

To: Unicode Technical Committee  
From: Deborah Anderson, Researcher, Dept. of Linguistics, UC Berkeley  
Re: Description of TLG Documents  
Date: Feb. 4, 2002

The following documents were created by the Thesaurus Linguae Graecae Project (TLG) at the University of California, Irvine. The TLG is a digital library containing more than 10,000 Greek literary texts (now in excess of 80 million words) from Homer to the end of Byzantium in AD 1453. TLG texts have been disseminated in CD-ROM format since 1985 and have recently become available online (<http://www.tlg.uci.edu>). The TLG is a major research tool used by thousands of scholars and students of Greek worldwide.

TLG texts are entered in the so-called Beta code, which is a character and formatting encoding convention developed by David W. Packard and his team at the Packard Humanities Institute in the late 70s. Each of the twenty-four Greek letters has an assigned ASCII position, and diacritics are indicated by non-alphabetic characters following the accented vowel. Beta-code was necessitated by the absence of a standard, cross-platform font table and keyboard for ancient (polytonic) Greek, a problem to be addressed and resolved by the establishment of a complete Unicode polytonic Greek set. Beta Code has been used by the TLG and a number of other electronic Greek corpora, such as the Perseus Project (Tufts University, the Duke Papyri Archive (containing documentary Greek papyri) and the Epigraphy Project (carried out by the Ohio State University and Cornell University) under the aegis of the Packard Humanities Institute. An annotated list of the TLG beta code symbols can be found at: <http://www.tlg.uci.edu/BetaCode.pdf>.

The proposals listed below include mostly non alphabetic characters found in the TLG corpus and missing from the Unicode Standard. Given the extensive range of texts present in the TLG collection, this list represents the majority of signs required in the electronic scholarly representation of Greek texts. It should be noted, however, that the TLG works exclusively from printed modern scholarly editions, therefore this list should not be regarded as exhaustive. For example, it does not contain any instances of medieval punctuation.

The characters are broken down into four groups:  
Measures (L2/02-029 <http://www.unicode.org/L2/L2002/02029-measures.pdf>),  
Metrical (L2/02-030 <http://www.unicode.org/L2/L2002/02030-metrical.pdf>),  
Miscellanea (L2/02-031 <http://www.unicode.org/L2/L2002/02031-misc.pdf>),  
and Musical (L2/02-032 <http://www.unicode.org/L2/L2002/02032-musical.pdf>).  
Each proposal includes a chart with the glyphs and a names list.

The last document, L2/02-033 (<http://www.unicode.org/L2/L2002/02033-tlg-backgnd.pdf>), is a lengthy draft which pulls together information on characters not currently in Unicode. The columns in the document identify the Beta Code number (corresponding to the Beta Code Manual), a representative glyph, a count of instances of each code in the present TLG corpus, a definition, comment, a possible Unicode equivalent, and a name. The four proposals above (Measures, Metrical, Misc., and Musical), were drawn from this source document.

The Miscellanea proposal includes two additional characters to be added to the Greek block: the Greek letter San and the Greek capital lunate sigma which is used extensively in papyrological texts in both upper- and lowercase. An addendum on page three of Miscellanea lists seven punctuation characters used in papyrological texts, though further details on their specific use are being sought at the present time from papyrologists. (The numbers on the left of page three refer to TLG Beta Codes).