



european lexicographic  
infrastructure

# Intro to Dictionary Encoding in TEI



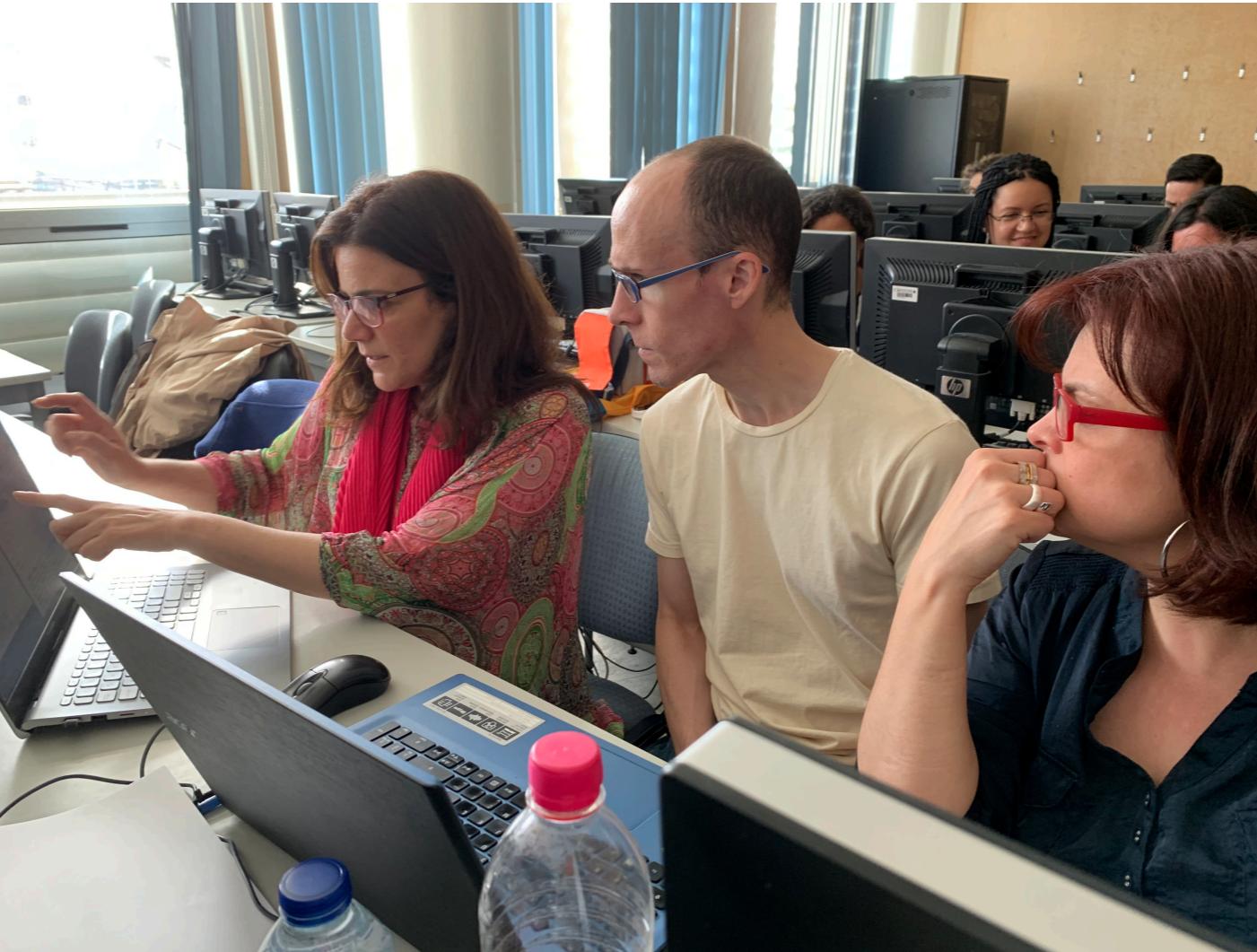
**TEI is one of the  
cornerstones of Digital  
Humanities practices**

«The Text Encoding Initiative (TEI) is a consortium which collectively develops and maintains a standard for the **representation of texts in digital form**. Its main deliverable is a set of **Guidelines** which specify **encoding methods for machine-readable texts**, chiefly in the humanities, social sciences and linguistics. The TEI Guidelines have become an accepted standard for digital text especially where there are concerns about **long-term preservation, interchange, or interoperability**.»

<div xml:id="DOB.ZP.sl.1" corresp="DOB.ZP.sr.1" n="1" type="chapter">  
    <pb n="1" facs="S-II-0832-010.jpg"/>  
    <head>Предисловіе</head>  
    <p xml:id="DOB.ZP.sl.1.1" corresp="DOB.ZP.sr.1.1">  
        <s xml:id="DOB.ZP.sl.1.1.1" corresp="DOB.ZP.sr.1.1.1">  
            <w xml:id="DOB.ZP.sl.1.1.1.1" corresp="DOB.ZP.sr.1.1.1.1">Сладка</w>  
            <w xml:id="DOB.ZP.sl.1.1.1.2" corresp="DOB.ZP.sr.1.1.1.2">е</w>  
            <w xml:id="DOB.ZP.sl.1.1.1.3" corresp="DOB.ZP.sr.1.1.1.3">стваръ</w>, <w  
                xml:id="DOB.ZP.sl.1.1.1.4" corresp="DOB.ZP.sr.1.1.1.4">и</w>  
            <w xml:id="DOB.ZP.sl.1.1.1.5" corresp="DOB.ZP.sr.1.1.1.5">пуна</w>  
            <w xml:id="DOB.ZP.sl.1.1.1.6" corresp="DOB.ZP.sr.1.1.1.6">ї</w>

# Why TEI when we already have XML?

- XML is a language "without words"
- XML is a method for building descriptive languages
- Extensibility: virtue or vice?



Ana, Juan and Raquel

# How to encode "Ana, Juan and Raquel"?

- <person>Ana</person>, <person>Juan</person> and <person>Raquel</person>
- <people>Ana, Juan and Raquel</people>
- <nonAlien>Ana</nonAlien>, <nonAlien>Juan</nonAlien> and <nonAlien>Raquel</nonAlien>
- <lexicographer>Ana</lexicographer>, <terminologist>Juan</terminologist> and <terminologist>Raquel</terminologist>
- <theFabulousThree>Ana, Juan and Raquel</theFabulousThree>

# Standardization

The chief deliverable of the Text Encoding Initiative (TEI) is a set of **Guidelines** which specify **encoding methods** for machine-readable texts in the humanities, social sciences and linguistics.

