Tecnologías digitales aplicadas al estudio de la poesía

Curso de Verano 2016 LINHD-UNED

Raúl Jiménez Martin

Tecnologías digitales aplicadas al estudio de la poesía: Curso de Verano 2016 LINHD-UNED

Raúl Jiménez Martin

Tabla de contenidos

Capítulo 1. Principios y problemas de la métrica medieval castellana e Introducción

Fernando Gómez Redondo, Catedrático de Teoría de la Literatura y Literatura Medieval, UAH . Uno de los problemas iniciales era la diferenciacion de los sistemas métricos que podemos encontrar en el estudio de esta materia. De esta manera, el autor propone los siguientes sistemas de versificación:

- Versificación isomélica: Llevada a cabo por juglares y trovadores la regularidad de la linea depende de la ejecución melçodica en el canto (cantus gestualis). El romancero.
- Versificación isosilábica: La regularidad la asegura el cómputo de sílabas mediante la aplicación de las reglas del ars grammatica; poemas recitados y discurso prosódico.Se define el concepto de pie rítmico a la unión de dos sílabas, siendo una sola sílaba un semipié.
- Versificación isorrítmica: En esta caso no importan tanto la sílaba como la posición del acento. El pie pasa a ser unidad métrica y soporte del acento rítmico. Reduce a dos medidas las clausulas (pies): binarias y ternarias (troqueos óo y dáctilos óoo), ya que nuestro lenguaje no distingue las sílabas luengas de las breves. En la isorritmia se contempla como desplazamientos acentuales el Ectasis (la sílaba breve se hace larga) y la Sístole (acortamiento de la sílaba por desplazamiento de la carga acentual). Otros methaplasmos que define Nebrija son la sinalefa, el Ectlisis, la sinéresis, la compensación y el incremento por aguda.

Las claves del análisis rítmico son la utilización de teorías vernáculas, la transformación de la sílaba fonológica en sílaba métrica y paso del sistema cuantitavio al acentual.

Cabe destacar el estudio de las pausas en la poesía para el incremento o pérdida de la cantidad silábica.

La teoría métrica medieval castellana se fija de modod definitivo en las artes que se componen en la segunda mitad del siglo XV. La tipología de versos castellanos es muy reducida y puede reducirse a solo tres esquemas: el verso clerical, sustituido por el arte mayor y el arte real o comun. La métrica medieval castellana es siempre de base par. El arte mayor (5+5 métricas) y el arte real se convertirán en los versos predilectos para el desarrollo de materias elevadas; el octosílabo será el metro que propicie la consolidación de la épica culta.

Capítulo 2. Generación computacional de poesía: 15 años de WASP (The Wishful Automatic Spanish Poet).

Pablo Gervás Gómez-Navarro, Director del Instituto de Cultura y Tecnología y del NIL Group, Universidad Complutense de Madrid .

El desafío inicial era la generación automática y creativa de poseía generada por ordenador. Los intentos, que iniciaron hace 15 años, consistían en el reuso de la experiencia para solucionar problemas.

Esto se podía aplicar para la creación poesía creando una base de datos de casos, y haciendo que el programa busque en un problema anterior que sea similar al actual e identifique la solución aportada en su momento. De esta manera el programa selecciona una solución y la propone, en este caso, a la persona encargada de revisar y aprobar la elección. Esto último es requisito indispensable.

En nuestro caso la solución es un verso. Y el problema es que cumpla los requisitos impuestos sobre métrica, gramática, significados... etc. Inicialmente se usaba una suerte de lenguaje sintaxis de lenguaje regular en Lisp que definiera los parámetros de cada creación.

Para el desarrollo de la Inteligencia artificial que se aplicaba en este proyecto se siguió un algoritmo evolutivo, basado en cuatro estadios (sistemas expertos): Generación de contenido, Poesía ó convertir esos textos generados en versos y rimas, Jueces que evalúen los diferentes aspectos y Revisadores que seleccionen los mejores resultados.

Posteriormente se aumentó la capacidad de cómputo para ampliar la efectividad del algoritmo. Se vio, con esto, que los resultados no mejoraron. Se intentó entonces estudiar qué esquemas hacían que los versos fueran mas interesantes y se encontraron una serie de rimas, independientemente de la longitud, que hacían destacar la poesía generada. De esta manera se le estaba dando a la máquina mas valores donde elegir para su composición. Curiosamente coincidían con construcciones populares.

