



UNIVERSIDAD AUTÓNOMA DE  
QUERÉTARO  
FACULTAD DE LENGUAS Y LETRAS

Aproximación a la literatura cubana desde la perspectiva de  
la lingüística de corpus

TESIS

Que como parte de los requisitos para obtener el Grado de

Maestro en Lingüística

Presenta

Danilo Orlando Vargas Nardiz

Dirigido por

Ignacio Rodríguez Sánchez

Querétaro, Qro a. 14 de abril de 2023



Dirección General de Bibliotecas y Servicios Digitales  
de Información



Aproximación a la literatura cubana desde la  
perspectiva de la lingüística de corpus

**por**

Danilo Orlando Vargas Nardiz

se distribuye bajo una [Licencia Creative Commons  
Atribución-NoComercial-SinDerivadas 4.0  
Internacional](#).

**Clave RI:** LLMAC-302229



UNIVERSIDAD AUTÓNOMA DE QUERÉTARO  
FACULTAD DE LENGUAS Y LETRAS

Aproximación a la literatura cubana desde la perspectiva de la lingüística de  
corpus

TESIS

Que como parte de los requisitos para obtener el Grado de

Maestro en Lingüística

Presenta

Danilo Orlando Vargas Nardiz

Dirigido por

Ignacio Rodríguez Sánchez

Ignacio Rodríguez Sánchez, Presidente

Eva P. Velásquez Upegui, Secretaria

Jeanie Murphy, Vocal

Antonio Reyes Pérez, Suplente

Ana R. Sánchez Barrera, Suplente

Centro Universitario, Querétaro, Qro.

Fecha de aprobación por el Consejo Universitario Abril 2023

México

*“Pero el hombre no está hecho para la derrota.  
Un hombre puede ser destruido, pero no derrotado.*

-El viejo y el mar-  
Ernest Hemingway.

A Carlos Alberto

A mi familia y mis amigos de toda la vida

A mi familia mexicana

## **Agradecimientos**

Mi más profundo agradecimiento a la Universidad Autónoma de Querétaro y a su personal docente por la oportunidad de superación profesional e intelectual. Agradezco al CONACYT por el apoyo que brinda a los estudiantes extranjeros, sin el cual sería virtualmente imposible continuar estudios de posgrado en el extranjero.

Agradezco al Dr. Ignacio Rodríguez Sánchez por su ayuda y apoyo incondicional desde el principio de la maestría. Por su paciencia y sus palabras de aliento. Y por la libertad que me dio para explorar temas y enfoques desconocidos para mí.

Igualmente a la Dra. Eva Velásquez Upegui, por su constante apoyo y motivación. Por fomentar mi curiosidad y expandir mis horizontes investigativos.

A mi familia y mis amigos sin cuyo apoyo nada de esto hubiera sucedido.

## Contenido

1	Introducción .....	10
2	Marco teórico.....	14
2.1	Lingüística de Corpus .....	15
2.1.1	Tipos de corpus: diseños y representatividad .....	17
2.1.2	Palabras Clave, nodos y colocaciones.....	23
2.1.3	Análisis de sentimientos .....	27
2.2	Lectura Distante .....	32
3	Antecedentes .....	35
3.1	Estudios literarios desde la lingüística de Corpus.....	36
3.2	Estudios sobre literatura cubana.....	40
4	Planteamiento .....	47
4.1	Planteamiento del problema .....	47
4.2	Hipótesis .....	48
4.3	Objetivos .....	48
4.4	Preguntas de investigación.....	48
5	Metodología .....	49
5.1	Descripción del Corpus .....	50
5.2	Preparación del corpus experimental .....	53
5.3	Extracción de palabras clave .....	55
5.4	Colocaciones. Líneas de concordancia. Sentimientos.....	59
5.5	Análisis de agrupación .....	64
6	Análisis y resultados.....	70
6.1	Determinación de palabras clave.....	71
6.2	La migración como tema central dentro de la literatura cubana post 1959.....	76
6.2.1	Líneas de concordancia. Análisis de sentimientos .....	83
6.2.2	Análisis de agrupación. Resultados .....	94
7	Discusión y conclusiones .....	99
8	Bibliografía .....	105
9	Apéndice 1. Listado de autores y novelas .....	110
10	Apéndice 2. Listados Palabras Clave determinadas en cada uno de los cruces. ....	112
11	Apéndice 3. Frecuencias relativas de los rasgos utilizados para el análisis de agrupación ..	129

## **Índice de Imágenes**

Imagen 1. Líneas de concordancia. Palabra clave CUBA.....	26
Imagen 2. Colocados de la palabra Cuba.....	27
Imagen 3. Texto recuperado en .pdf vs. texto procesado en .word .....	54
Imagen 4. Texto procesado y etiquetado en Notepad++.....	55
Imagen 5. Colocaciones más frecuentes con el sustantivo locativo Cuba. CLCXX+2 .....	61
Imagen 6. Líneas de concordancia sustantivo locativo Cuba y formas derivadas de regresar .....	62
Imagen 7. Procesamiento de los datos con el software Brezina ToolBox.....	68
Imagen 8. Colocaciones con el sustantivo locativo Cuba. Sección CLCXX+2 .....	77
Imagen 9. Dispersión de las colocaciones sal* de Cuba en la sección CLCXX+2 .....	78
Imagen 10. Colocaciones con el sustantivo locativo Cuba. Sección CLCXX+1 .....	81
Imagen 11. Agrupación de los textos literarios del CLCXX+ de acuerdo a los rasgos lingüísticos analizados.....	95

## **Índice de Tablas**

Tabla 1. Ejemplo de frecuencias relativas en dos corpus.....	24
Tabla 2. Listado de las primeras 60 palabras más frecuentes resultantes del cruce CLCXX+2 vs. CLCXX+1.....	59
Tabla 3. Frecuencia relativa de 17 rasgos en 9 novelas. Sección CLCXX+1 .....	66
Tabla 4. Verbos de movimiento .....	80
Tabla 5. Total de sentimientos y coocurrencias verbales.....	94
Tabla 6. Listado de novelas y autores que conforman el CLCXX+.....	111
Tabla 7. Primeras 300 Palabras Clave. Cruce CREA vs CLCXX+1 .....	115
Tabla 8. Primeras 300 Palabras Clave. Cruce Crea vs. CLCXX+2 .....	121
Tabla 9. Primeras 300 Palabras Clave. Cruce CLCXX+1 vs. CLCXX+2 .....	125
Tabla 10. Primeras 300 Palabras Clave. Cruce CLCXX+2 vs. CLCXX+1 .....	128



## Resumen

La literatura, como cualquier otra forma de arte, es parte de la identidad de una nación. En ella se recogen las costumbres, puntos de vista, temores y esperanzas de un pueblo. Gracias a los libros somos capaces de conocer y tal vez, entender la forma de vida de una determinada sociedad en una época determinada.

Esta investigación emplea metodología propia de la lingüística de corpus para plasmar el enfoque de *lectura distante* propuesto por Franco Moretti (2013). Este enfoque se utiliza para analizar un corpus de narrativa cubana contemporánea. Así, el objetivo de este trabajo es identificar diferencias temáticas en la narrativa cubana previa a 1959 y la posterior a esta fecha. Se toma como referencia 1959 porque es el año del triunfo de la revolución cubana.

Para ello se diseñó un corpus de 45 textos de narrativa (cuento y novela) cubana contemporánea, que se estudió con técnicas propias de la lingüística de corpus. En primer lugar, se realizaron tres análisis de palabras clave: cada una de las dos secciones se contrastaron con el CREA y posteriormente entre ellas.

Después, estas palabras clave se agruparon manualmente para identificar los temas característicos de cada período. Así, se identificaron temas como la migración, la etnicidad, la tierra, la revolución y género. Uno de los hallazgos más interesantes resultó ser la aparición del tema *migración* durante el segundo período, el cual no fue de interés para los escritores cubanos del período prerrevolucionario. Esto parece estar condicionado por los cambios sociales, económicos y políticos que se generaron a partir de 1959.

A continuación, se llevó a cabo un análisis cualitativo de líneas de concordancia y se identificaron una serie de sentimientos –no siempre positivos– vinculados a ese fenómeno: ansiedad, miedo, incertidumbre, alivio y tristeza, entre otros.

Finalmente, se realizó un análisis de agrupación jerárquica que permite evaluar cuantitativamente qué tan bien discriminan estos sentimientos las dos secciones del corpus. El dendrograma generado señala unos resultados muy prometedores por la precisión en la clasificación de los textos.

Todo esto confirma la pertinencia de combinar métodos propios de la lingüística de corpus en enfoques propios de los estudios literarios y culturales.

## **Abstract**

Literature, like any other form of art, is part of a nation's identity. It depicts the customs, points of view, fears and hopes of the people. Thanks to books we are able to know and perhaps understand the way of life of a certain society at a certain time.

This research uses the methodology of corpus linguistics to capture the distant reading approach proposed by Franco Moretti (2013). This approach is used to analyze a corpus of contemporary Cuban narrative. Thus, the objective of this work is to identify thematic differences in the Cuban narrative prior to 1959 and the one after this date. 1959 is taken as a reference because it is the year of the triumph of the Cuban revolution.

In order to accomplish this goal, a corpus of 45 contemporary Cuban narrative texts (story and novel) was designed, which was studied with techniques typical of corpus linguistics. First, three keyword analyses were carried out: each of the two sections were contrasted with the CREA and later between them.

These keywords were manually grouped to identify the characteristic themes of each period. Thus, issues such as migration, ethnicity, land, revolution and gender were identified. One of the most interesting findings turned out to be the appearance of the migration theme during the second period, which was not of interest to Cuban writers of the pre-revolutionary period. This seems to be conditioned by the social, economic and political changes that began in 1959.

Next, a qualitative analysis of the lines of agreement was carried out and a series of feelings –not always positive– linked to this phenomenon were identified: anxiety, fear, uncertainty, relief and sadness, among others.

Finally, a hierarchical clustering analysis in order to quantitatively assess how well these feelings are discriminated in the two sections of the corpus. The generated dendrogram shows very promising results for the precision in the classification of the texts.

All of this confirms the relevance of combining methods from corpus linguistics with approaches from literary and cultural studies.

# 1 Introducción

La literatura, como cualquier otra forma de arte, es parte de la identidad de una nación. En ella se recogen las costumbres, puntos de vista, temores y esperanzas de un pueblo. Gracias a los textos literarios somos capaces de conocer y, tal vez, entender la forma de vida de una determinada sociedad en una época determinada. La literatura se nutre del imaginario colectivo y los autores, por lo general y de diversas maneras, expresan sus puntos de vista particulares sobre lo que sea que esté aconteciendo en esa sociedad, en ese momento.

Según Roberto Fernández Retamar (2000), el historiador cubano Max Henríquez Ureña plantea que “el desenvolvimiento de la vida literaria en Cuba está tan estrechamente ligado al de la historia política, que se hace imposible disociarlo”. Así, no es de extrañar que los temas de la literatura cubana siempre hayan estado condicionados por la situación política, generalmente turbulenta, del país. Durante cuatro siglos Cuba estuvo dominada por alguna potencia extranjera, primero España y después Estados Unidos.

El 1ro de enero de 1959 triunfa la Revolución Cubana, un movimiento político-social liderado por Fidel Castro. El triunfo de la Revolución Cubana significó un cambio brusco, radical, en todos los aspectos de la vida en Cuba. El acceso de todo el pueblo a la educación y a la cultura fue una de las principales banderas enarboladas por el movimiento. El romance de los intelectuales de izquierda, estadounidenses, latinoamericanos y europeos, con la Revolución Cubana había comenzado incluso antes del triunfo de la Revolución. Figuras de la talla de Ernest Hemingway ya se habían manifestado a favor del movimiento liderado por Fidel Castro. Otras figuras importantes como Jean Paul Sartre y Simone de Beauvoir también visitaron la Habana a principios de los ´60 para ver de primera mano la concepción de “hombre nuevo”. Dentro de América Latina figuras como Gabriel García Márquez y Mario

Vargas Llosa comulgaron con la Revolución, al menos por un tiempo. El uno fue amigo de Fidel Castro hasta su muerte, el otro se alejó rotundamente del proceso.

A pocos años del triunfo revolucionario Fidel Castro dejó claro el derrotero cultural de la nación. En un encuentro de tres sesiones sostenido en 1961 en la Biblioteca Nacional de Cuba, Fidel Castro (1961) planteó:

“Esto significa que, dentro de la Revolución, todo; contra la Revolución, nada. Contra la Revolución nada, porque la Revolución tiene también sus derechos; y el primer derecho de la Revolución es el derecho a existir (...)

¿Cuáles son los derechos de los escritores y de los artistas, revolucionarios o no revolucionarios? Dentro de la Revolución, todo; contra la Revolución, ningún derecho (APLAUSOS)”.

A partir de estas ideas del líder revolucionario cubano se puede asumir que la política cultural cubana, en particular la literaria, estuvo supeditada a los intereses del nuevo gobierno. Igualmente se puede pensar que la libertad creativa y la libertad de expresión estuvieron limitadas y condicionadas. Muchos escritores que no estuvieron de acuerdo con lo planteado ese día en la Biblioteca Nacional fueron condenados al ostracismo o al exilio.

Partiendo de estas ideas de Fidel Castro y de las consecuencias que de ellas se derivaron, surge la idea de tratar de determinar y analizar los temas más frecuentes en la literatura cubana en el período post 1959. Sin embargo, cualquier análisis sobre esta temática parecería incompleto si se tomaran en cuenta solo los escritores que escribieron en Cuba durante este período. Muchos se exiliaron, más o menos voluntariamente, y continuaron su obra fuera de la isla; pero su marco de referencia siguió siendo Cuba. Por otra parte, sería imposible determinar si, en efecto, la Revolución Cubana tuvo un impacto real en la temática literaria si no se analizaran los temas de la literatura cubana antes de 1959.

Así, este trabajo se propone investigar rasgos lingüísticos que permitan determinar y describir los diferentes temas presentes en la literatura postrevolucionaria en Cuba con métodos propios de la Lingüística del Corpus; con el objetivo de precisar si, efectivamente, el triunfo de la Revolución Cubana tuvo un impacto en la vida cultural y literaria de la nación. Para ello se analizarán dos corpus de obras literarias (novelas y cuentos) de escritores cubanos que desarrollaron su obra dentro y fuera de Cuba. Estos dos corpus corresponden a dos períodos: 1902-1958 y 1959-2020.

La investigación está dividida en siete capítulos. En el Capítulo 2, Marco teórico, se tratan los temas relativos a las bases teóricas que sustentan esta investigación. Así, se hace un acercamiento a la Lingüística de Corpus, sus principales conceptos y los diferentes enfoques desde los que se puede abordar la creación de un corpus y su estudio. Además, se hace una revisión de los postulados relativos al análisis de sentimientos y a la lectura distante.

En el Capítulo 3, Antecedentes, se hace un repaso de diferentes investigaciones que probaron ser relevantes para este trabajo, ya sea porque aplicaron una metodología de análisis semejante o porque se ocuparon de temas similares. Es importante mencionar en este punto, que si bien es cierto que se registraron investigaciones literarias que utilizan la Lingüística de Corpus como metodología e investigaciones sobre literatura cubana, no se encontraron investigaciones sobre la literatura cubana desde la Lingüística de Corpus. En este aspecto esta es una investigación pionera.

El Capítulo 4 retoma lo relativo al planteamiento del problema y en el Capítulo 5 se describe la metodología seguida para el análisis. Así, en este capítulo se revisan a fondo los parámetros seguidos para la confección del corpus y su preparación. Se abordan, además, de una manera

práctica los conceptos discutidos en el Capítulo 2, a saber, palabras claves, colocaciones, líneas de concordancia y sentimientos.

Estos conceptos se tratan más a fondo en el Capítulo 6, en tanto se analizan como parte de los textos que interesan a esta investigación y se describe su funcionamiento dentro de los mismo. En el Capítulo 7 se discuten las implicaciones de los resultados obtenidos en el capítulo anterior.

Este trabajo, más que realizar un análisis exhaustivo de la literatura cubana del siglo XX y en especial, de la literatura cubana post 1959, pretende proponer una aproximación diferente a la de los estudios literarios previos en torno a este tema. Además, se pretende demostrar cómo con herramientas propias de la Lingüística de Corpus es posible realizar estudios cuyas repercusiones van más allá del campo lingüístico en tanto los patrones que esta metodología identifica pueden, en muchas ocasiones, tener implicaciones sociales y psicológicas. En definitiva, se espera demostrar qué cambios de índole social, política y económica suponen un impacto que trasciende las fibras más crudas de una sociedad y puede tocar las más sutiles, aquellas que atañen la cultura y la cosmovisión de una sociedad determinada en una época determinada.

## 2 Marco teórico

Diversos y abundantes son los estudios que se han realizado acerca de la literatura cubana; sin embargo, hasta donde se ha podido constatar a partir de las consultas bibliográficas realizadas para esta investigación, ninguno desde el enfoque metodológico de la Lingüística de Corpus (L.C de aquí en adelante). Abordar el análisis literario de una obra o de un corpus de obras es una tarea, a priori, harto compleja, sobre todo por lo tardado que pudiera resultar revisar manualmente el corpus en cuestión. Así, en este trabajo se pretende hacer un análisis de los temas dominantes en la literatura cubana después de 1959 desde la perspectiva de la Lingüística de Corpus. Pero, ¿qué es la Lingüística de Corpus?

En el presente capítulo se explicará precisamente qué es la Lingüística de Corpus y se hará una aproximación a los fundamentos teóricos propios de la misma y que se han tomado en cuenta para desarrollar esta investigación. Así, se revisarán algunos de los principales autores que han abordado el tema como McEnery & Hardie, Molina & Sierra y Rojo Sánchez. La intención de este capítulo es acercarse a los distintos tipos de corpus que existen, las intenciones con las que se confecciona cada uno y a la manera de hacerlo. Además, se abordarán conceptos fundamentales dentro de esta disciplina como *palabras clave*, *frecuencia*, *dispersión* y *representatividad*.

Asimismo, se examinarán los fundamentos teóricos relativos al análisis de sentimientos. Con respecto al análisis de sentimientos se revisarán los postulados de Liu (2015) relativos a la minería de opiniones, metodología que se adaptó a los propósitos de esta investigación. Asimismo se discutirán conceptos relevantes como las *palabras sentimientos* (Qiu et al., 2009). Además, se revisará la taxonomía de sentimientos propuesta por Marina & Penas (2000) la cual se utilizó con el fin de discretizar correctamente los sentimientos encontrados

en el corpus. Como se verá más adelante<sup>1</sup>, gracias al análisis de sentimientos se pueden determinar las posturas de los personajes con respecto a su contexto, las maneras en que las perciben y valoran. A partir de estas valoraciones se puede determinar el impacto de determinada situación en sus vidas. En el caso particular de esta investigación, de que manera influyen los temas que se analizarán en sus vidas y su entorno social.

Finalmente se revisará el concepto de lectura distante propuesta por Franco Moretti (2013) el cual es un pilar fundamental de esta investigación, en tanto es un enfoque que permite el estudio general de un gran número de textos en busca de patrones que los conecte en lugar de estudiar de manera exhaustiva solo un número pequeño de estos.

## **2.1 Lingüística de Corpus**

La L.C. puede definirse como un tipo de aproximación metodológica al análisis electrónico de uno o varios corpus para encontrar patrones lingüísticos o textuales. Por lo general el tamaño de la muestra<sup>2</sup> que se analiza es demasiado grande para hacerlo manualmente en un período de tiempo razonable, o sea, dentro de los plazos de la investigación. Y es esto, justamente, lo que justifica el uso de herramientas computarizadas. Las investigaciones que utilizan este enfoque metodológico lo hacen para validar, refutar o redefinir las teorías o hipótesis que se hayan planteado para determinada investigación (McEnery & Hardie, 2012).

Según McEnery & Hardie la L.C. no se enfoca en el estudio de ningún aspecto particular del idioma. Es, más bien, un área de investigación que abarca un grupo de procedimientos o métodos para estudiar el lenguaje. Es una metodología que facilita el estudio de un

---

<sup>1</sup> Ver sección 6.2.1

<sup>2</sup> Población o universo: Conjunto de todos los casos que concuerdan con determinadas especificaciones. Muestra: Subgrupo del universo o población del cual se recolectan los datos y que debe ser representativo de ésta. (Hernández Sampieri et al., 2014)



determinado aspecto del lenguaje. Los procedimientos de los que se vale la L.C. permiten realizar estudios desde perspectivas que no se han abordado anteriormente.

De acuerdo a Rojo Sánchez (2010), la L.C. es la “versión actual de la lingüística descriptiva, entendida en el sentido más amplio y más técnico posible”. Este autor señala que la L.C. se opone a la lingüística de orientación racionalista y a la lingüística descriptiva tradicional ya que, gracias a este enfoque metodológico empírico, las hipótesis acerca de los fenómenos lingüísticos parten de un conjunto mayor de datos objetivos, basados en el uso real de la lengua.

En definitiva, se podría plantear que la L.C. es un enfoque metodológico, aún en desarrollo, que no se encarga del estudio de ningún fenómeno particular del idioma; sino que a través del uso de datos reales permite al investigador realizar una aproximación más objetiva al fenómeno que se esté estudiando. Su uso está justificado por la necesidad de incorporar herramientas digitales al estudio del lenguaje, ya que sin ellas sería imposible analizar los corpus de una forma práctica, esto es, dentro de los límites temporales y materiales requeridos por una investigación.

Dentro de los estudios del lenguaje que utilizan corpus para desarrollar las investigaciones existe una distinción importante -la cual se abordará a continuación-, a saber la investigación basada en datos y la investigación guiada por datos. La presente investigación cae dentro de esta última categoría.

La investigación guiada por datos “rejects the characterisation of corpus linguistics as a method and claims instead that the corpus itself should be the sole source of our hypotheses about language” (McEnery & Hardie, 2012, p. 6). Así, se entiende que deben ser los datos

arrojados por el corpus los que marquen el camino de la investigación. Según Biber (2015) este tipo de enfoque es más inductivo en tanto los datos surgen del corpus y trata de reducir al mínimo las suposiciones que pueda tener el investigador antes de enfrentarse al corpus. Es justamente este enfoque el que se ha tomado en esta investigación. Si bien se tenía una hipótesis al principio de la misma, a medida que se obtuvieron los datos la investigación fue variando su intención original y derivando conclusiones que no se habían contemplado al principio de la misma.

Por otra parte, la investigación basada en datos se apoya en el corpus para confirmar, refinar o refutar una hipótesis. O sea, la hipótesis ya está formulada de antemano y los datos simplemente son el medio para su confirmación o refutación sin ir más allá. Una vez que la hipótesis ha sido resuelta se da por finalizada la labor del investigador.

En cualquier caso, ninguno de los dos enfoques es mejor o peor que el otro; son simplemente distintos y responden a los intereses de la investigación.

### **2.1.1 Tipos de corpus: diseños y representatividad**

Como se hace evidente de su nombre, la L.C. se vale del uso de corpus para conducir cualquier estudio o análisis. En este marco se entiende por corpus un conjunto de textos seleccionados a partir de criterios situacionales jerarquizados. Esto se hace para conseguir muestras estratificadas de diferentes muestras del habla y así lograr representatividad para explicar los comportamientos del rasgo lingüístico que se esté estudiando (Sánchez Barrera, 2019).

Así, no toda colección de textos puede considerarse como un corpus. Estos deben crearse con un objetivo específico y respondiendo a determinados criterios de confección. Por ejemplo,

la colección de textos contenidos en una biblioteca no responde a los criterios de confección de un corpus y por tanto no se puede considerar como tal.

Un corpus debería ser una selección de datos lingüísticos reales que constituya una buena aproximación a la realidad de la lengua estudiada, agrupados bajo ciertos criterios que garanticen que todas o algunas variedades lingüísticas estén representadas (dependiendo de la finalidad del corpus) (Molina Salinas & Sierra Martínez, 2015, p.309).

La confección es, tal vez, la parte más crítica de cualquier investigación que pretenda adoptar las líneas metodológicas de la L.C. pues “corpus studies should match their data to their research question” (McEnery & Hardie, 2012), o sea, el corpus se debe confeccionar teniendo en cuenta el fenómeno que se quiera estudiar. En el caso particular de la presente investigación, cuyo objetivo es estudiar la literatura cubana de 1902 a 2020 se ha confeccionado un corpus que responde a estos criterios temporales y geográficos, es decir, se han seleccionado textos narrativos escritos solamente por escritores cubanos que desarrollaron su obra durante este período. Este tema se tratará a profundidad en la sección 5.1.

McEnery y Hardie (2012) mencionan dos enfoques principales al momento de recolectar datos y conformar el corpus, a saber, el enfoque de *corpus monitor* y el enfoque *corpus balanceado* o *corpus muestra*.

Al confeccionar un corpus desde el enfoque de *corpus monitor*, se debe tener en cuenta que su propósito es crear un tipo de corpus abierto, que esté en constante crecimiento (Clear, 1992) y compuesto por una gran variedad de tipos de textos completos. El objetivo de este tipo de corpus es básicamente recolectar la mayor cantidad de muestras del lenguaje posibles; pero sin supeditarse a tipos de textos o límites temporales. Según este enfoque

The relative proportions of the different types of materials may vary over time. Monitor corpora could be said to balance any need to be precise about the composition of a corpus against sheer size – as the corpus grows, we might assume that any skew in the data naturally self-corrects, since there is no consistent skew in the data input.(McEnery & Hardie, 2012, p. 7)

Algunos de los ejemplos mencionados por estos autores que se adhieren a este tipo de enfoque son el Bank of English (BoE) desarrollado por la Universidad de Birgminham y al Corpus of Contemporary American English (COCA)(*English-Corpora: COCA*, 2023). Asimismo, plantean que la Web, *per se*, puede entenderse como un gran corpus que responde a este enfoque.

Dentro de este tipo de enfoque existen otros tipos de corpus. No se comentarán todos los que existen, pero sí los que son relevantes para esta investigación. Así, se debe mencionar el corpus diacrónico. Según McEnery & Hardie (2012) “A diachronic corpus samples texts across a span of time or from multiple time periods”, lo cual permite al investigador comparar la variación del lenguaje o determinado rasgo lingüístico en diferentes períodos temporales. Para esta investigación esta idea es hasta cierto punto relevante, en tanto se pretenden comparar elementos temáticos pertenecientes a dos períodos temporales distintos. Además, el corpus de referencia, (CREA)<sup>3</sup> está confeccionado de manera que permite realizar estudios diacrónicos.

Por su parte, el corpus balanceado, o corpus muestra es “a careful sample corpus, reflecting the language as it exists at a given point in time, is constructed according to a specific sampling frame” (McEnery & Hardie, 2012, p.6). Este enfoque busca representar un tipo particular de lenguaje a lo largo de un período de tiempo delimitado. Así no se busca crear

---

<sup>3</sup> ver página 52

un corpus que incorpore diversos tipos de textos y que esté en constante crecimiento, sino que se busca diseñar uno que se ajuste a los objetivos específicos de la investigación. Este enfoque propone el diseño cuidadoso de un corpus que refleje el lenguaje como existe en una época y lugar determinados. Se construye respetando un marco de muestreo muy específico.

In doing so, they seek to be balanced and representative within a particular sampling frame which defines the type of language, the population, that we would like to characterise. The population is the notional space within which language is being sampled (McEnery & Hardie, 2012, p.8).

Por ejemplo, si se quiere estudiar el uso de la ironía en las comedias de Shakespeare, se debe construir un corpus limitado solamente a las comedias de este autor. No se puede incluir ningún otro tipo de obra ni ningún otro autor. O si se quisieran estudiar rasgos estilísticos de la novela gótica en la época Victoriana, solo se podrían incluir obras que cumplieran con estos parámetros. Este último enfoque es el que se ha adoptado en esta investigación, o sea, se creó un corpus limitado a textos de la narrativa cubana de 1902 a 2020.

Pero, ¿qué ocurre cuando el muestreo que se desea realizar para la confección del corpus no puede realizarse con el rigor necesario? Hay factores como las circunstancias sociales, de salud mundial, política, e incluso, las limitaciones temporales que conlleva una investigación para el grado de maestría.

Así, es común que los criterios de muestreo no puedan cumplirse de manera precisa. Eso no debe impedir que el investigador deje de intentar cumplir su cometido de tener un corpus utilizable para los propósitos de su investigación. En este caso, el corpus de muestra se convertiría en un corpus ocasional. Según McEnery & Hardie, este tipo de corpus

make no pretension to adhere to a rigorous sampling frame, nor do they aspire to deal with issues of skew by the collection of an ever-larger body of data, as monitor

corpora may. Rather, they represent nothing more nor less than the data that it was possible to gather for a specific task.(McEnery & Hardie, 2012, p.11)

En el caso de esta investigación, se define el corpus con detalle en la página 51. Se considera que la divergencia entre este corpus ocasional y la muestra detallada es mínima y no afecta a los objetivos de la presente investigación<sup>4</sup>. Al diseñar este tipo de corpus se debe tener en cuenta algunos conceptos fundamentales, a saber, *representatividad, equilibrio (balance) y variabilidad*.

