

Standardising annotation of language data: TEI

Tomaž Erjavec

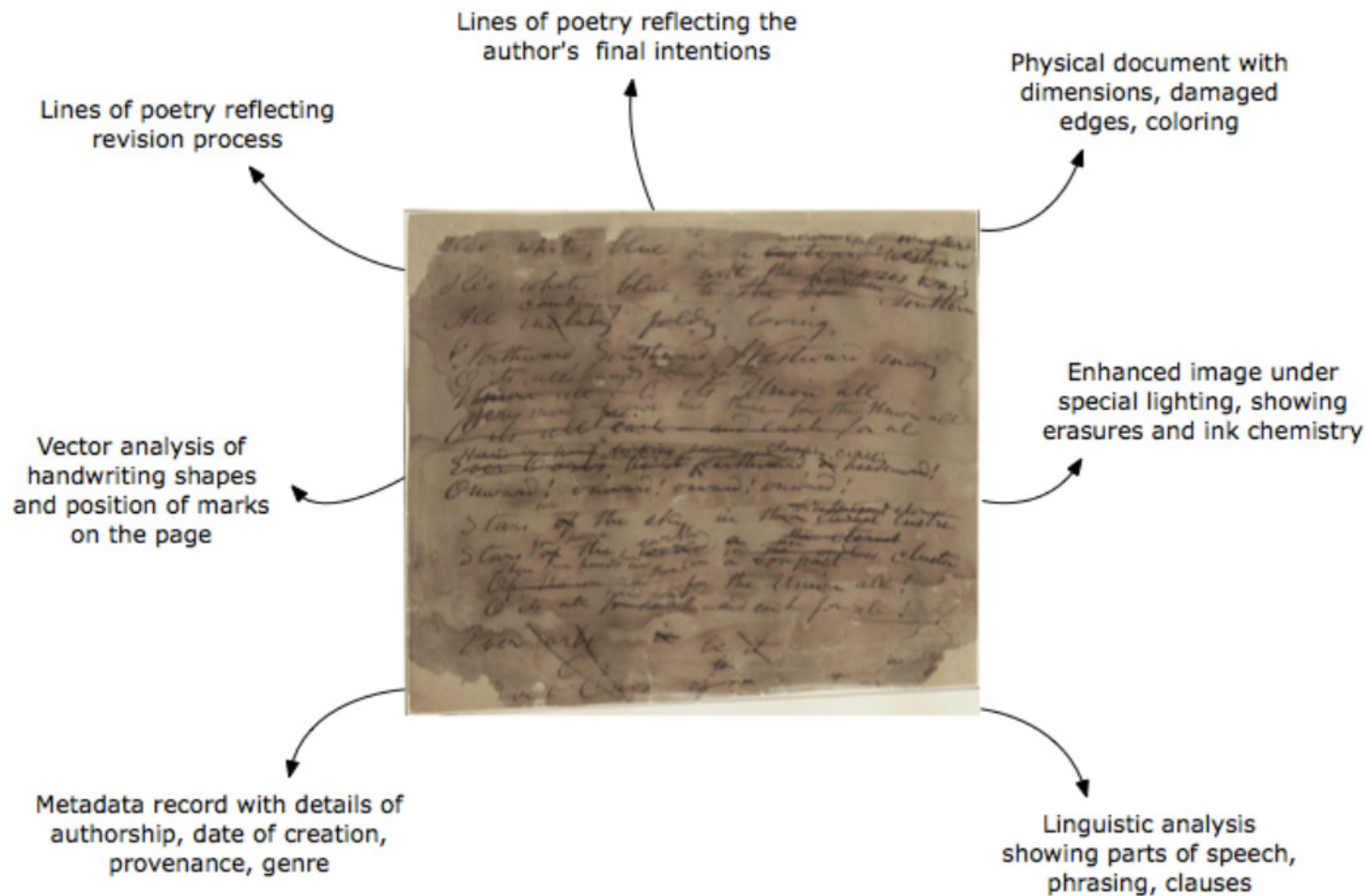
Department of Knowledge Technologies, Jožef Stefan Institute

FRI, 29th April, 2020

Overview of the lecture

- Introduction
- History, organisation and scope of TEI
- TEI modules
- TEI in Slovenia