- <http://www.tei-c.org/>

<div xml:id="DOB.ZP.sl.1" corresp="DOB.ZP.sr.1" n="1" type="chapter">  
  <pb n="1" facs="S-II-0832-010.jpg"/>  
  <head>Предисловіє</head>  
  <p xml:id="DOB.ZP.sl.1.1" corresp="DOB.ZP.sr.1.1">  
    <s xml:id="DOB.ZP.sl.1.1.1" corresp="DOB.ZP.sr.1.1.1">  
      <w xml:id="DOB.ZP.sl.1.1.1.1" corresp="DOB.ZP.sr.1.1.1.1">Сладка</w>  
      <w xml:id="DOB.ZP.sl.1.1.1.2" corresp="DOB.ZP.sr.1.1.1.2">е</w>  
      <w xml:id="DOB.ZP.sl.1.1.1.3" corresp="DOB.ZP.sr.1.1.1.3">стваръ</w>, <w  
        xml:id="DOB.ZP.sl.1.1.1.4" corresp="DOB.ZP.sr.1.1.1.4">и</w>  
      <w xml:id="DOB.ZP.sl.1.1.1.5" corresp="DOB.ZP.sr.1.1.1.5">пуна</w>  
    </s>  
  </p>

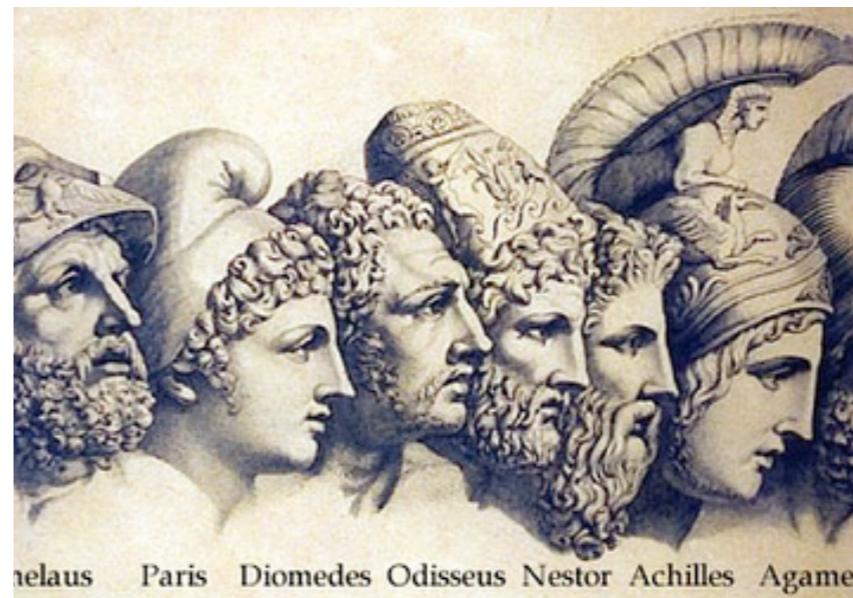
# What does TEI give you?

- **Vocabulary:** elements and attributes to describe certain feature
- **Grammar:** a schema which tells you how certain elements are combined

# Why do we need markup?

**Because language is  
both complex and  
ambiguous**

# He loves Paris.



- <placeName>Paris</placeName>
- <persName role="starlet">Paris</persName>
- <persName type="historical" role="troublemaker">Paris</persName>

# What's the point of markup?

- To make explicit (to a machine) what is implicit (to a person)
- To add value by supplying multiple (sometimes competing) annotations
- To facilitate re-use of the same material
  - in different formats
  - in different contexts
  - by different users

# So, TEI...

- makes explicit the distinctions **we want to make** when processing a string of bytes
- is a way of naming and characterizing parts of a text in a formalized, predictable way
- is (usually) more useful for marking up what we think things *are* than what they *look like*

# Separation of form and content in dictionaries

- Presentational markup cares more about fonts and layout than meaning
- Descriptive markup says what things are, and leaves the rendition of them for a separate step
- Separating the form of something from its content makes its re-use more flexible
- It also allows easy changes of presentation across a large number of documents

# TEI *is* legible

```
<entryFree xml:id="MZ.RGJS.катаруга">
    <form type="lemma">
        <orth>катарўга</orth>
    </form>
    <gramGrp>
        <pos>ж</pos>
    </gramGrp>
    <sense>
        <def>већа дрвена кука.</def>
        <cit type="example">
            <quote>„Ћерко, окачи кавдарé на катарўгу”</quote>
            <bibl>(<placeName>Горња Трница</placeName>).</bibl>
        </cit>
    </sense>
</entryFree>
```

# TEI is useful not just for describing but for studying dictionaries.

```
<entryFree xml:id="VSK.SR.авајлија" sortKey="авајлија">
    <form type="lemma"><orth>авајлија</orth></form>
    <etym source="#thirdEd">
        <lang value="турцизам">*</lang>
    </etym>
    <gramGrp><pos>f</pos></gramGrp>
    <sense xml:id="d1e1219">(<usg type="diatopic">y <rs type="region"
        ref="references/places.xml#uzickanahija">Уж. Н.</rs></usg>) некака
        јабука, <cit type="translation">
            <quote xml:lang="de">Art [Winter-]Apfel</quote>
        </cit><pc>, </pc>
        <cit type="translation">
            <quote xml:lang="la">mali genus [pirus malus L. var.]</quote>
        </cit><pc>. </pc></sense></entryFree>
```



# With a little help from your friends (XPath/XQuery etc.)

```
//cit[@type="translation"]/quote[@xml:lang="de"]  
[starts-with(., "Art") or starts-with(., "eine Art")]
```

**//cit[@type="translation"]/quote[@xml:lang="de"]**  
**[starts-with(., "Art") or starts-with(., "eine Art")]**

Description - 974 items	XPath location	Resource
Art [Winter-]Apfel	/TEI[1]/text[1]/body[1...]	VSK.SR.xml
Art Pökelrogen	/TEI[1]/text[1]/body[1...]	VSK.SR.xml
Art eines langen Oberkleides	/TEI[1]/text[1]/body[1...]	VSK.SR.xml
eine Art Kleid	/TEI[1]/text[1]/body[1...]	VSK.SR.xml
Art Kirschen	/TEI[1]/text[1]/body[1...]	VSK.SR.xml
eine Art Zwiebel	/TEI[1]/text[1]/body[1...]	VSK.SR.xml
Art Gebets der Türken	/TEI[1]/text[1]/body[1...]	VSK.SR.xml
eine Art harter, schwärzlicher Birnen	/TEI[1]/text[1]/body[1...]	VSK.SR.xml
eine Art Zwiebel zum Versetzen	/TEI[1]/text[1]/body[1...]	VSK.SR.xml
Art Frühkirsche [Herzkirsche]	/TEI[1]/text[1]/body[1...]	VSK.SR.xml
Art Zwiebel	/TEI[1]/text[1]/body[1...]	VSK.SR.xml
Art dehnenden Sprechens	/TEI[1]/text[1]/body[1...]	VSK.SR.xml
Art Frauenmütze	/TEI[1]/text[1]/body[1...]	VSK.SR.xml
Art Kirschen	/TEI[1]/text[1]/body[1...]	VSK.SR.xml
Art grösster Nüsse	/TEI[1]/text[1]/body[1...]	VSK.SR.xml
Art Vogel	/TEI[1]/text[1]/body[1...]	VSK.SR.xml
eine Art Pflanze	/TEI[1]/text[1]/body[1...]	VSK.SR.xml
eine Art Schwamm	/TEI[1]/text[1]/body[1...]	VSK.SR.xml
eine Art irdenen Ofens (ohne Kacheln)	/TEI[1]/text[1]/body[1...]	VSK.SR.xml
Art Erdäpfel (die aus Baiern gebracht sein	... /TEI[1]/text[1]/body[1...]	VSK.SR.xml
Art Pferdekrankheit [die Mauke]	/TEI[1]/text[1]/body[1...]	VSK.SR.xml
Art Getreidemass	/TEI[1]/text[1]/body[1...]	VSK.SR.xml
Art Meerfisch	/TEI[1]/text[1]/body[1...]	VSK.SR.xml
eine Art Pflanze	/TEI[1]/text[1]/body[1...]	VSK.SR.xml

