

Detección de sesgos en razón del género en decisiones judiciales utilizando PLN

Mariana Brocca¹, J. Andrés Díaz Pace², Diego García Olano³, Laura Giosa¹, Daniela Godoy², Camila Lospinoso¹, Alfonsina Soto Pineda¹, Silvia Schiaffino², Antonela Tommasel², Ezequiel Valicenti¹

¹ Facultad de Derecho, UNCPBA, Azul, Argentina
{marianabrocca@outlook.com; lauramariagiosa@gmail.com; camilalospinoso@gmail.com, alfonsinastop@gmail.com; ezequielvalicenti@gmail.com}.

² ISISTAN (CONICET-UNCPBA), Tandil, Argentina
{andres.diazpace, daniela.godoy, silvia.scchiaffino, antonela.tommasel}@isistan.unicen.edu.ar

³ University of Texas, Austin, USA
diegoolano@utexas.edu

Resumen. Desde hace algunos años se ha instalado con fuerza la preocupación por la adopción de perspectivas de género en los operadores del Poder Judicial. Los estereotipos de género se refieren a la construcción social y cultural de hombres y mujeres, en razón de sus diferentes funciones físicas, biológicas, sexuales y sociales. Se han realizado algunos estudios que analizan de qué manera los estereotipos inciden en el razonamiento judicial. En este trabajo, a partir de un proyecto de investigación interdisciplinar, se propone utilizar técnicas de PLN para detectar estos estereotipos en decisiones judiciales de manera semi-automática. El objetivo es asistir a la revisión y análisis de resoluciones con el fin de detectar situaciones y razonamientos que evidencien discriminación en razón del género. Los resultados preliminares obtenidos hasta el momento han permitido identificar conjuntos de frases asociadas típicamente a los diferentes estereotipos.

1 Introducción

La perspectiva de género en las decisiones judiciales es un área que ha recibido creciente atención en los últimos años [1]. De acuerdo a la Comité para la Eliminación de todas las Formas de Discriminación contra la Mujer (CEDAW), una de las obligaciones del Estado es “hacer frente a las relaciones prevalecientes entre los géneros y a la persistencia de estereotipos basados en el género (...)” [5]. Esta obligación de evitar los estereotipos de género alcanza a los/as jueces/zas y funcionarios/as judiciales, como integrantes de uno de los poderes del Estado. No obstante, en Argentina se evidencia todavía una falta de perspectiva de género en muchas sentencias judiciales.

Desde un enfoque socio-cultural, un estereotipo es “una visión generalizada o una preconcepción sobre los atributos o características de los miembros de un grupo en particular o sobre los roles que tales miembros deben cumplir” [4,7]. Los estereotipos de género se refieren a la construcción social y cultural de hombres y mujeres, en razón de sus diferentes funciones físicas, biológicas, sexuales y sociales [4]. Se han identificado diferentes estereotipos de género: a) estereotipos de sexo centrados en las diferencias físicas y biológicas entre hombres y mujeres (por ej. la fuerza física relativa de hombres y mujeres); b) estereotipos sexuales referidos a la interacción sexual entre hombres y mujeres (por ej. se asocia la sexualidad femenina a la procreación); c) estereotipos sobre los roles sexuales que aluden a los comportamientos que se atribuyen a (y se esperan de) los hombres y las mujeres con base en sus construcciones físicas, sociales y culturales (por ej. el hombre es el proveedor, y la mujer asume las tareas de cuidado); d) estereotipos compuestos, en los que interactúan con otros estereotipos que asignan atributos, características o roles a diferentes subgrupos de mujeres [4].

Partiendo de dichas construcciones, diferentes estudios analizan de qué manera los estereotipos inciden en el razonamiento judicial. En estos estudios de casos, los análisis han sido llevados a cabo individualmente, mediante un estudio particular (y manual) de los textos de cada una de las sentencias (ver en este sentido, los casos analizados en [1,15]). Si bien este análisis es eficaz, involucra un análisis laborioso y acotado de cada resolución, y no es fácilmente escalable a un cuerpo de sentencias, ya que involucraría un esfuerzo considerable por parte de expertos en derecho. La capacidad de analizar un cuerpo de sentencias permitiría contar con un panorama (estadísticamente) más representativo del uso de estereotipos y su rol en los resultados de las sentencias (por ej., sesgos), y más aún, proveería de herramientas al Poder Judicial para llevar adelante políticas tendientes a mejorar la situación actual de falta de perspectiva de género. En este contexto, las técnicas de Inteligencia Artificial (IA) y particularmente de Procesamiento de Lenguaje Natural (PLN) surgen como un enfoque prometedor para el problema de análisis semi-automático de una gran cantidad de textos judiciales. El PLN explora mecanismos computacionales para comprender y manipular el lenguaje natural escrito u oral, de manera que se pueda hacer uso del mismo.

