

拡張キリル文字群を含む Unicode フォントの 作成について

酒 井 純

Creating a Unicode font with "Cyrillic Extended" characters

SAKAI, Jun

要 旨

これまでの研究ではロシア諸年代記などで用いられる、初期キリル文字や特殊なキリル文字を含む Unicode Version 5.0 準拠のフォントを作成してきた。これに拡張キリル文字A表、および拡張キリル文字B表が加わった、新しいバージョンの Unicode の登場により、ロシア諸年代記のデジタル化に必要とされるグリフがほぼ Unicode に含まれることとなった。本研究では、まずコンピュータ上で初期キリル文字などを扱う環境について、これまでの研究の時点からどのように変化してきたのかを検証している。そしてこれを元に、これまでに作成してきた Preslav フォントの更新作業を行い、Version 1.0 としてリリースしている。

キーワード：初期キリル文字、フォント、ウェブフォント、Unicode Version 7.0.0

1. はじめに

ロシア諸年代記に代表される古ロシア語の文献では、現代ロシア語では用いられていない初期キリル文字、またその変遷過程で発生した多種の特殊なキリル文字が用いられている。これらの文字を含む文献を、コンコーダンスの作成などの目的で電子データ化するのに伴う問題について、酒井2004および酒井2009と研究を重ねてきた。このような中で、酒井2009ではその時点で最新であっ

た Unicode のバージョンに準拠したフォントを作成した。これには基本的なキリル文字に加えて、このとき研究協力をしていたロシア諸年代記コンコーダンスなどで必要とされる初期キリル文字についても含めて作成した。これにより、コンコーダンス上での特殊なキリル文字の表示の問題を、ある程度解決することができた。しかしながら、このときのフォントでは、Unicode フォントの外字領域を用いており、データの互換性について問題を残すこととなった。

本研究では、まず各種キリル文字をコンピューターで扱うための環境、及び問題点について整理し、酒井2009（以降前研究と記載）と比較しながら明らかにしてゆく。このために、まず Unicode の最新版でのキリル文字の収録状況、各種代表的 OS に搭載されているフォントの Unicode への対応状況、またウェブフォントと呼ばれるフォント形式について検証する。また前研究の後に見つかった、この研究の目的と合致する GNU Free Font についてもふれる。そしてこれに基づいて、最新版の Unicode に準拠するキリル文字フォントを、前研究で作成したフォントにグリフを追加して更新することで作成する。また作成したフォントから、ツールソフトを用いてウェブフォントも変換、作成する。

2. 初期キリル文字をコンピューターで扱う環境の変化

酒井2009の研究は、パソコンの OS の標準文字コードが、それまでのシフト JIS コードから Unicode へと転換する時期に進められた。実際には、世界的にマイクロソフト社の Windows 98 や Windows Me から Windows 2000 や Windows XP が OS シェアを中心となった時期にあたる。これらの OS の文字コードが Unicode を中心としたことにより、一つのファイル上で複数の言語が混在することが可能となった。これはキリル文字についても同様であり、これにより一つの文章ファイル中で日本語とロシア語が自由に混在できるようになった。

また、Unicode がそれまでのシフト JIS コードなどと異なったのが、多量の文字を含むことが可能となったことで、16 bit で表される基本多言語面で約 6 万字、追加多言語面 (SMP) で100万字の表現が可能となっている。そして、

それまでの文字コードで収録文字がほぼ固定化されていたのに対して、Unicode は様々な言語の文字に順次文字コードを割り当てて、バージョンアップしていくことで、収録する文字が増えるというのも特徴といえる。この文字数の増加により、それまでのデジタルデータ化では、いくつかのフォントを使い分けて表現した特殊キリル文字が、単一の文字コード、フォントで表現可能になった。このことにより、初期キリル文字などを含む文献の、本格的なデジタルデータ化が可能になったといえる。

一方、Unicode が順次文字を増やしていくということは、フォントデータも Unicode のバージョンアップにあわせてフォントにグリフを追加する作業の必要性につながる。また、コンコーダンスの表示に用いたのが HTML であり、フォントを内包して保存することが難しかったため、OS に標準搭載されているフォントに、表示データが限定されるという問題にもなった。

前研究では、MS-Windows XP に標準で搭載されていたフォントが Unicode Version 1.1.5 に準拠しており、コンコーダンス化で必要となる「**Ы**」や「**Ѡ**」などの文字が収録されていなかった。このため、当時 Unicode として最新であった Unicode Version 5.0 に準拠したフォントを作成することで、問題の解決を図った。そしてこの研究以後、何度かの Unicode の改訂を経て、キリル文字については、ほぼほとんどの文字が網羅される状況になってきている。

このため、まず前研究以降に更新された Unicode のバージョンアップによって、どのようなキリル文字が追加されたかを確認する。また、キリル文字を表示する環境として、パソコンの OS のシェアとその標準搭載フォントの変化と、新たな表示デバイスとして可能性の出てきたタブレット端末でのキリル文字の表示についても確認する。

2.1. Unicode のバージョン改訂について

Unicode のバージョンは、2014年現在 Version 7.0¹ が最新版となっている。前研究からは何度かのバージョンを経ており、それぞれのバージョンアップでの追加されたキリル文字関連の文字コードをまとめたのが次の表 1 である。

