

Database-oriented annotation of early modern plays: A proposal*



Anotación orientada a bases de datos para obras teatrales de la modernidad temprana: una propuesta

Jesús Tronch

Universitat de València, Spain

ABSTRACT

This essay proposes that the electronic texts of plays constituting a database-collection (in this case early modern drama) should be “annotated” by marking up not only its structural components but also the editorial annotations about a given feature or aspect of the play (usually included in the commentary notes of print editions), and that these annotations should be conceived having in mind the functionalities of a database. By marking up both the text's structural components and editor's information they constitute related data to be processed by the computer for searches and statistical analysis. This implies that texts should not be annotated individually and independently from the other anthologized works, but rather as part of an organized collection of data that, adequately encoded, will allow users to make queries into the whole database. A second section of the essay discusses three encoding mechanisms, based on the Guidelines of the Text Encoding Initiative, necessary to mark up these “annotations,” and possible *ad hoc* extensions of the TEI schema in order to represent the annotated features. Finally, a third section comments on practical examples showing how to encode a set of features: scene location, image, theme, allusion, proverb, wordplay, grammar, swearing expression, address form, as well as features covered by the TEI Guidelines such as roles, stage directions, names and place-names, verse form and textual issues.

RESUMEN

Este trabajo propone que los textos electrónicos de obras teatrales que forman parte de una colección en una base de datos (en este caso de teatro de la época moderna temprana) se deberían “anotar” marcando no sólo sus componentes estructurales, sino también las anotaciones editoriales con respecto a una característica o aspecto de la obra (generalmente incluidos en los comentarios anotados en las ediciones impresas), y que esas anotaciones deberían concebirse teniendo en cuenta las funcionalidades de la base de datos. Al marcar tanto los componentes estructurales del texto como la información del editor, se crean unos datos relacionados que se pueden procesar informáticamente por medio de búsquedas y análisis estadísticos. Esto implica que los textos no deberían anotarse individualmente y de forma independiente del resto de las obras de la antología, sino como parte de una serie organizada de datos que, con la codificación adecuada, permitirán a los usuarios hacer búsquedas en toda la base de datos. La segunda parte del artículo discute tres mecanismos de codificación, basados en las directrices de la Iniciativa para la Codificación de Textos (TEI), necesaria para marcar estas “anotaciones”, así como las posibles extensiones *ad hoc* del esquema TEI con el fin de representar las características anotadas. Finalmente, la tercera sección comenta algunos ejemplos prácticos que muestran cómo codificar determinados aspectos: localización de la escena, imagen, tema, alusión, proverbio, juego de palabras, gramática, juramentos, tratamiento, así como otras características recogidas por las directrices TEI, como los papeles, las direcciones escénicas, nombres y lugares, tipo de verso y aspectos textuales.

* Research for this essay has been carried out with the support of Research Project FFI2012-34347, funded by the Spanish government through the Plan Nacional I+D+I, and Research Project GVAICO2016-094, funded by the Conselleria d'Educació de la Generalitat Valenciana.

KEYWORDS: Database-oriented annotation of early modern plays database; encoding; markup; annotation; TEI; drama; early modern.

PALABRAS CLAVE: Anotación de obras de teatro de la edad moderna temprana orientada a bases de datos; codificación; marcación; anotación; TEI; teatro; época moderna temprana.

In this essay, I explore the idea of annotating the electronic text of early modern plays collected in a database so that a search for a given annotation could retrieve the corresponding segments in the play-texts. Although I refer to English plays of the sixteenth and seventeenth centuries, I do not envisage any actual database project for a specific corpus of plays; rather I reflect generally and theoretically on the concept of database-oriented annotation, which involves both interpreting aspects of the play (such as the use of proverbs, allusions, oaths, images, etc.) and associating these interpretations with the electronic text by encoding them in Extensible Markup Language (XML) conformant to the Guidelines of the Text Encoding Initiative (TEI). To this reflection, I devote a first section of this essay. Although I imagine users querying a database online, I am not concerned with defining a database management system, any specific database-driven web application or any user interface, but limit the technical description to discussing three TEI-conformant markup mechanisms, including possible *ad hoc* extensions of the TEI encoding scheme, for tagging texts that would be either transferred into, or already hosted in, a database. This discussion is the subject of a second section. Finally, in a third section I comment on examples of several aspects amenable to database use and that might be annotated using any of the three TEI-complying procedures described.

I. The concept of database-oriented annotation

Among the benefits of the new digital tools in textual scholarship is the fact that texts, as sets of data, can be compiled and interrelated in a database, a computerized arrangement of structured data that allows easier and faster search and retrieval of information than in printed collections of data. In that sense, a digital database-collection of early modern plays proves print-based collected editions to be limited. Users of *Open Source Shakespeare* (Johnson 2003–) can carry out searches for words or phrases in Shakespeare's works, and restrict their queries to specific genres or dates, in a speedier and

more convenient way than in printed concordances: for instance, in less than five seconds users can find which pre-1600 plays, or which tragedies, use the word “love” most. Readers of a database-collection of Shakespeare’s plays, such as *Internet Shakespeare Editions* (Best and Jenstad 1996–), can easily and rapidly make complex searches not only in the main text of all the Shakespeare plays but also in selected works, Shakespearean and non-Shakespearean, as well as within fields such as stage directions, speech prefixes, titles and headers, marginalia, front and back matter, page markings, compositors and document metadata. However, at present, the text of the annotations (“notes and commentary”) cannot be searched. Users of *Richard Brome Online* have search options for keywords in the “entire text,” “speech,” “stage directions” and in the “notes and glosses.” If one is interested in, for instance, the proverbs used by Brome, a search for the term “proverb” in the “notes and glosses” field returns 30 results (in all the plays).¹ Yet, this is misleading because the individual editors also used phrases such as “proverbial” and “proverbially” to annotate the use of a proverb or a variant of it. A more thorough search needs to insert the string “proverb*” (with the asterisk indicating a wildcard), which results in 169 hits. However, the results do not discriminate between an annotation on a proverbial phrase and an allusion to the biblical Book of Proverbs² since the search engine looks for matches for the queried string “proverb*” in the content of the note and not for a category of the annotated content. For querying a specific type of content across the collected plays, annotating a play (in the latter example, a Brome play) would entail that some kinds of annotation (in this case, proverbial uses) need to be “database-oriented,” that is, conceived and encoded having in mind that the play belongs to a database.

Similarly, if one is interested in finding occurrences of the musical instrument “recorder” used as a stage property, a search for “recorder” in the three digital resources mentioned above returns instances that do not refer to a prop, as in Hippolyta’s “like a child

¹ The results are displayed in groups of ten per page, with each occurrence showing the co-textual segment and a hyperlink to the segment in the corresponding modern edition, although the new page opens at the beginning of the act in which hit occurs rather than at the exact point (or speech) in the text.

² As in gloss number 1638 and note 372 for “I never was a gadder” in *Convent Garden Weeded*, 1.1.speech58.

on a recorder" (*Midsummer Night's Dream*, TLN 1920–1921), or "Go back to the recorder's" (in Brome's *The Demoiselle*, 2.1.speech224), where the "recorder" is a magistrate or judge.³ The term "recorder" should be annotated having in mind its category as a prop, a category that should have a stable identifier (e.g. the name "prop") for the computer to link it to similar items in the same category in the rest of the collected plays.

