

# Codifica XML-TEI: proposta di mark-up per i dizionari metodici

Barbara Patella

## 1. Vocabolari metodici: da prodotti “statici” a prodotti “dinamici”

Poiché il fulcro della ricerca di dottorato ricade sulla marcatura dei principali dizionari metodici italiani dell'Ottocento<sup>1</sup> – quindi su un'elaborazione informatica finalizzata a rendere interrogabili testi retrodigitalizzati<sup>2</sup> che, raccolti in una banca dati, saranno consultabili su una

- 1** Per un focus sulla tipologia della lessicografia metodica (o sistematica) si vedano in particolare MARELLO 1980 – riferimento fondamentale per questo filone di ricerca –, SESSA 1979 e APRILE 2023.
- 2** I dizionari inclusi nella banca dati sono i seguenti: 1) FRANCESCO TARANTO E CARLO GUACCI, *Vocabolario domestico italiano ad uso de' giovani, ordinato per categorie*, Napoli, Dalla Stamperia del Vaglio, 1849 (2<sup>a</sup> ed., 1851); 2) GIACINTO CARENA, *Prontuario di vocaboli attenenti a parecchie arti, ad alcuni mestieri, a cose domestiche, e altre di uso comune, per saggio di un vocabolario metodico della lingua italiana. Parte prima. Vocabolario domestico*, Torino, Stamperia Reale, 1846 (2<sup>a</sup> ed., 1851); 3) GIACINTO CARENA, *Prontuario di vocaboli attenenti a parecchie arti [...]. Parte seconda. Vocabolario metodico d'arti e mestieri*, Torino, Stamperia Reale, 1853; 4) GIACINTO CARENA, *Prontuario di vocaboli attenenti a parecchie arti [...]. Parte terza postuma contenente il vocabolario dei veicoli su terra, e dei veicoli su acqua, e frammenti relativi ai vocaboli mercantili, alla zecca, ed al cavalcare*, Torino, Stamperia Reale, 1860; 5) STEFANO PALMA, *Vocabolario metodico-italiano. Parte che si riferisce all'agricoltura e pastorizia, arti ed industrie che ne dipendono*, I-II, Milano, Libreria editrice di Educazione e d'Istruzione di Paolo Carrara, 1870; 6) STEFANO PALMA, *Prontuario di voci e maniere di dire del linguaggio mercantile, amministrativo ed economico*, Milano, Paolo Carrara Editore-Libraio, 1877; 7) ANGIOLINA BULGARINI, *Prontuario di voci concernenti i lavori donneschi*, Torino, Stamperia Reale di G.B. Paravia e comp., 1878; 8) AURELIO GOTTI, *Vocabolario metodico della lingua italiana. Casa*, Torino, Ditta

piattaforma online –, questo contributo vuole offrire una riflessione di metodo e, al contempo, una serie di applicazioni dimostrative riguardo a un sistema di mark-up in parte sperimentale. Parliamo di un sistema sperimentale perché il linguaggio che adoperiamo, l'XML-TEI<sup>3</sup>, è architettato e collaudato per funzionare prevalentemente su repertori semasiologici (in cui la ricerca procede da un significante noto al relativo significato ignoto) e non su repertori come i nostri che, oltre a fondarsi sul criterio opposto, quello onomasiologico (nei quali si parte da significati noti per giungere a significanti ignoti), presentano un ordinamento non alfabetico: le entrate, infatti, sono disposte all'interno di categorie tematiche (come “piante”, “abiti e accessori”, “cibi e bevande”, “casa e sue parti” ecc.) secondo associazioni concettuali e relazioni semantiche (soprattutto in base a rapporti di iponimia e meronimia).

Si è pensato, quindi, di informatizzare i dizionari metodici poiché rappresentano – per struttura, intenti e contenuti – fonti preziose per più rami della linguistica italiana (lessicografia, storia della lessicografia, storia della lingua, dialettologia), soprattutto per il tentativo rinnovatore dei compilatori, che furono, da un lato, attenti al recupero di quei settori lessicali trascurati dalla lessicografia tradizionale (dunque solleciti alla raccolta di voci tecniche di arti e mestieri, voci scientifiche, voci della conversazione) e, dall'altro, sensibili a forme ed espressioni dell'uso parlato (in controtendenza rispetto alla prassi lessicografica, fino ad allora fortemente letterariocentrica)<sup>4</sup> – non a caso, furono promotori di inchieste sul campo *ante litteram* (condotte principalmente in Toscana).

---

G.B. Paravia e comp., 1883; 9) FRANCESCO CORAZZINI, *La città e lo stato. La casa e la famiglia. Dizionario metodico*, Torino, Ermanno Loescher, 1885. Per una descrizione del progetto si rimanda a PATELLA 2023.