Un experimento curioso fue la generación de poesía a partir de textos periodísticos. El resultado no fue satisfactorio, se vio así que los textos periodísticos, sean del país que sean, no servían para generar poesía interesante. Se hizo la prueba de hacer un corpus base de textos periodísticos mexicanos junto con poesía mexicana y el resultado mejoró considerablemente.

El estado actual de la investigación comprende un interfaz gráfica para que el usuario pueda solicitar al sistema creaciones poéticas a medida, y de paso pueda comprobar la dificultad estadística a la que se enfrenta el modelo.

Capítulo 3. Mesa redonda sobre: Investigación, poesía y tecnología: un panorama de proyectos

Elena González-Blanco García profesora de literatura española y directora de LINHD-UNED. Introducción a *LINHD*, su labor y trabajos. Objetivo de este curso: Fusión de avances en humanidades con herramientas tecnológicas y el desarrollo de las "Humanidades digitales". Por ello es notable destacar el último proyecto llamado *Evelyn*, una suerte de espacio de trabajo orientado a acercar las humanidades digitales a investigadores de humanidades no relacionados con la tecnología, en constante fase de desarrollo.

Mª Gimena del Río Riande investigadora, Secrit-Conicet. Estudio y edición digital de poesía. Introducción al enfoque de estudio de la poseía medieval. Se contempla inicialmente el análisis teniendo en cuenta el origen de los manuscritos: Pre-parentético (tradición oral) y pos-parentético (reuso de material a partir de la invención de la imprenta). *Poemetca*: red de conocimientos y proyectos relacionados con la métrica medieval castellana. Su labor de divide en dos grandes proyectos: Artes poéticas medievales e Intervenciones de diálogo en la poesía.

Clara Isabel Martínez Cantón profesora departamento de Literatura Española y Teoría de la Literatura ,UNED. Trabajo conjunto con Poemetca en el estudio estadístico de la poesía medieval castellana. Estas estadísticas y estudios contemplan metadados, rimas, tipo de versos, acentos, cláusulas, figuras estilísticas y retóricas, nombres de autoridades y fechas (influencias entre autores). Estos proyectos dialogan gracias al uso de sistemas standard, como es el *XML-TEI* enlazado con servidores de vocabulario semántico.

Stefano Versace Laboratorio de Innovación en Humanidades Digitales, UNED. Se centra en la extracción de datos de los textos a analizar. Para ello se ha enfrentado a la problemática de que haya un mismo concepto para diferentes formas poéticas. Esto puede ser nombres distintos para la misma forma o distintas formas para nombres parecidos. Para ello se ha buscado la estandarización de conceptos para poder llevar a cabo la extracción de datos adecuada. Por ejemplo se ha hecho uso de bases de datos de textos poéticos de prestigio e internacionales como *Remetca*, *CSM*, *Skaldic* o *Lyric German Poetry of the Middle Ages*. Sobre esta base se pretende analizar las características métricas y de qué manera se reflejan. Finalmente se evidencia la necesidad de encontrar unos estándares que faciliten el estudio de textos de diferente índole usando las mismas herramientas o procesos.

Capítulo 4. Taller: Tecnologías de marcado específicas para poesía: TEI-XML

Helena Bermúdez Sabel, Universidade de Santiago de Compostela .

Para configurar un texto en lenguaje natural, y pueda ser entendido por una máquina, este debe pasar por un lenguaje de marcado. Marcar un texto es modelar la estructura inherente y las propiedades semánticas de los documentos culturales a través de jerarquías y estructuras ordenadas. XML aúna la cualidad de poder ser leído por una máquina y también por un ser humano

Nota: Como prueba del aprovechamiento de este video, comento que este documento resumen ha sido generado mediante un docbook

Es muy interesante el uso de expresiones regulares para automatizar el marcado de textos

Concretamente el standard TEI (*Text Encoding Initiative*) desarrolla un standard que ayuda a codificar documentos culturales, manteniendo una coherencia entre la comunidad de investigadores y participantes en los proyectos. En su web oficial hay documentación extensiva del funcionamiento de las etiquetas. Visita obligada como manual de referencia.

