

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

**Annotation and Computer Processing of Language Resources  
in Non-Latin Scripts and Phonetic Transcription**

課題番号 15202008  
Grant No. 15202008

平成15年度～平成17年度

科学研究費補助金基盤研究(A)研究成果報告書

2003–2005 Grants-in-Aid for Scientific Research (A) Research Report

研究成果報告書  
Research Report

平成 18 年 3 月  
March 2006

研究代表者 松村 一 登  
(東京大学大学院人文社会系研究科教授)

Project Leader: Kazuto Matsumura  
(Graduate School of Humanities and Sociology, University of Tokyo)



# ウラル系諸語の言語資料の電子化とマークアップ

## — 現状・課題・展望 —

松村一登

東京大学大学院人文社会系研究科

kazuto@tooyoo.l.u-tokyo.ac.jp

### 1. はじめに

私たちは、だれでも母語の使い手として、膨大な量のことばを毎日産出している。しかし、話されたことばは、ほとんどの場合産出されたその瞬間に、また、書かれたことばも、たとえば紙という物理的なメディアが切り刻まれたり、燃やされたりしたその瞬間に、そのとき何らかのことばが産出されたという事実もろとも、過去という時空のかなたに吸い込まれて、永久にアクセス不可能になる (=失われる)。これは、世界のいたるところで、毎日繰り返されている営みである。

ふつうはあとかたなく消えてしまう運命にあることばが、録音されたり、文字化されたりして、運よく保存されたものの1つ1つを「言語資料」(linguistic document) と呼ぶことにする。

言語の研究の基礎になるのは、研究対象とする言語の言語資料が一定量利用可能な状態にあることである。言語資料の少ない言語については、自分で、現地調査やインフォーマント調査を行うなどして、新たに言語資料を集めることから始めないと、そもそも研究が成り立たない。したがって、言語資料を集めることそのものが、言語の研究にとってきわめて重要な位置をしめる。また、そのようにして集めた言語資料は、文字化されることによって、研究資料とし

て、他の言語資料と同じステータスをもつようになる。

言語資料を集める方法と並んで、ひじょうに重要なことは、集めた言語資料を研究者の間で共有できるデータ形式にすることである。とくに、そうでなくとも人材的、資金的リソースにめぐまれない少数言語の場合、研究者ごとに、あるいは、研究目的ごとに（しばしば互換性のない）異なったデータ形式の言語資料を用意するよりは、研究目的や利用形態が異なっても、複数の研究者や研究プロジェクトの間で使い回しのきく言語資料になっているほうが、格段に好ましいことである。また g、現在では、集められた言語資料は、電子化され、いわゆる機械可読なデータ形式、すなわち、コンピュータで扱えるデータ形式に変換されて、蓄積・保管されるのが通例になっている。

科学研究費補助金・基盤研究 (A)(2) による本研究プロジェクトは、こういった視点を踏まえて行われたものである。以下では、研究代表者である筆者の専門領域、ウラル諸語の言語資料に焦点をあてて、研究の成果を報告する。

## 2. 言語資料の電子化

ウラル諸語の少数言語の言語資料の多くは、音声記号やキリル文字系の特殊文字を用いて表記されている。このため、言語資料をコンピュータによって処理・解析しようとするウラル諸語の研究者は、たいへんな苦勞を強いられているのが現状である。また、ウラル諸語の研究者に限らず、似たような境遇にある言語の研究者なら、だいたい同じような状況におかれているといっていよい。

この苦勞は、(a) 音声記号やキリル文字系の特殊文字等が過不足なく収録された Unicode 対応フォント、(b) コンピュータに関する高度な知識を持たなくても、音声記号ほかの特殊文字等を入力することができるようなツール、(c) 音声記号ほかの特殊文字で表記された文字資料を、コンピュータに関する特別な知識を持たなくても、容易にコンピュータ処理・解析のできる検索ツール — この3つの要素がそろって満たされるならば、原理的に解消できるはずである。

この3つの課題のうち、(b) については、筆者が研究代表者として 2000～2002 年度に行った特定領域研究 (2) 「消滅の危機に瀕した言語の言語資料のコ

ンピュータ処理のためのデータ構造・分析ツールの研究」において、完璧と言うにはほど遠いが、特殊文字の入力用の実用に耐える補助ツール [参考文献 1]、を開発し、特殊文字の入力に伴う手間をかなりの程度まで軽減することに成功した。

本研究の重点課題の 1 つは、(a) のフォントの問題の解決であったが、ラテン文字、キリル文字、音声記号に関するかぎり、目的はほぼ達成された（「5. ウラル音声記号を実装した Unicode フォントの開発」）。(c) についても、UTF-8 でエンコードされたラテン文字、キリル文字、音声記号で表記され、文レベルがマークアップされた言語資料ならば汎用で使える Perl 言語の検索ツールを使って、KWIC 索引の作成や、言語形式の出現頻度数の計算を、比較的容易に行うことができるようになった（「4. Perl 言語による検索プログラムの開発」）。

