

Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

Doc Type: Working Group Document**Title: Revised proposal for encoding the Glagolitic script in the UCS****Source: Michael Everson and Ralph Cleminson****Status: Expert Contribution****Action: For consideration by JTC1/SC2/WG2 and UTC****Date: 2002-12-04**

This document is based on the proposal written by Joe Becker and published in UTR#3, and the proposal written by Michael Everson in N1659. It is a revision of Michael Everson's N1931, and contains the proposal summary.

A. Administrative**1. Title**

Revised proposal for encoding the Glagolitic script in the UCS.

2. Requester's name

Michael Everson and Ralph Cleminson

3. Requester type (Member body/Liaison/Individual contribution)

Individual contribution.

4. Submission date

2002-12-04

5. Requester's reference (if applicable)**6. Choose one of the following:****6a. This is a complete proposal**

Yes.

6b. More information will be provided later

No.

B. Technical -- General**1. Choose one of the following:****1a. This proposal is for a new script (set of characters)**

Yes.

Proposed name of script

Glagolitic.

1b. The proposal is for addition of character(s) to an existing block

No.

1b. Name of the existing block**2. Number of characters in proposal**

88

3. Proposed category (see section II, Character Categories)

Category B.1.

4a. Proposed Level of Implementation (1, 2 or 3) (see clause 14, ISO/IEC 10646-1: 2000)

As an alphabetic script, Glagolitic requires Level 1.

4b. Is a rationale provided for the choice?

No.

4c. If YES, reference**5a. Is a repertoire including character names provided?**

Yes.

5b. If YES, are the names in accordance with the character naming guidelines in Annex L of ISO/IEC 10646-1: 2000?

Yes.

5c. Are the character shapes attached in a legible form suitable for review?

Yes.

6a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?

Michael Everson.

6b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:

Michael Everson, Fontographer.

7a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

Yes, see bibliography below.

7b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

Yes.

8. Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?

Yes, see below.

9. Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at <http://www.unicode.org> for such information on other scripts. Also see Unicode Character Database <http://www.unicode.org/Public/UNIDATA/UnicodeCharacterDatabase.html> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

Yes, see Unicode properties below.

C. Technical -- Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.

Yes, exploratory drafts in N1659 and N1931.

2a. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?

Yes.

2b. If YES, with whom?

Ralph Cleminson, Dr. Hinko Muren (Slovenia), ISO/TC46/SC2, ISO/TC46/SC4

2c. If YES, available relevant documents

N/A

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?

Scholarly, ecclesiastical, and library communities.

4a. The context of use for the proposed characters (type of use; common or rare)

Originally used to write Old Church Slavonic, and used in parts of medieval and early modern Croatia for both secular and ecclesiastical purposes.

4b. Reference

See below. ISO 6861:1996 is a coded character set for Glagolitic, as well as the bibliography below.

5a. Are the proposed characters in current use by the user community?

Yes.

5b. If YES, where?

By scholars, churchmen, and librarians.

6a. After giving due considerations to the principles in Principles and Procedures document (a WG 2 standing document) must the proposed characters be entirely in the BMP?

Yes.

6b. If YES, is a rationale provided?

Yes.

6c. If YES, reference

Accordance with the Roadmap, and close relationship of Glagolitic and Cyrillic texts.

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?

Yes, they should be encoded in a single block as presented here.

8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?

No.

8b. If YES, is a rationale for its inclusion provided?

8c. If YES, reference

9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?

No.

9b. If YES, is a rationale for its inclusion provided?

9c. If YES, reference

10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?

No.

10b. If YES, is a rationale for its inclusion provided?

10c. If YES, reference

11a. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)?

No.

11b. If YES, is a rationale for such use provided?

11c. If YES, reference

12a. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

No. Generic combining characters can occur with Glagolitic base forms.

12b. If YES, reference

13a. Does the proposal contain characters with any special properties such as control function or similar semantics?

No.

13b. If YES, describe in detail (include attachment if necessary)

14a. Does the proposal contain any Ideographic compatibility character(s)?

No.

14b. If YES, is the equivalent corresponding unified ideographic character(s) identified?