En este trabajo se propone explorar herramientas de análisis aplicables a una generalidad de casos, que faciliten al operador jurídico la tarea de evaluación mencionada. En base a conceptualizaciones existentes sobre los estereotipos de género, se presenta un enfoque y estrategias iniciales para detectar un conjunto de sesgos por las formas en que estos se presentan en la argumentación escrita que construyen los/as jueces/as en sus sentencias. La hipótesis es que los sesgos pueden identificarse a partir de un análisis contextual del texto, con asistencia de técnicas de PLN, que permitan analizar los contextos gramaticales, sintácticos y lingüísticos con presencia (o no) de un sesgo o preferencia discriminatoria.

Los resultados preliminares obtenidos hasta el momento con un conjunto pequeño de sentencias judiciales, han permitido identificar conjuntos de frases y palabras que se encuentran presentes en los diferentes estereotipos analizados. Esta información permitirá luego semi-automatizar el proceso de detección de estos estereotipos en nuevos documentos.

El artículo se organiza de la siguiente manera. En la sección 2 se analizan antecedentes del uso de PLN en el campo del derecho y los tribunales. En la sección 3 se presenta el concepto de estereotipos judiciales, que es central en nuestra propuesta. En la sección 4 se describe el enfoque propuesto. Luego, en la sección 5 se detallan algunos resultados preliminares. Finalmente, en la sección 6 se presentan las conclusiones y trabajos futuros.

2 Antecedentes

El desarrollo de técnicas de PLN ha impulsado diversas aplicaciones en el ámbito jurídico. Pueden mencionarse herramientas destinadas a: asistir en la revisión de documentos, permitiendo el trabajo con un gran volumen de datos [3]; automatizar el flujo de procesos, por ejemplo, de procesos judiciales o de sus etapas iniciales [5]; predecir posibles resultados, a partir del análisis y revisión de precedentes judiciales previos (vgr. Lexmachina.com¹); entre otras aplicaciones.

Se pueden mencionar también proyectos a nivel europeo como MIREL², que han usado PLN para la creación internacional de un marco formal de trabajo en el minado y comprensión de textos legales. La facultad de Derecho de Harvard³ ha digitalizado toda su base de jurisprudencia, la mayor de EE.UU. junto a la de la Biblioteca del Congreso de EE.UU. En materia de términos y condiciones, existen proyectos como Usable Privacy⁴, que usando PLN analiza la estructura de políticas de privacidad para facilitar al usuario la detección de las áreas que más le importan.

En el caso de nuestro país existen algunos desarrollos, por ejemplo Sherlock legal⁵, que es un buscador cognitivo que utiliza IA para responder preguntas realizadas en lenguaje natural desde la base de fallos de eDial.com. El buscador, luego de seleccionar un fallo judicial que resulte de interés, ofrece también documentos similares a la temática seleccionada. Otro antecedente, constituye el caso de Prometea, un sistema creado en conjunto por el Ministerio Público Fiscal de CABA y el Laboratorio de IA de la Facultad de Derecho de UBA (IALAB), que brinda asistencia para realizar tareas como la confección de documentos (pliegos de licitaciones, acuerdos homologatorios), y búsquedas de sentencias previas, normas y otros documentos⁶.

¹ <https://lexmachina.com/legal-analytics/>

² <http://www.mirelproject.eu/index.html>

³ <https://www.lawsitesblog.com/2016/06/harvard-ravel-digitization-project-adds-mass-delaware-cases.html>

⁴ <https://usableprivacy.org/>

⁵ http://www.albrematica.com.ar/sherlock/?utm_source=webjornada&utm_medium=web&utm_campaign=web_jornada_IA

⁶ Ver <https://ialab.com.ar/proyectos-de-impacto/>

3 Estereotipos judiciales

Según el Alto Comisionado de las Naciones Unidas para los Derechos Humanos (ACNUDH), el término **estereotipación judicial** se refiere a la “práctica mediante la cual los jueces/zas adjudican a una persona atributos, características o roles específicos solo por pertenecer a un grupo social determinado (por ejemplo, ser mujer). También se utiliza para referirse a la práctica de los jueces y juezas de perpetuar estereotipos perjudiciales al no cuestionar juzgar con estereotipos, por ejemplo, por tribunales inferiores o las partes en los procedimientos judiciales” [14].