言語資料の電子化において、もっとも重要で、もっとも時間と人手がかかるのが、言語資料の入力とマークアップである。ここで、言語テキストのいわゆる入力と、入力されたテキストのマークアップとは、はっきり区別しておいたほうがよい。なぜなら、入力は、使用文字によっては、OCR などによってかなりの程度まで自動化できるし、手入力する場合でも、文字に関する知識さえあれば、対象言語に関する知識がない人間でも行うことの出来る作業であるのに対し、マークアップは、テキストの中身に立ち入らないと出来ない部分が多いからである。ラテン文字やキリル文字、漢字かななどで表記されたテキストの場合は、業者に発注してOCR 作業まですませてもらい、納入されたOCR データを原典と付き合わせて校正する作業の段階になってから、当該言語の知識のある研究協力者に仕事を依頼するという工程にするという段取りが、おそらくもっとも効率のよい作業計画であろう。

### 3. ウラル諸語の言語資料の電子化とマークアップ

この研究プロジェクトでは、当初から XML でマークアップを行うことを目標にしてきた。その目標に近づくために、研究分担者の千葉庄寿氏（麗澤大学）を講師にして、「言語資料の XML によるマークアップ」と題する研究会を、

2004年12月11日に東京大学で開催した。この研究会は、この科研費プロジェクトの関係者以外にも公開し、参加者は22名になった。

そもそも、マークアップされた言語資料が一定量蓄積されるまでは、電子化された言語資料を使った研究は始めることができない。筆者がこの研究プロジェクトで目標としたのは、自分の研究対象であるウラル諸語に関して、Unicode (UTF-8) でエンコードされ、文レベルまでマークアップされた言語資料を、量的に整備することであった。具体的な成果については、それぞれの言語の言語資料に関する箇所で述べる。

言語資料のマークアップに関しては、XML (Extensible Markup Language) を用いることが当初からの目標だったが、Unicode でエンコードしたテキスト資料を、いわゆる well-formed な XML 文書としてマークアップする段階でよしとした。つまり、テキストのマークアップのための特定の規格 (たとえば、Text Encoding Initiative, TEI; Corpus Encoding Standard, CES ) を用いて、言語資料をいわゆる validate された XML 文書としてマークアップするという段階にまでは至らなかった。もっとも大きな理由としては、人文系の研究者には、最先端の LT 技術 (language technology) を十分に習得することが時間的に困難であったことである。

他方、コーパスの標準的なマークアップ規格が国際的に未だ確立しているとは言えない現状において、特定のマークアップ規格を採用するのはリスクが大きく、本格的なマークアップを今積極的に試みるよりは、先送りした方が無難であると判断したという事情は、ここで述べておくべきであろう。LTの世界では、技術の発達が早いので、2006年度から始める新しい科研費プロジェクトでは、テキストのマークアップに本格的に取り組むことができる程度まで、機が熟すだろうと考えている。

ウラル諸語の言語資料に関しては、文より上のレベルまでに限ってはあ  
るが、コーパス・ファイルが well-formed な XML 文書になるようにマークアップするためのノウハウが獲得できたと考えている。上で述べたように、ベプス語、マリ語、エストニア語等の言語資料で、利用可能な状態にあるコーパス・ファ

イルは、well-formedなXML文書のマークアップになっていることが、XMLエディタほかのチェック機能で確認した。

今後の課題は、言語資料に関する、いわゆるメタデータ（書籍の書誌情報にあたるもの）を体系的に付加するノウハウの確立と、文レベルより下位の要素、すなわち語のレベルで、品詞形態情報などの言語的情報を付加するためのツールの開発や、テキスト全体のマークアップのための規格の選定の作業を行うことである。品詞形態情報の付加にあたっては、自動的に品詞形態分析を行うプログラム（いわゆる POS tagger）が、少数言語に関しても開発されることが望まれる。

語レベルでの文法的情報の付加は、将来的には、おおむね次のような形式のものになるのではないかとイメージしている。

マリ語の原文とその語釈，翻訳

Йыландан чевер оръенже чолга шўдыр гай йолга.  
Jylanda-GEN beautiful bride-3sg clear star-NOM like shine.PAST-I.3sg  
「ヤランダの美しい花嫁は明るい星のごとく輝いていた」

```
<s>
<w lem="Йыланда" pos="prop. noun" mrf="gen.sg">Йыландан</w>
<w lem="чевер" pos="adj" mrf="nom.sg">чевер</w>
<w lem="оръен" pos="noun" mrf="nom.sg.3sg">оръенже</w>
<w lem="чолга" pos="adj" mrf="nom.sg">чолга</w>
<w lem="шўдыр" pos="noun" mrf="nom.sg">шўдыр</w>
<w lem="гай" pos="postp" mrf="">гай</w>
<w lem="йолгаш" pos="verb" mrf="3sg.pastI.ind">йолга</w>
</s>
```

