http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html



* [Home](http://www.tei-c.org/)
* [Guidelines](http://www.tei-c.org/guidelines/)
	+ [P5 Guidelines](http://www.tei-c.org/guidelines/p5/)
	+ [– Table of Contents](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html)
	+ [Customization](http://www.tei-c.org/guidelines/customization/)
	+ [Licensing and Citation](http://www.tei-c.org/guidelines/licensing-and-citation/)
	+ [TEI @ GitHub](https://github.com/TEIC/TEI)
	+ [Learn the TEI](http://www.tei-c.org/support/learn/)
* [Activities](http://www.tei-c.org/activities/)
	+ [Annual Meetings](http://members.tei-c.org/Events/meetings)
	+ [Technical Council](http://www.tei-c.org/activities/council/)
	+ [Board](http://www.tei-c.org/board/)
	+ [Workgroups](http://www.tei-c.org/activities/workgroups/)
	+ [SIGs](http://www.tei-c.org/activities/sig/)
	+ [Projects Using the TEI](http://www.tei-c.org/activities/projects/)
	+ [jTEI: Journal of the TEI](http://journal.tei-c.org/)
* [Tools](http://www.tei-c.org/tools/)
	+ [OxGarage](http://www.tei-c.org/oxgarage/)
	+ [Roma](http://www.tei-c.org/Roma/)
	+ [TAPAS](http://tapasproject.org/)
	+ [Stylesheets](http://www.tei-c.org/tools/stylesheets/)
	+ [Internationalization](http://www.tei-c.org/tools/i18n/)
	+ [Tools SIG](http://www.tei-c.org/activities/sig/tools/)
	+ [Tei Wiki Tools Page](http://wiki.tei-c.org/index.php/Category%3ATools)
* [Membership](http://members.tei-c.org/)
	+ [How to Join](http://members.tei-c.org/Join)
	+ [Benefits](http://members.tei-c.org/Join/benefits)
	+ [Meetings](http://members.tei-c.org/Events/meetings)
	+ [List of members](http://members.tei-c.org/Directory)
* [Support](http://www.tei-c.org/support/)
	+ [TEI-L Mailing List](http://www.tei-c.org/support/#tei-l)
	+ [TEI Wiki](http://wiki.tei-c.org/)
	+ [Learn the TEI](http://www.tei-c.org/support/learn/)
	+ [Bug Reports and Feature Requests](https://github.com/TEIC/TEI/issues)
* [About](http://www.tei-c.org/about/)
	+ [Frequently Asked Questions](http://www.tei-c.org/about/frequently-asked-questions/)
	+ [Goals and Mission](http://www.tei-c.org/about/mission/)
	+ [Charter](http://www.tei-c.org/about/charter/)
	+ [TEI Bylaws](http://www.tei-c.org/about/bylaws/)
	+ [Organization](http://www.tei-c.org/about/organization/)
	+ [– Board of Directors](http://www.tei-c.org/about/board-of-directors/)
	+ [– Technical Council](http://www.tei-c.org/activities/council/)
	+ [History](http://www.tei-c.org/about/history/)
	+ [Archive](http://www.tei-c.org/Vault/)
	+ [Badges](http://www.tei-c.org/about/badges/)
	+ [Contact the TEI](http://www.tei-c.org/about/contact/)
	+ [About the TEI Website](http://www.tei-c.org/about/website/)
* [News](http://www.tei-c.org/news/)

Top of Form

  

Bottom of Form

**P5: Guidelines for Electronic Text Encoding and Interchange**

Version 3.4.0. Last updated on 23rd July 2018, revision 1fa0b54

**9 Dictionaries**

**Table of contents**

* [9.1 Dictionary Body and Overall Structure](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIBO)
* [9.2 The Structure of Dictionary Entries](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIEN)
* [9.3 Top-level Constituents of Entries](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITP)
* [9.4 Headword and Pronunciation References](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIHW)
* [9.5 Typographic and Lexical Information in Dictionary Data](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIMV)
* [9.6 Unstructured Entries](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIFR)
* [9.7 The Dictionary Module](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#index-body.1_div.9_div.7)
* « [8 Transcriptions of Speech](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/TS.html)
* » [10 Manuscript Description](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/MS.html)
* [Home](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html)

This chapter defines a module for encoding lexical resources of all kinds, in particular human-oriented monolingual and multilingual dictionaries, glossaries, and similar documents. The elements described here may also be useful in the encoding of computational lexica and similar resources intended for use by language-processing software; they may also be used to provide a rich encoding for wordlists, lexica, glossaries, etc. included within other documents. Dictionaries are most familiar in their printed form; however, increasing numbers of dictionaries exist also in electronic forms which are independent of any particular printed form, but from which various displays can be produced.

Both typographically and structurally, print dictionaries are extremely complex. Such lexical resources are moreover of interest to many communities with different and sometimes conflicting goals. As a result, many general problems of text encoding are particularly pronounced here, and more compromises and alternatives within the encoding scheme may be required in the future.36 Two problems are particularly prominent.

First, because the structure of dictionary entries varies widely both among and within dictionaries, the simplest way for an encoding scheme to accommodate the entire range of structures actually encountered is to allow virtually any element to appear virtually anywhere in a dictionary entry. It is clear, however, that strong and consistent structural principles do govern the vast majority of conventional dictionaries, as well as many or most entries even in more ‘exotic’ dictionaries; encoding guidelines should include these structural principles. We therefore define two distinct elements for dictionary entries, one ([entry](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-entry.html)) which captures the regularities of many conventional dictionary entries, and a second ([entryFree](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-entryFree.html)) which uses the same elements, but allows them to combine much more freely. It is however recommended that [entry](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-entry.html) be used in preference to [entryFree](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-entryFree.html) wherever possible. These elements and their contents are described in sections [9.2 The Structure of Dictionary Entries](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIEN), [9.6 Unstructured Entries](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIFR), and [9.4 Headword and Pronunciation References](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIHW).

Second, since so much of the information in printed dictionaries is implicit or highly compressed, their encoding requires clear thought about whether it is to capture the precise typographic form of the source text or the underlying structure of the information it presents. Since both of these views of the dictionary may be of interest, it proves necessary to develop methods of recording both, and of recording the interrelationship between them as well. Users interested mainly in the printed format of the dictionary will require an encoding to be faithful to an original printed version. However, other users will be interested primarily in capturing the lexical information in a dictionary in a form suitable for further processing, which may demand the expansion or rearrangement of the information contained in the printed form. Further, some users wish to encode *both* of these views of the data, and retain the links between related elements of the two encodings. Problems of recording these two different views of dictionary data are discussed in section [9.5 Typographic and Lexical Information in Dictionary Data](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIMV), together with mechanisms for retaining both views when this is desired.

To deal with this complexity, and in particular to account for the wide variety of linguistic contexts within which a dictionary may be designed, it can be necessary to customize or change the schema by providing more restriction or possibly alternate content models for the elements defined in this chapter. Section [9.3.2 Grammatical Information](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPGR) illustrates this with the provision of a closed set of values for grammatical descriptors.

This chapter contains a large number of examples taken from existing print dictionaries; in each case, the original source is identified. In presenting such examples, we have tried to retain the original typographic appearance of the example as well as presenting a suggested encoding for it. Where this has not been possible (for example in the display of pronunciation) we have adopted the transliteration found in the electronic edition of the *Oxford Advanced Learner's Dictionary*. Also, the middle dot in quoted entries is rendered with a full stop, while within the sample transcriptions hyphenation and syllabification points are indicated by a vertical bar |, regardless of their appearance in the source text.

* » [9.2 The Structure of Dictionary Entries](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIEN)
* [Home](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html)

[**¶**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIBO)**9.1 Dictionary Body and Overall Structure**

Overall, dictionaries have the same structure of front matter, body, and back matter familiar from other texts. In addition, this module defines [entry](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-entry.html), [entryFree](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-entryFree.html), and [superEntry](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-superEntry.html) as component-level elements which can occur directly within a text division or the text body.

The following tags can therefore be used to mark the gross structure of a printed dictionary; the dictionary-specific tags are discussed further in the following section.

* [**text**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-text.html) contains a single text of any kind, whether unitary or composite, for example a poem or drama, a collection of essays, a novel, a dictionary, or a corpus sample.
* [**front**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-front.html) (front matter) contains any prefatory matter (headers, abstracts, title page, prefaces, dedications, etc.) found at the start of a document, before the main body.
* [**body**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-body.html) (text body) contains the whole body of a single unitary text, excluding any front or back matter.
* [**back**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-back.html) (back matter) contains any appendixes, etc. following the main part of a text.
* [**div**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-div.html) (text division) contains a subdivision of the front, body, or back of a text.
* [**entry**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-entry.html) contains a single structured entry in any kind of lexical resource, such as a dictionary or lexicon.
* [**entryFree**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-entryFree.html) (unstructured entry) contains a single unstructured entry in any kind of lexical resource, such as a dictionary or lexicon.
* [**superEntry**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-superEntry.html) groups a sequence of entries within any kind of lexical resource, such as a dictionary or lexicon which function as a single unit, for example a set of homographs.

As members of the classes [att.entryLike](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.entryLike.html) and [att.sortable](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.sortable.html), [entry](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-entry.html) and [entryFree](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-entryFree.html) share the following attributes:

* [**att.entryLike**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.entryLike.html) provides an attribute used to distinguish different styles of dictionary entries.

|  |  |
| --- | --- |
| **type** | indicates type of entry, in dictionaries with multiple types. Suggested values include: 1] main; 2] hom (homograph) ; 3] xref (cross reference) ; 4] affix; 5] abbr (abbreviation) ; 6] supplemental; 7] foreign |

* [**att.sortable**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.sortable.html) provides attributes for elements in lists or groups that are sortable, but whose sorting key cannot be derived mechanically from the element content.

|  |  |
| --- | --- |
| **sortKey** | supplies the sort key for this element in an index, list or group which contains it. |

The front and back matter of a dictionary may well contain specialized material such as lists of common and proper nouns, grammatical tables, gazetteers, a ‘guide to the use of the dictionary’, etc. These should be tagged using elements defined elsewhere in these Guidelines, chiefly in the core module (chapter [3 Elements Available in All TEI Documents](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CO.html)) together with the specialized dictionary elements defined in this chapter.

The [body](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-body.html) element consists of a set of *entries*, optionally grouped into one or several [div](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-div.html) elements. These text divisions might, for example, correspond to sections for different letters of the alphabet, or to sections for different languages in a bilingual dictionary, as in the following example:

**<body>**
 **<div>**
  **<head>**English-French**</head>**
  **<entry>**
   **<form>**
    **<orth>**cat**</orth>**
   **</form>**
*<!-- ... -->*
  **</entry>**
  **<entry>**
   **<form>**
    **<orth>**dog**</orth>**
   **</form>**
*<!-- ... -->*
  **</entry>**
  **<entry>**
   **<form>**
    **<orth>**horse**</orth>**
   **</form>**
*<!-- ... -->*
  **</entry>**
 **</div>**
 **<div>**
  **<head>**French-English**</head>**
  **<entry>**
   **<form>**
    **<orth>**chat**</orth>**
   **</form>**
*<!-- ... -->*
  **</entry>**
  **<entry>**
   **<form>**
    **<orth>**chien**</orth>**
   **</form>**
*<!-- ... -->*
  **</entry>**
  **<entry>**
   **<form>**
    **<orth>**cheval**</orth>**
   **</form>**
*<!-- ... -->*
  **</entry>**
 **</div>**
**</body>**

In a print dictionary, the entries are typically typographically distinct entities, each headed by some morphological form of the lexical item described (the *headword*), and sorted in alphabetical order or (especially for non-alphabetic scripts) in some other conventional sequence. Dictionary entries should be encoded as distinct successive items, each marked as an [entry](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-entry.html) or [entryFree](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-entryFree.html) element. The type attribute may be used to distinguish different types of entries, for example main entries, related entries, run-on entries, or entries for cross-references, etc.

