

WordSmith Tools, un primer encuentro

1.	WordSmith Tools.....	1
2.	Acceder al programa.....	1
3.	Registrar el programa.....	1
4.	Vuelta de reconocimiento.....	1
5.	Modificar la lengua de trabajo.....	2
6.	«Concord».....	4
a.	Búsqueda básica.....	5
b.	Búsqueda avanzada.....	6
c.	«collocates».....	8
d.	«patterns».....	8
e.	«clusters».....	9
7.	«WordList».....	9
a.	«Stop List».....	9
b.	«Compare WordLists».....	9

1. WordSmith Tools



WordSmith Tools es un programa de análisis de corpus textuales. Gracias a este tipo de herramientas se pueden explotar grandes conjuntos de textos mediante búsquedas de carácter contextual o estadístico.

2. Instalar el programa

- Haz Ctrl + clic sobre el enlace siguiente. Si no se abre ninguna ventana, corta y pega la dirección en el cajetín de búsqueda del explorador.
http://www.lexically.net/downloads/version4/all_wordsmith_4.exe
- Se abrirá una ventana automáticamente. Haz clic sobre «Run». Puede que salga otra ventana, si es así, haz clic de nuevo sobre «Run».
- El programa creará el icono amarillo the WordSmith Tools en el escritorio (Section 1) y también se encontrarán en el menú de Inicio de Windows () , bajo el submenú programas.

3. Registrar el programa

- Abre WordSmith Tools.
- En la ventana del «Controller»/ en el menú «File», haz clic sobre «Update from demo».
- Se abrirá una ventana. Introduce los datos siguientes:

```
Registered to: <YOUR NAME>
Other details:
Registration code: XXXX.XXXX.XXXX.XXXX.XXXX
```

4. Vuelta de reconocimiento

Al arrancar el programa se abre la ventana «Controller», que debe estar abierta en todo momento para el correcto funcionamiento del programa. Para más comodidad, puedes minimizar-la.

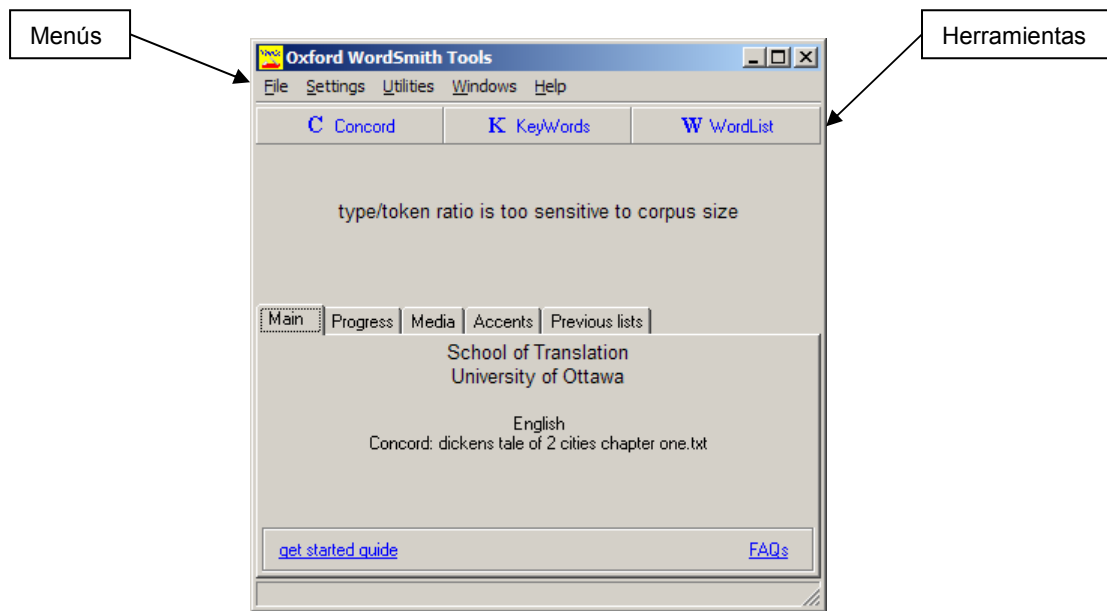


Imagen 2. Ventana «Controller»

Desde esta ventana podrás navegar entre las distintas herramientas del programa y los menús de opciones. Para este proyecto, nos centraremos en dos de las tres herramientas principales: «Concord» y «WorldList».

