

Cómo editar textos coloniales*

How to edit colonial documents

Manuel Contreras S.

Universidad Austral de Chile, Instituto de Lingüística y Literatura,
Apartado postal 567, Valdivia, Chile,
e-mail: manuelcontreras@uach.cl

Este artículo analiza la constitución de un corpus diacrónico del español de Chile (CorDECh), con el objeto de situar la labor de edición de textos coloniales. Se presenta una breve discusión sobre los métodos de lectura de textos en soporte de microforma, así como las implicancias del desarrollo tecnológico en la interpretación paleográfica, para lo cual se comparan y establecen criterios de transcripción del corpus; asimismo, se discuten algunos aspectos de la automatización del corpus y las implicancias y proyecciones para el análisis del mismo y se repasan algunos criterios de etiquetación, de acuerdo con los aspectos de la Text Encoding Initiative (TEI), adoptados por la RAE en el *Prontuario de Marcación SGML*, utilizado para el CORDE (Corpus Diacrónico del Español).

Palabras clave: edición de textos coloniales, filología hispánica, corpus diacrónico.

The conformation of Hispanic diachronic corpora is analysed in this paper with the purpose of describing the framework for the editing of colonial texts. A brief discussion of the methods of reading text in microform is presented here, as are the implications of technological development in palaeographic interpretation. Corpus transcription criteria are compared and established; in the same way, some aspects of corpora automatization and the implications and projections for its analysis are discussed here, and some criteria for corpus labelling are reviewed, according to the Text Encoding Initiative (TEI), adopted for the RAE in the *Prontuario de Marcación SGML*, used for the CORDE (Spanish Language Diachronic Corpus).

Key words: colonial text editing, Hispanic philology, diachronic corpus.

INTRODUCCIÓN

Cada vez que miramos hacia nuestros orígenes, en busca de las bases fundamentales de nuestra identidad, en cualquier sentido que entendamos este concepto, ineludiblemente tenemos que apelar al discurso escrito que testimonia este surgimiento, a fin de recurrir a las fuentes de primera mano que dan cuenta de la elaboración de ese universo primario, que da paso a la constitución –en nuestro caso– de la sociedad colonial del Reino de Chile.

* Este artículo es parte de los resultados obtenidos a través del proyecto Fondecyt N° 1040072.

Desde el punto de vista de los estudios documentales podría pensarse que la tradición historiográfica nacional ya ha dicho lo suficiente; sin embargo, pocos son quienes han perseguido el desarrollo de estas raíces desde sus bases, es decir, desde un texto distinto a los oficiales –reales cédulas, ordenanzas, bandos, etc.– no por descuido ni menosprecio de otros elementos históricos, sino por una dificultad objetiva en la lectura y transcripción de la documentación existente, sobre todo la que nos habla de aquellos nacientes días de este Reino, a mediados del siglo XVI y principios del XVII. En este sentido, como no, destacan los trabajos de Sergio Villalobos, quien nos entrega una visión de la sociedad colonial recurriendo a fuentes originarias invaluable.

Por otra parte, el ámbito de la filología nos ha entregado, por supuesto, una importante contribución al conocimiento de esta protoindianidad, aportando con la edición de innumerables textos coloniales, así como estudios importantes sobre la constitución del dialecto en estas tierras. Las normas para el trabajo filológico se fijaron ya, al menos en nuestro país, con el importante trabajo desarrollado por Mario Ferreccio en la *Biblioteca Antigua Chilena* (BACH), cuyos pasos han sido continuados felizmente por Raïssa Kordić, quien recientemente ha publicado el *Epistolario* de Sor Dolores Peña y Lillo.

Desde el campo de la lingüística, ha habido un retraso considerable si establecemos comparaciones con otros países que ya han desarrollado importantes conjuntos textuales diacrónicos y, sobre todo, considerando los avances de la lingüística del corpus. Si bien en un primer momento hubo un importante avance con la publicación en 1994 del capítulo de “Santiago de Chile”, en la colección de *Documentos para la Historia Lingüística de Hispanoamérica*, con 47 textos que abarcaron los siglos XVI al XVIII, la labor luego disminuyó y sólo fue posible ver una nueva colección documental publicada en el 2005, con un corpus de 221 transcripciones que abarcaban desde 1548 a 1651¹.

Entonces, es un problema fundamental que debe ser abordado en plenitud el poder establecer las bases de un trabajo paleográfico-lingüístico que permita el desarrollo de un corpus diacrónico de cierta relevancia, así como el poder dar cuenta de los mecanismos y herramientas que puedan ser utilizables hoy en día para una lectura y transcripción apropiada para los fines que se requieren en el ámbito de la investigación lingüística distintos, por supuesto, a los que persigue una edición crítica o una base documental para el conocimiento de la historia. Esto no significa, en modo alguno, que unos requerimientos sean mejores que otros, simplemente deben adecuarse al público objetivo al que se dirigen. Revisamos, a continuación, algunos de estos aspectos.

1. CONSTITUCIÓN DEL CORPUS

Jara señala un punto capital respecto de las fuentes documentales, extrapolable a los estudios filológicos y lingüísticos:

Sin fuentes primarias no es posible la creación histórica, el planteamiento de nuevos problemas /.../. Algunas de esas fuentes primarias constituyen para la Nueva Historia

¹ Esta colección corresponde al texto *Ilegibilidad y Cotidianeidad*, mencionado en las Obras Citadas.

canteras llenas de riquezas reveladoras y altamente indispensables a una ciencia que ha cambiado sus objetivos y que ha desarrollado nuevas técnicas, aptas para el tratamiento de sus materiales y de su documentación (Jara 1996:23).

Con todo, la constitución de un corpus, análogo en la sincronía al trabajo realizado por la geografía lingüística y la dialectología, presenta no pocas implicancias teóricas y metodológicas que trataremos de discutir a continuación.

Dentro del marco de este estudio histórico adquiere especial relevancia la determinación de los parámetros involucrados en la configuración del corpus por analizar y su posterior selección de acuerdo con dichos criterios. *Grosso modo*, existe la implicancia de una base textual lo más amplia posible como para constituirse en representativa de los diferentes tipos de texto, estilos, procedencias geográficas y sociales de sus autores, entre otras, y así poder obtener alguna información acerca de la ‘arquitectura’ del discurso social, esto es, de su constituirse como tal de acuerdo con una tradición histórica.

En primer lugar, como definición previa, tendremos que establecer la noción de corpus que está en la base de la discusión. Se entenderá por tal al ‘conjunto homogéneo de documentos lingüísticos de cualquier tipo /.../ que se toman como modelo de un estado o nivel de lengua predeterminado, al cual representan o se pretende que representen’ (Alvar *et al.* 1994: 9-10), ello dado que es prácticamente imposible contar con un corpus constituido por la totalidad de los textos producidos en una lengua. Esto implica, en primera instancia, que el corpus necesariamente es limitado, aunque no por ello deje de ser representativo y, en segundo término, los documentos que constituyen el corpus debieran reflejar una proporcionalidad vinculada con su aparición en el estado de lengua representado.

