



Onomázein

ISSN: 0717-1285

onomazein@uc.cl

Pontificia Universidad Católica de Chile
Chile

Martínez-Gamboa, Ricardo

Patrones cuantitativos en novelas chilenas de los siglos XIX a XXI

Onomázein, núm. 32, diciembre, 2015, pp. 239-253

Pontificia Universidad Católica de Chile

Santiago, Chile

Disponible en: <http://www.redalyc.org/articulo.oa?id=134544960014>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica

Red de Revistas Científicas de América Latina, el Caribe, España y Portugal

Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

Patrones cuantitativos en novelas chilenas de los siglos XIX a XXI

*Quantitative patterns in Chilean novels
of XIX-XXI centuries*

Ricardo Martínez-Gamboa

Universidad Diego Portales
Chile

ONOMÁZEIN 32 (diciembre de 2015): 239-253
DOI: 10.7764/onomazein.32.14



Ricardo Martínez-Gamboa: Escuela de Literatura Creativa, Facultad de Comunicación y Letras, Universidad Diego Portales, Chile. | Correo electrónico: ricardo.martinez@mail.udp.cl

Fecha de recepción: julio de 2014
Fecha de aceptación: abril de 2015

Resumen

El análisis cuantitativo de la literatura es un método de orientación discursiva y estadística que pretende determinar regularidades en la presencia de rasgos léxicogramaticales en las obras literarias que ha experimentado un fuerte desarrollo desde el trabajo de Franco Moretti (Goodwin, Holbo y Moretti, 2005; Moretti, 2011), quien propone la noción de “lectura distante” (*distant reading*), donde se considera que el procesamiento de los datos de las obras colabora a una interpretación más centrada en el contenido explícito de los textos que en la exégesis de los mismos.

Palabras clave: novelas chilenas; Análisis Lingüístico Computacional; modelo de Biber.

El presente artículo tiene por objetivo presentar una gestión de la naturaleza expuesta, acometida sobre una treintena de novelas chilenas publicadas entre 1862 (*Martín Rivas*) y 2011 (*Ruido*). Para ello se realiza una serie de procesamiento de lingüística computacional (ACL) inspirados en la investigación seminal de Biber (1988) y se establecen tres dimensiones en las que las novelas chilenas se despliegan que permiten agruparlas de acuerdo con sus propiedades lingüísticas facilitando su interpretación exegetica.

Abstract

Quantitative analysis of the literature is a discursive method and a statistical guidance that aims to determine the presence of regularities in lexicogrammatical features in literary works that has experienced strong growth since the work of Franco Moretti (Goodwin, Holbo & Moretti, 2005; Moretti, 2011), who proposed the notion of “distant reading”, which considers that data processing works collaborates to an interpretation more focused on the explicit content of the texts in exe-

gesis thereof. This article aims to present a work of the exposed nature realized on thirty Chilean novels published between 1862 (*Martín Rivas*) and 2011 (*Ruido*). For this, a series of processing of computational linguistics (ACL) inspired by the seminal research of Biber (1988) is carried down with the result of three dimensions in which the Chilean novels unfold that allow grouping according to their linguistic properties facilitating their exegetical interpretation.

Keywords: Chilean novels; Computational Linguistic Analysis; Biber’s model.

1. Introducción

Se denomina *Análisis Lingüístico Computacional* (ALC) a una línea de estudios en el contexto de la escuela de investigación de la literatura de orientación cognitiva, la que a su vez ha sido llamada de diversos modos, como *Teoría literaria cognitiva*, *Crítica cognitiva*, *Estilística cognitiva*, *Historicismo neural* y, sobre todo, *Poética cognitiva* (Stockwell, 2002; Gavins y Steen, 2003). El trabajo seminal en el campo de la ALC ha sido realizado por Lancashire (1993), donde el autor ha extraído computacionalmente patrones lingüísticos y de contenido desde la novela *El cuento de la sirvienta* de Margaret Atwood.

Recientemente, este tipo de aproximación a la literatura ha sido denominado “distant reading” (lectura remota) por Franco Moretti (2005; Goodwin, Holbo y Moretti, 2011), aduciéndose que

lo que es una fracción mínima del campo literario es el campo en el que [los teóricos literarios] trabajamos: un canon de 200 novelas, por ejemplo, suena muy grande para el siglo XIX en Gran Bretaña (y es mucho más grande que el actual), pero sigue siendo inferior al uno por ciento de las novelas que se publican en realidad: veinte mil, treinta mil, o más, nadie sabe en realidad —y el “close reading” no ayuda aquí—. Una novela al día, todos los días del año se tardaría un siglo más o menos (...). [Se propone que] [l]a gran masa [de novelas] sin leer deben ser estudiadas con el fin de ver los patrones de macroevolución en la historia literaria.

En el presente estudio se ha seguido los principios del ALC y del “distant reading” para abordar el estudio de una treintena de novelas chilenas producidas entre los siglos XIX y XXI. Las investigaciones sobre la novelística nacional y latinoamericana en esta perspectiva de periodización han venido llevándose a cabo al menos desde los trabajos de Cedomil Goic (1968, 1972) y Cánovas (1997), quien ha planteado que “establecer los rasgos específicos que adquiere la novela chilena en los nuevos narradores [y] despejar las imágenes estructurales que surgen de sus obras

novelísticas [hace necesario preguntarse por] la noción de generación” (Cuadros, 1997).

El presente estudio, en consecuencia, pretende dar cuenta del corpus mencionado de treinta novelas chilenas atendiendo a dicha orientación periodizante, ocupando a la vez el modelo de la ALC, particularmente siguiendo el modelo de Biber (1998; Biber, Connor y Upton, 2007), que ya ha sido implementado en un trabajo anterior (Martínez, 2012). El objetivo es realizar una serie de procesamiento de lingüística computacional (ACL) sobre las novelas para establecer en primer lugar si resulta posible determinar dimensiones de covariación y, de acuerdo con ello, agruparlas respecto de sus propiedades lingüísticas a fin de facilitar su interpretación exegética.

2. Los datos del análisis

El primer proceso llevado a cabo fue la selección de las obras. Para disponer de ellas se recurrió a la base de datos literaria Memoria Chilena, así como a repertorios de textos accesibles en Internet y a la colaboración de autores como Álvaro Bisama y Alejandro Zambra que facilitaron sus textos.