Text Encoding: Representing Research Objects



Digital formats for text

- Graphic formats (facsimile of text)
- Sound formats (speech)
- Video formats (movie of a speaker)
- Tabular formats (lexicons, some training sets)
- **XML** (annotated text and pointers)

Advantages of XML

- Supports hierarchical structures and pointers
- Allows mixing text and annotations
- A W3C standard
- Formal validation via a schema:
DTD, W3C schema, RelaxNG, Schematron
- Many associated standards:
XInclude, XPath, XSLT, Xquery
- Tool support:
Saxon (XSLT), eXist (XQuery), editors (Oxygen, emacs)
- “Simple” conversion to applications formats:
XML, HTML, JSON, tables, PDF, ...
- A good archive format

XML schemas for text annotation

- XML is a meta-language:
it is the schema defines the allowed elements and attributes, and allowed nestings
- In addition to the formal schema (syntax) we also need documentation of the meaning of the elements, attributes, values (semantics)
- We can develop a schema for a particular type of documents ourselves
- For data interchange (and tool support) it is better to use a standard schema
- For linguistic annotation there are (too) many standard and best-practice schemas:
 - ISO TC 37 SC4 standards for encoding language resources
 - National projects (TCF: Germany, FoLiA: Netherlands, ...)
 - and TEI, which is, however, much broader



< Text Encoding Initiative >

- Started as a research project in the humanities
 - supported by three professional associations
 - financed 1990-1994 (ZDA, EU)
- influential text types:
 - digital libraries, text collections
 - language corpora
 - scholarly datasets
- International consortium established 1999
- Web page: <http://www.tei-c.org/>

Goals of the TEI

- Better interchange and integration of scholarly data
- Support for all texts, in all languages, from all periods
- Guidance for the perplexed: what to encode:
a user-driven codification of existing best practice
- Assistance for the specialist: how to encode:
a loose framework into which unpredictable extensions
can be fitted

These apparently incompatible goals result in a highly flexible, modular, environment

Legacy of the TEI

- A way of looking at what 'text' really is
- A codification of current scholarly practice
- A set of shared assumptions and priorities about the digital agenda:
 - focus on content and function (rather than presentation)
 - identify generic solutions (rather than application-specific ones)

TEI Guidelines

- “TEI Guidelines for Electronic Text Encoding and Interchange” cf. <http://www.tei-c.org/Guidelines/>
- Cover generic structures as well as very specific fields
- A large collection of XML element and attribute definitions
- TEI Guidelines comprise documentation and formal definitions
Written in TEI ODD: One Document Does it all
- In print ~1.200 pages
- A modular system for creating personalized schemas
- Maintained on GitHub,
available in TEI/XML, HTML, PDF, EPUB,...

TEI Guidelines on the Web

Home Guidelines Activities Tools Membership Support About News

P5 Guidelines — English ▾

Search

P5: Guidelines for Electronic Text Encoding and Interchange

Version 4.0.0. Last updated on 13th February 2020, revision ccd19b0ba

[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](#)

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](#)

TEI sourcecode

- [Getting and Using the TEI Sources.](#)
- [TEI GitHub Repository](#)
- [Bug Reports, Feature Requests, etc.](#)

Chapters / modules

- 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, Formulæ, Graphics and Notated Music](#)
 - 15 [Language Corpora](#)
 - 16 [Linking, Segmentation, and Alignment](#)
 - 17 [Simple Analytic Mechanisms](#)
 - 18 [Feature Structures](#)
 - 19 [Graphs, Networks, and Trees](#)
 - 20 [Non-hierarchical Structures](#)
 - 21 [Certainty, Precision, and Responsibility](#)
 - 22 [Documentation Elements](#)
 - 23 [Using the TEI](#)
- 2 [The TEI Header](#)
 - 2.1 [Organization of the TEI Header](#)
 - 2.1.1 [The TEI Header and Its Components](#)
 - 2.1.2 [Types of Content in the TEI Header](#)
 - 2.1.3 [Model Classes in the TEI Header](#)
 - 2.2 [The File Description](#)
 - 2.2.1 [The Title Statement](#)
 - 2.2.2 [The Edition Statement](#)
 - 2.2.3 [Type and Extent of File](#)
 - 2.2.4 [Publication, Distribution, Licensing, etc.](#)
 - 2.2.5 [The Series Statement](#)
 - 2.2.6 [The Notes Statement](#)
 - 2.2.7 [The Source Description](#)
 - 2.2.8 [Computer Files Derived from Other Computer Files](#)
 - 2.3 [The Encoding Description](#)
 - 2.3.1 [The Project Description](#)
 - 2.3.2 [The Sampling Declaration](#)
 - 2.3.3 [The Editorial Practices Declaration](#)
 - 2.3.4 [The Tagging Declaration](#)
 - 2.3.4.1 [Rendition](#)
 - 2.3.4.2 [Tag Usage](#)
 - 2.3.5 [The Default Style Definition Language Declaration](#)
 - 2.3.6 [The Reference System Declaration](#)
 - 2.3.6.1 [Prose Method](#)
 - 2.3.6.2 [Search-and-Replace Method](#)
 - 2.3.6.3 [Milestone Method](#)
 - 2.3.7 [The Classification Declaration](#)
 - 2.3.8 [The Geographic Coordinates Declaration](#)
 - 2.3.9 [The Schema Specification](#)