- 3 Per la marcatura dei dizionari metodici siamo ricorsi allo standard di codifica internazionale XML-TEI e alle linee guida distribuite dal consorzio TEI – con particolare riguardo al nono capitolo, riservato proprio all'elaborazione dei repertori lessicali (cfr. TEI P5 2021, pp. 291-329) –, giacché forniscono un assetto di istruzioni capace di regolamentare il mark-up e, al tempo stesso, di personalizzarlo, favorendo inoltre una sostenibilità del progetto e un'interoperabilità dei dati marcati.
- 4 Tradizionalmente, «nei dizionari della lingua italiana si era dato largo spazio [...] al lessico “alto” della prosa e della poesia, attraverso gli esempi e le citazioni tratte

Addentrandoci, allora, nei processi di trasformazione che da dati “statici” (vocabolari metodici cartacei) conducono a prodotti interrogabili e dinamici (vocabolari metodici digitalizzati e marcati), mostriamo alcuni dei criteri e degli espedienti che abbiamo adottato con lo scopo di potenziare la consultabilità e la consultazione di tali risorse. Considerate le particolari caratteristiche (a cominciare dall’ordinamento per temi e dall’eterogeneità della microstruttura), è stato necessario – in funzione del lavoro informatico – uno studio preliminare sia dei repertori stessi sia delle regole dell’XML-TEI, dal momento che una certa disomogeneità si riscontra non solo fra un dizionario metodico e l’altro (variabilità esterna), ma anche all’interno di uno stesso vocabolario (variabilità interna).

## 2. Mark-up a livello di macrostruttura

Relativamente agli apparati, per i nostri dizionari abbiamo optato per una marcatura essenzialmente “leggera”, che ha interessato l’annotazione di tali sezioni di corredo mediante l’impiego dei tag <front>, <back>, <div>, <head>, <pb>, <p>: ci siamo serviti dell’elemento <div> (*text division*) per delimitare qualunque blocco di testo, ossia ciascun esotesto<sup>5</sup> (o apparato) che può trovarsi in apertura (nel *front matter*: <front><sup>6</sup>) o in chiusura (nel *back matter*: <back><sup>7</sup>) di un dizionario<sup>8</sup>;

---

dagli scrittori, ma non c’era posto per le parole “basse” e comuni della vita quotidiana, che non erano nominate, ovviamente, nelle opere letterarie. Proprio per colmare questa lacuna, e per andare incontro alle richieste di un pubblico che sempre più spesso cercava nei dizionari il termine “nazionale” con cui designare gli oggetti che avevano nomi diversi da regione a regione, nell’Ottocento [...] esplose la moda dei dizionari metodici o “sistematici”» (DELLA VALLE-PATOTA 2006, pp. 138-139).

<sup>5</sup> Cfr. RICCIO 2016, p. 64.

<sup>6</sup> Cfr. TEI P5 2021, pp. 193-196.

<sup>7</sup> Cfr. *ivi*, pp. 198-200.

<sup>8</sup> Oltre a ciò, il tag <div> (previsto non solo negli apparati ma anche all’interno del <body>) serve inoltre per etichettare le suddivisioni contenute nel corpo del dizionario, dal momento che i dizionari metodici, organizzando le entrate in sezioni

a esso è subordinato il tag <head>, usato per racchiudere i titoli dei blocchi testuali; <pb> è invece il marcatore dei numeri di pagina (*page beginning*), mentre <p> (*paragraph*) circostringe ogni porzione di testo contenuta all'interno di <div>. In base a ciò, mediante il lavoro di mark-up è possibile isolare e rendere interrogabili (nell'ottica di una piattaforma online) le seguenti sezioni:

- prefazione/premessa/introduzione
- avvertimenti
- dedica
- postfazione
- note a piè di pagina
- altra sezione (senza specificazione)

A sostegno di quanto detto, forniamo due esempi tratti dal *Pronuario di voci e maniere di dire del linguaggio mercantile, amministrativo ed economico* di Stefano Palma (1877).

Nella prima delle due immagini (figura 1) riportiamo la tipica modalità di marcatura delle prefazioni: contenuta negli apparati iniziali (quindi inclusa nella sezione <front><sup>9</sup>), la prefazione è marcata con <div> (che ne indica l'inizio e, mediante attributo, il relativo valore: *type="preface"*), sotto cui si trovano annidati il titolo (delimitato da <head>) e il testo della prefazione (contenuto in <p>); infine, il tag di chiusura col quale si segnala la fine della sezione (</div>). Per giunta, se lo si ritiene necessario, il testo può essere suddiviso in più parti (cioè si può inserire un elemento <p> per ogni porzione testuale che si vuole etichettare), secondo lo schema seguente:

---

tematiche, presentano il lemmario suddiviso e articolato in parti, capitoli e/o paragrafi.

- <sup>9</sup> Grazie alla modalità di visualizzazione dell'*outline* (osservabile nella colonna di sinistra sia della figura 1 sia della figura 2), si può notare l'impianto strutturale del *Pronuario* di Palma: vediamo, ad esempio, che il *front matter* si compone di due blocchi testuali (annotati con <div>) oppure che il corpo del dizionario (*body*) è articolato in più capitoli.

```

<div type = "preface">
  <head> TITOLO DELLA PREFAZIONE </head>
  <p> testo </p>
  <p> testo </p>
  [...] [...] [...]
  <p> testo </p>
</div>

