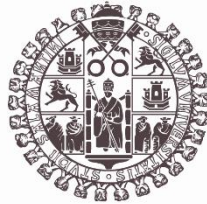


# Analizador Visual de Corpus Lingüísticos. Anexo V. Manuales de usuario

Trabajo de Fin de Grado

Ingeniería Informática



**VNiVERSiDAD  
D SALAMANCA**

Septiembre de 2024

David Prieto Santos

---

Roberto Therón Sánchez

---

Tabla de contenido

**1. Anexo V: Manuales de usuario ..... 1**

**1.1 Consideraciones previas..... 1**

**1.2 Manual de funcionamiento de la aplicación ..... 1**

## Tabla de figuras

Ilustración 1: Página principal .....	1
Ilustración 2: Página de carga de datos .....	2
Ilustración 3: Ventana de examinar archivo .....	2
Ilustración 4: Página con archivo examinado .....	3
Ilustración 5: Página de datos de muestra.....	3
Ilustración 6: Página con conjunto de datos de muestra elegido.....	3
Ilustración 7: Página de verificación .....	4
Ilustración 8: Página de verificación con error .....	4
Ilustración 9: Página de visualización .....	5
Ilustración 10: Gráfico de barras.....	5
Ilustración 11: Diagrama de sectores .....	6
Ilustración 12: Mapa de hexágonos.....	6
Ilustración 13: Estructura de un archivo con elementos anidados .....	7
Ilustración 14: Proyección solar (raíz).....	7
Ilustración 15: Proyección solar (nodo expandido) .....	8
Ilustración 16: Mapa de árbol (raíz).....	8
Ilustración 17: Mapa de árbol (nodo expandido) .....	9
Ilustración 18: Árbol de vectores .....	9

## 1. Anexo V: Manuales de usuario

### 1.1 Consideraciones previas

- No existe un manual de instalación ya que es una aplicación web, basta con navegar a la URL en el navegador.

### 1.2 Manual de funcionamiento de la aplicación

Primero se accederá a la aplicación a través del siguiente enlace:  
<https://davidprisan.observablehq.cloud/corpusling/>.



**Ilustración 1: Página principal**

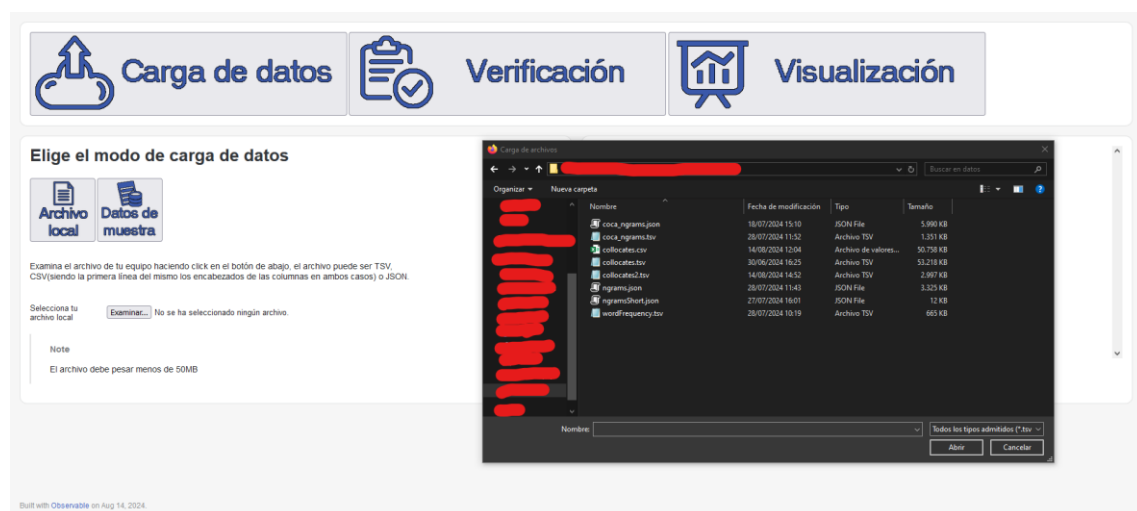
Una vez dentro, el usuario verá una página similar a la mostrada en la figura anterior.

