

Helena Bermúdez Sabel

Anotación multicamada externa e o enriquecemento de edicións dixitais

Multi-layered stand-off annotation and the enrichment of digital scholarly editions


Resumo: Desde o punto de vista tecnolóxico, non existen límites no que respecta á información que un editor/a pode acrecentar a un texto. Porén, a codificación dos múltiples e heteroxéneos datos propensos a seren integrados nunha edición dixital non é tan simple e directa para quen se inicia neste campo. Xeralmente, recórrese á práctica de codificación máis estendida: a etiquetaxe na liña. Esta estratexia consiste na introdución de elementos dentro do documento fonte que son, xunto cos seus atributos, os que conteñen a información estrutural e analítica do texto. Este método é perfectamente axeitado para os modelos ecdóticos máis simples. No entanto, de quereremos introducir múltiples capas de información, a marcación na liña é ineficiente. Nesta achega presentaremos un sistema de codificación externa que permitirá aplicar unha anotación editorial conformada por capas de datos múltiples e xerarquizadas, coa fin de deseñar un aparato crítico complexo, flexible e procesábel.

Palabras chave: edición dixital; análise cuantitativa; variación; XML; TEI.

Abstract: From a technological point of view, there are no limits to the amount of information an editor can add to a text. However, the encoding of the kind of multiple and miscellaneous data that is prone to be embedded in a digital edition might not be straightforward for inexperienced editors. Most beginners apply the most common method, that is, inline encoding. This technique entails the direct introduction of elements into the primary document which, together with their attributes, contain the structural and analytical information of the text. Such a method is perfectly adequate for simple ecdotic models. However, it becomes less efficient when we want to introduce multiple (and overlapping) layers of information.

Nota: Esta contribución foi realizada ao amparo do proxecto *Paleografía, Lingüística y Filología. Laboratorio on-line de la lírica gallego-portuguesa* (FFI2015-68451-P) financiado polo Ministerio de Economía, Industria y Competitividad.

Helena Bermúdez Sabel, USC – Laboratorio de Innovación en Humanidades Digitales, UNED

Open Access. © 2019 Helena Bermúdez Sabel, published by De Gruyter.  This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
<https://doi.org/10.1515/9783110585421-002>

Unauthenticated
Download Date | 6/1/19 2:08 PM

In this contribution, we present a stand-off annotation method which allows multi-layered and hierarchical annotation with the goal of designing a complex, flexible and processable critical apparatus.

Keywords: Digital Edition; Quantitative Analysis; Variation; XML; TEI.

1 Introducción: conceptos básicos

1.1 A marcación

Unha edición dixital está, xeralmente, construída sobre unha linguaxe de marcación, é dicir, os seus contidos están codificados de maneira que unha máquina poida lellos e interpretalos. Marcar, por tanto, consiste en introducir datos que serven para identificar certos trazos e características lóxicas ou físicas do obxecto textual, así como para controlar o seu procesamento posterior (Renear 2004: 219). Para que eses datos poidan ser interpretados correctamente, deben ser formalmente diferentes da secuencia de caracteres que conforma a transcripción dixital dun texto: isto é o que coñecemos como linguaxe de marcación.

Marcar consiste en crear unha estrutura informativa que formaliza e describe a fonte material. A persoa que codifica ten a capacidade de explicitar a natureza e a función de cada fragmento de información –segundo á interpretación que dela fai. Este paso fai posíbel que un computador sexa capaz de “comprender” e utilizar ese obxecto dixital de maneiras que aproveitan o seu autoconhecimento sobre a súa propia estrutura e contidos (Flanders 2011: 68).