5. Modificar la lengua de trabajo

Antes de proseguir, es necesario ajustar los parámetros del programa. Para ello, abre el menú «Settings» y haz clic sobre «Adjust settings....». Se abrirá una ventana con distintas pestañas. Activa la pestaña «Languages».

Como verás (Imagen 3), el español no consta en la lista de idiomas. Es por eso que hay que modificarla.

- Haz clic sobre el botón de la derecha «Edit Languages» y se abrirá una pestaña distinta (Imagen 4).
- Desplaza hacia abajo el menú «Available Languages» de la parte inferior izquierda hasta encontrar «Spanish» y selecciónalo.
- Una vez seleccionado, haz clic sobre la flecha azul a la derecha del menú y verás como se añade a la lista de la derecha.
- Al cerrar la ventana, el programa preguntará si quieres guardar la selección, acepta y a continuación haz clic sobre OK.
- Ahora, el programa está listo para empezar a trabajar.

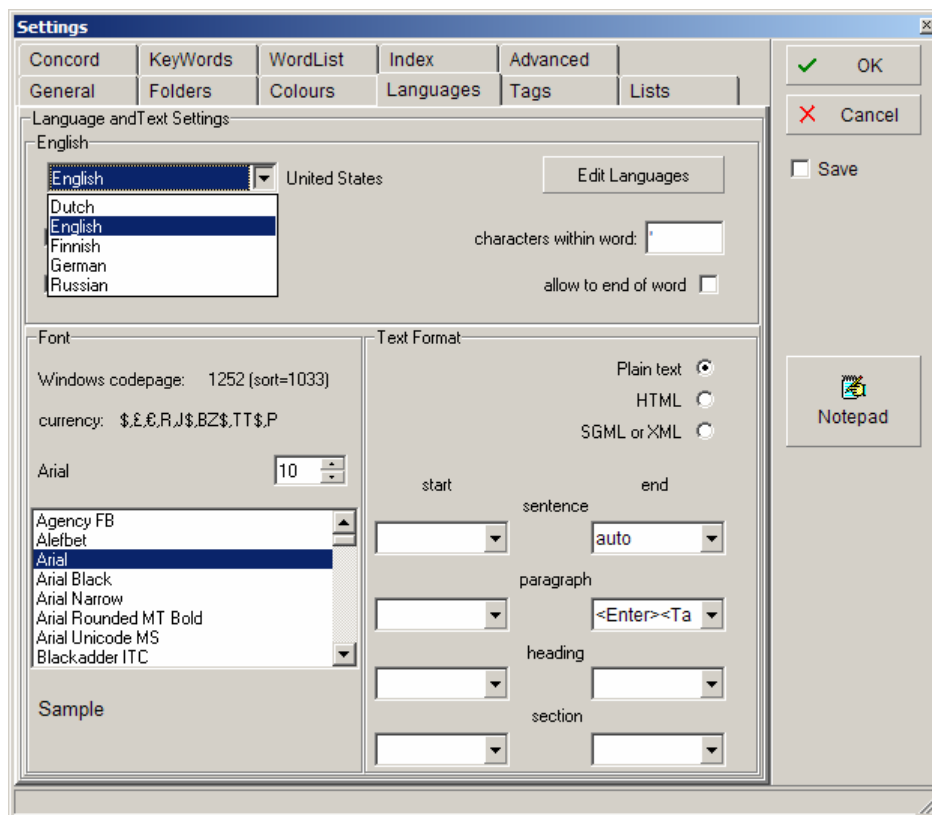


Imagen 3. Pestaña «Languages»

