$\mu o\nu o\vec{2}\pi o\lambda\upsilon$: Java-based Conversion of Monotonic to Polytonic Greek

Johannis Likos
ICT Consulting
Rusthollarintie 13 E 35
Helsinki 00910
Finland
likosjo@yahoo.com

Abstract

This paper presents a successfully tested method for the automatic conversion of monotonic modern Greek texts into polytonic texts, applicable on any platform. The method consists of combining various freely available technologies, which have much better results than current commercially available solutions. The aim of this presentation is to introduce a way of applying this method, in order to convert thousands of digitally available single-accented modern Greek pages into attractive artworks with multi-accented contents, which can be easily transferred either to the Web or a TeX-friendly printer. We will discuss the preparatory and postprocessing efforts, as well as the editing of syntax rulesets, which determine the quality of the results. These rulesets are embedded in extendable tables, functioning as flat databases.

1 Introduction

During the past centuries, Greek and Hellenic scholars have introduced and refined polytonism (multiple accenting) in the written word for the precise pronounciation of ancient Greek. Since spoken modern Greek is comparatively less complicated, the Greek government has officially replaced polytonism by monotonism (single accenting) for purposes of simplification, especially in the educational system. Also, Greek authors commonly use monotonism, since it is so much simpler to produce.

Classical, polytonic, Greek has three accents (acute, grave, and circumflex) and two *breathings* (rough and smooth—equivalent to an initial 'h' and lack thereof). Accents are lexically marked, but can change based on other factors, such as *clitics* (small, unstressed words that lean on another word to form a *prosodic word*—a single word for accent placement). In addition, two other symbols were used: diaeresis (to indicate two vowels that are not a diphthong) and iota subscript (a small iota that was once part of a diphthong but subsequently became silent).

Monotonic Greek retains only the acute accent, which was usually, though not always, the same as the classical acute. To make a graphic break with the past, the new acute accent was written as a new *tonos* glyph, a dot or a nearly vertical wedge, although this was officially replaced by a regular acute in 1986.

So, why bother with the complexities of polytonism? The benefits are increased manuscript readability and, even more important, reducing ambiguity. Despite the simplification efforts and mandates, the trend nowadays is *back to the roots*, namely to polytonism. More and more publishers appreciate, in addition to the content, the public impression of the quality of the printed work.

This paper discusses an innovative and flexible solution to polytonism with an open architecture, enabling the automatic multiple accenting of existing monotonic Greek digital documents.

2 Terminology

In this article, we will use the terms *polytonism* and *multiple accenting* interchangeably to mean the extensive usage of *spiritus lenis, spiritus asper, iota subscript, acute, gravis* and *circumflex*. Similarly, we use the terms *monotonism* and *single accenting* to mean the usage of simplified accenting rules in Modern Greek documents.

3 Historic Linguistic Development

During the last four decades the printed Greek word has undergone both minor and radical changes. Elementary school text during the late 1960s and early 1970s made Purified Greek (καθαρεύουσα) imperative, by strict government law of the time. The mid-1970s saw a chaotic transition period from Purified Greek to Modern Greek (δημοτική) with simplified

grammar, where some publications were printed with multiple accenting, some with single accenting and even some without any accenting at all!

Even after the government officially settled on monotonism in the early 1980s, Greek publishers were not able to switch immediately to the monotonic system. During the last decade, many computerized solutions have been invented for assistance in typing monotonic Greek. Today, there is a trend toward a mixture of simplified grammar with multiple accenting, decorated with Ancient Greek phrases. See table 1.

4 Polytonic Tools

There are two programs for Microsoft Word users, namely *ΤΟΝΙΣΜΟΣ* by DATA-SOFT and *ΑΥΤΟ-ΜΑΤΟΣ ΠΟΛΥΤΟΝΙΣΤΗΣ* (academic and professional version) by MATZENTA. A third is the experimental $\mu o \nu o \vec{2} \pi o \lambda \upsilon$, an open source project, which is the subject of this discussion.

These solutions are independent. The major difference between the commercial and open source programs is the control of the intelligence, such as logic, rule sets and integrated databases. In the case of the commercial solutions, users depend on the software houses; in the open source case, users depend on their own abilities. See table 2.

There is no absolutely perfect tool for polytonism, so the ultimate choice is of course up to users themselves.

5 Open Source Concept

$\mu o \nu o \vec{2} \pi o \lambda \upsilon$ implements a modular mechanism for multiple accenting of single-accented Greek documents. See figure 1.

5.1 Architecture

The $\mu o \nu o \vec{2} \pi o \lambda \upsilon$ architecture consists of (figure 2):

- methods: `DocReader`, `DocWriter`, `DBParser`, `Converter`
- configuration file: `*.cfg`
- flat database: `*.xml`
- document type definition: `*.dtd`
- optional spreadsheet: `*.csv`, `*.xls`