Un corpus, cualquiera que sea el objetivo de la investigación, debe cumplir con el principio de representatividad para poder ser utilizado adecuadamente como la base a partir de la cual se puedan realizar generalizaciones relativas al lenguaje (Biber, 1993). Existe la idea, errónea hasta cierto punto, de que la representatividad depende del tamaño del corpus; si bien es cierto que es importante compilar una muestra que sea lo suficientemente grande para cumplir con los objetivos de la investigación, el tamaño de la misma no es el factor determinante para lograr la representatividad. Biber (1993) plantea que otros elementos como la definición precisa de la población con la que se ha de trabajar y decisiones relativas al método de muestreo son consideraciones más relevantes.

Para definir la población se deben establecer límites claros; por ejemplo, para esta investigación la población a estudiar son textos narrativos (novela y cuento), escritos *solamente* por escritoras o escritores cubanos en un período de tiempo que abarca desde 1902 hasta 2020. Un texto que no cumpla con esas tres condiciones no puede ser incluido en el corpus que guía esta investigación.

---

<sup>4</sup> Otros tipos de corpus que se van a mencionar en otras secciones de esta tesis son los corpus de referencia.

Por su parte Molina y Sierra (2015), señalan que en la idea de representatividad está implícito el apego del corpus a la realidad de la población estudiada. Es decir, si se han tomado en cuenta elementos como la localidad geográfica de los textos o el tipo de texto o el momento en el que fue creado el texto. Asimismo, plantean que

“la representatividad tiene relación directa con el equilibrio en la medida de que un corpus es capaz de reproducir las diferencias de las variables que se supone representa, en la proporción en la que ocurren en el estado natural de la lengua” (Molina Salinas & Sierra Martínez, 2015, p.310).

O sea, el corpus debe ser capaz de reflejar el fenómeno lingüístico que se pretenda estudiar de la misma manera o lo más cercano posible a lo que ocurre en el lenguaje real. Asimismo, Biber (1993) plantea que la representatividad solo se obtiene si un corpus es capaz de capturar la variabilidad del lenguaje, entendida esta última como la posibilidad que ofrece el lenguaje de decir la misma cosa (o casi la misma cosa) de diferentes maneras. A partir de estas ideas se podría plantear que un corpus es representativo si es capaz de recoger los usos más cercanos al lenguaje natural en un período de tiempo y un lugar determinados.

De esta idea de representatividad se derivan otros dos conceptos que deben ser tenidos en cuenta, *la frecuencia* y *la dispersión*. La frecuencia se debe entender como la repetición de determinado rasgo lingüístico en el corpus; mientras más se repita, más frecuente será. La dispersión se entiende como la mayor o menor distribución de este rasgo por el corpus. Así podría suceder que un rasgo o concepto tenga una altísima frecuencia en un corpus; pero muy baja dispersión. Esto sucedería si es, por ejemplo, el tema de una novela; pero no aparece en otras. En ese caso, a pesar de la alta frecuencia este rasgo no sería de interés para la investigación, pues su dispersión en el corpus es muy baja. Debe existir un equilibrio entre la frecuencia y la dispersión del rasgo lingüístico en el corpus para que este sea de interés para

la investigación. Es justamente este equilibrio lo que posibilita determinar las palabras clave y los temas de estudio.

### **2.1.2 Palabras Clave, nodos y colocaciones**

Como se ha comentado anteriormente, la L.C. proporciona las herramientas necesarias para enfrentarse al estudio de algún fenómeno lingüístico específico. Es, básicamente, un enfoque metodológico. Una de sus grandes ventajas es que permite trabajar con datos del lenguaje real. Sin embargo, depende del investigador interpretar los datos que tiene a su disposición, determinar los patrones, descubrir la razón del fenómeno a partir de los datos obtenidos.

Al enfrentarse a una investigación de este tipo uno de los primeros datos que se obtienen son las *palabras clave*. De hecho, los análisis de palabras clave son uno de los ejes centrales de la L.C. (Rodríguez Sánchez & Bautista Botello, 2021). Estas se determinan a partir de la comparación del corpus experimental con un corpus de referencia. El primero es un corpus relativamente pequeño, creado por el investigador; el segundo es un corpus más grande normalmente creado para representar el uso general y variado de una lengua, el Corpus de Referencia del Español Actual (CREA) o el British National Corpus (BNC), por ejemplo. En la sección 5.1 Descripción del Corpus se profundizará en el concepto de corpus de referencia.

Al realizar esta comparación, se pueden determinar cuáles son las palabras que gozan de un status especial dentro del corpus, esto es, las palabras que en comparación con el otro corpus tienen una frecuencia mayor y una dispersión relevante. Al comparar la frecuencia de una palabra o rasgo lingüístico entre dos corpus se habla entonces de frecuencia relativa (o frecuencia normalizada), esto es, en qué medida una palabra que pudiera no presentar una alta frecuencia (que no es relevante) en un corpus si lo es en otro. La frecuencia relativa posibilita diferenciar el comportamiento de determinado rasgo lingüístico entre dos corpus y



a partir de ahí determinar el o los temas predominantes dentro de un texto o corpus. McEnery & Hardie definen la frecuencia relativa (o normalizada) como

A frequency expressed relative to some other value, as a proportion of the whole – for example, frequency of a word relative to the total number of words in the corpus. Normalised frequencies can be compared even if they arise from datasets of different sizes.(McEnery & Hardie, 2012, p. 268)

Básicamente, la frecuencia relativa es aquella que permite comparar la relevancia de determinado rasgo en dos corpus. La Tabla 1 ilustra la frecuencia relativa de la palabra *cabeza* en dos corpus. A priori pudiera pensarse que en el Corpus B esta palabra es más frecuente y por tanto más relevante; sin embargo, el Corpus B es casi diez veces más grande que el Corpus A. Así al realizar el cálculo de la frecuencia relativa se puede ver que, en proporción al número de palabras por corpus, la palabra *cabeza* es más frecuente en el Corpus A que en el Corpus B.

Nombre de corpus	Tamaño corpus	Frec. total "cabeza"	Frec. rel. por mill. "cabeza"
Corpus A	1,500,000	1,831	1,220
Corpus B	160,000,000	42,000	263

Tabla 1. Ejemplo de frecuencias relativas en dos corpus

Según Bondi (2010) todas las palabras poseen un status similar en el lenguaje; pero su importancia varía desde una perspectiva textual. Cada palabra puede contribuir a la formación del significado en el texto, pero solo a partir de las palabras claves se puede determinar cuáles son los elementos importantes dentro del mismo. Estos ítems léxicos son los que permiten identificar -además de rasgos lingüísticos- los rasgos sociales, culturales, incluso políticos o económicos que conforman el contexto del texto.

Desde la perspectiva de la L.C es importante tener en cuenta las coocurrencias de las palabras claves, o sea, con qué otros ítems léxicos tienden a aparecer dentro del corpus. Estas coocurrencias son las que ofrecen al investigador los datos lingüísticos que le permiten

realizar la interpretación del fenómeno que está estudiando. Así, se pueden establecer patrones de coocurrencias o concordancia gramatical o semántica, según sea el caso de interés. Desde el punto de vista semántico, estas coocurrencias pueden arrojar datos con implicaciones socioculturales (Bondi, 2010). Según esta autora existen dos tipos de palabras clave, que bien podrían dividirse en las que tienen una relevancia semántica dentro del texto y las que tienen una relevancia sintáctica. Las primeras dan cuenta de la estructura conceptual del texto, su “aboutness” o de qué trata. Las segundas son indicadores del propósito comunicativo del texto, su micro y macro estructura, de lo que hace el texto y cómo lo hace (Bondi, 2010).

Una vez que se han determinado las palabras clave, se pueden analizar sus coocurrencias dentro del texto. O sea, se determinan las líneas de concordancia, esto es, las líneas de texto donde estas palabras clave ocurren. La Imagen 1 muestra las líneas de concordancia donde coocurre la palabra clave *Cuba*. Como se ve, una de las coocurrencias más interesante es con el verbo, o formas derivadas del verbo *regresar*.

File	Left Context	Hit	Right Context
2.ACPCV3.txt	ismos represivos. Por lo tanto, bien avisado estaba de que no <b>debia regresar a</b>	<b>Cuba</b>	por ahora. Aun proseguían las persecuciones, llevadas a los extremos límites de
2.GCMEV2.txt	uerdo, aunque de dientes para afuera, con que su hermano <b>debía regresar a</b>	<b>Cuba</b> ,	y el visitante español se fue con la convicción de que su interlocutor era
2.JDSV2.txt	oche siberiana. Se desplazó hasta los lavabos preguntándose <b>cómo regresar a</b>	<b>Cuba</b>	sin Nadiezhda, se miró al espejo y comprendió que también en su cara estaba
2.GCMEV2.txt	de visita en Bélgica desde principio del invierno y se preparaba <b>para regresar a</b>	<b>Cuba</b>	vía Madrid, donde ya estaba mi hermano trabajando como agregado comerci
2.GCMEV2.txt	, pensar en Miriam Gómez, sola en Bruselas y a quien no <b>permitiría regresar a</b>	<b>Cuba</b> ,	pasara lo que pasara con él mismo. Solamente lo consolaba ahora o lo hacía
2.EACBV2.txt	rán ni en los centros espirituales, pero si logro escapar jamás <b>podré regresar a</b>	<b>Cuba</b> ;	si me agarran en el brinco, pasaré el resto de mi juventud en una
2.GCMEV2.txt	argumento y parecía que iba a decir que Miriam Gómez bien <b>podía regresar a</b>	<b>Cuba</b> ,	más fácilmente que él ir hasta Europa. Pero no dijo nada. Fue Alberto el
2.ACPCV3.txt	emos, / a la vez acariciado y maldito... "...Mi tía temerosa de <b>verme regresar a</b>	<b>Cuba</b> ,	me colmaba de giros. Al encontrarme, una mañana, con tres mil dólares en el
2.GCDDV2.txt	se ordenó conseguir un atadid para traer al comandante muerto <b>de regreso a</b>	<b>Cuba</b> .	Para mí fue un espectáculo (aunque sólo vi su terminación) grotesco. Me imp
2.GCMEV2.txt	, al decidirme a hacerme cargo de la misión, que Aldama estaría <b>de regreso a</b>	<b>Cuba</b>	en pocos días. Estos pocos días, hay que decirlo, se volvieron semanas primer
2.GCMEV2.txt	el ministerio de Relaciones Exteriores le habían enviado un pasaje <b>de regreso a</b>	<b>Cuba</b>	por avión, junto con un cable en que la conminaban a regresar. Él—cuidando
2.GCMEV2.txt	sía que tenía hasta el próximo día 3 de octubre para embarcarse <b>de regreso a</b>	<b>Cuba</b> ,	que ese era el último plazo que le había dado el ministerio y que
2.JDSV2.txt	: en su caso separación equivalía a pérdida. En unas horas volaría <b>de regreso a</b>	<b>Cuba</b>	y ella quedaría sola, sin trabajo, a merced de un borracho enloquecido y de
2.JDSV2.txt	dado en la troika y estuvo a punto de quebrarse y adelantar <b>el regreso a</b>	<b>Cuba</b> ;	no lo hizo porque le quedaban una pizca de orgullo, algo de esperanza y
2.JDSV2.txt	bastaría sola para derrotarlo y de que él mismo iba a pedir <b>el regreso a</b>	<b>Cuba</b> .	No fue así, por suerte y por desgracia, y hubo un momento en que
2.GCMEV2.txt	extraño en él como pueda parecer. Pero él aceptó la noticia de <b>su regreso a</b>	<b>Cuba</b>	sin muestras de violencia. Solamente pidió que le dieran más tiempo «para en
2.GCMEV2.txt	carta de su hermano Sabá en la que le decía que apuraban <b>su regreso a</b>	<b>Cuba</b>	y le preguntaba que cuándo regresaría él a Europa. La carta había venido en

Imagen 1. Líneas de concordancia. Palabra clave CUBA

A partir de este dato se analizan los colocados de la palabra *Cuba*. Es importante tener en cuenta que no todas las coocurrencias son relevantes, si se entiende por coocurrencia que cualquier palabra puede coocurrir con otra en un texto. En el marco de la L.C, al realizar búsquedas, se trabaja con nodos y colocados. McEnery & Hardie retoman a Sinclair quien define nodo y colocado como sigue:

A node is an item whose total pattern of co-occurrence with other words is under examination; a collocate is any one of the items which appears with the node within a specified span. Essentially there is no difference in status between node and collocate; if word A is a node and word B one of its collocates, when word B is studied as a node, word A will become one of its collocates. (McEnery & Hardie, 2012, p. 142)

O sea, el nodo es una palabra seleccionada arbitrariamente por el investigador, en este caso la palabra *Cuba*, y el colocado es la palabra que coocurre con este nodo en un espacio determinado y cuya frecuencia de coocurrencia es mayor que la de otras palabras. Esta frecuencia es lo que hace a este par relevante para la investigación. Como se ha comentado,

las líneas textuales en las que aparecen estos pares son las líneas de concordancia y son las que permiten contextualizar estas coocurrencias, desde un punto de vista lingüístico o paralingüístico, según el interés del autor. La *Imagen 2* muestra los colocados de la palabra *Cuba*. En este caso se tomó *Cuba* como nodo. La imagen permite observar cómo entre los colocados más frecuentes están formas verbales de *salir* y *regresar*.

The screenshot shows the AntConc software interface. The main window displays a table of collocates for the word 'Cuba'. The table has the following columns: Collocate, Rank, FreqL, FreqR, Range, Likelihood, and Effect. The search query is 'Cuba' and the results are sorted by Likelihood.

Collocate	Rank	FreqL	FreqR	Range	Likelihood	Effect	
1 en	1	836	640	196	24	453.945	1.189
2 santiago	2	54	53	1	17	309.025	5.52
3 cuba	3	68	36	32	12	164.824	3.007
4 venegas	4	17	0	17	1	142.054	7.394
5 de	5	1160	840	320	26	135.470	0.506
6 libre	6	24	1	23	8	88.177	3.996
7 provincia	7	17	16	1	2	81.134	4.823
8 españa	8	18	9	9	8	50.888	3.335
9 el	9	280	133	147	24	48.080	-0.553
10 regreso	10	17	16	1	5	45.806	3.229
11 viva	11	13	12	1	7	44.136	3.782
12 pueblo	12	22	17	5	8	43.443	2.632
13 salir	13	24	22	2	5	40.452	2.379
14 regresar	14	16	16	0	6	39.929	3.068
15 salido	15	14	13	1	8	39.722	3.343
16 cinemateca	16	7	7	0	3	38.834	5.394
17 villas	17	10	10	0	2	37.728	4.072
18 unidos	18	12	5	7	4	35.858	3.463
19 shiram	19	4	3	1	1	35.841	7.809

*Imagen 2. Colocados de la palabra Cuba*

Justamente es a partir de este análisis de palabras clave, colocados y líneas de concordancia que se ha conducido esta investigación.

### 2.1.3 Análisis de sentimientos

En el marco de esta investigación es de vital importancia abordar el tema del análisis de sentimientos, si bien las líneas metodológicas de esta disciplina se han acomodado a los

intereses de este trabajo en tanto se utilizaron solo los postulados teóricos propuestos por Liu (2015); pero no se utilizó ningún software especializado para realizar este análisis.

El análisis de sentimientos (A.S. a partir de ahora) es una tarea derivada del Procesamiento del Lenguaje Natural que identifica opiniones relacionadas con un objeto (Fernández et al., 2011). Según Liu (2015) el A.S., también llamado minería de opiniones, es un campo de estudio que se enfoca en el análisis de las opiniones, actitudes, valoraciones y sentimientos de las personas hacia una cosa. Por lo general el objeto del que se emite la opinión son productos, organizaciones, individuos o eventos.

Esta disciplina ha ido ganando terreno debido, entre otras razones, al auge del mundo digital (redes sociales) y a la posibilidad que tienen las personas de emitir sus valoraciones con respecto a algo de manera mucho más “fácil” y “libre” que hace 20-30 años. Entiéndase valoración en este caso como la **opinión** que emite un sujeto de las cualidades de un objeto; estas cualidades le son inherentes al objeto, o sea, existen independientemente del sujeto, por lo que los juicios valorativos que emite este son altamente subjetivos y están condicionados por el contexto del sujeto. Así, a través de estas valoraciones, el sujeto comunica sus sentimientos con respecto al objeto.

Además, el A.S. permite predecir el comportamiento de las personas con respecto al objeto o producto en cuestión a partir de sus valoraciones. Se enfoca principalmente en el texto escrito, sobre todo aquellos que se encuentran en las plataformas digitales como Facebook, Twitter o blogs en la web. Asimismo, tiene aplicaciones en campos que van desde el turismo y la hospitalidad hasta elecciones y campañas políticas.

El A.S. se realiza utilizando programas computacionales que funcionan basados en algoritmos. Estos algoritmos permiten reconocer y cuantificar las *palabras sentimiento*. Estas palabras comunican sentimientos con una polaridad positiva o negativa (Qiu et al., 2009). Son las que comunican las valoraciones que las personas hacen de los objetos (*bueno, óptimo, pésimo, bello, feo*). Por supuesto, deben analizarse dentro del contexto donde se realizan para poder determinar la posición del sujeto valorador con respecto al objeto valorado. Para esto Liu (2015) propone tres niveles de análisis que se deben tener en cuenta al trabajar con los textos, a saber, el análisis a nivel de documento, a nivel de oración y a nivel de aspecto.

El análisis a nivel de documento pretende determinar si el documento en general expresa un sentimiento positivo o negativo. A nivel de oración se trata de determinar si cada oración expresa una opinión positiva, negativa o neutral; esta neutralidad se debe entender como la ausencia de opinión o sentimiento. Finalmente, el nivel aspectual analiza directamente la opinión y el objeto de esta opinión. Este nivel se explica mejor a partir de un ejemplo de la propia autora:

“Although the service is not great, I still love this restaurant.” This sentence clearly has a positive tone, but we cannot say that this sentence is entirely positive. We can only say that the sentence is positive about the restaurant (emphasized), but it is still negative about its service (not emphasized). If someone reading the opinion cares a lot about the service, he probably will not go to eat at the restaurant. (Liu, 2015, p.10)

Estos dos últimos niveles de análisis (oración y aspecto) han sido los más relevantes para esta investigación, pues justamente se trata de analizar las *palabras sentimiento* en su contexto, esto es, las líneas de concordancia donde aparecen y determinar la valoración que el sujeto (en este caso *el personaje* o *el narrador*) realiza sobre el objeto valorado (*opinión target*).

Sin embargo, es importante aclarar, que, si bien se trabajó con conceptos y pautas metodológicas del A.S. tradicional, en esta investigación no se utilizó ninguno de los algoritmos tradicionalmente usados en esta disciplina, ni se pretendió hacer un análisis cuantitativo, sino más bien uno cualitativo desde una perspectiva tradicional dentro del ámbito de los estudios literarios.

Asimismo, es importante comentar que más que los sentimientos u opiniones que pudieran expresar los personajes en las narraciones, se analizaron las emociones que se comunicaban a partir de los contextos en los que aparecían las palabras clave y las *palabras sentimiento*.

Sin embargo, parece pertinente aclarar qué se entiende por emociones o sentimientos en esta investigación y desde que punto de vista se tratarán estas nociones.

Ambas nociones están estrechamente relacionadas con los estados de ánimo que puede experimentar un sujeto, los cuales pueden estar condicionados por diferentes factores, por lo general externos al sujeto. Según Marina y Penas (2000, p. 44)

Al ser humano le pasan muchas cosas. No actúa ni movido por la química ni movido por los conceptos. Conoce su entorno y es afectado por él de manera estrepitosa. Huye de unas experiencias y se precipita hacia otras.

Así, el contexto en el que está inmerso el sujeto es el que determina o condiciona estos estados de ánimo. El contexto provoca emociones, esto es, estados de ánimo intensos y pasajeros que pueden convertirse en sentimientos. Tanto las emociones como los sentimientos pueden ser negativos o positivos.

Llegados a este punto es necesario unificar estos conceptos. Valderrama (2015) propone una taxonomía de emociones; Marina y Penas (2000) por su parte presentan un diccionario de sentimientos. En ambos casos estos autores manejan nociones que se mueven

indistintamente entre los conceptos de sentimiento y emoción, a saber: *miedo, resignación, tristeza*, entre otras. Así, para los efectos de esta investigación se ha decidido unificar ambos conceptos dentro del campo de sentimientos.

Marina y Penas (2000) al analizar los sentimientos proponen partir de dos primitivos semánticos *bueno y malo*. Igualmente, Valderrama (2015) parte de la idea que existen emociones positivas y negativas.

Esta autora entiende ‘emoción’ como

una alteración del ánimo intensa y pasajera, agradable o penosa, que va acompañada de cierta conmoción somática. El término *emoción* proviene del latín *emovere* que significa poner en movimiento. Las emociones implican una predisposición a actuar en determinado sentido (Valderrama, 2015, p.15).

Las emociones positivas se perciben como agradables y generalmente están ligadas al logro de alguna meta. Las negativas “se perciben como desagradables, se relacionan con la frustración de objetivos, una amenaza o pérdida y orientan al organismo hacia la aversión y la evitación” (Valderrama, 2015). Así esta autora propone como emociones básicas, dentro de las positivas *confianza, esperanza, amor, gratitud, compasión, alegría, entusiasmo, autoestima y orgullo*. Dentro de las negativas *miedo, preocupación, ira, odio, envidia, tristeza, resignación, culpa y vergüenza*.

Como se verá más adelante, en la sección de metodología, esta investigación pretende combinar las líneas metodológicas del A.S., desde un punto de vista cualitativo y las ideas manejadas por Marina y Penas, y Valderrama. Desde una perspectiva más “artesanal” se analizan las líneas de concordancia donde aparecen las palabras claves para determinar las emociones que se comunican en el texto. Es importante señalar que todos los sentimientos



discretizados dentro de la investigación fueron descritos o definidos por estos autores en sus respectivas obras.

## **2.2 Lectura Distante**

Como se ha venido comentando, la razón principal de utilizar técnicas propias de la L.C. en esta investigación está justificada justamente por la imposibilidad de estudiar “artesanalmente” (obra a obra, lápiz en mano) un corpus de las dimensiones del que se utiliza aquí.

Por otra parte, sin importar qué tan aficionado a la lectura y a la literatura se pueda ser, un lector, aún con un interés más allá del lúdico, no leerá ninguna obra si esta no le resulta de alguna manera atractiva, independientemente de su status como clásico dentro del canon literario.

A partir de estas dos razones, es importante comentar la relevancia de la idea de *lectura distante* para esta investigación.

Lectura distante es una noción enunciada por primera vez, al menos formalmente, por Franco Moretti en su obra *Distant Reading* (Moretti, 2013). Este libro es una recopilación de diez ensayos en los que el autor aborda algunas problemáticas relativas al estudio de la literatura. Una de las cuestiones que trata es lo que él denomina como “the great unread”: dentro de la literatura existen obras canónicas, aquellas que marcan pautas dentro del género, las que los académicos estudian y los lectores leen. Sin embargo, más allá de estas, existen otras obras, menos conocidas, menos publicadas, pero que también forman parte de una corriente literaria y que retratan usos, costumbres, modos de hacer y escribir de una sociedad determinada en una época determinada. Este es “the great unread”, o sea, todas esas obras que han sido escritas pero que en su mayoría son desconocidas por el público general.

Según Moretti, es imposible para una persona (sea académico o solo lector por afición) conocer, ya no se diga estudiar, todas las obras escritas; así, propone que hay dos maneras de aproximarse al estudio de la literatura. La primera sería la tradicional, la lectura cercana “At bottom, it's a theological exercise-very solemn treatment of very few texts taken very seriously” (Moretti, 2013, p.57) y su principal crítica a esta aproximación es que permite solo el estudio de un número limitadísimo de textos a partir de los cuales es extremadamente difícil llegar a hacer generalizaciones.

En contrapartida, propone la lectura distante “where distance, let me repeat it, is a condition of knowledge: it allows you to focus on units that are much smaller or much larger than the text: devices, themes, tropes-or genres and systems” (Moretti, 2013, p.57). Así, a partir de la idea de lectura distante y usando herramientas computacionales<sup>5</sup> la idea de Moretti es distanciarse del texto en sí; ya no se trata de buscar elementos específicos o consagrarse al análisis de algún aspecto particular del texto, sino de buscar generalidades y establecer patrones (temas, figuras retóricas, maneras de decir) que sirvan para caracterizar y describir los textos. Para Moretti, esta es la forma de entender el sistema (en este caso la Weltliteratur) como un todo. Así, por ejemplo, en el caso específico de esta investigación se ha conformado un corpus de 45 textos narrativos de literatura cubana. Por razones prácticas es imposible analizar a profundidad uno por uno cada texto para determinar los temas que se tratan en cada uno, o sea es imposible realizar un análisis *micro* de los mismos. Es por esto que, siguiendo la propuesta de Moretti, se da un paso atrás y se decide analizarlos de manera macro, no ya

---

<sup>5</sup> La idea de lectura distante ha ido evolucionando en el tiempo. Al momento del primer ensayo “Conjectures of World Literature” (Moretti, 2013), el autor proponía estudiar la literatura a partir de lo que otros autores ya habían dicho con respecto al tema y comparar estos estudios.

cada texto independientemente, sino todo el corpus como un todo y determinar los patrones temáticos existentes en el corpus que permitan describir la literatura cubana del siglo XX.

Según López Fernández la lectura distante

se basa en la recopilación y el análisis de características formales y patrones hallados en obras literarias que pueden ser cuantificados, comparados y representados a través de métodos computacionales. Presenta así un nuevo paradigma en el que el canon ya no es un concepto válido para estudiar la historia de la literatura y que supone un salto cuantitativo en cuanto al conocimiento desde el prisma de la objetividad científica (López Fernández, 2015, p.267).

Esta noción de lectura distante es justamente uno de los pilares que sustentan este proyecto. Unida a las herramientas de la L.C esta filosofía de estudio de Moretti ha permitido el acercamiento a una muestra bastante considerable, por cierto, de la literatura cubana del siglo XX (y parte del XXI) y poder caracterizarla, poder establecer generalidades al respecto. Esto sin tener que enfrascarse en la lectura de las novelas una por una, lo cual, francamente hubiera sido imposible de hacer, al menos dentro de los plazos establecidos.

En este capítulo se ha tratado de abordar los principales postulados teóricos que sustentan esta investigación. Se han revisado las bases de la L.C y se examinaron elementos relativos al análisis de sentimientos y a la lectura distante. Estos fundamentos teóricos serán abordados indistintamente a lo largo de esta investigación. En el siguiente capítulo se hará una revisión de los antecedentes de este trabajo, esto es investigaciones que se han ocupado de temas similares y desde una perspectiva similar a la que propone en la presente investigación.

### **3 Antecedentes**

Las ideas de Franco Moretti comentadas en la sección anterior, unidas a las técnicas y metodología propias de la lingüística de corpus han servido como base teórica para diferentes investigaciones que se han aproximado a la literatura desde esta perspectiva.

El presente capítulo pretende, primero, revisar algunas investigaciones en las que justamente se emplean estas técnicas y que han servido de guía para esta investigación. Asimismo, se hace una brevísima revisión de investigaciones sobre literatura cubana, que si bien no se han realizado desde el enfoque de la L.C. ni tienen en cuenta los postulados de Moretti, sí contextualizan de manera muy precisa la evolución de la literatura cubana, sobre todo, el impacto que tuvo en esta el triunfo de la revolución cubana.

En la primera parte del capítulo se pretende hacer notar cómo sí es posible realizar análisis literarios desde la L.C. La pertinencia de las investigaciones revisadas es justamente la que justifica y valida la pertinencia de la presente investigación. En la segunda parte, a partir de la revisión de la literatura cubana se pretende establecer los temas abordados por esta en el siglo XX y lo que va del XXI para posteriormente compararlos con los temas discretizados en esta investigación y determinar si, efectivamente, son similares. De constatarse una similitud entre temas, se probaría una vez más la pertinencia de la L.C. para el análisis literario a la vez que supondría un nuevo camino para este tipo de análisis en Cuba, donde las investigaciones literarias, al menos las consultadas para este trabajo, se han llevado a cabo desde perspectivas pedagógicas, sociológicas o políticas; pero nunca desde un enfoque multidisciplinario que incluya a la L.C. como es el caso de esta investigación. Se debe mencionar que la presente es una investigación de carácter exploratoria en la que se pretende utilizar estas técnicas para comenzar a estudiar la literatura cubana desde esta perspectiva;

no se pretende realizar un análisis exhaustivo de la misma por razones, en principio temporales, ya que se trata aquí de una tesis de maestría. Por estas mismas razones, el corpus confeccionado no recoge toda la literatura cubana de este período, se trata más bien de un corpus ocasional<sup>6</sup> diseñado de acuerdo a los objetivos en cuestión.

