

TEI-Dict

un modulo per la codifica di Dizionari

Andrea Bellandi¹ - Fahad Khan²

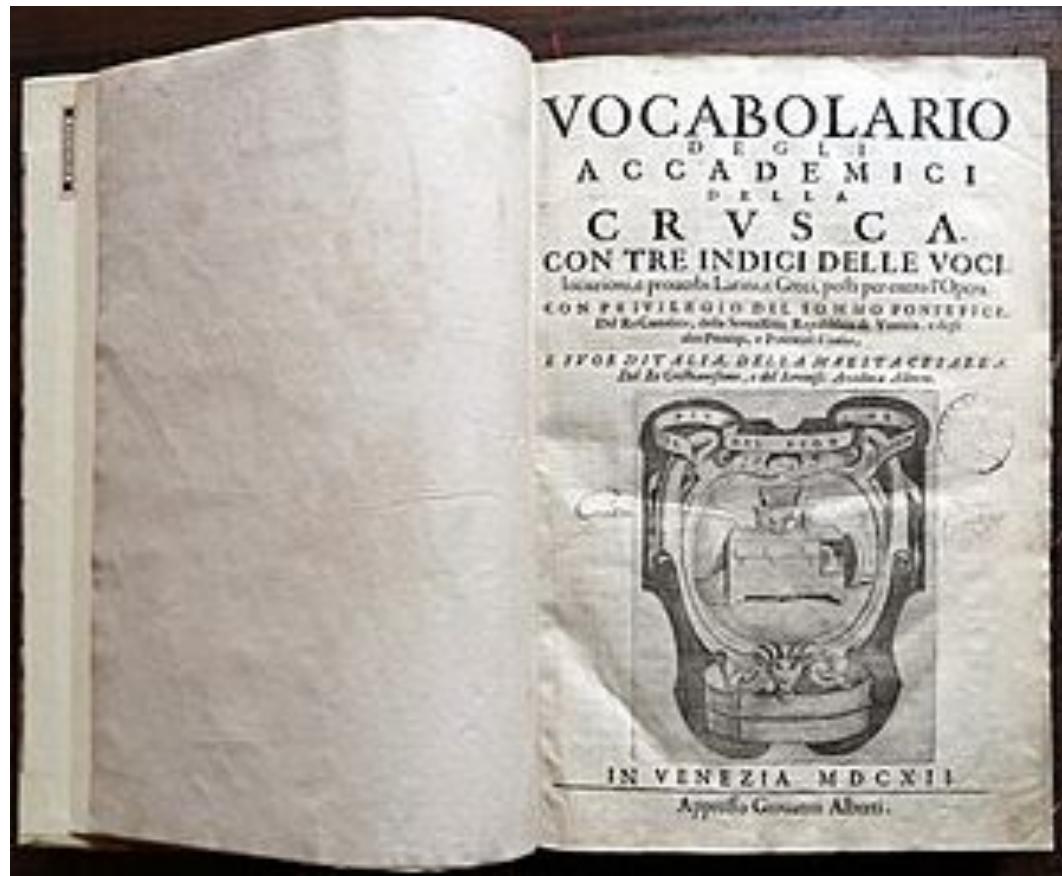
andrea.bellandi@ilc.cnr.it fahad.khan@ilc.cnr.it

¹ Literary Computing Group (<http://licolab.ilc.cnr.it/>)

² Language Resources and Infrastructures Group (<http://lari.ilc.cnr.it/it/>)

Istituto di Linguistica Computazionale “A. Zampolli”,
22/10/2018

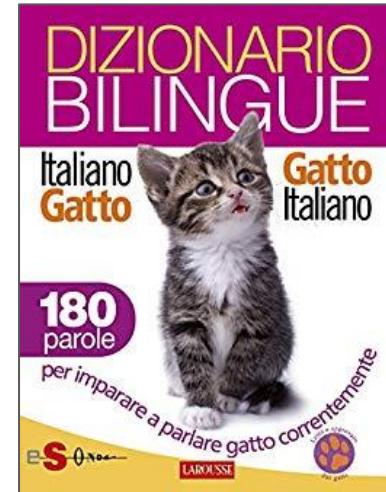
Il Dizionario



Il Vocabolario degli Accademici della Crusca è stato il primo vocabolario della lingua italiana. Uscito nel 1612, fu realizzato e pubblicato dall'Accademia della Crusca, istituzione culturale fondata a Firenze nel 1583.

Tipologie di Dizionario

- ❑ I dizionari si possono categorizzare in base a due aspetti principali:
 - ❑ al loro **contenuto**



- ❑ al **supporto fisico** che contiene il dizionario stesso

- ❑ **Cartaceo:** il dizionario è stampato su carta, nel classico formato di un libro, solitamente di generose dimensioni
- ❑ **Elettronico:** il dizionario è in forma digitale, contenuto su un CD-ROM o sull'hard disk di un computer; in questo caso l'ordinamento alfabetico dei vocaboli perde significato, in quanto l'accesso è solitamente diretto
- ❑ **Online:** un particolare dizionario elettronico che si può consultare collegandosi ad internet.

Tipologie di Dizionario: dettaglio (1)

Vediamo un po' più in dettaglio diversi tipi di dizionari:

Dizionari storici

registrano la tradizione letteraria scritta attraverso la citazione di esempi d'autori, per **testimoniare l'uso delle singole parole o delle locuzioni nelle varie epoche.**

Dizionari dell'uso

registrano la lingua contemporanea in una dimensione sincronica, nel funzionamento e nei caratteri attuali, ma prendendo in considerazione anche voci del passato, arcaiche o antiquate, varietà regionali, voci letterarie.

Dizionari etimologici

hanno il compito di tracciare la biografia di una parola, ripercorrendone la storia attraverso la documentazione scritta, dalla prima attestazione conosciuta fino ad oggi, e descrivendone le modificazioni di significato subite nel corso del tempo.

Dizionari dei sinonimi

registrano, per ogni lemma, i rispettivi **sinonimi**.

Tipologie di Dizionario: dettaglio (2)

Dizionari di neologismi

registrano le parole e le locuzioni nuovi d'una lingua, hanno lo scopo di testimoniare l'innovazione lessicale nelle sue varie manifestazioni e riportano con ampiezza anche voci legati a momenti ed episodi particolari, a mode e tendenze estemporanee.

Dizionari encyclopedici

condividono molti aspetti dei dizionari dell'uso e dei dizionari storici, perché **fondono in un'unica opera la parte encyclopedica e la parte più propriamente linguistica**. Questi dizionari comprendono anche i nomi propri.

Concordanze

sono repertori alfabetici delle parole contenute in una o più opere d'un autore, con l'indicazione e la citazione di tutti i luoghi in cui esse ricorrono.

Dizionari d'ortografia e pronuncia

si limitano a indicare la **corretta grafia e pronuncia delle parole della lingua italiana**.

Dizionari dialettali

In essi si cercano le **corrispondenze tra le voci di un dialetto e quelle italiane**, e si dà largo spazio alla fraseologia.

Tipologie di Dizionario: dettaglio (3)

□ Dizionari inversi e rimari

nei quali le parole sono disposte in ordine alfabetico rovesciato, partendo dall'ultima lettera. Se ne servono i **linguisti** (per individuare tutte le parole che hanno lo stesso suffisso o per analizzare la struttura delle parole composte), i **poeti**, i **parolieri** (che hanno bisogno di trovare con facilità tutte le parole che possono far rima) e gli **enigmisti**.

□ Dizionari analogici

permettono la **ricerca lessicale sulla base di approcci concettuali**. Hanno lo scopo di permettere il reperimento di una parola della quale si ignora (o non si ricorda) l'eventuale espressione lessicale. Il dizionario analogico risulta utile quando si conosce il significato associato alla parola da cercare o si conosce una parola con un significato correlato a quella cercata. Non contengono definizioni, e le parole sono raggruppate per campi semantici.

□ Dizionari delle collocazioni

permettono di trovare le collocazioni, quel particolare **tipo di combinazioni lessicali** (chiamate tecnicamente «co-occorrenze»). Esempi sono: mix esplosivo, vedova inconsolabile, lamiere contorte, criminali incalliti, sforzo spasmodico, svantaggio immeritato, ecc.

Un'entrata di dizionario



bambino

- 1a. s.m. , l'essere umano tra la nascita e l'inizio della fanciullezza: *i bambini giocano nel parco, libri, film, giochi per bambini | fare il bambino*, comportamento infantile; *non sono più un bambino*, sono adulto | fam., al pl. può indicare collettivamente figli e figlie: *ho due bambini, la stanza dei bambini*
- 1b. s.m. , l'essere umano maschio tra la nascita e l'inizio della fanciullezza | fam., figlio maschio: *ho un bambino e due bambine*
- 1c. s.m. , come appellativo affettuoso o ironico riferito ad adulti di sesso maschile: *bambino mio, cosa vuoi farci?*
2. s.m. , adulto molto ingenuo
3. s.m. , solo sing., per anton., con iniz. maiusc., Gesù Bambino
- 4a. agg. , molto giovane: *sposa bambina*
- 4b. agg. , fig., di qcн., ingenuo, semplice: *Carlo è ancora molto bambino | fig., di qcн., immaturo, non compiutamente sviluppato: mente bambina; civiltà bambina, agli albori*

Fonte: Nuovo De Mauro

bambino s. m. (f. -a) [da una radice onomatopeica *bamb-]. – 1. L'essere umano nell'età compresa tra la nascita e l'inizio della fanciullezza. Nell'uso com., spec. al plur. o con l'articolo indet., *bambino* può indicare genericam. maschi e femmine: *aspettare un b.*, riferito a gestanti; *allattare, svegliare i bambini*. Distingue i bambini da altri esseri umani. Altri significati: *bambino prodigo* (piuttosto frequente nella forma fr. *enfant prodige*), bambino precocità di sviluppo intellettuale ed eccezionali capacità artistiche, musicali, ecc.

Fonte: Treccani on-line

Data un'entrata, Il tipo di informazione e la sua rappresentazione, possono essere diversi in differenti vocabolari

- 2. l'essere umano nell'età che va dalla nascita all'inizio della fanciullezza: *allevare, educare un bambino, una bambina | Il Bambino*, per antonomasia, nella religione cristiana, Gesù quando era bambino (anche nelle espressioni *Gesù Bambino o il Bambinello*); l'immagine, la statuina che lo rappresenta dim. *bambinello, bambinetto, accr. bambinone, vezz. bambinuccio*
2. figlio: *aspettare un bambino; ha tre bambini, un maschio e due femmine*
3. persona adulta che si comporta in modo infantile: *ha trent'anni, ma è un bambino*
4. in espressioni affettuose o ironiche si rivolge anche a ragazzi o persone adulte: *bambino mio, devi avere pazienza!*

Fonte: GARZANTI on-line

Entrata: “codificare”

codificare

codificare v. tr. [dal fr. *codifier*, der. di *code* «codice»] (io codifico, tu codifichi, ecc.). – 1. a. Ridurre in codice, dare cioè un ordine sistematico a un complesso di norme giuridiche relative a una determinata materia: *c. il diritto civile, il diritto internazionale; c. le leggi sulla stampa, sui contratti di lavoro*. b. estens. Sistemare in modo ordinario e coerente un insieme di regole; dare valore normativo a usi o consuetudini *c. le regole grammaticali; norme ortografiche codificate dall'uso degli scrittori più autorevoli* 2. Esprimere informazioni e messaggi mediante le regole e i simboli di un sistema convenzionale (il *codice*) stabilito concordemente dall'emittitore e dal ricevitore dei messaggi allo scopo di trasmettere o elaborare automaticamente le informazioni o, talora, di mantenerle segrete: *c. un ordine, un messaggio, un'istruzione*. Riferito a calcolatori elettronici, convertire istruzioni del programma e dati nel codice di macchina 3. In biologia, inserire gli aminoacidi portati dall'RNA di trasferimento, per corrispondenza tra la tripletta di questo e le triplette dell'RNA messaggero, durante la sintesi proteica. ♦ Part. pass. **codificato**, anche come agg.: *il canone ormai codificato della figura umana è quello stabilito da Policleto.*

Fonte: Treccani on-line.