The idea of database-oriented annotation I propose in this essay is different from the conventional annotation found in commentary notes of critical editions in that the latter implies that a note or a gloss is generally phrased (and, in the case of digital-born editions, electronically encoded) to be read and used with reference to the specific point in the play-text being annotated, whereas the former understands that the play-text is related to other play-texts in a database and that the usefulness of the annotation increases because of database functionalities such as searching, quantifying and organizing related information and obtaining statistics.

In this context, I am using the term "annotation" both as the editorial analysis and interpretation of a given segment of the play-text (information usually expressed —but not always— in some commentary notes in "single-work-oriented" annotated editions) and as the encoding or markup of the electronic text (the process of inserting tags). These two activities of analysis and codification are not necessarily to be performed by the same person, but I envisage an annotator that would both decide on the interpretative aspects and insert markers or tags with their respective category identifiers in an electronic text, either by means of an XML editor, a text processor, or a user-friendly interface, if the text is to be transferred into a database, or by means of a user interface if the text is already hosted in the database.⁴

³ Also Buckingham's "to be spoke to but by the Recorder," in Shakespeare's *Richard III* (TLN 2243), where "Recorder" means the "officer appointed by the Mayor and Aldermen of London to regulate and preserve the City's customs and institutions" (Jowett 2001, 268).

⁴ The digital editors of the database-collection of Early Modern European Theatre (EMOTHE), which is being developed by the ARTELOPE research group at the University of Valencia, mark up an electronic text by first inserting preliminary tags of their own in a text or word processor, then introducing the pre-tagged text in a FileMaker-run database, which translates this preliminary markup into TEI-conformant XML. On the database, by means of a user-friendly interface, EMOTHE

To the question as to what kinds of annotations and what aspects of a play can be processed in a database-oriented way, an initial answer would be any aspect that the editors of the database-collection would consider worth investigating in relation to the rest of collected plays in the computerized corpus. In this essay I am considering some that are amenable to being searched and quantified by means of a database. Many of these aspects are already dealt with in commentary notes. For instance, a query could aim to find instances of a recurrent image (e.g. Fortune's wheel, as Mardock explains in his note to Pistol's "Fortune's furious fickle wheel," TLN 1477, in Shakespeare's *Henry V*), of culture-bound concepts such as the four elements (as Mardock annotates "O for a muse of fire" in the first line of *Henry V*), allusions to a given work (as Ostovich annotates "in the pig's palace of pleasure" in Brome's *Jovial Crew*, 3.1.speech367, an ironic allusion to Painter's *The Palace of Pleasure*), specific linguistic uses (of swearing expressions, forms of address, discourse markers, grammatical traits of early modern English, etc.), or uses of a given prop. Databases are more effective when searches or queries can combine different parameters: for instance, when the database is asked to find plays containing tyrants and that were written/performed/published before 1603; or plays containing non-Christian exclamations and non-European characters with scenes located in battlefields; or quotations in Latin, or more specifically quotations from Seneca, that are used in tragedies; or to find verbs with an -s inflection agreeing with plural subjects in plays that were written and/or performed and/or published after 1603.

Since the development of a database presupposes that the information must be organized and structured on the basis of pre-established unique identifiers, the kinds of annotations would also depend on the extent to which the database editors were able to build a taxonomy of categories and subcategories prior to the "annotation" of the plays for the database.⁵ Linguistic phenomena are well studied and classified, but other aspects such as images,

editors continue to fine-tune their markup and are able to insert notes of various types (all with the TEI <note> element). The editors do not actually key in TEI tags manually at any stage in this process.

⁵ The taxonomy of categories also needs to be standardized, in the sense that if, for instance, an image is to be called "Fortune's wheel" as an item in the database, it should be "Fortune's wheel" throughout and not "the wheel of Fortune."

culture-bound concepts or even motifs or themes are bound to be problematic, especially if the database contains a large corpus of early modern drama. The task of structuring a list of images, themes and motifs may well be a never-ending endeavour (and a vulnerable one, as it would be based on subjective judgements on which two editors may not agree),⁶ but a project could limit itself to a selection of the most recurrent images or motifs.⁷

As can be immediately inferred, the scope of the database-collection, as well as the human and technical resources devoted to the project, are elements that would condition the type and number of “annotations.” A corpus of relatively few plays, such as the 16 plays in *Richard Brome Online* or the 39 plays in the *Internet Shakespeare Editions* (which also plans to edit Shakespeare's poems) would be able to include more and more detailed annotations than a larger one (for instance, Tudor drama, or Restoration comedy). The development of a database of *all* the extant early modern plays in English would certainly need so many editors/annotators and so much time as to question the feasibility of the project.

However, these limitations should not be a hindrance to our entertaining, in a conceptual exploration, a possible database of marked up play-texts and to our discussing specific tagging procedures for them.

It should also be stressed that I am considering a database of marked up play-texts, and not only a database of information about plays. One could raise the objection that to have a play's component such as “prop” searchable in a database, one could simply list it in a general metadata section describing each play without the play-texts themselves, as if, for instance, the Lope de Vega database (Oleza *et al.* 2012–) would include “props” as another searchable item in its

⁶ I owe this comment to one of the anonymous referees of the *SEDERI* journal.

⁷ The editors of the Lope de Vega database began identifying motifs in the circa 400 plays in the collection but their classification proved so difficult that eventually this aspect was dropped. A complicated alternative in a project that would annotate images and other interpretations in a large collection of plays would be to develop the taxonomy synchronously to the annotation of the play-texts, that is, that an annotator that spots an image that is not included in the pre-established taxonomy would propose its inclusion to the project's editor(s), who should decide on its adequacy, its standardization, should update the information in the project's schema, and should revise already annotated plays that may have not contemplated that particular image.

“characterizations” section; or even more clearly, as it is expected in the announced electronic edition of Martin Wiggins’s *British Drama 1533–1642: A Catalogue*. In the present print-edition, Wiggins registers, for each play, its uses of props (further differentiated into categories such as musical instruments, weapons, money, food and drink, small and large portable objects, etc.)⁸ and provides a reference to the scene in which the term appears, although not the specific line in the scene. Ideally, the future electronic edition of this catalogue will allow searches for these items. But there are two answers to this objection. First, the aspects of interest to be annotated are not limited to those organized in the Lope de Vega database and in Wiggins’ catalogue, but can include other aspects usually covered in scholarly editions (such as recurrent images, culture-bound concepts, allusions, parallel instances in proverbial lore, as listed and exemplified above), or aspects, such as placenames or discourse markers, that are not usually dealt with in the commentary notes of in critical editions. Secondly, I am concerned with the annotation and markup of the electronic *text* of the plays, so that a search for an aspect of interest in a database may also retrieve the segment of text involved.⁹

An example from a line in Shakespeare’s *Hamlet* may be useful to explain the notion of database-oriented annotation (and its encoding) in more detail:

HAMLET Ah ha! Come, some music! Come, the
recorders! (3.2.283–84)¹⁰

The commentary note from Ann Thompson and Neil Taylor’s Arden edition reads as follows:

284 **recorders** wind instruments, flutes. Hamlet calls for music again at 287; the players eventually appear with recorders at 336.1. (If casting allows, an attendant should presumably leave the stage

⁸ Props are part of the “theatrical” broad group of components, also including staging needs, music and sound effects, costumes, and make-up (Wiggins 2012–2014, vol. 1, xxii–xxxix).