バージョン	リリース日	キリル文字関連の変更点
Unicode 5.0	2006/7/14	
Unicode 5.1	2008/3/15	U+0487 追加 U+0514~U+0523 追加 Cyrillic Extended-A 表追加 (U+2DE0~U+2DFF) Cyrillic Extended-B 表追加 (U+A640~U+A69F)
Unicode 5.2	2009/10/1	U+0524, U+0525 追加
Unicode 6.0	2010/10/11	U+0526, U+0527 追加 U+A660, U+A661 追加
Unicode 6.1	2012/1/31	U+A674~U+A67B 追加 U+A69F 追加
Unicode 6.2	2012/9/26	
Unicode 6.3	2013/9/26	
Unicode 7.0	2014/6/16	U+0514~U+052F 追加 U+A698~U+A69D 追加

表 1 Unicode 5.0~7.0までの変化

2.1.1. Unicode 5.1

Version 5.0 では基本キリル文字として U+0487 以外の U+0400~U+04FF のグリフ、補助キリル文字として U+0500~U+0513までグリフの割り当てがされていた。そしてこれ以降の Unicode の中で、キリル文字が大幅に増やされたのが、前研究の後発表された Version 5.1 からである。

Version 5.1 では、まず基本キリル文字の中で、唯一グリフの割り当てがなかった U+0487 に「[~]」の文字が割り当てられた。また補助キリル文字では、U+0514~U+0523までの16グリフ「**ӀӂӃӄӅӆӇӈӉӊӋӌӍӎӏӐӑӒӓӔӕӖӗӘәӚӛӜӝӞӟӠӡӢӣӤӥӦӧӨөӪӫӬӭӮӯӰӱӲӳӴӵӶӷӸӹӺӻӼӽӾӿӰӱӲӳӴӵӶӷӸӹӺӻӼӽӾӿ**」が新たに割り当てられた。そしてこのバージョンで、Extended-Cyrillic A と Extended-Cyrillic B (以下拡張キリル文字 A、および拡張キリル文字 B) が追加された。これは文字コードとして、U+2DE0~U+2DFF と U+A640~U+A690に割り当てられている。拡張キリル文字 A はすべて縮小補助文字で、他の文字の上

にルビのように付随して用いられる物である。

一方の拡張キリル文字 B には、初期キリル文字や、各種補助記号、また拡張キリル文字 A 同様の縮小補助文字および記号が含まれている。この拡張キリル文字 B の追加で、前研究では外字領域に割り当てざるを得なかった「Ѳ」「ѳ」および「Ѵ」「ѵ」も追加されている。ただし Version 5.1 の時点ではいくつかの部分のグリフについては未割り当て（各種キリル文字用の場所として文字コードは確保された状態）であった。

2.1.2. Version 5.2~6.3

Version 5.1 で大幅な文字コードとグリフの追加があった一方で、Version 5.2~6.2の改訂では、一部のグリフの追加にとどまっている。

まず Version 5.2 では U+0524 「Ѳ」 および U+5425 「ѳ」 の 2 つのグリフの追加、Version 6.0 では U+0526 「Ѵ」、U+0527 「ѵ」、U+A660 「Ѳ」、U+A661 「ѳ」 の 4 グリフの追加となっている。また Version 6.1 では U+A674~U+A67B および U+A69F の縮小補助文字が追加された。一方 Version 6.2 および 6.3 では、キリル文字関連では特に新たなコードの割り当ておよびグリフの追加はなかった。

2.1.3. Version 7.0

2014年12月現在、Unicode Version 7.0 が Unicode の最新版であるが、Version 7.0 でもいくつかのグリフの追加が行われた。ここでグリフの割り当てが進んだのが、補助キリル文字の U+0528~U+052F、および U+A698~U+A69D である。これで基本キリル文字、補助キリル文字、拡張キリル文字 A、拡張キリル文字 B の 4 つの中で、U+A69E をのぞきすべての文字コードにグリフが割り当てられることとなった。

このことから、今後キリル文字関連で Unicode へのグリフの追加は、U+A69E への 1 グリフの割り当て以外では、まず新たな文字コードの割り当てが必要となり、Version 5.1 での拡張キリル文字同様に大幅な変更となる可能性が

ある。古ロシア語文献をデータ化している立場からは、次の表 2 に見られるような、いわゆる初期キリル文字のグリフが追加されることが望まれる。

СЕ ЖЕ СОУТЬ ИМЕНА КНЯЗЕМЪ КИВЬСКИМ.

表 2 初期キリル文字の書体の例

2.1.4. これまでに割り当てられたグリフの一覧

これまでに Unicode に割り当てられた、キリル文字関連のグリフが次の表 3～6 である（すべて <http://www.unicode.org> より引用）。表の太枠部分は、Version 5.1 以降に追加された部分を表している（筆者による）。

それぞれの表の特徴としては、補助キリル文字で追加されたもの（特に後半部分）には、もともとは複数の文字だが、活字化する時点でグリフが結合して 1 グリフとして印刷されるグリフが多い。特に **ꙗ** や **ꙗꙗ** は 3 文字文の合字となっている。

