del original/Procedencia:	Fuente de la que procede el documento (hospital, sociedad médica...).
10. Ordenamiento o país:	País u ordenamiento jurídico en el que se utiliza dicho CI y en el que tiene validez legal.
11. Fuente web:	En caso de que el texto se obtuviese de internet se incluyó la URL original.

6. Resultados preliminares del análisis textual

Una vez compilado el corpus, se realizó el análisis (cuantitativo y cualitativo) para intentar identificar de forma empírica aquellos elementos que dificultan la comprensión y la legibilidad de este género textual, todo ello con el objetivo último de justificar ante la Administración la necesidad de contar con CI optimizados y de fácil comprensión.

Dado que los CI son documentos muy repetitivos y formulaicos, el índice de saturación de los corpus se alcanza pronto y, en principio, decidimos aplicar un enfoque minimalista. En la prueba piloto trabajamos con una muestra de 100 CI del corpus de español para identificar los rasgos morfológicos del mismo, siguiendo el enfoque de análisis del registro propuesto por Biber (2019). La primera fase consistió en formular hipótesis, diseñar una estrategia de análisis ajustada al género y proponer una batería de métricas que nos permitieran abordar el objeto de estudio con las herramientas disponibles para cada lengua. Trabajamos con la hipótesis inicial de que la frecuencia y la co-ocurrencia de algunas de las categorías gramaticales básicas que caracterizan el CI pueden identificarse utilizando datos sobre frecuencia de aparición de las distintas categorías gramaticales (Biber & Egbert, 2016). Para este primer acercamiento, se aplicaron técnicas de observación manual, así como diversas métricas del análisis tradicional de corpus, que nos ayudaron a definir la propuesta de métricas que aplicaríamos a todo el corpus, que clasificamos en dos grupos: métricas asociadas con la complejidad léxica y métricas asociadas con la legibilidad, y que mostramos en las siguientes tablas:

Tabla 2: Métricas asociadas con la complejidad léxica

Nº de tokens lenguaje jurídico especializado	Nº de types lenguaje jurídico especializado	Nº de tokens lenguaje médico especializado	Nº de types lenguaje médico especializado	Nº de tokens	Type/token ratio	Nº de lemmas en la banda > 5k	Nº Nombres (tokens)
Nº de hápax legomena (nombres)	Nº de tokens de adjetivos	Nº de types de adjetivos	Nº de hápax legomena (adjetivos)	Nº de tokens de adverbios	Nº de types de adverbios		

Fuente: Autoras (2024)

Tabla 3: Métricas asociadas con la legibilidad

Nº de conjunciones	Conjunciones coordinativas	Conjunciones subordinadas	Nº de oraciones	N de pronombres personales de 1ª persona	N de pronombres personales de 2ª persona	N de pronombres personales de 3ª persona	Nº de sintagmas verbales con modo subjuntivo
Nº de sintagmas verbales con participio pasado	Nº de sintagmas verbales con gerundio	Nº de sintagmas verbales con participio pasado	Nº de determinantes	Tiempo estimado de lectura. (ajustable, 250 palabras /minuto p. ej. Es frecuente)	Fernández Huerta (1954) Flesch-Reading-Ease Score (1948)		

Fuente: Autoras (2024)



En una segunda fase, todas estas herramientas de análisis se integraron en un prototipo para la anotación automática, procesado y extracción de métricas (legibilidad, frecuencia léxica y propiedades morfosintácticas) de textos médicos en castellano y catalán, adaptando algunas de las métricas ya existentes para el idioma inglés. Este sistema ha sido desarrollado con el Centro Tecnológico de las Tecnologías de la Información y las Comunicaciones (CENTIC) y el catedrático de lingüística aplicada Pascual Pérez-Paredes de la Universidad de Murcia, a partir de la propuesta científica del grupo de investigación GENTT. Este prototipo, denominado Protext GENTT, incorpora un conversor de textos a pdf y permite aplicar la batería de métricas que acabamos de exponer pudiendo seleccionar las que se consideren necesarias en cada momento. Es importante destacar que, a diferencia de lo que ocurre para el inglés, hemos encontrado muchas dificultades para identificar los listados léxicos especializados (médicos y jurídicos) en español y catalán que se incluyen en la herramienta, que son elementos imprescindibles para establecer comparaciones con corpus de referencia. A continuación, se incluye un listado con las principales funcionalidades de la herramienta.