En todo caso, aun cuando tengamos claridad en la noción de *corpus* a la cual nos referimos, debe haber también acuerdo en lo que debe ser el corpus diacrónico que pretendamos constituir. Según sus características constitutivas, la tipología de los corpus se presenta como una estructura que puede definirse según criterios de ordenamiento y conceptos asociados, tales como:

Según la modalidad de la lengua	Textuales y Orales ² .
Según el número de lenguas	Monolingüe, bilingüe o multilingüe ³ .
Según la cantidad y distribución de los textos	Grandes, equilibrados, piramidales, cerrados y abiertos o monitor ⁴ .

² *Textuales* (Escritos); *Orales* (Muestras de lengua hablada).

³ Monolingüe (Da cuenta de una lengua o variedad lingüística); Bilingüe o multilingüe (Muestra dos o más lenguas, cuyos corpus pueden ser *Comparables*, *Paralelos* o *Alineados*).

⁴ Grandes (No tienen límites de palabras o es muy elevado, no siguen criterios de representatividad o equilibrio), Equilibrados (La misma proporción de diferentes tipos de textos), Piramidales (Textos distribuidos en estratos o niveles, de manera que en cada nivel hay más variedad y menos textos), Cerrados (Constituidos por un número predeterminado de palabras), Abiertos o monitor (En constante crecimiento, gracias a la introducción periódica de nuevos textos según proporciones definidas. Suelen incluir textos completos en vez de muestras).

Según la especificidad de los textos	Generales, especializados, genéricos, canónicos, cronológicos, diacrónicos o históricos y sincrónicos ⁵ .
Según la representatividad del corpus	Textuales, de referencia y léxicos ⁶ .
Según el proceso al que se someta el corpus	Simples (de datos brutos, no anotados, no codificados), verticales y codificados o anotados ⁷ .

(STEL 2007).

Desde este punto de vista, el corpus *textual* que deseamos conformar, para su edición ulterior, debiera ser de carácter *monolingüe, grande, diacrónico, textual y simple*. Como la primera etapa del trabajo necesariamente ha de ser la selección y transcripción paleográfica de los documentos, en los inicios el corpus será *simple*, esto es, constará de los textos transcritos, sin anotación alguna, de acuerdo a criterios de transcripción que aquí señalaremos y que discutiremos en relación con otros usados en diversos ámbitos disciplinarios.

No es un dato menor el que nos entregan Kytö y Rissanen, cuando señalan que:

A historical linguist must therefore rely on a corpus, either in the old sense of the Word, that is a text or a collection of texts yielding linguistic evidence on the phenomenon studied, or in the new sense of the word, that is the computerized version of the same. And if (s) he wishes to avoid the study of first-hand textual evidence, (s)he has to rely on another scholar's earlier corpus work, or on some more refined outcome of it, such as dictionaries or concordances.

(Kytö y Rissanen 1997: 9)

Un factor importante de considerar es el *tipo de texto* que se recoge en el estudio. Los textos seleccionados tratarán de ser una muestra documental proporcionada del discurso informal, tanto oficial como no oficial, pues sobre todo en este último tipo de discursos el redactor tiende a preocuparse más por el contenido de tales documentos y mucho menos de la forma. Dentro de los textos seleccionados, la mayor parte de ellos se caracterizará, posiblemente, por su naturaleza más bien oficial que no oficial, ya

⁵ Generales (Pretenden reflejar la lengua o variedad lingüística de la forma más equilibrada posible), Especializados (Textos que pueden aportar datos para la descripción de un tipo particular de lengua), Genéricos (Textos pertenecientes a un único género; su objetivo es caracterizar dicho género), Canónicos (Obra completa de un autor), Cronológicos (Textos de una época concreta; su objetivo es estudiar la lengua producida durante ese período), Diacrónicos o históricos (Textos de diferentes etapas temporales. Permite observar la evolución de la lengua), Sincrónicos (Su finalidad es estudiar una o más variedades lingüísticas en un momento determinado, generalmente para establecer comparaciones entre variedades o dialectos).

⁶ Textuales (Formados por textos enteros), De referencia (Formados por fragmentos, porque interesa más el nivel de lengua que el texto en sí. Importa el equilibrio y la representatividad), Léxicos (*Sample corpus*. Formados por fragmentos muy breves de textos, de una longitud constante).

⁷ Simples (Textos guardados sin formato alguno y sin añadir ninguna información adicional), Verticales (El resultado de disponer en forma de columna las palabras de un texto ordenadas según criterios alfabéticos o frecuenciales), Codificados o anotados (Formados por textos a los que se han añadido, de forma manual o automática, determinadas informaciones referidas a la estructura de textos o aspectos puramente lingüísticos. Pueden ser: *analizados morfológicamente; parentizados; analizados o treebanks*).

que la experiencia señala que siguen siendo muy pocos los encontrados que respondan al requisito de ser no oficiales o privados.

Aun cuando se reconozca esta supremacía de los textos privados sobre los oficiales, por los autores citados anteriormente, este criterio no siempre es aplicable, debido a que la documentación celosamente resguardada, atesoradora de los más invaluable y significativos manuscritos, que dan testimonio de las diversas actividades de la vida pública nacional, mal que le pese al investigador, sólo dan cuenta de la vida privada y, junto con ello, del modo familiar o cotidiano de la expresión, cuando la emisión o recepción de ellos podía estar a cargo de algún personaje connotado o trascendente en el quehacer de la comunidad de la época. Tal es el caso, por ejemplo, de las cartas enviadas a y escritas por Juan de Ugarte Urrispurro. Esto dificulta aún más el poder lograr un continuo estilístico que represente a la sociedad de la época, pues debe recordarse que el porcentaje de alfabetismo no era muy alto, con lo cual el nivel de quienes escriben estará, necesariamente, sesgado en lo social. Sin embargo, para suplir estas posibles deficiencias se hará acopio de documentos lo suficientemente valiosos, aun cuando no se trate de correspondencia privada, para los fines de este estudio, siguiendo así lo expresado por Frago, en cuanto a que

la conveniencia de recurrir a un equitativo análisis de corpus no literarios a la hora de investigar en las diferentes facetas históricas del español se acentúa, y aun se convierte en necesidad esencial, cuando del español americano se trata, ya que el período de su configuración como habla regional diferenciada (...) es de una literatura hispanoamericana de escasa identidad (...).

En cambio, disponemos de un inabarcable acervo textual del todo válido para el más exigente análisis filológico (...).

(Frago 1987: 70)

En este sentido, dentro del marco informal, podremos distinguir aquellos que se hallen en los *textos privados*, correspondientes a cartas, testamentos propios, confesiones, etc., así como los enmarcados dentro de los *textos públicos no oficiales*, como actas, declaraciones, juicios, inventarios de bienes, cartas poder, de compraventa, etc.

2. ALMACENAMIENTO Y LECTURA

Kytö y Rissanen (1997: 10) señalan que: “In the world of the /.../ third millennium we might go as far as to say that the compilation and development of corpora and databases will be essential for the survival and continued flourish of historical studies of English”. Vale lo mismo para el caso del español, especialmente el de Chile, donde el desarrollo sistemático de corpus se ha concentrado en la sincronía.

En nuestro caso, y tal como definen dichos autores, se trata de un corpus cuya principal variable es el tiempo, por lo que “chronological distribution is one of the most important factors in selecting the text sample” (Kytö y Rissanen 1997: 13). El tamaño de la muestra que se pretende editar abarcaría, en un primer momento, 80 volúmenes de los Fondos Históricos del Archivo Nacional, con un promedio de 400 folios, recto y verso, por volumen.