The TEI ecosystem

- TEI Consortium (Technical Council, Board of Directors)
- Open source and collaborative development
- Guidelines and web pages regularly maintained
- A friendly and active tei-I mailing list
- jTEI journal and annual conferences of the TEI
- Support tools:
 - Roma: Web interface to parametrise TEI and generate XML schema
 - TEI Stylesheets: conversion to and from TEI for many formats (docx, open office, html, markdown, ePub, ...)

TEI in Slovenia (1998-)

- Reference corpora:
Fida, FidaPLUS, Gigafida; ssj500k; hr500k, ...
- EU projects (IJS et al.): MULTEXT-East, IMPACT, ELEXIS
- National projects ZRC SAZU + IJS:
Scholarly editions of Slovenian literature, Slovenian biography
- Projects with UL Dept. for Asian languages + IJS:
Japanese-Slovene learners' dictionary and corpora
- DARIAH-SI + CLARIN.SI:
digital library + corpora (e.g. siParl)
- National projects IJS:
 - JOS: Linguistic Annotation of Slovene
 - Janes: Linguistic Annotation of Non-Standard Slovene
 - KAS: Slovene Scientific Texts: Resources and Description

Examples of TEI use: “1984”

```

<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader xml:lang="en" xml:id="mteo-sl.teiHeader"> [340 lines]
  <text xml:id="mteo-sl." xml:lang="sl">
    <body xml:id="Osl">
      <div xml:id="Osl.1" n="1" type="part">
        <head>Prvi del</head>
        <div xml:id="Osl.1.2" n="1" type="chapter">
          <head>I</head>
          <p xml:id="Osl.1.2.2">
            <s xml:id="Osl.1.2.2.1">Bil je jasen, mrzel aprilski dan in ure so bile trinajst.</s>
            <s xml:id="Osl.1.2.2.2"><name>Winston Smith</name> je imel brado zakopano v prsi, da bi
              ušel strupenemu vetru, ko je stopil skozi steklena vrata bloka Zmaga, vendar ne dovolj
              hitro, da ne bi vrtinec peščenega prahu vstopil skupaj z njim.</s>
          </p>
          <p xml:id="Osl.1.2.3">
            <s xml:id="Osl.1.2.3.1">Veža je smrdela po kuhanem zelju in starih, cunjastih
              predpražnikih.</s>
            <s xml:id="Osl.1.2.3.2">Na eni strani je bil na steno pribit barven, za notranjo opremo
              prevelik plakat.</s>
            <s xml:id="Osl.1.2.3.3">Prikazoval je preprosto ogromen, več kot meter velik obraz:
              obraz moškega pri petinštiridesetih, s košatimi črnimi brki in z ostro začrtanimi,
              čednimi potezami.</s>
          </p>
        </div>
      </div>
    </body>
  </text>
</TEI>

```

Izidor Cankar: S poti

I. [BENETKE]{VENEZIA}.

§1 Težko si je misliti večjo razliko, nego je med ljubljanskim septembrskim jutrom in benečanskim septembrskim večerom. Zrak se mi vtihotaplja v sobo kakor lepa godba v srce, kadar sem obupal nad samim seboj in se povprašujem, čemu sem na svetu. In vendar je tudi današnje ljubljansko jutro bilo lepo, kljub tisti mrzli mokroti, ki peče oči. [Snoči sem bil pri izpovedi; prvo prebujenje po izpovedi je nekaj neizrečeno prijetnega. Človek odpre oči z zavestjo otroške dušne svežosti; sam s seboj je nežen kakor mati; preden prižge svečo, je tema vsa preprežena s svetlobo; svet je tako skladen in dober, vsakdo posebej poljuba vreden; kar je bilo srcu veliko, mu je majhno in le eno potrebno: ne delati krivice. Vrhutega so mi bili računi že v redu, obresti poravnane, drobne skrbi odložene. Ko sem se vozil na kolodvor, mi je cigareta neizmerno teknila.]