5.2 Configuration

The plain text configuration file defines (gray arrows in fig. 2) the necessary filenames and pathnames. The $\mu o \nu o \vec{2} \pi o \lambda \upsilon$ components read this during initialization to determine where to find the input files and where to write the output files (by default, the current working directory).
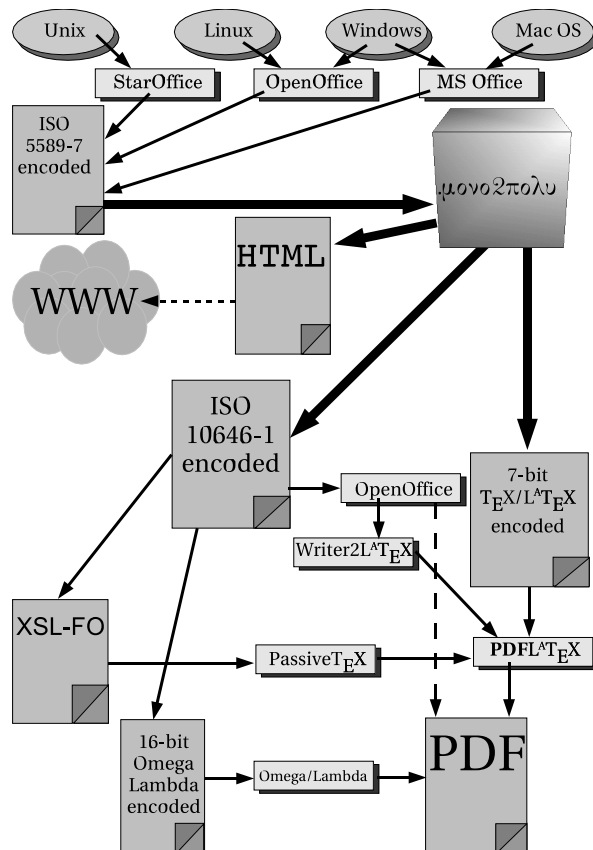

Figure 1: Overview of the overall multiple accenting concept, which involves many external tools.

5.3 Database Connectivity

The dotted arrows in the architecture figure (fig. 2) show the connection between a CSV spreadsheet, a Document Type Definition (DTD), the actual XML flat database, and the database parser.

5.4 Input Prerequisites

During the conversion process, invisible special control codes for formatting features (superscript, bold, etc.) make it difficult to coherently step through paragraphs, sentences and words. Therefore, plain text files serve best for polytonism input.

The `DocReader` component of $\mu o \nu o \vec{2} \pi o \lambda \upsilon$ expects the source document to be in the ISO 5589-7 encoding, and to be written according to the Modern Greek grammar, especially regarding the monotonic accenting rules.

Assistance for monotonic accenting while typing Modern Greek documents is provided by Microsoft's commercially bundled spellchecker, or any downloadable Open Source spellchecker.

Table 1: Selected examples of Greek publications from the last four decades.

Date	Publisher	Author	Subject/Title	Language	Remarks	Example
1968	ΕΚΔΟΣΕΙΣ Α. ΚΑΡΑΒΙΑ	S. TIMOSHENKO D.H. YOUNG	Ἀντοχὴ τῶν Ὑλικῶν	purified	polytonic translation including gravis	
1971	Ο.Ε.Δ.Β.	Μ. ΚΑΤΣΙΚΑΣ	ΓΕΩΓΡΑΦΙΑ ΣΤ' ΔΗΜΟΤΙΚΟΥ	purified	no dative at all and iota subscript seldom used	νὰ ἐπιδιωχϑῇ
1978	ΙΔΡΥΜΑ ΕΥΓΕΝΙΔΟΥ	Ι. ΧΑΣΤΑΣ	βιβλ. Τεχν. κ. Ἐπαγγ. Λυκείου	modern	polytonic without gravis	
1980	self-published	Ρ. ΓΡΑΙΚΟΥΣΗΣ	ΣΤΟΙΧΕΙΑ ΜΗΧΑΝΩΝ	modern	acute used instead of gravis in polytonic text, some feminine singular genitive in purified version	τῆς κλίσεως
1982	self-published	Ε.Σ.Μ.Α.	ΠΡΑΚΤΙΚΑ ΕΘΝ. ΑΕΡΟΠΟΡΙΚΟΥ ΣΥΝΕΔΡΙΟΥ	modern	monotonic typewriter text style	
1989	INTERBOOKS	ΓΡ. ΣΦΑΚΙΑΝΟΣ	ΕΜΠΟΡΙΚΗ ΑΛΛΗΛΟΓΡΑΦΙΑ	modern	monotonic with neutral accent and acute	εκτός από τόν
1998	ΠΑΡΑΤΗΡΗΤΗΣ	Α. ΣΥΡΟΠΟΥΛΟΣ	LᴬTEX	modern	acute only accent type in monotonic text	τό τόπι
2003	Α. ΦΩΤΙΕΡΗΣ	several authors	ἡ λέξη	modern	polytonic without gravis	καί ἀπ᾽ τό

Table 2: Comparison of polytonic tools.

