

## ソグド、マニ、ウイグル(SMU)文字コードの実装と開発する際の

## 難しさと解決法

## Difficulties and Solutions of Sogd, Manichaean, Uyghur (SMU) Character-code implementation and developing

ウメルジアン ウスマン†  
Omarjan Osman木村 哲也†  
Tetsuya Kimura山田 耕一†  
Yamada Koichi猿渡 隼斗†  
Hayato Saruwatari

## 概要

SMU 文字は表音文字“Phonogram”で構成されている文字である。中央アジアに住んでいるソグド人、マニ人、ウイグル人が使っていた文字はソグド、マニ、ウイグル“Sogd, Manichaean, Uyghur, (SMU)”文字と呼ばれ、現在のそれらとは若干異なっている。SMU 文字は古代と中期に使われていた文字で、現在使われていないという状況にある。SMU 文字で書かれた文献作品は、現在各国の博物館と大学に所蔵されている。古代と中期のチュルク語とアルタイ諸語(SMU 語)とそれらの文字の研究者の研究を助けるために、古代と中期 SMU 文献作品と様々な写本資料をデジタルテキストとしてデジタルアーカイビング“Digital archiving”が必要である。SMU 単語は母音、子音、分音符号の組み合わせで形成したグリフと複雑なりガチャーから構成されている。

本研究では、紀元前 2 世紀から紀元後 11 世紀までに右から左への横書きで書かれている古代と中期の SMU 写本資料(図 4 参照)を使用して分音符号、グリフと複雑なりガチャーの位置の設計を行った。SMU 文字コードの実装と開発の際の難しさは分音符号、グリフと複雑なりガチャーの形成である。本稿では、SMU 語のデジタルテキスト処理の基礎である分音符号、グリフ、複雑なりガチャー形成の設計を提示する。

**キーワード:** SMU、文字コード、グリフ、リガチャー

## 1 はじめに

古代と中期の SMU 文字を書式的に見ると、ソグドとウイグル文字の右から左の横書きと上から下の縦書きと二つの書き方がある。マニ文字の縦書きは存在しない。SMU 文書単語の構造は、分音符号、グリフ、複雑なりガチャーの構成から成っている。本稿では、SMU 文字の横書きの分音符号、グリフ、複雑なりガチャー形成の分析を行った。以下には分音符号、グリフ、複雑なりガチャーの定義を示す。

(1) **分音符号:** 補助記号、区別的発音符号は分音符号“Diacritics Marks”と呼ばれる。SMU 分音符号を上部と下部の分音符号と 2 つのグループに分ける事ができる。上部分音符号を  $\bar{D}$ 、下部分音符号を  $\underline{D}$ 、上部分音符号の組み合わせを  $\overline{D}$ 、下部分音符号の組み合わせを  $\underline{D}$ 、母音文字を  $V$ (Vowel)、子音文字を  $C$ (Consonant) で表現する。

母音、子音の分音符号と組み合わせを  $V.\bar{D}$ 、 $V.\underline{D}$ 、 $V.\overline{D}$ 、

$V.\underline{D}$ 、 $C.\bar{D}$ 、 $C.\underline{D}$ 、 $C.\overline{D}$ 、 $C.\underline{D}$  で表現する。

(2) **グリフ:** 符号、記号の抽象的な考えであり、図形的设计と表示ができる記号はグリフ“Glyph”と呼ばれる。

母音と子音文字コードを実装する際にグリフを応用する。1 つ文字のコードが 4 つ以上のグリフを応用する事ができる。SMU グリフの語頭形、語中形、語末形、独立形と 4 種類の表示形がある。母音と子音のグリフは、 $G$  で表現して、グリフの 4 種類の表示形は、語頭形を  $\underline{G}_l$ 、語中形を  $\underline{G}_m$ 、語末形を  $\underline{G}_r$ 、独立形を  $G_n$  で表現する。グリフの上部分音符号  $\bar{D}$  と組み合わせを  $\overline{G}$ 、下部分音符号  $\underline{D}$  と組み合わせを  $\underline{G}$  で表現して、上部分音符号の組み合わせ  $\overline{D}$  とグリフの組み合わせを  $\overline{G}$ 、下部分音符号の組み合わせ  $\underline{D}$  とグリフの組み合わせを  $\underline{G}$  で表現する。グリフの生成を以下のように定義する。

$$V, C = (G)_{V,C}$$

$$(V.\bar{D}, C.\bar{D}) = (\overline{G})_{V,C} \quad (V.\underline{D}, C.\underline{D}) = (\underline{G})_{V,C}$$

$$(V.\overline{D}, C.\overline{D}) = (\overline{\overline{G}})_{V,C} \quad (V.\underline{\underline{D}}, C.\underline{\underline{D}}) = (\underline{\underline{G}})_{V,C}$$