```

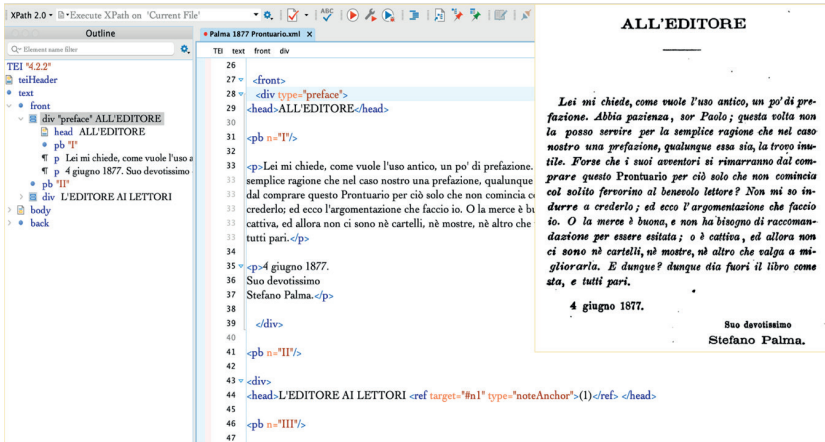


Figura 1 Schermata tratta da Oxygen XML Editor (a sinistra) con mark-up applicato alla prefazione contenuta in PALMA 1877 (riprodotta a destra, p. 1)

Nella seconda immagine (figura 2) osserviamo un altro apparato iniziale (che segue la prefazione appena vista), sempre delimitato da <div>, ma questa volta senza alcun attributo (@type) né valore di specificazione (come “preface”, “dedication” o simili). Qui, inoltre, notiamo che l’esotesto contiene un ulteriore elemento, cioè una nota a piè di pagina, che andrà annotata mediante l’uso di due stringhe:

```

<ref target="#n1" type="noteAnchor"> (1) </ref>
<note type="footnote" xml:id="n1"> testo della nota </note>

```

La prima stringa, a cui sono associati il riferimento al testo della nota (“#n1”) e il valore “noteAnchor”, è impiegata per marcare il numero – o, in altri casi, il simbolo – collegato alla nota; la seconda, contrassegnata dal valore “footnote” e da un identificatore univoco (“n1”), serve a marcare il testo della nota.

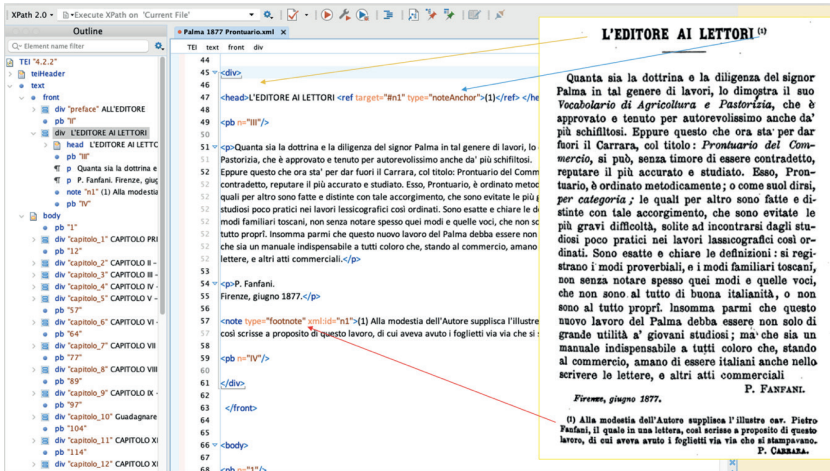


Figura 2 Schermata tratta da Oxygen XML Editor (a sinistra) con mark-up applicato a uno degli apparati iniziali contenuti in PALMA 1877 (riprodotto a destra, p. III)

## 2.1 Il tag <div> nel body

Considerato che tutti i vocabolari sistematici inclusi nella nostra banca dati prevedono – in base ai temi trattati – una suddivisione interna in più parti e/o in più capitoli (e talvolta persino in paragrafi) entro cui sono distribuite le entrate, ricorreremo all’elemento <div> anche nel corpo del dizionario (<body>), con lo scopo di isolare ogni suddivisione presente ed eventualmente specificarne il valore:

- <div> [...] </div> → sezione senza specificazione
- <div type = “part”> [...] </div> → parte (non numerata)

```

<div type = "part" n= "4"> [...] </div>    → parte 4
<div type = "chapter" n= "3"> [...] </div> → capitolo 3

<body>
<pb n="1"/>
<div type="chapter" n="1" xml:iid="capitolo_1">
<head> CAPO I. Dei lavori di maglia.</head>
<entry>
<form type="lemma" subtype="phrase"><orth>Lavori di maglia</orth></form> <sense>sono quelli che si fanno intrecciando il filo con ferri o
con altri arnesi per modo che formino tanti anelli l'uno attaccato all'altro; e sono: la <re
type="hyponym"></form><orth>Maglia</orth></form></re> propriamente detta, il <re
type="hyponym"></form><orth>Modano</orth></form></re>, l'<re type="hyponym"></form><orth>Uncinetto</orth></form></re>, il <re
type="hyponym"></form><orth>Chiacchierino</orth></form></re>, la <re type="hyponym"></form type="phrase"><orth>Trina a
tombolo</orth></form></re>.</sense>
</entry>
<entry>
<form type="lemma"><orth>Maglia</orth>
<form type="not-recommended"><orth> Punto </orth></form>
</form> <sense>Così chiamasi ciascun anello dei suddetti lavori.</sense>
</entry>

[...]   [...]   [...]

