

Género y salud en Italia y España: un análisis lingüístico-computacional de las leyes que regulan la medicina de género¹

Cristina CALÒ
"La Sapienza" Università di Roma

Resumen

El artículo aplica técnicas de PLN para analizar la incorporación de la perspectiva de género en la sanidad pública de Italia y España. Italia, con la Ley 3/2018 y el Plan de 2019, establece explícitamente una 'medicina de género' que introduce criterios sexo-genéricos en la investigación y práctica clínica. España, sin disponer de una ley específica, distribuye esta perspectiva en diferentes normativas, como la LO 3/2007 y la Ley 33/2011. Se observa así un modelo italiano sistemático y orientado a lo biomédico, frente a uno español más fragmentado pero basado en principios de equidad.

Palabras clave: medicina de género, lenguaje de género, lingüística computacional, género y salud, lenguaje jurídico.

Abstract

Using natural language processing, this article examines how gender perspectives are integrated into the public health systems of Italy and Spain. In Italy, 'gender medicine' is formally recognized through Law 3/2018 and the 2019 Plan, which systematically embeds sex-gender considerations into clinical practice. Spain, which lacks a dedicated statute, incorporates gender perspectives via broader regulatory instruments, including Organic Law 3/2007 and Law 33/2011. Consequently, Italy pursues a more systematic, biomedical approach, whereas Spain adopts a more fragmented, equity-oriented model.

Keywords: Gender Medicine, Gender Language, Computational Linguistics, Gender and Health, Legal Language.

¹ El presente trabajo es el resultado de una investigación realizada en el marco del proyecto PRIN 2022 "Lengua, género y ciudadanía en las sociedades democráticas. DiLeGIS - Observatorio sobre el discurso legislativo y de género en Italia y España" (Prot. 2022E5TBEM), financiado por el Ministerio italiano dell'Università e della Ricerca (MUR). La redacción del presente artículo y la investigación en la que se basa, en todas sus fases y partes, son fruto de la estrecha y equilibrada colaboración entre ambas autoras.

1. INTRODUCCIÓN

En las últimas décadas, el concepto de ‘género’ se ha consolidado significativamente en el ámbito de la investigación científica y las políticas públicas internacionales, evolucionando desde una categoría marginal hacia una perspectiva clave para entender fenómenos sociales, psicológicos y sanitarios.

En particular, las diferencias biológicas entre sexos se reconocen hoy como factores determinantes para el bienestar físico y psicológico. De hecho, diversas investigaciones científicas han demostrado ampliamente que las variaciones hormonales, genéticas y metabólicas propias de hombres y mujeres influyen de manera decisiva en la incidencia y evolución de numerosas patologías, incluyendo enfermedades cardiovasculares, oncológicas, metabólicas, autoinmunes, psiquiátricas y neurológicas (Institute of Medicine, 2011; Werling y Geschwind, 2013; Migeon, 2014; Klein y Flanagan, 2016; Franceschini y Fattore, 2021; Reckelhoff, 2023). Sin embargo, la aplicación efectiva de este reconocimiento en políticas sanitarias concretas dentro del contexto europeo ha resultado notablemente desigual.

En este sentido, Italia y España constituyen casos particularmente ilustrativos. Aunque culturalmente cercanos, estos países muestran enfoques claramente diferenciados en la incorporación de la perspectiva de género en sus marcos legislativos y sanitarios.

Desde una perspectiva jurídico-institucional, Italia ha implementado recientemente una legislación específica sobre la denominada ‘medicina de género’ mediante la Ley 3/2018 (Ley Lorenzin) y su correspondiente Plan Operativo Nacional (2019). Esta iniciativa reconoce explícitamente que las diferencias biológicas y socioculturales vinculadas al género afectan significativamente al estado de salud, a la respuesta terapéutica y a la adecuación de los servicios médicos. Así, se ha generado una atención renovada entre profesionales sanitarios, instituciones y ciudadanía sobre esta temática.

España, en cambio, ha seguido un camino normativo distinto. A pesar de carecer de una legislación específica sobre medicina de género, ha incorporado la perspectiva de género en la salud pública mediante iniciativas legislativas y administrativas más amplias, como la Ley Orgánica 3/2007 para la igualdad efectiva entre mujeres y hombres y la Ley 33/2011, que establece explícitamente principios de equidad y paridad de género en las políticas de salud pública. Complementariamente, la Orden PCI/462/2019 recuperó y reguló el Observatorio de Salud de las Mujeres, cuyo objetivo operativo es identificar, analizar y corregir desigualdades y sesgos de género en la provisión de servicios sanitarios.

En este contexto, el presente estudio propone examinar críticamente, mediante técnicas de lingüística computacional, los marcos legislativos y de implementación descritos. El objetivo es analizar cómo los textos legislativos italianos y españoles construyen discursivamente la relación entre género y salud. Para ello, se realizará un análisis cualitativo y cuantitativo profundo de las estrategias léxicas y lingüísticas empleadas, así como de las implicaciones pragmáticas e institucionales subyacentes a

estas decisiones textuales. De esta forma, se pondrá de relieve cómo las palabras contenidas en la legislación, más allá de su función normativa, modelan significativamente el imaginario colectivo, las expectativas y las prácticas concretas relacionadas con género y salud en ambos contextos nacionales analizados.

2. MARCO TEÓRICO DE REFERENCIA

2.1. *Medicina de género: bases biomédicas y normativas*

La medicina de género representa un enfoque emergente en el ámbito de la biomedicina cuyo objetivo es investigar cómo las diferencias biológicas, determinadas por el sexo, y las socioculturales, relacionadas con el género, influyen en el estado de salud y en las condiciones patológicas de los individuos (Organización Mundial de la Salud, 2018). Históricamente, la investigación médica ha adolecido de un notable sesgo androcéntrico, privilegiando el modelo masculino y considerándolo a menudo erróneamente como universalmente neutro. Este enfoque ha generado importantes lagunas en el conocimiento, con considerables repercusiones negativas en la salud femenina. De hecho, es bien conocido que diversas enfermedades cardiovasculares, oncológicas y hepáticas presentan diferencias significativas en los síntomas, la incidencia y la respuesta farmacológica entre hombres y mujeres. Sin embargo, estas diferencias han sido durante mucho tiempo descuidadas en la práctica clínica.