また、拡張キリル文字 A は前述の通りすべて縮小補助文字、拡張キリル文字 B は、単独の特殊キリル文字および記号、縮小補助文字などに区別することができる。

	040	041	042	043	044	045	046	047	048	049	04A	04B	04C	04D	04E	04F
0	È	А	Р	а	р	è	Ѡ	ѡ	Ѣ	Г	К	У	І	Ǻ	З	Û
	0400	0410	0420	0430	0440	0450	0460	0470	0480	0490	04A0	04B0	04C0	04D0	04E0	04F0
1	Ë	Б	С	б	с	ë	ѡ	ѣ	Ѥ	Г	К	У	Ж	ǻ	з	ü
	0401	0411	0421	0431	0441	0451	0461	0471	0481	0491	04A1	04B1	04C1	04D1	04E1	04F1
2	Ѣ	В	Т	в	т	ђ	Ѥ	Ѧ	ѧ	Ѩ	Х	Ж	Ǻ	Й	Ў	
	0402	0412	0422	0432	0442	0452	0462	0472	0482	0492	04A2	04B2	04C2	04D2	04E2	04F2
3	Ѓ	Г	У	г	у	ѓ	Ѧ	Ѩ	ѩ	Ѫ	Х	К	Ǻ	й	ў	
	0403	0413	0423	0433	0443	0453	0463	0473	0483	0493	04A3	04B3	04C3	04D3	04E3	04F3
4	Є	Д	Ф	д	ф	є	Ю	Ѧ	Ѩ	ѩ	Ѫ	ѫ	Ѭ	Й	Ї	Ї
	0404	0414	0424	0434	0444	0454	0464	0474	0484	0494	04A4	04B4	04C4	04D4	04E4	04F4
5	Є	Е	Х	е	х	є	Ю	Ѧ	Ѩ	ѩ	Ѫ	ѫ	Ѭ	æ	й	ї
	0405	0415	0425	0435	0445	0455	0465	0475	0485	0495	04A5	04B5	04C5	04D5	04E5	04F5
6	І	Ж	Ц	ж	ц	і	Ѧ	Ѩ	ѩ	Ѫ	ѫ	Ѭ	ѭ	Ё	Ӧ	Г
	0406	0416	0426	0436	0446	0456	0466	0476	0486	0496	04A6	04B6	04C6	04D6	04E6	04F6
7	Ї	З	Ч	з	ч	ї	Ѧ	Ѩ	ѩ	Ѫ	ѫ	Ѭ	ѭ	ё	ӧ	г
	0407	0417	0427	0437	0447	0457	0467	0477	0487	0497	04A7	04B7	04C7	04D7	04E7	04F7
8	Ј	И	Ш	и	ш	ј	Ѧ	Ѩ	ѩ	Ѫ	ѫ	Ѭ	ѭ	Э	Ө	Ы
	0408	0418	0428	0438	0448	0458	0468	0478	0488	0498	04A8	04B8	04C8	04D8	04E8	04F8
9	Љ	Й	Щ	й	щ	љ	Ѧ	Ѩ	ѩ	Ѫ	ѫ	Ѭ	ѭ	э	ө	ы
	0409	0419	0429	0439	0449	0459	0469	0479	0489	0499	04A9	04B9	04C9	04D9	04E9	04F9
A	НЬ	К	Ъ	к	ъ	њ	Ѧ	Ѩ	ѩ	Ѫ	ѫ	Ѭ	ѭ	Э	Ө	Ң
	040A	041A	042A	043A	044A	045A	046A	047A	048A	049A	04AA	04BA	04CA	04DA	04EA	04FA
B	Ѣ	Л	Ы	л	ы	ђ	Ѧ	Ѩ	ѩ	Ѫ	ѫ	Ѭ	ѭ	э	ө	Ң
	040B	041B	042B	043B	044B	045B	046B	047B	048B	049B	04AB	04BB	04CB	04DB	04EB	04FB
C	К	М	Ь	м	ь	ќ	Ѧ	Ѩ	ѩ	Ѫ	ѫ	Ѭ	ѭ	Э	Ө	Ң
	040C	041C	042C	043C	044C	045C	046C	047C	048C	049C	04AC	04BC	04CC	04DC	04EC	04FC
D	Й	Н	Э	н	э	й	Ѧ	Ѩ	ѩ	Ѫ	ѫ	Ѭ	ѭ	ж	э	Ң
	040D	041D	042D	043D	044D	045D	046D	047D	048D	049D	04AD	04BD	04CD	04DD	04ED	04FD
E	Ў	О	Ю	о	ю	ў	Ѧ	Ѩ	ѩ	Ѫ	ѫ	Ѭ	ѭ	З	У	Ң
	040E	041E	042E	043E	044E	045E	046E	047E	048E	049E	04AE	04BE	04CE	04DE	04EE	04FE
F	Ц	П	Я	п	я	ц	Ѧ	Ѩ	ѩ	Ѫ	ѫ	Ѭ	ѭ	з	у	Ң
	040F	041F	042F	043F	044F	045F	046F	047F	048F	049F	04AF	04BF	04CF	04DF	04EF	04FF

表 3 基本キリル文字

	050	051	052
0	Д	Е	Љ
	0500	0510	0520
1	д	е	љ
	0501	0511	0521
2	Ђ	Ј	Њ
	0502	0512	0522
3	ђ	ј	њ
	0503	0513	0523
4	Ѕ	ЈХ	Ц
	0504	0514	0524
5	ѕ	јх	ц
	0505	0515	0525
6	Ѕ	К	Ѓ
	0506	0516	0526
7	ѕ	к	ѓ
	0507	0517	0527
8	Љ	ЈЕ	Ј
	0508	0518	0528
9	љ	је	ј
	0509	0519	0529
A	Н	Q	ДК
	050A	051A	052A
B	н	q	дк
	050B	051B	052B
C	G	W	Д
	050C	051C	052C
D	G	w	д
	050D	051D	052D
E	Т	К	Л
	050E	051E	052E
F	т	к	л
	050F	051F	052F

表 4 補助キリル文字

	2DE	2DF
0	б	ц
	2DE0	2DF0
1	в	ч
	2DE1	2DF1
2	г	ш
	2DE2	2DF2
3	д	щ
	2DE3	2DF3
4	ж	е
	2DE4	2DF4
5	з	ст
	2DE5	2DF5
6	к	а
	2DE6	2DF6
7	л	е
	2DE7	2DF7
8	м	а
	2DE8	2DF8
9	н	ы
	2DE9	2DF9
A	о	ь
	2DEA	2DFA
B	п	ю
	2DEB	2DFB
C	р	я
	2DEC	2DFC
D	с	а
	2DED	2DFD
E	т	х
	2DEE	2DFE
F	х	и
	2DEF	2DFE