	ΤΟΝΙΣΜΟΣ	ΑΥΤΟΜΑΤΟΣ ΠΟΛΥΤΟΝΙΣΤΗΣ	μονο2πολυ
Greek language support			
Ancient	yes	yes	later
Hellenistic	yes	yes	later
Byzantine	no	no	later
Church/Biblical/NT	yes	yes	later
Purified	yes	yes	later
Modern	yes	yes	yes
Mixed (Ancient and Modern)	selectable	selectable	fixed
Editing assistance	no	yes	no
Database	fixed (binary)	fixed (binary), editable exception list	editable (XML)
Manual corrections		interactively	post-processed
Automatic hyphenation	unknown	yes	external task
Protection	hardlock	ID	no
File formats			
Input	Word (Win, Mac)	Word (Win, Mac) other Greek formats	ISO 5589-7 encoded ASCII on any platform
Output	Word (Win, Mac)	Word (Win, Mac) other Greek formats, HTML	ASCII (ISO 10646-1 encoded), HTML (ISO 10646-1 encoded)
Unicode support	yes	yes	yes
TEX-specific filters	`TeXto737.lex`, `737toTeX.lex`	no	Writer2LATEX
Requirements	Microsoft Word	Microsoft Word	JDK 1.4
Platforms			
Microsoft Windows	95, 98, ME NT, 2000, XP	95, 98, ME NT, 2000, XP	95, 98, ME NT, 2000, XP
Apple Macintosh	no	no	Mac OS 9, Mac OS X
Linux	no	no	Mandrake, Red Hat, SuSE
Unix	no	no	AIX, HP-UX, Sinix, Solaris
Availability	immediately	immediately	under development
Distribution	purchased license	purchased license	open source

5.5 Converter

The bold arrows in the architecture figure (fig. 2) show the data exchange between the internal components, the document reader, the database parser and the document writer to the converter. The conversion process does not include grammar analysis, since μονο2πολυ expects that monotonic proof reading has been done previously, with other tools.

6 External Interfaces

The output from μονο2πολυ (`DocWriter` method) is in the ISO 10646-1 encoding, in various formats, which are then post-processed. The dashed arrows in fig. 2 show the relationship between the external files.

6.1 Web Usage

For background, these web pages discuss polytonic Greek text and Unicode[1] (UTF) fonts:

---

[1] Concerning missing Greek characters and other linguistic limitations in Unicode, see *Guidelines and Suggested Amendments to the Greek Unicode Tables* by Yannis Haralambous at the 21st International Unicode Conference
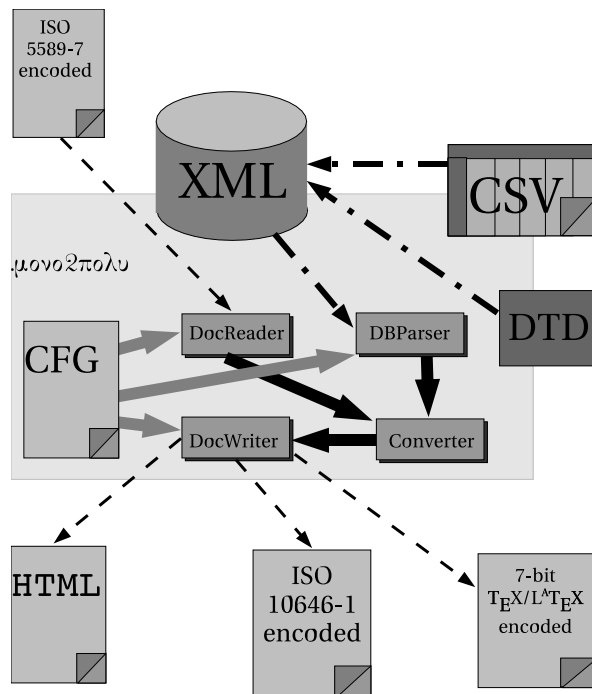
Johannis Likos

Figure 2: Overview of internal architecture, with external interfaces to existing standards.
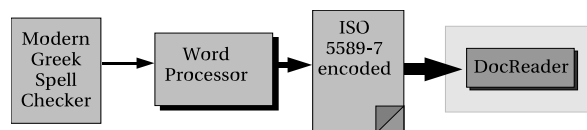

Figure 3: External creation of monotonic Greek with any word processor (e.g., OpenOffice) using a separate spellchecker (e.g., elspell).

- http://www.ellopos.net/elpenor/lessons/lesson2.asp
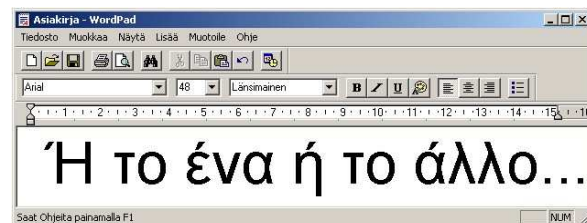- http://www.stoa.org/unicode/
- http://www.mythfolklore.net/aesopica/aphthonius/1.htm

on May 2002 in Dublin, Ireland (http://omega.enstb.org/yannis/pdf/amendments2.pdf).

Figure 4: Example of original monotonic input.

Using the direct HTML polytonic output from $\mu o\nu o\vec{2}\pi o\lambda\upsilon$ requires that the layout of the web page be done in advance, since manually editing the numeric Unicode codes in the *.html file is impractical (see figure 5). Dynamic web pages created through CGI scripts, PHP, etc. have not yet been tested.

```
<font
FACE="Arial Unicode MS"
SIZE="36">
&#7978;&#32;
&#964;&#8056;&#32;
&#7957;&#957;&#945;&#32;
&#7970;&#32;
&#964;&#8056;&#32;
&#7940;&#955;&#955;&#959;;
&#46;&#46;&#46;
</font>
```

Figure 5: Example of polytonic HTML output.

6.2 OpenOffice Usage

The ISO 10646-1 encoded polytonic output (fig. 6) from $\mu o\nu o\vec{2}\pi o\lambda\upsilon$ could be inserted into the OpenOffice Writer software, since the newest version can directly output polytonic Greek .pdf files. Unfortunately, the quality of the result leaves much to be desired. Better results can be produced by converting from Writer to LaTeX and doing further processing in the LaTeX environment.