En este sentido, varios estudios, entre ellos el de Baggio et al. (2013), han puesto de relieve la importancia de formular enfoques diagnósticos y terapéuticos diferenciados para garantizar equidad y eficacia en la atención sanitaria. Del mismo modo, Glezerman (2017) ha subrayado la necesidad de un cambio paradigmático en los protocolos clínicos, orientado a integrar sistemáticamente la variable sexo-género en las prácticas sanitarias. Este enfoque pretende evitar que se ignoren diferencias aparentemente mínimas, como el metabolismo farmacológico o la sintomatología cardiovascular, que pueden determinar resultados clínicos significativamente diferentes según el género.

En el plano normativo e institucional, la medicina de género ha obtenido un reconocimiento creciente durante las últimas décadas. El Instituto de Medicina de Estados Unidos reconoció formalmente desde 2001 la importancia de integrar la variable sexo-género en la investigación científica y en la práctica clínica (IOM, 2001). Paralelamente, la Organización Mundial de la Salud ha adoptado definiciones y directrices para favorecer la inclusión de la medicina de género en los sistemas sanitarios globales. En Europa, Italia ha sido pionera en introducir legislativamente la medicina de género con la Ley 3/2018, objeto del presente artículo, que obliga al Servicio Nacional de Salud a considerar sistemáticamente el género en todos los ámbitos clínicos y en la investigación farmacológica (Ley Lorenzin, art. 3). Esta medida, complementada en 2019 con un Plan Nacional de Aplicación coordinado por el Instituto Superior de Salud, refleja cómo la medicina de género ha evolucionado desde una perspectiva exclusivamente científica hacia una obligación jurídica y política (Ministerio de Salud, 2019).

2.2 *Género y lenguaje jurídico: la construcción discursiva del género en las normas*

Como es sabido, el lenguaje jurídico y los textos normativos desempeñan una función crucial en la construcción y representación social de los roles de género. Aunque tradicionalmente muchos idiomas hayan adoptado una aparente neutralidad lingüística, el llamado ‘masculino neutro’ ha sido objeto de críticas desde diversas perspectivas lingüístico-jurídicas que han evidenciado su carácter implícitamente discriminatorio (Pacella, 2016; Perra y Ruspini, 2015). El análisis del lenguaje jurídico, como señala Rodríguez Ruiz (2017), revela que las categorías ‘hombre’ y ‘mujer’ se han construido históricamente de manera asimétrica, perpetuando desigualdades sustanciales mediante elecciones terminológicas específicas.

En respuesta a esta situación, diversas iniciativas institucionales y supranacionales, como las Directrices del Parlamento Europeo (2008), han promovido activamente la adopción de un lenguaje inclusivo en las prácticas normativas y administrativas. En este sentido, se han propuesto y debatido críticamente estrategias lingüísticas como el doble género (‘el candidato o la candidata’), términos colectivos neutros (‘la ciudadanía’) y estructuras impersonales (Martínez Linares, 2022). A pesar de las inevitables controversias generadas, estas prácticas resultan fundamentales para garantizar una representación equilibrada y visible de los géneros en los textos jurídicos, contribuyendo implícitamente a la reducción de estereotipos y prejuicios.

2.3 *Lingüística computacional y análisis de textos legislativos*

Las tecnologías derivadas de la lingüística computacional, especialmente aquellas basadas en el procesamiento del lenguaje natural (PLN), están abriendo nuevas perspectivas metodológicas para el análisis científico de la representación del género en diversos tipos de textos, incluidos los legislativos. La disponibilidad de corpus digitalizados permite llevar a cabo análisis cuantitativos y cualitativos sin precedentes, facilitando la identificación precisa de frecuencias terminológicas y patrones sintácticos específicos (Pontrandolfo, 2019). Técnicas avanzadas como el etiquetado de partes del discurso (*part-of-speech tagging*), el reconocimiento de entidades nombradas (*named entity recognition*) y los algoritmos de modelado de temas (*topic modeling*) permiten descubrir asociaciones semánticas implícitas relacionadas con el género y evaluar de manera sistemática la coherencia y la inclusividad lingüística en los textos normativos (Mori, 2018; Agnoloni y Venturi, 2018).

Este enfoque metodológico no solo implica un progreso significativo respecto al análisis tradicional de la lingüística jurídica, sino que también representa una herramienta operativa valiosa para la redacción legislativa, ya que facilita la realización de revisiones lingüísticas basadas en datos empíricos objetivos. La integración de estas tecnologías responde eficazmente a los objetivos planteados desde la perspectiva de género, favoreciendo una transformación profunda del discurso jurídico y de las prácticas normativas hacia una mayor equidad e inclusividad.

En definitiva, la aplicación de la lingüística computacional a los textos legislativos ofrece dos ventajas fundamentales. Por una parte, permite medir empíricamente cómo

se representa el género en el lenguaje jurídico, por ejemplo cuantificando la presencia de formas femeninas, identificando estereotipos lingüísticos o analizando marcos semánticos relacionados con el género en corpus normativos específicos. Por otra parte, proporciona herramientas objetivas para orientar una revisión del lenguaje normativo desde una perspectiva inclusiva, basada en evidencias científicas verificables.

En este artículo utilizaremos precisamente estas herramientas computacionales para investigar cómo se construye y representa discursivamente el género en las normas italianas y españolas relativas a la medicina de género. Mediante el análisis lingüístico asistido por PLN trataremos de identificar posibles asimetrías discursivas, sesgos implícitos y estereotipos que aún podrían persistir en la formulación actual de las disposiciones legislativas.

3. METODOLOGÍA

3.1. *Construcción del corpus y criterios de selección de los textos*

Para la construcción del corpus se han seleccionado cinco textos normativos fundamentales para el desarrollo y la consolidación de la medicina de género en los contextos nacionales italiano y español. La elección se ha orientado específicamente hacia documentos oficiales que han introducido, promovido o ampliado significativamente el principio de integración de la perspectiva de género en el ámbito sanitario de los dos países.

En el caso italiano, el corpus incluye la Ley n.º 3 de 11 de enero de 2018, que constituye el primer documento legislativo italiano en introducir formalmente la perspectiva de género en todos los ámbitos de la medicina nacional. En particular, el artículo 3 de esta ley dispone la adopción de un Plan Nacional específicamente orientado hacia la difusión e integración sistemática de la perspectiva de género en la investigación, la prevención, el diagnóstico y el tratamiento de las enfermedades. Por esta razón, también se ha incluido en el corpus el ‘Plan Nacional para la Aplicación y Difusión de la Medicina de Género’ de 2019, elaborado por el Ministerio de Salud italiano, un instrumento estratégico que define objetivos específicos, acciones concretas y los actores involucrados en la efectiva integración de la perspectiva de género en los protocolos clínicos, la investigación científica, la innovación tecnológica, la formación profesional y la comunicación sanitaria.

