

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

酒 井 純

on making a Unicode font of early Cyrillic
alphabet used in old russian books

Jun SAKAI

キーワード：古ロシア語，初期キリル文字，True Type Font，Unicode

1. はじめに

本研究では，これまでコンピューター上で扱うことの難しかった初期キリル文字についての問題点を明らかにし，これらを解決したフォンを作成することを目指している。そして，Unicode に準拠した文字コードへのマッピングを行うことなどにより，単独のフォントでの古ロシア語文献データの印刷と表示が可能となった。

2. 研究の背景と目的

2.1. 初期キリル文字フォントの必要性

これまでの研究では、中條・酒井 2006 など古ロシア語によるロシア諸年代記のコンコードダンスを作成してきた。このような中で、そのデータ量および検索性から、電子メディアでの出版も考えるべきところにきており(『モスクワ年代記集成』コンコードダンス(中條 2004)で、3 分冊、総 2368 ページにおよぶ)、このために、『ノヴゴロド第 4 年代記』年代記コンコードダンス (中條・酒井 2005) 以降では CD-R によるコンコードダンス配布を行っている。

そして、このときに問題となったのが、初期キリル文字の表示についてである。これらのロシア諸年代記では現代ロシア語と共通する文字(以下、現代ロシア文字)以外に、「Ѹ」や「ѡ」といったいわゆる初期キリル文字と呼ばれる文字が混在している。これまでのコンコードダンスでは紙メディアを用いて発表してきたが、このときにはいくつかのフォントを用いて印刷を行ってきたが、中條・酒井 2006『イパチー年代記』コンコードダンスでは CD-R での配布をするために、html 形式でコンコードダンスのデータを収録したわけだが、表示に関していくつかの問題が生ずることとなった。

このときの方針としては、可能な限り MS-Windows XP に標準で添付されるフォントで大部分が表示できるようにし、特殊な文字(初期キリル文字など)については、追加のフォントを CD-R に収録し、ユーザーによって PC にインストールすることとした。ただし、これまでのコンコードダンスにおいて印刷時に用いてきた Izhitsa フォントについては、再配布に関する著作権等が明確ではないため、そのまま CD-R に付属データとして添付することができなかった。そこで、前回の CD-R 版『イパチー年代記』コンコードダンスでは、「ѠѡѸѹ」の 4 つの文字に特化した、追加フォントを収録することとなった。

しかしながら、この方法では 3 つのフォントを切り替えて表示をしており、html のデータ構造が煩雑になるとともに、今後の諸年代記のコーパス化を目指す XML との整合性の問題もあり、できれば一つのフォントですべてを表示で

きるようにするのが理想である。このために、本研究ではロシア諸年代記など初期キリル文字を用いた文章のデータを、単独で表示可能にするフォントを作成するに至った。

2.2. フォントの著作権

インターネット上には Izhitsa をはじめとしたいくつかの初期キリル文字フォントが存在している。しかし現在のところ、初期キリル文字を含み、再配布権等の著作権表示の明確なフォントの入手は難しい。また、代表的な初期キリル文字フォントといわれる Izhitsa さえも、いくつかの近似フォントがあり、どれがオリジナルなのかハッキリしないといった問題もある。

数々の初期キリル文字を含む文献のデータ化が進む中、初期キリル文字を扱う方法の確立が求められていることは確かである。しかし一方で現在までのところ初期キリル文字フォントについて、再配布等の著作権が完全にクリアされたものはない。このため、本研究で作成するフォントについては基本的に著作権フリー(厳密には GNU GPL)とすることとした。

2.3. 目的

本研究での目的は、次の3つの点にまとめることができる。

まず最初の目的は、現在 WindowsXP 上で利用可能な初期キリル文字のフォントについてまとめるとともに、これらに関する問題点について検討することである。現状で利用可能な環境と、その問題点を洗い出すことにより、製作すべきフォントの方針をハッキリさせることができると考えられる。

次の目的は現状の問題点を元に、ロシア諸年代記に求められる初期キリル文字フォントの設計および作成することである。具体的にはどの文字にどの文字コードを割り当てるかという、マッピングについて検討することが中心となる。詳細については後述するところであるが、本研究では Unicode の方針に即したフォントと、これまでの環境を考慮した、上位互換ともいえるマッピングの 2

種類について検討している。そしてこの方針に従って、フォントを作成している。

最期の目的は、本研究での成果の一つとなる、初期キリル文字をふくむフォントファイルの著作権がどのようにあるべきか検討することである。これは、前述のように、これまで用いられてきたフォントの著作権が明確でない点をふまえ、今回作成するフォントの著作権表示として GNU GPL を採用することを検討している。

3. 現在 Windows で利用可能なキリル文字について

現在パソコン上で初期キリル文字を含む文書を扱う場合、大きく二つの方法があるといえる。1つは、MS-Windows XP(以下 Windows と表記)や MacOS X など Unicode, つまり UTF-16LE を文字コード体系として採用した OS 上で、ワープロソフトなどのアプリケーションソフトを用いる方法である。またもう1つとしては、Tex と呼ばれるテキスト整形ソフトを用いる方法である。Tex については、OS や OS の文字コード体系に依存せず、また初期キリル文字を含むフォントもフリーソフトとして扱えるものであるが、Windows 上での扱いが煩雑であり一般化していないこと、また本研究では html 上からの利用も大きな目標となっていることから、これらについてはふれていない¹。