Quali informazioni dovrei codificare



*entrata/
forma
categoria
grammaticale*



etimologia



sensi



definizioni



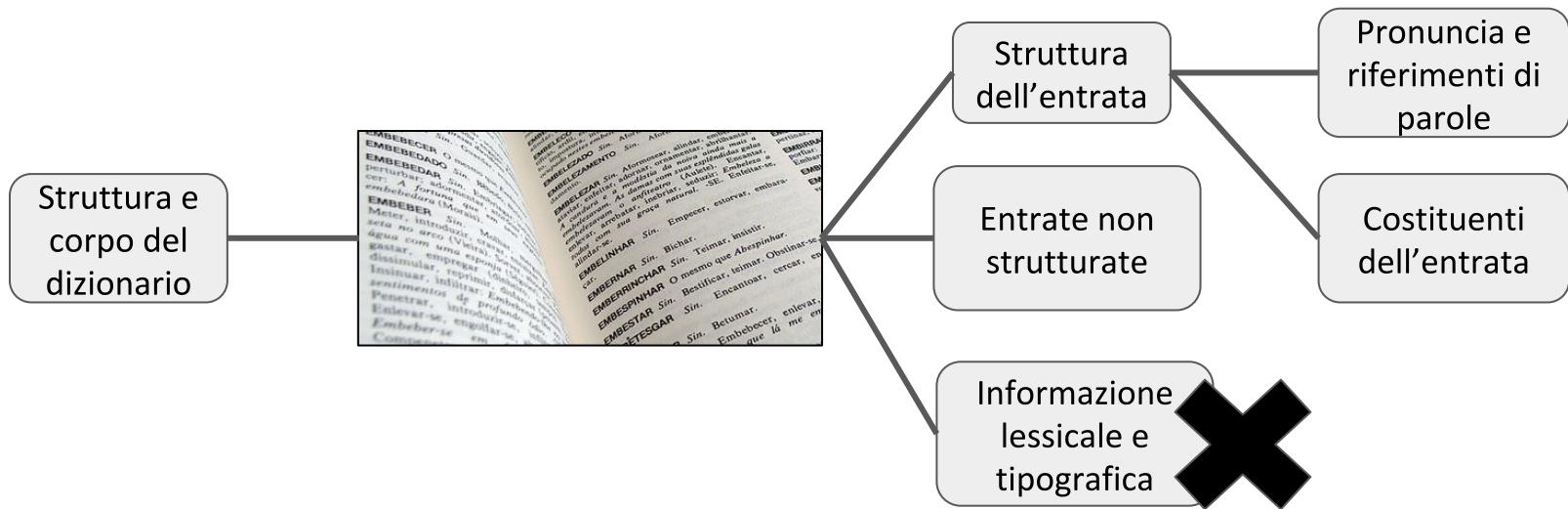
riferimento



esempio d'uso

TEI-Dict: Generalità

- TEI-Dict è un modulo specifico della TEI, ma in generale **non si limita alla sola codifica di dizionari**, può essere anche utilizzato per codificare risorse lessicali di vario tipo: glossari, lessici computazionali, ecc.



- Le guidelines di TEI-Dict sono reperibili sul sito della TEI al capitolo 9
<http://www.tei-c.org/release/doc/tei-p5-doc/en/html>

TEI-Dict: è solo uno dei moduli TEI ...



1 [The TEI Infrastructure](#)

2 [The TEI Header](#)

3 [Elements Available in All TEI Documents](#)

4 [Default Text Structure](#)

5 [Characters, Glyphs, and Writing Modes](#) → Rappresentazioni dei caratteri, caratteri non-standard, ecc...

6 [Verse](#) → Poesie con metriche, rime, ecc...

7 [Performance Texts](#) → Sceneggiature, drammi, trascrizioni., ecc.

8 [Transcriptions of Speech](#) → Trascrizioni di parlato

9 [Dictionaries](#)

10 [Manuscript Description](#) → Codifica delle descrizioni di manoscritti

Codifica manoscritti ←

Apparati critici per edizioni
critiche ←

Codifica tabelle, formule,
immagini, ecc... ←

11 [Representation of Primary Sources](#)

12 [Critical Apparatus](#)

13 [Names, Dates, People, and Places](#) → Codifica di entità nominate,
luoghi, date, ecc...

14 [Tables, Formulæ, Graphics and Notated Music](#)

15 [Language Corpora](#)

16 [Linking, Segmentation, and Alignment](#)

17 [Simple Analytic Mechanisms](#)

18 [Feature Structures](#)

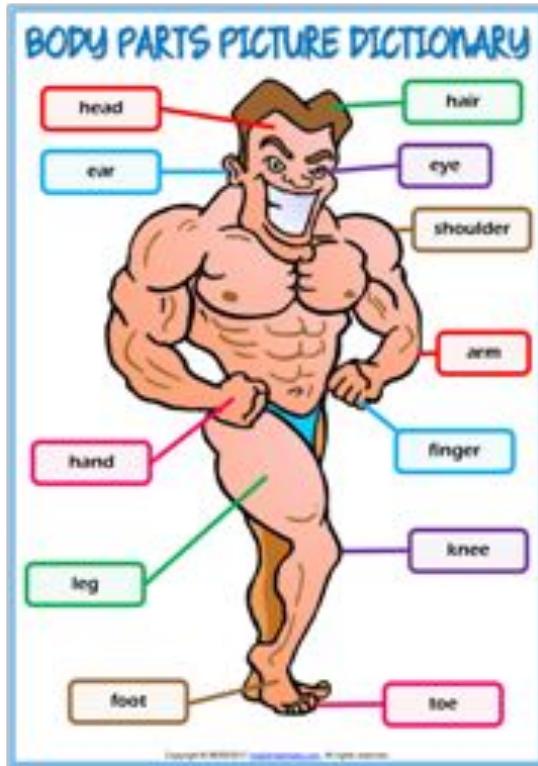
19 [Graphs, Networks, and Trees](#)

20 [Non-hierarchical Structures](#)

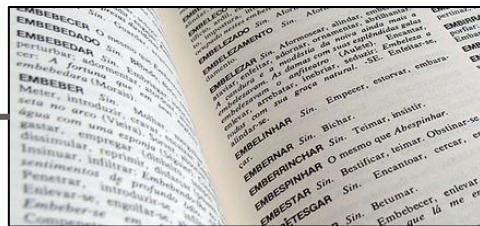
21 [Certainty, Precision, and Responsibility](#)

22 [Documentation Elements](#)

Struttura e corpo del dizionario



Struttura e corpo del dizionario



Struttura dell'entrata

Entrate non strutturate

Pronuncia e riferimenti di parole

Costituenti dell'entrata

TEI-Dict: Struttura del dizionario (1)

- ❑ Per la codifica della struttura di un dizionario (fronte, corpo e retro) si possono usare i tags TEI per tutti i testi:
 - ❑ <text> contiene un testo qualsiasi, un poema, un corpus, un dizionario, un romanzo, ecc..
 - ❑ <front> (front matter) contiene tutto ciò che si trova prima del corpo del testo (prefazioni, dediche, abstract, ecc...)
 - ❑ <body> (text body) contiene l'intero corpo del testo escluso tutto ciò che è nella parte iniziale e finale del dell'opera che contiene il testo
 - ❑ <back> (back matter) contiene tutto ciò che segue la parte principale del testo (ad esempio appendici)
 - ❑ <div> (text division) contiene una sub-divisione del fronte, del corpo o del retro di un testo
- ❑ TEI-Dict definisce dei tags che possono occorrere direttamente dentro un tag <text> o <body>:
 - ❑ <entry> contiene l'entrata del dizionario in forma strutturata.
 - ❑ <entryFree> contiene l'entrata del dizionario in forma libera
 - ❑ <superEntity> raggruppa più entrate.

TEI-Dict: Struttura del dizionario (2)

- ❑ <entry>, <entryFree> e <superEntry> hanno i seguenti attributi:
 - ❑ **att.entryLike** - classe utilizzata per distinguere differenti stili delle entrate del dizionario tramite:
@type specifica il tipo di entrata. I valori principali sono: 1] main; 2] hom (homograph) ; 3] xref (cross reference) ; 4] affix; 5] abbr (abbreviation) ; 6] supplemental; 7] foreign
 - ❑ **att.sortable** - classe per indicare il tipo ordinabile:
@sort la chiave di ordinamento per questa entrata, in un indice che la contiene

```
<index indexName="ALPHABETICAL_ORDERED_ENTRIES">
    <entry sortKey="codificare">
        <form>
            <orth>codificare</orth>
        </form>
    </entry>
    <!-- ... -->
</index>
```

TEI-Dict: Struttura del dizionario (3)

- ❑ In un dizionario possono essere contenuti, all'inizio e alla fine, materiale aggiuntivo, come ad esempio liste di nomi comuni e nomi propri, tavole grammaticali, guide all'uso del dizionario, ecc.
- ❑ Queste informazioni possono essere codificate usando anche elementi esterni al modulo TEI-Dict, ad esempio definite nel modulo “core”, capitolo 3 Elements Available in All TEI Documents
[\(http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CO.html\)](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CO.html), combinati con elementi di questo modulo.

TEI-Dict: Corpo del dizionario

```
<body>
  <div>
    <head>English-French</head>
    <entry>
      <form>
        <orth>cat</orth>
      </form>
    </entry>
    <entry>
      <form>
        <orth>dog</orth>
      </form>
    </entry>
  </div>
  <div>
    <head>French-English</head>
    <entry>
      <form>
        <orth>chat</orth>
      </form>
    </entry>
    <entry>
      <form>
        <orth>chien</orth>
      </form>
    </entry>
  </div>
</body>
```

Il corpo consiste in un insieme di entrate

Insiemi di entrate sono raggruppate dal tag `<div>`. In questo caso i raggruppamenti corrispondono alle lingue di un dizionario bilingua.

Sono possibili raggruppamenti arbitrari, ad esempio per lettera alfabetica.

TEI-Dict: Entrate omografiche (1)

- ❑ Gli omografi (dal greco omògraphos ‘dalla grafia uguale’) sono parole che hanno la stessa grafia, ma differiscono nella pronuncia.
- ❑ La diversa pronuncia può dipendere:
 - ❑ dalla diversa posizione dell’accento
àmbito (sostantivo) / **ambìto** (participio passato del verbo ambire).
viòla (sostantivo) / **viola** (3^a p. sing. dell’indicativo presente del verbo violare)
 - ❑ oppure dal diverso grado di apertura della vocale accentata
pèsca (= il frutto) / **pésca** (= azione del pescare).
pòrci (= maiali) / **pórci** (= porre noi stessi).

TEI-Dict: Entrate omografe (2)

- Esistono casi in cui anche pronuncia e grafia sono identiche.

càccia¹

VOCABOLARIO ON LINE

càccia1 s. f. [der. di **cacciare**] (pl. -ce). – 1. [...] è consentita dalla legge, contrario di c. chiusa; riserva di c., zona nella quale il diritto di **caccia** è esclusivo di un privato o di un ente (distinta dalla bandita, in cui la Leggi Tutto →

càccia²

VOCABOLARIO ON LINE

càccia2 s. m., invar. – Abbreviazione di cacciatorpediniere o di aeroplano da **caccia**: una flottiglia di c.; c. a reazione; i c.-bombardieri. ... Leggi Tutto →

Fonte: Treccani on-line.

TEI-Dict: Codifica di entrate omografe

```
<body>
  <superEntry>
    <entry type="hom" n="1">
      <form>
        <orth>caccia</orth>
      </form>
    </entry>
    <entry type="hom" n ="2">
      <form>
        <orth>caccia</orth>
      </form>
    </entry>
  </superEntry>
</body>
```

In genere gli omografi vengono considerati come entrate distinte, con un attributo “hom” che le contraddistingue.

