Slavisches Seminar
Freiburg University
Dr. Achim Rabus
achim.rabus@slavistik.uni-freiburg.de

Unicode and OpenType – a practical approach

27.10.2009

Outline

- n Presentation
- n integrated practical session

27.10.2009

Outline

- n *Presentation*
- n Multi-Purposing
- n What is Unicode and what is it good for?
- n What is OpenType and what is it good for?
- n Tools?
  - q InDesign
  - q Classical Text Editor
  - q XeLaTeX

Practical part

27.10.2009

Outline

- n *Presentation*
- n Multi-Purposing
- n What is Unicode and what is it good for?
- n What is OpenType and what is it good for?
- n Tools?
  - q InDesign
  - q Classical Text Editor
  - q XeLaTeX

27.10.2009

Multi-Purposing

- n Produce one file
- n Use it for multiple purposes

File

Printed edition   Website   Database   …

27.10.2009

Main questions in modern paleoslavistics

- n How do you (efficiently) create a file that can be used for multiple purposes?

- n How do you create a file that looks great when being used for printed editions and is correctly encoded when being used in databases etc.?

27.10.2009

Purists' solution

- Use a markup language, e.g. XML (eXtensible Markup Language)

```
<?xml version="1.0" encoding="UTF-8" standalone="yes" ?>
<!-- This XML document was generated by XCODEXML. -->
- <TRANDOC>
- <SOAP-ENV:Envelope
  xmlns:SOAP-ENV="http://schemas.xmlsoap.org/soap/"
  xmlns:uk="http://www.greenwichmeantime.co.uk"
  xmlns:us="http://www.easternstandardtime.com">
- <SOAP-ENV:Body>
- <TRAIN Date="18/03/2003" Time="13:00">
  <LOCOMOTIVE>
    <Name>Thomas</Name>
    <Length>12,500.00</Length>
    <Weight>3,400</Weight>
  </LOCOMOTIVE>
- <CARRIAGE>
    <Name>Annie</Name>
    <Length>10,000.00</Length>
    <Weight>2,000</Weight>
  - <Other_Information>
    Room for
    <count>100</count>
    standing & seated
    </Other_Information>
  </CARRIAGE>
- <CARRIAGE>
    <Name>Clarabel</Name>
    <Length>12,500.00</Length>
    <Weight />
  </CARRIAGE>
  </TRAIN>
  </SOAP-ENV:Body>
  </SOAP-ENV:Envelope>
  </TRANDOC>
```

-add tags
-make text machine-interpretable

27.10.2009

XML – Evaluation

Pro
- Complies with standards
- (e.g. Unicode)
- Structure of XML great to render manuscripts, texts, etc.
- Platform independent
- State of the art
- Mark up phenomena (e.g. ligatures) that are not included in the Unicode standard

Con
- My experience: Majority of philologists do not want to code in a text editor; they want to use a WYSIWYG tool such as Microsoft Word to prepare an edition

27.10.2009   XML is great, but **coding** XML can be painful…

XML: Practical solution

- Find a tool that allows you to enter text in an environment suitable for non computer-savvy philologists (preferably WYSIWYG)
- Find a way to use the documents produced with that tools for multiple purposes anyway

- Use XML without having to code XML

27.10.2009

Preferred characteristics of the tool

- Unicode
- WYSIWYG
- Export to XML

27.10.2009

## Outline

- *Presentation*
- Multipurposing
- What is Unicode and what is it good for?
- What is OpenType and what is it good for?
- Tools?
  - InDesign
  - Classical Text Editor
  - XeLaTeX

27.10.2009

Unicode

- What do you already know?

