

[Año]

Innovación Educativa

Ricardo Villafaña Figueroa

[ANÁLISIS ESTADÍSTICO DE TEXTOS]

Se muestra el uso del programa computacional "Simple Concordance Program" para el análisis estadístico de textos, tomando como ejemplo el Poema Número 20 de Pablo Neruda, del libro "Veinte poemas de amor y una canción desesperada"

Simple Concordance Program

Simple Concordance Program (SCP) es un programa de software libre, sin costo, que permite realizar análisis estadístico de textos.

El análisis estadístico de textos permite el estudio de corpus de textos literarios o políticos y la extracción de información de archivos históricos y de bases documentales.

De manera particular el SCP permite:

- Encontrar el contexto de una palabra (concordancia) y realizar búsqueda de palabras o frases
- Puede encontrar las líneas donde aparece una palabra y al mismo tiempo encontrar su contexto
- Crear listas de palabras, las cuales puede ordenar, imprimir o guardar en un archivo
- Especificar las palabras o frases usando prefijos, sufijos, palabras específicas o usando patrones. Puede seleccionar un grupo de palabras para producir concordancias o lista de palabras
- Excluir palabras usando *stoplist*
- Desplegar estadísticas de letras y frecuencia de palabras

A manera de introducción al análisis estadístico de textos, se presenta el análisis del Poema Número 20 de Pablo Neruda, del libro “Veinte poemas de amor y una canción desesperada”, utilizando el Simple Concordance Program.

POEMA 20

Pablo Neruda

1. Puedo escribir los versos más tristes esta noche.
2. Escribir, por ejemplo: «La noche está estrellada,
3. y tiritan, azules, los astros, a lo lejos.»
4. El viento de la noche gira en el cielo y canta.
5. Puedo escribir los versos más tristes esta noche.
6. Yo la quise, y a veces ella también me quiso.
7. En las noches como ésta la tuve entre mis brazos.
8. La besé tantas veces bajo el cielo infinito.
9. Ella me quiso, a veces yo también la quería.
10. Cómo no haber amado sus grandes ojos fijos.
11. Puedo escribir los versos más tristes esta noche.
12. Pensar que no la tengo. Sentir que la he perdido.
13. Oír la noche inmensa, más inmensa sin ella.
14. Y el verso cae al alma como al pasto el rocío.
15. Qué importa que mi amor no pudiera guardarla.
16. La noche está estrellada y ella no está conmigo.
17. Eso es todo. A lo lejos alguien canta. A lo lejos.
18. Mi alma no se contenta con haberla perdido.
19. Como para acercarla mi mirada la busca.
20. Mi corazón la busca, y ella no está conmigo.
21. La misma noche que hace blanquear los mismos árboles.
22. Nosotros, los de entonces, ya no somos los mismos.
23. Ya no la quiero, es cierto, pero cuánto la quise.
24. Mi voz buscaba el viento para tocar su oído.
25. De otro. Será de otro. Como antes de mis besos.
26. Su voz, su cuerpo claro. Sus ojos infinitos.
27. Ya no la quiero, es cierto, pero tal vez la quiero.
28. Es tan corto el amor, y es tan largo el olvido.
29. Porque en noches como ésta la tuve entre mis brazos,
30. Mi alma no se contenta con haberla perdido.
31. Aunque éste sea el último dolor que ella me causa,
32. y éstos sean los últimos versos que yo le escribo.

Estadísticas generales

Simple Concordance Program 4.09 (build 13), May 2008 - [Project1.scp]

File View Tools Help

KeyWords 133 words: a, acercarla, s Increasing Frequency Order Invert

Word List Statistics

All KeyWords **Statistics**

Word Profile Project Statistics Letter Frequencies

Word Frequency Profile of 133 words based on the KeyWords

Word Frequency	Number of Words	Cumulative Vocabulary	Cumulative Word Count	Percentage Vocabulary	Percentage Word Count
1	71	71	71	53.38346	24.39863
2	30	101	131	75.93985	45.01718
3	15	116	176	87.21805	60.48110
4	4	120	192	90.22556	65.97938
5	4	124	212	93.23308	72.85223
6	3	127	230	95.48872	79.03780
8	3	130	254	97.74436	87.28522
9	1	131	263	98.49624	90.37801
10	1	132	273	99.24812	93.81443
18	1	133	291	100.00000	100.00000