Por lo que respecta al contexto español, la selección incluye la Ley Orgánica 3/2007, principal ley marco sobre igualdad de género que introdujo el principio de transversalidad (*mainstreaming*) en todas las políticas públicas, incluida la salud. En concreto, el artículo 27 de esta ley establece explícitamente la obligación de aplicar la perspectiva de género en los programas y estrategias sanitarias, sentando así las bases para una gestión integrada, inclusiva y no discriminatoria de los servicios de salud. A esta normativa se suma la Ley 33/2011 de Salud Pública, que reforzó e institucionalizó aún más la consideración del género en las políticas sanitarias españolas, especificando que toda intervención en salud pública debe considerar explícitamente las diferencias y

necesidades específicas de mujeres y hombres. Finalmente, se ha seleccionado la Orden PCI/462/2019, por la que se crea el Observatorio de Salud de las Mujeres adscrito al Ministerio de Sanidad, consolidando un mecanismo institucional dedicado específicamente al análisis de las desigualdades de género en salud y a la promoción activa de la equidad sanitaria.

3.2 *Preprocesamiento lingüístico*

Para cada documento del corpus se llevó a cabo un análisis lingüístico automatizado mediante herramientas avanzadas de procesamiento del lenguaje natural (PLN). En particular, los textos fueron tokenizados, es decir, segmentados en unidades léxicas fundamentales (palabras y signos de puntuación), utilizando la biblioteca spaCy v3.7.2 con modelos específicos entrenados para italiano y español (it_core_news_lg y es_core_news_lg, respectivamente).

Además, se aplicó un proceso de lematización y etiquetado morfosintáctico con estos mismos modelos, lo que permitió contabilizar tipos léxicos basándose en los lemas. Siguiendo las prácticas estándar en lingüística computacional (Manning y Schütze, 1999), el texto fue posteriormente normalizado mediante la conversión a minúscula, eliminando caracteres numéricos y signos de puntuación superfluos.

Finalmente, se efectuó la segmentación de los textos en frases mediante el analizador sintáctico de spaCy. Este procedimiento facilitó análisis estructurales posteriores, como la identificación de colocaciones léxicas.

3.3 *Métricas computacionales*

Tras el preprocesamiento y la anotación descritos en la sección anterior, se abordó la fase cuantitativa propiamente dicha. Para cada documento y cada idioma se elaboraron listas de frecuencia (absolutas y normalizadas por 1.000 palabras) con el fin de aislar los términos clave más recurrentes y garantizar la comparabilidad entre textos de distinta extensión.

En segundo lugar, se evaluó la riqueza léxica mediante diversos índices de diversidad del vocabulario. Inicialmente, se calculó la relación tipo-token (*type-token ratio*, TTR), que mide la proporción entre los tipos léxicos únicos y el total de *tokens*. Dado que la TTR depende considerablemente de la longitud del texto, se utilizó también el índice Root TTR de Guiraud ($RTTR = V/\sqrt{N}$, donde V son los tipos y N los tokens) con el fin de obtener una medida comparativa de la diversidad léxica independientemente del tamaño del documento. Además, se optó por emplear el índice D de Herdan (definido técnicamente como $\log V/\log N$, donde V es el número de tipos léxicos distintos y N es el número total de tokens del corpus), en lugar del índice K de Yule (*Yule's K*)² debido a su mayor estabilidad en corpus de tamaños desiguales y a su

² Yule's K es un índice clásico introducido por George Yule (1944) para cuantificar la tendencia a la repetición léxica: valores bajos indican un vocabulario diversificado, mientras que valores altos reflejan un uso más restringido y repetitivo de los términos.

menor sensibilidad a la influencia de los *hapax legomena* (palabras que aparecen una sola vez), lo que brinda una medida más fiable de la riqueza léxica comparativa. El uso conjunto de TTR, RTTR e índice D de Herdan permitió obtener un perfil sólido y equilibrado sobre la variedad léxica de los textos analizados.

Un aspecto clave del análisis fue la identificación de colocaciones o coocurrencias significativas para destacar las asociaciones terminológicas características del discurso legislativo sobre medicina de género. Las colocaciones se definieron principalmente como bigramas y trigramas de palabras adyacentes con frecuencia y asociación estadísticamente relevantes (por ejemplo, ‘medicina di genere’ en italiano, ‘salud pública’ en español). Para extraer dichas colocaciones se utilizó el índice PMI (Pointwise Mutual Information), una métrica que compara la probabilidad observada de coocurrencia de dos palabras con la probabilidad esperada bajo condiciones de independencia estadística. En el cálculo del PMI, siguiendo el enfoque clásico de Church y Hanks (1990), se aplicaron umbrales específicos para asegurar la fiabilidad de los resultados: solo se consideraron relevantes aquellos pares de palabras que coocurrieran al menos 5 veces en el corpus, excluyendo combinaciones demasiado raras, susceptibles de asociaciones fortuitas o espurias. Además, se consideraron significativas únicamente aquellas colocaciones con un $PMI \geq 3$ (en logaritmo base 2)³, valor comúnmente aceptado como indicativo de una fuerte asociación estadística. Posteriormente, las colocaciones y los *n-grams* seleccionados fueron analizados cualitativamente mediante concordancias (concordance analysis, contexto KWIC, *Key Word in Context*), lo que permitió interpretar adecuadamente su significado dentro del contexto jurídico analizado.

4. ANÁLISIS CUANTITATIVO Y CUALITATIVO DE LOS TEXTOS NORMATIVOS ITALIANOS Y ESPAÑOLES

4.1. *Análisis cuantitativo*

El análisis cuantitativo realizado sobre los textos normativos italianos y españoles ha puesto en evidencia diferencias significativas en el uso del léxico y en las colocaciones características de ambos contextos nacionales.

4.1.1. *Riqueza léxica y densidad textual*

En primer lugar, para cada corpus, se calcularon los índices de diversidad léxica descritos en la metodología (TTR, RTTR y *D* de Herdan). La Tabla 1 resume estas métricas para cada texto analizado, mostrando claras diferencias entre los documentos españoles e italianos:

³ En términos simples, dos palabras se consideran significativamente colocadas si su coocurrencia es al menos $2^3 = 8$ veces más frecuente de lo que cabría esperar por casualidad.

