

APUNTES SOBRE LA ELABORACIÓN DE UN CORPUS ELECTRÓNICO DE DOCUMENTOS DEL ESPAÑOL DE AMÉRICA *

MICAELA CARRERA DE LA RED
ANDREA HERRÁN SANTIAGO
Universidad de Valladolid

1. LA LINGÜÍSTICA DE CORPUS Y LOS CORPUS HISTÓRICOS ESPECÍFICOS

El ideal en el análisis lingüístico es el carácter masivo de los datos, es decir, contar con un corpus cuantitativamente elevado en una tarea en la cual, si hace poco más de dos décadas existían aún las limitaciones propias de algo que comenzaba (Sinclair 1987; Sánchez 2004:334), hoy día no cabe aducir la falta de un instrumento de trabajo acumulativo moderno con el que poder llevar a cabo tanto el análisis sincrónico de la lengua como el estudio de la variación y cambio lingüístico. Este es el principio sobre el que descansa la moderna metodología de análisis de corpus lingüísticos (Lawler y Dry 1998; Biber, Conrad y Reppen 2000).

En nuestro caso no se trata de ceder ante la llegada de la última generación de corpus con medidas de cientos de millones de palabras, con la incorporación indiscriminada de textos de procedencias diversas y sin ningún nexo de unión más que poner al

* Las primeras ideas sobre cómo organizar el léxico del corpus colombiano se van a publicar en un artículo de la revista AFA (Carrera, en prensa).

alcance del público una base de datos de proporciones magnas, sino que más bien se trata de poner de relieve el interés de esta metodología en lo que puede suponer para la construcción de corpus históricos específicos, equilibrados y estructurados, referidos a un área concreta de una lengua o sobre un aspecto determinado de una comunidad lingüística. Por esto, no nos cabe duda de que la lingüística de corpus como herramienta útil en el proyecto que nos ocupa necesita resolver el dilema no tanto de la medida del corpus sino de la naturaleza intencional del material elegido. En nuestro caso, se trataría de tomar conciencia de que, grandes o pequeños, la creación de corpus debe tratar de conseguir que sean adecuados a los usos concretos que se desee alcanzar con su creación (Nurmi 2002:173-184).

Nos alineamos así con la concepción y modo de actuar, por ejemplo, de uno de los centros de referencia de la Lingüística de Corpus que radica en la Universidad de Helsinki y con la tarea realizada desde finales de los ochenta y a lo largo de la década de los noventa en el ámbito anglosajón, mediante la cual los integrantes de dicho Centro han ido mostrando cómo se puede trabajar con corpus diacrónicos de la lengua inglesa bien estructurados de acuerdo con tipos discursivos y géneros textuales bien precisos (Kytö 1996 [1991])¹. Basado en dos corpus (inglés y escocés) se han hecho estudios desde ángulos y perspectivas muy variadas: (1) procesos internos de cambio lingüístico, (2) sociolingüística histórica, (3) dialectología y variación regional, (4) convenciones textuales y evolución de género, (5) variabilidad pragmática. Este corpus se ha completado en Edimburgo con textos del siglo XVII y se ha creado un nuevo corpus denominado *ARCHER (A Representative Corpus of Historical English Registers)*, que ha sido objeto de análisis textuales con un planteamiento de estudio multidimensional de registros (Biber y Finegan 2001:66-83).

En el ámbito románico, una serie de reuniones celebradas periódicamente en Friburgo están trayendo a la luz tanto proyectos de creación de corpus y bancos de datos como estudios en los que la lingüística de corpus se perfila como instrumento esencial en el análisis de las lenguas románicas en la vertiente hablada (Pusch y

¹ La materialización y concreción de sus objetivos se ha plasmado en un proyecto llamado *English in Transition: Change through Variation* (Raumolin-Brunserg *et al.* 2002) y en la creación de una unidad de investigación llamada la *Research Unit for Variation and Change in English (VARIENG)*.

Raible 2002) y en la diacrónica como medio de análisis del cambio lingüístico dentro del dominio románico (Kabatek, Pusch y Raible 2005)². En esta última perspectiva, la que a nosotros nos incumbe, estamos esperanzados en que se siga por caminos parejos con los señalados para el ámbito anglosajón. Los principios parecen descansar en una organización de testimonios lingüísticos y textuales de gran diversidad de épocas, géneros, tradiciones discursivas, registros, útiles para la descripción y explicación del cambio lingüístico diacrónico.

Guiados por los mismos principios, aunque un poco más limitados en disponibilidad de recursos, deseamos contribuir a la preparación de un corpus específico útil para el estudio de la diacronía del español trasladado a estas zonas de Hispanoamérica. El entronque de este proyecto se encuentra en más de una década de trabajo en conexión con las líneas del *Proyecto Coordinado para el Estudio de la Historia del Español de América, Canarias y Andalucía*, impulsado por la *Comisión para el Estudio Histórico del Español de América* de la ALFAL (Asociación de Lingüística y Filología de la América Latina), que, bajo la coordinación de la profesora M.^a Beatriz Fontanella de Weinberg (Bahía Blanca, Argentina, m. 1995), según boletín publicado en San Juan de Puerto Rico, el año 1990, acordó reunir un corpus de textos de archivo sobre los que trabajar, de forma coordinada entre todos los equipos, con el fin de conseguir estudios globales de la historia en las distintas regiones del español de América. En el ya lejano año 1991, Beatriz Fontanella vio la conveniencia de que la responsable del proyecto que aquí se presenta, buena conocedora de los principios de la filología medieval, comenzara a trabajar desde Valladolid en la selección del corpus documental de los siglos XVI y XVII de Santo Domingo, que no recibía atención por parte de ningún investigador

² Relacionada con proyectos de elaboración de corpus para el léxico en el dominio hispánico, se recoge la contribución de Clavería y Torruella (2005:215-228), parte de un proyecto que conecta lexicografía diacrónica en forma de diccionario con tratamiento electrónico de datos en el proceso de informatización del *DCECH (Diccionario crítico etimológico castellano e hispánico)* de Corominas y Pascual, labor que se está haciendo en la Universidad Autónoma de Barcelona, cuyas líneas maestras son muy útiles para cualquier actuación con base de datos (Clavería, Sánchez y Torruella 1996:1631-1643). Y entre los estudios de aspectos diacrónicos del español americano hechos sobre corpus destaca el trabajo de Wesch (2005:519-530), que analiza pragmáticamente un corpus documental de los siglos XV al XVII.

(Fontanella 1993), y más adelante, desde 1997, con la profesora Elena Rojas como coordinadora, la colaboración se amplió a la historia documental del español de Colombia de los siglos XVI al XVIII (Rojas 2000), que por otro tipo de razones tampoco podía ser estudiada de manera fácil desde el propio país. No obstante, en aquellos momentos se produjo el mayor estancamiento nunca antes experimentado en la marcha del Proyecto auspiciado por la ALFAL. A pesar de lo cual, como la idea es buena y, por ello, la tarea ha de proseguir en el futuro, a continuación se exponen algunas tentativas o pruebas que se desean proseguir, esta vez ya como Grupo de Investigación Reconocido para el Estudio de la Historia Lingüística de Hispanoamérica, por la Universidad de Valladolid.

2. ETAPAS SUCESIVAS EN LA INVESTIGACIÓN

La investigación tiene como meta última una historia documental del español dominicano y colombiano. Pero este esfuerzo era hasta hace poco una tarea unipersonal y las herramientas de tipo técnico de las que disponíamos eran más bien pobres. Existen, no obstante, algunos signos de que esto pueda ir cambiando en el futuro.

En principio se trataba de la tarea filológica de recogida de un corpus de textos o “documentos” en los archivos correspondientes. De entrada, esta recopilación no tenía más limitaciones que los requisitos previstos por la Comisión del Proyecto Coordinado: 1) que fueran “documentos”, esto es, textos de archivo en principio de índole no literaria (aunque sobre esto habría mucho que matizar), vinculados a la administración y diplomacia colonial españolas, 2) procedentes de distintos focos de emisión (en nuestro caso, emitidos en ciudades y villas de las áreas geográficamente hoy incluidas en República Dominicana y Colombia) y 3) fechados allí en los siglos XVI, XVII y XVIII. Este último criterio entraña limitaciones en la investigación, que impide su seguimiento estricto (necesitamos fijarnos en los documentos emitidos en la metrópoli, para tener una visión cabal del contexto comunicativo). Al tiempo que se recogían los textos o documentos, se iban preparando transcritos según normas paleográficas muy restrictivas (de acuerdo con la escuela argentina de Paleografía) para publicar una muestra representativa, y se iban haciendo análisis de distintos aspectos lingüísticos. Se primaba la parte grafofonética, fonológica y morfosintáctica. Y los

estudios, con datos de uso de lengua mediante ejemplos extraídos de los textos, eran de índole cualitativa.