表 5 拡張キリル文字A

	A64	A65	A66	A67	A68	A69
0	З	Ы	Ц	☼	Д	Т
	A640	A650	A660	A670	A680	A690
1	з	ы	ц	☼	д	т
	A641	A651	A661	A671	A681	A691
2	З	Ь	Д	☼	С	Ч
	A642	A652	A662	A672	A682	A692
3	з	ь	д	☼	с	ч
	A643	A653	A663	A673	A683	A693
4	З	О	Л	е	ЗК	Һ
	A644	A654	A664	A674	A684	A694
5	з	о	л	е	зк	һ
	A645	A655	A665	A675	A685	A695
6	л	Я	М	и	Ч	Щ
	A646	A656	A666	A676	A686	A696
7	л	я	м	у	ч	щ
	A647	A657	A667	A677	A687	A697
8	Ӑ	А	О	ь	Д	∞
	A648	A658	A668	A678	A688	A698
9	ӑ	а	о	ы	д	∞
	A649	A659	A669	A679	A689	A699
A	ӕ	ӕ	⊙	ь	Ѳ	⊕
	A64A	A65A	A66A	A67A	A68A	A69A
B	ӗ	ӗ	⊙	w	ѳ	⊖
	A64B	A65B	A66B	A67B	A68B	A69B
C	Ɔ	Я	∞	˘	т	ь
	A64C	A65C	A66C	A67C	A68C	A69C
D	Ɔ	я	∞	˘	т	ь
	A64D	A65D	A66D	A67D	A68D	A69D
E	Ѳ	Ѳ	⊙	˘	Ц	▨
	A64E	A65E	A66E	A67E	A68E	
F	ѳ	Ѳ	˘	˘	Ц	и
	A64F	A65F	A66F	A67F	A68F	A69F

表 6 拡張キリル文字B

2.2. コンピューターの OS の変化

前研究の当時、コンピューターの OS としては Windows XP が主流となっており、このためコンコーダンスの表示環境としては Windows XP を主に想定して研究を進めた。しかし現在までに OS の種類や Windows のバージョンなどいくつかの点で大きな変化が起きている。

2.2.1. タブレット端末の普及と表示機器としての可能性

前研究以降の大きな変化の一つが、iPad や Android タブレットに代表されるタブレット端末の普及であり、これに伴うモバイル OS のシェア拡大である。

将来的にコンコーダンスをインターネット上へ公開することは、前研究当時も考えられていた。またその表示環境としては、Windows XP に搭載された標準ブラウザである Internet Explorer を想定していた。それが、その後の iPad または Android タブレットの登場により、インターネットの表示環境としてモバイル OS およびそのブラウザも考慮する必要が出てきた。このため、今回の研究でもモバイル OS への対応が考慮されるべきであると考え、次の 2 点について簡単ではあるが検証を行った。

まず一つ目が、モバイル OS に標準搭載されるフォントおよびブラウザが準拠している、Unicode のバージョンについてである。このために、基本キリル文字、補助キリル文字、拡張キリル文字 A、および拡張キリル文字 B の全グリフの一覧表を、html ファイルとしてフォントの指定無しで作成した。これを iOS 8.1 および Android 4.4 のブラウザで表示させ、どのグリフが表示可能かを確認した。次に示す図 1、図 2 は、この表示結果のスクリーンショットである。

この結果、iOS 8.1 および Android 4.4 とともに、標準ブラウザで利用されるフォントは、Unicode Version 5.0 相当であると考えられた。

次に 2 つ目の検証としては、両 OS の標準ブラウザが woff 形式のウェブフォントに対応しているかどうかを行った。後述するところであるが、両 OS のブラウザが対応している css 3.0 は woff 形式のウェブフォントが利用可能

であるとされている。キリル文字についてすべてが表示できるか確認するために、上記で作成したキリル文字グリフの一覧の html ファイルに、フォントの指定を css3.0 形式で加えて表示させた結果が、図 3、図 4 である。この表示結果から、woff 形式のフォントを利用することで、モバイル OS 上でこれまでに発表されているキリル文字のすべてのグリフを表示可能であることが確認できた。

2.2.2. MS-Windows

前述の通り、パソコン上でのキリル文字の表示については、ウェブフォント等を指定しないかぎり OS に標準搭載されるフォントに左右される。2014年現在、パソコンの OS のシェアとしては MS-Windows 7 がその中心であり、また MS-Windows 8 がシェアを伸ばしてきている。このため、ここでは Windows 7 および 8 に標準搭載されているフォントについてキリル文字の対応状況を確認する。

まず現在パソコン OS のシェアとして一番大きい Windows 7 についてである。マイクロソフト社の Windows シリーズに標準搭載されるセリフ体フォントの中で、欧文文字の収録グリフ数をもっとも多いのが Times New Roman フォントであり、これは Windows 7、Windows 8 においても同様である。前研究で対象とした Windows XP に含まれる Times New Roman フォントは、Unicode Version 1.1.5 準拠であると考えられたが、Windows 7 に含まれる同フォントは Version 5.0 準拠と考えられる。Version 5.0 ではキリル文字に関しては基本キリル文字と、補助キリル文字の U+0513 までのグリフを含むこととなる。ただし、Windows 7 の Times New Roman フォントでは、例外として U+051A ~ U+051D の「Q q Ww」の 4 グリフについても追加されている。

