

# マリ語の言語資料とその電子化

松村 一 登

## Summary

### Digitization of Mari linguistic resources

Kazuto MATSUMURA

This paper gives an overview of the current state of the digitization of Mari linguistic resources. After a brief account of the history of Mari literary languages, samples of different types of Mari text are illustrated and the procedural steps of character encoding conversion are presented. There is also an account of the Turku-Tokyo project to create UTF-8 encoded corpora of the two Mari literary languages. Finally, the future prospects of corpus-based Mari linguistics are discussed briefly.

#### 1 マリ語について

ロシアのボルガ川中流域のマリ・エル共和国やその東方のウラル地方のバシコルトスタン共和国とその周辺地域で話されているマリ語 (марий йылме/мары йылме; марийский язык; Mari) は、19世紀の半ばには、フィンランドのM・A・カストレーン (Castrén 1845)、エストニアのF・J・ヴィーデマン (Wiedemann 1847) らの研究によって、その文法構造が明らかにされるようになった。

1872年に最初の初等教本 (букварь) が出されたことが示すように、マリ語は、ロシア革命前から書きことばが使われていた少数言語の1つである。1907年から1911年まで、カザンで印刷された7冊の『マリ語の暦』 (Марла календарь) は、当時のマリ人の知識人層の水準の高さを物語っている。マリ語の文章語の洗練は、スターリンによって民族的知識人たちが処刑された1930年代の後半からしばらく中断するが、第2次大戦後も続けられ、今日に至っている。

革命後に書きことばが作られた少数言語が、まずラテン文字表記を採用し、1930年代の終わりに強制的にキリル文字表記に移行したのとは異なり、革命前に採用されたキリル文字によるマリ語の表記法は、細部では変更もあったものの、ソビエト時代を通じて使われ、今日のキリル文字の正書法に受け継がれている。

マリ語には、現在2つの文章語がある。1つは、東マリ語 (ないし、牧地マリ語) と呼ばれ、もうひとつは、西マリ語 (ないし、山地マリ語) と呼ばれる。西マリ語は、おおむね、マリ・エル共和国の西南部分、ボルガ川の右岸の地域で話され、東マリ語はそれ以外の地域で話される方言に基づく文章語である。出版活動はそれぞれの文章語で行われている。図1と図2は、それぞれ同じ小説の東マリ語版 (Шабдар Осып 1975) と西マリ語版 (Шабдар Осып 1978) の最初のページである。

マリ語に関するより詳しい日本語の解説は、松村（1992）を参照されたい。

## 2 マリ語の言語資料

マリ語は、ロシアの少数言語の中では、文字化された言語資料が比較的豊富に存在する言語といってよい。たとえば、ヘルシンキのフィンランド内国語研究センター（Kotimaisten kielten tutkimuskeskus；Forskningscentralen för de inhemska språken）の図書室には、マリ語の研究書や出版物が豊富にそろっており、研究者なら誰でも利用できる。

現在、研究者が利用できるマリ語の言語資料は、おおむね、次の3つの大きなタイプに分類できる。

(1) 印刷物としてではなく、最初から電子的な文書として入手できるもの。雑誌・新聞の記事のワープロ原稿や、Web上でオンライン公開されている新聞記事などがある。このタイプの言語資料は、今後増えていくと考えられる。

(2) キリル文字表記の正書法で印刷された学校教科書、読本、小説、新聞・雑誌等のテキスト類。革命後のものが主体だが、革命前の印刷物も残っている。このタイプの言語資料がもっとも多く存在すると考えられる。図1と図2は、このタイプの言語資料の典型的な例である。

(3) フィンランドやハンガリーなど、マリ人でない研究者によって、ウラル音声記号（UPA：Uralic Phonetic Alphabet；FUT：Finno-Ugric Transcription）や、ラテン文字表記で記録されたもの。19世紀後半から20世紀初頭にかけて、すなわち、ロシア革命以前から革命直後の時代に、現地調査によって収集された言語資料や、戦時の捕虜や亡命者などとしてロシア国外にやってきたマリ人をインフォーマントにして収集した言語資料が主体。手書き草稿で残っているものも多いが、清書されて、ドイツ語などとの対訳で出版された資料もすくなく存在する。図3はこのタイプの言語資料のサンプルである。