La digitalización del corpus se efectuará por fotograma, a razón de 600 dpi, con un zoom de 900%, lo que entrega un promedio de “peso” de imagen de 5 Mb, en formato *tiff*. La imagen se escaneará como “película positiva b/n”, para que una vez digitalizado el fotograma se invierta la imagen a “negativo” y se visualice como ‘positiva’, dado que al hacerlo directamente como “película negativa b/n” se pierde parte importante del documento. Una vez digitalizadas, las imágenes se procesarán a fin de corregir brillo y contraste, fundamentalmente, para visualizarlas con óptimas características para su lectura y transcripción. Los volúmenes serán transcritos y archivados en un DVD, así como en disco duro, debido a la permanencia del material digitalizado, cuya durabilidad es mayor en este tipo de unidad que en el soporte portátil. Se calcula que, como mínimo, los archivos digitales de los 80 volúmenes ocuparían un espacio de alrededor de 320 GB.

No es menor, técnicamente hablando, lo que aquí se señala, por cuanto la incidencia en la lectura y transcripción de la disponibilidad del material documental no deja de ser esencial para llevar a buen puerto la edición de textos de esta naturaleza. Para que sea más claro lo que aquí se expone y su relevancia, daré cuenta de los diversos aspectos por los que se tiene que pasar para llegar a una correcta lectura del documento. Ejemplificaré, para mayor abundancia, con documentos del siglo XVI, de aquí en adelante. En esta primera parte, se trata del primer folio del volumen 12 de la colección de Fondos Históricos “Escribanos de Santiago”, del Archivo Nacional.

No se trata de un fallo en la imagen, simplemente es lo que se vería en una lectora de microfichas o de microfilm con lentes de muy buena resolución y de un aumento adecuado. Sin embargo, esta característica no es el común de las disponibles. Además, al carecer de la disponibilidad del documento original, tenemos que acceder al texto microfichado/microfilmado en negativo, lo que sin duda dificulta la lectura, por cuanto cualquier elemento que oscurezca el soporte original inmediatamente será tratado como un “blanco” en la imagen en negativo, iluminando el sector donde se encuentra y entorpeciendo la lectura del manuscrito. Hasta hace poco no contábamos con más posibilidades que efectuar la transcripción directamente desde la lectora o, en su defecto, fotocopiar algún documento a partir de la revisión en la lectora, lo cual también traía restricciones, ya que la calidad de la fotocopia tampoco era óptima y los costos involucrados restringían el acceso a un volumen considerable de documentación. Como se aprecia en la imagen siguiente, el proceso de lectura y transcripción se hacía de suyo lento, incómodo y, en la práctica, hacía poco viable el acceso a fondos documentales más antiguos, particularmente a los del siglo XVI, esenciales para la determinación de la manera como se constituye nuestra identidad como pueblo, desde las más variadas perspectivas que dicha aseveración implica.

Para nuestro caso, desde el punto de vista lingüístico, se hacía de suma urgencia contar con un corpus diacrónico que involucrase a las primeras etapas de la constitución de la variable dialectal, por lo que se llegó a una selección documental de volúmenes que implicara que la mayor parte del mismo representaría estas etapas cronológicas. Pero aún quedaba el escollo de la lectura de los documentos.

Aun cuando hace tiempo ya que los más modernos archivos y bibliotecas de Europa y Estados Unidos, principalmente, contaban con la tecnología apropiada para realizar la digitalización automática del soporte de microficha/microfilm, los altos costos que involucraban para cualquier inversión en moneda nacional (entre 12 y 15 millones de pesos como mínimo), nuevamente parecía alejar las posibilidades de

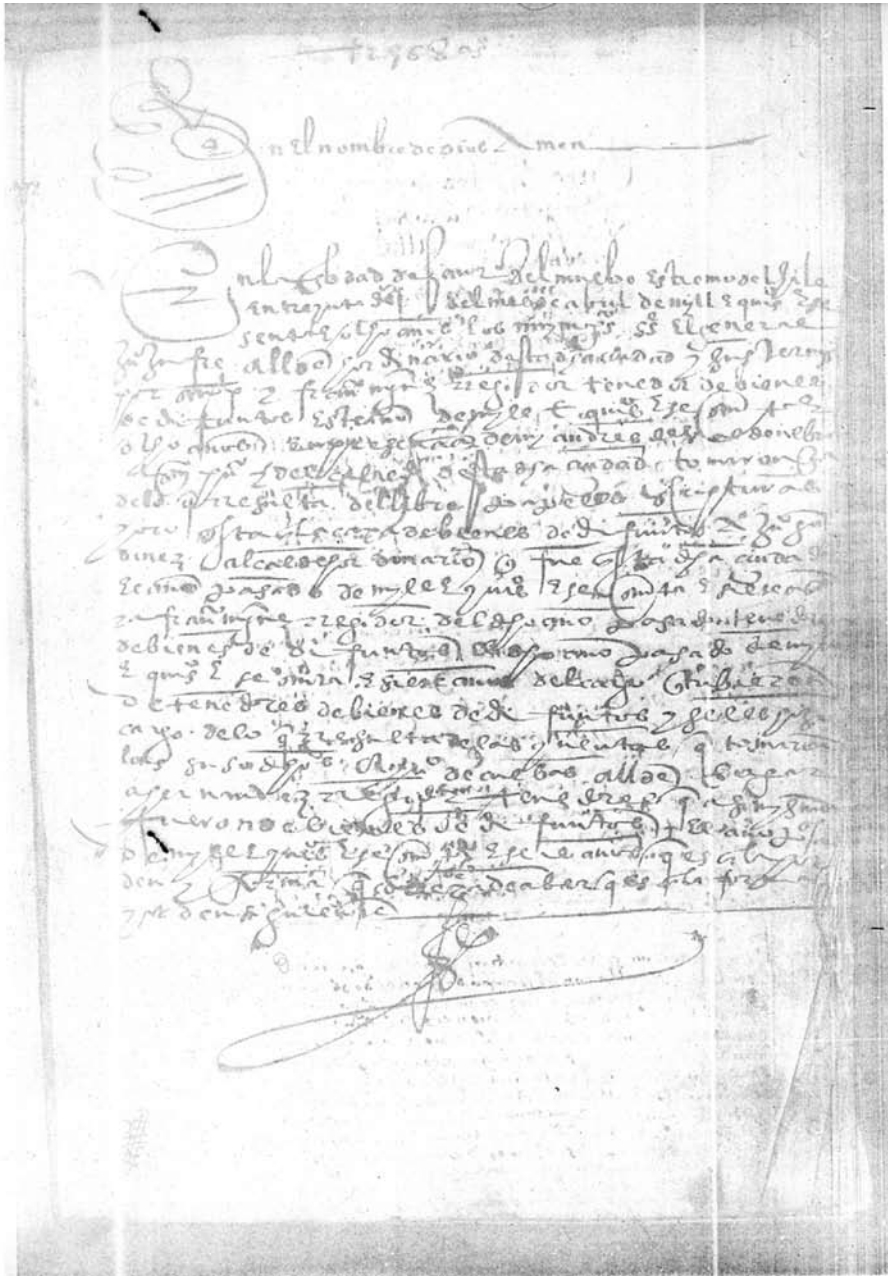


Imagen 2: Bienes de Difuntos (digitalización de fotocopia)