3.1. Windows でのキリル文字フォントの状況

まず、日本語版 WindowsXP (SP2)に標準で収録されているフォントの、キリル文字の対応状況についてである。

現在キリル文字を使用可能な Windows のフォントとしては、Unicode 1.1.5 以降対応の MS 明朝フォントなどいわゆる和文フォントと呼ばれるものと、TimesNewRoman, Century などいわゆる欧文フォントと呼ばれるものの2種

¹ Tex でのキリル文字の扱いについては、Nox Insomniae (<http://www.ceres.dti.ne.jp/~i-yasuda/>)に詳しい。

類に大きく分類することができる。どちらも現在は Unicode²フォントとして実装されているが、Unicode へ移行する前まではそれぞれ ISO2022, ISO646 と異なった符号化文字集合に基づいているため、上記のような呼び方が日本国内では定着している。

3.2. Unicode1.1.5 以前の欧文フォント

まず先に取り上げるのが、TimesNewRoman や Century など、いわゆる欧文フォントと従来呼ばれてきたフォント群についてである。まず欧文フォント例として、TimesNewRoman フォントのキリル文字収録状況について次の表 1 にあげる。

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
040		Ё	Ѣ	Ѓ	Є	Ѕ	І	Ї	Ј	Љ	Њ	Ѧ	Ќ		Ў	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		ё	ђ	ѓ	є	ѕ	і	ї	ј	љ	њ	ѧ	ќ		ў	ц
046																
047																
048																
049	Ѐ	ѐ	Ɔ	џ			Ж	ж			К	к	К	к		
04A			Ѓ	џ											У	У
04B	Ў	ў	Х	х					Ѣ	Ѣ	Н	н				
04C																
04D									Ѣ	Ѣ						
04E									Ѣ	Ѣ						
04F																

表1 Times New Roman フォント キリル文字一覧

² 符号化文字集合として Unicode, 実際の文字エンコーディングスキームとして UTF-16 Little Endian が用いられている。

この表 1 は、Times New Roman フォントで、Unicode のキリル文字領域である、U+0400 ~ U+04FF³までに割り当てられた文字をまとめたものである。この表では、左側 1 列目のラベルが Unicode 番号の上 3 桁、上側 1 行目のラベルが Unicode 番号の下 1 桁を表している。例えば、2 列 041 行にある「B」の Unicode 番号は U+0412 となる。

この表からわかるように、Times New Roman フォントでは、U+0460 以降の文字については収録されている文字が非常に限定されている。このため、初期キリル文字の「Ѣ」「Ѧ」「Ѯ」などの文字をこれらの欧文フォントで扱うことはできない。

ここにあげた文字など、U+0460 ~ U+04FF の Unicode 番号に当たる文字は、Unicode 1.1.5 の時点においてほとんどのものが定義されている。しかしながら欧文フォントにこれらの文字の多くが収録されていないことは、このフォントが作成されたのが 1992 年であることから考えると、Unicode 1.0 に従ったものかとも考えられる。ただし、Unicode 1.0 のマッピングを現時点では入手できず、この点については今後の追跡調査が必要であると考えられる。

3.3. Unicode 1.1.5.対応フォント

次に、MS 明朝や MS ゴシックなどのいわゆる和文フォントと、Arial Unicode MS⁴についてである。これらのフォントでは、Unicode 1.1.5 に則り文字が収録されており、上記のいわゆる欧文フォントより多くの文字が収録されている。次の表 2 は、MS 明朝フォントのキリル文字収録状況である。

³ Unicode 符号化文字集合では、文字番号を通常 U+xxxx(x は 16 進数)の形で表す。ただし、本来の Unicode 文字体系は 32bit であり、正式には U+0xxxxxxxx が正しい表記であり、上記 U+xxxx は U+0000xxxx を表す簡略表記である。

⁴ Arial Unicode MS については、一時 Microsoft 社により無料配布されていたが、現在は同社サイトからのダウンロードは中止されており、再開が望まれる。

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
040		Ё	Ѣ	Ѧ	Ѹ	Ѻ	Ѽ	Ѿ	ѿ	ѿ	ѿ	ѿ	ѿ		ѿ	ѿ
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		ё	ѣ	ѧ	ѹ	ѻ	ѽ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ		ѿ	ѿ
046	Ѡ	ѡ	Ѣ	ѣ	Ѥ	ѥ	Ѧ	ѧ	Ѩ	ѩ	Ѫ	ѫ	Ѭ	ѭ	Ѯ	ѯ
047	Ѱ	ѱ	Ѳ	ѳ	Ѵ	ѵ	Ѷ	ѷ	Ѹ	ѹ	Ѻ	ѻ	Ѽ	ѽ	Ѿ	ѿ
048	Ѡ	ѡ	Ѣ	ѣ	Ѥ	ѥ	Ѧ	ѧ	Ѩ	ѩ	Ѫ	ѫ	Ѭ	ѭ	Ѯ	ѯ
049	Г	г	Г	г	Б	б	Ж	ж	З	з	К	к	К	к	К	к
04A	К	к	Н	н	Н	н	П	п	Q	q	С	с	Т	т	У	у
04B	У	у	Х	х	Ц	ц	Ч	ч	Ч	ч	h	h	Ѣ	ѣ	Ѣ	ѣ
04C	І	Ѫ	ѫ	Ѭ	ѭ			Н	н			Ч	ч			
04D	Ă	ă	Ā	ā	Æ	æ	Ė	ė	Ə	ə	Ė	ė	Ж	ж	Ž	ž
04E	З	з	Й	й	Й	й	Ö	ö	Ө	ө	Ө	ө			Ÿ	ÿ
04F	Ÿ	ÿ	Ÿ	ÿ	Ÿ	Ÿ			Ы	ы						

