

PUBLICACIONES DE LA
ASOCIACIÓN ARGENTINA
DE HUMANIDADES DIGITALES



¿Aparece dignidad? Experiencia de lectura distante con Voyant Tools aplicada a sentencias

Does Dignity Appear? Distant Reading Experience with Voyant Tools Applied to Sentences

Helga María LELL

helgalell@conicet.gov.ar

Conicet - Universidad Nacional de La Pampa

RESUMEN

Este trabajo presenta una experiencia de usuario de la herramienta Voyant Tools. El objetivo es mostrar su posible contribución a una investigación que analiza los usos del concepto de dignidad por la Corte Interamericana de Derechos Humanos (IDH) entre 1982-2018. Se espera ver qué puede aportar una lectura distante de los documentos seleccionados al estudio propuesto.

PALABRAS CLAVE

Voyant Tools, lectura distante, dignidad, Corte Interamericana de Derechos Humanos.

ABSTRACT

This review presents a user experience of Voyant Tools. The goal is show possible contribution to an investigation analyzing the uses of the concept of dignity by the Inter-American Court of Human Rights (IDH) between 1982-2018. It is hoped to see what a distant reading of the selected documents can contribute to the proposed study.

KEYWORDS

Voyant Tools, Distant Reading, Dignity, Inter-American Court of Human Rights.

PublicAAHD -
2020

Vol. 1

ISSN

2718- 7470



1. CORPUS¹

El corpus se conforma con las sentencias de casos contenciosos dictadas por la Corte IDH en el año 2018 y en las que se tratan cuestiones de fondo. Se excluyen, por lo tanto, aquellas que son interpretativas o que homologan soluciones amistosas. Dentro de cada sentencia solo se incluyen los argumentos que establece la Corte IDH sobre el fondo de la cuestión tratada. Es decir, ante un caso controvertido, los textos suelen incluir comentarios sobre las excepciones preliminares, los hechos y las pruebas; estas partes han sido borradas en los archivos a cargar para evitar que se incluyan textos ajenos al objeto. De esta manera, se incluye únicamente el discurso de la Corte IDH sobre cómo interpretar, a la luz del caso concreto, los derechos humanos del sistema interamericano. Todos los documentos se han incorporado en texto plano (codificados en UTF-8).

El corpus se compone de 19 documentos (con un total de 541.742 palabras y 18.200 palabras únicas). Cada documento ha sido nombrado de manera tal que se inicia con el año de la sentencia, con la letra c que se corresponde con el indicador que utiliza la Corte para identificar a los casos contenciosos y el número del caso conforme al registro de la Corte. Eventualmente, de acuerdo con los resultados que quieran obtenerse según la pregunta de investigación, podrían incluirse otros datos como la carátula del caso, qué Estado ha sido demandado y la fecha (mes y día).

2. EXPERIENCIA DE USUARIA EN EL ANÁLISIS DEL CORPUS²

El texto más largo tiene 45.372 palabras, y el más corto 10.298. Si bien en principio se podría pensar que la Corte varía mucho, aún en un mismo año y con la misma composición, respecto de cuánto espacio argumental dedica a la cuestión de fondo comparativamente entre sentencias, esta sería una conclusión apresurada puesto que la extensión depende principalmente de la cantidad de presuntas violaciones que hay bajo controversia. Para poder profundizar en esta hipótesis se debería recurrir a una lectura cercana.

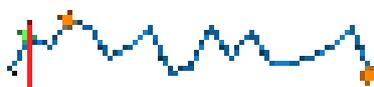


Figura 1. Extensión de los documentos. Elaboración propia.

El documento más corto es también el que tiene mayor densidad de palabras (0.201). El menos denso (0.113) es el más largo. Es curioso que los extremos coincidan puesto que esta asimetría no se da en el resto de los documentos.

¹ Sobre las distintas acepciones que hemos sistematizado como marco teórico, ver Lafferriere y Lell (2020). El proyecto de investigación es PICTO-UCA 2017-0032 El concepto de Dignidad Humana según la Corte Interamericana de Derechos Humanos. Análisis de los casos contenciosos y de las opiniones consultivas. Cofinanciado por ANPCyT y UCA, Argentina.

