

PUBLICACIONES DE LA  
ASOCIACIÓN ARGENTINA  
DE HUMANIDADES DIGITALES

## **AntConc y visualización de datos: relevancia de un análisis lingüístico en procesos de recolección y comparación**

### ***AntConc and Data Visualization: Relevance of a Linguistic Analysis in Collection and Comparison Processes***

Natalí CARDOZA ROJAS

[ncardoza@pucp.edu.pe](mailto:ncardoza@pucp.edu.pe)

Pontificia Universidad Católica del Perú

<https://orcid.org/0000-0002-7889-4812>

#### **RESUMEN**

El presente trabajo pretende mostrar algunos aspectos relevantes del uso del programa AntConc en las Humanidades Digitales. Para ello, se analizará la recolección y comparación de datos de *La Argentina Manuscrita* por medio de sus herramientas, tales como *Concordance* y *Word List*. De esta manera, es posible llegar a un entendimiento óptimo de los resultados sobre la conexión de palabras con el texto, además de permitir plantear interrogantes en torno al contenido histórico del corpus.

**PALABRAS CLAVE:** Humanidades Digitales, AntConc, *La Argentina Manuscrita*, análisis lingüístico, corpus.

#### **ABSTRACT**

This paper aims to show some relevant aspects of the use of the AntConc program in Digital Humanities. To do this, the collection and comparison of data from *La Argentina Manuscrita* will be analyzed through its tools, such as *Concordance* and *Word List*. In this way, it is possible to reach an optimal understanding of the results on the connection of words with the text, in addition to allowing questions to be raised about the historical content of the corpus.

**KEYWORDS:** Digital Humanities, AntConc, *La Argentina Manuscrita*, linguistic analysis, corpus



Publicaciones de la Asociación Argentina de  
Humanidades Digitales PublicAAHD - 2022 Vol. 3.

<https://doi.org/10.24215/27187470e040>

## 1. INTRODUCCIÓN

Las Humanidades Digitales ofrecen una serie de herramientas para una lectura distante a los corpus, textos o documentos que buscamos analizar de manera global. Con ello se pretende afirmar que no solo se poseen capacidades de comprensión de lectura sino que también enriquecen aquellas lecturas que capturan la frecuencia o patrones de palabras, o la conexión entre estas, para observar aspectos y visualizar narrativas que una lectura cercana por sí sola no nos puede proporcionar:

Si, por ejemplo, dispones de una colección de documentos, es posible que desearas encontrar patrones de uso gramatical o frases de aparición recurrente en la misma. También puede ser que quisieras hallar frases cuya probabilidad de aparición fuese más alta o más baja en la obra de un autor, o bien en un tipo determinado de textos; clases particulares de estructuras gramaticales; o muchos ejemplos de un concepto particular en una gran cantidad de documentos que se encuentran enmarcados en cierto contexto. En este sentido, el análisis de corpus resulta muy útil para demostrar hipótesis sobre textos, o para triangular resultados obtenidos a través de otras metodologías de análisis textual basadas en herramientas digitales (Froehlich, 2018).

El objetivo de este trabajo es, pues, ofrecer una aproximación de análisis de un texto por medio de la frecuencia o patrones de sus palabras, frases o conceptos. Para ello, se utilizará una de las herramientas que aún se mantiene en permanente actualización: AntConc<sup>1</sup>. Esta herramienta informática, creada por Laurence Anthony, proporciona un análisis lingüístico a los textos para evidenciar relaciones o patrones desde la visualización de datos.

## 2. METODOLOGÍA

Para dar un ejemplo de uso de la herramienta se analizará la edición digital de *La Argentina Manuscrita* de Rui Díaz de Guzmán, en formato texto plano o txt preparado por HD Lab (CAICYT, CONICET) (del Rio

---

<sup>1</sup> Accesible en <http://www.laurenceanthony.net/software/antconc/>.

Laurence Anthony, el creador de AntConc, sigue actualizándolo y compartiendo tutoriales de sus versiones en su cuenta de Youtube: <https://www.youtube.com/user/AntlabJPN>.

Riande et al., 2020)<sup>2</sup>. Consideremos que Díaz de Guzmán fue un conquistador mestizo del siglo XVII y sirvió como militar en Río de Plata; además, fue el primer escritor que ofreció una exploración y colonización de este mismo territorio, tal como lo describe, en el Libro I, sobre las provincias del Río de Plata en 1512<sup>3</sup>. Ahora bien, lo que se pretende analizar es la información contenida en el primer libro. Para ello, toda su data se visualizará por medio de AntConc, versión 3.2.4w (2011) para Windows.

Cabe resaltar que AntConc solo analiza textos planos (txt.); por ende, nuestro libro I del Manuscrito se descargará como *Rui-Diaz-La-Argentina-manuscrita-LIBRO-I.txt* para que, finalmente, el programa pueda leerlo como un archivo txt.:



Figura 1. Componente *Corpus Files* en AntConc. Fuente: Elaboración propia.

Una vez que se sube el archivo en *File - Open File(s)*, nuestro texto aparecerá en el *Corpus Files* ubicado en la parte izquierda del programa. A continuación, procederemos a la construcción e interpretación de los datos a través de las herramientas que nos ofrece AntConc.

Empecemos, entonces, por la herramienta *Concordance*. Esta “muestra lo que se conoce como *keyword in context view* (vista de palabras clave en contexto [KWIC, por sus iniciales en inglés]), cuyos resultados se obtienen mediante la barra de búsqueda” (Froehlich, 2018,

<sup>2</sup> Accesible desde:

<http://hdlab.space/La-Argentina-Manuscrita/argentina-manuscrita-acerca-de/>.