Capítulo 5. Taller: TEI para el etiquetado de poesía. Otros módulos (edición crítica) y posibilidades

Helena Bermúdez Sabel, Universidade de Santiago de Compostela .

En este taller se profundiza en el uso del etiquetado TEI con varios ejemplos. Podemos adaptar TEI a nuestros propósitos modelando el XML schema, mediante la definición de vocabulario y la aplicación formal de restricciones.

Se hace hincapié en los pros del uso de TEI integrando todo lo necesario en un solo documento. Esto se consigue incluyendo los enlaces pertinentes, como al espacio de nombres.

Hay una serie de herramientas de corte internacional para la adaptación de TEI, como *Roma*(interfaz web) Hojas de transformación incluidas en *oXygen*, *OxGarage* (multitransformador desarrollado por la Universidad de Oxford), y *roma* (herramienta en linea de comandos).

Otros módulos son "Personografías y referencias geográficas y la descripción de manuscritos.

Se comenta en la exposición el uso de los identificadores y referencias de las etiquetas de TEI-XML para obtener listas y datos estadísticos. Otras utilidades es la obtención de referencias cruzadas y el desarrollo del aparato crítico.

Capítulo 6. Taller: Transformaciones y visualizaciones. Del etiquetado XML a la web

Juan José Escribano Santiago, Ingeniero Técnico en Informática de Sistemas, Universidad Politécnica de Madrid .

Este módulo introduce los conceptos relativos a la manipulación de archivos XML y su presentación visual. El XML como todo lenguaje de marcado, se le puede aplicar de una transformación para por ejemplo convertirlo a HTML, PDF, TXT u otros formatos que necesitemos, ya sea para una representación mas atractiva para lectores como transformaciones en otros XML para que sean usados por otros programas. La transformación standard de XML es el *XSLT*. Este standard se basa en las recomendaciones del consorcio *W3C*.

Este taller trabaja sobre una serie de ejemplos para comprobar la efectividad de las transformaciones. Primero para XML y posteriormente para las diferentes versiones de XML-TEI.

Cada archivo XML, se le desea aplicar una transformación, debe estar declarada en el encabezado del archivo, donde aparezca un link al archivo XSLT (ya sea en local, o remoto). Este archivo XSLT está compuesto por la serie de instrucciones que "transforman" el XML según el deseo del autor.

Capítulo 7. Taller: Aprovechamiento de las tecnologías semánticas y los recursos enlazados en el análisis de textos multilingües

Primer video por Víctor Rodríguez Doncel Ontology Engineering Group, UPM

Inicialmente se plantea la problemática actual de una web de documentos (Internet) que está orientada a la lectura por parte de los humanos y no por parte de las máquinas. Esto hace que la interconexión entre documentos en la web no esté bien desarrollada.

Para superar estas limitaciones se propuso la web semántica. De esta manera se podrían cruzar datos de diferentes servidores con búsquedas complejas. En este punto es importante comentar la importancia de la dbpedia, que es la wikipedia formateada para ser consumida por una computadora.

Un concepto base para la web semántica son los *Triples RDF* (esto es similar a los enfoques de modelado conceptual clásicos como entidad-relación o diagramas de clases, ya que se basa en la idea de hacer declaraciones sobre los recursos (en particular, recursos web) en forma de expresiones sujeto-predicado-objeto.)

Gracias a los Triples RDF podemos crear consultas que generen datasets compuestos de información de diferentes bases de datos distribuidas. La web semántica ha tenido una gran evolución en los últimos años, además esta compuesta siempre de datos abiertos que son legibles por computadora.

En este punto hay que comentar la importancia de las *Uris*, ya que conforman los RDFs que generan esta web semántica. La estandaricación de os RDFs es gestionada por el consorcio W3C. Los triples RDF son un modelo de datos. De esta manera se puede gestionar como una estructura escalable y ordenada mediante relaciones entre los datos. Los documentos nos pueden hablar de datos o nos pueden transmitir un conocimiento.

Ontología (computacional OWL) es la mejor manera de expresar el conocimiento de un dominio. Esto es, una explicación explícita y formal de una conceptualización compartida.