14c. If YES, reference

D. Proposal

User community

Glagolitic is considered to be the alphabet devised by St Cyril, probably in 862, for his translation of the the Scriptures and liturgical books into Slavonic. The Glagolitic inscriptions and manuscripts which survive from the early period are relatively few in number, but of great philological importance. Amongst the Slavs of the Eastern Rite, Glagolitic was subsequently supplanted by the alphabet now known as Cyrillic (based on Greek uncials but following the order and function of the Glagolitic letters), which probably arose in late tenth-century Bulgaria. In those parts of Croatia where a vernacular liturgy was used, Glagolitic continued in use until modern times: the last Glagolitic missal was printed in Rome in 1893. In these areas Glagolitic is still occasionally used as a decorative alphabet.

Structure

In the encoding Glagolitic is treated as a separate alphabet from Cyrillic, both because of its historical primacy and because in almost every case the shapes of the letters in the two alphabets are completely dissimilar: the one can in no sense be regarded as a visual variant of the other. Glagolitic itself exists in two styles, known as round and square. Round Glagolitic is the original style and more geographically widespread (though surviving examples are less numerous); square Glagolitic (and the cursive style derived from it) was used in Croatia from the thirteenth century. There are a few documents written in a style intermediate between the two. The letter-forms used in the charts are round Glagolitic, since a number of characters found in this style are not represented in square Glagolitic, while the reverse (with the exception of the stapić) is not the case.

Directionality

Like Cyrillic, the Glagolitic script is written in linear sequence from left to right with no contextual modification of the letterforms.

Ordering

The ordering of the Glagolitic alphabet is derived from that of Greek, though nearly half the Glagolitic characters have no equivalent in Greek and not every Greek letter has its equivalent in Glagolitic. The original (Cyrillo-methodian) alphabet is not recoverable, there being no extant documents from this period; its content and order can nevertheless be reconstructed with reasonable certainty on the basis of later abecedaria and a number of acrostics of extremely early origin. It should however be recognised that the order of the alphabet as represented in the surviving documents is not entirely stable, particularly towards the end.

Variant Glyph Forms

Several of the letters have variant glyph forms. These are not given separate character codes here. The character € SMALL YUS should be distinguished from ꞥ (or Ꞧ) by means of VARIANT SELECTOR-1. It occurs sporadically in a few early manuscripts, but is clearly not a random variant, since it occurs only in masculine participles (in this case, though, it is worth considering that if it isn't a random variant it might be worth encoding). In ISO 6861:1996 five variant forms are explicitly encoded, and were in N1659. Discussion with the convener of ISO/SC4/WG2 indicated that in bibliographical contexts, probably *most* current use transliterates Glagolitic into Cyrillic in any case, and it is suggested that the variant characters in ISO 6861:1996 be unified with the basic letters.

Punctuation and diacritics

Glagolitic uses the following punctuation:

,	002C	COMMA
.	002E	FULL STOP
;	037E	GREEK QUESTION MARK
·	0387	GREEK ANO TELEIA
:	0589	ARMENIAN FULL STOP
∴	10FB	GEORGIAN PARAGRAPH SEPARATOR

The punctuation mark ANO TELEIA is used, flanking a letter on either side, to show the numeric use of the letter (Faulmann 1880). Glagolitic uses numerous diacritical marks, many of them in common with Cyrillic. It is suggested that research on the use of these in both Cyrillic and Glagolitic be initiated, and that any characters missing from the UCS be added at a future date to the general Combining Diacritical Marks block if necessary. It has been observed that a line above or a tilde above a letter or letters can be used to show the numeric use of the letter.

Names

The names of the letters present a number of problems. The first is that regular phonological developments have caused some of them to diverge in the modern languages; Ѡ Ѡ Ѡ Ѡ *črivi* CHRIVI, for example, becomes *črv* in Croatian, *чрѣв* *črāv* in Bulgarian and *черѣв* *červ* in Russian. Secondly, the names of some letters used only in very early documents (such as Ѣ SPIDERY HA) have not survived. This particularly affects the nasal vowels, all of which are known by the name *jus* (which corresponds to Old Church Slavonic Ѣ Ѣ Ѣ *josŭ* YUSU), differentiated by various modern adjectives. For the letter Ѧ even the phonetic value is uncertain: *ö*, *jö* or *jo* have been suggested. We have used YO, which was also used to name this character in ISO 6861:1996.