⁹ If the electronic version of Wiggins’ catalogue merely reproduces the information in the print edition, readers will miss not only the full line number reference but also the text’s lines that contain the term or aspect in question.

¹⁰ Reference keyed to Thompson and Taylor’s Arden edition (2006).

to convey Hamlet's request, but it has not been customary to add a SD.) (2006, 318)

In order to represent this line and its note in an electronic text, a specific standard or format of markup language is necessary, as explained in the next two paragraphs (which readers already familiarized with markup languages or with the TEI Consortium Guidelines may skip). The typographical layout of this line and its note as printed contains tacit or implicit codes that allow readers to understand that (a) the line is the beginning of a character's speech, in this case Hamlet's, a fact indicated by (b) the speech prefix "hamlet," in small capitals and slightly separated from the rest of the line; (c) the line "Ah ha! . . ." is a prose line because it does not start on a separate typographical line below its predecessor; (d) "284" is the line number to which the annotation is keyed; (e) "recorders" points back to the same word in that line of dialogue; and (f) "wind instruments, flutes" is the gloss explaining "recorders." As human readers in a given culture, we have learned to read an annotated edition of a play so that we can distinguish what a speech prefix is, what a prose line is, and so on; we have learnt the cultural codes of text genres and the typographical layout implicit in the graphic representation of this line. But computers need to be informed of these codes in order to process an electronic text; we inform computers by means of markers inserted in the digitized text that make explicit these implicit codes. Otherwise processors would render digitized texts as "nothing but a sequence of undifferentiated bits" ("About these Guidelines").¹¹

The TEI Guidelines recommend representing this line with explicit markers that tell the computer what segments of the text constitute structural components and other implicit features of the play-text:¹² for instance, in order to signal when a speech prefix

¹¹ Or as Alan Galey puts it, "a structured series of alphanumeric symbols that most humans never see, and that few would wish to see, but which we rely on machines to manage the drudgery of processing" (2004, par. 3). In his "A gentle introduction to the TEI markup language," Mueller usefully reminds us "computers do not think, do not talk, and do not understand anything. They are machines that carry out instructions to the letter" ("A very gentle").

¹² The Guidelines "make recommendations about suitable ways of representing those features of textual resources which need to be identified explicitly in order to facilitate processing by computer programs. In particular, they specify a set of markers (or tags)

begins, TEI uses the <speaker> element (elements are enclosed in angular brackets <>), with a start-tag "<speaker>" and an end-tag "</speaker>" framing the segment in question ("hamlet"):

Example no. 1:

```
<sp>
<speaker>hamlet</speaker>
<p>Ah ha! Come, some music! Come, the recorders!</p>
[ . . . ]
</sp>
```

The markers "<sp>" and "</sp>" enclose the whole speech, and the tags "<p>" and "</p>" enclose a prose line.¹³ It should be pointed out that this is very basic TEI markup for performance texts. In a fuller codification, <sp> and <p> would contain values associated to attributes indicating the person to whom the content of the element <sp> is ascribed (a @who attribute) and a unique identifier for the element (an @xml:id attribute): e.g. <sp who="#Hamlet">, <p xml:id="HAM_3.2.284">. However, in some examples I will omit these units for the sake of clarity.¹⁴

As for the commentary note on "recorders," TEI recommends the <note> element, inserted either after the word(s) in question or elsewhere (with appropriate linking markers), as explained in chapter 3.8 (also in 11.3 and 12.1) of the TEI Guidelines.

Example no. 2

```
<sp>
<speaker>hamlet</speaker>
```

which may be inserted in the electronic representation of the text, in order to mark the text structure and other features of interest" ("About these Guidelines," par. 2).

¹³ If the line division of the printed text is to be represented, the TEI element <lb/> can be inserted between "the" and "recorders." A visualization on a computer screen of this encoded text could have the same appearance as in a print-based edition. The tags are invisible. The tags are like "secret annotations" inserted by the digital editor that speak to the computer, but not to human readers.

¹⁴ For instance, the identifier used in the Folger digital text of *Hamlet* for the prose line "Ah ha! Come, some music! [...]" (3.2.317) is "xml:id="ftln-2218" n="3.2.317". The encoded xml file for *Hamlet* is freely available at <http://www.folgerdigitaltexts.org/downloads/sourcecode.html>.

<p>Ah ha! Come, some music! Come, the recorders!<note
n="3.2.284" type="commentary"><term rend="bold">recorders
</term><gloss>wind instruments, flutes</gloss>. Hamlet calls for
music again at 287 [...] </note></p>

[...]

</sp>

However, the database-oriented annotation I propose does not simply involve this operation of encoding the content of a note by using <note>. If Thompson and Taylor's edition of *Hamlet* is encoded as in the above example, the computer would know that the note is a "commentary" (as different from a textual note, for instance) and that the note contains a "gloss" of the term "recorders"; but the computer would not know that this word is a call in the dialogue for a stage property or prop (as recorders eventually are brought onstage at 3.2.336.1); that the prop in question belongs to the category of "musical instruments"; or that the phrase could be interpreted as an implicit stage direction for an attendant to "leave the stage to convey Hamlet's request" (Thompson and Taylor 2006, 318). For the purpose of turning an editor's annotation into categorized data for a database, the content of the note should be marked up. But, as pointed out earlier, I am concerned with the play-text itself. If the text to be encoded were an electronic transcript of the Second Quarto of *Hamlet*, or a modern-spelling edition without any commentary notes, the category or interpretation of "prop" attached to the word "recorders" should still be encoded. Besides, it is more efficient to provide specific tags for each aspect of interest than to subsume all the specific issues of the editor's annotation into just one note whose content only human readers can process. In the case of "recorder," the categories "prop," "musical instrument," and "implicit stage direction" need to be codified in order to be part of the markup code so that the computer can process this data. An encoding of this sort would allow searches for all props in the plays collected in the database (and would return the text involved as a result), searches for all musical instruments used as a prop, and for other instances of "recorders" or "flutes" used as a prop.¹⁵

¹⁵ If the search is restricted to the text of the stage directions and is performed in a database of English drama between 1580 and 1642, the results would be similar to the ones compiled by Alan C. Dessen and Leslie Thomson (1999).

II. Encoding mechanisms

In this section, I discuss possible TEI-conformant procedures to encode the “annotations” by using the example of “recorders” in *Hamlet* 3.2.284 as a prop.

The TEI Guidelines do not include “prop” as an element, but as a value of the @type attribute in the <tech> element (e.g. <tech type=“prop”>recorders</tech>), but <tech> is used for technical and special-purpose stage-directions that are not meant for the actors.¹⁶ For an aspect that is not covered in the TEI Guidelines, the elements <ab> (anonymous block) and <seg> (arbitrary segment) constitute “two neutral or ‘anonymous’ elements to which the encoder can add any meaning not supplied by other TEI defined elements,” thus “extending the semantics of the TEI markup scheme” (chapter 16.3). Since “recorders” is, on one level, just a string of characters below the “chunk” level, <seg> should be used, as in

Example no. 3:

```
<p>Ah ha! Come, some music! Come, the <seg type=“prop”
  subtype=“musical_instrument”ana=“#in-call”>recorders</seg>!
</p>
```

Here I have represented the call for a prop as the value “in-call” of the analysis attribute (@ana).