```

**Figura 3** Schermata tratta da *Oxygen* con esempio di mark-up applicato al corpo del dizionario di BULGARINI 1878 (p. 1)

### 3. Mark-up a livello di microstruttura

Quanto alla microstruttura dei vocabolari metodici, alquanto composita, la vera sfida del progetto consiste nel valorizzare e nel far emergere, da un lato, i numerosissimi sottoelementi che sono presenti all'interno degli articoli lessicografici – quali parole, locuzioni e proverbi non registrati come entrate autonome oppure i commenti dei compilatori – e, dall'altro, le relazioni semantiche sottese sia allo sviluppo verticale (costituito dal lemmario) sia a quello orizzontale (rappresentato dal materiale lessicale annidato sotto le voci a espone), vale a dire i rapporti di iponimia-iperonimia, olonimia-meronimia e sinonimia-antonimia, sui quali la lessicografia sistematica fonda la concatenazione di entrate e sottoentrate. Per tali ragioni, accanto agli elementi – per così dire – canonici (quali lemmi, campo

del significato e rinvio ad altre voci)<sup>10</sup> ci siamo prefissati di restituire informaticamente anche una serie di informazioni che differenziano la lessicografia metodica: i commenti, la giustapposizione di voci a esponente, la tipologia delle entrate registrate (forme flesse, locuzioni, proverbi), le voci correlate (al lemma o a voce interna) annidate all'interno degli articoli.

Alla luce di ciò, secondo lo schema configurato per la microstruttura, prevediamo la marcatura dei seguenti elementi:

lemmi	
Proposta di mark-up a livello di <b>microstruttura</b> :	campi correlati (collocati, sinonimi, meronimi, polirematiche, proverbi ecc.)
	commenti dei lessicografi
	indicatori d'uso / di registro
	rinvio ad altre voci

Nelle tabelle che presentiamo di seguito illustriamo, più in dettaglio, la serie degli elementi lessicali, semantici o testuali che sono stati oggetto di mark-up. Nella colonna di sinistra sono indicate le stringhe XML-TEI (con tag, attributi e valori), mentre a destra i corrispondenti elementi marcati a livello di microstruttura e di mediostruttura<sup>11</sup>:

<sup>10</sup> All'opposto, nel nostro modello di marcatura non abbiamo contemplato elementi considerati abituali negli articoli lessicografici, quali categorie grammaticali, esempi, etimologie, in quanto informazioni fornite sporadicamente o in modo irregolare e incoerente nei dizionari metodici, nei quali la redazione delle voci è spesso eterogenea e non predicibile.

<sup>11</sup> Cfr. RICCIO 2016, pp. 65-68.



Serie di tag, attributi e valori	Elementi marcati
<code>&lt;form type="lemma"&gt;&lt;orth&gt;&lt;/orth&gt;&lt;/form&gt;</code>	<i>Lemma</i> (unità lessicale semplice)
<code>&lt;form type="lemma" subtype="inflected"&gt;&lt;orth&gt;&lt;/orth&gt;&lt;/form&gt;</code>	<i>Lemma</i> (unità lessicale in forma flessa)
<code>&lt;form type="lemma" subtype="phrase"&gt;&lt;orth&gt;&lt;/orth&gt;&lt;/form&gt;</code>	<i>Lemma</i> (unità lessicale di due o più parole)
<code>&lt;form type="variant"&gt;&lt;orth&gt;&lt;/orth&gt;&lt;/form&gt;</code>	<i>Variante del lemma</i>
<code>&lt;form type="obsolete"&gt;&lt;orth&gt;&lt;/orth&gt;&lt;/form&gt;</code>	<i>Forma obsoleta</i>
<code>&lt;form type="not-recommended"&gt;&lt;orth&gt;&lt;/orth&gt;&lt;/form&gt;</code>	<i>Parola/ Locuzione sconsigliata dal lessicografo</i>
<code>&lt;re&gt;&lt;form&gt;&lt;orth&gt;&lt;/orth&gt;&lt;/form&gt;&lt;/re&gt;</code>	<i>Voce correlata</i> al lemma o a <code>&lt;re&gt;</code>
<code>&lt;re type="colloc"&gt;&lt;form&gt;&lt;orth&gt;&lt;/orth&gt;&lt;/form&gt;&lt;/re&gt;</code>	<i>Parola co-occorrente</i> con il lemma o con <code>&lt;re&gt;</code>
<code>&lt;re type="hyperonym"&gt;&lt;form&gt;&lt;orth&gt;&lt;/orth&gt;&lt;/form&gt;&lt;/re&gt;</code>	<i>Iperonimo</i> (del lemma o di <code>&lt;re&gt;</code> )
<code>&lt;re type="hyponym"&gt;&lt;form&gt;&lt;orth&gt;&lt;/orth&gt;&lt;/form&gt;&lt;/re&gt;</code>	<i>Iponimo</i> (del lemma o di <code>&lt;re&gt;</code> )
<code>&lt;re type="meronym"&gt;&lt;form&gt;&lt;orth&gt;&lt;/orth&gt;&lt;/form&gt;&lt;/re&gt;</code>	<i>Meronimo</i> (del lemma o di <code>&lt;re&gt;</code> )
<code>&lt;re type="derivation"&gt;&lt;form&gt;&lt;orth&gt;&lt;/orth&gt;&lt;/form&gt;&lt;/re&gt;</code>	<i>Base/Derivato</i> del lemma o di <code>&lt;re&gt;</code>
<code>&lt;re type="antonym"&gt;&lt;form&gt;&lt;orth&gt;&lt;/orth&gt;&lt;/form&gt;&lt;/re&gt;</code>	<i>Antonimo</i>
<code>&lt;re type="synonym"&gt;&lt;form&gt;&lt;orth&gt;&lt;/orth&gt;&lt;/form&gt;&lt;/re&gt;</code>	<i>Sinonimo</i>
<code>&lt;re type="geo-synonym"&gt;&lt;form&gt;&lt;orth&gt;&lt;/orth&gt;&lt;/form&gt;&lt;/re&gt;</code>	<i>Geosinonimo</i>
<code>&lt;re type="foreign"&gt;&lt;form&gt;&lt;orth&gt;&lt;/orth&gt;&lt;/form&gt;&lt;/re&gt;</code>	<i>Forestierismo</i>