<sup>3</sup> Véase la edición digital con mayor detalle en:

[http://hdlab.space/La-Argentina-Manuscrita/argentina\\_manuscrita\\_pages/pg\\_0005](http://hdlab.space/La-Argentina-Manuscrita/argentina_manuscrita_pages/pg_0005).

s/p). Con esta herramienta podremos considerar, entonces, la relación entre palabras clave y su frecuencia en un contexto del corpus. Así, procederemos a visualizar, en la siguiente imagen, los resultados que nos arrojan si en la búsqueda colocamos la palabra *indio*:

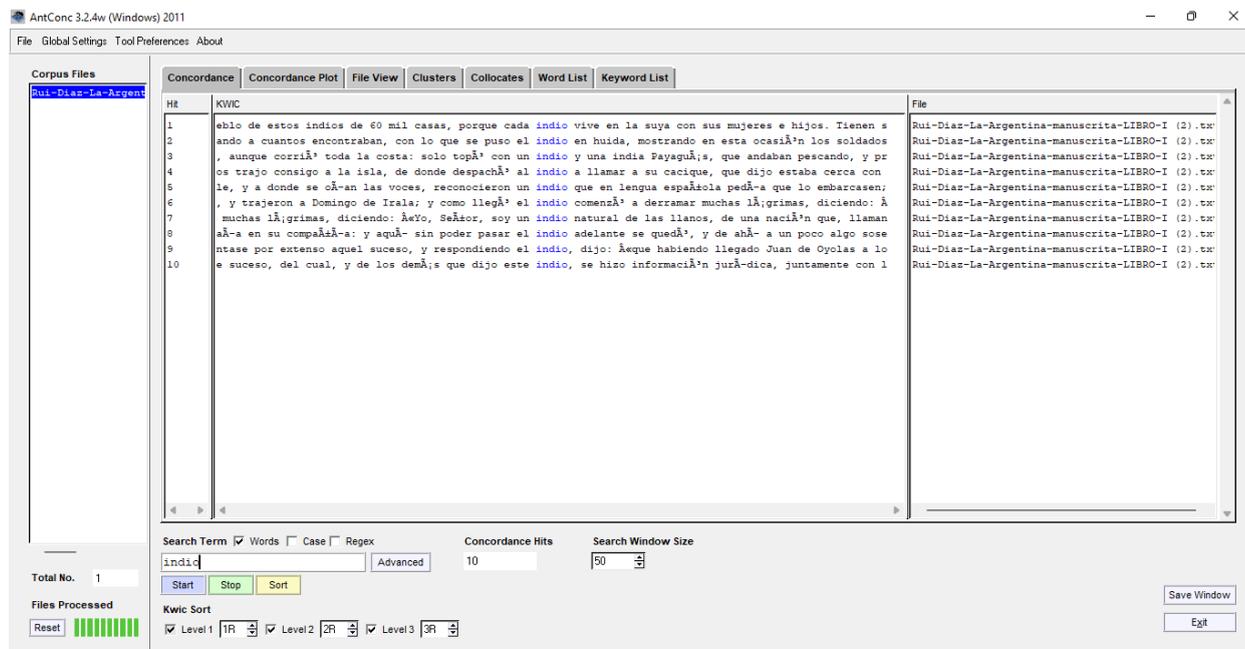


Figura 2. Datos que arroja la herramienta *Concordance*. Fuente: Elaboración propia.

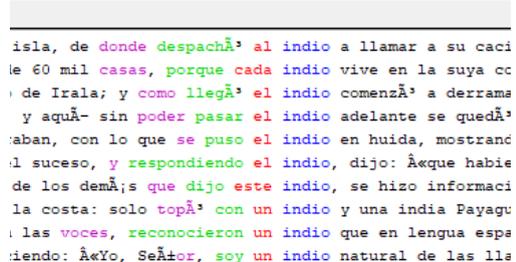
Como se puede observar en la imagen, la palabra *indio* aparece en diez ocasiones en todo el texto y está, además, relacionado a oraciones tales como las que se mostrará a continuación:

- (a) "cada indio vive en la suya (casa) con sus mujeres e hijos"
- (b) "con lo que se puso el indio en huida"
- (c) "con un indio y una india Payaguás, que andaban pescando"
- (d) "despachó al indio a llamar a su cacique"
- (e) "reconocieron un indio que en lengua española pedía que lo embarcasen"
- (f) "y como llegó el indio comenzó a derramar muchas lágrimas"
- (g) "soy un indio natural de los llanos"
- (h) "sin poder pasar el indio adelante se quedó"

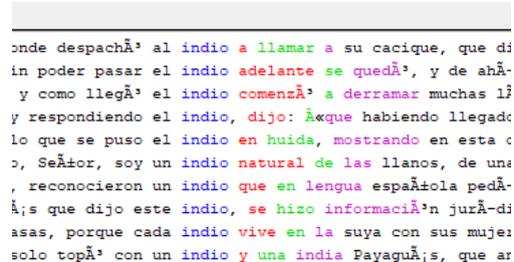
(i) “y respondiendo el indio, dijo...”