En 1997 se propuso la idea de trabajar todos los equipos coordinados sobre léxico. Ya entonces se hablaba de trabajar con el programa informático TACT (*Text Analysis Computing Tools*), para medir el índice de frecuencias de las palabras. Pero tardó aún un tiempo, hasta el 2001, en incorporarse la idea de trabajar con más programas y herramientas informáticas que ayudaran a progresar de forma más rápida y exhaustiva en el análisis lexicográfico de estos textos.

3. FIJACIÓN DEL CORPUS

La fase inicial implica fijación y ordenamiento del corpus. De entre los muchos criterios posibles, hemos seguido tres: 1) diacrónico, 2) diatópico y 3) discursivo. Este último es el más complejo. En él se enlaza variación textual y variación discursiva, ya que en estos textos, según los principios instaurados por la ciencia auxiliar histórica de la diplomática (Real Díaz 1991 [1970]), la tipología se basa en la relación función/forma y varía de acuerdo con la materia tratada. Esta tipología textual trasladada al dominio del análisis del discurso se traduce en un tipo particular de variedad discursiva para cada asunto.

3.1. *Criterio cronológico*

El criterio diacrónico es el más evidente, ya que la meta es trazar el avance del cambio lingüístico en el tiempo. Tratamos de cubrir los tres siglos fragmentados en décadas. El corpus comienza con un documento dominicano de 1509-1510 como fecha de inicio y termina con un documento colombiano de 1803 como fecha final. En la Tabla 1 se puede ver el número de documentos reunidos por siglos y por países, con una extensión en folios muy desigual, desde aquellos con dos o tres folios a otros que tienen cerca de los trescientos folios.

PAÍS	Siglo XVI	Siglo XVII	Siglo XVIII
Santo Domingo	49	55	30
Colombia	144	70	58

Tabla 1. *Número de documentos reunidos en el corpus.*

Aumenta el interés el componente añadido de la posibilidad de comparación y contraste entre dos zonas, distintas pero no tanto que no podamos intuir que al menos en parte formen entre sí una unidad lingüística, bajo la denominación de zona caribeña, que toca de lleno a Santo Domingo así como a una buena parte de la llamada en orígenes Tierra Firme y que hoy día conforma la llamada costa atlántica del país colombiano (con Santa Marta y Cartagena de Indias históricamente descubiertas y pobladas desde Santo Domingo).

3.2. Criterios de selección y transcripción del corpus

El carácter inédito de los textos, ya que se trata de documentos recogidos directamente de los archivos, la mayoría de ellos transcritos por vez primera para este corpus, es otro factor de interés y relevancia de esta investigación. En cuanto al modo de transcripción, hemos optado por hacer una transcripción de tipo paleográfico y otra con un notable grado de modernización (Tabla 2), sobre todo en la indicación de pausas, oraciones y párrafos mediante puntuación actual y también con la supresión de algunas grafías superfluas (dobles consonantes meramente gráficas, <v> e <j> vocálicas, etc.). Este último modo de transcribir, sin entorpecer posibles trabajos de grafémica o fonético-fonológicos, facilita los análisis lingüísticos de niveles como el sintáctico o pragmalingüístico. Y también se muestra más útil para la futura edición electrónica prevista.

TRANSCRIPCIÓN PALEOGRÁFICA	PRESENTACIÓN CRÍTICA
<p>+</p> <p>yo el dean de sto. santo* domingo de la ysla española digo que lo que me parece de çiertos capitulos que por mandado de los señores. presidente. y oydores del Consejo de las yndias me fueron mostrados y mandado diese mi paresçer encargandome la conciencia es lo siguiente/.</p> <p>1+ al primero capitulo que dize que se den los yndios perpetuos digo que me paresçe bien y que es lo que conbiene al seruicio de dios y de vra magt vuestra* magestad* y bien y conservaçion de los dichos yndios y</p>	<p>+</p> <p>Yo, el deán de Santo Domingo de la isla Española, digo que lo que me parece de çiertos capítulos que, por mandado de los señores presidente y oidores del Consejo de las Indias, me fueron mostrados y mandado diese mi paresçer, encargándome la conciencia, es lo siguiente:</p> <p>1+ Al primero capítulo, que dice que se den los indios perpetuos, digo que me parece bien, y que es lo que conviene al servicio de Dios y de Vuestra Magestad y bien y conservación de los dichos indios y</p>

<p>tierra /.</p> <p>2 + al segundo que dize que el rrepartimiento de los yndios se haga lo mass breve que posible sea. digo que todo Remedio breve es el que conbiene y me paresçe que asy deve hazer</p> <p>3 + al terçero capitulo que dize que qualquiera provincia que se conquistare o de nuevo se poblare se Repartan las tierras e yndios a los conquistadores y a cada vno segun sus seruiçios y para ello se pongan las personas que el Capitulo dize y que no se los puedan quitar por ningun caso mi delito hasta que sea visto y determinado por su magt. magestad* digo que me paresçe que conviene/.</p> <p>4 + al quarto capitulo que dize que el ynteres de los pueblos de los yndios que a su magestad*. cupieren. se distribuyan y gasten en passar labradores a aquellas partes con todo lo demass que en este capitulo dize digo que me paresçe que conviene /.</p> <p>5+ al quinto capitulo digo que se deve de hazer y guardar como en el se que contiene* y que en ninguna manera se permita vender ni traspasar ningun Repartimiento de yndios /.</p>	<p>tierra.</p> <p>2+ Al segundo, que dice que el repartimiento de los indios se haga lo más breve que posible sea, digo que todo remedio breve es el que conviene y me parece que así debe hazer.</p> <p>3+ Al terçero capitulo, que dice que qualquiera provincia que se conquistare o de nuevo se poblare se repartan las tierras e indios a los conquistadores y a cada uno según sus servicios y para ello se pongan las personas que el capítulo dice y que no se los puedan quitar por ningún caso mi delito hasta que sea visto y determinado por Su Magestad, digo que me parece que conviene.</p> <p>4+ Al quarto capitulo, que dice que el interese de los pueblos de los indios que a Su Magestad cupieren, se distribuyan y gasten en pasar labradores a aquellas partes con todo lo demás que en este capitulo dice, digo que me parece que conviene.</p> <p>5+ Al quinto capítulo, digo que se debe de hazer y guardar como en él se contiene y que en ninguna manera se permita vender ni traspasar ningún repartimiento de indios.</p>
---	--

Tabla 2. Ejemplo de transcripción doble: SD-16-s/f 1529(ca), AGI, Indiferente General 1624, R.4, N° 7, folios 784r-784v.

3.3. Criterio diatópico

Además de seguir un criterio de datación temporal, hay que procurar abarcar diferentes focos urbanos de irradiación de documentos. Y así hemos hecho en Santo Domingo, con documentos fechados en las villas de Ázua, Santiago de los Caballeros, Concepción de la Vega, Boyacá, etc. y en el Nuevo Reino de Granada, con Santafé de Bogotá como centro, los territorios colindantes de la provincia de Santa Marta y las gobernaciones de Cartagena y Popayán, con numerosos focos de emisión de documentos, algunos de ellos fundados y desaparecidos en distintos momentos de la historia colonial o de independencia, otros hoy día con una situación política diferente, con fuertes cambios históricos,

como sucede con la fragmentación de la isla de La Española, hoy día dividida en dos países, Haití y República Dominicana, o como la cambiante y bastante reciente en el tiempo situación de frontera entre Colombia y Venezuela, en torno a la sierra de Los Andes.

3.4. Variedad de registros según variedad de textos

También, como ya hemos dicho antes, se pretende introducir en la organización del corpus y la clasificación de los textos un criterio de tipología textual. Los “documentos” indianos son un tipo especial de textos que, ligados a la esfera de la gobernación y administración de la América hispánica, portan testimonios históricos de primer orden. Los propios teóricos de la lingüística de corpus señalan que una conjunción que se muestra especialmente útil en el modo de actuar la lingüística de corpus es precisamente la que une tipo especial de discurso y análisis lexicográfico (Biber, Conrad y Reppen 2000:54).