27.10.2009

Unicode

- … is an encoding system to encode all characters of the wold's language
- … is an organisation taking care of this task
- *„Unicode provides a unique number for every character,*
  no matter what the platform,
  no matter what the program,
  no matter what the language"
- http://www.unicode.org/standard/WhatIsUnicode.html
- No font dependency

27.10.2009   Section inspired by Kempgen 2008

Unicode and (O)CS

- New version of Unicode 5.1 (2008): Most (but not all) CS characters encoded
- Exceptions: see below

27.10.2009

Pre-Unicode times

- 2^8 = 256 glyphs per font
- Different codepages: either write Latin diacritica (č, ę etc.) oder Cyrillic
- OCS: Idiosyncratic encodings, special fonts

- Problem: Compatibility

27.10.2009

Problematic examples

- Birchbark documents
- http://gramoty.ru

- поклоно ѿ карпа · к осподину · моєму фоми

27.10.2009

Further examples

- АНУСКРНПТЪ
- http://manuscripts.ru
- н годнна бѣ ꙗко шестаꙗ. прнде жена ѿ самарнꙗ

- и година бѣ ꙗко шестаꙗ. приде жена ѿ самариꙗ

- Historical reasons

27.10.2009

What is the problem here?

- Idiosyncratic encodings
- No fall-back font, you cannot work with the texts unless you have the apropriate font
- You cannot put the idiosyncratically encoded data in a database together with standard-conform texts

-     Use Unicode to avoid problems!

27.10.2009

Why is Unicode so important?

Unicode: Conclusion

- Unicode is…
- platform independent
- font independent
- an indispensable prerequisite for the preparation of modern editions/databases etc.

- But: Does not solve all problems

Outline

- *Presentation*
- Multi-Purposing
- What is Unicode and what is it good for?
- What is OpenType and what is it good for?
- Tools?
  - InDesign
  - Classical Text Editor
  - XeLaTeX

Why is Unicode sometimes insufficient?

- New version of Unicode 5.1 (2008): Most (but not all) CS characters encoded
- Missing: Some superscripts
- Missing: Glyph variants
- Missing: Ligatures

- Unicode encodes characters, not glyphs
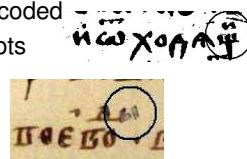
What is a character?

- "[A character is t]he smallest component of written language that has *semantic value*"
- http://www.w3.org/TR/charmod/

What is a glyph?

- "[Glyphs are] the basic units of organization of the *visual rendering* of text"
- http://www.w3.org/TR/charmod/

- A set of glyphs makes up a font

- Mnemonic: Character ≈ langue
                      Glyph ≈ parole

So what?

- Unicode encodes *characters*, not *glyphs*. It is interested in the *encoding level*, not in the *presentation level*
- Problem: Philologists are interested in the *presentation level*

- ѣ ƀ ꙗ etc. ꙗ ꙗ, б б etc.

What is OpenType?

- A font technology
- One font for all platforms
- Unicode

- OpenType features: Fonts are smart!
- Character (encoding) remains untouched, presentation form (glyphs) may change
- Several glyph forms for one character

Thanks to Adam Twardoch

Examples

- Ligatures: fl, ffi
- ѣ encoding: Ѣ+в
- ꙗ encoding: a+y ← important for electronic processing! (Multi purposing)
- Alternates: б vs. б
- *Russian vs. Serbian italics: n vs. ū; г vs. ī*

- (here faked)

Tech stuff

- OpenType features are organised in tables
- lig
- hist
- salt
- aalt
- locl etc.

- Access to the tables?

Problem

- OpenType is smarter than most applications
- Most unicode-savvy applications can handle OpenType fonts, but not their smart features

Your turn:

- What are the main benefits of OpenType?

- Preserve digital semantics while having optically faithful presentation at the same time.

- OpenType can solve many problems when preparing modern editions

27.10.2009

OpenType: applications

- Full OpenType support: InDesign and other Adobe applications
- Full OpenType support: Classical Text Editor (Windows only)
- Full OpenType support: XeLaTeX
- Partly OpenType support: Mellel (Mac only) (+ ligatures, OSF, smallcaps; - stylistic sets, historical alternates)

- Poor OpenType support: MS Word (may change in 2010), OpenOffice

27.10.2009

OpenType: applications

- Full OpenType support: InDesign and other Adobe applications
- Full OpenType support: Classical Text Editor (Windows only)
- Full OpenType support: XeLaTeX
- Partly OpenType support: Mellel (Mac only) (+ ligatures, OSF, smallcaps; - stylistic sets, historical alternates)

- Poor OpenType support: MS Word (may change in 2010), OpenOffice

27.10.2009

Browser support?