facilitar la labor investigativa en este campo. Con todo, la adaptación de periféricos computacionales ideados para otra labor ha permitido obtener resultados más que satisfactorios, utilizando el scanner CanoScan 9950F, con adaptador para películas y diapositivas, que cuenta con una resolución óptica de 4.800 x 9.600 dpi e interpolada

de 19.200 x 19.200 dpi. Sin embargo, esta última característica no es la primordial, si bien es importante. Para nuestra opción importa que el escáner, a través de su software *ScanGear*, efectúa un zoom de hasta 900% sin dificultades para la memoria del computador, lo que permite, posteriormente, agrandar el tamaño de la imagen para leer cómodamente el texto, como veremos luego. Además, el software asociado, *ArcSoft Photo Studio 5.5*, entrega la posibilidad de realizar algunas mejoras automáticas que le dan mayor nitidez al texto.

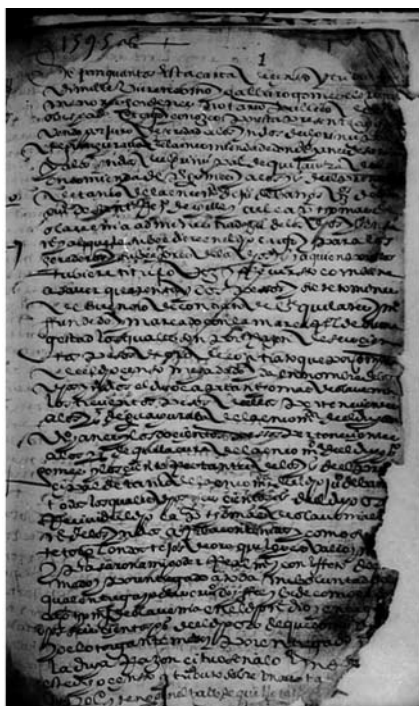


Imagen 3: texto digitalizado, convertido a imagen “positiva”

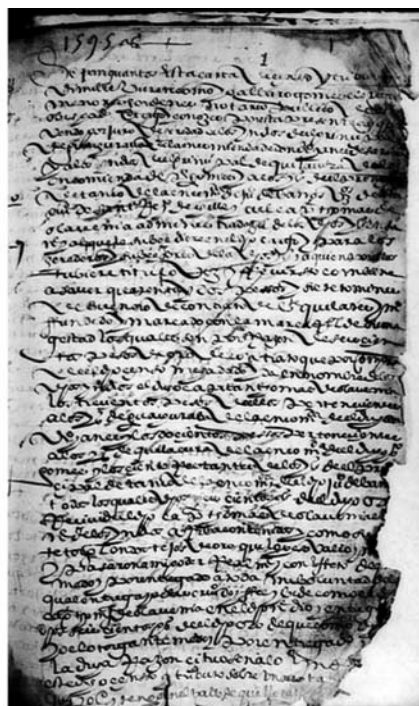


Imagen 4: texto con mejora automática

Lo anterior se traduce en la posibilidad de ampliar una línea del texto sin perder nitidez y resolución para leer, ahora, en el computador, lo que implica que, con el advenimiento y masificación de las *Wide screen*, se facilita el poder disponer de la línea completa y no de partes de ella, para leer y transcribir alternativamente el texto.

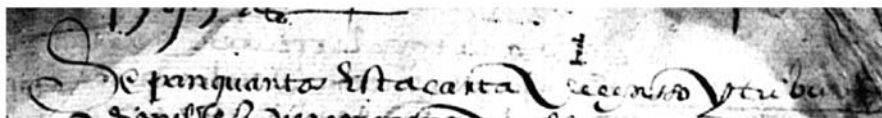


Imagen 5: 1ª línea, del folio 1, del volumen 12 de Escribanos de Santiago

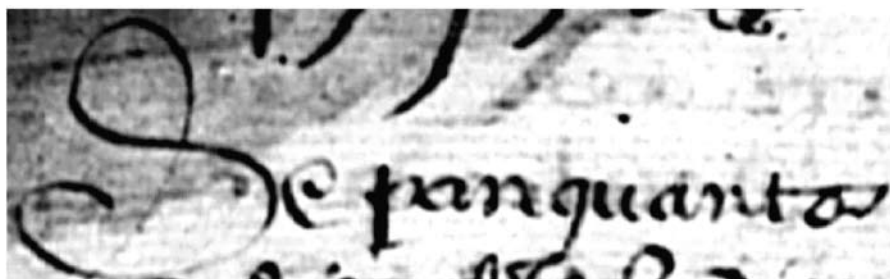


Imagen 6: primeras palabras del folio citado, con agrandamiento de imagen manual.

Con todo, la lectura de los textos, si bien facilitada por los medios tecnológicos disponibles, sigue siendo una dificultad objetiva para la constitución de corpus diacrónicos. Trataremos en las siguientes líneas de este asunto.

3. ASPECTOS DE LA TRANSCRIPCIÓN DEL CORPUS

El objetivo de la transcripción paleográfica lo resume Cencetti con estas palabras:

Scopo della paleografia è, infatti, non solo interpretare esattamente gli antichi manoscritti, ma anche datarli, localizzarli e, in generale, trarre dal loro aspetto esteriore tutti gli elementi utili allo Studio del loro contenuto e, su un piano più ampio, alla storia della cultura in genere. Il suo studio comprende pertanto: /.../quello della storia della scrittura alfabetica (paleografia in senso stretto); quello dei segni accessori della scrittura alfabetica (interpunzione, numerali, segni ortografici e critici, ecc.) /.../.

(Cencetti 1978: 7-8)

Desde el punto de vista metodológico, se han empleado en nuestro ámbito tres formas para la transcripción paleográfica documental: las vinculadas con la historia, la filología y la lingüística. La primera, dice relación con el método empleado por historiadores para transmitir el acervo documental, donde se moderniza todo el documento (grafías, puntuación, léxico, etc.). El caso más representativo es, quizás, la *Colección de Documentos Inéditos para la Historia de Chile*, de José Toribio Medina. Respecto de las normas de transcripción paleográfica, de acuerdo con esta óptica, dice Jara (1996: 22) que son las empleadas por Medina, en la obra citada, esto es, “modernizar la ortografía, pero conservando el sonido original, para mantener el sabor arcaico. En consecuencia, esta publicación no es útil para filólogos, salvo en lo que respecta a saber el lugar en que está ubicado el documento. Es una advertencia que debe ser tenida en cuenta”.