Abreviatura	Tipología diplomática	Lugar expedición	Fecha expedición	Autor o escribano
Col.-16-1533	Carta (Informe)	San Miguel (Pop.)	11-11-1533	Sebastián de Belalcázar
Col.-16-1534	Probanza	Cartagena	10-12-1534	Diego de Santa Cruz (es.)
Col.-16-1539	Capitulación	Santafé de Bogotá	1539	Giménez de Quesada
Col.-16-1541	Ordenanzas	Cartagena	20-06-1541	Francisco Nieto (es.)
Col.-16-1541a	Carta (Informe)	Cartagena	20-08-1541	Cristóbal de Latosilla
Col.-16-1542	Carta (Informe)	Cartagena	19-04-1542	Alonso de Heredia
Col.-16-1543	Carta (Informe)	Santafé de Bogotá	30-09-1543	Rodrigo de Correa (es.)
Col.-16-1543a	Carta (Queja)	Tunja	17-11-1543	Escribano de cabildo
Col.-16-1544	Memorial	Anzerma (Pop.)	1544	Antonio de Oliva (es.)
Col.-16-1546	Carta (Informe)	Cartagena	02-01-1546	Miguel Díez Armendáriz
Col.-16-1547	Carta (Informe)	Santafé de Bogotá	15-02-1547	Justicia y Regimiento
Col.-16-1549	Carta (Informe)	Santa Marta	12-01-1549	Varios sacerdotes
Col.-16-1549a	Consulta (Interr.)	Cartagena	07-12-1549	Alonso Téllez (es.)
Col.-16-1553	Carta (Poder)	Santafé de Bogotá	28-11-1553	Escribano de cabildo
Col.-16-1561	InformaciónVisita	Cartagena	1560-1561	Diego Suárez (es.)
Col.-16-1567	Memorial	Río Hacha	15-12-1567	Lcdo. de Heredia (es.)
Col.-16-1577	Carta (Informe)	Santafé de Bogotá	08-02-1577	Luis Zapata (arzobispo)
Col.-16-1591	Acta fundación	Sevilla (NRG)	25-06-1591	Luis Pérez de Andrada (es.)
Col.-16-1599	Carta (Informe)	Cartagena	10-05-1599	Arzobispo Bartolomé
Col.-17-1604	Información	Santafé de Bogotá	19-06-1604	Gzo. Sez. Robledo (es.)

Tabla 3. Ficha de catalogación de los textos con reflejo de la tipología diplomática.

Tan solo con observar la segunda columna de la Tabla 3, donde se inserta la tipología diplomática de los documentos de Colombia del siglo XVI, nos damos cuenta del predominio de un tipo textual sobre los demás. Sin duda alguna, la carta es el “documento indiano por excelencia”, como han estudiado los diplomatas (Heredia 1985) y hemos corroborado desde la lingüística (Carrera 2005), al ser el eje de la comunicación entre dos realidades geográficas distantes, Europa y América. Es una línea de investigación en la que se ha de profundizar en el futuro. En la Tabla 4 se da una tipología muy sencilla con dos variables, emisor-destinatario y tipo de carta, como base de la interrelación:

EMISOR-DESTINATARIO	TIPO DE CARTA
Autoridad soberana-autoridades delegada	Carta real
Autoridades delegadas-autoridad soberana	Carta oficial
Particular-autoridad constituida	Particular
Particular-particular	Privada

Tabla 4. *Tipología de las cartas indianas (Heredia 1985:138-141).*

4. LEXICOGRAFÍA DIACRÓNICA AUTOMATIZADA

En el estudio de la historia documental del español de Santo Domingo y Colombia, una de las líneas en la que más esfuerzo se ha puesto es aquella que conecta lexicografía diacrónica³ con tratamiento electrónico de textos.

4.1. *Algunos referentes en el dominio hispánico*

La lexicografía diacrónica hispánica conoció una etapa de propuestas de renovación en los últimos 70 y los primeros 80 del siglo pasado. Surgieron entonces diferentes proyectos con la característica común de centrar su interés en la etapa medieval del español, con obras literarias como base del corpus y con tratamiento

³ Esta parte del proyecto está financiada por la Consejería de Educación de la JCYL (VA021/04).

automático de datos⁴. Fue, sin embargo, en fechas previas a este ambiente de renovación lexicográfica, a fines de los sesenta y principios de los setenta, cuando la documentación americana de archivo y otras fuentes subsidiarias de lengua escrita (crónicas, escritos de viajeros, etc.) se convirtió en el terreno más propicio para la aplicación de un método de análisis masivo de datos por ordenador. En el *Hispanic Seminary of Medieval Studies* de Madison (Wisconsin), Boyd-Bowman (1971, 1982, 1987) hizo una recopilación sistemática del léxico de este tipo de textos durante cuatro siglos de hispanización americana en una obra magnífica, de la cual hizo precisamente un análisis crítico Fontanella de Weinberg (1994:53-65). No cabe duda de que este proyecto ha inspirado otras propuestas que se han acometido con posterioridad.

En fechas más cercanas a nosotros contamos con la propuesta de Melis y Company (1996:1659-1666), dos discípulas de Lope Blanch y seguidoras de su propuesta (Lope Blanch 1979:105-112) de estudiar mediante documentos la evolución histórica del léxico del español de América y, más concretamente, del español de México. Este análisis se ha convertido en el primer inventario léxico de una zona hispanoamericana con base en documentos (Company y Melis 2002 [1998]). La idea de elaboración de lexicografía histórica con fuentes documentales como base ha fructificado en otra zona en el

⁴ Como ejemplos, el Proyecto *DEM* (*Diccionario del español medieval*) de Müller (1980:175-193) –en cuya exposición del proyecto se mencionan como referente los primeros frutos del *Hispanic Seminary of Medieval Studies* en Madison, Wisconsin, o el Proyecto *LMH* (*Léxique Médiéval Hispanique*) de Pottier (1980:194-247), quien afirmaba aún: “Nous nous réjouissons du projet de *Dictionnaire de l'espagnol médiéval* présenté ici-même par M. B. Müller. Il manque un instrument de travail cumulatif moderne” (1980:197). El Seminario de Madison comenzó su tarea de recopilación electrónica de léxico hispánico medieval en el ya lejano 1946, cuando se publicó el *Dictionary of the Old Spanish Language (DOSL)* y ha continuado editando su material léxico hasta la actualidad, con fines y propósitos distintos a aquellos pero en la misma línea de uso del soporte electrónico (Kasten y Cody 2001). Aún tardará la Real Academia en adoptar esa metodología de corpus aplicada a la recogida masiva de material lexicográfico, sobre todo desde la perspectiva diacrónica. De hecho se describía en aquel momento la situación de estancamiento y avance dificultoso del estudio del léxico histórico del español en el seno de la Academia (Alvar Ezquerro 1976:39). Esta situación ha cambiado un tanto desde el año 1994, momento de puesta en marcha por parte de la Real Academia Española de un eficaz instrumento para la recogida del léxico diacrónico de la lengua española, reconocido con las siglas *CORDE* (*Corpus Diacrónico del Español*), al cual hay que ir incorporando registros diferentes al estrictamente literario de las diferentes etapas del español.

español de las Islas Canarias, gracias a la labor conjunta y coordinada de los profesores Corrales y Corbella (2001), esta vez en una obra con el rango de diccionario, ya que contiene comentarios de los autores, alusiones a otros repertorios, marcas de procedencia y definiciones. Es un modelo para la tarea que aquí intentamos realizar, con la salvedad de que este diccionario llega hasta el siglo XX, período en el que se sirven de obras literarias costumbristas, históricas o etnográficas. Los volúmenes de México y de las Canarias enlazan directamente tanto con el Proyecto de Madison, llevado a cabo por Boyd-Bowman como con los objetivos previstos en el *Proyecto de Estudio Histórico del Español de América, Canarias y Andalucía* antes mencionado, porque sus autores se han inspirado en el impulso que se dio al estudio del componente léxico a través de documentos.