§2 Na peronu {me} je že čakal moj mladi prijatelj Fritz, poet in umetnostni zgodovinar iz rajha. Prezebal je kljub svoji topli športni suknji. Pogled mu je bil v kolodvorskih meglah, parah in sajah kakor vlažen protest. Bilo je gotovo, da je slabo spal.

§3 „Ne, sploh nisem spal. Vaši hoteli so brlogi za najbolj pokore potrebne [izpokornike]{spokornike}. Večerjal sem žlico juhe, v postelji sem si zlomil vrat in dobil revmatizem. Bagage.“

```
<p id="p.3" n="3">„Ne, sploh nisem spal. Vaši hoteli so brlogi za najbolj
pokore potrebne <app>
  <rdg wit="DS">izpokornike</rdg>
  <rdg wit="KN">spokornike</rdg>
</app>. Večerjal sem žlico juhe, v postelji sem si zlomil vrat in dobil
revmatizem. Bagage.“</p>
```


Slovene biography: on the web



Slovenska biografija



[Abecedno kazalo](#)

[Obdobja](#)

[Poklici in dejavnosti](#)

[Skupine oseb](#)

[Na današnji dan](#)

[Rodbine](#)

[Zemljevid](#)

Strgar, Jan (1881–1955)

★ 8. maj 1881

Nemški Rovt, Slovenija

† 9. november 1955

Jesenice (Jesenice, obč.), Slovenija

Imena

ime: Strgar Jan

ime: Strgar Janez

ime: Stergar Jan

Poklic ali dejavnost

- čebelar
- trgovec

► [Podatki v zapisu Text Encoding Initiative](#)

Slovenski biografski leksikon

Strgar (Stergar) Jan(ez), čebelar in trgovec s čebelami, r. 8. maja 1881 v Nem. rovtu št. 18 (Boh. Bistrica)

Slovene biography: TEI

```

<person xmlns="http://www.tei-c.org/ns/1.0" xml:id="sbi619828"
  corresp="sbl-text.xml#sbl03314" role="main">
  <idno type="URL">http://www.slovenska-biografija.si/oseba/sbi619828/</idno>
  <sex value="1"/>
  <persName>
    <forename>Jan</forename>
    <surname>Strgar</surname>
  </persName>
  <persName>
    <forename>Janez</forename>
    <surname>Strgar</surname>
  </persName>
  <persName>
    <forename>Jan</forename>
    <surname>Stergar</surname>
  </persName>
  <occupation scheme="#occupation" code="#cebelar"/>
  <occupation scheme="#occupation" code="#trgovec"/>
  <birth>
    <date when="1881-05-08">8. maja 1881</date>
    <placeName>
      <settlement>Nemški Rovt</settlement>
      <country>Slovenija</country>
      <geo>46.2713028 13.9787442</geo>
    </placeName>
  </birth>

```

jaSlo: Japanese-Slovene dictionary

bakeru ばける【化ける】(V1)

pojavititi se pod krinko; preleviti se; prevzeti obliko; spremeniti se (v pošast)

- 狐(きつね)が女の子(おんなのこ)に化けた。

Lisica se je spremenila v deklico.

- 化け猫(ばけねこ)

(čarobna) mačka, ki se je spremenila [v človeka ipd.]