(j) “y de los demás que dijo este indio, se hizo información jurídica”

Lo anterior nos puede mostrar, en líneas generales, el contexto, datos y vivencias de los indios que vivieron en territorio de Río de Plata en el siglo XVII, según el autor de esta crónica. Esto quiere decir, en otras palabras, que, con (a), (c), (d) y (g), se puede inferir que un indio vive y crea vínculos familiares; se dedica a trabajos como la pesca; tiene relaciones de subordinación, como un cacique al mando; y se ubica en un determinado espacio geográfico, como es en los llanos. Por otro lado, con (b) y (f), se evidencia también que los indios viven experiencias de huida o sufrimiento por la condición de subordinación a la que se ven sometidos. Esto también lo podremos resaltar, si optamos por un análisis relacional entre las palabras, con el uso de *Kwic Sort* de distintos niveles, y que puede ir tanto por el lado izquierdo o derecho de las oraciones:



isla, de donde despachã al indio a llamar a su cacique, que di  
le 60 mil casas, porque cada indio vive en la suya co  
de Irala; y como llegã el indio comenzã a derrama  
y aquã- sin poder pasar el indio adelante se quedã  
aban, con lo que se puso el indio en huida, mostrand  
el suceso, y respondiendo el indio, dijo: Àque habie  
de los demás; que dijo este indio, se hizo informaci  
la costa: solo topã con un indio y una india Payagu  
las voces, reconocieron un indio que en lengua espa  
iendo: ÀYo, Señor, soy un indio natural de las lla



onde despachã al indio a llamar a su cacique, que di  
in poder pasar el indio adelante se quedã, y de ahã  
y como llegã el indio comenzã a derramar muchas lã  
y respondiendo el indio, dijo: Àque habiendo llegad  
lo que se puso el indio en huida, mostrando en esta c  
o, Señor, soy un indio natural de las llanos, de una  
, reconocieron un indio que en lengua espaãtola pedã  
ã;s que dijo este indio, se hizo informaciã'n jurã-di  
asas, porque cada indio vive en la suya con sus mujer  
solo topã con un indio y una india Payaguã;s, que ar

Figuras 3 y 4. El empleo de *Kwic Sort on left* y *on right*. Fuente: Elaboración propia.

Asimismo, cabe resaltar que en la búsqueda podremos conseguir información que considera tanto la palabra *indio* en singular y plural, como para ambos géneros. Esto será posible si se coloca la palabra *indi\**. Con esta modificación, nos mostrará toda la frecuencia de oraciones en las que aparecen *Indias*, *indios*, *indignación*, *india* e *indio*. Esto es así porque al tabular el operando \*, este nos permitirá visualizar todas las palabras compuestas por *indi...*, tal como se observa a continuación:

1 Excelencia, de donde el año de 1540 pasó a las Indias con el Adelantado Alvar Núñez Cabeza de  
 2 bres y guerras, de cuantas se han padecido en las Indias; no quedando de ellos más memoria  
 3 olán volvió a España del descubrimiento de las Indias. Este capitán Américo llegó  
 4 ueta, y algunas esmeraldas que hallaron entre los indios, de donde llevaron para Portugal  
 5 e Solís, vecino de la villa de Lebrija, para las Indias Occidentales: este era piloto  
 6 de Cristóbal Colón en el descubrimiento de las Indias; y continuando su derrota llegó  
 7 que se reducen en el principal, y la multitud de indios naturales de diversas naciones  
 8 las mayores que su Majestad tiene y posee en las Indias, porque además de habiéndose  
 9 de una y otra parte están poblados más de 20000 indios Guaranás, que los de aquella  
 10 a y corpulenta, y tienen guerra ordinaria con los indios Charrás del río de la Plata  
 11 corriendo la costa arriba, hay algunos pueblos de indios de esta misma nación; es toda  
 12 isto al unos portugueses que han estado entre los indios, y por lo que se ha descubierto  
 13 ados: hay en este asiento y comarca más de 10000 indios Guaranás, tratables y amigos  
 14 y otra al Nordeste; fue esta isla muy poblada de indios Guaranás, y en este tiempo es  
 15 uertos y ríos. El de la Cananea está poblado de indios Caribes del Brasil; tiene un  
 16 territorio hay muchas provincias y poblaciones de indios de diversas naciones, por medio  
 17 na de mucha pesquería; corren toda esta isla los indios Charrás de aquella costa, que  
 18 ste río a pasar por una población muy grande de indios Guaranás, que llaman Tapas, que  
 19 re doscientas leguas por gran suma de naciones de indios: los primeros y más altos son  
 20 cuya jurisdicción y comarca hay más de 200 mil indios Guaranás, poblados así por ríos  
 21 ue dicen Atiuajua, que contienen más de 100 mil Indios poblados de esta nación. Nace  
 22 San Pablo, en cuya ribera está poblada; no tiene indios ningunos, porque los que habitan  
 23 , entrando por la gobernación de Tucumán. Estos indios fueron repartidos con los demás  
 24 ra. De esta ciudad arriba hay algunas naciones de indios, y aunque tienen diferentes lenguas  
 25 den bien la lengua española: fueron más de 8000 indios antiguamente, y ahora han queda

Search Term  Words  Case  Regex      Concordance Hits      Search Window Size  
      Advanced      177      50

Figura 5. Búsqueda de *indi\**. Fuente: Elaboración propia.

Nótese, con esto último, que será distinto buscar en *Search Term* esta palabra en comparación a *?*. Por su parte, al realizar la búsqueda *indi?*, se excluyen las palabras como *indignación*, que pueden abarcar por ser compuesta por las primeras cuatro palabras que comienzan con *indi-*. Al tabular con el operador *?*, este nos mostrará tanto la palabra *indio* como *india*, ubicados en distintas partes del texto, tal como se evidencia en la siguiente imagen:

Estos indios de 60 mil casas, porque cada indio vive en la suya con sus mujeres e hijos. Tiene algunos de la casa, y en especial de un india, mujer que había sido muy estimada de Siripo, y repudiada por la española: la cual india movida de rabiosos celos, le dijo al Siripo cuando ellos se encontraban, con lo que se puso el indio en huida, mostrando en esta ocasión los soldados que corrían toda la costa: solo topó con un indio y una india Payaguás, que andaban pescando, y en la costa: solo topó con un indio y una india Payaguás, que andaban pescando, y preguntándole cómo se fue a la isla, de donde despachó al indio a llamar a su cacique, que dijo estaba cerca cuando se oyeron las voces, reconocieron un indio que en lengua española pedía que lo embarcasen para ir a Domingo de Irala; y como llegó el indio comenzó a derramar muchas lágrimas, diciendo muchas lágrimas, diciendo: «Yo, Señor, soy un indio natural de las llanuras, de una nación que, llamo a mi compañera: y aquí sin poder pasar el indio adelante se quedó, y de ahí a un poco algo se extendió aquel suceso, y respondiendo el indio, dijo: «que habiendo llegado Juan de Oyolá del cual, y de los demás que dijo este indio, se hizo información jurada, juntamente con el capitán Salazar, de proveer el remedio por vía de una india que tenía en su servicio el capitán Salazar,

Figura 6. Búsqueda de *indi*?. Fuente: Elaboración propia.