Una vez que se obtuvieron las versiones en textos electrónicos de una cincuentena de novelas, se procedió a la selección de treinta de ellas, procurando cubrir de la manera más poblada posible el periodo analizado: *2666*, *Alhué*, *Bonsái*, *Condell*, *El loco Estero*, *El lugar sin límites*, *El niño que enloqueció de amor*, *El socio*, *El último grumete de la Baquedano*, *Eloy*, *Estrellas Muertas*, *Formas de volver a casa*, *Hijo de Ladrón*, *La amartajada*, *La última niebla*, *La casa de los espíritus*, *Los detectives salvajes*, *Lumpérica*, *Mala onda*, *Martín Rivas*, *Música Marciana*, *Palomita Blanca*, *Papelucho detective*, *Papelucho en vacaciones*, *Papelucho historiador*, *Papelucho perdido*, *Papelucho y el marciano*, *Papelucho*, *Patas de perro*, *Ruido*. La más antigua de las cuales resulta ser *Martín Rivas* (Alberto Blest Gana, 1862), y la más reciente, *Ruido* (Álvaro Bisama, 2011).

Estas obras han sido convertidas al formato de texto plano (.txt), a partir de sus formatos originales en Acrobat (.pdf) y Microsoft Word (.doc, .docx). A continuación, con un programa llamado KfNgram (Fletcher, 2002), se ha realizado un conteo de palabras (*tokens*, esto es la cantidad de palabras totales en un texto; Peirce, 1932), para finalmente ser llevados los datos al programa de conteo de características en el modelo de “bag of words” (LSM, 2011).

3. Rasgos analizados

El análisis de rasgos se ha llevado a cabo ocupando la herramienta Linguistic Inquiry and Word Count (LIWC), que “se compone de cerca de 4.500 palabras y palabras derivadas. Cada raíz de palabra o palabra define una o más categorías. Por ejemplo, la palabra *cried* forma parte de cinco categorías de palabras: tristeza, emoción negativa, afecto, verbo y el verbo en tiempo pasado” (Pennebaker y otros, 2007). El programa LIWC se creó para calcular el grado en que las personas utilizan diferentes categorías de palabras a través de una amplia variedad de textos, incluyendo correo electrónico, discursos, poemas o habla cotidiana transcrita.

Para la versión en castellano de LIWC existen 7463 palabras etiquetadas para formar 68 categorías gramaticales y semánticas que permiten, subsecuentemente, realizar un análisis robusto de contenidos de los textos.

La tabla 1 muestra las categorías disponibles en aquel Spanish Dictionary.

En la tabla 2 se presentan los resultados para las algunas categorías que arroja el *software* para el corpus (los valores están expresados como porcentajes respecto del total de palabras de cada texto).

Siguiendo el procedimiento propuesto por Biber (1988), para cada una de las 68 categorías con sus respectivos porcentajes para cada una de las treinta novelas, se calcula su correlación Pearson respecto de todas las categorías, lo que

TABLA 1

Rasgos LIWC

LINGÜÍSTICAS	ARTÍCULO	RELATIVIDAD	ABAJO
	ASENTIR		ARRIBA
	NEGACIÓN		ESPACIO
	NOSOTROS		EXCL
	NÚMERO		FUTURO
	OTRO		INCL
	PREPO		MOCIÓN
	PRONOM		PASADO
	TU		PRESENTE
	UNOMISMO		TIEMPO
	YO		ASEARSE
PROCESOS PSICOLÓGICOS	APECTIVA	ASUNTOS PERSONALES	CASA
	AMIGOS		COMER
	ANSIEDAD		CUERPO
	CAUSA		DEPORTES
	CERTEZA		DINERO
	COMU		DORMIR
	DISCREP		ESCUELA
	EMONEG		FÍSICO
	EMOPOS		LOGRO
	ENOJO		METAFO
	FAMILIA		MUERTE
	HUMANOS		MÚSICA
	INHIB		OCUPA
	INSIGHT		PLACER
	MECCOG		RELIG
	OÍR		SEXUAL
	OPTIMI		TRABAJO
	REFOTRO		TV
	SENTIDOS	EXPERI- MENTALES	FILLERS
	SENTIR		MALDECIR
	SENTPOS		NONFL
	SOCIAL		
	TENTAT		
	TRISTEZA		
	VER		

TABLA 2

Rasgos en las novelas

	2666	ALHUÉ	BONSÁI	CONDELL	EL LOCO ESTERO	EL LUGAR SIN LÍMITES	EL NIÑO QUE ENLOQUECIÓ DE AMOR	EL SOCIO
Pronombre	7,03	7,32	6,13	7,21	8,07	9,03	13,14	7,93
Yo	0,66	0,45	0,43	1,84	1,53	1,74	5,86	1,14
Nosotros	0,1	0,35	0,05	1,21	0,16	0,16	0,2	0,07
Unomismo	0,75	0,79	0,48	3,03	1,67	1,89	6,05	1,2
Tú	1,26	2,54	0,99	0,2	2,29	1,39	1,13	1,67
Otro	4,5	3,31	4,25	3,14	3,78	5,28	5,8	4,62
Negación	1,9	1,56	2,09	1,82	1,6	2,84	2,99	2,25
Asentir	0,32	0,19	0,39	0,22	0,27	0,36	0,55	0,46
Artículo	12,81	13,26	9,5	14,07	12,8	11,87	8,53	12,03
Preposición	14,21	13,55	12,25	14,05	15,44	13,23	11,09	13,06
Número	1,05	0,51	1,42	1,64	0,48	0,43	0,42	0,86
Afectiva	3,14	4,64	3,75	2,52	5,13	3,43	6,34	4,62
Sentipos	1,66	2,42	2,09	1,51	2,71	1,85	2,79	2,14
Sentfeel	0,47	0,77	0,8	0,36	1,03	0,82	1,41	0,76
Optimismo	0,48	0,74	0,53	0,54	0,78	0,46	0,45	0,58
Negemo	1,15	1,72	1,42	0,76	1,95	1,41	3,14	2,1
Ansiedad	0,22	0,3	0,35	0,11	0,59	0,24	0,71	0,48
Enojo	0,33	0,32	0,24	0,22	0,51	0,33	0,68	0,5
Tristeza	0,32	0,75	0,39	0,07	0,46	0,37	1,15	0,62

da una matriz de medidas 68 x 68, de la que se presenta un segmento en la tabla 3.

Las correlaciones Pearson orientan sobre qué rasgos gramaticales y semánticos covarían en un texto, tomando valores entre -1 y 1, donde el segmento que va de -1 a -0,33 indica una correlación negativa (uno de los rasgos está presente en la novela cuando el otro está ausente y viceversa) y el segmento que va de 0,33 a 1 indica una correlación positiva (cuando uno de los rasgos está presente el otro también está presente). Por su parte, el segmento que va de -0,33 a 0,33 indica no correlación. Por ejemplo, en la tabla hay una

alta correlación entre “unomismo” y “pronombre” (0,85).