Conforme lo señalado anteriormente, dentro de los estereotipos podemos encontrar a los estereotipos de género que constituyen prejuicios generalizados sobre las características que poseen o deberían poseer los hombres o las mujeres. Para ilustrar el concepto, en la Tabla 1 se detallan un conjunto de estereotipos considerados en este trabajo: a) de sexo; b) sexual; c) roles de género; d) compuestos. Para cada uno de ellos se detallan ejemplos concretos y cómo se reflejan en algunas prácticas. Ciertamente, este conjunto de estereotipos puede ser refinado y ampliado.

A modo de ejemplo, se incluyen algunas frases extraídas de una sentencia⁷, donde se puede observar el estereotipo sexual: “(...) la institución del matrimonio como cédula básica de la sociedad y establece las condiciones mediante las cuales ha de ser legítima la unión intersexual entre un hombre, y una mujer en el sentido en que ha de ser reconocida y protegida como tal con el fin de perpetuar la especie en el orden natural”; “Se trata de un acto cuyo objeto atenta contra los intereses generales o colectivos”; “dos personas del mismo sexo contrajeron matrimonio, en tanto su objeto atenta contra los intereses colectivos y, en especial, contra el orden público que impone el régimen del matrimonio”. Es nuestra hipótesis que la utilización de ciertas palabras y frases que aluden directa o indirectamente al estereotipo, podrían ser detectadas con un modelo de PLN entrenado adecuadamente para este tipo de contextos.

4 Enfoque propuesto

En el presente trabajo, el problema a resolver se concentra en el procesamiento de resoluciones judiciales con el objetivo de detectar distintos tipos de sesgos que impliquen una posible discriminación en razón del género. Los estereotipos son un ejemplo común de este tipo de sesgos. El objetivo a largo plazo de la investigación apunta a definir un método basado en adaptaciones de técnicas de PLN al dominio jurídico, para asistir a la revisión y análisis de resoluciones con el fin de detectar situaciones y razonamientos que evidencien sesgos en razón del género. Se espera que este método, en conjunto con una herramienta prototipo semi-automática, permitirán reducir los tiempos de análisis de los pronunciamientos judiciales a efectos de comprobar si estos cuentan con perspectiva de género, que es obligatoria en virtud de diversas normas de derecho internacional y derecho interno. En base a estos

⁷ Juzgado Nacional de 1a Instancia en lo Civil Nro. 106, “Lamuedra, Ernesto Ricardo c. L., A. L. y otro”, del 04/05/2010, en La Ley Online, cita: TR LALEY AR/JUR/11707/2010.

desarrollos, un experto podría contar con una asistencia para encarar estudios sistemáticos sobre organismos determinados (por ej., un juzgado o una fiscalía en particular), sobre una materia determinada (por ej., procesos vinculados a violencia doméstica), o sobre un tipo de resolución determinada (por ej., sentencias definitivas).

Tabla 1. Estereotipos según [10].

Tipos	Ejemplos	Refleja en algunas prácticas
De sexo Centrados en atributos y diferencias físicas y psicológicas existentes entre hombres y mujeres	Los hombres son más fuertes. Son impulsivos, necesitan usar fuerza física. Son racionales, saben negociar porque no son emocionales como las mujeres. Son firmes. Las mujeres son irracionales, sensibles, emocionales.	Prohibición de hecho o de derecho a trabajar que requiera fuerza física, horarios nocturnos, manejo de armas de fuego. Las mujeres no saben negociar.
Sexual Características sexuales que son o deberían ser poseídas por hombres y mujeres y la interacción sexual entre ambos	La sexualidad está vinculada a la procreación	Prohibición de matrimonio entre personas del mismo sexo.
Roles de género Comportamientos atribuidos y esperados de hombres y mujeres a partir de construcciones sobre su físico	En las familias, los hombres son los proveedores primarios y las mujeres quienes cuidan a los hijos/as	Norma escrita que distribuye derechos y obligaciones. Cuando se las supera, las mantienen las prácticas
Compuesto Aquel que interactúa con otros estereotipos de género atribuyendo roles o características a grupos diversos de mujeres	Mujeres solteras, las mujeres lesbianas no son buenas madres.	Negar la posibilidad de adopción. Negar la guarda o tenencia de hijos/as

Como metodología de trabajo, se propone un proceso de descubrimiento de conocimiento a partir de datos, KDD (del inglés *Knowledge Discovery in Databases*) [8, 12]. En particular, este proceso de KDD busca determinar patrones que

identifican textos jurídicos con ciertos patrones o pistas comúnmente asociadas a sesgos en razón de género, por ejemplo, en base a operacionalizaciones de los conceptos de la Tabla 1.