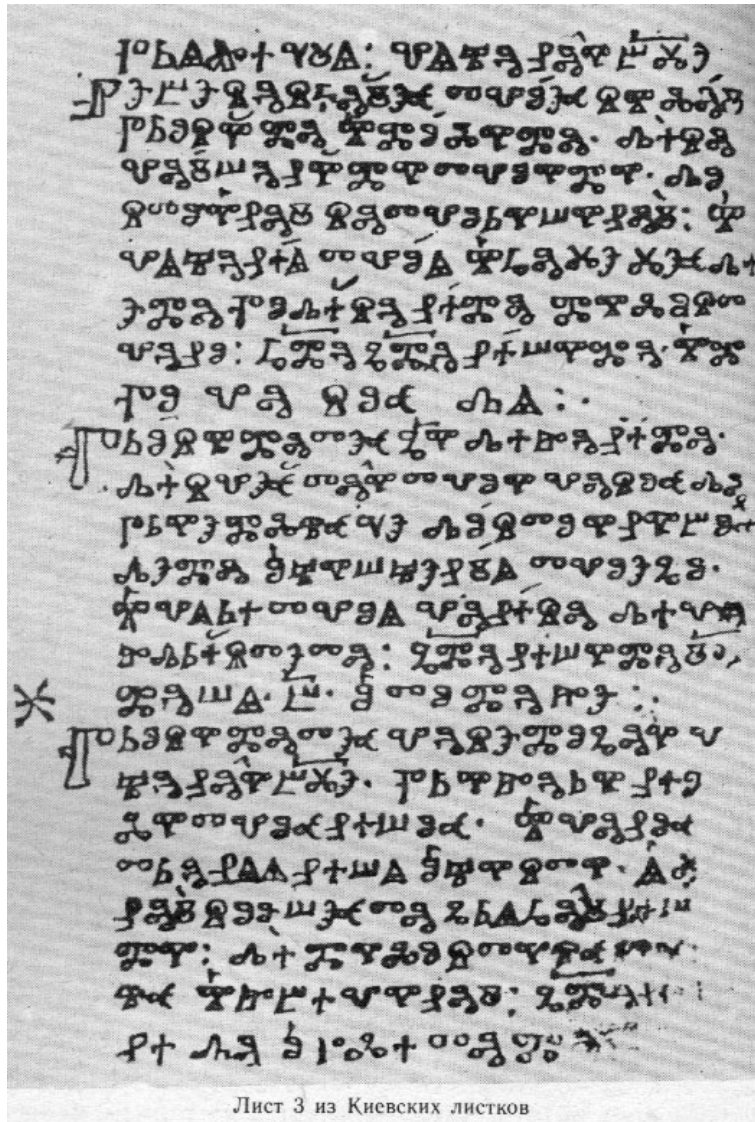
The letters have inherent numerical values. Those from 1 to 1000 are well attested; the values 2000, 3000 and 5000 for 𐌆 SHA, 𐌺 YERU and 𐌿 YU respectively are supported by fifteenth-century evidence; the higher values assigned to some letters by modern authors are not reliable.