Para empezar el proceso de representación de datos, debe pulsar el botón de Carga de datos, viendo a continuación lo mostrado en la figura siguiente:



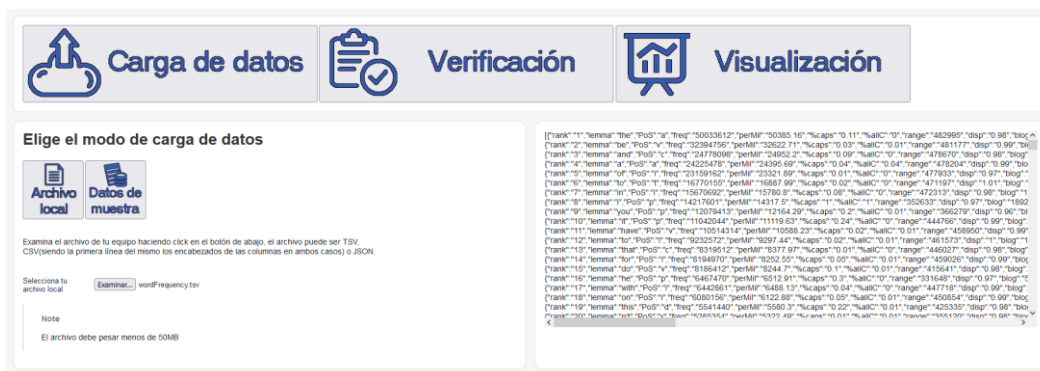
**Ilustración 2: Página de carga de datos**

Si quiere subir un archivo local pulsará el botón de Archivo local y posteriormente el de Examinar archivo:



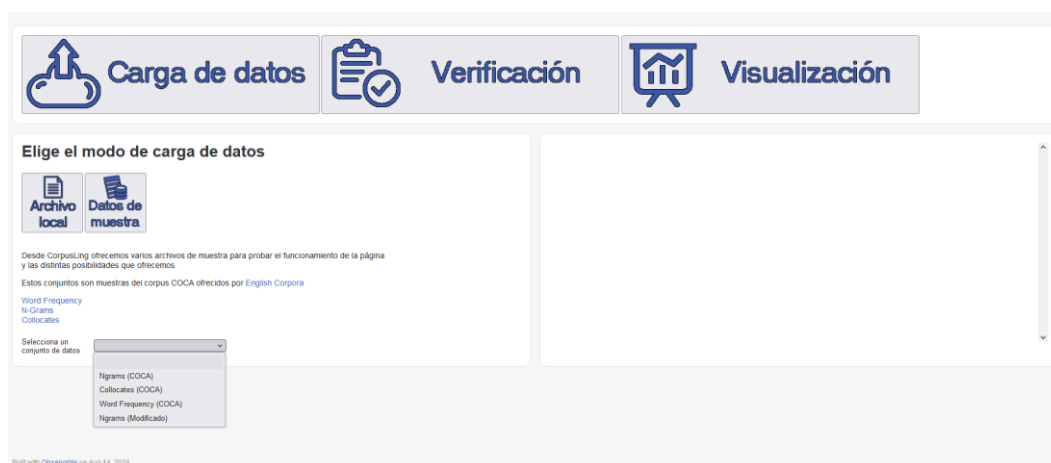
**Ilustración 3: Ventana de examinar archivo**

Una vez seleccionado el archivo correspondiente al corpus con el que quiere trabajar, verá su archivo procesado en formato JSON en la parte derecha de su pantalla:



**Ilustración 4: Página con archivo examinado**

Si, por el contrario, prefiere usar uno de los conjuntos de datos de muestra que proponemos desde CorpusLing, pulse el botón de Datos de muestra:



**Ilustración 5: Página de datos de muestra**

En esta ventana, puede elegir el conjunto de datos de muestra que prefiera para trabajar con él, así como acceder a las páginas de English-Corpora de donde proceden estos conjuntos de muestra (de uso gratuito).