この3つのタイプの区別は、厳格なものではない。(1)は、正書法で表記され、紙のメディアに印刷されて読まれることを前提としているのが通例であるから、(2)のような印刷されたテキストからコンピュータ入力されて作成された二次的な電子化言語資料は、事実上(1)と区別できない。この論文の著者は、マリ語の書籍を数冊、手入力で電子化してコーパスとして利用している。(3)は話しことばが主体であるが、清書されて印刷物として出版されれば、正書法で書かれていないという点を除き、事実上(2)のタイプとみなすこともできる。さらに、Unicodeでエンコードした電子化資料にすれば、(3)のタイプの資料もコンピュータ処理が可能になる。

このように、現在入手できるすべてのマリ語の言語資料は、Unicodeでエンコーディングすることによって、原理的に、すべてコンピュータ処理が可能な文字資料となる。ただし、マリ語の文字化された資料を実際に電子化し、コーパス、ないしは全文検索型のテキストデータベースとして利用できるようにするためには、少し工夫が必要である。実例を用いて、ロシアから入手した(1)のタイプの言語資料を実際に処理する手順の概略を示そう。

ロシアで作られたマリ語のファイルは、たいていMS Wordの文書である。まず、これをいわゆるテキスト・ファイルとして保存し直したものを、そのまま文字コードを指定せずにエディ

ターで読み込むと図4のように画面に表示される。これは、MS Windowsでキリル文字言語の表示に使われるフォントでは、キリル文字を、西欧語の表示のための標準的なフォントでウムラウト文字他の特殊文字が収納されている領域に収納しているためである。これをキリル文字表示用のフォントに切り替えれば、図5のようにキリル文字が画面に現れる。

この段階ですぐ気づくことは、ロシア語では使われない5つのキリル文字系の特殊文字(Һ, ä, ö, ӱ, ё)の代わりに、% (=H) や ε (=ö) など、別の文字や記号が現れていることである。これは、現時点で、ロシアにおけるマリ語のコンピュータ処理にUnicodeが普及しておらず、これらの特殊文字を外字で表したMS Windows用のキリル文字フォントを用いて文書を作成しているためである。すなわち、ロシアでマリ語の特殊な文字の表示のために使われているUnicode対応でないフォントがインストールされていないシステムでは、ロシアで作成されたマリ語の文書を表示すると、問題の5つのキリル系文字(Һ, ä, ö, ӱ, ё)が、画面上ではマリ語の文字としてではなく、別の文字や記号として表示されてしまう。

この場合、当該の5つの特殊文字の代わりに使われている記号・文字を、Unicodeに対応したコードポイントでそっくり置き換えてしまえば問題は解消する。この作業は、ともにUnicodeによる文字処理に対応したスクリプト言語(たとえばPerl 5.8)とテキスト・エディタ(たとえばEmEditor)を用いておこなうと効率的に進めることができる。

(2)のタイプの言語資料、すなわち、キリル文字表記の正書法で印刷されたマリ語のテキストは、特定のOCRソフト(たとえばFineReader)を使えば、非常に高い精度で読み込むことが可能である。認識結果の校正の際には、Unicode対応のエディタで作業を行う。なお、校正の過程で、マリ語で使われる特殊文字の入力を効率よく行うためには、やはり特殊文字入力用の補助ツールがあったほうがいい。

(3)のタイプのマリ語言語資料の中で、ウラル音声記号ないしその変種で表記された資料は、言語学的な研究を目的とする録音資料などの文字転写であることが多い。この種のテキストでは、広く普及しているコンピュータのOSやアプリケーションに付属する標準的なフォントが実装していない特殊な文字や、補助記号の複雑な組み合わせがふんだんに用いられている。このため、入力が困難を極めることが多く、もっとも電子化しにくいタイプの言語資料であるが、歴史的な資料として、また、話しことばの記録として、(1)や(2)のタイプの言語資料では代替できない貴重な資料であり、電子化された場合の付加価値が高い。