En el segundo caso, se responde, como señala Kordić (2005: 27), a “las normas textológicas de la serie Biblioteca Antigua Chilena (BACH)”, establecidas por Mario Ferreccio, que han sido pilar de numerosas ediciones críticas generadas por el Seminario de Filología Hispánica de la Universidad de Chile, como por ejemplo la reciente edición del *Cautiverio Feliz* de Pineda y Bascuñán. Estas normas las sintetiza más adelante como sigue:

La reducción fonografemática realizada contempla el principio básico del respeto y conservación de todo rasgo gráfico que implique efectiva o eventualmente la representación de un rasgo fónico diferencial, significativo; todo aquel recurso que, tras el examen del comportamiento gráfico del escriba, demuestre ser inoperante, se moderniza, con el fin de evitar en el texto editado la presencia de inútiles grafías exóticas que sobrecarguen visualmente el texto y confundan al lector

(Kordić 2005: 27-28)

En este sentido se simplifican las geminadas *ff*, *ss*, *rr*; se simplifican usos como *qu-cu* (quales > cuales), uso de nasales según uso actual (*enpezar* > *empezar*), restitución de vocal protética (*scribano* > *escribano*); se aplican normas actuales de puntuación, acentuación y uso de mayúsculas y minúsculas.

Finalmente, con el tercer criterio, se pretende entregar un reflejo lo más fiel posible del texto, ya que como señala Alvar y Alvar (1981: 9):

La edición paleográfica de un texto tiene sus propias peculiaridades: trata de hacer asequible con signos actuales lo que resultaría de otro modo de penosa o imposible lectura para quien no tenga cierto tipo de conocimientos. Pero, por otra parte, trata de presentar ese material de la manera más fiel con respecto al original que transcribe. No es –como se ha dicho erróneamente– algo que pueda suplir a la fotografía, sino lo que la fotografía no puede dar: la sencillez, sin transgredir en nada de lo que consta en el original.

Para este trabajo se han adoptado los criterios del grupo de trabajo de ALFAL, con algunas modificaciones que ya explicaremos. La pauta original, que puede encontrarse en <http://pizarro.fll.urv.es/proyecto/transcripcion.htm>, se observa a continuación:

NORMAS DE TRANSCRIPCIÓN DE LOS DOCUMENTOS

Las transcripciones serán de carácter literal estricto, ajustándose a las siguientes normas:

1. Se respetará en un todo la grafía original del texto.
2. Sólo se apartará del mismo en cuanto contemplará la separación gramatical de palabras.
3. Esto regirá no sólo en cuanto se separarán las palabras unidas, sino también en cuanto se unirán las letras de una palabra que estén separadas.
4. Se conservarán las abreviaturas.
5. La *s* larga y la *s* de doble curva (redonda) se transcribirán con *s* redonda (la usual en la grafía moderna).
6. La *c* con cedilla se transcribirá literalmente: *ç*.
7. Se respetará el uso de *i* e *y*, ya sea como vocales o consonantes según el texto original: *myll*, *maior*, *vyo*.
8. La *r* mayúscula con valor fonético de *rr*, se mantendrá: *Río*, *Rodrigo*, *coRe*.
9. Se respetará la duplicación de letras: *cozza*, *cappitan*, *ottra*, *ffecha*.
10. Se conservarán las contracciones: *del*, *della*, *desta*, *ques*.
11. Se respetará el signo copulativo *e*.
12. Se conservará la puntuación del original.
13. Se conservará el uso de mayúsculas y minúsculas del original.
14. Se respetará la acentuación (o su ausencia) del original.

Imagen 7: Normas ALFAL

En cuanto a los fines que aquí nos proponemos –entregar una transcripción lo más fiel posible al original– cabe hacer algunos alcances. Para superar los escollos que implica la labor de transliteración paleográfica, lo primero que debe tenerse en cuenta es el conocimiento que se tiene sobre las formas alfabéticas de la época en que el texto estudiado se ha escrito, esenciales y variantes, así como el sistema lingüístico al que corresponde. Además, el detenerse particularmente en las características de la escritura que debe leerse sirve para identificar sus rasgos internos, junto con refrendar o no el carácter autógrafo del texto. En este sentido cabe hacerse eco de las palabras de Morales, quien señala:

Averiguar hasta qué punto la letra corresponde a los principios distintivos de su tiempo, estudiándola, en lo posible, con independencia de su valor textual, es una faena del mayor interés histórico, incluíble en el orden de las investigaciones morfológicas de la cultura. Espíritu de la letra, podría nombrarse semejante trabajo, género de grafología de las épocas en la que se evidenciaría, una vez más, que el estilo no es el hombre, sino el tiempo. Los textos, apreciados con ese criterio, adquirirían nuevo sentido, convirtiéndose en expresión viva y directa de su momento, actualizándose así aquello que, por su dibujo, parecía letra muerta.

(Morales 1994: 30)

Nunca será bastante el insistir en que la lectura de un texto que hay que transcribir paleográficamente debe realizarse pausadamente y sin recurrir excesivamente a la memorización de las características de los grafemas ni al parecido que se pudiese observar entre los de diversos textos, pues el trazado, aun cuando responda a idénticas matrices de puntos, también conlleva una variación morfológica que es propia de la concreción de dicha estructura. Cuando nos adentramos en el “reino de la palabra”, ya no sólo es necesario haber reconocido el sistema ortográfico y alfabético del texto que hemos de analizar –de cada uno de ellos, pues cada documento, muchas veces siendo de un mismo autor, varía escriturariamente dependiendo de las circunstancias que están involucradas en el proceso de su construcción–, sino que ahora se hace imprescindible el dominar el tipo de reducción lingüística que opera en el sistema de abreviaturas empleado, pues generalmente nos encontramos ante formas plurilaterales, en las cuales se escribe no sólo la primera y última letra de la palabra, sino que también involucra a otras letras, lo cual –aun cuando con el tiempo llegó a ser un sistema estandarizado– daba los indicios necesarios para identificar el morfema al cual se hacía referencia.

Con todo ello, y aunque se pudiera recurrir a alguna transcripción realizada con anterioridad a la propia, la labor de lectura e interpretación ha de realizarse de manera individualizada, sin mediatizaciones que entorpezcan el libre acceso al documento, sino que, por el contrario, nuestra lectura podrá compararse con otras pero siempre partiendo de una base asentada sobre los criterios señalados de observación, estructuración y análisis comparativo del desarrollo interno del texto en cuestión.

Finalmente, tal como se lee en Núñez:

Resultan muy operativas la observación y la fijación en la memoria de las formas ocasionales de cada letra considerada aisladamente, con independencia una de otras. Lo mismo con los nexos de letras que se dan en cualquier escritura y que, por su reiteración, contribuyen a deformar las formas paradigmáticas de aquéllas. Otro tanto puede decirse de los demás signos gráficos que no sean letras ni nexos, pero que se usan para completar el sentido de la escritura: signos de puntuación, numerales, de interrogación, etc.

(Núñez 1994: 36)

De las grafías que conforman la tradición escrituraria del romance castellano, llegó hasta el período en estudio una serie de signos y abreviaturas que facilitaban la labor del escribano en la producción textual. La mayor parte de ellas nos son ahora desconocidas y de difícil interpretación, pues el criterio con el cual se realizaba la reducción de una palabra en su respectiva abreviatura se ha perdido con el tiempo. Por ello, junto con efectuar la transcripción documental, se dedicará un apartado a recopilar aquellas abreviaturas utilizadas en el período.