Hoy día, superados los escollos de tipo técnico, para conseguir una visión coherente del léxico, habría que intentar abarcar el mayor número de zonas hispanohablantes del amplio continente que es América e incluir áreas como Canarias y Andalucía (Frago 1999:74-100). Una corroboración de la idea de visión de conjunto y de análisis comparativo la encontramos en la presentación del diccionario histórico canario por parte de Corrales y Corbella (2000:190):

También desde el punto de vista de la propia lengua española como lengua histórica, estos diccionarios parciales de una determinada lengua funcional se muestran indispensables si se pretende conocer en toda su complejidad cómo fue el proceso de expansión atlántica de nuestra lengua (...) La noción misma de canarismo es una abstracción que no tiene razón de ser más que en comparación con otras variedades del español y, en este sentido, como indica Juan Antonio Frago, el criterio comparativo resulta indispensable “para que la identificación dialectal sea científicamente válida”.

4.2. *Estudios de concordancias: análisis del léxico y colocaciones*

En nuestra investigación la fase siguiente a la transcripción incluye las tareas de concordancia y lematización. Como ya he defendido con anterioridad (Carrera, en prensa), mi propuesta de análisis abarca todo tipo de vocabulario y he advertido también del peligro de que, ante la ingente tarea que supone transcribir, vaciar,

ordenar y catalogar los vocablos de un sinnúmero de “documentos” con los que contamos, nos dejemos llevar por el camino reductor de rastrear solo aquello que pueda considerarse específico, incluso si se trata de países como República Dominicana y Colombia, que destacan por la vitalidad de la lengua española y por la creación de nuevos términos que han aportado su especial fisonomía al español dominicano y colombiano.

De todos los programas de concordancias que existen disponemos del *Concordance*, hoy en su versión 3.2, el mismo utilizado en sus investigaciones lexicográficas, hace tiempo al menos, por el CILUS (*Centro de Investigaciones Lingüísticas de la Universidad de Salamanca*). Este programa trabaja con el texto en formato .txt para hacer las concordancias, plenas o en forma de lista rápida de frecuencia de aparición.

En ejemplos como *gouernacion*, en la Figura 1, se recogen variantes textuales que responden a un tipo de transcripción paleográfica, estricta, conservadora de rasgos gráficos, algunos relevantes, como el caso de la consonante bilabial con sus variantes , <u> y <v>, de interés grafemático, fonético y fonológico, otros como meras variantes grafémicas <c, ç> y, por último, casos con mantenimiento de las abreviaturas.

Context...	Word	...Context	L...
Y que luego que el soradeltantado se ...	gouernacion	y que les aga entender lo que an de ...	189
// En la çibdad de popayan desta proui...	gouernacion	de popayan a catorze dias del mes d...	304
(Yo Juan negrete de santander esrn...	gouernacion	es nueva mente poblada y ay mucha...	307
(Yo Juan negrete de santander esrn...	gouernacion	como cabeçera que [e]sta çibdad es ...	307
(Ansinismo que los v ^o s de los pueblo...	gouernacion	syenpre an sustentado10 a su costa ...	318
(Ansinismo que su magt no tiene en [...	gouernacion	se les haga muy creçidas mçds pues...	320
(Ansymismo que su mgt tiene fecha ...	gouernacion	y deve gozar y pues ella fue justa y ...	322
en las <clítico> dhas provisiones Real...	gouernacion	esta muy neçesitada de [e]spañoles ...	327
en las <clítico> dhas provisiones Real...	gouernacion	no tiene mas de veynte y siete v ^o s e l...	327
en las <clítico> dhas provisiones Real...	gouernacion	porque en ella no Residen ofiçales nj...	327
en las <clítico> dhas provisiones Real...	gouernacion	hasta tanto que su mgt sea ynformad...	327
Por la mucha priesa que me dieron los...	gouernacion	ha sido y a damian de peralta de peñ...	339
// Temo reçeuir mucho daño en reçeui...	gouernacion	para que juntandolos con veinte e çin...	341
v.magt. me manda que le de verdader...	gouernacion	por tenje de gouor. y capt. general y ...	349
v.magt. me manda que le de verdader...	gouernacion	que el los a echado a perder y puest...	349
// yn liçençiado llamado diego hernand...	gouernacion	el qual esta en antiochja y avnque a t...	353
(En la çibdad de cartagena a diez y n...	gouernacion	e otras aviendo enrntendido el notable...	487
(En la çibdad de cartagena a diez y n...	gouernacion	y de la de santa martha parezcan ant...	487
// En la çibdad de cartajena de la cost...	gouernacion	de cartajena para en esta ynformacio...	490
(Yo Luis Perez de andrada scriu ^o ma...	gouernacion	de santa marta y de las <clítico> dha...	562
// dixo que por quanto su md salio Por ...	gouernacion	y Prouinçia de santa marta por El Rey ...	565
// dixo que por quanto su md salio Por ...	gouernacion	Para cuyo efecto ansi deste <clítico> ...	565

Figura 1. Concordancias de “gouernacion” en textos colombianos.

Está comprobado que el mantenimiento de estas variantes provoca unas disfunciones serias en el momento del recuento de apariciones y en la fijación de las palabras claves, porque aparecen variantes como *gouern* (2), *gouernacion* (2), *gouernaon* (1), que hacen muy difícil la tarea de lematización y de selección de entradas en todos sus contextos. Se concluye que también para esta tarea de concordancia es preferible acudir a un tipo de transcripción con ciertas regularizaciones gráficas, sin repercusión en lo fonético, con puntuación actualizada, se pueden hacer concordancias sobre este tipo de texto. La ventaja de esta última decisión se verá en el momento de extraer los lemas, más fácil con este tipo de apariencia textual.

Context...	W...	...Context	Line
Dos «lema»cartas«lema» he Reçebido de v.s. con las quales he R...	haga	fe / y no «clítico» se de credito a tantas mentiras / tan fuera de Ra...	6
suplica a vra. mt. que en Remuneración de los serujios que ha he...	haga	md. de «clítico» le mandar dar salario con la «clítico» dha gouerna...	75
(noble señor diego de viana alld ordinario en esta çibdad de carta...	haga	fee para que el «clítico» dho governador «clítico» lo presente ante ...	82
Los capitulos que esta çibdad de santa fe y nuevo rreyno de gran...	haga	mçd son las sigujes	137
Yten supcan a su magt mande hazer e	haga	a su costa vn espital para los hombres nescitados y doljentes p...	141
Yten supcan a su magt	haga	mçd a esta çibdad que quando mandare rrepartir el «clítico» dho n...	144
Yten supcan a su magt que mande dar liçençia para que en esta «...	haga	vna fortaleza y en otras partes qual al governador «clítico» le pare...	145
// Yten supcan a su magt «clítico» les	haga	mçd que todos los que hallaren sepulturas en sus rrepartim ^o s o ot...	150
Yten supcan a su magt que por quanto por menos molestia de los ...	haga	mçd que la «clítico» dha encomienda tenga fuerza de rrepartim ^o s ...	156
Alonso de heredia thenje de governador por el adelantado Don p ^o ...	haga	mçd que de las condenaçiones que se an hecho e hizieren de lo q...	232
la Justjia y rregim ^o desta çibdad de santa fee del nuevo Regno d...	haga	tanto bien y mçd de lo confirmar e aver por bu ^o con mas todo lo q...	247
[Margen izdo.: Anzerma Ynstruçion] Primeramente suplicar a su ma...	haga	mçd de Reponer e dar por ningunas las ordenaçones que su magt f...	284
Yten pedir e suplicar a su magt nos	haga	mçd de nos dar los yndios perpetuos por tres vidas porque de otr...	286
Yten que su magt nos	haga	mçd de que el oro de minas que en nuestra tierra se sacare no no...	287
Yten hazer Relaçion a su magt como ay muchas personas que tien...	haga	mçd que del oro que dellas se sacare no llevar mas del quinto por ...	268
Yten suplicar a su magt nos	haga	mçd de nos dar liçen ^a e facultad para que en toda esta gouernacion podamo...	269
Yten que su magt nos	haga	mçd de nos dar liçen ^a e facultad para que de su Real caxa se pue...	290
Yten que su magt nos	haga	mçd a esta villa para alguna ayuda ypagar/ de las deudas que an h...	291
Yten que su magt nos	haga	mçd de que el cabildo e Regim ^o desta «clítico» dha villa pueda libr...	293
Yten que su magt nos	haga	mçd de dar liçen ^a e facultad para que si algunos de los conquistad...	294
Yten que su magt nos	haga	mçd a esta villa de quantos Regim ^o s perpetuos que sean de los v...	295
Yten que su magt nos	haga	mçd de mandar que si alg ^o s yndios de aquí adelante vacaren en e...	296
Yten que su magt nos	haga	mçd de mandar que ningun conquistador ni vez ^o desta «clítico» dh...	297
Yten que su magt nos	haga	mçd de mandar que qu ^o quier y ^o e conquistador desta «clítico» dh...	298
(Ansimismo que su magt no tiene en [estas partes otras fortaleza...	haga	muy creçidas mçds pues las mereçen por lo que [esta «clítico» dh...	320
// Temu reçequir mucho daño en reçequirlo toda la tierra que v.mt. fu...	haga	para paçificar la tierra donde no vaya çerrijo y no es justo que el ...	341
// Temu reçequir mucho daño en reçequirlo toda la tierra que v.mt. fu...	haga	luego pues esta en tanta nescisidad he mandado que se le de ad...	341
// En la villa de santa cruz de mopox: de la gouernacion de cartagen...	haga	buen tratamiento y provean de comida para la «clítico» dha jornad...	459
// En la villa de santa cruz de mopox: de la gouernacion de cartagen...	haga	fee y se ynbie a su mgd l para que vista provea lo que mas sea se...	459
En la çibdad de ssanta fee deste nuevo rreyno de granada martes ...	haga	todos los autos y diligencias judiciales y estrajudiciales que conve...	470
Sean quantos esta carta de poder vieren como nos la Justjia y R...	haga	bien y diligentemente todo lo que all ende esta «clítico» dha çibdad...	474
Primeramente que por quanto su magt tiene conçeçidas por termino...	haga	ni quebrante a esta çibdad los privilegios e mrd ^s que por su magt l...	519
que su magt les	haga	mrd de quatroçientos o quinientas liçençias de esclavos para que l...	529
[Margen izdo.: Peton] Don fernando çaçique de çipaquira y don alo...	haga	fee ynterponiendo en ellos vmd su autoridad y decreto judicial y p...	666
[Margen izdo.: Pet.on] Don fernando çaçique de çipaquira y don alo...	haga	fee y los demas traslados que hubieremos de menester para el ef...	673
en otra oçaçion A dado qta A V.Magd esta çjud de Sta ffee cabez...	Haga	nro. en vna plasa de asIENTO por el conosido zelo que tiene del ser...	714
Vra mag.d Por su Rl çedula de 25 de Hen ^o deste año me manda le ...	Haga	y assi es forosso hazerla yo por Ron en la man ^a sigte	762
Y porque e entendido que de tpos atras a ynstançia de los gouern...	Haga	el «clítico» dho creçimiento de soldados y que Vra magd tenga por...	765
La Çibdad de Sancta fee cabeça del nueuo R.no de Granada haue...	haga	merçed particular mente en cosas donde exerçite su talento y çelo...	823