音声記号によって表記された言語資料の場合、OCRによる認識の精度は一般にあまり高くなく、キーボードからの手入力のほうが効率があがることも多い。手入力の場合、特殊文字入力用の補助ツールが必須である。図6は、この論文の著者が開発したUnicode対応の特殊文字入力用補助ツール(UniInput)によって、図3の言語資料の冒頭部分を手入力した際の画面である(鈴木・小野・松村2003)。

### 3 マリ語のコーパス

この論文の著者は、2004年10月、フィンランドのトゥルク大学ボルガ諸語研究部門のJ・ルートネン(Jorma Luutonen)氏と共同で、マリ語のコーパスを整備するプロジェクトを立ち上げた。このプロジェクトは、トゥルク大学が作成し所有するマリ語の言語資料と、この論文の著者が作

成し所有するマリ語の言語資料 (4.5 MB) を合体させて Unicode (UTF-8) でエンコードされた中規模のマリ語コーパスを作成するとともに、あわせてその言語資料の言語学的利用のためのツールを開発することを目標にしている。

この論文の著者が、トゥルク大学から使用を許されたマリ語の文章語の言語資料は、概略次の通りである。

a. Кугарня (新聞)	0.4MB (東マリ語)
b. Марий Эл (雑誌)	4.5MB (東マリ語)
c. Marko コーパス	4.3MB (東マリ語)、1.0MB (西マリ語)
d. Ончыко (雑誌)	29.4MB (東マリ語)
e. У сем (雑誌)	6.0MB (西マリ語)
f. その他	1.5MB (東マリ語)

このプロジェクトにおける私の役割分担は、テキスト資料の全体を Unicode (UTF-8) に変換し、コーパスとして利用できるデータ形式に整形する作業であるが、文字コード変換の段階はすでに終わり、校正の段階にはいつている。なお、c. は現在トゥルク大学のサーバーに置かれているマリ語のコーパスで、以前、私が入力して寄贈したテキストが含まれている。c. 以外の言語資料は、トゥルク大学がロシアから入手し、まだ未整形のまま置かれていたものである。

トゥルク大学のマリ語のコーパスは、以下のサンプルに示すように、たとえば s1 でキリル文字の小文字шを表し、+m でキリル文字の大文字Mを表すというふうに ASCII の小文字と数字や記号の組み合わせによってマリ語のテキストをエンコードしている。このエンコード方式は、1980年代ころまでのもので、今では時代遅れになっている。

```
2 : 1  2  uremyslte    **l**
2 : 2  -  poro kecle !
2 : 3  -  poro kecle ! kuze ilet ?
2 : 4  -  tau , elrkyn-elrkyn . a slkecle ?
2 : 5  -  myalt sajynak ilem . elrla tyj mom ysltet ?
2 : 6  -  myj elrla +morkysl kaem .
2 : 7  -  molan +morkysl kaet ?
2 : 8  -  +morkyslto avam ila . tudo ynde slonlgo . tudlan polslasl kaem .
2 : 9  tudyn izirak sadsle ulo . sadpakclaslte olmapuzlo-vlak . olmam pogasl
2 : 10 kaem . olmazle pesl saj . avam mylamat olmam pua .
```

プロジェクトの成果としてできあがるコーパスの最終的な大きさは、言語資料全体の整形が完了するまで不明だが、校正がかなり進んでいる d. の雑誌 Ончыко の 4 年分 (1996-1999) のテキストだけでおよそ 180~190 万語ある。

このプロジェクトでは、マリ語の言語資料を文ごとに分割して、XML 文書としてのマークアップ

プを行う（図7）。そのあと、開発中の品詞形態分析プログラムによって、語のレベルで品詞形態情報を付加することを目指している。

品詞形態情報が付加されていないコーパス、いわゆる plain text のコーパス用の Unicode 対応の KWIC 索引作成ツールを、この論文の著者は Perl 5.8 で試作して、マリ語のコーパス検索で実際に使用している。このツールを用いて、上の d. のマリ語資料から名詞 ен 「人間」の用例を検索した結果の KWIC 索引の一部が図8である。英語などラテン文字系の言語のコーパス用に多機能のコーパス検索用ツールが多数開発されているが、この種のツールがキリル文字の特殊文字や、音声記号を含むテキストでも利用できるのか、まだ未確認である。