表2 MS 明朝フォント キリル文字一覧

表からもわかるように、これらのフォントでは初期キリル文字の多くが収録されている。実際にこれまでにデータ化されたロシア諸年代記において出現する初期キリル文字で、MS 明朝フォントに収録のない文字は、現在までのところ「Ѡ」「ѡ」「Ѣ」「ѣ」「Ѥ」の5つのみである⁵。

文字の収録状況から考えると、現在の標準的な Windows 環境においてキリル文字を表示するのに用いるフォントとして、これらの和文フォントがベストであることは確かである。

しかしながら、これらの和文フォントには大きな欠点がある。というのは、U+0410 ~ U+044F の文字の幅が、MS 明朝のプロポーショナルフォントであ

⁵ これ以外にも「Ѡ」「ѡ」「Ѣ」などがあるが本研究以後の課題とし、ここではふれない。

る MS P 明朝フォントにおいても，等幅になっているということである。次にあげる表 3 は，同じ文を Times New Roman フォントと MS P 明朝フォントで表示したものである。

Times New Roman	Се же соуть имена княземъ Киевськимъ.
MS P 明朝	С е ж е со у т ь и м е н а к н а з е м ь К и е в ь с к ы м .

表 3 Times New Roman フォントと MS P 明朝フォントでの表示の違い

この表からもわかるように，MS P 明朝フォントを使って現代ロシア文字を表示した場合，通常のプロポーショナルフォントに比べて可読性が非常に悪くなる。これは，Shift-Jis 系の MS 明朝フォント⁶ではキリル文字が全角文字として扱われており，この上位互換をとるために現在の Unicode 系の MS P 明朝フォントでも等幅に設定されていると考えられる。

このため，実際にコンコードンスを実際に印刷したり，html として表示を行う場合に，現代ロシア文字については MS P 明朝フォントを用いることはできず，これまでの研究ではこの部分には Times New Roman フォントを用いてきた。

3.4. Unicode 5.0

現在まで，まだその実装としてのフォントは出てきていないが，Unicode の最新バージョンは Unicode 5.0.0 となっており，これに対応するキリル文字の一覧が次の表 4 である。

⁶ Shift-Jis 文字エンコードスキームは MS-Windows Me 以前の 16bit 処理系で (MS-DOS, Windows95, 98, Me など) で採用されていたもの。これらの OS には Shift-Jis コードの MS 明朝フォントが付属しており，キリル文字については全角文字として扱われていた。

	040	041	042	043	044	045	046	047	048	049	04A	04B	04C	04D	04E	04F		050	051	052
0	Ё	А	Р	а	р	ё	Ѡ	ѡ	Ѣ	ѣ	Ѥ	ѥ	Ѧ	ѧ	Ѩ	ѩ	Ѫ	ѫ	Ѭ	ѭ
1	Ѯ	ѯ	Ѱ	ѱ	Ѳ	ѳ	Ѵ	ѵ	Ѷ	ѷ	Ѹ	ѹ	Ѻ	ѻ	Ѽ	ѽ	Ѿ	ѿ	ѿ	ѿ
2	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ
3	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ
4	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ
5	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ
6	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ
7	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ
8	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ
9	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ
A	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ
B	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ
C	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ
D	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ
E	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ
F	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ

表4 Unicode キリル文字領域 (The Unicode Standard 5.0 , Cyrillic より抜粋)⁷

この表に見るように、U+0400 ~ U+04FF について Unicode 5.0 と 1.1.5 の違いは U+0487 を残してすべての文字コード番号に文字が当てられている点にある。その中で、新たに初期キリル文字として追加になっているのが U+04CF の「I」の文字⁸である。

これ以外に、Unicode 5.0 では U+0500 ~ U+052F が「Cyrillic Supplement」として割り当てがある。ただし、これらに初期キリル文字に含まれるものはない。また、U+0514 ~ U+052F までは現在は割り当てがない状態である。

⁷ この表のみ、列ラベルが Unicode 番号上 3 桁、行ラベルが Unicode 番号下 1 桁となっている点に注意。

⁸ 上記表 4 では、通常の大文字の「I」となっているが、Unicode Database では、「04CF;Cyrillic small letter palochka」と表記されており、本来はスモールキャピタルで、表 4 のミスであると考えられる。

3.5. Izhitsa

一方、初期キリル文字フォントとしてよく用いられるのが Izhitsa⁹といわれる一群のフォントで、インターネットからいくつかのバージョンを入手することができる。これらは、現在 Windows 環境で初期キリル文字を扱うために用いられる、主要なフォントの一つとなっている。しかしながら、このフォントについてはその出所、および著作権など現在までのところ明確にできておらず、扱いに注意を要する。

現在、Izhitsa 系とも呼べるフォントは Izhitsa, IzhitsaCTT, Izhitza, Ukrainian?Izhitsa, IzhitsaC, YU C Izhitsa の 6 種類が確認される。その特徴を次の表 5 にまとめる。

書体名	FileSize	Version	Copyright 表記	文字数	文字形
Izhitsa	53KB	1.0	Generated by Fontographer 3.5	167	同じ？
Izhitsa CTT	50KB	1.000.000	Copyright 1991, 1994 ParaGraph. All rights reserved. Technical realization by TeamAXis Corp. Hinted with AXisType ToolKit: Copyright 1992, 1994 by Dmitry Komissarov.	168	
Izhitza	46KB	001.000	Copyright (C) ParaGraph JV	167	同じ？
Ukrainian?Izhitsa	48kB	001.000	Copyright (C) ParaGraph JV	162	
YU C Izhitsa	90KB	001.000	Copyright (C) ParaGraph JV	256	
Izhitsa C	44KB	001.000	Copyright (C) ParaGraph International	167	