### 3.1. ベプス語・ロシア語対訳テキスト集の電子化

ベプス語 (Vepsian, Veps; вепский язык) は、ウラル系バルト・フィン諸語に属する言語で、2002年のロシア連邦国勢調査によれば、ベプス人8240人のうち、5753人(69.8%)が話すと答えている ([参考文献3])。ベプス人の居住地域は、

オネガ湖 (Онежское озеро) の南西岸地域 (シェルトゼロ Шелтозеро 周辺), およびオネガ湖の南のペプソフスカヤ台地 (Вепсовская возвышенность) である。1930年代に文語を成立させる試みがなされたが定着しなかった。ソ連邦崩壊後の1990年代から、文語を復活しようとする動きがおこり、学校教科書、文法書などが出版されている。

バルト・フィン系の言語の場合、ロシア正教系の民族の言語の場合でも、フィンランド語やエストニア語にならって、ラテン文字表記を採用する傾向が強い。

ベプス語の現在の文章語のテキストの例

Vepsläižed — nece om üks' vähäluguižiš rahvhišpäi, kudamb eläb Venän pohjoižpoles. Istorii sanub, miše hö oma elänuded neniš tahoiš jo VI. voz'sadalpäi; nene tahod oma vepsän rahvhan etnine kodima. Venän valdkundan rahvhan kirjutamine ozutab, miše 1897 v. oli völ 25,5 tuhad vepšläst, no möhempa rahvhanlugu väheni, i nügüd' jäl'gmäine 1989 v. rahvhan kirjutamine ozutab, miše vepsläižid om jänu 12.5 tuhad mest. Vepsän rahvhal ei ole ičeze valdkundad. Hö eläba Karjalas (Petroskoi i Änižjärven agj), Vologdan oblastiš (Babajevon agj) i Peterin oblastiš (Podporožjen, Tihvinan, Boksitan i Lodeinian Pöudon agjoiš).

Vepsän kel' om üks' Baltian meren suomalaisižis kelišpäi. Baltian meren suomalaisižed keled oma: suomen kel', karjalan kel', estin kel', ižoran kel', vodin kel', livin kel'. Lähemba kaikiid vepsän kel'he om karjalan kel' (hänen livvin i lüdin paginad). (Zaiceva 1995:3)

本研究プロジェクトでは、ロシア語訳が添えられて1969年に出版された、ウラル音声記号表記のベプス語のテキスト集 (Зайцева и Муллонен 1969; 296 ページ) をそっくり電子化し、well-formed な XML 文書としてマークアップした。ベプス語とロシア語のテキストの入力は、ロシア科学アカデミー言語学研究所 (ペテルブルク) の E・V・ゴロフコ氏 (Evgeniy V. Golovko) (入力作業: M・Z・ムスリモフ氏 [Mekhmet Z. Muslimov]) に依頼して行った。入力作業には、この科研費プロジェクトで開発したフォントが使われた (本稿「5. ウラル音声記号を実装した Unicode フォントの開発」参照)。XML によるのマークアップは、筆者が東京で行った。

資料1として、入力したテキスト集のページのサンプル (ウラル音声記号で転写したベプス語テキストとそのロシア語対訳) , 資料2として、同じベプス語テキストに XML のマークアップを試みたもののサンプルを載せた。

ベプス語は言語資料の少ない言語なので、入力したテキスト集のほかに、辞書や文法書・文法記述も、参考のためにここにあげておきたい。

#### ベプス語のテキスト集 (入力したテキスト集)

М. Зайцева и М. Муллонен 1969. *Образцы вепсской речи*. Издательство “Наука”, Ленинградское отделение.

#### ベプス語の辞書

Зайцева М.И. и Муллонен М.И. 1972. *Словарь вепского языка*. Ленинград: Наука

#### ベプス語の文法記述

Зайцева М.И. 1981. *Грамматика вепского языка*. Ленинград: Наука

Хямяляйнен М.М. 1966. “Вепский язык.” - *Языки народов СССР III*. Москва: Наука

Зайцева М.И. 1993. “Вепский язык.” - *Языки мира. Уральские языки*. Москва: Наука.

