

Análisis de datos textuales por Excel VBA: LETRAS

ver. 2013.07.17

Este documento es un manual de uso para LETRAS.xlsm (en adelante LETRAS), conjunto de programas para el análisis de datos lingüísticos y filológicos. LETRAS está en desarrollo continuo, de modo que este mismo documento también cambia continuamente sin previo aviso. Para el detalle de modificaciones, véase la primera hoja (L) de LETRAS. Le rogamos al usuario que al notar inconvenientes o funciones mejorables, nos los comunique a través del correo electrónico puesto en la etiqueta [Top](Portada) de la interfaz de LETRAS. Le agradecemos su colaboración.

<http://lecture.ecc.u-tokyo.ac.jp/~cueda/gengo/>

Hiroto Ueda

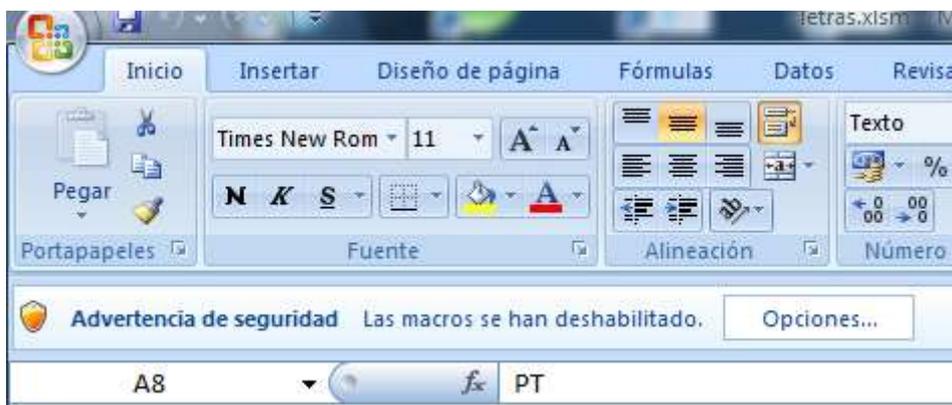
Colaboración:

Ana Isabel García (2013)

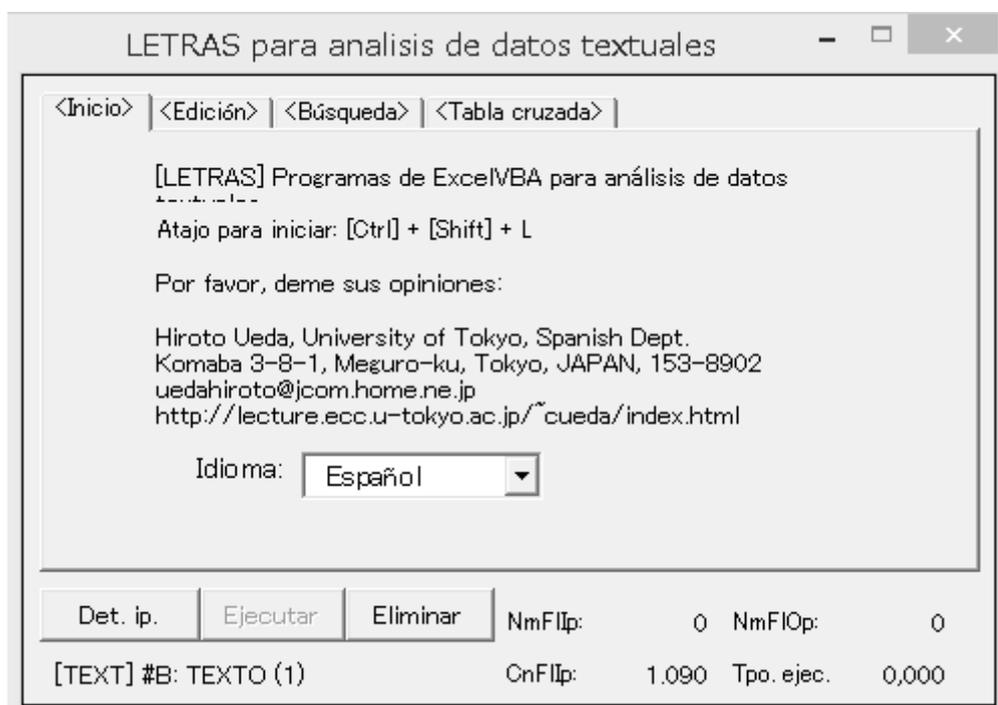
Leonardo Campillos (2014)

1. Inicio

Active LETRAS y habilite el programa de Macros contestando positivamente a «Advertencia de seguridad. Opciones...»:



Al pulsar [Ctrl] + [Shift] + L, aparecerá la interfaz de la figura siguiente. Esta figura muestra la etiqueta [Top] (Portada) arriba a la izquierda. Si el usuario desea analizar su Libro (Book) de Excel, después de abrir el Libro, puede analizarlo al activar la Macro con el atajo [Ctrl] + [Shift] + L.



En la barra de título, aparece el nombre del programa: «LETRAS for textual data analysis», seguido de los botones de minimización (_), maximización (□) y finalización (X). Para maximizar la interfaz minimizada o iniciarla de nuevo, utilice las teclas de atajo: [Ctrl] + [Shift] + L.

Las columnas de la primera hoja [L] presentan la información acerca de la lengua, el sistema, y el color de la lengua en la columna [A]. Cambie la asignación de idioma en [A6]=2, [A10]=0 y seleccione el color de fondo en [A8]:

English	Español	日本語
LETRAS for textual data analysis	LETRAS para análisis de datos textuales	LETRAS: テキストデータ分析
ver. 2014.03.05	«	«
Select language in the cell [A6]: English=1; Spanish=2; Japanese=3.	Seleccione el idioma en la celda [A6]: inglés = 1; español = 2; japonés = 3.	言語を選択してください。英語=1; スペイン語=2; 日本語=3 をセル[A6]に書き込んでください。
2	«	«
Select background color in the cell [A8].	Seleccione el color de fondo en la celda [A8].	背景色を[A8]に指定してください。
Background color Color de fondo 背景色	«	«
Select display (1) or no display (0) in the cell [A10].	Seleccione la opción de presentar (1) o no presentar (0) cero en la celda [A10].	ゼロ表示(1)・非表示(0)のオプションを[A10]に指定してください。
0		

Al iniciar de nuevo LETRAS la interfaz cambia al idioma español, con la coma (,) como separador decimal.

En esta interfaz se encuentran varias pestañas: [Portada], [Preparar], [Uni/Sep], [Ordenar], [Unifocal], [Bifocal] y [Etiqueta], las cuales iremos explicando en este documento.

La figura siguiente muestra la parte inicial de una muestra de datos textuales, que está en la hoja Tx1 de LETRAS.

ID	TEXTO	Tema	Línea
Madrid	A la recepción de un hotel madrileño llega un profesor extranjero para participar como conferenciante en un seminario sobre Nutrición organizado por una universidad de verano con sede en El Escorial. El profesor hablará con el conserje, pidiéndole información sobre los servicios del hotel, así como sobre posibles visitas turísticas por la región.	[A] Hotel	1
Madrid	– ¡Buenos días! Desearía una habitación individual para estar tres noches. ¿Qué precio tiene?	[A] Hotel	2

Se trata de unos ejemplos de conversaciones traducidas al español de cinco lugares distintos: Madrid, Sevilla, Ciudad de México, Lima y Buenos Aires. Los datos están organizados en dos partes de filas: la primera fila de títulos y las restantes de los datos, divididos según la clasificación de títulos. La organización de los datos es libre solo con una condición: los datos de ID (Identificador), si los hay, siempre en la primera columna [A].

En la parte inferior de la interfaz se encuentra los siguientes botones, casillas, etc.:

[Det. ip](Determinar input) actualiza el contenido de la casilla [Ip sheet] en el estado de la hoja actual. Cuando desee analizar los datos de la hoja con la columna de datos textuales seleccionada que está viendo, pulse este botón. Lo puede hacer directamente pulsando la parte correspondiente de la casilla de la lista.

[EJECUTAR] Se ejecuta. La primera ejecución puede durar un poco. Si se desea detener la ejecución, pulse la tecla [Esc] del teclado.

[Borrar] Borra sin confirmación la hoja seleccionada en Excel. Se pueden seleccionar las etiquetas de múltiples hojas (clicando con [Shift] (Mayúscula) o [Ctrl]).

[Nm.Fl.Ip] (Número procesado de filas de input) Se representa en cada momento el número de la línea en proceso.