### **3.1 Estudios literarios desde la lingüística de Corpus**

Cada vez es más común realizar análisis literarios usando técnicas de la L.C. combinadas con el enfoque de *lectura distante* propuesto por Moretti (2013). Muchos de estos estudios se ven enmarcados en subdisciplinas como la estilística de corpus y las humanidades digitales (Hoover et al., 2014).

Harvey y Shalom (1997) recogen algunos de los primeros trabajos de análisis literarios realizados desde la perspectiva de la L.C. En las investigaciones compiladas por estos autores se analiza la organización y la estructura lingüística de narrativas en torno al tema del deseo. Así, se realizan varios análisis del habla natural y de obras de diversos géneros de la literatura popular con técnicas y herramientas propios de la lingüística de corpus para indagar (mediante el análisis de, por ejemplo, verbos recíprocos) cómo en esas obras los personajes manifiestan sus deseos y establecen relaciones amorosas.

Entre los trabajos más interesantes de esta compilación está el de Alice Deignan (1997). Deignan explora cómo la metáfora, como figura retórica, juega un papel fundamental en la creación de imágenes estereotipadas relativas al deseo sexual en sus diferentes manifestaciones. Para esto trabaja con el *Bank of English*, un corpus de 211 millones de palabras que recoge muestras del lenguaje natural del inglés actual (conversaciones, programas de radio, periódicos, libros, cartas, etc.). Como se comentó en la página 19, el

---

<sup>6</sup> Ver página 20.

*Bank of English* es un corpus monitor, es decir, un corpus abierto que crece en el tiempo y que no se enfoca en ningún género específico, sino que busca recopilar la mayor cantidad de textos posibles.

Al realizar el análisis de estos textos la autora determina los estados de ánimo que más frecuentemente se asocian a la palabra *DESIRE*. Para ello estudia las palabras clave referentes a *DESIRE* y sus líneas de concordancia más frecuentes, a partir de las cuales determina que algunas de las nociones que más se asocian (metafóricamente) a *DESIRE* son *FIRE*, *ANGER*, *ILLNESS*, *INSANITY*, *PAIN*. Debajo algunos ejemplos (Deignan, 1997 p.26)<sup>7</sup>:

1. In Spain, where they had lived for three years, he met a beautiful blonde Norwegian and they fell **madly** in love. (insanity)
2. I had felt a **twinge** of desire. (pain)
3. I am **sick** with love. (illness)

El análisis de ejemplos como los anteriores le permite concluir a Deignan que

(...) we fear desire, possibly for its potential to disrupt the established patterns of our lives; desire is talked of metaphorically as a wild animal, and as the dangerous and elemental forces of water, fire and electricity. (...) Thus, desire appears uninvited and takes us over; we are not responsible. This both reflects our physical perception of desire and allows us to disclaim responsibility for 'sinful' desire. (Deignan, 1997, p.40)

Otro de los trabajos que forma parte de esta compilación es *You would if you loved me: Language and Desire in teen novel* (Knowles, 1997). En este trabajo, el autor estudia el lenguaje del deseo sexual desde la perspectiva de un adolescente (varón) poco atractivo, esto es no deseado por sus coetáneas. Para ello analiza el soliloquio de Tubs Caspar, protagonista

---

<sup>7</sup> Las palabras fueron subrayadas y destacadas en negritas por el autor de esta investigación para evidenciar las palabras a partir de las cuales Deignan determina sus metáforas.

de la novela *The Chocolate War*. Esta novela se inscribe dentro de la *ficción realista*, un género que empezó a ganar popularidad en los Estados Unidos en las décadas de los '60 y los '70; uno de los principales temas de este género era la sexualidad.

Knowles toma como base el modelo de transitividad de Halliday para analizar las líneas de concordancia y colocaciones relativas<sup>8</sup>, a lo que denomina procesos “Material (doing), Mental (sensing), Verbal (saying) and Relational (being)” (Knowles, 1997, p.127). Así, determina las relaciones que se establecen entre los personajes de la novela y las cargas emocionales de cada uno. Este análisis le permite determinar el contexto en el que se presenta el deseo a los adolescentes. Según Knowles

That context is one of corruption. The transitivity and lexical patternings that structure the soliloquy demonstrate clearly the powerlessness of the main protagonist. In fact, the whole of *The Chocolate War* is a book where there is ‘no longer any trace of the happiness and simplicity that traditionally are associated with childhood’ (Egoff1981:44). (Knowles, 1997, p.139)

En su trabajo *Graphs, Maps and Trees* (2005) Franco Moretti retoma el concepto de lectura distante, un poco más elaborado tal vez, y plantea que este es un enfoque “where distance is however not an obstacle, but a specific form of knowledge: fewer elements, hence a sharper sense of their overall interconnection. (Moretti, 2005, p.10)

En esta obra Moretti analiza, a través de modelos abstractos, diferentes géneros, tramas, subtramas y otros aspectos de la literatura que le permitan establecer patrones y características comunes dentro de los géneros estudiados. Si bien se aproxima al estudio de la novela en Dinamarca, Japón, Francia e Italia, el centro de “Shapes, relations, structures. Forms. Models” es el desarrollo de la novela inglesa de 1740 a 1900. Así determina la

---

<sup>8</sup> Ver sección Tipos de corpus: diseños y representatividad

existencia de 44 subgéneros dentro de la novela inglesa y trata de explicar los procesos políticos, históricos y sociales que pueden desembocar en el surgimiento de un nuevo género o el reemplazo de un género establecido por uno nuevo. Así, por ejemplo, plantea que

When one genre replaces another, it's reasonable to assume that the cause is internal to the two genres, and historically specific: amorous epistolary fiction being ill-equipped to capture the traumas of the revolutionary years, say-and gothic novels being particularly good at it. (Moretti, 2005, p.28)

En definitiva, este trabajo de Moretti es un intento de clasificación y agrupación de subgéneros de la novela a partir de una serie de rasgos lingüísticos procesados computacionalmente que le posibilita establecer patrones que a su vez permiten interpretar los contextos políticos, históricos, sociales y geográficos en los que estos textos fueron producidos.

En el ámbito mexicano se han desarrollado también una serie de trabajos que emplean técnicas de la L.C. para conducir análisis literarios.

Sánchez Barrera (2019) realiza un análisis del discurso de la novela rosa mexicana contemporánea. Sánchez Barrera usa dos listados de palabras clave y realiza un análisis similar al comentado en la sección Palabras Clave, nodos y colocaciones. Uno de los listados se obtiene del contraste entre un corpus de novela romántica mexicana contemporánea (CNRNM) y un corpus de novela romántica tradicional; otro listado que se obtiene del corpus CNRNM y el CREA. Así determina 6 rasgos léxicos relevantes dentro del corpus de interés, los cuales le permiten analizar las representaciones de género tradicionales dentro de este tipo de literatura, las cuales, a su vez, según la autora, van en detrimento de las mujeres, si bien ellas son las principales consumidoras de estas novelas. Los 6 rasgos léxicos analizados



en este trabajo fueron los sustantivos corporales *mirada, sonrisa y beso*, y los sustantivos espaciales *baño, habitación y cama*. Este análisis le permite concluir que

Por los referentes corporales, es posible afirmar que las autoras mexicanas de novelas románticas contemporáneas narran actos comunicativos que van de la agresión al romance; también que el romance puede ser apasionado y erótico, pero que la castidad puede ser valorada. Por los sustantivos espaciales, es posible saber que la intimidad es tópico importante en este tipo de novelas, y aún más importante aquella que tiene que ver con el desarrollo de actividades eróticas. (Sánchez Barrera, 2019, p.78)

Por su parte, vinculado con el análisis de los espacios, Rodríguez y Bautista (2021) realizan un trabajo sobre la novela contemporánea mexicana escrita por mujeres mediante el análisis de palabras clave y colocaciones. Según los autores, este análisis de palabras clave les permite señalar rasgos diferenciales entre los textos experimentales al compararlos con el corpus de referencia. Rodríguez y Bautista analizan un corpus de su diseño, el cual denominan Corpus de Narrativa escrita por Mujeres Mexicanas (CNEM-Mex) y lo contrastan en oposición al CREA. Así, toman en cuenta las 200 palabras con mayor representatividad dentro del corpus estudiado; la mayoría de estas estuvieron vinculadas a temas como los espacios y el cuerpo. Al analizar las líneas de concordancia relativas a las palabras claves, los autores determinan que existen muchas referencias textuales a la violencia hacia la mujer en México y a la reafirmación de la posición social dominante de los hombres.

Asimismo, un análisis más pormenorizado de la obra de escritoras como Amparo Dávila o Elena Garro permite identificar ciertos sentimientos (el miedo, la angustia, la soledad) como propios del universo personal de su narrativa.

### **3.2 Estudios sobre literatura cubana**

Como se ha estado comentando, esta investigación pretende realizar un análisis de la literatura cubana del siglo XX-XXI e intenta abordar cuestiones relativas a los temas

literarios y al impacto que pudieran tener asuntos políticos, sociales y económicos en la misma. Así, hasta cierto punto esta investigación que trata de incorporar diversas disciplinas y novedosa en tanto, hasta donde se tiene noticia, esto es, la bibliografía que se ha consultado hasta ahora, no se ha realizado ninguna aproximación a la literatura cubana desde la L.C. Sí son abundantes los estudios acerca de esta desde perspectivas puramente literarias, sociológicas o pedagógicas. Estos estudios anteriores han servido de base para discretizar los temas tratados por la literatura cubana, en tanto estos autores han podido aislar temas de gran impacto en el imaginario cubano y analizar su pertinencia en la vida de la nación. En esta sección se trata de realizar una revisión de algunas de las investigaciones que han servido de pauta para la presente.

Un trabajo relevante y novedoso sobre literatura cubana es el de Mirta Martínez (1999), ya que aborda el tema del homosexualismo dentro de la literatura cubana, el cual era un tabú para el régimen y de alguna manera trató de obliterar todo lo relativo al mismo no solo en las esferas culturales sino también en las sociales de la nación.

En este trabajo *Apuntes sobre el homoerotismo masculino y femenino en la literatura cubana de los 90* la autora analiza cómo la literatura de los años '90 del siglo pasado en Cuba retoma el tema del homoerotismo/homosexualismo, que había sido eliminado, no solo de la literatura sino de la vida pública e institucional del país por considerarse contrario a los valores e ideas del hombre nuevo que pretendía construir la revolución cubana. Sin embargo, a partir del artículo de Martínez, sí es importante señalar que este no es un tema ajeno al quehacer de los escritores cubanos, pues ya se había tratado en períodos anteriores. Como señala la autora, otras novelas ya habían abordado esta temática (*El ángel de Sodoma-1928-* de Alfonso Hernández Cata, *Hombres sin mujer -1938-* de Carlos Montenegro, *Paradiso -1960-* y

Oppiano Licario, -1977- de José Lezama Lima)<sup>9</sup>, la cual fue posteriormente obliterada por la política cultural estatal.

Teodosio Fernández (2000) hace un recuento de la evolución de la literatura cubana del siglo XX. Fernández plantea que entre los principales temas de la narrativa de la isla durante la primera mitad del XX figuraban la narrativa de la tierra, inserta en corrientes como el naturalismo, el realismo y el costumbrismo. Además, uno de los rasgos distintivos de la literatura de la época, especialmente durante la década del '30 del siglo XX, fue el compromiso de los escritores con la realidad política y social del país<sup>10</sup>. Fernández señala como un hito en la literatura cubana la inserción de esta en el ámbito de la literatura hispanoamericana a partir de la conceptualización de *lo real maravilloso americano* por Alejo Carpentier en novelas como *Ecue-Yamba-O*, *El reino de este mundo* o *Los pasos perdidos*.

El triunfo de la Revolución Cubana cambió el panorama cultural del país; de acuerdo a Fernández su período romántico con los intelectuales había concluido para mediados de los años '60. Así, comienza a ser notable una preferencia por el “realismo realista” consecuente con el proceso seguido por el gobierno de Fidel Castro, aunque la mayoría de las novelas publicadas durante los primeros años de la Revolución “pueden interpretarse como pervivencia de la novela de la tierra o de la narrativa costumbrista de crítica social” (Fernández, 2000, p.87).

---

<sup>9</sup> Vale la pena señalar que todas las obras mencionadas por la autora, excepto *El ángel de Sodoma* son parte del corpus de esta investigación.

<sup>10</sup> Los últimos años de la década del '20 y los primeros de la década del '30 estuvieron caracterizados en la Isla por una gran inestabilidad política, derivada de la *Revolución del '30* contra el gobierno de Gerardo Machado y de la sucesión de gobiernos derivados de la caída del de Machado.

Entre los temas más recurrentes de la narrativa de la época revolucionaria menciona la crítica a la época republicana y el ensalzamiento de las bondades derivadas de los procesos sociales impulsados por el gobierno revolucionario, así como los cambios de mentalidad que estos exigían. Además, señala el género policiaco como uno de los vehículos de propaganda del gobierno en tanto se centraba el conflicto en la lucha de los servicios de seguridad revolucionarios contra la CIA o enemigos internos de la Revolución.

Para la década de los ´80 Fernández precisa que

la narrativa cubana de la isla se incorporaba al proceso que la narrativa seguía en otros países de Latinoamérica, impulsado por escritores que habían sufrido la represión y el exilio a causa de dictaduras sangrientas, o simplemente se habían sentido afectados por el desencanto que se extendió por entonces ante el fin de las utopías (Fernández, 2000, p.90).

Asimismo, sintetiza la situación cubana durante la década de los ´90, caracterizada por las privaciones y escasez económicas y el desencanto social, lo cual impactó fuertemente en la narrativa de la isla.

En su tesis doctoral, Behar (2007) hace un revisión del contexto social, político y cultural cubano después de 1959 a la vez que se enfoca principalmente en la década de los ´90 del siglo XX, conocida como *Período Especial en Tiempos de Paz*. Esta fue una década particularmente compleja en la isla. En 1989 cae el Muro de Berlín y comienza así el derrumbe del bloque socialista de Europa del Este que concluye con la disolución de la Unión Soviética en 1991. Así el gobierno cubano pierde a sus principales socios económicos. Esto unido al endurecimiento del bloqueo económico de Estados Unidos (Ley Torricelli-1992-, Ley Helms-Burton-1996-) contra Cuba configuró un contexto socio-económico muy complejo, cuyas consecuencias aún perduran.

Behar analiza los principales temas de la literatura cubana en los '60,'70 y '80, a saber, la insurrección contra Batista, la angustia y la enajenación del ser humano en la sociedad contemporánea como una expresión de la culpabilidad por un pasado burgués, y el interés por exponer la decadencia social de las épocas prerrevolucionarias. Asimismo, afirma que desde 1971 hasta el comienzo de la década de los '90 “el objetivo literario se torna popularista y de formación ideológica” (Behar, 2007), siendo la novela policiaca o detectivesca el vehículo ideal para estos fines<sup>11</sup>. Para la década de los '90 Behar señala que comienzan a aparecer en la literatura cubana temas antes “mal vistos” como el exilio, la droga, la corrupción, la desesperanza y el suicidio. Asimismo, comienza a aceptarse la idea de que los escritores exiliados cubanos también forman parte del patrimonio literario de la isla.

La narrativa finisecular cubana, escrita dentro o fuera de Cuba, y moldeada por el gran caos que caracteriza a la década de los noventa, posee un carácter innovador. La producción literaria de esta época muestra una serie de características nunca antes vistas en la literatura de las décadas anteriores, ni siquiera antes del periodo revolucionario que comienza en 1959 (Behar, 2007, p.43).

Por su parte Barrial (2015), desde una perspectiva pedagógica, esto es, la literatura aplicada a la enseñanza de ELE, analiza el contexto cubano después de 1959 y cómo este determinó el desarrollo de la narrativa cubana. Esta autora reconoce que la literatura en particular, y la vida cultural cubana en general han sido controladas por el Estado Cubano desde 1959, ya que este es el único propietario de los bienes públicos. Como consecuencia de esto todas las instituciones culturales cubanas (MinCult, UNEAC, ICAIC, UPEC, Casa de las Américas)<sup>12</sup> no son instituciones independientes del Estado.

---

<sup>11</sup> Ver Sonia Behar 2007 p. 21 para entender los requisitos que debe cumplir la novela policiaca revolucionaria.

<sup>12</sup> MinCult: Ministerio de Cultura.

UNEAC: Unión Nacional de Escritores y Artistas de Cuba.

ICAIC: Instituto Cubano de Arte e Industria Cinematográficas.

UPEC: Unión de Periodistas de Cuba.

Asimismo, Barrial comenta acerca de la censura en Cuba

(...) si no se escribe al acorde con lo que la Revolución te ofrece, el Consejo Nacional de Cultura se ocupará de destruir o, por lo menos, de menospreciar aquello que los autores expresen libremente, dentro de esa “libertad” que ofrece la Revolución (Río Barrial, 2015, p.13).

Si bien su aproximación es eminentemente pedagógica, Barrial hace una caracterización muy precisa del contexto político y cultural que se vivió durante los últimos 40 años del siglo XX en Cuba. Además, aborda las posturas de la intelectualidad internacional, dividida entre el apoyo y el rechazo a las posturas culturales del gobierno revolucionario, dirigido por Fidel Castro.

En su artículo *La moda en la literatura cubana: 1960-1979: Tejiendo y destejiendo al hombre nuevo* (2019) Cabrera Arús analiza la representación de la moda en la literatura post-59. Su investigación recorre las representaciones del vestir y su significado dentro de la ideología del socialismo cubano y su relación con la idea del hombre nuevo. Esta autora entiende *moda* el sistema de significados asociados a las prácticas vestimentarias y de adornos del cuerpo y las políticas de modulación del gusto colectivo en lo que refiere a la vestimenta. A partir de textos literarios, esta autora concluye que la moda adquirió un uso simbólico para comunicar valores y afiliaciones políticas relativas al concepto del hombre nuevo que se pretendía crear en Cuba. Argumenta que en esta nueva sociedad la moda sirvió para distinguir a los revolucionarios comprometidos de los “gusanos”<sup>13</sup>, en tanto los primeros vestían de acuerdo a los patrones catalogados como deseables por el Estado.

---

<sup>13</sup> Utilizada despectivamente por el gobierno revolucionario cubano para referirse a todos aquellos miembros de la sociedad cubana que no apoyaran, abiertamente o no, con las doctrinas revolucionarias.

Si bien es cierto que ninguna de estas investigaciones se aproxima a la literatura cubana desde la metodología de la L.C., sí es interesante destacar que los temas y posturas registradas por estas autoras están en consonancia con los resultados de esta investigación, los cuales se abordarán más adelante.

En definitiva, los trabajos presentados hasta aquí permiten ilustrar dos puntos importantes para esta investigación. Por una parte, si bien es cierto que no existen estudios sobre la literatura cubana desde la lingüística de corpus, los referenciados en esta investigación permiten anticipar que el triunfo de la revolución cubana sí tuvo un impacto en el quehacer literario de los escritores cubanos, tanto dentro como fuera de la isla. Además, *a priori*, se puede pensar que, a pesar de la variedad de temas trabajados dentro de la literatura cubana de este período, hubo algunos que predominaron dentro de este quehacer artístico.

Por otra parte, las investigaciones literarias realizadas desde la perspectiva de la L.C. permiten constatar que, efectivamente, existe una lógica y una metodología científica que es lo suficientemente precisa para determinar patrones literarios y cuyo impacto va más allá de la lingüística, ya que a partir de estos patrones (que, por lo general, reflejan rasgos políticos, sociales, culturales, etc.) posibilitan la caracterización de un espacio temporal en un espacio geográfico determinados.

Si las técnicas de la L.C. son lo suficientemente finas para corroborar o refutar lo planteado por estos investigadores sobre la literatura cubana de la época en cuestión, es algo que se verá más adelante.

## 4 Planteamiento

### 4.1 Planteamiento del problema

Como se ha venido comentando, la literatura, además de ser expresión artística, es un reflejo de la identidad de los pueblos en tanto está condicionada por el contexto en el que se escribe y por la subjetividad propia de quien escribe la obra. Así, a partir del análisis de textos literarios se puede formar una idea de las características de una sociedad determinada en una época determinada.

El caso de Cuba en este sentido es *sui generis* dentro de la literatura latinoamericana, ya que, si bien se adscribió, con sus particularidades, a los cánones literarios latinoamericanos durante la primera mitad del siglo XX, esto cambió a partir de 1959, debido en gran medida a la política cultural oficial (Castro, 1961).

En este sentido, se debe hacer notar que también la crítica y los estudios literarios dentro del país estuvieron y están supeditados a los lineamientos de las instituciones oficiales. Por otra parte, los estudios sobre literatura realizados desde fuera de la isla, si bien no están controlados oficialmente no dejan de estar sesgados por la idea de confrontar el paradigma estatal.

En este contexto surge la idea de la presente investigación. Hasta donde se tiene noticia no se ha realizado antes un estudio sobre literatura cubana desde la perspectiva de la Lingüística de Corpus y llevarlo a cabo parece una empresa pertinente si se tiene en cuenta que a pesar de la subjetividad que implica cualquier investigación, una aproximación desde la L.C. parece ser la más neutral ya que se trata de un tipo de investigación guiada estrictamente por los datos que se van recopilando.



## **4.2 Hipótesis**

A raíz de este problema surgen dos hipótesis fundamentales que son la base de esta investigación, a saber:

1. El triunfo de la Revolución Cubana y el consecuente planteamiento de una nueva política cultural, alineada con los intereses de la nueva clase dominante significaron un cambio total en las temáticas tratadas por la literatura cubana hasta 1959.
2. Una de los temas centrales dentro de la literatura cubana post 1959 será el tema de la emigración.

## **4.3 Objetivos**

Así, a partir de las hipótesis se plantearon los siguientes objetivos:

1. Determinar, utilizando métodos propios de la Lingüística de Corpus, cuáles fueron los principales temas manejados por la literatura cubana antes y después de 1959 para así establecer si estos variaron.
2. Determinar si, en efecto, la migración se convirtió en un tema fundamental de la literatura cubana después de 1959.

## **4.4 Preguntas de investigación**

En este sentido se trató de dar respuesta a las siguientes preguntas:

1. ¿Hubo cambios en las principales temáticas literarias en Cuba después de 1959? ¿Cuáles?
2. ¿Se comenzó a tratar la migración como uno de los principales temas de la literatura cubana después de 1959?

## 5 Metodología

Como se ha comentado anteriormente, el presente trabajo bien podría considerarse una propuesta multidisciplinaria, si se tiene en cuenta que el objeto de nuestro análisis son textos literarios cuyo análisis, desde la L.C, puede tener implicaciones, sociales, políticas y culturales, que bien podrían ser explotados en campos relativos a las ciencias sociales.

Así, sí es importante puntualizar que la propuesta metodológica que se ha seguido está determinada por las pautas de la Lingüística de Corpus. En este sentido, se ha realizado un estudio guiado por datos y no determinado por estos.

Las hipótesis, los objetivos y preguntas de investigación planteadas han servido como guía a esta investigación; pero se ha dejado que sean los datos los que confirmen o refuten estas, sin tratar de ajustarlos a la conveniencia de la investigación.

En este capítulo se pretende describir la metodología seguida a lo largo de esta investigación. Las primeras secciones están dedicadas a la descripción de los corpus usados y los pasos seguidos para su preparación (en el caso de corpus experimental). Asimismo, se retoman conceptos discutidos en secciones anteriores<sup>14</sup> como palabras clave, colocaciones, frecuencia y dispersión. De la misma manera se retoman las ideas relativas al análisis de sentimientos. Finalmente se analizarán los pasos seguidos para realizar la agrupación jerárquica, una idea que no estaba contemplada originalmente en los propósitos de esta investigación; pero que sirvió para demostrar de manera muy efectiva que sí es posible, utilizando técnicas propias de la L.C., establecer patrones que permiten agrupar y por tanto determinar diferencias en corpus compuestos por obras literarias.

---

<sup>14</sup> Ver secciones 2.1.2, 2.1.3.

## 5.1 Descripción del Corpus

Para esta investigación se trabajó con dos corpus: uno de creación propia y el CREA. El Corpus de Literatura Cubana del Siglo XX+ (a partir de ahora CLCXX+) abarca un período temporal de 1902 a 2020<sup>15</sup>. Todos los autores contemplados son cubanos nacidos en Cuba. No se tomó en cuenta ningún autor extranjero, si bien es cierto que algunos escritores no cubanos han vivido y desarrollado su obra en Cuba (Ernest Hemingway, Daniel Chavarría). Un caso excepcional es el del escritor Pablo de la Torriente Brau, cuyas obras sí fueron incluidas en el corpus, ya que a pesar de haber nacido en Puerto Rico vivió en la isla desde su niñez y en su biografía oficial su nacionalidad figura como cubana<sup>16</sup>.

Para su estudio y ajustándose a los objetivos y preguntas de investigación, este gran corpus de literatura cubana se dividió en dos. Esta división se realizó basada en un criterio estrictamente temporal. Se tomó 1959, año del Triunfo de la Revolución Cubana, como centro deíctico temporal, pues justamente la hipótesis plantea que es a partir de este momento que comienzan a producirse los cambios temáticos en la literatura cubana y que la migración entra a jugar un papel central en esta. La primera sección del corpus (a partir de ahora CLCXX+1) abarca desde 1902 hasta 1958 y la segunda sección del corpus (a partir de ahora CLCXX+2) abarca desde 1959 hasta 2020.

Como se puede consultar en el Apéndice 1. Listado de autores y novelas, no se tomó en cuenta el hecho de que las obras fueran o no publicadas en Cuba. Esta determinación responde a varios factores. En primer lugar, la escasez de material disponible, específicamente el relativo a la época pre revolucionaria (CLCXX+1) ya que muchos autores

---

<sup>15</sup> Todos los textos de narrativa que componen el corpus fueron recogidos en su totalidad, no fragmentos de los mismos y se acogen a los criterios de selección descritos en esta sección.

<sup>16</sup> El listado completo de autores, textos, fechas y lugar de publicación se puede consultar en la sección Apéndice 1. Listado de autores y novelas, Tabla 6. Listado de novelas y autores que conforman el CLCXX+.

de esta época no se volvieron a publicar en Cuba después de 1959. En segundo lugar, el hecho de que muchos autores fueran censurados o tuvieran que exiliarse de la isla, lo cual tuvo como consecuencia directa que no se publicaran más en el país.

Así, el CLCXX+1 está compuesto por 19 textos de narrativa: 13 novelas y 6 antologías de cuentos, todos de autores cubanos. Estos textos comprenden un período de 1902 a 1958.

El CLCXX+2 está compuesto, a su vez, por 26 textos de narrativa: 24 novelas y 2 antologías de cuentos, todos de autores cubanos. Este corpus comprende un período de 1959 a 2020.

Existe un desbalance en la cantidad de textos entre ambos corpus, el CLCXX+1 está compuesto por 7 textos menos que el CLCXX+2. Como se ha comentado anteriormente esto se debe a la disponibilidad de textos, siendo mucho más difícil de encontrar textos pertenecientes al primer período. El CLCXX+1 tiene un total de 1,202,234 *tokens* y el CORPUS-POST 2,527,811 *tokens*. En este caso se trata de un “*opportunistic corpus*” o corpus ocasional, ver página 20 (McEnery & Hardie, 2012), cuya confección estuvo supeditada a la disponibilidad de los datos.

La confección del corpus estuvo supervisada y avalada por un especialista en la materia. De la misma manera, se procuró asesoría especializada en los asuntos relativos a la literatura cubana para asegurar representatividad en la confección del corpus. Otro aspecto importante relacionado, de nuevo, con la disponibilidad de textos es el desbalance entre escritores y escritoras. En total, los corpus están compuestos por obras de 21 escritores, de los cuales, solo 3 son mujeres, ver Apéndice 1. Listado de autores y novelas.

Como corpus de referencia para las comparaciones se utiliza el CREA.

El Corpus de Referencia del Español Actual (CREA) es un conjunto de textos de diversa procedencia, almacenados en soporte informático, del que es posible extraer información para estudiar las palabras, sus significados y sus contextos. El CREA cuenta, en su última versión (3.2, junio de 2008), con algo más de ciento sesenta millones de formas. Se compone de una amplia variedad de textos escritos y orales, producidos en todos los países de habla hispana desde 1975 hasta 2004. Los textos escritos, seleccionados tanto de libros como de periódicos y revistas, abarcan más de cien materias distintas. La lengua hablada está representada por transcripciones de documentos sonoros, obtenidos, en su mayor parte, de la radio y la televisión. (*Corpus de Referencia del Español Actual (CREA) / Portal del Hispanismo*)

Se debe señalar que el CREA no es un corpus perfecto y ha sido objeto de diversas críticas, positivas y negativas por parte de diversos autores. Según Sánchez Rojo (2010) el CREA (y el CORDE) son corpus que ofrecen muchas ventajas a los investigadores. Primeramente, codifica la información de acuerdo a país, área y sub-área temática, año y tipo de texto. Esta codificación, además permite trabajar con períodos específicos de tiempo (20, 30, 120 años); esto posibilita recuperar y analizar la evolución temporal de un determinado ítem léxico, o sea, permite realizar estudios diacrónicos de determinados rasgos lingüísticos, ver página 19.