- Firefox 3 supports some OpenType features
- cf. http://mymapofjapan.com/styles.html

27.10.2009

Practical part

- Use OpenType-savvy applications to produce documents for multiple purposes

27.10.2009

InDesign

- powerful DTP application
- WYSIWYG
- easy access to OpenType features

- Please open the application

27.10.2009

InDesign: Showcase

- Old style figures
- stylistic alternates (л, д)
- Ligatures
- Russian vs. Serbian italics

Your turn

- Exercise I

InDesign: RTF export

- What we learn: OpenType glyph variants disappear, but information (= Unicode character) remains intact

What we learned yesterday

- Main goal: produce file that can be used for multiple purposes
- Unicode: encoding level (characters)
- OpenType: presentation level (glyphs)
- InDesign: Good access to OpenType features

Indesign: XML export

- Tag Text/Map paragraph styles to tags
- Import/create Tags <text>, <page> etc.
- View – Structure; Window – Tags
- Either: Full WYSIWYG markup or let the pros take care of further processing
- Markup of all OpenType features with e.g. <aalt> etc.
- Stylesheet will take care of glyphs

InDesign: Evaluation

Pro
- Unique typographic control (OpenType)
- decent XML functionality
- Academic license exists
- Platforms: Windows, Mac
- large user community

Con
- Besides OpenType-functionality: not smart
- Similar to manual typesetting
- no independent footnote streams
- no easy line numbering
- etc.
- not cheap…

Alternative: Classical Text Editor (CTE)

- Specialised tool for critical editions
- New version: OpenType capabilities!
- Exports TEI XML

- www.oeaw.ac.at/kvk/cte/

- Please open the application

CTE: Showcase

- Create notes for apparatus
- Ligatures
- Alternates
- locl-, style-features

- Problem: CampusRoman does not work properly in CTE
- Use OldStandard instead

CTE: Where to go

- Format – Document – Characters (choose OpenType)
- Options – preferences – Keyboard (choose set of alternates)
- Format – Font (choose OT features, stylistic sets, language)
- Edit – Alternate Glyph (or CTR-.) (choose glyphs)

CTE

- XML export: Automatically uses tei.dtd, adds some basic tags
- Insert tags manually: Insert: XML/TEI tags

Your turn

- Exercise II

CTE: Evaluation

Pro
- Many special functions that are helpful to produce critical editions
- Decent export capabilities

Con
- Windows only
- commercial license
- proprietary
- Does not display all OpenType features correctly

and still: great tool to produce

Conclusion

- n Why use OpenType?
- n Faithful encoding
- n Nice optical presentation
- n OpenType-savvy applications can be helpful tools to produce state-of-the-art documents for multiple purposes without having to code

Thank you very much!

Alternative: XeLaTeX

- n What is LaTeX?

- n XeLaTeX = LaTeX + Unicode + OpenType

XeLaTeX

- n You need: modern TeX distribution (e.g. TeXLive, MikTeX etc.)
- n Unicode-capable editor (e.g. TeXMaker, TeXShop, TeXworks etc.)

XeLaTeX: How it works

- n Simply use „xelatex document.tex" instead of „latex document.tex" or „pdftex document.tex" to compile your document

XeLaTeX: What to do

- n some additional commands in the preamble

- n \usepackage{fontspec}
- n \defaultfontfeatures{Mapping=tex-text}
- n \usepackage{xunicode}
- n \usepackage{xltxtra}

- n do NOT use \inputenc or \fontenc

XeLaTeX: Working with fonts

- Use fontspec to change fonts
- Define font features
- Use OpenType features

(Xe)LaTeX: Typing editions

- Numerous possibilities
- My choice: package ledmac
- Documentation: see http://tug.ctan.org/tex-archive/macros/latex/contrib/ledmac/ledmac.pdf

XeLateX

- Exercise III