Errors XPath - VSK.SR.xml X

```
//entryFree[descendant::cit[@type="translation"]/
quote[@xml:lang="de"]][starts-with(., "Art") or starts-
with(., "eine Art")]/form[@type="lemma"]/orth
```

ава́јлија	/TEI[1]/text[1]/body[1... VSK.SR.
авгу́та́р	/TEI[1]/text[1]/body[1... VSK.SR.
а́здија	/TEI[1]/text[1]/body[1... VSK.SR.
а́лет	/TEI[1]/text[1]/body[1... VSK.SR.
а́лица	/TEI[1]/text[1]/body[1... VSK.SR.
а́льма	/TEI[1]/text[1]/body[1... VSK.SR.
а́ма́з	/TEI[1]/text[1]/body[1... VSK.SR.
намаз	/TEI[1]/text[1]/body[1... VSK.SR.
А́рапка	/TEI[1]/text[1]/body[1... VSK.SR.
арпа́цик	/TEI[1]/text[1]/body[1... VSK.SR.
аршла́ма	/TEI[1]/text[1]/body[1... VSK.SR.
а́че́ње	/TEI[1]/text[1]/body[1... VSK.SR.
а́ша́њка	/TEI[1]/text[1]/body[1... VSK.SR.
ба́баје	/TEI[1]/text[1]/body[1... VSK.SR.
ба́бац	/TEI[1]/text[1]/body[1... VSK.SR.
ба́бин ко́ко́т	/TEI[1]/text[1]/body[1... VSK.SR.
ба́бини зу́би	/TEI[1]/text[1]/body[1... VSK.SR.
ба́бино у́хо	/TEI[1]/text[1]/body[1... VSK.SR.
ба́бура	/TEI[1]/text[1]/body[1... VSK.SR.
баве́рац	/TEI[1]/text[1]/body[1... VSK.SR.
ба́га	/TEI[1]/text[1]/body[1... VSK.SR.
ба́га́ш	/TEI[1]/text[1]/body[1... VSK.SR.
ба́гра	/TEI[1]/text[1]/body[1... VSK.SR.
ба́даљ	/TEI[1]/text[1]/body[1... VSK.SR.
ба́дањ	/TEI[1]/text[1]/body[1... VSK.SR.
ба́дељ	/TEI[1]/text[1]/body[1... VSK.SR.
ба́дњара	/TEI[1]/text[1]/body[1... VSK.SR.

Арапка

Press F2 for focus

# TEI *is* customizable

- using ODD you can create your own subset of TEI to use in a project
- you can add, change, alter... (which doesn't necessarily mean you should)
- TEI is localizable

<одредница>  
    <облик тип="основни">маѣла</облик> , <облик тип="наставак">е</облик>  
    <граматика>  
        <врстаречи>ж.</врстаречи>  
    </граматика> – <белешка> У В. <упућеница речник="вук">маѣла</упућеница> ж., у RJA.  
        <упућеница речник="RJA">mähala</упућеница> ж. </белешка>: <значење>  
        <пример>  
            <цитат> Отѣшла по <курзив>маѣле.</курзив>  
            </цитат>  
        </пример> – <пример>  
            <цитат> У <курзив>маѣлу</курзив> ми ваља ѫс да трâжим. </цитат>  
        </пример>  
    </значење>  
    <етимологија><језик>Арап.</језик>  
        <странареч>mahalle</странареч> им.</етимологија>  
    <значење>  
        <дефиниција> одељак вароши који обично има своју џамију и имама. </дефиниција>  
    </значење>  
</одредница>  
<одредница>

**TEI, however, is not what  
most end users will  
interact with**

# raskovnik.org

- Dictionary platform
- Institute of Serbian Language of the Serbian Academy of Arts and Sciences + Belgrade Center for Digital Humanities
- Dictionary source files: TEI
- Backend: eXist (native XML database)
- Frontend: PHP

## Лепа реч и гвоздена врата отвара

Откријте и истражите богатство српске лексике

Претражи речнике..



Претрагу покреће Algolia

## Речники

- Сви речници
  - Вук Караџић, *Српски рјечник* ?
  - Глиша Елезовић, *Речник косовско-мештакохиској дијалекта* ?
  - Момчило Златановић, *Речник говора јужне Србије* ?
  - М. Ђојанић и Р. Тривунац, *Речник дубровачког говора* ?

## Стари речници у новом руку

Први пут сте на Расковнику? Не зните одакле да почнете? Препоручујемо да се упознавате со основним функцијама наше речничке платформе.

**Претрага**

Какво функционише инстант-претрага?



**Мапе**

Какве везе имају мапе са речницима?

поп|



**пòп**

Караџић

**поп**

Златановић

**пòп**

Еlezović

**пòп злàтâр, пòп зûк**

Караџић

**пóпа**

Караџић

**попа**

Еlezović

**пòпадати**

Караџић

**попадýка**

Златановић

**попадýка**

Златановић

**попàдија**

Караџић

**попадѝја**

Еlezović

**попàдијин**

Караџић

## Српски рјечник

Претражи...



**дे॑рт,\*** **м.** vide [брига](#): Од *дे॑рта* ме забољела глава.

← [дे॑рњава](#)[↔ 10](#) [☒ 1](#) [⤒ 1](#)[де॑ртли](#) →

## Речник косовско-метохиског дијалекта

Претражи...



**дे॑рт,** а **м.** мука, невоља: Дошǎо сам, ћа, да ти кाजем мој *дे॑рш*. — Б Ѹд је м Ѣка ту и *гे॑рш*. — Ти још не знаш шта је *дे॑рш*. Од пер. *derd* им. туга, брига; душевни, унутрашњи бол. — У В. [де॑рт](#) м. RJA. *dërt* м.

← [дे॑рнут](#)[↔ 10](#) [☒ 1](#)[дертоват](#) →

## Речник говора јужне Србије

Претражи...



**дерт** **м** (перс.) (тур. *dert*) јад, бол (најчешће због љубави). „Изеде га дерт, не помагав ни кафáне“ ([Врање](#)).

← [Дерéп](#)[↔ 10](#) [☒ 1](#)[дертлија](#) →

[ници](#)[Исеџци](#)[Претрага](#)[Мапа](#)[О нама](#)

## Српски рјечник

Претражи...



дे०т,\* **м.** vide [брига](#): Од *герша* ме забољела глава.

[← дे०тава](#)[↔ 10](#)[✖ 1](#)[↓ 1](#)[де०тли →](#)

## Речник косовско-метохиског дијалекта

Претражи...



дे०т, а **м.** мука, невоља

*гे०рш*. — Ти још не знаш

де०т,\* **м.** vide [брига](#): Од *герша* ме забољела глава.

у и

унутрашњи бол. — У В. [де०т](#) м. КЈА. *aert* м.

[← де०тнут](#)[↔ 10](#)[✖ 1](#)[дертоват →](#)

## Речник говора јужне Србије

Претражи...