前研究では、Unicode Version 5.0 対応のフォントを作っているが、標準搭載フォントで同等以上のキリル文字の表示が可能となっており、この意味では前研究で作成したフォントの役割は終わったといえる。

次に Windows 8 に搭載される Times New Roman フォントについてであるが、基本キリル文字、および補助キリル文字については Unicode Version 6 準拠と

考えられる。グリフの収録状況としては Windows 7 と同様に、基本キリル文字のグリフについてはすべて網羅されている。次に補助キリル文字のグリフについては、Windows 7 のものが U+0513 までの収録であったのに対して、U+0527 まで収録されており、この点で Version 6 対応と言うことができる。一方で、拡張キリル文字 A および拡張キリル文字 B のグリフについては、収録されておらず、この点は Version 5.0 準拠ともいえる状況である。

またサンセリフ体のフォントとして、Windows 8 で標準 UI フォントとなっている SeGoe UI フォントについても確認をしている。このフォントは Windows 8 標準搭載フォント中、欧文文字に関して最大の収録グリフ数を備えている。このフォントでは、基本キリル文字、補助キリル文字に加えて、すべての拡張キリル文字 A および U+A697 までの拡張キリル文字 B についてもグリフが割り当てられている。このことから、このフォントは Unicode の Version 6.0~6.3 のいずれかに、正式に準拠しているといえるだろう。

この点で、前研究および本研究のきっかけである「**Ѳ ѳ Ѵ ѵ**」といったキリル文字が、OS 搭載の標準フォントで可能になったことが確認できた。ただし、SeGoe UI フォントは前述の通りサンセリフ体のフォントである。欧文の文章の本文にはセリフ体を用いることが通例とされており、本文中では SeGoe フォントは用いにくい。このため、セリフ体のグリフを持つ、最新版の Unicode のバージョンに準拠したフォントの必要性は、いまだなくなっていないといえるだろう。

2.3. ウェブフォント

前研究の当時、html および css でフォントを指定するには、OS の標準搭載フォントを選ぶか、追加のフォントをユーザーにインストールしてもらう必要があった。このため、中條、酒井2006などで作成した CD-ROM 版コンコーダンスには Windows 用フォントを収録しており、これをコンコーダンス利用初回時にインストールするように指示をしていた。これが2011年に w3c による勧告が行われた css 3.0² により、html および css ファイルでのフォントの扱い

に道が開かれた。それが html 上でフォントを扱う標準規格であるウェブフォントである。これにより、css 3.0 に対応したいいわゆるモダンブラウザでは、統一的にウェブフォントを扱うことが可能となった。

ウェブフォントの中心となる woff (Web Open Font Format) フォントは、Windows などでも標準となっている True Type Font 形式を基礎として、付加情報をつけた上で容量を圧縮した形式といえる。このため、実際の woff ファイルは、True Type Font からツールを用いて簡単に作成することが可能である。この woff フォントファイルを html ファイルや css ファイルとともにウェブ上にあげることによって、そのフォントファイルを利用することが可能である。

ウェブフォントという多くの場合、Google Fonts や Adobe Typekit に代表されるような、インターネット上に固定的に公開されているフォントが利用されることが多いが、オリジナルのウェブフォントを作成して、これを用いることも可能である。

また、実際の css のコーディングとしては、Web Open Font Format と True Type Font を併用することで、多くのウェブブラウザでオリジナルのフォントを利用することが可能であるとされている。以前はインターネットエクスプローラーの独自形式である EOT ファイル、および古いブラウザに対応するため SVG ファイルも必要とされていたが、モダンブラウザの普及により現在は上記 2 種類のファイルのみの形式で十分になっている。

このことにより、これまでの CD-ROM 版コンコードランスの初回利用時に必要とされていた、ユーザーによるフォントのインストール作業を省略することが可能となった。これは CD-ROM 版ユーザーの利便性を大きく向上することになるとともに、将来的にインターネット上のコーパスとして、コンコードランスを公開するときにも用いることのできる技術であるといえる。

2.4. GNU Free Font

前研究では良く用いられるキリル文字フォントとして、Izhizsa フォントを取り上げて、その詳細について分析した。その中で、このフォントの著作権者

があいまいである点や、ライセンス形態が不明な点から、CD-ROM 版コンコードダンスに収録することは断念していた。そして、このフォントのかわりにライセンス形態を明確にした Preslav フォントを作成して CD-ROM に収録した。

このような状況でフォントのライセンス形態について検討する中で、新たにフリーソフトとして扱われている GNU Free Font というフォントを発見することとなった。このフォントは、GNU GPL+FE というフォント用のライセンスを持って公表されているもので、2014年現在で Unicode Version 6.0 に準拠して作成されている。またこのフォントの配布形式は Open Type Font、True Type Font、web open font format となっており、この点でも本研究の目的とほぼ合致している。また、このプロジェクトは現在も継続されており、今後も Unicode のバージョンアップに準拠したフォントが作成されるだろうことが期待される。