Segundo video por Elena Montiel Ponsoda, Ontology Engineering Group, UPM

Esta charla profundiza en el concepto del modelo de datos, que da lugar a la nube de datos enlazados. Para el propósito de este curso se comenta sobre la existencia de un subgrupo de esta gran nube de datos enlazados, los correspondientes a los datos lingüísticos (*LLOD cloud*). Existe una motivación de enlazar diferentes informaciones que hablen sobre un mismo concepto, para llevar a cabo este objetivo existe un modelo concreto llamando *lemon-ontolex*, que se usa para la representación de información lingüística multilingüe con respecto a una ontología. Este modelo es la estandarización y unificación de conceptos enlazables, después de varios años de desarrollo de diferentes proyectos, por parte de la W3C.

Lemon-Ontolex pretende superar una serie de limitaciones: describir morfosintacticamente entradas léxicas, capturar matices de significado y dar cuenta de la variación denominativa y de las traducciones.

Para llevar cabo consultas multilingües cabe destacar la base de datos *BabelNet*, y la aplicación *BabelFly*.

Capítulo 8. Taller: Posibilidades de los Linked Data para el análisis de textos y especialmente para la visualización geográfica: el caso del proyecto Pelagios

Pau de Soto, Institut de Estudis Catalans, IEC. La base de *Pelagios* es crear vínculos entre diferentes proyectos a partir de localizaciones geográficas. Para ello se han desarrollado una serie de herramientas específicas, aunque la principal baza del proyecto es la implicación de una amplia comunidad muy activa que crea estos etiquetados.

Para ello se realizan anotaciones (en RDF) y verificaciones de cada topónimo encontrado. La anotación se compone de una uri que identifica el topónimo de manera única. Este proceso se puede llevar a cabo tanto en textos escritos como en mapas. Estas anotaciones también sirven para enlazar diferentes documentos mediante consultas complejas por topónimos.

Los datasets pueden hacer uso de *archivos tipo VoiD* (vocabulary of Interlinked Datasets) compuestos por RDF como puente entre Linked Data Projects.

Para crear y editar las anotaciones por parte de investigadores y usuarios se creó la aplicación *Recogito*. El programa permite la implementación de anotaciones geográficas tanto en textos como en imágenes. También se ha creado Pelagios API como un servicio libre para hacer uso de las funcionalidades de Pelagios por herramientas externas.

Se ha extendido la funcionalidad del sistema para vincular no solo topónimos, si no también hechos y personas. Finalmente se invita a la comunidad de interesados a participar en la inclusión de datos usando el sistema Recogito.

Capítulo 9. Taller: Cómo crear un perfil de aplicación de datos para el análisis poético

Mariana Curado Malta, Universidad de Oporto, Portugal. Un perfil de aplicación de metadatos(MAP) es un conjunto de elementos(de schemas de metadatos), de restricciones (reglas sobre los datos y de guías de aplicación para un contexto específico de aplicación.

Un schema de metadatos es lo mismo que un vocabulario RDF, o sea, un conjunto de términos para describir cosas, que pueden ser clases (una entidad, resultado de modelar la realidad mediante abstracciones) o propiedades.

Un objetivo del MAP es la interoperatibilidad, comunicaciones entre los datos sin intervención humana. Esto hace dinámica la web de datos y resultan utilizaciones no esperadas de los datos. Inicialmente no se sabe quienes van a contribuir al desarrollo de esta web de datos interconectada y semántica, por ello hay que establecer un método estandarizado de colaboración.

Para desarrollar primero necesitamos definir los Requisitos Funcionales. Esto es las funcionalidades del sistema, por ejemplo para listar poemas podemos empezar por gestionar la funcionalidad Idioma y la funcionalidad Rima. Para listar una serie de requisitos funcionales (RF) podemos hacer uso de las bases de datos y documentación existentes, hacer entrevistas u observar el trabajo de otro y hacer casos de uso.

Existe actualmente un ejemplo aplicado a bibliografías, esto es el Functional Requirements for bibiographic Records, es interesante estudiar este método para ver como se ha desarrollado un caso de éxito de implementación de un modelo conceptual para representar datos, en este caso, recursos bibliográficos. Otros recursos importantes de consulta son el TEI o la Biblioteca Nacional de España.

Como segundo paso debemos definir un modelo de datos, también conocido como modelo de dominio. Para ello debemos identificar las cosas que lo componen. Para ser un componente del dominio las cosas deben tener propiedades propias y relaciones entre ellas. También se deben identificar las restricciones que definen cada cosa.