Cando se lle presenta un poema a unha persoa, esta pode recoñecer que se trata dunha composición poética antes de comezar a súa lectura: existe unha serie de convencións na presentación do texto que fornecen esa información, como os saltos de liña delimitando os versos ou un incremento do interliñado para separar as diferentes estrofas. Porén, para un computador, un texto apenas é unha sucesión de caracteres (no fin transformados a código binario). O que a marcación nos ofrece é a posibilidade de explicitar os contidos que corresponden con cada verso e cales son os versos que conforman cada unha das estrofas. Ademais, podemos introducir calquera tipo de información analítica como, por exemplo, a descomposición sintáctica do texto. Isto permitiría que o poema puidese ser procesado como un conxunto de sintagmas con diferentes funcións que, á súa vez, se estruturan en versos e estes en estrofas. Cada un deses sintagmas pode ser accedido sistematicamente, polo que podemos considerar que un texto marcado é semellante a unha base

de datos en que as marcas funcionan como os diferentes campos de pesquisa habilitados.

Por tanto, a través das linguaxes de marcación podemos representar dixitalmente modelos altamente complexos de artefactos textuais, sobre os cales se poden realizar procedementos de análise informática: lematización, anotación morfosintáctica, modelado de tópicos, estudos estilométricos e un longo etcétera.¹

1.2 XML

XML (*eXtensive Markup Language*, Linguaxe de Marcación Extensible) é unha das linguaxes de marcación de uso máis estendido. Prové de SGML (ISO 8879) e é regulado polo World Wide Web Consortium (W3C), unha comunidade internacional cuxa misión consiste no desenvolvemento de estándares para a Web (Acera García 2011: 32).

XML consiste nun modelo formal baseado nunha xerarquía ordenada, ou o que é o mesmo, nunha árbore. Na Figura 1.1, preséntase un documento XML² xunto coa representación gráfica da súa estrutura. Todos os elementos que o conforman teñen unha posición e unha orde na xerarquía. Este modelo fornece un autocoñecemento inherente que permite que, por exemplo, poidamos preguntar cal é a última unidade frásica de cada verso, sen necesidade de explicitar nin o número de verso nin a orde das unidades porque a posición é constitutiva das linguaxes de marcado.

O modelo en árbore obriga a que cada elemento teña que estar perfectamente aniñado, é dicir, non pode haber contidos que comecen nunha “rama” e terminen noutra. Esta limitación fai que, por veces, quen codifica teña que procurar estratexias que eviten o *solapamento* de elementos.³ Como se verá na seguinte subsección, a anotación externa é unha desas técnicas.

¹ A modo de exemplo, véxase *DanteSearch* (<http://perunaenciclopediadantescadigitale.eu:8080/dantesearch>, accedido 21/05/2018) como modelo dun corpus XML que incorpora unha análise lingüística de corte morfosintáctico e léxico. Para ilustrar unha das múltiples posibilidades dentro do campo literario, véxase *Spectacular Intersections of Place in Southey's "Thalaba the Destroyer"*, en que se utiliza a información semántica codificada no XML para implementar análise de redes sobre os paratextos (<https://rebrand.ly/thalaba>, accedido 21/05/2018).

² Os contidos do poema correspóndense cos primeiros versos do poema “María Soliña”, de Celso Emilio Ferreiro (1962).

³ Véxase Renear *et al.* (1993) para unha aproximación teórica da definición de texto vinculada á utilización de linguaxes de marcación. Desde a década de 90 producíronse enriquecedores debates que colocan a cuestión de que é un texto en relación coa súa modelación no soporte dixital.

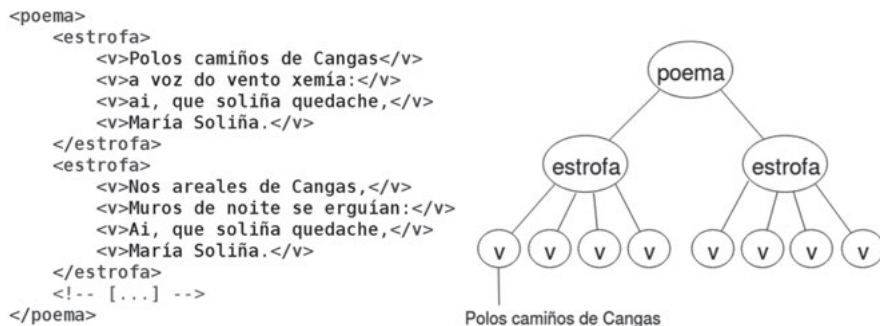


Figura 1.1: Exemplo dun poema marcado en XML e a súa representación gráfica.