Continúa Sánchez Rojo argumentando las ventajas del CREA, así plantea que permite realizar búsquedas complejas de las frecuencias de un ítem determinado al combinar tres o más parámetros (tipo de texto, tema, período de tiempo) y que además permite exportar las concordancias.

Sin embargo, el CREA ha sido también objeto de duras críticas negativas, como su sesgo eurocentrista, ya que el 50% del material utilizado para su confección procede de España y el otro 50% de las “zonas lingüísticas tradicionales: andina, antillas (caribeña), caribe

continental, chilena, Estados Unidos, México y Centroamérica y Río de la Plata” (*Parámetros habituales (CREA)*, 2023).

Además Davies (2009) plantea que debido a su extensión, el CREA solo devuelve resultados cuando el ítem léxico ocurre menos de 1000 veces en el corpus. Asimismo, presenta la frecuencia no normalizada de los casos obtenidos correspondiente a algunos años de mayor acumulación; pero no la presenta por siglos. Otro de los grandes problemas que Davies señala es el hecho de que el CREA no está lematizado, lo cual limita la posibilidad de realizar búsquedas complejas.

## **5.2 Preparación del corpus experimental**

Debido a limitaciones temporales y logísticas (se debe tener en cuenta que esta investigación se desarrolló en tiempos de la pandemia de CoVid-19) se procuró que todos los textos ya estuvieran en formato digital, ya fuera .doc o .pdf porque no fue posible desplazarse a bibliotecas y obtener copias de ejemplares para poder digitalizarlas. Además, debido al carácter exploratorio de esta investigación, se consideró suficiente la cantidad de textos digitales encontrados, no se necesitó recurrir a textos físicos. Todos se convirtieron a formato .txt y con el software Notepad++ se realizó una corrección ortotipográfica de carácter semiautomática con el fin de poder procesar los textos.

La recopilación de textos supuso un reto para esta investigación, en parte por la escasez de los mismos, sobre todo los pertenecientes al CLCXX+1 y por el estado en el que fueron hallados los mismos. En ninguno de los casos los formatos digitales permitían el procesamiento directo con el software AntConc<sup>17</sup> Versión 3.5.9. Este software es

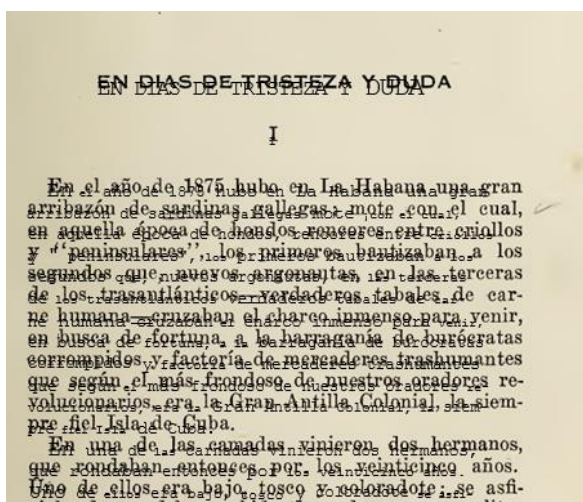
---

<sup>17</sup> Los softwares utilizados para esta investigación son herramientas que permiten el procesamiento, anotación, etiquetado y análisis de un corpus. Estas herramientas exigen una preparación de los textos como la descrita en esta sección.

una herramienta gratuita y multiplataforma que se utiliza para el aprendizaje basado en datos y para llevar a cabo investigaciones que se fundamentan en la lingüística de corpus. Funciona en cualquier ordenador con Microsoft Windows (probado en Win 98/Me/2000/NT, XP, Vista, Win 7), Macintosh OS X (probado en 10.4.x, 10.5.x, 10.6.x) y Linux (probado en Ubuntu 10, Linux Mint). AntConc se ha desarrollado en Perl, se han utilizado diversos compiladores para generar ejecutables para los diferentes sistemas operativos. (Anthony, 2023.)

En la mayoría de los casos se debió llevar a cabo un proceso de formateo de hasta tres pasos, a saber: (1) Formato .pdf a formato .jpg, (2) formato .jpg a formato .docx, (3) Formato .docx a formato .txt.

La Imagen 3 muestra el problema que representa el texto en formato .pdf para su procesamiento, a saber, ilegibilidad, líneas de texto superpuestas y la imposibilidad de cargar este formato en AntConc. Además, se puede observar el mismo texto una vez que ha sido procesado. Sin embargo, el formato .docx tampoco permite procesar el texto en AntConc (v.3.5.9), por lo que debe hacer una tercera conversión de formato a .txt, ver Imagen 4.



#### EN DIAS DE TRISTEZA Y DUDA

En el año de 1875 hubo en La Habana una gran arribazón de sardinas gallegas; mote con el cual, en aquella época de hondos, rencores entre criollos y “peninsulares”, los primeros bautizaban a los segundos que, nuevos argonautas, en las terceras de los trasatlánticos verdaderos tabales de carne humana cruzaban el charco inmenso para venir, en busca de fortuna, a la barraganía de burócratas corrompidos y factoría de mercaderes trashumantes que según el más frondoso de nuestros oradores revolucionarios, era la Gran Antilla Colonial, la siempre fiel Isla de Cuba.

En una de las camadas vinieron dos hermanos, que rondaban entonces por los veinticinco años. Uno de ellos era bajo, tosco y coloradote; se asfixiaba dentro del consabido terno de pana

Imagen 3. Texto recuperado en .pdf vs. texto procesado en .word

Una vez que todos los textos estuvieron en formato .txt se etiquetaron con el software TagAnt (L. Anthony, 2022) lo cual permitió realizar una lematización y un análisis de categoría gramatical.

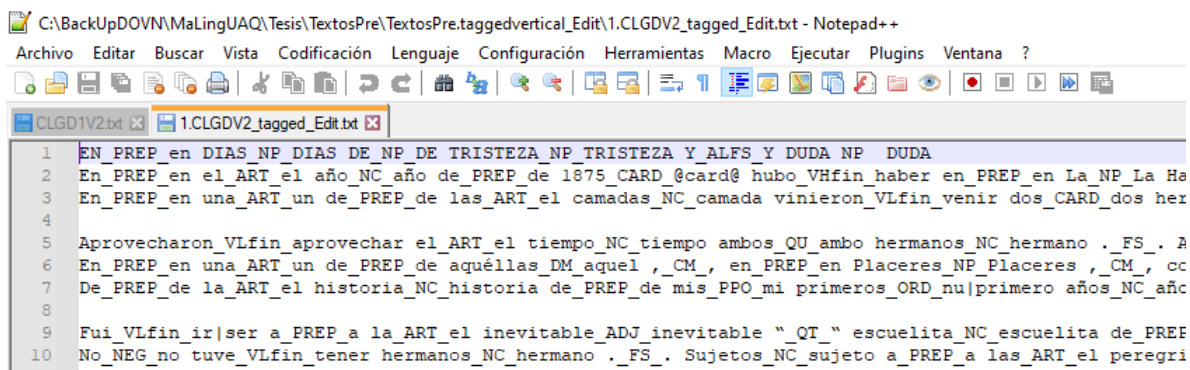


Imagen 4. Texto procesado y etiquetado en Notepad++

Al concluir los procesos de corrección y etiquetado, todos los textos fueron procesados con el software AntConc (L. Anthony, 2022.).

### 5.3 Extracción de palabras clave

Como se ha comentado anteriormente, en la sección Palabras Clave, nodos y colocaciones, las palabras claves son aquellas que tienen más frecuencias relativas al realizarse la comparación entre dos o más corpus y tienen un status especial dentro del texto ya que son portadoras de un importante significado social y desde el punto de vista lingüístico contribuyen a la búsqueda de unidades de significado (Stubbs, 2010). A partir de un análisis de palabras clave se pueden señalar los rasgos fundamentales que marcan las diferencias entre un corpus y otro (Rodríguez Sánchez & Bautista Botello, 2021). Así, a los efectos de esta investigación, las palabras claves permitieron discretizar los temas característicos de cada período estudiado y establecer las diferencias entre cada uno, las cuales se comentarán más adelante.



Para determinar las palabras claves y de ahí los temas característicos de cada época se realizaron una serie de comparaciones. Primeramente, se comparó el CLCXX+1 con el CREA y se determinaron las palabras claves de esta sección del corpus. Un proceso similar se realizó con el CLCXX+2. Este primer cruce se hizo para determinar las palabras clave de cada sección del corpus basadas en su frecuencia relativa<sup>18</sup>.

Una vez que se determinaron las palabras clave pertenecientes a cada una de estas secciones del corpus a partir de la comparación con la totalidad del CREA, se procedió a comparar todos los textos de ambas secciones del corpus entre sí<sup>19</sup>.

En esta investigación se trabajó solo con las palabras clave positivas ya que al realizarse dos cruces las palabras que en el primero aparecerían como negativas, van a aparecer como positivas en el segundo cruce. O sea, al realizar un cruce CLCXX+1 como corpus de experimental y el CLCXX+2 como corpus de referencia se obtiene un set determinado de palabras más frecuentes a partir del cual se pueden determinar las palabras clave. Los valores de frecuencia obtenidos en este cruce van desde X hasta 0, estas son las palabras claves del corpus CLCXX+1, o sea las más frecuentes en relación al CLCXX+2. Si se decidiera continuar el conteo después de 0 se obtendría un grupo de palabras con frecuencia negativa; para no trabajar con frecuencia negativa es justamente que se realiza el segundo cruce, en este caso invirtiendo los papeles de cada corpus, esto es el CLCXX+1 se toma como corpus de referencia y el CLXX+1 como corpus experimental (ver Tabla 2). Así se determinan las

---

<sup>18</sup> (ver página 24)

<sup>19</sup> Se realizan dos cruces porque de realizarse un solo cruce se obtendrían los mismos resultados. O sea, un listado completo de palabras clave de un cruce arroja palabras positivas y negativas: las palabras con un índice positivo alto son las palabras que ocurren más que en el corpus de referencia; las palabras con un índice cercano a 0 se llaman Lockwords y ocurren con la misma frecuencia relativa en ambos corpus; finalmente las palabras negativas ocurren menos que en el corpus de referencia.

palabras claves de cada corpus. En la Tabla 2, a manera de ejemplo, se pueden observar las primeras 60 palabras clave relativas al CLCXX+2. En este listado abundan los nombres propios, lo cual es lógico ya que se trata de personajes de las novelas. Estos nombres propios no son relevantes para el análisis desarrollado en esta investigación, por lo tanto se desechan. Sin embargo, en estas primeras 60 aparecen algunas que resultan interesantes en tanto no son frecuentes en el CLCXX+1 y que podrían ser representantes, debido a su alta frecuencia de nuevos temas presentes en el CLCXX+2, a saber, **expletivos** (41) *mierda*, (44) *carajo*; **locativos** (53) *Cuba*, (58) *apartamento*.

Este proceso de decantación se realizó con las primeras 300 palabras clave en cada uno de los cruces, es decir, se realizaron 4 cruces *ergo*, se analizaron 1,200 palabras claves, ver Apéndice 2. Listados Palabras Clave determinadas en cada uno de los cruces. De estas 1,200 se mantuvieron aquellas que en realidad mostraran anomalías en cuanto a las frecuencias y los temas a los que se pudieran referir. Posteriormente se agruparon desde un punto de vista temático y semántico.

Así se discretizaron los campos temáticos presentes en los corpus de interés. De todos los campos temáticos discretizados se decidió trabajar a profundidad con uno solamente, en tanto abarcarlos todos redundaría en un imposible por cuestiones temporales. Sobre estos campos temáticos se comentará en la sección Análisis y resultados. El siguiente paso del análisis, una vez determinadas las palabras clave y los temas de interés fue analizar las líneas de concordancia en las que ocurrían estas palabras, sección Colocaciones. Líneas de concordancia. Sentimientos.

Rango	Frecuencia	Keyness	Keyword
1	7543	1085.34	dijo
2	1214	917.57	conde
3	1414	738.78	carlos
4	14438	717.71	había
5	716	618.12	fronesis
6	594	512.79	cemí
7	571	492.93	blen
8	592	488.49	irving
9	611	472.51	elisa
10	517	433.91	horacio
11	13056	404.13	pero
12	453	391.06	bobby
13	511	381.65	coño
14	484	379.81	esteban
15	425	366.89	nadiezdha
16	490	364.1	bernardo
17	448	356.97	walter
18	408	352.21	foción
19	1578	346.19	dije
20	12159	341.09	era
21	586	336.96	flaco
22	582	333.88	incluso
23	362	312.5	loreta
24	3082	303.06	esa
25	9445	300.79	yo
26	344	296.96	sofia
27	367	296.15	darío
28	465	294.59	adela
29	363	292.74	fidel
30	5044	287.73	porque
31	333	287.46	licario
32	360	282.31	david
33	5712	279.66	estaba
34	319	275.38	manolo
35	4231	275.34	donde
36	325	269.08	adriano
37	34176	268.42	no
38	500	267.64	alberto
39	4756	266.55	ahora
40	7799	265.76	él

41	442	263.01	mierda
42	299	258.11	franqui
43	422	256.06	lino
44	433	251.81	carajo
45	290	250.34	ofelia
46	494	245.46	gordo
47	290	239.09	castro
48	4086	232.73	eso
49	397	232.28	cué
50	266	218.55	magistrado
51	379	214.33	bárbaro
52	242	208.9	gisela
53	1423	202.94	cuba
54	795	202.43	clara
55	234	202	antoni
56	345	200.36	oso
57	232	200.27	orestes
58	324	196.82	apartamento
59	1716	192.06	sabía
60	378	191.3	inmediato

Tabla 2. Listado de las primeras 60 palabras más frecuentes resultantes del cruce CLCXX+2 vs. CLCXX+1

#### 5.4 Colocaciones. Líneas de concordancia. Sentimientos

Una vez completada y analizada la lista de palabras clave se procedió a agruparlas por campos temáticos, para poder establecer los temas abordados por la literatura cubana en ambos períodos (ver sección 6.1 Determinación de palabras clave). Si bien algunos temas fueron comunes a ambas secciones del corpus, hubo uno que destacó precisamente por ser central en el CLCXX+2 y no aparecer tratado en el CLCXX+1, a saber, la *migración*. Así se decidió enfocarse en este tema. A partir de las palabras clave relativas a este se decidió analizar las colocaciones y las líneas de concordancia (ver sección Palabras Clave, nodos y colocaciones). Según Bondi

the cultural and ideological implications of a lexical element can be illuminated by an analysis of its collocation and semantic preference – the tendency of the word to

co-occur with other words and with words belonging to a specific semantic category or field. (Bondi, 2010, p.3)

A partir de esta idea de Bondi y teniendo en cuenta el tema de análisis elegido, esto es, la migración, se enfocó el análisis en algunos sustantivos locativos (específicamente *Cuba* u otros sustantivos relativos a Cuba como *isla, país, etc.*). Así se tomaron estos sustantivos como centro deíctico locativo y siguiendo la lógica que la migración es un movimiento de seres vivos, por tanto, cualquier referencia a esta va a implicar verbos de movimiento, se decidió analizar los verbos de movimiento que más frecuentemente coocurrían con estos sustantivos. Finalmente se determinaron los cinco verbos más frecuentes (*salir, volver, regresar, quedarse, irse*), los cuales serán analizados en la sección 6 Análisis y resultados. Se debe aclarar que no solo se trabajó con estos verbos de movimiento *per se*; por cuestiones cuantitativas, se decidió incluir en la contabilidad las palabras derivadas de las raíces verbales que coocurrían con sustantivos locativos relativos a Cuba (el *regreso a la Isla, la salida de Cuba*). En la Imagen 5, se pueden apreciar justamente las colocaciones más frecuentes con el sustantivo Cuba. Así en los rangos *11, 12, 13* se encuentran las formas no personales del verbo, *salir, regresar y salido* respectivamente. En el rango *15* aparece *regreso*, y en el *18* aparece *volver*. Una vez que se constató que, efectivamente, estos son los verbos de movimiento que más coocurren con el sustantivo locativo Cuba, se procedió a determinar las líneas de concordancia en las que estos ocurrían a fin de precisar y analizar el contexto en el que el autor los utilizó y la intención con la que lo hizo, ver Imagen 6. Líneas de concordancia sustantivo locativo Cuba y formas derivadas de regresar.

Este mismo proceso se repitió con los mismos verbos y otros sustantivos locativos utilizados como “sinónimos” de Cuba en las obras, a saber, *isla, país, patria, etc.* cómo se comentó previamente en este párrafo.

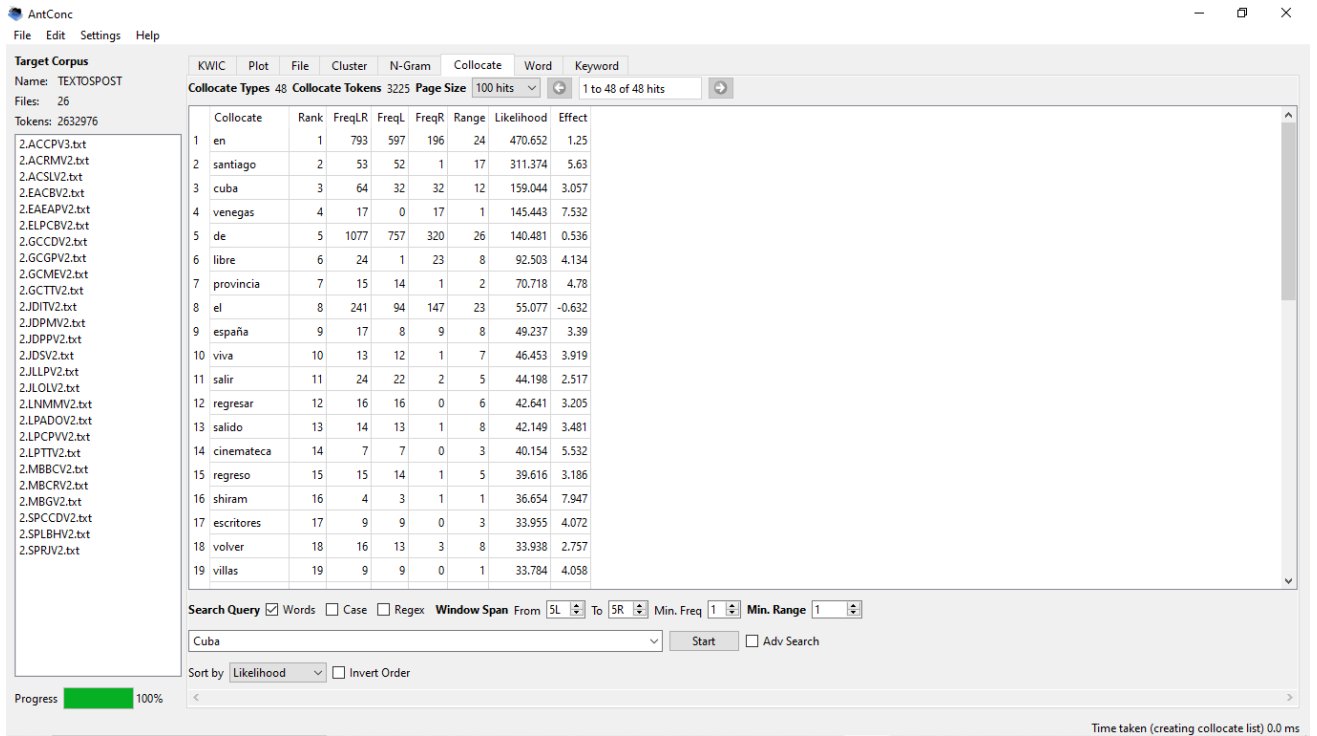


Imagen 5. Colocaciones más frecuentes con el sustantivo locativo Cuba. CLCXX+2

Es importante señalar que no solo se analizó el comportamiento de estas coocurrencias en la sección CLCXX+2 del corpus. También se analizaron en la sección CLCXX+1 con el fin de constatar que expresaban los autores con ellas y poder establecer una comparación en el uso entre ambas secciones. Esta comparación permitió establecer la variación presente entre CLCXX+1 y CLCXX+2. Los resultados derivados de esta comparación se comentarán en la sección 6 Análisis y resultados.

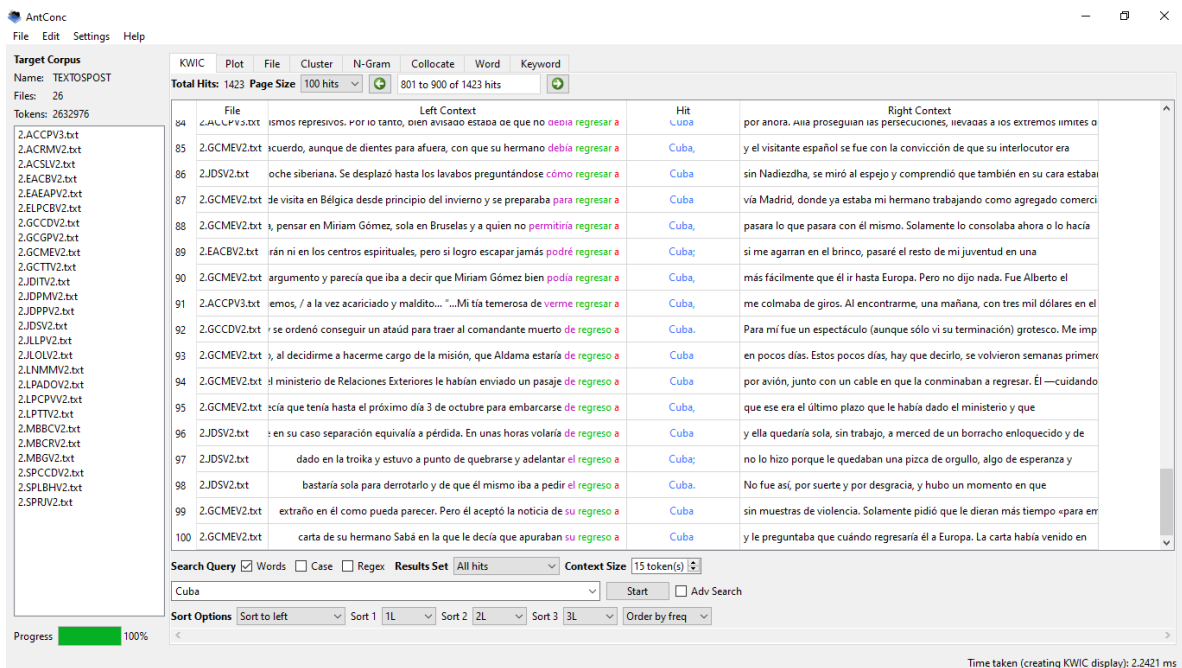


Imagen 6. Líneas de concordancia sustantivo locativo Cuba y formas derivadas de regresar

Al establecer las líneas de concordancia se pudo determinar el contexto (lingüístico) en el cual se desarrollaba el tema -migración-. A partir de un cuidadoso análisis manual, esto es analizando cada línea de concordancia en la que aparecían sustantivos relativos a *Cuba* con verbos de movimiento<sup>20</sup>, se establecieron una serie de rasgos socioculturales, económicos, políticos y sobre todo anímicos que posibilitaron formar una idea del contexto (literario) en el que transcurría la obra. Teniendo en cuenta que el panorama sociocultural, político y económico de la época postrevolucionaria ha sido abordado desde diferentes perspectivas de estudio pareció más interesante y novedoso conducir una aproximación a los rasgos anímicos<sup>21</sup> expresados por los personajes de las novelas.

<sup>20</sup> En la mayoría de los casos fue necesario expandir la lectura más allá de las líneas de concordancia ofrecidas por AntConc para poder determinar con mayor precisión el contexto en que se realizaban los rasgos de interés.

<sup>21</sup> Entiéndase por rasgos anímicos los sentimientos y emociones reflejadas por los personajes a través de sus diálogos o por el narrador a través de descripciones situacionales.

El análisis de sentimientos se realizó igualmente de manera manual. Como se comentó anteriormente, si bien se siguió la propuesta teórica de Liu (2015) no se recurrió a ningún software para la clasificación de los sentimientos. Así, a partir de las líneas de concordancia, se realizó un análisis en dos niveles: el oracional y el de aspecto<sup>22</sup>. Asimismo, se utilizó la taxonomía de sentimientos propuesta por Valderrama (2015) para determinar la naturaleza de los sentimientos reflejados en los textos. Para entender mejor el análisis en estos niveles véase el ejemplo 1.

1. Miedo: Eliseo Alberto. Caracol Beach

(...) que consultó el asunto con su «superior», el falso teniente Lázaro Samá. Fue un diálogo difícil. **Beto había decidido no volver a Cuba por temor a represalias. Tenía miedo.** Estaba convencido de que lo considerarían un traidor. Ya era un (...)

En el ejemplo 1 es evidente el sentimiento que experimenta el personaje ante la perspectiva de regresar a Cuba. *Temor y miedo* se pueden catalogar como *palabras sentimientos* de acuerdo a Liu. Además, en consonancia con la taxonomía de Valderrama el *miedo* clasificado como uno de los sentimientos negativos que puede experimentar el ser humano.

Un tipo de análisis similar se llevó a cabo con todas las líneas de concordancia en las que se constató la ocurrencia de sustantivos locativos relativos a Cuba y verbos de movimiento (181 en total), si bien en muchos casos no fue suficiente el nivel oracional, y hubo que expandir el análisis al nivel aspectual, esto es, ir más allá de la línea de concordancia específica y consultar una sección más amplia del texto para poder determinar el sentimiento que se pretendía expresar.

---

<sup>22</sup> Ver sección 2.2



## 5.5 Análisis de agrupación

Hasta aquí se han expuesto los diferentes pasos del análisis seguido en esta investigación. Se ha visto cómo se determinaron las palabras clave y cómo los datos guiaron todo el proceso hasta decidir poner el enfoque en los verbos de movimiento que coocurren con el sustantivo locativo Cuba y sus correferentes. Asimismo, se describió el proceso analítico relativo a las líneas de concordancia y cómo se determinaron las *palabras sentimientos* que permitieron determinar los estados de ánimo provocados por la migración y que fueron reflejados en las distintas obras. El análisis de las líneas de concordancia es, tal vez, el punto más importante de esta investigación en tanto es donde confluyen todas las partes de la misma, a saber, (i) la parte lingüística reflejada por las palabras clave analizadas y la determinación de las palabras sentimientos (ii) la parte psicológica, en tanto estas palabras de sentimientos permiten establecer los estados de ánimo reflejados en los textos y (iii) la parte sociológica ya que justamente a partir de estos estados de ánimo se puede establecer una especie de panorama de la sociedad cubana de la época. Se debe mencionar, por supuesto, el aspecto literario de la investigación, en tanto todo este análisis se hizo a partir de textos literarios.

En un principio se diseñó la investigación para abarcar sólo hasta el análisis de las líneas de concordancia y las palabras sentimientos; pero se decidió llevarla un paso más allá y realizar un trabajo de agrupación, que permitiera, a través del análisis cualitativo de una serie de rasgos lingüísticos, agrupar temáticamente las novelas.

Este proceso de agrupación es una consecuencia natural del análisis que se ha descrito hasta aquí. Los rasgos lingüísticos tomados en cuenta fueron precisamente las palabras sentimientos identificadas gracias a las líneas de concordancia. Así, se establecieron campos semánticos que permitieran crear redes y agrupar conceptos derivados de los originales.

Este tipo de trabajo de agrupación permite, si los rasgos se han determinado correctamente, establecer si es posible afirmar que existen diferencias temáticas y/o temporales relevantes entre el CLCXX+1 y el CLCXX+2. Además, es posible agrupar temáticamente los textos trabajados. En este caso específico, determinar en qué sección del corpus es más recurrente este tema. Sobre los resultados de este proceso de agrupación se hablará en la sección 6.2.2 Análisis de agrupación. Resultados.

Como se comentó, el análisis de agrupación es de tipo cuantitativo. Se realizó con la herramienta #LancsBox: Lancaster University corpus toolbox. Esta herramienta en realidad es una interface web que permite hacer un análisis estadístico (en este caso un análisis de agrupación jerárquica) en la paquetería R. Esta herramienta trabaja basada en las frecuencias relativas de las palabras (rasgos) de interés. La Tabla 3 es solo un ejemplo de algunos de los rasgos analizados en 9 de las novelas de la sección CLCXX+1 del corpus y sus frecuencias relativas. Como se explica a continuación, todos los rasgos se discretizaron construyendo campos semánticos a partir de las palabras sentimientos. Después se buscaron sus frecuencias relativas con la herramienta AntConc. Una lista completa de los rasgos y sus frecuencias en cada novela del corpus se puede ver en la sección 11.