(3) **リガチャー:** グリフの抽象的な考えであり、文字の組み合わせと相互作用から合成して、設計することができる図形記号集合はリガチャー“Ligature”と呼ばれる。リガチャーとは、 $(G)_{V,C}$  と  $(G)_{V,C}$  の結合によって生成した合字である。SMU 合字の形成規則には、 $V$  と  $C$  の組み合わせを  $VC$ 、 $VCC$ 、 $VCV$ 、 $VVCC$ 、 $VCCV$  で表現して、 $C$  と  $V$  の組み合わせを  $CV$ 、 $CVV$ 、 $CVC$ 、 $CCVV$ 、 $CVVC$  で表現する。SMU リガチャーの語頭形、語中形、語末形、独立形と 4 種類の表示形がある。語頭形を  $\underline{L}_l$ 、語中形を  $\underline{L}_m$ 、語末形を  $\underline{L}_r$ 、独立形を  $L_n$  で表現する。リガチャーのグリフと組み合わせを以下のように定義する。グリフ  $(G)_{(V,C)l,m,r,n}$  を  $Y$ 、不明グリフを  $X$  から表現する。

$$Y + (X)_{l,m,r,n}, (X)_{l,m,r,n} + Y = (L)_{l,m,r,n}$$

$$\overline{Y} + (X)_{l,m,r,n}, (X)_{l,m,r,n} + \overline{Y} = (\overline{L})_{l,m,r,n}$$

$$\underline{Y} + (X)_{l,m,r,n}, (X)_{l,m,r,n} + \underline{Y} = (\underline{L})_{l,m,r,n}$$

† 長岡技術科学大学  
新潟県長岡市上富岡町 1603-1

$$\overline{Y} + (X)_{l,m,r,n}, (X)_{l,m,r,n} + \overline{Y} = (\overline{L})_{l,m,r,n}$$

$$\underline{Y} + (X)_{l,m,r,n}, (X)_{l,m,r,n} + \underline{Y} = (\underline{L})_{l,m,r,n}$$

SMU 単語の構造は簡単なリガチャーと複雑なリガチャーから構成している。グリフとグリフの組み合わせで形成したリガチャーは、簡単なリガチャーと呼ばれる。複雑なリガチャーは、リガチャー同士または、リガチャーとグリフの組み合わせで、形成される。リガチャーは、必ず意味のある単語になる。単語を  $W(Word)$  で表現する。リガチャーの組み合わせを  $\overline{L} + \underline{L} = (\overline{L})_{l,m,r,n}$  で表現する。

SMU 文字コードの実装する際には、分音符号、グリフ、リガチャーの問題が存在しているが、問題の源泉を以下のように提示する。

上部分音符号  $\overline{D}$ , 下部分音符号  $\underline{D}$  の実装の問題。

上部分音符号の組み合わせ  $\overline{D}$  のグリフと組み合わせ、下部分音符号の組み合わせ  $\underline{D}$  のグリフと組み合わせの実装の問題。

上部分音符号の組み合わせ  $\overline{D}$  のリガチャーと結合、下部分音符号の組み合わせ  $\underline{D}$  のリガチャーと結合の実装の問題。

分音符号の組み合わせ  $\overline{D}$  と  $\underline{D}$  の実装と複雑なリガチャーの組み合わせ  $(\overline{L})_{l,m,r,n}$  の実装は、右から左の横書きで書かれたアラビア文字“Arabic Script”、シリア文字“Syriac Script”では存在しない。

## 2 SMU 分音符号、グリフとリガチャー形成の分析とグリフの設計

### 2.1 SMU 分音符号形成の分析

SMU 分音符号形成の分析を 2 つのグループの位置決めを行った。

第1グループは(上部 “Top”  $\overset{\circ}{\circ}, \overset{\circ}{\circ}, \overset{\circ}{\circ}, \overset{\circ}{\circ}, \overset{\circ}{\circ}, \overset{\circ}{\circ}, \overset{\circ}{\circ}, \overset{\circ}{\circ}, \overset{\circ}{\circ}, \overset{\circ}{\circ}, \overset{\circ}{\circ}$ )である。

第2グループは(下部 “Bottom”  $\underset{\circ}{\circ}, \underset{\circ}{\circ}, \underset{\circ}{\circ}, \underset{\circ}{\circ}, \underset{\circ}{\circ}, \underset{\circ}{\circ}, \underset{\circ}{\circ}, \underset{\circ}{\circ}, \underset{\circ}{\circ}, \underset{\circ}{\circ}, \underset{\circ}{\circ}$ )である。