Figura 2. Forma verbal “haga” en textos colombianos.

Este programa proporciona un contexto anterior y posterior al vocablo seleccionado muy amplio, como se ve en la Figura 2, con la forma verbal *haga*. Además de forma verbal lematizable bajo el infinitivo *hacer*, este tipo de concordancias, mediante el análisis de las colocaciones, esto es, de los vocablos inmediatamente anteriores y posteriores a la palabra clave, permite el salto a un análisis proposicional, con una cuantificación simple, por ejemplo, del tipo

de OD que porta, de forma que de los 39 ítems que se recogen en esta pantalla 19 tienen *merced* como vocablo inmediatamente posterior: la locución *haga merced* es abrumadora en esta selección textual y puede llegar a ser un rasgo lingüístico significativo en la variedad de lengua que tratamos de estudiar.

El lematizador de este programa, del que proporcionamos una muestra en la Figura 3, está poco automatizado. La revisión, búsqueda e introducción de las formas diferentes de la palabra, que el programa trata como distintas, debe hacerse manualmente y consignar la forma base de una palabra en el correspondiente archivo *lemma, desechando cambios gramaticales tales como tiempo o pluralidad. Carece de la posibilidad de recuento total de la frecuencia del lema.

El lema se introducirá modernizado. Las variantes de los textos se conservarán entre paréntesis como parte de la entrada léxica. Esto no coincide con la manera de proceder sobre este asunto por Boyd-Bowman (1971, 1987), quien coloca entradas del tipo *eçeuto* (*excepto*), *nueveçientos*, *gouernación* al lado de *gubernación*.

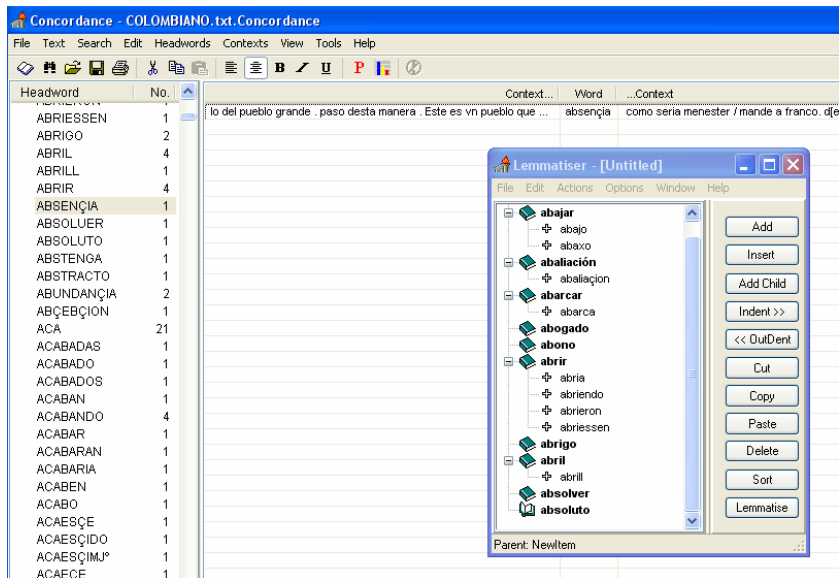


Figura 3. Captura de pantalla del archivo *lemma de documentos de Colombia.

5. PRIMERAS PRUEBAS DE ETIQUETADO ELECTRÓNICO

Al buscar herramientas con las que organizar el corpus y elaborar el proyectado diccionario, nos encontramos con los avances actuales en modelos de tratamiento de textos y documentos, y nos dimos cuenta de que todas las tareas previamente programadas podrían concentrarse en el marco de una edición digital. No obstante, hay que precisar el sentido en el que empleamos el concepto de edición digital. En este caso no se trata de informática aplicada a la crítica textual, una perspectiva poco adecuada al tipo de corpus con el que trabajamos, textos de índole no literaria, sin ediciones previas ni repercusiones en otros textos, es más, concebidos sin afán de publicación, sin el carácter de escritos para un lector, etc.

En este caso, la incursión en el mundo de la edición digital con el corpus de “documentos” del que hemos hablado puede ayudar a compartir la información recogida con otros equipos y con otros centros universitarios que trabajen con el mismo tipo de editores de textos⁵, capaces de proporcionar el máximo nivel de estándar y de posibilidad de intercambio.

También se plantea la edición digital como una herramienta útil para la vertiente de catalogación así como para los procedimientos de análisis textual, esto último la meta más alejada y laboriosa de alcanzar.

Tras diversas pruebas de etiquetado de textos para análisis sintácticos, si bien con un etiquetado no estándar, en concreto para el estudio de los clíticos (Carrera 2003), hemos comenzado la etapa de edición con la catalogación sistemática de los textos transcritos. Contamos con un editor de textos en formato XML/XSLT, con diversas plantillas. De ellas nos interesan de forma especial dos: *TEI P5 corpus* y *TEI P5 dictionaries*.

Con el primer tipo de plantilla hemos abierto un proyecto *teiCorpus* como cabecera de la edición electrónica del corpus (*vid.* Figura 4).