Hemos evaluado el caso de la palabra *indio* anteriormente. Valdría la pena mencionar que, en el caso del término *india*, obtendremos lo siguiente:

- (a) “y en especial de una india, mujer que había sido muy estimada de Siripo”
- (b) “y repudiada por la española: la cual india movida de rabiosos celos...”
- (c) “y una india Payaguás, que andaban pescando”
- (d) “proveer el remedio por vía de una india que tenía en su servicio el capitán”

Es interesante evidenciar en principio que, lo que se escribe en el caso de las mujeres indias y, además, la frecuencia de aparición que es menor en comparación al indio. Las características para el caso de estas, quienes son (a) estimadas, (b) movidas de rabiosos celos o que (d) sirven al capitán, no hace más que marcar las diferencias entre las experiencias que vivió cada indio e india, en el antiguo Río de Plata. Así, se puede evidenciar la ausencia de un reconocimiento o consideración para el caso de las indias.

Si bien en una lectura cercana podremos saber sobre las experiencias de las y los indios, un aspecto a destacar de este tipo de herramienta de análisis es que permite visualizar aquellos patrones que nos resultaría difícil evidenciar en una lectura corrida del texto. Así, una

lectura lejana nos permite plantearnos interrogantes tales como la invisibilización de la mujer en la historia y narrativas del pasado.

Ahora bien, obtener estos datos es posible también a través de la opción de *Save Output to Text File*. Esto con el objetivo de guardar el archivo txt:

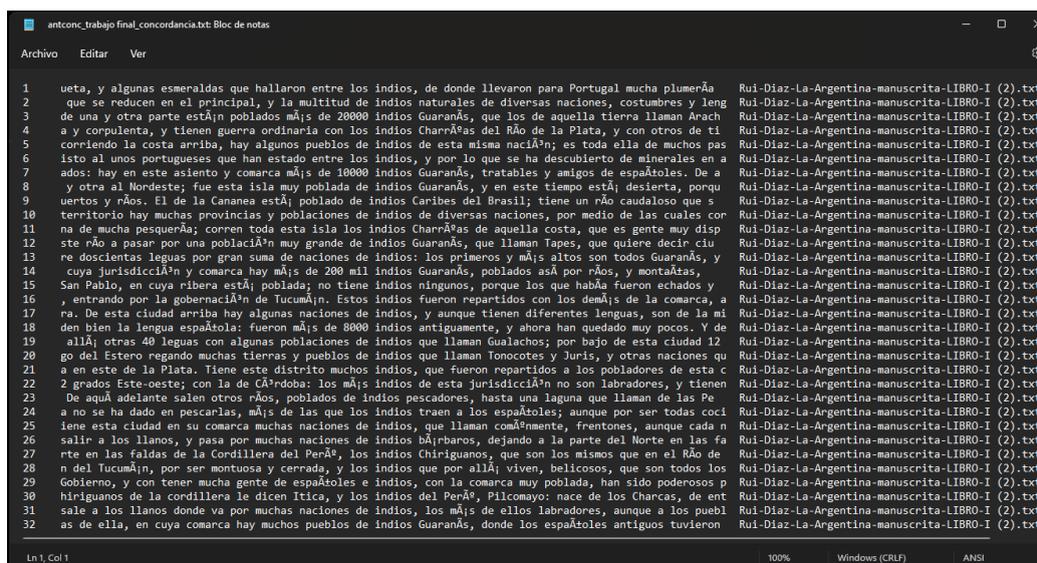


Figura 7. Descarga de los datos de *concordance* en formato txt. Fuente: Elaboración propia.

Para el siguiente caso, con la pestaña *File View*, cabe resaltar que existen dos formas de dirigirnos a ella. Cuando se da clic a la palabra *india*, como se muestra a continuación, el cursor cambia y nos dirige al archivo en el que se encuentra la concordancia de esa palabra en la oración.

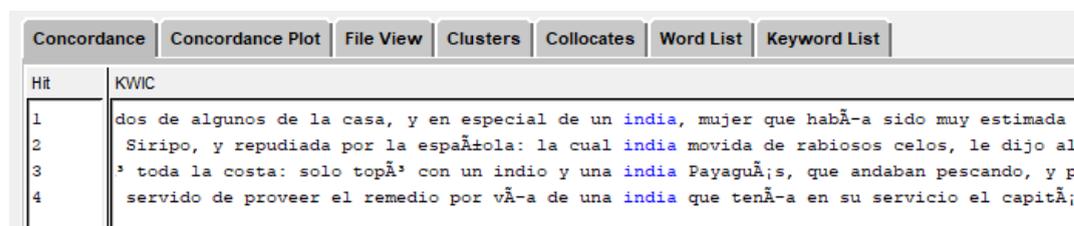


Figura 8. *File View* y *Concordance*. Fuente: Elaboraci3n propia.