Los pasos del enfoque son los siguientes:

1. Identificar el objetivo del proceso de descubrimiento desde la perspectiva del consumidor del conocimiento y también de los expertos que realizan el análisis, en este caso, expertos en perspectiva de género y derechos humanos.
2. Entender el dominio de aplicación, los datos disponibles y el conocimiento que se pretende extraer a partir de ellos. Los datos iniciales para el presente proyecto serán obtenidos de estudios existentes y de sitios de instituciones como Centro de Información Judicial (CIJ), Sistema Argentino de Información Jurídica (SAIJ), entre otros.
3. Seleccionar un conjunto de datos de muestra sobre los cuales realizar el proceso de descubrimiento de conocimiento
4. Pre-procesar los datos utilizando técnicas de PLN, hasta tener una representación de los documentos con la cual las técnicas de IA (por ej., aprendizaje supervisado, o no supervisado) puedan trabajar.
5. Decidir acerca de los algoritmos de aprendizaje a ser aplicados para el descubrimiento de patrones de acuerdo a los objetivos identificados inicialmente en el proceso.
6. Aplicar los algoritmos en busca de patrones de interés en los datos.
7. Interpretar el conocimiento extraído a partir de los patrones identificados y realizar su validación con nuevos datos y con el juicio de expertos.

Éste es un proceso iterativo ya que usualmente es necesario volver sobre los pasos mencionados, aprovechando retroalimentaciones, hasta que el conocimiento extraído pueda ser efectivamente utilizado.

Las técnicas de PLN pueden ser organizadas en diferentes niveles de procesamiento. Cada uno de estos niveles representa un tipo de análisis que se puede efectuar al texto de entrada para extraer información específica. Estos niveles son: fonológico, morfológico, léxico, sintáctico, semántico, del discurso y pragmático. El esquema de PLN planificado para el problema se muestra en la Figura 1.

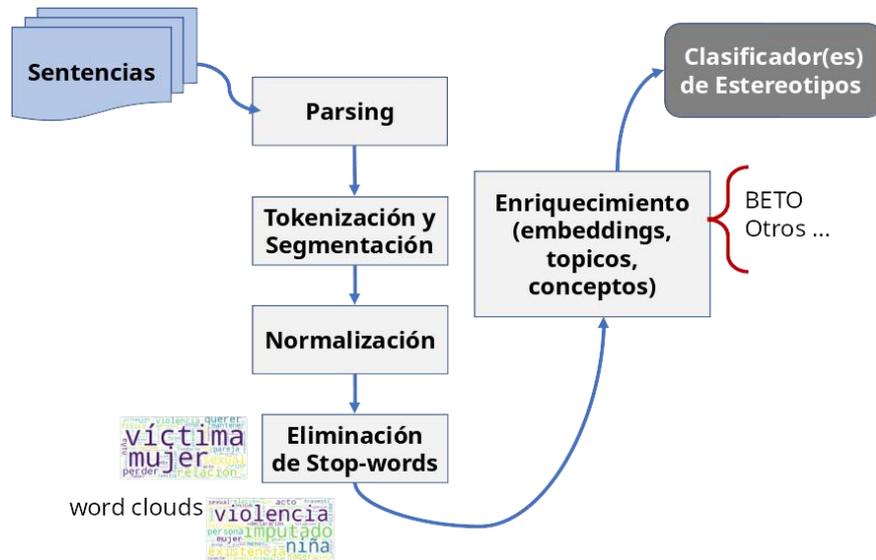


Figura 1. Etapas de procesamiento del enfoque

5 Resultados preliminares

Hasta el momento se ha avanzado en ciertas etapas del enfoque. Un activo importante para el trabajo es la disponibilidad de datos (sentencias) donde se observen sesgos en razón de género. Un aspecto relacionado es el contar con un equipo interdisciplinario, compuesto tanto por expertos en Derecho y género como en técnicas computacionales y PLN. En este sentido, se estableció un equipo formado por investigadores de la Facultad de Cs. Exactas y de la Facultad de Derecho de la Universidad UNCCPBA.

Obtención del conjunto de datos: Inicialmente, se avanzó en el análisis de documentos judiciales (sentencias) para identificar en ellos la presencia de los estereotipos planteados. Para ello, en una primera etapa fue necesaria la obtención de los documentos oficiales a texto completo, para lo cual se utilizaron las bases de datos provenientes del Centro de Información Judicial (CIJ), del Sistema Argentino de Información Jurídica (SAIJ), del Poder Judicial de la Provincia de Buenos Aires (JUBA), del Colegio de Abogados del Departamento Judicial de Azul y del Observatorio de Jurisprudencia del Equipo Latinoamericano de Justicia y Género (ELA). Asimismo, las búsquedas se complementaron con acceso a otras bases provistas por editoriales comerciales (Información Legal de Thomson Reuters; MicroJuris Argentina; elDial.com).