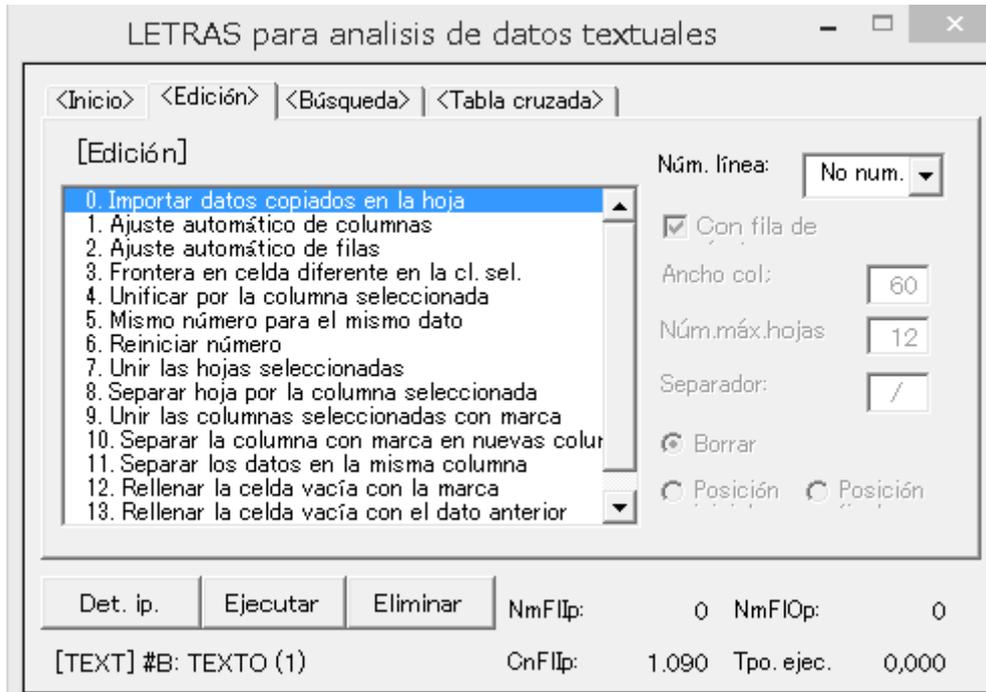
[Nm.Fl. Op] (Número procesado de filas de output)

[Cn.Fl.Ip]: (Cuenta de filas de input) Se representa la suma de líneas de input.

[Tpo.ejec.] (Tiempo de ejecución) Se presenta el tiempo de proceso en milisegundos.

2. Edición

Pulse la pestaña [Edición]. La función de esta pestaña es preparar los datos lingüísticos que serán objeto de los análisis textuales.



2.1. Importar

Vamos a preparar los datos (o «corpus») lingüísticos para someterlos a distintos procesamientos analíticos. Importemos nuestros datos a una hoja de Excel junto con la fila de títulos y la columna de los números identificadores de las líneas.

Por ejemplo vamos a copiar un texto como el siguiente. Por ejemplo, con “#B: Tema”, “Tema” se coloca en la columna B en la primera línea [B1]. El texto siempre va en la columna A. La anotación de #A: no es obligatoria. Cuando hay cambios en los datos en las columnas especificamos sus datos de la misma manera:

#A: Texto
#B: Tema
#C: Ciudad

#A: A la recepción de un hotel madrileño llega un profesor extranjero para participar como conferenciante en un seminario sobre Nutrición organizado por una universidad de verano con sede en El Escorial. El profesor hablará con el conserje, pidiéndole información sobre los servicios del hotel, así como sobre posibles visitas turísticas por la región.

#B: [A] Hotel

#C:(a) Madrid

#A: – ¡Buenos días! Desearía una habitación individual para estar tres noches. ¿Qué precio tiene?

El resultado es el siguiente:

Texto	Tema	Ciudad	Línea
A la recepción de un hotel madrileño llega un profesor extranjero para participar como conferenciante en un seminario sobre Nutrición organizado por una universidad de verano con sede en El Escorial. El profesor hablará con el conserje, pidiéndole información sobre los servicios del hotel, así como sobre posibles visitas turísticas por la región.	[A] Hotel	(a) Madrid	1
– ¡Buenos días! Desearía una habitación individual para estar tres noches. ¿Qué precio tiene?	[A] Hotel	(a) Madrid	2

Se puede seleccionar en [Núm.línea]: [No núm.] (No se agrega el número de fila), [Secuencial] (Se agrega número secuencial de fila), [Reinicio] (Se agrega número secuencial con reinicio en el punto de cambio del título).

2.2. Ajustar las columnas

Al seleccionar [Ajuste automático de columnas], la anchura de todas las columnas utilizadas se ajusta de acuerdo con los datos contenidos, y también con la máxima anchura asignada en [Tamaño:].

2.3. Linde en la celda diferente

La función del [Linde en la celda diferente] es trazar una línea en el borde de las celdas cuando aparece un dato diferente del anterior. Conviene utilizarlo después de la operación [Ordenar] (→4.).

2.4. Unir múltiples hojas

La función de [Unir las hojas seleccionadas] es unir las hojas múltiples seleccionadas en una hoja única. Seleccione los nombres de las hojas, con [Shift] ([Ctrl]) + clic en la casilla de abajo. En el ejemplo de arriba, hemos seleccionado [Tx1] y [Tx2]. La figura siguiente muestra la parte inicial de [Sample-2]:

Tx2

Text	Title-1	Title-2	Row
Varios matrimonios jóvenes, amigos todos ellos, van a reunirse para celebrar el cumpleaños de María, mujer de Pablo. Éste quiere prepararle una fiesta-sorpresa con ayuda de las otras parejas de amigos con los que, desde hace tiempo, se reúnen para divertirse. Por eso, es Pablo quien llama por teléfono a su amigo Marcos, casado con Mabel, para invitarles a la fiesta y hablar sobre la organización de la misma y sobre las cosas que llevará cada uno.	[F] Fiesta	(a) Madrid	461
¡¡Ring, ring!!	[F] Fiesta	(a) Madrid	462
- ¿Sí? ¿Dígame?	[F] Fiesta	(a) Madrid	463

Output:

Text	Title-1	Title-2	Row
A la recepción de un hotel madrileño llega un profesor extranjero para participar como conferenciante en un seminario sobre Nutrición organizado por una universidad de verano con sede en El Escorial. El profesor hablará con el conserje, pidiéndole información sobre los servicios del hotel, así como sobre posibles visitas turísticas por la región.	[A] Hotel	(a) Madrid	1

2.5. Separar hoja

La función de [Separate sheet by selected column] es separar la hoja seleccionada en varias hojas según el dato encontrado en la columna seleccionada. Como se crean tantas hojas de acuerdo con los datos de la columna, el máximo permitido está establecido por defecto en 12. Si desea aumentar el número puede cambiar la cifra.

Input: Hoja de output de 3.1. (Seleccionamos la columna [B].)

Output: Hoja de input de 3.1. (La última es Sample-2 de 3.1.)

2.6. Unir columnas

La función de [Unit selected columns with the mark] es unir las columnas múltiples seleccionadas con una marca asignada en [Mark]. En la figura siguiente las columnas [B] y [C] están seleccionadas

	A	B	C	D
1	Text	Title-1	Title-2	Row
2	A la recepción de un hotel madrileño llega un profesor extranjero para participar como conferenciante en un seminario sobre Nutrición organizado por una universidad de verano con sede en El Escorial. El profesor hablará con el conserje, pidiéndole información sobre los servicios del hotel, así como sobre posibles visitas turísticas por la región.	[A] Hotel	(a) Madrid	1
3	– ¡Buenos días! Desearía una habitación individual para estar tres noches. ¿Qué precio tiene?	[A] Hotel	(a) Madrid	2

Output:

Text	Title-1/Title-2	Row
A la recepción de un hotel madrileño llega un profesor extranjero para participar como conferenciante en un seminario sobre Nutrición organizado por una universidad de verano con sede en El Escorial. El profesor hablará con el conserje, pidiéndole información sobre los servicios del hotel, así como sobre posibles visitas turísticas por la región.	[A] Hotel/(a) Madrid	1
– ¡Buenos días! Desearía una habitación individual para estar tres noches. ¿Qué precio tiene?	[A] Hotel/(a) Madrid	2

2.7. Separar columna

La función de [Selected column separation by the mark] es separar la columna en varias, utilizando una marca de separación. Se crean nuevas columnas según el número de la misma marca encontrada en la celda.

Input: Output de 3.3.

Output: Input de 3.3.