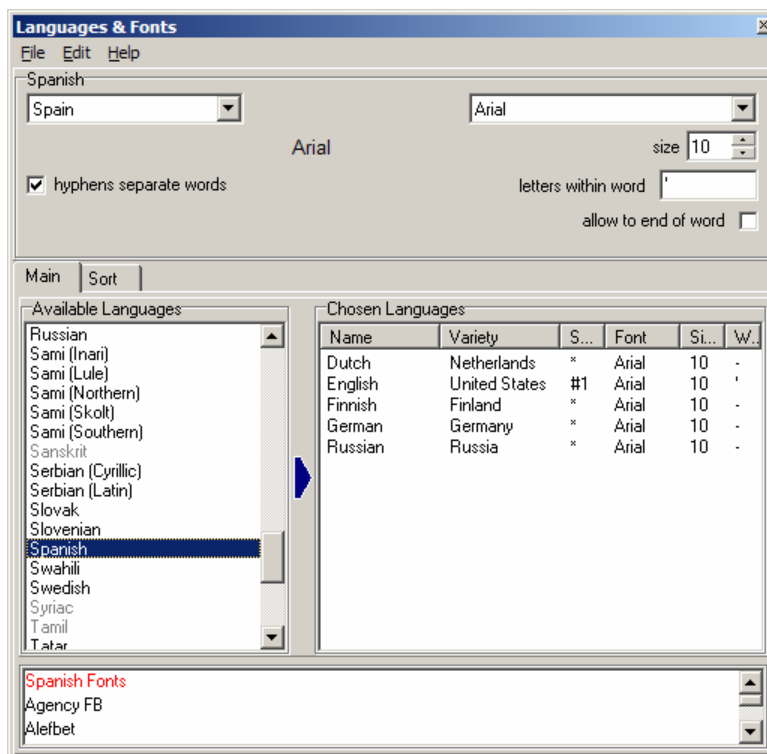


Imagen 4. Ventana de modificación de lenguas

6. «Concord»

«Concord» es un programa que extrae listas de ejemplos del uso de palabras en contexto (concordancias), presentes en el corpus. Por ejemplo, la Imagen 5 muestra una lista de ejemplos del uso de la forma de tratamiento «señor» en distintos textos de Benito Pérez Galdós.

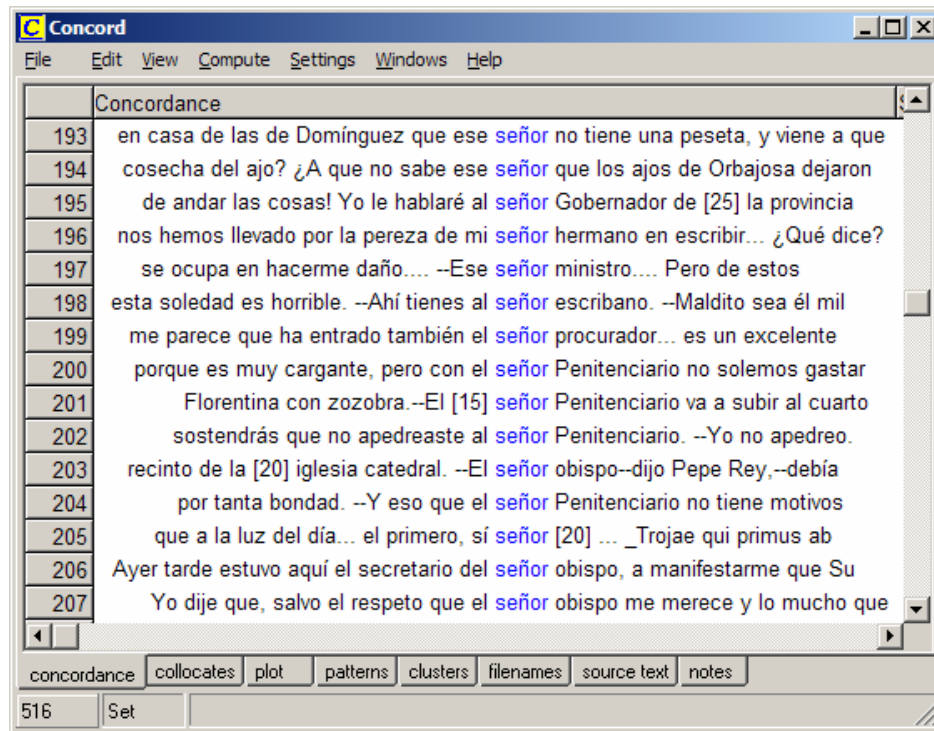


Imagen 5. Lista de concordancias

Uso del «Concord» paso a paso:

- Para abrir la herramienta, haz clic sobre el botón «Concord» de la ventana «Controller».
- Para realizar una búsqueda, abre el menú «File» y haz clic sobre «New». O simplemente usa el atajo Ctrl + N.
- En la primera pestaña «Texts» es dónde puedes escoger en qué textos se llevará a cabo la búsqueda.
- Haz clic en el botón «Choose Texts Now» y se abrirá una ventana nueva. En la parte izquierda verás la estructura de la raíz de tu ordenador. Usa el botón  para subir niveles o haz clic en las distintas carpetas para abrirlas.
- Una vez hayas encontrado la carpeta donde guardes tu corpus, selecciona los documentos que quieras usar.
- A continuación, haz clic sobre la flecha azul a la derecha del menú y verás como se añade a la lista de la derecha.
- Haz clic sobre  en la esquina superior derecha.