A more direct encoding could be achieved if “prop” becomes an element in itself. In fact, the *Internet Shakespeare Editions* guidelines do include a tag for props and account for the possibility of describing that element by means of a “desc” attribute, and of distinguishing whether the prop is mentioned or implied (Best 2014, 4.4.10). Adapted to TEI style, a <prop> element could be used as in the following example:

Example no. 4:

```
<p>Ah ha! Come, some music! Come, the <prop
  type=“musical_instrument” subtype=“recorder” desc=“flute”
  ana=“#in-call”>recorders</prop>!</p>
```

¹⁶ See <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-tech.html>.

This encoding does not conform to the TEI Guidelines, but their recommendations are expressed in terms of XML, and one of the advantages of XML is that it is, precisely, “extensible.” In the course of this essay I will be using TEI elements (<>), attributes (@) and values, but also pointing out situations in which the Guidelines might be extended in order to better represent certain implicit features and therefore facilitate the functionalities of a database of early modern plays.¹⁷

In example no. 4, I have dealt only with the marking-up of the prop “recorders” mentioned in the play-text. If a <note> element for a conventional commentary note were inserted instead of a <prop> element, the text would be excessively cluttered. One solution is to insert this <note> not at its point of attachment, as I did in example no. 2, but at its point of appearance in the printed source edition (see TEI, 3.8 and 16.2). Another elegant way around the matter is to place the “annotations” elsewhere in the document (even in a different document as “stand-off markup”)¹⁸ with a mechanism to secure the linking of the annotation to the exact point or segment of the text in question (see TEI, sections 3.8, 16.2, and 17.3). This system of separate annotations has certain advantages: a single annotation may cover several segments of text occurring at different points, and different “annotators” may work simultaneously on different kinds of annotations on the same textual segment (TEI, 16.9); and if categories are modified, they could be more easily altered if grouped together rather than scattered throughout the play-text. The XML files of *Folger Digital Texts* use stand-off markup to encode interpretations such as emendations, verse/prose, and stanzas: the <back> section of the file contains grouped interpretation elements (<interp> within <interpGrp>) defining the kinds of emendations, lines and stanzas; then pointer elements (<ptr/>) contain the @ana value identifying the interpretation, and the @target value identical to the @xml:id value of the textual segment in question; in the case of emendations, these <ptr/> are grouped with a subdivision element

¹⁷ Prior studies have pointed out how the TEI model fails to answer adequately all the encoding demands posed by early modern dramatic texts. See, among others, Ian Lancashire’s *Renaissance Electronic Texts: Guidelines for Encoding* (1994) and Galey’s “Dizzying the Arithmetic of Memory” (2004, par. 1, 2 and 22).

¹⁸ One of the anonymous readers for *SEDERI* has pointed out that this method has the advantage of allowing for multiple annotations of the same word or passage, besides being the cleanest way to do the kind of annotations proposed in this essay.

(<div>), while in the case of verse and stanzas, <ptr/> are placed in the running text.

A stand-off annotation following the Folger model could mark the word “recorders” with a <seg> element, as in this example:

Example no. 5:

```
<sp>
<speaker>hamlet</speaker>
<p>Ah ha! Come, some music! Come, the <seg
xml:id="HAM_3.2.317_8">recorders</seg>!</p>
</sp>
```

and in the <back> section,

```
<interpGrp type="prop">
<interp xml:id="musical_instrument">any musical instrument
used as a stage property</interp>
<interp xml:id="in-call">a call in the dialog for a prop</interp>
</interpGrp>

<div type="prop">
<ptr type="prop" ana="#musical_instrument #in-call" target=
"#HAM_3.2.317_8"/>
<ptr type="prop" ana="#musical_instrument" target="#HAM_
3.2.373.1 #HAM_3.2.374_3/>
```

Note that <div> contains another pointer to other instances of “recorders” (in the stage direction at 3.2.373.1 and in Hamlet’s line 374) with their respective @target values and that this pointer does not include “in-call.”

The three encoding procedures described so far can be classified by using two binary oppositions: in-line or stand-off annotation; with TEI-compliant elements or with newly created *ad hoc* elements.

	in-line	stand-off
TEI element	(1) <seg type="prop"> </seg>	(3) <interpGrp type="prop"> <interp xml:id=" "> <div type="prop"> <ptr type="prop"/>
<i>ad hoc</i> element	(2) <prop> </prop>	

III. Examples of "annotations"

I will comment on examples of several aspects of a play that might be annotated with a view to being searched and quantified in a database by specifying how they could be represented in the markup, characterized into types or values that would be part of a taxonomy of category identifiers.¹⁹ I will first discuss aspects whose encoding require *ad hoc* extension of the TEI schema. Then I will tackle those that are, in one way or another, included in the TEI Guidelines (roles, stage directions, names and place-names, verse form and textual issues).

III.1

For aspects that would need extending the TEI recommendations, I will discuss alternative markup procedures. For the sake of economy I will omit examples of tagging, but these can be imagined if the terms in bold heading each paragraph are taken, unless otherwise stated, as *ad hoc* categories to be encoded as

- (1) the @type value in the <seg> or <ab> elements,
- (2) the name of the *ad hoc* element created;
- or (3) the @type value in <interpGrp> and <div>, if a stand-off markup is chosen.

Then the subcategories that I will specify will constitute

¹⁹ As the TEI Guidelines indicate, the values should be taken from a pre-established taxonomy, preferably listed in the TEI header (<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.typed.html>).

- (1) the @subtype value of <seg> or <ab>;
- (2) the @type, @subtype and/or @ana values of a presupposed new element;
- or (3) the @xml:id value of <interp> element, and the @ana value of <ptr/>.

Scene location: In many critical editions, the first commentary note for a scene defines its location, a practice also indicated in the ISE guidelines (Best 2014, 5.2.10). For instance, in Wray’s Arden edition of *The Tragedy of Mariam*, the beginning of the first commentary note is “1.1 Location: a public space in Herod’s palace in Jerusalem.” At least four components in this note deserve annotation. Three of them are geographical: in descending order of their relative size, a city (“Jerusalem”), a building (“palace”), and a specific area in this building (“a public space”). Implied in the notion of city are larger geographical units such as country or region. The fourth component is the person with which the building is associated (“Herod”). But the characteristics of the scene location, while not stated in this commentary note, may well include binary oppositions such as public/private, urban/rural, outdoors/indoors, imaginary/real, or implied/explicit. Having these characteristics in a database may allow analyses of the kinds of space used in a play, group of plays, or in the whole corpus, combined with aspects such as genre, plays dealing with specific kinds of characters or certain themes, plays by a given author, plays written, produced, or published within a certain period, etc. Thus, a database-oriented annotation that defines the first scene in *The Tragedy of Mariam* as taking place in a “public” space will allow connections with similar scenes and contrasts with scenes in “private” spaces, such as Pheroras’s room in 2.1, or in a “palace.” Identifying the palace as being associated with “Herod” would allow searches for similar scenes in *Mariam* (all scenes except 2.3, 2.4, 4.1, 4.6, and 4.8) and in other plays featuring Herod.