分音符号形成の上部グループの中には“( $\overset{\circ}{\circ} + \overset{\circ}{\circ}$ ) =  $\overset{\circ}{\circ}$  ta”, “( $\overset{\circ}{\circ} + \overset{\circ}{\circ}$ ) =  $\overset{\circ}{\circ}$  gh”, “( $\overset{\circ}{\circ} + \overset{\circ}{\circ}$ ) =  $\overset{\circ}{\circ}$  ha”, “( $\overset{\circ}{\circ} + \overset{\circ}{\circ}$ ) =  $\overset{\circ}{\circ}$  za”, 下部グループの中には“( $\underset{\circ}{\circ} + \underset{\circ}{\circ}$ ) =  $\underset{\circ}{\circ}$  gha”, “( $\underset{\circ}{\circ} + \underset{\circ}{\circ}$ ) =  $\underset{\circ}{\circ}$  h” のコードを実装する際のグリフの形成は難しい。グリフの形成を誤ると、単語の意味が変わる。アラビア文字、シリア文字にはこの6つの分音符号の組み合わせは存在しない。難しい原因は分音符号の上部の組み合わせ  $\overline{D}$  の実装と下部の組み合わせ  $\underline{D}$  の実装であり、 $\overline{D}$  と  $\underline{D}$  はグリフ、リガチャーと結合して実装する。 $\overline{D}$  と  $\underline{D}$  の分析を以下のように示す。

$$(\overline{D} + \overline{D}) = \overline{\overline{D}}, (\underline{D} + \underline{D}) = \underline{\underline{D}},$$

例:  $\overset{\circ}{\circ} + \overset{\circ}{\circ} = gh \overset{\circ}{\circ}, \underset{\circ}{\circ} + \underset{\circ}{\circ} = h \underset{\circ}{\circ}$

### 2.2 SMU グリフ形成の分析

SMU グリフを語頭形、語中形、語末形、独立形と4種類に分けて分析を行った。SMU の1つ文字の  $G_l, G_m, G_r, G_n$  と4つの異なるグリフがある。4つの異なるグリフを以下のように示す。(表1参照)

表1. SMU グリフタイプ

Name	$G_n$	$G_r$	$G_m$	$G_l$	実装
S	س	س	س	س	س
M	م	م	م	م	م
U	و	و	و	و	و

語頭形( $X_l$  “Left-joining glyph form”) =  $G_l$

語中形( $X_m$  “Dual-joining glyph form”) =  $G_m$

語末形( $X_r$  “Right-joining glyph form”) =  $G_r$

独立形( $X_n$  “Nominal glyph form”) =  $G_n$

母音と子音の  $G_l, G_m, G_r, G_n$  と4つの異なるグリフを実装して、グリフ形成の難しさの分析を行った。以下のようにグリフと分音符号の組み合わせ  $\overline{D}$  と  $\underline{D}$  を実装した際にはグリフ  $(\overline{G})_{(V,C)l,m,r,n}$  と  $(\underline{G})_{(V,C)l,m,r,n}$  を実装する。例: 文字“Dah”を入力した際にはグリフの“ $D G_l$ ”, “ $a G_m$ ”, “ $h G_r$ ”を実装すると、“ $\overset{\circ}{\circ}$  dag”という単語になる。

$$G_l + G_m + (G_r + \overline{D}) = G_l + G_m + \overline{G_r}, \overset{\circ}{\circ} \text{ dag}”$$

$$G_l + G_m + (G_m + \underline{D}) = G_l + G_m + \underline{G_m}, \underset{\circ}{\circ} \text{ doh}”$$

$$(G_l + \overline{D}) + G_m + G_r = \overline{G_l} + G_m + G_r, \overset{\circ}{\circ} \text{ gat}”$$

$$(G_l + \underline{D}) + G_m + G_r = \underline{G_r} + G_m + G_r, \underset{\circ}{\circ} \text{ hat}”$$

グリフ  $(\overline{G})_{l,m,r,l}$  と  $(\underline{G})_{l,m,r,l}$  の生成は他の言語では存在しない。グリフ実装の結果、 $(\overline{G})_{l,m,r,l}$  と  $(\underline{G})_{l,m,r,l}$  の実装の難しさはコードを実装した際に2つの分音符号と  $(X)_{l,m,r,n}$  の組み合わせになることである。(図1参照)

### 2.3 SMU リガチャー形成の分析

SMU リガチャーにも語頭形、語中形、語末形、独立形と4種類に分けて分析を行った。SMU リガチャー形成の  $L_l, L_m, L_r, L_n$  と4つの異なるリガチャーがある。4つの異なるリガチャーを以下のように示す。(表2参照)