La búsqueda de documentos se circunscribió a sentencias emitidas en los últimos 20 años. Para hallarlas, se utilizaron palabras clave referidas a tópicos jurídicos tradicionalmente relacionados con cuestiones de género como “femicidio”, “abuso sexual”, “violencia de género”, “violencia doméstica”, “legítima defensa”, “responsabilidad parental”, “tenencia de hijos”, “cuidados”, entre otras. A partir de la recopilación de las sentencias, se procedió a su lectura para determinar si el pronunciamiento poseía o no estereotipos de género. Sobre el grupo de sentencias

sobre las que se consideraban “con estereotipos de género”, se efectuó un segundo análisis más pormenorizado con el objetivo de determinar qué tipos de estereotipos contenían e identificar puntualmente qué expresiones o frases del pronunciamiento permitían advertir el sesgo o estereotipo.

En la recopilación de estas resoluciones judiciales existieron ciertos obstáculos. Una primera dificultad en el acceso fue que, en razón de la protección de la intimidad de las personas involucradas en los procesos judiciales, en sistemas muchas bases de datos abiertas (como la de la Suprema Corte de Justicia de la Prov. de Bs. As.) las causas que versan sobre materia penal y de familia no son publicadas masivamente o su acceso requiere de una autorización especial. Asimismo, se encontraron varias sentencias de segunda instancia que advirtieron la presencia de sesgos en sentencias de primera instancia. Sin embargo, no fue posible acceder en todos los casos a estas últimas, dado que, o no son publicadas, o no son fácilmente hallables en los buscadores ya que la carátula (identificación) del caso suele variar entre la primera y la segunda instancia. No obstante, se pudo finalmente construir un conjunto de datos inicial de 33 sentencias, en las cuales se identificaron los estereotipos mencionados, marcando (manualmente) los textos/frases donde se evidencian estos sesgos. Esta información constituye el principal insumo para la aplicación de técnicas de PLN.

A continuación se detallan las técnicas/fases de PLN aplicadas a los textos hasta el momento por el grupo de trabajo.

Parsing de las sentencias: Cada documento (sentencia) se encuentra almacenado en formato PDF, y su texto fue anotado por los expertos de derecho para marcar las secciones que indicaban sesgos y presencia de estereotipos. Por esta razón, se implementó un script para extraer automáticamente el texto de cada archivo, dividirlo en párrafos, y adicionalmente extraer del mismo oraciones que estuvieran anotadas junto con sus respectivos comentarios.

Tokenización y segmentación: Esta fase consiste en dividir el texto en entidades significativas llamadas tokens, tales como palabras u oraciones, dados los espacios en blanco presentes y las puntuaciones. En esta tarea se debe definir si se utilizarán los signos de puntuación como token, si se dará importancia o no a las mayúsculas y si se unificarán palabras similares en un mismo token. En nuestro caso particular, los signos de puntuación no serán utilizados y tampoco se va a distinguir entre mayúsculas y minúsculas: Inicialmente no se van a unificar palabras, aunque se está analizando la posibilidad de incorporar de un diccionario de sinónimos.

Normalización: Esta etapa refiere a una serie de tareas relacionadas destinadas a poner todo el texto en igualdad de condiciones: convirtiendo el texto en mayúsculas o minúsculas, eliminando la puntuación, convirtiendo los números a sus equivalentes de palabras etc. Este proceso en nuestro caso (y por lo general para cualquier dominio) está muy apegado al realizado de forma estándar.

Eliminación de stop-words: Las stop-words o “palabras vacías” son las palabras que no agregan mucho significado a una oración ya sea por su frecuencia o por su semántica. Las stop-words se refieren generalmente a artículos, preposiciones, adverbios, pronombres, e incluso algunos verbos que no representan nada relevante para el contenido. Por lo tanto, pueden ignorarse sin sacrificar el significado de la

oración. Existen listas de *stop-words* para cada idioma e incluso temáticas, pero a su vez se pueden agregar a estas listas aquellas palabras que se consideren que no aportan información en el dominio en particular de trabajo. En este trabajo se usó la lista para español.