Diversidad léxica

Documento	TTR (%)	RTTR	Herdan-D	Tokens	Tipos	Densidad léxica (%)
Ley Orgánica 3/2007 (Esp.)	9,5	18,5	0,78	38.112	3.615	52,8
Ley 33/2011 (Esp.)	13,5	19,1	0,80	19.915	2.691	52,3
Orden PCI/2019 (Esp.)	24,5	13,6	0,83	3.082	756	59,7
Legge 3/2018 (Ita.)	13,7	17,0	0,79	15.242	2.093	58,0
Piano 2019 (Ita.)	20,9	33,2	0,85	25.247	5.274	61,5

Tabla 1 - Métricas de diversidad y densidad léxica de los textos legislativos sobre medicina de género en España e Italia.

Del análisis cuantitativo se desprende que los textos italianos y de menor extensión exhiben una mayor variedad léxica y densidad en comparación con los textos españoles más extensos, que presentan menor diversidad y mayor repetitividad. En concreto, el Piano italiano de 2019 (25.247 tokens) destaca por su riqueza léxica: contiene 5.274 tipos distintos (TTR \approx 20,9 %), la más elevada del corpus, junto con la mayor densidad léxica (61,5 % de palabras de contenido). Esto significa que, a pesar de no ser el texto más largo, el Piano incorpora un vocabulario notablemente amplio y un alto porcentaje de términos informativos. Por su parte, la Legge italiana 3/2018 (15.242 tokens) también muestra valores relativamente altos de diversidad (TTR 13,7 %; densidad 58,0 %), acordes con su extensión intermedia.

En contraste, los textos legislativos españoles de gran longitud revelan una diversidad léxica más limitada. La Ley Orgánica 3/2007, con 38.112 tokens, alcanza solo 3.615 tipos únicos, lo que se traduce en un TTR de 9,5 %, el más bajo entre los documentos analizados. De modo similar, la Ley 33/2011 (19.915 tokens) presenta un TTR de 13,5 % con 2.691 tipos. Estas cifras indican que, aunque estos textos españoles son muy extensos, no introducen un vocabulario proporcionalmente variado, sino que recurren con frecuencia a los mismos términos y fórmulas legales. Ello se refleja igualmente en su densidad léxica, cercana al 52 % en ambas leyes, significativamente inferior a la de los documentos italianos. Un caso particular es la breve Orden ministerial Orden PCI/462/2019 de España (3.082 tokens), que alcanza un TTR elevado (24,5 %) y una densidad léxica del 59,7 %. Este valor alto de diversidad, comparable al del *Piano* italiano, se explica por la menor extensión de la orden y su carácter específico, que limita la repetición excesiva de términos.

Las métricas normalizadas confirman esta tendencia. El índice RTTR y Herdan-D, más independientes de la longitud textual, son notablemente mayores en el *Piano* italiano de 2019 (RTTR = 33,2; D = 0,85) que en las leyes españolas más largas (por ejemplo, RTTR \approx 18,5 y D = 0,78 en la LO 3/2007). Esto sugiere que incluso al corregir el efecto del tamaño, el texto italiano mantiene la mayor riqueza de vocabulario. En cambio, los valores de Herdan-D para las leyes españolas (\approx 0,78–0,80) denotan una repetición léxica más alta relativa a su longitud. En suma, los textos españoles extensos muestran una menor variedad léxica y utilizan un conjunto más restringido de palabras

pese a su gran tamaño, mientras que los documentos italianos (y el español más breve) incorporan un vocabulario más diverso.

Desde el punto de vista estilístico y discursivo, estas diferencias en diversidad léxica y densidad sugieren distintas estrategias redaccionales. Los textos españoles, especialmente las leyes generales de gran alcance, tienden a un estilo jurídico formulaico y exhaustivo, con abundantes reiteraciones de términos legales y construcciones fijas que garantizan la coherencia normativa pero reducen la variación léxica.

En contraste, el *Piano* italiano 2019, de naturaleza programática y orientado a la implementación de la medicina de género, emplea un discurso más expositivo y técnico, introduciendo numerosos términos especializados y conceptos diversos (médicos, organizativos, formativos, etc.). Esto se traduce en una mayor riqueza léxica y en una densidad informativa superior, ya que el texto concentra una proporción más alta de palabras con carga conceptual. Un efecto similar se observa en la orden española de 2019, cuyo enfoque específico conlleva menos redundancias y más contenido léxico en poco espacio.

En definitiva, la mayor diversidad y densidad léxica de los documentos italianos (y del texto español breve) señalan un estilo más variado y preciso en la expresión de contenidos, posiblemente orientado a detallar aspectos técnicos y contextuales de la medicina de género. Por otro lado, la menor variedad y mayor repetitividad de las extensas leyes españolas reflejan un registro legal homogéneo y sistemático, que privilegia la claridad y uniformidad terminológica en la formulación de principios generales. Estas características léxico-discursivas diferenciadas tienen implicaciones en la accesibilidad y en el tono comunicativo de las normas: los textos con alta densidad y diversidad pueden resultar más especializados y cargados de información, mientras que aquellos con vocabulario más restringido tienden a ser más redundantes y formulaicos, algo propio de la tradición legislativa a la que pertenecen.

4.1.2. *Frecuencia y distribución de los términos clave*

En el corpus italiano se observa una clara centralidad del término ‘genere’ (género), que aparece con la mayor frecuencia relativa (aproximadamente 350 ocurrencias). Le siguen en frecuencia ‘salute’ (salud) y ‘medicina’, con alrededor de 200 apariciones, reflejando el foco temático de los textos. También destacan términos como ‘nazionale’ (nacional), ‘ricerca’ (investigación) y ‘donne’ (mujeres), en rangos entre 100 y 150 ocurrencias, mientras que ‘differenze’ (diferencias) figura con unas 80 apariciones. La presencia de vocablos administrativos-legales frecuentes como ‘comma’ (que en italiano designa los apartados de un artículo legal) o ‘presente’ (usado en expresiones como ‘la presente ley’) subraya el carácter normativo-formal de estos documentos. En conjunto, el léxico italiano combina términos de la esfera sanitaria (salute, medicina, ricerca) con otros propios del discurso legislativo (nazionale, comma, presente), evidenciando que se trata de una política de salud pública enmarcada en un plan nacional:

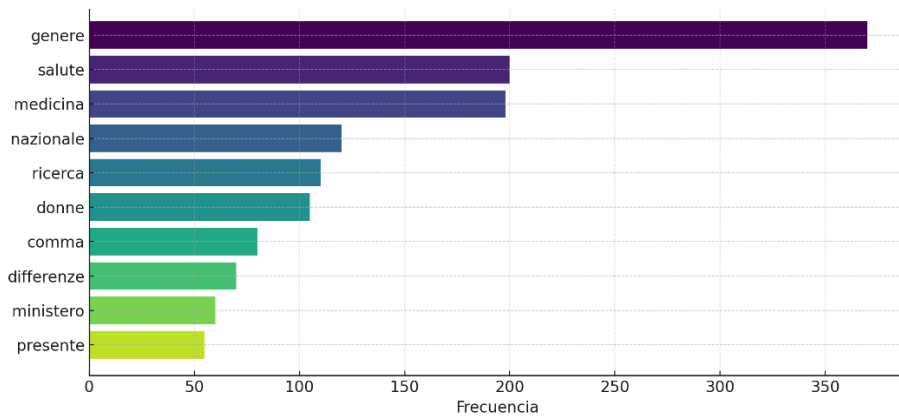


Figura 1 - Palabras más frecuentes en el corpus italiano

Por su parte, en el corpus español, las palabras más frecuentes reflejan tanto el contenido jurídico-administrativo como el énfasis específico en la igualdad de género. ‘Salud’ es el término más recurrente, seguido por vocablos genéricos del ámbito jurídico, como ‘artículo’ y ‘ley’, junto con el término clave ‘igualdad’. Además, destaca la presencia explícita de ‘mujeres’ (más de 200 ocurrencias), acompañada, aunque en menor medida, por ‘hombres’:

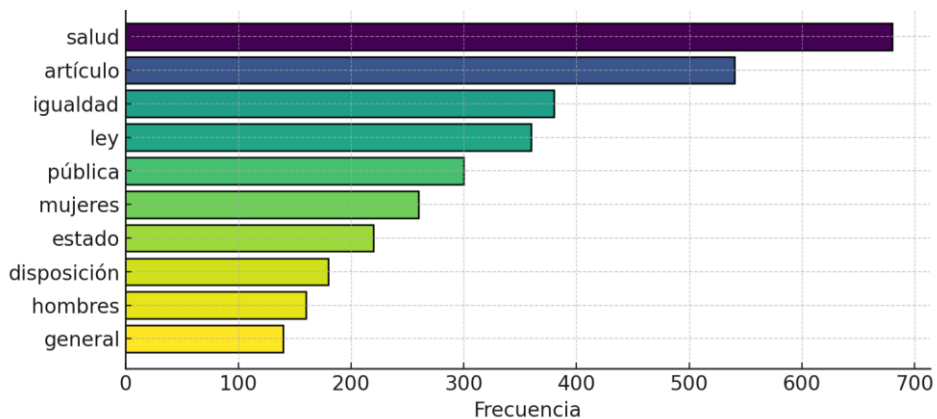


Figura 2 - Palabras más frecuentes en el corpus español

La presencia simultánea de ‘mujeres’ y ‘hombres’ en posiciones prominentes confirma el enfoque explícitamente binario adoptado en los textos normativos españoles sobre igualdad. Resulta relevante observar que la frecuencia de ‘mujeres’ es casi el doble que la de ‘hombres’ (unas 230 frente a aproximadamente 120 apariciones). Esto revela que, aunque la retórica normativa insiste en la visibilización simultánea de ambos sexos en cada enunciado relevante, garantizando que las disposiciones se apliquen de forma equitativa, se concede mayor atención al sujeto femenino,

coherentemente con políticas destinadas a corregir disparidades que tradicionalmente han afectado más a las mujeres.

En síntesis, mientras que el léxico italiano privilegia principalmente la noción abstracta de ‘género’ y términos técnico-sanitarios, los textos españoles enfatizan el concepto de ‘igualdad’, mencionando explícitamente y con frecuencia ambos sexos.

La distribución interna de estos términos clave en los textos analizados confirma y refuerza las diferencias de enfoque señaladas. El gráfico de dispersión KWIC ilustra claramente la posición y concentración relativa de apariciones de términos clave como ‘genere/género’, ‘donne/mujeres’ y ‘uomini/hombres’ a lo largo de cada documento:

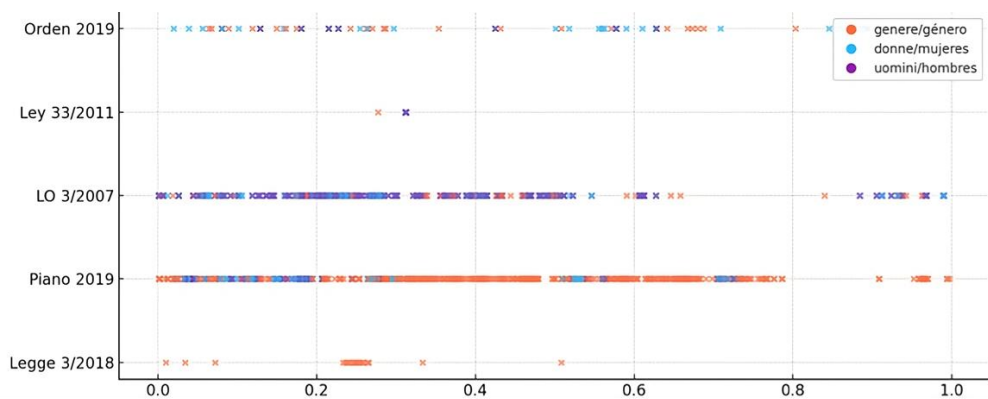


Figura 3 - Dispersión KWIC – ocurrencias de términos clave en los textos

En concreto, como se aprecia en la Figura 3, el término ‘genere/género’ aparece de manera difundida en prácticamente todos los documentos de cada corpus, evidenciando su rol conceptual primordial. En cuanto a los sustantivos que designan a las poblaciones por sexo, el contraste es marcado: en el corpus italiano, el término ‘donne’ (mujeres) aparece exclusivamente en el Plan Nacional de 2019, mientras que la Ley Marco 3/2018 no lo menciona en absoluto. Esta ausencia de referencias explícitas a mujeres (y también a hombres) en el texto legal italiano de 2018 parece obedecer a un afán deliberado por mantener un lenguaje neutro e inclusivo, centrado en la noción general de ‘género’ y evitando destacar un sexo sobre otro. En contraposición, la presencia prominente de ‘donne’ en el Plan de 2019 sugiere una voluntad política de visibilizar problemáticas sanitarias específicas de las mujeres, posiblemente como respuesta para corregir desigualdades históricas. Este contraste léxico (de la neutralidad a la mención explícita) refleja cómo las instituciones italianas adaptaron su discurso: primero con una ley marco de tono generalista, luego con un plan detallado que pone el acento en diferencias sexo-específicas. Por el contrario, en los textos españoles las ocurrencias de mujeres y hombres están más equilibradas y muy a menudo aparecen juntas en la formulación (“mujeres y hombres”).