Figura 2: Principales funcionalidades de ProText GENTT

- Script de PDF a textos limpios
- El script funciona con un PDF o con varios PDF a la vez
- Generación de alerta o warnings sobre aquellos PDF que no se convirtieron o no se limpiaron adecuadamente
- Etiquetado morfológico (POS tagging) con Freeling en Python, ya que permite varios idiomas y genera análisis muy detallados
- Métricas básicas morfológicas de los listados anteriores
- Métricas implementadas desde el etiquetado morfológico antes citado
- Métricas que indican la legibilidad
- Implementación en Python de fórmulas ya establecidas que determinan la legibilidad de un texto
 - Métrica Fernández Huerta (1954) para textos en español y métrica *Flesch Reading Ease score* (1948) para textos en inglés
- Métricas que cotejan con listas de vocabulario externas para sacar porcentajes
 - Vocabulario médico
 - Vocabulario legal
 - Vocabulario de frecuencia general (palabras más comunes de un idioma)
- Análisis comparativo entre textos
- Reducción dimensional para visualizar todos los textos
- *Clustering* para encontrar textos similares
- Generación de métricas anteriores por grupos de textos, (en lugar de por textos individuales)
- Generación de resultados/informe/gráficas con Streamlit
- Posibilidad de exportar la tabla de métricas y las gráficas a Excel
- *Machine Learning* supervisado (en proceso de revisión)
- Etiquetado de los datos (en proceso de revisión)
- Entrenamiento del modelo de *Machine Learning* (en proceso de revisión)

Fuente: Proyecto Hipócrates. Grupo Gentt (2022)

Aplicando esta herramienta y los métodos de observación y análisis manual que enumeramos a continuación, se han realizado ya informes preliminares para el corpus de catalán “*A preliminary analysis of a corpus of consent forms in Spanish and Catalan*” y “*A preliminary analysis of a corpus of consent forms in Spanish*” para el corpus de español, que resumimos a continuación.

Progresión temática (moves). La determinación automática de la macroestructura es, sin duda, una cuestión compleja, pero en este caso, al tratarse de un género de corta extensión y formato recurrente, ha sido posible identificar los tipos básicos de organización de la información. Se han

etiquetado automáticamente las secuencias y se han separado para analizarlas conjuntamente y poder comparar los resultados de manera manual. Este proceso nos ha permitido constatar la gran diversidad de macroestructuras que dificultan la localización de la información fundamental por parte del lector.

- 1 *Legibilidad lingüística*: además de las cuestiones relacionadas con los índices de densidad léxica, a partir de los resultados obtenidos con ProText hemos constatado la abundancia de oraciones complejas, el uso excesivo de la pasiva, el alto nivel de subordinación, la presencia de demasiados sintagmas complejos, el escaso uso de pronombres personales o los conectores, la presencia de frases/estructuras formulaicas fosilizadas, las colocaciones exclusivas de este tipo de género, la abundancia de términos de especialidad, el poco uso de procedimientos de desteminologización (o reformulación del discurso mediante paráfrasis, sinonimia, etc.);
- 2 *Variabilidad en la redacción de las proposiciones performativas en las que el paciente da el consentimiento*: la fórmula jurídica que aparece como parte indispensable del CI se ha extraído automáticamente de todos los consentimientos analizados para organizar sus distintas realizaciones en *clusters*. Aquí también hemos constatado una importante diversidad de formulaciones que aparecen en distintos lugares del documento, lo que contribuye a dificultar la comprensión por parte del lector;
- 3 *El uso de color e imágenes y la tipografía*: también ha sido objeto de nuestro análisis. Hemos comprobado que el color y la imagen son recursos muy poco utilizados, especialmente si lo comparamos con los CI para ensayos clínicos en los que se utilizan para facilitar la lectura: subrayar la información importante o indicar los cambios de sección. En cuanto a la tipografía, hemos detectado la presencia en los textos de elementos tipográficamente diferenciados: cursivas, negritas, mayúsculas, diferentes tipos de letra. Sin embargo, su aplicación no siempre resulta coherente y, por tanto, no facilita la lectura.