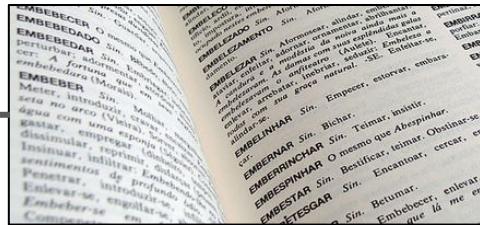
```
<body>
  <superEntry>
    <entry type="hom" n="1"/>
    <entry type="hom" n ="2"/>
    <form>
      <orth>caccia</orth>
    </form>
  </superEntry>
</body>
```

Se, ad esempio, l'informazione relativa alla forma è identica, può essere codificata una sola volta

Struttura dell'entrata



Struttura e
corpo del
dizionario



Struttura
dell'entrata

Entrate non
strutturate

Pronuncia e
riferimenti di
parole

Costituenti
dell'entrata

TEI-Dict: Le entrate di un dizionario

- ❑ Ogni entrata di un dizionario può contenere diverse informazioni, detti costituenti dell'entrata:

- ❑ una **definizione**
 - ❑ *Specifica il significato della parola*
- ❑ aspetti **grammaticali**
 - ❑ *Categoria grammaticale, genere, ecc....*
- ❑ **Sinonimi / contrari**
 - ❑ *Sinonimi e/o contrari della parola*
- ❑ **Traduzione**
 - ❑ *La traduzione in altra lingua*
- ❑ **Etimologia**
 - ❑ *Origine e storia della parola*
- ❑ **Riferimenti ad altre entrate**
- ❑ **Esempi di uso**
 - ❑ *Frasi che utilizzano la parola*

La presenza di alcune informazioni varia a seconda del tipo dizionario

Livelli di un'entrata (1)

- ❑ Gli elementi che compongono la struttura gerarchica di un'entrata sono i seguenti:



- ❑ <entry> contiene un'entrata del dizionario
 - ❑ È scoraggiato l'utilizzo di questo elemento
- ❑ <entryFree> (*unstructured entry*) contiene una singola entrata “non strutturata” dove è permesso l'utilizzo di qualsiasi elemento tei-dict violando le regole di utilizzo (annidamenti, contenimenti, ecc...)
- ❑ <hom> (homograph) raggruppa l'informazione relativa ad un omografo dentro un'entrata.
- ❑ <sense> raggruppa tutto ciò che riguarda un senso di una parola (definizione, esempi, traduzioni, ecc..).

Livelli di un'entrata (2)

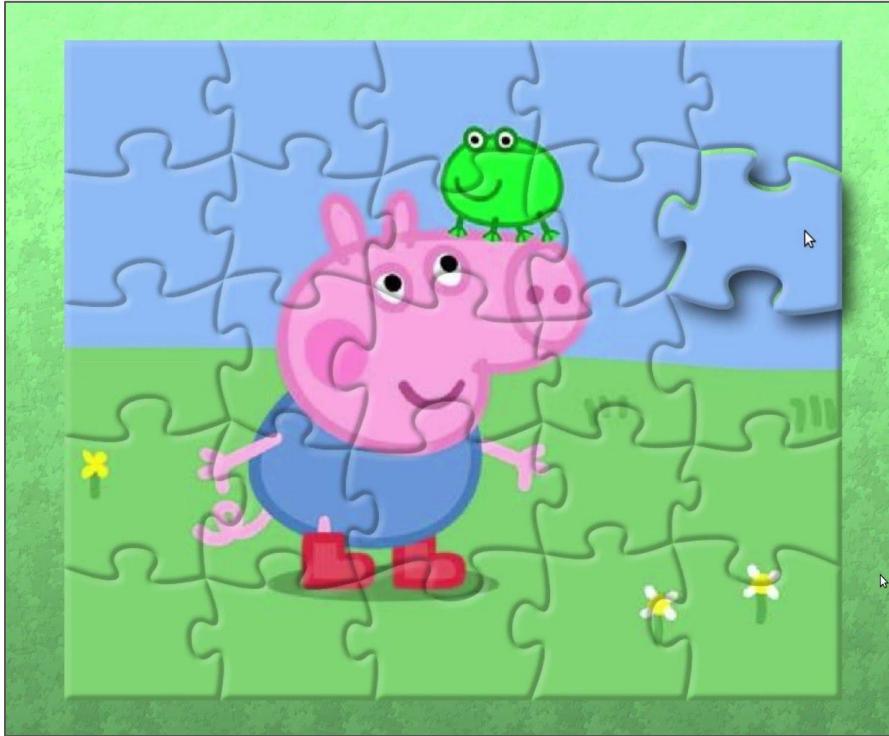
Esempio di entrata con due omografi, il secondo dei quali ha un senso strutturato in due sotto sensi

```
<entry>
  <hom n="1">
    <sense n="1">
      <!-- ... -->
    </sense>
  </hom>
  <hom n="2">
    <sense n="1">
      <sense n="a">
        <!-- ... -->
      </sense>
      <sense n="b">
        <!-- ... -->
      </sense>
    </sense>
    <sense n="2">
      <!-- ... -->
    </sense>
  </hom>
</entry>
```

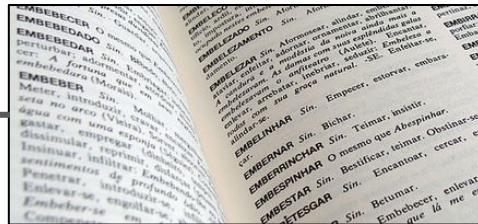
```
<superEntry>
  <entry n="1" type="hom">
    <sense n="1">
      <!-- ... -->
    </sense>
  </entry>
  <entry n="2" type="hom">
    <sense n="1">
      <sense n="a">
        <!-- ... -->
      </sense>
      <sense n="b">
        <!-- ... -->
      </sense>
    </sense>
    <sense n="2">
      <!-- ... -->
    </sense>
  </entry>
</superEntry>
```

i due omografi possono essere trattati come entry separate raggruppate come superEntry

Struttura dell'entrata



Struttura e
corpo del
dizionario



Struttura
dell'entrata

Entrate non
strutturate

Pronuncia e
riferimenti di
parole

Costituenti
dell'entrata

Livelli di un'entrata (3)

- ❑ Le regole definite nel modulo TEI-Dict specificano che:
 - ❑ L'elemento **<entry>** non può contenere altri elementi **<entry>**.
 - ❑ L'elemento **<hom>** è contenuto dentro l'elemento **<entry>**.
 - ❑ L'elemento **<sense>** può essere contenuto dentro **<entry>**, **<hom>** e **<sense>** stesso quando riferisce a sotto sensi senza limiti di profondità.
- ❑ Ognuno degli elementi suddetti può contenere i seguenti costituenti (*top-level constituents*) di un'entrata:
 - ❑ **<form>** raggruppa le informazioni ortografiche, di pronuncia di sillabazione
 - ❑ **<gramGrp>** raggruppa le informazioni morfo-sintattiche (**<pos>**, **<gen>**, **<number>**, **<case>**, ...)
 - ❑ **<def>** contiene la definizione di un'entrata
 - ❑ **<cit>** contiene una quotazione da altri documenti con riferimento bibliografico, un esempio di occorrenza in un testo, o una traduzione della forma
 - ❑ **<usg>** contiene informazioni sull'utilizzo della forma
 - ❑ **<xr>** contiene una porzione di testo che rimanda il lettore ad un'altra entrata o a un altro testo
 - ❑ **<etym>** contiene informazione etimologica
 - ❑ **<re>** entrata che appare nel corpo di un'altra entrata per qualche motivo (es: derivazione)
 - ❑ **<note>** contiene annotazioni generiche

I costituenti delle entrate: visione generale



Informazioni di forma (1)

- ❑ Le possibili informazioni relative a una specifica forma di una parola sono definite dai seguenti elementi:
 - ❑ <orth> (orthographic form) codifica l'ortografia della forma di un'entrata.
 @type rappresenta i tipi di spelling.
 - ❑ <pron> (pronunciation) codifica la pronuncia (o le pronounce) di una parola
 @extent indica se la pronuncia o l'ortografia specificate si applicano a tutta o parte della parola. I possibili valori sono: 1] full (full form) ; 2] pref (prefix) ; 3] suff (suffix) ; 4] inf (infix) ; 5] part (partial)
 - ❑ <hyph> (hyphenation) codifica le parti di una parola in termini di stem e prefissi/suffissi.
 - ❑ <syll> (syllabification) codifica la sillabazione della parola.
 - ❑ <lbt> (label) permette di inserire un testo negli elementi form, example, translation, contenente un abbreviazione, una forma contratta, un sinonimo , ecc.

Informazioni di forma (2)

- ❑ <iType> (inflectional class) indica una classe associata con un'entrata del dizionario.
@type indicates il tipo di indicatore usato per specificare la classe. I valori possibili includono: 1] abbrev; 2] verbTable

Un po' di esempi (1)

soucoupe [sukup] ...

Esempio di codifica della pronuncia (“piattino”)

```
<form>
<orth>soucoupe</orth>
<pron>sukup</pron>
</form>
```

ar.ea ...

Esempio di codifica della sillabazione

```
<form>
<orth>area</orth>
<hyph>ar|ea</hyph>
<syll>ar|e|a</syll>
</form>
```

brag ... vb brags, bragging, bragged

```
<form>
<orth>brag</orth>
</form>
<gramGrp>
<pos>vb</pos>
</gramGrp>
<form type="inflected">
<orth>brags</orth>
<orth>bragging</orth>
<orth>bragged</orth>
</form>
```

Esempio di codifica di informazioni grammaticali e forme flesse esplicite (“vantarsi”)

Un po' di esempi (2)

horrier [ORifje] (7) vt

```
<form>
<orth>horrier</orth>
<pron>ORifje</pron>
<gramGrp>
  <iType type="vble">7</iType>
<!-- ... --&gt;
&lt;/gramGrp&gt;
&lt;/form&gt;
<!-- ... --&gt;</pre>
```

Esempio di codifica di forme flesse
tramite la specifica di patterns flessionali
elencati in una tabella esterna

MTBF abbreviation for mean time between failures

```
<entry>
<form type="abbrev">
<orth>MTBF</orth>
</form>
<form type="full">
<lbl>abbreviation for</lbl>
<orth>mean time between failures</orth>
</form>
</entry>
```

Esempio di codifica di acronimo

Un po' di esempi (3)

biryani or biriani (,bɪrɪ'a:nɪ)

```
<form>
<orth>biryani</orth>
<orth>biriani</orth>
<pron notation="ipa">,bɪrɪ'a:nɪ</pron>
</form>
```

Esempio di codifica della pronuncia di due forme ortografiche differenti della stessa parola

mackle ('mæk'l) or macule ('mækju:l)

```
<form>
<orth>mackle</orth>
<pron notation="ipa">'mæk'l</pron>
</form>
<form>
<orth>macule</orth>
<pron notation="ipa">'mækju:l</pron>
</form>
```

Esempio di codifica delle pronunce di due forme ortografiche differenti della stessa parola

hospitaller or US hospitaler ('hɒspɪtələ)

```
<form>
<orth>hospitaller</orth>
<form>
<usg type="geo">US</usg>
<orth>hospitaler</orth>
</form>
<pron notation="ipa">'hɒspɪtələ</pron>
</form>
```

Esempio di codifica della stessa pronuncia a due forme ortografiche diverse

Informazioni grammaticali (1)

- ❑ Le possibili informazioni relative a una specifica forma di una parola sono definite dai seguenti elementi:
 - ❑ <gen> (gender) indica il genere associato alla forma dell'entrata del dizionario.
 - ❑ <number> indica il numero associato alla forma dell'entrata del dizionario.
 - ❑ <case> indica il caso associato alla forma dell'entrata del dizionario.
 - ❑ <per> indica la persona (1st, 2nd, 3rd, etc.) associata alla forma flessa relativa all'entrata del dizionario.
 - ❑ <tns> (tense) indica il tempo verbale associato alla forma flessa relativa all'entrata del dizionario.
 - ❑ <mood> indica l'informazione relativa al modo associato alla forma flessa relativa all'entrata del dizionario (e.g. indicative, subjunctive, imperative).
 - ❑ <pos> (part of speech) indica la categoria grammaticale dell'entrata del dizionario.
 - ❑ <subc> (subcategorization) contiene informazioni di sotto categorizzazione (transitive/intransitive, countable/non-countable, etc.)
 - ❑ <colloc> (collocate) contiene le parole che co-occorrono più frequentemente con il lemma relativo all'entrata del dizionario.