表 5 Izhitsa フォントの特徴

この 6 つのフォントについてであるが、YU C Izhitsa を除いて文字のマッピングについてはほぼ共通しており、文字形やバージョン情報等から考えると、前者 2 つと後者 4 つにまとめることができる。この点から、これらのフォントの 2 群は ParaGraph 社により、異なる時期に作成されたものではないかと考えられる。次にあげる図 1 は、2 つの文字形の違いを比較したものである。

⁹ Izhitsa とはもともと初期キリル文字の「V」の文字名である。

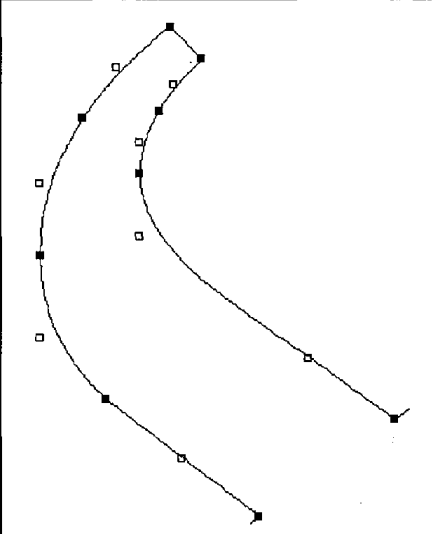
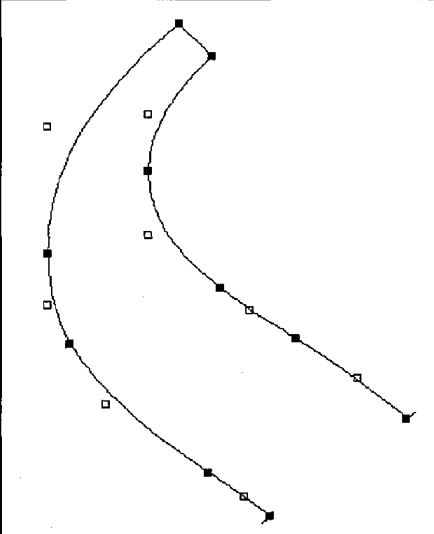
「8」左上部分拡大	
	
Izhitsa CTT	YU C Izhitsa

図1 Izhitsa CTT フォントと、YU C Izhitsa フォントの比較

一方、YU C Izhitsa のみが文字数が多くなっているが、これは他のフォントが ISO646 に適合した文字のマッピングを持っているのに対して、YU C Izhitsa のみ Unicode に不完全ながら、対応した文字のマッピングを持っているためである¹⁰。

次にあげる表は、Izhitsa フォントと YU C Izhitsa フォントに収録されている文字を一覧にしたものである。

¹⁰ この Unicode 対応版フォントが、ParaGraph 社による正式版かどうか現時点では確認できていない。

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
002		!	а	б					()	*		,		.	
003	о	1	2	3	4	5	6	7	8	9	:	;	Δ		Ѡ	?
004	ж	А	Б	Г	Д	Е	Ф		Н	И	Ψ	Ξ		М		О
005	Р	ΟΥ	Ы	С	Т	У	В	W	Х	У	Ζ	[]	Ѡ	ι
006		а		с	д	е	ф	г	х	и	ψ	ξ		ш		о
007	ρ	ου	ы	с		у	в	w	х	у	ζ	ξ	ы	ѓ	κ	
008		Г	,	г					Φ					К	ћ	
009			'			.	Δ	—						к	ћ	
00A				Ψ	Δ		:		ѐ							
00B		Ѡ	Ι	ι					ё				ψ	Σ	ς	ī
00C	А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П
00D	Р	Г	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я
00E	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п
00F	ρ	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я

表6 Izhitsa フォント収録文字一覧

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
002		!	а	б					()	*		,		.	
003	о	1	2	3	4	5	6	7	8	9	:	;	Δ		Ѡ	?
004	Ж	А	Б	Ц	Д	Е	Ф	Г	Х	И	Ј	К	Л	М	Н	О
005	П	Љ	Р	Г	Т	У	В	Њ	Ц	У	З	Ш	П	Ђ	Ч	ι
006	ж	а	б	ц	д	е	ф	г	х	и	ј	к	л	м	н	о
007	п	љ	ρ	с	т	у	в	њ	ц	у	з	ш	ђ	ѣ	ч	
00A					Δ		:		ѐ							
00B		Ѡ														
00D											V					
00E								!	а	б					()
00F	*		,	-	.		о	1	2	3	4	5	6	7	8	9
010	:	;	Δ		Ѡ	?	ж	А	Б	Г	Д	Е	Ф		Н	И
011	Ψ	Ξ		М		О	Р	ΟΥ	Ы	С	Т	У	В	W	Х	У
012	Ζ	[Ѡ]	Ѡ	ι		а		с	д	е	ф	г	х	и