2.8. Separar columna

La función de [Vertical column separation by the mark] es separar la columna de manera vertical utilizando la marca asignada. La marca puede ser borrada ([Delete]), o estar en la posición inicial ([Initial position]) o en la final ([Last position]). La figura siguiente muestra la marca coma (,) en la posición final:

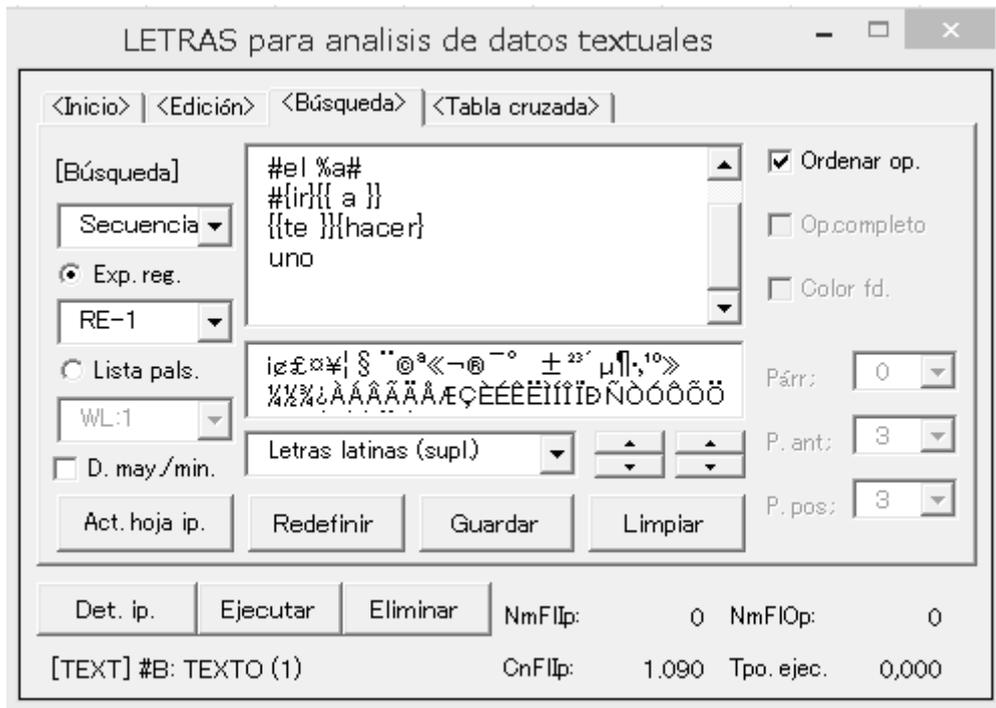
Text	Title-1/Title-2	Row
A la recepción de un hotel madrileño llega un profesor extranjero para participar como conferenciante en un seminario sobre Nutrición organizado por una universidad de verano con sede en El Escorial. El profesor hablará con el conserje, pidiéndole información sobre los servicios del hotel, así como sobre posibles visitas turísticas por la región.	[A] Hotel	1
A la recepción de un hotel madrileño llega un profesor extranjero para participar como conferenciante en un seminario sobre Nutrición organizado por una universidad de verano con sede en El Escorial. El profesor hablará con el conserje, pidiéndole información sobre los servicios del hotel, así como sobre posibles visitas turísticas por la región.	(a) Madrid	1

2.9. Ordenación reversa

Con el signo [r] se realiza la ordenación con el criterio reverso, que consiste en buscar el primer término en la última letra del texto, seguida de la penúltima, seguida de la tercera letra contando desde la posición final, y así sucesivamente. La figura siguiente muestra el signo de admiración, seguido de la letra 'e' (*padre*), 'l' (*chaval*), 'o' (*mano*), etc.:

Text	Title-1	Title-2	Row
– ¡Mamá! ¿Ya viste qué chida está esta sudadera? ¡Y me queda muy padre !	[H] Compras	(c) México	747
– ¡Hola, Juanito ! ¡Qué es de tu vida, chaval !	[I] Fútbol	(a) Madrid	835
– ¡Hola, Juanito! ¡Qué es de tu vida, mano !	[I] Fútbol	(c) México	876
– ¡Bueno, pues muchas gracias, mi amor !	[J] Despedida	(c) México	1006
¡¡Ring, ring!!	[F] Fiesta	(a) Madrid	462
¡¡Ring, ring!!	[F] Fiesta	(b) Sevilla	480

3. Búsqueda



En la pestaña <Búsqueda> se ofrecen funciones para distintos análisis textuales. Esta función permite realizar la búsqueda de un patrón lingüístico en un texto (p. ej. combinaciones de palabras, sílabas o caracteres). La función admite los siguientes tipos de búsqueda:

- [Secuencia] (Secuencia sin texto): obtiene únicamente los resultados del patrón buscado, sin el texto en que aparece.
- [Sec. en tx.](Secuencia en texto): obtiene los resultados del patrón de búsqueda en el propio texto (el patrón se resalta para su identificación).
- [Sec. fr. tx.](Secuencia fuera del texto): obtiene los resultados del patrón de búsqueda en dos columnas: en la primera, cada ocurrencia del patrón buscado, y en la segunda, el contexto en que aparece.
- [Sec. en contexto] (Secuencia en contexto): se trata de un modo de visualización de los resultados semejante al formato de concordancias o *Key-Word in Context* (KWIC). En la primera columna, aparece el contexto anterior al patrón buscado; en la segunda columna, cada

ocurrencia del patrón (resaltado en negrita); y en la tercera, el contexto inmediatamente posterior.

- [Incluir sec.] (Incluir secuencia): obtiene en una hoja nueva todos los contextos en que se registra el patrón buscado, excluyendo los contextos en que no aparece.
- [Excluir sec.] (Excluir secuencia): obtiene en una hoja nueva todos los contextos en que **no** aparece el patrón buscado, excluyendo aquellos en que se registra.
- [Reemplazar sec.] (Reemplazar secuencia): reemplaza secuencias según especifica el patrón.

Además, se puede seleccionar una de las dos fórmulas de tratamiento textual: [Expresión regular] y [Lista de palabras] (véase figura)

De las dos casillas de texto, colocadas verticalmente, la superior es para escribir la fórmula simplificada de búsqueda y reemplazo en expresión regular (en adelante, ER), La casilla inferior contiene varias letras especiales, que se selecciona o bien en la casilla pequeña con Unicode o en el grupo especificado en la casilla combo. De facto está seleccionado el grupo de «Letras latinas (supl[emento])». Se pueden utilizar estas letras copiándolas a la casilla superior. Los dos botones de triángulos, hacia arriba y abajo, son para cambiar el tamaño de las dos casillas de texto. A su derecha [en la versión actual] se encuentran otros dos botones con triángulos, que son para cambiar el tamaño de las fuentes de las letras utilizadas en las mismas casillas de texto. Al pulsar el botón de [Redefinir] aparece una caja de mensaje con expresiones regulares predefinidas. El botón [Guardar] es para guardar las expresiones en la hoja RE (Expresión regular) o en la WL (Lista de palabras). Al pulsar el botón de [Borrar], se borra todo el texto tanto de la casilla superior como de la inferior. Normalmente no se selecciona [D. may./min.] (búsqueda sensible a la distinción entre minúsculas y mayúsculas). Si desea distinguirlas, active la casilla. Con [En tx.] (Secuencia en texto), se puede seleccionar [Color fd.] (Color de fondo) y [Full op. (output)] (Output completo). El primero es para colorear el fondo de la celda reemplazada o buscada y el segundo es para reproducir todas las filas, inclusive las filas no correspondientes. Al seleccionar [Full op.], devuelve solo los datos reemplazados y/o buscados.

3.1. Secuencia sin texto

Mediante la función [F.w.c.: Focus without context], se reúnen los números identificadores de las filas que contienen los datos buscados, sin su contexto, de la manera siguiente (ER: %d#):

Fim. búsqueda	Título:1	Título:2	Hoja	Fila
antigüedad	[C] Funcionar	(a) Madrid	Tx1	178
barbaridad	[B] Camino	(b) Sevilla	Tx1	123
barbaridad	[B] Camino	(e) B.A.	Tx1	165
barbaridad	[E] Consultori	(a) Madrid	Tx1	328
bondad	[H] Compras	(a) Madrid	Tx2	238

También es posible realizar la ordenación según el orden de ocurrencias en el texto (ER: %d#):

universidad	[A] Hotel	(a) Madrid	1
usted	[A] Hotel	(a) Madrid	3
ciudad	[A] Hotel	(a) Madrid	10
ciudad	[A] Hotel	(a) Madrid	11

De esta tabla se puede realizar la clasificación cruzada con la Función [Pivot-table] de Excel.

