An Encoding Model for Librettos: the Opera Liber DTD

Elena Pierazzo *(pierazzo@ital.unipi.it)*
University of Pisa

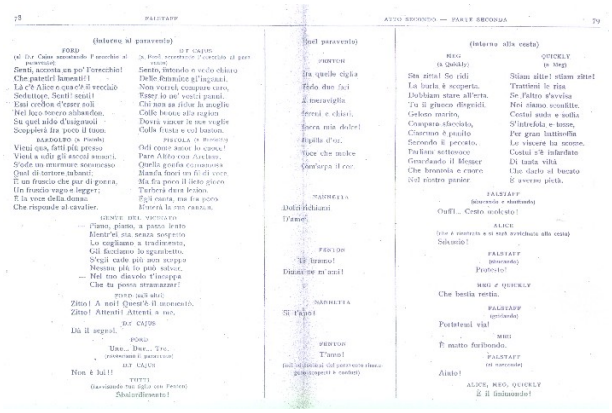
Opera librettos are a very peculiar literary genre. Often considered an ancillary part of the opera, merely the plot through which the music can express its power and its beauty or the pretext for singers to show their capabilities and the potential of their voices, the libretto is a little studied aspect of the literature.

A considerable number of web sites are currently presenting collections of librettos in several formats (`doc`, `pdf`, `html`, `gif`, `txt`). However, they generally do not cite their source or even which version of the text they are based upon; furthermore, in most cases they do not respect editorial traditions of the libretto.

Librettos have some peculiar structural characteristics: they can be considered a subcategory of drama texts, but they distinguish themselves form the non-musical drama texts mainly in two ways:

- the presence of *Concertato* sections
- the extreme fragmentation of the versification.

The *Concertato* is a musical term that is passed on in the librettos tradition to mean normally a scene or part of a scene performed simultaneously by different characters, each singing different texts, including several cues and stage directions. The number of simultaneous sequences can range from a minimum of two, to a maximum of seven/eight, as in the following example taken from Falstaff (music by Giuseppe Verdi and libretto by Arrigo Boito).



Pages 77-78 of the Falstaff libretto

In the libretto the versification is extremely fragmentary: as for the drama, verses and stanzas are usually split according to the different cues; furthermore, *fin de siècle* librettos admit different metres that can change at any moment, even within a cue.

An important point is that usually the libretto that is printed and distributed to the public can be markedly different from the one that is sung on the stage. In the score, verses and words are adapted to the musical progression and for that reason they can be stretched, repeated, modified, cut and added. The libretto is often conceived as a support for the spectator; in the libretto, indeed, portions of text suppressed in the score, stage directions, comments, notes that have no match with the performed opera, can help the spectator to follow the plot. All these peculiarities need to be seriously taken into consideration before starting any encoding. Firstly, this is because the librettos' printing tradition has fixed some conventions to represent the different characteristics. Second, the public to which a digital collection of librettos is addressed will expect its habits to be taken into account.

In the last two years a research project named "L'Opera prima dell'Opera" (*The text/literary source before the staging of Opera*) has carried out the creation of a digital library of librettos called *Opera Liber*, freely available on the Net (currently at `<http://80.19.150.245/operaliber/>` but will be soon transferred to `<http://www.operaliber.it>`). *Opera Liber* is a portal for the study and the documentation of the Italian librettos for the period 1870 - 1920, including works of the main Italian composers such as Verdi, Puccini, Leoncavallo, Mascagni, Ponchielli and many others. The main resource of the web site is represented by the collection of texts, available both for reading and for linguistic querying. The texts have been encoded in XML TEI format and are managed and queried using the native XML database eXist. The *Opera Liber DTD* is a customization of the TEI DTD P4, fully documented on the web site, and it is constituted by a mixed base set (verse and drama) and additional tag sets such as figure, transcriptions of primary sources, linking, and names and dates. Some customizations of

the DTD have been made, following the prescription of the Chapter 29 of the *Guidelines for Text Encoding and Interchange* (Sperberg-McQueen & Burnard).

In creating the encoding model the main problem was to find a correct encoding for *Concertatos*. The Concertato can surely be considered a sort of structural division, even if not at the same level of usual structural divisions (such as acts and scenes, encoded by the TEI `<div>` elements). A milestone approach that was also considered would miss the consistency of the *Concertato* sequences. For that reason we decided to create a new element `<sequences>` that will include a number of `<sequence>` elements, according to the number of columns in the printed form.

We decided also to consider as source physical copies of librettos, and not so called *ideal copies* and that because it is often difficult to determine the belonging of a copy to a particular edition or issue. Publishers, in fact, usually printed a large amount of librettos, storing unsold copies, just changing the front matters to fit the libretto they have in their repositories to a particular *mise en scène*, sometimes mixing copies from different printings. Furthermore, some of the copies we have considered for encoding contain manuscript notes or dedications. For all these reasons we settled on recording the provenance of the encoded copies, creating the element `<copyStmt>` (and the child elements `<settlement>` and `<repository>`) inside the `<sourceDesc>` element.

Another problem was given by the encoding of the name of characters. One of the peculiar characteristics of 'classic' drama (from Greek tradition till the beginning of the twentieth century) is the so called *agnition*, i.e. a character that is believed to be a certain person, is recognised to be someone else, often determining the unravelling of the plot. That means that a character may have two names, the supposed and the real one, but it is not two persons and that’s why the possibility of using two nested `<persName>` was refused. We decided, instead, to create a new attribute (called *alias*) for the `<persName>` element to record supposed or virtual names, reserving the *reg* attribute for real names.

The different kinds of metre have not been semantically encoded because in many case the difficulty of understanding the rationale invites caution. For that reason the metric divisions have been marked only in really obvious cases, while in other cases only the physical appearance of the verse has been encoded by the usage of the `<hi>` element. In such way we have recorded the presence of:

- particular indentations, normally, but not always, representing a changing of metre;
- inverted commas, normally representing not sung verses;
- dashes normally representing the alternation of voices in choral singing.

A number of minor implementations of the DTD have been operated, fully documented in the web site.

The web site *Opera Liber* collects the experience of two years' work in the field of encoding opera librettos and offers itself as a point of reference for analogous experiences.

Bibliography

Opera Liber. `<http://80.19.150.245/operaliber/>`

Sperberg-McQueen, C.M., and L. Burnard, eds. *TEI P4: Guidelines for Electronic Text Encoding and Interchange.* Text Encoding Initiative Consortium, 2002. Accessed 2004-10-09. `<http://www.tei-c.org/P4X/>`

Verdi, Giuseppe, and Arrigo Boito. *Falstaff.* New York: G. Schirmer, 1963.