表2. SMU リガチャータイプ

Name	$L_n$	$L_r$	$L_m$	$L_l$	G	実装
S	س	س	س	س	س	س

M	آ	أ	ع	ع	ب	ق
U	ع	ع	ع	ع	ع	ع

語頭形  $(X.Y)_l$  “Left-joining ligature” =  $L_l$

語中形  $(X.Y)_m$  “Dual-joining ligature” =  $L_m$

語末形  $(X.Y)_r$  “Right-joining ligature” =  $L_r$

独立形  $(X.Y)_n$  “Nominal ligature” =  $L_n$

リガチャーの  $L_n, L_r, L_m, L_l$  と 4 つの異なるリガチャーを実装して、リガチャーの分析を行った。右から左の横書きで書かれた他の言語でもリガチャーがある。右から左の横書きの他の言語とソグド、マニリガチャーの形成方法は同じである。それらのリガチャーの形成は以下に示す。

$$X_r.Y_l, Y_r.X_l, X_r.Y_m.X_l, Y_r.X_m.Y_l$$

$$X_r.Y_l, X_r.Y_l, X_r.Y_m.X_l, X_r.Y_m.X_l$$

$$Y_r.X_l, Y_r.X_l, Y_r.X_m.Y_l, Y_r.X_m.Y_l$$

ウイグルリガチャーはそれらとは異なっている。異なっているリガチャーの形成は以下に示す。

$$X_r.Y_l, X_r.Y_m.X_l, X_r.Y_m.X_l$$

$$Y_r.X_l, Y_r.X_m.Y_l, Y_r.X_m.Y_l$$

ウイグルリガチャーの組み合わせの形成は以下に示す。

$$X_r.Y_l, Y_r.X_l, X_r.Y_m.X_l, Y_r.X_m.Y_l$$

ウイグルリガチャーの形成は他の言語では存在しない。

リガチャーの実装の結果、 $(\bar{L})_{l,m,r,l}$ 、 $(\underline{L})_{l,m,r,l}$  と  $(\bar{L})_{l,m,r,n}$  の難しさはコードの実装した際に分音符号の組み合わせになることである。

### 2.4 SMU グリフの設計

SMU グリフ実装の  $G_l, G_m, G_r, G_n$  と 4 つの実装式がある。ソグド、マニ、アラビア文字、シリア文字の 4 つにはそれぞれ 1 つのグリフ実装式がある。ウイグル文字のグリフの実装式は異なり、1 つの文字に対して 4 つ以上のグリフ実装式がある。例：ウイグルグリフ “A” の 1)  $X_l^A$ , 2)  $X_l^A$ , 3)  $X_m^A$ , 4)  $X_m^A$ , 5)  $X_r^A$ , 6)  $X_r^A$ , 7)  $X_r^A$ , 8)  $X_n^A$  と実装式がある。

SMU グリフの設計では、古代写本を参考にして、文字の図形記号様式を写本と同じにした。設計した際は、Adobe® Photoshop® 7.0 と FontCreator 5.6 を使ってグリフの設計を行った。SMU グリフの設計は以下に行った。

**ソグドグリフ**：ソグドの 40 個の母音と 78 個の子音グリフがある。以下のように設計した。

ソグド 40 個の母音グリフは 1(“A”  $X_l^A, X_m^A, X_r^A, X_n^A$ ), 2(“Ah”  $X_l^A, X_m^A, X_r^A, X_n^A$ ), 3(“E”  $X_l^E, X_m^E, X_r^E, X_n^E$ ), 4(“Ey”  $X_l^E, X_m^E, X_r^E, X_n^E$ ), 5(“I”  $X_l^I, X_m^I, X_r^I, X_n^I$ ), 6(“Iy”  $X_l^I, X_m^I, X_r^I, X_n^I$ ), 7(“O”  $X_l^O, X_m^O, X_r^O, X_n^O$ ), 8(“Ov”  $X_l^O, X_m^O, X_r^O, X_n^O$ ), 9(“U”  $X_l^U, X_m^U, X_r^U, X_n^U$ ), 10(“Uv”  $X_l^U, X_m^U, X_r^U, X_n^U$ ) である。

ソグド 78 個の子音グリフは 11(“Beth”  $X_l^B, X_r^B, X_n^B$ ), 12(“Pe”  $X_l^P, X_n^P$ ), 13(“Teth”  $X_l^T, X_m^T, X_r^T, X_n^T$ ), 14(“Zhe”  $X_l^Z, X_m^Z, X_r^Z, X_n^Z$ ), 15(“Che”  $X_l^C, X_m^C, X_r^C, X_n^C$ ), 16(“Cheth”  $X_l^C, X_m^C, X_r^C, X_n^C$ ), 17(“D”  $X_l^D, X_n^D$ ), 18(“R”  $X_l^R, X_m^R, X_r^R, X_n^R$ ), 19(“Z”  $X_l^Z, X_m^Z, X_n^Z$ ), 20(“Zhee”  $X_l^Z, X_m^Z, X_r^Z, X_n^Z$ ), 21(“S”  $X_l^S, X_m^S, X_r^S, X_n^S$ ), 22(“Shi”  $X_l^S, X_m^S, X_r^S, X_n^S$ ), 23(“Ghe”  $X_l^G, X_r^G$ ), 24(“F”  $X_n^F$ ), 25(“Koph”  $X_l^K, X_m^K, X_r^K, X_n^K$ ), 26(“K”  $X_l^K, X_m^K, X_r^K, X_n^K$ ), 27(“G”  $X_l^G, X_m^G, X_r^G, X_n^G$ ), 28(“L”  $X_l^L, X_m^L, X_r^L, X_n^L$ ), 29(“M”  $X_l^M, X_m^M, X_r^M, X_n^M$ ), 30(“N”  $X_l^N, X_m^N, X_r^N, X_n^N$ ), 31(“Hee”  $X_n^H$ ), 32(“V”  $X_l^V, X_m^V, X_r^V, X_n^V$ ), 33(“Jod”  $X_l^J, X_m^J, X_r^J, X_n^J$ ) である。