英語や日本語などの大言語においては、コーパスを用いた語彙研究、文法研究、談話研究が一般化し、言語研究の現場で日常化しつつあるが、マリ語のような少数言語でも、十分な量の言語資料が電子化されさえすれば、近い将来、コーパスを用いた研究がふつうに行えるようになるであろう。マリ語のテキストに品詞形態情報を付加するツールは、ハンガリーで開発されているようだが、残念ながら Unicode にはまだ対応していないようである。トゥルク大学との共同プロジェクトでも、この課題の解決を試みることになっている。この論文の著者による試みについては、松村（2003）で簡単に紹介した。

### [引用文献]

Alhoniemi, Alho und Sirkka Saarinen, ed. (1983) *Timofej Jevsejevs Folklore-Sammlungen aus dem Tscheremissischen, I. Märchen, Sagen und Volkserzählungen*. Helsinki : Sociéte Finno-Ougrienne.

Castrén, M. A. (1845) *Elementa grammatices Tscheremissæ*. Kuopio.

Wiedemann, F. J. (1847) *Versuch einer Grammatik der tscheremissischen Sprache*. Reval.

Шабдар Осып (1975) *Ўдырамаш корно*. Йошкар-Ола: Книгам лукшо марий идательство.

————— (1978) *Ыдырамаш корны*. Йошкар-Ола: Книгам лукшы мары издательство.

鈴木麗爾、小野智香子、松村一登 (2003) 『フィールド言語学者のための Unicode ツール』

(環太平洋の「消滅に瀕した言語」にかんする緊急調査研究 成果報告書 B010)

松村一登 (1992) 「マリ語」— 『言語学大辞典』第4巻 (三省堂)、pp.155-164.

————— (2003) 「少数言語のコーパス」— 『月刊言語』2003年11月号、pp.86-87.

\*本稿は、科学研究費補助金による研究「音声記号等で表記された言語資料のマークアップとコンピュータ処理」(基盤研究(A)(2) #15202008、2003~2005年度、研究代表者・松村一登)の研究成果の報告の一部をなすものである。

## Первый книга

### СЕМЫК

#### 1

Чевер кече чолга шинчаж дене ужар мландым раш онча. Мланде ўмбалне: уремыште, олык лапыште, мардеж логалдыме моло верыште — лўп-лўп шокшын чучеш, ласкан шўлалташ юж ок сите. Уржа пасушто гына изишак йонгыдо: ала-кушеч чодыра велым толын лекше умыр мардеж, выж-ж пуалын, вуйым лукшо уржам Юл вўд ловык гай койыктен тарвата... Адак тыгай годым ўмылан чодыраште кожер йымалан шылын кодшо эр юалге юж, чодыра кишын, пеледышын ўпшен, айдемым, тугак моло чонанымат куандара.

Ужар посто гай олык мучко, южо вере, нарынче лаштыкын койын, семыкшудо<sup>1</sup> пеледеш. Кашын мардеж почеш мўяк-гутло олык саска пуш нерым шыман чыгылта. Тиде лопка олык покшечын ший кудыр гай кадыргылын, коклан ужар чашкерыш пурен-лектын, энгер вўд чурийончыш гай йылгыжеш, мотор кенгеж пўртўсым каласен моштыдымын сўрастара.

Олык мучко южо вере, тўшкан-тўшкан пўдыра ош пеледышла койын, ала-мо тарванымеш: ала кўшыч кайыше ош йўксў-влак канаш волен улыт, ала ашныме комбо, игыжым вўден, нўргў шудым пурлаш лектын, ала ош комбо гай ошын чийыше марий ўдыр-влак шинчаланшудым погат — мўндырчын палашат ок лий. Шукыжо ўдыр-влак докан. Молан манаш гын, таче яра кече. Яра кечат веле огыл — таче семык... Ужар олык покшелне ўдыр-каче-влак кок тўшка дене коштыт. Теве ик тўшка

<sup>1</sup> Семыкшудо — нарынче пеледышан сурепица шудо.