013	Ѱ	ѱ		Ѳ		ѳ	Ѵ	ѵ	Ѷ	ѷ	Ѹ	ѹ			
014	Ѻ	ѻ	Ѽ	ѽ	Ѿ										
040		Ѣ		Ѥ		Ѧ	ѧ		Ѩ		Ѭ	ѭ			
041	Ѯ	ѯ	Ѱ	ѱ	Ѳ	ѳ	Ѵ	ѵ	Ѷ	ѷ	Ѹ	ѹ	Ѻ	ѻ	Ѽ
042	ѽ	Ѿ	ѿ	Ѱ	ѱ	Ѳ	ѳ	Ѵ	ѵ	Ѷ	ѷ	Ѹ	ѹ	Ѻ	ѻ
043	Ѽ	ѽ	Ѿ	ѿ	Ѱ	ѱ	Ѳ	ѳ	Ѵ	ѵ	Ѷ	ѷ	Ѹ	ѹ	Ѻ
044	ѿ	Ѱ	ѱ	Ѳ	ѳ	Ѵ	ѵ	Ѷ	ѷ	Ѹ	ѹ	Ѻ	ѻ	Ѽ	ѽ
045		Ѣ		Ѥ		Ѧ	ѧ	Ѩ	ѩ	Ѭ	ѭ				
201	-			—	Δ					,	,				
202			.												
203	Ѱ														

表7 YU C Izhitsa フォント文字一覧

この二つの表から YU C Izhitsa のみに収録された文字を抜き出したのが次の表 8 である。

文字コード	004A	0051	0057	0058	005C	005D	006A	0071	0077	0078	007C	007D
文字	Ј	Љ	Њ	Ћ	Ќ	Ќ	ј	љ	њ	ћ	ќ	ќ

表8 YU C Izhitsa フォントにのみ収録されている文字

3.6. これまでのロシア諸年代記コンコードダンスでの文字の扱い

これまでのロシア諸年代記コンコードダンスを作成するとき、初期キリル文字については、通常のラテン文字に置き換えてコンコードダンスへの変換処理を行い、最終段階において印刷用に Izhitsa などの初期キリル文字に置き換えていた。また、印刷時や html への出力時に用いるフォントは、前述のいくつかのフォントを組み合わせてもちいた。例えば、イパチー年代記コンコードダンスにおいては、現代ロシア文字については Times New Roman フォント、Unicode 1.1.5に含まれる初期キリル文字のほとんどを MS P 明朝、残りの文字を Izhitsa

フォント¹¹を用いている。次の表 9 は、イパチー年代記コンコードンスで実際に用いた文字とフォントとの対応についてまとめたものである。

フォント名	使用文字
Times New Roman	SIİsiïАБВГДЕЖЗИЙКЛМНОПРСТУФХЦЧШЩЪЫЬЭЮЯ абвгдежзийклмнопрстуфхцчшщъыьэюя
MS P 明朝	ω esijbŵŵbкaлǵvoyō
Izhitsa	Иa88

表 9 イパチー年代記コンコードンスに用いたフォントの一覧

4. フォント作成の方針

以上に見てきたように、初期キリル文字を含む文献のデータ化においては、文字コードとして Unicode を基本としながらも、Unicode の指針(一つの文字コードには一つの文字)からはずれた方法により初期キリル文字を入力している。また、実際に印刷したり html データとして表示をする場合には、Times New Roman, MS P 明朝, Izhitsa の 3 つのフォントを混在して使用している。

このような状態から、1 つのフォントでほぼすべての初期キリル文字を網羅できるようにするため、本研究ではフォント作成の方針を以下のようにまとめる。

- ① 文字ピッチについて、すべての文字をプロポーショナルフォントし、印刷時、表示時の可読性の向上を目指す。
- ② 収録文字は Unicode 5.0 を基本とし、U+0400 ~ U+0513 に割り当てのあるすべての文字を収録することを最終的な目標とする。
- ③ Unicode に収録のない文字について解決するため、文字コードマッピング

¹¹ CD-R 版コンコードンスでは、前述のように「Иa88」のみのオリジナルのフォントを用いている。

グについて2種類の方針をとる。

- ④ 著作権表示を明確にし、再配布を可能にする。

4.1. 文字のマッピング

フォントを作成する時に問題となるのが、どの文字コードにどの文字を割り当てるか、という文字コードのマッピングである。本研究では、その使用目的に合わせて、2つの文字コードマッピングを検討している。1つめは、Unicodeの指針に合致した方法でマッピングするものである。Unicode 5.0 にしたがって文字をマッピングしていき、Unicode にないものについては外字領域にマッピングするものである。そして、2つめのマッピングでは、Izhitsa フォントで用いられてきた手法を取り込むことで、Izhitsa と Unicode 両方の折衷的なマッピングを用いている。後者については、これまでよく用いられてきた Izhitsa フォントに対して上位互換をとることを目的としている。

4.1.1. Unicode の指針に合致した方法

本来の Unicode の指針に合致した方法で、初期キリル文字に対応したフォントを作成する場合、次のような方針が考えられる。

- ① 現代のキリル文字とほぼ同型で Unicode に割り当てがある文字(U+0400 ~ U+045F)は、その文字コードに割り当てる。例：А~Я, а~я など
- ② 初期キリル文字として Unicode に割り当てがある文字(U+0460 ~ U+0486)は、その文字コードを割り当てる。例：Ѡ, ѡ, Ѣ, ѣ など
- ③ 現代のキリル文字とは異形であり、上記①と重複する文字は、Unicode 外字領域に割り当てる。例：Ѡ, ѡ, Ѣ, ѣ
- ④ 現代のキリル文字ではすでに消滅しており、Unicode に割り当てもない文字についても、Unicode 外字領域に割り当てる。例：Ѹ, ѹ

この方針に従って作成した文字コードのマッピングが次の表 10 である。

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
040 ~ 051	U+0400 ~ U+0513 については, Unicode 5.0 に従う															
F000	И	а	Ѹ	ѹ												