⁵ Para avanzar en la buena dirección, nos hemos puesto en contacto con otros grupos más consolidados en el área de la lingüística computacional, como el de la Universidad de Deusto, con distintos proyectos de marcación informática y edición digital de textos en su haber (Abaitua e Isasi 2003).

```

<?xml version="1.0" encoding="UTF-8"?>
<?oxygen                                RNGSchema="http://www.tei-
c.org/release/xml/tei/custom/schema/relaxng/tei_corpus.rng" type="xml"?>
<teiCorpus xmlns="http://www.tei-c.org/ns/1.0">
  <TEI xmlns="http://www.tei-c.org/ns/1.0">
    <teiHeader>
      <fileDesc>
        <titleStmt>
          <title> ORGANIZACIÓN DEL LÉXICO EN UN CORPUS DOCUMENTAL PARA
            EL ESTUDIO DEL ESPAÑOL DE AMÉRICA: SANTO DOMINGO Y
            COLOMBIA </title>
          <author> GRUPO PARA EL ESTUDIO DE LA HISTORIA LINGÜÍSTICA
            IBEROAMERICANA <!-- ... --> </author>
          <sponsor> UNIVERSIDAD DE VALLADOLID </sponsor>
          <funder> JUNTA DE CASTILLA Y LEÓN (VA021/04) </funder>
          <principal> MICAELA CARRERA DE LA RED </principal>
        </titleStmt>
        <editionStmt>
          <edition> Transcripción paleográfica y transcripción de imágenes de
            documentos digitalizados y fotocopiados en los
            archivos </edition>
          <source> <p> Catalogación según referencia archivística </p>
          </source>
          <!-- ... -->
        </editionStmt>
      </fileDesc>
    </teiHeader>
  </teiCorpus>

```

Figura 4. Cabecera de l Proyecto, tipo <teiCorpus>.

La gramática o DTD (*Document Type Definition* o Declaración de Tipo de Documento) que permita describir la estructura de textos o documentos de la naturaleza de los que aquí se han descrito precisa de una gran dosis de revisión y de adaptación del etiquetado, para que pueda ser aceptado como verdadero y a la vez cubrir los objetivos previstos para el análisis. Hemos hecho diversas pruebas de anotación de estructura de diferentes textos. Aquí ofrecemos dos, para discusión entre expertos:

- 1) el primero del siglo XVI, en concreto de 1520, fechado en Santa María de la Antigua del Darién, es una carta-informe perteneciente al expediente archivístico de la residencia de Pedrarias Dávila;
- 2) el segundo del siglo XVII, de 1694, es una carta-informe del sargento Alonso Cortés, militar que sofocó las sublevaciones de negros cimarrones y de servicio de la ciudad de Cartagena de Indias, perteneciente a unos autos contra un mulato declarado traidor.

En la Figura 5 se ve la cabecera TEI (<teiHeader>), con el apartado de descripción del archivo (<fileDesc>). Entre otras, señalamos la duda de la etiqueta <title>, que podría reemplazarse por <subject>, porque se refiere al contenido del texto y en los atributos se especifica “resumen”.

```

<TEI xml:id="MyTextNumber4" >
  <teiHeader >
    <fileDesc >
      <titleStmt >
        <title sameAs="NomenclColec"> Col-16-1520a </title>
        <title type="resumen">Residencia de Pedrarias Dávila: Carta
          de Rodríguez de Alarconcillo
        </title>
        <author xml:id="autégrafa">Rodríguez de Alarconcillo
          <!-- ... -->
        </author>
      </titleStmt >
      <editionStmt >
        <msIdentifier >
          <settlement > Sevilla </settlement >
          <repository > A.G.I. </repository >
          <altIdentifier > Patronato, 197, R.1 </altIdentifier >
        </msIdentifier >
      </editionStmt >
      <history > <p>Documento parte del expediente de la residencia
        tomada a Pedrarias Dávila, emprendida por
        mandato de las autoridades
        metropolitanas </p> </history >
      <origin > <p>Escrito en la ciudad de Santa María La Antigua
        del Darién </p> </origin >
      <doc: date value="año-mes-día"> 1520-07-07 </doc: date >
      <recordHist >
        <source > <p>Catalogado directamente desde el manuscrito
          original </p> </source >
      </recordHist >
    </fileDesc >
  </teiHeader >
  <!-- ... -->
  </p >
</sourceDesc >
</fileDesc >
</teiHeader >

```

Figura 5. Cabecera de un texto colombiano de 1520.

En la Figura 6 hemos puesto una síntesis de la estructura total del mismo texto, con la división tripartita (<front>, <body> y <back>) que se adapta bien a la estructura de una carta.

```

<text>
  <pb xml:id="folio1r">
    <front>
      <div type="dirección">
        <p> <?> + </?>
          Cesaria y catolica majestades
          y muy altos y muy poderosos señores
        <!-- ... -->
      </p>
    </div>
  </front>
  <body type="dispositivo">
    <div1 type=ack corresp="acatamiento"><p>El liçençiado
    Johan Rodriguez de Alarçonçillo, vezino de la çibdad de Cordova, veso
    los pies y muy reales manos de V.M.</p></div1>, <div2
    type="motivación"><p>a la qual suplico plega saver que, visto por
    Lope de Sosa, que Dios haya, como v.a. le havra proveido del cargo de
    estos sus Reinos de Castilla del Oro, ovo por vien de me traer para alcalde
    mayor dellos y para tomar la residençia que v.a. mandaria que se tomasse
    a Pedrarias Davila e sus ofiçiales e a los ofiçiales de v.a.,[...] y en
    remedialla e proueerla lo mejor que pudo. habló a Johan Alonso de Sosa,
    el sobrino y teniente de Lope de Sosa, para que estuviese en el mismo
    cargo que aya, y ansi se hizo, e asentole el salario.</p></div2> [...]
    Suplico a v.a. lo mande proveer y remediar por manera que yo me pueda
    sostener en su real serviçio.
  </body>
  <back>
    <div type=ack corresp="fórmuladespedida">Nuestro Señor la cesárea y catolicas
    magestades de vuestra alteza guarde y prospere bien aventuradamente con
    acreçentamiento de más reinos y señoríos.</div>
    <dateline>De la çibdad de Santa María de la Antigua del
    Darien vii de Junio de 1520 años</dateline>
    <div type=ack corresp="fórmula acatamiento">de vuestra imperial
    y real magestad muy humile servidor y vasallo </div>
    <signed> El Licenciado de Alarçonçillo</signed>
  </back>
  </pb>
</text>
</TEI>

```

Figura 6. Anotaciones estructurales de una carta-informe del siglo XVI.

La Figura 7 contiene un fragmento de la parte central (<body>) de la carta-informe del sargento Alonso Cortés, de 1694. En este caso, hemos ido instalando etiquetas de diferentes modos de abreviar en los documentos, un punto que creemos puede ser de gran utilidad para textos manuscritos. Bien es cierto que quizás este tipo de etiquetado resulte extremadamente complejo y satura el aspecto del texto, dejando poco espacio para otro etiquetado de cualquier tipo, aspecto que también aconsejan los expertos cuidar.

```
<body TEIform="carta">
  <div part="protocolo"> <?> + </?>
    <salute> Señor </salute> </p>
  </div>
  <div1 part="dispositivo">
    <pb id='1r'>En Primero de Julio de <reg resp="vocal">e</reg> ste año puso
    en la <abbr type=suspension="ea" , expan="Real">Rl.</abbr> noticia de
    <abbr type=title, expan="Vuestra Magestad">V.M.</abbr> <abbr
    type=suspension="en", expan="Sarxento">Sarxto</abbr> <abbr
    type=suspension="era" , expan="General"> Genl.</abbr> de Batalla <abbr
    type=title , expan="Don"> Dn.</abbr> Martín de Çevallos y la Zerda, <abbr
    type= initial , expan="que">q</abbr> governava esta plaza, los justos rezelos
    en que estava, viendo respirar el fuego de los ynsendios passados que tanto an dado
    que hazer ponderando la dificultad de su averiguación juridica solizita mi
    ocupassion poner en la <abbr type=suspension="ea" ,
    expan="Real">Rl.</abbr> noticia de <abbr type=title , expan="Vuestra
    Magestad">V.M.</abbr> el que estos rezelos ban passado a priesa a evidenzia
    de algun riesgo, volbiendose a ensender. Y oy se hubiera ensendido si no le uviera
    atajado a tiempo, por el cuydado con que me <diacritics
    id=acirc>â</diacritics> tenido la pretenzion que <abbr
    expan="dicho">â</abbr> Governador apunta en su representación sobre
    querer marchar con picas los Capitanes de las <abbr type= superscription="as" ,
    expan="Compañías">Compas.</abbr> de Pardos libres de vatallón y milizia a
    ymitazion de los del Presidio y demas <abbr type=superscription="nes",
    expan="Cappitanes">Cappnes.</abbr> de españoles, y aunque por entonzes se
    les aparto de <reg resp="vocal">e</reg> ste yntento</pb> <pb
    id="1v">con lo apretado de la orden que tube para hazerles marchasen en la forma
    que se hauia acostumbrado y mandandoles no hablasen en otra, Y aunque
    obedezieron, no fue tan enteramente que no lo dejase de hazer el <abbr
    type=superscription="n" , expan="Capitan" > Cappn.</abbr> de Pardos libres
    Juan Bezerra con el fribolo pretexto de dezir estaua enfermo no estandolo , pues se
    paseaua, y asistio al Puesto <abbr type=initial , expan="que">q</abbr>
    ocupaua su <abbr type= superscription="a" , expan="Compañía "
    >Compa.</abbr> fue preziso por entonzes tolerar admitiendo la escusa y esta
    toleranzia que le hauia de seruir de <abbr type=superscription="or" , expan=
    "mayor">mor.</abbr> estímulo para apartarse de las ocaçiones que pudieran
    perturbar la vnion entre los militares y <abbr type=superscription="nos" ,
    expan="vezinos"> veznos.</abbr> cedio osadia para sentarse en una silla de las
```