Figura 4 Tabella con tipologie di lemmi e di voci correlate

Serie di tag, attributi, valori	Elementi marcati
<code>&lt;xr type="hyperonym"&gt;&lt;ref target="#lemma"&gt; &lt;/ref&gt;&lt;/xr&gt;</code>	Rinvio a un <i>iperonimo</i> (lemmatizzato)
<code>&lt;xr type="hyponym"&gt;&lt;ref target="#lemma"&gt; &lt;/ref&gt;&lt;/xr&gt;</code>	Rinvio a un <i>iponimo</i> (lemmatizzato)
<code>&lt;xr type="meronym"&gt;&lt;ref target="#lemma"&gt; &lt;/ref&gt;&lt;/xr&gt;</code>	Rinvio a un <i>meronimo</i> (lemmatizzato)
<code>&lt;xr type="holonym"&gt;&lt;ref target="#lemma"&gt; &lt;/ref&gt;&lt;/xr&gt;</code>	Rinvio a un <i>olonimo</i> (lemmatizzato)
<code>&lt;xr type="derivation"&gt;&lt;ref target="#lemma"&gt; &lt;/ref&gt;&lt;/xr&gt;</code>	Rinvio al <i>derivato</i> o alla <i>base</i> del lemma
<code>&lt;xr type="antonym"&gt;&lt;ref target="#lemma"&gt; &lt;/ref&gt;&lt;/xr&gt;</code>	Rinvio a un <i>antonimo</i> (lemmatizzato)
<code>&lt;xr type="synonym"&gt;&lt;ref target="#lemma"&gt; &lt;/ref&gt;&lt;/xr&gt;</code>	Rinvio a un <i>sinonimo</i> (lemmatizzato)
<code>&lt;entry type="holonym"&gt;</code>	<i>Olonimo</i> (di altre <i>entries</i> " <i>meronym</i> ")
<code>&lt;entry type="meronym"&gt;</code>	<i>Meronimo</i> (di una <i>entry</i> " <i>holonym</i> ")
<code>&lt;entry type="hyperonym"&gt;</code>	<i>Iperonimo</i> (di altre <i>entries</i> " <i>hyponym</i> " annidate)
<code>&lt;entry type="hyponym"&gt;</code>	<i>Iponimo</i> (di una <i>entry</i> " <i>hyponym</i> ")
<code>&lt;entry type="xref"&gt;</code>	Entrata con rinvio a un'altra voce
<code>&lt;sense&gt; &lt;/sense&gt;</code>	Campo del significato

Serie di tag, attributi, valori	Elementi marcati
<sense value="comment" n="35"> </sense>	Note di commento del lessicografo numerate
<sense value="comment"> </sense>	Note di commento del lessicografo non numerate
<usg> </usg>	Indicazioni d'uso, di registro o simili
<superEntry></superEntry>	Gruppo di voci annidate
<span type="escludi"> </span>	Porzione di testo esclusa dalla marcatura

Figura 5 Tabella con tipologie di rinvii, entrate e altri elementi

### 3.1 Voci sconsigliate dai lessicografi: il caso del *Prontuario* di Bulgarini

Un'ulteriore caratteristica che contraddistingue i vocabolaristi meto-  
dici – in linea col fatto di concepire i propri dizionari come “manuali  
di guida” – riguarda la registrazione di parole e locuzioni vitande per  
mettere in guardia il lettore. Per questo genere di voci, un caso partico-  
lare è rappresentato dal *Prontuario di voci concernenti i lavori donneschi*, in  
cui Angiolina Bulgarini allestisce un apparato a piè di pagina destinato  
a un uso particolare, come si legge in apertura del volume:

Ho poi notato a piè di pagina i vocaboli stranieri o inesatti che un uso,  
secondo me men buono, o certi scrittori di guide di lavori donneschi ci fanno  
adoperare invece di quelli più italiani e più propri che pur il buon uso vivente  
ci porge, e del quale appunto sono le voci da me segnate<sup>12</sup>.

In tal caso è stato necessario progettare una soluzione alternativa,  
rispetto all'originale cartaceo, per trattare informaticamente queste  
voci, così da recuperarle e restituirle in funzione delle ricerche sulla  
piattaforma: nel mark-up abbiamo perciò rinunciato a etichettarle  
come note a piè di pagina, perché ciò che ci interessava non era il luogo  
occupato nella pagina, bensì il valore di questo materiale lessicale; da  
qui si è deciso di farle “risalire” in fase di marcatura accanto ai lemmi,  
etichettandole come “voci non raccomandate” tramite la stringa

<sup>12</sup> BULGARINI 1878, *Avvertenza*, p. IV.