TEI provides detailed descriptions (markers) of the components of a place or location, but their markers are at the level of elements, which are units that have to frame content of the text indicating the place.²⁰ While often this content is absent (in the form of a heading or

²⁰ Elements such as <place>, <location>, <placeName>, <country>, <region>, <settlement>, <district>, <geoName>, <geoFeat>, <locale>. Another complication lies in the incompatibilities to nest certain elements within other elements.

a stage direction). Even in the case of explicit locations (e.g. in the stage direction “*Enter a Soldier in the woods, seeking Timon*” [Shakespeare, *Timon of Athens*, 5.3.0]), the annotation refers to a whole scene and not to the textual segment describing this location. These situations prompt me to propose “scene_location” as a category to be encoded as attribute values in the different markup methods explained above, with the characteristics I have just pointed out as subcategories. If stand-off annotation were used, the @target value of <ptr/> could refer back to an @xml:id value identifying the <div1> or <div2> element that frames the scene in question.

Image: The analysis or interpretation of a textual segment as an “image” is covered in chapter 17.3, “Spans and Interpretations,” in relation to and <interp>, not to <seg>, but nothing explicitly prevents us from inserting the category “image” as the @type value of a <seg> element. Let us consider Rosencrantz’s image of “a massy wheel [...] when it falls” in 3.3.12–15, which George Hibbard annotates in his Oxford edition of the play as:

The main idea here is that of Fortune’s wheel, also referred to at 2.2.484–88. The relation between kingly power and the caprices of Fortune is a common theme in Shakespeare. See especially *Lear* 2.4.71–73 and *Timon* 1.1.66–97. (1987, 271)

Hibbard’s note points out the interrelationship of the image in question to other instances in the play and in other Shakespeare plays, but not to other early modern plays, such as *The Spanish Tragedy* (3.1.5) or *The Tragedy of Mariam* (4.4.48).²¹ The specificity of “Fortune’s wheel” as a sub-category could be represented, as indicated above, in the values of @subtype attribute of the <seg> element, of the xml:id attribute of the <interp> element, or of the analysis attribute (@ana) of the <ptr> element. Further distinctions as to the association of this image to royal power as different from other values such as “the vanity of success and the folly of self-

²¹ To which could be added Christopher Marlowe’s *Tamburlaine, Part One*, (1.2.174), John Webster’s *Duchess of Malfi* (3.5.94); *A Cure of a Cuckold* (4.1.401) by John Webster and William Rowley, possibly with Thomas Heywood; and William Alexander’s *The Alexandrian Tragedy* (M3v), among others. For the sake of economy, it is very likely that Hibbard’s note could not include other Shakespearean instances of this theme, such as in *Henry VI Part 3* (4.3.48–49), *Henry V* (3.6.27), *As You Like It* (1.2.31–32), and *Antony and Cleopatra* (4.14.88–89, 4.15.51).

advancement” could well be left outside the scope of the database (Wray 2012, 162).

Theme: I use the term “theme” since this is included as one of the sample values of the @type attribute in chapter 17.3 of the TEI Guidelines, together with “image,” “character,” and “allusion.” Motifs could also be considered under this term, or perhaps a new element or value could be created *ad hoc*. (In section I of this essay I comment on the practical difficulties of encoding this aspect if the database were to include a large corpus of plays.) I will discuss an example related to the phenomenon of suicide as part of the action of a play. The term “suicide” could be encoded as a subcategory associated with scenes, passages, and lines in *The Spanish Tragedy*, in *Hamlet*, in *The Tragedy of Mariam*, etc. Specific contrasts of aspects of suicide could be obtained if further characteristics (for instance, the gender and age of the suicidal characters) are considered in the categorization. Characteristics could be taken from Bernard Paulin’s *Du couteau à la plume*. This comprehensive study of suicide in English literature from 1580 to 1625 distinguishes the moral approval or disapproval of suicide (even if it is ambiguous), motivations for suicide,²² method and instruments of suicide,²³ if it is shown onstage or reported,²⁴ as well as cases of intended and failed suicides. Within the TEI schema, these subcategories could be represented as @ana values.

Allusion: This is another value, mentioned in chapter 17.3 of the TEI Guidelines, of the and <interp> elements and its associated grouping elements. Allusions are usually dealt with in commentary notes, as with the *Hamlet* line “For if the King like not the comedy” (3.2.319), taken as an echo of “And if the world like not this tragedy [...]” in *The Spanish Tragedy* (4.1.188). This indirect intertextual reference could be marked-up so as to allow searches

²² “guilt” or “remorse” in the case of the Butler in *Mariam*, or Homes in *Sir Thomas Wyatt*, Atis in *Croesus*, Dymnus in *Philotas*, etc.; “revenge” in Isabella in *The Spanish Tragedy*, although “madness” could also be attributed to Isabella, as in the case of Ophelia in *Hamlet*; “love” in the case of Bel-imperia in *The Spanish Tragedy*.

²³ “hanging” in *Mariam* and in the case of Fronto in *Caesar and Pompey*, and perhaps Homes in *Sir Thomas Wyatt*; “drowning” in Ophelia; “dagger” in Bel-imperia.

²⁴ “reported” in *Mariam*, as is the case of Ophelia in *Hamlet*, 2 Consuls in *Caesar and Pompey*, or of Roxane in *Alexandrian Tragedy*. Isabella, Bel-imperia, and Hieronimo commit suicide onstage in *The Spanish Tragedy*.

for, and quantifications of, other allusions in *Hamlet* to *The Spanish Tragedy*, all allusions used in *Hamlet*, and allusions to *The Spanish Tragedy* in all the plays of the database. If the allusions are subcategorized as parodic, neutral, or questionable, these parameters could also be included in the query. In contrast to image and theme, the annotation of the allusion has to include an intertextual reference to the alluded instances. TEI provides the <ref> and <ptr> elements for cross-references, which can be used to point to texts other than the current text (3.6 and 16.1). It should be noted, however, that <ref> and <ptr>, cannot be contained by <interp>, <interpGrp> and <ptr>. One solution is to include the cross-reference nested in <seg> surrounding the textual segment in question:

```
<|><seg xml:id="HAM_3.2.319"> <ptr type="allusion"
ana="#parodic" target="#SPAN_TRAG_4.1.188"/>And if the
world like not this tragedy</|>
```

Here I have used @ana for the subcategory "parodic." As an alternative to @target, the canonical reference attribute (@cRef) could be used,²⁵ with references defined in the TEI header of the corpus. As with the case of images (and also themes), a project undertaking the annotation of a corpus of plays should be ready to centralize information and decisions and to update references and identifiers to be used throughout the collection.

Proverb: Hibbard annotates the use of proverbial language for "a thousand pound" in Hamlet's "I'll take the Ghost's word for a thousand pound" (3.2.278–79), and for Rosencrantz's "you deny your griefs to your friend" (3.2.330). If encoded with database-oriented annotation in mind, proverbs can be searched, quantified and related to uses in other plays in the corpus, such as "a thousand pound" being also used in *Much Ado About Nothing* 1.1.88, or *Arden of Faversham* 8.137, and the proverbial notion of grief becoming lessened when imparted to others being also used in *The Spanish Tragedy* (1.3.32). Since we have a typology of proverbial language in Dent's index, its conventional references ("T248.1" in the case of "a thousand pound," and "G447" in the second case) could well be used as the value of a @n attribute. Besides, since Dent regards "a thousand pound" as of questionable relevance (2) or "doubtful

²⁵ If this is reserved for canonical works such as the Bible, perhaps an @iRef attribute (for intertextual or indirect reference) could be created.

legitimacy" (43) and Rosencrantz's "you deny your griefs to your friend" as sententious, this status could be specified as a subcategory.