1.3 Anotación externa

Chamamos anotación externa (en inglés, *stand-off annotation*) á anotación dun documento primario que non se insire no fluxo do texto (*inline*). Xeralmente, desenvólvese nun arquivo (ou arquivos) diferente do que contén os datos textuais primarios (Bański 2010). O proceso mediante o cal se relaciona o documento primario coa súa anotación presupón a vinculación de diferentes localizacións da fonte primaria coa información que a describe, for esta ligazón a través de *byte offsets*, elementos, atributos ou outros métodos (Ide/Romary 2004: 218).

Este método presenta varias vantaxes. Por una parte, pódense crear conxuntos de anotacións moi complexos tomando como referencia calquera unidade lingüística (desde o fonema á frase, por exemplo), e mesmo combinando varios tipos de análises sen que haxa ningún problema de solapamento. Por outra parte, ao separar o texto primario da súa anotación, facilitamos a reutilización do primeiro: trátase dun documento lexíbel, apenas cos metadatos esenciais, que podemos considerar máis “neutro” ao carecer da información máis analítica e interpretativa. Favorécese así que outros membros da comunidade científica poidan facer outras interpretacións ou completar un traballo a partir do mesmo documento base.

Unha das máis recentes define o texto como un grafo, modelo para o cal nin solapamentos nin transposicións resultan problemáticos (Dekker/Birnbaum 2017).

2 Presentación dun caso de uso: análise da variación textual

De estarmos editando unha tradición textual con máis dun testemuño, a referencia ás diferenzas textuais entre os relatores é ineludíbel. Evidentemente, dependendo do marco teórico en que nos coloquemos, definiremos estas variantes dunha ou outra maneira polo que tamén se desenvolverán tarefas diferentes para a súa detección e análise. O que presentaremos nas seguintes páxinas aspira a ser un método flexíbel para a codificación de edicións dixitais sen entrar nunha definición específica de texto e variante nin na defensa de determinadas correntes ecdóticas.

O motivo polo que nos centraremos na problemática das variantes entre testemuños é, precisamente, pola complexidade do seu modelado nalgúns tradicións. Ademais, é un aspecto da edición de textos que presenta unha grande potencialidade na súa publicación en soporte dixital (Pierazzo 2015: 25–29).

A crítica textual é moi rica en relación co papel das variantes no resultado final da edición, mais, xeralmente, tanto escolas que defenden a materialidade do texto –por exemplo, Cerquiglioni (1989)– como aqueles métodos que procuran un texto crítico –por exemplo, Chiesa (2002)– requiren da análise de variantes. Porén, a información que queiramos recoller a respecto da propia existencia e (posterior) descrición das variantes pode ser moi heteroxénea. Por exemplo, podemos establecer un recoñecemento de variantes centrado na categorización de erros conxuntivos e separativos coa fin de establecer as relacións de filiación entre os testemuños e, consecuentemente, da construción do *stemma* cando for posíbel ou de *stemmata* locais en tradicións con altos niveis de contaminación (Mink 2000). Noutros casos, o noso estudo pode estar centrado na procura de pistas para unha identificación de copistas, editores e outros axentes que participaron durante a transmisión da obra ou na alteración dos seus contidos, coa vontade de recoller a motivación que hai detrás desas alteracións. Noutros casos, ante a elaboración dunha edición xenética, a variación é chave para analizar o proceso creativo de quen participa na súa autoría e/ou edición.

Apenas se forneceu unha pequena mostra dos posíbeis intereses filolóxicos ante a análise de variantes coa fin de mostrar o multifacética que pode ser a codificación destes fenómenos. No caso de uso que presentaremos, falarase dunha taxonomía de variantes. Porén, a información que se pode retirar da análise de variantes é tan dependente das circunstancias do texto e dos intereses do ou da investigadora, que unha taxonomía xenérica que englobe esa complexidade e

recolla todos os matices sería inconcibíbel. Por tanto, o obxectivo desta contribución é proporcionar un mecanismo de codificación que formalice, para calquera fin, as variantes textuais.