C:\Users\Ricardo\Documents\Poemas\Project1.scp Project1 1 file, 32 lines, 291 words, 133 word vocabulary

Estadísticas generales

Líneas	32
Número total de palabras	291
Número de palabras distintas	133
Palabra más larga (10 caracteres)	estrellada
La oración más larga	Aunque éste sea el último dolor que ella me causa

Frecuencia de las palabras en orden alfabético

The screenshot shows the Simple Concordance Program 4.09 (build 13, May 2008) interface. The window title is "Simple Concordance Program 4.09 (build 13), May 2008 - [Project1.scp]". The menu bar includes File, View, Tools, and Help. The toolbar contains icons for file operations and search. The main interface is divided into three panes: Concordance, Word List, and Statistics. The Word List pane is active, showing a list of 133 words sorted in ascending alphabetical order. The list is displayed in a grid format with columns for frequency and word. The status bar at the bottom indicates the current file path, project name, and statistics: "C:\Users\Ricardo\Documents\Poemas\Project1.scp Project1 1 file, 32 lines, 291 words, 133 word vocabulary".

Frequency	Word	Frequency	Word	Frequency	Word	Frequency	Word	Frequency	Word	Frequency	Word
5	a	1	acercarla	2	al	1	alguien	3	alma	1	amado
2	amor	1	antes	1	astros	1	aunque	1	azules	1	árboles
1	bajo	1	besé	1	besos	1	blanquear	2	brazos	2	busca
1	buscaba	1	cae	2	canta	1	causa	2	cielo	2	cierto
1	claro	5	como	2	con	2	conmigo	2	contenta	1	corazón
1	corto	1	cómo	1	cuánto	1	cuerpo	5	de	1	dolor
1	ejemplo	9	el	6	ella	3	en	1	entonces	2	entre
5	es	4	escribir	1	escribo	1	eso	3	esta	4	está
2	estrellada	2	ésta	1	éste	1	éstos	1	fijos	1	gira
1	grandes	1	guardarla	1	haber	2	haberla	1	hace	1	he
1	importa	1	infinito	1	infinitos	2	inmensa	18	la	1	largo
1	las	1	le	3	lejos	3	lo	8	los	4	más
3	me	6	mi	1	mirada	3	mis	1	misma	2	mismos
10	no	8	noche	2	noches	1	nosotros	1	oído	1	oír
2	ojos	1	olvido	2	otro	2	para	1	pasto	1	pensar
3	perdido	2	pero	1	por	1	porque	1	pudiera	3	puedo
6	que	1	quería	1	qué	3	quiero	2	quise	2	quiso
1	rocío	2	se	1	sea	1	sean	1	sentir	1	será
1	sin	1	somos	3	su	2	sus	1	tal	2	también
2	tan	1	tantas	1	tengo	1	tiritan	1	tocar	1	todo
3	tristes	2	tuve	1	último	1	últimos	3	veces	1	verso
4	versos	1	vez	2	viento	2	voz	8	y	3	ya
3	yo										

Frecuencia de las palabras en orden creciente

Simple Concordance Program 4.09 (build 13), May 2008 - [Project1.scp]