težavnostna stopnja 1

konkordance za ばける: L2 ([3](#)), L0 ([9](#)), jpWaC ([174](#))

konkordance za 化ける: L1 ([24](#)), L0 ([108](#)), jpWaC ([743](#))

```
<entry xml:id="jaslo.8547">
  <form xml:lang="ja" type="hw">
    <orth type="roma">bakeru</orth>
    <orth type="kana">ばける</orth>
    <orth type="kanji">化ける</orth>
  </form>
  <gramGrp>
    <pos>V1</pos>
  </gramGrp>
  <cit type="translation" xml:lang="sl">
    <quote xml:lang="sl">pojavititi se pod krinko</quote>
```

ssj500k v2.2: training corpus

```

<s xml:id="ssj1.1.1">
  <pc ana="mte:Z" msd="UposTag=PUNCT" xml:id="ssj1.1.1.t1">"</pc>
  <w ana="mte:Pd-msg" lemma="tisti" xml:id="ssj1.1.1.t2">Tistega</w><c> </c>
  <w ana="mte:Ncmsg" lemma="večer" xml:id="ssj1.1.1.t3">večera</w><c> </c>
  <w ana="mte:Va-rls-n" lemma="biti" xml:id="ssj1.1.1.t4">sem</w><c> </c>
  <w ana="mte:Rgp" lemma="preveč" xml:id="ssj1.1.1.t5">preveč</w><c> </c>
  <w ana="mte:Vmep-sm" lemma="popiti" xml:id="ssj1.1.1.t6">popil</w>
  <pc ana="mte:Z" xml:id="ssj1.1.1.t7">,</pc><c> </c>
  <w ana="mte:Vmep-sn" lemma="zgoditi" xml:id="ssj1.1.1.t8">zgodilo</w><c> </c>
  <w ana="mte:Px-----y" lemma="se" xml:id="ssj1.1.1.t9">se</w><c> </c>
  <w ana="mte:Va-r3s-n" lemma="biti" xml:id="ssj1.1.1.t10">je</w><c> </c>
  <w ana="mte:Ncmsan" lemma="mesec" xml:id="ssj1.1.1.t11">mesec</w><c> </c>
  <w ana="mte:Ncmpg" lemma="dan" xml:id="ssj1.1.1.t12">dni</w><c> </c>
  <w ana="mte:S1" lemma="po" xml:id="ssj1.1.1.t13">po</w><c> </c>
  <w ana="mte:Pd-nsl" lemma="ta" xml:id="ssj1.1.1.t14">tem</w>
  <pc ana="mte:Z" xml:id="ssj1.1.1.t15">,</pc><c> </c>
  <w ana="mte:Cs" lemma="ko" xml:id="ssj1.1.1.t16">ko</w><c> </c>
  <w ana="mte:Va-rls-n" lemma="biti" xml:id="ssj1.1.1.t17">sem</w><c> </c>
  <w ana="mte:Vmep-sm" lemma="izvedeti" xml:id="ssj1.1.1.t18">izvedel</w>
  <pc ana="mte:Z" xml:id="ssj1.1.1.t19">,</pc><c> </c>
  <w ana="mte:Cs" lemma="da" xml:id="ssj1.1.1.t20">da</w><c> </c>
  <w ana="mte:Pp1-sa--y" lemma="jaz" xml:id="ssj1.1.1.t21">me</w><c> </c>
  <w ana="mte:Ncfsn" lemma="žena" xml:id="ssj1.1.1.t22">žena</w><c> </c>
  <w ana="mte:Vmpr3s" lemma="varati" xml:id="ssj1.1.1.t23">vara</w>
  <pc ana="mte:Z" xml:id="ssj1.1.1.t24">.</pc>
</s>

```

ssj500k v2.2: syntax

```
<linkGrp corresp="#ssj1.1.1" targFunc="head argument" type="UD-SYN">
  <link ana="ud-syn:punct" target="#ssj1.1.1.t6 #ssj1.1.1.t1"/>
  <link ana="ud-syn:det" target="#ssj1.1.1.t3 #ssj1.1.1.t2"/>
  <link ana="ud-syn:obl" target="#ssj1.1.1.t6 #ssj1.1.1.t3"/>
  <link ana="ud-syn:aux" target="#ssj1.1.1.t6 #ssj1.1.1.t4"/>
  <link ana="ud-syn:advmod" target="#ssj1.1.1.t6 #ssj1.1.1.t5"/>
  <link ana="ud-syn:root" target="#ssj1.1.1 #ssj1.1.1.t6"/>
  <link ana="ud-syn:punct" target="#ssj1.1.1.t8 #ssj1.1.1.t7"/>
  <link ana="ud-syn:parataxis" target="#ssj1.1.1.t6 #ssj1.1.1.t8"/>
  <link ana="ud-syn:expl" target="#ssj1.1.1.t8 #ssj1.1.1.t9"/>
  <link ana="ud-syn:aux" target="#ssj1.1.1.t8 #ssj1.1.1.t10"/>
  <link ana="ud-syn:obl" target="#ssj1.1.1.t8 #ssj1.1.1.t11"/>
  <link ana="ud-syn:nmod" target="#ssj1.1.1.t11 #ssj1.1.1.t12"/>
  <link ana="ud-syn:case" target="#ssj1.1.1.t14 #ssj1.1.1.t13"/>
  <link ana="ud-syn:nmod" target="#ssj1.1.1.t11 #ssj1.1.1.t14"/>
```