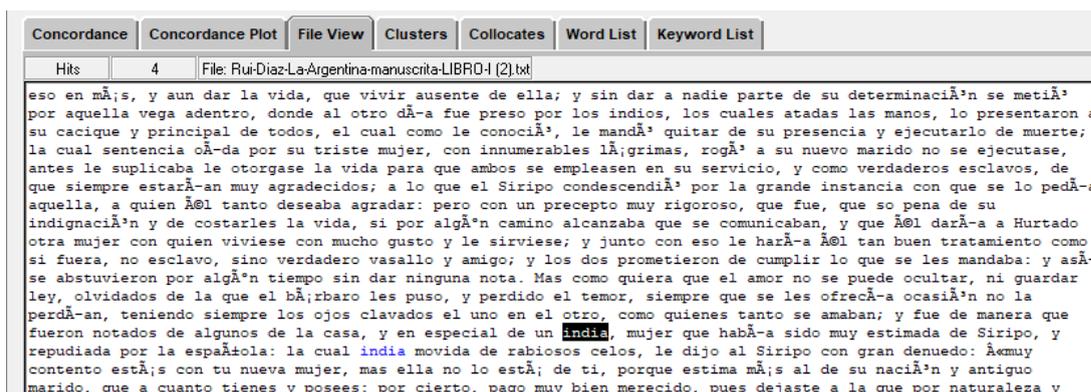


Figura 9. *File View*. Fuente: Elaboración propia.

Esta es una forma útil para contextualizarnos y entender el sentido de la palabra en la oración. Así, en este caso, podríamos afirmar que la india, estimada por Siripo, le advirtió sobre los amantes que se encontraban a escondidas de él.

La otra forma es ir directamente a la pestaña *File View* para mostrarnos el texto de la siguiente forma:

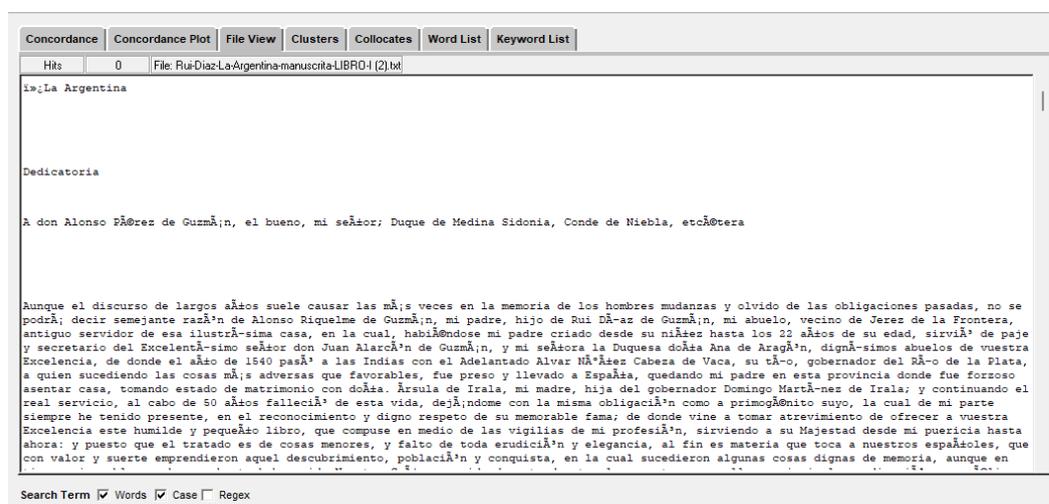


Figura 10. *File View*. Fuente: Elaboración propia.

Para el caso de la herramienta *Clusters*, consideremos analizar ahora la palabra *hij\**. Esta herramienta nos ofrece la frecuencia de la palabra que aparece en el texto. Mientras que *Concordance* nos ofrecía el contexto en el que aparecía la palabra *indi\** y su conexión con otras palabras, *Clusters* nos indica, de mayor a menor frecuencia, la cantidad

de veces en que la palabra *hij\** es utilizada. Así, tenemos en primer lugar a la palabra *hijo* acompañada de la preposición *de*. Esta palabra se repite 3 veces en todo el texto, mientras que *e hijas*, *e hijos*, *hijas y*, *hijos e* y *sus hijas* aparecen dos veces cada una en todo el texto:

Rank	Freq	Cluster
1	3	hijo de
2	2	e hijas
3	2	e hijos
4	2	hijas y
5	2	hijos e
6	2	sus hijas
7	1	casa, hijos
8	1	hija de
9	1	hija del
10	1	hijas para
11	1	hijas, a
12	1	hijos y
13	1	hijos. Los
14	1	hijos. Tienen
15	1	madre, hija
16	1	muchos hijos
17	1	o, hijo
18	1	padre, hijo
19	1	Salazar, hija
20	1	sus hijos
21	1	un hijo

Figura 11. Resultados de *Clusters*. Fuente: Elaboración propia.

Asimismo, *clusters* nos muestra el número total de *Cluster types*, que es 21. Esta cantidad es el número de tipos de palabras en las que está agrupada la palabra *hij\**. Por su parte, el número total de *Cluster tokens* es 28. Esto último quiere decir la suma de todas las frecuencias en las que ha aparecido relacionada la palabra *hij\** con preposiciones u otras palabras.

Ahora, si esto lo comparamos con la palabra *indi\**, obtendremos lo siguiente:

Concordance	Concordance Plot	File View	Clusters	Collocates	Word List	Keyword List
Total No. of Cluster Types: 115 Total No. of Cluster Tokens: 330						
Rank	Freq	Cluster				
1	59	los indios				
2	35	de indios				
3	26	indios de				
4	17	indios que				
5	11	algunos indios				
6	10	indios, y				
7	9	indios Guaraná				
8	6	indios Payaguá				
9	6	mil indios				
10	5	e indios				
11	5	indios, los				
12	5	indios, que				
13	5	muchos indios				
14	5	unos indios				
15	4	el indio				
16	4	indios con				
17	3	indignaciã				
18	3	indios amigos				
19	3	indios del				
20	3	indios llamados				
21	3	un indio				
22	2	estos indios				
23	2	indios Charrã				
24	2	indios comarcanos				

Figura 12. Búsqueda de *indios* en *Clusters*. Fuente: Elaboración propia.