Eso sí, existen grafías difíciles de encontrar, ya que sólo se asentarán con la imposición escolar de la *Ortografía* académica del siglo XVIII, como por ejemplo el caso de ñ, la cual es de escasísima figuración y siempre con la indicación de una tilde que supone una superescritura de otra *n*, como en *ano*. Otro problema frecuente en las transcripciones paleográficas es la interpretación de la grafía correspondiente a la vibrante múltiple, ya que ésta puede figurar con carácter simple mayúsculo incluso en posición intervocálica –como en *se Pramas*– o con un tipo invertido, ya sea inicial o interior de palabra, como en los casos de *figuimienta*, *se ha* o *figuibi*.

Asimismo, la tradicional ortografía de los escribanos públicos y de algunos personajes de gran formación escolar conserva, junto a la *s* simple de doble curva, las medievales *ese larga* (ʃ) y la de carácter sigmático (σ), como puede leerse en *ſito* ‘sancto’ y en *ſta* o en *ſyres*.

Finalmente, el carácter de las abreviaturas está principalmente determinado por la contracción de la palabra, en muchos casos de manera relativamente estandarizada, que permite inferir el desarrollo de la misma. En los casos en que existe una combinación de grafía y símbolo –como en los de *per*, *ser* y *ver* en *pertenece*, *servido* o *verdad*– el trazado para este último es claramente reconocible, siendo aislable del resto de la palabra y aplicable a situaciones y contextos similares. Por último, la variedad que pueden presentar las posibles combinaciones de una abreviatura, en general, estará determinada más por el tipo de letras que se emplee en el documento (procesal, humanística, cortesana, redondilla, etc.) que de divergencias morfológicas o idiolectales.

Habida consideración de lo dicho anteriormente se seguirá, en la transliteración documental, el siguiente conjunto de normas:

- Esencialmente, se respetarán las grafías originales del texto. Por razones de inteligibilidad, se realizará la separación gramatical de las palabras encadenadas. Por otra parte, también se unirán aquellos vocablos que se encuentren separados dentro de la línea. En caso de haber contracciones, sólo se separarán aquellas en que una letra es utilizada para indicar dos términos, sin que el final de una sea el inicio de la otra, v.gr., *êl* “en el”, **D**^{or} **on** “Doctor Don”, etc.

- Las abreviaturas se desplegarán para facilitar la lectura. Dicho despliegue se realizará, cuando sea posible verificarlo, conforme a las pautas internas del texto o del autor; de no ser así, según la forma actualmente aceptada. El desarrollo se indicará con letra cursiva, mientras que el texto original se mantendrá con la letra común. En cuanto a las abreviaturas *xpo* y *xpianos* u otras variantes, éstas se transcribirán como *Christo* y *christianos*, según la tradición escrituraria de la época.
- Los casos de R mayúscula en posición interna y en posición inicial, siempre que no correspondan a nombre propio, se transcribirán como *rr*. Se mantendrán las letras dobles del texto, tanto en posición inicial como interna, por ejemplo, *ſſer, mill, cappitán*. La ñ sólo aparecerá cuando la nasal lleve dicha tilde o en el despliegue de la abreviatura que lo requiera. El uso de mayúsculas se observará conforme a la ortografía de los documentos.
- Se conservará la ese alta (ſ) y la sigmática (σ). El signo copulativo tironiano se representará por *e* en los casos en que se verifique en los documentos, como se hace por lo demás dentro de la tradición filológica hispánica.

Se utilizarán, además, algunos signos especiales que indican características de texto, como

- Los puntos suspensivos [...] indicarán algún tipo de omisión en el documento original.
- El paréntesis angular < > contendrá una rectificación o adición al texto realizada por sobre la caja de escritura.
- El paréntesis de corchetes [] señalará que se ha efectuado una restitución en la transcripción del texto, ya sea porque éste presentaba algún daño físico (rotura, mancha) o debido a un final de línea que ha quedado inconcluso.
- La barra oblicua sencilla (/) marcará la separación de líneas en el documento, llevando como superíndice su número correlativo. La barra doble (//) indicará cambio de foja, ya sea el reverso de una o el paso a otra.
- La indicación [**ilegible**] señalará pasajes que no han podido ser transcritos debido a su dificultad de lectura; al contrario, los legibles, pero incorrectos o de dudosa comprensión, irán acompañados de un [**sic**].

Por otra parte, al inicio de la transliteración de cada documento se hallará una referencia que, en términos generales, indicará lo siguiente: Archivo y fondo del cual procede el manuscrito original, donde la sigla A.G.I. corresponde a *Archivo General de Indias* y la sigla A.N.S. a la de *Archivo Nacional de Santiago de Chile*. Localización de las fojas dentro del volumen respectivo, fecha, tema del texto y autor.

Al final de la transcripción se entregará otra clase de información: Reproducción de palabras o frases tarjadas en el documento, con indicación de la línea donde se hallan, y referencia a una posible dificultad de lectura o interpretación de algún pasaje del texto, así como los lugares donde se halle roto el manuscrito.

No hay que dejar de tener presente, en todo caso, las indicaciones que hace Kordić respecto de las ediciones críticas, válidas para cualquier ámbito de la edición de textos:

En primer lugar, nos enfrentamos, en el caso de las zonas marginales o semimarginales, como Chile, a una tradición textual casi exclusivamente manuscrita y de un solo testimonio, borradores o copias en limpio /.../. Resultado de lo anterior, estos textos no plantean problemas como los de intervención de impresores, deturpación de los testimonios o contaminación textual entre ediciones o ejemplares distintos./.../

Por otra parte, al no contar con ediciones críticas de muchas de estas obras, y no existir otros ejemplares o testimonios, la resolución de los pasos oscuros o lecturas inciertas se torna muy compleja, ya que no hay punto de comparación o referencia posible.

(Kordić 2006: 196-197)

El trabajo documental, con todo, debe tener presente que existen otras formas de editar documentos históricos, tal como lo señalan Steven y Burg, cuando se refieren a los métodos de transcripción:

There is no single agreed-on method of transcription. Rather, editors use different types of editorial apparatus and different degrees of emendation depending on the nature of the documents and the intended audience. Editors choose among five majors forms of document presentation (photographic facsimile, typographical facsimile, diplomatic transcription, expanded transcription, or clear text transcription) to find a style that best suits the needs of their audiences, the purposes of their editions, and their personal preferences.⁸

(Stevens y Burg 1997: 72)

4. INTERCAMBIABILIDAD

En la actualidad es de suyo indispensable pensar que, en una creciente sociedad globalizada, el intercambio de información se hace cada vez más necesario, por lo que es de necesidad insoslayable plantearse la codificación del material transcrito a fin de considerar dos criterios: la *intercambiabilidad* de manera independiente de los recursos tecnológicos (software o hardware) y la *conservación* de la información, es decir, que producto de esta codificación no haya pérdidas o cambios en los datos. Nos referimos, en estos términos, a la edición electrónica del material documental.