En el corpus, las correlaciones más altas son las que corresponden a palabras relacionadas con “unomismo” y “yo” (0,98), palabras relacionadas con “religión” y “metáforas” (0,91) y palabras relacionadas con lo “físico” y el “cuerpo” (0,96), mientras que las más altas correlaciones negativas son las relacionadas con “artículos” y “pronombres” (-0,74), “preposiciones” y “pronombres” (-0,76), “unomismo” y “preposiciones” (-0,70), “artículos” y “cognición-mecánica” (-0,71) y “artículos” y “causas” (-0,71).

TABLA 3

Correlaciones de los rasgos en las novelas

	PRONOMBRE	YO	NOSOTROS	UNOMISMO	TÚ	OTRO	NEGACIÓN	ASENTIR
PRONOMBRE	1	0,8658867	0,1385612	0,8471758	-0,2958883	0,6887749	0,644976	0,4258525
YO	0,8658867	1	0,1729795	0,9811295	-0,4742539	0,3120801	0,5687597	0,341161
NOSOTROS	0,1385612	0,1729795	1	0,3600361	-0,490643	-0,1640612	0,056919	-0,1541862
UNOMISMO	0,8471758	0,9811295	0,3600361	1	-0,5454468	0,2631697	0,5504759	0,2914488
TÚ	-0,2958883	-0,4742539	-0,490643	-0,5454468	1	-0,1521624	-0,4043303	-0,2140606
OTRO	0,6887749	0,3120801	-0,1640612	0,2631697	-0,1521624	1	0,532781	0,4517415
NEGACIÓN	0,644976	0,5687597	0,056919	0,5504759	-0,4043303	0,532781	1	0,5365714
ASENTIR	0,4258525	0,341161	-0,1541862	0,2914488	-0,2140606	0,4517415	0,5365714	1

Al disponerse de la tabla 68 x 68 de correlaciones, según el procedimiento biberiano, se realiza una reducción de variables, esto es, agrupar las variables en factores de covariación. “El análisis factorial es una herramienta matemática que se puede utilizar para examinar una amplia gama de datos” (Reese y Lochmüller, 1994) y su principal objetivo consiste reducir la gama de datos a un pequeño número de factores que resuman y representen diferentes áreas de variables relacionadas (Biber, 1988: 79 ss.). “La utilidad más relevante de este tipo de análisis es el hecho de que provee a los investigadores de una representación de la estructura subyacente de los datos estadísticamente fundamentada. Esta estructura viene dada por un número lo más pequeño posible de dimensiones de covariación de factores” (Martínez, 2012: 154).

En el procedimiento llevado a cabo para el corpus de treinta novelas los resultados han sido la reducción a tres dimensiones de covariación, tal como muestra el resultado arrojado por el programa SPSS con rotación PROMAX (véase la tabla 4):

“Dos condiciones que debe cumplir todo análisis factorial son la parsimonia y la interpretabilidad. La parsimonia exige el menor número posible de factores, la interpretabilidad exige la posibilidad de una interpretación significativa de las dimensiones definidas” (Martínez, 2012: 156).

TABLA 4

Componentes PROMAX

COMPONENT	TOTAL	% OF VARIANCE	CUMULATIVE %
1	23,696	35,903	35,903
2	16,053	24,323	60,227
3	9,322	14,125	74,351

En el caso presente, las tres dimensiones que han ocurrido parecen interpretables. Se presentan los valores máximos y mínimos de las categorías encontradas para cada una de las tres dimensiones.

La primera dimensión (véase la tabla 5) distribuye los rasgos en un eje donde los valores positivos mayores corresponden a los “pronombres”, las “causas”, la “cognición-mecánica” y el “insight”, mientras que en el otro extremo se encuentran los “artículos”, las “preposiciones”, las “metáforas”.

Estos rasgos permiten considerar a este eje como uno que se despliega hacia un Foco Narrativo (presencia de pronombres y tiempos verbales) y otro Foco Descriptivo (presencia de artículos, preposiciones y vocablos descriptivos como “ocupaciones”, “colegio”, “música”, “TV”), siguiendo la nomenclatura de Biber (1988), así como Parodi (2004; cf. Marinkovich y Cademartori, 2004).

TABLA 5

Dimensión 1

	1
Pronombre	0,965
Presente	0,918
Causa	0,898
Otros	0,879
Negaciones	0,865
Yo	0,863
Sentidos	0,851
Asentir	0,85
Cogmec	0,846
Mismo	0,821
Futuro	0,786
Insight	0,785
Incl	0,765
Otref	0,753
Excl	0,736
Com	0,711
Oir	0,689
Social	0,68
Discrep	0,666
Certeza	0,633
Comer	0,61
Posem	0,575
Dormir	0,558
Familia	0,542
Ver	0,53
Ira	0,492
Movimiento	0,434
Negemo	0,353
Posemo	0,351
Número	-0,369
Humanos	-0,412

Hogar	-0,522
Espacio	-0,525
Amigos	-0,529
TV	-0,613
Música	-0,618
Muerte	-0,72
Escuela	-0,728
Relig	-0,759
Trabajo	-0,762
Placer	-0,778
Ocup	-0,795
Metaf	-0,814
Prepos	-0,884
Artículo	-0,94

TABLA 6

Dimensión 2

Groom	0,96
Discrep	0,954
Sentir	0,931
Oir	0,922
Com	0,915
Insight	0,88
Número	0,852
Social	0,833
Ocup	0,71
Ansiedad	0,656
Negar	0,598
Incl	0,559
Bajo	0,544
Otref	0,54
Metaf	0,516
Afecto	0,421
Trabajo	0,41

Movimiento	-0,372
Tiempo	-0,383
Sexual	-0,395
Humanos	-0,418
Artículo	-0,418
Físico	-0,42
Certeza	-0,429
Tentativa	-0,448
Nosotros	-0,46
Arriba	-0,46
Familia	-0,53
Causa	-0,548
Pasado	-0,552
Futuro	-0,56
Tristeza	-0,569
Sexual	-0,62
Hogar	-0,653
Posemo	-0,659
Ver	-0,67
Escuela	-0,836

En la segunda dimensión (véase la tabla 6) se agrupan en un extremo rasgos que están relacionados con lo cognitivo (“sentir”, “oir”, “insight”), mientras que en el otro extremo aparecen rasgos relacionados con la órbita de lo emocional (“tristeza”, “posemo”). A esta dimensión, en consecuencia, se la puede denominar “eje Cognición-Emoción”.