que estauan en el cuerpo de Guardia Prinzipal que e puesto en la Puerta del Puente por justos motiuos del <abbr type=suspension="çi", expan="seruiçio">seruio.</abbr> de <abbr type=title, expan="Vuestra Magestad">V.M.</abbr> Y pidiendo la silla el Alferes de la guardia y yendo el Cauo de esquadra de ella a pedirlas y dadola el Alferes <abbr type=title, expan="Don">Dn. Juan Polo del Aguila, quien hauia llamado a este <abbr type=superscription="n", expan="Capitan">Cappn.</abbr> para que se sentase, no vasto esta demostrazion para que la diese ademas de las ynstanzias que sobre darla le hizo, de que resultado tener alguna question con el Cauo de esquadra de guardia </pb> <pb id="2r">Passando à alguna demostrazion yndebida <diacritics id=acirc>â</diacritics> que acudio el <abbr type=suspension="en" type=superscription="to", expan="Sargento">Sargto. con su ynsignia y el Alferes con la espada <diacritics id=acirc>â</diacritics> atajar el desacato que a uista de las Banderas de <abbr type=title, expan="Vuestra Magestad">V.M.</abbr> se hazia, de que se siguió darle el Alferes al <abbr expan="dicho">dho</abbr> <abbr type=superscription="n", expan="Capitan">Cappn.</abbr> Juan Bezerra una cuchillada en la caueza, y llegando el <abbr type=superscription="n" , expan="Capitan">Cappn.</abbr> de la Guardia <abbr type=title, expan="Don">Dn.</abbr> Diego Beltran de Cayzedo que al presente se hallaua en la plaza de armas [...]

Figura 7. Propuesta de etiquetado de las abreviaturas y de otros signos diacríticos del texto de 1694.

Hemos probado también la utilidad del editor de XML/XSLT para realizar búsquedas lexicográficas. Aún no hemos tenido ocasión de utilizar la función de lematización, que sabemos que tiene esta herramienta. Como muestra la palabra *seruiçio* en el texto colombiano de 1520, en la Figura 8.

Project

- Col-corpus.xpr
- Col-16-1520.txt
- Col-16-1520.xml

```

54 <front>
55 <div type="dirección">
56 <p>
57 +
58 Cesaria y catolica majestades
59 y muy altos y muy poderosos señores
60 <!-- ... -->
61 </p>
62 </div>
63 <front>
64 <body type="dispositivo">
65 <div1 type="ack corresp="acatamiento"><p>El liçenciado Johan Rodríguez de Alarconçill
66 veso los pies y muy reales manos de V.M.</div1>, <div2 type="motivación">a la qual sup
67 como v.a. le havra proveido del cargo de estos sus Reinos de Castilla del Oro,
68 a la qual suplico plega saver que, visto por Lope de Sosa, que Dios haya,
69 como v.a. le havra proveido del cargo de estos sus Reinos de Castilla del Oro,
70 ovo por vien de me traer para alcalde mayor dellos
71 y para tomar la residencia que v.a. mandaria que se tomasse a Pedrarias Davila e si
72 y por ser Lope de Sosa natural de Cordova y persona
73 </div2>

```

Info	Description - 6 items	Resource
-	y por ser el viaje en seruiçio de V.M.,	Col-16-1520.xml
-	<p> Las otras cosas de estos reinos que al real seruiçio de v.a.	Col-16-1520.xml
-	estén en la horden policía e buen Regimiento que deven e como al seruiçio de v.a. conviene	Col-16-1520.xml
-	puede tener tanta noticia de las cosas de acá como Pedrarias ni mayor respecto a su real seruiçio	Col-16-1520.xml
-	a seruiçio de v.mt. y en ellas tendra grand <pb>	Col-16-1520.xml
-	por manera que yo me pueda sostener en su real seruiçio.	Col-16-1520.xml

Figura 8. Concordancia del vocablo "seruiçio".

Y por último, la Tabla 5 contiene una ficha de los aspectos que podrían tomarse en cuenta para un etiquetado TEI del estudio discursivo y pragmático. La ficha está elaborada sobre un documento de 1530, una carta-informe de García de Lerma, gobernador de Santa Marta, dictada a un escribano, dirigida al presidente y oidores de la Audiencia de Santo Domingo, escrita a raíz de la recepción de otra carta en la que se le exigía explicaciones sobre lo acontecido con una entrada en el río Magdalena.

```

< documento > Col-16-1530 < /documento >
< definición diplomática > Carta-informe < /definición >
< autor id= "escribano" > heterógrafa < /autor >
< emisor id= "autoridad-delegada" > gobernador de Santa Marta < /emisor >
< receptor id= "destinatario" > Real Audiencia de Santo Domingo < /receptor >
< tratamiento > V.S. (Vuestra Señoría), "a esos señores" < /tratamiento >
< dirección id= "fórmula de intitulación" > Reverendisimo y muy magnifico
Señor < /dirección >
< ámbito discursivo de producción > institucional/oficial < /ámbito >
< registro > discurso epistolar < /registro >
< modo discursivo > narrativo/argumentativo < /modo >
< acto comunicativo > discurso dialógico/respuesta < /acto >
< acto de habla perlocutivo > dar quenta a v.s. de todo lo que ha pasado < /acto
habla >

```

Tabla 5. *Algunos aspectos textuales y discursivos para convertir en etiquetas TEI.*

6. CONSIDERACIONES FINALES

Hasta aquí algunas reflexiones surgidas de nuestra ya dilatada experiencia y dedicación a los textos calificados como "de Indias". Para nosotros las herramientas informáticas son nada más (y nada menos) que eso, herramientas. Nos han de ayudar a sistematizar nuestro modo de acercarnos al abigarrado mundo de la diplomática y de la historia lingüística. Los lenguajes informáticos, las siglas, los programas deben servir de acicate para comprobar si somos capaces de lograr arrancarles el máximo de sus funciones, pero no nos han de hacer creer que hemos de abandonar las ancilares e imprescindibles tareas de acudir a los archivos, desempolvar aún muchos legajos, leer sus letras ininteligibles a veces, escudriñar entre sus manchas y agujeros, saborear su estilo discursivo y atrapar las estructuras y vocablos de una lengua que germinará en las variedades del español colombiano y dominicano. Ni más ni menos.