4.2 Colocaciones/n-grams y temas emergentes en los textos

El análisis de colocaciones y *n-grams* ha permitido identificar patrones particularmente interesantes en ambos corpora. En el corpus italiano, la colocación más frecuente es claramente ‘medicina di genere’ (alrededor de 150 ocurrencias), lo que confirma que el tema principal es la institucionalización de una ‘medicina’ orientada a las necesidades específicas derivadas del género. A continuación, destaca ‘differenze di genere’ (aproximadamente 100 ocurrencias), expresión clave que enfatiza la importancia atribuida a las diferencias biológicas y socioculturales entre hombres y mujeres en la investigación y la práctica clínica.

Otras colocaciones relevantes incluyen ‘professioni sanitarie’ (~70 apariciones), término relacionado con las disposiciones normativas relativas a la formación y participación de los profesionales de la salud, así como ‘sperimentazione clinica’ e ‘ricerca clinica’ (más de 25 apariciones cada una), destacando la necesidad explícita de considerar el género en los ensayos clínicos y la investigación científica.

Finalmente, la frecuente aparición de la colocación ‘dispositivi medici’ (~50 apariciones), aunque no directamente vinculada al género, refleja la naturaleza ‘ómnibus’ de la Ley italiana 3/2018, que regula múltiples aspectos sanitarios, incluyendo capítulos específicos sobre las profesiones sanitarias y la normativa de dispositivos médicos. Esta circunstancia diluye ligeramente el enfoque temático del corpus italiano, aunque este permanece claramente centrado en la perspectiva de género:

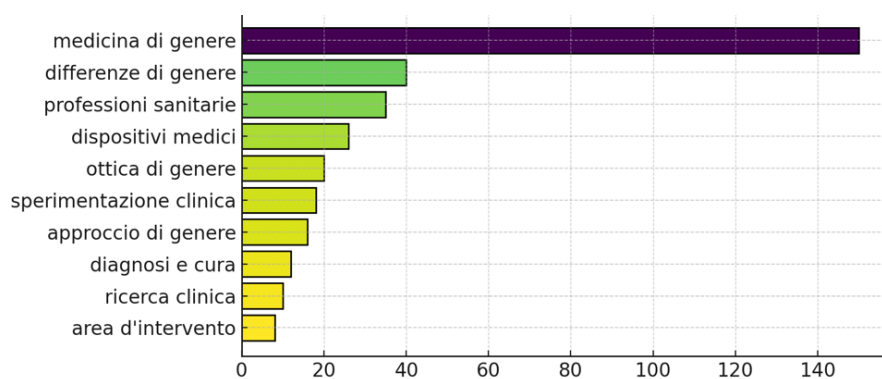


Figura 4 Colocaciones más frecuentes en el corpus italiano

En el corpus español, las colocaciones más frecuentes destacan claramente dos ejes temáticos principales: por un lado, la igualdad de género (en términos binarios), y por otro, la salud pública. La colocación más recurrente es ‘salud pública’ (más de 200 ocurrencias), lo cual refleja directamente el objeto central de la Ley 33/2011. A continuación, encontramos el bigrama ‘mujeres y hombres’ (~170 apariciones), expresión que materializa explícitamente el principio de igualdad entre los sexos, reiterada constantemente sobre todo en la Ley Orgánica 3/2007.

Otras colocaciones significativas incluyen ‘principio de igualdad’ (~120 ocurrencias), referencia directa al principio jurídico consagrado en la legislación sobre igualdad; así como ‘por razón de sexo’ (~60 ocurrencias), fórmula jurídica específica utilizada para prohibir explícitamente la discriminación basada en el género. La

presencia recurrente de estas frases formularias refleja cómo el lenguaje español está fuertemente anclado en un registro jurídico-institucional que enfatiza los derechos y obligaciones de los ciudadanos desde una perspectiva de igualdad.

La expresión ‘administraciones públicas’ (~80 ocurrencias) también emerge con frecuencia, subrayando la importancia asignada a las instituciones como responsables fundamentales de promover tanto la igualdad como la salud pública. Asimismo, colocaciones como ‘igualdad de trato’ (~50 ocurrencias) e ‘igualdad entre mujeres y hombres’ (~50 ocurrencias) reafirman la igualdad formal entre los sexos como núcleo temático de las leyes analizadas.

Es notable además el *n-gram* ‘salud de la población’ (~45 ocurrencias), indicando una perspectiva inclusiva y amplia sobre la salud pública. Finalmente, llama la atención la frecuente aparición de ‘lactancia natural’ (~35 ocurrencias), colocación peculiar probablemente asociada a medidas específicas para promover la lactancia materna, contempladas en la LO 3/2007. Este último aspecto señala que, en el discurso legislativo español, se considera también la dimensión reproductiva de la salud:

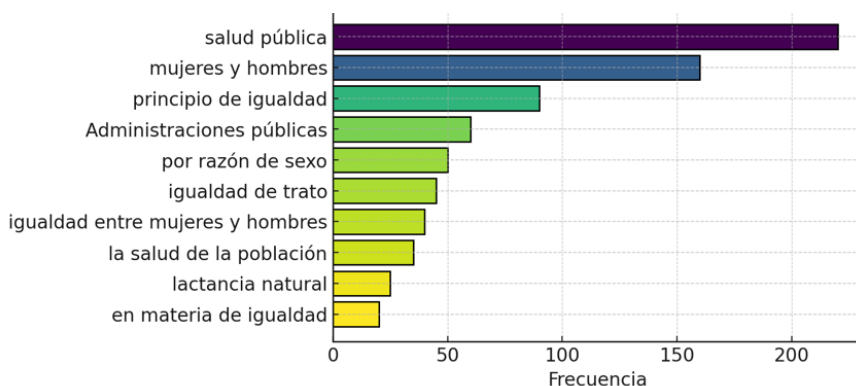


Figura 5 – Colocaciones más frecuentes en el corpus español

5. ANÁLISIS CUALITATIVO

Las diferencias léxicas identificadas mediante el análisis cuantitativo delimitan claramente dos perspectivas discursivas diferenciadas en los corpus analizados.