Some dictionaries provide distinct entries for homographs, on the basis of etymology, part-of-speech, or both, and typically provide a numeric superscript on the headword identifying the homograph number. In these cases each homograph should be encoded as a separate entry; the [superEntry](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-superEntry.html) element may optionally be used to group such successive homograph entries. In addition to a series of [entry](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-entry.html) elements, the [superEntry](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-superEntry.html) may contain a preliminary [form](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-form.html)group (see section [9.3.1 Information on Written and Spoken Forms](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPFO)) when information about hyphenation, pronunciation, etc., is given only once for two or more homograph entries. If the homograph number is to be recorded, the global attribute n may be used for this purpose. In some dictionaries, homographs are treated in distinct parts of the same entry; in these cases, they may be separated by use of the [hom](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-hom.html) element, for which see section [9.2.1 Hierarchical Levels](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIENHI).

A sort key, given in the sortKey attribute, is often required for superentries and entries, especially in cases where the order of entries does not follow the local character-set collating sequence (as, for example, when an entry for ‘3D’appears at the place where ‘three-D’ would appear).

A dictionary with no internal divisions might thus have a structure like the following; a [superEntry](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-superEntry.html) is shown grouping two homograph entries.

**<body>**
 **<entry>**
  **<form>**
   **<orth>**manifestation**</orth>**
*<!-- demonstration -->*
  **</form>**
 **</entry>**
 **<entry>**
  **<form>**
   **<orth>**émeute**</orth>**
*<!-- riot -->*
  **</form>**
 **</entry>**
 **<superEntry>**
  **<entry type="hom" n="1">**
   **<form>**
    **<orth>**grève**</orth>**
*<!-- strike -->*
   **</form>**
  **</entry>**
  **<entry type="hom" n="2">**
   **<form>**
    **<orth>**grève**</orth>**
*<!-- shore -->*
   **</form>**
  **</entry>**
 **</superEntry>**
**</body>**

* « [9.1 Dictionary Body and Overall Structure](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIBO)
* » [9.3 Top-level Constituents of Entries](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITP)
* [Home](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html)

[**¶**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIEN)**9.2 The Structure of Dictionary Entries**

A simple dictionary entry may contain information about the form of the word treated, its grammatical characterization, its definition, synonyms, or translation equivalents, its etymology, cross-references to other entries, usage information, and examples. These we refer to as the *constituent parts* or *constituents* of the entry; some dictionary constituents possess no internal structure, while others are most naturally viewed as groups of smaller elements, which may be marked in their own right. In some styles of markup, tags will be applied only to the low-level items, leaving the constituent groups which contain them untagged. We distinguish the class of *top-level constituents* of dictionary entries, which can occur directly within the [entry](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-entry.html) element, from the class of *phrase-level* constituents, which can normally occur only within top-level constituents. The top-level constituents of dictionary entries are described in section [9.2.2 Groups and Constituents](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIENGP), and documented more fully, together with their phrase-level sub-constituents, in section [9.3 Top-level Constituents of Entries](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITP).

In addition, however, dictionary entries often have a complex hierarchical structure. For example, an entry may consist of two or more sub-parts, each corresponding to information for a different part-of-speech homograph of the headword. The entry (or part-of-speech homographs, if the entry is split this way) may also consist of senses, each of which may in turn be composed of two or more sub-senses, etc. Each sub-part, homograph entry, sense, or sub-sense we call a *level*; at any level in an entry, any or all of the constituent parts of dictionary entries may appear. The hierarchical levels of dictionary entries are documented in section [9.2.1 Hierarchical Levels](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIENHI).

* » [9.2.2 Groups and Constituents](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIENGP)
* [Home](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html)

[**¶**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIENHI)**9.2.1 Hierarchical Levels**

The outermost structural level of an entry is marked with the elements [entry](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-entry.html) or [entryFree](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-entryFree.html). The [hom](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-hom.html) element marks the subdivision of entries into homographs differing in their part-of-speech. The [sense](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-sense.html) element marks the subdivision of entries and part-of-speech homographs into senses; this element nests recursively in order to provide for a hierarchy of sub-senses of any depth. It is recommended to use the [sense](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-sense.html) element even for an entry that has only one sense to group together all parts of the definition relating to the word sense since this leads to more consistent encoding across entries. All of these levels may each contain any of the constituent parts of an entry. A special case of hierarchical structure is represented by the [re](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-re.html) (related entry) element, which is discussed in section [9.3.6 Related Entries](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPRE). Finally, the element [dictScrap](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-dictScrap.html) may be used at any point in the hierarchy to delimit parts of the dictionary entry which are structurally anomalous, as further discussed in section [9.6 Unstructured Entries](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIFR).

* [**entry**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-entry.html) contains a single structured entry in any kind of lexical resource, such as a dictionary or lexicon.
* [**entryFree**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-entryFree.html) (unstructured entry) contains a single unstructured entry in any kind of lexical resource, such as a dictionary or lexicon.
* [**hom**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-hom.html) (homograph) groups information relating to one homograph within an entry.
* [**sense**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-sense.html) groups together all information relating to one word sense in a dictionary entry, for example definitions, examples, and translation equivalents.

|  |  |
| --- | --- |
| **level** | gives the nesting depth of this sense. |

* [**dictScrap**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-dictScrap.html) (dictionary scrap) encloses a part of a dictionary entry in which other phrase-level dictionary elements are freely combined.

For example, an entry with two senses will have the following structure:

**<entry>**
 **<sense n="1"/>**
 **<sense n="2"/>**
**</entry>**

bibliography

An entry with two homographs, the first with two senses and the second with three (one of which has two sub-senses), may have a structure like this:

**<entry>**
 **<hom n="1">**
  **<sense n="1">**
*<!-- ... -->*
  **</sense>**
  **<sense n="2">**
*<!-- ... -->*
  **</sense>**
 **</hom>**
 **<hom n="2">**
  **<sense n="1">**
   **<sense n="a">**
*<!-- ... -->*
   **</sense>**
   **<sense n="b">**
*<!-- ... -->*
   **</sense>**
  **</sense>**
  **<sense n="2">**
*<!-- ... -->*
  **</sense>**
  **<sense n="3">**
*<!-- ... -->*
  **</sense>**
 **</hom>**
**</entry>**

bibliography

In some dictionaries, homographs have separate entries; in such a case, as noted in section [9.1 Dictionary Body and Overall Structure](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIBO), the two homographs may be treated as entries, optionally grouped in a [superEntry](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-superEntry.html):

**<superEntry>**
 **<entry n="1" type="hom">**
  **<sense n="1">**
*<!-- ... -->*
  **</sense>**
  **<sense n="2">**
*<!-- ... -->*
  **</sense>**
 **</entry>**
 **<entry n="2" type="hom">**
  **<sense n="1">**
   **<sense n="a">**
*<!-- ... -->*
   **</sense>**
   **<sense n="b">**
*<!-- ... -->*
   **</sense>**
  **</sense>**
  **<sense n="2">**
*<!-- ... -->*
  **</sense>**
  **<sense n="3">**
*<!-- ... -->*
  **</sense>**
 **</entry>**
**</superEntry>**

bibliography

The hierarchic structure of a dictionary entry is enforced by the structures defined in this module. The content model for [entry](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-entry.html) specifies that entries do not nest, that homographs nest within entries, and that senses nest within entries, homographs, or senses, and may be nested to any depth to reflect the embedding of sub-senses. Any of the top-level constituents ([def](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-def.html), [usg](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-usg.html), [form](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-form.html), etc.) can appear at any level (i.e., within entries, homographs, or senses).

* « [9.2.1 Hierarchical Levels](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIENHI)
* [Home](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html)

[**¶**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIENGP)**9.2.2 Groups and Constituents**

As noted above, dictionary entries, and subordinate levels within dictionary entries, may comprise several constituent parts, each providing a different type of information about the word treated. The *top-level constituents* of dictionary entries are:

* information about the form of the word treated (orthography, pronunciation, hyphenation, etc.)
* grammatical information (part of speech, grammatical sub-categorization, etc.)
* definitions or translations into another language
* etymology
* examples
* usage information
* cross-references to other entries
* notes
* entries (often of reduced form) for related words, typically called *related entries*

Any of the hierarchical levels ([entry](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-entry.html), [entryFree](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-entryFree.html), [hom](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-hom.html), and [sense](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-sense.html)) may contain any of these top-level constituents, since information about word form, particular grammatical information, special pronunciation, usage information, etc., may apply to an entire entry, or to only one homograph, or only to a particular sense. The examples below illustrate this point.

The following elements are used to encode these top-level constituents:

* [**form**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-form.html) (form information group) groups all the information on the written and spoken forms of one headword.
* [**gramGrp**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-gramGrp.html) (grammatical information group) groups morpho-syntactic information about a lexical item, e.g. [pos](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-pos.html), [gen](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-gen.html), [number](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-number.html), [case](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-case.html), or [iType](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-iType.html) (inflectional class).
* [**def**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-def.html) (definition) contains definition text in a dictionary entry.
* [**cit**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-cit.html) (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.
* [**usg**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-usg.html) (usage) contains usage information in a dictionary entry.
* [**xr**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-xr.html) (cross-reference phrase) contains a phrase, sentence, or icon referring the reader to some other location in this or another text.
* [**etym**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-etym.html) (etymology) encloses the etymological information in a dictionary entry.
* [**re**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-re.html) (related entry) contains a dictionary entry for a lexical item related to the headword, such as a compound phrase or derived form, embedded inside a larger entry.
* [**note**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-note.html) contains a note or annotation.

In a simple entry with no internal hierarchy, all top-level constituents can appear as children of [entry](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-entry.html).

**com.peti.tor** **/k@m"petit@(r)/** n person who competes. OALD

**<entry>**
 **<form>**
  **<orth>**competitor**</orth>**
  **<hyph>**com|peti|tor**</hyph>**
  **<pron>**k@m"petit@(r)**</pron>**
 **</form>**
 **<gramGrp>**
  **<pos>**n**</pos>**
 **</gramGrp>**
 **<def>**person who competes.**</def>**
**</entry>**

For the elements which appear within the [form](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-form.html) and [gramGrp](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-gramGrp.html) elements of this and other examples, see below, section [9.3.1 Information on Written and Spoken Forms](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPFO), and section [9.3.2 Grammatical Information](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPGR).

Any top-level constituent can appear at any level when the hierarchical structure of the entry is more complex. The most obvious examples are [def](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-def.html) and [cit](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-cit.html), which appear at the [sense](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-sense.html) level when several senses or translations exist:

**disproof** **(dɪsˈpru:f)** *n* **1** facts that disprove something **2** the act of disproving. CED

**<entry>**
 **<form>**
  **<orth>**disproof**</orth>**
  **<pron notation="ipa">**dɪsˈpru:v**</pron>**
 **</form>**
 **<gramGrp>**
  **<pos>**n**</pos>**
 **</gramGrp>**
 **<sense n="1">**
  **<def>**facts that disprove something**</def>**
 **</sense>**
 **<sense n="2">**
  **<def>**the act of disproving**</def>**
 **</sense>**
**</entry>**

For ease of processing of such entries containing multiple senses along with those containing only a single sense, it is recommended to use [sense](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-sense.html) in all entries to wrap those elements relating to a particular word sense.