Representación computacional de documentos: Cuando se desea procesar computacionalmente textos escritos de forma libre, estos deben primero convertirse a un formato que sea entendible por la computadora y apto para la realización de cálculos. Un formato estándar para esta tarea es conocido como *bags-of-words*, o “bolsa de palabras”, cuya representación computacional es a través de vectores. En este modelo los documentos, en nuestro caso, el texto de las sentencias, se mapean a un espacio de vectores de alta dimensionalidad donde cada documento consiste en una secuencia de términos. Los términos únicos en un conjunto de documentos determinan las dimensiones del espacio. En estos vectores, pueden aplicarse “pesos” que representan la relevancia de la palabra para el usuario o dentro del tema, generalmente representada por la frecuencia con que aparecen. En nuestro caso, se usó TF-IDF ya que permite minimizar el impacto de palabras que se repiten habitualmente en una gran proporción de los documentos, a la vez que incrementa la importancia de aquellas palabras que aparecen en una baja proporción de los documentos



Figura 2. Nubes de palabras para diferentes estereotipos

En el conjunto de sentencias analizadas se identificaron un total de 136 frases que denotan alguno de los estereotipos de la Tabla 1. En estas frases, 34.56% tienen el estereotipo Sexual, 33.82% el estereotipo Roles de género y 31.62% el Compuesto. Usando la representación de documentos de las frases pertenecientes a cada estereotipo, la Figura 2 muestra las nubes de palabras para tres estereotipos y el peso relativo de las palabras en cada uno de ellos. Esta forma simple de visualización permite observar algunas relaciones interesantes entre las palabras. Por ejemplo, en el

estereotipo “Compuesto” las palabras salientes son “mujer”, “deber” y “haber hecho”, asociando a la mujer cierta responsabilidad siendo que en varias de estas frases denotan el estereotipo de “mala madre”.

Análisis de tópicos: Dentro del PLN, un modelo de tópicos es un algoritmo de aprendizaje no supervisado del lenguaje que identifica tópicos o temáticas en las cuales las palabras que comparten un contexto similar aparecen juntas. Con el fin de descubrir los tópicos latentes en el conjunto de sentencias analizadas, se usó BERTopic [8], un enfoque de modelado de tópicos que extrae representaciones coherentes de tópicos. Este enfoque genera embeddings de documentos con modelos de lenguajes basados en transformaciones, agrupa estos embeddings y luego genera representaciones con TF-IDF basado en clases. Un embedding es un modelo en donde las palabras o frases del lenguaje natural son representadas como vectores de números reales. Los embeddings apuntan a cuantificar y categorizar las semejanzas semánticas entre elementos lingüísticos basándose en sus propiedades distribucionales en muestras grandes de datos. Este análisis sería de utilidad para descubrir tópicos latentes que no hayan sido previamente identificados, incluso dentro de los grandes estereotipos, como así también observar la cercanía y superposición de tópicos. Por otro lado, podrían utilizarse para la detección de tópicos en nuevos documentos.

En la Figura 3 se muestra un ejemplo de los tópicos resultantes al procesar las sentencias analizadas, habiéndose agrupado automáticamente estas sentencias en 4 grupos. Para cada tópico se detallan las palabras más frecuentes en esos documentos.

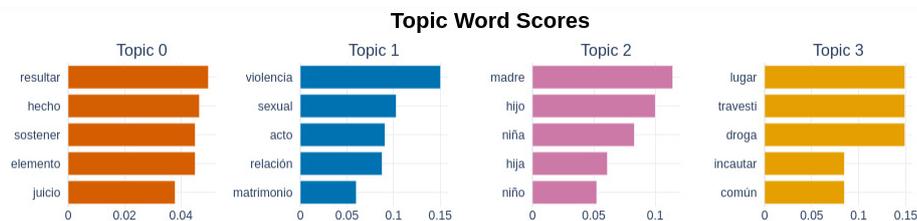


Figura 3. Análisis de tópicos sobre documentos judiciales

La Figura 4 muestra una representación espacial de párrafos de sentencias judiciales en los que se han detectado estereotipos versus aquellos párrafos en los que no hay estereotipos. Cada texto fue transformado en un vector en función de sus tokens más representativos, y luego enriquecidos con información de tópicos latentes. El gráfico superior permite observar una clara separación de los estereotipos (independientemente de su tipo). El gráfico inferior amplifica la región de los estereotipos y muestra cómo se distribuyen internamente los distintos tipos. Estos resultados evidencian las posibilidades de aprovechar distintas representaciones de vectores y de aplicar técnicas de clasificación de texto para la detección de estereotipos.