Wordplay: A regular issue in editors' commentary notes is to point out the meanings involved in puns. For instance, Thompson and Taylor annotate "distempered" in *Hamlet* (3.2.328), distinguishing the sense "out of temper," as intended by Guildenstern, and "drunk" as taken by Hamlet (319). Users of the database could quantify, for instance, the punning habits of different characters or different genres. Wordplay could be subcategorized by using types of wordplay (for instance, defined in eight categories and seven functions by Delabastita 1993, 78–86, 137–51) and by indicating if the secondary meaning contains a sexual innuendo or a recurrent topic (e.g., "drunkenness").

Grammar: Grammatical features peculiar to early modern English, such as the use of the verbal -s inflection agreeing with plural subjects (Blake 2002, 4.4.2; Hope 2003, 2.1.8a) or the use of "double comparative" (Blake 2002, 3.2.3.4; Hope 2003, 1.2.4), are generally dealt with in commentary notes. In a database-oriented annotation, these features could become subcategories with the terms used by either Blake or Hope as identifiers in the taxonomy.

Oath: Most editors gloss the interjection "perdie" in Hamlet's "Why then belike he like it not, perdie" (3.3.286), in the sense of "certainly," "indeed," and as derived from the French word "*pardieu*" (OED *pardie*, *int.*). In a database, the use of "perdie" could be analyzed in connection not only to other asseverations meaning "indeed" in other plays, but also to other oaths, and within this category, to emphatic expressions swearing by God and His attributes, and to constructions using an introductory "by" (here in French-derived "per") followed by the sworn phrase. The search would bring out not only expressions in *Hamlet*, such as "faith" (1.5.150), "God willing" (1.5.208), "For God's love" (1.2.205), "by Gis" (4.5.63), "by my fay" (2.2.284). These might be possible to include as cross-references in a gloss or commentary note, although they are usually not included—but also in other plays, e.g., "perdie" in *King Lear* (2.4.91), *Comedy of Errors* (4.4.76); "God's will" in *Henry V* (4.3.26, 76), *Spanish Tragedy* (3.12A.76); "by my faith" in *As You Like It* (3.5.43, 4.1.23, 5.4.65), *Henry IV Part 1* (1.2.144, 2.1.94, 4.1.135, 5.4.125), *Henry V* (3.7.112), and *The Spanish Tragedy* (4.1.59). The TEI

Guidelines (chapter 17.1) provide markers for the word class “interjection” as a @type attribute in units down to the word level (sentence <s>, clause <cl>, phrase <phr>, and word <w>). Given the subcategorization of the exclamation in different values, it would seem appropriate to create an element for each lexical word class (<noun>, <verb>, <adjective>, etc.) in order to allow for type and subtype attributes to specify categories and subcategories respectively: for instance, <oath>; then, swearing by Christian terms (type=“Christian”), Non-Christian terms, human notion; then, within swearing by Christian terms, expressions swearing by God (subtype=“God”), by Christ, by Mary, etc., as David Crystal and Ben Crystal classify them (2009, 435–39).

Address_form: Thompson and Taylor annotate “Good my lord” (3.2.322) to indicate that “Guildenstern’s mode of address is carefully deferential” (319) and refer to other uses of the phrase addressed to Hamlet. Database-oriented annotations of the expression “Good my lord” would allow the display of a concordance of other uses of this form of address not only in *Hamlet* (2.2.548, Hamlet to Polonius; 3.1.99, Ophelia to Hamlet) but also in the whole collection (e.g. *Spanish Tragedy*, 4.3.5, Hieronimo to the Duke of Castile). And if the speakers involved in the exchange are encoded, and in their turn, characters are encoded with values indicating s such as their sex, age, and rank, users of the database may carry out sociolinguistic analyses of the forms of address.

A similar treatment could be given to other fixed expressions in the categories of **discourse marker**, **greeting**, and **exclamation** (not often included in commentary notes) which would allow database searches for uses of “go to” (a discourse marker in *Hamlet* 1.3.121 that Thompson and Taylor annotate “a contemptuous or dismissive expression” [199]), or “how now” (an exclamation of surprise or reproach in *Hamlet* 1.1.81) or in different situations and spoken by different characters. Crystal and Crystal provide typologies that can be used to subcategorize them (2009, 127–29, 206, 158–59).

III.2

This subsection covers aspects for which the TEI Guidelines provide a markup procedure.

Roles: Conventional editions offer commentary notes on the roles of the play, usually attached to the list of characters. Even when marking up an early text without a *dramatis personae*, the editor-encoder must define the roles, speaking and non-speaking, within the <castList> element (TEI, 7.1.4). Annotations of the play's roles in parameters such as sex, age, nationality, geographical or ethnic affiliation, socio-economic status, religion, status, or quality may be useful as criteria to select aspects to be analysed through the database. The TEI Guidelines provide for elements and attributes to mark up this information (13.3.2 and 15.2.2), which is to be included in description of the participants (<particDesc>) in the profile of the text (<profileDesc>) within the header (2.4). As the following example from the Folger *Hamlet* XML file shows, the editors include data about sex, state, and his or her death:

```
<person xml:id="CLAUDIUS-HAM">
  <persName><name>King Claudius</name></persName>
  <state><p>brother to the late King Hamlet</p></state>
  <sex value="1">male</sex>
  <death when-custom="ftln-4076"/>
</person>
```

Other kinds of information can be encoded with <age>, <state>, <socecStatus>, <nationality>, <faith>, and <trait>, or with the @role attribute of the <person> element (TEI, 13.3.2 and 15.2.2). An analysis from a feminist approach, for instance, might find it useful to have female characters categorized according to their status in a patriarchal order (maid, wife, widow, mother, mistress).

It should be noted that the @xml:id value of <person> ("CLAUDIUS-HAM" in the above example) is the specific identifier for a given role. This same identifier is to be used

- (1) in <castList>, as the @xml:id value of <role> (TEI, 7.1.4)

```
<castList>
  <castItem>
    <role xml:id="CLAUDIUS-HAM">King Claudius</role>...
```

- (2) in every speech the role speaks, as @who value in <sp> (TEI, 7.2.2)

```
<sp xml:id="sp-0191" who="#CLAUDIUS-HAM">
  <speaker>KING</speaker>
```

And optionally (3) in identifications of the role's participation in a stage action, usually as the @ana value of a stage direction in the <stage> and <move/> elements (TEI, 7.2.4), as in the first stage direction in the second scene of *Hamlet* as encoded by the *Folger Digital Texts* editors:

```
<stage xml:id="stg-0190.2b" type="entrance" ana="#CLAUDIUS-
HAM #GERTRUDE #POLONIUS #LAERTES #HAMLET
#VOLTEMAND #CORNELIUS-HAM">
```

It should also be noted that the identifier for Hamlet's uncle that the *Folger Digital Texts* editors chose includes the suffix "-HAM" in order to distinguish Claudius in *Hamlet* from other Claudiuses. This is important for the purpose of database-oriented encoding, since the identifier for a given role in a given play should be unique in the whole corpus of plays. Thus, for a messenger in *Hamlet* (in 4.5.108 and 4.7.39) to be distinguished from a messenger in *The Spanish Tragedy*, we would, for instance, need identifiers such as "messenger-HAM" and "messenger-SPAN_TRAG," or in the case of various messengers in the same play, as in *Henry VI Part Three*, identifiers such as "messenger_to_Henry_VI-3H6," "1_messenger_to_Warwick-3H6," and "2_messenger_to_Warwick-3H6."²⁶ Decisions for defining these role identifiers presuppose that all roles in the corpus are mapped.