表 10 初期キリル文字のマッピング①

また、この文字コードマッピングの特徴は次のようにまとめられる。

- ① Unicode の指針に従っているため、Unicode 準拠としておけば、特に文字コードマッピングについての注釈が必要ない。
- ② Unicode からはずれる文字については、すべて外字文字領域にまとめられているため、外字領域をみれば特殊な文字の一覧することができる。
- ③ これまでよく用いられた Izhitsa フォントとの互換性はなく、これまでのデータを再利用するためには、何らかのツールが必要となる。
- ④ 外字領域に文字を割り当てているため、システム側に通常のフォントとの関連づけを設定する必要がある。
- ⑤ 外字領域に文字を割り当てているため、これまでの Izhitsa 等を用いた方法に比べて、文字の入力が煩雑となる。

4.1.2. Izhitsa Font との整合性を考慮した方法

本来であれば、Unicode の方針からははずれることとなるが、従来の Izhitsa フォントを用いたデータの再利用を考え、Izhitsa フォントと Unicode を合わせたマッピングを 2 つめとしている。

この文字コードマッピングの方針は、つぎのようにまとめることができる。

- ① 現代のキリル文字については、Unicode の指定通りのマッピングを行い、

複数の書体が考えられる場合には、現代の形を優先する（4.1.1.の方針に同じ）。

② 初期キリル文字として Unicode に割り当てがある文字は、その文字コードを割り当てる（同上）。

③ 現代のキリル文字とは異形であり、上記①と重複する文字で、Izhitsa フォントに存在する文字は、Izhitsa フォントに合わせて割り当てる。

例：Ѡ, Ѣ, Ѥ

④ 現代のキリル文字とは異形であり、上記①と重複する文字で、Izhitsa フォントにも存在しない文字は、新たに Izhitsa フォントに合わせて ASCII 領域に割り当てる。例：Ѧ

この方針に従って作成したのが、次の表 11 である(書体については変更予定)。

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
002		!	Ѡ	ѡ	Ѣ				()	*		,		.	
003	Ѧ	ѧ	Ѩ	ѩ	Ѫ	ѫ	Ѭ	ѭ	Ѯ	ѯ	:	;	Ѱ		ѱ	?
004	Ѳ	ѳ	Ѵ	ѵ	Ѷ	ѷ	Ѹ		Ѻ	ѻ	Ѽ	ѽ		ѿ		ѿ
005	Ѱ	ѱ	Ѳ	ѳ	Ѵ	ѵ	Ѷ	ѷ	Ѹ	ѹ	Ѻ	[]	Ѽ	ѽ
006		ѿ		ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ		ѿ		ѿ
007	ѿ	ѿ	ѿ	ѿ		ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ	ѿ
040 ~ 051	U+0400 ~ U+0513 については、Unicode 5.0 に従う															

表 11 初期キリル文字のマッピング②

また、この文字コードマッピングの特徴は次のようにまとめられる。

① Unicode の指針に従っていない部分があるため、文字コードの割り当て

について付属マニュアル等に文字コードマッピングについての注釈が必要となる。

- ② Unicode からはずれる文字については, Izhitsa フォントなどを参考に割り当てているため, これを一覧する方法がない。
- ③ これまでよく用いられた Izhitsa フォントとの上位互換性をとっているため, これまでのデータを再利用することが容易である。
- ④ 通常のラテン文字領域に初期キリル文字を割り当てているため, 入力は容易となるが, その入力時の文字割り当て一覧表などが必要となる。

4.2. フォントフェイス

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, 基本的にセリフ体(Century フォントや TimesNewRoman に用いられている, Roman 体といわれるもの)に統一して作成することとする。

セリフ体	Се же соуть имена княземъ Кіевьскым.
Izhitsa	се же соуть имена княземъ кивьскым.

表 12 セリフ体と初期キリル文字の書体との比較

初期キリル文字は, 表 12 のように, 本来は Izhitsa の小文字¹²と同じ書体で表記するべきとも考えられるが, 現在よく用いられるセリフ体に比べると可読性が低いため, セリフ体で作成する。ただし, やはり初期キリル文字の書体による表記も可能なように, 全く同じ文字コードマッピングをもつ, 別書体のフォントを作成する予定でいる(5.3.参照)。

また, 今回作成する書体のデザインは, イパチー年代記の書体や, 他のフォント等をを参考にしつつ, オリジナルのデザインとして作成する。

¹² 本来の初期キリル文字には大文字・小文字の区別はない。

4.3. フォントの著作権

今回のフォント作成の理由の1つであるフォントの著作権問題についてであるが、本研究の成果となるフォントについてはフリーソフトとして配布する予定である。そのときのライセンスとして、GNU GPL(一般公衆利用許諾契約書)を用いる予定である。このライセンスは、オープンソースソフトを公開するために作成されたものであるが、現在は上記以外のソフトおよび著作物にも適用可能となっている。

GNU GPL の大きな特徴としては以下のようにまとめられる。

- ✓ 著作物の使用目的を限定しないこと
- ✓ 著作物を複製して再配布ができること
- ✓ 著作物を改良したり、改良したものを再配布できること

このライセンスに従うことで、本研究での成果となるフォントについては、著作権は筆者に属するものとするが、自由にフォントの再配布や改良を行うことができることとなる。また、フォントファイルの著作権表示部分、および添付マニュアルにこのことを明記することで、再利用、再配布を促進できればと考えている。

5. フォントの製作

5.1. Preslav.ttf

実際に作成する 4.1.1.の Unicode の指針に則った文字コードマッピングのフォントとしては、Preslav¹³フォントという名称にてこれを作成している。