Na sección 3, presentarase un proposta de marcación de variantes en que se asocia a súa descrición a través do uso dunha taxonomía de variantes. Esta proposta é compatíbel coas recomendacións da *Text Encoding Initiative*. TEI é un consorcio internacional dedicado a manter as directrices para a codificación de textos, coñecidas como *TEI Guidelines*, e na actualidade deseñadas como un vocabulario XML.⁴

3 Metodoloxía

3.1 TEI e as edicións filolóxicas

As directrices de TEI non teñen un forte carácter prescriptivo, polo que presentan unha grande flexibilidade, permitindo que se poidan marcar os mesmos fenómenos de diferentes maneiras segundo a interpretación e o obxectivo de quen codifica.

Na hora de marcar as diferentes lecturas dos testemuños, tamén conviven diferentes formas de identificalas, así como diversos métodos para vincular a información do aparato co texto (TEI Consortium 2017a). Dos métodos dispoñíbeis, o chamado método de segmentación en paralelo (*parallel segmentation method*) parece ser o máis popular entre a comunidade científica (Burghart 2016: §1.3).

O método de segmentación en paralelo consiste en inserir un elemento <app> por cada unha das unidades de variación,⁵ isto é, por cada *locus* no texto en que polo menos concorren dúas leccións. Cada unha das leccións é, á súa vez, inserida en cadanseu elemento <rdg>⁶ (ou <lem> se se desexa establecer ou emendar unha das leccións)⁷ cun atributo que identifica o testemuño ou testemuños que presentan esa determinada lección (Figura 1.2).

⁴ Para máis información, visítese o seu portal web (<http://www.tei-c.org>) e véxase Cummings (2007).

⁵ O concepto de unidade de variación é tomado de De Vos *et al.* (2012: 113).

⁶ Neste documento, seguindo a convención máis habitual, cítanse os elementos entre parénteses angulares e os atributos precedidos da arroba.

⁷ Tamén existe a posibilidade de que dentro de <app> se poidan abrir un ou varios <rdgGrp> que agruparían as leccións dos testemuños.

```

<l n="1">
  <app>
    <lem wit="#El #Hg">Experience</lem>
    <rdg wit="#La">Experiment</rdg>
    <rdg wit="#Ra2">Eryment</rdg>
  </app> though noon Auctoritee </l>
<l>Were in this world ...</l>

```

Figura 1.2: Exemplo de marcación de variantes en TEI (TEI Consortium 2017a: §12.2.3).

De querer especificar a motivación atrás dunha variante, os exemplos dispoñíbeis nas *TEI Guidelines* presentan o atributo `@type` para clasificar a lección (con valores como “substantiva” ou “gráfica”) e o atributo `@cause` para categorizar a orixe da variante (suxerindo valores do tipo “homeoteleuton” ou “falsa emenda”). Porén, como se expuxo na sección 2, o estudo da variación pode facerse desde múltiples puntos de vista, polo que estes dous atributos poden resultar insuficientes para modelar esa complexidade. Esta circunstancia motiva a presente achega.

3.2 Taxonomía de variantes

Unha estrutura de trazos (*feature structure*) consiste nun conxunto de pares *atributo:valor*, no que os valores poden ser atómicos ou outras estruturas de trazos (aniñadas). É dicir, un elemento é descrito a partir dunha serie de propiedades ás que se lles dá un valor. No entanto, este valor pode estar conformado, á súa vez, por diferentes propiedades cos seus correspondentes valores individuais. Podemos considerar, pois, as estruturas de trazos como un método xenérico para a estruturación de datos, que presenta numerosas vantaxes derivadas do seu uso como formato de meta-representación (Witt/Stegmann 2009).⁸

A Figura 1.3 mostra un exemplo dunha estrutura de trazos, conformado polas propiedades “caso”, “xénero” e “singular”, cuxos valores son “acusativo”, “feminino” e “falso”. Neste simple exemplo, cada un dos trazos está ao mesmo nivel, mais poderíamos crear unha xerarquía moito máis complexa, de aí que as estruturas de trazos sexan un modelo especialmente axeitado para a codificación de taxonomías.