Un po' di esempi (1)

médire v.t. ind. (de)

```
<form>
  <orth>médire</orth>
</form>
<gramGrp>
  <pos>v</pos>
  <subc>t ind</subc>
  <colloc>de</colloc>
</gramGrp>
```

Esempio di codifica grammaticale
usando i singoli elementi
("calunniare")

```
<form>
  <orth>médire</orth>
</form>
<gramGrp>
  <gram type="pos">v</gram>
  <gram type="subc">t ind</gram>
  <gram type="collocPrep">de</gram>
</gramGrp>
```

Esempio di codifica grammaticale
usando solo l'elemento `<gram>`

Un po' di esempi (2)

isotope adj. et n. m.

```
<form>
  <orth>isotope</orth>
</form>
<gramGrp>
  <pos>adj</pos>
</gramGrp>
<gramGrp>
  <pos>n</pos>
  <gen>m</gen>
</gramGrp>
```

Esempio di ripetizione dell'elemento
<gramGrp>

wits (wits) pl n 1 (sometimes singular) the ability to reason and act, esp quickly

```
<entry>
  <form>
    <orth>wits</orth>
    <pron notation="ipa">wits</pron>
  </form>
  <gramGrp>
    <number>pl</number>
    <pos>n</pos>
  </gramGrp>
  <sense n="1">
    <gramGrp>
      <number>sometimes singular</number>
    </gramGrp>
    <def>the ability to reason and act, esp quickly ...</def>
  </sense>
</entry>
```

Esempio di utilizzo dell'elemento
<gram> a livello di senso

Informazioni di senso: definizione

- Il significato di ogni parola in un dizionario può essere descritto in diversi modi differenti: con sinonimi, con esempi, con traduzioni in altre lingue, con definizioni.

demigod (...) n. 1.a. a being who is part mortal, part god. b. a lesser deity. 2. a godlike person.

```
<entry>
  <form>
    <orth>demigod</orth>
    <pron> ... </pron>
  </form>
  <gramGrp>
    <pos>n</pos>
  </gramGrp>
  <sense n="1">
    <sense n="a">
      <def>a being who is part mortal, part god.</def>
    </sense>
    <sense n="b">
      <def>a lesser deity.</def>
    </sense>
  </sense>
  <sense n="2">
    <def>a godlike person.</def>
  </sense>
</entry>
```

Esempio di definizione di sensi
(semidio/divinità minore - essere divino)

Informazioni di senso: traduzione (1)

- ❑ Per la traduzione vengono introdotti i seguenti elementi:
 - ❑ <cit> (cited quotation) contains a quotation from some other document, together with a bibliographic reference to its source. In a dictionary it may contain an example text with at least one occurrence of the word form, used in the sense being described, or a translation of the headword, or an example.
 - ❑ <lbl> (label) contains a label for a form, example, translation, or other piece of information, e.g. abbreviation for, contraction of, literally, approximately, synonyms:, etc
- ❑ Information about the individual translation equivalents within a sense is grouped using <cit type="translation">. This information may include the translation text (tagged <quote>), morphological information (gen, case, etc.), usage notes (usg), translation labels (lbl), and definitions (def). When bibliographic data is provided, the quote element should be used.

Informazioni di senso: traduzione (2)

- Nei dizionari multilingua si distingue la traduzione (<quote>), dalla definizione dentro la quale essa appare.

rémoulade [Remulad] nf remoulade, rémoulade (dressing containing mustard and herbs).

```
<entry>
  <form>
    <orth>rémoulade</orth>
    <pron>Remulad</pron>
  </form>
  <gramGrp>
    <pos>n</pos>
    <gen>f</gen>
  </gramGrp>
  <cit type="translation" xml:lang="en">
    <quote>remoulade</quote>
    <quote>rémoulade</quote>
    <def>dressing containing mustard and herbs</def>
  </cit>
</entry>
```

Esempio di definizione di sensi con traduzione e definizione

Informazioni di senso: traduzione (3)

O.A.S. ... nf (abr  v de **Organisation de l'Arm  e secr  te**) OAS (illegal military organization supporting French rule of Algeria).

```
<entry>
<!-- ... -->
<cit type="translation" xml:lang="en">
  <quote>OAS</quote>
  <def>illegal military organization supporting French rule
of
  Algeria</def>
</cit>
</entry>
```

Esempio di utilizzo di `<cit>` anche in dizionari monolingua, quando viene fornita una traduzione di un prestito

Esempio di definizione di sensi con traduzione e definizione

havdalah or havdoloh Hebrew (*havda'la*; *Yiddish hav'dol  *) *n* Judaism the ceremony marking the end of the sabbath or of a festival, including the blessings over wine, candles and spices [literally: separation]

```
<entry type="foreign">
  <form>
    <orth>havdalah</orth>
    <orth>havdoloh</orth>
    <gramGrp>
      <gram type="pos">n</gram>
    </gramGrp>
  </form>
  <sense>
    <usg type="dom">Judaism</usg>
    <def>the ceremony marking the end of the sabbath or of a
festival,
      including the blessings over wine, candles and
spices</def>
  </sense>
  <cit type="translation" xml:lang="en">
    <usg type="style">literally</usg>
    <quote>separation</quote>
  </cit>
</entry>
```

Informazioni di senso: etimologia (1)

- ❑ Etymologies may contain highly structured lists of words in an order indicating their descent from each other, but often also include related words and forms outside the direct line of descent, for comparison. Not infrequently, etymologies include commentary of various sorts, and can grow into short (or long!) essays with prose-like structure. Gli elementi principali sono:
 - ❑ **<etym>** (etymology) encloses the etymological information in a dictionary entry.
 - ❑ **<lang>** (language name) contains the name of a language mentioned in etymological or other linguistic discussion.
 - ❑ **<date>** contains a date in any format.
 - ❑ **<mentioned>** marks words or phrases mentioned, not used.
 - ❑ **<gloss>** identifies a phrase or word used to provide a gloss or definition for some other word or phrase.
 - ❑ **<pron>** (pronunciation) contains the pronunciation(s) of the word.
 - ❑ **<usg>** (usage) contains usage information in a dictionary entry.
 - ❑ **<lbt>** (label) contains a label for a form, example, translation, or other piece of information, e.g. abbreviation for, contraction of, literally, approximately, synonyms:, etc.

Informazioni di senso: etimologia (2)

abismo m. (del gr. *a* priv. y *byssos*, fondo). Sima, gran profundidad. ...

```
<entry>
<form>
<orth>abismo</orth>
</form>
<etym>del <lang>gr.</lang>
<mentioned>a</mentioned> priv. y
<mentioned>byssos</mentioned>,
<gloss>fondo</gloss>
</etym>
<!-- ... -->
</entry>
```

neume \n(y)üm\ n [F, fr. ML *pneuma*, *neuma*, fr. Gk *pneuma* breath — more at **pneumatic**: any of various symbols used in the notation of Gregorian chant ... [WNC]

```
<entry>
<!-- ... -->
<etym>
<lang>F</lang> fr. <lang>ML</lang>
<mentioned>pneuma</mentioned>
<mentioned>neuma</mentioned> fr. <lang>Gk</lang>
<mentioned>pneuma</mentioned>
<gloss>breath</gloss>
<xr type="etym">more at <ptr target="#pneumatic"/>
</xr>
</etym>
<sense>
<def>any of various symbols used in the notation of
Gregorian chant
<!-- ... -->
</def>
</sense>
</entry>
<!-- ... -->
<entry xml:id="pneumatic">
<etym>
<!-- ... -->
</etym>
</entry>
```

Informazioni di senso: informazioni di riferimento ad altre entrate

- L'elemento `<re>` codifica tutte quelle entrate che appaiono nel corpo di altre entrate per motivi derivazionali, forme flesse, o composti, frasi, collocazioni e idiomì contenti l'entrata.
- In linea di principio ogni un elemento `<re>` può contenere tutto ciò che può contenere l'elemento `<entry>`, ma un `<re>` non può essere annidato dentro un altro elemento `<re>`

bevvy ('bəvɪ) *informal* n, pl -vies 1 a drink, esp an alcoholic one: *we had a few bevvies last night* 2 a session of drinking. ▷ *vb -vies, -vying, -vied (intr)* 3 to drink alcohol [probably from Old French *bevee, buvee*, drinking] > **'bevvied** adj

```
<etym>probably from <lang>Old French</lang>
<mentioned>bevee</mentioned>, <mentioned>buvee</mentioned>
<gloss>drinking</gloss>
</etym>
<re type="derived">
<form>
<orth>bevvied</orth>
</form>
<gramGrp>
<pos>adj</pos>
</gramGrp>
</re>
</entry>
```

Informazioni di senso: altre informazioni (1)

```
<cit type="example">
  <quote>the multiplex eye of the fly.</quote>
</cit>
```

Codifica di un testo che riferisce a un esempio di utilizzo dell'entrata “multiplex”

horrifier ... vt to horrify. **elle était horrifiée par la dépense** she was horrified at the expense.

```
<entry>
<! -- ... -->
<cit type="translation" xml:lang="en">
  <quote>to horrify</quote>
</cit>
<cit type="example">
  <quote>elle était horrifiée par la dépense</quote>
  <cit type="translation" xml:lang="en">
    <quote>she was horrified at the expense.</quote>
  </cit>
</cit>
</entry>
```

Traduzione ed esempio di uso

valeur ... n. f. ... 2. Vx. Vaillance, bravoure (spécial., au combat). 'La valeur n'attend pas le nombre des années' (Corneille). ...

```
<sense n="2">
  <usg type="time">Vx.</usg>
  <def>Vaillance, bravoure (spécial., au combat)</def>
  <cit type="example">
    <quote>La valeur n'attend pas le nombre des années</quote>
    <bibl>
      <author>Corneille</author>
    </bibl>
  </cit>
</sense>
```

<bibl> è utilizzato per marcare una fonte

Informazioni di senso: altre informazioni (2)

- ❑ L'elemento <usg> specifica il tipo di utilizzo di un particolare senso. I tipi <usg> si riferiscono a:
 - ❑ temporal use (archaic, obsolete, etc.)
 - ❑ register (slang, formal, taboo, ironic, facetious, etc.)
 - ❑ style (literal, figurative, etc.)
 - ❑ connotative effect (e.g. derogatory, offensive)
 - ❑ subject field (Astronomy, Philosophy, etc.)
 - ❑ national or regional use (Australian, U.S., Midland dialect, etc.)
- ❑ Tra i valori di tali tipi abbiamo (per una lista esaustiva fare riferimento alle guidelines):
 - ❑ geo (geographic area)
 - ❑ time (temporal, historical era ‘archaic’, ‘old’, etc.)
 - ❑ dom (domain)
 - ❑ reg (register)
 - ❑ style (figurative, literal, etc.)
 - ❑ lang language for foreign words, spellings pronunciations, etc.