Stage directions: TEI provides an open taxonomy of stage directions by specifying type values of the <stage> element (7.2.4). What TEI calls "business," the *Internet Shakespeare Editions* (ISE) guidelines name "action," but they both share the terms "entrance," "exit," "setting," "location," and in part "delivery," since ISE uses "whoto" to indicate to whom the dialogue is addressed and for an aside (Best 2014, 4.4.5). The Lope de Vega and the *EMOTHE* collections, which follow TEI, demarcate the segment of dialogue that constitutes an "aside" by means of <seg>, with "aside" as its type:

²⁶ The *Folger Digital Texts* used the value "MESSENGERS" (established in the participants description in the header) in 4.5.108 and 4.7.39 even though only one messenger speaks.

```
<speaker>polonius</speaker>
<p xml:id="en-1232"><stage type="delivery"> [Aside]</stage>
<seg type="aside">Though this be madness yet there is method
in't.</seg> — Will you walk out of the air, my lord?</p>
</sp>
```

This procedure allows highlighting of the content of the “aside” in the online display of the text. By including “aside” in the markup, a database could be able to quantify the use of this special kind of stage direction and could allow analyses in relation to the characters that use it. If the demarcation of the content of the “aside” is not carried out, I would suggest encoding “asides” more directly with the term “aside” as @type value of <stage>.²⁷

If it were of interest to search for significant kinds of stage action, the directions marked-up with the “business” or “delivery” values could contain a @subtype specifying that kind, perhaps taken from the terms in Dessen and Thomson’s dictionary (1999, 257–58), as in the following example from *The Tragedy of Mariam* 1.1.65–68 (here quoted from Bevington 2002):

```
<l>How canst thou then so faintly now lament</l>
<l>Thy truest lover’s death, a death’s disgrace? </l>
<l><stage type="delivery" subtype="weep"> [She weeps.]</stage>
Ay, now, mine eyes, you do begin to right</l>
<l>The wrongs of your admirer and my lord! </l>
```

Names and place-names: These are usually not covered in commentary notes. For practical reasons of limited paper space, an editor of a print-based edition will not insert a note whenever a toponym, such as “Norway,” “Wittenberg,” or “Judaea,” occurs, but these limitations do not affect the electronic text. A toponym such as “Judaea” (4.4.44) in *The Tragedy of Mariam* can be tagged with <placeName>, <country>, <region>, <settlement> and other elements related to geographical spaces (TEI, 13.2.3), and could be subcategorized to specify details, including if the place-name belongs to the action of the play (diegetic) or if it is simply alluded to:

²⁷ It would be also useful for a TEI-conformant markup to use “whoto” rather than “delivery,” and the “optional” and “uncertain” values, as the ISE editions do.

<l>Thou shouldst the wonder of <placeName><country key="PS IL" type="diegetic">Judaea</country></placeName> be,</l>²⁸

Verse form: The TEI Guidelines provide specific recommendations in their sixth chapter, but in the case of early modern plays, annotations could also make explicit the functions of these metrical features, such as, the conventional use of couplets signaling the end of a scene, of a soliloquy, of a speech, or of part of a set-speech.

Textual issue: Critical editions usually include comments on textual problems and editorial decisions. For instance, with respect to Guildenstern's "put your discourse into some frame and start not so wildly from my affair" (3.2.336), Thompson and Taylor explain "start" as an emendation of Q2's "stare" because "e/t is an easy misreading" (319). If the TEI element for emendations, <corr> in combination with <sic>, both nested in the <choice> element (chapter 12), is supplemented with attributes to indicate the type and origin of error, database users could relate this case to other errors derived from the confusion of final -t for final -e, with errors made by Roberts's compositor Y (James Roberts is the printer of the second quarto of *Hamlet*), and with other cases where the Folio (and not a modern editor) provides the received emendation. All these aspects could be combined with the selection of multiple-text plays (*Hamlet*, *King Lear*, *Othello*, *Richard III*, *Henry IV Part 2*) or single-text plays. This particular alternate reading could be represented as follows:

"[...] put your discourse into some frame and
 <choice type="substitution" subtype="misreading" ana="#t-for-e-misreading">
 <sic source="HQ2" resp="Roberts-Y">stare</sic>
 <corr resp="HF1" source="Shakespeare-F1">start</corr>
 </choice> not so wildly from my affair."²⁹

²⁸ The values I have added in the key to identify the country are the ISO 3166 code for Palestine and Israel (see <http://www.iso.org/iso/home/standards/country_codes.htm>. This is admittedly problematic, but the ISO standard for countries follows the names given by the United Nations and does not contemplate historic countries or regions.

²⁹ "HQ2" is the abbreviation for the Second Quarto of *Hamlet* recommended in the *Guidelines for Editors of Internet Shakespeare Editions* (Best 2014, 7.1.1). The two compositors of Q2 *Hamlet* printed by Roberts are identified as X and Y (Brown 1955).

Conclusion

To sum up, database-oriented annotation entails, to a large extent, turning the analyses and interpretations found in some commentary notes, as well as other aspects not usually dealt with in these commentary notes, as explained in section III, into categorized marked up data that are operational for a database of play-texts. It also implies that, as the database models aspects of the “universe” of the data it collects, the types of queries users would be able to make condition the categories of annotation and the way the elements are to be marked up. In this essay, I have explained possible annotations for certain aspects in early modern plays. Of course, the set of aspects I have dealt with is not finite (linguists would perhaps like to see other aspects such as register, speech-act, etc. encoded in the text). The possibilities could be “dangerously” infinite, so it is worth taking into account John Lavagnino’s warning that “Extra markup is costly, and it is essential that a project decide just which features need to be marked in order to serve its scholarly ends” (2007, par. 3). The limits are not so much in the digital media employed but imposed at best by the editors’ own limitations.

References

- Alexander, William, Earl of Stirling. 1994 (1607) *The Alexandrian Tragedy*. In *The Monarchic Tragedies*. London; rpt. *Recreations with the Muses*. London, 1637. In *English Verse Drama*. Cambridge: Chadwyck-Healey.
- . 1994 (1604). *The Tragedy of Croesus*. In *The Monarchic Tragedies*. London; rpt. *Recreations with the Muses*. London, 1637. In *English Verse Drama*. Cambridge: Chadwyck-Healey.
- Arden of Faversham*. 1990 (1592). Edited by Martin White. London: A & C Black.
- Bevington, David, gen. ed. 2002. *English Renaissance Drama: A Norton Anthology*. New York: W. W. Norton.
- Best, Michael. 2014. “Editorial Guidelines.” *The Internet Shakespeare Editions*. University of Victoria. Last modified 21 March 2014. <http://internetshakespeare.uvic.ca/Foyer/Guidelines/index.html>.
- , and Janelle Jenstad, coord. 1996–. *The Internet Shakespeare Editions*. University of Victoria. Accessed September 2016. <http://internetshakespeare.uvic.ca/>
- Blake, Norman F. 2002. *A Grammar of Shakespeare’s Language*. Houndmills: Palgrave.