#### 3

# ПЫТАРИШ КНИГА

## СЕМЫК

### 1

Цевеp кечы соты сѣнзажы доно ыжар мўландым тўрўк анжа. Мўланды вѣлны: ўлицашты, алык лапышты, мардеж пуалдымы молы вўрышты — пишок шокшын чучеш, йонгын шўлалташ воздухшат ак ситы. Ыржа нырышты веле изиш йонгата: тагышецын шўргы дорцын толшы изи мардеж, выж-ж Ыфылалын, вуйым лыкшы ыржам Йыл вѣлныш коэла тўрваталеш... Техень годым Ымылан шўргы лоштыш кож лѣвалан шўлын кодшы киш дон пеледыш лышан ир юалгы воздух веле эдемым, тенгеок молы йўнганымат ладнангдара.

Ыжар посто гань алык мычки, кыды вўрежы, нарынзы лаштыкын кайын, семыкшуды\* пеледеш. Шоэн-шоэн мардеж семынь толшы алыкыштыш мўак-тотлы шуды пыш нерым шўман цўгылта. Ти кымда алык покшецын ши лентала айынылын, кынамжы ыжар тўшкавлў лошкы пырен-лўктын, ўнгыр вўд тўгыр ганьы йўлгыжеш, цевеp кўнгыжын цўрежым желесен мыштыдымыла сўлнештара.

Алык мычки кыды вўрежы, кого ош пеледышла кайын, аран-аран тама-ўнят тўрваньлеш: аль кўшыц чўнгештышы ош йўксывлў кўнаш валеныт, аль тоны урдым комбы, игывлўжым потен, нўргы шудым пырылаш лўктын, аль ош комбы гань ошым чишы мары Ыдырвлў охырашудым погат — мўндырцын пўлашат ак ли. Шукыжок Ыдырвлў, векат. Малын манаш тўнь, тагачы йўрсым кечы. Йўрсым кечы велеат агыл — тагачы семык... Ыжар алык покшалны Ыдыр-рвезывлў кож партын каштыт. Те-

\* Семыкшуды — нарынзы пеледышан сурепица шуды.

### 3

kuze ožnê<sup>1</sup> Jêβaška štêkten ialan čerkê m i polat-  
šamêčê m Čarla olaš

Pulanur mari-šamêč mlanem kalasen ult têγaj jomakê m:  
nemnan kurê m teč jatê r ončê č Čarla olaštê ilen ik  
oza. tuðê n lümšê ulmaš Jêβaška. kêzê t Čarlaštê ulê  
ik čerkê i polat-šamêč. Čarla mari-šamêč i Pulanur  
mari Miron mêlanem têγe kutê ra:

βoznesenskij čerkê m, kaznačej polatê m, ola učičê l  
pörtê m i (ateka) lekê r pörtê m Jêβaška štêkten ialan.  
tiðê ozažê Jêβaška šočê n kuštê ulmêžê m raš oγê t šin-  
čê. lüm kê čšê<sup>2</sup> šonat tuðê mari.

Jêβaška ulmaš ia ðen roðê kučêšê. ian polšê mê ðen  
tuðê šukê kuyu čapê m koden. olat, tiðê čerkê m i kü  
polat-šamêčê m tuðê štêkten ia-šamêčlan. čort-šamêč  
Jêβaškam<sup>3</sup> peš kolštê n ult i mom nunê lan Jêβaška šü-  
ðen čê la šten ult.

Jêβaška ške ilen kêzê tsê ola učičê l pörtê štê. tiðê  
učičê l pört šoγa kuyu Kakšan eγer βokten. tiðê Kak-  
šan joyen kaja kuyu Jul βüðê škê. Kakšan βüt kê č pört  
jümak ulmaš ian kornêžê. tiðê kornê m sajn kermê č ten  
saβê rmê ulmaš. a ia-šamêč tiðê kornê ðen Jêβaška ðek  
unalê keš<sup>4</sup> koštê t ulmaš.

tiðê Jêβaška očê ní torγoβuj jê η ulmaš. tutlan Ozan  
olaš kajaš küleš ê lê γê n, tuðê koštê n Kakšan βüt üm-  
pač. tuðê m koštêkten ult ia-šamêč. tuðê Kakšan βüt  
ümpak šara ulmaš portê šê m, a portê š ümpak pê štê ul-  
maš küpčê kê m, škežê Jêβaška tê š βozeš ta Ozan olaš  
kaja i toleš ulmaš.