Según Brezina (2018)

By virtue of being in a text together, many linguistic variables are related in some way. For example, in English the relative frequency of adjectives in texts is related to the relative frequency of nouns because adjectives modify nouns and therefore typically occur together with them (...)

Así, este software analiza la frecuencia relativa de los rasgos lingüísticos en un texto y sus coocurrencias más frecuentes para realizar el proceso de agrupación. Siguiendo la metodología propuesta por Brezina (2018) para este tipo de análisis, se decidió trabajar con

41 rasgos lingüísticos, esto para lograr una agrupación óptima. Los rasgos fueron determinados a partir de la taxonomía de sentimientos propuesta por Valderrama (2015) y de los conceptos manejados por Marina y Penas (2000). Así, tomando como núcleo cada uno de los sentimientos discretizados por Valderrama y constatados en el corpus se procedió a la elaboración de campos semánticos.

Según Lewandowski (1982) un campo semántico es un grupo de palabras o expresiones que abarcan un determinado grupo de significados, que dependiendo de sus significados léxicos pertenecen a un mismo sistema lingüístico y que comparten al menos un rasgo semántico.

		AC_EA	AC_EYO	AC_PP	AC_REM	CL_GD	CM_HSM	DML_J	ED_D	LCC_N
1	ansi*	0.0	0.2	1.0	2.0	1.1	0.4	0.3	3.0	0.8
2	conform*	0.0	0.0	0.2	0.0	0.9	0.9	0.7	7.0	1.8
3	curios*	0.0	0.2	1.0	0.0	0.6	1.9	1.4	14.9	3.9
4	desacuerdo	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	desengañ*	0.0	0.0	0.3	0.0	0.2	0.0	0.0	0.0	0.0
6	esperanza*	0.4	0.4	1.0	0.7	0.7	2.2	1.6	16.9	4.4
7	impoten*	0.0	0.0	0.2	1.0	0.1	1.8	1.3	13.9	3.6
8	incredul*	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9	miedo*	1.1	0.6	3.9	2.3	2.2	3.9	2.9	29.8	7.7
10	neutr*	0.0	0.0	0.1	0.0	0.1	0.4	0.3	3.0	0.8
11	nostalgi*	0.0	0.0	0.1	0.3	0.3	0.3	0.2	2.0	0.5
12	obliga*	0.4	0.9	1.9	1.3	0.6	2.7	2.0	20.9	5.4
13	religi*	0.4	1.3	1.1	0.7	0.3	0.5	0.4	4.0	1.0
14	resenti*	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
15	resigna*	0.4	0.0	0.3	0.3	0.2	0.3	0.2	2.0	0.5
16	separa*	0.0	0.4	1.4	0.3	1.0	3.1	2.3	23.9	6.2
17	*migr*	0.4	1.1	1.0	0.3	0.8	0.0	0.0	0.0	0.0

Tabla 3. Frecuencia relativa de 17 rasgos en 9 novelas. Sección CLCXX+J<sup>23</sup>

<sup>23</sup> Se utiliza un asterisco (\*) precediendo y/o sucediendo a la raíz de la palabra para asegurar que el software AntConc encuentre todas las ocurrencias en el corpus de las palabras, cual sea que sea su tipo, que se formen a partir de esta raíz.

A partir de esta idea, como se ha comentado, se decidió tomar como punto de partida los sentimientos determinados en el corpus. Asimismo, se construyeron relaciones de sinonimia y antonimia para establecer los rasgos que se procesarían con el software Brezina ToolBox.

Paradis (2012) entiende estas relaciones de sinonimia y antonimia como sigue:

there are (ii) words that evoke similar meanings but have different forms, such as synonyms and hyponyms, and there are (iii) words that have different forms and different meanings and are semantically related through opposition such as antonyms. (Carita Paradis, 2012, p.6).

A partir de las ideas de elementos lingüísticos que comparten al menos un rasgo semántico y que pueden expresar significados similares u opuestos se determinaron 41 rasgos, los cuales se analizaron con el software Brezina ToolBox.

Anteriormente se mencionó que los rasgos lingüísticos fueron determinados a partir de la creación de campos semánticos derivados de las palabras de sentimientos determinadas en el corpus. Además de estos rasgos se decidió incluir en el análisis las frecuencias de los verbos de movimiento analizados en las líneas de concordancia y los sustantivos locativos referentes a Cuba. Esto con el fin de establecer a ciencia cierta si el uso de todos estos rasgos era en realidad diferente entre ambas secciones del corpus y si esto permitía hacer una agrupación más precisa.

Una vez que se han determinado los rasgos a analizar, se calcularon sus frecuencias relativas por novela, como se muestra en la sección 11 y se introdujeron estos datos en el software. La Imagen 7 muestra cómo luce la interfaz del software Brezina Toolbox una vez que se han cargado los datos pertinentes en el sistema. Este software realiza los cálculos para el proceso de agrupación sin ninguna intervención del usuario lo cual permite obtener un dendograma que, de haberse realizado bien el procedimiento anterior, permite graficar las diferencias que

pueden existir entre ambas secciones del corpus. En la sección 6.2.2 se comentará al respecto con más profundidad.

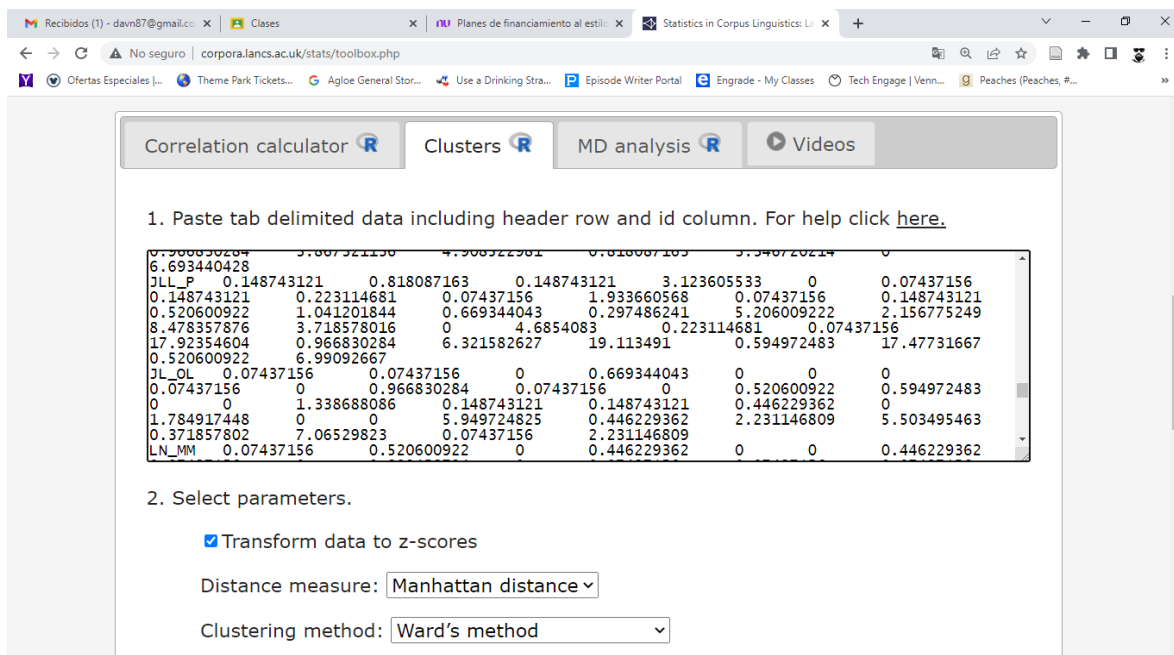


Imagen 7. Procesamiento de los datos con el software Brezina ToolBox

En este capítulo se ha revisado la metodología utilizada para desarrollar esta investigación. El análisis que se propone a partir de la misma está dividido en cuatro partes, a saber, la creación del corpus, la determinación de las palabras clave para cada sección del corpus y a partir de estas los ejes temáticos que se manejaron en cada sección. Posteriormente la extracción y análisis de las líneas de concordancia en las que aparecen las palabras clave relativas al tema de interés, o sea, la *migración* y el análisis de los sentimientos reflejados en estas líneas de concordancia. Finalmente, a partir de la creación de campos semánticos creados a partir de estos sentimientos, un trabajo de agrupación que permite determinar de manera muy precisa si, en efecto, son relevantes las diferencias temáticas y si estas son lo suficientemente fuertes para dividir el corpus temporal y temáticamente.

En el Capítulo 6 Análisis y resultados se mostrarán los resultados obtenidos a partir de la metodología descrita.

## 6 Análisis y resultados

Como se ha comentado, la literatura, más allá de su relevancia artística o lúdica, permite conocer y comprender las dinámicas de funcionamiento de determinadas sociedades en determinados momentos históricos. Así, si se analizara en conjunto la obra de Ernest Hemingway, Scott Fitzgerald, T. S. Elliot y John Dos Passos, es lógico afirmar que se podría tener una idea bastante acertada de la sociedad estadounidense de los años '20-'30 del siglo XX. De igual manera, si se realizara ese análisis a las obras de Arthur Conan Doyle, Robert Louis Stevenson, Charles Dickens y Jane Austen se obtendría una idea de los valores y dinámicas de la Inglaterra victoriana. La literatura, más allá de entretenimiento o arte, es memoria histórica.

El presente capítulo pretende mostrar, a través del análisis de palabras claves, colocaciones y líneas de concordancia los temas principales que impactaron la sociedad cubana post revolucionaria y cómo estos reflejaron, a través de la literatura, los puntos de vista, intereses y sentires de un pueblo. Específicamente se analizará la *migración*, en tanto las limitaciones temporales de esta investigación no permiten analizar otros temas a fondo y, como se verá durante este capítulo, la *migración* es un tema que ha tenido un impacto profundísimo en la sociedad cubana de la segunda mitad del siglo XX y lo que va del XXI.

De acuerdo a la metodología propuesta en el capítulo 5 Metodología, se realizó un análisis dividido en cuatro etapas. Primero se determinaron las palabras clave características de cada sección del corpus y a partir de estas, los temas predominantes en cada sección. Posteriormente, se estableció la *migración* como tema de estudio y se analizaron las palabras claves y colocaciones más frecuentes con estas palabras clave. Así se pudieron determinar y analizar las líneas de concordancia donde estas aparecen. En tercer lugar, a partir de estas

líneas de concordancia se realizó un análisis de sentimientos. Finalmente se realizó un trabajo de agrupación a partir de rasgos semánticos derivados de los sentimientos presentes en el corpus que permitió agrupar las novelas de acuerdo al tratamiento que, individualmente, le daban al fenómeno migratorio.

## **6.1 Determinación de palabras clave**

Como se comentó en la sección 2.1.2, las palabras clave son uno de los ejes centrales de la L.C y por su frecuencia y dispersión son las que gozan de un status especial dentro del corpus. Además, son las más relevantes tanto lingüística como extralingüísticamente pues justamente a partir de esta frecuencia y dispersión es que se hacen evidentes los rasgos de interés para la investigación.

Así, el primer paso del análisis fue determinar las palabras clave para cada sección del corpus. Estas palabras clave permitieron establecer y comparar los temas característicos de cada sección y por tanto de cada período histórico. Al analizar diacrónicamente el corpus se pudo establecer, en un primer momento, que los temas sí variaron entre las secciones, siendo la *migración* el elemento más notable en esta variación. De acuerdo a la metodología propuesta en la sección 5.3 Extracción de palabras clave, para determinar estas palabras clave se realizaron cuatro cruces, a saber:

1. CREA (corpus de referencia) vs. CLCXX+1 (corpus experimental)
2. CREA (corpus de referencia) vs. CLCXX+2 (corpus experimental)
3. CLCXX+1(corpus de referencia) vs. CLCXX+2 (corpus experimental)
4. CLCXX+2 (corpus de referencia) vs. CLCXX+1(corpus experimental)

Una vez obtenida la lista de palabras clave para cada una de las secciones de corpus se decidió tener en cuenta para el análisis solamente las primeras 300 palabras obtenidas en cada cruce,



es decir se revisaron un total de 1200 palabras clave a fin de determinar y agrupar los temas característicos de cada sección del corpus (ver sección Apéndice 2. Listados Palabras Clave determinadas en cada uno de los cruces. Así, se obtuvieron los siguientes resultados:

**1. Cruce CREA vs. CLCXX+1**

Temas/Palabras clave:

- a. Etnia/esclavitud: Negro, negra, negrero, mulata, barracones, esclavos.
- b. Género (femenino): Loca, mamá, niña, mujer, harén, hermana.
- c. Tierra: Jíbaro, manigua, batey, bohío, tierra.
- d. Cuerpo: ojos, carne, brazos, manos, alma, cara, corazón, cabeza, labios, pecho, boca, rostro, dedos, pies.
- e. Navegación: Mar, marinero, barco, capitán, cruceros.

**2. Cruce CREA vs. CLCXX+2**

Temas/Palabras clave:

- a. Locativos/nación: Cuba, Habana, Vedado, Fontanar, cubana, cubano, Malecón, mar.
- b. Etnia: Negro, negra, mulato, mulata, rubito.
- c. Género (femenino): Loca, madre, virgen, abuela, muchacha, mamá, puta.
- d. Expletivos: Mierda, carajo, coño, comemierda.

A partir de los resultados de este primer cruce se puede comenzar a observar que existen algunas diferencias entre las palabras clave de ambas secciones del corpus. En el caso de la sección CLCXX+1 se puede apreciar dos temas específicos que no se encuentran en la sección CLCXX+2, a saber, el tema de la tierra y el de la navegación. La presencia de estos dos temas en esta sección es lógica, en tanto estos están en sintonía con las corrientes existentes en Cuba durante la primera mitad del XX: el naturalismo, el costumbrismo y el

realismo (Fernández, 2000). Además, es una muestra que las tendencias literarias de Cuba seguían más o menos los mismos caminos que el resto de América Latina en esa época. De esta idea, es decir, de la sintonía de la literatura insular con lo que estaba ocurriendo en el continente se puede deducir que la literatura cubana no estaba supeditada a ningún interés particular, más allá de los intereses propios de los autores y los sesgos propios de sus posturas estéticas, políticas o sociales. Además, las palabras claves determinadas a partir del cruce con el CREA siguen la línea de lo planteado por Fernández López (2022), en tanto las que hacen referencia a la tierra, al género o a la esclavitud son un indicio, por una parte de los temas y por otra de la manera en la que los escritores cubanos trataron de combinarlos en sus obras enfocándose, como plantea Fernández López no solo en el hombre sino en la manera en la que este interactúa con la naturaleza.

La novela realista y la naturalista prestaron atención, además de al hombre, al entorno y la naturaleza que constituían su ámbito vital. En la novela hispanoamericana «de la tierra» puede parecer que el hombre no interesa en cuanto tal, sino en cuanto vive en un mundo natural que acaba por convertirse en protagonista verdadero. (Fernández López, 2022)

Asimismo, es interesante la aparición de temas como género y etnicidad en ambas secciones del corpus, aunque es de suponer que el acercamiento a estos temas es diferente en ambos períodos, en tanto los puntos de vista relativos a estas cuestiones han ido variando a lo largo del siglo XX. También es lógico pensar que, efectivamente, el tratamiento de estos temas en ambas secciones del corpus va a ser muy diferente si se tiene en cuenta que la Revolución Cubana sí ha tratado, al menos institucionalmente, de empoderar a la mujer cubana y atacar al racismo secular existente en Cuba. Sin embargo, por cuestiones temporales, estos temas no se tocaron en esta investigación, aunque valdría la pena realizar algún tipo de aproximación en investigaciones futuras.

Los comentarios anteriores son, sin embargo, superficiales ya que parten solamente de los datos obtenidos a partir de la comparación del CREA con los corpus experimentales. El interés real de esta investigación se centra en el contraste (que debería existir) al comparar ambas secciones del corpus. Así, el siguiente paso fue determinar las palabras clave y a partir de ellas los temas característicos de cada una de las secciones.

### **3. Cruce CLCXX+1 vs. CLCXX+2**

Temas/Palabras clave:

- a. Nación/Locativos: Cuba, Habana, cubano, cubana, Malecón, país, Fontanar, Vedado, nacional, barrio.
- b. Migración: Miami, aeropuerto, dólares, exilio, regresar.
- c. Revolución: Fidel, Castro, revolución, Batista, comunista, ministerio, Che, buró, Camilo.
- d. Expletivos: Coño, mierda, culo, puta, cojones, comemierda, cabrón, maricón, jodido.

### **4. Cruce CLCXX+2 vs. CLCXX+1**

Temas:

- a. Etnia/esclavitud: Negrero, negro, trata, negrada, esclavos, rancho.
- b. Cuerpo: Carne, ojos, corazón, alma, brazos, manos, cabellos, vientre.
- c. Navegación: Marineros, barco, capitán, armadores, piratas, marinería, contramaestre, mar, cabotaje.
- d. Tierra: Jíbaro, manigua, batey, bohío, tierra.
- e. Presidio: Presidio, galera, preso, cautivo, celdas, rejas, alcaide.
- f. Género: Harén, chiquilla, mujeres, hermana.

Al revisar las palabras claves obtenidas al realizar una comparación entre ambas secciones del corpus se pudo hacer una discriminación más precisa de los temas en cada una. Como se esperaba (y aquí se confirma la primera hipótesis) los temas sí variaron de un período a otro. En lo relativo a los ejes temáticos pertinentes a la sección CLCXX+1, este segundo cruce confirmó lo que se había planteado a partir del cruce de esta sección con el CREA, o sea, los temas de esta sección y por tanto de este período temporal están en consonancia con los temas manejados por la literatura hispanoamericana en la época. Este cruce permitió constatar que temas como la esclavitud, la tierra, la navegación y el género fueron recurrentes para los escritores cubanos. Un análisis más profundo del tratamiento de estos temas dentro de la obra es altamente recomendable y pudiera ser el objetivo de investigaciones futuras.

En lo relativo a los temas que caracterizan a la sección CLCXX+2 se pudo constatar que estos parecen ser un poco más “criollos”, en tanto tocan temas más propios de la sociedad cubana de la época, la cual estuvo marcada por profundos cambios políticos, sociales y culturales que experimentaba el país. Así, es de esperar que la literatura cubana por 1959 se distanciara de las corrientes imperantes en América Latina ya que lo que estaba sucediendo en Cuba en la década de los ´60 y lo que continuó sucediendo es un fenómeno ajeno al resto de países latinoamericanos. Así, fue posible determinar temas como *revolución*, *nación* y *migración*, los cuales, por supuesto, suponen un cambio radical al compararlos con los temas del CLCXX+1.

Hasta aquí se ha podido corroborar, con técnicas propias de la L.C., lo acertado de las investigaciones revisadas en la sección 3.2 Estudios sobre literatura cubana, relativas a los temas de la literatura en la Cuba post revolución. Como se ha venido comentando a lo largo de esta investigación, las limitaciones temporales suponen un desafío para la amplitud de la

misma, en tanto se trata solo de una investigación vinculada a un programa de maestría, así de los temas discretizados se eligió solamente uno para analizarlo a profundidad, a saber, la *migración*. Sin embargo, se debe aclarar que, si bien se decidió poner el foco de la investigación en la *migración*, el desarrollo de este tema va a estar estrechamente relacionado con los otros ya que la migración es básicamente una consecuencia de los cambios políticos, sociales y económicos que ha estado experimentado la sociedad cubana durante las últimas seis décadas.

## **6.2 La migración como tema central dentro de la literatura cubana post 1959**

Desde la misma década de los '60 comenzó el flujo migratorio desde Cuba, sobre todo, hacia los Estados Unidos y especialmente por razones políticas y económicas, razones que se han mantenido hasta la fecha (Duany, 2005). Ya que este fenómeno ha permeado la sociedad cubana por décadas, es lógico que se convirtiera en una de los temas centrales de la literatura cubana de este período.

Esta afirmación se pudo constatar en el corpus a través del análisis de palabras clave. Una vez confirmada la *migración* como uno de los ejes temáticos dentro de la literatura, el siguiente paso del análisis fue explorar cómo se manifiesta este tema dentro del corpus, más allá de las palabras clave, o sea, se analizaron los aspectos lingüísticos de los que se valían los autores para abordar este tema.

La *migración* es el “desplazamiento geográfico de individuos o grupos, generalmente por causas económicas o sociales” (ASALE & RAE, 2022). Desde una perspectiva netamente lingüística la migración es básicamente movimiento desde y hacia un centro deíctico locativo. Como esta investigación se enfoca en el tratamiento literario del fenómeno migratorio cubano, se decidió tomar el sustantivo locativo *Cuba* como este centro deíctico, en tanto fue

una de las palabras clave que aparecieron al cruzar el CLCXX+2 como corpus experimental con el CLCXX+1 como corpus de referencia. Asimismo, se decidió a partir de criterios metonímicos, o sea, utilizar otros sustantivos locativos referentes a Cuba como *país*, *isla*, *patria* o *nación*. Una vez que se determinó el punto de partida del análisis, el siguiente paso fue determinar las colocaciones más frecuentes con estos sustantivos locativos.

Collocate	Rank	FreqL	FreqR	Range	Likelihood	Effect
1 en	1	793	597	196	24	470.652 1.25
2 santiago	2	53	52	1	17	311.374 5.63
3 cuba	3	64	32	32	12	159.044 3.057
4 venegas	4	17	0	17	1	145.443 7.532
5 de	5	1077	757	320	26	140.481 0.536
6 libre	6	24	1	23	8	92.503 4.134
7 provincia	7	15	14	1	2	70.718 4.78
8 el	8	241	94	147	23	55.077 -0.632
9 españa	9	17	8	9	8	49.237 3.39
10 viva	10	13	12	1	7	46.453 3.919
11 salir	11	24	22	2	5	44.198 2.517
12 regresar	12	16	16	0	6	42.641 3.205
13 salido	13	14	13	1	8	42.149 3.481
14 cinemateca	14	7	7	0	3	40.154 5.532
15 regreso	15	15	14	1	5	39.616 3.186
16 shiram	16	4	3	1	1	36.654 7.947
17 escritores	17	9	9	0	3	33.955 4.072
18 volver	18	16	13	3	8	33.938 2.757
19 villas	19	9	9	0	1	33.784 4.058

Imagen 8. Colocaciones con el sustantivo locativo Cuba. Sección CLCXX+2

La Imagen 8 muestra justamente las colocaciones más frecuentes con *Cuba*. Este proceso de análisis se repitió no solo con este sustantivo sino también con otros que, como se ha mencionado anteriormente, funcionan como correferentes locativos de Cuba. Así, se trabajó solo con las primeras 50 colocaciones obtenidas para cada uno de estos sustantivos. Si bien los resultados obtenidos en esta primera etapa del análisis fueron interesantes, lo más llamativo fue la aparición de 5 verbos de movimiento, altamente específicos, los cuales se repitieron con los otros sustantivos, a saber, *ir(se)*, *quedar(se)*, *salir*, *volver* y *regresar*. Estos

cinco verbos aparecieron un total de 181 veces asociados a sustantivos locativos que hacían referencia a Cuba (isla, país, patria).

Es importante hacer notar que, independientemente de los otros usos que puedan tener dentro del corpus y que no son objeto de interés para esta investigación, la presencia de estos verbos comunica la idea de que sí existe un flujo humano que tiene como centro déictico locativo al país. Su frecuencia de uso, así como su dispersión en el corpus corroboran la idea de que se trata de un tema recurrente y común a todos los escritores incluidos en el corpus. Por ejemplo, la Imagen 9 muestra precisamente la dispersión del este rasgo lingüístico en el CLCXX+2. El hecho de que aparezca en 11 de las 26 novelas del corpus con este uso asociado al movimiento hacia afuera de la Isla es un indicio de la recurrencia de la migración dentro del corpus.

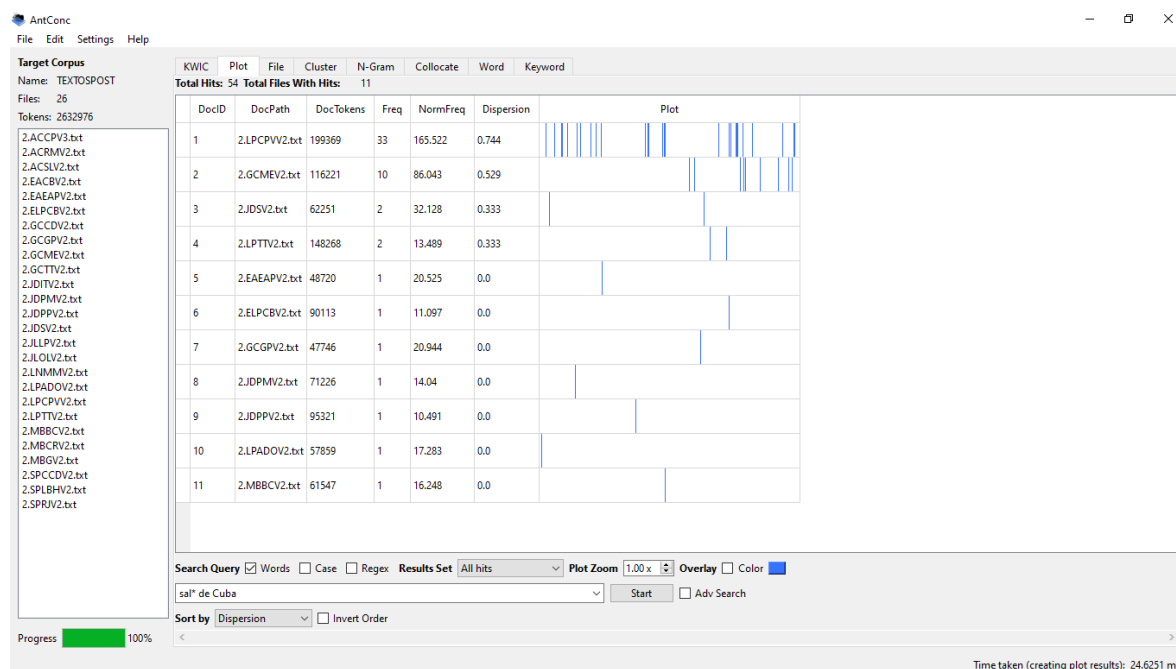


Imagen 9. Dispersión de las colocaciones sal\* de Cuba en la sección CLCXX+2

Asimismo, es interesante el tipo de movimiento que indican los verbos discretizados. Dos de estos indican movimiento hacia afuera de Cuba (*salir, ir(se)*), dos que indican movimiento hacia adentro (*volver, regresar*) y uno en el que de cierta manera está implícita la noción de no movimiento o permanencia en un lugar (*quedar(se)*).

Como se observa en la Tabla 4 el uso de estos verbos de movimiento no está balanceado en el corpus, o sea, es mucho más prominente el uso de los verbos que indican el movimiento hacia afuera de la isla. El hecho de que *salir e ir(se)* -61.1% de las coocurrencias- sean los verbos que más coocurren con sustantivos locativos relativos a Cuba sugiere, *a priori*, un escenario social en Cuba que está marcado por una tendencia migratoria, que como se verá el revisar los datos no existía antes de la revolución. O sea, la dirección del flujo migratoria, es mayormente, hacia afuera de la isla. Las coocurrencias con los otros verbos, especialmente con *volver y regresar*, son esperadas ya que es natural que exista un movimiento bidireccional, sin embargo, está desbalanceado su uso siendo el movimiento desde Cuba, hacia afuera de la isla, el más prominente. Si se tiene en cuenta que, como se ha venido planteando, la literatura es básicamente el reflejo de una sociedad dada en un momento dado, el hecho de que *volver/regresar* representen solo la mitad de las coocurrencias de *salir/ir(se)* da una idea de los intereses y motivaciones experimentadas por la sociedad cubana de la época. En esta primera etapa del análisis, a partir del uso de estos verbos de movimiento, ya se hace evidente que emigrar de Cuba es una motivación presente en el imaginario de la sociedad cubana.



Verbos	Total	%
salir	71	61.1
ir(se)	39	
regresar	42	32.8
volver	17	
quedar(se)	11	6.1
Total	180	100

Tabla 4. Verbos de movimiento

Este mismo procedimiento se realizó con la sección CLCXX+1. En la Imagen 10 se pueden observar las colocaciones más frecuentes con el sustantivo locativo *Cuba*. Así, se buscaron las colocaciones más frecuentes con sustantivo locativo *Cuba* (y con sus correferentes) en el corpus. En contraste con los resultados obtenidos para este procedimiento en la sección CLCXX+2, en este caso se constató que ninguna de las primeras colocaciones fue con verbos de movimiento<sup>24</sup>. Resulta interesante que el movimiento con Cuba como centro locativo no fuera relevante para estos autores ya que este hecho contribuye a confirmar que la migración no era de gran interés para la literatura de la época y, por tanto, tampoco era un fenómeno social relevante. Por supuesto sí existe un uso de estos verbos de movimiento dentro del corpus; pero este está asociado al movimiento dentro de la isla y no hacia o desde ella.

---

<sup>24</sup> La interfaz del programa AntConc sólo muestra las primeras 19 colocaciones; pero se analizaron las primeras 50 organizadas por su frecuencia de uso.