En particular, en el contexto italiano emerge con fuerza una retórica de la diferencia, articulada especialmente a través de la expresión recurrente ‘diferencias de género’. Este sintagma aparece asociado frecuentemente con términos y expresiones técnicas relacionadas con la investigación clínica, los ensayos farmacológicos y los procedimientos diagnósticos y terapéuticos, tal y como muestran los ejemplos:

1. [...] è predisposto un Piano formativo nazionale per la medicina di genere, volto a garantire la conoscenza e l'applicazione dell'orientamento alle differenze di genere nella ricerca, nella prevenzione, nella diagnosi e nella cura. (Legge 11 gennaio 2018, n. 3)

2. Le malattie respiratorie costituiscono oggi uno degli ambiti di indagine più interessanti rispetto alle differenze di genere. Differenze di genere in ambito respiratorio sono infatti presenti già nel periodo embrionale. (Piano per l'applicazione e la diffusione della Medicina di Genere, in attuazione dell'articolo 3, comma 1, Legge 3/2018)
3. In compenso, se l'uomo si ammala, ad esempio di malattia di Graves o lupus eritematoso sistemico, ha una prognosi peggiore. Pertanto, ogni sforzo di ricerca deve considerare le differenze di genere sia dal punto di vista preclinico (con modelli che rispecchino entrambi i sessi), sia da quello clinico. (Piano per l'applicazione e la diffusione della Medicina di Genere, in attuazione dell'articolo 3, comma 1, Legge 3/2018)
4. In campo oncologico sono state descritte molte differenze di genere, tuttavia le donne rimangono ancora sottorappresentate nelle sperimentazioni cliniche in aree importanti come l'oncologia non genere-specifica dove, ad esempio, costituiscono appena il 38,8% (Piano per l'applicazione e la diffusione della Medicina di Genere, in attuazione dell'articolo 3, comma 1, Legge 3/2018)

Dentro de este marco discursivo, el concepto de 'diferencia' adquiere una función explicativa fundamental, basada en la premisa de que los cuerpos masculinos y femeninos, junto con los factores socioculturales asociados, presentan características distintas en términos de incidencia, sintomatología y respuesta farmacológica ante diversas patologías. De esta constatación se deriva la prescripción normativa contenida en la Ley italiana 3/2018, que establece explícitamente la necesidad de considerar las diferencias de género para asegurar la calidad y la idoneidad de los servicios sanitarios.

En cuanto al estilo comunicativo, los textos italianos se caracterizan predominantemente por un lenguaje clínico-científico orientado hacia los profesionales sanitarios e instituciones de investigación. Asimismo, la sintaxis relativamente breve y lineal del Plan italiano de 2019 refleja una intencionalidad claramente operativa, orientada a facilitar la claridad y eficacia comunicativa en la implementación práctica de las directrices.

Por otro lado, en el caso español, la estrategia discursiva es distinta, en buena medida influida por la presencia de la Ley Orgánica 3/2007, que actúa como ley marco sobre la igualdad de género. Aquí la retórica es explícitamente binaria, articulándose en torno a la igualdad formal entre mujeres y hombres, fórmula reiterada insistentemente en todos los textos analizados. Con este propósito, el corpus español menciona frecuentemente ambos sexos, enfatizando así explícitamente el principio de equivalencia y no discriminación, como muestran los ejemplos:

5. En especial, queda prohibida toda discriminación entre mujeres y hombres en las actuaciones de salud pública, de acuerdo con lo establecido por la Ley Orgánica 3/2007, de 22 de marzo, para la igualdad efectiva de mujeres y hombres, así como por la demás normativa existente en esta materia. (Ley 33/2011, de 4 de octubre)
6. Las Administraciones públicas garantizarán un igual derecho a la salud de las mujeres y hombres, a través de la integración activa, en los objetivos y en las actuaciones de la política de salud, del principio de igualdad de trato, evitando que por sus diferencias biológicas o por

los estereotipos sociales asociados, se produzcan discriminaciones entre unas y otros. (Ley Orgánica 3/2007, de 22 de marzo)

7. [...]se pretende dotar a la Administración General del Estado de un mecanismo que permita fortalecer el enfoque de género en las políticas sanitarias, contribuyendo a la consecución de la equidad en salud, mediante la mejora del conocimiento de las causas y determinantes de la salud de las mujeres, la difusión de conocimientos que contribuyan a promocionar servicios de calidad adecuados, la promoción de análisis y estudios en colaboración con los agentes sociales implicados, la promoción del debate acerca de cómo eliminar las desigualdades, la cooperación con otras instituciones, la formación y la investigación. (Orden PCI/462/2019, de 17 de abril)

En consecuencia, incluso cuando el discurso aborda la salud pública (como en la Ley 33/2011 o en la Orden 2019), lo hace siempre en clave de igualdad de derechos entre mujeres y hombres, o de manera orientada a la eliminación de las desigualdades históricas existentes entre ambos sexos.

6. CONCLUSIONES

En conclusión, el análisis conjunto cuantitativo y cualitativo realizado en este estudio pone claramente de manifiesto que, aunque los textos normativos italianos y españoles comparten el objetivo general de integrar la dimensión de género en el ámbito sanitario, las estrategias lingüísticas y narrativas empleadas en ambos contextos son notablemente diferentes.

En particular, Italia adopta explícitamente el paradigma de la medicina de género, centrandó su atención en las diferencias biológicas y sanitarias con el fin de mejorar la práctica clínica. Para ello, utiliza un lenguaje predominantemente técnico-científico, menos connotado políticamente, con escasas referencias directas a ‘mujeres’ y ‘hombres’ como grupos sociales específicos.

Por el contrario, España inscribe claramente la salud pública dentro del marco de las políticas de igualdad de género, adoptando una estrategia discursiva en la que el dualismo explícito mujeres-hombres aparece constantemente. El léxico empleado se centra en conceptos relacionados con derechos, deberes y equidad (igualdad, discriminación, trato), en línea con una tradición normativa que subraya la justicia social y la falta de discriminación.

Estas diferencias discursivas quedan plenamente corroboradas por los resultados obtenidos a partir del análisis computacional: desde las frecuencias léxicas (por ejemplo, predominio del término ‘género’ en italiano frente a ‘igualdad’ en español), pasando por las colocaciones más destacadas (‘medicina de género’ en italiano frente a ‘salud pública’ y ‘mujeres y hombres’ en español), hasta la distribución interna de los términos clave en los textos normativos.

En definitiva, mientras que el estilo normativo italiano parece más orientado hacia una reforma técnico-científica del sistema sanitario desde una perspectiva de género, el enfoque español privilegia un marco jurídico-igualitario que impregna transversalmente las políticas sanitarias. Esta distinción discursiva refleja claramente trayectorias político-

culturales diferenciadas en ambos países: en Italia predomina la perspectiva del género entendido como diferencia biológica y sanitaria, mientras que en España prevalece la perspectiva del género como igualdad formal y sustantiva.