⁸ Entre as súas vantaxes poderíase falar tamén da estabilidade e éxito do modelo, o cal foi adoptado en 2012 como estándar internacional (ISO 24612). Para máis información sobre este aspecto, véxase Romary (2015). No que se refire ás estruturas de trazos e TEI, o módulo relativo a este modelo fai parte das guías desde a súa versión P3 (Sperberg-McQueen/Burnard 1994). Para unha discusión sobre as vantaxes de usar as estruturas de trazos, véxase Pose *et al.* (2014: 9–10).

```

<fs>
  <f name="case">
    <symbol value="accusative"/>
  </f>
  <f name="gender">
    <symbol value="feminine"/>
  </f>
  <f name="singular">
    <binary value="false"/>
  </f>
</fs>

```

Figura 1.3: Exemplo dunha estrutura de trazos (TEI Consortium 2017b: § 18.3).

Nas Figuras 1.4 e 1.5, temos dúas entradas dunha posíbel taxonomía de variantes en que se describen sendas variantes de lingua, concretamente, unha nasalización progresiva (Figura 1.4) e unha prótese vocálica (Figura 1.5). Á diferenza do exemplo na Figura 1.3, algúns dos trazos utilizados para describir estes fenómenos están, á súa vez, conformados por outras estruturas de trazos. As características utilizadas para a descrición dos fenómenos de variación dependerán dos elementos que se queiran pesquisar ou cuantificar posteriormente. O feito de que nestes exemplos das Figuras 1.4 e 1.5 se utilicen os trazos do proceso fonético (por exemplo, adición, redución, alteración), a posición así como a presenza ou

```

<fs xml:id="pro-nas">
  <f name="taxonomia">
    <fs type="linguistica">
      <f name="categoria">
        <fs type="fonetica">
          <f name="proceso">
            <fs type="alteracion">
              <f name="modo">
                <fs type="asimilacion">
                  <f name="procedemento">
                    <symbol value="nasalizacion"/>
                  </f>
                  <f name="direccion">
                    <symbol value="progresiva"/>
                  </f></fs>
                </f></fs></f>
              </f></fs></f>
            <f name="posicion">
              <symbol value="final"/></f>
            <f name="constriccion">
              <binary value="si"/></f>
          </fs></f>
        </fs></f>
      <f name="descripcion">
        <string xml:lang="gl">Nasalización progresiva</string>
      </f>
    </fs>
  </fs>

```

Figura 1.4: Exemplo de entrada de taxonomía de variantes (producción propia).


```

<fs xml:id="protese">
  <f name="taxonomia">
    <fs type="linguistica">
      <f name="categoria">
        <fs type="fonetica">
          <f name="proceso">
            <symbol value="adicion"/>
          </f>
          <f name="posicion">
            <symbol value="inicial"/>
          </f>
          <f name="constriccion">
            <binary value="non"/>
          </f>
        </fs>
      </f>
    </fs>
  </f>
  <f name="descripcion">
    <string xml:lang="gl">Prótese</string>
  </f>
</fs>

```

Figura 1.5: Exemplo de entrada de taxonomía de variantes (producción propia).

non de constricións poden ser relevantes para o estudo dunha tradición poética verso-silábica xa que todos eles poden alterar o cómputo silábico.

A través dos exemplos que ilustran esta sección preténdese transmitir a grande potencialidade do modelo. Por unha parte, permite chegar a un grande nivel de concreción a base da superposición de elementos mínimos. Cada entrada da taxonomía pode estar conformada por tantas características foren precisas, podendo mudar o nivel de granularidade segundo as nosas necesidades. Por exemplo, para a descrición do proceso fonético de adición (Figura 1.5) non foi preciso acrecentar máis trazos para describir esta propiedade. Pola contra, na Figura 1.4, en que o proceso fonético afectado é a asimilación, o valor é representado como unha estrutura de trazos con dúas propiedades. Por outra parte, calquera destas descricións pode ser reutilizada en entradas posteriores, factor que facilita tanto o proceso de conformación da taxonomía como a súa coherencia interna. Tomando como referencia as entradas presentadas nas Figuras 1.4 e 1.5 poderían simplificarse referenciando un trazo que significase “variante lingüística de tipo fonético”, xa que ambas as dúas entradas partillan os valores desas propiedades. Aínda máis, se temos unha serie de trazos para describir un fenómeno fonético por adición dunha vogal, poderíamos describir a prótese e a epéntese só tendo que engadir un trazo como o de “posición” para diferenciarlos.