Con la palabra *indi\**, podremos evidenciar que es una de las palabras más frecuentes en todo el texto, con un total de 330 veces en la que aparece en el texto y siendo, además, la palabra compuesta *los indios* la de mayor frecuencia (59 son las veces que se ha repetido); le sigue *de indios* (35), *indios de* (26), *indios que* (17), y así.

Aún más, si le damos clic a, por ejemplo, la palabra *de indios* o *indios de*, nos dirigirá nuevamente a la sección *Concordance* para observar de forma más detallada en qué parte del texto se encuentra:

Hit	KWIC
1	alojarse en ella. Allí- le vinieron cuatro canoas de indios que llaman Gujara
2	estando en ella le acometieron más de 300 canoas de indios que llaman Agases, que en aq
3	-a como a las dos de la tarde vinieron dos canoas de indios Payaguá; de parte de su caci
4	con sus picas y otras armas mataron gran cantidad de indios, de manera que fueron desbar
5	aban durmiendo, dieron sobre ellos mucha cantidad de indios, siendo repartidos en buen or
6	, como lo hicieron por muchas regiones y comarcas de indios de lenguas diferentes, y tam
7	en demanda de esta tierra, llevando consigo copia de indios amigos, y bajando en canoas y
8	e este río a pasar por una poblaciã muy grande de indios Guaraná-s, que llaman Tapes,
9	quella costa, en la cual murieron algunos a manos de indios de ella, y otros de hambre y
10	co, donde les salieron al encuentro gran multitud de indios Charcas; por lo cual dieron v
11	-os que se reducen en el principal, y la multitud de indios naturales de diversas nacione
12	, el cual hallã en aquella tierra mucha multitud de indios labradores en grandes pueblos
13	llera. De esta ciudad arriba hay algunas naciones de indios, y aunque tienen diferentes l
14	corre doscientas leguas por gran suma de naciones de indios: los primeros y más altos sc
15	dida en alto y bajo; hay en ambas muchas naciones de indios que todos son labradores: los
16	: tiene esta ciudad en su comarca muchas naciones de indios, que llaman comúnmente, frer
17	ene es, que por aquella parte hay muchas naciones de indios que poseen oro y plata, en es
18	a salir a los llanos, y pasa por muchas naciones de indios bárbaros, dejando a la parte
19	al sale a los llanos donde va por muchas naciones de indios, los más de ellos labradores
20	a a mano izquierda como vamos, hay otras naciones de indios que llaman Napabes y Payagua
21	hay allí; otras 40 leguas con algunas poblaciones de indios que llaman Gualachos; por baj
22	ãrmino de Tomina; y hallando algunas poblaciones de indios, vasallos del poderoso Inga,
23	cabo, corriendo su derrota por muchas poblaciones de indios que les agasajaron y dieron y
24	te territorio hay muchas provincias y poblaciones de indios de diversas naciones, por me
25	te, y otra al Nordeste; fue esta isla muy poblada de indios Guaraná-s, y en este tiempo e

Figura 13. *Concordance* y *Clusters on left*. Fuente: Elaboración propia.

río Paraná; los indios de aquel territorio, con la misma malicia  
 acián con algunos indios de aquella costa, para lo cual despachá  
 y calidad. Son los indios de aquella isla de buena voluntad y amigos  
 n mezclado con los indios de aquella tierra, la cual es muy dilatada  
 oncertaron algunos indios de aquella tierra de matarle, y así lo pu  
 de los Guaraná-s, indios de aquella tierra, y términos de las otr  
 ndo le mataron los indios de aquella tierra. Con esto Sebastián Gab  
 muchos pueblos de indios de diversas lenguas y naciones, con quien  
 s y poblaciones de indios de diversas naciones, por medio de las cui  
 algunos a manos de indios de ella, y otros de hambre y cansancio, y  
 Córdoba: los más indios de esta jurisdicción no son labradores, y  
 algunos pueblos de indios de esta misma nación; es toda ella de mu  
 , más de tres mil indios de guerra, con mucha flechera, dardos, y  
 se defender de los indios de guerra, que luego que sintieron a los  
 a la gente por los indios de la comarca llamados Chayos y Charrás,  
 ños de lo cual los indios de la provincia del Paraguay se juntaron  
 convocados muchos indios de la provincia, salieron por tercios y pe  
 habló con algunos indios de las islas, de quienes recibí alguna  
 ones y comarcas de indios de lenguas diferentes, y también en costu  
 ay pueblo de estos indios de 60 mil casas, porque cada indio vive en  
 scates hubo de los indios de quienes tomé lengua de lo que por allí  
 y Caaguazá, y los indios de río arriba Jeruquisaba y Carayazapera  
 unicación con los indios de su comarca, con quienes trabé amistad;

Figura 14. *Concordance y Clusters on right*. Fuente: Elaboración propia.

Al comparar, podremos observar que en el caso de *de indios* se relaciona con la cantidad o el grupo: *vinieron cuatro / dos canoas de indios, mataron gran cantidad de indios, regiones y comarcas de indios, murieron algunos a manos de indios, encuentro gran multitud de indios, hay muchas naciones de indios, hallando algunas poblaciones de indios, isla muy poblada de indios*. Así, la relación con *de indios* es directamente con la ubicación, el hallazgo de estos, sus muertes o matanzas, encuentros. En el caso de *indios de*, podremos ver que su frecuencia se relaciona con *indios de aquellas tierras / costa / isla / territorio, indios de diversas lenguas / naciones, indios de guerra, indios de 60 mil casas*.