**Ilustración 6: Página con conjunto de datos de muestra elegido**

Una vez seleccionado un conjunto de datos de muestra, puede ver dicho conjunto en formato JSON, igual que examinando un archivo de su equipo.

Con el conjunto de datos ya elegido, el siguiente paso es la verificación de datos, a la que puede acceder una vez pulse en el botón de Verificación, mostrando una página similar a la siguiente:

Carga de datos

Verificación

Visualización

Comprueba que tus datos se ven bien

Ordenar por rank Ascendente

Comprueba que tus datos se han procesado correctamente. En la tabla de la derecha puedes ver los números de color azul y el texto en verde. Una **error** de color rojo indica que hay un dato que no se corresponde con la columna o que hay un problema con el conjunto de datos.

Puedes editar el contenido de las celdas haciendo click en cada una de ellas, realiza todos los cambios que consideres antes de avanzar a la parte de visualización.

rank	lemma	PoS	freq	perM	%caps	%altC	range	disp	blog	web	TVM	spok	fic	mag	news	acad	blogPM	webPM	TVMPM
1	the	a	58033612	50385.16	0.11	0	482995	0.98	6266654	7095508	3781291	5760948	6367838	6801589	6579270	7440931	50434.35	55189.32	29524.14
2	be	v	32394756	32622.71	0.03	0.01	481177	0.99	5594081	5325829	8029744	7025941	4412160	4193720	4102481	3689062	45020.81	41409.63	62695.87
3	and	c	24778088	24952.2	0.09	0	478678	0.98	3205178	3453140	1824964	3314232	3059827	3300767	2993081	3627686	25795.44	26849.01	14249.23
4	a	a	24225478	24395.69	0.04	0.04	478204	0.99	3098338	3182822	2649922	2948384	2970752	3472793	3299770	2602697	24935.58	24747.22	20696.47
5	of	i	23159162	23321.89	0.01	0	477933	0.97	2897295	3419616	1398417	2542569	2280164	3252696	2867922	4500485	23317.58	26588.36	10918.77
6	to	t	16770155	16887.99	0.02	0	471197	1.01	2395858	2286438	2442742	2364762	1794557	1950363	1854109	1681326	19281.99	17777.62	19072.82
7	in	i	15670682	15788.8	0.08	0	472313	0.98	1890236	2145004	1097130	1896531	1561200	2213407	2276571	2590613	15212.72	16677.93	8566.34
8	i	p	14217681	14317.5	1	1	352633	0.97	1892542	1450100	4814503	2215625	2106849	848473	661408	240858	15231.28	11274.88	37591.42
9	you	p	12079413	12164.29	0.2	0.01	366279	0.96	1228877	1097675	5050382	2234800	1240860	728340	388496	111372	9890.07	8534.69	39433.15
10	s	p	11042044	11119.63	0.24	0	444766	0.99	1491385	1343709	2316063	1897943	1408985	1058450	949681	576423	12002.74	10447.67	18083.71
11	have	v	10514314	10588.23	0.02	0.01	458950	0.99	1633029	1525515	1761082	1902874	1807548	1335863	1411100	937640	13142.7	11861.25	13750.45

Ilustración 7: Página de verificación

Ahora se le mostrará una tabla con los datos de su corpus, pudiendo ordenar los datos por el valor de cada columna, de forma ascendente o descendente (por defecto: primera columna, ascendente).

Como se le explica en la página, las columnas se colorearán de verde para datos de tipo texto y de azul para datos de tipo numérico, si hubiera algún error se colorearía la celda correspondiente de rojo:

Carga de datos

Verificación

Visualización

Comprueba que tus datos se ven bien

Ordenar por rank Ascendente

Comprueba que tus datos se han procesado correctamente. En la tabla de la derecha puedes ver los números de color azul y el texto en verde. Una **error** de color rojo indica que hay un dato que no se corresponde con la columna o que hay un problema con el conjunto de datos.