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
040	È	É	Ђ	Ѓ	Е	Ѕ	Ї	Ј	Љ	Њ	Ћ	К	И	У	Ц	
041	А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П
042	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я
043	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п
044	р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я
045	è	é	ђ	ѓ	е	ѕ	ї	ј	љ	њ	ћ	к	и	у	ц	
046	Ґ	ґ	Ғ	ғ	Ҕ	ҕ	Җ	җ	Ҙ	ҙ	Ӏ	Ӂ	ӂ	Ӄ	ӄ	Ӆ
047	Ҧ	ҧ	Ҩ	ҩ	Ҫ	ҫ	Ү	ү	Ұ	ұ	Ӧ	ӧ	Ӱ	ӱ	Ӳ	ӳ
048	Ҷ	ҷ	Ҹ	ҹ	Һ	һ	Ҽ	ҽ	Ҿ	ҿ	Ӫ	ӫ	Ӭ	ӭ	Ӯ	ӯ
049	Г	г	Ґ	ғ	Ғ	ғ	Ҕ	ҕ	Җ	җ	Ҙ	ҙ	Ӏ	Ӂ	ӂ	Ӄ
04A	К	к	Ғ	ғ	Ҕ	ғ	Җ	җ	Ҙ	ҙ	Ӏ	Ӂ	ӂ	Ӄ	ӄ	Ӆ
04B	Ұ	ұ	Ӏ	Ӂ	ӂ	Ӄ	ӄ	Ӆ	ӆ	Ӈ	ӈ	Ӊ	ӊ	Ӌ	ӌ	Ӎ
04C	Ӑ	ӑ	Ӓ	ӓ	Ӕ	ӕ	Ӗ	ӗ	Ә	ә	Ӛ	ӛ	Ӝ	ӝ	Ӟ	ӟ
04D	Ӡ	ӡ	Ӣ	ӣ	Ӥ	ӥ	Ӧ	ӧ	Ө	ө	Ӫ	ӫ	Ӭ	ӭ	Ӯ	ӯ
04E	ӱ	Ӳ	ӳ	Ӵ	ӵ	Ӷ	ӷ	Ӹ	ӹ	Ӻ	ӻ	Ӽ	ӽ	Ӿ	ӿ	
04F	Ӱ	ӱ	Ӳ	ӳ	Ӵ	ӵ	Ӷ	ӷ	Ӹ	ӹ	Ӻ	ӻ	Ӽ	ӽ	Ӿ	ӿ
050	đ	đ	đ	đ	đ	đ	đ	đ	đ	đ	đ	đ	đ	đ	đ	đ
051	Є	є	Ј	ј	Љ	љ	Њ	њ	Ћ	ќ	Ќ	ќ	Ў	ў	Ѡ	ѡ
052	Љ	љ	Њ	њ	Ћ	ќ	Ќ	ќ								
2DE	°	°	°	°	°	°	°	°	°	°	°	°	°	°	°	°
2DF	°	°	°	°	°	°	°	°	°	°	°	°	°	°	°	°
A64	Ƶ	ƶ	Ʒ	Ƹ	ƹ	ƺ	ƻ	Ƽ	ƾ	ƿ	ƺ	ƻ	Ƽ	ƾ	ƿ	ƺ
A65	Ƶ	ƶ	Ʒ	Ƹ	ƹ	ƺ	ƻ	Ƽ	ƾ	ƿ	ƺ	ƻ	Ƽ	ƾ	ƿ	ƺ
A66	Ƶ	ƶ	Ʒ	Ƹ	ƹ	ƺ	ƻ	Ƽ	ƾ	ƿ	ƺ	ƻ	Ƽ	ƾ	ƿ	ƺ
A67	Ƶ	ƶ	Ʒ	Ƹ	ƹ	ƺ	ƻ	Ƽ	ƾ	ƿ	ƺ	ƻ	Ƽ	ƾ	ƿ	ƺ
A68	Ƶ	ƶ	Ʒ	Ƹ	ƹ	ƺ	ƻ	Ƽ	ƾ	ƿ	ƺ	ƻ	Ƽ	ƾ	ƿ	ƺ
A69	Ƶ	ƶ	Ʒ	Ƹ	ƹ	ƺ	ƻ	Ƽ	ƾ	ƿ	ƺ	ƻ	Ƽ	ƾ	ƿ	ƺ

表7 GNU Free Font に収録されているキリル文字関連グリフ一覧

本研究で今回作成するフォントは、Unicode Version 7.0 対応とするため、現在のところはこの GNU Free Font と差別化できているが、グリフの完成度の高さや、他の言語の文字への対応（表 8）状況からも、Preslav フォントの開発を続ける意義は今後薄くなると考えられる。このため、今後の研究方針としては、GNU Free Font の開発状況を踏まえての対応が必要になるであろう。

文 字 種	収録数	文 字 種	収録数
Basic Latin	98	Mathematical Operators	391
Latin-1 Supplement	96	Miscellaneous Technical	113
Latin Extended-A	128	Control Pictures	39
Latin Extended-B	208	Enclosed Alphanumerics	10
IPA and Phonetic Extensions	288	Box Drawing	116
Spacing Modifier Letters	112	Block Elements	32
Combining Diacritical Marks (+suppl.)	145	Geometric Shapes	96
Greek and Coptic	134	Miscellaneous Symbols	191
Coptic	66	Dingbats	191
Cyrillic (+suppl., +ext.-A, -B)	408	Private Use Area	2448
Armenian	87	Alphabetic Presentation Forms	58
Hebrew	87	Arabic Presentation Forms-A	169
Arabic (+suppl.)	300	Combining Half Marks	7
Devanagari (+ext.)	155	Arabic Presentation Forms-B	141
Bengali	92	Specials	5
Gurmukhi	79	Thaana	50
Gujarati	84	Sinhala	80
Oriya	90	Ethiopic (+suppl., +ext.)	345
Tamil	72	Cherokee	85
Malayalam	98	Tagalog Hanunoo Buhid Tagbanwa	23
Thai	87	Gothic	27
Georgian (+suppl.)	83	Byzantine & Western Musical Symbols	466
Latin Extended (Additional,C,D)	379	Mathematical Alphanumeric Symbols	996
Greek Extended	233	Tai Le	35
General Punctuation (+suppl.)	138	Buginese	30
Superscripts and Subscripts	34	Glagolitic	86
Currency Symbols	26	Ancient Greek Numbers	5
Combining Diacritical Marks for Symbols	33	Domino and Mahjong Tiles	144
Letterlike Symbols	77		
Number Forms	58	total in font	10537
Arrows (+suppl.)	183	total in Private Use	2448