A fin de poder lograr ambas metas, se recurrirá al metalenguaje SGML, en versión del *Prontuario de Marcación SGML* que la Real Academia Española ha adaptado para la informatización del Corpus Diacrónico del Español (CORDE), ya que se

⁸ Dicen los autores, respecto de cada uno de estos tipos de edición: "Photographic facsimile editions reproduce original documents and present them on the page, screen, or microform frame free from alterations caused by converting manuscripts or original print into modern type" (p. 72). "Editors who use typographical facsimile as a method of transcription transfer original documents into type so that the appearance and physical arrangement of words on the typescript page mimic that of the original" (p. 74). "Editors using diplomatic transcription transfer the text of a document into modern type but use symbols such as arrows, carets, and brackets to indicate the document's stylistic details and appearance /.../. This method is often used for documents for which the reader may be interested in the composition process" (p. 76). "Expanded transcription encompasses a wide spectrum of editing styles, all of which standardize accidentals, datelines, and signatures; mark paragraphs with indentations; and do not attempt to reproduce the excessive spacing and physical layout of the text of documents." (p. 76). "Clear text editions use silent emendation and limited editorial apparatus to produce clean typescript transcription of documents." (p. 79).

consideran allí las marcas para las diversas tipologías diacrónicas de los textos que pudieran encontrarse, tanto en prosa como en verso.

Algunas marcas de este tipo son de carácter *intratextual*, como las siguientes:

prosa	la marca <p> señala, en prosa, el comienzo de un párrafo nuevo en el texto; los textos en prosa de los corpus se segmentan en unidades <s>, que se corresponden aproximadamente con los enunciados y con las oraciones ortográficas
verso	la marca <verse> se utilizará para iniciar una tirada de líneas dispuestas a modo de versos. Dichas líneas, que coincidirán casi siempre con lo que entendemos por “verso”, irán precedidas de la marca <lb>, debemos cerrarlo al finalizar una tirada de versos; la marca <lb> se utiliza para señalar el comienzo de cada línea de una composición en verso
números de página	la marca <pb> indica el comienzo de una nueva página en la edición que se toma como fuente para el texto electrónico del corpus. Su atributo <i>n</i> tiene como valor el número de la nueva página. El elemento <pb> debe aparecer justo antes de la nueva página, no al final, independientemente del lugar donde se encuentre el número en la página de la edición original. La marca <pb> debe estar separada del texto precedente y siguiente por espacios en blanco. Si el cambio de página no coincide con salto de párrafo, la marca <pb> se deja en la misma línea
marcación de texto resaltado	se utiliza la marca <hi> para señalar fragmentos de texto resaltado, ya sea entrecorillado, en cursiva, en negrita, en versalitas, en mayúsculas o subrayado. La marca <hi> debe rodear el texto tipográficamente resaltado. La marca que indica el final de texto resaltado es </hi>. Para señalar las diferencias entre tipos de resalte tipográfico, se utiliza el atributo <i>rend</i> dentro del elemento <hi>”
marcación de las citas del texto	el elemento <quote> se emplea para marcar citas. La finalidad de la incorporación de esta marca al esquema de codificación del CORDE es que no se mezclen, en la fase de recuperación de la información, fragmentos de texto de épocas o autores diferentes
expresiones no castellanas	la marca <foreign> sirve para diferenciar del resto del discurso las palabras, expresiones o frases en lengua no castellana, con la finalidad de constituir un subgrupo dentro del proceso de lematización que permita la recuperación de las palabras del corpus en otros idiomas
glosas del texto	el elemento <gloss> sirve para marcar las anotaciones que aparezcan en los márgenes de las ediciones de cita, ya sean del propio autor o de autoría diferente a la del texto principal
cambios de autoría	el elemento <change> sirve para marcar todo fragmento distinto de una cita que corresponda a una autoría diferente de la principal

Otras marcas dicen relación con la *estructura del corpus*: “La estructura global de cada corpus una vez codificado consta de dos grandes partes: el *prólogo* y el *corpus textual*, articuladas a su vez también de manera bipartita. El prólogo reúne todo un conjunto de declaraciones (declaración SGML, DTD) que informan sobre el tipo de etiquetas empleadas en la codificación del texto. El corpus, segunda parte del documento SGML, es el documento textual completo. No incluye ningún tipo de declaración, sino texto, codificación y referencias de entidad. Se divide en dos grandes partes: la cabecera del corpus y la serie de elementos <TEI.2>, esto es, textos codificados con sus cabeceras correspondientes”. (TEI: *Text Encoding Initiative*).

Esta es una propuesta que será adoptada para el trabajo con el corpus, ya que hasta el momento nos hemos concentrado en la transcripción documental, por cuanto la constitución de un equipo de trabajo integral que aborde todos los aspectos del proyecto ha sido, cuando menos, dificultosa.

A continuación podemos observar la plantilla de la cabecera del proyecto del *Corpus Diacrónico del Español de Chile* (CorDECh). Se ha utilizado el software *Cooktop 2.5*, que es un editor de XML/XSLT de distribución libre, desarrollado por Víctor Pavlov (<http://xmlcooktop.com>). Para verificar la sintaxis de la codificación, el programa cuenta con un validador que permite detectar las “partidas falsas” o aquellos elementos que no cuentan con un cierre en su descripción.

```
<?xml version="1.0" encoding="UTF-8"?>
<?cooktop 2.5 RINGSHEMA="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>CORPUS DIACRÓNICO DEL ESPAÑOL DE CHILE (CorDECh)</title>
          <author>Instituto de Lingüística y Literatura</author>
          <sponsor>Universidad Austral de Chile</sponsor>
          <principal>Manuel Contreras Seitz</principal>
        </titleStmt>
        <editionStmt>
          <edition>Transcripción paleográfica de documentos digitalizados por el Archivo Nacional de Chile</edition>
          <source>Fondos Históricas del Archivo Nacional de Chile</source>
        </editionStmt>
      </fileDesc>
    </teiHeader>
  </TEI>
</teiCorpus>
```

Imagen 8: Cabecera proyecto

La gramática o DTD (*Document Type Definition*) es un lenguaje mediante el cual se definen con precisión aquellos elementos que son necesarios en la elaboración de un documento o un grupo de documentos estructurados de manera similar. Estas “declaraciones de elementos” aportan el nombre oficial de las etiquetas que aparecerán dentro de los delimitadores (por ejemplo, <change>), y describen lo que cada elemento puede contener (modelo de contenido). Para nuestras consideraciones, la DTD considerará elementos como los que se muestran en la imagen 9. Esta estructura corresponde a un texto de Luis de Cartagena, de 25 de septiembre de 1548, donde se otorga un traslado de la merced hecha por Pedro de Valdivia a Inés Suárez

```

<TEI xml:id="MyTextNumber1">
  <teiHeader>
    <fileDesc>
      <subjectStmt>
        <subject sameAs="NomeclColec">Chile-16-1548</subject>
        <subject type="resumen">Traslado de merced de don Pedro de Valdivia a
doña Inés Suárez </subject>
        <author xml:id="autógrafa">Luis de Cartagena</author>
      </subjectStmt>
      <editionStmt>
        <msIdentifier>
          <settlement>Santiago de Chile</settlement>
          <repository>A.N.S.</repository>
          <altIdentifier>Real Audiencia, 340, fs.123-v</altIdentifier>
        </msIdentifier>
      </editionStmt>
      <history><p>Documento en el que se describen los trabajos pasados por Inés
Suárez en la batalla de Santiago, por cuyos méritos se le concede merced de tierras e
indios</p></history>
      <origin><p>Escrito en la ciudad de Santiago de Chile</p></origin>
      <doc>1548-09-25</doc>
      <recordHist>
        <source><p>Catalogado directamente desde el manuscrito
original</p></source>
      </recordHist>
    </fileDesc>
  </teiHeader>
</TEI>

```

Imagen 9: DTD

5. PALABRAS FINALES

Dentro de todas las consideraciones que pueden establecerse para la edición de textos coloniales hispanoamericanos, inclusive tomando en cuenta los destinatarios potenciales de tales colecciones, debiera fijarse un conjunto de criterios fundamentales que estuviesen en la base de las transcripciones, con el fin de disponer de un acervo textual que no comprometiese la fidelidad de los originales, no con el fin de reproducir facsimilarmente el manuscrito, sino para conservar la estructura documental de las fuentes, con el objeto de que luego ellas puedan ser utilizadas para finalidades diversas, lo que sin duda sería difícil de realizar si se posee una colección diplomática ya editada.