- Brome, Richard. *Covent Garden Weeded*. Modern text edited by M. Leslie. Accessed 29 April 2016. <http://www.hrionline.ac.uk/brome>.
- . *The Demoiselle*. Modern text edited by Lucy Munro. <http://www.hrionline.ac.uk/brome>.
- Brown, John Russell. 1955. "The Compositors of *Hamlet Q2* and *The Merchant of Venice*." *Studies in Bibliography* 7: 17–40.
- Chapman, George. 1994 (1631). *Caesar and Pompey*. In *English Verse Drama*. Cambridge: Chadwyck-Healey.
- Crystal, David, and Ben Crystal. 2002. *Shakespeare's Words: A Glossary and Language Companion*. London: Penguin. Accessed September 2016. <http://www.shakespeareswords.com>.
- Daniel, Samuel. 1994 (1623). *Philotas*. In *English Verse Drama*. Cambridge: Chadwyck-Healey.
- Dekker, Thomas, and John Webster. 1994 (1604). *The Famous History of Sir Thomas Wyatt*. In *English Verse Drama*. Cambridge: Chadwyck-Healey.
- Delabastita, Dirk. 1993. *There's a Double Tongue: An Investigation into the Translation of Shakespeare's Wordplay with Special Reference to 'Hamlet'*. Amsterdam: Rodopi.
- Dent, Robert William. 1981. *Shakespeare's Proverbial Language: An Index*. Berkeley: University of California Press.
- Dessen, Alan, and Leslie Thomson. 1999. *Dictionary of Stage Directions in English Drama, 1580–1642*. Cambridge: Cambridge University Press.
- Folger Digital Texts. Texts of Shakespeare's plays and poems edited by Barbara A. Mowat, and Paul Werstine. Digital text editors, Michael Poston and Rebecca Niles. Folger Shakespeare Library. Accessed September 2016. <http://www.folgerdigitaltexts.org/>.
- Galey, Alan. 2004. "Dizzying the Arithmetic of Memory: Shakespearean Document as Text, Image, Code." *Early Modern Literary Studies* 9 (3) / Special Issue 12 (January): 4.1–28. Accessed September 2016. <http://extra.shu.ac.uk/emls/09-3/galedizz.htm>.
- Hibbard, George, ed. 1987. *Hamlet*. By William Shakespeare. Oxford Shakespeare. Oxford: Oxford University Press.
- Hope, Jonathan. 2003. *Shakespeare's Grammar*. London: Thomson Learning.
- Johnson, Eric. 2003–. *Open Source Shakespeare: An Experiment in Literary Technology*. George Mason University. Accessed September 2016. <http://www.opensourceshakespeare.org/>.
- Jowett, John, ed. 2001. *Richard III*. By William Shakespeare. Oxford: Oxford University Press.

- Kyd, Thomas. 2013 (1592). *The Spanish Tragedy*. Edited by Clara Calvo and Jesús Tronch. Arden Early Modern Drama. London: Bloomsbury.
- Lancashire, Ian. 1994. *Renaissance Electronic Texts: Guidelines for Encoding*. University of Toronto. Accessed September 2016. <http://www.library.utoronto.ca/utel/ret/guidelines/guidelineso.html>.
- Lavagnino, John. 2007. "Electronic Textual Editing: When Not to Use TEI." In *Electronic Textual Editing*, edited by John Unsworth, Katherine O'Brien O'Keefe, and Lou Burnard. Accessed September 2016. http://www.tei-c.org/About/Archive_new/ETE/Preview/lavagnino.xml.
- Mardock, James, ed. *Henry V*. By William Shakespeare. Internet Shakespeare Editions. University of Victoria. Accessed September 2016. <http://internetshakespeare.uvic.ca/Library/Texts/H5/>.
- Marlowe, Christopher. 1981 (1590). *Tamburlaine, Part One*. In *Revels Plays*, edited by J. S. Cunningham. Manchester: Manchester University Press.
- Marston, John. 1986. *Antonio and Mellida*. In *The Selected Plays of John Marston*, edited by MacDonald P. Jackson and Michael Neill. Cambridge: Cambridge University Press.
- Middleton, Thomas. 1969 (1630). *A Chaste Maid in Cheapside*. In *Revels Plays*, edited by R. B. Parker. Manchester: Manchester University Press.
- Mueller, Martin. "A Very Gentle Introduction to TEI Markup Language." *Text Encoding Initiative*. Accessed September 2016. <http://www.tei-c.org/Support/Learn/mueller-index.htm>.
- OED. 2013. *Oxford English Dictionary Online*. Oxford: Oxford University Press. Accessed September 2016. <http://www.oed.com/>.
- Oleza, Joan et al. 2012. *Base de Datos y Argumentos del teatro de Lope de Vega*. ARTELOPE. Universitat de València. Accessed September 2016. <http://artelope.uv.es/baseArteLope.html>.
- Ostovich, Helen, ed. *Jovial Crew*. By Richard Brome. In *Richard Brome Online*. Royal Holloway, University of London, and University of Sheffield. Accessed September 2016. <http://www.hrionline.ac.uk/brome>.
- Paulin, Bernard. 1977. *Du couteau à la plume: Le Suicide dans la littérature anglaise de la Renaissance, 1580–1625*. Lyon: L'Hermès & Publications de la Université de Saint-Etienne.
- Richard Brome Online*. Royal Holloway, University of London, and University of Sheffield. Accessed September 2016. <http://www.hrionline.ac.uk/brome>.
- TEI Consortium. 2014 (2007). *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Ed. Lou Burnard and Syd Bauman. Oxford: TEI Consortium. Accessed September 2016. <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html>.

Tronch

- Thompson, Ann, and Neil Taylor, eds. 2006. *William Shakespeare: Hamlet: The Second Quarto Text*. The Arden Shakespeare [3rd series]. London: Thomson Learning.
- Webster, John, and William Rowley (possibly with Thomas Heywood). 1994 (1661). *A Cure of a Cuckold*. In *English Verse Drama*. Cambridge: Chadwyck-Healey.
- Webster, John. 2009 (1614). *The Duchess of Malfi*. Edited by Leah Marcus. Arden Early Modern Drama. London: A & C Black.
- Wiggins, Martin, in association with Catherine Richardson. 2012–2014. *British Drama 1533–1642: A Catalogue*. 4 vols. Oxford: Oxford University Press. [Vol. 1 and 2 in 2012. Vol. 3 in 2013. Vol. 4 in 2014.]
- Wray, Ramona, ed. 2012. *The Tragedy of Mariam*. By Elizabeth Cary. Arden Early Modern Drama. London: Bloomsbury.

How to cite this article:

Tronch, Jesús. "Database-oriented annotation of early modern plays: a proposal." *SEDERI* 26 (2016): 129–156.

Author's contact: tronch@uv.es

Postal address: Departament de Filologia Anglesa i Alemanya – Facultat de Filologia, Traducció i Comunicació – Avinguda Blasco Ibáñez, 32 – 46010 València, Spain

Submission: 20/05/2016

Acceptance: 04/07/2016