Figure 6: Example of polytonic output.

6.3 (LA)TEX Usage

The most likely scenario for (LA)TEX users is using the Greek babel package, and adding the $\mu o\nu o\vec{2}\pi o\lambda\upsilon$ 7-bit polytonic output text into the source .tex file. See figures 7 and 8.

The 7-bit output from $\mu o\nu o\vec{2}\pi o\lambda\upsilon$ could presumably also be inserted into .fo files, and processed through PassiveTEX, but this has not yet been tested. Likewise, the ISO 10646-1 output could presumably be processed directly with Ω/Λ, but this has not been tested, either.
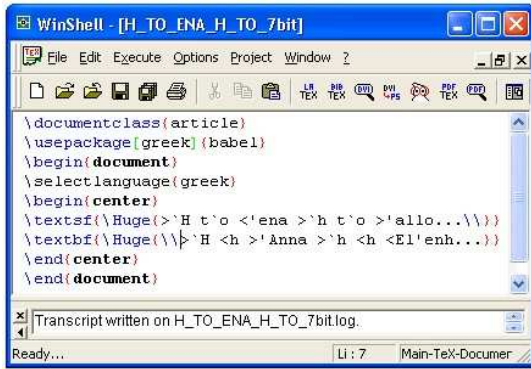
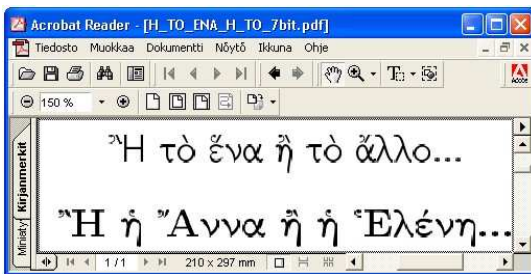Figure 7: Example of polytonic TₑX output, either from μονο2̇πολυ or Writer2LᴬTₑX.

Figure 8: Polytonic PDF output from TₑX.

7 Technologies Used in μονο2̇πολυ

After some evaluation, we chose to focus on Java, Unicode and XML, due to their flexibility in processing non-Latin strings, obviously a critical requirement of μονο2̇πολυ.

7.1 Programming Language

Two major reasons for choosing Java (J2SE) as the implementation language of μονο2̇πολυ were the capabilities for handling XML and Unicode through widely-available and well-documented libraries. The Java SDK provides extremely useful internationalization features, with the ability to easily manipulate string values and files containing wide characters.

In order to concentrate on μονο2̇πολυ's essential features, no graphical user interface has been designed.

7.2 Character Set

The choice of Unicode/ISO 10646-1 for the character set should be clear. It combines monotonic and polytonic Greek letters, is known worldwide and standardized on most platforms, and contains most (though not all) variations of Greek vowels and

consonants, in the Greek and the Greek Extended tables.[2]

For further information on writing polytonic Greek text using Unicode, see `http://www.stoa.org/unicode/`.

7.3 Text Parsing Libraries

Most helpful for the parsing of XML-based database entries are the SAX and DOM Java libraries.

The following Java source code, taken from the μονο2̇πολυ class `DBparse`, serves to demonstrate usage of SAX and DOM. The code counts and then outputs the total amount of all available entries in the XML database file.

```java
import java.io.*;
import org.xml.sax.SAXException;
import org.xml.sax.SAXParseException;
import javax.xml.parsers.DocumentBuilder;
import javax.xml.parsers.DocumentBuilderFactory;
import javax.xml.parsers.FactoryConfigurationError;
import javax.xml.parsers.ParserConfigurationException;
import org.w3c.dom.*;
public class DBparse{
  static Document document;
  String warn="No XML database filename given...";
    public static void main(String param[]){
        if (param.length!=1){
            System.out.println(warn);
            System.exit(1);}
        File mydbfile=new File(param[0]);
        boolean load=mydbfile.canRead();
        if (load){
          try{
            DocumentBuilderFactory fct
              = DocumentBuilderFactory.newInstance();
            DocumentBuilder builder
              = fct.newDocumentBuilder();
            document = builder.parse(mydbfile);}
          catch (SAXParseException error){
            System.out.println("\nParse error at line: "
                + error.getLineNumber() + " in file: "
                + error.getSystemId());
            System.out.println("\n" + error.getMessage() );}
          catch (ParserConfigurationException pce)
                {pce.printStackTrace();}
          catch (IOException ioe){ioe.printStackTrace();}
          catch (Throwable t){t.printStackTrace();}}
        else{System.out.println("XML database missing!");}
      String mytag='\u03C3'+"";
      NodeList taglist=document.getElementsByTagName(mytag);
      int amount=taglist.getLength();
      System.out.println("amount of entries:\n" + amount );}}
```

Notice particularly the fourth-last line, where `mytag` is assigned `'\u03C3'`, namely the character σ, used as the search string.

8 Database Structure

The XML standard from the W3C has proven to be a simpler choice for storing either monotonic or polytonic Unicode text than the alternatives, such as spreadsheets or even SQL databases. The quality of the final polytonic result depends on the precision of the XML content, where ambiguities have to

---

[2] `http://www.unicode.org/versions/Unicode4.0.0/ch07.pdf`

be marked with special symbols for manual post-processing.