7. Conclusión

Los datos cuantitativos y cualitativos, tanto de investigaciones anteriores del grupo GENTT como los del proyecto HIPÓCRATES, coinciden con los de numerosos estudios previos (Dubay, 2004; Göpferich, 2009; Pilegaard & Ravn, 2012; García-Izquierdo & Montalt, 2017, Ribeiro-Alves & Ferreira Cabrera, 2018) que afirman que, en muchos casos, no se proporciona información previa a la firma del CI y que la información que se proporciona (oralmente o por escrito) no es adecuada ni comprensible, por lo que es absolutamente necesario mejorar la legibilidad y la comprensibilidad de la documentación que se ofrece a los pacientes, que es muchas veces la única que recibe. Estos resultados, pues, avalan la necesidad de mejorar los documentos de CI y la pertinencia de sensibilizar a los colectivos profesionales vinculados a este género (profesionales de la medicina, traductores y juristas) durante su formación, tarea esta en la que también trabaja nuestro equipo en la actualidad con un proyecto financiado². Los resultados refuerzan nuestra hipótesis de que el CI, como género

² Proyecto CIAICO/2022/032, financiado por la Consellería de Educación, Universidades y Empleo. Dirección General de Ciencia e Investigación. Generalitat Valenciana.



textual que debe cumplir unos propósitos comunicativos legalmente exigidos, podría beneficiarse de una formulación estandarizada que facilitara al paciente la localización e interpretación de la información (es decir, las diferentes secuencias retóricas). También ha quedado demostrada la necesidad de adaptar la terminología específica a los niveles de alfabetización de los destinatarios; reducir la extensión del CI y trabajar con formatos de texto más atractivos.

El diseño de un analizador morfosintáctico propio (ProText GENTT)³, específicamente creado para la determinación de la legibilidad y comprensibilidad del CI en las tres lenguas de trabajo, y el análisis complementario de los tres corpus (catalán, castellano e inglés) que hemos descrito nos han permitido objetivar y modelar lingüísticamente la intuición de la necesidad de mejora de la redacción y del procedimiento de administración de los textos de CI. Esta herramienta nos permitirá en el futuro analizar otros géneros médicos, puesto que ya disponemos de recursos lingüísticos electrónicos, especialmente en español y en inglés, del ámbito de la medicina.

Para lograr una mejora real de la eficacia comunicativa del género, trabajamos actualmente en el diseño de una app, que incluirá modelos optimizados en varios idiomas y niveles de complejidad, a partir de estos resultados y de las recomendaciones que se incluyen en la guía de buenas prácticas HIPÓCRATES⁴ relativas al contexto real de administración del consentimiento. Esta herramienta permitirá, además, seleccionar el CI por tipo de procedimiento y nivel de legibilidad (nivel bajo: infografías; nivel medio: vocabulario controlado y sintaxis sencilla; nivel experto dirigido a profesionales de la sanidad: vocabulario altamente especializado) a partir de plantillas uniformes fácilmente personalizables. Para ofrecer a los pacientes formatos más atractivos, al tratarse de documentos en formato electrónico, se podrán incluir hipervínculos, infografías, imágenes, glosarios e incluso formularios para consultas específicas.

Otra iniciativa en la que trabajamos en la actualidad es la creación de una plataforma de gestión de recursos para personal sanitario y traductores del género CI, que estamos desarrollando a partir de la experiencia en proyectos anteriores del grupo, como la Enciclopedia electrónica de géneros GENTT⁵ o las Plataformas TransTools GENTT⁶ de recursos de documentación para la traducción especializada (Borja Albi, 2013; Borja Albi & García-Izquierdo, 2015, 2016).

Referencias

- Biber, D. (2019). Text-linguistic approaches to register variation. *Register Studies*, 1(1), 42–75. <https://doi.org/10.1075/rs.18007.bib>
- Biber, D., & Egbert, J. (2016). Register Variation on the Searchable Web: A MultiDimensional Analysis. *Journal of English Linguistics*, 44(2), 95–137. <https://doi.org/10.1177/0075424216628955>
- Borja Albi, A. (2012). Aproximación traductológica a los textos médico-jurídicos. *Panace@*, 13(36), 167–175.
- Borja Albi, A. (2013). A genre analysis approach to the study of the translation of court documents. *Linguistica Antverpiensia, New Series – Themes in Translation Studies*, 33–53.