In the following example, [gramGrp](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-gramGrp.html) is used to distinguish two homographs:

**bray** **/breI/** n cry of an ass; sound of a trumpet. ∙ vt [VP2A] make a cry or sound of this kind. OALD

**<entry>**
 **<form>**
  **<orth>**bray**</orth>**
  **<pron>**breI**</pron>**
 **</form>**
 **<hom>**
  **<gramGrp>**
   **<pos>**n**</pos>**
  **</gramGrp>**
  **<sense>**
   **<def>**cry of an ass; sound of a trumpet.**</def>**
  **</sense>**
 **</hom>**
 **<hom>**
  **<gramGrp>**
   **<pos>**vt**</pos>**
   **<subc>**VP2A**</subc>**
  **</gramGrp>**
  **<sense>**
   **<def>**make a cry or sound of this kind.**</def>**
  **</sense>**
 **</hom>**
**</entry>**

Information of the same kind can appear at different levels within the same entry; here, grammatical information occurs both at entry and homograph level.

**ca.reen** **/k@"ri:n/** vt,vi 1 [VP6A] turn (a ship) on one side for cleaning, repairing, etc. 2 [VP6A, 2A] (cause to) tilt, lean over to one side. OALD

**<entry>**
 **<form>**
  **<orth>**careen**</orth>**
  **<hyph>**ca|reen**</hyph>**
  **<pron>**k@"ri:n**</pron>**
 **</form>**
 **<gramGrp>**
  **<pos>**vt**</pos>**
  **<pos>**vi**</pos>**
 **</gramGrp>**
 **<sense n="1">**
  **<gramGrp>**
   **<subc>**VP6A**</subc>**
  **</gramGrp>**
  **<def>**turn (a ship) on one side for cleaning, repairing, etc.**</def>**
 **</sense>**
 **<sense n="2">**
  **<gramGrp>**
   **<subc>**VP6A**</subc>**
   **<subc>**VP2A**</subc>**
  **</gramGrp>**
  **<def>**(cause to) tilt, lean over to one side.**</def>**
 **</sense>**
**</entry>**

Alone among the constituent groups, [form](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-form.html) can appear at the [superEntry](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-superEntry.html) level as well as at the [entry](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-entry.html), [hom](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-hom.html), and [sense](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-sense.html) levels:

**a.ban.don** 1**/@"band@n/**v [T1] 1 to leave completely and for ever; desert: The sailors abandoned the burning ship. 2 …**abandon** 2 n [U] the state when one's feelings and actions are uncontrolled; freedom from control...LDOCE

**<superEntry>**
 **<form>**
  **<orth>**abandon**</orth>**
  **<hyph>**a|ban|don**</hyph>**
  **<pron>**@"band@n**</pron>**
 **</form>**
 **<entry n="1">**
  **<gramGrp>**
   **<pos>**v**</pos>**
   **<subc>**T1**</subc>**
  **</gramGrp>**
  **<sense n="1">**
   **<def>**to leave completely and for ever … **</def>**
  **</sense>**
  **<sense n="2"/>**
 **</entry>**
 **<entry n="2">**
  **<gramGrp>**
   **<pos>**n**</pos>**
   **<subc>**U**</subc>**
  **</gramGrp>**
  **<sense>**
   **<def>**the state when one's feelings and actions are uncontrolled; freedom
       from control…**</def>**
  **</sense>**
 **</entry>**
**</superEntry>**

* « [9.2 The Structure of Dictionary Entries](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIEN)
* » [9.4 Headword and Pronunciation References](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIHW)
* [Home](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html)

[**¶**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITP)**9.3 Top-level Constituents of Entries**

This section describes the top-level constituents of dictionary entries, together with the phrase-level constituents peculiar to each.

* the [form](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-form.html) element, which groups orthographic information and pronunciations, is described in section [9.3.1 Information on Written and Spoken Forms](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPFO)
* the [gramGrp](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-gramGrp.html) element, which groups elements for the grammatical characterization of the headword, is described in section [9.3.2 Grammatical Information](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPGR)
* the [def](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-def.html) element, which describes the meaning of the headword, is described in section [9.3.3 Sense Information](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPSE)
* the [etym](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-etym.html) element and its special phrase-level elements are documented in section [9.3.4 Etymological Information](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPET)
* the [cit](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-cit.html) element and its specific applications are described in section [9.3.3 Sense Information](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPSE) and section [9.3.5 Other Information](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPMI)
* the [usg](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-usg.html), [lbl](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-lbl.html), [xr](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-xr.html), and [note](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-note.html) elements are described in section [9.3.5 Other Information](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPMI)
* the [re](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-re.html) element, which marks nested entries for related words, is described in section [9.3.6 Related Entries](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPRE)
* » [9.3.2 Grammatical Information](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPGR)
* [Home](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html)

[**¶**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPFO)**9.3.1 Information on Written and Spoken Forms**

Dictionary entries most often begin with information about the form of the word to which the entry applies. Typically, the orthographic form of the word, sometimes marked for syllabification or hyphenation, is the first item in an entry. Other information about the word, including variant or alternate forms, inflected forms, pronunciation, etc., is also often given.

The following elements should be used to encode this information: the [form](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-form.html) element groups one or more occurrences of any of them; it can also be recursively nested to reflect more complex sub-grouping of information about word form(s), as shown in the examples.

* [**form**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-form.html) (form information group) groups all the information on the written and spoken forms of one headword.

|  |  |
| --- | --- |
| **type** | classifies form as simple, compound, etc. Suggested values include: 1] simple; 2] lemma; 3] variant; 4] compound; 5] derivative; 6] inflected; 7] phrase |

* [**orth**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-orth.html) (orthographic form) gives the orthographic form of a dictionary headword.

|  |  |
| --- | --- |
| **type** | gives the type of spelling. |
| **extent [att.partials]** | indicates whether the pronunciation or orthography applies to all or part of a word. Suggested values include: 1] full (full form) ; 2] pref (prefix) ; 3] suff (suffix) ; 4] inf (infix) ; 5] part (partial) |

* [**pron**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-pron.html) (pronunciation) contains the pronunciation(s) of the word.

|  |  |
| --- | --- |
| **extent [att.partials]** | indicates whether the pronunciation or orthography applies to all or part of a word. Suggested values include: 1] full (full form) ; 2] pref (prefix) ; 3] suff (suffix) ; 4] inf (infix) ; 5] part (partial) |

* [**hyph**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-hyph.html) (hyphenation) contains a hyphenated form of a dictionary headword, or hyphenation information in some other form.
* [**syll**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-syll.html) (syllabification) contains the syllabification of the headword.
* [**stress**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-stress.html) contains the stress pattern for a dictionary headword, if given separately.
* [**lbl**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-lbl.html) (label) contains a label for a form, example, translation, or other piece of information, e.g. abbreviation for, contraction of, literally, approximately, synonyms:, etc.

In addition to those listed above, the following elements, which encode morphological details of the form, may also occur within [form](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-form.html) elements:

* [**gram**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-gram.html) (grammatical information) within an entry in a dictionary or a terminological data file, contains grammatical information relating to a term, word, or form.

|  |  |
| --- | --- |
| **type** | classifies the grammatical information given according to some convenient typology—in the case of terminological information, preferably the dictionary of data element types specified in [ISO 12620](http://www.isocat.org/). Sample values include: 1] pos (part of speech) ; 2] gen (gender) ; 3] num (number) ; 4] animate; 5] proper |

* [**gen**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-gen.html) (gender) identifies the morphological gender of a lexical item, as given in the dictionary.
* [**number**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-number.html) indicates grammatical number associated with a form, as given in a dictionary.
* [**case**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-case.html) contains grammatical case information given by a dictionary for a given form.
* [**per**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-per.html) (person) contains an indication of the grammatical person (1st, 2nd, 3rd, etc.) associated with a given inflected form in a dictionary.
* [**tns**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-tns.html) (tense) indicates the grammatical tense associated with a given inflected form in a dictionary.
* [**mood**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-mood.html) contains information about the grammatical mood of verbs (e.g. indicative, subjunctive, imperative).
* [**iType**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-iType.html) (inflectional class) indicates the inflectional class associated with a lexical item.

|  |  |
| --- | --- |
| **type** | indicates the type of indicator used to specify the inflection class, when it is necessary to distinguish between the usual abbreviated indications (e.g. *inv*) and other kinds of indicators, such as special codes referring to conjugation patterns, etc. Sample values include: 1] abbrev; 2] verbTable |

* [**pos**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-pos.html) (part of speech) indicates the part of speech assigned to a dictionary headword such as noun, verb, or adjective.
* [**subc**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-subc.html) (subcategorization) contains subcategorization information (transitive/intransitive, countable/non-countable, etc.)
* [**colloc**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-colloc.html) (collocate) contains any sequence of words that co-occur with the headword with significant frequency.

Of these, the [gram](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-gram.html) element is most general, and all of the others are synonymous with a [gram](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-gram.html) element with appropriate values (gen, number, case, etc.) for the type attribute.

The use of these elements as children of [form](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-form.html) is deprecated; instead, they should always be children of a [gramGrp](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-gramGrp.html) within [form](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-form.html) when describing that particular form of the word.

Different dictionaries use different means to mark hyphenation, syllabification, and stress, and they often use some unusual glyphs (e.g., the ‘middle dot’ for hyphenation). All of these glyphs are in the Unicode character set, as discussed in [Character References](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/SG.html#SG-er). When transcribing representations of pronunciation the International Phonetic Alphabet should be used. It may be convenient (as has been done in the text of this chapter) to use a simple transliteration scheme for this; such a scheme should however be properly documented in the header.

In the simplest case, nothing is given but the orthography:

**<form>**
 **<orth>**doom-laden**</orth>**
**</form>**

Often, however, pronunciation is given.

**soucoupe** [sukup] … DNT

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

For a variety of reasons including ease of processing, it may be desired to split into separate elements information which is collapsed into a single element in the source text; orthography and hyphenation may for example be transcribed as separate elements, although given together in the source text. For a discussion of the issues involved, and of methods for retaining both the presentation form and the interpreted form, see section [9.5 Typographic and Lexical Information in Dictionary Data](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIMV).

This example splits orthography and hyphenation, and adds syllabification because it differs from hyphenation:

**ar.ea** … W7

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

Multiple orthographic forms may be given, e.g. to illustrate a word's inflectional pattern:

**brag** … *vb* **brags, bragging, bragged** … CED

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

Or the inflectional pattern may be indicated by reference to a table of paradigms, as here:

**horrifier** **[ORifje]** (7) vt … [C/R]

**<form>**
 **<orth>**horrifier**</orth>**
 **<pron>**ORifje**</pron>**
 **<gramGrp>**
  **<iType type="vbtable">**7**</iType>**
*<!-- ... -->*
 **</gramGrp>**
**</form>**
*<!-- ... -->*

Explanatory labels may be attached to alternate forms:

**MTBF** *abbreviation for* mean time between failures CED

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

When multiple orthographic forms are given, a pronunciation may be associated with all of them, as here:

**biryani** *or* **biriani** **(ˌbɪrɪˈa:nɪ)** … CED

**<form>**
 **<orth>**biryani**</orth>**
 **<orth>**biriani**</orth>**
 **<pron notation="ipa">**ˌbɪrɪˈa:nɪ**</pron>**
**</form>**

In other cases, different pronunciations are provided for different orthographic forms; here, the [form](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-form.html) element is repeated to associate the first orthographic form explicitly with the first pronunciation, and the second orthographic form with the second pronunciation:

**mackle** **(ˈmækᵊl)** *or* **macule** **(ˈmækju:l)** … CED

**<form>**
 **<orth>**mackle**</orth>**
 **<pron notation="ipa">**ˈmækᵊl**</pron>**
**</form>**
**<form>**
 **<orth>**macule**</orth>**
 **<pron notation="ipa">**ˈmækju:l**</pron>**
**</form>**

Recursive nesting of the [form](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-form.html) element can preserve relations among elements that are implicit in the text. For example, in the CED entry for ‘hospitaller’, it is clear that ‘U.S.’ is associated only with ‘hospitaler’, but that the pronunciation applies to both forms. The following encoding preserves these relations:

**hospitaller** *or US* **hospitaler** **(ˈhɒspɪtələ)** … CED

**<form>**
 **<orth>**hospitaller**</orth>**
 **<form>**
  **<usg type="geo">**US**</usg>**
  **<orth>**hospitaler**</orth>**
 **</form>**
 **<pron notation="ipa">**ˈhɒspɪtələ**</pron>**
**</form>**

* « [9.3.1 Information on Written and Spoken Forms](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPFO)
* » [9.3.3 Sense Information](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPSE)
* [Home](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html)

[**¶**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPGR)**9.3.2 Grammatical Information**

The [gramGrp](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-gramGrp.html) element groups grammatical information, such as part of speech, subcategorization information (e.g., syntactic patterns for verbs, count/mass distinctions for nouns), etc. It can contain any of the morphological elements defined in section [9.3.1 Information on Written and Spoken Forms](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPFO) for [form](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-form.html) and can appear as a child of [entry](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-entry.html), [form](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-form.html), [sense](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-sense.html), [cit](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-cit.html), or any other element containing content about which there is grammatical information. For example, in the entry ‘**pinna** **(ˈpɪnə)** *n*, *pl* **-nae** (-ni:) *or* **-nas** CED’, the word defined can be either singular or plural; the ‘pl.’ specification applies only to the inflected forms provided. Compare this with ‘pants (paents) pl. n.’, where ‘pl.’ applies to the headword itself.