現在までのところ、ロシア諸年代記に用いられている文字を優先したため、U+0400 ~ U+0513 の間は、初期キリル文字を除く文字のほとんどが未収録となっており、作成途中である。本校執筆時点では、これを Version 0.1 としており、すべての文字がそろった時点で Version 1.0 として正式に発表する予定

¹³ Preslav(プレスラフ)という名称は、ブルガリアにある都市名で、ここの城市遺跡から初期キリル文字で表記されたブルガリア皇帝に仕えた高官の墓碑銘が発見されている(言語学大事典 世界文字辞典)。

である。ただし、ロシア諸年代記のコンコードانس作成に必要な文字はそろっているともいえ、次期コンコードانسの作成にはこれを利用する予定である。

次の表 13 は、これまでに作成した Preslav フォントの文字をまとめたものである。

	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	:	;				
005												[\]		
007												{		}		
040					€		ı	İ								
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					€	ѕ	ı	İ								
046	Ѡ	ѡ	Ѣ	ѣ	Ѧ	ѧ	Ѩ	ѩ	Ѭ	ѭ	Ѯ	ѯ	Ѱ	ѱ	Ѳ	ѳ
048			Ѵ	ѵ	Ѷ											
04C	I															I
205							∴		∴							
F000	Ѡ	ѡ	Ѣ	ѣ												

表 13 Preslav フォント収録文字一覧

5.2. Preslav_iz.ttf

上記 Preslav フォントを、4.1.2.で述べた方針に則って改変したものが Preslav_iz フォントである。文字コードのマッピングで、Izhitsa との上位互換性をとっていることを、名称の後ろに_iz とつけることで表している。

こちらのフォントについても文字の作成状況は Preslav フォントとほぼ同様である。これは、文字のマッピングを入れ替えるだけのためである。

次の表 14 は Preslav_iz フォントの文字をまとめたものである。

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
002		!	Ѧ		Ѧ			'	()			,	-	.	/
003	0	1	2	3	4	5	6	7	8	9	:	;				
005						8						[]		
006	"															
007						8										
00A						\						{				
00B												}				
040					Є		I	İ								
041	A	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	a	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					є	s	i	ı								
046	W	w	Ѣ	ѣ	Ї	ї	А	А	І	І	Љ	Љ	Њ	Њ	Ќ	ќ
048			Ѧ	Ѧ	Ѧ											
04C	I															I
205							∴		∴							

表 14 Preslav_iz フォント収録文字一覧

5.3. Dobrogea.ttf

この Dobrogea¹⁴フォントは、上記の二つのフォントと、文字コードマッピングの互換性をとった、初期キリル文字フォントであり、初期キリル文字の書体での再現するために作成予定のものである(本論執筆時点で未着手)。初期キリル文字の書体で Unicode に対応したフォントとしては YU C Izhitsa もあるが、Dobrogea フォントのが従来の Izhitsa フォントとの互換性は高いと考えられ、

¹⁴ Dobrogea (ドブルジャ)とは、ルーマニアドナウ川下流右岸低地の地方名で、1950 年にブルガリア大守ディミタルの事績を初期キリル文字で記した石片が出土した場所である。(言語学大事典 世界文字辞典)

また再配布などの著作権についても明確にするため、十分に作成する意義があると考えられる。

また、このフォントも Preslav フォントと同様に 4.1.1.と 4.1.2.のそれぞれの方針に従った2つのフォントを、Dobrogea, Dobrogea_iz として発表する予定である。

次の表 15 は、現在作成中の Dorrogea_iz フォントのテスト版(Version 0.0.9)に収録する文字をまとめたものである。ただし、文字の書体データは YU C Izhitsa のものを使用している。

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
002		!	а	Ѡ	Ѣ				()	*		,		.	
003	о	1	2	3	4	5	6	7	8	9	:	;	Δ		Ѧ	?
004	Ѧ	А	Б	Г	Д	Е	Ѧ		Н	І	Ψ	Ѓ		М		О
005	Р	ОУ	Ы	Ѕ	Т	Ў	Ѧ	Ѧ	Х	У	Ѕ	І		Ј	Ѧ	Љ
006		а		с	д	е	Ѧ	Ѧ	Ѧ	і	Ѧ	Ѧ		Ѧ		о
007	р	оу	ы	ѕ		Ѧ	Ѧ	Ѧ	х	у	ѕ	Ѧ	Ѧ	Ѧ	Ѧ	
040		Ѧ		Г		Ѕ	І		Ψ			Ѧ	К			
041	А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П
042	Р	Г	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		Ѧ		г		ѕ	і	ї	Ѧ			Ѧ	к			
046 ~ 051	U+0460 ~ U+0513 については未作成															

表 15 Dobrogea_iz フォント収録文字一覧

6. 考察

6.1. 本研究の成果と問題点

今回の研究において明らかにしたように、初期キリル文字を含む文献をコンピュータに取り込んで処理を行うことは、3つのフォントを混在させざるを得ない点など、文字・フォントの点にいくつか困難な点があったといえる。しかし、Windows NT以降のUnicodeを文字コードとして採用したOSの普及により、以前に比較して状況が改善されてきていることも確かである¹⁵。

このような中で、本研究では現在の状況を各フォントの文字収録状況や著作権に関して、問題点を明らかにすることができたと考えている。また、これまでのIzhitsaフォントなどに頼った状況から、完全にUnicodeの指針には合致していない過渡的なものではあるが、Preslavフォントとしてこれまでのいくつかの問題点を解決したフォントを作成できたと考えている。具体的には、