TABLA 7

Dimensión 3

Escuela	0,696
Cuerpo	0,654
Logro	0,65
Negar	0,643
Excl	0,6

Ira	0,59
Dinero	0,577
Espacio	0,56
Oir	0,529
Sentir	0,492
Discrep	0,478
Posemo	0,367
Música	-0,362
Mismo	-0,43
Futuro	-0,476
Tentativa	-0,512
Muerte	-0,557
Sexual	-0,588
Asentir	-0,591
Físico	-0,681
Com	-0,684
Familia	-0,727
Certez	-0,764
Ver	-0,798

En la tercera dimensión (véase la tabla 7), que, como en todo caso de reducciones dimensionales, presenta menos rasgos, es posible advertir en uno de los extremos aspectos de orden social como “escuela” o “dinero”, mientras que en el otro aparecen elementos de orden individual como “familia”, “físico”, “sexual”, por lo que se puede denominar como “eje Social-Individual”.

4. Las novelas en los ejes

Tal como plantea Biber (1988), las novelas, así como en su caso las clases de textos, se pueden distribuir en cada uno de los tres ejes o dimensiones:

- Dimensión 1: Narración-Descripción
- Dimensión 2: Cognición-Emoción
- Dimensión 3: Social-Individual

Para llevar a cabo esta distribución es necesario en primer lugar normalizar los puntajes que cada novela obtiene en cada uno de los 68 rasgos. A este procedimiento se le denomina “normalización Z”, que consiste en medir cuántas desviaciones estándar se aleja del promedio en cada rasgo cada uno de los resultados de cada una de las novelas. Por ejemplo, en el rasgo “pronombres”, el promedio es un 9,33%, mientras que la desviación estándar resulta de 2,08%. En este caso, por ejemplo, la novela 2666 obtiene un puntaje Z de -1,105. Ello porque su porcentaje de pronombres es de 7,03%, lo que resulta un poco más de una desviación estándar por debajo del promedio ($Puntaje\ Z = (Puntaje - Promedio) / (Desviación\ Estándar)$).

Ranqueadas de esta manera las novelas se distribuyen del siguiente modo para cada una de las dimensiones.

TABLA 8

Dimensión 1: Narración-Descripción

El niño que enloqueció de amor	35,06
Papelucho detective	23,34
Papelucho	22,95
Lumpérica	22,22
Palomita Blanca	19,99
Mala onda	16,41
Papelucho y el marciano	13,53
Formas de volver a casa	11,18
Papelucho perdido	9,35
Patas de perro	9,12
Papelucho en vacaciones	7,28
Papelucho historiador	6,50
Martín Rivas	5,57
Los detectives salvajes	3,44
El lugar sin límites	2,45
La ultima niebla	1,04
Eloy	0,17

Estrellas Muertas	-1,97
Hijo de Ladrón	-2,55
La amortajada	-7,93
El socio	-8,09
El loco Estero	-9,37
Bonsái	-10,65
Música Marciana	-16,78
2666	-17,81
La casa de los espíritus	-19,38
Condell	-25,89
El último grumete de la Baquedano	-25,94
Alhué	-30,62
Ruido	-32,61

TABLA 9

Dimensión 2: Cognición-Emoción

La amortajada	27,90
La ultima niebla	27,26
Eloy	27,02
El niño que enloqueció de amor	17,58
Patas de perro	13,84
Alhué	5,32
El loco Estero	4,81
Martín Rivas	4,08
El lugar sin límites	3,81
El socio	3,69
La casa de los espíritus	3,52
Papelucho y el marciano	1,71
Hijo de Ladrón	-1,43
Papelucho en vacaciones	-1,72
Papelucho	-2,33
Papelucho perdido	-2,65
El último grumete de la Baquedano	-3,60

Mala onda	-4,69
Palomita Blanca	-4,98
Papelucho detective	-5,25
Música Marciana	-8,46
Bonsái	-8,64
Papelucho historiador	-8,80
Ruido	-9,14
Los detectives salvajes	-9,62
Formas de volver a casa	-10,31
2666	-10,68
Lumpérica	-12,76
Estrellas Muertas	-17,26
Condell	-18,20

Papelucho perdido	-0,40
La casa de los espíritus	-0,68
Estrellas Muertas	-0,76
Mala onda	-0,77
Música Marciana	-0,97
El lugar sin límites	-2,71
Patas de perro	-3,63
Hijo de Ladrón	-3,63
Ruido	-4,43
Alhué	-4,77
El loco Estero	-5,45
El socio	-5,85
Martín Rivas	-7,23
Lumpérica	-26,70

TABLA 10

Dimensión 3: Social-Individual

	Factor 3
Condell	14,18
Palomita Blanca	9,87
Papelucho	8,83
La última niebla	7,11
Bonsái	5,60
El último grumete de la Baquedano	4,15
Los detectives salvajes	3,93
Formas de volver a casa	3,40
Papelucho detective	3,10
Papelucho y el marciano	2,99
El niño que enloqueció de amor	1,95
La amortajada	1,27
Papelucho historiador	1,01
Eloy	0,77
Papelucho en vacaciones	0,02
2666	-0,19

Estas tablas merecen cierta interpretación exegética; sin embargo, resulta más apropiado para el presente trabajo desarrollar el análisis con procedimientos computacionales. Se puede hacer un análisis de las dimensiones entendiendo los valores en cada una de ellas como una distribución en un espacio cartesiano, donde los guarismos que se alcanzan sitúan a cada novela en ejes x, y, z. De esta manera resulta posible agrupar las obras por las configuraciones de aglomerados que generan.

