

Sorting Unicode Greek in MS Word

By Steven Craig Miller

Email: stevencraigmiller@comcast.net

October 2006

Table of Contents

| | |
|----------------------------------|------------|
| Introduction | page 2 |
| Instructions on How to Use Macro | pages 3-5 |
| How To Install Macro in MS Word | page 6 |
| Copy of Macro | pages 7-21 |

Introduction

While MS Word now works with unicode and allows one to work with Greek texts, Word cannot (at the time of this writing) sort accented Greek text alphabetically. Because of this I have developed a macro which will help facilitate this.

Actually all the macro does is convert accented Greek text to uppercase Greek text without accents. Since the unaccented uppercase Greek letters are consecutive in unicode, they can be sorted easily in alphabetical order.

For example, take the following vocabulary list:

| | |
|----------|-------------------|
| γυνή | woman |
| υῖός | son |
| μήτηρ | mother |
| ἄνθρωπος | man |
| ἄνῆρ | man |
| ἄναξ | lord, king |
| βασιλεύς | king |
| δῶμα | house, room |
| ἵππος | horse |
| θυμός | heart, soul, life |
| χείρ | hand |
| πατήρ | father |
| ἄστυ | city |
| παῖς | child |
| αἷμα | blood |

If one tried to sort the Greek text using MS word one would get the following result:

| | |
|----------|-------------------|
| αἷμα | blood |
| βασιλεύς | king |
| γυνή | woman |
| δῶμα | house, room |
| θυμός | heart, soul, life |
| μήτηρ | mother |
| ἄναξ | lord, king |
| ἄνθρωπος | man |
| ἄνῆρ | man |
| παῖς | child |
| πατήρ | father |
| ἵππος | horse |
| ἄστυ | city |
| υῖός | son |
| χείρ | hand |

This is obviously not what one wants. I've come up with a very simple solution.

Instructions On How To Use My Macro

Step One:

Duplicate the column you want sorted and place the duplicate column first in a table. My macro is limited to working with Greek texts in MS Word's tables.

Warning: The table must also be the first table in a Word document! If need be, open a new document, copy and paste your table in that new document, and run the macro from that new document.

This macro looks for the first table in an active Word document and converts the text in the first column of that table. It will only convert unicode Greek characters (all other characters will be left as they were).

For example, take the vocabulary list we saw on the last page, duplicate the first column, like so:

| | | |
|----------|----------|-------------------|
| γυνή | γυνή | woman |
| υῖός | υῖός | son |
| μήτηρ | μήτηρ | mother |
| ἄνθρωπος | ἄνθρωπος | man |
| άνήρ | άνήρ | man |
| ἄναξ | ἄναξ | lord, king |
| βασιλεύς | βασιλεύς | king |
| δῶμα | δῶμα | house, room |
| ἵππος | ἵππος | horse |
| θυμός | θυμός | heart, soul, life |
| χείρ | χείρ | hand |
| πατήρ | πατήρ | father |
| ἄστυ | ἄστυ | city |
| παῖς | παῖς | child |
| αἷμα | αἷμα | blood |

Step two:

Run the macro: The macro converts Greek text to unaccented uppercase Greek text. Like so:

| | | |
|----------|----------|-------------------|
| ΓΥΝΗ | γυνή | woman |
| ΥΙΟΣ | υῖος | son |
| ΜΗΤΗΡ | μήτηρ | mother |
| ΑΝΘΡΩΠΟΣ | ἄνθρωπος | man |
| ΑΝΗΡ | ἄνῆρ | man |
| ΑΝΑΞ | ἄναξ | lord, king |
| ΒΑΣΙΛΕΥΣ | βασιλεύς | king |
| ΔΩΜΑ | δῶμα | house, room |
| ΙΠΠΟΣ | ἵππος | horse |
| ΘΥΜΟΣ | θυμός | heart, soul, life |
| ΧΕΙΡ | χείρ | hand |
| ΠΑΤΗΡ | πατήρ | father |
| ΑΣΤΥ | ἄστυ | city |
| ΠΑΙΣ | παῖς | child |
| ΑΙΜΑ | αἷμα | blood |

Warning: A long vocabulary list (let's say one with over a thousand words) will take minutes to convert. This macro runs very slow on such long lists.