File View Tools Help

KeyWords 133 words: a, acercarla, s Increasing Frequency Order Invert

Concordance Word List Statistics

Layout
Columns Left Aligned Frequencies All KeyWords Word List

Word List: 133 words

1	acercarla	1	alguien	1	amado	1	antes	1	astros	1	aunque
1	azules	1	árboles	1	bajo	1	besé	1	besos	1	blanquear
1	buscaba	1	cae	1	causa	1	claro	1	corazón	1	corto
1	cómo	1	cuánto	1	cuerpo	1	dolor	1	ejemplo	1	entonces
1	escribo	1	eso	1	éste	1	éstos	1	fijos	1	gira
1	grandes	1	guardarla	1	haber	1	hace	1	he	1	importa
1	infinito	1	infinitos	1	largo	1	las	1	le	1	mirada
1	misma	1	nosotros	1	oído	1	oír	1	olvido	1	pasto
1	pensar	1	por	1	porque	1	pudiera	1	quería	1	qué
1	rocío	1	sea	1	sean	1	sentir	1	será	1	sin
1	somos	1	tal	1	tantas	1	tengo	1	tiritan	1	tocar
1	todo	1	último	1	últimos	1	verso	1	vez	2	al
2	amor	2	brazos	2	busca	2	canta	2	cielo	2	cierto
2	con	2	conmigo	2	contenta	2	entre	2	estrellada	2	ésta
2	haberla	2	inmensa	2	misimos	2	noches	2	ojos	2	otro
2	para	2	pero	2	quise	2	quiso	2	se	2	sus
2	también	2	tan	2	tuve	2	viento	2	voz	3	alma
3	en	3	esta	3	lejos	3	lo	3	me	3	mis
3	perdido	3	puedo	3	quiero	3	su	3	tristes	3	veces
3	ya	3	yo	4	escribir	4	está	4	más	4	versos
5	a	5	como	5	de	5	es	6	ella	6	mi
6	que	8	los	8	noche	8	y	9	el	10	no
18	la										

C:\Users\Ricardo\Documents\Poemas\Project1.scp Project1 1 file, 32 lines, 291 words, 133 word vocabulary

Frecuencia de las palabras en orden decreciente

The screenshot shows the Simple Concordance Program 4.09 interface. The 'KeyWords' field contains '133 words: a, acercarla, s'. The 'Decreasing Frequency Order' dropdown is selected, and the 'Invert' checkbox is unchecked. The 'Word List' tab is active, displaying a list of 133 words sorted by frequency. The status bar at the bottom indicates '1 file, 32 lines, 291 words, 133 word vocabulary'.

Frequency	Word	Frequency	Word	Frequency	Word	Frequency	Word	Frequency	Word	Frequency	Word
1	acercarla	1	alguien	1	amado	1	antes	1	astros	1	aunque
1	azules	1	árboles	1	bajo	1	besé	1	besos	1	blanquear
1	buscaba	1	cae	1	causa	1	claro	1	corazón	1	corto
1	cómo	1	cuánto	1	cuerpo	1	dolor	1	ejemplo	1	entonces
1	escribo	1	eso	1	éste	1	éstos	1	fijos	1	gira
1	grandes	1	guardarla	1	haber	1	hace	1	he	1	importa
1	infinito	1	infinitos	1	largo	1	las	1	le	1	mirada
1	misma	1	nosotros	1	oído	1	oír	1	olvido	1	pasto
1	pensar	1	por	1	porque	1	pudiera	1	quería	1	qué
1	rocío	1	sea	1	sean	1	sentir	1	será	1	sin
1	somos	1	tal	1	tantas	1	tengo	1	tiritan	1	tocar
1	todo	1	último	1	últimos	1	verso	1	vez	2	al
2	amor	2	brazos	2	busca	2	canta	2	cielo	2	cierto
2	con	2	conmigo	2	contenta	2	entre	2	estrellada	2	ésta
2	haberla	2	inmensa	2	misimos	2	noches	2	ojos	2	otro
2	para	2	pero	2	quise	2	quiso	2	se	2	sus
2	también	2	tan	2	tuve	2	viento	2	voz	3	alma
3	en	3	esta	3	lejos	3	lo	3	me	3	mis
3	perdido	3	puedo	3	quiero	3	su	3	tristes	3	veces
3	ya	3	yo	4	escribir	4	está	4	más	4	versos
5	a	5	como	5	de	5	es	6	ella	6	mi
6	que	8	los	8	noche	8	y	9	el	10	no
18	la										

Frecuencia de las palabras por longitud de palabra, orden creciente

Simple Concordance Program 4.09 (build 13), May 2008 - [Project1.scp]