Currently, the entries of the basic database consist of tags with parameters and values. The tag name indicates the type of the expression: a single character, a prefix, a suffix, a substring, a word or a chain of words. The five parameters are as follows:

1. The monotonic ISO 5589-7 encoded source expression to be converted.

2. The Unicode output text.

3. A 7-bit output text for (LA)TEX usage with the Greek babel package.

4. The equivalent numeric value according to the *Extended Greek* Unicode table for HTML usage.

5. An explanatory comment or example, in case of ambiguities or linguistic conflicts.

Here, I have built on the work of prior Greek TEX packages, such as GreekTEX (K. Dryllerakis), Scholar TEX (Y. Haralambous), and greektex (Y. Moschovakis and G. Spiliotis), for techniques of using the iota subscript, breathings and accents in 7-bit transliterated .tex source files.

In the following examples, note carefully the different bases used: '074 is octal, #8172 is decimal and '03D1' is hexadecimal.

8.1 Data Type Definition

The required basic *Data Type Definition* is currently located in the experimental namespace xmlns:β = http://koti.welho.com/ilikos/TeX/ LaTeX/mono2poly/mono2poly.dtd. It contains the following information:

```
<!ELEMENT β (σ+)>
<!ELEMENT σ (#PCDATA)>
<!ATTLIST σ
  μ CDATA #REQUIRED
  π CDATA #REQUIRED
  τ CDATA #REQUIRED
  δ CDATA #REQUIRED
  ξ CDATA #REQUIRED>
```

Thus, we have one element, called β (βάση δεδομένων = *database*). It contains multiple element sets, called σ (στοιχεῖα συλλαβῆς = *syllable data*). Each element set has, at present, five attributes, namely μ for monotonic expressions, π for polytonic expressions, τ for 7-bit (LA)TEX code, δ for HTML code, and finally ξ for comments.

The DTD can be overridden by a local .dtd file, which must be specified in the header of the .xml database file; for example:

```
<!DOCTYPE β SYSTEM "my_own_mono2poly.dtd">
```

Both the .dtd and .xml must reside in the same directory.

8.2 Data Entries

Here is an example database entry, showing the only Greek capital consonant with spiritus asper:

```
<σ
    μ="P"
    π="'P"
    τ="\char'074 R"
    δ="&#8172;"
    ξ="'Ρόδος"
></σ>
```

The *slash* symbol indicates the closing element tags in XML, while the *backslash* symbol is used for (LA)TEX commands. Both appear in the .xml database file.

8.3 Header and Body

Although not explicitly documented, exotic characters may be used in .dtd and .xml files as long as the appropriate encoding is declared:

```
<?xml version="1.0" encoding="UTF-16"?>
```

The header should include other information as well. Schematically:

```
<?xml version="1.0" encoding="UTF-16"?>
<!DOCTYPE β SYSTEM "mono2poly.dtd">
<!-- author: ... -->
<!-- affiliation: ... -->
<!-- creation date: ... -->
...
<!-- notes: ... -->
<β>
 <σ μ="..." π=".." τ=".." δ=".." ζ=".."></σ>
 ...
 <σ μ="..." π=".." τ=".." δ=".." ζ=".."></σ>
</β>
```

For quality assurance, after database creation and after each update a validation and verification test should be run, to detect XML syntax errors and linguistic content mistakes.

This concept of the database as a lookup/mapping table allows differentiating between initial and intermediate consonants. For example:

$$ϐ\ (03D0) \leftrightarrow β\ (03B2)$$
$$ϑ\ (03D1) \leftrightarrow θ\ (03B8)$$
$$ϱ\ (03F1) \leftrightarrow ρ\ (03C1)$$
$$φ\ (03D5) \leftrightarrow \phi\ (03C6)$$

Therefore, by updating the XML file, post-processing may be reduced. Experienced linguists may wish to use different tools for the correcting and the updating of the flat database. Rows with multiple columns from spreadsheets can be inserted directly into XML data files, as long as the columns are sorted in the expected order.

8.4 Expression Types

In each database entry, there is one source expression, at least three target expressions, and possibly one explanation. The ISO 5589-7 encoded source expression and the first ISO 10646-1 encoded target expression may be a:

- single uppercase or lowercase character with or without spiritus and/or accent
- partial word, such as prefix, intermediate syllable, suffix
- complete word
- chain of combined words
- combination of partial word pairs, such as a suffix followed by a prefix
- mixture of complete and partial words, such as a complete word followed by a prefix or a suffix, followed by a complete word

The rest of the target expressions represent the same information as the first in other output formats, namely for 7-bit Greek (LA)TEX and HTML as well. The intelligence of the $\mu o\nu o\vec{2}\pi o\lambda\upsilon$ system currently lies in the database, so while creating and editing entries, it is crucial to write them correctly.

8.5 Editing Tools

One of the most powerful Unicode editors is the Java-based Simredo 3.x by Cleve Lendon, which has a configurable keyboard layout, and is thus suitable for this sort of task. The latest version of Simredo, 3.4 at this writing, can be downloaded from `http://www4.vc-net.ne.jp/~klivo/sim/simeng.htm`, and installed on any platform supporting JDK 1.4.1 from Sun. Simredo can be started by typing `java Simredo3` or perhaps `java -jar Simredo3.jar` in the shell window (Linux) or in the command window (Windows). Unicode/XML with Simredo has been successfully tested on Windows XP and on SuSE Linux 8.1 Professional Edition.