■ Análisis secuencial

Las grafías y los fonemas están ordenados de acuerdo con unas determinadas reglas específicas de la lengua. Para estudiar sus reglas, uno de los métodos utilizados es el «análisis secuencial». Empecemos con una secuencia: «la escuela». Con la fórmula de expresión regular «..», que significa dos letras seguidas, tanto utilizadas como no utilizadas en la palabra, obtenemos el resultado siguiente:

Input:

Row	Text
1	la escuela

Output [Exp. reg.: ..; orden de ocurrencia]:

Row	Key word
1	La

- 1 E
- 1 Sc
- 1 Ue
- 1 La

De esta manera, el programa reproduce el resultado de la búsqueda de dos letras secuenciales incluyendo el espacio sucesivamente. Por otra parte, nos interesan otras posibilidades de la secuencia de dos unidades: es, cue, el. Para la búsqueda preparamos las tres fórmulas siguientes:

```
..
(.+)=>#$1
..
```

La primera reproduce las dos letras como en el caso anterior. La segunda fórmula significa que más de una ocurrencia de cualquier letra se convierte en «#» más la referencia anterior, que quiere decir la reproducción de la secuencia correspondiente a «(...)», en este caso a «.+)», es decir, la totalidad de las letras en la celda: «#la escuela». Luego, de nuevo «..» reproduce ahora «#l», «a », «es»,... El resultado es:

Output [Exp. reg.: ..; (.+)=>#\$1; ..; orden de ocurrencia):

Row	Key word
1	la
1	e
1	sc
1	ue
1	la
1	#l
1	a
1	es
1	cu
1	el

Mediante esta tabla, junto con sus clasificadores, podemos realizar una clasificación cruzada de sus frecuencias. El cuadro siguiente muestra la frecuencia relativa de «f» más una letra en los documentos notariales emitidos en la Castilla del siglo XII al XVII. Se observan unas concentraciones de curva de campana, con subidas y bajadas en la frecuencia relativa. La data cronológica

de la cumbre no coincide y la forma de la curva también varía según la secuencia tratada:

年代	<fa>	<fc>	<fe>	<ff>	<fh>	<fi>	<fj>	<fl>	<fm>	<fn>	<fo>	<fr>	<ft>	<fu>	<fy>	<fz>
1120	27	0	6	0	0	21	0	0	0	0	25	7	0	15	0	0
1160	34	0	6	0	0	27	0	0	0	0	13	7	0	5	0	0
1200	48	0	0	0	0	29	0	28	0	0	9	18	0	20	0	0
1220	16	0	18	0	0	16	0	0	0	0	0	10	0	0	0	0
1240	48	0	63	0	0	70	18	0	0	147	8	27	10	28	0	0
1260	51	24	65	52	0	104	0	28	11	119	51	20	0	43	0	0
1280	82	0	66	134	0	102	0	4	0	0	30	31	16	71	0	0
1300	56	0	47	309	0	91	8	0	7	0	9	21	0	48	0	0
1320	48	0	39	485	0	68	0	32	0	0	5	8	4	41	6	0
1340	16	17	20	172	0	24	0	0	0	0	0	11	10	21	0	33
1360	48	8	47	238	0	75	0	3	8	0	34	49	10	55	0	0
1380	72	9	65	122	0	92	10	0	17	0	17	68	65	62	0	0
1400	124	0	129	83	0	120	57	17	28	10	63	133	50	128	0	0
1420	163	16	197	35	0	163	27	39	10	0	51	98	59	168	24	20
1440	167	44	105	15	13	113	0	29	7	6	44	64	22	122	0	0
1460	133	50	113	71	0	82	0	8	0	0	69	69	17	130	26	0
1480	148	6	147	7	80	102	0	106	0	0	49	144	0	121	6	0
1500	135	0	94	22	4	102	0	12	0	6	36	96	0	138	41	44
1520	124	7	94	4	51	104	0	36	6	11	60	145	12	178	12	0
1540	52	0	63	5	34	86	10	37	71	39	40	113	0	120	68	0
1560	15	0	31	0	31	36	0	28	17	0	34	34	0	109	0	0
1580	22	0	51	1	50	65	15	11	26	0	43	50	0	99	13	0
1600	48	0	45	8	16	67	14	26	0	0	69	49	0	89	0	0
1620	39	0	47	0	25	39	0	0	20	0	35	24	0	46	0	0
1640	44	0	29	4	8	73	14	5	0	0	30	73	0	93	0	0
1660	26	0	40	0	25	39	0	33	0	0	30	49	0	40	0	0
1680	19	0	24	0	21	18	0	0	0	0	0	23	0	41	0	0
1700	20	0	40	3	41	47	0	0	0	0	37	95	0	52	0	0

Para obtener datos de secuencia de tres letras se utilizan las fórmulas siguientes:

...
 (.+)=>#\$1
 ...
 (.+)=>#\$1
 ...

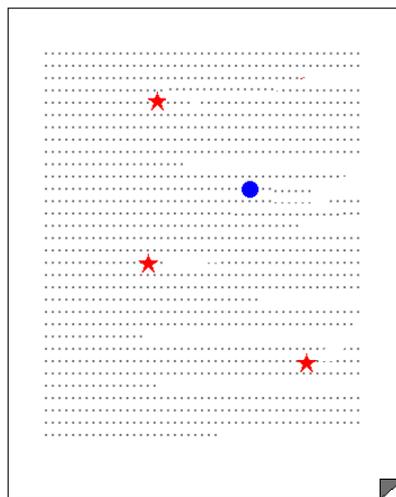
Para obtener datos de secuencia de dos palabras se utilizan las fórmulas siguientes, con el signo «&» que es una secuencia de una o más de una letras continuas:

&&
 (.+)=>x \$1
 &&

*OJO: hay que escribir espacio entre & y & para que funcione.

3.2. Secuencia en el texto

Al analizar datos lingüísticos y filológicos, con frecuencia buscamos en el texto unas secuencias de palabras, unas palabras o una parte de las palabras en cuestión. Con LETRAS podemos buscarlas utilizando expresiones regulares (ER) o una lista de palabras y encontrarlas con una marca especial. Para entrenarnos en el manejo de ER, de esta sección en adelante, nos concentraremos en el uso de las mismas con una fórmula simple, dejando de lado por el momento la lista de palabras (que puede ser abundante). La figura siguiente muestra una imagen esquemática de la función [En tx.] (Secuencia en el texto):



¥c¥c (consonante + consonante)

*OJO: para escribir ¥, pulse ALT GR + la tecla para escribir ° a o barra lateral.

A la rece{**pc**}ión de un hotel ma{**dr**}ileño {**ll**}ega un {**pr**}ofesor e{**xt**}ra{**nj**}ero para pa{**rt**}icipar como co{**nf**}ere{**nc**}ia{**nt**}e en un seminario so{**br**}e Nu{**tr**}ición o{**rg**}anizado por una unive{**rs**}idad de verano con sede en El E{**sc**}orial.

Al seleccionar [Color fd.] (Color de fondo), se marca con el mismo color que el título de Texto de la hoja de input, previamente asignado.

#r% (Seleccionado [Color fd.]

Texto	Título:1	Título:2	Hoja	Fila
A la {*recepción*} de un hotel madrileño llega un profesor extranjero para participar como conferenciante en un seminario sobre Nutrición organizado por una universidad de verano con sede en El Escorial. El profesor hablará con el conserje, pidiéndole información sobre los servicios del hotel, así como sobre posibles visitas turísticas por la {*región*}.	[A] Hotel	(a) Madrid	Tx1	2
– ¡Buenos días! Desearía una habitación individual para estar tres noches. ¿Qué precio tiene?	[A] Hotel	(a) Madrid	Tx1	3

La función de reemplazo de Excel está limitada. Con LETRAS la hemos ampliado con fórmulas de ER y listas de palabras para obtener la forma adecuada para análisis lingüísticos y filológicos. Para realizar la operación de reemplazo, utilizamos la fórmula de búsqueda en forma de "[A]=>[B]", con la cual la secuencia correspondiente a la fórmula [A] se reemplaza por la forma de [B]. Es imprescindible el uso de dos signos, [=] de igualdad y [>] de desigualdad superior. Aplicamos por ejemplo la fórmula «de=>DE» al texto siguiente:

A la recepción de un hotel madrileño llega un profesor extranjero para participar como conferenciante en un seminario sobre Nutrición organizado por una universidad de verano con sede en El Escorial.

Output-1:

A la recepción **DE** un hotel madrileño llega un profesor extranjero para participar como conferenciante en un seminario sobre Nutrición organizado por una universidad **DE** verano con se**DE** en El Escorial.

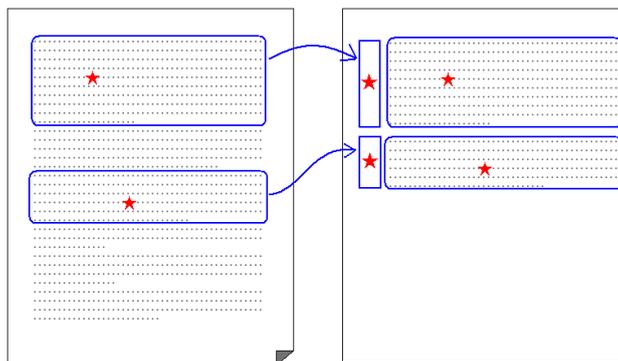
Para cambiar solo la preposición de, a exclusión de una parte de palabra como «sede», se utiliza el signo de linde léxico «#» en la parte izquierda de la fórmula: «#de#=>DE»:

Output-2:

A la recepción **DE** un hotel madrileño llega un profesor extranjero para participar como conferenciante en un seminario sobre Nutrición organizado por una universidad **DE** verano con sede en El Escorial.