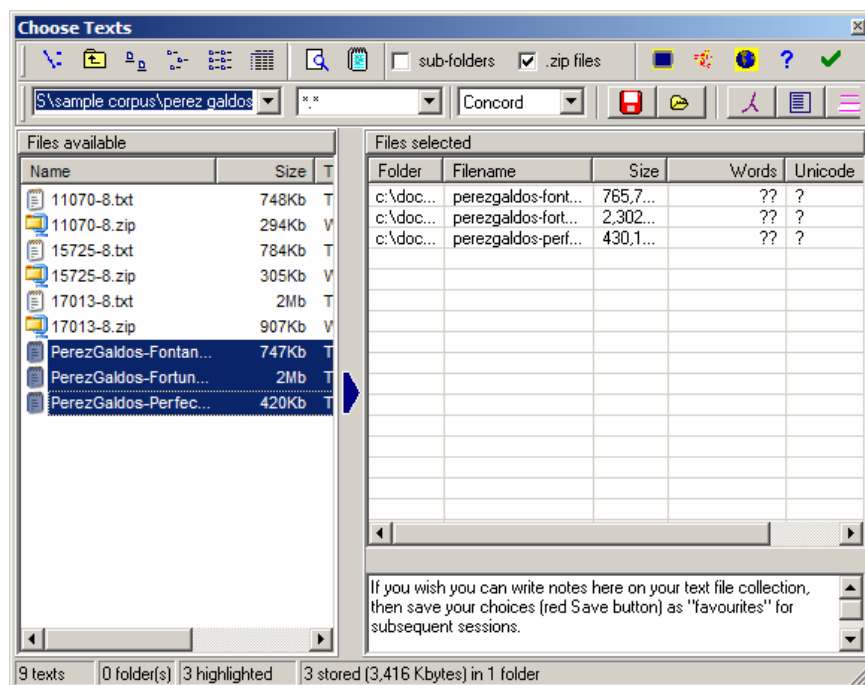


Imagen 6. Ventana de selección de textos.

a. Búsqueda básica

La ventana de búsqueda se abre automáticamente tras escoger los textos con los que quieres trabajar. Si no fuera así, abre el menú «File» de la ventana «Concord» y haz clic en «New» o pulsa Ctrl + N.

«Concord» dispone de dos interfaces de consulta: una básica (pestaña «Search Word») y otra avanzada (pestaña «Advanced»). En el cajetín de búsqueda pueden introducirse palabras o expresiones (Imagen 7). Por defecto, el programa no diferencia entre mayúsculas o minúsculas. Una de las virtudes de esta herramienta es su sintaxis de consulta:

Símbolo	Descripción	Ejemplos
*	ignorar el final de la palabra, ignorar el principio de la palabra, ignorar el principio y el final de la palabra, ignorar toda una palabra	tele* *ness *happi* book * hotel
?	aceptar cualquier carácter, incluso signos de puntuación	Engl???
#	aceptar cualquier cifra, del 0 al 9	?50.00 \$### £##.00
^	aceptar cualquier letra del alfabeto	Fr^nc^
==	respetar mayúsculas o minúsculas	==French== ==Fr*==
/	Separar palabras alternativas.	may/can/will