дерт **м** (перс.) (тур. *dert*) јад, бол (најчешће због лубави). „Изеде га дерт, не помагав ни кафана“ ([Врање](#)).

[← Дерён](#)[↔ 10](#)[✖ 1](#)[дертлија →](#)

**бàјалица**, f. die Zauberin, incantatrix. [cf. [бахорица](#), [врачара 2](#), [врачарица](#), [гатара 2](#)]. Пјевају како се у сватовима (као [ђевојком](#)) разбољео купус на сланини, поглавар. Проја баје, купуса нестаје.

← **Баја**

⇄ 10

☒ 1

**бàхорица**, f. (у [Боци](#)) die Zauberin, incantatrix, vide [бајалица](#), cf. [врачара](#): Нешто ме је забољела глава, | Нег' ми зови *бахорицу*, мајко, | Да би мене младу *бахорила* — Дозвала је другу *бахорицу* — Нег' ми зови трећу *бахорицу*.

## Српски рјечник

нога



**кракат**, а, о, adj. langbeinig, longis cruribus.

**нòга**, f. (acc. нòгу pl. нòге, нòгâ и. нòгû, ) 1. der Fuß,

доћи коме *на ноге*, selbst zu ihm kommen. 3. ићи који

Schritt, pedetentim. 4. (у Ц. г.) н. п. из свега је Спужа

свију Спужана. 5. (у Јадру) eine gewisse Anzahl, certus

разрезују порезу, онда кажу: ово село има две ноге, ово ногу и по, и т. д.

нòга



нòгавица



ногáри, нòгаче



ногáчи



Додај на овој страни

## Речник косовско-метохиског дијалекта

Претражи...



**кракàт** и **кракаš**, **кракàша**, о прид. који има дугачке ноге и прави велики корак: Његòва је жена по *кракаša*. — У В. **кракат**, а, о, у RJA. 1. *krakat* прид. b.

← крај пàмук

⇄ 10 ⌂ 1

Краљаne →

## Изгубљени у преводу

Лексеме из Вуковог *Српског речника* са непрециznим еквивалентима на немачком или латинском, нпр. *шрёскавица* → нем. Art Tanzes („врста игре“) или *бабини зуби* → лат. herbae genus („врста биљке“).

Прикажи:  списак речи  целокупне одреднице

УКУПНО: **977**

1	2	3	4	5	6	7	8
---	---	---	---	---	---	---	---

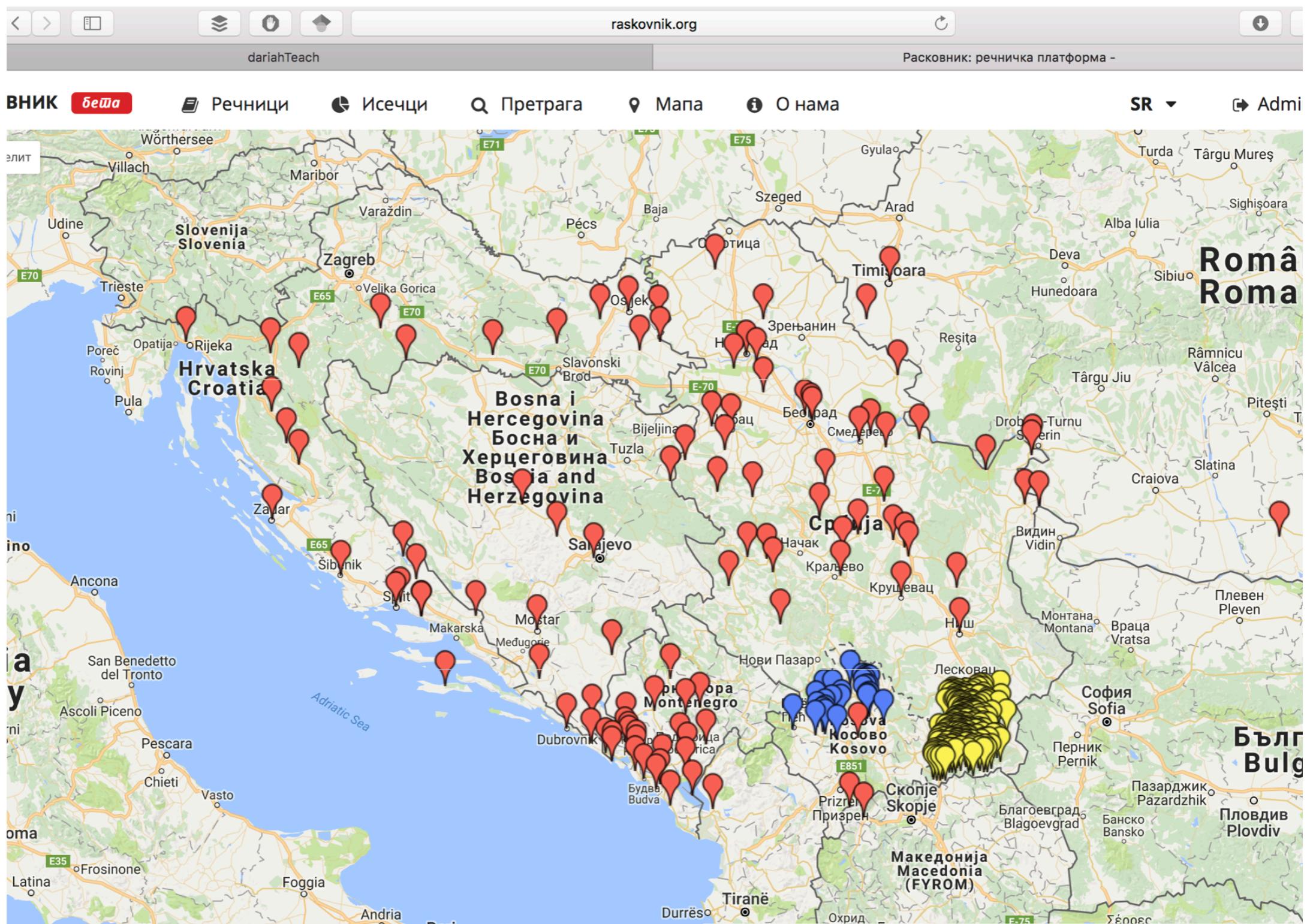
авáлија	бâр	бòбут	буздòвân
авгùтâр	бàрак	бòгаз	бùјân
àздија	бàрбежа	бôдалъ	бùква
àлет	бàрбûн	бòдâч	бùрача
àлица	бардàклија	бòдачкî вез	бурлíкâње
àльма	бàсма	бòдълика	бùћи
àмâз	бâцањ	бòжjâ плàхтица	бúцов
Àрапка	бàша	бокèљица	бùчевина
арпàџик	бèг	бòкêлька	бùшина
аршлàма	бегàница	бòровић	вàвежњици
áченоје	бèча	Бòсânка	вáгôв
àшâњка	бïјелô зëље	брàдâш	вàкавица
бàбаје	бијèлька	брáна	вáкља
бáбац	бïло	брàтичина	вàргâň