En este punto definimos *Clase* como un constructor que representa cosas en el mundo, y las *Relaciones* entre ellas. Para diseñar estas clases y relaciones que definen cada dominio hay una serie de técnicas y herramientas como UML, ER, ORM, Grafo RDF, entre otras.

El tercer paso es definir un Description Set Profile (DSP). Esto se hace explorando el entorno, agrupando vocabulario técnico (ontologías) basándonos en standards internacionales previamente y testeando la matriz de restricciones expresamente diseñada para el DSP

Capítulo 10. Taller: Quantitative research on versification: the corpus of czech verse

Petr Plechac, Institute of Czech Literature, Czech Academy of Sciences. Corpus of Czech Verse (CCV) es una base de datos interactiva con la poesía de la República Checa del siglo 19 y el principio del siglo 20. Tiene diferentes niveles de anotaciones como métrica, rima, transcripciones fonéticas y morfológicas entre otras. Ofrece también una serie de herramientas online con diferentes propósitos: por ejemplo una da acceso directo mediante consola de comandos, otra para búsqueda de estadísticas sobre rimas o métricas o la búsqueda de palabras clave. Dos herramientas importantes son Gunstick y Hex:

Gunstick es una herramienta especializada de CCV para la búsqueda de rimas por palabra, pudiendo mostrar frecuencia, distribución en el tiempo y lista de los versos donde la rima aprece junto con los enlaces a los poemas encontrados.

Hex es una herramienta para búsqueda de palabras concretas en la base de datos. También puede mostrar estadísticas de aparición y uso

Capítulo 11. Taller Análisis de textos poéticos y estilometría con R

Salvador Ros Muñoz profesor del Departamento de Sistemas de Comunicación y Control, UNED y Antonio Robles Gómez profesor del Departamento de Sistemas de Comunicaciones y Control, UNED .

- Introducción al trabajo práctico

  Para este taller se tiene disponible un servidor privado con RStudio para trabajar con los ejercicios proporcionados. Inicialmente se usará la librería *CoreNLP* (que se encuentra instalada en el servidor. Primero se carga e inicializa la librería y después se carga el archivo PoesíaIngles.txt

- Introducción al procesamiento del lenguaje natural (Natural Language Processing, NLP)

  NLP consiste en la interacción persona-ordenador a través del lenguaje. Entre los objetivos y retos a los que se enfrenta están la Ambigüedad y variabilidad del lenguaje. También es complicado tratar con la escalabilidad del lenguaje natural cuando sumamos idiomas, frases largas, dominios ... etc

- Pasos de NLP:

  - Tokenización y sentence splitting. Consiste en que partiendo de nuestro texto, lo dividimos en tokens que consisten en palabras. El siguiente paso es dividir el texto en frases.

  - Lemmatization y POS tagging. Lematizar es producir la forma canónica de las palabras, lo que se conoce como su lema. El proceso de lematización es distinto para cada palabra según sea nombre, verbo, ... El POS Tagging consiste en asignar a cada token una etiqueta que identifica su categoría. Por ejemplo verbo, Pronombre personal, ... Es importante comentar los estandars ya activos como son Penn Tree Bank Project y Universal tag-set.

  - Dependencias. Aquí se analica de forma gramatical de las frases, creando una estructura que enlace las partes identificadas anteriormente. El resultado de esta fase es un árbol de dependencias. Las relaciones son binarias identificando un Governor Y un Dependent. Además, cada relación está etiquetada con un código que identifica dicha relación. Estos códigos también son un standard e identifica el tipo de relación.

  - Reconocimiento de entidades con nombre. Esta fase consiste en identificar los elementos del texto y asignarles una categoría semántica. Por ejemplo, fecha, duración, persona, lugar ...

  - Correferencias. Sirven para identificar las palabras que se refieren a la misma persona o al mismo objeto. De esta manera se extraen relaciones semánticas entre los tokens.

- Introducción al análisis de datos textuales. Partiendo de las anotaciones de NLP, se pueden aplicar una serie de técnicas para explorar y visualizar un corpus de documentos de texto.