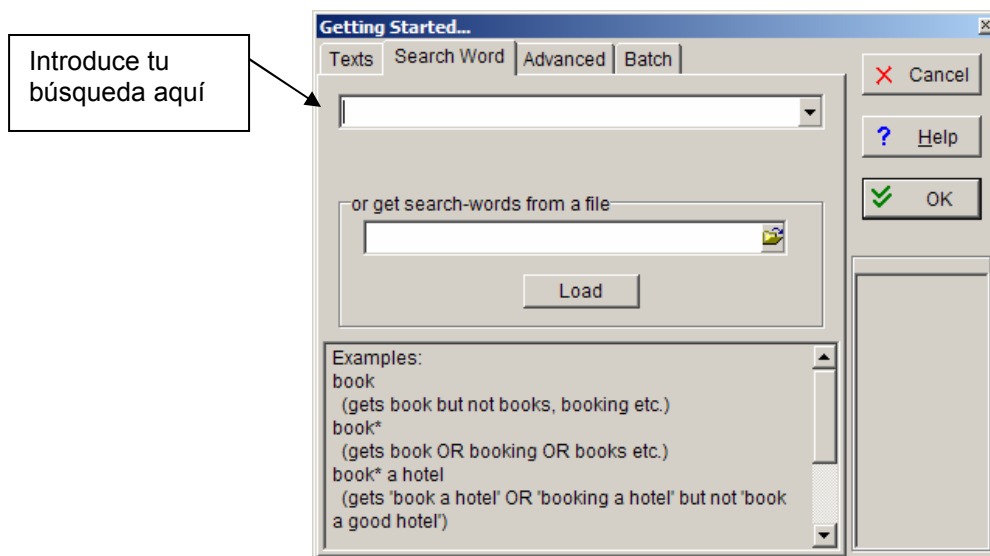


Imagen 7. Búsqueda básica.

b. Búsqueda avanzada

En la siguiente pestaña «Advanced», las opciones de búsqueda avanzada permiten introducir dos condiciones a la búsqueda básica.

- 1) Buscar ejemplos de expresiones que contengan una palabra o expresión principal, introducida en el cajetín de texto de la búsqueda básica, a cierto número de palabras hacia la derecha (R) o izquierda (L) de otra palabra o expresión, que debe introducirse en esta misma pestaña «Advanced».
- 2) Excluir ejemplos que contengan cierta palabra o expresión.

Recuerda: Puedes servirte de la misma sintaxis de consulta.

En la parte inferior derecha verás un cuadro resumen de los parámetros de consulta que has introducido.

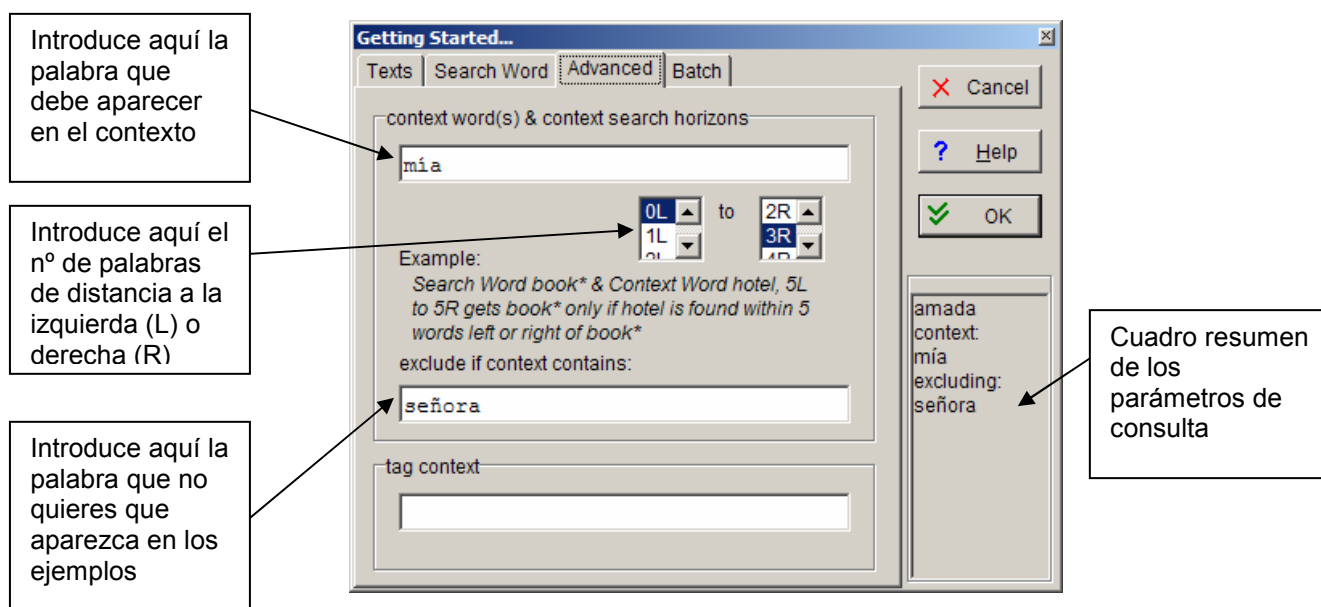


Imagen 10. Ventana de búsqueda avanzada

A menudo se obtienen cantidades ingentes de ejemplos, que aunque pueden contener una gran riqueza de información, puede resultar imposible de analizar. Para facilitar la explotación de los resultados, el programa ofrece una gran variedad de maneras de ordenar estos últimos. Para ello, haz clic sobre la barra «Concordance» sobre la lista de ejemplos y se abrirá una ventana nueva llamada «Concordance sort», en la que podrás establecer tres criterios de ordenación. Generalmente, un solo criterio bastará.