Step Three:

Sort the table by using MS Word's sort function. The result is below:

| | | |
|----------|----------|-------------------|
| ΑΙΜΑ | αἷμα | blood |
| ΑΝΑΞ | ἄναξ | lord, king |
| ΑΝΗΡ | ἄνῆρ | man |
| ΑΝΘΡΩΠΟΣ | ἄνθρωπος | man |
| ΑΣΤΥ | ἄστυ | city |
| ΒΑΣΙΛΕΥΣ | βασιλεύς | king |
| ΓΥΝΗ | γυνή | woman |
| ΔΩΜΑ | δῶμα | house, room |
| ΘΥΜΟΣ | θυμός | heart, soul, life |
| ΙΠΠΟΣ | ἵππος | horse |
| ΜΗΤΗΡ | μήτηρ | mother |
| ΠΑΙΣ | παῖς | child |
| ΠΑΤΗΡ | πατήρ | father |
| ΥΙΟΣ | υἱός | son |
| ΧΕΙΡ | χείρ | hand |

Step Four:

Delete the first column (which should no longer be needed) and you are done.

How To Install Macro in MS Word

I'm distributing this macro as a text file. You should have received two files: a PDF file entitled "Sorting Unicode Greek in MS Word" and a text file named "Macro ConvertUnicodeGreek2Uppercase."

The text file should be identical to what appears on pages 7-21 of this file.

By comparing the text file with pages 7-21 of this file, you should be able to check if anyone has tampered with the text file.

All you need to do is to copy and paste the text file into the Microsoft Visual Basic Editor.

Open a Word document. Go to Tools -> Macro -> Visual Basic Editor. (Or use the short cut Alt-F11.) Copy and past the text file. If you have other macros already present, you might want to open a new module. Insert -> Module and then paste in this macro.

I hope I have left comments in the macro which will explain everything. Someone with even limited understanding of VBA and Word's macros should be able to make sense of it. Please fill free to modify this macro as needs be.

If you have any questions, please feel free to contact me at:

stevencraigmiller@comcast.net

```
'  
' Macro: ConvertUnicodeGreek2Uppercase  
'  
' Created by Steven Craig Miller  
'  
' Email: stevencraigmiller@comcast.net  
'  
' October 2006  
'  
' This macro is being distributed with a PDF file which explains how to use it.  
' The PDF file is entitled: Sorting Unicode Greek in MS Word  
'  
'  
Sub ConvertUnicodeGreek2Uppercase()  
  
    Dim str As String  
    Dim numofrows As Long  
    Dim row As Long  
  
    ' Change the following value if you want to convert a different column  
    ' Otherwise, this macro will only convert the first column in a table  
    Const col As Long = 1  
  
    If ActiveDocument.Tables.Count >= 1 Then  
        numofrows = ActiveDocument.Tables(1).Rows.Count  
  
        For row = 1 To numofrows  
  
            With ActiveDocument.Tables(1).Cell(row, col).Range  
                str = ActiveDocument.Tables(1).Cell(row, col).Range  
            End With  
  
            str = StripString(str)  
            str = ConvertString(str)  
  
            With ActiveDocument.Tables(1).Cell(row, col).Range  
                .Delete  
                .InsertBefore str  
            End With  
  
        Next row  
  
    End If  
End Sub
```

```

'
' Function: StripString
'
' When I copied information (namely a string) from a table cell
' and then put that text into another cell, I noticed that an extra
' paragraph marker was added, which added an unwanted blank line after
' the text. This function corrects that problem. (Perhaps there is a
' better solution?)
'
'
Function StripString(ByVal str As String) As String
    Dim strlen As Long
    strlen = Len(str)
    If strlen > 1 Then
        StripString = Left$(str, strlen - 2)
    Else
        StripString = str
    End If
End Function

'
' Function: ConvertString
'
' This function does the following:
' (1) It steps through a string, character by character,
' (2) changing each character into a number.
' (3) It then sends that number to the ConvertGreekCharacter function
'     to be evaluated and possibly changed.
' (4) It takes the number returned from ConvertGreekCharacter function,
' (5) changes it back into a character and
' (6) sticks it back into the string at the same spot.
'
Function ConvertString(ByVal str As String) As String
    Dim strpos As Long
    Dim strlen As Long
    Dim charnum As Long

    strlen = Len(str)
    For strpos = 1 To strlen
        charnum = ConvertGreekCharacter(AscW(Mid(str, strpos, 1)))
        Mid(str, strpos, 1) = ChrW(charnum)
    Next strpos
    ConvertString = str
End Function

'
' Function: ConvertGreekCharacter
'