**マニグリフ**：マニグリフの 8 個の母音と 30 個の子音グリフがある。以下のように設計した。

マニ 8 個の母音グリフは 1(“A”  $X_n^A$ ), 2(“Ah”  $X_n^A$ ), 3(“E”  $X_n^E$ ), 4(“I”  $X_n^I$ ), 5(“O”  $X_n^O$ ), 6(“Ov”  $X_n^O$ ), 7(“U”  $X_n^U$ ), 8(“Uv”  $X_n^U$ )。

マニ 30 個の子音グリフは 9(“Be”  $X_l^B, X_m^B, X_r^B, X_n^B$ ), 10(“Pe”  $X_n^P$ ), 11(“Te”  $X_n^T$ ), 12(“Zhe”  $X_n^Z$ ), 13(“Che”  $X_n^C$ ), 14(“He”  $X_n^H$ ), 15(“De”  $X_n^D$ ), 16(“Re”  $X_n^R$ ), 17(“Ze”  $X_n^Z$ ), 18(“Zhee”  $X_n^Z$ ), 19(“Si”  $X_n^S$ ), 20(“Shi”  $X_n^S$ ), 21(“Ghe”  $X_n^G$ ), 22(“Fi”  $X_n^F$ ), 23(“Khi”  $X_n^K$ ), 24(“Ke”  $X_n^K$ ), 25(“Ge”  $X_n^G$ ), 26(“Ngi”  $X_n^N$ ), 27(“Le”  $X_n^L$ ), 28(“Me”  $X_n^M$ ), 29(“Ne”  $X_n^N$ ), 30(“Hee”  $X_n^H$ ), 31(“Ve”  $X_n^V$ ), 32(“Ye”  $X_n^Y$ )

**ウイグルグリフ**：ウイグルグリフの 32 個の母音と 96 個の子音グリフがある。以下のように設計した。

ウイグル 32 個の母音グリフは 1(“A”  $X_l^A, X_m^A, X_r^A, X_n^A$ ), 2(“Ah”  $X_l^A, X_m^A, X_r^A, X_n^A$ ), 3(“E”  $X_l^E, X_m^E, X_r^E, X_n^E$ ), 4(“I”  $X_l^I, X_m^I, X_r^I, X_n^I$ ), 5(“O”  $X_l^O, X_m^O, X_r^O, X_n^O$ ), 6(“Ov”  $X_l^O, X_m^O, X_r^O, X_n^O$ ), 7(“U”  $X_l^U, X_m^U, X_r^U, X_n^U$ ), 8(“Uv”  $X_l^U, X_m^U, X_r^U, X_n^U$ )。

ウイグル 96 個の子音グリフは 9(“Be”  $X_l^B, X_m^B, X_r^B, X_n^B$ ), 10(“Pe”  $X_l^P, X_m^P, X_r^P, X_n^P$ ), 11(“Te”  $X_l^T, X_m^T, X_r^T, X_n^T$ ), 12(“Zhe”  $X_l^Z, X_m^Z, X_r^Z, X_n^Z$ ), 13(“Che”  $X_l^C, X_m^C, X_r^C, X_n^C$ ), 14(“He”  $X_l^H, X_m^H, X_r^H, X_n^H$ ), 15(“De”  $X_l^D, X_m^D, X_r^D, X_n^D$ ), 16(“Re”  $X_l^R, X_m^R, X_r^R, X_n^R$ ), 17(“Ze”  $X_l^Z, X_m^Z, X_r^Z, X_n^Z$ ), 18(“Zhee”  $X_l^Z, X_m^Z, X_r^Z, X_n^Z$ ), 19(“Si”  $X_l^S, X_m^S, X_r^S, X_n^S$ ), 20(“Shi”  $X_l^S, X_m^S, X_r^S, X_n^S$ ), 21(“Ghe”  $X_l^G, X_m^G, X_r^G, X_n^G$ ), 22(“Fi”  $X_l^F, X_m^F, X_r^F, X_n^F$ ), 23(“Khi”  $X_l^K, X_m^K, X_r^K, X_n^K$ ), 24(“Ke”  $X_l^K, X_m^K, X_r^K, X_n^K$ ), 25(“Ge”  $X_l^G, X_m^G, X_r^G, X_n^G$ ), 26(“Ngi”  $X_l^N, X_m^N, X_r^N, X_n^N$ ), 27(“Le”  $X_l^L, X_m^L, X_r^L, X_n^L$ ), 28(“Me”  $X_l^M, X_m^M, X_r^M, X_n^M$ ), 29(“Ne”  $X_l^N, X_m^N, X_r^N, X_n^N$ ), 30(“Hhe”  $X_l^H, X_m^H, X_r^H, X_n^H$ ), 31(“Ve”  $X_l^V, X_m^V, X_r^V, X_n^V$ ), 32(“Ye”  $X_l^Y, X_m^Y, X_r^Y, X_n^Y$ )