Puedes editar el contenido de las celdas haciendo click en cada una de ellas, realiza todos los cambios que consideres antes de avanzar a la parte de visualización.

rank	lemma	PoS	freq	perM	%caps	%altC	range	disp	blog	web	TVM	spok	fic	mag	news	acad	blogPM	webPM	TVMPM
<b>a</b>	the	a	58033612	50385.16	0.11	0	482995	0.98	6266654	7095508	3781291	5760948	6367838	6801589	6579270	7440931	50434.35	55189.32	29524.14
2	be	v	32394756	32622.71	0.03	0.01	481177	0.99	5594081	5325829	8029744	7025941	4412160	4193720	4102481	3689062	45020.81	41409.63	62695.87
3	and	c	24778088	24952.2	0.09	0	478678	0.98	3205178	3453140	1824964	3314232	3059827	3300767	2993081	3627686	25795.44	26849.01	14249.23
4	a	a	24225478	24395.69	0.04	0.04	478204	0.99	3098338	3182822	2649922	2948384	2970752	3472793	3299770	2602697	24935.58	24747.22	20696.47
5	of	i	23159162	23321.89	0.01	0	477933	0.97	2897295	3419616	1398417	2542569	2280164	3252696	2867922	4500485	23317.58	26588.36	10918.77
6	to	t	16770155	16887.99	0.02	0	471197	1.01	2395858	2286438	2442742	2364762	1794557	1950363	1854109	1681326	19281.99	17777.62	19072.82
7	in	i	15670682	15788.8	0.08	0	472313	0.98	1890236	2145004	1097130	1896531	1561200	2213407	2276571	2590613	15212.72	16677.93	8566.34
8	i	p	14217681	14317.5	1	1	352633	0.97	1892542	1450100	4814503	2215625	2106849	848473	661408	240858	15231.28	11274.88	37591.42
9	you	p	12079413	12164.29	0.2	0.01	366279	0.96	1228877	1097675	5050382	2234800	1240860	728340	388496	111372	9890.07	8534.69	39433.15
10	s	p	11042044	11119.63	0.24	0	444766	0.99	1491385	1343709	2316063	1897943	1408985	1058450	949681	576423	12002.74	10447.67	18083.71
11	have	v	10514314	10588.23	0.02	0.01	458950	0.99	1633029	1525515	1761082	1902874	1807548	1335863	1411100	937640	13142.7	11861.25	13750.45

Ilustración 8: Página de verificación con error

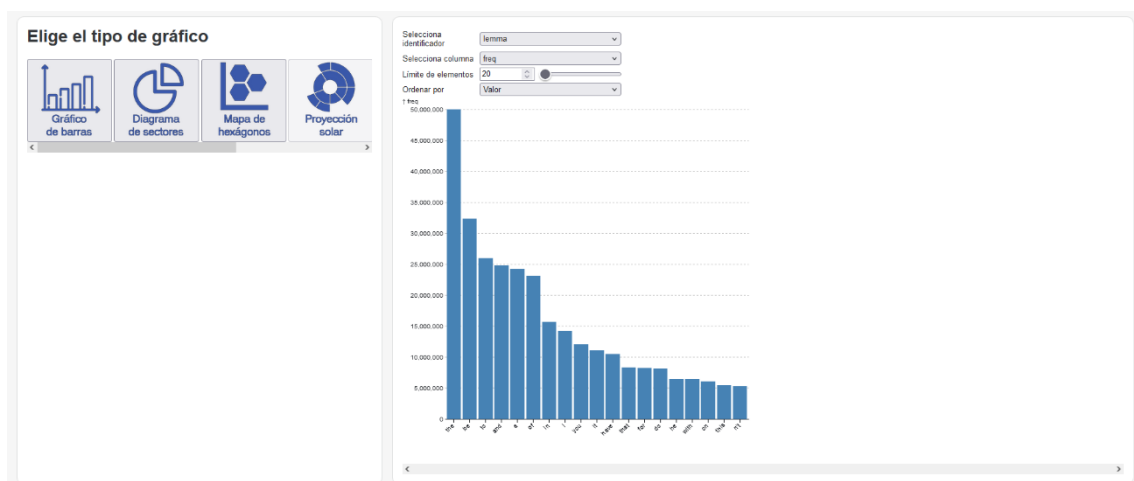
Una vez haya comprobado que sus datos se han procesado correctamente y no hay errores, puede pasar al apartado de visualización, pulsando el botón correspondiente:



**Ilustración 9: Página de visualización**

En esta ventana, podrá elegir el tipo de gráfico con el que quiere representar su corpus pulsando el botón para cada uno de ellos en la botonera.

Vea como se representa cada uno de los gráficos usando el conjunto de datos de muestra de Word Frequency:

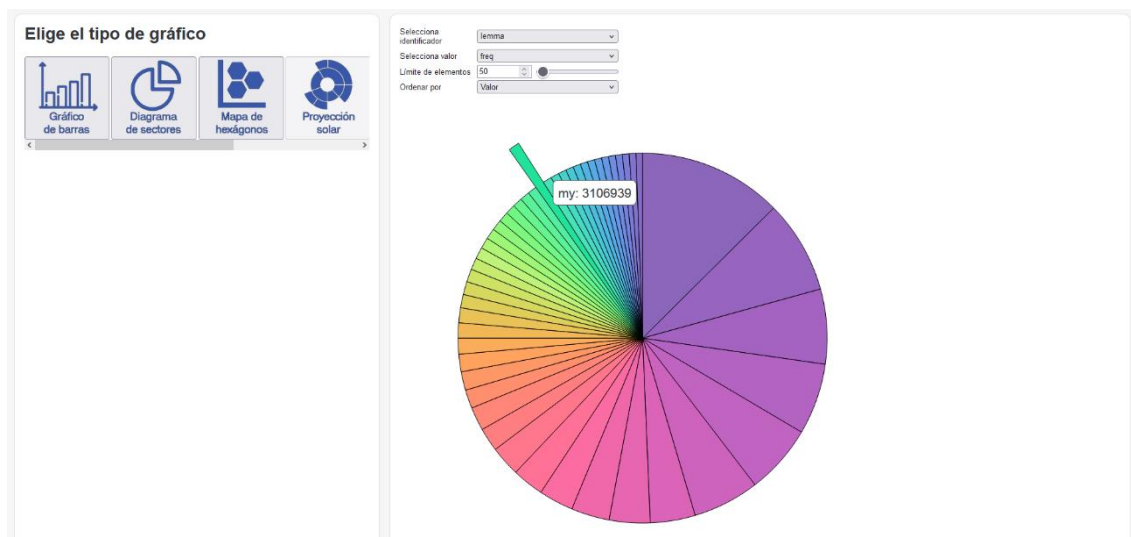


**Ilustración 10: Gráfico de barras**

El primero es el clásico gráfico de barras, simple pero eficaz, Se le permite elegir el identificador de entre las columnas de tipo texto para el eje X y el valor de entre las columnas de tipo numérico para el eje Y.