```

```
' This function converts Greek Letters into their appropriate Uppercase letter
' Non-Greek letters and punctuation are left unchanged!
'
' Based on the Unicode Standard 5.0 (2006)
'
' All constants are Uppercase Greek Letters
' Unicode documents give all values of characters in Hex (Base 16)
' This programming language (VBA) uses the affix "&H" before a Hex number
'
```

```
Function ConvertGreekCharacter(ByVal charnum As Long) As Long
```

```
    Const alpha As Long = &H391
    Const beta As Long = &H392
    Const gamma As Long = &H393
    Const delta As Long = &H394
    Const epsilon As Long = &H395
    Const zeta As Long = &H396
    Const eta As Long = &H397
    Const theta As Long = &H398
    Const iota As Long = &H399
    Const kappa As Long = &H39A
    Const lambda As Long = &H39B
    Const mu As Long = &H39C
    Const nu As Long = &H39D
    Const xi As Long = &H39E
    Const omicron As Long = &H39F
    Const pi As Long = &H3A0
    Const rho As Long = &H3A1
    ' Unicode 03A2 is reserved
    Const sigma As Long = &H3A3
    Const tau As Long = &H3A4
    Const upsilon As Long = &H3A5
    Const phi As Long = &H3A6
    Const chi As Long = &H3A7
    Const psi As Long = &H3A8
    Const omega As Long = &H3A9
```

```
    Select Case charnum
```

```
        Case Is = &H386
            charnum = alpha
        ' Unicode 0387 is punctuation
        Case Is = &H388
            charnum = epsilon
        Case Is = &H389
            charnum = eta
        Case Is = &H38A
            charnum = iota
        ' Unicode 038B is reserved
```

Case Is = &H38C
charnum = omicron
' Unicode 038D is reserved
Case Is = &H38E
charnum = upsilon
Case Is = &H38F
charnum = omega
Case Is = &H390
charnum = iota
' Unicodes 0391-03A9 are uppercase Greek Letters, here used as constants
Case Is = &H3AA
charnum = iota
Case Is = &H3AB
charnum = upsilon
Case Is = &H3AC
charnum = alpha
Case Is = &H3AD
charnum = epsilon
Case Is = &H3AE
charnum = eta
Case Is = &H3AF
charnum = iota
Case Is = &H3B0
charnum = upsilon
Case Is = &H3B1
charnum = alpha
Case Is = &H3B2
charnum = beta
Case Is = &H3B3
charnum = gamma
Case Is = &H3B4
charnum = delta
Case Is = &H3B5
charnum = epsilon
Case Is = &H3B6
charnum = zeta
Case Is = &H3B7
charnum = eta
Case Is = &H3B8
charnum = theta
Case Is = &H3B9
charnum = iota
Case Is = &H3BA
charnum = kappa
Case Is = &H3BB
charnum = lambda
Case Is = &H3BC
charnum = mu

Case Is = &H3BD
 charnum = nu
 Case Is = &H3BE
 charnum = xi
 Case Is = &H3BF
 charnum = omicron
 Case Is = &H3C0
 charnum = pi
 Case Is = &H3C1
 charnum = rho
 Case Is = &H3C2
 charnum = sigma
 Case Is = &H3C3
 charnum = sigma
 Case Is = &H3C4
 charnum = tau
 Case Is = &H3C5
 charnum = upsilon
 Case Is = &H3C6
 charnum = phi
 Case Is = &H3C7
 charnum = chi
 Case Is = &H3C8
 charnum = psi
 Case Is = &H3C9
 charnum = omega
 Case Is = &H3CA
 charnum = iota
 Case Is = &H3CB
 charnum = upsilon
 Case Is = &H3CC
 charnum = omicron
 Case Is = &H3CD
 charnum = upsilon
 Case Is = &H3CE
 charnum = omega