N. Zaiceva 1995. *Vepän kelen grammatik*. Petroskoi: Petroskoin valtion yliopiston kustantamo.

### 3.2. マリ語のコーパスの拡充

マリ語 (Mari; марий йылме; марийский язык)は、ロシアのボルガ川中流域のマリ共和国、バシコルトスタン共和国、キーロフ州などで話されているウラル系ボルガ諸語に属する言語である。2002年のロシア連邦の国勢調査では、マリ人は604,298人、マリ語の使用人口は母語話者は約487,855人(2002)。マリ語には、19世紀の終わり頃からの伝統で、方言差に由来する文章語が2つ(東マリ

語 451,033 人, 西マリ語 36,822 人) 使われている。

フィンランドのトゥルク大学のボルガ諸語研究を専門とする S・サーリネン教授 (Sirkka Saarinen) と研究者の J・ルートネン氏 (Jorma Luutonen) を相手方として, マリ語のコーパスを整備する共同プロジェクトの合意文書を 2004 年 8 月に交わした。トゥルク大学のマリ語の言語資料を Unicode でエンコードし直し, 最低限の XML マークアップを施す役割を引き受ける代わりに, トゥルク大学のマリ語コーパスをそっくり寄贈してもらうことになった。以前, 筆者の側から寄贈したマリ語のテキストが, 現在トゥルク大学のマリ語コーパスの一部をなしているが, 言語資料の収集の面ではフィンランド側の方が圧倒的にめざまれているので, こういった相補う形で協力体制が組めたことは幸運といていい。このプロジェクトは最初, 2006 年 3 月までのものだったが, 2006 年 4 月以降も更新されている。

トゥルク大学の側から提供された言語資料は, ロシアで MS Word 文書として作成されたマリ語のテキスト (以下「原資料」) と, 原資料をトゥルク大学のマリ語コーパスのエンコード方式として使われてきた ASCII の転写方式に変換したファイル (以下「ASCII 転写」) である。原資料が添えられた ASCII 転写は, トゥルク大学の側で比較的最近入手した言語資料であるが, これまでトゥルク大学のマリ語コーパスとして利用されてきた言語資料については, ASCII 転写のみが提供された。

MS Word 文書になっている原資料は, マリ語の新聞や雑誌の原稿として作成されたものを, 編集部または著者から買い取ったもののようである。マリ語の文章語の表記に用いられるキリル文字アルファベットには, ロシア語のアルファベットにはない特殊文字 (äöÿЫи) が含まれる。マリ語の特殊文字の画面表示・印刷には, Unicode 対応でない独自のキリル文字フォントを使うことがマリ人たちの間で行われてきた。このため, そのフォントがインストールされていない Windows マシンで原資料を読み込むと, 資料 3 のように, 特殊文字が別のキリル文字や記号に化けてしまう (例: тÿ%o [тÿн], йєратыме [йöратыме]) 。

原資料をフィンランド、トゥルク大学のマリ語コーパスで使ってきた ASCII 転写に変換すると、資料4のようになるが、これは人間には非常に読みにくい。資料5は、同じテキストを Unicode (UTF-8) でエンコードし直し、well-formed な XML 文書としてマークアップしたものである。(1) ASCII 転写だけのものは、ASCII 転写から Unicode に変換する、(b) 原資料があるものは、まず MS Word 文書をテキストファイルに変換し、特殊文字を Unicode の文字に置き換える、(c) すべての言語資料を XML 文書としてマークアップする、この3つの作業が共同プロジェクトにおける筆者の役割である。原資料から Unicode 変換されたテキストは、トゥルク大学にいるマリ人の客員研究員が、母語話者として校正した。

共同プロジェクトの対象となったすべてのファイル、言語資料は Unicode に変換された。このうち、校正がひとり済んだのは、次の2つの言語資料である。なお、原資料のほんの一部だが、ファイルが破損していてテキストが取り出せなかったため破棄されたものがある。

Ончыко	1996～1999	月刊誌	32.0 MB (224 万語)
Марий Эл	1990 年代	新聞	4.5 MB (XML タグなし)

マリ語のコーパスについては、松村 (2006) [参考文献 2] でよりくわしく論じてある。

### 3.3. 19 世紀エストニア語のテキストのマイクロ化と電子画像化

タリン大学学術図書館 (旧エストニア科学アカデミー図書館) 所蔵の 19 世紀のエストニア語の新聞記事、学校用読本などをマイクロ化するとともに、デジタルスキャニングを行って TIF 画像データ化した。同時に閲覧が容易になるように PDF 版も作成した。

文献名	年代	ジャンル
Perno Postimees	1857～1863	新聞
Jakobson: Kooli Lugesemise raamat I	1867	学校読本
Jakobson: Kooli Lugesemise raamat II	1875	学校読本
Jakobson: Kooli Lugesemise raamat III	1876	学校読本
Sakala	1878～1880	新聞
Hermann: Eesti Keele Grammatik	1884	文法書
Hermann: Eesti Keele Lauseõpetus	1896	文法書
Georg Müller	1891	教会文書

新聞 Perno Postimees (資料6) は、エストニア語の最初の新聞とっていいものだが、19世紀の半ばころに行われた正書法改革以前の表記法 (旧表記) で書かれた言語資料である。Carl Robert Jakobson (1884-1882) の学校読本 (資料8, 資料9) は、新表記で書かれたものとして広く読まれ、20世紀に入ってから読本として用いられた。