<form type="not-recommended"> <orth> voce scongiata </orth> </form>

Mostriamo allora un'immagine del *Prontuario* (BULGARINI 1878, p. 7) con la corrispondente resa nell'editor XML (*Oxygen*), in cui è possibile osservare due casi di voci non raccomandate: il primo riguarda l'entrata *maglie sode*, a cui si accompagna una numerazione (posta fra parentesi tonde) che rimanda a due voci scongiate (*briglie*, *barrette*) – e per questo collocate a piè di pagina dalla lessicografa –; il secondo caso concerne il lemma *pippiolini*, al quale è associato – mediante la stessa modalità di *maglie sode* – il termine *picot* (francesismo da evitare). In entrambi i casi, in fase di mark-up, abbiamo fatto risalire le voci non raccomandate, affiancandole al lemma a cui sono connesse:

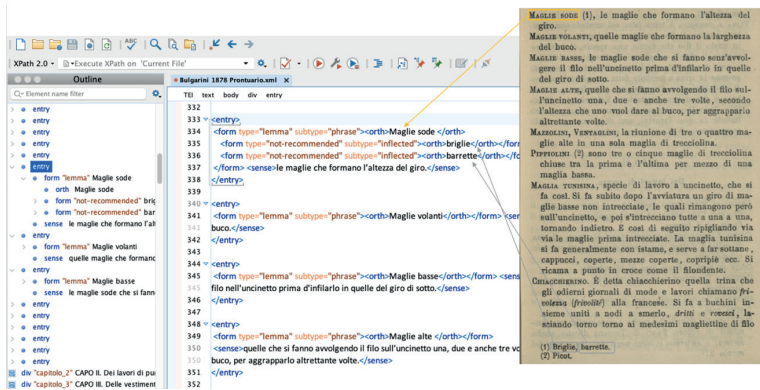


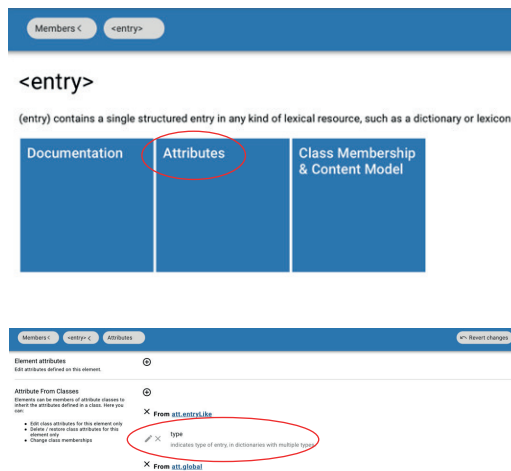
Figura 6 Schermata tratta da *Oxygen* (a sinistra) con mark-up applicato a BULGARINI 1878 (riprodotto a destra, p. 7)

#### 4. Personalizzazione dello schema con *Roma web application*

Come abbiamo accennato a più riprese, poiché i vocabolari metodici presentano una larga eterogeneità strutturale, è stato fondamentale poter personalizzare il modello di mark-up (lo schema XML-TEI) e adattarlo alle nostre esigenze al fine di estrapolare informazioni ed elementi linguisticamente interessanti: per questo ci siamo serviti di

un'applicazione web fornita dal consorzio TEI, ossia *Roma web application*<sup>13</sup>, che consente di aggiungere, rimuovere o modificare elementi, attributi e valori prestabiliti rispetto allo standard di riferimento; questo strumento permette di diversificare il mark-up in base ai nostri scopi e potenziare la consultabilità dei vocabolari metodici (a partire dalla marcatura delle relazioni semantiche), rimanendo tuttavia aderenti alle norme imposte dalle linee guida (*P5 Guidelines*)<sup>14</sup>.

Presentiamo allora alcuni passaggi dell'intero procedimento con cui è possibile personalizzare lo schema<sup>15</sup>, prendendo come esempio le modifiche relative a <entry>, tag che delimita le entrate. Dopo aver selezionato l'elemento di nostro interesse (<entry>), tramite l'opzione "Attributes", possiamo scegliere relativi attributi e valori da modificare; per questa simulazione scegliamo l'attributo @type:



13 Specifichiamo che abbiamo testato la nuova versione di questo tool (raggiungibile all'indirizzo <https://romabeta.tei-c.org>), che, pur essendo ancora in versione beta, funziona correttamente.

14 Cfr. TEI P5 2021.

15 Di seguito riportiamo alcune schermate esemplificative tratte dall'applicazione *Roma*.

Decidiamo, quindi, se aggiungere o eliminare i valori associati a *entry* (mediante *@type*) in base al disegno di tagset prestabilito e alle funzionalità di ricerca che vorremo rendere disponibili. Mostriamo il caso in cui aggiungiamo dei valori non contemplati dall'XML-TEI, fra cui ad esempio “meronym” e “hyperonym” (per i quali forniamo anche una descrizione della funzione che assumono nel sistema di marcatura):

Values  
Set values for this attribute.

**hom** (homograph) groups information relating to one homograph within an entry.

---

**Description**  
Contains a brief description of the object documented by its parent element, typically a documentation element or an entity.

```
1 <desc xmlns="http://www.tei-c.org/ns/1.0"
  versionDate="2008-02-01" xml:lang="en"
  >groups information relating to one
  homograph within an entry.</desc>
```

---

**xref** (cross reference) a reduced entry whose only function is to point to another main entry (e.g. for forms of an irregular verb or for variant spellings: was pointing to be, or esthete to aesthete).