,

' Now we skip to the Extended (Precomposed Polytonic) Greek Letters

' Range: 1F00-1FFF

,

Case Is = &H1F00
 charnum = alpha
 Case Is = &H1F01
 charnum = alpha
 Case Is = &H1F02
 charnum = alpha
 Case Is = &H1F03

charnum = alpha

Case Is = &H1F04

charnum = alpha

Case Is = &H1F05

charnum = alpha

Case Is = &H1F06

charnum = alpha

Case Is = &H1F07

charnum = alpha

Case Is = &H1F08

charnum = alpha

Case Is = &H1F09

charnum = alpha

Case Is = &H1F0A

charnum = alpha

Case Is = &H1F0B

charnum = alpha

Case Is = &H1F0C

charnum = alpha

Case Is = &H1F0D

charnum = alpha

Case Is = &H1F0E

charnum = alpha

Case Is = &H1F0F

charnum = alpha

Case Is = &H1F10

charnum = epsilon

Case Is = &H1F11

charnum = epsilon

Case Is = &H1F12

charnum = epsilon

Case Is = &H1F13

charnum = epsilon

Case Is = &H1F14

charnum = epsilon

Case Is = &H1F15

charnum = epsilon

' Unicode 1F16 & 1F17 are reserved

Case Is = &H1F18

charnum = epsilon

Case Is = &H1F19

charnum = epsilon

Case Is = &H1F1A

charnum = epsilon

Case Is = &H1F1B

charnum = epsilon

Case Is = &H1F1C

charnum = epsilon
Case Is = &H1F1D
charnum = epsilon
' Unicode 1F1E & 1F1F are reserved

Case Is = &H1F20
charnum = eta
Case Is = &H1F21
charnum = eta
Case Is = &H1F22
charnum = eta
Case Is = &H1F23
charnum = eta
Case Is = &H1F24
charnum = eta
Case Is = &H1F25
charnum = eta
Case Is = &H1F26
charnum = eta
Case Is = &H1F27
charnum = eta
Case Is = &H1F28
charnum = eta
Case Is = &H1F29
charnum = eta
Case Is = &H1F2A
charnum = eta
Case Is = &H1F2B
charnum = eta
Case Is = &H1F2C
charnum = eta
Case Is = &H1F2D
charnum = eta
Case Is = &H1F2E
charnum = eta
Case Is = &H1F2F
charnum = eta

Case Is = &H1F30
charnum = iota
Case Is = &H1F31
charnum = iota
Case Is = &H1F32
charnum = iota
Case Is = &H1F33
charnum = iota
Case Is = &H1F34
charnum = iota