新聞 Sakala は、エストニアのタルトにある文学博物館 (Kirjandusmuuseum) から TIF 画像データで寄贈を受けた言語資料のため、この資料に限り、筆者はマイクロ版を保有していない。画像データ化のもとになったマイクロフィルムは、フィンランドのヘルシンキ大学図書館が撮影したものである。

Karl August Hermann (1851-1909) の2つの文法書は、エストニア語で書かれた最初の文法書として有名なものである。それ以前の文法書としてはドイツ語で書かれた Eduard Ahrens (1803-1863) の文法書 (1843, 1853) と Ferdinand Johann Wiedemann (1805-1887) の文法書 (1875) が研究史的に重要である。

最後にあげた Georg Müller (?1570-1608) のテキストは、書かれたのが17世紀初め (1600-1606) であり、厳密には19世紀のエストニア語の研究資料ではないが、編集されて出版されたのが1891年ということで、この表に含めた。

この時代のエストニアの印刷物は、ほとんどがいわゆるドイツ文字 (Fraktur) の活字で印刷されているため、コンピュータによる光学的文字認識 (OCR) による文字資料化の方法をとることは現状では不可能で、文字資料化は、エスト

ニアの学生や研究者の協力による手作業で、少しずつ進めてきた。いまのところ、Perno Postimees 紙の最初の1年分(創刊号～第52号)の記事の入力が済んではいる(資料7)ものの、本格的な文字資料化の作業は、まだ今後の課題となっている。

### 3.4. エストニア語のコーパスの拡充

タルト大学の一般言語学講座のコンピュータ言語学の研究室の研究者 Heiki-Jaan Kaalep 氏, Kadri Muischnek 氏から、エストニア語の新聞記事、文学作品のテキストを新たに入手し、XMLによるマークアップを行い、エストニア語のコーパスに追加した。現在、筆者のもとにあるエストニア語コーパスの構成は次のようになっている。

松村が入力したテキスト	67万語
タルト大学の基本コーパス(1980年代)	101万語
タルト大学の歴史コーパス(1890年代～1970年代)	177万語
タルト大学の1990年代のコーパス	199万語
新聞記事コーパス(新たに追加されたもの)	3214万語
文学作品コーパス(新たに追加されたもの)	404万語

タルト大学のコーパスについては、公式サイトを参照されたい。

<http://www.cl.ut.ee/korpused/index.php?lang=en>

### 3.5. その他のバルト・フィン少数言語の言語資料

1999年11月～2001年3月の期間に、エストニアの研究者の協力で、カレリア語ジョルジャ方言、ベプス語、イジョール語、ボート語、エストニア語ヴォル方言、エストニア語セト方言の録音資料(ベプス語のみ約3時間、他の言語は約10時間)を入手した。その録音資料は、エストニア人たちによって大部分がすでに音声記号で転写されて、文字資料化されている。その文字化されてい

た言語資料を、もう一度整形し、XML でマークアップし直した。なお、ボート語だけは、この資料に加えて、印刷されたボート語テキスト集から電子化した言語資料があるが、それも一緒に XML 文書に変換した。

### 3.6. サミー語の新聞記事資料のマイクロ化と電子画像化

サーミ語 (Sami) は、ノルウェー、スウェーデン、フィンランド3国の隣接する地域、およびロシアのコラ半島で話されているウラル系の言語で、話者はおよそ3万5千人と推定されている。方言の差が大きく、それぞれの方言を独立した言語とみなして「サーミ諸語」という言い方がされることが多い。もっとも話者の多い北サーミ語 (30000人) の文章語は、ノルウェー、フィンランド、スウェーデンにまたがる比較的広い範囲で用いられる。

以前バックナンバーを購入してあったノルウェーで発行されている北サーミ語の新聞2紙 (Áššu 紙, Min Áigi 紙) の 1998年1月～2001年12月までの4年間の紙面をマイクロ化するとともに、デジタルスキャニングを行って TIF 画像データ化した。閲覧が容易になるように PDF 版も作成した。

新聞名	データ	マイクロフィルム	TIFデータ	PDF
Áššu	1998～2001	8巻	CD 6枚	CD 6枚
Min Áigi	1998～2001	8巻	CD 6枚	CD 6枚

将来的には、TIF 画像を使って OCR し、文字データ化したものを XML でマークアップすることができるなら理想的である (資料 10 は Áššu 紙の紙面)。

## 4. Perl 言語による検索プログラムの開発

この研究では、文字化された言語資料の整形やマークアップ、コーパスの検索、コーパス検索結果の再処理など、すべての工程で Perl 言語のプログラムを Windows XP 上で走らせて作業を行ってきた。英語のコーパス検索用のプログ