En la Imagen 8 puedes ver la lista de criterios disponibles. Cada opción ordenará los ejemplos alfabéticamente en función del elemento seleccionado. Por ejemplo, si se realiza una búsqueda de ejemplos donde aparezca las preposiciones «para» o «por» y se ordenan según «Centre», todos los ejemplos que contengan «para» se listarán antes que los que contengan «por». Las opciones L1, L2, L3, L4 y L5, listarán los ejemplos en orden alfabético según la palabra que se encuentre a 1, 2, 3, 4 o 5 palabras de distancia de la palabra o expresión introducida en la búsqueda básica. R1, R2, R3, R4 y R5 harán lo mismo pero hacia la derecha. Finalmente, la opción «Context Word» ordenará los ejemplos alfabéticamente según la palabra o expresión introducida en la pestaña de búsqueda avanzada.

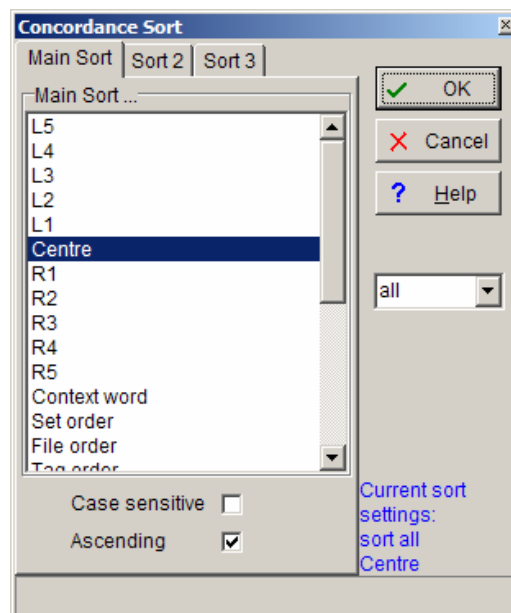


Imagen 8. Criterios de ordenación.

Como se puede observar en la Imagen 9, ordenar los ejemplos ayuda a descubrir tendencias y construcciones. Por ejemplo, la Imagen 9 ilustra el uso de la locución «por mucho que».

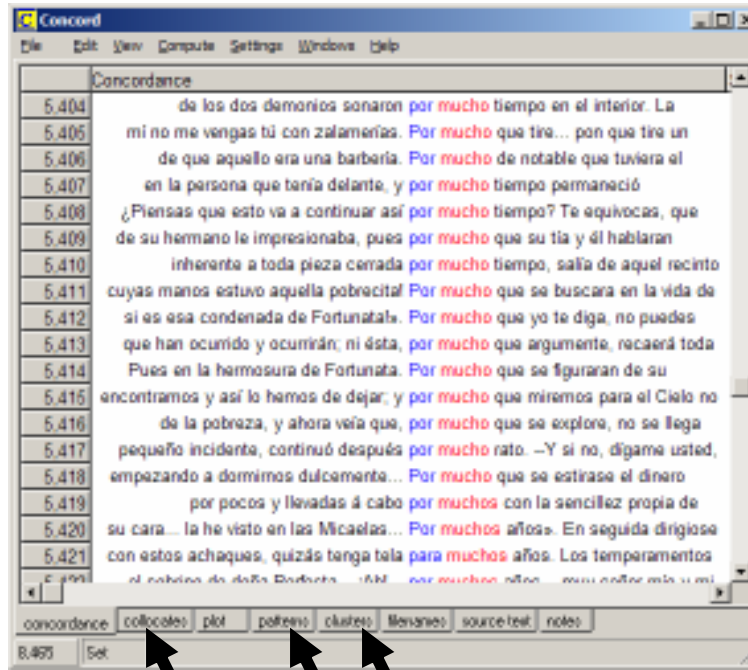


Imagen 9. Ejemplos ordenados según L1

El programa todavía tiene más herramientas de análisis. Bajo la lista de ejemplos se encuentran una serie de pestañas. Hasta ahora, hemos trabajado sólo en la pestaña «Concordance». Sin embargo, encontrarás información muy interesante en las pestañas «Collocates», «Patterns» y «Clusters» (Imagen 9).

c. «collocates»

La información que contiene esta pestaña pretende ayudar a identificar tendencias y estructuras. La tabla de datos muestra las veces que la palabra de la primera columna aparece junto a la palabra de la segunda y en qué posición se encuentra (una, dos, tres palabras a la derecha, a la izquierda etc.).