# Enriching Serbian dictionaries

- Data normalization
- Geolocation
- Additional annotation

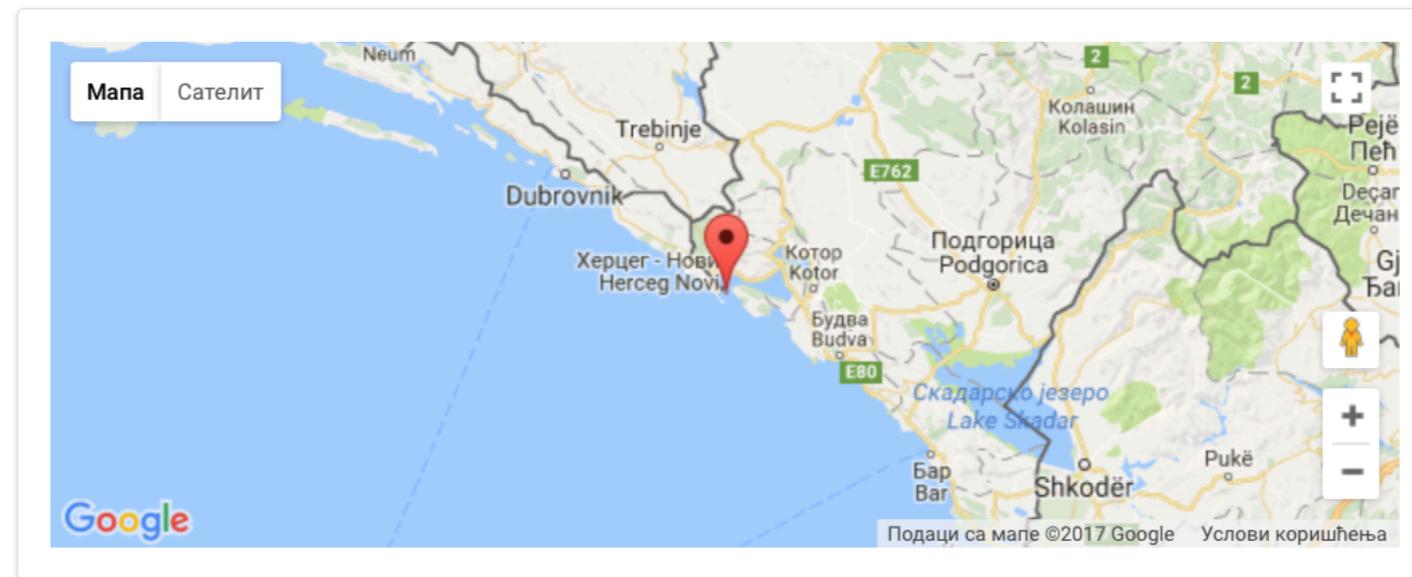
```
<entryFree xml:id="VSK.SR.харамбаша">
    <form type="lemma">
        <orth norm="харамбаша">хàрамбаша</orth>
    </form> [<xr type="cf" source="#thirdEd">
        <ref target="#VSK.SR.арамбаша" type="variant" rend="regular">
            <rs type="lemma">арамбаша</rs>
        </ref>
    </xr>], <etym>
        <lang value="tr">*</lang>
    </etym>
    <gramGrp>
        <pos>m.</pos>
    </gramGrp>
    <sense xml:id="VSK.SR.харамбаша.1">
        <def>поглавица хајдучки</def>, <cit type="translation">
            <quote xml:lang="de">der Räuberhauptmann</quote>
        </cit>, <cit type="translation">
            <quote xml:lang="la">dux latronum</quote>
        </cit>. </sense>
    </entryFree>
```

```
<entryFree xml:id="VSK.SR.чаратан">
  <form type="lemma">
    <orth norm="чаратан">чарàтан</orth>
  </form>,
  <gramGrp>
    <pos>m.</pos>
  </gramGrp>
  <sense xml:id="VSK.SR.чаратан.1">
    <usg type="diatopic">(y <rs type="geo" subtype="region" numRef="Бока Которска" key="89"
>Боци</rs>)</usg>
    <def>човјек који се роди у <ref target="#VSK.SR.марчана_свијећа"
rend="italic"><rs type="lemma">марчаној свијећи</rs> </ref> (<ref target="#VSK.SR.луна"
rend="extended"> <rs type="lemma">луни</rs> </ref>), па прије него је крштен пријеђе
преко њега мачка или друга какова погана животиња</def>; <dictScrap>за то послије остане
несрећан, те не може имати никакога заната нити постојанога посла, него се смуца
којекуда, и својом мајсторијом, игром и шалом људе забавља и вара и новце од њих мами,
али му ни ови новци не могу бити пробитачни, него како дођу тако и отиду</dictScrap>.
  </sense>
</entryFree>
```



## Бока Которска

географска област



Прикажи:  списак речи  целокупне одреднице

УКУПНО: 673

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

âк	бàтûн	блёска	брёме
Âмат	бàхорити	бльёснути	брòкљача
àцâл	бàхорица	бобòтати	брòњица
бáбовић	бèдрени прág	богàтâш	брòхан
бàбушка	бéрба	богòраднîк	брòцнути
бâдаљ	бèстîjnскî	Бòжић	брòч
бâдељ	бìльешка	бокельски	брòча
бâк	бјегùница	бòкîн	брòчñй
бакàлâj	бјегúнче	брàтанић	Бùгар
бàндижâн	блàтîште	брàтаница	бûран
бàнђен	блёка	брàче	Вàра
бàња	блéка	брòдûн	вàрица

```
<entryFree xml:id="VSK.SR.адумац">
  <form type="lemma">
    <orth norm="адумац">а́думац</orth>
  </form>, <form type="ending">
    <orth>мца</orth>
  </form>, <gramGrp>
    <pos>т.</pos>
  </gramGrp>
  <sense xml:id="VSK.SR.адумац.1">
    <xr type="vide">
      <lbl>vide</lbl>
      <ref target="#VSK.SR.хадумац" type="variant">
        <rs type="lemma">хадумац</rs>
      </ref>
    </xr>. <cit type="example" source="#firstEd" subtype="intentionallyOmittedFromSecondEd">
      <quote xml:lang="sr">Ја кажем адум сам, а он дријеши гаће да ј*б*</quote>
    </cit>. </sense>
  </entryFree>
```

## Без цензуре

Под притиском Српске православне цркве и значајног дела тадашње српске елите, Вук је из другог издања свог речника изоставио низ вулгарних речи и примера. Опсцене лексике нема ни у постхумним издањима *Српског рјечника*, а неке од ових речи нису обрађене ни у савременијим речницима српског језика.

Прикажи:  списак речи  целокупне одреднице

УКУПНО: **128**

1 2 3 4

àдумац	главић	дркати	изјёпсти
арàтос	глâс	дрљати	ѝсписнâ
бруцâње	глòгиња	дùпльâш	јàрмењâк
брùцати	góвно	запíздити	јêб
брùце	голòкур	зарèћи се	јебàнија
бùнгурац	гráдити	зезâње	јебâње
вárêње	гùзица	зèзати	јёбаоница
вјёштица	долàктâш	зèзнути	јёбати
гвоздèњâк	дркâње	изјёбати	јёбац

## Без цензуре

Под притиском Српске православне цркве и значајног дела тадашње српске елите, Вук је из другог издања свог речника изоставио низ вулгарних речи и примера. Опсцене лексике нема ни у постхумним издањима *Српској рјечници*, а неке од ових речи нису обрађене ни у савременијим речницима српског језика.