ラムにはすぐれたものがないわけではないが、英語以外の言語のコーパスにはそのまま利用できなかつたりする。また、コーパスとして整備するまでのテキスト加工の段階で必要とされるツールは自前で用意しなければならない。グラフィックなインターフェースには欠けるが、文字データ処理に強い Perl 言語は、こういった場面で、小回りが利くので便利である。

マリ語のコーパスを検索して作った KWIC 索引のサンプルを、資料 11 として掲げる。この KWIC 索引を作る際に使った Perl 言語のプログラムは、コーパスファイルが XML でマークアップされているという特性を利用しないで検索しているものであるが、将来的には、XML のマークアップを利用した検索ができるツールが開発される必要があり、また、そのように期待もされている。

## 5. ウラル音声記号を実装した Unicode フォントの開発

フィンランド内国語研究センター (通称 Kotus, ヘルシンキ) の研究員 K・ルッペル氏 (Klaas Ruppel) の全面協力で作成した OpenType の Unicode フォントである。さまざまな事情で、JLOT-Fluralic という作業名がそのままフォントの名称になった。グリフ作成をはじめ、フォント製作の実際の作業は、フィンランドのフォント業者 (JL-types Ky) に発注して行ったもので、市販のソフト並みの高品位の印刷も可能である。フォントの仕様など、技術的な点については、資料 12 を参照されたい。

このフォントは、無料で配布している。入手を希望する場合は、この論文の著者、松村まで E メールで問い合わせられたい。

## 6. 19 世紀のフィンランド語文法書等のマイクロ化

文字化には至っていないが、今後の電子文書化の可能性を念頭に置きつつ、19 世紀～20 世紀初頭のフィンランド語とエストニア語の文法書を中心とする一連の歴史的な文献を、マイクロフィルム化し、デジタル画像化する作業を行った。デジタル画像 (TIF) をもとに PDF 版も同時に作成したので、マイクロフィ

フィルム・リーダーを用いることなく、PCの画面でこれらの文献を参照することができるほか、必要なページの紙へのコピーも容易になった。

なお、現在のデジタル画像技術は、マイクロフィルム並みの解像度にはるかに及ばないようなので、文献を複製・保存するための方法としてのマイクロ化の需要はここしばらくはなくなるであろう。しかし、マイクロ資料は、閲覧、とくに必要なテキスト部分の検索がたいへんやっかいである。マイクロ資料をデジタル画像経由でPDF文書に変換して、こちらを閲覧に利用すれば、マイクロ資料より格段に便利な利用環境が生まれる。

フィンランド語関連の文献は、フィンランド内国語研究センター(ヘルシンキ)の図書室所蔵のものを現地で一時借り出し、ヘルシンキ大学図書館のマイクロフィルム・センターでマイクロフィッシュ化するという手続きで入手することができた。エストニア語関連の文献は、タリン大学学術図書館所蔵のものを、同図書館のマイクロ化部門担当者に直接依頼してマイクロフィルム化して入手した。どちらの場合も、デジタルスキニング以降の工程は、すべて東京で行った。

マイクロ化とデジタル画像化、PDF文書化を行ったフィンランド語・エストニア語関連の歴史的文献は、以下の一覧表の通りである。

## 6.1. フィンランド語に関する歴史的文献

Judén, Jacob: *Försök till Utredande af Finska Språkets Grammatik*. Wiborg, 1818.

And. Joh. Sjögren: *Ueber die finnische Sprache und ihre Literatur*. St. Petersburg, 1821.

Becker, Reinhold von: *Finsk Grammatik*. Åbo, 1824.

Renvall, Gustavus: *Finsk språklära*. Åbo, 1840.

Koranteri, H. K. *Suomalainen kieli-oppi, koulujen tarpeeksi*. Wiipuri, 1845.

Collan, Fab. *Finsk språklära*. Helsingfors, 1847.

Eurén, Gustaf Erik. *Finsk språklära*. Åbo, 1849.

Eurén, Gustaf Erik. *Finsk språklära i sammandrag*. Åbo, 1851.

Eurén, Gustaf Erik. *Suomalainen kielioppi suomalaisille*. Turku, 1852.