Inicialmente la tabla está organizada por orden de frecuencia. Sin embargo, haciendo clic sobre el título de cada columna los datos pueden reorganizarse alfabéticamente o de mayor a menor y viceversa.

d. «patterns»

Esta ventana ofrece la misma información que la anterior pero de forma distinta. Mientras que la anterior expresa la relación de las palabras en cifras, ésta muestra listas de palabras ordenadas de mayor a menor frecuencia. Aunque mostrar dos veces la misma información puede parecer redundante, recuerda que hay quien interpreta mejor la información en cifras y quien la interpreta mejor en palabras.

El valor de frecuencia varía en función del corpus que usemos. Por ejemplo, en un corpus pequeño (10 000 palabras) 20 ocurrencias de una determinada estructura pueden representar una frecuencia elevada mientras que si se tratara de un corpus grande (100 millones de palabras), 20 ocurrencias indicarían una estructura de uso marginal. Por eso puede personalizarse el valor de frecuencia mínimo:

- a. Ve a la ventana «Controller», menú «Settings», «Adjust Settings....» y a la pestaña «Concord».
- b. En la parte inferior de ésta, en el cuadro «Collocates», puedes modificar el valor de frecuencia mínimo, el número de palabras por unidad, el número de palabras que se considerarán alrededor y los límites de análisis, por ejemplo, se puede detener el análisis siempre que se llegue al final de una frase.

e. «clusters»

Esta pestaña presenta grupos de tres palabras que aparecen con frecuencia en la lista de ejemplos. Este análisis saca a relucir estructuras fraseológicas.

7. «WordList»

Como el nombre indica, en esta pestaña puedes elaborar listas de palabras según la frecuencia con la que aparecen en el corpus. Puedes reordenar la lista alfabéticamente, de mayor a menor frecuencia o por archivo con sólo hacer clic sobre el título de cada columna.

a. «Stop List»

Si lo que te interesa son palabras de contenido, puede que preposiciones, artículos y demostrativos, (las palabras, por su función, más frecuentes en español) entorpezcan tu trabajo.

Para solucionar este problema puedes crear una lista con las palabras que quieres excluir del análisis.

- a. Escribe la lista en la Libreta de notas o en cualquier otro programa siempre y cuando guardes el archivo como sólo texto (.txt).
- b. Separa las palabras con comas o escribe una palabra por línea.
- c. Una vez tengas la lista, ve a la ventana «Controller», al menú «Settings», en la opción «Adjust Settings...», en la pestaña «Lists».
- d. Una vez ahí, en el cajetín «stop list» busca la lista que acabas de crear en tu ordenador pulsado el botón  y cuando la encuentres, selecciónala.
- e. Haz clic sobre el botón «Load».

La próxima vez que crees una lista de palabras, el programa ignorará las palabras que se encuentren en esta lista.

Para modificar la lista, haz los cambios en el documento de texto y luego ven a esta misma ventana y haz clic sobre «Edit».

b. «Compare WordLists»

La herramienta «WordList» tiene una opción que permite comparar dos listas de palabras para extraer las palabras más frecuentes que difieren en una y otra lista. Este tipo de información sirve para comparar estilos o contenido. Por ejemplo, si se comparan dos listas una a partir de un corpus técnico sobre motores y la otra sobre textos literarios del siglo XX, aparecerán grandes diferencias terminológicas. O si se comparan listas hechas a partir textos del siglo XVI y del siglo XX, los resultados mostrarán diferencias de grafía.

Para usar esta opción, crea una lista de palabras de cada texto o grupo de textos que quieras comparar y guárdalas. Para guardar una lista ve al menú «File» de la ventana «WordList» y haz clic sobre la opción «Save» () o usa el atajo Ctrl + F2.