File View Tools Help

KeyWords 133 words: a, acercarla, s Increasing Length Order Invert

Concordance Word List Statistics

Layout
Columns Left Aligned Frequencies All KeyWords Word List

Word List: 133 words

5 a	8 y	2 al	5 de	9 el	3 en
5 es	1 he	18 la	1 le	3 lo	3 me
6 mi	10 no	2 se	3 su	3 ya	3 yo
1 cae	2 con	1 eso	1 las	8 los	4 más
3 mis	1 oír	1 por	6 que	1 qué	1 sea
1 sin	2 sus	1 tal	2 tan	1 vez	2 voz
3 alma	2 amor	1 bajo	1 besé	5 como	1 cómo
6 ella	3 esta	4 está	2 ésta	1 éste	1 gira
1 hace	1 oído	2 ojos	2 otro	2 para	2 pero
1 sean	1 será	1 todo	2 tuve	1 amado	1 antes
1 besos	2 busca	2 canta	1 causa	2 cielo	1 claro
1 corto	1 dolor	2 entre	1 éstos	1 fijos	1 haber
1 largo	3 lejos	1 misma	8 noche	1 pasto	3 puedo
2 quise	2 quiso	1 rocío	1 somos	1 tengo	1 tocar
3 veces	1 verso	1 astros	1 aunque	1 azules	2 brazos
2 cierto	1 cuánto	1 cuerpo	1 mirada	2 mismos	2 noches
1 olvido	1 pensar	1 porque	1 quería	3 quiero	1 sentir
1 tantas	1 último	4 versos	2 viento	1 alguien	1 árboles
1 buscaba	2 conmigo	1 corazón	1 ejemplo	1 escribo	1 grandes
2 haberla	1 importa	2 inmensa	3 perdido	1 pudiera	2 también
1 tiritan	3 tristes	1 últimos	2 contenta	1 entonces	4 escribir
1 infinito	1 nosotros	1 acercarla	1 blanquear	1 guardarla	1 infinitos
2 estrellada					

C:\Users\Ricardo\Documents\Poemas\Project1.scp Project1 1 file, 32 lines, 291 words, 133 word vocabulary

Frecuencia de las palabras por longitud de palabra, orden decreciente

Simple Concordance Program 4.09 (build 13), May 2008 - [Project1.scp]