Outra das vantaxes das estruturas de trazos é a posibilidade de declarar como debe ser a xerarquía de cada entrada, que trazos son obrigatorios e cales optativos e que valores posibles poden ter cada un deles (TEI Consortium 2017b: §18.11). A través desta declaración, a coherencia e validez da taxonomía no seu conxunto están garantidos.

3.3 Implementación

Como xa se anunciou, a ligazón da taxonomía á edición consiste nunha anotación externa. A edición é o documento primario mentres que na taxonomía se recolle a anotación secundaria.

Na Figura 1.6, móstrase un exemplo de como pode facerse a ligazón entre ambos os documentos. Neste caso, optouse por una vinculación semántica a través dun atributo: cada entrada da taxonomía conta cun identificador único (o valor do atributo `@xml:id` nas Figuras 1.4 e 1.5) que despois é referenciado no aparato de variantes en que ese fenómeno acontece a través do atributo `@ana` (*analysis*). Por tanto, no verso presentado na Figura 1.6, temos tres unidades de variación en relación con dous testemuños se tomarmos a palabra (ou *token*) como unidade de referencia, de aí que haxa un elemento `<app>` por cada unha delas. No entanto, nese primeiro *locus* de discordancia, encontramos dúas variantes: unha de lingua no testemuño identificado como *A*, valor do atributo `@wit` (*witness*), e outra gráfica no testemuño *B*. Na primeira, cuxa variación é referenciada como “pro-nas”, remítese para a entrada da taxonomía relativa á nasalización progresiva. O elemento `<seg>` é utilizado para segmentar o *token* e especificar que caracteres son motivo da variación, co fin de evitar ambigüidades. Neste caso, a presenza de `<seg>` marcando a letra `⟨n⟩` informa de que a variante identificada no atributo `@ana` radica unicamente na presenza dese carácter, mais se houberse dous ou máis fenómenos no mesmo *token*, cada un dos elementos `<seg>` tería que conter un atributo co identificador correspondente para poder discernir que compoñentes da cadea gráfica se corresponden con cada elemento de variación. Cumpre esa mesma función o elemento marcando o `⟨j⟩` na lectura de *B* (presenza de `⟨j⟩` para a representación da vogal fechada anterior seguida de *minims*).

```
<l n="3">
  maior ca
  <app>
    <rdg wit="#A" ana="#pro-nas">mi<seg>n</seg></rdg>
    <rdg wit="#B" ana="#j-minim">m<seg>j</seg></rdg>
  </app>
  <app>
    <rdg wit="#A" ana="#reg">senhor</rdg>
    <rdg wit="#B">senhor</rdg>
  </app>
  nen outra
  <app>
    <rdg wit="#A">ren</rdg>
    <rdg wit="#B" ana="#abb">rē</rdg>
  </app>
</l>
```

Figura 1.6: Exemplo de edición dun verso con variantes (producción propia).

4 Conclusión

A modo de conclusión, expóranse as vantaxes do método presentado nestas páxinas.

Por unha parte, algúns dos beneficios do método concretízanse durante o proceso de elaboración dos materiais fonte, é dicir, da edición e da taxonomía de variantes. A separación do texto primario da súa anotación axuda a garantir a sustentabilidade do proxecto e a súa actualización. Por exemplo, é posíbel aumentar o nivel de granularidade, de profundidade, con que se describe cada unha das entradas da taxonomía, sen que iso implique facer modificacións na edición. O método de anotación externa presentado na Figura 1.5 é dos máis “intrusivos” no documento primario, mais, aínda así, este é lexíbel grazas a que a carga analítica está apenas presente no valor dun atributo e, en consecuencia, a edición sería facilmente reutilizábel por persoas alleas á súa elaboración inicial. Este factor, xunto co uso dun estándar como TEI, incrementa a sustentabilidade, recíclaxe e reutilización do código.