				Value	ISO 6861
𐌰	𐌰𐌶𐌿	azъ	AZU	1	1
𐌱	𐌱𐌴𐌹𐌺	buky	BUKY	2	2
𐌲	𐌲𐌳𐌰𐌳	vědě	VEDE	3	3
𐌳	𐌳𐌶𐌰𐌴𐌹𐌺	glagoli	GLAGOLI	4	4
𐌴	𐌴𐌶𐌰𐌴𐌹	dobro	DOBRO	5	5
𐌵	𐌵𐌶𐌰𐌴𐌹	jestъ	JESTU	6	6
𐌶	𐌶𐌶𐌰𐌴𐌹𐌺	živěte	ZHIVETE	7	7
𐌷	𐌷𐌳𐌰𐌴	zělo	DZELO	8	8
𐌸	𐌸𐌶𐌰𐌴𐌹𐌺	zemlja	ZEMLJA	9	9
𐌹	𐌹𐌶𐌰	iže	IZHE	10	20
𐌺	𐌺𐌶𐌰	iže	INITIAL IZHEI	10	20
𐌻	𐌻	i	I	20	10
𐌼	𐌼𐌳𐌰𐌴𐌹𐌺	djervъ	DJERVI	30	—
𐌽	𐌽𐌰𐌴𐌹	kako	KAKO	40	40
𐌾	𐌾𐌶𐌰𐌴𐌹𐌺	ljudъje	LJUDIJE	50	50
𐌿	𐌿𐌶𐌰𐌴𐌹𐌺𐌰𐌴𐌹𐌺	myslite	MYSLITE	60	60
𐊀	𐊀𐌰𐌴𐌹𐌺	našъ	NASHI	70	70
𐊁	𐊁𐌰𐌴𐌹	onъ	ONU	80	80
𐊂	𐊂𐌶𐌰𐌴𐌹𐌺	pokoјъ	POKOJI	90	90
𐊃	𐊃𐌶𐌰𐌴𐌹𐌺	rici	RICI	100	100
𐊄	𐊄𐌶𐌰𐌴𐌹𐌺	slovo	SLOVO	200	200
𐊅	𐊅𐌶𐌰𐌴𐌹𐌺𐌰𐌴𐌹𐌺	tvrdō	TVRIDO	300	300
𐊆	𐊆𐌰𐌴𐌹𐌺	ukъ	UKU	400	400
𐊇	𐊇𐌶𐌰𐌴𐌹𐌺	frъtъ	FRITU	500	500
𐊈	𐊈𐌳𐌰𐌴𐌹𐌺	hęgъ	HERU	600	6000/600
𐊉	𐊉𐌰𐌴𐌹𐌺	otъ	OTU	700	700
𐊊			PE		
𐊋	𐊋𐌰𐌴𐌹	šta	SHTA	800	800
𐊌	𐊌𐌰𐌴𐌹	ci	CI	900	900
𐊍	𐊍𐌶𐌰𐌴𐌹𐌺	črъvъ	CHRIVI	1000	1000
𐊎	𐊎𐌰𐌴𐌹	ša	SHA	2000	2000
𐊏	𐊏𐌶𐌰𐌴𐌹𐌺	jerъ	YERU	3000	4000
𐊐	𐊐𐌶𐌰𐌴𐌹𐌺	jerъ	YERI		3000
𐊑	𐊑𐌰𐌴𐌹𐌺	jatъ	YATI		5000
𐊒			SPIDERY HA		
𐊓	𐊓	ju	YU	5000	8000
𐊔			SMALL YUS		—
𐊕			YO		7000
𐊖			IOT. SMALL YUS		—
𐊗			BIG YUS		9000
𐊘			IOT. BIG YUS		—
𐊙	𐊙𐌰𐌴𐌹𐌺	fitā	FITA		9/500
𐊚	𐊚𐌶𐌰𐌴𐌹𐌺	ižica	IZHICA		400
𐊛	𐊛𐌰𐌴𐌹𐌺𐌰𐌴𐌹𐌺	stapic	STAPIC		

Unicode Character Properties

Spacing letters, category “Lo”, bidi category “L” (strong left to right)

xx00-xx2B, xx30-xx5B



Note the use of the ARMENIAN FULL STOP, the GEORGIAN PARAGRAPH SEPARATOR and the GREEK ANO TELEIA here.

TABLE XX - Row 2C: GLAGOLITIC

	2C0	2C1	2C2	2C3	2C4	2C5
0	Ɱ	Ɱ̅	Ɱ̇	Ɱ	Ɱ̅	Ɱ̇
1	Ɱ̈	Ɱ̉	Ɱ̊	Ɱ̈	Ɱ̉	Ɱ̊
2	Ɱ̋	Ɱ̌	Ɱ̍	Ɱ̋	Ɱ̌	Ɱ̍
3	Ɱ̎	Ɱ̏	Ɱ̐	Ɱ̎	Ɱ̏	Ɱ̐
4	Ɱ̑	Ɱ̒	Ɱ̓	Ɱ̑	Ɱ̒	Ɱ̓
5	Ɱ̔	Ɱ̕	Ɱ̖	Ɱ̔	Ɱ̕	Ɱ̖
6	Ɱ̗	Ɱ̘	Ɱ̙	Ɱ̗	Ɱ̘	Ɱ̙
7	Ɱ̚	Ɱ̛	Ɱ̜	Ɱ̚	Ɱ̛	Ɱ̜
8	Ɱ̝	Ɱ̞	Ɱ̟	Ɱ̝	Ɱ̞	Ɱ̟
9	Ɱ̠	Ɱ̡	Ɱ̢	Ɱ̠	Ɱ̡	Ɱ̢
A	Ɱ̣	Ɱ̤	Ɱ̥	Ɱ̣	Ɱ̤	Ɱ̥
B	Ɱ̦	Ɱ̧	Ɱ̨	Ɱ̦	Ɱ̧	Ɱ̨
C	Ɱ̩	Ɱ̪		Ɱ̩	Ɱ̪	
D	Ɱ̫	Ɱ̬		Ɱ̫	Ɱ̬	
E	Ɱ̭	Ɱ̮		Ɱ̭	Ɱ̮	
F	Ɱ̯	Ɱ̰		Ɱ̯	Ɱ̰	