As noted above in section [9.3.1 Information on Written and Spoken Forms](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPFO), the elements for morphological information are simply shorthand for the general purpose [gram](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-gram.html) element. Consider this entry for the French word *médire*:

**médire** v.t. ind. (de) … PLC

This entry can be tagged using specialized grammatical elements:

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

Or using the [gram](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-gram.html) element:

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

Like [form](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-form.html), [gramGrp](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-gramGrp.html) can be repeated, recursively nested, or used at the [sense](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-sense.html) level to show relations among elements.

**isotope** adj. et n. m. … DNT

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

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

**<entry>**
 **<form>**
  **<orth>**wits**</orth>**
  **<pron notation="ipa">**wɪts**</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>**

* « [9.3.2 Grammatical Information](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPGR)
* » [9.3.4 Etymological Information](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPET)
* [Home](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html)

[**¶**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPSE)**9.3.3 Sense Information**

Dictionaries may describe the meanings of words in a wide variety of different ways—by means of synonyms, paraphrases, translations into other languages, formal definitions in various highly stylized forms, etc. No attempt is made here to distinguish all the different forms which sense information may take; all of them may be tagged using the [def](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-def.html) element described in section [9.3.3.1 Definitions](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPDE).

As a special case it is frequently desirable to distinguish the provision of translation equivalents in other languages from other forms of sense information; the use of <cit type="translation"> (which groups a translation equivalent with related information such as its grammatical description) for this purpose is described in section [9.3.3.2 Translation Equivalents](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPTR).

* » [9.3.3.2 Translation Equivalents](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPTR)
* [Home](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html)

[**¶**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPDE)**9.3.3.1 Definitions**

Dictionary definitions are those pieces of prose in a dictionary entry that describe the meaning of some lexical item. Most often, definitions describe the headword of the entry; in some cases, they describe translated texts, examples, etc.; see <cit type="translation">, section [9.3.3.2 Translation Equivalents](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPTR), and <cit type="example">, section [9.3.5.1 Examples](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPEG). The [def](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-def.html) element directly contains the text of the definition; unlike [form](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-form.html) and [gramGrp](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-gramGrp.html), it does not serve solely to group a set of smaller elements. The close analysis of definition text, such as the tagging of hypernyms, typical objects, etc., is not covered by these Guidelines.

Definitions may occur directly within an entry; when multiple definitions are given, they are typically identified as belonging to distinct senses, as here:

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

**<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>**

In multilingual dictionaries, it is sometimes possible to distinguish translation equivalents from definitions proper; here a [def](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-def.html) element is distinguished from the translation information within which it appears.

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

**<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>**

* « [9.3.3.1 Definitions](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPDE)
* [Home](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html)

[**¶**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPTR)**9.3.3.2 Translation Equivalents**

Multilingual dictionaries contain information about translations of a given word in some source language for one or more target languages. Minimally, the dictionary provides the corresponding translation in the target language; other material, such as morphological information (gender, case), various kinds of usage restrictions, etc., may also be given. If translation equivalents are to be distinguished from other kinds of sense information, they may be encoded using <cit type="translation">. The global xml:lang attribute should be used to specify the target language.

As in monolingual dictionaries, the [sense](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-sense.html) element is used in multilingual dictionaries to group information (forms, grammatical information, usage, translation(s), etc.) about a given sense of a word where necessary. Information about the individual translation equivalents within a sense is grouped using <cit type="translation">. This information may include the translation text (tagged [q](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-q.html) or [quote](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-quote.html)), morphological information ([gen](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-gen.html), [case](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-case.html), etc.), usage notes ([usg](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-usg.html)), translation labels ([lbl](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-lbl.html)), and definitions ([def](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-def.html)).When bibliographic data is provided, the [quote](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-quote.html) element should be used.

* [**cit**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-cit.html) (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**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-lbl.html) (label) contains a label for a form, example, translation, or other piece of information, e.g. abbreviation for, contraction of, literally, approximately, synonyms:, etc.

Note how in the following example, different translation equivalents are grouped into the same or different senses, following the punctuation of the source and the usage labels:

**dresser** … (a) (Theat) habilleur m, -euse f; (Comm: window ~) étalagiste mf. she's a stylish ~ elle s'habille avec chic; V hair. (b) (tool) (for wood) raboteuse f; (for stone) rabotin m. CR

**<entry n="1">**
 **<form>**
  **<orth>**dresser**</orth>**
 **</form>**
 **<sense n="a">**
  **<sense>**
   **<usg type="dom">**Theat**</usg>**
   **<cit type="translation" xml:lang="fr">**
    **<quote>**habilleur**</quote>**
    **<gramGrp>**
     **<gen>**m**</gen>**
    **</gramGrp>**
   **</cit>**
   **<cit type="translation" xml:lang="fr">**
    **<quote>**-euse**</quote>**
    **<gramGrp>**
     **<gen>**f**</gen>**
    **</gramGrp>**
   **</cit>**
  **</sense>**
  **<sense>**
   **<usg type="dom">**Comm**</usg>**
   **<form type="compound">**
    **<orth>**window **<oRef/>**
    **</orth>**
   **</form>**
   **<cit type="translation" xml:lang="fr">**
    **<quote>**étalagiste**</quote>**
    **<gramGrp>**
     **<gen>**mf**</gen>**
    **</gramGrp>**
   **</cit>**
  **</sense>**
  **<cit type="example">**
   **<quote>**she's a stylish **<oRef/>**
   **</quote>**
   **<cit type="translation" xml:lang="fr">**
    **<quote>**elle s'habille avec chic**</quote>**
   **</cit>**
  **</cit>**
  **<xr type="see">**V. **<ref target="#hair">**hair**</ref>**
  **</xr>**
 **</sense>**
 **<sense n="b">**
  **<usg type="category">**tool**</usg>**
  **<sense>**
   **<usg type="hint">**for wood**</usg>**
   **<cit type="translation" xml:lang="fr">**
    **<quote>**raboteuse**</quote>**
    **<gramGrp>**
     **<gen>**f**</gen>**
    **</gramGrp>**
   **</cit>**
  **</sense>**
  **<sense>**
   **<usg type="hint">**for stone**</usg>**
   **<cit type="translation" xml:lang="fr">**
    **<quote>**rabotin**</quote>**
    **<gramGrp>**
     **<gen>**m**</gen>**
    **</gramGrp>**
   **</cit>**
  **</sense>**
 **</sense>**
**</entry>**
*<!-- ... -->*
**<entry xml:id="hair">**
 **<sense>**
*<!-- ... -->*
 **</sense>**
**</entry>**

In the following example, a distinction is made between the translation equivalent (‘OAS’) and a descriptive phrase providing further information for the user of the dictionary.

**O.A.S.** ... nf (abrév de **Organisation de l'Armée secrète**) OAS (illegal military organization supporting French rule of Algeria). CR

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

Note that <cit type="translation"> may also be used in monolingual dictionaries when a translation is given for a foreign word:

***havdalah*** *or* ***havdoloh*** *Hebrew* (**havdaˈla;** *Yiddish* **havˈdɔlə)** *n Judaism* the ceremony marking the end of the sabbath or of a festival, including the blessings over wine, candles and spices [literally: separation] CED

**<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>**

* « [9.3.3 Sense Information](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPSE)
* » [9.3.5 Other Information](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPMI)
* [Home](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html)

[**¶**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPET)**9.3.4 Etymological Information**

The element [etym](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-etym.html) marks a block of etymological information. 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. This variation in structure makes it impracticable to define tags which capture the entire intellectual structure of the etymology or record the precise interrelation of all the words mentioned. It is, however, feasible to mark some of the more obvious phrase-level elements frequently found in etymologies, using tags defined in the core module or elsewhere in this chapter. Of particular relevance for the markup of etymologies are:

* [**etym**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-etym.html) (etymology) encloses the etymological information in a dictionary entry.
* [**lang**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-lang.html) (language name) contains the name of a language mentioned in etymological or other linguistic discussion.
* [**date**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-date.html) contains a date in any format.
* [**mentioned**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-mentioned.html) marks words or phrases mentioned, not used.
* [**gloss**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-gloss.html) identifies a phrase or word used to provide a gloss or definition for some other word or phrase.
* [**pron**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-pron.html) (pronunciation) contains the pronunciation(s) of the word.
* [**usg**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-usg.html) (usage) contains usage information in a dictionary entry.
* [**lbl**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-lbl.html) (label) contains a label for a form, example, translation, or other piece of information, e.g. abbreviation for, contraction of, literally, approximately, synonyms:, etc.

As in other prose, individual word forms mentioned in an etymological description are tagged with [mentioned](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-mentioned.html) elements. Pronunciations, usage labels, and glosses can be tagged using the [pron](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-pron.html), [usg](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-usg.html), and [gloss](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-gloss.html) elements defined elsewhere in these Guidelines. In addition, the [lang](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-lang.html) element may be used to identify a particular language name where it appears, in addition to using the xml:lang attribute of the [mentioned](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-mentioned.html) element.

Examples:

**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>**

* « [9.3.4 Etymological Information](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPET)
* » [9.3.6 Related Entries](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPRE)
* [Home](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html)

[**¶**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPMI)**9.3.5 Other Information**

[**¶**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPEG)**9.3.5.1 Examples**

Dictionaries typically include examples of word use, usually accompanying definitions or translations. In some cases, the examples are quotations from another source, and are occasionally followed by a citation to the author.

The <cit type="example"> element contains usage examples and associated information; the example text itself should be enclosed in a [q](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-q.html) or [quote](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-quote.html) element. The [cit](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-cit.html) element associates a quotation with a bibliographic reference to its source.

* [**q**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-q.html) (quoted) contains material which is distinguished from the surrounding text using quotation marks or a similar method, for any one of a variety of reasons including, but not limited to: direct speech or thought, technical terms or jargon, authorial distance, quotations from elsewhere, and passages that are mentioned but not used.
* [**quote**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-quote.html) (quotation) contains a phrase or passage attributed by the narrator or author to some agency external to the text.
* [**cit**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-cit.html) (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.

Examples frequently abbreviate the headword, and so their transcription will frequently make use of the [oRef](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-oRef.html) element described below in section [9.4 Headword and Pronunciation References](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIHW).