Informazioni di senso: altre informazioni (3)

colour or US color ... [CED](#)

```
<form>
<orth>colour</orth>
<form>
<usg type="geo">US</usg>
<orth>color</orth>
</form>
</form>
```

palette [palet] nf (a) (Peinture: lit, fig) palette. (b) (Boucherie) shoulder. (c) (aube de roue) paddle; (battoir à linge) beetle; (Manutention, Constr) pallet. [CR](#)

```
<sense n="a">
<usg type="dom">Peinture</usg>
<usg type="style">lit</usg>
<usg type="style">fig</usg>
<cit type="translation" xml:lang="en">
<quote>palette</quote>
</cit>
</sense>
<sense n="b">
<usg type="dom">Boucherie</usg>
<cit type="translation" xml:lang="en">
<quote>shoulder</quote>
</cit>
</sense>
<sense n="c">
<sense>
<usg type="syn">aube de roue</usg>
<cit type="translation" xml:lang="en">
<quote>paddle</quote>
</cit>
</sense>
<sense>
<usg type="syn">battoir à linge</usg>
<cit type="translation" xml:lang="en">
<quote>beetle</quote>
</cit>
</sense>
<sense>
<usg type="dom">Manutention</usg>
<usg type="dom">Constr</usg>
<cit type="translation" xml:lang="en">
<quote>pallet</quote>
</cit>
</sense>
</sense>
```

Informazioni di senso: altre informazioni (4)

- L'elemento <xr> raggruppa le informazioni che si riferiscono a un riferimento a un'altra entrata

hostellerie Syn. de hôtellerie (sens 1).

```
<xr type="syn">
<lbl>Syn. de</lbl>
<ref>hôtellerie (sens 1)</ref>.
</xr>
```

Riferimento ad un sinonimo di un senso
specifico di un'altra entrata

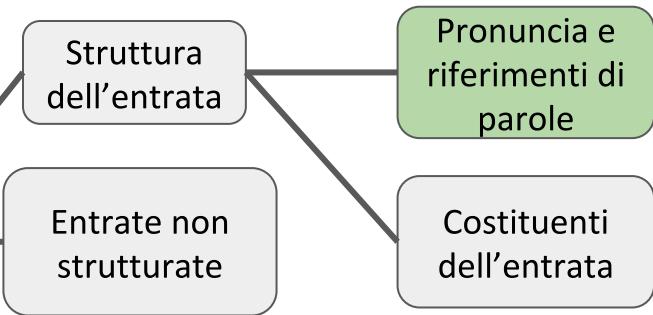
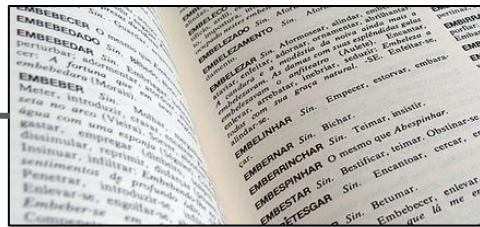
- Per maggiori approfondimenti si rimanda alle guidelines

Struttura dell'entrata

Pronunciation



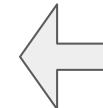
Struttura e corpo del dizionario



Riferimenti di parole (1)

- Spesso elementi di un'entrata possono contenere dei riferimenti sia alla forma che alla pronuncia dell'entrata stessa

bambino s. m. (f. -a) [da una radice onomatopeica *bamb-]. – 1. L'essere umano nell'età compresa tra la nascita e l'inizio della fanciullezza. Nell'uso com., spec. al plur. o con l'articolo indet., *bambino* può indicare genericam. maschi e femmine: *aspettare un b.*, riferito a gestanti; *allattare, svezzare, allevare, educare un b., i b.; l'igiene del b.; c'era una nidiata di*



Esempio di riferimento alla forma con la sua iniziale

- I tipi di riferimento sono:

- ~ indicates a reference to the full form of the headword
- pref~ gives a prefix to be affixed to the headword
- ~suf gives a suffix to be affixed to the headword
- A~ gives the first letter in uppercase, indicating that the headword is capitalized
- pref~suf gives a prefix and a suffix to be affixed to the headword
- a. gives the initial of the word followed by a full stop, to indicate reference to the full form of the headword
- A. refers to a capitalized form of the headword

Riferimenti di parole (2)

- ❑ Gli elementi necessari per la codifica di questi aspetti sono:
 - ❑ <oRef> (orthographic-form reference) in a dictionary example, indicates a reference to the orthographic form(s) of the headword.

@type indicates the kind of typographic modification made to the headword in the reference. Sample values include: 1] cap (capital) ; 2] noHyph (no hyphen)
 - ❑ <pRef> (pronunciation reference) in a dictionary example, indicates a reference to the pronunciation(s) of the headword.
 - ❑ l'attributo **@target** specifies the destination of the reference by supplying one or more URI References

Riferimenti di parole (3)

academy ... The Royal A~ of Arts

```
<q>The Royal <oRef type="cap"/> of Arts</q>
```

vag- or vago- comb form ... : vagus nerve < vagal > < vagotomy > [W7](#)

```
<entry>
  <form>
    <orth xml:id="di-o1">vag-</orth>
    <orth xml:id="di-o2">vago-</orth>
  </form>
  <def>vagus nerve</def>
  <cit type="example">
    <quote>
      <oRef target="#di-o1" type="noHyph"/>al</quote>
    <quote>
      <oRef target="#di-o2" type="noHyph"/>tomy</quote>
    </cit>
  </entry>
```

Riferimenti di parole (4)

take ... < Mr Burton took us for French > [NPEG](#)

```
<cit type="example">
<quote>Mr Burton <oRef type="pt">took</oRef> us for
French</quote>
</cit>
```

take ... < was quite ~n with him > [NPEG](#)

```
<cit type="example">
<quote>was quite <oRef type="pp">
n</oRef> with him</quote>
</cit>
```

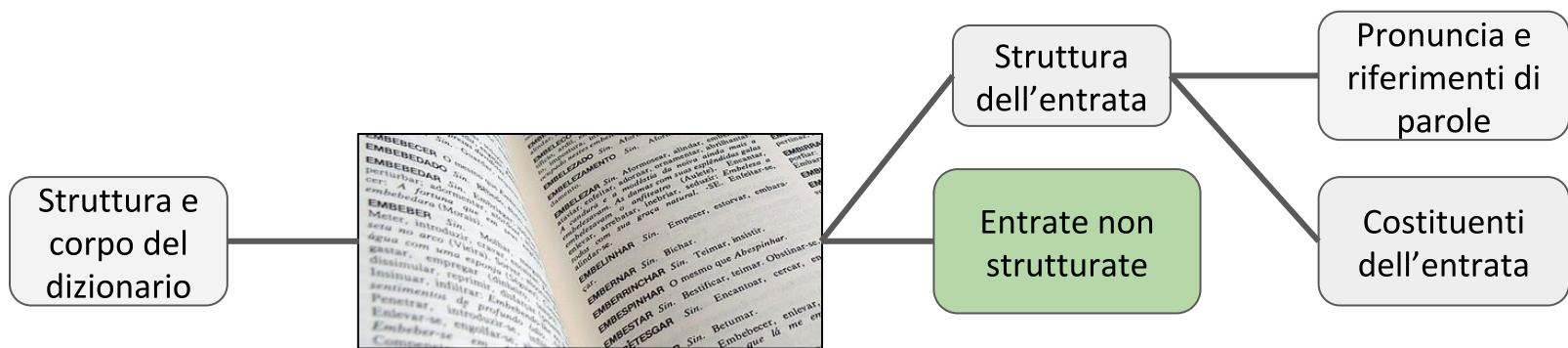
Riferimenti di forme
ortografiche

hors d'oeuvre /,aw'duhv/ (Fr O:r dœvr)/ n, pl hors d'oeuvres also hors d'oeuvre
'duhv(z) (Fr ~)/ [NPEG](#)

Riferimento di pronuncia

```
<form>
<orth>hors d'oeuvre</orth>
<pron>%aU'dUv</pron>
<form>
<usg type="lang">Fr</usg>
<pron xml:id="di-p2">OR d0vR</pron>
</form>
</form>
<form type="inflected">
<number>pl</number>
<orth>hors d'oeuvres</orth>
<orth>hors d'oeuvre</orth>
<pron extent="part">"dUv(z)</pron>
<form>
<usg type="lang">Fr</usg>
<pron>
<pRef target="#di-p2"/>
</pron>
</form>
</form>
```

Entrate senza struttura



Entrata libera

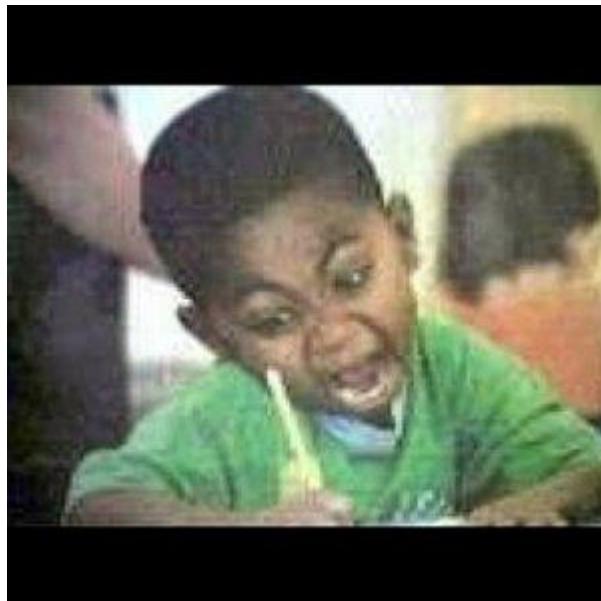
- ❑ L'elemento `<entryFree>` non pone restrizioni riguardo alla struttura dell'entrata. Ogni elemento che abbiamo visto può apparire in qualsiasi ordine e struttura. Chi codifica un'entrata con questo elemento è libero di usare qualsiasi elemento a proprio piacimento.
- ❑ L'elemento `<entryFree>` permette di violare la struttura tipica di un elemento `<entry>`. Ad esempio, la seguente codifica non usa elementi di raggruppamento e mantiene tutto il rendering del testo uguale a quello dell'entrata di esempio.

biryani or biriani (,bɪrɪ'a:nɪ) *n* any of a variety of Indian dishes ... [from Urdu]

```
<entryFree>
<orth>biryani</orth> or <orth>biriani</orth>
<pron notation="ipa">(,bɪrɪ'a:nɪ)</pron>
<def>any of a variety of Indian dishes ...</def>
<etym>[ from <lang>Urdu</lang>]</etym>
</entryFree>
```

Parte II - Fare un'entrata

In questa parte guarderemo due entrate di TEI e le codifichiamo insieme.



Lo Scheletro di un Documento TEI

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>titolo</title>
    </titleStmt>
    <publicationStmt>
      <p>informazioni sulla pubblicazione</p>
    </publicationStmt>
    <sourceDesc>
      <p>Informationi sulla sorgente</p>
    </sourceDesc>
  </fileDesc>
</teiHeader>
<text>
  <body>
    <p>Some text here.</p>
  </body>
</text>
</TEI>
```



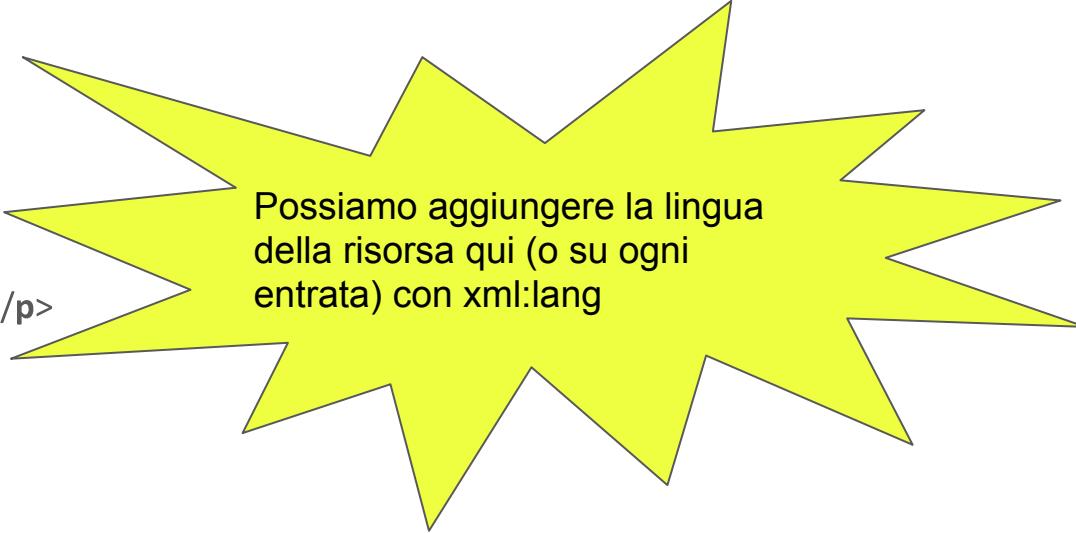
Lo Scheletro di un Documento TEI

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>titolo</title>
    </titleStmt>
    <publicationStmt>
      <p>informazioni sulla pubblicazione</p>
    </publicationStmt>
    <sourceDesc>
      <p>Informationi sulla sorgente</p>
    </sourceDesc>
  </fileDesc>
</teiHeader>
<text>
  <body>
    <p>Some text here.</p>
  </body>
</text>
</TEI>
```