G = 00
P = 00

TABLE XX - Row 2C: GLAGOLITIC

dec	hex	Name	dec	hex	Name
000	00	GLAGOLITIC CAPITAL LETTER AZU	089	59	GLAGOLITIC SMALL LETTER FITA
001	01	GLAGOLITIC CAPITAL LETTER BUKY	090	5A	GLAGOLITIC SMALL LETTER IZHITSA
002	02	GLAGOLITIC CAPITAL LETTER VEDE	091	5B	GLAGOLITIC SMALL LETTER STAPIC
003	03	GLAGOLITIC CAPITAL LETTER GLAGOLI	092	5C	(This position shall not be used)
004	04	GLAGOLITIC CAPITAL LETTER DOBRO	093	5D	(This position shall not be used)
005	05	GLAGOLITIC CAPITAL LETTER YESTU	094	5E	(This position shall not be used)
006	06	GLAGOLITIC CAPITAL LETTER ZHIVETE	095	5F	(This position shall not be used)
007	07	GLAGOLITIC CAPITAL LETTER DZELO	096	60	(This position shall not be used)
008	08	GLAGOLITIC CAPITAL LETTER ZEMLJA	097	61	(This position shall not be used)
009	09	GLAGOLITIC CAPITAL LETTER IZHE	098	62	(This position shall not be used)
010	0A	GLAGOLITIC CAPITAL LETTER INITIAL IZHE	099	63	(This position shall not be used)
011	0B	GLAGOLITIC CAPITAL LETTER I	100	64	(This position shall not be used)
012	0C	GLAGOLITIC CAPITAL LETTER DERVI	101	65	(This position shall not be used)
013	0D	GLAGOLITIC CAPITAL LETTER KAKO	102	66	(This position shall not be used)
014	0E	GLAGOLITIC CAPITAL LETTER LJUDIJE	103	67	(This position shall not be used)
015	0F	GLAGOLITIC CAPITAL LETTER MYSLITE	104	68	(This position shall not be used)
016	10	GLAGOLITIC CAPITAL LETTER NASHI	105	69	(This position shall not be used)
017	11	GLAGOLITIC CAPITAL LETTER ONU	106	6A	(This position shall not be used)
018	12	GLAGOLITIC CAPITAL LETTER POKOJI	107	6B	(This position shall not be used)
019	13	GLAGOLITIC CAPITAL LETTER RITSI	108	6C	(This position shall not be used)
020	14	GLAGOLITIC CAPITAL LETTER SLOVO	109	6D	(This position shall not be used)
021	15	GLAGOLITIC CAPITAL LETTER TVRIDO	110	6E	(This position shall not be used)
022	16	GLAGOLITIC CAPITAL LETTER UKU	111	6F	(This position shall not be used)
023	17	GLAGOLITIC CAPITAL LETTER FRITU	112	70	(This position shall not be used)
024	18	GLAGOLITIC CAPITAL LETTER HERU	113	71	(This position shall not be used)
025	19	GLAGOLITIC CAPITAL LETTER OTU	114	72	(This position shall not be used)
026	1A	GLAGOLITIC CAPITAL LETTER PE	115	73	(This position shall not be used)
027	1B	GLAGOLITIC CAPITAL LETTER SHTA	116	74	(This position shall not be used)
028	1C	GLAGOLITIC CAPITAL LETTER TSI	117	75	(This position shall not be used)
029	1D	GLAGOLITIC CAPITAL LETTER CHRIVI	118	76	(This position shall not be used)
030	1E	GLAGOLITIC CAPITAL LETTER SHA	119	77	(This position shall not be used)
031	1F	GLAGOLITIC CAPITAL LETTER YERU	120	78	(This position shall not be used)