File View Tools Help

KeyWords 133 words: a, acercarla, s Decreasing Length Order Invert

Concordance Word List Statistics

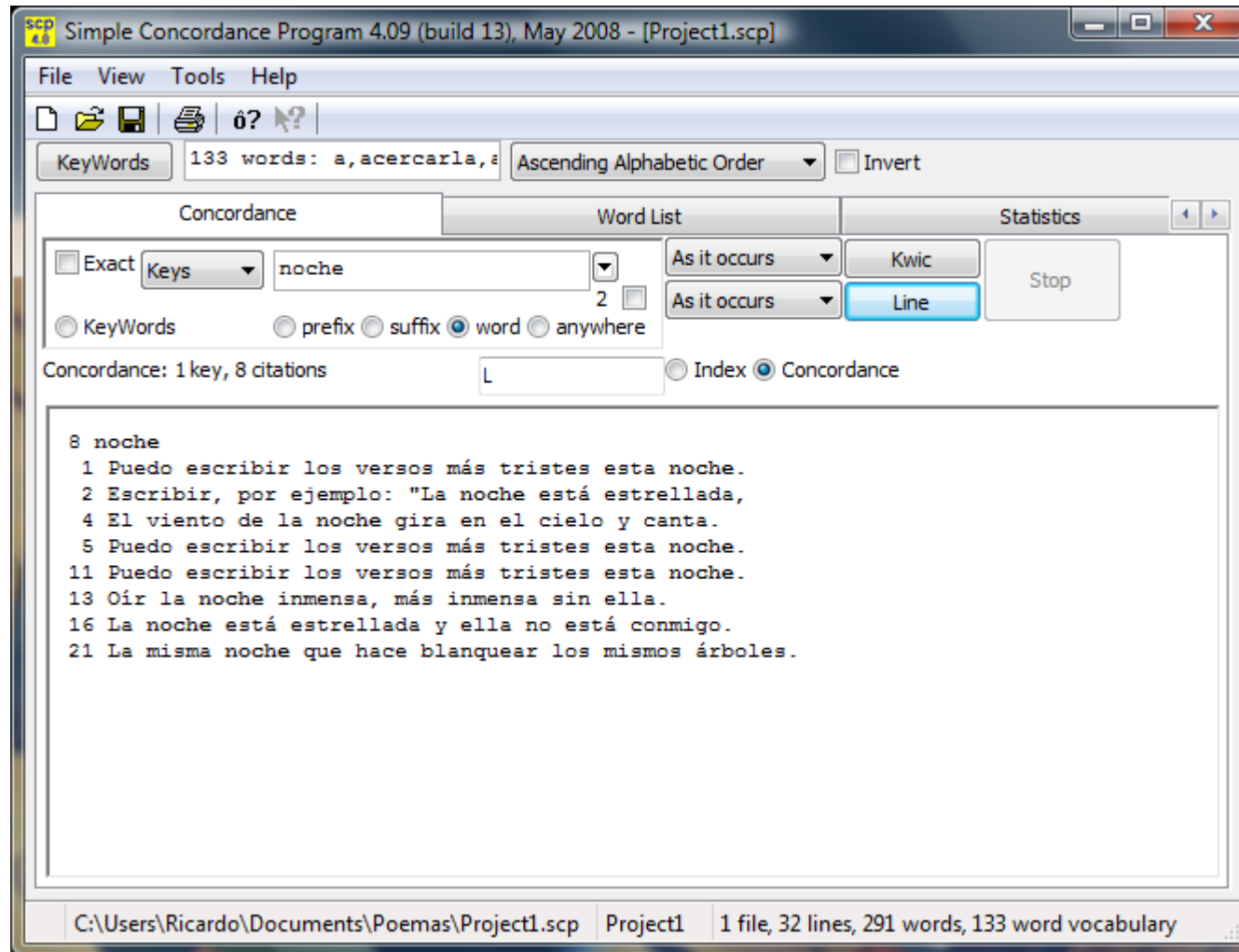
Layout
Columns Left Aligned Frequencies All KeyWords Word List