3.3. Secuencia fuera del texto

Con [F.o.c. (Focus out of context)] (Secuencia fuera del contexto) se realiza la búsqueda con un resultado donde aparecen tanto la palabra clave como su contexto en distintas columnas, de la manera siguiente:



Output [Exp. reg.: &[ai]d[ao]#]

Fm. búsqueda	Texto	Título:1	Título:2	Hoja	Fila
acelerado	- ¡Hombre, José! ¿Qué tal? ¿Cómo te va? Que vas <i>{*acelerado*}</i> , para un poco, oye, ¿tú sabes dónde está <i>***</i> ?	[B] Camino	(a) Madrid	Tx1	98
acompañado	- También es fácil, la cosa es conseguir un buen pescado, puede ser corvina, lenguado, cojinova o también sale bien de toyo o pejerrey... Tiene que ser blanco, ah. Si lo quieres mixto, le puedes agregar mariscos: pulpo, camarones o langostinos, conchas... Entonces, por un kilo de pescado o pescado y mariscos se le echa una taza y un poquito más de jugo de limón; picas el pescado en daditos de un centímetro más o menos, le echas el limón, una cebolla grande picada a la pluma, sal y ají al gusto y lo dejas unos veinte minutos. Lo sirves rodeado de una hoja de lechuga y <i>{*acompañado*}</i> con un trozo de camote y otro de choclo . Si no te gustan las cosas picantes, pones el ají aparte. Eso es todo.	[J] Despedida	(d) Lima	Tx2	588

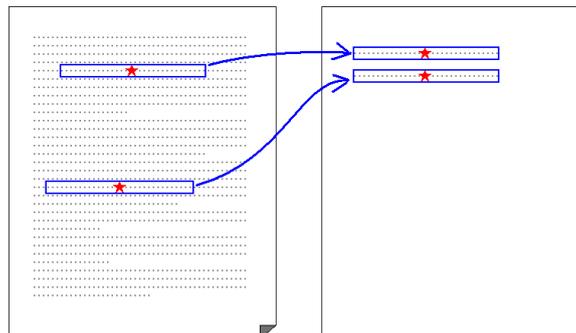
Opción: Agregar párrafos

Se puede agregar con [Add para (paragraphs)] cuantos párrafos se desee. Lo normal es agregar uno o dos, si son necesarios. La figura siguiente muestra el resultado con un párrafo agregado.

Film. búsqueda	Texto
acelerado	En el centro de Madrid se tropiezan por casualidad dos amigos. Uno de ellos quiere ir al supermercado de ***, para hacer unas compras pero no sabe cómo llegar. El otro –que llega tarde a la oficina– le dice por dónde tiene que ir. \$\$ – ¡Hombre, José! ¿Qué tal? ¿Cómo te va? Que vas {*acelerado*}, para un poco, oye, ¿tú sabes dónde está ***? \$\$ – Hombre, claro. Mira, tiras a la derecha por la calle Burgos y después por San Fernando. Llegas a Cuatro Caminos y allí, en la rotonda, giras a la izquierda en dirección a la Plaza de Toros. Sigues recto hasta el parque de la Marga. Ahí ya tomas la autovía dirección Burgos y enseguida verás la indicación de...

3.4. Secuencia en su contexto

Al seleccionar [Sec. en contexto] (Secuencia en contexto) obtenemos los datos de Secuencia con el contexto anterior a la izquierda y el posterior a la derecha. Es útil para analizar el entorno sintáctico de las formas en cuestión:



Output [ER:&[ai]d[oa]s?#]

Texto	Foco	Texto
ofe de Estadísticas, ¿viste qué buen tipo? Sus clases no son	aburridas	. Uno entiende todo y en los parciales es más flexible...
– ¡Hombre, José! ¿Qué tal? ¿Cómo te va? Que vas	acelerado	, para un poco, oye, ¿tú sabes dónde está ***?
veinte minutos. Lo sirves rodeado de una hoja de lechuga y	acompañado	con un trozo de camote y otro de choclo . Si no te gustan la:
rio son las clases de Computación. Tiene un programa más	adecuado	a nuestros gustos, ¿no te parece?
– Sí señor, estamos	afiliados	a varios programas de las líneas aéreas. Ahí está la lista. Si

La ER«&[ai]d[oa]s?#» utilizada en esta ocasión es para buscar la secuencia de más de cero veces de la letra utilizada en la palabra (%), más 'a' o 'i' ([ai]), más cero o una vez de 's' (s?), con la cual se buscan las palabras terminadas en la forma correspondiente al participio pasado tanto en singular, como en plural.

■ Sinónimos

En la versión vulgata del «Génesis» encontramos tres formas verbales: «ait», «dixit», «inquit», cuyo significado corresponde a «(él) dijo». Si buscamos su origen hebreo notamos que estas tres formas latinas proceden de un mismo verbo. Nos preguntamos si son puras variantes intercambiables, que el traductor Jerónimo seleccionó al azar, o si se pueden encontrar algunas razones del uso de cada verbo.

Al analizar todos los capítulos del «Génesis», con la función de [KWIC], encontramos el uso de «ait» al final de oración, aunque la misma forma se encuentra también en otros sitios. Parece ser que «ait» se utiliza como una añadidura «..., dijo él».

Egressus ad eos Loth post tergum accludens ostium	ait	
Jacob vero convocata omni domo sua	ait	Abjicite deos alienos qui in medio vestri sunt: et mandamina ac man-
ae ad vesperam Jacob de agro egressa est in occursum ejus Lia: et	ait	Ad me inquit intrabis. Quia mercede conduxi te pro mandragoris f
uae. Pariebant omnes oves varios foetus. Quando vero e contrario	ait	Alba quaeque accipies pro mercede. Omnes greges alba pepererit
itis Esau sermonibus patris: irrugit clamore magno. Et consternatus	ait	Benedic etiam et mihi pater mi.
benedixitque ei et	ait	Benedictus Abram Deo excelsis: qui creavit caelum et terram:
Quos cum vidisset	ait	Castra Dei sunt haec. Et appellavit nomen loci illius Manaim id est

El verbo «inquit», en cambio, es utilizado dentro de la oración, como una inserción «..., dijo él, ...»:

At ille. Servus	inquit	Abraam sum:
Ad quos ille clementer. Accedite	inquit	ad me. Et cum accessissent prope. Ego sum ait Joseph frater veste
At ille. Septem	inquit	agnas accipies de manu mea: ut sint mihi in testimoniam: quoniam e
Cui Judas. Quid respondebimus	inquit	domino meo? Vel quid loquemur aut juste poterimus obtendere? E
espondit. Filii mei sunt quos donavit mihi Deus in hoc loco. Adduc	inquit	eos ad me: ut benedicam illis.
unt paululum. Tunc Joseph accessit dispensatore domus ait. Surge	inquit	et persequere viros. Et apprehensis dicit. Quare reddidistis manum

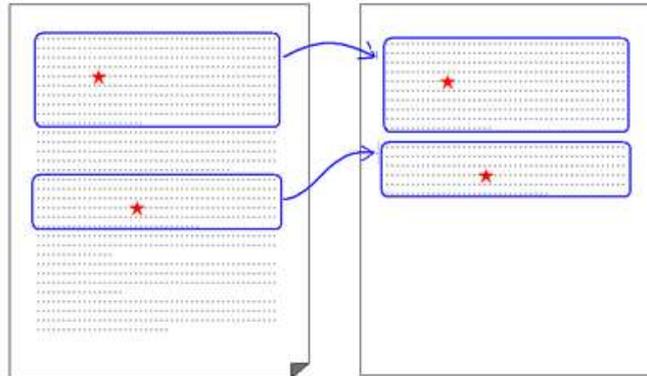
«Dixit», por otra parte, se usa preferiblemente al inicio de la oración, lo que no implica que siempre al inicio de la oración se utilice «dixit». Suponemos que cuando se trata de una afirmación importante se utiliza esta forma, cuya forma correspondiente al español es «lo que dijo él es que ...»:

	Dixit	que Esau. Quatenam sunt istae turmae quas obviam habui? Respon-
	dixit	que fratribus suis. Afferte lapides. Qui congregantes fecerunt turmas
	Dixit	que illi Isaac. Quis enim es tu? Qui respondit. Ego sum filius tuus pr
	Bene dixit	que illis Deus: et ait. Crescite et multiplicamini et replete terram et n
ergo semper Esau Jacob pro benedictione qua benedixerat ei pater.	Dixit	que in corde suo. Venient dies luctus patris mei: et occidam Jacob f

De las tres formas, el español (y otras lenguas románicas) heredó la última.