- Estilometría. Hay dos características comunes en los estudios de estilometría: los textos se interpretan numéricamente y los números se analizan estadísticamente. El lenguaje R ofrece una serie de paquetes que llevan a cabo esta labor, como Stylo(), que permite cargar y procesar un corpus de textos. También realiza un análisis multivariable estilométrico y permite visualizar y evaluar los resultados por frecuencias de palabras. Esto es procesar: reconocer palabras y asignarles una representación (información) que permita poder tomar decisiones. Representación es añadir (substituir) a la cadena de caracteres que forman una palabra (o secuencia de palabras) información explícita de sus características para una tarea determinada.

Capítulo 12. Taller: Procesamiento del lenguaje natural (PLN) y sus aplicaciones en poesía

Nuria Bel, Universitat Pompeu Fabra. El objetivo base es realizar tareas del tipo leer, entender y extraer información. Ya hay herramientas que pretenden interpretar un texto y resultar si es positivo o negativo. El programa que identifica las palabras positivas de las negativas se basan en un diccionario inicial donde aparece cada palabra con su correspondiente etiqueta que la valora.

En PLN, las tareas mas conocidas son:

- Análisis de opinión
- Corrección gramatical
- Traducción automática
- Búsqueda y recuperación de información
- Extracción de información
- Resumen automático
- Respuesta a preguntas y asistentes virtuales
- Análisis lingüístico

Para alcanzar estos objetivos es necesario dominar una diferentes técnicas. En formato y codificación de textos por ejemplo, se debe identificar qué es una palabra, un carácter, un código o un espacio. Es importante basarse en el standard UTF8 que representa todos los caracteres de los idiomas a nivel internacional, siendo el formato txt el más seguro ante posibles incompatibilidades. Una vez que tenemos el texto en formato utf8 podemos programar un algoritmo que cuente palabras, o que las combinaciones (bigramas) de palabras mas habituales o significativas, reconocer nombres propios y clasificarlos, anotaciones morfosintácticas.

Destacan las herramientas libres y online Contawords y Voyant-Tools.

Capítulo 13. Análisis del soneto castellano del Siglo de Oro con métodos computacionales

Borja Navarro Colorado, Universidad de Alicante . El análisis computacional de textos ya empezó a principios del siglo XX, aunque es ahora cuando está tomando madurez y presencia en círculos académicos humanistas. La mayor novedad y motivación de análisis computacional es la escalabilidad que podemos alcanzar con la tecnología hoy en día.

El proceso de estudio se compone de una primera parte donde se compila y anota el corpus, y una segunda parte donde se implementa el modelo de análisis métrico y semántico. El marco metodológico aplicado es una aproximación tradicional con un análisis en profundidad de autores canónicos, y una aproximación computacional con un análisis automático de todos los autores, siendo este último un método distante y a gran escala.

Para analizar textos tenemos el ejemplo de Distant Reading de Moretti donde se busca lo común en la Historia de la Literatura para analizar amplios periodos como un todo. Aquí se busca la objetividad de los datos usando un método cuantitativo.

Otra aproximación al análisis computacional la encontramos en el Macroanálisis de Jocker, que usa la técnica Topic Modeling (Text Mining) para un análisis inmamentista en la novela del siglo XIX en inglés.

La idea base del análisis a gran escala es la Búsqueda de los general y los rasgos comunes a todo el periodo de estudio. El proceso consiste en una extracción de datos con objetividad y anotados usando un corpus de referencia. Después se trata de llevar a cabo un análisis de datos sobre frecuencias y técnicas Text Mining.

Un corpus es una amplia colección de textos digitales, compilado en función de unos criterios, representativo de un hecho, periodo, fenómeno, etc, y anotado para preservar aspectos lingüístico-literarios profundos. Para el caso del Corpus de sonetos del Siglo de Oro (SdO), los criterios de compilación son que sean sonetos en castellano y del siglo XVI y XVII. Los criterios de anotación son que sean metadatos, anotación estructural y anotación métrica.

Este estudio en concreto se ha llevado a cabo extrayendo los sonetos de una base de datos HTML e implementándolos en XML-TEI mediante expresiones regulares para que fuera un proceso automático. Este proceso de anotación se debe hacer de manera consistente, para ello se debe elaborar una guía de anotación y esta debe ser similar a la que podría llevar a cabo otros investigadores. Para la anotación semiautomática llevada a cabo para este proyecto, se inició con una anotación automática(Python , XML y expresiones regulares), después se sigue con una revisión manual y finalmente con una revisión de los versos erróneos.