

Interoperability of an 18th century Italian-Latin-Croatian dictionary

Petra Bago, Damir Boras

Department of information and communication sciences,
Faculty of humanities and social sciences,
University of Zagreb
{pbago, dboras}@ffzg.hr

Introduction

- An increasing interest in digitizing process of historical texts
- A lack of communication between the community members
 - The digitization projects are usually isolated to the project teams, universities, institutions and individuals
- A demand for standardization of technologies and processes
 - A key concept emerges: Interoperability

Reaction to lack of communication

- International cooperation
- Sustainable Interoperability for Language Technology (SILT) (USA) and Fostering Language Resources Network (FlaReNet) (EU)
- The main goal: to create a consensus related to sharing data and technologies for language resources and applications
 - working towards the interoperability of existing data, and promote standards for markup and resource creation

Interoperability

- For computer systems:
 - Syntactic interoperability
 - to enable communication and data exchange, relying on specific data formats, communication protocols, and the like
 - important that information is exchanged
 - Semantic interoperability*
 - to automatically interpret exchanged information meaningfully and accurately in order to produce useful results via compliance to a common information exchange reference model
 - important that information is interpreted the same on both sides

TEI encoding scheme

- TEI (Text Encoding Initiative) encoding scheme for dictionaries
- enables semantic interoperability
 - exchange without or with minimal information loss, and correct interpretation of information
- the Guidelines recommend how to encode implicit features of textual resources, thereby making them explicit (*a de facto* standard)
- based on XML
- manual process

Encoding dictionaries

- The structure of entries
 - varies among and within dictionaries
 - a scheme should be suitable for various entry structures
 - complex but consistent structure
- The information found within entries
 - most information is implicit or compressed (lexicographical metadata)
 - to encode precise typographic form of the source text or the underlying structure of the information it presents

About della Bella's dictionary

- 1. volume of a second edition of “Dizionario italiano-latino-illirico” (Italian-Latin-Croatian dictionary) compiled by Ardelio della Bella and printed in Dubrovnik in 1785
- very complex entry structure (examples)
- intended for Italian Jesuist missionaries
- Croatian grammar in the preamble
- 899 pages, 2 parts (preamble + dictionary), 2 volumes (preamble, A-H + I-Z), ~19,000 headwords

Encoding of della Bella's dictionary

- to keep sequence of information found in the original text
- all additional information is encoded through attributes of elements

Abate . *Abbas* , *tis* . *m*. Opat , ta . *m*. Igu-|men , ena . *m*.
Dignità d'Abate . Opat-|stvo , va . *n*. Igumenstvo , tva .
m.

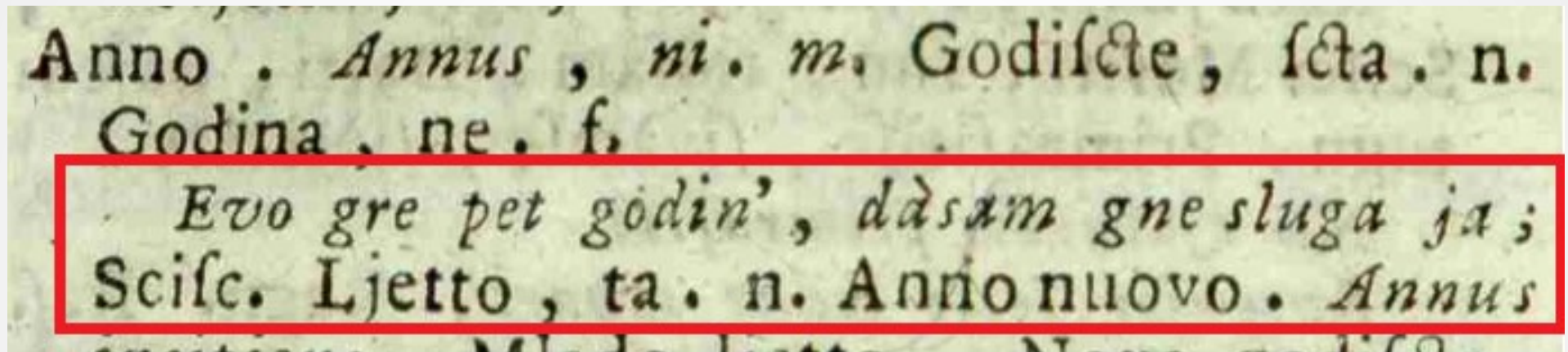
Encoding of della Bella's dictionary

Abate . *Abbas* , *tis* . *m* .
Opat , ta . *m* . Igu-|men , ena .
m . Dignità d'Abate . Opat-|
stvo , va . *n* . Igumenstvo ,
tva . *m* .

```
<form type="lemma"  
xml:lang="it">  
<orth>Abate</orth>  
<pc>.</pc>  
</form>
```

```
<quote>Abbas  
<pc>,</pc></quote>  
<form type="inflected">  
<gramGrp>  
<case value="genitive"/>  
<number  
value="singular"/>  
tis  
<pc>.</pc>  
</gramGrp>  
</form>
```

Encoding of della Bella's dictionary



```
<cit type="example" xml:lang="hr">  
<quote>Evo gre pet godin' , dàsam gne sluga ja  
<pc>;</pc></quote><lb/>  
<bibl>Scisc<pc>.</pc></bibl>  
</cit>
```

Conclusion

- To enable semantic interoperability of digitized historical dictionaries -> the dictionaries have to be encoded using some standard
- Successful encoding of della Bella's 18th century dictionary entries using a TEI (Text Encoding Initiative) encoding scheme (a *de facto* standard)
- Automatization of the encoding process?
- Linking to external resources (i.e. online encyclopedias)?

Thank you!