- ① Izhitsaでは明確ではなかった著作権の問題を、オリジナルフォントを作成することで解決できたこと。
- ② 文字の収録状況・等幅フォントの問題から1つのコンコードダンスに3つのフォントを混在して使用しなければならなかった点が改善されたこと。
- ③ Izhitsaフォントと互換性をとることで、これまでに作成したデータの再利用を可能にできたこと

などが考えられる。

一方で、いくつかの問題が残っており、

- ① Preslav_izフォントでは、Izhitsaフォントの上位互換性をとるためにUnicodeの指針からははずれるものとなっていること。
- ② 収録文字が現時点では限られており、用途が限られること

などがあげられるであろう。

¹⁵ MS-DOSからWindows 3.1, Windows NT 3.5の時期における初期キリル文字の状況については、酒井 1996 に詳しい。

6.2. 初期キリル文字フォントを用いた、文字列処理の問題

この研究に関連して、これらのフォントを用いる上で必ず問題となるのが、これらのフォントを用いたデータの文字列処理である。

その中でも特に問題になるのが、データの並べ替えについてである。通常のラテン文字=いわゆるアルファベットや、現代ロシア文字については多くのプログラミング言語で大文字小文字間の変換は、専用の文字列処理関数が用意されており、文字の並べ替えについても大文字小文字を無視したり、活かしたりすることが可能である。しかしながら、今回作成したフォントや、Izhitsa フォントなどを用いたデータの場合、プログラミング言語に最初から用意されている並べ替え関数では問題が発生する。それは、現代ロシア文字以外についての大文字小文字の対応情報を持っていないことや、現代ロシア文字と初期キリル文字との間での対応関係、また並べ替え順序についての情報を持っていないことによる。

このため、これらのフォントを用いたデータを処理するためには、専用のソートプログラムが必要になることがわかっており、現在これについても検討中である(できればDLLのような形でフォントと同時に GNU GPL に則って配布ができればと考えている)。

7. 展望

考察のところで述べたように、これまでの初期キリル文字に関する問題点をほぼ解決できたとはいえ、まだいくつかの問題が残っている。

まず_iz 系フォントが Unicode の指針からははずれるものとなっていることについてであるが、このフォントは Unicode 準拠のフォントが普及するまでの過渡的な役割として作成しているものである。これは、過去の初期キリル文字を含むデータを Unicode 対応データに変換する文字列処理プログラムが完成した時点でその役割を Preslav 等の Unicode 対応フォントに譲る予定である。ただし、ラテン文字として初期キリル文字を入力できるという点では優れてお

り、Preslav_iz から Preslav への変換プログラムを組み合わせて使うことで、入力専用フォントとしての利用価値を検討する必要があると考えられる。

一方で Unicode の外字領域に文字を収録したフォントについては、本来は同じ文字の異体字である¹⁶ものも多く、これは OpenType とよばれるフォントの規格を用いることで異体字として一つの文字コードに割り当てることが可能であるが、現在までのところ一般的なアプリケーションソフトでは利用ができない¹⁷ため、今後 Open Type の普及状況をみながら対応を検討していく予定である。

各フォントの収録文字数についてであるが、現在までに Preslav フォントおよび Preslav_iz フォントの主要部分については製作を終えているが、Unicode 5.0 に収録されている文字にはまだまだ足りない文字が多い。このため、まずは初期キリル文字から、その後ロシア語以外の言語で用いられる文字へと広げて、順次追加していく予定である。

Dobrogea フォントについては、Izhitsa フォントを用いたテスト版ができあがってはいるが、このままでは著作権に問題があるため、公開することはできない状態である。

また、これらの完成したフォントについては <http://linguistics.factory-js.com/> 上で発表していく事を予定している。

そして、6.2.で述べているように、これらのフォントを用いたデータ用の処理プログラムを作成する必要がある、次なるコンコーダンス作成のために、現在計画中である。

¹⁶ 「Ѧ」「Ѧ」は「Я」「я」の異体字、「Ѧ」「Ѧ」は「0y」「0y」の異体字である。

¹⁷ 現在 OpenType の異体字切り替えに対応しているのは、Adobe InDesign, Quark XPress, Canon System Solutions DICOLORE などいわゆる DTP アプリケーションソフトに限定されている。

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ド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
- 中條直樹編『『ノヴゴロド第一年代記』(シノド本) コンコードダンス』日本古代ロシア研究会, 1998
- 中條直樹編『『ロシア原初年代記』(ラヴレンチー本) コンコードダンス』日本古代ロシア研究会, 1998
- 中條直樹編『『モスクワ年代記集成 (XV 世紀末)』語彙集』日本古代ロシア研究会, 2003
- 中條直樹編『『モスクワ年代記集成 (XV 世紀末)』コンコードダンス』日本古代ロシア研究会, 2004
- 中條直樹, 酒井純『『ノヴゴロド第 4 年代記』コンコードダンス』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』コンコードダンス』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
- 西田龍雄『講座言語第 5 巻 世界の文字』大修館書店, 1981
- 深沢千尋『文字コード超研究』ラトルズ, 2003

参考サイト

- GNU GPL, <http://www.gnu.org/licenses/gpl.ja.html>, (2007/01/08)
- Microsoft Typography, <http://www.microsoft.com/typography/>, (2007/01/08)
- Nox Insomnie, <http://www.ceres.dti.ne.jp/~i-yasuda/>, (2007/01/07)
- ParaType (Izhitsa フォント販売元), <http://www.paratype.com/>, (2007/01/08)
- Unicode, <http://www.unicode.org/>, (2007/01/08)