Case Is = &H1F35

charnum = iota

Case Is = &H1F36

charnum = iota

Case Is = &H1F37

charnum = iota

Case Is = &H1F38

charnum = iota

Case Is = &H1F39

charnum = iota

Case Is = &H1F3A

charnum = iota

Case Is = &H1F3B

charnum = iota

Case Is = &H1F3C

charnum = iota

Case Is = &H1F3D

charnum = iota

Case Is = &H1F3E

charnum = iota

Case Is = &H1F3F

charnum = iota

Case Is = &H1F40

charnum = omicron

Case Is = &H1F41

charnum = omicron

Case Is = &H1F42

charnum = omicron

Case Is = &H1F43

charnum = omicron

Case Is = &H1F44

charnum = omicron

Case Is = &H1F45

charnum = omicron

' Unicode 1F46 & 1F47

Case Is = &H1F48

charnum = omicron

Case Is = &H1F49

charnum = omicron

Case Is = &H1F4A

charnum = omicron

Case Is = &H1F4B

charnum = omicron

Case Is = &H1F4C

charnum = omicron

Case Is = &H1F4D

charnum = omicron

' Unicode 1F4E & 1F4F are reserved

Case Is = &H1F50

charnum = upsilon

Case Is = &H1F51

charnum = upsilon

Case Is = &H1F52

charnum = upsilon

Case Is = &H1F53

charnum = upsilon

Case Is = &H1F54

charnum = upsilon

Case Is = &H1F55

charnum = upsilon

Case Is = &H1F56

charnum = upsilon

Case Is = &H1F57

charnum = upsilon

' Unicode 1F58 is reserved

Case Is = &H1F59

charnum = upsilon

' Unicode 1F5A is reserved

Case Is = &H1F5B

charnum = upsilon

' Unicode 1F5C is reserved

Case Is = &H1F5D

charnum = upsilon

' Unicode 1F5E is reserved

Case Is = &H1F5F

charnum = upsilon

Case Is = &H1F60

charnum = omega

Case Is = &H1F61

charnum = omega

Case Is = &H1F62

charnum = omega

Case Is = &H1F63

charnum = omega

Case Is = &H1F64

charnum = omega

Case Is = &H1F65

charnum = omega

Case Is = &H1F66

charnum = omega

Case Is = &H1F67

charnum = omega

Case Is = &H1F68

charnum = omega
Case Is = &H1F69
charnum = omega
Case Is = &H1F6A
charnum = omega
Case Is = &H1F6B
charnum = omega
Case Is = &H1F6C
charnum = omega
Case Is = &H1F6D
charnum = omega
Case Is = &H1F6E
charnum = omega
Case Is = &H1F6F
charnum = omega

Case Is = &H1F70
charnum = alpha
Case Is = &H1F71
charnum = alpha
Case Is = &H1F72
charnum = epsilon
Case Is = &H1F73
charnum = epsilon
Case Is = &H1F74
charnum = eta
Case Is = &H1F75
charnum = eta
Case Is = &H1F76
charnum = iota
Case Is = &H1F77
charnum = iota
Case Is = &H1F78
charnum = omicron
Case Is = &H1F79
charnum = omicron
Case Is = &H1F7A
charnum = upsilon
Case Is = &H1F7B
charnum = upsilon
Case Is = &H1F7C
charnum = omega
Case Is = &H1F7D
charnum = omega
' Unicode 1F7E & 1F7F are reserved

Case Is = &H1F80
charnum = alpha

Case Is = &H1F81
charnum = alpha
Case Is = &H1F82
charnum = alpha
Case Is = &H1F83
charnum = alpha
Case Is = &H1F84
charnum = alpha
Case Is = &H1F85
charnum = alpha
Case Is = &H1F86
charnum = alpha
Case Is = &H1F87
charnum = alpha
Case Is = &H1F88
charnum = alpha
Case Is = &H1F89
charnum = alpha
Case Is = &H1F8A
charnum = alpha
Case Is = &H1F8B
charnum = alpha
Case Is = &H1F8C
charnum = alpha
Case Is = &H1F8D
charnum = alpha
Case Is = &H1F8E
charnum = alpha
Case Is = &H1F8F
charnum = alpha

Case Is = &H1F90
charnum = eta
Case Is = &H1F91
charnum = eta
Case Is = &H1F92
charnum = eta
Case Is = &H1F93
charnum = eta
Case Is = &H1F94
charnum = eta
Case Is = &H1F95
charnum = eta
Case Is = &H1F96
charnum = eta
Case Is = &H1F97
charnum = eta
Case Is = &H1F98

charnum = eta
 Case Is = &H1F99
 charnum = eta
 Case Is = &H1F9A
 charnum = eta
 Case Is = &H1F9B
 charnum = eta
 Case Is = &H1F9C
 charnum = eta
 Case Is = &H1F9D
 charnum = eta
 Case Is = &H1F9E
 charnum = eta
 Case Is = &H1F9F
 charnum = eta