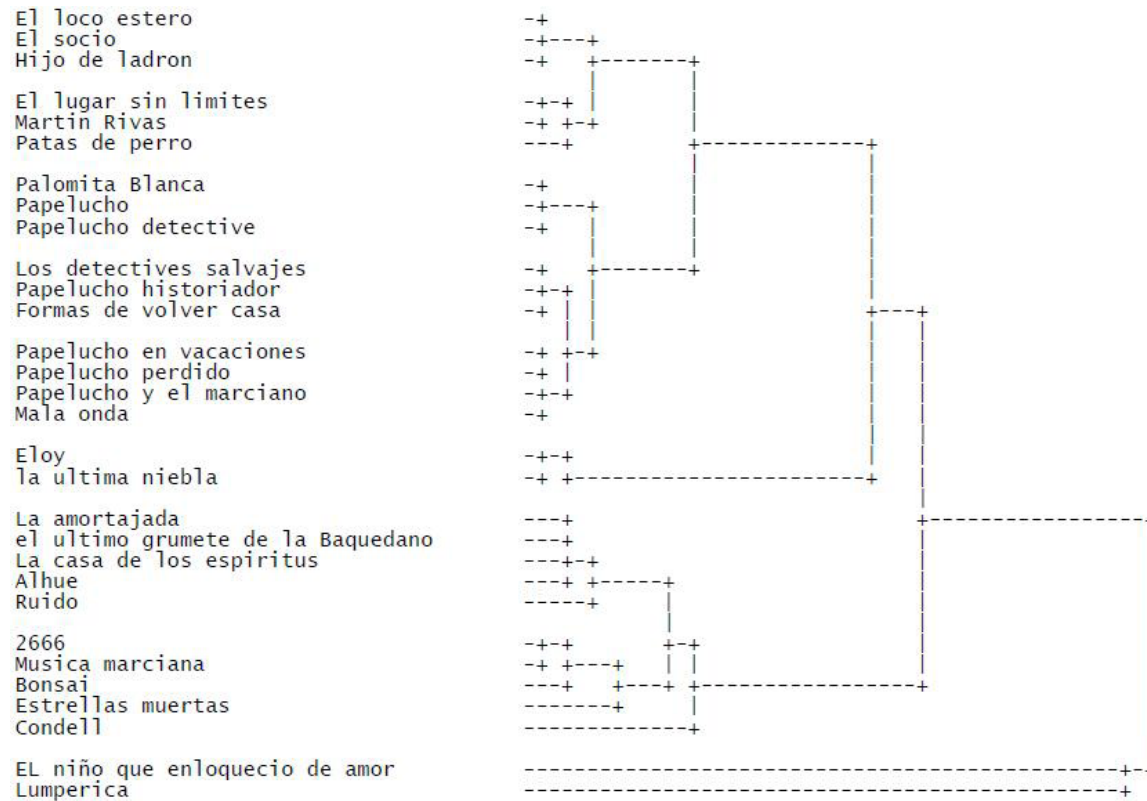
Este tipo de análisis se puede llevar a cabo realizando un proceso de *clustering*, donde los puntajes dimensionales permitan a un programa estadístico como SPSS agrupar a las novelas por su cercanía en las tres dimensiones.

Para ello se ha elaborado el dendrograma siguiente que agrupa las novelas en conjuntos pares en que la tabla indica, con un signo "+", el nivel de cercanía de parejas de novelas o grupos de ellas (véase la tabla 11).

En el dendrograma es posible observar dos conjuntos de agrupaciones bastante claros. El

TABLA 11

Análisis de cluster



primero, desde *El loco estero* hasta *Mala onda*, y el segundo, desde *La amortajada* hasta *Condell*. Quedan fuera de estos dos grupos dos novelas consideradas unánimemente por la crítica literaria como anómalas: *El niño que enloqueció de amor* (cf. Vásquez, 1963) y *Lumpérica* (cf. Donoso, 2009); la primera por su exceso de emocionalidad y la segunda por su explícita y expresa prosa rupturista.

Dentro de estos dos conjuntos mayores se observa que el primero corresponde casi íntegramente a obras del siglo XX, mientras que en el segundo se aglomeran las obras del bolañismo y el posbolañismo (2666, *Ruido*, *Música marciana*, *Bonsái*, etc.). Del mismo modo, es posible advertir que en un conjunto de relaciones dendrográficas bastante estrechas (lo que es marcado en el diagrama por la cercanía de las marcas “+” al lado izquierdo del mismo) entre *El loco Estero*, *El*

socio e *Hijo de ladrón*, obras todas de la primera mitad del siglo XX, y más abajo, entre *El lugar sin límites*, *Martín Rivas* y *Patas de perro*, todas obras, ellas y las tres anteriores, de la “masculinidad”. Los siguientes tres grupos (esto es, desde *Palomita Blanca* hasta *Mala onda*) constituyen un grupo que se podría denominar de novelas de formación (*Bildungsroman*, cf. Arango, 2009), nucleadas en torno a los libros de *Papelucho*...

Además de los rasgos de análisis exegéticos anteriores, es posible advertir que las obras tienden a agruparse de modo cronológico (las obras hasta la primera mitad del siglo XX se agrupan en el dendrograma, tal como las que van desde la segunda mitad del siglo XX hasta la actualidad).

Por ejemplo, si se toman individualmente y en dadas las distancias en el espacio cartesiano, medidas con la fórmula cos alfa:

TABLA 12

Fórmula distancia coseno alfa

$$\cos \alpha = \frac{u_1 \cdot v_1 + u_2 \cdot v_2 + u_3 \cdot v_3}{\sqrt{u_1^2 + u_2^2 + u_3^2} \cdot \sqrt{v_1^2 + v_2^2 + v_3^2}}$$

donde un valor que se aproxima a 1 significa una alta similitud (por ejemplo, entre 2666 y *Ruido*, 0,96) y un valor que se aproxima a -1 significa una alta disimilitud (por ejemplo, entre *Mala onda* y *Alhué*, -0,97).

La tabla que se presenta a continuación expresa las similitudes entre todas las obras y en ella es posible dilucidar que las obras que se aproximan epocalmente tienden a tener valores cercanos, mientras que las que se alejan epocalmente se distancian en los cosenos de sus vectores.

Estos resultados son similares a los que han obtenido Hughes, Foti, Krakauer y Rockmore (2012) en un estudio sobre patrones cuantitativos en narrativa anglosajona, como muestra la tabla.