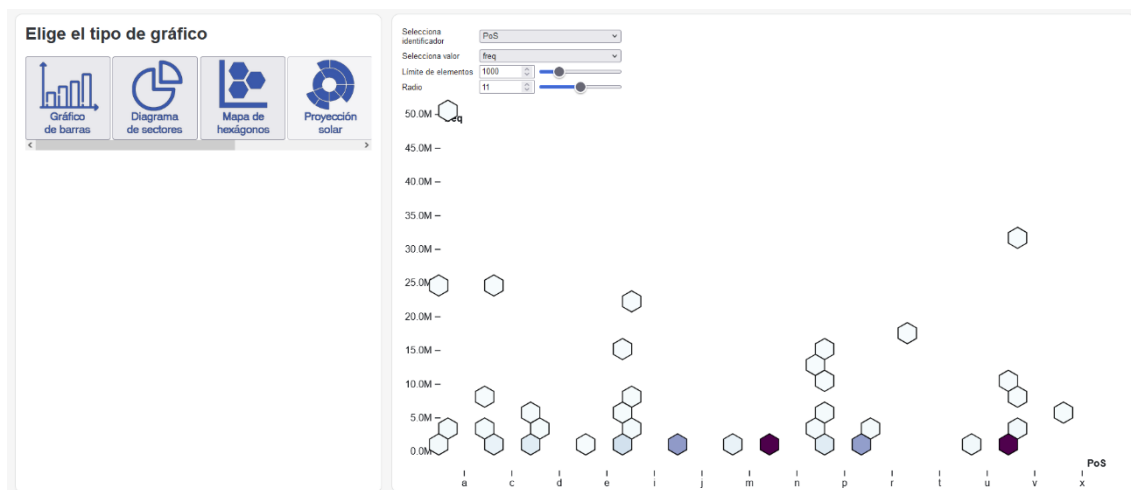
También se ofrecen las posibilidades de limitar el número de elementos a mostrar y la forma en que se ordenan dichos elementos: por valor o por identificador.





**Ilustración 11: Diagrama de sectores**

El siguiente es el gráfico de sectores, ofrece los mismos filtros que el gráfico de barras. Cuando el ratón pasa por encima de un sector, dicho sector se resalta y se le muestra el identificador y valor del sector, para facilitar la legibilidad.



**Ilustración 12: Mapa de hexágonos**

Este gráfico es el mapa de hexágonos, cuenta con los mismos filtros que los anteriores, a excepción de la opción de ordenar, sustituida por un selector de radio para aumentar o disminuir el tamaño de los hexágonos.

Se recomienda el uso de este gráfico si el corpus cuenta con alguna característica que permita agrupar elementos. El corpus de Word Frequency que se usa de ejemplo cuenta con la columna de *PoS* (tipo de palabra), por lo que, aunque seleccionemos 1.000 elementos, el gráfico sigue siendo legible.

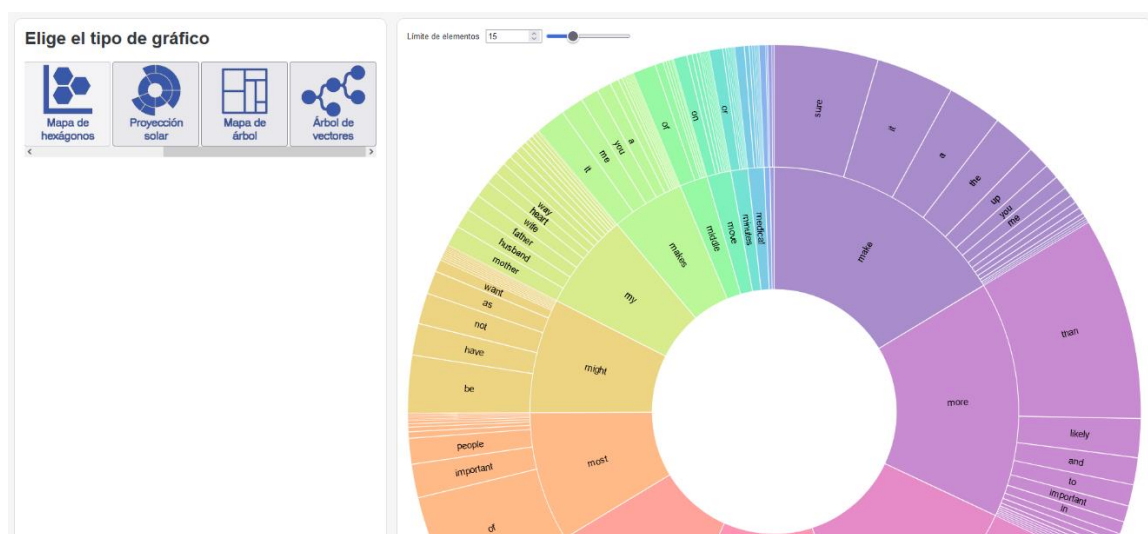
Cabe destacar que el gráfico se representará aunque la columna seleccionada no agrupe los elementos, pero el eje X no será legible con muchos elementos.