Examples:

**multiplex** **/…/** adj tech having many parts: the multiplex eye of the fly. LDOCE

**<quote>**the multiplex eye of the fly.**</quote>**

Or when one wants a more comprehensive representation of examples:

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

As the following example shows, [cit](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-cit.html) can also contain elements such as [pron](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-pron.html), [def](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-def.html), etc.

**some** … 4. (S~ and any are used with more): Give me ~ more**/s@'mO:(r)/** OALD

**<sense n="4">**
 **<usg type="colloc">**
  **<oRef type="cap"/>** and **<mentioned>**any**</mentioned>** are used with
 **<mentioned>**more**</mentioned>**
 **</usg>**
 **<cit type="example">**
  **<quote>**Give me **<oRef/>** more**</quote>**
  **<pron extent="part">**s@'mO:(r)**</pron>**
 **</cit>**
**</sense>**

In multilingual dictionaries, examples may also be accompanied by translations:

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

**<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>**

When a source is indicated, the example should be marked with a [bibl](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-bibl.html) element:

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

**<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>**

* « [9.3.5.1 Examples](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPEG)
* » [9.3.5.3 Cross-References to Other Entries](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPXR)
* [Home](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html)

[**¶**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPUS)**9.3.5.2 Usage Information and Other Labels**

Most dictionaries provide restrictive labels and phrases indicating the usage of given words or particular senses. Other phrases, not necessarily related to usage, may also be attached to forms, translations, cross-references, and examples. The following elements are provided to mark up such labels:

* [**usg**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-usg.html) (usage) contains usage information in a dictionary entry.
* [**lbl**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-lbl.html) (label) contains a label for a form, example, translation, or other piece of information, e.g. abbreviation for, contraction of, literally, approximately, synonyms:, etc.

As indicated in the following section ([9.3.5.3 Cross-References to Other Entries](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPXR)), the [lbl](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-lbl.html) element may be used for any kind of significative phrase or label within the text. The [usg](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-usg.html) element is a specialization of this to mark usage labels in particular. Usage labels typically indicate

* 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.)

Many dictionaries provide an explanation and/or a list of such usage labels in a preface or appendix. The type of the usage information may be indicated in the type attribute on the [usg](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-usg.html) element. Some typical values are:

**geo**

geographic area

**time**

temporal, historical era (‘archaic’, ‘old’, etc.)

**dom**

domain

**reg**

register

**style**

style (figurative, literal, etc.)

**plev**

preference level (‘chiefly’, ‘usually’, etc.)

**acc**

acceptability

**lang**

language for foreign words, spellings pronunciations, etc.

**gram**

grammatical usage

In addition to this kind of information, multilingual dictionaries often provide ‘semantic cues’ to help the user determine the right sense of a word in the source language (and hence the correct translation). These include synonyms, concept subdivisions, typical subjects and objects, typical verb complements, etc. These labels may also be marked with the [usg](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-usg.html) element; sample values for the type attribute in these cases include:

**syn**

synonym given to show use

**hyper**

hypernym given to show usage

**colloc**

collocation given to show usage

**comp**

typical complement

**obj**

typical object

**subj**

typical subject

**verb**

typical verb

**hint**

unclassifiable piece of information to guide sense choice

In this entry, one spelling is marked as geographically restricted:

**colour** *or US* **color** … CED

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

In the next example, usage labels are used to indicate domains, register, and synonyms associated with different senses:

**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>**

When the usage label is hard to classify, it may be described as a ‘hint’:

**rempaillage** […] nm reseating, rebottoming (with straw). CR

**<entry>**
 **<cit type="translation" xml:lang="en">**
  **<quote>**reseating**</quote>**
  **<quote>**rebottoming**</quote>**
  **<usg type="hint">**with straw**</usg>**
 **</cit>**
**</entry>**

* « [9.3.5.2 Usage Information and Other Labels](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPUS)
* » [9.3.5.4 Notes within Entries](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPNO)
* [Home](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html)

[**¶**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPXR)**9.3.5.3 Cross-References to Other Entries**

Dictionary entries frequently refer to information in other entries, often using extremely dense notations to convey the headword of the entry to be sought, the particular part of the entry being referred to, and the nature of the information to be sought there (synonyms, antonyms, usage notes, etymology, an illustration, etc.)

Cross-references may be tagged in dictionaries using the [ref](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-ref.html) and [ptr](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-ptr.html) elements defined in the core module (section [3.6 Simple Links and Cross-References](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CO.html#COXR)). In addition, the [xr](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-xr.html) element may be used to group all the information relating to a cross-reference.

* [**xr**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-xr.html) (cross-reference phrase) contains a phrase, sentence, or icon referring the reader to some other location in this or another text.
* [**ref**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-ref.html) (reference) defines a reference to another location, possibly modified by additional text or comment.
* [**ptr**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-ptr.html) (pointer) defines a pointer to another location.
* [**lbl**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-lbl.html) (label) contains a label for a form, example, translation, or other piece of information, e.g. abbreviation for, contraction of, literally, approximately, synonyms:, etc.

As in other types of text, the actual pointing element (e.g. [ref](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-ref.html) or [ptr](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-ptr.html)) is used to tag the cross-reference target proper (in dictionaries, usually the headword, possibly accompanied by a homograph number, a sense number, or other further restriction specifying what portion of the target entry is being referred to). The [xr](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-xr.html) element is used to group the target with any accompanying phrases or symbols used to label the cross-reference; the cross-reference label itself may be tagged as a [lbl](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-lbl.html) or may remain untagged. Both of the following are thus legitimate:

**glee** … Compare **madrigal** (1) CED

**<entry>**
 **<form>**
  **<orth>**glee**</orth>**
 **</form>**
 **<xr>**Compare **<ptr target="#madrigal.1"/>**
 **</xr>**
**</entry>**
**<entry xml:id="madrigal.1">**
 **<form>**
*<!-- ... -->*
 **</form>**
**</entry>**

**hostellerie** Syn. de hôtellerie (sens 1). DNT

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

In addition to using, or not using, [lbl](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-lbl.html) to mark the cross-reference label, the two examples differ in another way. The former assumes that the first sense of *madrigal* has the identifier madrigal.1, and that the specific form of the reference in the source volume can be reconstructed, if needed, from that information. The latter does not require the first sense of ‘hôtellerie’ to have an identifier, and retains the print form of the cross-reference; by omitting the target attribute of the [ref](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-ref.html) element, however, the second example does assume implicitly either that some software could usefully parse the phrase tagged as a [ref](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-ref.html) and find the location referred to, or else that such processing will not be necessary.

The type attribute on the pointing element or on the [xr](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-xr.html) element may be used to indicate what kind of cross-reference is being made, using any convenient typology. Since different dictionaries may label the same kind of cross-reference in different ways, it may be useful to give normalized indications in the type attribute, enabling the encoder to distinguish irregular forms of cross-reference more reliably:

**rose**2 … *vb* the past tense of **rise** CED

**<entry n="2">**
 **<form>**
  **<orth>**rose**</orth>**
 **</form>**
 **<xr type="inflectedForm">**
  **<lbl>**the past tense of**</lbl>**
  **<ref target="#rise">**rise**</ref>**
 **</xr>**
**</entry>**
*<!-- ... -->*
**<entry xml:id="rise">**
 **<form>**
  **<orth>**rise**</orth>**
 **</form>**
*<!-- main entry for "rise" as verb -->*
**</entry>**

from cross-references for synonyms and the like:

**antagonist** … syn see **adverse**W7

**<xr type="synonym">**
 **<lbl>**syn see**</lbl>**
 **<ref target="#adverse">**adverse**</ref>**
**</xr>**
*<!-- ... -->*
**<entry xml:id="adverse">**
 **<form>**
  **<orth>**adverse**</orth>**
 **</form>**
*<!-- list of synonyms for "adverse" -->*
**</entry>**

Strictly speaking, the reference above is not to the entry for *adverse*, but to the list of synonyms found within that entry. In some cases, the cross-reference is to a particular subset of the meanings of the entry in question:

**globe** …V. **armillaire** (sphère) PR

**<xr>**V. **<ref target="#armillaire">**armillaire**</ref>**
 **<lbl type="sense-restriction">**sphère**</lbl>**
**</xr>**

Cross-references occasionally occur in definition texts, example texts, etc., or may be free-standing within an entry. These may typically be encoded using [ref](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-ref.html) or [ptr](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-ptr.html), without an enclosing [xr](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-xr.html). For example:

**entacher** … Acte entaché de nullité, contenant un vice de forme ou passé par un incapable\*. DNT

The asterisk signals a reference to the entry for *incapable*.

**<def>**contenant un vice de forme ou passé par un **<ptr target="#incapable"/>**.**</def>**

In some cases, the form in the definition is inflected, and thus [ref](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-ref.html) must be used to indicate more exactly the intended target, as here:

**justifier** …4. IMPRIM Donner a (une ligne) une longeur convenable au moyen de blancs (2, sens 1, 3). DNT

**<sense n="4">**
 **<usg type="dom">**imprim**</usg>**
 **<def>**Donner a (une ligne) une longeur convenable au moyen de
 **<ref target="#blanc-2.1.3">**blancs (2, sens 1, 3)**</ref>**
 **</def>**
**</sense>**
**<entry xml:id="blanc" n="2">**
*<!-- ... -->*
 **<sense n="1">**
*<!-- ... -->*
  **<def xml:id="blanc-2.1.3">**...**</def>**
*<!-- ... -->*
 **</sense>**
*<!-- ... -->*
**</entry>**

* « [9.3.5.3 Cross-References to Other Entries](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPXR)
* [Home](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html)

[**¶**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPNO)**9.3.5.4 Notes within Entries**

Dictionaries may include extensive explanatory notes about usage, grammar, context, etc. within entries. Very often, such notes appear as a separate section at the end of an entry. The standard [note](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-note.html)element should be used for such material.

* [**note**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-note.html) contains a note or annotation.

For example:

**neither** **(ˈnaɪðə, ˈni:ðə)** *determiner* **1a** not one nor the other (of two); not either: *neither foot is swollen* … **usage** A verb following a compound subject that uses *neither*… should be in the singular if both subjects are in the singular: *neither Jack nor John has done the work* CED

**<entry>**
 **<form type="contraction">**
  **<orth>**neither**</orth>**
  **<pron notation="ipa">**ˈnaɪðə**</pron>**,
 **<pron notation="ipa">**ˈni:ðə**</pron>**
 **</form>**
*<!-- ... -->*
 **<cit type="example">**
  **<quote>**neither foot is swollen**</quote>**
 **</cit>**
 **<note type="usage">**A verb following a compound subject
   that uses **<hi rend="italic">**neither**</hi>**… should be
   in the singular if both subjects are in the singular:
 **<hi rend="italic">**neither Jack nor John has done the work**</hi>**
 **</note>**
**</entry>**

The formal declaration for [note](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-note.html) is given in section [3.8 Notes, Annotation, and Indexing](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CO.html#CONO).

* « [9.3.5 Other Information](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPMI)
* [Home](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html)

[**¶**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPRE)**9.3.6 Related Entries**

The [re](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-re.html) element encloses a degenerate entry which appears in the body of another entry for some purpose. Many dictionaries include related entries for direct derivatives or inflected forms of the entry word, or for compound words, phrases, collocations, and idioms containing the entry word.

Related entries can be complex, and may in fact include any of the information to be found in a regular entry. Therefore, the [re](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-re.html) element is defined to contain the same elements as an [entry](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-entry.html) element, with the exception that it may not contain any nested [re](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-re.html) elements.