 Case Is = &H1FA0
 charnum = omega
 Case Is = &H1FA1
 charnum = omega
 Case Is = &H1FA2
 charnum = omega
 Case Is = &H1FA3
 charnum = omega
 Case Is = &H1FA4
 charnum = omega
 Case Is = &H1FA5
 charnum = omega
 Case Is = &H1FA6
 charnum = omega
 Case Is = &H1FA7
 charnum = omega
 Case Is = &H1FA8
 charnum = omega
 Case Is = &H1FA9
 charnum = omega
 Case Is = &H1FAA
 charnum = omega
 Case Is = &H1FAB
 charnum = omega
 Case Is = &H1FAC
 charnum = omega
 Case Is = &H1FAD
 charnum = omega
 Case Is = &H1FAE
 charnum = omega
 Case Is = &H1FAF
 charnum = omega

Case Is = &H1FB0
charnum = alpha
Case Is = &H1FB1
charnum = alpha
Case Is = &H1FB2
charnum = alpha
Case Is = &H1FB3
charnum = alpha
Case Is = &H1FB4
charnum = alpha
' Unicode 1FB5 is reserved
Case Is = &H1FB6
charnum = alpha
Case Is = &H1FB7
charnum = alpha
Case Is = &H1FB8
charnum = alpha
Case Is = &H1FB9
charnum = alpha
Case Is = &H1FBA
charnum = alpha
Case Is = &H1FBB
charnum = alpha
Case Is = &H1FBC
charnum = alpha
' Unicode 1FBD & 1FBE & 1FBF are sometype of accents

' Unicode 1FC0 & 1FC1 are sometype of accents
Case Is = &H1FC2
charnum = eta
Case Is = &H1FC3
charnum = eta
Case Is = &H1FC4
charnum = eta
' Unicode 1FC5 is reserved
Case Is = &H1FC6
charnum = eta
Case Is = &H1FC7
charnum = eta
Case Is = &H1FC8
charnum = epsilon
Case Is = &H1FC9
charnum = epsilon
Case Is = &H1FCA
charnum = eta
Case Is = &H1FCB
charnum = eta

Case Is = &H1FCC

charnum = eta

' Unicode 1FCD & 1FCE & 1FCF are sometype of accents

Case Is = &H1FD0

charnum = iota

Case Is = &H1FD1

charnum = iota

Case Is = &H1FD2

charnum = iota

Case Is = &H1FD3

charnum = iota

' Unicode 1FD4 & 1FD5 are reserved

Case Is = &H1FD6

charnum = iota

Case Is = &H1FD7

charnum = iota

Case Is = &H1FD8

charnum = iota

Case Is = &H1FD9

charnum = iota

Case Is = &H1FDA

charnum = iota

Case Is = &H1FDB

charnum = iota

' Unicode 1FDC is reserved

' Unicode 1FDD & 1FDE & 1FDF are sometype of accents

Case Is = &H1FE0

charnum = upsilon

Case Is = &H1FE1

charnum = upsilon

Case Is = &H1FE2

charnum = upsilon

Case Is = &H1FE3

charnum = upsilon

Case Is = &H1FE4

charnum = rho

Case Is = &H1FE5

charnum = rho

Case Is = &H1FE6

charnum = upsilon

Case Is = &H1FE7

charnum = upsilon

Case Is = &H1FE8

charnum = upsilon

Case Is = &H1FE9

charnum = upsilon

Case Is = &H1FEA
 charnum = upsilon
Case Is = &H1FEB
 charnum = upsilon
Case Is = &H1FEC
 charnum = rho
' Unicode 1FED & 1FEE & 1FEF are sometype of accents

' Unicode 1FF0 & 1FF1 are reserved
Case Is = &H1FF2
 charnum = omega
Case Is = &H1FF3
 charnum = omega
Case Is = &H1FF4
 charnum = omega
' Unicode 1FF5 is reserved
Case Is = &H1FF6
 charnum = omega
Case Is = &H1FF7
 charnum = omega
Case Is = &H1FF8
 charnum = omicron
Case Is = &H1FF9
 charnum = omicron
Case Is = &H1FFA
 charnum = omega
Case Is = &H1FFB
 charnum = omega
Case Is = &H1FFC
 charnum = omega
' Unicode 1FFD & 1FFE are sometype of accents
' Unicode 1FFF is reserved

End Select
ConvertGreekCharacter = charnum
End Function