² El análisis realizado en Voyant Tools para esta reseña se encuentra accesible desde <https://voyant-tools.org/?panels=corpus%2Creader%2Ctrends%2Csummary%2Ccontexts&corpus=5cbe53b62347b2c3937896c9179944e8>.



Figura 2. Densidad del vocabulario. Elaboración propia.

El promedio de palabras por oración oscila entre 22.9 a 49.7. Estos promedios deben considerarse aproximados y debe tenerse en cuenta un posible margen de error originado en que Voyant Tools considera el punto como el fin de una oración y en las sentencias existen diversas abreviaturas y también puntos que separan incisos de artículos de la norma que se invoca (por ej. 1.2.).



Figura 3. Promedio de palabras por oración. Elaboración propia.

A fin de observar la frecuencia bruta de palabras, se ha elaborado un listado de palabras vacías a ser filtradas. El cirrus resultante con las 55 palabras más frecuentes es el siguiente:



Figura 4. Cirrus 1. Elaboración propia.

En cuanto a las palabras más frecuentes, luego de filtrar las vacías, estas son: derecho (2015); derechos (1829); víctimas (1486); tribunal (1276); violación (1113). Esto resulta lógico con la labor de la Corte, esto es, analizar los derechos de las víctimas tras una presunta violación cometida por los Estados. Claramente ello se refleja en los términos. Es interesante que una de las palabras más frecuentes sea *tribunal*, algo que cuando se analiza en el contexto de uso, muestra que la Corte la utiliza para referirse a sí misma. Por lo tanto, podemos pensar en una fuerte autorreferencialidad y una aparición en primer plano del organismo respecto de lo que resuelve y analiza (algo que no suele ser común en organismos judiciales que intentan mantener un discurso impersonal³).

Cuando se eliminan las palabras frecuentes que tienen que ver con la competencia y labor de la Corte, las palabras más frecuentes son: personas (926); familiares (907); investigación (850); medidas (793); proceso (757).

³ Esta característica es resaltada por Duarte i Montserrat y Martínez (1995).

2018_c354: martínez (149), casación (134), meléndez (63), rojas (83), archbold (42).
 2018_c355: munárriz (149), escobar (155), walter (93), desaparición (105), forzada (79).
 2018_c356: coc (45), quip (32), cajbón (30), pop (27), morente (24).
 2018_c359: vih (205), antirretroviral (64), alvarez (60), msf (56), viral (55).
 2018_c360: antezana (140), cueto (105), aparicio (92), tenicela (88), tello (87).
 2018_c362: linda (206), loaiza (194), soto (149), lópez (161), sexual (111).
 2018_c363: isaza (104), uribe (72), paramilitares (51), nare (34), desaparición (97).
 2018_c364: uva (77), velandia (76), albeiro (58), villamizar (55), bárcenas (53).
 2018_c368: miraval (230), álvarez (216), solano (71), carrascal (60), noel (38).
 2018_c370: alvarado (115), desaparición (114), espinoza (36), nitza (33), chihuahua (26).
 2018_c371: folios (83), mujeres (188), sexual (156), hernández (111), jaramillo (42).
 2018_c372: prescripción (38), barraza (12), lesa (21), humanidad (27), cortés (10).

3. LA DIGNIDAD EN EL CORPUS

Ahora bien, lo anterior permite tener una contextualización respecto de cuáles han sido los tópicos principales a los que se ha abocado la Corte. A continuación, me centraré en ver, a partir de esta lectura distante, si aparece y cómo aparece el término *dignidad* (y otros derivados o conexos como *digno/a*, *dignificante*, etc.) en este corpus. Para ello, utilizaré las búsquedas de la cadena de texto *dign**.

El conjunto de palabras relacionadas con la dignidad aparece 93 veces (frecuencia bruta) y la frecuencia normalizada relativa es de 172.

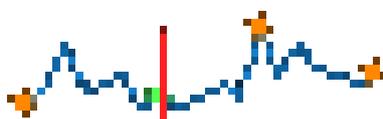


Figura 9. Tendencia de la frecuencia relativa del término dignidad (y relacionados) en el corpus. Elaboración propia.

A su vez, la asimetría estadística arroja un resultado de 1, por lo que está debajo de la media y su aparición es ligeramente irregular.