---

**meronym**

**Description**  
Contains a brief description of the object documented by its parent element, typically a documentation element or an entity.

```
1 <desc xmlns="http://www.tei-c.org/ns/1.0"
  versionDate="2023-02-23" xml:lang="en"
  >an entry that is a meronym of another
  entry </desc>
```

---

**hyperonym**

**Description**  
Contains a brief description of the object documented by its parent element, typically a documentation element or an entity.

```
1 <desc xmlns="http://www.tei-c.org/ns/1.0"
  versionDate="2023-02-23" xml:lang="en"
  >an entry that is a hyperonym for one or
  more entries </desc>
```

Una volta completato il procedimento, l'applicazione permette di generare l'XML *schema* personalizzato e salvare il documento in più formati (*RelaxNG schema*, *W3C schema*, *DTD*, ecc.); a questo punto lo schema può essere associato a un determinato testo all'interno di un editor XML – nel nostro caso specifico in *Oxygen*, dove la validità è confermata dai valori che (correttamente inseriti con *Roma* e riconosciuti dall'editor) appaiono nel menù a tendina accanto all'attributo *@type* (collegato a *<entry>*). Come si vede dalla schermata, vengono suggeriti sia i valori preimpostati dalla TEI (“abbr”, “affix”, “foreign”, “hom”, “main”, “supplemental”, “xref”) sia quelli aggiunti da noi (“hyperonym”, “derivation”, “holonym”, “hyponym”, “meronym”):

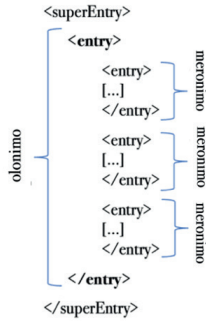


Figura 7 Schermata di Oxygen contenente il testo del *Vocabolario domestico di Carena* (1851)

#### 4.1 Il tag <superEntry>

Restando nell'ambito della personalizzazione dello schema di marcatura, segnaliamo che in qualche caso siamo andati a reinventare o a ridefinire la funzione dei tag, come ad esempio per <superEntry>. Secondo le norme, <superEntry> può essere impiegato per raggruppare «a sequence of entries within any kind of lexical resource, such as a dictionary or lexicon with function as a single unit, for example a set of homographs»<sup>16</sup>; noi ne abbiamo rideterminato il ruolo – estendendolo – in conformità all'ordine ideologico dei vocabolari metodici, impiegandolo come tag “contenitore” per raggruppare non solo eventuali omografi, ma anche entrate legate da rapporti semantici (es. olonimo/meronimi o iperonimo/iponimi) o lessicali (es. corradicali; base/derivato), con lo scopo di poterle isolare e individuare nelle opzioni di ricerca del sito. Pertanto, per un olonimo e i relativi meronimi (come pure per iperonimi e iponimi) il criterio gerarchico da rispettare sarà il seguente:

<sup>16</sup> TEI P5 2021, p. 292.



In conclusione mostriamo la modalità d'impiego di `<superEntry>`, applicandola a uno dei repertori della banca dati; nella fattispecie, proponiamo un caso di meronimia presente nel *Vocabolario domestico* di Carena (1851). Nel primo capitolo, dedicato al *Vestire e alle sue accompagnature*, il lessicografo fa seguire all'olonimo *anello* i meronimi che ne dipendono (*gambo*, *castone*, *fondo del castone*, *fascia del castone*, *foglia*); grazie al mark-up, siamo in grado di restituire la relazione implicata e raggruppare le entrate semanticamente connesse all'interno di `<superEntry>`, in base al seguente schema di annidamento:

```

<entry>
  <entry>
    [...] [...] [...]
  </entry>
</entry>

```

Osserviamo allora da vicino, relativamente agli elementi in questione, la marcatura che abbiamo realizzato all'interno dell'editor:

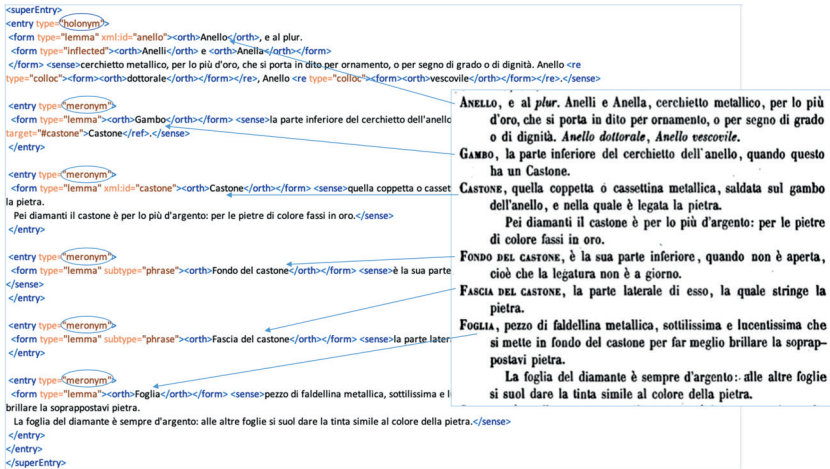


Figura 8 Schermata di Oxygen (a sinistra) con mark-applicato a CARENA 1851 (riprodotto a destra, p. 32)

## Conclusioni

In queste pagine, che si inseriscono nel solco della lessicografia digitale, abbiamo voluto offrire un piccolo saggio dei criteri e dei procedimenti di mark-up adottati per i dizionari metodici, una particolare tipologia lessicografica che, per il criterio onomasiologico e per l'ordinamento non alfabetico, ha rappresentato un banco di prova per il linguaggio XML-TEI. Attraverso esempi e applicazioni di marcatura, abbiamo dato dimostrazione delle operazioni con cui è possibile trasformare dizionari metodici cartacei in prodotti digitali interrogabili, mostrando peraltro alcune delle soluzioni necessariamente escogitate qualora lo standard TEI non contemplasse condizioni analoghe a quelle dei nostri repertori.