Por ello, nos hemos decantado hasta ahora por la transcripción paleográfica, ya que se estima que ésta sirve de base para cualquier tipo de estudio lingüístico, histórico o sincrónico, permite la proyección hacia diccionarios históricos, así como, por qué no, como fuente para la difusión del material transcrito, reeditándolo para alcanzar públicos diversos. Es esta última perspectiva metodológica por la cual nos decantamos. Como se ha dicho, la digitalización del documento presenta invaluable ventajas a la hora de transcribir y de conformar los ‘abecedarios’ internos de cada documento y de cada autor. Este trabajo puede tener proyecciones, inclusive, para el desarrollo de software de reconocimiento óptico de caracteres manuscritos, en la medida en que, habiendo recopilado un corpus documental amplio, se puedan establecer los rasgos fundamentales de cada grafema –así como de su conjunto de alógrafos–, independientemente del individuo que los traza, por medio de las teorías específicas dentro del campo del modelamiento matemático.

6. IN MEMORIAM

Se dice que la vida en el sur es más lenta y el giro de la rueda en la periferia tiene una armonía distinta a la del epicentro. Recientemente nos hemos enterado del fallecimiento, el 9 de julio, del destacado filólogo y académico Mario Ferreccio Podestá. El recuerdo de su nombre y de su obra hará que su espíritu permanezca eternamente entre nosotros.

La obra de Mario Ferreccio estuvo ligada a las letras coloniales, siendo creador, en la Universidad de Chile, del Seminario de Filología Hispánica. Además, nos entregó buena parte de sus desvelos a través de la *Biblioteca Antigua Chilena* (BACH), otra ‘marca registrada’ de su factura, colaborando en estas iniciativas –volúmenes 4 y 5– quien fuera su destacada discípula, Raïssa Kordić Riquelme. La BACH se inició con la publicación del *Purén Indómito* (1984), el que marca el hito para las ediciones críticas de textos de la época. A partir de allí, otros trabajos darán paso y consolidarán la obra de Ferreccio: la *Relación Autobiográfica de Úrsula Suárez* (BACH 2) también de 1984; *El Ignacio de Cantabria*, de Pedro de Oña, BACH 3 (1992); *La Guerra de Chile*, BACH 4 (1996); el *Cautiverio Feliz*, BACH 5 (2001), último volumen en que aparece la figura del maestro. Desde allí, su obra se prolonga a través de los pasos de su discípula, quien publica los *Testamentos coloniales chilenos*, BACH 6 (2005) y el *Epistolario de Sor Dolores Peña y Lillo (Chile, 1763-1769)*, BACH 7 (2008).

En este mismo campo, dedicó sus labores de investigación a las *Fuentes de la Filología Chilena* y al *Registro General de los Gentilicios Nacionales de Chile*, trabajos que cada vez nos entregan una luz más brillante en este intrincado camino.

Hoy día, los caminos que recorrió Mario Ferreccio, ya no le conocen y, por desgracia, aquellos caminos que él recorre hoy día, nosotros no los conocemos. La luz que lo iluminara algún día lo ha recibido en el oriente, por toda la eternidad.

OBRAS CITADAS

- Alvar, Manuel y Elena Alvar. 1981. *Cancionero de Estúñiga*, edición paleográfica, Institución “Fernando el Católico”, Zaragoza.
- Alvar Ezquerro, Manuel y Juan Andrés Villena Ponsoda [coord.]. 1994. *Estudios para un corpus del español*, Universidad de Málaga, España.
- Cencetti, Giorgio. 1978. *Paleografía Latina*, Jouvence, Roma.
- Contreras Seitz, Manuel E. 2005. *Ilegibilidad y Cotidianeidad. Paleografía y colección diplomática de documentos chilenos del período colonial (1548-1651)*, Programa de Estudios y Documentación en Ciencias Humanas (PEDCh), Colección Fuentes Documentales, Universidad de Los Lagos, Osorno, 475 págs.
- Frago Gracia, Juan Antonio. 1987. “Una introducción filológica a la documentación del Archivo General de Indias”, *Anuario de Lingüística Hispánica*, Vol. 3º, Univ. de Valladolid; pp. 67-97.
- Jara, Álvaro. 1996. *Protocolos de los Escribanos de Santiago*, transcripción paleográfica de Álvaro Jara y Rolando Mellafe, Centro de Investigaciones Diego Barros Arana, Archivo Nacional, Fuentes para el Estudio de la Colonia, Ediciones DIBAM, Santiago.
- Kordić, Raïssa. 2005. *Testamentos coloniales chilenos*, estudio preliminar de Cedomil Goić, Biblioteca Indiana, Centro de Estudios Indianos (CEI), Universidad de Navarra, Iberoamericana/Vervuert, Madrid.

- Kordić, Raïssa. 2006. "La crítica textual hispanoamericana: Algunas especificaciones metodológicas", en *Onomázein* N° 13/1; Pontificia Universidad Católica de Chile, Instituto de Letras, Santiago; pp. 191-202.
- Kytö, Merja y Matti Rissanen. 1997. "Language analysis and diachronic corpora", en *Tracing the trail of time. Proceedings from the Second Diachronic Corpora Workshop*, Raymond Hickey, Merja Kytö, Ian Lancashire y Matti Rissanen (eds.), Rodopi, Amsterdam; pp. 9-22.
- Morales, José Ricardo. 1994. *Estilo y paleografía de los documentos chilenos*, DIBAM y Centro de Investigaciones Diego Barros Arana, Santiago.
- Núñez Contreras, Luis. 1994. *Manual de Paleografía*, Cátedra, Madrid.
- Peña y Lillo, Sor Dolores. 1736-1769. *Epistolario*. Prólogo y edición crítica de Raïssa Kordić Riquelme, Iberoamericana/Vervuert, Biblioteca Indiana, Barcelona, 2007. 520 págs.
- Real Academia Española. 1993. *Documentos para la historia lingüística de Hispanoamérica. Siglos XVI a XVIII*, BRAE, Anejo LIII, ALFAL, Comisión de Estudio Histórico del Español de América, M^a Beatriz Fontanella de Weinberg (comp.), Espasa-Calpe, Madrid.
- Real Academia Española. 2001. *Corpus Diacrónico del Español. Prontuario de Marcación SGML*, Madrid.
- Servei de Tecnologia Lingüística. 2007. *Introducción a la lingüística del corpus*, http://www.ub.edu/stel/esp_support.htm
- Stevens, Michael y Steven Burg. 1997. *Editing Historical Documents*, AltaMira Press, California, USA; 265 págs.