### 3 SMU 文字コードを実装する際の問題と解決手法

#### 3.1 SMU 文字コードを実装する際の問題

SMU 文字コードの実装する際の問題は以下のように三つある。

上部分音符号  $\bar{D}$  と下部分音符号  $\underline{D}$  の実装である。

分音符号  $\bar{D}$  と  $\underline{D}$  を実装する際には、 $\bar{D}$  と  $\underline{D}$  を直接的にリガチャーと結合している。(図1参照)

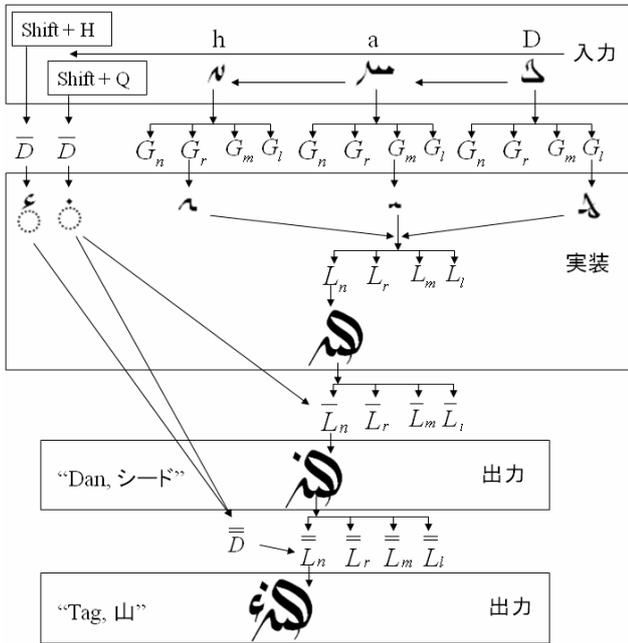


図1. SMU文字コードの実装式

分音符号と複雑なりガチャーの間にはグリフが設定されていない。この状況では  $X.(L)_{l,m,r,n}, (L)_{l,m,r,n}.X$  の位置が変位する際に、分音符号の位置が正しく入力できない。単語の意味も変わる事になる。(表3参照)

表3. SMU文字コードの実装の誤り

正式実装 (A)	誤実装 (B)	誤実装 (C)
$(\bar{L})_n$		

分音符号とリガチャーの間には、分音符号とグリフとを組み合わせる設計が必要である。

上部分音符号の組み合わせ  $\bar{D}$  と下部分音符号の組み合わせ  $\underline{D}$  の実装である。上部と下部分音符号の組み合わせ  $\bar{D}$  と  $\underline{D}$  を母音、子音との組み合わせが設定されていない。

$\bar{D}$  と  $\underline{D}$  の母音、子音と組み合わせの設定がされていない際には、単語の誤実装は(表3B)のようになる。

上部分音符号の組み合わせ  $\bar{D}$  の上部リガチャー  $(\bar{L})_{l,m,r,n}$  と結合、下部分音符号の組み合わせ  $\underline{D}$  の下部リガチャー  $(\underline{L})_{l,m,r,n}$  と結合の実装である。 $\bar{D}$ 、 $\underline{D}$  の

$(\bar{L})_{l,m,r,n}, (\underline{L})_{l,m,r,n}$  と結合は複雑なりガチャーの形成  $(\bar{L})_{l,m,r,n}$  になった。

#### 3.2 SMU 文字コードの問題の解決手法

分音符号の問題を解決するためには、分音符号とグリフの組み合わせを行って、グリフを設計をした。グリフの設計現状を以下に示す。

##### 3.2.1 分音符号の母音、子音との組み合わせの設計:

上部分音符号  $\bar{D}$  について母音と子音と組み合わせを行った。まず、 $\bar{D}$  の母音と組み合わせ  $\overline{DV}, \overline{DVV}, \overline{VD}, \overline{VVD}$  と  $\bar{D}$  の子音と組み合わせ  $\overline{DC}, \overline{DCC}, \overline{CD}, \overline{CCD}$  の設計と下部分音符号  $\underline{D}$  の母音と組み合わせ  $\underline{DV}, \underline{DVV}, \underline{VD}, \underline{VVD}$  と  $\underline{D}$  の子音と組み合わせ  $\underline{DC}, \underline{DCC}, \underline{CD}, \underline{CCD}$  の設計した。(表4参照)