Questi elementi sono obbligatori nel header

Lo Scheletro di un Documento TEI

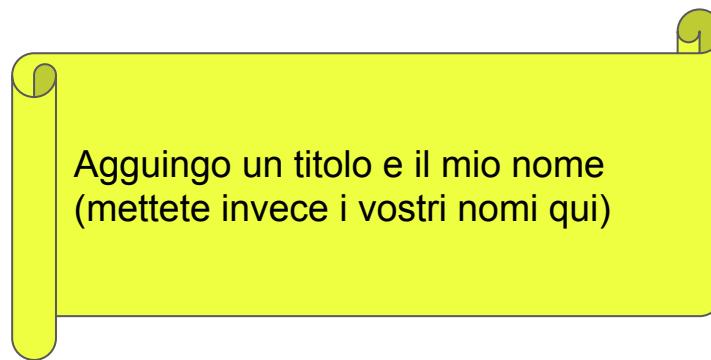
```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
<teiHeader>
  <fileDesc xml:lang = "it">
    <titleStmt>
      <title>titolo</title>
    </titleStmt>
    <publicationStmt>
      <p>informazioni sulla pubblicazione</p>
    </publicationStmt>
    <sourceDesc>
      <p>Informationi sulla sorgente</p>
    </sourceDesc>
  </fileDesc>
</teiHeader>
<text>
  <body>
    <p>Some text here.</p>
  </body>
</text>
</TEI>
```



Possiamo aggiungere la lingua della risorsa qui (o su ogni entrata) con `xml:lang`

Lo Scheletro di un Documento TEI

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
<teiHeader>
  <fileDesc xml:lang = "it">
    <titleStmt>
      <title>il mio primo dizionario</title>
      <author>Fahad Khan</author>
    </titleStmt>
    <publicationStmt>
      <p>informazioni sulla pubblicazione</p>
    </publicationStmt>
    <sourceDesc>
      <p>Informazioni sulla sorgente</p>
    </sourceDesc>
  </fileDesc>
</teiHeader>
<text>
  <body>
    <p>Some text here.</p>
  </body>
</text>
</TEI>
```



La nostra prima entrata

lessicografo s. m

IPA: /lessi'kografo/

Sillabazione: les | si | cò | gra | fo

Etimologia

dal greco tardo *λεξικογράφος*, composto di *λεξικόν* ossia "lessico" e *γράφος* ovvero "-grafo"

- **1)** (linguistica) chi si occupa di lessicografia

Francesco Sabatini è un linguista, lessicografo e presidente onorario dell'Accademia della Crusca

- **2)** colui che contribuisce alla compilazione di un dizionario

Tullio de Mauro è stato un lessicografo

Sinonimi

dizionario, vocabolario

Parole derivate

- *lessicografia*

(Adatta da <https://it.wiktionary.org/wiki/lessicografo>)



Creare l'entrata

- Creiamo la nostra entrata dentro l'elemento **<body>** e la diamo l'id numerico con l'attributo @n:

```
<text>
  <body>
    <entry n= "1">
      ...
    </entry>
  </body>
</text>
```

Informazioni sulla forma

lessicografo s. m

IPA: /lessi'kografo/

Sillabazione: les | si | cò | gra | fo

Etimologia

dal greco tardo *λεξικογράφος*, composto di *λεξικόν* ossia "lessico" e *γράφος* ovvero "-grafo"

- **1)** (linguistica) chi si occupa di lessicografia

Francesco Sabatini è un linguista, lessicografo e presidente onorario dell'Accademia della Crusca

- **2)** colui che contribuisce alla compilazione di un dizionario

Tullio de Mauro è stato un lessicografo

Sinonimi

dizionario, vocabolario

Parole derivate

- *lessicografia*

(Adatta da <https://it.wiktionary.org/wiki/lessicografo>)

Informazioni sulla forma

Lemma, headword

lessicografo s. m

IPA: /lessi'kografo/

Sillabazione: les | si | cò | gra | fo

Etimologia

dal greco tardo *λεξικογράφος*, composto di *λεξικόν* ossia "lessico" e *γράφος* ovvero "-grafo"

- 1) (linguistica) chi si occupa di lessicografia

Francesco Sabatini è un linguista, lessicografo e presidente onorario dell'Accademia della Crusca

- 2) colui che contribuisce alla compilazione di un dizionario

Tullio de Mauro è stato un lessicografo

Sinonimi

dizionario, vocabolario

Parole derivate

- *lessicografia*

(Adatta da <https://it.wiktionary.org/wiki/lessicografo>)

Informazioni sulla forma

lessicografo s. m

IPA: /lessi'kografo/

Sillabazione: les | si | cò | gra | fo

Etimologia

dal greco tardo *λεξικογράφος*, composto

Lemma, headword

La pronuncia della parola nel alfabeto IPA (International Phonetic Alphabet) - uno standard internazionale che permette di rappresentare i suoni di lingue diverse in maniera univoca

grafo"

- 1) (linguistica) chi si occupa di lessicografia

Francesco Sabatini è un linguista, lessicografo e presidente onorario dell'Accademia della Crusca

- 2) colui che contribuisce alla compilazione di un dizionario

Tullio de Mauro è stato un lessicografo

Sinonimi

dizionario, vocabolario

Parole derivate

- *lessicografia*

(Adatta da <https://it.wiktionary.org/wiki/lessicografo>)

Informazioni sulla forma

Lemma, headword

lessicografo s. m

IPA: /lessi'kografo/

Sillabazione: les | si | cò | gra | fo

Etimologia

dal greco tardo λεξικόν

La pronuncia della parola nel alfabeto IPA (International Phonetic Alphabet) - uno standard internazionale che permette di rappresentare i suoni di lingue diverse

- 1) (linguistica) c

Francesca

Crusca

Una sillabazione - che divide una parola in sillabe

aniera univoca

- 2) colui che contribuisce alla compilazione di un dizionario

Tullio de Mauro è stato un lessicografo

Sinonimi

dizionario, vocabolario

Parole derivate

- lessicografia

(Adatta da <https://it.wiktionary.org/wiki/lessicografo>)

Informazioni sulla forma

Lemma, headword

lessicografo s. m

IPA: /lessi'kografo/

Sillabazione: les | si | cò | gra | fo

Etimologia

dal greco tardo λεξικόν

La pronuncia della parola nel alfabeto IPA (International Phonetic Alphabet) - uno standard internazionale che permette di rappresentare i suoni di lingue diverse

- 1) (linguistica) c

Francesca

Crusca

Una sillabazione - che divide una parola in sillabe

aniera univoca

- 2) colui che contribuisce alla compilazione di un dizionario

Tullio de Mauro è stato un lessicografo

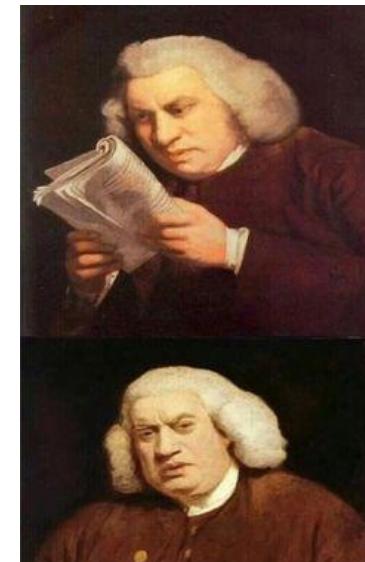
Sinonimi

dizionario, vocabolario

Parole derivate

- lessicografia

(Adatta da <https://it.wiktionary.org/wiki/lessicografo>)



Informazioni sulla forma

```
<entry n= "1">
  <form type = "lemma">
    <orth>lessicografo</orth>
    <pron notation = "ipa">lessi|kɔ|gra|fɔ</pron>
    <syll>les|si|cò|gra|fo</syll>
  </form>
</entry>
```



<form>: raggruppa tutte le informazioni sulle form scritte e orali di un lemma

Informazioni sulla forma

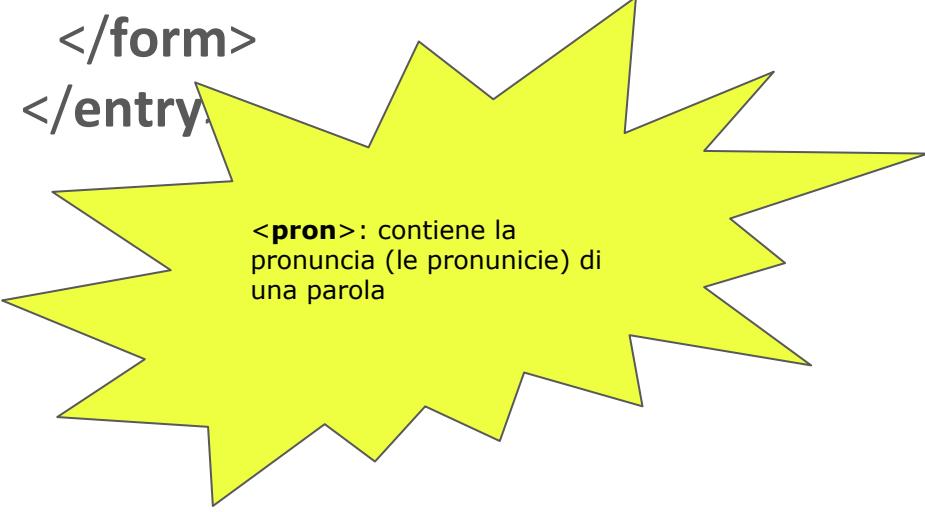
```
<entry n= "1">  
  <form type = "lemma">  
    <orth>lessicografo</orth>  
    <pron notation = "ipa">lessi'kɒfə</pron>  
    <syll>les|si|cò|gra|fo</syll>  
  </form>  
</entry>
```



<orth>: fornisce la forma ortografica di un lemma.

Informazioni sulla forma

```
<entry n= "1">
  <form type = "lemma">
    <orth>lessicografo</orth>
    <pron notation = "ipa">lessi'kografo</pron>
    <syll>les | si | cò | gra | fo</syll>
  </form>
</entry>
```



<pron>: contiene la pronuncia (le pronuncie) di una parola

Informazioni sulla forma