³ De momento, el prototipo solo es accesible para los miembros del grupo, puesto que estamos pendientes de su inscripción en el Registro de lo Propiedad Intelectual o en la generación de una patente.

⁴ <https://hipocratesgentt.uji.es/>.

⁵ <http://www.gentt.uji.es/corpus-y-arbol-de-generos/>.

⁶ <http://www.gentt.uji.es/plataformas/>.



- Borja Albi, A., & García-Izquierdo, I. (2015). Corpus-based knowledge management systems for specialized translation: bridging the gap between theory and professional practice. In M. T. Sánchez Nieto (Ed.), *Corpus-based Translation and Interpreting Studies: From description to Application*. (pp. 191–209). Frank & Timme.
- Borja Albi, A., & García-Izquierdo, I. (2016). Web-based tools and resources for legal translators: the JudGENTT translation-oriented glossaries for criminal courts translators. *Onomázein*, 33, 226–250.
- Carrillo Guerrero, L. (2005). Actualización retórica de la lengua: el Registro. *Tonos digital: Revista de estudios filológicos*, 9.
- Casado, M. (2008). *Manual de Documentos médico-legales*. Consejería de Sanidad y Dependencia.
- DuBay, W. H. (2004). *The principles of readability*. Impact Information.
- Epstein, R. M., Franks, P., Shields, C. G., Meldrum, S. C., Miller, K. N., Campbell, T. L., & Fiscella, K. (2005). Patient-centered communication and diagnostic testing. *The Annals of Family Medicine*, 3(5), 415–421. <https://doi.org/10.1370/afm.348>
- Epstein, R. M., & Street, Richard. L. (2011). The values and value of patient-centered care. *The Annals of Family Medicine*, 9(2), 100–103. <https://doi.org/10.1370/afm.1239>
- Fernández Huerta, J. (1954). Evaluación de la composición escrita. *Revista Española de Pedagogía*, 12(47), 337–348.
- Flesch, R. F. (1948). A new readability yardstick. *Journal of Applied Psychology*, 1948, 32, 221- 223.
- García-Izquierdo, I., & Montalt, V. (2013). Equigeneric and intergeneric translation in patient-centred care. *HERMES: Journal of Language and Communication in Business*, (51), 39–51. <https://doi.org/10.7146/hjlc.v26i51.97436>
- García-Izquierdo, I., & Montalt, V. (2017). Understanding and enhancing comprehensibility in text for patients in an institutional health care context in Spain: A mixed methods analysis. *RESLA*, 30(2), 592–611. <https://doi.org/10.1075/resla.00008.gar>
- García-Izquierdo, I., & Bellés, B. (en prensa). Improving clinical communication: a qualitative study on the informed consent. *Revista de Lenguas y Lingüística Aplicada*.
- GENNT (2022). Proyecto HIPÓCRATES. *Creación de recursos multilingües para la mejora y la humanización de la comunicación médico-paciente en los servicios públicos de salud (HIPOCRATES)*. Ministerio de Ciencia, Innovación y Universidades Proyectos de I+D de Generación de Conocimiento y Proyectos de I+D+i Retos Investigación.
- Gobierno de España. Ley 41/2002, de 14 de noviembre, básica reguladora de la autonomía del paciente y de derechos y obligaciones en materia de información y documentación clínica. Ministerio de la Presidencia, Justicia y Relaciones con las Cortes. <https://www.boe.es/eli/es/l/2002/11/14/41/con>
- Göpferich, S. (2009). Comprehensibility assessment using the Karlsruhe comprehensibility concept. *The Journal of Specialised Translation*, 11, 31–52.
- Martínez-Carrasco, R., & Ordóñez-López, P. (2023). El consentimiento informado en la comunicación médico-paciente: análisis crítico del marco legislativo. *HERMES - Journal of Language and Communication in Business*, 63, 100–117. <https://doi.org/10.7146/hjlc.vi63.134391>
- Montalt, V., & García-Izquierdo, I. (2016). ¿Informar o comunicar? Algunos temas emergentes en comunicación para pacientes. *Panace@*, 17(44), 81–84.
- Muñoz Miquel, A., & García-Izquierdo, I. (2020). El Consentimiento informado y la comunicación centrada en el paciente: reflexiones desde la perspectiva de profesionales sanitarios y pacientes. In B. Tejera, J. F. Aixelá & Gheorghiu, C. I. (Eds.), *Translatum Nostrum. La traducción y la interpretación en el ámbito especializado* (pp. 137–153). Comares.