TABLA 13

Distancias coseno alfa entre las novelas

	2666	Alhue	Bonsai	Condell	El loco estero	El lugar sin limites	EL niño que enloquecio de amor	El socio	El ultimo grumete de la Baquedano	Eloy	Estrellas muertas	Formas de volver casa	Hijo de ladrón	La amortajada	la ultima niebla	La casa de los espíritus	Los detectives salvajes	Lumperica	Mala onda	Martin Rivas	Musica marcialna	Palomita Blanca	Papelucho detective	Papelucho en vacaciones	Papelucho historiador	Papelucho perdido	Papelucho y el marciano	Papelucho	Patas de perro	Ruido
2666	1,00	0,75	0,91	0,91	0,47	-0,76	-1,00	0,48	0,91	-0,52	0,61	-0,28	0,63	-0,26	-0,53	0,75	0,18	-0,33	-0,68	-0,68	1,00	-0,64	-0,72	-0,72	-0,10	-0,68	-0,90	-0,75	-0,88	0,96
Alhue	0,75	1,00	0,54	0,58	0,91	-0,25	-0,80	0,88	0,91	0,16	-0,05	-0,84	0,60	0,42	0,09	0,99	-0,51	-0,53	-0,97	-0,36	0,80	-0,95	-1,00	-0,99	-0,73	-0,98	-0,96	-0,98	-0,35	0,90
Bonsai	0,91	0,54	1,00	1,00	0,16	-0,95	-0,88	0,14	0,84	-0,58	0,64	-0,05	0,28	-0,35	-0,50	0,59	0,42	-0,50	-0,55	-0,91	0,88	-0,34	-0,52	-0,56	0,08	-0,55	-0,69	-0,48	-0,94	0,79
Condell	0,91	0,58	1,00	1,00	0,19	-0,94	-0,88	0,16	0,87	-0,52	0,59	-0,10	0,25	-0,28	-0,43	0,63	0,37	-0,56	-0,59	-0,93	0,88	-0,36	-0,56	-0,60	0,02	-0,59	-0,70	-0,50	-0,92	0,80
El loco estero	0,47	0,91	0,16	0,19	1,00	0,16	-0,55	0,99	0,65	0,39	-0,29	-0,94	0,67	0,59	0,25	0,87	-0,77	-0,28	-0,85	0,06	0,55	-0,98	-0,91	-0,86	-0,83	-0,85	-0,81	-0,94	0,00	0,71
El lugar sin limites	-0,76	-0,25	-0,95	-0,94	0,16	1,00	0,71	0,18	-0,63	0,71	-0,75	-0,26	-0,08	0,54	0,59	-0,31	-0,67	0,40	0,27	0,92	-0,71	0,03	0,23	0,28	-0,35	0,27	0,43	0,18	0,95	-0,57
EL niño que enloquecio de amor	-1,00	-0,80	-0,88	-0,88	-0,55	0,71	1,00	-0,55	-0,93	0,45	-0,55	0,36	-0,66	0,19	0,48	-0,80	-0,10	0,35	0,73	0,64	-1,00	0,71	0,77	0,77	0,17	0,73	0,93	0,80	0,83	-0,98
El socio	0,48	0,88	0,14	0,16	0,99	0,18	-0,55	1,00	0,61	0,33	-0,23	-0,89	0,74	0,52	0,17	0,83	-0,74	-0,18	-0,80	0,12	0,55	-0,98	-0,88	-0,82	-0,78	-0,80	-0,81	-0,94	-0,01	0,70
El ultimo grumete de la Baquedano	0,91	0,91	0,84	0,87	0,65	-0,63	-0,93	0,61	1,00	-0,14	0,24	-0,58	0,45	0,14	-0,13	0,93	-0,13	-0,65	-0,91	-0,71	0,93	-0,76	-0,90	-0,92	-0,46	-0,91	-0,93	-0,84	-0,67	0,95
Eloy	-0,52	0,16	-0,58	-0,52	0,39	0,71	0,45	0,33	-0,14	1,00	-0,99	-0,65	-0,33	0,96	0,97	0,17	-0,87	-0,36	-0,27	0,39	-0,46	-0,20	-0,21	-0,22	-0,79	-0,27	0,13	-0,08	0,81	-0,28
Estrellas muertas	0,61	-0,05	0,64	0,59	-0,29	-0,75	-0,55	-0,23	0,24	-0,99	1,00	0,57	0,40	-0,92	-0,98	-0,06	0,82	0,31	0,17	-0,44	0,55	0,10	0,10	0,12	0,72	0,16	-0,24	-0,03	-0,86	0,38
Formas de volver casa	-0,28	-0,84	-0,05	-0,10	-0,94	-0,26	0,36	-0,89	-0,58	-0,65	0,57	1,00	-0,36	-0,82	-0,56	-0,83	0,89	0,50	0,86	-0,03	-0,35	0,87	0,87	0,85	0,97	0,86	0,66	0,81	-0,20	-0,54
Hijo de ladrón	0,63	0,60	0,28	0,25	0,67	-0,08	-0,66	0,74	0,45	-0,33	0,40	-0,36	1,00	-0,18	-0,51	0,51	-0,18	0,34	-0,41	0,13	0,67	-0,75	-0,56	-0,46	-0,15	-0,41	-0,73	-0,76	-0,38	0,71
La amortajada	-0,26	0,42	-0,35	-0,28	0,59	0,54	0,19	0,52	0,14	0,96	-0,92	-0,82	-0,18	1,00	0,93	0,44	-0,91	-0,53	-0,53	0,21	-0,19	-0,43	-0,47	-0,49	-0,93	-0,53	-0,14	-0,33	0,63	0,00
la ultima niebla	-0,53	0,09	-0,50	-0,43	0,25	0,59	0,48	0,17	-0,13	0,97	-0,98	-0,56	-0,51	0,93	1,00	0,13	-0,75	-0,49	-0,24	0,23	-0,48	-0,07	-0,14	-0,19	-0,73	-0,24	0,21	0,03	0,75	-0,33
La casa de los espíritus	0,75	0,99	0,59	0,63	0,87	-0,31	-0,80	0,83	0,93	0,17	-0,06	-0,83	0,51	0,44	0,13	1,00	-0,48	-0,63	-0,99	-0,45	0,80	-0,91	-0,99	-1,00	-0,73	-0,99	-0,94	-0,94	-0,38	0,90
Los detectives salvajes	0,18	-0,51	0,42	0,37	-0,77	-0,67	-0,10	-0,74	-0,13	-0,87	0,82	0,89	-0,18	-0,91	-0,75	-0,48	1,00	0,23	0,53	-0,44	0,10	0,62	0,54	0,51	0,92	0,53	0,27	0,50	-0,62	-0,11
Lumperica	-0,33	-0,53	-0,50	-0,56	-0,28	0,40	0,35	-0,18	-0,65	-0,36	0,31	0,50	0,34	-0,53	-0,49	-0,63	0,23	1,00	0,70	0,72	-0,34	0,29	0,56	0,66	0,57	0,70	0,39	0,33	0,20	-0,39
Mala onda	-0,68	-0,97	-0,55	-0,59	-0,85	0,27	0,73	-0,80	-0,91	-0,27	0,17	0,86	-0,41	-0,53	-0,24	-0,99	0,53	0,70	1,00	0,46	-0,73	0,88	0,98	1,00	0,78	1,00	0,89	0,90	0,30	-0,84
Martin Rivas	-0,68	-0,36	-0,91	-0,93	0,06	0,92	0,64	0,12	-0,71	0,39	-0,44	-0,03	0,13	0,21	0,23	-0,45	-0,44	0,72	0,46	1,00	-0,64	0,09	0,36	0,45	-0,06	0,45	0,44	0,22	0,79	-0,55
Musica marcialna	1,00	0,80	0,88	0,88	0,55	-0,71	-1,00	0,55	0,93	-0,46	0,55	-0,35	0,67	-0,19	-0,48	0,80	0,10	-0,34	-0,73	-0,64	1,00	-0,70	-0,77	-0,76	-0,17	-0,73	-0,93	-0,80	-0,84	0,98
Palomita Blanca	-0,64	-0,95	-0,34	-0,36	-0,98	0,03	0,71	-0,98	-0,76	-0,20	0,10	0,87	-0,75	-0,43	-0,07	-0,91	0,62	0,29	0,88	0,09	-0,70	1,00	0,95	0,90	0,73	0,88	0,91	0,99	0,20	-0,83