<sup>1</sup> Im Original oшнŷ.

<sup>2</sup> Im Original лумнŷчшŷ.

<sup>3</sup> Im Original Jŷвашнем.

<sup>4</sup> Im Original уналкеш.

Ī'ëäiûøðä êîêðúí óëüià. Âàøðàððàøàí èàèøèðàè øçðãûâûëüøàí, içñ°  
 øéí÷âíí÷àèðùøàí èèàèøà á%. Òèää — èóãøí-âèàèèèáí èóíääí iàèüià iðíçàèè  
 Àèäèñàé Ôããíðíàè÷ Èðãñííãðíã. Òóãúí âíçúñ "Ñððãää", "Ñúñãüü", "Ääëüièà  
 çàðíèèü" ðñiáí, "Īñíáúé ñíèää", "Īñèããíèèè iàððíí", "Òðóáèà èñíáàðà" äà øóéí ñèí  
 ñããñòü, íéèüiàøùæã-âèàèèè ÷üèèiàð èóãúíüð.  
 Àèäèñàé Ôããíðíàè÷ Èèðíã íáèãñòüñã Ííèññíèà èçíáí ýèãø 1921 èèüøðä iàðèè  
 áøãø øí÷úí. Ìàðèèæã øèã äèàèèèèèøã äáíá ðçèíüää íóðèià, íí ðãíð÷ãñèèè  
 iàøàøðùæã ðóø èúèiüèèáí ý%ãððà. Òóãã äüiàð Èðãñííãðíãúí ðç%ããðíèèè-âèàè  
 ÷üèèiàð iàðèè èíéüø-øíèèèèèè, ðàðàèèèèèè óéüð.

図4 ロシアから入手した MS Word のファイルをテキスト・ファイルとして保存したもの

Пелемыште коктын улына. Ваштарешем какширак шўргывылышан, пўсе  
 шинчаончалтышан илалше е%. Тиде — лудшо-влаклан кумдан палыме  
 прозаик Алексей Федорович Красноперов. Тудын возымо "Страда",  
 "Сыновья", "Дальние зарницы" роман, "Особый сплав", "Последний патрон",  
 "Трубка комбата" да шуко моло повесть, ойлымашыже-влакым чыланат  
 лудыныт.  
 Алексей Федорович Киров областысе Ониссола лўман ялеш 1921 ийыште  
 марий ешеш шочын. Марлаже шке диалектше дене тўкныде мутлана, но  
 творческий пашаштыже руш йылмылан э%ерта. Туге гынат Красноперовын  
 тў%о геройжо-влак чыланат марий койыш-шоктышан, характеран улыт.