Прикажи:  списак речи  целокупне одреднице

Укупно: 124

1

2

3

4

**àдумац**, мца, **m.** vide [хадумац](#). Ја кажем адум сам, а он дријеши гаће да j\*б\*.

**арàтос** [арантос] (άράτος) verflucht, hol's der Teufel, in malam rem! *Арапос* га било! *Арапос* ти кирије, сиђи ми с кола. Аратос ти [поскура](#), не прди ми по цркви. [cf. [2 куртала 1](#)].

**брुцâње**, **n.** das Bekommen der Schamhaare, nactio pilorum circa pudenda, pubertas, pubescentia (?).

**брùцати**, **ам**, **v. impf.** und **pf.** Schamhaare bekommen, nancisci pilos circa pudenda.

**брùце**, **f. pl.** die Schamhaare, pili circa pudenda.

**бùнгурац**, бùнгурца, **m.** dim. **v.** [бунгур](#). Ручку бунгурца, вечери ни к\*\*ца.

**вáрêње**, **n.** 1. das Kochen, coctio. 2. варéње (у приповијеци) пече ли се печење и вари ли се *варење* , das

вјештица. Овако су истраживали вјештице и за Карађорђијева времена у Србији. Гдјекоји на бијеле покладе изврђу вериге наопако од вјештица. Гдјекоји у попрет метну какав рог, јер кажу да вјештица особито бјежи од смрада овога. Гдјекоји мрве јајиње љуске да се не би вјештице у њима могле возити преко вода. — Куд ће вјештица до у свој род?

**гвоздèњâк**, гвоздењáка, **m.** 1. [котао](#) од гвожђа, eisener Kessel, crater ferreus. [cf. [котлача](#)]. 2. гвозден буздан: узе кадија *ївоздењак* , те удари Мују над прдењак . cf. [оклиџнути](#).

**главîћ**, **m.** die Sichel, mentulae caput.

**глâс**, **m.** (loc. *глáсу*, pl. *глâси*, *глâсови*)

1. die Stimme, vox [cf. [грло 3](#), [аваз](#)]: има лијеп *лас* ; познајем га по *ласу* .  
2. die Nachricht, nuncius [cf. [погласје](#), [хабер 1](#) ([абер](#))]: Ради бисмо добре

**TEI is a way to encode data. How  
you will present that data to  
your users is a separate matter.**

# TEI Guidelines

<https://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html>

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

# < Text Encoding Initiative >

e Guidelines Activities Tools Membership Support About News

P5 Guidelines — English

## P5: Guidelines for Electronic Text Encoding and Interchange

Version 2.8.0. Last updated on 6th April 2015, revision 13

[English] [Deutsch] [Español] [Italiano] [Français] [日本語] [한국어] [中文]

### Front Matter

**Title**

- + i. [Releases of the TEI Guidelines](#)
- + ii. [Dedication](#)
- + iii. [Preface and Acknowledgments](#)
- + iv. [About These Guidelines](#)
- + v. [A Gentle Introduction to XML](#)
- + vi. [Languages and Character Sets](#)

### Back Matter

[Appendix A Model Classes](#)  
[Appendix B Attribute Classes](#)  
[Appendix C Elements](#)  
[Appendix D Attributes](#)  
[Appendix E Datatypes and Other Macros](#)  
+ Appendix F [Bibliography](#)  
+ Appendix G [Prefatory Notes](#)  
+ Appendix H [Colophon](#)

### Text Body

- + 1 [The TEI Infrastructure](#)
- + 2 [The TEI Header](#)
- + 3 [Elements Available in All TEI Documents](#)
- + 4 [Default Text Structure](#)
- + 5 [Characters, Glyphs, and Writing Modes](#)
- + 6 [Verse](#)
- + 7 [Performance Texts](#)
- + 8 [Transcriptions of Speech](#)
- + 9 [Dictionaries](#)
- + 10 [Manuscript Description](#)
- + 11 [Representation of Primary Sources](#)
- + 12 [Critical Apparatus](#)
- + 13 [Names, Dates, People, and Places](#)
- + 14 [Tables, Formulae, Graphics and Notated Music](#)
- + 15 [Language Corpora](#)
- + 16 [Linking, Segmentation, and Alignment](#)
- + 17 [Simple Analytic Mechanisms](#)
- + 18 [Feature Structures](#)
- + 19 [Graphs, Networks, and Trees](#)
- + 20 [Non-hierarchical Structures](#)
- + 21 [Certainty, Precision, and Responsibility](#)
- + 22 [Documentation Elements](#)
- + 23 [Using the TEI](#)

### TEI sourcecode

- [Using the TEI Sourceforge Repository](#)
- [Sourceforge Subversion Repository](#)
- [Bug Reports, Feature Requests, etc.](#)



# TEI Chapters (1)

- **The TEI Infrastructure**
  - Transcriptions of Speech
- **The TEI Header**
  - **Dictionaries**
- Elements Available in All TEI Documents
  - Manuscript Description
  - Representation of Primary Sources
- Default Text Structure
  - Critical Apparatus
- Representation of Non-standard Characters and Glyphs
  - Names, Dates, People, and Places
- Verse
  - Tables, Formulæ, and Graphics
- Performance Texts

# TEI Chapters (2)

- Simple Analytic Mechanisms
- Feature Structures
- Graphs, Networks, and Trees
- Non-hierarchical Structures
- Certainty and Responsibility
- Documentation Elements
- Using the TEI

# TEI Elements

<http://bit.ly/tei-elements>

**TEI** < Text Encoding Initiative >

Home Guidelines Activities Tools Membership Support About News

P5 Guidelines – English

## P5: Guidelines for Electronic Text Encoding and Interchange

Version 2.8.0. Last updated on 6th April 2015, revision 13197

### Appendix C Elements

Sorted alphabetically

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t
u	v	w	x	z	<a href="#">Show all</a>	<a href="#">Show by module</a>													

#### a

[ab](#) [abbr](#) [abstract](#) [accMat](#) [acquisition](#) [activity](#) [actor](#) [add](#) [additional](#) [additions](#) [addName](#) [address](#) [addrLine](#) [addSpan](#) [adminInfo](#) [affiliation](#) [age](#) [alt](#) [alternate](#) [altGrp](#) [altIdent](#) [altIdentifier](#) [am](#) [analytic](#) [anchor](#) [app](#) [appInfo](#) [application](#) [arc](#) [argument](#) [att](#) [attDef](#) [attList](#) [attRef](#) [author](#) [authority](#) [availability](#)

#### b

[back](#) [bibl](#) [biblFull](#) [biblScope](#) [biblStruct](#) [bicond](#) [binary](#) [binaryObject](#) [binding](#) [bindingDesc](#) [birth](#) [bloc](#) [body](#) [broadcast](#) [byline](#)

#### c

[c](#) [caesura](#) [calendar](#) [calendarDesc](#) [camera](#) [caption](#) [case](#) [castGroup](#) [castItem](#) [castList](#) [catchwords](#) [catDesc](#) [category](#) [catRef](#) [cb](#) [cell](#) [certainty](#) [change](#) [channel](#) [char](#) [charDecl](#) [charName](#) [charProp](#) [choice](#) [cit](#) [citedRange](#) [cl](#) [classCode](#) [classDecl](#) [classes](#) [classRef](#) [classSpec](#) [climate](#) [closer](#) [code](#) [collation](#) [collection](#) [colloc](#) [colophon](#) [cond](#) [condition](#) [constitution](#) [constraint](#) [constraintSpec](#) [content](#) [corr](#) [correction](#) [correspAction](#) [correspContext](#) [correspDesc](#) [country](#) [creation](#) [cRefPattern](#) [custEvent](#) [custodialHist](#)