Con lo anterior, podremos afirmar lo que anteriormente resaltamos; es decir, la relación entre la palabra y el contexto en que se encuentra nos evidencia el modo en el que el texto ordena y muestra las características y experiencias que viven los indios en un territorio, su relación con caciques, comarcas, sus experiencias de sufrimiento, guerras o matanzas; así también, observamos con qué se les asocia: tierras, lenguas, trabajo, cantidad poblacional. Con ello también cabe resaltar que el autor otorga mayor relevancia a las vivencias de los indios frente a las indias.

Ahora bien, lo que hemos analizado al dirigirnos a la pestaña *Concordance*, también podría desarrollarse en el mismo *Cluster*. Para ello, es necesario observar lo que ofrece en la parte inferior:

The screenshot shows the Clusters Tool interface with the following settings:

- Search Term:** indios de
- Advanced:**  Words,  Case,  Regex,  N-Grams
- Cluster Size:** Min. Size 4, Max. Size 5
- Min. Cluster Frequency:** 1
- Sort by:** Sort by Word
- Search Term Position:**  On Left,  On Right
- Invert Order:**

Figura 15. *Clusters Tool*. Fuente: Elaboración propia.

Esta herramienta permite ordenar por palabra o frecuencia en *Sort by*, además de buscar el término dependiendo de la posición *On Left* u *On Right*. Asimismo, podríamos añadir el mínimo y máximo del tamaño de palabras que queremos considerar para la búsqueda. Con esto último nos referimos a cuántas son las palabras que podríamos permitir agregar para visualizar su relación con estas:

Total No. of Cluster Types: 48		Total No. of Cluster Tokens: 52
Rank	Freq	Cluster
1	1	indios de 60 mil casas
2	1	indios de 60 mil casas, porque
3	1	indios de aquel territorio
4	1	indios de aquel territorio, con
5	1	indios de aquella costa
6	1	indios de aquella costa, para
7	1	indios de aquella isla
8	1	indios de aquella isla de
9	4	indios de aquella tierra
10	1	indios de aquella tierra de
11	1	indios de aquella tierra, la
12	1	indios de aquella tierra, y
13	1	indios de aquella tierra. Con
14	1	indios de diversas lenguas
15	1	indios de diversas lenguas y
16	1	indios de diversas naciones
17	1	indios de diversas naciones, por
18	1	indios de ella u

Figura 16. *Clusters Results*. Fuente: Elaboración propia.

Esta información, en *Clusters*, puede ser más detallada respecto a su frecuencia y el número total de las veces que ha aparecido en el texto la palabra *indios de* (48), junto con las siguientes palabras.

Ahora bien, pasemos finalmente a la herramienta *Word List*, la cual nos muestra las palabras que se han usado de mayor a menor frecuencia:

Concordance				Concordance Plot				File View				Clusters				Collocates				Word List				Keyword List			
Hits		1		Total No. of Word Types: 3898				Total No. of Word Tokens: 27706																			
Rank	Freq	Word																									
1	1924	de																									
2	1376	y																									
3	1000	que																									
4	928	a																									
5	751	la																									
6	555	en																									
7	531	los																									
8	519	el																									
9	436	con																									
10	433	n																									
11	370	por																									
12	300	se																									
13	257	s																									
14	256	del																									
15	254	su																									
16	212	o																									
17	180	lo																									
18	157	al																									
19	156	las																									
20	152	mã																									
21	148	indios																									
22	146	donde																									
23	139	para																									
24	137	no																									

Figura 17. *Word List*. Fuente: Elaboración propia.

Resaltamos aquí lo siguiente: las palabras como *de, y, que, a, la, en, los, con, por, del*, son las de mayor frecuencia. Esto es evidente por el uso necesario de preposiciones, artículos, conectores, para un entendimiento óptimo de las oraciones. Sin embargo, podríamos descartar estas palabras vacías (también conocidas como *stopwords*) para que, de esta manera, se identifiquen cuáles son las palabras claves del texto con mayor frecuencia. En ese sentido, se modificará en *Tool Preferences - Category - Word List* para eliminar estas palabras de la

lista. Para ello, se dará clic a *Use a stoplist listed below* para agregar las palabras vacías que se desea eliminar<sup>4</sup>.

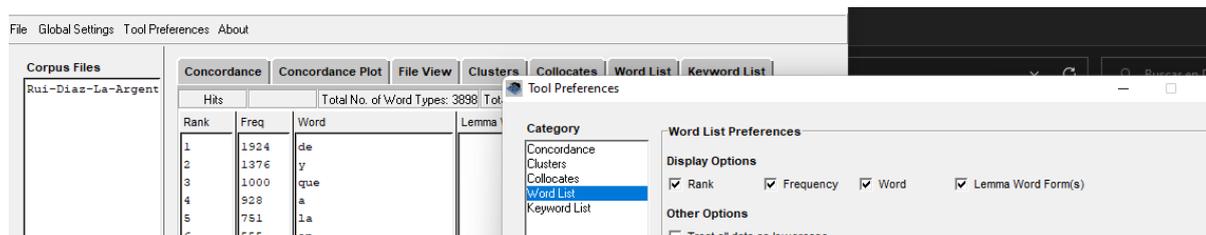


Figura 18. *Tool Preferences*. Fuente: Elaboración propia.