- Korander, H. Konst. *Suomalaisten nimuke- ja lausukka-(eli tehdikkö-)sanojen jakautumisesta taivutuksen suhteen*. Helsingissä, 1859.
- Koskinen, Yrjö. *Finska språkets satslära*. Åbo, 1860.
- Länkelä, Jaakko. *Suomen kielen kielioppi. Alkeisopetuksen tarpeeksi*. Hämeenlinnassa, 1867.
- Jahnsson, A. W. *Finska språkets satslära: för läroverkens behof*. Helsingfors, 1871.
- Setälä, Emil Nestor. *Suomen kielen lauseoppi. Oppikirjan koe*. Helsingissä, 1880.
- Genetz, Arvid. *Suomen kielen äänne- ja muoto-oppi ynnä runousoppi*. Helsingissä, 1881.
- Hämäläinen, Konstantin. *Suomenkielen muoto-oppi*. Helsingissä, 1883.
- Petersen, F. J. *Lauseopillinen esimerkki-kokoelma*. Helsingissä, 1885.
- Mietintö kieliopillisten koulu-oppikirjojen yhtämukaisiksi järjestämisestä*. Helsinki, 1888.
- Kallio, Augusti Herman. *Suomen kielioppi. Alkeisopetusta varten*. Toinen, muutettu painos. Helsingissä, 1896.
- Setälä, E. N. *Suomen kielioppi. Äänne- ja sanaoppi*. Helsingissä, 1898.

## 6.2 エストニア語に関する歴史的文献

- Beiträge zur genauern Kenntniß der esthnischen Sprache*. Herausgegeben von Joh. Heinr. Rosenplänter. 1813-32.
- Eduard Ahrens. *Grammatik der Ehstnischen Sprache Revalschen Dialektes. Erster Theil: Formenlehre*. Reval, 1843.
- Eduard Ahrens. *Grammatik der Ehstnischen Sprache. Erster Theil: Formenlehre*. Zweite umgearbeitete Ausgabe. Reval, 1853.
- Eduard Ahrens. *Grammatik der Ehstnischen Sprache. Zweiter Theil: Satzlehre*. Reval, 1853.
- F. J. Wiedemann. *Grammatik der ehstnischen Sprache*. St.-Pétersbourg, 1875.

## 参考文献

- [1] 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)
- [2] 松村一登 2006 「マリ語の言語資料とその電子化」 *Uralica*, Vol. 14 [印刷中]
- [3] *Национальный состав и владение языками, гражданство. Итоги всероссийской переписи населения 2002 года*. Том 4, книга 1. Москва: «Статистика России», 2004.

## 資料一覧

- 1 ウラル音声記号で転写したベプス語テキストとそのロシア語対訳 (Зайцева и Муллонен 1969)
- 2 資料1と同じベプス語テキストに XML のマークアップを試みたもの
- 3 ロシアで MS Word 文書として作成されたマリ語のテキスト (「原資料」)
- 4 資料3の原資料をトゥルク大学のマリ語コーパスのエンコード方式として使われてきた ASCII の転写方式に変換したもの (「ASCII 転写」)
- 5 資料3の原資料を Unicode (UTF-8) でエンコードし直し, XML 文書としてマークアップしたもの
- 6 1857年創刊のエストニア語の新聞 *Perno Postimees* の創刊号の紙面
- 7 資料6と同じ *Perno Postimees* 紙の記事をマークアップしたもの
- 8 C. R. Jakobson のエストニア語の学校読本第1巻のタイトルページ
- 9 資料8と同じ読本のテキストページ
- 10 ノルウェーで発行されている北サーミ語の新聞 *Áššu* 紙の紙面
- 11 マリ語のコーパスを検索して作った KWIC 索引のサンプル
- 12 OpenType フォント JLOT-Fluralic に関する資料

Макарова Маланья Ивановна, 80 л.  
Mag. зап. 166/4. М. Муллонен, 1961.

### 15. kut kalitkad pašttas

otan jouhod ruģiništ, paņēn padēižēhe vēt, paņēn jouhod, iēcīī skalnu šegotān, stolon tahtān iēcēn pēksan. tegesē kovem-  
beņē, miāna očan, turutan niņgomiid valuskeižid, katkeidan, niņga  
turutan, turutan, kāžiī ģārōtan. šid miāna ajan, tēgen skančaižēn,  
šegotan šūdāmudēn. šūdāmudēn miāna skančaižēhe, skančaižēn umbi-  
štān da i pāche. paņēn kaudatēst, paņēn muņēižēn šūdamehe,  
mišē magedēmb da šūiāna voihtut da kalitkeižēn-sē umbištān da  
pālē-sē kaudatēsut, nu da i pāche i šigā paštub. paštub ka miāna  
vēdan da voidān i šlakēidan sōda. čōma mujub.

### 15. Как калитки пекут

Возьму ржаную муку, налью в горшок воды, положу  
муки, этой скалкой размешаю, на столе мешу это тесто. [Когда]  
станет потверже, я беру, раскагаю такие валики, разломаю их,  
так катаю, катаю, руками расплющу. Потом я раскагаю, оде-  
лаю сканец, размешаю начинку. Начинку я [кладу] на сканец,  
сканец закрою — и в печь. Положу сметаны, положу яичко  
в начинку, чтобы вкуснее [была], да масла туда и закрою ка-  
литку, да сверху-то сметанки, ну да и в печь, и печется там. Ис-  
печется, так я выташу и смажу, и ем да притмокиваю. Вкусной  
кажется.