- Organización Médica Colegial. (2016). *Decálogo del Consentimiento Informado*. Comisión de ética de la OMC. https://www.cgcom.es/noticias/2016/06/16_06_21_decologo_consentimiento_informado
- Pilegaard, M., & Ravn, H. B. (2012). Readability of patient information can be improved. *Danish Medical Journal*, 59(5), 1–5.
- Ribeiro-Alves, A., & Ferreira Cabrera, A. (2018). Estudio de corpus: estructura y legibilidad en el documento de consentimiento informado en el ámbito académico profesional de las ciencias biomédicas. *RLA: Revista de Lingüística Teórica y Aplicada*, 56(2), 91–116. <http://dx.doi.org/10.4067/S0718-48832018000200091>
- Simón Lorda, P. (2000). *El Consentimiento Informado. Historia, Teoría y Práctica*. Triacastela.
- Suojanen, T., Koskinen, K., & Tuominen, T. (2014). *User-centered translation*. Routledge.

Notas

Contribución de autoría

Concepción y elaboración del manuscrito: I. García-Izquierdo, A. Borja Albi

Recolección de datos: I. García-Izquierdo, A. Borja Albi

Análisis de datos: I. García-Izquierdo, A. Borja Albi

Discusión de los resultados: I. García-Izquierdo, A. Borja Albi

Revisión y aprobación: I. García-Izquierdo, A. Borja Albi

Datos de la investigación

No se aplica.

Financiación

Esta publicación se ha realizado en el contexto del proyecto de investigación HIPÓCRATES. PGC2018-098726-B-I00. Creación de recursos multilingües para la mejora y la humanización de la comunicación médico-paciente en los servicios públicos de salud (Hipócrates) (2019-2022). Financiado por el Ministerio de Ciencia e Innovación. Investigadoras principales: Isabel García-Izquierdo y Anabel Borja Albi.

Derechos de uso de imagen

No se aplica.

Aprobación de comité de ética en investigación

No se aplica.

Conflicto de intereses

No se aplica.

Licencia de uso

Los autores ceden a *Cadernos de Tradução* los derechos exclusivos de primera publicación, con el trabajo simultáneamente licenciado bajo la [Licencia Creative Commons](#) Atribución 4.0 Internacional (CC BY). Esta licencia permite a terceros remezclar, adaptar y crear a partir del trabajo publicado, otorgando el crédito adecuado de autoría y publicación inicial en esta revista. Los autores están autorizados a celebrar contratos adicionales por separado para distribuir de manera no exclusiva la versión del trabajo publicado en esta revista (por ejemplo, publicarlo en un repositorio institucional, en un sitio web personal, en redes sociales académicas, realizar una traducción o republicar el trabajo como un capítulo de libro), siempre y cuando se reconozca la autoría y la publicación inicial en esta revista.

Publisher

Cadernos de Tradução es una publicación del Programa de Posgrado en Estudios de Traducción de la Universidad Federal de Santa Catarina. La revista *Cadernos de Tradução* está alojada en el [Portal de Periódicos UFSC](#). Las ideas expresadas en este artículo son responsabilidad de sus autores y no representan necesariamente la opinión del equipo editorial o de la universidad.



Editores del número especial

Andréia Guerini – Fernando Ferreira Alves – Orlando Grossegese

Editor de sección

Willian Moura

Corrección de normas

Alice S. Rezende – Ingrid Bignardi – João G. P. Silveira – Kamila Oliveira

Histórico

Recibido el: 03-06-2023

Aprobado el: 17-02-2024

Revisado el: 28-02-2024

Publicación: 04-2024