goo300k: Historical Slovene

```

<choice>
  <orig>
    <w>ludy</w>
  </orig>
  <reg>
    <w lemma="človek" ana="#Ncm">ljudi</w>
  </reg>
</choice>
<c> </c>
<lb/>
<choice>
  <orig>
    <w>memujete</w>
  </orig>
  <reg>
    <w lemma="mimo" ana="#Rgp">mimo</w>
    <c> </c>
    <w lemma="iti" ana="#Vmb">iti</w>
  </reg>
</choice>
<pc>,</pc>
<c> </c>
<w lemma="biti" ana="#Va">je</w>
<c> </c>

```

Parla-CLARIN and siParl 2.0

- Parliamentary data: openly available and very useful for a number of disciplines
- Many countries have already made available corpora of parliamentary debates, but each one encoded differently
- CLARIN organised several events and initiatives to promote and study parliamentary corpora (PC)
- In 2019 CLARIN also organised a workshop and financed work to develop a common encoding schema for PCs
- Cooperation between DARIAH-SI (Andrej Pančur) and CLARIN.SI (Tomaž Erjavec):
 - TEI based Parla-CLARIN schema
 - siParl 2.0: first corpus encoded in Parla-CLARIN

Parla-CLARIN

- TEI based Parla-CLARIN schema: <https://github.com/clarin-eric/parla-clarin>
- ODD document + XML schema + example folders
- Readable documentation in HTML: <https://clarin-eric.github.io/parla-clarin/>

clarin-eric / parla-clarin

Code Issues 5 Pull requests 0 Actions Projects 0 Wiki

Schema for modelling parliamentary debates <https://clarin-eric.github.io/parla-c...>

Manage topics

94 commits 1 branch 0 packages 1 release

Branch: master New pull request Create new branch

TomazErjavec Add link to GitHub page.

Examples	AKN2TEI conversion: HTML-like elements from AKN to TEI
Schema	Add link to GitHub page.
bin	Add validation and some changes.
docs	Correct some typos.
README.md	Fix and add READMEs.

Parla-CLARIN

A TEI Schema for Corpora of Parliamentary Proceedings

v0.1

Table of contents

1. [Introduction](#)
 - 1.1. [Scope and purpose](#)
 - 1.2. [Background](#)
2. [General requirements](#)
 - 2.1. [Characters](#)
 - 2.2. [Documenting the encoding process](#)
 - 2.3. [Languages](#)
 - 2.4. [Identifiers and referencing](#)
 - 2.5. [Temporal information](#)
 - 2.6. [Files](#)
3. [Overall document structure](#)
 - 3.1. [Corpus structure](#)
 - 3.2. [Text divisions](#)
 - 3.3. [Document variants](#)
4. [Corpus metadata](#)
 - 4.1. [Speaker metadata](#)
 - 4.2. [Party metadata](#)
 - 4.3. [Relationships between people and parties](#)
5. [Transcriptions](#)
 - 5.1. [Utterances and commentary](#)

siParl 2.0

- First corpus encoded in Parla-CLARIN
<http://hdl.handle.net/11356/1300>
- Slovene parliament 1990-2018
- Speaker meta-data
- Mandates, Parties, Committees, ...
- Speeches
- CLARIN.SI version: linguistic annotation
- Available in the repository and through CLARIN.SI concordancers