REFERENCIAS BIBLIOGRÁFICAS

- ABAITUA, J. e ISASI, C. (eds.) (2003): *La edición digital*, Letras de Deusto, 100/vol.33.
- ALVAR EZQUERRA, M. (1976): *Proyecto de lexicografía española*, Barcelona: Planeta.
- BIBER, D. y FINEGAN, E. (2001): “Diachronic relations among speech-based and written registers in English”, en S. Conrad y D. Biber (eds.), *Variation in English: Multi-dimensional Studies*, Edimburgo: Longman, 13-42.
- BIBER, D.; CONRAD, S. y REPPEN, R. (2000 [1998¹]): *Corpus Linguistics. Investigating Language Structure and Use*, Cambridge: Cambridge University Press.
- BOYD-BOWMAN, P. (1971): *Léxico hispanoamericano del siglo XVI*, Londres: Tamesis Books.
- BOYD-BOWMAN, P. (1982): “El léxico hispanoamericano del siglo XVI”, en E. de Bustos Tovar (ed.), *Actas del IV Congreso Internacional de Hispanistas*, Salamanca: Universidad de Salamanca, 191-198.
- BOYD-BOWMAN, P. (1987): *Léxico hispanoamericano del siglo XVI*, Madison: Hispanic Seminary of Medieval Studies (microfichas).
- CARRERA DE LA RED, M. (2000): “Reino de Nueva Granada. Siglos XVI a XVIII”, en E. Rojas Mayer (comp.), *Documentos para la historia lingüística de Hispanoamérica. Siglo XVI a XVIII*, Madrid, BRAE, Anejo LVIII, 304-425.
- CARRERA DE LA RED, M. (2003): “Los clíticos en textos colombianos de los siglos XVI al XVIII: evolución del orden y función”, en F. Sánchez Miret (ed.), *Actas del XXIII Congreso Internacional de Lingüística y Filología Románica (Salamanca 2001)*, Tübingen: Niemeyer, II/1, 193-208.
- CARRERA DE LA RED, M. (2005): “Análisis de situaciones comunicativas en el documento indiano por excelencia: la carta”, en M. Sedano, A. Bolívar y M. Shiro (eds.), *Haciendo lingüística. Homenaje a Paola Bentivoglio*, Caracas: Universidad Central de Venezuela, 625-641.
- CARRERA DE LA RED, M. (2006): “Escribir para gobernar: análisis pragmalingüístico del ‘discurso diplomático’ en la etapa colonial española”, en M. Casado, R. González y M.^a V. Romero (eds.), *Análisis del discurso: lengua, cultura y valores. Actas del I Congreso Internacional (Universidad de Navarra, Pamplona, noviembre de 2002)*, Madrid: Arco/Libros, II, 1711-1732.
- CARRERA DE LA RED, M. (en prensa): “Hacia una organización del léxico colombiano extraído de documentos. Siglos XVI a XVIII”, *AFA, Homenaje a Manuel Alvar*.

- CARRERA DE LA RED, M. y ZAMORA SALAMANCA, F. J. (1993): "Santo Domingo. Siglos XVI y XVII", en M^a. B. Fontanella de Weinberg (comp.), *Documentos para la historia lingüística de Hispanoamérica. Siglo XVI a XVIII*, Madrid, BRAE, Anejo LIII, 7-71.
- CLAVERÍA NADAL, G. y TORRUELLA I CASAÑAS, J. (2005): "Base de datos para un corpus de documentaciones léxicas", en J. Kabatek, C. D. Pusch y W. Raible (eds.), *Romanistische Korpuslinguistik II: Korpora und diachrone Sprachwissenschaft (Romance Corpus Linguistics II: Corpora and Diachronic Linguistics)*, Tübingen: Gunter Narr, 215-228.
- CLAVERÍA, G.; SÁNCHEZ, C. y TORRUELLA, J. (1996): "Sobre la informatización del DCECH", en A. Alonso et al. (eds.), *Actas del III Congreso Internacional de Historia de la Lengua Española*, Madrid: Arco/Libros, 2, 1631-1645.
- COMPANY, C. y MELIS, CH. (2002 [1998]): *Léxico histórico del español de México. Régimen, clases funcionales, usos sintácticos, frecuencias y variación gráfica*, México: Universidad Autónoma de México, 2 vols. (un volumen en 2002).
- CONRAD, S. y BIBER, D. (2001): "Multi-dimensional methodology and the dimensions of register variation in English", en S. Conrad y D. Biber (eds.), *Variation in English: Multi-dimensional Studies*, Edimburgo: Longman, 66-83.
- CORRALES, C. y CORBELLA, D. (2000): "Sobre el Diccionario histórico del español de Canarias (DHECan)", en C. Corrales y D. Corbella (eds.), *Estudios de Dialectología dedicados a Manuel Alvar con motivo del XL aniversario de la publicación de El español hablado en Tenerife*, La Laguna: Instituto de Estudios Canarios, 187-205.
- CORRALES, C. y CORBELLA, D. (2001): *Diccionario histórico del español de Canarias (DHECan)*, Tenerife: Instituto de Estudios Canarios.
- FONTANELLA DE WEINBERG, M. B. (1994): "El léxico de cuatro siglos de vida americana", *Boletín de la Academia Argentina de Letras*, LIX (enero-junio), 53-65.
- FONTANELLA DE WEINBERG, M.^a B. (comp.) (1993): *Documentos para la historia lingüística de Hispanoamérica. Siglo XVI a XVIII*, Madrid: BRAE, Anejo LIII.
- FRAGO GRACIA, J. A. (1999): *Historia del español de América*, Madrid: Gredos.
- HEREDIA HERRERA, A. (1985): *Recopilación de estudios de diplomática indiana*, Sevilla: Diputación Provincial.
- KABATEK, J.; PUSCH, C. D. y RAIBLE, W. (eds.) (2005): *Romanistische Korpuslinguistik II: Korpora und diachrone Sprachwissenschaft (Romance Corpus Linguistics II: Corpora and Diachronic Linguistics)*, Tübingen: Gunter Narr.

- KASTEN, LL. A. y CODY, F. J. (comps.) (2001): *Tentative Dictionary of Medieval Spanish (second edition, greatly expanded)*, Nueva York: Hispanic Seminary of Medieval Studies.
- KYTÖ, M. (comp.) (1996 [1991]): *Manual to the Diachronic Part of the Helsinki Corpus of English Texts: Coding Conventions and Lists of Source Texts*, Helsinki: Department of English, University of Helsinki, (3ª ed.).
- LAWLER, J. y DRY, H. A. (1998): *Using Computers in Linguistics. A Practical Guide*, Londres/Nueva York: Routledge.
- LOPE BLANCH, J. M. (1979): "Tareas más urgentes de la lingüística iberoamericana", *Atas do Simposio de PILEI*, São Paulo, 105-112.
- MELIS, CH. y COMPANY, C. (1996): "Incorporación de la informática a un proyecto de lingüística: el léxico histórico de la Nueva España", en A. Alonso *et al.* (eds.), *Actas del III Congreso Internacional de Historia de la Lengua Española*, Madrid: Arco/Libros, 2, 1659-1666.
- MÜLLER, B. (1980): "El proyecto de un diccionario del español medieval (DEM) y el estado de la investigación en el campo del léxico del español antiguo", *Cahiers de Linguistique Hispanique Médiévale*, 5, 175-194.
- NURMI, A. (2002): "Does size Matter? The Corpus of Early English Correspondence and Its Sampler", en H. Raumlönn-Brunberg *et al.* (eds.), *Variation Past and Present. VARIENG studies on English for Terttu Nevalainen* (Mémoires de la Société Néophilologique de Helsinki 61), Helsinki: Société Néophilologique, 173-184.
- POTTIER, A. (1980): "Léxique médiéval hispanique", *Cahiers de Linguistique Hispanique Médiévale*, 5, 195-247.
- PUSH, C. D. y RAIBLE, W. (2002): *Romance Corpus Linguistics and Spoken Language Studies*, Tübingen: Gunter Narr.
- RAUMLÖNN-BRUNBERG, H. *et al.* (eds.) (2002): *Variation Past and Present. VARIENG studies on English for Terttu Nevalainen* (Mémoires de la Société Néophilologique de Helsinki), Helsinki: Société Néophilologique.
- REAL DÍAZ, J. J. (1991 [1970]): *Estudio diplomático del documento indiano*, Madrid: Dirección de Archivos Estatales.
- ROJAS MAYER, E. (comp.) (2000): *Documentos para la historia lingüística de Hispanoamérica. Siglo XVI a XVIII*, Madrid: BRAE, Anejo LVIII.
- SÁNCHEZ, A. (2004): "Gestión informática de grandes recopilaciones de textos y extracción de información", en P. Fuertes (coord.), *Lengua y sociedad. Investigaciones recientes en lingüística aplicada*, Valladolid: Servicio de Publicaciones de la Universidad, 333-367.
- SINCLAIR, J. M. (ed.) (1987): *Looking up. An account of the Cobuild Project in lexical computing*, Londres: Collins.

WESCH, A. (2005): “Spanish documents of the 15th, 16th, and 17th centuries from a pragmatic point of view”, en J. Kabatek, C. D. Pusch y W. Raible (eds.), *Romanistische Korpuslinguistik II: Korpora und diachrone Sprachwissenschaft (Romance Corpus Linguistics II: Corpora and Diachronic Linguistics)*, Tübingen: Gunter Narr, 519-530.