表4. SMU分音符号の母音と子音との組み合わせ

分音符号	母音	グリフ
$\bar{D} \overset{\circ}{+}$	$\underline{G}_m \overset{\Delta}{=}$	$\overline{G}_m \overset{\Delta}{+}$
分音符号	子音	グリフ
$\underline{D} \overset{\circ}{+}$	$\underline{G}_l \overset{\wedge}{=}$	$\underline{G}_l \overset{\wedge}{+}$

##### 3.2.2 上部と下部分音符号の組み合わせの設計:

分音符号  $\bar{D}$  と  $\underline{D}$ 、 $\bar{D}$  と  $\underline{D}$  の組み合わせをして分音符号の組み合わせ  $\bar{D}$  と  $\underline{D}$  を設計した。その後、分音符号の組み合わせを行って、上部分音符号の組み合わせ  $\bar{D}$  と下部分音符号の組み合わせ  $\underline{D}$  を設計した。(表5参照)

表5. SMU分音符号とグリフの組み合わせ

分音符号の組み合わせ	子音	グリフ
$\bar{D} \overset{\circ}{+}$	$\underline{G}_l \overset{\vee}{=}$	$\overline{G}_l \overset{\vee}{+}$
分音符号の組み合わせ	母音	グリフ
$\underline{D} \overset{\circ}{+}$	$\underline{G}_l \overset{\wedge}{=}$	$\underline{G}_l \overset{\wedge}{+}$

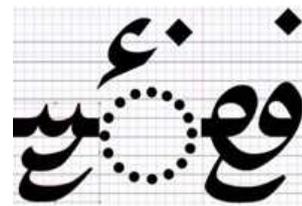


図2. SMU 上部分音符号の組み合わせ  $\bar{D}$  の設計

上部分音符号の組み合わせ  $\bar{D}$  の母音と組み合わせの設計  $\overline{DV}, \overline{DVV}, \overline{VD}, \overline{VVD}$  と  $\bar{D}$  の子音と組み合わせの設計  $\overline{DC}, \overline{DCC}, \overline{CD}, \overline{CCD}$  を設計して、下部分音符



字は使われていないが、ウイグル文字は、モンゴル族の一部が使っている。

SMU 語は天山山脈とアルタイ山脈で集まって生活している人々の言語として広大な地域で話されて、書かれていた。古代トルコ語と呼ばれている文献作品の大部分がウイグル語で書かれた。SMU 文献作品の大部分も古代ウイグル語で記録されている。古代と中期の中央アジアにおける出土した SMU 写本資料の最初のデータは、1902 年 11 月から 1903 年 3 月にトルファン“Turfan”に、ドイツの考古学者と遠征グループ“German Turfan expedition”が、発見した。トルファンは、天山山脈のトルファン盆地の中にあつた古代国名である。トルファンウイグル王国“Turfan Uyghur Handanliki, Turfan Uyghur Doliti”と呼ばれる。トルファン盆地に、コーチョー“Khočo”、イディクト“Idikut”などの有名な古代都市がある。

筆者は、ソグド語、マニ語文書のデジタル処理の研究の参考文献として、現在ベルリントルファンコレクション“Berlin Turfan Collection [1]”に所蔵され、1904 年に発表されたソグド、マニ文字の資料を使用して、本研究を行った。(図 4 A, B 参照)



図 4. 古代と中期の SMU 文献作品

筆者は、ウイグル語文書のデジタル処理の研究においてクダトクピリク“Kutadgu Bilig(幸福をもたらす知恵)”の右から左の横書きの資料を使用した。(図 4 C 参照)

クダトクピリクは、カラハン朝“Karahanolilar(広大王国)”時代に書かれたトルキスタン“Turkistan”の文献作品で、ウイグル族の哲学者、思想者、詩人、ユスブハスハジブ“Yusup Has Hajip”によって、紀元後 10 世紀に書かれた作品である。クダトクピリクのウイグル文字は、右から左の横書きと上から下の縦書きで書かれていた。現在、ウィーンキング図書館に所蔵されている。[2] それ以外にも、クダトクピリクのウイグル語アラビア“Arabic”文字で書かれた書式の異なる本もある。

#### 4.1 ソグド (ソグド人と言語)

ソグディアナ“Sogdiana”とは、紀元前 2 世紀から紀元後 8 世紀まで中央アジアにあつた古代国名である。現在のウズベキスタン、アフガニスタン、タジキスタンは古代ソグディアナ地区である。現在、ソグド語とその文字はすでに消滅している。この領域に、サマルカンド“Samarkand”、パンジケント“Panjakent”、フェルガナ“Fergana”などの有名な古代都市がある。ソグド語は中央イラン系の言語の一種で、ソグド文字は、アラム文字“Aramaic”から得られて、ウイグル文字の直接の先祖であり、モンゴル“Monggol”文字、満州“Manchu”文字、ブリヤ