The author would be happy to assist in the preparation of a polytonic Greek keymap file (`.kmp`) for Simredo, but the manual may prove sufficient. The creation of such a keymap file is easily done by simply writing one line for each key sequence definition. For instance, given the sequence `2keys;A˝A` using the desired Unicode character, or the equivalent sequence `2keys;A\u1F0D` with the big endian hexadecimal value, one can produce an uppercase *alpha with spiritus asper and acute accent* by pressing the ⌷;⌷ and ⌷A⌷ keys simultaneously. According to the Simredo manual, other auxiliary keys such as `Alt` can be combined with vowel keys, but not `Ctrl`. Some other Unicode editors:
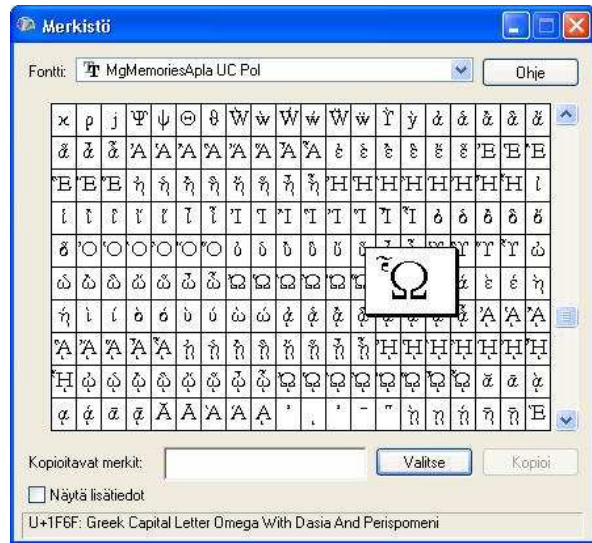

Figure 9: Another useful tool is a character mapping table like this accessory on Windows XP, which displays the shape and the 16-bit big endian hexadecimal code of the selected character.

- For Windows: `http://www.alanwood.net/unicode/utilities_editors.html`.
- For Linux: `http://www.unicodecharacter.com/unicode/editors.html`.
- For Mac OS: `http://free.abracode.com/sue/`.

Unfortunately, XMLwriter and friends neither support configurable keyboard layouts nor display 16-bit Unicode.

8.6 Polytonic Keyboard Drivers

Instant interactive multi-accenting while editing Greek documents is available either through plug-ins for some Windows applications, such as SC Unipad (`http://www.unipad.org/main/`) and Antioch (`http://www.users.dircon.co.uk/~hancock/antioch.htm`), or with the help of editable keyboard mapping tables, such as the Simredo Java program described above. Regrettably, the *Hellenic Linux User Group* (HEL.L.U.G., `http://www.hellug.gr` and `http://www.linux.gr`) has no recommendations for polytonic Greek keybord support.

Whatever polytonic keyboard driver has been installed and activated may be useful for new documents, but does not much help the author who is not familiar with the complicated rules of polytonism!
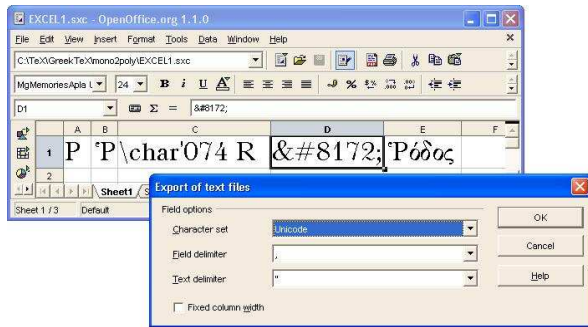
Figure 10: Using a spreadsheet to produce a long extendable list with five columns, which then can be saved as a `.csv` file. Be careful with the parametrization!
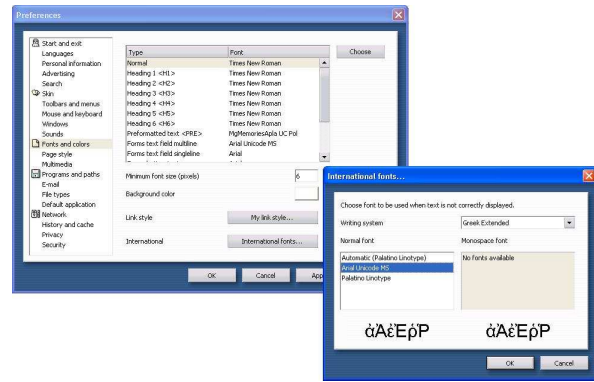
8.7 Auxiliary Tables

Preparation and periodic updates of auxiliary tables can of course be done with any software supporting Unicode. Spreadsheets have the advantage of putting the entries into cells row-by-row and thus organizing the parameters by column. This may prove easier than directly writing the XML file. See figure 10.

A row in such a `.csv` file looks like this:

`"P","῾P","\char'074 R","Ῥ","῾Ρόδος"`

Of course it then must be re-shaped with element and attribute tags to make an XML-syntax database entry.

8.8 Viewing Tools

Users without any programming knowledge may find it useful to open and inspect the header and the body of the XML database before using it in polytonic documents. Here is a procedure for doing that.

First, set the Unicode font in the preferences of the desired browser (fig. 11). These days, most browsers support this, including Internet Explorer, Konqueror, Netscape Navigator and Opera.