図5 図4のファイルを MS Windows のキリル文字コードで再表示したもの

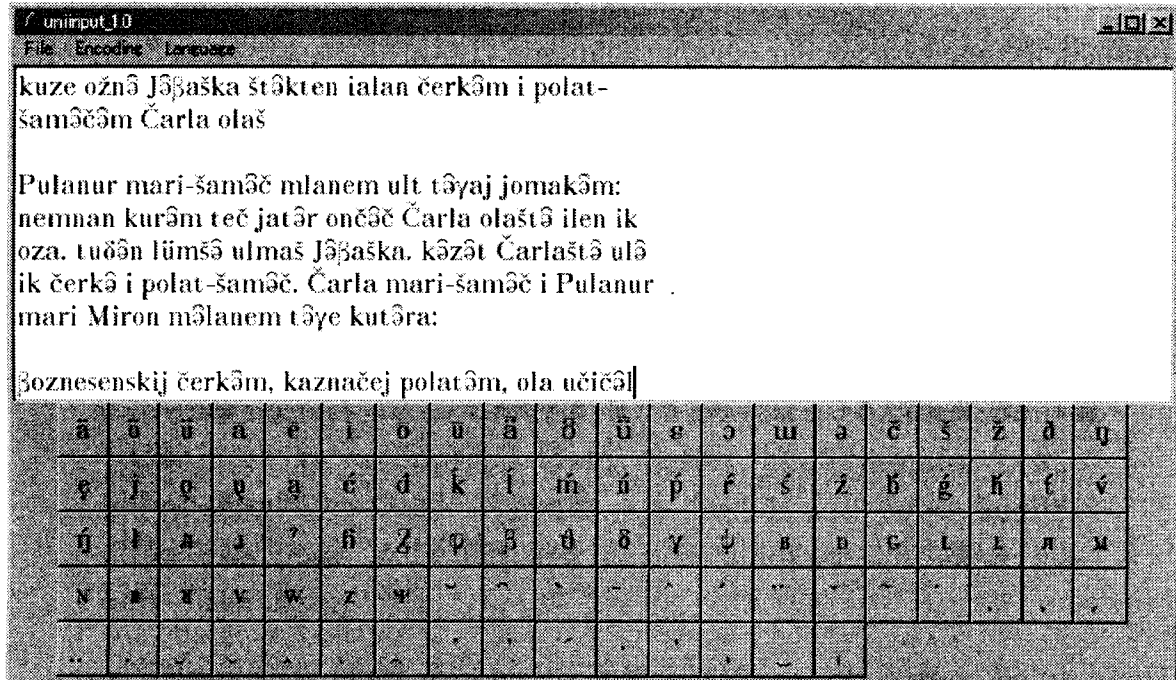


図6 特殊文字入力ツールの入力画面

<p>  
 <s> Пӧлемыште коктын улына. </s>  
 <s> Ваштарешем какширак шӱргывылышан, пӧсӧ шинчаончалтышан илалше ен. </s>  
 <s> Тиде — лудшо-влакклан кумдан палыме прозаик Алексей Федорович Красноперов. </s>  
 <s> Тудын возымо "Страда", "Сыновья", "Дальние зарницы" роман, "Особый сплав", "Последний патрон", "Трубка комбата" да шуко моло повесть, ойлымашыже-влакым чыланат лудыныт. </s>  
 </p>  
 <p>  
 <s> Алексей Федорович Киров областысе Ониссола лӱман ялеш 1921 ийыште марий ешеш шочын. </s>  
 <s> Марлаже шке диалектше дене тӱкныде мутлана, но творческий пашаштыже руш йылмылан энгерта. </s>  
 <s> Туге гынат Красноперовын тӱн геройжо-влак чыланат марий койыш-шоктышан, характеран улыт. </s>  
 </p>  
 <p>  
 <s> Писатель пӱрымашат моло айдеме семьянак вучыдымо савыртышан лиш. </s>  
 <s> Алексей Федоровичат сылнымут корным ойырен налме нерген ончычшо шоненат огыл. </s>  
 <s> 1940 ийыште Ярангысе педучилищым тунем пытарымеке, ялысе школышто йоча-влакым туныкташ тӱнгалын, но ик-кок тылзе гыч тудым салтак радамыш ушат. </s>  
 <s> Ончычсо самырык туныктышо связистыш савырна. </s>  
 <s> Сар тӱнгалмым элнан касвел чекыштыже вашлиеш, Прибалтикем арала, вара — мондалтдыме, орлыккан да шучко 900 йӱд-кече, Ленинград блокадым тӱрыснек чытен лекмаш. </s>  
 </p>  
 <p>  
 <s> Шуко совет салтаклан 1945 ий 9 Май пытаргыш сар кече лийын гын, Алексей Федоровичлан тудо утларак шуйнен. </s>  
 <s> Пӱрымаш Ониссола рвезым ынде Эрвелыш лупшалеш, садлан война пагытат кужемеш. </s>  
 <s> Тыге марий самырык рядовой салтак гыч тӱнгалын, ынде 40 ияш майор, сур шинельым 1961 ийыште веле, Йошкар-Олашке пашам ышташ толмекше, кудашеш. </s>  
 </p>

図7 マークアップされたマリ語テキスト