Pertanto, con l'informatizzazione dei vocabolari metodici si prospettano notevoli vantaggi – in termini sia quantitativi sia qualitativi – in previsione delle ricerche che verranno predisposte sulla piattafor-



ma online e, di conseguenza, dei risultati che il supporto informatico consentirà di estrapolare ed elaborare.

## Bibliografia

- APRILE 2023 = MARCELLO APRILE, *I dizionari metodici nell'Ottocento*, in Emiliano Picchiorri e Maria Silvia Rati (a cura di), *La lessicografia italiana dell'Ottocento. Bilanci e prospettive di studio*, Atti del Convegno, Università "G. D'Annunzio" di Chieti-Pescara (Chieti, 24-25 maggio 2022), Firenze, Cesati, 2023, pp. 101-123.
- BULGARINI 1878 = ANGIOLINA BULGARINI, *Prontuario di voci concernenti i lavori donneschi*, Torino, Stamperia Reale di G.B. Paravia e comp., 1878.
- CARENA 1851 = GIACINTO CARENA, *Prontuario di vocaboli attenenti a parecchie arti, ad alcuni mestieri, a cose domestiche, e altre di uso comune, per saggio di un vocabolario metodico della lingua italiana. Parte prima. Vocabolario domestico*, Torino, Stamperia Reale, 1846 (2<sup>a</sup> ed., 1851).
- DELLA VALLE-PATOTA 2006 = VALERIA DELLA VALLE E GIUSEPPE PATOTA, *L'italiano. Biografia di una lingua*, Milano, Sperling & Kupfer, 2006.
- MARELLO 1980 = CARLA MARELLO, *Lessico ed educazione popolare. Dizionari metodici dell'800*, con introduzione di Giovanni Nencioni, Roma, Armando, 1980.
- PALMA 1877 = STEFANO PALMA, *Prontuario di voci e maniere di dire del linguaggio mercantile, amministrativo ed economico secondo il buon uso toscano, in servizio delle scuole tecniche e commerciali. Con parole preliminari di P. Fanfani*, Milano, Paolo Carrara Editore-Libraio, 1877.
- PATELLA 2023 = BARBARA PATELLA, *Dizionari metodici dell'Ottocento: verso una piattaforma interrogabile*, in Emiliano Picchiorri e Francesco Montuori (a cura di), *In fieri, 4. Ricerche di linguistica italiana*, Atti della IV Giornata dell'ASLI per i dottorandi (Firenze, Accademia della Crusca, 2-4 dicembre 2021), Firenze, Cesati, 2023, pp. 241-247.
- RICCIO 2016 = ANNA RICCIO, *Gli strumenti per la ricerca linguistica. Corpora, dizionari e database*, Roma, Carocci, 2016.
- SESSA 1979 = MIRELLA SESSA, *Arti e mestieri nei vocabolari metodici*, in Atti del "Convegno Nazionale sui lessici tecnici delle arti e dei mestieri" (Cortona «Il Palazzone», 28-30 maggio 1979), 1979, pp. 37-78.

## Barbara Patella

TEI P5 2021 = *Guidelines for Electronic Text Encoding and Interchange* (April, 2022), Text Encoding Initiative Consortium (consultabile all'indirizzo <https://tei-c.org/release/doc/tei-p5-doc/it/html/index.html>).

**Riassunto** Il contributo, incentrato sull'informatizzazione dei principali dizionari metodici italiani dell'Ottocento, si focalizza in particolar modo sulle scelte di mark-up (a livello di macrostruttura e di microstruttura) che, basate sull'XML-TEI, sono state adottate – e in parte personalizzate – per rendere interrogabili questa tipologia di repertori onomasiologici non ordinati alfabeticamente, così da potenziarne la consultazione e garantirne modalità di ricerca avanzate (nell'ottica di una piattaforma che, realizzata con la collaborazione degli informatici, sarà fruibile in rete); inoltre, ai criteri teorici si accompagnano alcune applicazioni dimostrative che sono alla base della costruzione della banca dati dedicata alla lessicografia metodica, in modo tale da offrire un piccolo saggio sia dell'attività di marcatura sia dei procedimenti che dal dato cartaceo conducono al dato digitale.

**Abstract** This contribution, centred on the computerisation of the main Italian non-alphabetic dictionaries of the 19th century, focuses in particular on the mark-up choices (at the level of macrostructure and microstructure) which, based on XML-TEI, have been adopted – and partly customised – to make this type of onomasiological repertories searchable, so as to enhance its consultation and guarantee advanced research options (with a view to an online platform that will be created with the collaboration of computer scientists). Moreover, the theoretical criteria are accompanied by some demonstrative applications which are at the basis of the construction of the database dedicated to non-alphabetic lexicography, in order to offer a small sample of the mark-up activity as well as of the procedures that lead from the paper data to the digital data.