表 8 GNU Free Font に含まれる文字の数 (GNU Free Font サイトより引用)

2.5. インターネット上のコーパスの増加

ICTの向上とネットサービスの多様化により、最近インターネット上のコーパスが増加してきている。日本語でも国立国語研究所による書き言葉コーパス「少納言」のサービス提供により、気軽にだれでもがコーパス、またその展開形としてのコンコーダンスを利用できるようになっている。

ロシア語についても同様に18世紀～現代ロシア語の文章を収録した“Russian National Corpus”など、様々なコーパスが出現してきている。このような点から、今後ロシア諸年代記のコンコーダンスについてもインターネット上のコーパスとしての展開が必要であると考えられ、その点でも Unicode Version 7.0 に対応したウェブフォントの必要性が高まっているといえるだろう。

2.6. まとめ

前研究では必要とされるグリフを、基本キリル文字の U+0400～U+04FF および補助キリル文字の U+0500～U+052F までの範囲としていた。しかしながら、これまでに見てきた状況の変化により、新たにフォントを更新するのであれば、基本キリル文字と補助キリル文字のグリフに加えて、拡張キリル文字 A および拡張キリル文字 B のすべてのグリフを含む必要があると考えられた。

また、これまでの研究では Windows に対応する True Type Font 形式でフォントを作成してきたが、今後は True Type Font 形式でフォントを開発した上で、ウェブフォントである woff 形式への変換をする必要があることがわかった。また同時に、フォントの表示確認対象としては、パソコンの OS である MS 社の Windows シリーズのウェブブラウザに加えて、iOS や Android といったモバイル OS 上のウェブブラウザについても必要であると考えられた。

3. フォントの作成について

3.1. フォント作成の方針

本研究で作成している Preslav フォントは、酒井2009で作成したものに更新を加えたものである。実際の収録文字としては、Unicode Version 7.0 に準拠し

て、基本ラテン文字94文字、基本キリル文字256文字、補助キリル文字48文字、拡張キリル文字A32文字、拡張キリル文字B95文字を収録することとしている。

基本ラテン文字については、前回の研究では作成していなかったが、基本的なフォント名などを表すためにも必要なため、94文字すべてのグリフを作成している。これは基本キリル文字からコピーする部分も多い。またフォントフェイス、いわゆるフォントの書体のデザインについては、前回と同様にセリフ体として統一している。

一方、フォントの著作権については、GNU GPL font exception clause に沿ったライセンス形態を取るものとする。これは、通常の GPL に加えて、フォントを用いる場合の例外条項を追加したものである。通常の GPL だけの場合文章中に GPL のフォントを用いるとその文書ファイルも GPL として公開する必要が出てきてしまう。これに対して FE (font exception clause) が追加されると、文書中にフォントとして利用する場合には、その文書ファイルは GPL として扱う必要がないというものである。このため、フォントを GPL として配布する場合には通常この GPL+FE が用いられるため、このライセンス形態を取ることにした。

3.2. Preslav.ttf Version 1.0

次の表 9 が、True Type Font 形式で作成した Preslav フォントの一覧である。前述のとおり、Unicode Version 7.0 で規定される基本ラテン文字およびキリル文字をすべて収録している。フォントの作成には武蔵システム社の TTEdit³ を用いた。

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
002		!		#	\$	%	&		()	*	+	,	-	.	/
003	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
004	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
005	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
006	`	a	b	c	d	e	f	g	h	I	j	k	l	m	n	o
007	р	q	р	s	t	u	v	w	x	Y	z	{		}	~	
040	È	É	Ê	Ë	€	Š	İ	Í	J	Љ	Њ	Ћ	Ќ	Й	Û	Ц
041	А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П
042	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я
043	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п
044	р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я
045	é	ë	ñ	í	é	s	i	í	j	љ	њ	ќ	ќ	й	Û	ц
046	Ɔ	w	Ɔ	Ɔ	Ɔ	к	А	А	Ѓ	Ѓ	Ж	ж	Ѓ	Ѓ	Ѓ	Ѓ
047	Ψ	ϕ	Ѳ	ѳ	V	v	Ŵ	ŵ	Oy	oy	О	о	Ѓ	Ѓ	Ѓ	ŵ
048	Ѓ	с	Ѳ	ѳ	Ŵ	ŵ	Ŵ	ŵ	Ŵ	ŵ	Ŵ	ŵ	Ŵ	ŵ	Ŵ	ŵ
049	Г	г	Г	г	Г	г	Ж	ж	З	з	К	к	К	к	К	к
04A	К	к	Н	н	Н	н	П	п	О	о	С	с	Т	т	У	у
04B	У	у	Х	х	Ц	ц	Ч	ч	Ч	ч	Ч	ч	Ч	ч	Ч	ч
04C	І	Ж	ж	К	к	Л	л	Н	н	Н	н	Ч	ч	М	м	І
04D	Ă	ă	Ă	ă	Ă	ă	Ě	ě	Ě	ě	Ě	ě	Ě	ě	Ě	ě
04E	З	з	Й	й	Й	й	Ŏ	ŏ	Ŏ	ŏ	Ŏ	ŏ	Ŏ	ŏ	Ŏ	ŏ
04F	Ÿ	ÿ	Ÿ	ÿ	Ÿ	ÿ	Г	г	Б	б	Г	г	Х	х	Ж	ж
050	d	a	d	d	Ů	ů	Ź	ź	Љ	љ	Њ	њ	G	g	Т	т
051	Е	e	Ј	ј	Ѓ	ѓ	Р	р	Я	я	Q	q	W	w	К	к
052	Љ	љ	Њ	њ	Ц	ц	Ч	ч	Н	н	Д	д	Я	я	Л	л
2DE	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
2DF	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
A64	Ž	ž	Ž	ž	Ž	ž	Š	š	Š	š	Š	š	Š	š	Š	š
A65	Ы	ы	Ы	ы	Ы	ы	Н	н	А	а	Ж	ж	Н	н	А	а
A66	И	и	И	и	И	и	М	м	О	о	О	о	О	о	О	о
A67	⋯	□	⋯	□	⋯	□										
A68	Д	д	С	с	Ж	ж	Ц	ц	Д	д	Т	т	У	у	Ц	ц
A69	Т	т	Ч	ч	Б	б	Ш	ш	О	о	Ф	ф	Ъ	ъ	Ъ	ъ