Papelucho detective	-0,72	-1,00	-0,52	-0,56	-0,91	0,23	0,77	-0,88	-0,90	-0,21	0,10	0,87	-0,56	-0,47	-0,14	-0,99	0,54	0,56	0,98	0,36	-0,77	0,95	1,00	0,99	0,76	0,98	0,94	0,97	0,32	-0,88
Papelucho en vacaciones	-0,72	-0,99	-0,56	-0,60	-0,86	0,28	0,77	-0,82	-0,92	-0,22	0,12	0,85	-0,46	-0,49	-0,19	-1,00	0,51	0,66	1,00	0,45	-0,76	0,90	0,99	1,00	0,76	1,00	0,92	0,93	0,33	-0,87
Papelucho historiador	-0,10	-0,73	0,08	0,02	-0,83	-0,35	0,17	-0,78	-0,46	-0,79	0,72	0,97	-0,15	-0,93	-0,73	-0,73	0,92	0,57	0,78	-0,06	-0,17	0,73	0,76	0,76	1,00	0,78	0,49	0,66	-0,36	-0,36
Papelucho perdido	-0,68	-0,98	-0,55	-0,59	-0,85	0,27	0,73	-0,80	-0,91	-0,27	0,16	0,86	-0,41	-0,53	-0,24	-0,99	0,53	0,70	1,00	0,45	-0,73	0,88	0,98	1,00	0,78	1,00	0,89	0,90	0,30	-0,84
Papelucho y el marciano	-0,90	-0,96	-0,69	-0,70	-0,81	0,43	0,93	-0,81	-0,93	0,13	-0,24	0,66	-0,73	-0,14	0,21	-0,94	0,27	0,39	0,89	0,44	-0,93	0,91	0,94	0,92	0,49	0,89	1,00	0,97	0,58	-0,99
Papelucho	-0,75	-0,98	-0,48	-0,50	-0,94	0,18	0,80	-0,94	-0,84	-0,08	-0,03	0,81	-0,76	-0,33	0,03	-0,94	0,50	0,33	0,90	0,22	-0,80	0,99	0,97	0,93	0,66	0,90	0,97	1,00	0,35	-0,91
Patitas de perro	-0,88	-0,35	-0,94	-0,92	0,00	0,95	0,83	-0,01	-0,67	0,81	-0,86	-0,20	-0,38	0,63	0,75	-0,38	-0,62	0,20	0,30	0,79	-0,84	0,20	0,32	0,33	-0,36	0,30	0,58	0,35	1,00	-0,70
Ruido	0,96	0,90	0,79	0,80	0,71	-0,57	-0,98	0,70	0,95	-0,28	0,38	-0,54	0,71	0,00	-0,33	0,90	-0,11	-0,39	-0,84	-0,55	0,98	-0,83	-0,88	-0,87	-0,36	-0,84	-0,99	-0,91	-0,70	1,00

5. Discusión

Por todo lo anterior, es posible señalar que el modelo biberiano, aplicado de manera similar a Martínez (2012), permite una interpretabilidad literaria, sobre todo en lo que respecta a las interpretaciones de carácter literario formal, temático y cronológico, de la literatura narrativa chilena. No obstante lo anterior, y en un desarrollo posterior de la metodología aplicada en el presente trabajo, deberá profundizarse en los principios de análisis literario que se puedan llegar a clarificar con ese estudio.

En efecto, si se considera la cronología en este tipo de análisis, surgen nuevas informaciones.

Se ha creado, en consecuencia, una nueva columna en las tablas, con los años de edición de las obras, y se ha calculado todas las correlaciones Pearson entre estos valores y los de los indicadores originales. Se asume que los valores cercanos y superiores (de modo absoluto) a 0,70 pueden ser buenos indicadores de la “evolución de las obras”. Solo tres valores han quedado en ese rango, con una correlación negativa:

- Porcentaje de palabras afectivas (-0,729) (acariciar, aburrir, fracaso, etc.)

- Porcentaje de emociones positivas (-0,726) (ganar, inspiración, pasión, etc.)
- Porcentaje de sensaciones positivas (-0,681) (éxtasis, goce, humor)

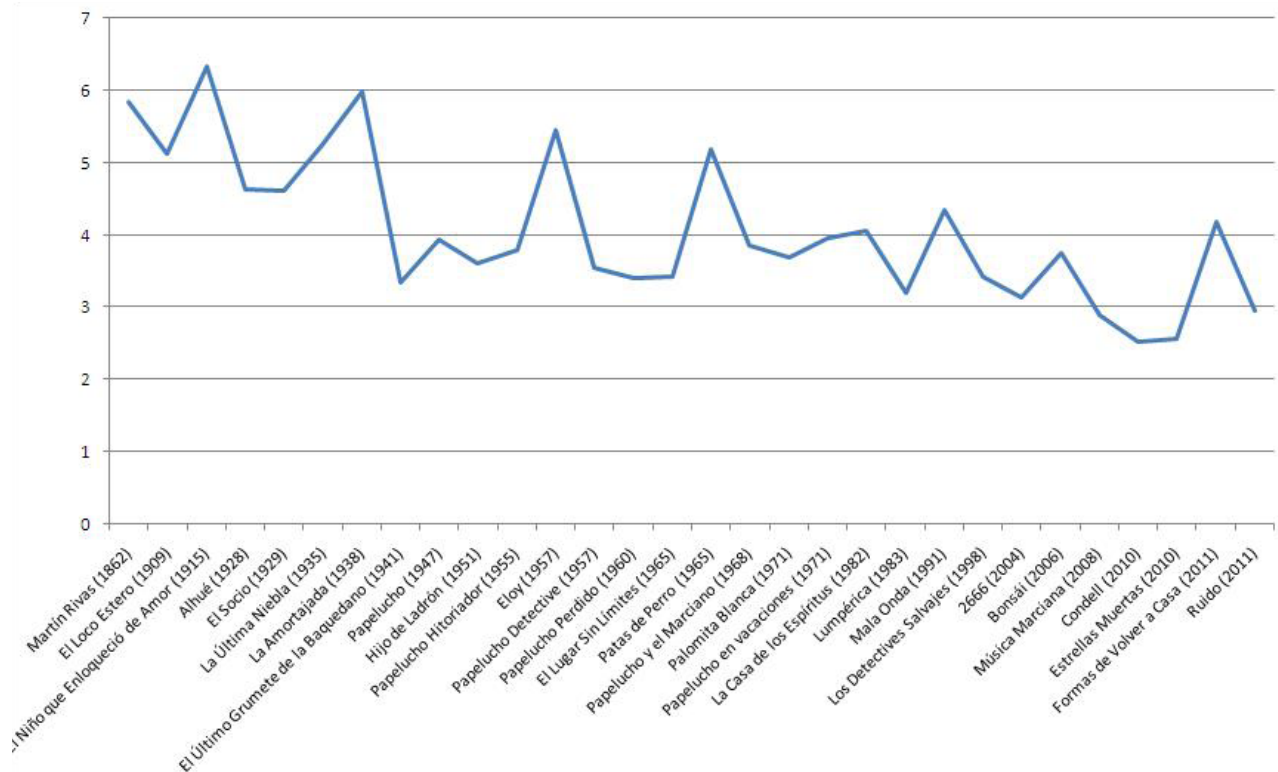
Esto significa que, a medida que pasan los años, el porcentaje de palabras afectivas (y de emociones y sensaciones positivas) disminuye sistemáticamente. Se debe indicar que también obtienen altos valores las emociones negativas (que se incluyen asimismo en las palabras afectivas).