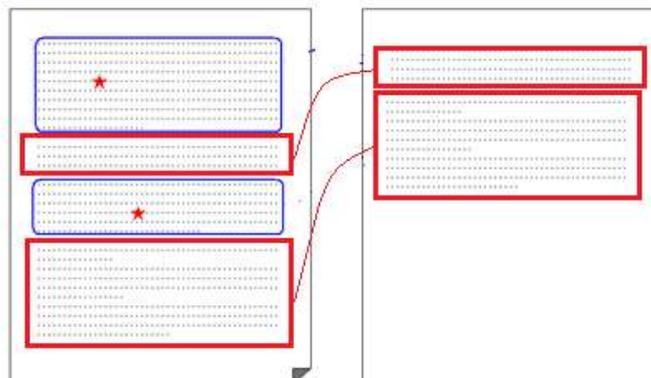
3.5. Texto con la secuencia

La función de [Texto con sec.] es extraer las filas que incluyen los datos buscados y copiarlas en una hoja, de la manera siguiente:



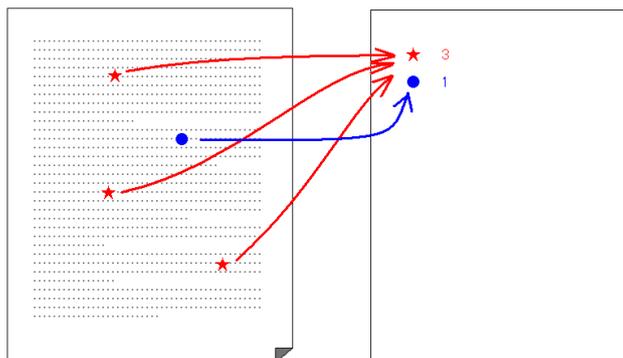
3.6. Texto sin la Secuencia

Al seleccionar [Texto sin sec.], se copian las filas que no incluyen los datos buscados, de la manera siguiente:



4. Tabla cruzada

En la pestaña de <Tabla cruzada> se calcula la frecuencia de las secuencias. También es posible elaborar listados de secuencias correspondientes en la tabla cruzada.



Hay dos opciones de ordenación: orden alfabético y orden de aparición. Se puede seleccionar una columna, que es la clave de la clasificación. Al seleccionar [Sch. fm.] (Search formula) se calcula la frecuencia por unidad de la fórmula de búsqueda. Al seleccionar [Secuencia], la unidad de frecuencia va a ser las formas o patrones (el “Secuencia”) correspondientes a la fórmula de búsqueda. Bajo este menú, en la casilla [Cl.] (Clasificar), con [Total], se calcula la totalidad de frecuencias, sin hacer la clasificación. [Hoja] se utiliza para calcular la frecuencia de la forma en cada hoja, a la cual siguen las letras alfabéticas, que son columnas (A, B, C, ...), una de las cuales es la columna clave de la clasificación. En la casilla [It.] (Intervalo), hay opciones de [No intervalo], [10], [20], [25], ..., que van a ser cifras de intervalo cuando se trata de números enteros. Al elegir 10, por ejemplo, la clasificación de cifras va a ser de 0 a 9, de 10 a 19, ..., y así sucesivamente. La casilla [Fr] se utiliza para seleccionar uno de los tres modos de frecuencias: [Frec. (frecuencia) absoluta], [Frec. (frecuencia) relativa] y [Por mil palabras]. La última frecuencia se refiere a la frecuencia absoluta dividida por el total de palabras en la misma categoría, multiplicado por mil. Las siguientes figuras son resultados de los cálculos de las hojas [TX1] y [TX2]. Se explican los distintos resultados según cada opción.

Output-1 [P(atrón): &ando#; &iendo#; Vertical: F(orma).Total; Horizontal: C(lase).Todo]: al seleccionar F.Total, se obtiene el recuento total tras aplicar las fórmulas de patrón, sumando los resultados de todos los patrones de búsqueda.

&ando# □ &iendo#	Todo
Total	153

Output-2 [P(atrón): &ando#; &iendo#; Vertical: F(orma).Patón; Horizontal: C(lase).Todo]: con la opción Fm: fml se obtiene el recuento total, pero desglosando los recuentos de cada fórmula o patrón de búsqueda.

&ando# □ &iendo#	Todo
&iendo#	24
&ando#	129
Total	153

Output-3[P(atrón): &ando#; &iendo#; Vertical: F(orma).Secuencia; Horizontal: C(lase).Todo]: mediante Fm: Foc se extraen los recuentos de cada una de las palabras diferentes en que se registra el patrón de búsqueda.

&ando# □ &iendo#	Todo
saliendo	1
fernando	5
cuando	56
cogiendo	1
fijando	2
conversando	1

Output-4 [P(atrón): &ando#; &iendo#; Vertical: F(orma).Secuencia; Horizontal: C(lase).ID]

&ando# □ &iendo#	Madrid	Sevilla	México	Lima	B.A.	Total
saliendo			1			1
fernando	1	3		1		5
cuando	12	14	15	6	9	56
cogiendo		1				1
fijando		1			1	2

Output-5 [P(atrón): &ando#; &iendo#; Vertical: F(orma).Secuencia; Horizontal: C(lase).Línea, Int(ervalo): 100]

&ando# □ &iendo#	0	100	200	300	400	500	600	700	800	900	1000	Total
saliendo	1											1
fernando	1		1	3								5
cuando		7	12	9	5	2	1	2	8	7	3	56
cogiendo		1										1
fijando		2										2

Output-6 [P(atrón): &ando#; &iendo#; Vertical: F(orma).Patrón; Horizontal: C(lase).ID] F.R.V. (Frecuencia relativa vertical)

&ando# □ &iendo#	Madrid	Sevilla	México	Lima	B.A.
&iendo#	19,4%	6,3%	19,0%	9,5%	22,7%
&ando#	80,6%	93,8%	81,0%	90,5%	77,3%

Output-7 [P(atrón): &ando#; &iendo#; Vertical: F(orma).Patrón; Horizontal: C(lase).ID] P.M.P. (Por mil palabras)

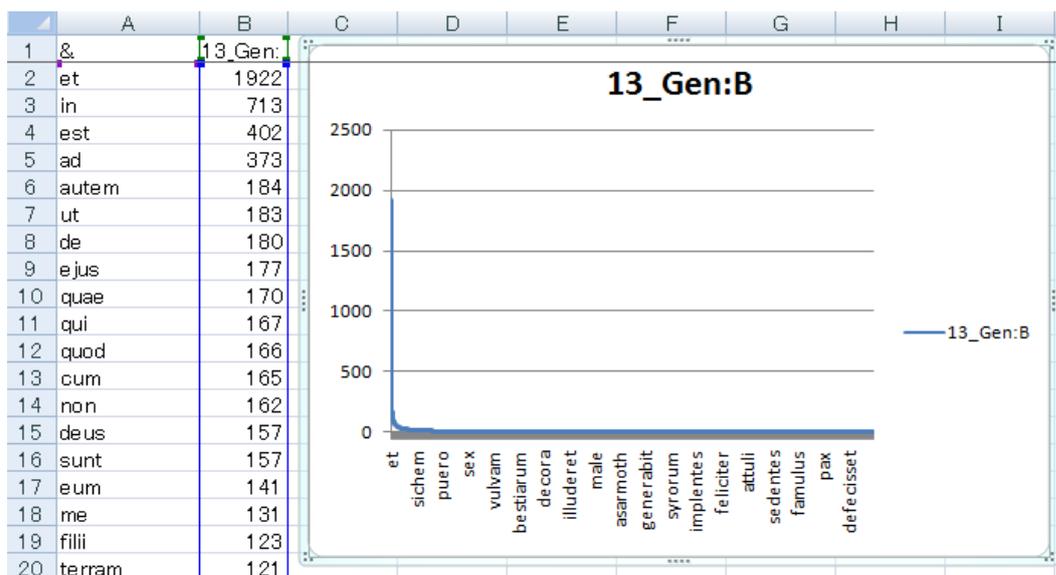
&ando# □ &iendo#	Madrid	Sevilla	México	Lima	B.A.
&iendo#	1,272	0,382	1,247	0,480	0,948
&ando#	5,271	5,737	5,302	4,563	3,224

Output-8 [P(atrón): &ando#; &iendo#; Vertical: F(orma).Patrón; Horizontal: C(lase).ID] Lista de secuencias

&ando# &iendo#	Madrid	Sevilla	México	Lima	B.A.
&iendo#	viniendo-173	cogiendo-116	saliendo-50	corriendo-432	haciendo-157
&iendo#	insistiendo-244	haciendo-580	viniendo-203	viendo-529	compartiendo-227
&iendo#	doliendo-337		insistiendo-274		sufriendo-316
&iendo#	haciendo-556		haciendo-377		viviendo-451
&iendo#	viendo-561		haciendo-603		viendo-669
&iendo#	manteniendo-570		manteniendo-617		
&iendo#	añadiendo-948		subiendo-624		
&iendo#			haciendo-627		
&ando#	Fernando-99	cuando-113	cambiando-208	conversando-143	fijando-155
&ando#	Cuando-168	fijando-119	cuando-208	mirando-147	Cuando-161

■ Distribución de forma L

Al calcular la frecuencia de las palabras encontradas en la versión latina del *Génesis* con la función de Distribución, y elaborar el gráfico, obtenemos la figura siguiente:



La palabra más frecuente es «et» ('y') con 1922 ocurrencias, a la cual siguen 713 de «in» ('en'), 402 de «est», Se nota la curva descendiente muy pronunciada. En realidad se hallan muy pocas palabras de alta frecuencia y, por otra parte, las palabras de poca frecuencia son numerosas. Hemos encontrado 3.480 palabras de una sola ocurrencia (frecuencia), lo que se representa en el gráfico de la figura anterior. Este tipo de curva se denomina la «distribución de forma L», y casi en todos los textos de cierta longitud, presenta la misma distribución. En esta ocasión, no hemos hecho una operación llamada «lematización» que consiste en

agrupar todas las formas flexivas en una forma representativa (→7.). Al realizar la lematización, sin embargo, también se presenta la misma tendencia de curva.