#### d

[« Appendix B](#)  
[Attribute Classes](#)  
[» Appendix D](#)  
[Attributes](#)  
[Home](#)

 TEXT ENCODING INITIATIVE

Home Guidelines Activities Tools Membership Support About News

P5 Guidelines – English

## P5: Guidelines for Electronic Text Encoding and Interchange

Version 2.8.0. Last updated on 6th April 2015, revision 13197

---

### Appendix C Elements

Sorted alphabetically

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t
u	v	w	x	z	Show all		Show by module												

**[analysis] Simple analytic mechanisms**

[c](#) [cl](#) [interp](#) [interpGrp](#) [m](#) [pc](#) [phr](#) [s](#) [span](#) [spanGrp](#) [w](#)

**[certainty] Certainty and uncertainty**

[certainty](#) [precision](#) [respons](#)

**[core] Elements common to all TEI documents**

[abbr](#) [add](#) [addrLine](#) [address](#) [analytic](#) [author](#) [bibl](#) [biblScope](#) [biblStruct](#) [binaryObject](#) [cb](#) [choice](#) [cit](#) [citedRange](#) [corr](#) [date](#) [del](#) [desc](#) [distinct](#) [divGen](#) [editor](#) [email](#) [emph](#) [expan](#) [foreign](#) [gap](#) [gb](#) [gloss](#) [graphic](#) [head](#) [headItem](#) [headLabel](#) [hi](#) [imprint](#) [index](#) [item](#) [i](#) [label](#) [lb](#) [lg](#) [list](#) [listBibl](#) [measure](#) [measureGrp](#) [media](#) [meeting](#) [mentioned](#) [milestone](#) [monogr](#) [name](#) [note](#) [num](#) [orig](#) [p](#) [pb](#) [postBox](#) [postCode](#) [ptr](#) [pubPlace](#) [publisher](#) [q](#) [quote](#) [ref](#) [reg](#) [relatedItem](#) [resp](#) [respStmt](#) [rs](#) [said](#) [series](#) [sic](#) [soCalled](#) [sp](#) [speaker](#) [stage](#) [street](#) [teiCorpus](#) [term](#) [textLang](#) [time](#) [title](#) [unclear](#)

**[corpus] Corpus texts**

[activity](#) [channel](#) [constitution](#) [derivation](#) [domain](#) [factuality](#) [interaction](#) [locale](#) [particDesc](#) [preparedness](#) [purpose](#) [setting](#) [settingDesc](#) [textDesc](#)



Home Guidelines Activities Tools Membership Support About News

 P5 Guidelines — English 

## P5: Guidelines for Electronic Text Encoding and Interchange

Version 2.8.0. Last updated on 6th April 2015, revision 13197

### Appendix C Elements

Sorted alphabetically

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t
u	v	w	x	z	<a href="#">Show all</a>		<a href="#">Show by module</a>												

e

[edition](#) [editionStmt](#) [editor](#) [editorialDecl](#) [education](#) [eg](#) [egXML](#) [eLeaf](#) [elementRef](#) [elementSpec](#) [email](#) [emph](#) [encodingDesc](#) [entry](#) [entryFree](#) [epigraph](#) [epilogue](#) [equipment](#) [equiv](#) [eTree](#) [etym](#) [event](#) [ex](#) [exemplum](#) [expan](#) [explicit](#) [extent](#)

[\[English\]](#) [\[Deutsch\]](#) [\[Español\]](#) [\[Italiano\]](#) [\[Français\]](#) [\[日本語\]](#) [\[한국어\]](#) [\[中文\]](#)

↑ [TEI P5](#)  
[Guidelines](#) «  
[Appendix B](#)  
[Attribute Classes](#)  
» [Appendix D](#)  
[Attributes](#)  
[Home](#)

« [Appendix B](#)  
Attribute Classes  
» [Appendix D](#)  
Attributes  
Home

[TEI Consortium](#) | [Feedback](#)

TEI Guidelines Version 2.8.0. Last updated on 6th April 2015, revision 13197. This page generated on 2015-04-06T13:19:40Z.

P5 Guidelines – English

## P5: Guidelines for Electronic Text Encoding and Interchange

Version 2.8.0. Last updated on 6th April 2015, revision 13197

### <entry>

[Home](#)  
[C Elements](#)

<entry> contains a single structured entry in any kind of lexical resource, such as a dictionary or lexicon. [[9.1 Dictionary Body and Overall Structure](#) [9.2 The Structure of Dictionary Entries](#)]

<b>Module</b>	dictionaries — Dictionaries
<b>Attributes</b>	<a href="#">att.global</a> (@xml:id, @n, @xml:lang, @xml:base, @xml:space) ( <a href="#">att.global.rendition</a> (@rend, @style, @rendition)) ( <a href="#">att.global.linking</a> (@corresp, @synch, @sameAs, @copyOf, @next, @prev, @exclude, @select)) ( <a href="#">att.global.analytic</a> (@ana)) ( <a href="#">att.global.facs</a> (@facs)) ( <a href="#">att.global.change</a> (@change)) ( <a href="#">att.global.responsibility</a> (@cert, @resp)) <a href="#">att.entryLike</a> (@type) <a href="#">att.sortable</a> (@sortKey)
<b>Member of</b>	<a href="#">model.entryLike</a>
<b>Contained by</b>	<b>dictionaries:</b> <a href="#">superEntry</a> <b>drama:</b> <a href="#">castList</a> <a href="#">epilogue</a> <a href="#">performance</a> <a href="#">prologue</a> <a href="#">set</a> <b>figures:</b> <a href="#">figure</a> <b>textstructure:</b> <a href="#">argument</a> <a href="#">body</a> <a href="#">div</a> <a href="#">div1</a> <a href="#">div2</a> <a href="#">div3</a> <a href="#">div4</a> <a href="#">div5</a> <a href="#">div6</a> <a href="#">div7</a> <a href="#">epigraph</a> <a href="#">postscript</a>
<b>May contain</b>	<b>analysis:</b> <a href="#">interp</a> <a href="#">interpGrp</a> <a href="#">span</a> <a href="#">spanGrp</a> <b>certainty:</b> <a href="#">certainty</a> <a href="#">precision</a> <a href="#">respons</a> <b>core:</b> <a href="#">cb</a> <a href="#">cit</a> <a href="#">gap</a> <a href="#">gb</a> <a href="#">index</a> <a href="#">lb</a> <a href="#">milestone</a> <a href="#">note</a> <a href="#">pb</a> <a href="#">ptr</a> <a href="#">ref</a> <b>ictionaries:</b> <a href="#">def</a> <a href="#">dictScrap</a> <a href="#">etym</a> <a href="#">form</a> <a href="#">gramGrp</a> <a href="#">hom</a> <a href="#">re</a> <a href="#">sense</a> <a href="#">usg</a> <a href="#">xr</a> <b>figures:</b> <a href="#">figure</a> <a href="#">notatedMusic</a> <b>iso-fs:</b> <a href="#">fLib</a> <a href="#">fs</a> <a href="#">fvLib</a> <b>linking:</b> <a href="#">alt</a> <a href="#">altGrp</a> <a href="#">anchor</a> <a href="#">join</a> <a href="#">joinGrp</a> <a href="#">link</a> <a href="#">linkGrp</a> <a href="#">timeline</a> <b>spoken:</b> <a href="#">incident</a> <a href="#">kinesic</a> <a href="#">pause</a> <a href="#">shift</a> <a href="#">vocal</a> <a href="#">writina</a>