Examples:

**bevvy** **(ˈbɛvɪ)** *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* CED

**<entry>**
 **<form>**
  **<orth>**bevvy**</orth>**
  **<pron notation="ipa">**ˈbɛvɪ**</pron>**
 **</form>**
 **<usg type="reg">**informal**</usg>**
 **<hom>**
  **<gramGrp>**
   **<pos>**n**</pos>**
  **</gramGrp>**
  **<sense n="1">**
   **<def>**a drink, esp. an alcoholic one: we had a few bevvies last night.**</def>**
  **</sense>**
 **</hom>**
*<!-- ... sense 2 ... -->*
 **<hom>**
  **<gramGrp>**
   **<pos>**vb**</pos>**
  **</gramGrp>**
  **<sense n="3">**
   **<def>**to drink alcohol**</def>**
  **</sense>**
 **</hom>**
 **<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>**

* « [9.3 Top-level Constituents of Entries](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITP)
* » [9.5 Typographic and Lexical Information in Dictionary Data](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIMV)
* [Home](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html)

[**¶**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIHW)**9.4 Headword and Pronunciation References**

Examples, definitions, etymologies, and occasionally other elements such as cross-references, orthographic forms, etc., often contain a shortened or iconic reference to the headword, rather than repeating the headword itself. The references may be to the orthographic form or to the pronunciation, to the form given or to a variant of that form. The following elements are used to encode such iconic references to a headword:

* [**oRef**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-oRef.html) (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**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-pRef.html) (pronunciation reference) in a dictionary example, indicates a reference to the pronunciation(s) of the headword.

These elements all inherit the following attributes from the class [att.pointing](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.pointing.html) which may optionally be used to resolve any ambiguity about the headword form being referred to.

* [**att.pointing**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.pointing.html) provides a set of attributes used by all elements which point to other elements by means of one or more URI references.

|  |  |
| --- | --- |
| **target** | specifies the destination of the reference by supplying one or more URI References |

Headword references come in a variety of formats:

**~**

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

The [oRef](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-oRef.html) element should be used for iconic or shortened references to the orthographic form(s) of the headword itself. It is an empty element and replaces, rather than enclosing, the reference. Note that the reference to a headword is not necessarily a simple string replacement. In the example ‘**colour**1, (US = color) …~ films; ~ TV; Red, blue and yellow are ~s.’ OALD, the tilde stands for either headword form (*colour*, *color*).

Examples:

**colonel** … army officer above a lieutenant-~. OALD

**<def>**army officer above a lieutenant-**<oRef/>**
**</def>**

**academy** … The Royal A~ of Arts OALD

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

The following example demonstrates the use of the target attribute to refer to a specific form of the headword:

**vag-** or **vago-** comb form … : vagus nerve < **vag**al > < **vago**tomy > 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>**

In many cases the reference is not to the orthographic form of the headword, but rather to another form of the headword—usually to an inflected form. In these cases, the element [oRef](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-oRef.html) should be used; this element may take as its content the string as it appears in the text.

**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">**
   **<oRef/>**n**</oRef>** with him**</quote>**
**</cit>**

The next example shows a discontinuous reference, using the attributes next and prev, which are defined in the additional module for linking, segmentation, and alignment (see chapter [16 Linking, Segmentation, and Alignment](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/SA.html)) and therefore require that that module be selected in addition to that for dictionaries.

**mix up**… < it's easy to **mix** her **up** with her sister > NPEG

**<cit type="example">**
 **<quote>**it's easy to **<oRef next="#ov2" xml:id="ov1">**mix**</oRef>**
   her **<oRef prev="#ov1" xml:id="ov2">**up**</oRef>** with her sister**</quote>**
**</cit>**

In addition, some dictionaries make reference to the pronunciation of the headword in the pronunciation of related entries, variants, or examples. The [pRef](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-pRef.html) element should be used for such references.

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

**<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>**

Because headword and pronunciation references can occur virtually anywhere in an entry, the [oRef](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-oRef.html) and [pRef](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-pRef.html) elements may appear within any other element defined for dictionary entries.

Since existing printed dictionaries use different conventions for headword references (swung dash, first letter abbreviated form, capitalization, or italicization of the word, etc.) the exact method used should be documented in the header.

* « [9.4 Headword and Pronunciation References](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIHW)
* » [9.6 Unstructured Entries](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIFR)
* [Home](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html)

[**¶**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIMV)**9.5 Typographic and Lexical Information in Dictionary Data**

Among the many possible views of dictionaries, it is useful to distinguish at least the following three, which help to clarify some issues raised with particular urgency by dictionaries, on account of the complexity of both their typography and their information structure.

* (a) the *typographic view*—the two-dimensional printed page, including information about line and page breaks and other features of layout
* (b) the *editorial view*—the one-dimensional sequence of tokens which can be seen as the input to the typesetting process; the wording and punctuation of the text and the sequencing of items are visible in this view, but specifics of the typographic realization are not
* (c) the *lexical view*—this view includes the underlying information represented in a dictionary, without concern for its exact textual form

For example, a domain indication in a dictionary entry might be broken over a line and therefore hyphenated (‘naut-’ ‘ical’); the typographic view of the dictionary preserves this information. In a purely editorial view, the particular form in which the domain name is given in the particular dictionary (as ‘nautical’, rather than ‘naut.’, ‘Naut.’, etc.) would be preserved, but the fact of the line break would not. Font shifts might plausibly be included in either a strictly typographic or an editorial view. In the lexical view, the only information preserved concerning domain would be some standard symbol or string representing the nautical domain (e.g. ‘naut.’) regardless of the form in which it appears in the printed dictionary.

In practice, publishers begin with the lexical view—i.e., lexical data as it might appear in a database—and generate first the editorial view, which reflects editorial choices for a particular dictionary (such as the use of the abbreviation ‘Naut.’ for ‘nautical’, the fonts in which different types of information are to be rendered, etc.), and then the typographic view, which is tied to a specific printed rendering. Computational linguists and philologists often begin with the typographic view and analyse it to obtain the editorial and/or lexical views. Some users may ultimately be concerned with retaining only the lexical view, or they may wish to preserve the typographic or editorial views as a reference text, perhaps as a guard against the loss or misinterpretation of information in the translation process. Some researchers may wish to retain all three views, and study their interrelations, since research questions may well span all three views.

In general, an electronic encoding of a text will allow the recovery of at least one view of that text (the one which guided the encoding); if editorial and typographic practices are consistently applied in the production of a printed dictionary, or if exceptions to the rules are consistently recorded in the electronic encoding, then it is *in principle* possible to recover the editorial view from an encoding of the lexical view, and the typographic view from an encoding of the editorial view. In practice, of course, the severe compression of information in dictionaries, the variety of methods by which this compression is achieved, the complexity of formulating completely explicit rules for editorial and typographic practice, and the relative rarity of complete consistency in the application of such rules, all make the mechanical transformation of information from one view into another something of a vexed question.

This section describes some principles which may be useful in capturing one or the other of these views as consistently and completely as possible, and describes some methods of attempting to capture more than one view in a single encoding. Only the editorial and lexical views are explicitly treated here; for methods of recording the physical or typographic details of a text, see chapter [11 Representation of Primary Sources](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/PH.html). Other approaches to these problems, such as the use of repetitive encoding and links to show their correspondences, or the use of feature structures to capture the information structure, and of the ana and inst attributes to link feature structures to a transcription of the editorial view of a dictionary, are not discussed here (for feature structures, see chapter [18 Feature Structures](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/FS.html). For linkage of textual form and underlying information, see chapter [17 Simple Analytic Mechanisms](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/AI.html)).

* » [9.5.2 Lexical View](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIMVLV)
* [Home](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html)

[**¶**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIMVTV)**9.5.1 Editorial View**

Common practice in encoding texts of all sorts relies on principles such as the following, which can be used successfully to capture the editorial view when encoding a dictionary:

1. All characters of the source text should be retained, with the possible exception of *rendition text* (for which see further below).
2. Characters appearing in the source text should typically be given as character data content in the document, rather than as the value of an attribute; again, rendition text may optionally be excepted from this rule.
3. Apart from the characters or graphics in the source text, nothing else should appear as content in the document, although it may be given in attribute values.
4. The material in the source text should appear in the encoding in the same order. Complications of the character sequence by footnotes, marginal notes, etc., text wrapping around illustrations, etc., may be dealt with by the usual means (for notes, see section [3.8 Notes, Annotation, and Indexing](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CO.html#CONO)).37

In a very conservative transcription of the editorial view of a text, *rendition characters* (e.g. the commas, parentheses, etc., used in dictionary entries to signal boundaries among parts of the entry) and *rendition text* (for example, conjunctions joining alternate headwords, etc.) are typically retained. Removing the tags from such a transcription will leave all and only the characters of the source text, in their original sequence.38

Consider, for example, the following entry:

**pinna** (ˈpɪnə) *n*, *pl* **-nae** (-ni:) *or* **-nas** **1** any leaflet of a pinnate compound leaf *2* *zoology* a feather, wing, fin, or similarly shaped part **3** another name for **auricle** (2). [C18: via New Latin from Latin: wing, feather, fin] CED

A conservative encoding of the editorial view of this entry, which retains all rendition text, might resemble the following:

**<entry>**
 **<form>**
  **<orth>**pinna**</orth>**
  **<pron notation="ipa">**ˈpɪnə**</pron>**
 **</form>**
 **<gramGrp>**
  **<pos>**n**</pos>**, **</gramGrp>**
 **<form type="inflected">**
  **<number>**pl**</number>**
  **<form>**
   **<orth type="lat" extent="part">**-nae**</orth>**
   **<pron extent="part">**(-ni:)**</pron>**
  **</form>** or **<orth type="std" extent="part">**-nas**</orth>**
 **</form>**
 **<sense n="1">**1 **<def>**any leaflet of a pinnate compound leaf**</def>**
 **</sense>**
 **<sense n="2">**2 **<usg type="dom">**zoology**</usg>**
  **<def>**a feather, wing, fin, or similarly shaped part**</def>**
 **</sense>**
 **<sense n="3">**3 **<xr type="syn">**
   **<lbl>**another name for**</lbl>**
   **<ref target="#auricle.2">**auricle (2)**</ref>**
  **</xr>**
 **</sense>**
 **<etym>**[**<date>**C18**</date>**: via **<lang>**New Latin**</lang>** from **<lang>**Latin**</lang>**:
 **<gloss>**wing**</gloss>**, **<gloss>**feather**</gloss>**,
 **<gloss>**fin**</gloss>**]**</etym>**
**</entry>**
**<entry xml:id="auricle.2">**
 **<form>**
*<!-- ... -->*
 **</form>**
**</entry>**

A somewhat simplified encoding of the editorial view of this entry might exploit the fact that rendition text is often systematically recoverable. For example, parentheses consistently appear around pronunciation in this dictionary, and thus are effectively implied by the start- and end-tags for [pron](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-pron.html).39 In such an encoding, removing the tags should exactly reproduce the sequence of characters in the source, minus rendition text. The original character sequence can be recovered fully by replacing tags with any rendition text they imply.