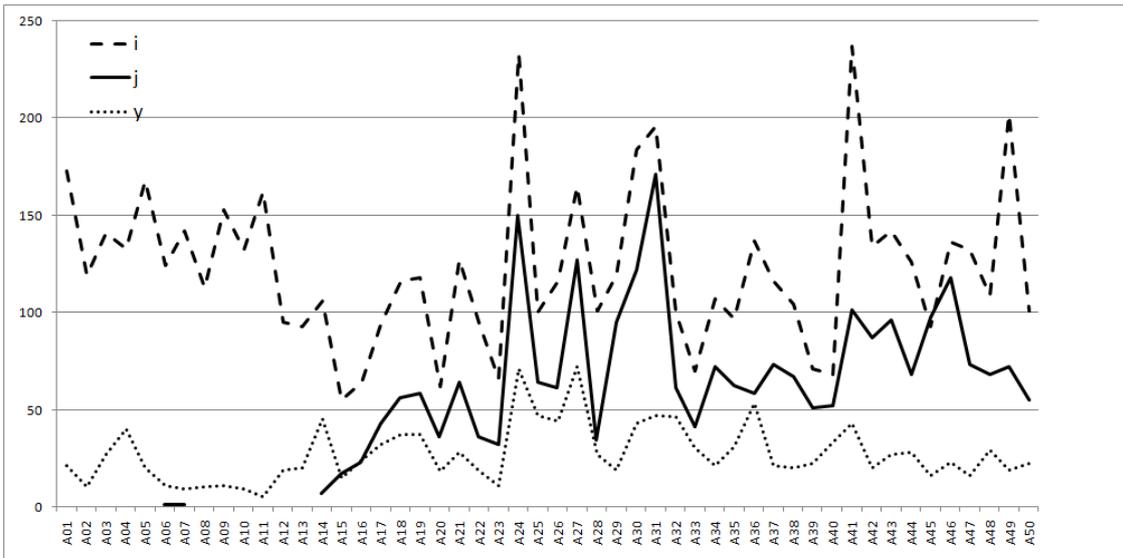
La lengua en general posee unos pocos vocablos funcionales de alta frecuencia como conjunciones, preposiciones, pronombres, etc. y gran cantidad de léxico de contenido como sustantivos, adjetivos, verbos, adverbios, etc., lo cual presenta aparentemente un desequilibrio de frecuencia. Pensamos, no obstante, que este tipo de distribución es razonable desde el punto de vista funcional. Si la frecuencia de todos los léxicos fuera igual, el usuario tendría la dificultad de tener que aprender numerosos vocablos funcionales, que en realidad no se utilizarían con frecuencia, y el mismo usuario se encontraría continuamente con vocablos de contenido desconocidos. De esta manera la lengua no podría funcionar como medio de comunicación.

Por otra parte, al buscar el léxico de contenido más frecuente, encontramos en efecto «Deus» (Dios) con 157 ocurrencias, mucho más frecuente que los pronombres, de hecho, ocupa el decimoquinto lugar. Tal léxico se denomina como palabra característica y debemos prestarle su debida atención.

■ **Lista contrastiva de frecuencias**

El resultado de la [Distribución] se presenta como «lista contrastiva de frecuencias», objeto de múltiples análisis estadísticos. La observación misma de la lista contrastiva también es interesante para estudiar el texto.

La figura siguiente muestra la distribución de las tres grafías de la vocal palatal, «i», «j» e «y», que se encuentran en 50 capítulos del *Génesis* de una biblia romanceada (traducida a la lengua hablada en la Edad Media):



Al leer el texto, notamos que estas tres letras se utilizan en el mismo léxico: *mi*, *mj*, *my*; *amigo*, *amjgo*, etc. Este tipo de distribución se llama «variación libre», y parece ser que estas letras eran variantes del mismo grafema.

Por otra parte, al elaborar una tabla de frecuencias, notamos que la letra «j» no aparece hasta la mitad del capítulo 14. Si su uso estuviera en variación libre, debería aparecer en cualquier sitio. Se observa el cambio de cuaderno en el mismo capítulo y se nota la diferencia de mano del escriba. Por ello pensamos que en el texto hubo una variación no colectiva sino individual del uso de estas letras medievales.

5. [Bis] Expresión regular

La expresión regular (ER en adelante) es útil para el procesamiento de textos. Sus reglas son simples y con la práctica se aprende fácilmente. Compruebe sus funciones con varios experimentos. LETRAS posee todas las funciones de expresión regular en su versión general.

5.1. Signos especiales

¥t Tabulación.

¥b Frontera léxica inglesa

¥B Cualquier posición fuera de frontera léxica inglesa, es decir, «complemento» de ¥b (en adelante utilizamos el término «complemento»)

¥n Cambio de línea

*OJO: para escribir ¥, pulse ALT GR + la tecla para escribir ° a o barra lateral.

Input:

The Universal Declaration of Human Rights Article 1. All human beings are born free and equal in dignity and rights. They are endowed with reason and conscience and should act towards one another in a spirit of brotherhood.

Output: ER=¥bin¥b («in» entre dos fronteras léxicas inglesas)

The Universal Declaration of Human Rights Article 1. All human beings are born free and equal {*in*} dignity and rights. They are endowed with reason and conscience and should act towards one another {*in*} a spirit of brotherhood.

La secuencia «in» en «beings» no corresponde a la ER por no estar entre dos fronteras léxicas.

5.2. Grupo de letras

[xyz] Cualquier letra que está dentro de los corchetes [...]. Se puede utilizar el guion para abarcar las letras que se encuentran entre dos letras, por ejemplo [a-d] significa lo mismo que [abcd].

[^xyz] Complemento de [xyz].

.(punto) Cualquier letra excepto el signo de cambio de línea (¥n).

¥w	Cualquier carácter del alfabeto inglés más numeral más un signo de raya inferior (_). Es igual a [a-z0-9_].
¥W	Complemento de ¥w
¥d	Cualquier numeral [0-9].
¥D	Complemento de ¥d
¥s	Espacios [¥t¥r¥n¥f]
¥S	Complemento de ¥s

Input:

All human beings are born free and equal in dignity and rights.

Output: ER=[e-h]

All {*h*}uman b{*e*}in{*g*}s ar{*e*} born {*f*}r{*e*}{*e*} and {*e*}qual in di{*g*}nity and ri{*g*}{*h*}ts.

Output: ER=[^a-v]

All {* *}human {* *}beings {* *}are {* *}born {* *}free {* *}and {* *}equal {* *}in {* *}dignit{*y*}{* *}and {* *}rights {*. *}

5.3. Selección, agrupamiento, repetición

- (X|Y) Agrupamiento de expresiones regulares X e Y. Por ejemplo, «(ment|tion|sion)» corresponde a una de las tres formas agrupadas.
- + Una o más de una veces de repetición. Por ejemplo, «a+h» corresponde a *ah, aah, aaah*, etc.
- * Cero o más de cero veces de repetición. Por ejemplo, a*h corresponde a *h, ah, aah, aaah*, etc.
- {n} n veces de repetición. Por ejemplo, a{3}h corresponde a *aaah*.
- {n,} n veces o más de n veces de repetición. Por ejemplo, a{3, }h corresponde a *aaah, aaaah, aaaaaah*, etc.
- {n,p} n veces o más de n veces y menos de p veces o menos de p veces de repetición. Por ejemplo, a{3, 4}h corresponde a *aaah, aaaah*.

Output. ER=(free|equal)

All human beings are born {*free*} and {*equal*} in dignity and rights.

Output. ER=[e-h]+

All {*h*}uman b{*e*}in{*g*}s ar{*e*} born {*f*}r{*ee*} and {*e*}qual in di{*g*}nity and ri{*gh*}ts.

Output. ER=[aeoiu]{2}

All human b{*ei*}ngs are born fr{*ee*} and eq{*ua*}l in dignity and rights.

Output. ER=[^aeoiu]{2,}

A{*ll h*}uma{*n b*}ei{*ngs *}are{* b*}o{*rn fr*}ee a{*nd *}equa{*l *}i{*n d*}i{*gn*}i{*ty *}a{*nd r*}i{*ghts.*}

5.4. Escape de letras

Para buscar los signos especiales, (,), [,], {, }, ?, !, .(punto), +, *, |, ¥. se pone «¥>> delante para escapar de su función especial.

Input

¿Cómo está usted?

Output. ER=¥?

¿Cómo está usted?

5.5. Referencia posterior

Utilizando la función de [back reference] «referencia posterior» de ER, podemos reproducir la forma correspondiente a la ER de búsqueda.

¥n Significa n veces de repetición de la ER anterior. Por ejemplo (.)¥1 corresponde a la repetición de cualquier letra, inclusive el espacio.

\$n Corresponde a n-ésima ER en paréntesis (...). Por ejemplo «(¥bc.*)=>[[\$1]]» significa que se busca la palabra que empieza por «c» y lo reproduce entre [[...]].