Then, select Unicode UTF-16 as the default encoding (fig. 12). The browser can now detect syntax errors, giving immediate feedback (fig. 13).

8.9 Priorization of Database Entries

Polytonic exceptions (e.g., οὔτε and ὥστε without circumflex) and especially ambiguities (e.g., που → ποὺ or πού, πού → ποῦ; πως → πὼς or πώς, πὼς → πῶς) have the highest priority in the database, then the special expressions, while the simple, casual and obvious accented syllables or particles have the lowest priority. In order to avoid mis-accented and

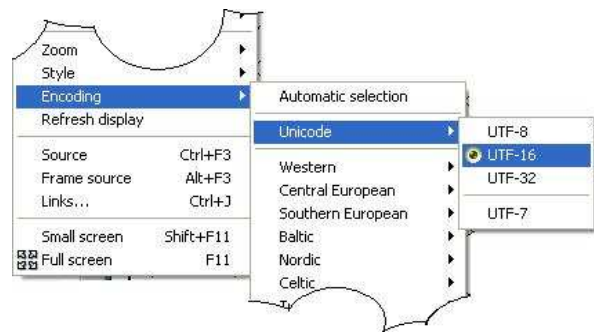Figure 11: Choosing the Unicode font for viewing in a browser.

Figure 12: Selecting UTF-16 for the default encoding.

mis-spirited syllables as much as possible, entries must be in the appropriate order.

For example, table 3 shows lexical rules defining eight variations of the Greek interrogative pronoun τί (= "which") as a single monotonic expression:

- with and without neutral accent
- with and without Greek question mark
- standalone
- leading word in the sentence
- intermediate word in the sentence
- trailing word in the sentence

Database entries like these are needed to account for the variations shown in table 4. As a rough analogy in English, it is as if table 3 shows variations on "I": Initial position ("I went to the store"); after

Figure 13: Example error message from browser.

Table 3: Eight variations of τί as a monotonic expression.

```
<!--   ἐρωτηματικὲς ἀντωνυμίες   -->
<σ μ="Τι;" π="Τί;" τ="Τ'i;" δ="&#932;&#8055;&#894;" ξ="Τί;"> </σ>
<σ μ="Τι " π="Τί " τ="Τ'i " δ="&#932;&#8055;&#32;" ξ="Τί λές;"> </σ>
<σ μ=" τι;" π=" τί;" τ=" t'i;" δ="&#32;&#964;&#8055;&#894;" ξ="Κι᾽ ἔγραψε τί;"> </σ>
<σ μ=" τι " π=" τί " τ=" t'i " δ="&#32;&#964;&#8055;&#32;" ξ="Καὶ τί ἔγραψε;"> </σ>
<σ μ="Τι;" π="Τί;" τ="Τ'i;" δ="&#932;&#8055;&#894;" ξ="Τί;"> </σ>
<σ μ="Τι " π="Τί " τ="Τ'i " δ="&#932;&#8055;&#32;" ξ="Τί λές;"> </σ>
<σ μ=" τι;" π=" τί;" τ=" t'i;" δ="&#32;&#964;&#8055;&#894;" ξ="Κι᾽ ἔγραψε τί;"> </σ>
<σ μ=" τι " π=" τί " τ=" t'i " δ="&#32;&#964;&#8055;&#32;" ξ="Καὶ τί ἔγραψε;"> </σ>
```

Table 4: Similar but unrelated syllables treated differently.

Position	Syllable	Polytonic examples
leading	τι-, Τι-	τιμή, Τισσαφέρνης
intermediate	-τι-	ἀδυνάτισμα
trailing	-τι	κάτι, πράγματι
leading	τί-, Τί-	τίποτα, Τίγρης
intermediate	-τί-	ἐκτίμηση
trailing	-τί	γατί, ψωμί, τυρὶ

a verb ("What do I know?"); etc., and then table 4 shows that "I" isn't always capitalized: "It looks good" vs. "Can you see it?"

The above does not cover all cases related to τί. The monotonic text γιατι may be accented in two different ways in polytonic Greek (namely, γιατὶ and γιατί); additional entries would be required to handle this.

9   Polytonic Printing Press

The author has found several Greek newspapers, magazines, and books, including university presses, using polytonic Greek:

- daily newspapers:
  - *Η ΚΑΘΗΜΕΡΙΝΗ*
  - *ΕΣΤΙΑ*
- monthly magazines:
  - *ΝΕΜΕΣΙΣ*
- book publishers:
  - *ΕΚΔΟΤΙΚΗ ΑΘΗΝΩΝ*
  - *ΚΥΡΙΑΚΙΔΗΣ*
  - *ΓΕΩΓΡΙΑΔΗΣ*
  - *ΠΑΠΑΝΙΚΟΛΑΟΥ*
  - *ΙΝΔΙΚΤΟΣ*
  - *ΓΝΩΣΗ*
  - *ΚΑΛΟΦΩΛΙΑΣ*

- academic, polytechnic and university presses:
  - Academy of Athens
  - Polytechnics of Athens
  - University Publications of Crete
  - University of Ioannina
  - Democritian University of Thrace
- private educational organizations:
  - *ΚΟΡΕΛΚΟ*
  - *ΜΩΡΑΪΤΗ*
- others:
  - Hellenic Parliament
  - military press
  - Orthodox Church

10   Testing

The following testing procedure was used for μονο$\vec{2}$-πολυ development. The author worked on SuSE Linux and Windows XP, but any platform (Linux, Unix, Mac or Windows) should work as well, as long the JDK is installed.