siParl 2.0: metadata

```
<org xml:id="party.Levica.1" role="political_party">
  <orgName full="yes" xml:lang="sl">Združena levica</orgName>
  <orgName full="yes" xml:lang="en">United Left</orgName>
  <orgName full="init">Levica</orgName>
  <event from="2014-03-01" to="2017-06-24">
    <label xml:lang="en">existence</label>
  </event>
  <idno type="wikimedia" xml:lang="sl">https://sl.wikipedia.org/wiki/Levica_(politi%C4%8Dna_stranka)</idno>
  <idno type="wikimedia" xml:lang="en">https://en.wikipedia.org/wiki/United_Left_(Slovenia)</idno>
</org>
```

```
<person xml:id="BavčarIgor">
  <persName>
    <surname>Bavčar</surname>
    <forename>Igor</forename>
  </persName>
  <sex value="M"/>
  <birth when="1955-11-28">
    <placeName ref="https://www.geonames.org/3196359">Ljubljana</placeName>
  </birth>
  <affiliation role="MP" ref="#DZ" from="1992-12-23" to="1996-11-27" ana="#DZ.1"/>
  <affiliation role="member" ref="#party.D" from="1992-12-23" to="1994-03-11" ana="#DZ.1"/>
  <affiliation role="member" ref="#party.LDS.2" from="1994-03-12" to="1996-11-27" ana="#DZ.1"/>
  <affiliation role="MP" ref="#DZ" from="1996-11-28" to="2000-10-26" ana="#DZ.2"/>
  <affiliation role="member" ref="#party.LDS.2" from="1996-11-28" to="2000-10-26" ana="#DZ.2"/>
  <affiliation role="MP" ref="#DZ" from="2000-10-27" to="2004-10-21" ana="#DZ.3"/>
  <affiliation role="member" ref="#party.LDS.2" from="2000-10-27" to="2004-10-21" ana="#DZ.3"/>
  <idno type="wikimedia" xml:lang="sl">https://sl.wikipedia.org/wiki/Igor_Bav%C4%8Dar</idno>
</person>
```

siParl: speeches

```
<div>
```

```
  <note type="time">Seja se je pričela ob 9.30 uri.</note>
```

```
  <note type="speaker">PRESEDNIK RAFAEL KUŽNIK:</note>
```

```
  <u who="#KužnikRafael" xml:id="KPZONJFSPD-Redna-021-1998-11-25.u1" ana="#chair">
```

```
    <seg xml:id="KPZONJFSPD-Redna-021-1998-11-25.seg1">
```

```
      <gap reason="inaudible"/> in potem še točko razno. Ali ima kdo kakšno pripombo k  
      dnevnemu redu? Če nima, bi prosil, da bi šli kar na obravnavo 1. točke.</seg>
```

```
    <seg xml:id="KPZONJFSPD-Redna-021-1998-11-25.seg2">Prehajamo na 1. TOČKO DNEVNEGA
```

```
      REDA - OBRAVNAVA SKLEPA DRŽAVNEGA ZBORA REPUBLIKE SLOVENIJE ŠT. 412.01/93-9/12,  
      EPA 619-II Z DNE 13.11.1998 O PREDLOGU ODLOKA O IMENOVANJU PREDSEDNIKA IN DVEH  
      ČLANOV UPRAVNEGA ODBORA SKLADA ZA FINANCIRANJE RAZGRADNJE NUKLEARNE ELEKTRARNE
```

```
      ..
```

```
      ..
```

```
<u who="#KužnikRafael" xml:id="KPZONJFSPD-Redna-021-1998-11-25.u1" ana="#chair">
```

```
  <seg xml:id="KPZONJFSPD-Redna-021-1998-11-25.seg1">
```

```
    <s xml:id="KPZONJFSPD-Redna-021-1998-11-25.seg1.s1">
```

```
      <gap reason="inaudible"/>
```

```
      <w xml:id="KPZONJFSPD-Redna-021-1998-11-25.seg1.s1.t1" ana="mte:Cc" msd="UposTag=CCONJ" lemma="in">in</w>
```

```
      <w xml:id="KPZONJFSPD-Redna-021-1998-11-25.seg1.s1.t2" ana="mte:Rgp" msd="UposTag=ADV|Degree=Pos" lemma="potem">potem</w>
```

```
      <w xml:id="KPZONJFSPD-Redna-021-1998-11-25.seg1.s1.t3" ana="mte:Q" msd="UposTag=PART" lemma="še">še</w>
```

```
      <w xml:id="KPZONJFSPD-Redna-021-1998-11-25.seg1.s1.t4" ana="mte:NcfSa" msd="UposTag=NOUN|Case=Acc|Gender=Fem|Number=Sing" lemma="pripombo">pripombo</w>
```

```
      <w join="right" xml:id="KPZONJFSPD-Redna-021-1998-11-25.seg1.s1.t5" ana="mte:Rgp" msd="UposTag=ADV|Degree=Pos" lemma="razno">razno</w>
```

```
      <pc xml:id="KPZONJFSPD-Redna-021-1998-11-25.seg1.s1.t6" ana="mte:Z" msd="UposTag=PUNCT">.</pc>
```

```
    <linkGrp corresp="#KPZONJFSPD-Redna-021-1998-11-25.seg1.s1" targFunc="head argument" type="UD-SYN">
```

```
      <link ana="ud-syn:cc" target="#KPZONJFSPD-Redna-021-1998-11-25.seg1.s1.t4 #KPZONJFSPD-Redna-021-1998-11-25.seg1.s1.t1"/>
```

```
      <link ana="ud-syn:root" target="#KPZONJFSPD-Redna-021-1998-11-25.seg1.s1 #KPZONJFSPD-Redna-021-1998-11-25.seg1.s1.t2"/>
```

```
      <link ana="ud-syn:root" target="#KPZONJFSPD-Redna-021-1998-11-25.seg1.s1 #KPZONJFSPD-Redna-021-1998-11-25.seg1.s1.t3"/>
```

```
      <link ana="ud-syn:root" target="#KPZONJFSPD-Redna-021-1998-11-25.seg1.s1 #KPZONJFSPD-Redna-021-1998-11-25.seg1.s1.t4"/>
```

```
      <link ana="ud-syn:acl" target="#KPZONJFSPD-Redna-021-1998-11-25.seg1.s1.t4 #KPZONJFSPD-Redna-021-1998-11-25.seg1.s1.t5"/>
```