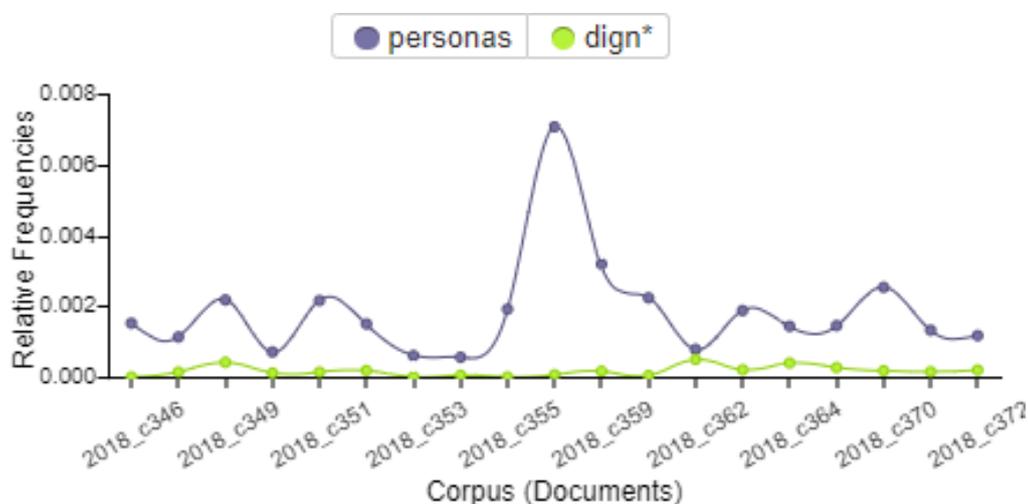


Figura 10. Tendencia de la frecuencia relativa del término dignidad (y relacionados) en el corpus en comparación con el término personas. Elaboración propia.

En comparación con el término que más aparece (*personas*), podemos ver que la dignidad tiene una aparición significativamente menor, algo que llama la atención si se considera el peso argumental que tiene la dignidad en los sistemas de protección de derechos humanos como una nota de las personas humanas⁴. Un ejemplo de ello se puede encontrar en cuánto y cómo aparece este término en los tratados internacionales, entre ellos, en la Convención Americana de Derechos Humanos.

Ahora bien, si analizamos el peso de cada documento dentro del corpus y la frecuencia con la que aparece la dignidad en ellos, podemos encontrar como sorpresa que en la tercera barra (ver figura 11) muestra un documento con poco peso, pero un alto grado de aparición del término *dignidad* (y relacionados). Al detenernos en este dato visual podemos ver que se trata del caso que se refiere al derecho a la salud. Esto permitiría pensar que es necesario indagar a partir de otras lecturas distantes comparativas con otros años si el tema ha estado presente o bien si este es un fallo *hito* y que acentúa la relación salud-dignidad por no haberse explicitado antes. También una lectura cercana podría permitir extraer algunas conclusiones sobre la relación entre estos términos.



Figura 11. Peso de los documentos en el corpus y frecuencia del término *dignidad* y relacionados. Elaboración propia.

Del análisis en contexto surge que en 16 de los 19 documentos aparece el término *dignidad* o relacionados (o sea, que la Corte invoca la dignidad en la mayoría de los casos). Se puede advertir que principalmente la dignidad: 1) califica a un derecho (usualmente, la vida), o 2) aparece como un sustantivo en el marco de una enumeración (por ejemplo libertad y dignidad o dignidad y autonomía), o 3) se vincula como una nota de los seres humanos.

4. ANÁLISIS DE LA HERRAMIENTA

A partir de esta experiencia como usuaria cabe reflexionar sobre qué puede aportar esta herramienta a una investigación como la descrita al iniciar el trabajo.

Respecto de las ventajas y aportes posibles:

- Voyant Tools permite una visualización muy rápida de un corpus grande, no solo por año, sino que también permitiría a través de exportar tablas, comparar los distintos corpus. En este caso, el corpus solo contó con 19 documentos, pero podría ser aún mucho mayor y la ventaja se mantendría igualmente (lo cual aumenta su beneficio al optimizar el trabajo).
- El elemento de contexto permite ver los términos que están próximos en los documentos. Esto es sumamente útil para reconocer cómo y cuándo está usado el término.