1. Visit any web site with rich Modern Greek content, for example, news sources such as `http://www.pathfinder.gr`.

2. Open a new document with a word processor supporting spell checking of monotonic Greek.

3. Copy a long excerpt of continuous text from the web site.

4. Paste the selected and copied text into the word processor window.

5. Correct any misspelled words, but do not use any style or font effects.

6. Save the document as plain ISO 5589-7 encoded text file.

7. Process the document with μονο$\vec{2}$πολυ, as a Java application from a console window.

8. Take the 7-bit TeX result and add it to the LaTeX template file in your favourite environment (LyX, Kile, MiKTeX, etc.).

9. Produce a `.ps` or a `.pdf` file and check the final result with GSview or some other reader.

The results improve as the database is enriched. However, some manual editing is inevitable, depending on the complexity of the document to be multi-accented, because authors may mix Ancient Greek phrases into Modern Greek sentences.

11 Future Developments

One important improvement would be to relocate some of the intelligence to external script files, for defining and modifying the polytonic grammar rule sets.

Another avenue is to integrate $\mu o\nu o\vec{2}\pi o\lambda\upsilon$ with the source and data files of the open source Writer2LaTeX project (by Henrik Just, `http://www.hj-gym.dk/~hj/writer2latex/`). That would provide a reverse conversion, from UTF-16BE/LE encoded Greek documents into 7-bit (La)TeX.

12 Related links

Here we list some further readings on the complexity of Greek multiple accenting and related subjects. First, these articles (mainly written in Greek) on the importance of the spiritus lenis and especially of the spiritus asper:

- `http://www.typos.com.cy/nqcontent.cfm?a_id=4681`

- `http://www.kairatos.com.gr/polytoniko.htm`

- `http://www.krassanakis.gr/tonos.htm`

- `http://www.mathisis.com/nqcontent.cfm?a_id=1767`

  Further general sources are the following:

- Ministry of National Education and Religion Affairs — `http://www.ypepth.gr/en_ec_home.htm`

- Institute for Language and Speech Processing — `http://www.ilsp.gr`

- `http://www.ekivolos.gr`

References

Blomqvist, J., Toivanen, A., *Johdatus Uuden testamentin Kreikkaan.* Yliopistopaino: Helsinki (1993), 15–25.

Bornemann, E., Risch, E., *Griechische Grammatik.* Verlag Moritz Diesterweg: Frankfurt a./M. (1978), 1–25.

Deitsch, A., Czarnecki, D., *Java Internationalization.* O'Reilly: Beijing, Cambridge, Köln, Paris, Sebastopol, Taipei, Tokyo (2001), 171–198.

Harold, E.R., *Java I/O.* O'Reilly: Beijing, Cambridge, Köln, Paris, Sebastopol, Taipei, Tokyo (1999), 387-413

Kaegi, A., Bruhn, E., *Griechische Schulgrammatik.* Verlag Weidmann: Zürich, Hildesheim (1989), 2–10.

Dr.Maier, F., Weiss, M., Zeller, A., *ORGANON Grammatik I+II.* Bayerischer Schulbuch-Verlag: München, C.C. Buchners Verlag, Salzburg (1979), 1–12.

Οἰκονόμου, Μ. Χ., *ΓΡΑΜΜΑΤΙΚΗ ΤΗΣ ΑΡΧΑΙΑΣ ΕΛΛΗΝΙΚΗΣ. ΙΝΣΤΙΤΟΥΤΟ ΝΕΟΕΛΛΗΝΙΚΩΝ ΣΠΟΥΔΩΝ*: Θεσσαλονίκη (1993), 12–21.

Pakarinen, W., *Kreikan kielioppi.* Suomalaisen Kirjallisuuden Seura: Helsinki (1993), 1–13.

Seeboerger-Weichselbaum, M., *Java/XML. DAS bhv TASCHENBUCH.* verlag moderne industrie: Bonn Landsberg (2002), 181–208.

Syropoulos, A., Tsolomitis, A., Sofroniou, N., *Digital Typography Using LaTeX.* Springer Professional Computing, Springer-Verlag: Berlin, Heidelberg, New York (2003), 315–319.

Τσάρτζανος, A., *ΓΡΑΜΜΑΤΙΚΗ ΤΗΣ ΑΡΧΑΙΑΣ ΕΛΛΗΝΙΚΗΣ ΓΛΩΣΣΗΣ. ΕΚΔΟΤΙΚΟΣ ΟΙΚΟΣ ΑΔΕΛΦΩΝ ΚΥΡΙΑΚΙΔΗ*: Θεσσαλονίκη, (1993), 9–14.

Τσολάκης, Χρ. Λ., *ΝΕΟΕΛΛΗΝΙΚΗ ΓΡΑΜΜΑΤΙΚΗ. ΕΚΔΟΣΕΙΣ ΚΩΔΙΚΑΣ*, Θεσσαλονίκη: (1988), 28–35.

Dr. Wendt, H.F., *Langenscheidts Praktisches Lehrbuch Neugriechisch. Ein Standardwerk für Anfänger*, Langenscheidt: Berlin, München, Wien, Zürich, New York (1993), 20–35.