AntConc

File Edit Settings Help

Target Corpus

Name: TEXTOSPRES

Files: 19

Tokens: 1421083

1.ACEAV2.txt  
1.ACEV2.txt  
1.ACPV2.txt  
1.ACRMV2.txt  
1.CLGDV2.txt  
1.CMHSV2.txt  
1.DMLV2.txt  
1.EDDV3.txt  
1.LCCNV3.txt  
1.LCEMV2.txt  
1.LNENV2.txt  
1.LRGLV2.txt  
1.MCLHV2.txt  
1.MCLV2.txt  
1.OJCECV3.txt  
1.PTBASDV2.txt  
1.PTBCCV2.txt  
1.VPCFV2.txt  
1.VPCRV2.txt

Collocate Types 19 Collocate Tokens 715 Page Size 100 hits 1 to 19 of 19 hits

Collocate	Rank	FreqLR	FreqL	FreqR	Range	Likelihood	Effect
1 santiago	1	23	22	1	9	179.299	7.027
2 en	2	201	130	71	12	123.849	1.277
3 de	3	285	206	79	13	40.178	0.558
4 libre	4	8	0	8	2	35.365	4.565
5 cuba	5	10	5	5	3	31.003	3.553
6 rumbo	6	7	7	0	2	29.849	4.448
7 habana	7	9	4	5	5	27.091	3.482
8 platanillo	8	3	3	0	1	24.342	7.259
9 popular	9	4	3	1	3	23.671	5.674
10 viva	10	6	4	2	2	22.360	4.039
11 bejuco	10	6	4	2	1	22.360	4.039
12 esclavitud	12	4	3	1	1	20.821	5.15
13 decimos	13	3	3	0	2	20.165	6.259
14 literatura	14	4	2	2	1	18.379	4.696
15 campos	15	4	3	1	2	17.678	4.565
16 a	16	121	83	38	12	15.841	0.547
17 sabe	17	7	4	3	4	13.594	2.605
18 muchos	18	7	4	3	4	13.494	2.592
19 p	19	3	0	3	3	13.013	4.504

Search Query  Words  Case  Regex Window Span From 5L To 5R Min. Freq 3 Min. Range 1

Start  Adv Search

Sort by Likelihood  Invert Order

Progress 100%

Imagen 10. Colocaciones con el sustantivo locativo Cuba. Sección CLCXX+1

Así, al confrontar la Imagen 8 con la Imagen 10 se puede apreciar el contraste que existe entre las dos secciones del corpus. Esto es, la frecuencia del uso de verbos de movimiento con Cuba y sus correferentes en la sección CLCXX+2 y la poca frecuencia de este uso en la sección CLCXX+1. Así, el hecho de que no se registren colocaciones de estos verbos de movimiento con el locativo *Cuba* corrobora dos ideas: (i) para los escritores cubanos de la primera mitad del siglo XX la *migración* no era un tema relevante, a pesar de que para la época Cuba y particularmente La Habana eran un puerto importante a nivel mundial. El fenómeno migratorio sí estaba presente en Cuba; pero el flujo era, principalmente, hacia la isla y no desde ella.

Esta idea, que se hace evidente desde un punto de vista lingüístico es confirmada por Ajá Díaz (2002), quien plantea que “hasta el primer tercio del siglo pasado, Cuba mantuvo saldos migratorios externos positivos debido principalmente al aporte de la inmigración española y

antillana”. O sea, no se percibe en la literatura cubana de la época un “ímpetu” social por salir de la isla, sino más bien se ve la migración como un fenómeno normal marcado especialmente por el ingreso de personas al país, no por la salida de este.

(ii) Desde el punto de vista social, el fenómeno migratorio comenzó a ganar prominencia después de 1959, cuando comenzó a ser visto como un eje temático del quehacer literario en la isla. Ajá Díaz continúa diciendo que

El año 1959 marca la modificación de los componentes migratorios tradicionales de Cuba, al cobrar un papel central, tanto los elementos políticos y económicos motivados por la propia evolución del proceso revolucionario, como por la contradicción entre los Estados Unidos y Cuba, entre los cuales el tema migratorio ocupa particular espacio. (Díaz, 2002, p.4)

Teniendo en cuenta los datos revisados hasta este punto se puede afirmar que:

1. La *migración* no era un tema de relevancia para la literatura cubana antes de 1959, por lo que se puede especular que socialmente no era percibido como un fenómeno impactante. Además, se percibía a la isla como un país económicamente próspero por lo que los movimientos migratorios eran mayormente hacia la isla y no desde ella. Para las décadas posteriores a 1959 sí derivó en un tema de gran impacto social, al punto de convertirse en uno de los ejes temáticos de la literatura de la época.
2. Directa o indirectamente el triunfo de la revolución cubana sí tuvo un impacto tanto en la vida social del país como en el quehacer literario. Las dimensiones de este impacto quedan, sin embargo, para otro tipo de estudio.

En la siguiente sección se analizará el funcionamiento de estas colocaciones dentro del texto. O sea, se analizarán las líneas de concordancia donde estas ocurren, lo cual permite identificar el contexto en el cual están funcionando y cómo los autores se valen de estos

recursos lingüísticos para transmitir los estados de ánimo de sus protagonistas y contextualizar las acciones.

### **6.2.1 Líneas de concordancia. Análisis de sentimientos**

Como se comentó en la sección 2.1.2, las líneas de concordancia que se van a analizar son aquellas en las que ocurren las palabras clave. Estas líneas de concordancia son las que permiten precisar el contexto y determinar la relevancia real de estas palabras en el texto.

Aunque los datos revisados hasta este punto son claros y contribuyen a arrojar alguna luz sobre el impacto del fenómeno migratorio en la sociedad cubana de la época revolucionaria, no son suficientes para confirmar las hipótesis que guían esta investigación. Para comprender más a fondo el reflejo del fenómeno migratorio en la literatura cubana post '59 se analizaron “artesanalmente” las 181 líneas de concordancia en las que coocurrían los cinco verbos de movimiento antes mencionados con sustantivos locativos referentes a Cuba. El análisis de estas líneas de concordancia permitió conformar un panorama relativo a los estados de ánimos, sentimientos y pensamientos reflejados por los escritores cubanos de la época, lo cual puede interpretarse básicamente como un reflejo de la cosmovisión social.

A partir del análisis de estas líneas de concordancia se discretizaron 20 conceptos/nociones<sup>25</sup> presentes dentro de las diferentes obras. Estos conceptos/nociones se analizaron ligados a los verbos con los que coocurrían. Los más recurrentes dentro del corpus fueron:

1. Ansiedad- 55 coocurrencias
2. Neutro- 26 coocurrencias
3. Separación- 16 coocurrencias

---

<sup>25</sup> Se decidió utilizar concepto/noción pues no se pueden englobar todos como sentimientos, ya que se pudieron determinar conceptos como “separación” la cual no es un sentimiento per se, sino más bien se interpretó como una circunstancia a partir de la cual se pueden generar sentimientos.

4. Resignación 2<sup>26</sup>- 13 coocurrencias

5. Miedo- 13 coocurrencias

Dentro de estas nociones más frecuentes se debe hacer notar que tres de ellas son altamente negativas, a saber, *ansiedad*, *resignación* y *miedo*:

- Ansiedad (C): Deseo intenso acompañado de vehemencia, **miedo** o apresuramiento. anhelo, ansia, ansiedad, avidez, codicia, mono (síndrome de abstinencia), voracidad. Antónimos: saciedad, desgana. (Marina & Penas, 2000)
- Resignación (C): La pérdida del objeto de nuestros deseos o proyectos, o cualquier experiencia dolorosa que aceptamos negándonos a luchar para evitarla, provoca un sentimiento negativo, frecuentemente acompañado de calma y desesperanza. Antónimo: rebeldía. (Marina & Penas, 2000)
- Miedo (C): La percepción de un peligro o la anticipación de un mal posible provoca un sentimiento desagradable, acompañado de deseos de huida. aprensión, canguelo, hipocondría, miedo, pánico, pavor, temor, terror. Antónimos: esperanza, confianza, impavidez. (Marina & Penas, 2000)

Estos conceptos de Marina y Penas (2000) descubren la existencia de un tono, hasta cierto punto sombrío y desesperanzador al tratar el tema de la *migración* en la literatura cubana. En cuanto a la ansiedad, los siguientes ejemplos contribuyen a ilustrar como se ve reflejada en los textos analizados:

---

<sup>26</sup>Se codificaron tres tipos de resignación a partir de los contextos en los que se encontraron, a saber:  
Resignación 1: tener que regresar a Cuba  
Resignación 2: no poder regresar a Cuba  
Resignación 3: no poder salir (tener que quedarse en) de Cuba.

1. Y mi padre es cubano, como tú... **Él salió de Cuba hace treinta años... ¡en un bote!  
Es un balsero, como tú.** - 2.LPCPVV2.txt<sup>27</sup>-
2. Y menos comprendía por qué le había dado a Marcos **por salir de Cuba y hasta lo  
había intentado en la siempre peligrosa travesía por el estrecho de la Florida (...)-  
2.LPCPVV2.txt-**
3. Hace poco la pobre Marilú, **más soplada que una cafetera, salió de Cuba por fin.-  
2.ELPCBV2.txt-**
4. (...) se caían de los árboles y hasta por qué volaban los aviones que veían desde  
la casa de Fontanar, **había salido de Cuba con otras decenas de miles de balseros y  
desde hacía mucho tiempo vivía en Puerto Rico, casado (...)- 2.LPCPVV2.txt-**
5. (...) ella a una simple L desde su niñez—. **Elisa le confió que había debido  
ocultar su nombre real para poder salir de Cuba con el pasaporte a nombre de  
Loreta, estampado con un visado inglés** que, años atrás, le había (...)-  
2.LPCPVV2.txt-
6. (...) la promesa. **Aquella era mi primera actuación en cine y el apoyo de Ofelia  
me resultaba imprescindible para triunfar y salir de Cuba acompañando la película.  
Si lo lograba no regresaría a ningún precio; pediría asilo en cualquier país (...)-  
2.JDPMV2.txt-**
7. (...) se iba sin la visa belga, que resolvería su entrada en Bélgica desde España,  
**pero eso demostraría un ansia por salir de Cuba que quizá revelara sus planes.** Así,  
no dijo más que «Deja, yo lo voy a resolver con (...)-2.GCMEV2.txt-

---

<sup>27</sup> Así se codificaron los nombres de los archivos (textos) analizados. El primer número (2 en este caso) corresponde a la sección del corpus a la que pertenece la obra. Las dos primeras letras corresponden al autor, ej. LP-Leonardo Padura; las siguientes tres letras corresponden al título de la obra, ej. CPV-Como Polvo en el Viento; la V2 corresponde a la versión que se analizó, en este caso la versión 2, ya corregida con el software Notepad++.

Antes de intentar analizar estos ejemplos se debe precisar que ningún análisis y mucho menos un análisis literario está desprovisto de los sesgos propios del analista. Estos sesgos están condicionados por factores extralingüísticos entre los que se pueden contar desde posturas políticas hasta el nivel socioeconómico. Sin embargo, en el caso de esta investigación el conocimiento del contexto cubano probó ser una herramienta útil y no un obstáculo. Habiendo dicho esto también es pertinente mencionar que todo el proceso analítico estuvo basado en la información recuperable de las líneas de concordancia.

El caso de los ejemplos 1 y 2 es relativamente transparente en tanto existen pistas lingüísticas como la expresión *la siempre **peligrosa** travesía* – en este caso no existe manera de rebatir la existencia de un peligro, en tanto es el propio personaje quien lo afirma- y pistas extralingüísticas *en un bote, el estrecho de la Florida, balseo* que permiten entender la idea de peligro asociada a salir de Cuba.

Cuba es una isla, *ergo*, está rodeada de mar; existen solo dos formas de salir del país en avión o en barco. Hasta 2013 las salidas por avión de Cuba siempre estuvieron controladas por el gobierno y supeditadas a una serie de requisitos a los que la mayoría de la población no tenía acceso. Por tanto, las salidas por este medio estaban reservadas a personas de altos niveles de ingresos económicos o con familiares fuera de la isla.

Por otra parte, las salidas marítimas siempre fueron más “accesibles” aunque también ilegales. Más allá del peligro penal que implica incurrir en una ilegalidad, se debe tener en cuenta el peligro físico que implica lanzarse a mar abierto sin las condiciones necesarias para hacerlo. La palabra *balseo* da un indicio de la precariedad de las embarcaciones y las ansias de abandonar el país por cualquier medio necesario. *Balseo* es un sustantivo derivado de *balsa*, esto es, una embarcación pequeña y rústica, más apropiada para la navegación en ríos

o lagos que en mar abierto. Estas ideas son un indicio de ese *deseo intenso, acompañado de vehemencia o miedo* (Marina & Penas, 2000) por abandonar la isla.

El tercer ejemplo es tal vez menos transparente en el sentido de que apela a un coloquialismo cubano: *más soplado que una cafetera*. En el argot cubano esta expresión hace referencia al apuro o la urgencia con la que una persona abandona un lugar. En los tres casos, sin embargo, es palpable la necesidad o urgencia de salir del país por cualquier medio necesario.

El ejemplo 4 es similar a los ejemplos 1 y 2 en tanto se retoma la idea del *balseo* asociada a la necesidad de salir de Cuba por cualquier medio necesario. Igualmente el ejemplo 5 refleja esta necesidad al describirse una situación en la que la protagonista apela a la falsificación de documentos para abandonar la isla con el consiguiente riesgo legal que esto implica tanto en Cuba como en otro país. Justamente de esta idea, o sea de no hacer caso de las consecuencias, se deriva la interpretación de la *ansiedad* o necesidad imperativa de salir de Cuba por cualquier medio necesario. Es poco probable que una persona, a no ser que se encuentre al borde de la desesperación llegue, a tales límites por un objetivo.

El ejemplo 6 presenta una situación similar en cuanto al sentir ya que el protagonista también recurre a medios no convencionales para salir del país, en este caso como funcionario estatal acompañando una película<sup>28</sup>. El hecho de que esté en disposición de solicitar asilo político en el país de destino es una pista de su resolución de salir de Cuba y no regresar -ambas intenciones están explícitas en la línea en cuestión-. Por su parte el ejemplo 7 no supone un

---

<sup>28</sup> Dentro de la literatura y el cine existen varios ejemplos de las llamadas “deserciones” de funcionarios estatales de países socialistas al bando contrario. Hechos que se reflejan a través del arte; pero que tienen su origen en la vida real. Obras como “1984”, “La insoportable levedad del ser” o “La vida de los otros” son un reflejo preciso de la vida en los países del extinto bloque socialista. Aún si el primero, 1984, es en realidad una distopía.



gran reto al analista en tanto la palabra *ansia* contenida dentro de la línea de concordancia permite determinar fácilmente el sentimiento que se está reflejando.

Analizar la *resignación* es, tal vez, un poco más subjetivo que analizar la *ansiedad* en tanto requiere la capacidad del investigador no solo de encontrar las pistas lingüísticas que están en las líneas de concordancia, sino también de empatizar con lo que pudiera estar sintiendo el personaje. Es importante tener en cuenta, de nuevo, que la migración es un proceso condicionado por factores sociales, políticos y económicos; en el caso específico de la migración cubana, por factores políticos y económicos, así los emigrantes salen o tratan de salir de la isla movidos más por una postura contraria al gobierno o por necesidades económicas que por el deseo de desarraigarse de su tierra. En los ejemplos siguientes se evidencian posturas que si bien se decidieron catalogar como la resignación de no poder volver más a Cuba -entendiendo resignación desde la perspectiva de Marina y Penas (2000)- se pueden percibir también sentimientos como el enojo y el resentimiento. Como se ha comentado en este trabajo se identificaron tres tipos de resignación en los textos. Los ejemplos reflejan la resignación 2, esto es, la resignación a no poder volver más a Cuba.

8. (...) rumbo a la casa de su madre, **me dijo que no pensaba volver más a Cuba, que ahora escribiría en inglés y que cada día sentía menos propio su país** (...) -

2.GCCDV2.txt-

9. (...) una cabeza contemporánea de Unamuno, sino porque sentía que los engañaba a todos: **ya él sabía que de irse no regresaría jamás a Cuba.** -

2.GCMEV2.txt-

10. Estaba decidido a seguir esperando y **no volver a su patria mientras la vida no recuperara lo que para él debía ser su curso normal** (...) - 2.LPCPVV2.txt-

11.¿Los eternos retornos? ¿Los ciclos inquebrantables? Cuando Fabio y Liuba salieron de Cuba con la secreta intención de no regresar y el empeño de sacar en algún momento (...) -2.LPCPVV2.txt-

12. (...) pero no lo dijo: ni siquiera a los amigos más íntimos les decía sus intenciones de no regresar más a Cuba. Tal vez íntimamente él no admitía todavía esta posibilidad. - 2.GCMEV2.txt-

Los ejemplos anteriores tienen en común que en todos ellos los personajes han resuelto abandonar su país para no regresar. Una resolución de este tipo, motivada por las razones que sean implica siempre un cierto grado de resignación, que puede ser más o menos intenso, dependiendo de la persona. Desde la perspectiva de Marina y Penas (2000) se acepta la “experiencia dolorosa” de no volver más al país de origen.

En el ejemplo 8, ideas como *no pensar en volver más a Cuba, escribir en inglés y sentir cada día menos su propio país* son indicadores de un proceso de aceptación por parte del personaje de su nueva realidad fuera de su país, una nueva realidad que parece ser permanente y definitiva. O sea, el personaje acepta el hecho de que no regresará más a Cuba.

Asimismo, en los ejemplos 9 y 10 se puede percibir esta sensación de resignación. En el ejemplo 9 el uso del adverbio temporal *jamás* es una pista contundente de la resolución tomada por el personaje de no regresar a Cuba.

En el ejemplo 10 ocurre algo similar en tanto el regreso a Cuba está supeditado a la condición de que la *vida retome su curso normal*, lo cual implica un alto nivel de incertidumbre y subjetividad, pues ¿qué es lo que considera normal este personaje?, ¿qué circunstancias son condición *sine qua non* para su regreso? Cualquiera que sea la respuesta a estas preguntas, el personaje se ha resignado a no volver si estas condiciones no se cumplen.

Los ejemplos 11 y 12 son similares e interesantes ya que en ambos se presenta una cierta dosis de engaño y secretismo. Y en estos ejemplos se pueden identificar más de un sentimiento, por una parte la ansiedad que puede generar tener que ocultar un secreto, incluso de las personas más queridas o cercanas. Además, el miedo a que se frustren los proyectos, el cual es la causa de este secreto. Y en ambos casos la intención de no regresar más a Cuba.

El análisis relativo al miedo fue bastante explícito y evidente ya que las pistas lingüísticas fueron muy claras, como se observa en los siguientes ejemplos:

13. **Beto había decidido no volver a Cuba por temor a represalias. Tenía miedo.**

Estaba convencido de que lo considerarían un traidor. - 2.EACBV2.txt-

14. **no entiendo por qué no te quedaste en Italia, como otros que fueron al viaje, y te arriesgaste a volver a Cuba. ¿Y si te estaban esperando?... ¿No podías quedarte y después, no sé, irte a España con tu (...)** -2.LPCPVV2.txt-

15. **Irving había tenido la misma pesadilla de todos los cubanos lanzados al exilio: alguna vez regresaba a la isla y.... no lo dejaban volver a salir.** - 2.LPCPVV2.txt-

16. **(...) Walter se sentía acosado (no dijo por quién) y por ello buscaba una vía para salir de Cuba,** constituyó la mayor novedad para los que aún no sabían de tal pretensión (...)-2.LPCPVV2.txt-

17. **Lo siento por ti, pero, como te imaginarás, cuando tenga un pasaporte para viajar, no puedo arriesgarme a volver Cuba.** Así es la cosa, china: no hay regreso para Johnny. Estoy condenado al exilio eterno. -2.LPCPVV2.txt

18. **volvió a repetirse el sueño recurrente que tenía, en el que estaba en Cuba y, al tratar de salir estando ya en el aeropuerto, comprobaba que no tenía pasaporte (...)**

El ejemplo 13 es bastante claro en cuanto al sentimiento que refleja. Sustantivos como *temor*, *represalias*, *miedo* tienen una connotación negativa y no dejan lugar a otra interpretación que no sea que el personaje tiene miedo a regresar a Cuba.

En el caso 14 también la presencia del verbo *arriesgarse* es una pista bastante clara en tanto un *riesgo* representa un daño o amenaza para la persona que se *arriesga* y la respuesta lógica ante una amenaza es la de temerla y por tanto protegerse de ella. Lo mismo ocurre en el ejemplo 17.

Igualmente, en el ejemplo 15 la *pesadilla* es una experiencia negativa que genera miedo en las personas que la experimentan. En este caso específico es, de nuevo, ese temor a no poder salir de la isla. En el ejemplo 18 se observa una idea similar, en tanto no se utiliza el sustantivo *pesadilla* sino *sueño*. Sin embargo, es interesante la recurrencia de estos sueños en los personajes y cómo este temor está arraigado en su subconsciente.

Finalmente, en el ejemplo 16, el *acoso* es una acción que puede llegar a generar miedo en la persona que está siendo acosada. En este caso, la persona se siente llevada al límite de tener que abandonar su país para poder poner fin a este acoso. Otro caso interesante es el de la neutralidad o la ausencia de sentimientos, como se verá en los próximos ejemplos. En estos casos se podría especular que el fenómeno migratorio está tan normalizado en la cosmovisión cubana que se utiliza simplemente como un punto de referencia, un trasfondo para la historia. O sea, se utiliza un evento migratorio puntual para contextualizar- en la mayoría de los casos temporalmente- la acción descrita, sea por el narrador o por los personajes, en el texto.

19.(...) que en verdad resultaron orgánicas e intentos más o menos exitosos de ingenio verbal criollo, el joven recién salido de Cuba era, en realidad, una persona que con su fusión de inocencia cósmica (...) -2.LPCPVV2.txt-

20. Inclinado sobre la agenda, el Rubito anotó los datos con una letra clara, de escribano. —El mes que viene voy a Cuba de vacaciones —dijo mientras devolvía pluma y agenda al bolsillo interior del saco—. -2.JDPPV2.txt-

21. Pronto dejaría de hacer siquiera esto. Finalmente, Arcos regresó a Cuba y la mujer de Pipo Carbonell regresó con él, quedándose Pipo en la embajada por un tiempo (...) -2.GCMEV2.txt-

22. las caras que pondrían algunos antiguos clientes míos, asiduos del famoso “Templo del Daiquirí”, al darse cuenta de que había regresado a la patria.- 2.ACCPV3.txt-

23. También le había jurado a Nadiezdha la noche anterior en el buffet del Cosmos que regresaría a Cuba en la fecha programada, para la que apenas faltaba una semana.- 2.JDSV2.txt-

A diferencia de los ejemplos del 1-18 en estos casos no se puede determinar un sentimiento específico a partir del contexto. Los autores utilizan los eventos migratorios para contextualizar temporalmente su historia, sin intención de transmitir ningún estado de ánimo. De nuevo, esto podría sugerir una normalización del fenómeno migratorio en Cuba; esto es, a diferencia de lo que ocurría antes de 1959, después del triunfo de la revolución el flujo migratorio del país comenzó poco a poco a aceptarse como algo normal e incluso comprensible, parte de la cotidianeidad en la isla.

Hasta aquí se ha tratado de ilustrar cómo se realizó el análisis de sentimientos transmitidos en las novelas a partir de las líneas de concordancia donde aparecen los verbos de movimiento más frecuentes asociados a sustantivos locativos referentes a Cuba. Si bien se pudiera argumentar en contra de la alta subjetividad relativa a este análisis, lo cierto es que ningún análisis está desprovisto de determinado grado de subjetividad y si bien el contexto

extralingüístico es considerado aquí como importantísimo, se trabajó basándose en las pistas lingüísticas provistas por las líneas de concordancia.

Los ejemplos revisados ilustran las nociones o sentimientos más frecuentes dentro del corpus.

La Tabla 5 muestra el total de estos y las coocurrencias verbales.

Sentimientos	ir*	qued*	regres*	sal*	volv*	Total general
Alivio	2					2
Ansiedad	15	1	8	27	4	55
conformismo	1					1
curiosidad	2					2
desacuerdo	1		5	3		9
desengaño					1	1
esperanza					1	1
impotencia				4		4
incredulidad	5					5
miedo		3	4	2	4	13
neutro	4	1	10	14		29
nostalgia	3		2	1	1	7
obligación	1				1	2
religiosidad				3		3
resentimiento				2		2
resignación 1			7			7
resignación 2	3		5	1	4	13

Sentimientos	ir*	qued*	regres*	sal*	volv*	Total general
resignación 3		5		4		9
separación	3	1	1	10	1	16
<b>Total general</b>	<b>40</b>	<b>11</b>	<b>42</b>	<b>71</b>	<b>17</b>	<b>181</b>

Tabla 5. Total de sentimientos y coocurrencias verbales

Una vez que se determinaron todas las nociones/sentimiento a partir de las líneas de concordancia, se decidió hacer otro tipo de análisis, esto es, un análisis de agrupación para determinar, desde un punto de vista cuantitativo, si, en efecto, existían diferencias significativas al abordar estas nociones entre ambas secciones del corpus.

### 6.2.2 Análisis de agrupación. Resultados

Como se ha comentado en la sección 5.5 a partir de los sentimientos discretizados artesanalmente en el primer análisis, se decidió realizar un segundo análisis, desde una perspectiva cuantitativa, con la intención de determinar si en efecto los rasgos con los que se trabajó se tratan de manera diferente entre ambas secciones del corpus. A partir de la elaboración de campos semánticos y tomando como base los rasgos originales se elaboraron otros rasgos hasta llegar a 41. La Tabla 3 muestra algunos de los rasgos que se elaboraron y la frecuencia relativa en 9 de las novelas de la sección CLCXX+1 del corpus. En el Apéndice 3. Frecuencias relativas de los rasgos utilizados para el análisis de agrupación se pueden observar la totalidad de los rasgos y sus frecuencias relativas en cada una de las novelas. Estas frecuencias relativas fueron calculadas para luego introducir los datos al software Brezina ToolBox, el cual realiza un cálculo que permite agrupar las novelas de acuerdo a los rasgos discretizados.





más; después se agrupa ese par, con un tercer texto, que es el que tiene mayor parecido con ese par, posteriormente se calculan las similitudes de ese trío con un cuarto texto, y así sucesivamente. Los textos se van agrupando, subsumiendo en ramas de un árbol que pueden leerse según los elementos que las componen. Por ejemplo, la Imagen 11 tiene unos recuadros rojos que señalan las tres ramas principales en las que se agruparon los textos. De acuerdo a los rasgos analizados se puede establecer una clara diferencia entre los textos de ambas secciones del corpus, en tanto casi todos los textos del CLXX+1 se encuentran agrupados a la izquierda del dendrograma y los de la sección CLCXX+2 a la derecha. La única excepción la constituyen las novelas de Alejo Carpentier, en las cuales no se realizó una agrupación temporal, y esta es una excepción lógica, ya que es el único autor del que se utilizaron obras de ambos períodos y el hecho de que aparezcan juntas sugiere que su estilo no varió temporalmente ni estuvo influenciado por cuestiones “administrativas”.

Asimismo, se observa un tercer cuadro a la extrema izquierda en el que aparecen las obras de Eliseo Diego y Onelio Jorge Cardoso. Este también es un elemento interesante: Eliseo Diego y Onelio Jorge Cardoso son cuentistas, no novelistas; sin embargo, el software es capaz de realizar un análisis lo suficientemente fino para agruparlos a ellos en un grupo diferente al de los novelistas.

Los rasgos utilizados probaron ser fundamentales para establecer diferencias entre ambas secciones del corpus. Esta clara división que existe entre ambas secciones, permite plantear que, en efecto, los sentimientos, la movilidad y demás cuestiones asociadas al tema de la migración se trataron de manera diferente en ambos casos. Así se corrobora lo que se ha venido planteando a lo largo de la investigación, en el sentido de que el enfoque que se le dio a las temáticas no fue el mismo y desde un punto de vista temporal su relevancia varió. Sin

embargo, este software solo permite establecer eso: que su peso temático o su tratamiento entre ambas partes del corpus fue diferente; pero no es lo suficientemente refinado para determinar cómo fue este tratamiento o desde qué posiciones se abordó el fenómeno.

Como se ha comentado, se observan, además, otras dos cuestiones interesantes. Como se ha notado, las obras de Alejo Carpentier se agrupan juntas, en el recuadro que contiene a las obras de la sección CLCXX+2 lo cual es indicio de que no se puede establecer una separación temática ni temporal, a partir de los rasgos utilizados, entre las obras que se escribieron antes de 1959 y aquellas posteriores a esta fecha. Esto puede responder a varias razones. En primer lugar, de todos los escritores recopilados en el corpus, Carpentier es el único que contribuye a ambas secciones del mismo. Así se podría afirmar que en su obra, al menos cuando se trata de estas cuestiones migratorias, existe un tratamiento consecuente del tema, independientemente de las cuestiones políticas a las que pudo haber estado supeditada la literatura de los años posteriores a 1959. Asimismo, hace suponer la existencia de un estilo muy marcado dentro de la obra de Carpentier que trasciende estos límites temporales, hasta cierto punto artificiales, determinados por eventos sociopolíticos.