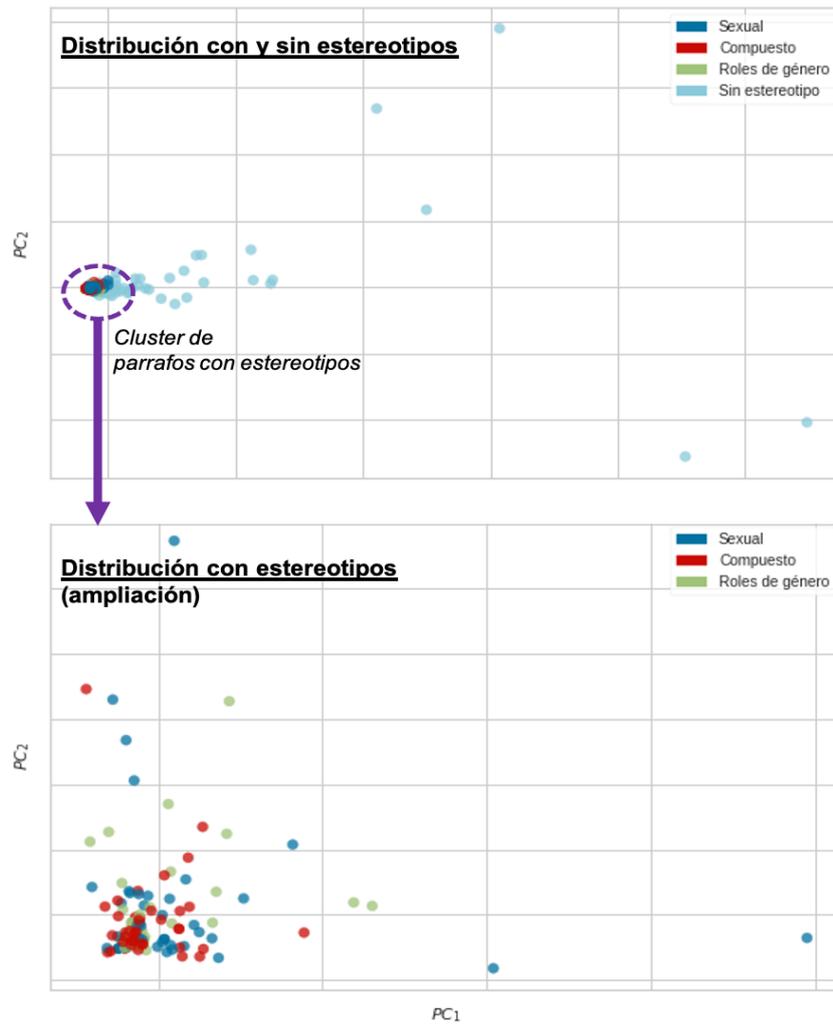


Figura 4. Distribución de párrafos con y sin estereotipos marcados.

6 Conclusiones y trabajos futuros

En este trabajo se detalla un enfoque para detectar sesgos en razón del género en sentencias judiciales. Se presentaron algunos resultados preliminares del proyecto, donde se partió de un conjunto de sentencias donde se habían identificado frases que corresponden a estereotipos judiciales, y se aplicaron técnicas de PLN para analizar estos textos. Se trabajó con diferentes alternativas, realizando un análisis de tópicos y

describiendo los diferentes estereotipos judiciales según las palabras más representativas, según el conjunto de datos.

En las etapas que restan del proyecto se espera aumentar el conjunto de datos, de manera de contar con mayor cantidad de sentencias con estereotipos identificados. Luego, se analizarán técnicas de Machine Learning (ML) para poder construir un modelo que permita catalogar sentencias judiciales según el estereotipo que presente. Para realizar reconocimiento de entidades en los textos, se trabajará con diferentes ontologías, tanto de propósito general como DBpedia o EuroWordNet y de propósito específico como LIKIF Core Legal Ontology [10] o LegalRuleML ontology [2]. Por otro lado, la aplicación de PLN en idioma español para el dominio legal ha cobrado interés recientemente, con desarrollos de modelos de lenguaje como BETO y RoBERTa⁸, que ofrecen perspectivas interesantes para este proyecto. Finalmente, se trabajará con técnicas de clasificación de manera de, teniendo en cuenta los resultados de la representación de cada estereotipo, asignar a nuevos documentos el o los estereotipos a los que pertenezca. Se espera que la herramienta obtenida constituya una ayuda para el especialista, dada la importancia de contar siempre con la supervisión humana para confirmar las conclusiones definitivas sobre la existencia o no de sesgo en las sentencias.

Como principales resultados del proyecto, se espera obtener: i) un estudio empírico que permita determinar el grado de coincidencia de los resultados de un análisis de sentencias realizado con PLN contra un análisis similar realizado manualmente por expertos humanos, ii) una metodología de análisis semi-automatizado apoyada por una herramienta prototipo, iii) un reporte de hallazgos en relación a sesgos de género en sentencias judiciales, iv) un taller de sensibilización sobre el tema de sesgo en razón de género para distintos actores e instituciones judiciales, así como para la comunidad. En resumen, se espera que la detección de sesgos pueda llevar a mejorar la educación y formación de los/as jueces/zas, fiscales y otros operadores/as judiciales -a partir de su utilización en actividades de formación, coadyuvando a la tarea de “entrenar” a los/as funcionarios en el trabajo con perspectiva de género-, así como también proveer herramientas de evaluación a organismos estatales y no-estatales -por ej., su utilización permitiría facilitar el análisis de las sentencias emitidas por un tribunal en particular durante un período de tiempo, a fin de evaluar si la actividad se realiza o no con la utilización de sesgos-.