Encoding in this way, the example given above might resemble the following. The [tagUsage](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-tagUsage.html) element in the header would be used to record the following patterns of rendition text:

* parentheses appear around [pron](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-pron.html) elements
* commas appear before inflected forms
* the word ‘or’ appears before alternate forms
* brackets appear around the etymology
* full stops appear after [pos](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-pos.html), inflection information, and sense numbers
* senses are numbered in sequence unless otherwise specified using the global n attribute

**<entry>**
 **<form>**
  **<orth>**pinna**</orth>**
  **<pron>**"pIn@**</pron>**
 **</form>**
 **<gramGrp>**
  **<pos>**n**</pos>**
 **</gramGrp>**
 **<form type="inflected">**
  **<number>**pl**</number>**
  **<form>**
   **<orth type="lat" extent="part">**-nae**</orth>**
   **<pron extent="part">**-ni:**</pron>**
  **</form>**
  **<orth type="std" extent="part">**-nas**</orth>**
 **</form>**
 **<sense n="1">**
  **<def>**any leaflet of a pinnate compound leaf.**</def>**
 **</sense>**
 **<sense n="2">**
  **<usg type="dom">**Zoology**</usg>**
  **<def>**a feather, wing, fin, or similarly shaped part.**</def>**
 **</sense>**
 **<sense n="3">**
  **<xr type="syn">**
   **<lbl>**another name for**</lbl>**
   **<ref>**auricle (sense 2).**</ref>**
  **</xr>**
 **</sense>**
 **<etym>**
  **<date>**C18**</date>**: via **<lang>**New Latin**</lang>** from **<lang>**Latin**</lang>**:
 **<gloss>**wing**</gloss>**, **<gloss>**feather**</gloss>**, **<gloss>**fin**</gloss>**
 **</etym>**
**</entry>**

When rendition text is omitted, it is recommended that the means to regenerate it be fully documented, using the [tagUsage](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-tagUsage.html) element of the TEI header.

If rendition text is used systematically in a dictionary, with only a few mistakes or exceptions, the global attribute rend may be used on any tag to flag exceptions to the normal treatment. The values of the rend attribute are not prescribed, but it can be used with values such as no-comma, no-left-paren, etc. Specific values can be documented using the [rendition](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-rendition.html) element in the TEI header.

In the following (imaginary) example, no left parenthesis precedes the pronunciation:

**biryani** or **biriani** %bIrI"A:nI) any of a variety of Indian dishes … [from Urdu]

This irregularity can be recorded thus:

**<entry>**
 **<form>**
  **<orth>**biryani**</orth>**
  **<orth>**biriani**</orth>**
  **<pron rend="noleftparen">**%bIrI"A:nI**</pron>**
 **</form>**
 **<def>**any of a variety of Indian dishes … **</def>**
 **<etym>**from **<lang>**Urdu**</lang>**
 **</etym>**
**</entry>**

* « [9.5.1 Editorial View](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIMVTV)
* » [9.5.3 Retaining Both Views](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIMVBO)
* [Home](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html)

[**¶**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIMVLV)**9.5.2 Lexical View**

If the text to be interchanged retains only the lexical view of the text, there may be no concern for the recoverability of the editorial (not to speak of the typographic) view of the text. However, it is strongly recommended that the TEI header be used to document fully the nature of all alterations to the original data, such as normalization of domain names, expansion of inflected forms, etc.

In an encoding of the lexical view of a text, there are degrees of departure from the original data: normalizing inconsistent forms like ‘nautical’, ‘naut’., ‘Naut.’, etc., to ‘nautical’ is a relatively slight alteration; expansion of ‘delay -ed -ing’ to ‘delay, delayed, delaying’ is a more substantial departure. Still more severe is the rearranging of the order of information in entries; for example:

* reorganizing the order of elements in an entry to show their relationship, as in

**clem** **(klɛm)** *or* **clam** *vb* **clems, clemming, clemmed** *or* **clams, clamming, clammed** CED

where in a strictly lexical view one might wish to group ‘clem’ and ‘clam’ with their respective inflected forms.

* splitting an entry into two separate entries, as in

**celi.bacy** /"selIb@sI/ n [U] state of living unmarried, esp as a religious obligation. celi.bate /"selIb@t/ n [C] unmarried person (esp a priest who has taken a vow not to marry). OALD

For some purposes, this entry might usefully be split into an entry for ‘celibacy’ and a separate entry for ‘celibate’.

An encoding which captures the lexical view of the example given in the previous section might look something like the following. In this encoding:

* abbreviated forms have been silently expanded
* some forms have been moved to allow related forms to be grouped together
* the part of speech information has been moved to allow all forms to be given together
* the cross-reference to ‘auricle’ has been simplified

**<entry>**
 **<form>**
  **<orth>**pinna**</orth>**
  **<pron>**"pIn@**</pron>**
  **<form type="inflected">**
   **<number>**pl**</number>**
   **<form>**
    **<orth type="lat">**pinnae**</orth>**
    **<pron>**'pIni:**</pron>**
   **</form>**
   **<orth type="std">**pinnas**</orth>**
  **</form>**
 **</form>**
 **<gramGrp>**
  **<pos>**n**</pos>**
 **</gramGrp>**
 **<sense n="1">**
  **<def>**any leaflet of a pinnate compound leaf.**</def>**
 **</sense>**
 **<sense n="2">**
  **<usg type="dom">**Zoology**</usg>**
  **<def>**a feather, wing, fin, or similarly shaped part.**</def>**
 **</sense>**
 **<sense n="3">**
  **<xr type="syn">**
   **<ptr target="#auricle.2"/>**
  **</xr>**
 **</sense>**
 **<etym>**
  **<date>**C18**</date>**: via **<lang>**New Latin**</lang>** from **<lang>**Latin**</lang>**:
 **<gloss>**wing**</gloss>**, **<gloss>**feather**</gloss>**, **<gloss>**fin**</gloss>**
 **</etym>**
**</entry>**

Whether the given dictionary encoding focusses on the lexical view and thus approaches the status of lexical databases, or uses the typographic/editorial view approach and needs to communicate the sometimes informally stated values for the particular descriptive features, the issue of ‘interoperability’ of the content and of the container objects becomes relevant, in view of the growing tendency to interlink pieces of information across Internet resources. In such cases, it becomes crucial to be able to encode the fact that whether the information on, for instance, the value of the grammatical category of Number is provided as "sg.", "sing.", "Singular", or equivalently "poj." in Polish, or "Ez." in German, etc., what is actually referred to is always the same grammatical value that can be rendered with a plethora of markers, depending on the publisher, language, or lexicographic tradition. In order to signal that this variety of surface markers in fact indicate the same underlying value, it is possible to align them with an external inventory of standardized values. The TEI provides means to align grammatical categories as well as their content with the ISOcat reference, which is a Web implementation of ISO 12620.

In the example below, a fragment of the entry for *isotope* cited in section [9.3.2 Grammatical Information](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPGR) is adorned by references to ISOcat definitions for "part of speech" (dcr:datcat) and "adjective" (dcr:valueDatcat). Depending on the status and extent of the dictionary, various strategies may be used to reduce the redundancy of the repeated ISOcat references.

**<entry>**
*<!--...-->*
 **<form>**
  **<orth>**isotope**</orth>**
 **</form>**
 **<gramGrp>**
  **<pos dcr:datcat="http://www.isocat.org/datcat/DC-1345"
   dcr:valueDatcat="http://www.isocat.org/datcat/DC-1230">**adj**</pos>**
 **</gramGrp>**
*<!--...-->*
**</entry>**

* « [9.5.2 Lexical View](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIMVLV)
* [Home](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html)

[**¶**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIMVBO)**9.5.3 Retaining Both Views**

It is sometimes desirable to retain both the lexical and the editorial view, in which case a potential conflict exists between the two. When there is a conflict between the encodings for the lexical and editorial views, the principles described in the following sections may be applied.

* » [9.5.3.2 Recording Original Locations of Transposed Elements](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIMVOL)
* [Home](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html)

[**¶**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIMVAV)**9.5.3.1 Using Attribute Values to Capture Alternate Views**

If the order of the data is the same in both views, then both views may be captured by encoding one ‘dominant’ view in the character data content of the document, and encoding the other using attribute values on the appropriate elements. If all tags were to be removed, the remaining characters would be those of the dominant view of the text.

The attribute class [att.lexicographic](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.lexicographic.html) is used to provide attributes for use in encoding multiple views of the same dictionary entry. These attributes are available for use on all elements defined in this chapter when the base module for dictionaries is selected.

When the editorial view is dominant, the following attributes may be used to capture the lexical view:

* [**att.lexicographic**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.lexicographic.html) provides a set of attributes common to all elements in the dictionary module.

|  |  |
| --- | --- |
| **norm** | (normalized) gives a normalized form of information given by the source text in a non-normalized form |
| **split** | gives the list of split values for a merged form |

When the lexical view is dominant, the following attributes may be used to record the editorial view:

* [**att.lexicographic**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.lexicographic.html) provides a set of attributes common to all elements in the dictionary module.

|  |  |
| --- | --- |
| **orig** | (original) gives the original string or is the empty string when the element does not appear in the source text. |
| **mergedIn** | gives a reference to another element, where the original appears as a merged form. |

One attribute is useful in either view:

* [**att.lexicographic**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.lexicographic.html) provides a set of attributes common to all elements in the dictionary module.

|  |  |
| --- | --- |
| **opt** | (optional) indicates whether the element is optional or not |

For example, if the source text had the domain label ‘naut.’, it might be encoded as follows. With the editorial view dominant:

**<usg norm="nautical" type="dom">**naut.**</usg>**

The lexical view of the same label would transcribe the normalized form as content of the [usg](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-usg.html) element, the typographic form as an attribute value:

**<usg orig="naut." type="dom">**nautical**</usg>**

If the source text gives inflectional information for the verb *delay* as ‘delay, -ed, -ing’, it might usefully be expanded to ‘delayed, delayed, delaying’. An encoding of the editorial view might take this form:

**<form>**
 **<orth>**delay**</orth>**
 **<form type="inflected">**
  **<orth norm="delayed" extent="part">**-ed**</orth>**
  **<tns norm="pst,pstp"/>**
 **</form>**
 **<form type="inflected">**
  **<orth norm="delaying" extent="part">**-ing**</orth>**
  **<tns norm="prsp"/>**
 **</form>**
**</form>**

Note the use of the [tns](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-tns.html) tag with null content, to enable the representation of implicit information even though it has no print realization.

The lexical view might be encoded thus:

**<form>**
 **<orth>**delay**</orth>**
 **<form type="inflected">**
  **<orth orig="-ed">**delayed**</orth>**
  **<tns orig="">**pst**</tns>**
  **<tns orig="">**pstp**</tns>**
 **</form>**
 **<form type="inflected">**
  **<orth orig="-ing">**delaying**</orth>**
  **<tns orig="">**prsp**</tns>**
 **</form>**
**</form>**

A particular problem may be posed by the common practice of presenting two alternate forms of a word in a single string, by marking some parts of the word as optional in some forms. The following entry is for a word which can be spelled either ‘thyrostimuline’ or ‘thyréostimuline’:

**thyr(é)ostimuline** [tiR(e)ostimylin] …

With the editorial view dominant, this entry might begin thus:

**<form>**
 **<orth split="thyrostimuline, thyréostimuline">**thyr(é)ostimuline**</orth>**
 **<pron split="tiRostimylin, tiReostimylin">**tiR(e)ostimylin**</pron>**
**</form>**

With the lexical view dominant, however, two [orth](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-orth.html) and two [pron](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-pron.html) elements would be encoded, in order to disentangle the two forms; the orig attribute would be used to record the typographic presentation of the information in the source.

**<form>**
 **<orth xml:id="dic-o1"
  orig="thyr(é)ostimuline">**thyrostimuline**</orth>**
 **<pron xml:id="dic-p1"
  orig="tiR(e)ostimylin">**tiRostimylin**</pron>**
**</form>**
**<form>**
 **<orth mergedIn="#dic-o1">**thyréostimuline**</orth>**
 **<pron mergedIn="#dic-p1">**tiReostimylin**</pron>**
**</form>**

This example might also be encoded using the opt attribute combined with the attributes next and prev defined in chapter [16 Linking, Segmentation, and Alignment](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/SA.html).