También se ha comentado la existencia de dos obras al margen de las demás. Estas obras pertenecen a Eliseo Diego y Onelio Jorge Cardoso, ambos cuentistas, de los más destacados en Cuba. Los rasgos analizados fueron los mismos para todos los textos, así que este hecho sugiere que el tratamiento estilístico que se le da a los mismos depende del género de escritura que se esté analizando. Además, la existencia de este subgrupo al margen de la agrupación total puede ser indicio de otras dos cuestiones (i) el fenómeno migratorio cubano no era de interés para estos cuentistas o (ii) existe una diferencia estilística entre el cuento y la novela,

y por tanto una manera distinta de abordar determinados temas (migración) que puede estar supeditada al género literario desde el cual se trate el tema.

En definitiva, estas cuestiones estilísticas y temáticas (relativas a Carpentier y al género literario) parecen ser material interesante para futuras investigaciones.

## 7 Discusión y conclusiones

La literatura, además de expresión artística, es un reflejo de las ambiciones, frustraciones, intereses y deseos de una sociedad determinada en un período determinado. Es un registro escrito que permite conocer una sociedad y una época. Al menos esta ha sido la premisa que ha guiado esta investigación. En este caso se trató de explorar el impacto de un movimiento político, social y económico tan relevante como la Revolución Cubana en uno de los aspectos más delicados de una nación, esto es, su cultura y particularmente su producción literaria.

Como se ha comentado, hasta donde se tiene conocimiento, las investigaciones que existen y que se han consultado sobre literatura cubana son por lo general desde enfoques pedagógicos, sociológicos o puramente literarios, y por lo general se enfocan en períodos o temas muy específicos. Además, lo hacen desde una perspectiva diferente a la de *lectura distante* propuesta por Franco Moretti (2013)<sup>29</sup>. Esta investigación es una investigación pionera en tanto, hasta donde se ha investigado, es la primera vez que se emplean técnicas propias de la L.C para aproximarse al estudio de la literatura cubana. Asimismo es una investigación multidisciplinaria que apela a enfoques lingüísticos para su desarrollo; pero que a la misma vez busca vincular aspectos de estudios literarios, sociales y psicológicos a la misma ya que tiene en cuenta el impacto social de un fenómeno político y la manera en la que, psicológicamente, puede influir en un fenómeno tan sensible en el contexto cubano como lo es la migración.

---

<sup>29</sup> Es importante resaltar la relevancia de la *lectura distante*, en tanto este enfoque significa un cambio en la manera de abordar el estudio de la literatura pues permite el análisis de un gran número de textos en un período de tiempo relativamente corto. De no haberse adoptado este enfoque propuesto por Moretti hubiera sido imposible concluir esta investigación.

Desarrollar una investigación totalmente libre de los sesgos del investigador es virtualmente imposible, sin embargo, el enfoque desde la lingüística de corpus permitió acercarse al estudio de la literatura cubana de la manera más neutral posible, en tanto esta investigación ha sido enteramente guiada por los datos arrojados en cada una de las etapas de la misma.

Desde el mismo diseño del corpus se trató de balancear los periodos a estudiar y si bien es cierto que esto no se logró enteramente debido a las limitaciones impuestas por la disponibilidad de material, especialmente el relativo a la sección CLCXX+1, sí se logró una muestra representativa que permitiera la obtención de datos suficientes para el estudio.

Las palabras claves determinadas a partir de las comparaciones realizadas entre las diferentes secciones que conforman el corpus son, por supuesto, el pilar de la investigación en tanto permitieron establecer los temas característicos de cada etapa y a partir de estos identificar los intereses y preocupaciones de la sociedad cubana en cada una de las etapas. Es pertinente decir que este hecho tiene implicaciones sociológicas y pone de manifiesto la trascendencia de la L.C en otros campos además de lingüístico.

Asimismo, el funcionamiento de estas palabras claves en sus respectivas líneas de concordancia posibilitó comenzar a identificar elementos psicológicos de la sociedad cubana, expresados a través de los sentimientos que comunican los personajes de las novelas.

Así, a partir del análisis y resultados anteriores se pueden corroborar las hipótesis postuladas para esta investigación, a saber: (i) El triunfo de la Revolución Cubana y el consecuente planteamiento de una nueva política cultural, alineada con los intereses de la nueva clase dominante significaron un cambio total en las temáticas tratadas por la literatura cubana hasta 1959. (ii) Una de los temas centrales dentro de la literatura cubana post 1959 será el tema de

la migración. Asimismo, estos resultados revelan algunas cuestiones que no se contemplaron en la hipótesis, pero que vale la pena comentar.

En el caso de la primera hipótesis parece ser evidente que el triunfo de la Revolución Cubana sí tuvo un impacto en el quehacer cultural cubano y específicamente en la literatura. A partir del análisis de palabras clave se pudo constatar que los temas trabajados por los escritores cubanos variaron de un período a otro.

En el caso del primer período, pre 1959, los temas literarios estaban en consonancia con los del canon literario latinoamericano de la época. La novela costumbrista a principios de siglo y posteriormente la novela de la tierra caracterizaron este período, siempre adecuándose a la realidad cubana y a su geografía, de ahí que a partir de las palabras claves los principales temas discretizados fueran aquellos relativos a la tierra, la navegación y la esclavitud<sup>30</sup>.

Un tema interesante es el de género, aún si no parece realista pensar que en épocas tan tempranas como la primera mitad del XX la literatura pudiera ocuparse de cuestiones de género desde las perspectivas actuales, tal vez valdría la pena analizar de qué manera o desde que perspectiva se trató este tema por los escritores cubanos de la época. Lo que sí es un hecho es que de alguna manera fue una cuestión de interés narrativo, como lo demuestran las palabras claves extraídas del corpus. Por otra parte, la migración no fue un tema relevante para la literatura pre revolucionaria lo cual es indicio de que tampoco era un tema de relevante desde el punto de vista social.

Sin embargo, al analizar las palabras claves relativas a los temas tratados por la literatura post revolucionaria, se puede constatar que sí hubo un cambio en las temáticas. Por razones

---

<sup>30</sup> La esclavitud se abolió oficialmente en Cuba en 1886, así que nos es de extrañar que fuera este un tema literario recurrente durante las primeras décadas del XX.

obvias, si la literatura retrata una sociedad determinada en una época determinada, es lógico pensar que durante los primeros años a partir del triunfo las temáticas fueron relativas a la revolución y al cambio social. Y pudiera pensarse que este primer acercamiento a esta temática fue espontáneo; sin embargo, las “*Palabras a los intelectuales*” de Fidel Castro en 1961, página 11, son indicio suficiente de que el nuevo gobierno no dejaría mucho espacio a la espontaneidad. Así es igualmente predecible que a partir de las palabras claves se pudieran determinar cómo temas centrales aquellos relativos a la revolución y la nacionalidad.

Hasta aquí se puede corroborar la primera hipótesis que guía esta investigación, ya que sí cambiaron las temáticas literarias a partir de los primeros años de la revolución, lo cual es constatable a partir del análisis de las palabras claves.

La segunda hipótesis también se confirma ya que sí apareció la *migración* como nueva temática a partir del triunfo revolucionario. Esta es, tal vez, la parte más relevante de esta investigación. Con técnicas propias de la L.C. se pudo realizar esta distinción entre los temas respectivos a cada período. Además, se pudo constatar el impacto de un evento sociopolítico en la cultura y en la vida social de un pueblo. Así, la *migración* solo ganó relevancia temática en la época post '59, lo cual sugiere dos cosas (i) la *migración* estuvo y está condicionada por una variable político-económica ya que esta fue la única que cambió entre ambos períodos<sup>31</sup>, (ii) la *migración* se convirtió en una cuestión lo suficientemente relevante desde el punto de vista social como para ser tratada en la literatura.

---

<sup>31</sup> Esta no es una tesis dedicada a temas socio-políticos por tanto no se profundizará más de lo estrictamente necesario en estas cuestiones. Está claro que esto es solo una generalización y que los cambios, de toda índole derivados de este evento son mucho más profundos y complejos.

Como se ha comentado anteriormente en este trabajo, una vez que se determinó la *migración* como uno de los temas centrales de la literatura post '59 en Cuba, se analizaron las líneas de concordancia donde se refleja este tema y los sentimientos que se producen a partir de esta.

Es notable el hecho de que ninguno de los sentimientos que se discretizaron tienen una connotación positiva *per se*. Un ejemplo interesante es el de la *ansiedad*. La mayoría de las ocurrencias de este sentimiento (42 de un total de 55) se registraron en líneas de concordancia donde aparecía formas de los verbos *irse* (15 coocurrencias) y *salir* (27 coocurrencias). Así, es lógico pensar que, para los personajes<sup>32</sup> el hecho de salir de Cuba se ha convertido en su objetivo principal, un objetivo alrededor del cual girará su vida hasta que sea cumplido. De ahí la ansiedad que les produce lograrlo y, retomando la definición de *ansiedad* propuesta por Marina y Penas (2000), el impacto negativo que tiene esta necesidad en sus vidas.

Asimismo, el miedo asociado a *volver/regresar* a Cuba (8 ocurrencias de 13) es un claro indicio de la hostilidad que representa para estos personajes volver a su propio país. Si bien el miedo solo se constató “explícitamente” en 13 líneas de concordancia, parece ser un sentimiento general distribuido por todo el corpus. El miedo es un componente semántico dentro de la definición de *ansiedad* que se ha venido trabajando en esta investigación. De la misma manera está presente en la definición de *resignación* (Marina & Penas, 2000) en tanto la aceptación de una experiencia dolorosa y la negación de una lucha para evitarla implica el miedo a las consecuencias que esta lucha podría tener. O sea, parece ser que las obras se desarrollan en un contexto en el cual el miedo es una constante; aún si no está presente de

---

<sup>32</sup> Es plausible entender a estos personajes como el vehículo del que se valen los autores para comunicar a sus lectores sus posturas acerca de la sociedad en la que se desarrollan las acciones de sus obras.



una manera explícita, sí parece ser parte, de una manera más o menos, consciente del imaginario popular cubano.

De la misma manera es interesante la presentación de la *resignación* desde perspectivas distintas, una resignación a no poder regresar a Cuba y la otra a tener que quedarse en la Isla. En ambos casos está presente, sin embargo, el miedo. Por una parte, el miedo al desarraigo y a la alienación del país propio y por otra el miedo a no cumplir con este objetivo de salir de Cuba.

Si bien a lo largo de la investigación se han tratado, por conveniencia analítica, separadamente los sentimientos encontrados, es una realidad que el ser humano no es capaz de aislar estos sentimientos a medida que los experimenta. O sea, el *miedo, la ansiedad, la resignación, la impotencia, el desengaño*, incluso *la esperanza* (solo 1 ocurrencia) parecen ser parte de la cosmovisión cubana del período post revolucionario, derivadas del fenómeno migratorio.

De nuevo, se debe acotar que esta no es una investigación sociopolítica; pero se cree pertinente señalar que sentimientos de este tipo parecen estar condicionado por temas socioeconómicos, los cuales están directamente supeditados al funcionamiento político de un país. Así, es lógico pensar que estos sentimientos negativos que se reflejan en la literatura son consecuencia directa de un sistema político que, desde la perspectiva de los personajes (no ya de los autores pues siempre existen puntos de vista antagónicos en cualquier obra) no responde a los intereses del pueblo sino a los propios del sistema.

## 8 Bibliografía

- Anthony, D. L. (s. f.). *AntConc (Windows, Macintosh OS X y Linux)*.
- Anthony, L. (2022). *TagAnt (2.0.4)*. <http://www.laurenceanthony.net/>
- ASALE, R.-, & RAE. (s. f.). *Migración / Diccionario de la lengua española*. «Diccionario de la lengua española» - Edición del Tricentenario. Recuperado 21 de septiembre de 2022, de <https://dle.rae.es/migración>
- Behar, S. (2007). *La caída del hombre nuevo: Narrativa cubana del Período Especial*. [Tesis Doctoral (Doctor en Filosofía)]. Florida International University.
- Biber, D. (1993). Representativeness in Corpus Desing. En *Literary and Linguistic Computing* (Vol. 4, pp. 243-257). . <https://doi.org/10.1093/lc/8.4.243>
- Biber, D. (2015). Corpus-based and corpus-driven analyses of language variation and use. En *The Oxford Handbook of Linguistic Analysis* (Bernd Heine, Heiko Narrog).
- Bondi, M. (2010). Perspectives on keywords and keyness. An introduction. En *Keyness in texts*. John Benjamins Publishing Company.
- Brezina, V. (2018). *Statistics in Corpus Linguistics: A practical guide*. Cambridge University Press.
- Cabrera Arús, M. A., & Suquet, M. (2019). La moda en la literatura cubana, 1960-1979:Tejiendo y destejiendo al hombre nuevo. *Cuban Studies*, 47, 195-221.
- Carita Paradis. (2012). Lexical semantics. En *The encyclopedia of applied linguistics*. Wiley-Blackwell.
- Castro, F. (1961). *Discurso pronunciado por el comandante Fidel Castro Ruz, Primer Ministro del gobierno revolucionario y Secretario del PURSC, como conclusión de las reuniones con los intelectuales cubanos, efectuadas en la Biblioteca Nacional el*

16, 23 y 30 de junio de 1961.

<http://www.cuba.cu/gobierno/discursos/1961/esp/f300661e.html>

Clear, J. (1992). Corpus Sampling. En *New Directions in English Language Corpora: Methodology, Results, Software Developments* (pp. 21-31). Walter de Gruyter.  
*Corpus de Referencia del Español Actual (CREA) | Portal del Hispanismo*. (s. f.).

Recuperado 7 de octubre de 2022, de

<https://hispanismo.cervantes.es/recursos/corpus-referencia-del-espanol-actual-crea>

Deignan, A. (1997). Metaphors of desire. En *Language and desire: Encoding sex, romance and intimacy* (pp. 21-42). Routledge. Taylor and Francis Group.

Díaz, A. A. (2002). *La emigración cubana. Balance del siglo XX*.

<http://bibliotecavirtual.clacso.org.ar/Cuba/cemi-uh/20120821040024/emig.pdf>

Duany, J. (2005). *La migración cubana*. 164-181.

*English-Corpora: COCA*. (s. f.). Recuperado 11 de abril de 2023, de <https://www.english-corpora.org/coca/>

Fernández López, J. (2022, recuperado). *Novela regionalista o de la tierra*.

<http://www.hispanoteca.eu/Literatura%20LA/Novela%20regionalista%20o%20de%20la%20tierra.htm>

Fernández Retamar, R. (2000). Introducción a la literatura cubana. *América sin nombre*, 2, 5. <https://doi.org/10.14198/AMESN2000.2.02>

Fernández, T. (2000). La narrativa cubana del siglo XX: Notas para la reconstrucción de un proceso. *América sin nombre: boletín de la Unidad de Investigación de la Universidad de Alicante «Recuperaciones del mundo precolombino y colonial en el siglo XX hispanoamericano»*, ISSN 1577-3442, N.º. 2, 2000 (Ejemplar dedicado a:

*Revisiones de la literatura cubana*), pags. 84-91.

<https://doi.org/10.14198/AMESN2000.2.10>

- Guillermo Rojo Sánchez. (2010). Sobre codificación y explotación de corpus textuales: Otra comparación del Corpus del Español con el CORDE y el CREA. *Lingüística*, 24, 11-50.
- Harvey, K., & Shalom, C. (1997). *Language and Desire. Encoding Sex, Romance and Intimacy*. Routledge.
- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2014). *Metodología de la investigación* (6.<sup>a</sup> ed.). McGraw Hill Education.
- Hoover, D. L., Culpeper, J., & O'Halloran, K. (2014). *Digital Literary Studies: Corpus Approaches to Poetry, Prose, and Drama*. Routledge.
- Javi Fernández, Ester Boldrini, José Manuel Gómez, & Patricio Martínez-Barco. (2011). Análisis de sentimientos y Minería de opiniones: El corpus EmotiBlog. *Procesamiento del Lenguaje Natural*, 47, 179-187.
- Knowles, M. (1997). You would if you loved me: Language and desire in the teen novel. En *Language and desire: Encoding sex, romance, and intimacy*.
- Liu, B. (2015). *Sentiment Analysis. Mining opinions, sentiments, and emotions*. Cambridge University Press.
- López Fernández, P. S. (2015). MORETTI, Franco. Lectura distante. Trad. Lilia Mosconi. Fondo de Cultura Económica. *Revista de Humanidades digitales*, 6, 266-271.
- Marina, J. M., & Penas, M. L. (2000). *Diccionario de los sentimientos*. Anagrama.
- Mark Davies. (2009). Creating useful historical corpora: A comparison of CORDE, the Corpus del Español, and the Corpus do Português. En *Diacronía de las lenguas*

- iberorrománicas. Nuevas aportaciones desde la lingüística de corpus.* (Andrés Enrique-Arias, pp. 137-166). Iberoamericana/Vervuert.
- Martínez, M. S. (1999). Apuntes sobre el homoerotismo masculino y femenino en la literatura cubana de los 90. *Lectora: revista de dones i textualitat*, 37-48.
- McEnery, T., & Hardie, A. (2012). *Corpus Linguistics*. Cambridge University Press.
- Molina Salinas, C., & Sierra Martínez, G. (2015). Hacia una normalización de la frecuencia de los corpus CREA y CORDE. *Revista signos*, 48(89), 307-331.  
<https://doi.org/10.4067/S0718-09342015000300002>
- Moretti, F. (2005). *Graphs, Maps, Trees: Abstract Models for a Literary History*. Verso.
- Moretti, F. (2013). *Distant reading*. Verso Books.
- Parámetros habituales (CREA)*. (s. f.). Real Academia Española. Recuperado 20 de marzo de 2023, de <https://www.rae.es/banco-de-datos/crea/parametros-habituales-crea>
- Qiu, G., Liu, B., Bu, J., & Chen, C. (2009). Expanding Domain Sentiment Lexicon through Double Propagation. *JCA*, 9, 1199-1204.
- Río Barrial, I. (2015). *La literatura en el contexto de la Revolución Cubana: Análisis histórico-literario en el aula de ELE*. [Máster en Español como Lengua Extranjera]. Universidad de Oviedo.
- Rodríguez Sánchez, I., & Bautista Botello, E. (2021). Palabras clave en la literatura mexicana contemporánea escrita por mujeres. En *Humanidades Digitales: Corpus y Literatura en México*. (Rodríguez Sánchez, I&Bautista Botello, E, pp. 17-52). Bonilla Artigas Editores y Red de Humanidades Digitales.
- Sánchez Barrera, A. R. (2019). *Análisis del discurso de la Novela Rosa Contemporánea Mexicana desde la Lingüística de Corpus*. [Maestría en Lingüística]. Universidad Autónoma de Querétaro.

- Stubbs, M. (2010). Three concepts of keywords. En *Three concepts of keywords* (M. Bondi&M. Scott). John Benjamins Publishing Company.
- Theodor Lewandowski. (1982). *Diccionario de Lingüística*. Cátedra.
- Valderrama, B. (2015). Emociones: Una taxonomía para el Desarrollo Emocional. *Arandu UTIC*, 2(1). <http://www.utic.edu.py/revista.ojs/index.php/revistas/article/view/14>

## 9 Apéndice 1. Listado de autores y novelas

<b>Autores</b>	<b>Género</b>	<b>Publicación</b>	<b>País de publicación</b>	<b>Título</b>
Miguel de Carrión	novela	1917	Cuba	Las Honradas
Miguel de Carrión	novela	1919	Cuba	Las Impuras
Carlos Loveiras	novela	1920	Cuba	Generales y Doctores
Luis Rodríguez Embil	novela	1908	Cuba	Gil Luna Artista
Lydia Cabrera	cuento	1940	Cuba	Cuentos Negros de Cuba
Lydia Cabrera	cuento	1954	Cuba	El Monte
Carlos Montenegro	novela	1937	Cuba	Hombres sin mujer
Pablo de la Torriente Brau	novela	1940	Cuba	Aventuras del soldado desconocido cubano
Pablo de la Torriente Brau	cuento	1930	Cuba	Cuentos Completos
Dulce Ma. Loynaz	novela	1951	España	Jardín
Lino Novas Calvo	novela	1933	Cuba	El Negrero
Alejo Carpentier	novela	1956	Cuba	El Acoso
Alejo Carpentier	novela	1953	México	Los pasos perdidos
Alejo Carpentier	novela	1949	México	El reino de este mundo
Alejo Carpentier	novela	1934	España	Ecué-Yamba-O
Alejo Carpentier	novela	1974	México	El recurso del método
Alejo Carpentier	novela	1978	México	La consagración de la primavera
Alejo Carpentier	novela	1962	México	El siglo de las Luces
José Lezama Lima	novela	1966	Cuba	Paradiso
José Lezama Lima	novela	1977	Cuba/México	Oppiano Licario
Virgilio Piñera	novela	1952	Cuba	La carne de René
Virgilio Piñera	cuento	1956	Cuba	Cuentos Fríos
Onelio Jorge Cardoso	cuento	1958	Cuba	El Cuentero
Eliseo Diego	cuento	1946	Cuba	Divertimentos
Guillermo Cabrera Infante	cuento	1960	Cuba	Así en la Guerra como en la Paz
Guillermo Cabrera Infante	novela	2010	España	Cuerpos divinos
Guillermo Cabrera Infante	novela	2013	España	verbo auxiliar
Guillermo Cabrera Infante	novela	1967	España	Tres tristes tigres
Miguel Barnet	novela	1983	Cuba	Gallego
Miguel Barnet	novela	1969	Cuba	Canción de Rachel
Miguel Barnet	novela	1966	Cuba	Biografía de un Cimarrón

Jesús Díaz	novela	1996	España	La piel y la Máscara
Jesús Díaz	novela	2000	España	Siberiana
Jesús Díaz	novela	1992	España	Las Palabras perdidas
Jesús Díaz	novela	1987	Cuba/España	Las iniciales de la Tierra
Luis Rogelio Noguera	novela	1980	Cuba	Y si muero mañana
Senel Paz	novela	1988	Cuba	Un rey en el Jardín
Senel Paz	novela	1991	Cuba	El lobo, el bosque y el hombre nuevo
Senel Paz	novela	2007	Cuba	En el cielo con Diamantes
Eliseo Alberto	novela	2005	España	Esther en alguna parte
Eliseo Alberto	novela	1998	España	Caracol Beach
Leonardo Padura	novela	2020	España	Como Polvo en el viento
Leonardo Padura	novela	2018	España	La transparencia del tiempo
Leonardo Padura	cuento	2015	España	Aquello estaba deseando ocurrir
Ena Lucía Portela	novela	2002	España	Cien botellas en una pared

*Tabla 6. Listado de novelas y autores que conforman el CLCXX+*



## 10 Apéndice 2. Listados Palabras Clave determinadas en cada uno de los cruces.

Rango	Frecuencia	Keyness	Keyword
1	42046	409383.36	la
2	38042	370267.6	que
3	37685	366761.05	y
4	28715	279256.44	a
5	28670	278797.96	en
6	18527	179958.29	los
7	13841	134423.09	con
8	12610	122453.09	las
9	12610	122417.94	las
10	10575	102670.74	del
11	9365	90911.56	lo
12	6963	67575.47	como
13	4775	46326.99	es
14	3284	31852.11	si
15	3184	30881.44	ya
16	2792	27076.58	qué
17	2095	20297.54	dos
18	1481	14340.08	nos
19	1269	12296.9	tiempo
20	907	8760.17	tú
21	871	8435.55	todas
22	839	8125.11	esto
23	759	7336.57	tu
24	675	6534.21	misma
25	592	5729.12	mar
26	566	5476.94	contra
27	553	5350.84	teresa
28	7659	5153.73	me
29	5009	5153.06	mi
30	449	4342.17	primera

31	438	4213.76	aquellos
32	713	3779.11	rené
33	4879	3672.64	había
34	361	3467.81	nombre
35	343	3314.22	soy
36	1383	3183.45	pedro
37	4456	2941.57	era
38	307	2904.15	pascasio
39	464	2772.7	á
40	1666	2651.19	ojos
41	491	2387.28	bárbara
42	234	2221.83	da
43	3373	2196.25	yo
44	225	2182.99	menegildo
45	223	2150.79	cha
46	209	2015.08	has
47	209	2015.08	pasa
48	2481	1949.7	ella
49	5939	1759.85	le
50	356	1757.2	rogelio
51	578	1678.27	negros
52	172	1656.48	fui
53	2706	1654.37	él
54	169	1639.66	mármolo
55	1265	1623.48	aquel
56	1165	1621.73	aquella
57	156	1501.44	flores
58	182	1424.08	rigoletto
59	553	1414.83	teresa
60	152	1411.32	tanta
61	223	1370.46	cha
62	709	1343.2	carne
63	1916	1286.97	estaba
64	130	1261.27	hicotea
65	129	1230.91	brai
66	266	1226.73	marineros
67	126	1210.8	lecho
68	1161	1178.84	mis
69	889	1176.53	pronto

70	1501	1174.24	tenía
71	117	1135.15	#####
72	118	1133.31	jóvenes
73	121	1114.1	mouche
74	111	1076.93	powlavski
75	147	1075.33	baldomero
76	109	1037.53	the
77	104	997.73	mas
78	128	994.13	negreros
79	129	990.45	negrero
80	152	987.67	dalia
81	1244	982.39	noche
82	102	978.37	mes
83	104	974.86	mas
84	199	966.88	graciela
85	806	954.19	hacia
86	2792	952.09	qué
87	384	935.96	barco
88	97	929.96	loca
89	97	929.96	portal
90	96	920.28	d
91	141	911.85	wong
92	96	904.56	debo
93	251	872.88	susana
94	108	856	morita
95	87	844.08	dolé
96	9365	837.62	lo
97	290	820.64	alicia
98	869	819.24	iba
99	86	815.32	medias
100	93	787.12	cochón
101	1075	784.9	usted
102	1461	783.28	hombre
103	1521	782.92	casa
104	16396	778.52	un
105	81	775.08	fe
106	480	771.26	brazos
107	92	768.55	cañizo
108	1024	759.77	mí
109	883	741.79	manos
110	84	734.81	mongo

111	76	726.7	leche
112	907	721.54	tú
113	141	713.73	nene
114	1882	710.03	te
115	483	705.76	alma
116	638	690.27	sol
117	37685	663.81	y
118	3709	660.39	sin
119	744	657.62	hubiera
120	382	657.2	capitán
121	154	645.07	teodoro
122	798	644.02	voz
123	350	640.44	jardín
124	949	637.07	eran
125	5153	634.45	sus
126	138	614.29	presidio
127	678	607.99	cara
128	985	606.96	habían
129	382	594.03	instante
130	1665	593.95	dijo
131	63	593.4	deber
132	147	591.11	candela
133	870	583.11	hombres
134	523	580.95	corazón
135	60	571.93	pienso
136	592	571.87	mar
137	352	569.87	sombra
138	62	556.34	baro
139	61	551.52	collum
140	89	546.63	oscuro
141	63	538.43	ormond
142	698	537.26	puerta
143	804	535.53	cabeza
144	218	535.03	gritó
145	55	533.61	trebijo
146	1662	533.3	tan
147	112	525.19	galera
148	540	516.31	parecía
149	328	512.19	labios
150	59	511.71	palo
151	482	500.93	negro

152	116	499.83	muñeco
153	876	496.44	allí
154	65	495.6	jíbaro
155	51	494.8	capinche
156	139	491.09	gallinas
157	274	481.16	habana
158	595	480.83	aire
159	178	478.99	exclamó
160	106	475.87	chiquito
161	299	474.83	pecho
162	548	468.45	allá
163	48	465.7	kénde
164	332	461.2	sentía
165	496	460.53	boca
166	64	459.05	manigua
167	510	456.73	silencio
168	10748	448.42	su
169	842	441.88	mano
170	419	439.89	vio
171	47	439.32	platos
172	308	437.32	miró
173	269	434.23	muchacho
174	63	433.5	batey
175	48	432.66	palos
176	70	430.95	georgina
177	278	425.07	hablaba
178	92	423.95	mulata
179	268	422.8	joaquín
180	3184	419.01	ya
181	43	417.19	llillina
182	43	417.19	rudesindo
183	43	417.19	ñogumá
184	862	414.28	hizo
185	400	413.22	muerto
186	326	411.91	vieja
187	319	409.77	caer
188	43	407.66	arere
189	43	407.66	usebio
190	55	404.92	burón
191	312	401.42	acaso
192	61	401.31	obscura