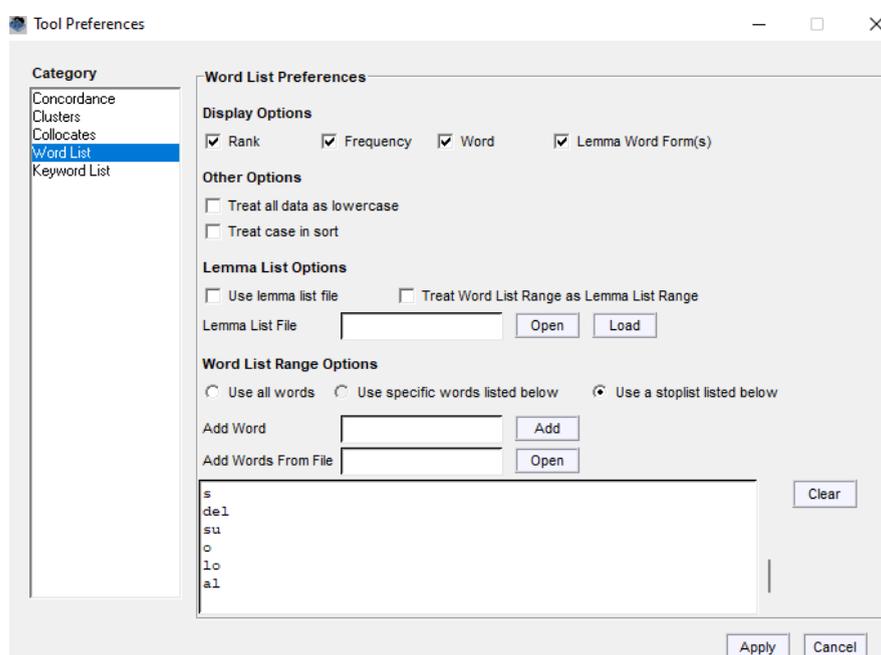


Figura 19. Eliminación de palabras vacías. Fuente: Elaboración propia.

Asimismo, para una adecuada eliminación de las palabras vacías, se ha descargado en txt. todas las palabras correspondientes de *Word List* con el objetivo de conseguir una revisión más minuciosa. Una vez que se han identificado estas palabras, se procede a colocarlas en la lista de *Tool Preferences* y darle clic a *Apply*.

Finalmente, conseguiremos quedarnos con las palabras relacionadas al texto, como es el caso de *indios* (con una frecuencia de

<sup>4</sup> Para la elaboración de esta última parte, ha sido de mucha utilidad revisar el vídeo de YouTube de Anthony Laurence: "AntConc 3.4.0 Tutorial 9: Word List Tool". En <https://www.youtube.com/watch?v=zHYrfxijbly>

148), *tierra* (112), *ellos* (80), *gente* (80), *don* (76), *Juan* (65), *costa* (57), *soldados* (44), entre otros:

Rank	Freq	Word	Lemma Word Form(s)
1	148	indios	
2	112	tierra	
3	80	ellos	
4	80	gente	
5	76	don	
6	73	parte	
7	69	gran	
8	68	llaman	
9	66	leguas	
10	65	Juan	
11	65	todos	
12	62	puerto	
13	60	muchos	
14	57	algunos	
15	57	costa	
16	54	sin	
17	52	otros	
18	50	fueron	
19	48	dos	
20	48	muchas	
21	44	ser	
22	44	soldados	
23	42	arriba	
24	41	Plata	

Figura 20. *Word List Results*. Fuente: Elaboración propia.

De esta manera, hemos explorado cómo podemos obtener información de una palabra, en su frecuencia, en su relación con otras palabras, en su ubicación en el texto y en su contexto, gracias al uso de las pestañas que proporciona AntConc.

Las tres últimas herramientas restantes, *Concordance Plot*, *Keyword List* y *Collocates*, permiten, respectivamente, visualizar los resultados de las palabras claves del texto, comparar con otro corpus y presentar altas probabilidades de las palabras que están juntas.

### 3. CONCLUSIONES

Para el uso de la herramienta AntConc se ha procedido a descargar un corpus; buscar palabras claves en su contexto; y, con ello, identificar patrones para, luego, establecer criterios y plantear preguntas a partir de lo que se descubre en los resultados de la data. Por ende, podríamos

afirmar que AntConc es una herramienta útil de las Humanidades Digitales al permitir emplear una lectura distante en un análisis de corpus como se ha demostrado en el caso del libro I de *La Argentina Manuscrita*. Ha permitido analizar las frecuencias, patrones y comparaciones entre palabras, tales como *hij\**, *indi\** e *indi?*, junto a la conexión de oraciones que están ligadas a estas palabras claves.

Esto, finalmente, enriquece nuestra lectura si consideramos aspectos que, de manera global, no podríamos capturar si no se recolecta y se analizan los datos. De esta manera, una lectura cercana se complementa con una distante al permitir visualizar datos que muestran patrones, comparaciones, tendencias y, además, nos permite cuestionar o interrogar aspectos de análisis lingüístico del corpus.

## REFERENCIAS BIBLIOGRÁFICAS

- del Rio Riande, Gimena, Hernández, Nidia y De León, Romina (eds.). (2019-2020). Ruy Díaz de Guzmán. Historia Argentina del descubrimiento, población y conquista de las provincias del Río de la Plata. HD CAICYT Lab. <http://hdlab.space/La-Argentina-Manuscrita/> [16 de Junio de 2022]
- Froehlich, H. (2018). Análisis de corpus con AntConc, traducido por Carlos Manuel Varón Castañeda. *Programming Historian en español*, 2. <https://doi.org/10.46430/phes0032>
- Laurence, A. [Antlab by Laurence Anthony] (16 de junio de 2022). *AntConc 3.4.0 Tutorial 10: Keyword List Tool*. [Vídeo]. <https://www.youtube.com/watch?v=tMNm5Kb9LLE>