Output. ER=(.)¥1

A{*ll*} human beings are born fr{*ee*} and equal in dignity and rights.

Output. ER=#(c%)=>[[\$1]]

Esta expresión regular reemplaza toda palabra que comienza por *c* añadiendo dos corchetes antes y después de la palabra. Para aplicar expresiones regulares de reemplazo, es preciso seleccionar la opción “8. Reemplazar” en el menú “Análisis unifocal”.

Rumi: Hola, profesor.
Prof. Rubio: Buenos días.
Rumi: Buenos días. Nos encontramos ahora en la Universidad [[Complutense]] de Madrid. ¿Dónde nos vamos ahora?
Prof.: Bueno, vamos a iniciar hoy el [[Camino]] del [[Cid]], la primera parte.

Input:

Otro(22)ssí mando que los menestrales non echen suerte en el judgado por seer juezes, ca el juez deve tener la seña, e tengo que si <a> afruenta viniesses o a logar de periglo e omne vil o rafez toviesses la seña que podrié (23) caer el concejo en grant onta e en grant vergüença.

Output. ER=¥(¥d+¥)(&)=>\$2 \$1

Otrosí (22)mando que los menestrales non echen suerte en el judgado por seer juezes, ca el juez deve tener la seña, e tengo que si <a> afruenta viniesses o a logar de periglo e omne vil o rafez toviesses la seña que podrié (23) caer el concejo en grant onta e en grant vergüença.

5.6. Redefinición

LETRAS.xlsm redefine las expresiones regulares (ER) de la forma siguiente:

#	Frontera léxica, inclusive en lenguas europeas
¥l	Cualquier letra europea y su complemento ¥L
%	Cero o más de cero veces de una letra europea [¥l*]
&	Cero o más de cero veces de una letra europea [¥l+]
¥g	Letra griega y su complemento ¥G
¥e	Letra coreana y su complemento ¥E
¥y	Carácter chino y su complemento ¥Y
¥v	Letra vocálica y su complemento ¥V
¥c	Letra consonántica y su complemento ¥C

- ¥h Letra japonesa hiragana y su complemento ¥H
- ¥k Letra japonesa hiragana y su complemento ¥K
- ¥z Letra japonesa kanji y su complemento ¥Z
- ¥j Letra japonesa y su complemento ¥J

Output: ER=#(m%)=>[\$1]

Otro(22)ssí [mando] que los [menstrales] non echen suerte en el judgado por seer juezes, ca el juez deve tener la seña, e tengo que si <a> afruenta viniesse o a logar de periglo e omne vil o rafez toviessa la seña que podrié (23) caer el concejo en grant onta e en grant vergüença.

Output: ER=(%[rs])#=>[\$1]

Otro(22)ssí mando que [los] [menstrales] non echen suerte en el judgado [por] [seer] [juezes], ca el juez deve [tener] la seña, e tengo que si <a> afruenta viniesse o a [logar] de periglo e omne vil o rafez toviessa la seña que podrié (23) [caer] el concejo en grant onta e en grant vergüença.

5.7. Substitución

Seleccionamos [Expresión regular] en [Fórmula de búsqueda]. Por ejemplo si utilizamos la fórmula «r¥w*» (el signo de «¥» puede ser también la barra invertida «\»). «¥w» significa una letra utilizada en las palabras inglesas (alfabeto inglés). El asterisco «*» significa la repetición de más de cero inclusive veces de la expresión anterior. Por ejemplo, la fórmula siguiente convierte las secuencias que empieza con «r» en «<R>».

r¥w*=><R>

A la <R>ón de un hotel mad<R>ño llega un p<R> ext<R> pa<R> pa<R> como confe<R> en un semina<R> sob<R> Nut<R>ón o<R> po<R> una unive<R> de ve<R> con sede en El Esco<R>.

La expresión regular normal no reconoce, por ejemplo la «o» acentuada «ó» como letra utilizada en la palabra. Para salvar el problema LETRAS amplía las funciones de la expresión regular con varias redefiniciones (→ 8.). Por ejemplo «¥l» incluye letras utilizadas en las principales lenguas europeas (español, francés, alemán, etc.). De modo que «r¥w*» corresponde a «rile», mientras que «¥l*» de LETRAS corresponde a «rileño», incluye también «ñ»:

r¥l*=><R>

A la <R> de un hotel mad<R> llega un p<R> ext<R> pa<R> pa<R> como confe<R> en un semina<R> sob<R> Nut<R> o<R> po<R> una unive<R> de ve<R> con sede en El Esco<R>.

LETRAS redefine «%» en «¥l*», de modo que en vez de escribir «r¥l*», puede utilizar la expresión más simple: «r%».

La «back reference» (referencia posterior) es una función útil para reproducir con «\$1» la secuencia correspondiente a la fórmula entre paréntesis «(...)»:

(r%)=><\$1>

A la <recepción> de un hotel mad<rileño> llega un p<rofesor> ext<ranjero> pa<ra> pa<rticipar> como confe<renciante> en un semina<rio> sob<re> Nut<rición> o<rganizado> po<r> una unive<rsidad> de ve<rano> con sede en El Esco<rial>.

Hemos puesto la secuencia correspondiente entre dos corchetes angulares <...>.

Para buscar la palabra que empieza con «r», se utiliza el signo de «#», que corresponde al linde léxico:

#(r%)=><\$1>

A la <recepción> de un hotel madrileño llega un profesor extranjero para participar como conferenciante en un seminario sobre Nutrición organizado por una universidad de verano con sede en El Escorial.

Las letras europeas que no están en el alfabeto inglés, se interpretan con una combinación establecida por el usuario, por ejemplo: «ñ» con «n@» y «ó» con «ñ»:

(%n@%)=><\$1>

(%ó%)=><\$1>

A la <recepción> de un hotel <madrileño> llega un profesor extranjero para participar como conferenciante en un seminario sobre <Nutrición> organizado por una universidad de verano con sede en El Escorial.

Las fórmulas arriba mencionadas pueden agruparse con «(...|...)» como: «(%(n@|ó)%)=><\$1>». Con las fórmulas agrupadas se acelera el procesamiento considerablemente.

5.8. Condiciones anterior y posterior

Con los signos de {{...}} se pueden especificar condiciones antes y después del patrón de búsqueda:

{{te }} va

Texto	Foco	Texto
...no puedo llevarle algún extra a los críos. Vete allí, que te	va	a encantar!
...ropa deportiva □sudaderas y pants□ pa' tus chavos. ¡Te	va	a encantar! Ya verás. Yo el otro día vi batas para hombre y
...y faldas guapas para tu mujer o chándal para los críos. ¡Te	va	a encantar! Ya verás. Yo el otro día vi camisas de hombre
...hasta chandal preciosos para los niños. ¡Tú verás cómo te	va	a gustar! Yo he visto allí unos monos para hombres, y delar
...mujer muy buena para tu esposa o buzos para los chicos. Te	va	a gustar. He visto también mandiles de colegio...
...dejés engañar por las apariencias. En cualquier momento te	va	a hacer pisar el palito y vas a caer en la trampa. Después n

5.9. Distinción de mayúscula y minúscula

Al seleccionar [D. may./min.] (Distinción de mayúscula y minúscula), se realiza la operación con distinción de ambos tipos de letras.

(%e%)=><\$1> (Sin distinción de mayúscula y minúscula)

A la <recepción><de> un <hotel><madrileño><llega> un <profesor><extranjero> para participar como <conferenciante><en> un <seminario><sobre> Nutrición organizado por una <universidad><de><verano> con <sede><en> <El><Escorial>.

(%e%)=><\$1> (Con distinción de mayúscula y minúscula)

A la <recepción><de> un <hotel><madrileño><llega> un <profesor><extranjero>

para participar como <conferenciante><en> un <seminario><sobre> Nutrición organizado por una <universidad><de><verano> con <sede><en>El Escorial.

(%E%)=><\$1> (Con distinción de mayúscula y minúscula)

A la recepción de un hotel madrileño llega un profesor extranjero para participar como conferenciante en un seminario sobre Nutrición organizado por una universidad de verano con sede en <El><Escorial>.

5.10. Lista léxica

Al seleccionar [Lista léxica] en [Análisis unifocal], se realiza el reemplazo en unidades de palabras. En el ejemplo siguiente, «a» o «A» como palabra se convierte en «A», mientras que en la «a» de «la», por ejemplo, no se realiza el reemplazo. La lista de palabras es útil cuando se hace la conversión de una gran cantidad de palabras en un tiempo mínimo.

a=>A

de=>DE

en=>EN

A la recepción DE un hotel madrileño llega un profesor extranjero para participar como conferenciante EN un seminario sobre Nutrición organizado por una universidad DE verano con sede EN El Escorial.

Como no se trata de expresiones regulares, no podemos agrupar las tres fórmulas en una: «(a|de|en)». Si desea utilizar la expresión regular agrupada, debe formular: «#(a|de|en)#=><\$1>».

FIN