Por outra parte, existen todas as vantaxes que xorden durante o procesamento posterior, tanto na hora da publicación como da análise.

No que se refire á publicación, a taxonomía de variantes pode ser explorada para presentar equivalentes ao aparato crítico en soporte dixital, así como para a tradución da información da taxonomía en elementos visuais. Cada unha das camadas que conforman unha entrada da taxonomía pode ser accedida e utilizada para enriquecer a edición. Na Figura 1.7 ofrecemos un exemplo en que a presenza de determinados trazos é transformada en cambios no formato do texto. Por exemplo, aquelas variantes que fosen clasificadas na taxonomía como variantes gráficas aparecen en negriña, e as variantes de lingua en anil. Polo tanto, para a inserción de elementos visuais na edición, apenas estamos a utilizar a información dun dos trazos (tipo de taxonomía). Porén, poderíamos servirmos de calquera dos elementos que conforman a definición da variante. Como se pode observar na Figura 1.7, a descrición máis específica do fenómeno

1.	A	Pois	contra	uos	non	me	ual	mia	sennor
	B	Poys	contra	uos	non	mj	ual	mha	senhor
2.	A	de	uus	seruir	nen	de	uus	querer	ben
	B	de	u9	fuir	nen	de	u9	q̄rer	bem
3.	A	mayor	ca	mī	sennor	nen	outra	ren	
	B	mayor	ca	mj	Nasalización progresiva				

Figura 1.7: Exemplo de edición enriquecida con elementos visuais (producción propia).

é convertida nun diálogo emerxente que aparece ao pasar o cursor por riba do texto afectado.

No relativo á análise, un método que fai uso das estruturas de trazos tal e como se presentou permite a realización de estudos cuantitativos e pesquisas dunha maneira moi eficiente. Faise posíbel medir a concentración dunha determinada variante, ou apenas daquelas que respondan a un tipo en particular utilizando como criterio para a cuantificación calquera dos metadatos dispoñíbeis. Tendo en mente unha taxonomía con entradas que seguen unha estrutura semellante á presentada nas Figuras 1.4 e 1.5, poderíanse extraer da edición todas as variantes de lingua ou limitarse a aquelas fonético-fonolóxicas. Tamén é posíbel facer pesquisas máis específicas e recuperar todas as instancias en que houbo unha adición nalgunhas das lecturas dos relatores, for a través dun fenómeno lingüístico ou dun erro de copia. Con este tipo de datos pódese calcular de maneira inmediata a frecuencia dunha variante ou tipo de variante utilizando como variables elementos codicolóxicos como o manuscrito, caderno ou folio. Por exemplo, na Figura 1.8 observamos como alén de recuperar as ocorrencias dun determinado fenómeno, podemos estudar a distribución desa mesma variante por testemuño. É evidente que estudos que analizan a frecuencia de determinados fenómenos para resolver incógnitas sobre o proceso de transmisión dunha obra non son algo novidoso. Porén, o que se quere defender nesta achega é o deseño de modelos de edición nativos ao soporte dixital que permitan elaborar este tipo de análises cuantitativas dunha maneira rigorosa e eficiente.



Figura 1.8: Exemplo de resultados de pesquisa por fenómeno (producción propia).