```
<entry n= "1">
  <form type = "lemma">
    <orth>lessicografo</orth>
    <pron notation = "ipa">lessi'kografo</pron>
    <syll>les|si|cò|gra|fo</syll>
  </form>
</entry>
```

<pron>: contiene la pronuncia (le pronuncie) di una parola

Possiamo specificare ipa come il tipo di notazione con l'attributo @notation

Informazioni Grammaticale

lessicografo *s. m.*

IPA: /lessi'kografo/

Sillabazione: les | si | cò | gra | fo

Etimologia

dal greco tardo λεξικογράφος, composto di λεξικόν ossia "lessico" e γράφος ovvero "-grafo"

- **1)** (linguistica) chi si occupa di lessicografia

Francesco Sabatini è un linguista, lessicografo e presidente onorario dell'Accademia della Crusca

- **2)** colui che contribuisce alla compilazione di un dizionario

Tullio de Mauro è stato un lessicografo

Sinonimi

dizionario, vocabolario

Parole derivate

- *lessicografia*

(Adatta da <https://it.wiktionary.org/wiki/lessicografo>)

Informazioni Grammaticale

lessicografo *s. m.*

IPA: /lessi'kografo/

Sillabazione: les | si | cò | g

Etimologia

dal greco tardo *λεξικού*

Informazioni sulla
categoria
grammaticale e il
genere della parola

sto di *λεξικό*

e *γράφος* ovvero "-grafo"

- **1)** (linguistica) chi si occupa di lessicografia

Francesco Sabatini è un linguista, lessicografo e presidente onorario dell'Accademia della Crusca

- **2)** colui che contribuisce alla compilazione di un dizionario

Tullio de Mauro è stato un lessicografo

Sinonimi

dizionario, vocabolario

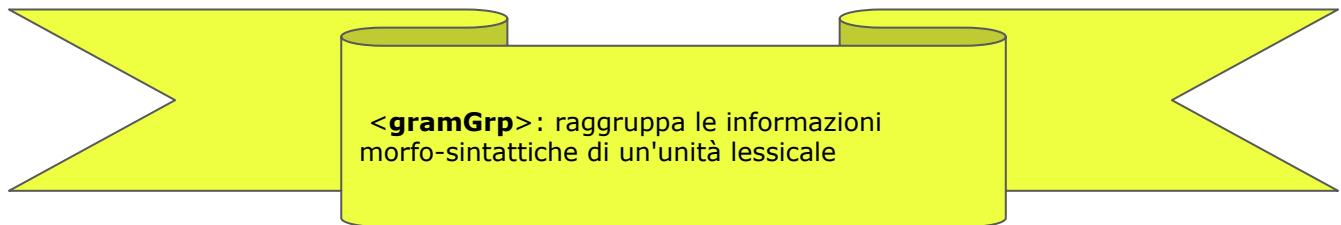
Parole derivate

- *lessicografia*

(Adatta da <https://it.wiktionary.org/wiki/lessicografo>)

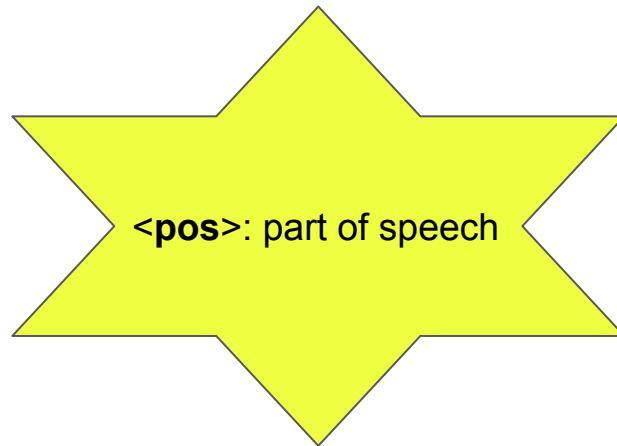
Informazioni Grammaticale

```
<entry n="1">
  <form>
    ...
  </form>
  <gramGrp>
    <pos>s.</pos>
    <gen>m.</gen>
  </gramGrp>
</entry>
```



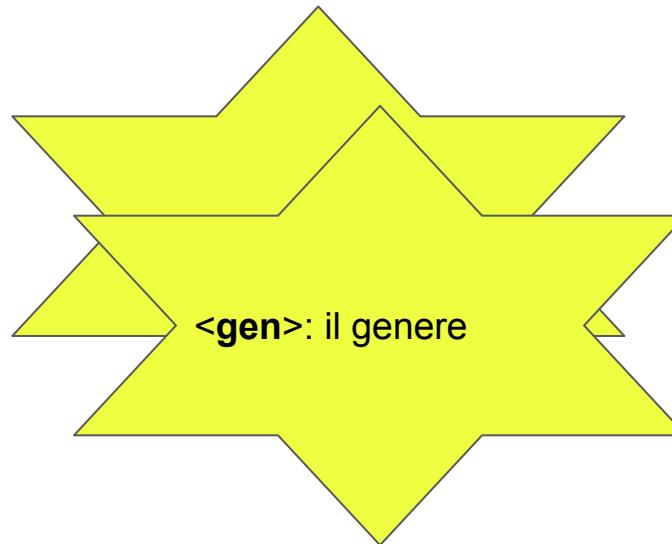
Informazioni Grammaticale

```
<entry n="1">
  <form>
    ...
  </form>
  <gramGrp>
    <pos>s.</pos>
    <gen>m.</gen>
  </gramGrp>
</entry>
```



Informazioni Grammaticale

```
<entry n="1">
  <form>
    ...
  </form>
  <gramGrp>
    <pos>s.</pos>
    <gen>m.</gen>
  </gramGrp>
</entry>
```



L'etimologia

lessicografo s. m

IPA: /lessi'kografo/

Sillabazione: les | si | cò | gra | fo

Etimologia

dal greco tardo λεξικογράφος, composto di λεξικόν ossia "lessico" e γράφος ovvero "-grafo"

- **1)** (linguistica) chi si occupa di lessicografia

Francesco Sabatini è un linguista, lessicografo e presidente onorario dell'Accademia della Crusca

- **2)** colui che contribuisce alla compilazione di un dizionario

Tullio de Mauro è stato un lessicografo

Sinonimi

dizionario, vocabolario

Parole derivate

- *lessicografia*

(Adatta da <https://it.wiktionary.org/wiki/lessicografo>)

L'etimologia - un primo passo

```
<entry n= "1">
  <form>
    ...
  </form>
  <gramGrp>
    ...
  </gramGrp>
  <etym>dal greco tardo λεξικογράφος, composto di λεξικόν
  ossia "lessico" e γράφος ovvero "-grafo"</etym>
</entry>
```

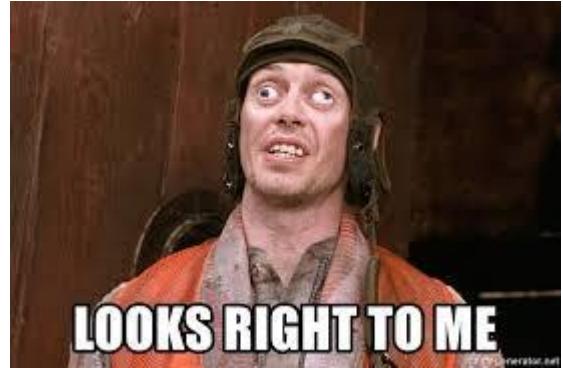
L'etimologia - un primo passo

```
<entry n= "1">
  <form>
    ...
  </form>
  <gramGrp>
    ...
  </gramGrp>
  <etym>dal greco tardo λεξικογράφος, composto di λεξικόν
  ossia "lessico" e γράφος ovvero "-grafo"</etym>
</entry>
```



L'etimologia - un primo passo

```
<entry n= "1">
  <form>
    ...
  </form>
  <gramGrp>
    ...
  </gramGrp>
  <etym>dal greco tardo λεξικογράφος, composto di λεξικόν
  ossia "lessico" e γράφος ovvero "-grafo" </etym>
</entry>
```



Aggiungiamo più struttura

<etym>dal <lang>greco tardo</lang>
<mentioned>λεξικογράφος</mentioned>, composto di
<mentioned>λεξικόν</mentioned> ossia
<gloss>"lessico"</gloss> e
<mentioned>γράφος</mentioned> ovvero
<gloss>"-grafo"</gloss></etym>

Agguingiamo più struttura

<etym>dal <lang>greco tardo</lang>
<mentioned>λεξικογράφος</mentioned>, composto di
<mentioned>λεξικόν</mentioned> ossia
<gloss>"lessico"</gloss> e
<mentioned>γράφος</mentioned> ovvero
<gloss>"-grafo"</gloss></etym>



Aggiungiamo più struttura

<etym>dal <lang>greco tardo</lang>
<mentioned>λεξικογράφος</mentioned>, composto di
<mentioned>λεξικόν</mentioned> ossia
<gloss>"lessico"</gloss> e
<mentioned>γράφος</mentioned> ovvero
<gloss>"-grafo"</gloss></etym>

Aggiungiamo più struttura

<etym>dal <lang>greco tardo</lang>
<mentioned>λεξικογράφος</mentioned>, composto di
<mentioned>λεξικόν</mentioned> ossia
 <gloss>"lessico"</gloss> e
<mentioned>γράφος</mentioned> ovvero
 <gloss>"-grafo"</gloss></etym>

<lang>: il nome di una lingua menzionata nell'etimologia o in altra discussione linguistica.

Aggiungiamo più struttura

<etym>dal <lang>greco tardo</lang>
<mentioned>λεξικογράφος</mentioned>, composto di
<mentioned>λεξικόν</mentioned> ossia
<gloss>"lessico"</gloss> e
<mentioned>γράφος</mentioned> ovvero
<gloss>"-grafo"</gloss></etym>

Aggiungiamo più struttura

<etym>dal <lang>greco tardo</lang>

<mentioned>λεξικογράφος</mentioned>, composto di

<mentioned>λεξικόν</mentioned> ossia

<gloss>"lessico"</gloss> e

<mentioned>γράφει</mentioned> ovvero

<gloss>codifica</gloss></etym>

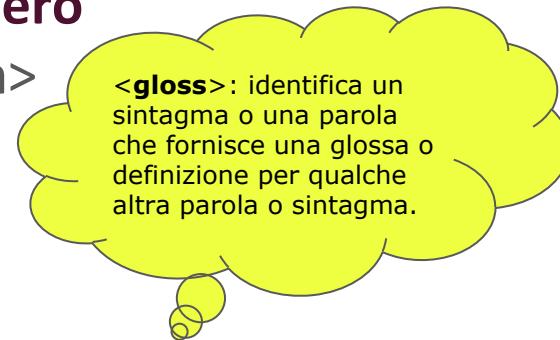
<mentioned>: codifica
parole o sintagmi citati o
riportati, non usati.

Aggiungiamo più struttura

<etym>dal <lang>greco tardo</lang>
<mentioned>λεξικογράφος</mentioned>, composto di
<mentioned>λεξικόν</mentioned> ossia
 <gloss>"lessico"</gloss> e
<mentioned>γράφος</mentioned> ovvero
 <gloss>"-grafo"</gloss></etym>

Aggiungiamo più struttura

<etym>dal <lang>greco tardo</lang>
<mentioned>λεξικογράφος</mentioned>, composto di
<mentioned>λεξικόν</mentioned> ossia
 <gloss>"lessico"</gloss> e
<mentioned>γράφος</mentioned> ovvero
 <gloss>"-grafo"</gloss></etym>



Codifichiamo i due sensi...

```
<entry n="1">
  <form>
    ...
  </form>
  <etym>...
  </etym>
  <sense n="s1">
    ...
  </sense>
  <sense n = "s2">
    ...
  </sense>
</entry>
```

Il primo senso

lessicografo s. m

IPA: /lessi'kografo/

Sillabazione: les | si | cò | gra | fo

Etimologia

dal greco tardo *λεξικογράφος*, composto di *λεξικόν* ossia "lessico" e *γράφος* ovvero "-grafo"

- **1)** (linguistica) chi si occupa di lessicografia

Francesco Sabatini è un linguista, lessicografo e presidente onorario dell'Accademia della Crusca

- **2)** colui che contribuisce alla compilazione di un dizionario

Tullio de Mauro è stato un lessicografo

Sinonimi

dizionario, vocabolario

Parole derivate

- *lessicografia*

(Adatta da <https://it.wiktionary.org/wiki/lessicografo>)

Il primo senso

lessicografo s. m

IPA: /lessi'kografo/

Sillabazione: les | si | cò | gra | fo

Etimologia

dal greco tardo *λεξικογράφος*, composto di *λεξικόν* ossia "lessico" e *γράφος* ovvero "-grafo"

- 1) (linguistica) chi si occupa di lessicografia

Francesco Sabatini è un linguista, lessicografo e presidente onorario dell'Accademia della Crusca

- 2) colui che contribuisce alla compilazione di un dizionario

Tullio de Mauro è stato un lessicografo

Sinonimi

dizionario, vocabolario

DOBBIAMO CODIFICARE:

- IL DOMINIO
"LINGUISTICA")"
- LA DEFINIZIONE
- LA CITAZIONE

Parole derivate

- lessicografia

(Adatta da <https://it.wiktionary.org/w/index.php?title=Lessicografo&oldid=10000000>)

Il primo senso

```
<sense n="s1">
  <usg type = "dom">(linguistica)</usg>
  <def>chi si occupa di lessicografia</def>
  <cit type= "example">
    <quote>Francesco Sabatini è un linguista, lessicografo e
presidente onorario dell'Accademia della Crusca</quote>
  </cit>
</sense>
```

Il primo senso