**<form>**
 **<orth next="#dict-o2" xml:id="dict-o1">**thyr**</orth>**
 **<orth next="#dict-o3" prev="#dict-o1"
  xml:id="dict-o2" opt="true">**é**</orth>**
 **<orth prev="#dict-o2" xml:id="dict-o3">**ostimuline**</orth>**
 **<pron next="#dict-p2" xml:id="dict-p1">**tiR**</pron>**
 **<pron next="#dict-p3" prev="#dict-p1"
  xml:id="dict-p2" opt="true">**e**</pron>**
 **<pron prev="#dict-p2" xml:id="dict-p3">**ostimylin**</pron>**
**</form>**

Note that this transcription preserves both the lexical and editorial views in a single encoding. However, it has the disadvantage that the strings corresponding to entire words do not appear in the encoding uninterrupted, and therefore complex processing is required to retrieve them from the encoded text. The use of the opt attribute is recommended, however, when long spans of text are involved, or when the optional part contains embedded tags.

For example, the following gives two definitions in one text: ‘picture drawn with coloured chalk made into crayons’, and ‘coloured chalk made into crayons’:

**pas.tel** /"pastl US: pa"stel/ n 1 (picture drawn with) coloured chalk made into crayons. 2… OALD

A simple encoding solution would be to leave the definition text unanalysed, but this might be felt inadequate since it does not show that there are two definitions. A possible alternative encoding would be:

**<sense n="1">**
 **<def>**coloured
   chalk made into crayons**</def>**
 **<def>**picture drawn with coloured chalk
   made into crayons**</def>**
**</sense>**

This transcribes some characters of the source text twice, however, which deviates from the usual practice. The following encoding records both the editorial and lexical views:

**<sense n="1">**
 **<def next="#d2" xml:id="d1" opt="true">**picture drawn
   with**</def>**
 **<def prev="#d1" xml:id="d2">**coloured chalk made into
   crayons**</def>**
**</sense>**

* « [9.5.3.1 Using Attribute Values to Capture Alternate Views](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIMVAV)
* [Home](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html)

[**¶**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIMVOL)**9.5.3.2 Recording Original Locations of Transposed Elements**

The attributes described in the previous section are useful only when the order of material is the same in both the editorial and the lexical view. When the two views impose different orders on the data, the standard linking mechanisms may be used to show the original location of material transposed in an encoding of the lexical view.

If the original is only slightly modified, the [anchor](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-anchor.html) element may be used to mark the original location of the material, and the location attribute may be used on the lexical encoding of that material to indicate its original location(s). Like those in the preceding section, this attribute is defined for the attribute class [att.lexicographic](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.lexicographic.html):

* [**att.lexicographic**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.lexicographic.html) provides a set of attributes common to all elements in the dictionary module.

|  |  |
| --- | --- |
| **opt** | (optional) indicates whether the element is optional or not |

For example:

**pinna** **(ˈpɪnə)** *n*, *pl* **-nae** (-ni:) *or* **-nas** CED

**<form>**
 **<orth>**pinna**</orth>**
 **<pron notation="ipa">**ˈpɪnə**</pron>**
 **<anchor xml:id="p01"/>**
 **<form type="inflected">**
  **<number>**pl**</number>**
  **<form>**
   **<orth extent="part">**-nae**</orth>**
   **<pron extent="part">**-ni:**</pron>**
  **</form>**
  **<orth extent="part">**-nas**</orth>**
 **</form>**
**</form>**
**<gramGrp>**
 **<pos location="#p01">**n**</pos>**
**</gramGrp>**

* « [9.5 Typographic and Lexical Information in Dictionary Data](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIMV)
* » [9.7 The Dictionary Module](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#index-body.1_div.9_div.7)
* [Home](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html)

[**¶**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIFR)**9.6 Unstructured Entries**

The content model for the [entry](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-entry.html) element provides an entry structure suitable for many average dictionaries, as well as many regular entries in more exotic dictionaries. However, the structure of some dictionaries does not allow the restrictions imposed by the content model for [entry](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-entry.html). To handle these cases, the [entryFree](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-entryFree.html) and [dictScrap](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-dictScrap.html) elements are provided to support much wider variation in entry structure. The [dictScrap](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-dictScrap.html) element offers less freedom, in that it can only contain phrase level elements, but it can itself appear at any point within a dictionary entry where any of the structural components of a dictionary entry are permitted. As such, it acts as a container for otherwise anomalous parts of an entry.

The [entryFree](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-entryFree.html) element places no constraints at all upon the entry: any element defined in this chapter, as well as all the normal phrase-level and inter-level elements, can appear anywhere within it. With the [entryFree](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-entryFree.html) element, the encoder is free to use any element anywhere, as well as to use or omit grouping elements such as [form](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-form.html), [gramGrp](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-gramGrp.html), etc.

The [entryFree](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-entryFree.html) element allows the encoding of entries which violate the structure specified for the [entry](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-entry.html) element. For example, in the following entry from a dictionary already in electronic form, it is necessary to include a [pron](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-pron.html) element within a [def](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-def.html). This is not permitted in the content model for [entry](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-entry.html), but it poses no problem in the [entryFree](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-entryFree.html) element.

<ent

h="demigod"> <hwd>demi|god</hwd> <pr> <ph>"demIgQd</ph> </pr> <hps

ps="n"> <hsn> <def>one who is partly divine and partly human</def>

<def>(in Gk myth, etc) the son of a god and a mortal woman,

eg<cf>Hercules</cf> <pr> <ph>"h3:kjUli:z</ph> </pr> </def> </hsn>

</hps> </ent>

(OALD)

**<entryFree>**
 **<form>**
  **<orth>**demigod**</orth>**
  **<hyph>**demi|god**</hyph>**
  **<pron>**"demIgQd**</pron>**
 **</form>**
 **<gramGrp>**
  **<pos>**n**</pos>**
 **</gramGrp>**
 **<def>**one who is partly divine and partly human**</def>**
 **<def>**(in Gk myth, etc) the son of a god and a mortal woman, eg
 **<mentioned>**Hercules**</mentioned>**
 **</def>**
 **<pron>**"h3:kjUli:z**</pron>**
**</entryFree>**

The [entryFree](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-entryFree.html) element also makes it possible to transcribe a dictionary using only phrase-level (‘atomic’) elements—that is, using no grouping elements at all. This can be desirable if the encoder wants a completely ‘flat’ view, with no indication of or commitment to the association of one element with another. The following encoding uses no grouping elements, and keeps all rendition text:

**biryani** *or* **biriani** **(ˌbɪrɪˈa:nɪ)** *n* any of a variety of Indian dishes … [from Urdu] CED

**<entryFree>**
 **<orth>**biryani**</orth>** or **<orth>**biriani**</orth>**
 **<pron notation="ipa">**(ˌbɪrɪˈa:nɪ)**</pron>**
 **<def>**any of a variety of Indian dishes …**</def>**
 **<etym>**[from **<lang>**Urdu**</lang>**]**</etym>**
**</entryFree>**

Here is an alternative way of representing the same structure, this time using [dictScrap](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-dictScrap.html):

**<entry>**
 **<dictScrap>**
  **<orth>**biryani**</orth>** or **<orth>**biriani**</orth>**
  **<pron notation="ipa">**(ˌbɪrɪˈa:nɪ)**</pron>**
  **<def>**any of a variety of Indian dishes …**</def>**
  **<etym>**[from **<lang>**Urdu**</lang>**]**</etym>**
 **</dictScrap>**
**</entry>**

* « [9.6 Unstructured Entries](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DIFR)
* [Home](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html)

[**¶**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#index-body.1_div.9_div.7)**9.7 The Dictionary Module**

The module defined in this chapter makes available the following components:

**Module dictionaries: Dictionaries**

* Elements defined: [case](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-case.html) [colloc](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-colloc.html) [def](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-def.html) [dictScrap](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-dictScrap.html) [entry](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-entry.html) [entryFree](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-entryFree.html) [etym](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-etym.html) [form](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-form.html) [gen](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-gen.html) [gram](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-gram.html) [gramGrp](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-gramGrp.html) [hom](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-hom.html) [hyph](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-hyph.html) [iType](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-iType.html) [lang](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-lang.html) [lbl](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-lbl.html) [mood](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-mood.html) [number](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-number.html) [oRef](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-oRef.html) [orth](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-orth.html) [pRef](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-pRef.html) [per](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-per.html) [pos](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-pos.html) [pron](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-pron.html) [re](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-re.html) [sense](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-sense.html) [stress](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-stress.html) [subc](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-subc.html) [superEntry](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-superEntry.html) [syll](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-syll.html) [tns](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-tns.html) [usg](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-usg.html) [xr](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-xr.html)
* Classes defined: [att.entryLike](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.entryLike.html) [att.lexicographic](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.lexicographic.html) [model.entryLike](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-model.entryLike.html) [model.formPart](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-model.formPart.html) [model.gramPart](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-model.gramPart.html) [model.lexicalRefinement](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-model.lexicalRefinement.html) [model.morphLike](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-model.morphLike.html) [model.ptrLike.form](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-model.ptrLike.form.html)

The selection and combination of modules to form a TEI schema is described in [1.2 Defining a TEI Schema](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ST.html#STIN).

↑ [TEI P5 Guidelines](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html) « [8 Transcriptions of Speech](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/TS.html) » [10 Manuscript Description](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/MS.html)

**Notes**

36

We refer the reader to previous and current discussions of a common format for encoding lexical resources. For example, Amsler and Tompa (1988); Calzolari et al. (1990);Fought and Van Ess-Dykema; Ide and Veronis (1995); Ide et al. (1993); Ide et al. (1992); DANLEX Group (1987); and Tutin and Veronis (1998); Ide et al. (2000).

 [**↵**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#Note72_return)

37

Complications of sequence caused by marginal or interlinear insertions and deletions, which are frequent in manuscripts, or by unconventional page layouts, as in concrete poetry, magazines with imaginative graphic designers, and texts about the nature of typography as a medium, typically do not occur in dictionaries, and so are not discussed here.

 [**↵**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#Note73_return)

38

This is a slight oversimplification. Even in conservative transcriptions, it is common to omit page numbers, signatures of gatherings, running titles and the like. The simple description above also elides, for the sake of simplicity, the difficulties of assigning a meaning to the phrase ‘original sequence’ when it is applied to the printed characters of a source text; the ‘original sequence’ retained or recovered from a conservative transcription of the editorial view is, of course, the one established during the transcription by the encoder.

 [**↵**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#Note74_return)

39

The omission of rendition text is particularly common in systems for document production; it is considered good practice there, since automatic generation of rendition text is more reliable and more consistent than attempting to maintain it manually in the electronic text.

 [**↵**](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#Note75_return)

[[English](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html)] [[Deutsch](http://www.tei-c.org/release/doc/tei-p5-doc/de/html/DI.html)] [[Español](http://www.tei-c.org/release/doc/tei-p5-doc/es/html/DI.html)] [[Italiano](http://www.tei-c.org/release/doc/tei-p5-doc/it/html/DI.html)] [[Français](http://www.tei-c.org/release/doc/tei-p5-doc/fr/html/DI.html)] [[日本語](http://www.tei-c.org/release/doc/tei-p5-doc/ja/html/DI.html)] [[한국어](http://www.tei-c.org/release/doc/tei-p5-doc/ko/html/DI.html)] [[中文](http://www.tei-c.org/release/doc/tei-p5-doc/zh-TW/html/DI.html)]

[**TEI Consortium**](http://www.tei-c.org/About/) | [**Feedback**](http://www.tei-c.org/About/contact.xml)

 *TEI Guidelines*[*Version*](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/AB.html#ABTEI4)[*3.4.0*](http://www.tei-c.org/release/doc/tei-p5-doc/readme-3.4.0.html)*. Last updated on 23rd July 2018, revision*[*1fa0b54*](https://github.com/TEIC/TEI/commit/1fa0b54)*. This page generated on 2018-07-23T09:39:25Z.*