Agradecimientos

Este trabajo fue realizado en el marco del proyecto PEIDyT “Perspectiva de género en decisiones judiciales: Aplicando técnicas de Procesamiento de Lenguaje Natural (PNL) para detectar sesgos en razón del género” (Cod. 03-PEIDYT-J), financiado por SECAT-UNCPBA, en el marco del Programa de Fortalecimiento en Acciones

⁸<https://www.iic.uam.es/procesamiento-del-lenguaje-natural/primer-modelo-lenguaje-en-espanol-adaptado-sector-legal/>

Sustantivas de Ciencia y Tecnología en Universidades Nacionales de la Secretaría de Políticas Universitarias (SPU).

Referencias

1. Asenio, R. et. al. (2010). Discriminación de Género en las Decisiones Judiciales: justicia penal y violencia de género. Defensoría General de la Nación.
2. Athan, T., Governatori, G., Palmirani, M., Paschke, A., Wyner, A.: LegalRuleML: Design principles and foundations. In: Wolfgang Faber and Adrian Pashke (ed.) The 11th Reasoning Web Summer School. pp. 151–188. Springer, Berlin, Germany (jul 2015)
3. Bonina, N. (2020). Inteligencia artificial y derecho: ¿las máquinas van a reemplazar a los abogados?, Thomson Reuters La Ley, AR/DOC/3809/2020.
4. Cook, Rebecca; Cusack, Simone (2010). Estereotipos de género. Perspectivas legales transnacionales. Profamilia, Bogotá.
5. Comité para la Eliminación de la Discriminación contra la mujer (CEDAW) (2004). Recomendación general N° 25, sobre el párrafo 1 del artículo 4 de la Convención sobre la eliminación de todas las formas de discriminación contra la mujer, referente a medidas especiales de carácter temporal. Disponible en <https://www.refworld.org/es/publisher,CEDAW,GENERAL,,52d905144,0.html>
6. Corvalan, Juan G. (2019). Prometea. Inteligencia artificial para transformar organizaciones públicas. Astrea. U. del Rosario. DPI Cuántico. IMODEV.
7. Cusack, Simone. (2013) Gender Stereotyping as a Human Rights Violation. Research Report Submitted to the Office of the High Commissioner for Human Rights. [Online]. Available: <https://www.ohchr.org/Documents/Issues/Women/WRGS/2013-Gender-Stereotyping-as-HR-Violation.docx>.
8. Fayyad, U., Gregory Piatetsky-Shapiro, Padhraic Smyth - Knowledge Discovery and Data Mining: Towards a Unifying Framework - KDD-96 Proceedings - AAAI Press - 1996
9. Grootendorst, Maarten- BERTopic: Neural topic modeling with a class-based TF-IDF procedure - arXiv preprint arXiv:2203.05794 - 2022
10. “Guía para el Poder Judicial sobre estereotipos de género y estándares internacionales sobre derechos de las mujeres” elaborada por la Fiscalía General de la Nación, disponible en https://acnudh.org/load/2020/03/Poder-Judicial_PDF-2.pdf
11. Hoekstra, Rinke , Breuker, Joost, Di Bello, Marcello, Boer, Alexander - The LKIF Core Ontology of Basic Legal Concepts - Proceedings of the 2nd Workshop on Legal Ontologies and Artificial Intelligence Techniques, 2007
12. Hillol Kargupta, Anupama Joshi, Krishnamoorthy Sivakumar and Yelena Yesha - Data Mining Next Generation Challenges and Future Directions - AAAI Press - (2004)
13. Malacalza, Laurana (Dir) (2020). Diagnóstico regional sobre las políticas públicas y las violencias de género en el centro de la Provincia de Buenos Aires, Facultad de Derecho (UNICEN)-Editorial Azul, Azul.
14. Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos sobre la estereotipación de género (2017). Documento de antecedentes sobre el papel del Poder Judicial en el abordaje de los estereotipos nocivos de género en casos relativos a la salud y los derechos sexuales y reproductivos. Disponible en <https://www.ohchr.org/es/women/gender-stereotyping>.

15. Papalia, N. (2018). ¿Cómo juezas y jueces resuelven los casos de violencia doméstica? Un estudio sobre el fuero penal de la Ciudad Autónoma de Buenos Aires, Universidad de Palermo.