<b>Declaration</b>	<div style="text-align: right; margin-bottom: 5px;"><a href="#">XML syntax</a></div> <pre>element entry {     att.global.attributes,     att.global.rendition.attributes,     att.global.linking.attributes,     att.global.analytic.attributes,     att.global.facs.attributes,     att.global.change.attributes,     att.global.responsibility.attributes,     att.entryLike.attributes,     att.sortable.attributes,     ( hom   sense   model.entryPart.top   model.global   model.ptrLike )+ }</pre>
<b>Example</b>	<pre>&lt;entry&gt; &lt;form&gt; &lt;orth&gt;disproof&lt;/orth&gt; &lt;pron&gt;dɪs'pru:f&lt;/pron&gt; &lt;/form&gt; &lt;gramGrp&gt; &lt;pos&gt;n&lt;/pos&gt; &lt;/gramGrp&gt; &lt;sense n="1"&gt; &lt;def&gt;facts that disprove something.&lt;/def&gt; &lt;/sense&gt; &lt;sense n="2"&gt; &lt;def&gt;the act of disproving.&lt;/def&gt; &lt;/sense&gt; &lt;/entry&gt;</pre> <div style="text-align: right; margin-top: -10px;"><a href="#">Show all</a></div>
<b>Note</b>	<p>Like all elements, <code>entry</code> inherits an <code>@xml:id</code> attribute from the class <code>global</code>. No restrictions are placed on the method used to construct <code>@xml:ids</code>; one convenient method is to use the orthographic form of the headword, appending a disambiguating number where necessary. Identification codes are sometimes included on machine-readable tapes of dictionaries for in-house use.</p> <p>It is recommended to use the <code>sense</code> 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.</p>

# Structure of a TEI document

- **<TEI>** (TEI document) contains a single TEI-conformant document, comprising a TEI header and a text, either in isolation or as part of a **teiCorpus** element.
- **<teiCorpus>** contains the whole of a TEI encoded corpus, comprising a single corpus header and one or more TEI elements, each containing a single text header and a text.

```
<TEI>
  <teiHeader>
    <!-- required -->
  </teiHeader>
  <facsimile>
    <!-- optional, new in TEI P5 -->
  </facsimile>
  <text>
    <!-- required if no facsimile -->
  </text>
</TEI>
```

```
<teiCorpus>
  <teiHeader>
<!-- required -->
  </teiHeader>
  <TEI>
<!-- required -->
  </TEI>
</teiCorpus>
```

# You tell me

- what goes into <front>?
- what goes into <back>?

```
<text>
  <front>
    <!-- optional -->
  </front>
  <body>
    <!-- required -->
  </body>
  <back>
    <!-- optional -->
  </back>
</text>
```

**What does a dictionary  
entry written in TEI look  
like?**

**disproof** (dɪs"pru:f) n. 1. facts that disprove something. 2. the act of disproving

**disproof** (dɪs"pru:f) n. 1. facts that disprove something. 2. the act of disproving

## <entry>

</entry>

**disproof** (dɪs"pru:f) n. 1. facts that disprove something. 2. the act of disproving

```
<entry>
  <form type="lemma">
    ...
  </form>
</entry>
```

**disproof** (dɪs"pru:f) n. 1. facts that disprove something. 2. the act of disproving

```
<entry>
  <form type="lemma">
    <orth>disproof</orth>
  </form>
```

```
</entry>
```

**disproof** (dɪs"pru:f) n. 1. facts that disprove something. 2. the act of disproving

```
<entry>
  <form type="lemma">
    <orth>disproof</orth>
    <pron>dls"pru:f</pron>
  </form>
```

```
</entry>
```

**disproof** (dɪs"pru:f) n. 1. facts that disprove something. 2. the act of disproving

```
<entry>
  <form type="lemma">
    <orth>disproof</orth>
    <pron>dls"pru:f</pron>
  </form>
  <gramGrp>

  </gramGrp>

</entry>
```

**disproof** (dɪs"pru:f) n. 1. facts that disprove something. 2. the act of disproving

```
<entry>
  <form type="lemma">
    <orth>disproof</orth>
    <pron>dls"pru:f</pron>
  </form>
  <gramGrp>
    <pos>n</pos>
  </gramGrp>

</entry>
```

**disproof** (dɪs"pru:f) n. 1. facts that disprove something. 2. the act of disproving

```
<entry>
  <form type="lemma">
    <orth>disproof</orth>
    <pron>dls"pru:f</pron>
  </form>
  <gramGrp>
    <pos>n</pos>
  </gramGrp>
  <sense n="1">

  </sense>
  <sense n="2">

  </sense>
</entry>
```

**disproof** (dɪs"pru:f) n. 1. facts that disprove something. 2. the act of disproving

```
<entry>
  <form type="lemma">
    <orth>disproof</orth>
    <pron>dls"pru:f</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>
```

# Some Elements You Will Learn to Love (and Hate)

- **form** (form information group) groups all the information on the written and spoken forms of one headword.
- **gramGrp** (grammatical information group) groups morpho-syntactic information about a lexical item, e.g. pos, gen, number, case, or iType (inflectional class).
- **def** (definition) contains definition text in a dictionary entry.
- **cit** (cited quotation) contains a quotation from some other document, together with a bibliographic reference to its source. In a dictionary it may contain an example text with at least one occurrence of the word form, used in the sense being described, or a translation of the headword, or an example.
- **usg** (usage) contains usage information in a dictionary entry.
- **xr** (cross-reference phrase) contains a phrase, sentence, or icon referring the reader to some other location in this or another text.
- **etym** (etymology) encloses the etymological information in a dictionary entry.

# Time to do some preparatory work on your own data

- Go to our Class Notes ([bit.ly/TEI-Lisbon2019](https://bit.ly/TEI-Lisbon2019)) and find the Exercise 1 for Tuesday, July 2nd

## 1. a. If you have a dictionary sample

- select an entry to analyze:
  - the best entry to start with is neither too short nor too long
  - the entry should be complex enough yet typical enough
- take a screenshot of your entry, save it on your computer, then drag it to the “black editor” of the Class Notes in your section (below) – after a few seconds, it should siplay in the preview to the right.
- look back at the exercises we did yesterday (*coffee*) and do the same for your selected entry:
  - take notes (after the image in your sections below) outlining the structural and typographic features
  - think about microstructure as a hierarchy: which elements contain other elements?

## 1. b. If you have corpus data

- think about the data contained in your corpus and how you could extract lexical information from it
- think about the structure of a prototypical dictionary entry based on your data
- take notes (in your section below) and start describing your dictionary entry
  - think about microstructure as a hierarchy: which elements contain other elements?

# Good luck!