Word List: 133 words

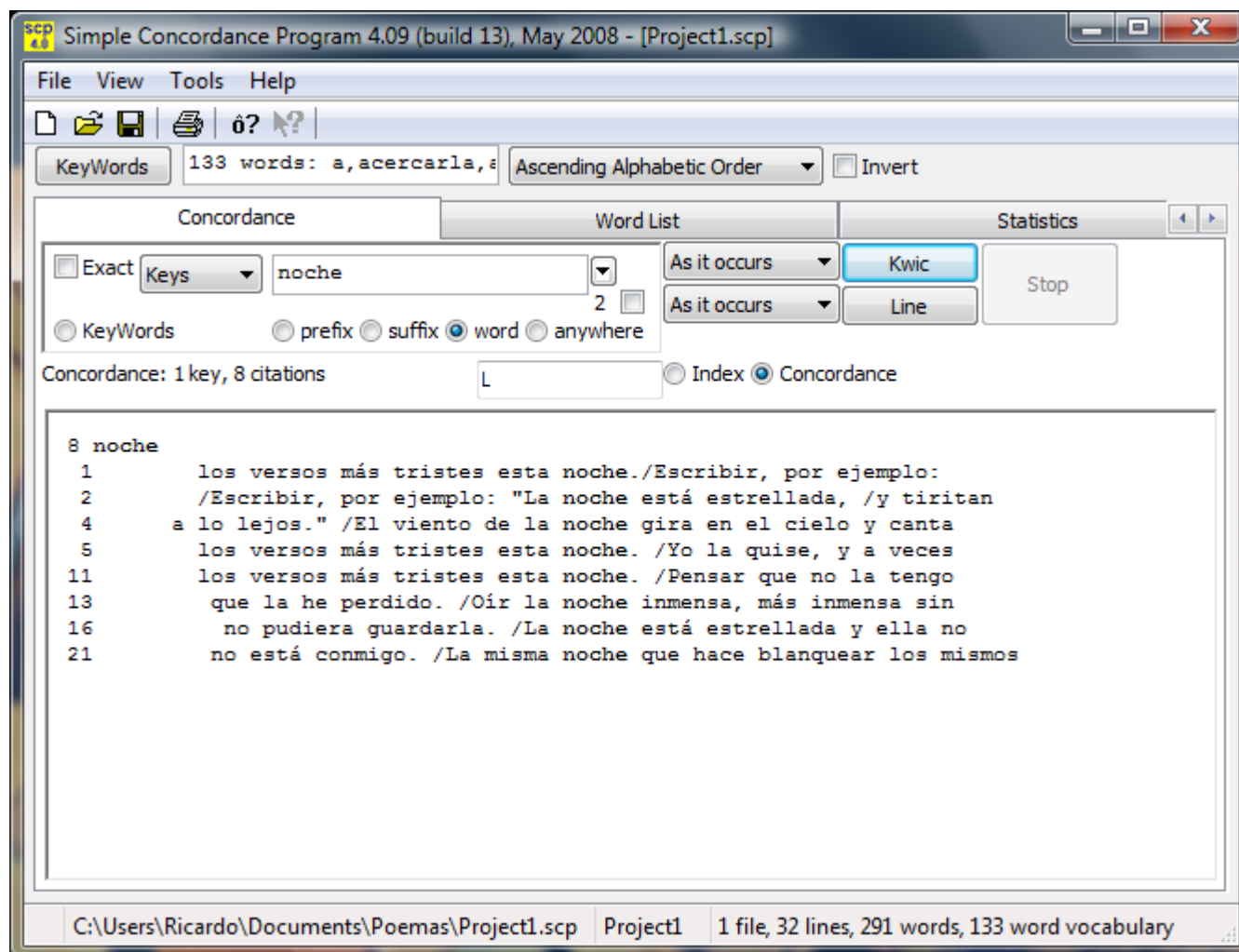
2	estrellada	1	acercarla	1	blanquear	1	guardarla	1	infinitos	2	contenta
1	entonces	4	escribir	1	infinito	1	nosotros	1	alguien	1	árboles
1	buscaba	2	conmigo	1	corazón	1	ejemplo	1	escribo	1	grandes
2	haberla	1	importa	2	inmensa	3	perdido	1	podiera	2	también
1	tiritan	3	tristes	1	últimos	1	astros	1	aunque	1	azules
2	brazos	2	cierto	1	cuánto	1	cuerpo	1	mirada	2	misimos
2	noches	1	olvido	1	pensar	1	porque	1	quería	3	quiero
1	sentir	1	tantas	1	último	4	versos	2	viento	1	amado
1	antes	1	besos	2	busca	2	canta	1	causa	2	cielo
1	claro	1	corto	1	dolor	2	entre	1	éstos	1	fijos
1	haber	1	largo	3	lejos	1	misma	8	noche	1	pasto
3	puedo	2	quise	2	quiso	1	rocío	1	somos	1	tengo
1	tocar	3	veces	1	verso	3	alma	2	amor	1	bajo
1	besé	5	como	1	cómo	6	ella	3	esta	4	está
2	ésta	1	éste	1	gira	1	hace	1	oído	2	ojos
2	otro	2	para	2	pero	1	sean	1	será	1	todo
2	tuve	1	cae	2	con	1	eso	1	las	8	los
4	más	3	mis	1	oír	1	por	6	que	1	qué
1	sea	1	sin	2	sus	1	tal	2	tan	1	vez
2	voz	2	al	5	de	9	el	3	en	5	es
1	he	18	la	1	le	3	lo	3	me	6	mi
10	no	2	se	3	su	3	ya	3	yo	5	a
8	y										

C:\Users\Ricardo\Documents\Poemas\Project1.scp Project1 1 file, 32 lines, 291 words, 133 word vocabulary

Concordancia. Líneas en que aparece la palabra *noche*



Concordancia. El contexto y la línea donde aparece la palabra *noche*



Concordancia. El contexto de la palabra *noche* (como prefijo)