ート“Buryat”文字、カルムイク“Kalmyk”文字の前身である。ソグド文字は上から下の縦書きと右から左の横書きと二つの書き方がある。本研究では、ソグド語文書の右から左の横書きの文献作品のテキストの構造を使用した。

#### 4.1.1 ソグドアルファベット

ソグド文字の紀元後 4 世紀から紀元後 8 世紀頃間で使用され、そのアルファベットは 33 個の文字“10 個の母音と 23 個の子音”から成っていた。(表 7 参照)

表 7. ソグドアルファベット

𐰀	𐰁	𐰂	𐰃
Ey /ē/ (4)	E /e/ (3)	Ah /ā/ (2)	A / a / (1)
𐰄	𐰅	𐰆	𐰇
Ov /ō/ (8)	O /o/ (7)	Iy /ī/ (6)	I /i/ (5)
𐰈	𐰉	𐰊	𐰋
Pe /p/ (12)	Beth /b/ (11)	Uv /ü/ (10)	U /u/ (9)
𐰌	𐰍	𐰎	𐰏
Cheth /x/ (16)	Che /ç/ (15)	Zhe/ ʒ / (14)	Teth /t/ (13)
𐰐	𐰑	𐰒	𐰓
Zhee / / (20)	Z /z/ (19)	R /r/ (18)	D /d/ (17)
𐰔	𐰕	𐰖	𐰗
F /f/ (24)	Ghe /ǰ / (23)	Shi / / (22)	S /s/ (21)
𐰘	𐰙	𐰚	𐰛
L /l/ (28)	G /g/ (27)	K /k/ (26)	Koph /q/ (25)
𐰜	𐰝	𐰞	𐰟
V /v/ (32)	Hee /h/ (31)	N /n/ (30)	M /m/ (29)
			𐰠
			Jod /j/ (33)

#### 4.1.2 ソグド母音、子音、その他の符号

ソグド文字には 10 個の母音、23 個の子音 4 個の分音符号、3 個の数字、2 個の句読点がある。以下に示す。

母音は 1(“A”  $X_n^a$ ), 2(“Ah”  $X_n^ā$ ), 3(“E”  $X_n^ē$ ), 4(“Ey”  $X_n^ē$ ), 5(“T”  $X_n^t$ ), 6(“Iy”  $X_n^ī$ ), 7(“O”  $X_n^ō$ ), 8(“Ov”  $X_n^ō$ ), 9(“U”  $X_n^u$ ), 10(“Uv”  $X_n^ü$ ) である。

子音は 11(“Beth”  $X_n^b$ ), 12(“Pe”  $X_n^p$ ), 13(“Teth”  $X_n^t$ ), 14(“Zhe”  $X_n^ʒ$ ), 15(“Che”  $X_n^ç$ ), 16(“Cheth”  $X_n^x$ ), 17(“D”  $X_n^d$ ), 18(“R”  $X_n^r$ ), 19(“Z”  $X_n^z$ ), 20(“Zhee”  $X_n^ʒ$ ), 21(“S”  $X_n^s$ ), 22(“Shi”  $X_n^ʃ$ ), 23(“Ghe”  $X_n^ǰ$ ), 24(“F”  $X_n^f$ ), 25(“Koph”  $X_n^q$ ), 26(“K”  $X_n^k$ ), 27(“G”  $X_n^g$ ), 28(“L”  $X_n^l$ ), 29(“M”  $X_n^m$ ), 30(“N”  $X_n^n$ ), 31(“Hee”  $X_n^h$ ), 32(“V”  $X_n^v$ ), 33(“Jod”  $X_n^j$ ) である。

分音符号は(𐰀, 𐰁, 𐰂, 𐰃)である。

数字は(I, II, III)である。

句読点は(𐰔, 𐰕)である。

#### 4.2 マニ (マニ人と言語)

マニは古代と中期に信仰された宗教の名前である。東トルキスタン“East Turkistan”の東ウイグル可汗国“Sherqiy Uyghur Kagan Dowliti”の中期に記録されたマニ文字の多く資料が、天山山脈のヤルクンタク“Yalkun tag(火炎の山)”の中にあつたベゼクリク千仏洞“Bezekli Ming Oy(装飾した千仏の洞)”で発見された。東ウイグル可汗国時代のヤルクンタク地区に生活していたトルファン人の一部は、紀元後 4 世紀と紀元後 8 世紀までマニ教の影響を受けてマニ文字を使っていた。トルファン人によって使われたマニ文字は、トルファンマニ文字“Turfan Manichaean Yeziki”と