```
<sense n="s1">  
  <usg type = "dom">(linguistica)</usg>  
  <def>chi si occupa di lessicografia</def>  
  <cit type= "example">  
    <quote>Francesco Sabatini è un linguista.  
    presidente onorario dell'Accademia della Crusca</quote>  
  </cit>  
</sense>
```

<usg> (uso) in una voce di dizionario, contiene informazioni sull'uso.

presidente onorario dell'Accademia della Crusca

Il primo senso

```
<sense n="s1">  
  <usg type = "dom">(linguistica)</usg>  
  <def>chi si occupa di lessicografia</def>  
  <cit type= "example">  
    <quote>Francesco Sabatini è  
presidente onorario dell'Accademia  
  </cit>  
</sense>
```

<usg> (uso) in una voce di dizionario, contiene informazioni sull'uso.

L'assegnazione di "dom" a @type ci permette di specificare che stiamo codificando il dominio

Il primo senso

```
<sense n="s1">  
  <usage> "don  
  <definition>  
    <entity>
```

preside

CON <USO> SI POSSONO SPECIFICARE

- GEOGRAPHIC AREA
- TEMPORAL, HISTORICAL ERA ('ARCHAIC', 'OLD', ETC.)
- DOMAIN
- REGISTER
- STYLE (FIGURATIVE, LITERAL, ETC.)
- PREFERENCE LEVEL ('CHIEFLY', 'USUALLY', ETC.)
- ACCEPTABILITY
- LANGUAGE FOR FOREIGN WORDS, SPELLINGS PRONUNCIATIONS, ETC.
- GRAMMATICAL USAGE

se

ce di
sull'uso.

Il primo senso

```
<sense n="s1">
    <usg type = "dom">(linguistica)</usg>
    <def>chi si occupa di lessicografia</def>
    <cit type= "example">
        <quote>Francesco Sabatini è un linguista, lessicografo e
presidente onorario dell'Accademia della Crusca</quote>
    </cit>
</sense>
```

Il primo senso

<sense n="s1">

<usg type = "dom">(linguistica)</usg>

<def>chi si occupa di lessicografia</def>

<cit type= "example">

<quote>**Francesco Sabatini è**

<def> (definizione) contiene una definizione in una voce di dizionario.

e

presidente onorario dell'Accademia della Crusca</quote>

</cit>

</sense>

Il primo senso

```
<sense n="s1">
    <usg type = "dom">(linguistica)</usg>
    <def>chi si occupa di lessicografia</def>
    <cit type= "example">
        <quote>Francesco Sabatini è un linguista, lessicografo e
presidente onorario dell'Accademia della Crusca</quote>
    </cit>
</sense>
```

Il primo senso

```
<sense n="s1">
```

```
    <usg type = "dom">(linguistica)</usg>
```

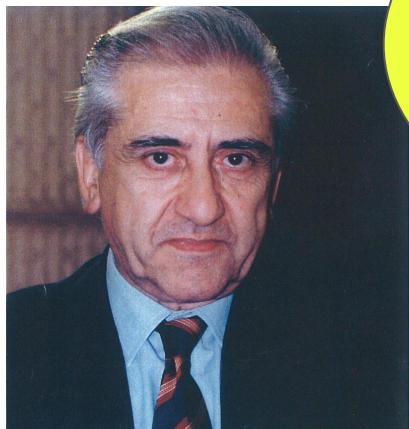
```
    <def>chi si occupa di lessicografia</def>
```

```
    <cit type= "example">
```

```
        <quote>Francesco Sabatini è un linguista, lessicografo e  
presidente onorario dell'Accademia della Crusca</quote>
```

```
    </cit>
```

```
</sense>
```



La citazione va dentro un elemento
<quote> che a sua volta va dentro
l'elemento <cit> con @type uguale
a "example"

Il secondo senso

lessicografo s. m

IPA: /lessi'kografo/

Sillabazione: les | si | cò | gra | fo

Etimologia

dal greco tardo *λεξικογράφος*, composto di *λεξικόν* ossia "lessico" e *γράφος* ovvero "-grafo"

- **1)** (linguistica) chi si occupa di lessicografia

Francesco Sabatini è un linguista, lessicografo e presidente onorario dell'Accademia della Crusca

- **2)** colui che contribuisce alla compilazione di un dizionario

Tullio de Mauro è stato un lessicografo

Sinonimi

dizionario, vocabolario

Parole derivate

- *lessicografia*

(Adatta da <https://it.wiktionary.org/wiki/lessicografo>)

Il secondo senso

lessicografo s. m

IPA: /lessi'kografo/

Sillabazione: les | si | cò | gra | fo

Etimologia

dal greco tardo *λεξικογράφος*, composto di *λεξικόν* ossia "lessico" e *γράφος* ovvero "-grafo"

- **1)** (linguistica) chi si occupa di lessicografia

Francesco Sabatini è un linguista, lessicografo e presidente onorario dell'Accademia della Crusca

- **2)** colui che contribuisce alla compilazione di un dizionario

Tullio de Mauro è stato un lessicografo

Sinonimi

dizionario, vocabolario

Parole derivate

- *lessicografia*

(Adatta da <https://it.wiktionary.org/wiki/lessicografo>)

DOBBIAMO CODIFICARE:

- **LA DEFINIZIONE**
- **LA CITAZIONE**
- **I SINONIMI**

Il secondo senso

```
<sense n = "s2">
  <def>colui che contribuisce alla compilazione di un dizionario</def>
  <cit type= "example">
    <quote>Tullio de Mauro è stato un lessicografo</quote>
  </cit>
  <xr type= "syn">
    <ref>dizionario</ref>
  </xr>
  <xr type= "syn">
    <ref>vocabolario</ref>
  </xr>
</sense>
```

Il secondo senso

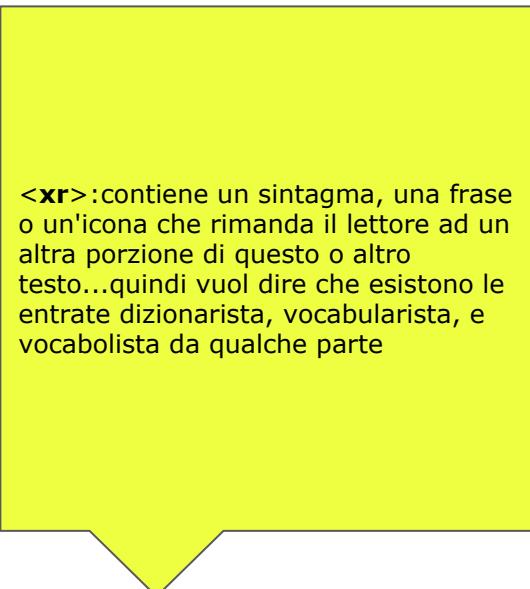
```
<sense n = "s2">
  <def>colui che contribuisce alla compilazione di un dizionario</def>
  <cit type= "example">
    <quote>Tullio de Mauro è stato un lessicografo</quote>
  </cit>
  <xr type= "syn">
    <ref>dizionario</ref>
  </xr>
  <xr type= "syn">
    <ref>vocabolario</ref>
  </xr>
</sense>
```

Usiamo gli elementi `<xr>` e `<ref>` per codificare i sinonimi di *lessicografo*



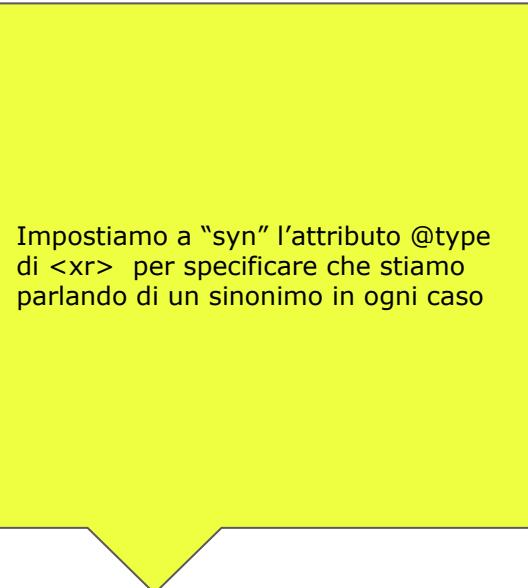
Il secondo senso

```
<sense n = "s2">  
  <def>colui che contribuisce alla compilazione</def>  
  <cit type= "example">  
    <quote>Tullio de Mauro è stato un lessicografo, un  
  </cit>  
  <xr type= "syn">  
    <ref>dizionario</ref>  
  </xr>  
  <xr type= "syn">  
    <ref>vocabolario</ref>  
  </xr>  
</sense>
```



Il secondo senso

```
<sense n = "s2">
  <def>colui che contribuisce alla compilazione</def>
  <cit type= "example">
    <quote>Tullio de Mauro è stato un lessicografo e un
  </cit>
  <xr type= "syn">
    <ref>dizionario</ref>
  </xr>
  <xr type= "syn">
    <ref>vocabolario</ref>
  </xr>
</sense>
```



Il secondo senso

```
<sense n = "s2">  
  <def>colui che contribuisce alla compilazione</def>  
  <cit type= "example">  
    <quote>Tullio de Mauro è stato un lessicografo e  
    </cit>  
  <xr type= "syn">  
    <ref>dizionario</ref>  
  </xr>  
  <xr type= "syn">  
    <ref>vocabolario</ref>  
  </xr>  
</sense>
```



COMPITO!!!!!!



Codificate le parole “dizionarioista”, “vocabolarioista” e “vocabolista”: prendendo le informazioni per ogni entrata dalla sua pagina wikionario:

- <https://it.wiktionary.org/wiki/dizionarioista>
- <https://it.wiktionary.org/wiki/vocabolarioista>
- <https://it.wiktionary.org/wiki/vocabolista>

(Codificate solo i tipi di informazione che non abbiamo considerato in questa lezione e non “Note” o “Traduzione”
-- a meno che non vogliate)

UN ALTRO COMPITO!!!



Completate la codifica dell'entrata “lessicografo”, cioè la parte che riguarda **parole derivate** usando le linee guida di TEI come riferimento. In particolare dovete guardare ciò che è scritto sull'elemento <re>.

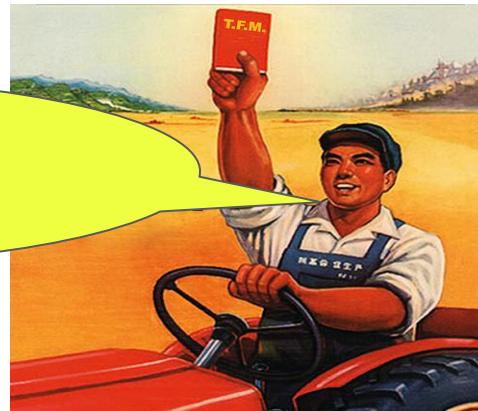


UN ALTRO COMPITO!!!



Completate la codifica dell'entrata "lessicografo", cioè la parte che riguarda **parole derivate** usando le linee guida di TEI come riferimento. In particolare dovete guardare ciò che è scritto sull'elemento <re>.

RTF TEI Guidelines:
<<http://it.guidelines.tei-c.org/html/DI.html>>



Tocca a voi

Ora tocca a voi fare un esempio...



...fregare

fregare vtr.

IPA: /fre' gare/

Sillabazione: fre | gà | re

Etimologia

dal latino *frīcare*

1) fare dei freghi

- *Avendo preso male le misure, fregò la portiera dell'automobile contro la colonna dell'ingresso del garage*

2) (popolare) tirare un bidone

- *Ti ha fregato!*

- *Il baro lo fregò di brutto calando un poker d'assi alla fine dell'ultima mano*

3) rubare

- *Mi sono distratta un attimo e mi hanno fregato la borsa da sotto il naso!*

4) interessare, importare

- *Non me ne frega niente*

5) (per estensione) imbrogliare o ingannare, quindi in modo ipocrita

- *oh, tu non mi freghi!*

Parole derivate

fregarsi, fregatura, sfregare



Grazie per la vostra attenzione!!!

Se avete delle domande (soprattutto sulla validazione dei documenti TEI)...

Grazie per la vostra attenzione!!!

.....chiedete al Prof. Del Grosso !!!