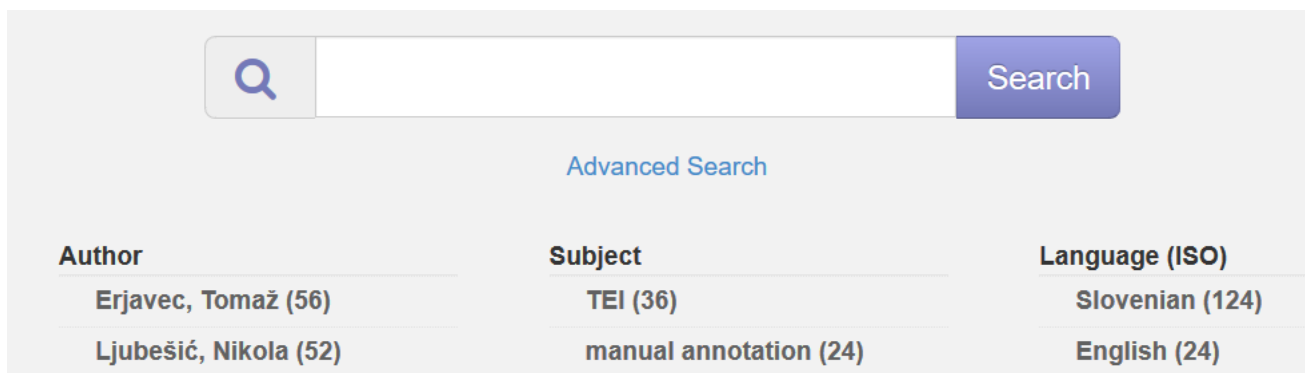
```
      <link ana="ud-syn:root" target="#KPZONJFSPD-Redna-021-1998-11-25.seg1.s1 #KPZONJFSPD-Redna-021-1998-11-25.seg1.s1.t6"/>
```

```
  </linkGrp>
```

```
</s>
```

Conclusions

- TEI is a large collections of elements and attributes for annotating various types of texts and their analyses
- Documented, maintained, with lots of support tools and a large user community
- Relatively popular in Slovenia in Digital Humanities, Corpus and Computational Linguistics
- Many CC TEI language resources (with down-conversions) are available in the CLARIN.SI repository



The screenshot shows the search interface of the CLARIN.SI repository. At the top, there is a search bar with a magnifying glass icon on the left and a blue 'Search' button on the right. Below the search bar, the text 'Advanced Search' is displayed in blue. The search results are presented in a table with three columns: Author, Subject, and Language (ISO). The table contains two rows of data.

Author	Subject	Language (ISO)
Erjavec, Tomaž (56)	TEI (36)	Slovenian (124)
Ljubešić, Nikola (52)	manual annotation (24)	English (24)