Como seguramente haya apreciado, hasta este momento, el resto de los botones están desactivados dado que el corpus no es adecuado para utilizarlos. Si quiere usar su propio corpus para representarlo con estos gráficos, el archivo debe tener la estructura que se muestra a continuación:

```
[{"name":"m","children":[{"name":"may","children":[{"name":"not","children":[{"name":"be","children":[{"name":"able","children":[{"name":"to","value":"2352"}]},{name":"the","children":[{"name":"best","value":"480"}]},{name":"most","value":"259"}, {"name":"same","value":"173"}]},{name":"case","value":"165"}]},{name":"only","value":"149"}]},{name":"right","value":"102"}, {"name":"answer","value":"48"}]},{name":"last","value":"46"}]},{name":"first","value":"41"}]},{name":"one","value":"35"}, {"name":"way","value":"34"}]},{name":"end","value":"30"}]},{name":"perfect","value":"29"}]},{name":"greatest","value":"25"}, {"name":"biggest","value":"24"}]},{name":"worst","value":"23"}]},{name":"ideal","value":"18"}]},{name":"whole","value":"17"}, {"name":"primary","value":"15"}]},{name":"smartest","value":"15"}]},{name":"time","value":"14"}]},{name":"kind","value":"13"}, {"name":"ones","value":"12"}]},{name":"world","value":"12"}]},{name":"place","value":"11"}]},{name":"sole","value":"11"}, {"name":"correct","value":"10"}]},{name":"easiest","value":"10"}]},{name":"exact","value":"10"}]},{name":"fastest","value":"10"}, {"name":"final","value":"10"}]},{name":"real","value":"10"}]},{name":"sharpest","value":"10"}]},{name":"truth","value":"10"}, {"name":"actual","value":"9"}]},{name":"man","value":"9"}]},{name":"result","value":"9"}]},{name":"brightest","value":"8"}, {"name":"cheapest","value":"8"}]},{name":"entire","value":"8"}]},{name":"top","value":"8"}]},{name":"ultimate","value":"8"}]},{name":"enough","children":[{"name":"to","value":"225"}]},{name":"for","value":"33"}]},{name":"time","value":"21"}, {"name":"of","value":"8"}]},{name":"aware","children":[{"name":"of","value":"224"}]},{name":"that","value":"47"}]},{name":"in","children":[{"name":"the","value":"157"}]},{name":"a","value":"45"}]},{name":"its","value":"43"}]},{name":"your","value":"16"}]},{name":"my","value":"12"}, {"name":"their","value":"10"}]},{name":"good","value":"8"}]},{name":"our","value":"8"}]},{name":"a","children":[{"name":"good","value":"142"}]},{name":"bad","value":"58"}]},{name":"problem","value":"47"}]},{name":"great","value":"39"}, {"name":"big","value":"37"}]},{name":"perfect","value":"31"}]},{name":"very","value":"25"}]},{name":"coincidence","value":"23"}, {"name":"household","value":"20"}]},{name":"lot","value":"20"}]},{name":"major","value":"19"}]},{name":"part","value":"14"}, {"name":"direct","value":"13"}]},{name":"matter","value":"12"}]},{name":"more","value":"12"}]},{name":"fan","value":"10"}, {"name":"true","value":"10"}]},{name":"valid","value":"10"}]},{name":"huge","value":"9"}]},{name":"real","value":"9"}]
```

**Ilustración 13: Estructura de un archivo con elementos anidados**

También puede usar el corpus N-grams (modificado).



**Ilustración 14: Proyección solar (raíz)**