表9 Preslav Version 1.0 に収録したグリフ一覧

3.3. Preslav.woff Version 1.0

Presrav.ttf ファイルを元として、武蔵システム社の WOFF コンバーター⁴を用いて woff ファイルを作成している。収録したグリフについては、Presrav.ttf と同様である。

4. 考察

今回の Unicode Version 7.0 に準拠したキリル文字を含む Preslav フォントの作成により、酒井2009で問題となっていた、3つのフォントを混在して利用しなければならない問題は解決したといえる。また、ライセンス形態についても GPL+FE を利用することを明示しており、著作権の問題もクリアできたと考えられる。

一方で、酒井2009で今後の課題として残した Izhitsa 上位互換フォントや、初期キリル文字のフォントフェイスを用いたフォントについては、今回も課題として残っている。これらについては完成次第、ウェブサイト上で公開したいと考えている。

また、これまでの研究の成果として作成したフォントについては、<http://www.scll6.info/>上にて順次公開の予定である。

5. おわりに

今回の研究では、古ロシア語の文献で用いられる初期キリル文字など、特殊なキリル文字をコンピューター上で扱う環境について、酒井2009と比較してどのような点に変化があったのかを確認してきた。そしてそれに基づいて、Unicode Version 7.0 に準拠した、特殊キリル文字を含む Preslav フォントを作成した。このフォントは、GNU GPL + FE ライセンスでの公開を予定している。これにより、デジタル化した古ロシア語文献のコンピューター上での表示に関するほとんどの問題を、今回の研究で解決することができたと考えている。

同時に今回の研究の過程で、その目的をほぼ同じくする GNU Free Font の存在を知ることとなり、今後の研究方針を変更する必要性も出てきた。フォントの独自開発には区切りをつけ、今後は従来のコンコードダンスデータの Unicode 7.0 対応や、インターネット上へのコーパスとしての公開などを研究の主な目的とすべきであろうと考えている。

6. 謝辞

本研究は、文部科学省科学研究費補助金基板研究（C）「古ロシア語文献デジタル化の為の、初期キリル文字を含むウェブフォントの開発について」（課題番号：25370390）の助成を受けたものです。

註

- 1 本来 Unicode のバージョンは、7.0.0 のように3つの数字で表現されるが、本稿ではこれ以降、最後の1桁を省略して7.0のように表記する。
- 2 <http://www.w3.org/TR/css3-fonts/>
- 3 <http://opentype.jp/ttedit.htm>
- 4 <http://opentype.jp/woffconv.htm>

参考文献

- ・川俣晶『パソコンにおける日本語処理文字コードハンドブック』技術評論社、1999
- ・黒田龍之助『羊皮紙に眠る文字たち—スラヴ言語文化入門—』現代書館、1998
- ・小林潔『ユーラシア・ブックレット No. 57 ロシアの文字の話』東洋書店、2004
- ・小林龍生、安岡孝一、戸村哲、三上喜貴編『インターネット時代の文字コード』共立出版、2002
- ・酒井純『古ロシア語文献のコンコーダンス化における技術的問題について』岡崎女子短期大学紀要、第38号、2004
- ・酒井純『古ロシア語文献に用いられる初期キリル文字の Unicode フォントの作成について』親和言語文化研究、創刊号、2009
- ・酒井智宏『特殊文字のコンピュータ処理について —古ロシア語の場合の例—』名古屋大学大学院国際開発研究科修士論文、1996
- ・中條直樹、酒井純『『ノヴゴロド第4年代記』コンコーダンス』日本古代ロシア研究会、2005
- ・中條直樹、酒井純『『イパチー年代記』コンコーダンス』日本古代ロシア研究会、2006
- ・長瀬眞理、西村弘之共著『コンピュータによる文章解析入門 — OCP への招待—』オーム社、1986
- ・深沢千尋『文字コード超研究』ラトルズ、2003

参考サイト

- ・GNU Free Font、<https://www.gnu.org/software/freefont/>、(2014/12/21)
- ・GNU GPL、<http://www.gnu.org/>、(2014/12/21)
- ・Microsoft Typography、<http://www.microsoft.com/typography/>、(2014/12/21)
- ・ParaType (Izhitsa フォント販売元)、<http://www.paratype.com/>、(2014/12/21)
- ・Unicode、<http://www.unicode.org/>、(2014/12/21)