⁴ La relación entre *persona* y *dignidad* en los documentos normativos que componen el sistema americano de Derechos Humanos se puede encontrar en Lell (2018).

Podría ser útil también para descartar algunas menciones como, por ejemplo, cuando se trate de una cita textual y dependiendo de qué cita se trate (cabe recordar que se trata de una investigación sobre el discurso de la Corte).

- En el segmento de *Términos*, al seleccionar uno y expandir la información, se obtienen otros datos que pueden ser relevantes. Me parece significativo que traiga a colación las frases y que puedan ordenarse. Esto permitiría muy fácilmente detectar fórmulas recurrentes que utiliza el tribunal.
- La herramienta sistematiza formas de comparar la frecuencia bruta y normalizada de las palabras en un corpus para ver el peso de ellas.
- La herramienta es sumamente valiosa para hacer posible un trabajo que manualmente sería muy dificultoso, que acarrearía un gran tiempo y que demandaría otros conocimientos (por ejemplo, saber cómo elaborar gráficos). La visualización de los gráficos es atractiva y permite realizar múltiples comparaciones.

Respecto de las desventajas o limitaciones, cabe decir:

- El armado del corpus debe ser realizado con mucho cuidado, no solo en cuanto al formato sino también en relación con lo que no debe estar presente porque generaría distorsiones (por ejemplo, en el caso que presenté, haber dejado el análisis sobre excepciones preliminares hubiera provocado una distorsión), y por otro lado se debe tener mucho cuidado en la selección de las palabras vacías. No todas las palabras que se incluyen en el listado son realmente vacías (como podrían ser un artículo o un conector), pero podrían serlo en relación con la pregunta de investigación. Entonces, si bien Voyant Tools acelera mucho el trabajo, la depuración del texto y lo que le pedimos a la herramienta debe ser elaborado con mucho cuidado y revisión.
- Algunas instrucciones o explicaciones en los botones de ayuda no son tan claras para un usuario lego.
- Los conteos de palabras podrían traer algunos resultados con imprecisiones. Se mencionó el caso de las oraciones y los puntos, pero otro caso relevante para la investigación que menciono pueden ser las citas a las normas jurídicas. Estas no son parte del discurso de la Corte y, si apareciera dignidad en esas citas, Voyant Tools las contaría.
- Hay algunos conceptos que se componen de más de una palabra (por ejemplo: dignidad humana, derechos humanos, vida digna, integridad sexual, tratos crueles, etc.) que, en el cirrus, aparecen separadas. Es decir, es una herramienta preparada para contabilizar palabras y no conceptos. Esto puede generar algunas complicaciones cuando se necesita hacer distinciones en la lectura distante (por ejemplo, no es lo mismo que alguien esté con vida que alguien tenga derecho a una vida digna).

5. CONCLUSIONES

En respuesta a la pregunta del título, queda claro que la palabra *dignidad* (y otras relacionadas) aparece en el corpus. Mediante Voyant Tools podemos tener una foto o una imagen visual respecto de cómo está presente en el corpus y en qué contexto. Incluso es posible tener una idea, aunque sea un poco superficial, sobre qué sentido tiene al hacer el análisis en contexto. No obstante, a los efectos de la investigación marco es necesario hacer una lectura cercana, pero puede ser una tarea más sencilla a raíz de la contextualización que proporciona esta herramienta. Con solo algunos clics se tiene una imagen de un amplio corpus y se pueden pensar algunas regularidades (pasar de los datos a la información).

REFERENCIAS BIBLIOGRÁFICAS

- Duarte i Montserrat, C. y Martínez, A. (1995). *El lenguaje jurídico*. A-Z Editora.
- Lafferriere, J. N. y Lell, H. (2020). Hacia una sistematización de los usos semánticos del concepto de dignidad humana en la protección internacional de derechos humanos: una revisión doctrinaria. *Cuestiones Constitucionales*, 1(43), 129-167. <http://dx.doi.org/10.22201/ijp.24484881e20204315181>
- Lell, H. M. (2018). El concepto jurídico de persona como fundamento de los Derechos Humanos. Controversias semánticas en el sistema americano. En D. Herrera, N. Lafferriere, G. Maino y D. Ranieri (Coords.), *Estado de Derecho y Derechos Humanos* (pp. 67-82). EdUCA, Konrad Adenauer Stiftung.