032	20	GLAGOLITIC CAPITAL LETTER YERI	121	79	(This position shall not be used)
033	21	GLAGOLITIC CAPITAL LETTER YATI	122	7A	(This position shall not be used)
034	22	GLAGOLITIC CAPITAL LETTER SPIDERY HA	123	7B	(This position shall not be used)
035	23	GLAGOLITIC CAPITAL LETTER JU	124	7C	(This position shall not be used)
036	24	GLAGOLITIC CAPITAL LETTER SMALL YUS	125	7D	(This position shall not be used)
037	25	GLAGOLITIC CAPITAL LETTER YO	126	7E	(This position shall not be used)
038	26	GLAGOLITIC CAPITAL LETTER IOTATED SMALL YUS	127	7F	(This position shall not be used)
039	27	GLAGOLITIC CAPITAL LETTER BIG YUS			
040	28	GLAGOLITIC CAPITAL LETTER IOTATED BIG YUS			
041	29	GLAGOLITIC CAPITAL LETTER FITA			
042	2A	GLAGOLITIC CAPITAL LETTER IZHITSA			
043	2B	GLAGOLITIC CAPITAL LETTER STAPIC			
044	2C	(This position shall not be used)			
045	2D	(This position shall not be used)			
046	2E	(This position shall not be used)			
047	2F	(This position shall not be used)			
048	30	GLAGOLITIC SMALL LETTER AZU			
049	31	GLAGOLITIC SMALL LETTER BUKY			
050	32	GLAGOLITIC SMALL LETTER VEDE			
051	33	GLAGOLITIC SMALL LETTER GLAGOLI			
052	34	GLAGOLITIC SMALL LETTER DOBRO			
053	35	GLAGOLITIC SMALL LETTER YESTU			
054	36	GLAGOLITIC SMALL LETTER ZHIVETE			
055	37	GLAGOLITIC SMALL LETTER DZELO			
056	38	GLAGOLITIC SMALL LETTER ZEMLJA			
057	39	GLAGOLITIC SMALL LETTER IZHE			
058	3A	GLAGOLITIC SMALL LETTER INITIAL IZHE			
059	3B	GLAGOLITIC SMALL LETTER I			
060	3C	GLAGOLITIC SMALL LETTER DERVI			
061	3D	GLAGOLITIC SMALL LETTER KAKO			
062	3E	GLAGOLITIC SMALL LETTER LJUDIJE			
063	3F	GLAGOLITIC SMALL LETTER MYSLITE			
064	40	GLAGOLITIC SMALL LETTER NASHI			
065	41	GLAGOLITIC SMALL LETTER ONU			
066	42	GLAGOLITIC SMALL LETTER POKOJI			
067	43	GLAGOLITIC SMALL LETTER RITSI			
068	44	GLAGOLITIC SMALL LETTER SLOVO			
069	45	GLAGOLITIC SMALL LETTER TVRIDO			
070	46	GLAGOLITIC SMALL LETTER UKU			
071	47	GLAGOLITIC SMALL LETTER FRITU			
072	48	GLAGOLITIC SMALL LETTER HERU			
073	49	GLAGOLITIC SMALL LETTER OTU			
074	4A	GLAGOLITIC SMALL LETTER PE			
075	4B	GLAGOLITIC SMALL LETTER SHTA			
076	4C	GLAGOLITIC SMALL LETTER TSI			
077	4D	GLAGOLITIC SMALL LETTER CHRIVI			
078	4E	GLAGOLITIC SMALL LETTER SHA			
079	4F	GLAGOLITIC SMALL LETTER YERU			
080	50	GLAGOLITIC SMALL LETTER YERI			
081	51	GLAGOLITIC SMALL LETTER YATI			
082	52	GLAGOLITIC SMALL LETTER SPIDERY HA			
083	53	GLAGOLITIC SMALL LETTER YU			
084	54	GLAGOLITIC SMALL LETTER SMALL YUS			
085	55	GLAGOLITIC SMALL LETTER YO			
086	56	GLAGOLITIC SMALL LETTER IOTATED SMALL YUS			
087	57	GLAGOLITIC SMALL LETTER BIG YUS			
088	58	GLAGOLITIC SMALL LETTER IOTATED BIG YUS			