Parece ser, en consecuencia, que, a medida que avanzan los años, las novelas van desensibilizando su escritura, evitando el ámbito emocional (valor mayor para *El niño que enloqueció de amor* -6,34-; valor menor para *Condell* -2,52-), como se muestra en el siguiente gráfico de la evolución de los porcentajes de palabras afectivas ordenadas por años de las novelas (véase la tabla 14).

El resultado de este trabajo deberá ser ampliado en el futuro aplicando el diseño metodológico esta vez para cubrir un número de obras más alto y también para ocupar nuevas clasificaciones léxicas emanadas de la misma investigación.

TABLA 14

Desensibilización de las novelas



6. Bibliografía citada

ARANGO RODRÍGUEZ, Selen, 2009: "La novela de formación y sus relaciones con la pedagogía y los estudios literarios", *Folios* 30, 127-146.

BIBER, Douglas, 1988: *Variation Across Speech and Writing*, Cambridge: Cambridge University Press.

BIBER, Douglas, Ulla CONNOR y Thomas UPTON, 2007: *Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure*, Amsterdam: John Benjamins Publishing.

CANOVAS, Rodrigo, 1997: *Novela chilena, nuevas generaciones, el abordaje de los huérfanos*, Santiago de Chile: Ediciones Universidad Católica de Chile.

CUADROS, Ricardo, 1997: "Crítica literaria y fin de siglo: (Rodrigo Cánovas, Novela Chilena, Nuevas

generaciones, el abordaje de los huérfanos)", *Literatura y lingüística* 10, 232-242.

DONOSO, Jaime, 2009: "Práctica de la Avanzada: Lumpérica y la figuración de la escritura como fin de la representación burguesa de la literatura y el arte" en Rubí CARREÑO (ed.): *Diamela Eltit: redes locales, redes globales*, Madrid: Iberoamericana, 239-260.

FLETCHER, William, 2002: "KFNGram" [<http://www.kwicfinder.com/>, fecha de consulta: 25 de julio de 2014].

GAVINS, Joanna y Gerard STEEN (eds.), 2003: *Cognitive Poetics in Practice*, Londres: Routledge.

HUGHES, James, Nicholas FOTI, David KRAKAUER y Daniel ROCKMORE, 2012: "Quantitative patterns of stylistic influence in the evolution of literature",

Proceedings of the National Academy of Sciences, 109, 20.

LANCASHIRE, Ian, 1993: "Computer-Assisted Critical Analysis: A Case Study of Margaret Atwood's *Handmaid's Tale*" en George LANDOW y Paul DELANEY (eds.): *The Digital Word: Text-based Computing in the Humanities*, Cambridge, Mass.: The MIT Press, 293-318.

GOIC, Cedomil, 1968: *La novela chilena: los mitos degradados*, Santiago de Chile: Editorial Universitaria.

GOIC, Cedomil, 1972: *Historia de la novela hispanoamericana*, Valparaíso: Ediciones Universitarias de Valparaíso.

GOODWIN, Jonathan, John HOLBO y Franco MORETTI (eds.), 2011: *Reading Graphs, Maps, and Trees: Responses to Franco Moretti*, Anderson, SC.: Parlor Press.

LANGUAGE IN SOCIAL MEDIA, 2011: *Workshop on Language in Social Media – LSM 2011*, Portland, Oregon: The Association for Computational Linguistics.

MARINKOVICH, Juana y Yanina CADEMARTORI, 2004: "Foco narrativo y foco informativo: Dos dimensiones para una descripción de los manuales en la formación técnico-profesional", *Revista Signos* 37(55), 31-40.

MARTÍNEZ, Ricardo, 2012: "Can you read my mind?: la historia de la mente en cinco novelas anglosajonas modernas" en Mike WILSON (ed.): *Where is my mind? Cognición, Literatura y Cine*, Santiago: Cuarto propio, 135-170.

MORETTI, Franco, 2005: *Graphs Maps Trees. Abstract Models for a Literary Theory*, London, New-York: Verso.

MORETTI, Franco, 2011: "Network analysis, plot analysis", *Literary Lab* 2, 1-42.

PARODI, Giovanni, 2004: "Lingüística de corpus y análisis multidimensional: exploración de la variación en el corpus PUCV 2003", *Revista Española de Lingüística* 31, 1, 45-76.

PEIRCE, Charles Sanders, 1932: *Collected Papers of Charles Sanders Peirce*, Cambridge, MA.: Harvard University Press.

PENNEBAKER, James, Cindy CHUNG, Molly IRELAND, Amy GONZALES y Roger BOOTH, 2007: *The Development and Psychometric Properties of LIWC2007*, Austin, Texas: LIWC.net.

REESE, Charles y Charles LOCHMÜLLER, 1994: "Introduction to Factor Análisis" [www.chem.duke.edu/~reese/tutor1/factucmp.html]. 1998, fecha de consulta: 25 de julio de 2014].

STOCKWELL, Peter, 2002: *Cognitive Poetics: an Introduction*, Routledge: Londres.

VÁSQUEZ, Ángel, 1963: "Los tres planos de la creación artística de Eduardo Barrios", *Revista Iberoamericana* XXIX, 55, 125-37.