De cara al futuro, resultaría fructífero extender este análisis a otros países y contextos para verificar si los patrones identificados son generales o dependen de las tradiciones jurídico-culturales nacionales. Asimismo, la inclusión de nuevos tipos de textos, como guías clínicas, normativas regionales o debates parlamentarios, permitiría una visión más holística de cómo se articula la perspectiva de género en la praxis sanitaria más allá de las leyes estatales. Metodológicamente, el empleo de otras técnicas avanzadas de PLN (Procesamiento de Lenguaje Natural) ofrecería un valor añadido: *topic modeling* para descubrir temas latentes recurrentes en los discursos normativos, o reconocimiento de entidades para mapear actores e instituciones clave, podrían complementar el análisis léxico-semántico tradicional, proporcionando una comprensión más profunda de las narrativas de género en la legislación.

BIBLIOGRAFÍA

- AGNOLONI, Tommaso; VENTURI, Giulia (2018): “Semantic Processing of Legal Texts”, in Visconti, Jacqueline (ed.) (2018): *Handbook of Communication in the Legal Sphere*, Berlin/Boston: De Gruyter Mouton, pp. 109–138.
- BAGGIO, Giovannella; CORSINI, Alberto; FLOREANI, Annarosa; GIANNINI, Sandro; ZAGONEL, Vittorina (2013): “Gender medicine: a task for the third millennium”, *Clinical Chemistry and Laboratory Medicine*, 51(4), pp. 713-727.
- BIBER, Douglas; CONRAD, Susan; REPPEN, Randi (1998): *Corpus Linguistics: Investigating Language Structure and Use*, Cambridge: Cambridge University Press.
- CAVAGNOLI, Stefania (2013): *Linguaggio giuridico e lingua di genere: una simbiosi possibile*, Alessandria: Edizioni dell’Orso.
- CHURCH, Kenneth W.; HANKS, Patrick (1990): “Word Association Norms, Mutual Information, and Lexicography”, *Computational Linguistics*, 16(1), pp. 22-29.
- EUROPEAN PARLIAMENT (2008): *Gender-Neutral Language in the European Parliament: Guidelines*, Luxembourg: Office for Official Publications of the European Communities.
- FELICI, Annarita; MORI, Laura (2019), “Corpora di italiano legislativo a confronto: dall’Unione europea alla Cancelleria svizzera”, in *Le tendenze dell’italiano contemporaneo rivisitate. Atti del LII Congresso Internazionale della SLI*, Milano: Officinaventuno, pp. 287–304.
- FRANCESCHINI, Anna; FATTORE, Liana (2021): “Gender-specific approach in psychiatric diseases: Because sex matters”, *European Journal of Pharmacology*, 896, p. 173895.
- FRANCESCONI, Enrico; MONTEMAGNI, Simonetta; PETERS, Wim; TISCORNIA, Daniela (eds.) (2010): *Semantic Processing of Legal Texts: Where the Language of Law Meets the Law of Language*, Berlin/Heidelberg: Springer (Lecture Notes in Computer Science, 6036).
- GLEZERMAN, Marek (2017): *Medicina de género: La nueva revolución de la medicina: el papel del sexo en el diagnóstico y el tratamiento de enfermedades*, Barcelona: Plataforma Editorial.
- INSTITUTE OF MEDICINE, COMMITTEE ON UNDERSTANDING THE BIOLOGY OF SEX AND GENDER DIFFERENCES (2001): *Exploring the Biological Contributions to Human Health: Does Sex Matter?*, Washington: National Academies Press.
- INSTITUTE OF MEDICINE (2011): *Sex Differences and Implications for Translational Neuroscience Research: Workshop Summary*, Washington: National Academies Press.
- KLEIN, Sabra L.; FLANAGAN, Kathryn L. (2016): ‘Sex differences in immune responses’, *Nature Reviews Immunology*, 16(10), pp. 626-638.
- MANNING, Christopher D.; SCHÜTZE, Hinrich (1999): *Foundations of Statistical Natural Language Processing*, Cambridge (MA): MIT Press.
- MARTÍNEZ LINARES, M. Antonia (2022): “Sobre los dobles de género y cuestiones gramaticales conexas”, *CLAC*, 89, pp. 71-88.
- MCENERY, Tony; HARDIE, Andrew (2012): *Corpus Linguistics: Method, Theory and Practice*, Cambridge: Cambridge University Press.

- MIGEON, Barbara R. (2014): *Females Are Mosaics: X Inactivation and Sex Differences in Disease* (2^a ed.), Oxford: Oxford University Press.
- MORRA, Lucia; PASA, Barbara (eds.) (2015): *Questioni di genere nel diritto: implicite e crittotipi*, Torino: Giappichelli.
- ORGANIZACIÓN MUNDIAL DE LA SALUD/WORLD HEALTH ORGANIZATION (2011): *Gender Mainstreaming for Health Managers: A Practical Approach, Facilitators' Guide & Participant's Notes*, Geneva: WHO.
- PACELLA, Gemma (2016): "Il linguaggio giuridico di genere: la rappresentazione sessuata dei soggetti nel diritto e nella regolamentazione lavoristica", *Lavoro e diritto*, 3, pp. 481-500.
- PERRA, Margherita Sabrina; RUPINI, Elisabetta (2015): "La società del maschile 'neutro'. Alle radici dell'ostilità verso un linguaggio sessuato e 'non umano'", *in Genere.it*, 21 aprile 2015.
- PONTRANDOLFO, Gianluca (2019): "Corpus Methods in Legal Translation Studies", in Biel, Lucja; Engberg, Jan; Martín Ruano, Raquel; Sosoni, Vilelmini (eds.): *Research Methods in Legal Translation and Interpreting: Crossing Methodological Boundaries*, London/New York: Routledge, pp. 13-28.
- PONTRANDOLFO, Gianluca (2019): "Legal Linguistics in Italy", in Vogel, Friedrich (ed.): *Legal Linguistics Beyond Borders: Language and Law in a World of Media, Globalisation and Social Conflicts*, Berlin: Duncker & Humblot, pp. 59-79.
- RECKELHOFF, Jane F. (2023): "Mechanisms of sex and gender differences in hypertension", *Journal of Human Hypertension*, 37(8), pp. 596-601.
- RODRÍGUEZ RUIZ, Blanca (2017): *Género y Constitución. Mujeres y Varones en el Orden Constitucional Español*, Lisboa: Juruá Editorial.
- WERLING, Donna M.; GESCHWIND, Daniel H. (2013): "Sex differences in autism spectrum disorders", *Current Opinion in Neurology*, 26(2), pp. 146-153.
- WORLD HEALTH ORGANIZATION (2018): *Gender Equality Must Be at the Core of Health for All: Statement by the Director-General on International Women's Day (7 March 2018)*, Geneva: WHO.