Referencias bibliográficas

- Acera García, Miguel Ángel (2011): *Manual imprescindible de XML. Edición 2012*. Madrid: Anaya.
- Bański, Piotr (2010): “Why TEI stand-off annotation doesn’t quite work: and why you might want to use it nevertheless”. En: *Proceedings of Balisage: The Markup Conference 2010*. Vol. 5. [En liña, <https://doi.org/10.4242/BalisageVol5.Banski01>, consultado en 14/01/2018].
- Burghart, Marjorie (2016): “The TEI Critical Apparatus Toolbox: Empowering Textual Scholars through Display, Control, and Comparison Features”. En: *Journal of the Text Encoding Initiative*, 10. [En liña, <https://doi.org/10.4000/jtei.1520>, consultado en 14/01/2018].
- Cerquiglini, Bernard (1989): *Éloge de la variante: Histoire critique de la philologie*. Paris: Éditions du Seuil.
- Chiesa, Paolo (2002): *Elementi di critica testuale*. Bologna: Pàtron Editore.
- Cummings, James (2007): “The Text Encoding Initiative and the Study of Literature”. En: Siemens, Raymond George/Schreibman, Susan (eds.): *A Companion to Digital Literary Studies*. Malden, MA: Blackwell Pub, pp. 451–476.
- De Vos, Ilse/Macé, Caroline/Geuten, Koen (2012): “Comparing Stemmatalogical and Phylogenetic Methods to Understand the Transmission History of the *Florilegium Coislinianum*”. En: Bucossi, Alessandra/Kihlman, Erika (eds.): *Ars Edendi Lecture Series*. Stockholm: Stockholm University Publications, vol. II, pp. 107–129.
- Dekker, Ronald Haentjens/Birnbaum, David J. (2017): “It’s more than just overlap: Text As Graph”. En: *Proceedings of Balisage: The Markup Conference 2017*. Vol. 19. [En liña, <http://www.balisage.net/Proceedings/vol19/html/Dekker01/BalisageVol19-Dekker01.html>, consultado en 14/01/2018].
- Ferreiro, Celso Emilio (1962): *Longa noite de pedra*. Vigo: Salnés.
- Flanders, Julia (2011): “Collaboration and Dissent: Challenges of Collaborative Standards for Digital Humanities”. En: Deegan, Marilyn/McCart, Willard (eds.): *Collaborative Research in the Digital Humanities*. Farnham, Surrey, UK/Burlington, VT: Ashgate Pub, pp. 67–80.
- Ide, Nancy/Romary, Laurent (2004): “International Standard for a Linguistic Annotation Framework”. En: *Natural Language Engineering*, 10. 3–4, pp. 211–225.
- Mink, Gerd (2000): “Editing and genealogical studies: the New Testament”. En: *Literary and Linguistic Computing*, 15.1, pp. 51–56.
- Pierazzo, Elena (2015): *Digital scholarly editing: theories, models and methods*. Farnham, Surrey: Ashgate.
- Pose, Javier/Lopez, Patrice/Romary, Laurent (2014): “A Generic Formalism for Encoding Stand-off Annotations in TEI”. [En liña, <https://hal.inria.fr/hal-01061548>, consultado en 14/01/2018].
- Renear, Allen H. (2004): “Text Encoding”. En: Schreibman, Susan/Siemens, Raymond George/Unsworth, John (eds.): *A companion to digital humanities*, Malden, Mass.: Blackwell Pub., pp. 218–239.
- Renear, Allen/Mylonas, Elli/Durand, David (1993): “Refining Our Notion of What Text Really Is”. [En liña, <http://cds.library.brown.edu/resources/stg/monographs/ohco.html>, consultado en 14/01/2018].
- Romary, Laurent (2015): “Standards for language resources in ISO – Looking back at 13 fruitful years”. [En liña, <http://arxiv.org/abs/1510.07851>, consultado en 14/01/2018].
- Sperberg-McQueen, Michael/Burnard, Lou (eds.) (1994): *TEI Guidelines for Electronic Text Encoding and Interchange (TEI P3)*. Chicago: Text Encoding Initiative. [Versión en liña (1999): <http://www.tei-c.org/Vault/GL/P3/index.htm>, accedido: 30/06/2018].

- TEI Consortium (2017a): “Critical Apparatus”. En: *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. [En liña, <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/TC.html>, consultado en 14/01/2018].
- TEI Consortium (2017b): “Feature Structures”. En: *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. [En liña, <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/FS.html>, consultado en 14/01/2018].
- Witt, Andreas/Stegmann, Jens (2009): “TEI Feature Structures as a Representation Format for Multiple Annotation and Generic XML Documents”. En: *Proceedings of Balisage: The Markup Conference 2009*. [En liña, <http://www.balisage.net/Proceedings/vol3/html/Stegmann01/BalisageVol3-Stegmann01.html>, consultado en 14/01/2018].