193	92	399.63	cruceros
194	42	398	existe
195	368	396.71	rostro
196	324	395.36	mamá
197	188	395.19	echó
198	55	393.58	bohío
199	208	393.51	enseguida
200	94	392.95	noel
201	1326	392.65	poco
202	240	392.3	patio
203	449	391.68	cuarto
204	682	389.6	cosa
205	69	389.58	dominga
206	41	388.35	función
207	41	388.35	longina
208	173	386.02	esclavos
209	68	385.07	barracones
210	291	382.75	mío
211	13841	378.62	con
212	106	378.38	factoría
213	39	378.38	hiliodomiro
214	343	377.51	marido
215	13695	375.98	no
216	270	374.67	dedos
217	1352	371.23	nada
218	478	371.16	decía
219	465	370.14	viejo
220	441	369.6	sabía
221	38	368.68	FALSO
222	38	368.68	tatabisako
223	40	366.37	mezy
224	51	365.85	torriente
225	451	363.64	pie
226	184	362.35	selva
227	679	360.79	podía
228	690	359.73	tierra
229	37	358.98	rompemontes
230	37	358.98	shiakar
231	48	358.52	mackandal
232	353	356.93	detrás
233	964	356.73	mujer

234	292	351.47	viento
235	1110	348.08	entonces
236	28715	344.76	a
237	39	343.13	ayere
238	52	342.49	guajiros
239	311	341.94	niña
240	109	340.15	súbito
241	41	338.38	pa
242	61	336.48	harén
243	340	332.86	daba
244	37	332.82	ten
245	348	332.27	cielo
246	35	330.45	os
247	34	329.87	canot
248	34	329.87	dekín
249	66	329.32	viejito
250	40	329.23	lenormand
251	40	329.23	ogún
252	312	328.35	pobre
253	185	328.27	detuvo
254	50	323.2	neófitos
255	177	323.07	oh
256	381	322.32	volvió
257	283	321.81	brazo
258	42	321.34	aviadora
259	904	321.29	he
260	34	320.8	afirmaba
261	34	320.8	oeste
262	34	320.8	t
263	33	320.17	krumen
264	33	320.17	nieburg
265	121	318.43	ja
266	48	318.26	matienzo
267	37	316.98	ten

268	197	313.11	miraba
269	297	312.75	gesto
270	380	311.87	mirada
271	33	311.16	sacaba
272	33	311.16	velo
273	4650	307.97	pero
274	759	307.73	tu
275	56	307.12	clavé
276	263	306.56	hermana
277	55	306.43	florinda
278	215	306.42	mía
279	110	305.97	compadre
280	201	304.91	hombros
281	314	304.76	pies
282	115	303.63	coco
283	35	301.83	walo
284	89	301.17	matanzas
285	277	299.69	veía
286	59	298.1	salomé
287	1582	296.05	ahora
288	159	293.85	negras
289	31	291.88	crear
290	31	291.88	loro
291	30	291.06	ariyénye
292	1950	290.13	vez
293	32	290.04	cera
294	2842	288.22	cuando
295	185	287.24	acababa
296	359	286.93	puso
297	174	286.19	sentí
298	31	285.7	do
299	31	285.7	soyán
300	29	281.36	mambiala

Tabla 7. Primeras 300 Palabras Clave. Cruce CREA vs CLCXX+1

rango	freq	keyness	keyword
1	97754	800585	la
2	87255	714248	que
3	83768	685570	y
4	64701	529065	a
5	61696	504400	en
6	34692	283218	los
7	28397	231803	con
9	24383	198959	las
10	23069	188261	lo
11	22656	184887	del
12	16114	131455	como
13	10239	83498	es
14	7928	64641	si
15	6507	53048	qué
16	5721	46637	ya
17	3933	32038	dos
18	3194	26011	nos
19	2891	23527	tú
20	2554	20808	tiempo
21	14438	16515	había
22	17775	13064	me
23	1429	11621	tu
24	12159	11301	era
25	1364	11105	todas

26	1289	10494	esto
27	1165	9483	contra
28	7543	9407	dijo
29	9445	9009	yo
30	1101	8961	primera
31	1080	8790	misma
32	7799	7391	él
33	888	7200	nombre
34	880	7159	soy
35	858	6980	mar
36	15358	6974	le
37	6610	6831	ella
38	5712	6067	estaba
39	626	5066	aquellos
40	576	4682	fui
41	571	4539	blen
42	716	4422	fronesis
43	531	4315	pasa
44	594	3984	cemí
45	592	3845	irving
46	2891	3752	tú
47	23069	3712	lo
48	3786	3617	tenía
49	438	3517	da
50	425	3465	nadiezdha

51	1578	3249	dije
52	383	3109	flores
53	6873	3060	mi
54	397	2990	cué
55	1103	2983	habana
56	13056	2922	pero
57	362	2875	loreta
58	1716	2825	sabía
59	1214	2786	conde
60	34176	2785	no
61	611	2737	elisa
62	408	2690	foción
63	327	2653	has
64	6507	2585	qué
65	2230	2557	iba
66	4777	2514	te
67	3795	2495	casa
68	306	2481	víctor
69	586	2476	flaco
70	300	2372	tanta
71	2250	2344	podía
72	4756	2245	ahora
73	1423	2219	cuba
74	273	2213	jóvenes
75	960	2212	sintió
76	270	2178	the
77	453	2164	bobby
78	1795	2135	hacía
79	2705	2083	noche

80	1206	2023	preguntó
81	299	1973	franqui
82	333	1968	licario
83	511	1952	coño
84	238	1940	VERDADERO
85	465	1927	adela
86	1949	1924	aquella
87	967	1887	miró
88	2445	1857	habían
89	228	1846	fe
90	517	1839	horacio
91	1467	1822	parecía
92	83768	1799	y
93	221	1789	mes
94	2275	1760	eran
95	217	1757	leche
96	379	1755	bárbaro
97	4086	1679	eso
98	422	1665	lino
99	2011	1657	aquel
100	7422	1632	cuando
101	3550	1576	nada
102	2207	1573	mí
103	190	1537	pienso
104	1268	1521	quería
105	185	1496	teresa
106	325	1490	adriano
107	433	1472	carajo
108	1521	1454	pronto

109	176	1423	loca
110	174	1419	raydel
111	2021	1416	madre
112	512	1396	sonrió
113	290	1390	ofelia
114	484	1379	esteban
115	494	1371	gordo
116	168	1358	portal
117	182	1341	rialta
118	490	1326	bernardo
119	1281	1321	decía
120	4231	1304	donde
121	1048	1298	debía
122	682	1298	virgen
123	448	1297	walter
124	1695	1274	sé
125	751	1264	hubiese
126	1900	1260	ojos
127	1065	1241	volvió
128	151	1231	tolia
129	64701	1207	a
130	242	1207	gisela
131	713	1189	pensó
132	647	1181	abuela
133	148	1177	debo
134	34665	1176	un
135	785	1162	sentía
136	1966	1161	mano
137	654	1150	negros

138	190	1145	ramsés
139	345	1136	oso
140	139	1133	gofia
141	367	1119	linda
142	367	1111	darío
143	774	1090	rojo
144	133	1073	función
145	1906	1053	allí
146	232	1053	orestes
147	158	1047	alix
148	956	1043	vio
149	755	1041	veía
150	146	1014	rine
151	124	991	medias
152	442	986	mierda
153	121	975	d
154	1181	971	allá
155	168	966	vivian
156	23517	965	su
157	179	956	galeb
158	3377	951	sí
159	1293	946	bueno
160	117	942	sacaba
161	4390	936	ni
162	141	929	elizardo
163	529	915	muchacha
164	1391	897	puerta
165	795	881	clara
166	5044	880	porque

167	539	878	cubano
168	108	853	yoyi
169	985	850	cuarto
170	112	839	palo
171	792	831	daba
172	1346	825	calle
173	940	817	negro
174	1562	813	cabeza
175	243	809	cigarro
176	101	804	platos
177	324	802	apartamento
178	7928	800	si
179	103	799	walterio
180	108	794	normand
181	143	778	ynaca
182	681	778	mamá
183	185	777	mulato
184	271	776	ron
185	95	775	FALSO
186	112	768	hugues
187	112	768	rubito
188	2685	766	algo
189	2438	763	entonces
190	459	760	miraba
191	466	754	esquina
192	2207	752	ver
193	93	747	existe
194	699	746	empezó
195	3256	745	siempre

196	661	740	sabes
197	2353	739	casi
198	371	739	gritó
199	1195	735	cara
200	572	729	venía
201	3258	726	tan
202	2196	724	nunca
203	719	716	café
204	4374	716	vez
205	683	715	vi
206	234	713	antoni
207	191	707	larry
208	537	707	hablaba
209	86	701	fitzberg
210	2322	698	cómo
211	136	697	espejuelos
212	472	694	llamaba
213	576	694	fui
214	86	690	mas
215	86	690	putas
216	804	690	cama
217	189	690	hemingway
218	441	682	elena
219	1030	678	voy
220	83	677	guesty
221	86	669	mas
222	1352	668	voz
223	1127	668	estaban
224	100	667	comemierda



225	959	663	silencio
226	174	661	malecón
227	319	660	manolo
228	947	659	viejo
229	83	651	bustrófedon
230	901	648	amigo
231	80	642	sarusky
232	251	641	tom
233	391	637	gustaba
234	102	634	mandy
235	1251	633	solo
236	344	629	sofía
237	468	628	pasaba
238	857	628	miedo
239	90	627	des
240	160	626	vedado
241	527	625	respondió
242	548	621	seguía
243	477	606	muchacho
244	607	606	rato
245	83	601	pa
246	76	594	mayorala
247	75	593	cubanas
248	690	589	rostro
249	311	589	carro
250	1427	589	verdad
251	1189	588	hubiera
252	436	583	loco
253	72	576	puigventós

254	452	575	conocía
255	115	575	mirta
256	75	574	palos
257	568	572	pelo
258	70	571	baldoquina
259	73	570	cayito
260	223	570	silvestre
261	2484	568	hacia
262	137	568	barral
263	74	566	liuba
264	90	566	fontanar
265	3294	565	después
266	875	563	boca
267	2057	561	aquí
268	71	561	deber
269	70	560	crear
270	70	560	oeste
271	342	559	cubanos
272	100	558	kindelán
273	366	555	enseguida
274	213	555	murmuró
275	2136	554	luego
276	1150	553	hablar
277	351	553	decían
278	343	548	volvía
279	1444	547	todavía
280	67	546	yúnior
281	1145	545	dio
282	351	544	cubana

283	139	544	campbell
284	101	544	silvina
285	68	544	actual
286	85	542	branly
287	575	542	vieja
288	255	540	rené
289	214	539	iris
290	605	537	llevaba
291	99	536	lúmino

292	520	534	negra
293	194	534	silvio
294	69	532	zun
295	145	531	augusta
296	16114	530	como
297	1514	527	padre
298	509	527	labios
299	138	525	mulata
300	548	525	acaso

Tabla 8. Primeras 300 Palabras Clave. Cruce Crea vs. CLCXX+2

Rango	Frecuencia	Keyness	Keyword
1	7543	1085.34	dijo
2	1214	917.57	conde
3	1414	738.78	carlos
4	14438	717.71	había
5	716	618.12	fronesis
6	594	512.79	cemí
7	571	492.93	blen
8	592	488.49	irving
9	611	472.51	elisa
10	517	433.91	horacio
11	13056	404.13	pero
12	453	391.06	bobby
13	511	381.65	coño
14	484	379.81	esteban
15	425	366.89	nadiezda
16	490	364.1	bernardo
17	448	356.97	walter
18	408	352.21	foción
19	1578	346.19	dije
20	12159	341.09	era
21	586	336.96	flaco
22	582	333.88	incluso
23	362	312.5	loreta

24	3082	303.06	esa
25	9445	300.79	yo
26	344	296.96	sofía
27	367	296.15	darío
28	465	294.59	adela
29	363	292.74	fidel
30	5044	287.73	porque
31	333	287.46	licario
32	360	282.31	david
33	5712	279.66	estaba
34	319	275.38	manolo
35	4231	275.34	donde
36	325	269.08	adriano
37	34176	268.42	no
38	500	267.64	alberto
39	4756	266.55	ahora
40	7799	265.76	él
41	442	263.01	mierda
42	299	258.11	franqui
43	422	256.06	lino
44	433	251.81	carajo
45	290	250.34	ofelia
46	494	245.46	gordo
47	290	239.09	castro
48	4086	232.73	eso

49	397	232.28	cué
50	266	218.55	magistrado
51	379	214.33	bárbaro
52	242	208.9	gisela
53	1423	202.94	cuba
54	795	202.43	clara
55	234	202	antoni
56	345	200.36	oso
57	232	200.27	orestes
58	324	196.82	apartamento
59	1716	192.06	sabía
60	378	191.3	inmediato
61	271	188.66	miami
62	539	177.56	cubano
63	3135	176.51	ese
64	251	172.31	tom
65	647	168.78	abuela
66	473	168.36	decidió
67	569	168.09	quizás
68	306	166.59	víctor
69	214	166.22	martin
70	191	164.88	larry
71	190	164.01	ramsés
72	17775	163.63	me
73	2250	160.49	podía
74	390	159.78	marcos
75	2891	159.19	tú
76	549	158.34	revolución
77	1048	158.06	debía
78	182	157.11	rialta
79	194	157.02	silvio
80	282	156.08	ricardo
81	960	156.03	sintió
82	441	155.09	elena
83	179	154.52	galeb
84	774	153.97	rojo
85	189	152.76	hemingway
86	174	150.2	raydel
87	6610	149.08	ella
88	391	147.27	director
89	1695	144.1	sé

90	1206	142.57	preguntó
91	1268	142.45	quería
92	696	141.84	final
93	408	136.62	cine
94	158	136.39	alix
95	168	134.86	aeropuerto
96	168	134.86	vivian
97	751	134.32	hubiese
98	243	134.21	cigarro
99	512	133.35	sonrió
100	166	133.16	lidia
101	618	132.89	tipo
102	151	130.35	tolia
103	400	129.44	teléfono
104	599	129.39	imagen
105	161	128.91	carteles
106	177	128.57	batista
107	148	127.76	televisión
108	713	127.25	pensó
109	248	126.91	jorge
110	146	126.03	peralta
111	146	126.03	rine
112	311	125.51	carro
113	165	124.95	culo
114	97754	124.12	la
115	143	123.44	ynaca
116	163	123.27	ramiro
117	141	121.71	elizardo
118	200	121.6	enrique
119	1021	120.29	creo
120	139	119.99	campbell
121	139	119.99	gofia
122	223	119.11	silvestre
123	195	117.68	fotos
124	209	116.13	silvia
125	682	115.45	virgen
126	621	115.14	policía
127	2328	114.17	aunque
128	132	113.95	gaspar
129	130	112.22	maruja
130	346	111.62	relación

131	129	111.36	joel
132	351	110.35	cubana
133	147	109.87	jean
134	306	109.46	pablo
135	214	108.83	iris
136	126	108.77	miriam
137	137	108.51	barral
138	391	107.7	parís
139	260	107.43	puta
140	262	106.05	bar
141	174	105.62	malecón
142	713	105.61	país
143	133	105.12	gustavo
144	206	102.96	madrid
145	292	102.91	regresó
146	253	102.69	trago
147	23069	102.36	lo
148	236	102.32	revista
149	3376	102.27	sí
150	248	102.05	lucía
151	115	99.27	mirta
152	275	99.23	laura
153	134	99.02	gil
154	205	98.89	radio
155	1103	98.31	habana
156	146	97.53	cojones
157	145	96.72	augusta
158	112	96.68	hugues
159	278	96.38	quizá
160	164	93.63	comunista
161	108	93.23	beto
162	108	93.23	normand
163	108	93.23	yoyi
164	4777	91.7	te
165	106	91.5	milller
166	4290	89.08	fue
167	103	88.91	margaret
168	103	88.91	walterio
169	8710	88.58	o
170	146	88.06	información
171	102	88.05	mandy

172	102	88.05	tamara
173	101	87.19	silvina
174	139	87.04	you
175	1882	86.98	mejor
176	100	86.32	comemierda
177	99	85.46	lúmino
178	207	85.27	terrazza
179	391	85.11	gustaba
180	4870	84.27	hasta
181	719	83.94	café
182	128	83.03	embajada
183	96	82.87	champollion
184	211	82.57	ramos
185	121	82.48	gustado
186	278	81.22	película
187	104	80.57	miss
188	104	80.57	oscar
189	141	80.01	moisés
190	124	79.83	héctor
191	92	79.42	duke
192	102	78.89	carmelo
193	360	78.57	parque
194	90	77.69	fontanar
195	170	77.07	mario
196	174	76.86	ministerio
197	88	75.96	olaya
198	106	75.78	esther
199	150	75.64	cabrón
200	98	75.51	rio
201	163	75.2	dólares
202	87	75.1	lezama
203	257	74.89	nacional
204	2376	74.8	años
205	86	74.24	fitzberg
206	148	74.17	vera
207	96	73.83	siberia
208	85	73.37	agnes
209	85	73.37	brantly
210	84	72.51	restaurant
211	83	71.65	bustrófedon
212	83	71.65	guesta

213	1427	71.32	verdad
214	205	71.24	foto
215	2322	70.87	cómo
216	82	70.78	ingrid
217	82	70.78	show
218	81	69.92	livia
219	245	69.85	decisión
220	321	69.48	regresar
221	111	69.48	rusa
222	80	69.06	sam
223	80	69.06	sarusky
224	131	68.64	duque
225	1251	68.49	solo
226	3550	68.29	nada
227	79	68.19	ismael
228	135	68.07	bruno
229	908	67.99	alguien
230	120	67.92	che
231	78	67.33	bélgica
232	78	67.33	mai
233	373	67.33	respuesta
234	78	67.33	theo
235	160	67.19	vedado
236	88	67.09	arístides
237	113	66.63	catalina
238	77	66.47	buró
239	77	66.47	claudé
240	118	66.39	paul
241	967	66.27	miró
242	137	66.17	asintió
243	87255	65.93	que
244	112	65.85	maricón
245	122	65.58	tetas
246	100	65.48	beach
247	100	65.48	kindelán
248	244	65.45	poeta
249	86	65.41	bolero
250	699	65.3	empezó
251	75	64.74	camilo
252	212	64.52	problemas
253	160	64.43	gómez

254	361	64.38	barrio
255	92	64.25	ballet
256	92	64.25	eugenio
257	74	63.88	liuba
258	74	63.88	rafaela
259	2096	63.81	mientras
260	6507	63.4	qué
261	2685	63.05	algo
262	73	63.01	cayito
263	83	62.89	jodido
264	149	62.48	conejo
265	72	62.15	aldama
266	72	62.15	jesse
267	72	62.15	leticia
268	72	62.15	puigventós
269	131	61.85	poesía
270	112	61.82	rubito
271	1831	61.32	decir
272	249	61.03	coronel
273	304	60.87	necesitaba
274	125	60.78	camión
275	661	60.75	sabes
276	70	60.43	baldoquina
277	70	60.43	lucinda
278	110	60.31	ruso
279	530	60.14	estás
280	214	59.69	universidad
281	69	59.56	zun
282	79	59.54	vide
283	336	59.38	irse
284	103	58.9	exilio
285	68	58.7	berto
286	68	58.7	gipsy
287	78	58.7	profesora
288	737	58.22	amigos
289	826	58.2	realidad
290	67	57.84	gundín
291	67	57.84	karla
292	67	57.84	sabá
293	67	57.84	yúnior
294	96	57.67	nelson

295	106	57.29	catalán
296	120	57.17	certeza
297	66	56.97	candito

298	66	56.97	guevara
299	66	56.97	titón
300	66	56.97	totó

Tabla 9. Primeras 300 Palabras Clave. Cruce CLCXX+1 vs. CLCXX+2

Rango	Frecuencia	Keyness	Keyword
1	1383	2716.87	pedro
2	464	924.93	á
3	713	729.74	rené
4	491	714.9	bárbara
5	307	712.23	pascasio
6	356	664.94	rogelio
7	5009	614.44	mi
8	553	591.05	teresa
9	225	521.98	menegildo
10	709	511.36	carne
11	251	499.3	susana
12	268	496.15	joaquín
13	1075	467.07	usted
14	290	448.9	alicia
15	199	440.75	graciela
16	266	409.89	marineros
17	384	405.32	barco
18	182	393.8	rigoletto
19	169	392.06	mármolo
20	223	370.23	cha
21	1666	366.45	ojos
22	68115	320.96	de
23	1405	320.47	ha
24	152	318.43	dalia
25	523	311.86	corazón
26	130	301.58	hicotea
27	1161	301.07	mis
28	129	299.26	brai
29	18527	294.85	los
30	147	289.34	baldomero
31	121	280.7	mouche
32	141	275.99	nene
33	129	272.88	negrero

34	128	270.61	negreros
35	138	269.32	presidio
36	154	263.23	teodoro
37	483	261.39	alma
38	350	258.48	jardín
39	111	257.5	powlavski
40	108	250.54	morita
41	106	245.9	factoría
42	112	234.28	galera
43	435	218.92	éste
44	93	215.75	cochón
45	92	213.43	cañizo
46	3709	211.13	sin
47	904	209.59	he
48	141	205.22	wong
49	870	204.35	hombres
50	92	203.12	cruceros
51	87	201.83	dolé
52	173	199.37	emoción
53	638	191.19	sol
54	84	184.75	gastón
55	95	179.38	factores
56	477	172.22	señor
57	130	169.73	rosario
58	197	169.3	andrés
59	116	164.57	muñeco
60	75	164.09	ó
61	94	163.77	noel
62	331	160	don
63	84	155.29	mongo
64	65	150.79	jíbaro
65	69	150.34	dominga
66	151	148.89	ignacio
67	184	147.63	selva

68	63	146.15	ormond
69	62	143.83	baro
70	1206	142.08	este
71	61	141.51	collum
72	5153	140.16	sus
73	382	139.06	capitán
74	126	134.95	lecho
75	172	133.84	pensamiento
76	480	133.34	brazos
77	1247	133.23	sólo
78	292	132.41	viento
79	907	131.44	gran
80	80	130.32	portugués
81	131	130	presos
82	578	129.7	negros
83	70	129.41	georgina
84	12610	127.97	las
85	55	127.59	burón
86	55	127.59	trebijo
87	173	126.98	esclavos
88	2687	125.83	sobre
89	343	124.62	estas
90	343	123.16	marido
91	193	120.13	ramón
92	352	120.12	sombra
93	61	119.55	harén
94	51	118.31	capinche
95	227	118.15	ésta
96	89	117.6	oscuro
97	50	115.99	neófitos
98	59	115.1	salomé
99	84	113.34	luisa
100	48	111.35	kénde
101	48	111.35	matienzo
102	883	110.99	manos
103	175	110.92	trata
104	85	109.98	repuso
105	54	109.52	cautivos
106	56	108.45	clavé
107	58	107.9	armadores
108	279	105.07	duda

109	75	104.17	placeres
110	1612	104.11	todos
111	58	103.38	evaristo
112	79	101.17	arturo
113	382	100.54	instante
114	1240	99.87	esta
115	43	99.75	llillina
116	43	99.75	rudesindo
117	43	99.75	usebio
118	43	99.75	ñogumá
119	125	99.03	barcos
120	66	98.23	piratas
121	1461	98.13	hombre
122	51	97.4	torriente
123	41	95.11	longina
124	56	95	carlota
125	192	94.83	costa
126	90	94.4	adelantado
127	110	94.05	seda
128	1326	93.34	poco
129	40	92.79	lenormand
130	40	92.79	mezy
131	128	92.36	nuevamente
132	46	91.59	leona
133	64	91.2	pirata
134	81	90.91	semblante
135	54	90.77	venado
136	50	90.48	cuca
137	39	90.47	ayere
138	39	90.47	hiliodomiro
139	145	90.14	torno
140	321	89.99	oro
141	57	89.76	sinfín
142	115	89.72	coco
143	55	89.14	bohío
144	475	88.91	embargo
145	42	88.69	aviadora
146	253	88.59	rey
147	38	88.15	camacho
148	38	88.15	tatabisako
149	139	86.68	gallinas

150	106	86.31	éstos
151	50	86.24	marinería
152	37	85.83	rompemontes
153	37	85.83	shiakar
154	223	84.9	causa
155	43	84.89	arere
156	72	83.47	celda
157	293	83.43	cabo
158	73	83.08	dulzura
159	688	82.13	aún
160	44	82.02	christophe
161	48	81.99	mackandal
162	135	81.34	amo
163	35	81.19	walo
164	41	80.44	factorías
165	54	80.37	carlo
166	76	80.07	ademán
167	81	79.09	eduardo
168	34	78.87	canot
169	34	78.87	clelia
170	34	78.87	dekín
171	311	78.85	niña
172	690	78.63	tierra
173	3184	78.55	ya
174	175	78.51	ideas
175	53	78.36	derredor
176	1265	77.46	aquel
177	33	76.55	krumen
178	33	76.55	nieburg
179	76	76.21	mozo
180	39	75.99	chiquilla
181	119	74.77	pájaros
182	614	73.91	mujeres
183	105	73.7	pensamientos
184	1857	73.53	entre
185	40	73.29	ogún
186	35	72.81	curita
187	35	72.81	ranchero
188	75	72.7	marinero
189	31	71.91	soyán
190	100	71.74	ligero

191	63	71.66	ingreso
192	48	71.44	aguardaba
193	59	71.16	julia
194	65	70.98	fray
195	147	70.97	candela
196	81	70.56	carmela
197	426	70.5	han
198	167	70.35	rosa
199	97	70.05	cabellos
200	30	69.59	ariyénye
201	128	68.08	prisa
202	769	68.04	agua
203	43	67.83	poe
204	43	67.83	portugueses
205	46	67.43	alcoba
206	29	67.27	mambiala
207	29	67.27	marekén
208	29	67.27	yannes
209	35	67.14	curador
210	35	67.14	tiniebla
211	177	66.9	oh
212	201	66.81	efecto
213	37	66.78	serafín
214	595	66.59	aire
215	154	65.85	vientre
216	158	65.33	seguida
217	155	65.3	preciso
218	64	65.27	manigua
219	28	64.95	chéggue
220	28	64.95	serapio
221	118	64.69	cadáver
222	122	64.63	angustia
223	36	64.62	ruth
224	72	64.34	hondo
225	54	64.3	x
226	106	64.21	chiquito
227	546	63.8	amor
228	17896	63.78	se
229	41	63.73	celdas
230	119	62.9	querida
231	303	62.89	dolor



232	70	62.78	reja
233	27	62.63	argensola
234	27	62.63	beruá
235	27	62.63	ramira
236	53	62.47	gobernador
237	121	62.4	ja
238	49	62.21	dulcemente
239	839	61.97	esto
240	122	61.68	paisaje
241	43	61.46	goleta
242	79	61.41	brigada
243	187	60.59	lleno
244	26	60.31	alcaide
245	26	60.31	masilla
246	66	59.67	despecho
247	327	59.48	estos
248	29	59.26	gobernadora
249	46	59.07	islote
250	146	58.68	cubierta
251	43	58.61	crucero
252	61	58.39	obscura
253	63	58.19	batey
254	871	58.06	todas
255	25	58	amarar
256	25	58	wila
257	25	58	yé
258	592	57.81	mar
259	95	57.48	sentimientos
260	380	57.27	mirada
261	28	57.01	estuario
262	168	56.98	cartas
263	263	56.76	hermana
264	106	56.74	fiebre
265	73	56.69	inquietud
266	63	56.51	bruscamente

267	63	56.51	veneno
268	68	56.45	impaciencia
269	66	56.43	viejito
270	82	56.37	paco
271	30	56.14	negrada
272	32	56.02	carruaje
273	24	55.68	alvareda
274	24	55.68	contra maestre
275	24	55.68	montalbán
276	24	55.68	prendes
277	200	55.6	deseo
278	58	55.2	cuñado
279	85	54.41	espanto
280	29	53.95	salgari
281	31	53.88	poza
282	56	53.66	aquella
283	23	53.36	chichiriche
284	115	53.36	colegio
285	23	53.36	impuras
286	23	53.36	mordazo
287	23	53.36	sepulturero
288	245	53.04	luna
289	168	52.91	acabó
290	40	52.83	interno
291	40	52.83	zas
292	388	52.55	demás
293	26	52.51	cabotaje
294	312	51.87	pobre
295	89	51.8	matanzas
296	28	51.76	caravanas
297	60	51.73	vago
298	140	51.55	desconocido
299	1165	51.54	aquella
300	22	51.04	euryanthe

Tabla 10. Primeras 300 Palabras Clave. Cruce CLCXX+2 vs. CLCXX+1

## 11 Apéndice 3. Frecuencias relativas de los rasgos utilizados para el análisis de agrupación

[https://docs.google.com/spreadsheets/d/1QpCk\\_yelpdLgGf9dUI6ZWrpqm1ffBfg/edit#gid=1151508973](https://docs.google.com/spreadsheets/d/1QpCk_yelpdLgGf9dUI6ZWrpqm1ffBfg/edit#gid=1151508973)