Иняков Василий, 45 л.  
Mag. зап. 166/3. М. Муллонен, 1961.

### 16. pīha osti bazarou žerklon

eī endē ūks pīha. fēdan, mān hān bazaralē i osti šigā  
žerklon. kodhe tuī, kaslēb žerkloho: «čōma onēn». ak iēcēn homēič,  
dumējib: «minak hān šinna kaslēb?». konz hān ģāks (pīha) kodšipei;  
ak kasob: «a, sanob, lūgūde fēdan, keda kaslēb. hān, sanob,  
bazarou ģūzi iēcēn da sēn kartiņaine om. turēške, sanob, mamoi,  
kacō». mamazē tuī, kacuhī: «ka, sanob, om babka iēcīt, minun  
vuišē». šit tuī tatazē: «ka min tii pagūžēti, om iēcīt mužik,  
bardaņke da furāskois». šid vāhāne da tuī iče hān pīha. he  
kūžētasē: «kedak šina bazarou ģūžid?». — «a kedak om?». — «ka  
mii naku kacumēi, sanotas, da em tea, ken om». a hān šūžuti heit  
kaiktid ūhēš da ozufi: «vot, sanob, tii iče onēžēgi».

### 16. Парень купил на базаре зеркало

Жил раньше один парень. Пошел он на базар и купил  
там зеркало. Пришел домой, посматривает в зеркало: «Красивый  
я». Жена заметила это, думает: «Что он туда посматривает?».  
Когда он ушел из дому, жена смотрит: «А, говорит, генер я знаю,  
да это же карточка. Или-ка, говорит, мама, посмотри». Ее мать  
пришла, посмотрела: «Да это, говорит, бабка, похожая на меня».  
Потом пришел ее отец: «Да что вы говорите, это мужик, с боро-  
дой да в фуражке». Потом через некоторое время пришел сам  
парень. Они спрашивают: «Кого ты на базаре нашел?». — «А  
кого?». — «Да мы тут смотрели, говорят, да не знаем, кто». А он  
поставил их всех вместе и показал: «Вот, говорит, сами вы и есть».

### ПОДВИННИКИ (rihalaine)

Ершова Марфа Ксенофонтовна, 74 г.  
Mag. зап. 162/10. М. Муллонен, 1961.

### 17. voik

poručiba minun vouktan voudežēn-sē,  
krasņijan da krasoleižēn  
rođimijad da rođitelad,  
kaļhēd da kandjeihudēd  
minun vouktan da voudežēn-sē.  
miā krasņimoi i ģikuimoi  
ičiī dubovijou da laveižū-sē  
rođimjoide rođitelnoi  
čōmas da čogeižēs-sē.  
rođimijad tii roditelad,  
kaļhēd da kandjeihudēd,  
kacuhtagatēške minun goŕo-gotkijan pālē-sē.  
poručēičitei vouktan voudežēn ka  
vēhālē vīnulē da randēižēn-sē,  
ģiki edahēižēn-sē ka.  
vēhād vīnud armoižēd-nē ģūšas,  
ii rođimijan roditelan armod-nē ģūgoi,  
ģūšas vēhād vīnud armoižēd-nē,  
ķiven da ķiņčēd-nē,  
roudan da kovučēd-nē.  
sōtēi-mamoihudēm da tatoihudēm,  
kutag miņei ģūnēb ūgu eģigata-sē  
ičēmoi kaļhež da kanzeižēs-sē.  
sōtēi šinā da čūžoihudēm,

ウラル音声記号で転写されたベプス語のテキストに最小限のタグ付けを試みたもの

<div3 id="16">

<informant>Иняков Василий, 15 л.</informant>

<doc\_info>Маг. зап. 166/3. М. Муллонен, 1961.</doc\_info>

<div4 id="16\_1" lang="vep">

<h>16. príha ost'í bazarou źerkлон</h>

<p>

<s no="0671"> e'li ende ük's príha. </s>

<s no="0672"> t'edan, mäń hän bazarale i ost'í śigä źerkлон. </s>

<s no="0673"> kod'he tu'í, kasle'b źerkлоho: «čoma o'leń». </s>

<s no="0674"> ak néceń homęič, dumę ib: «minak hän śinna kasle'b?». </s>

<s no="0675"> konz hän lák's (príha) kod'išpe'i, ak kacob: «a, sanob, nügüde t'edan, keda kasle'b. </s>

<s no="0676"> hän, sanob, bazarou líuži néičen da šeń kart'ingine om. </s>

<s no="0677"> tulęške, sanob, mamoi, kaco». </s>

<s no="0678"> mamaze tu'í, kacuht'i: «ka, sanob, om babka néčit', minun vuit't'e». </s>

<s no="0679"> śit\_tu'í tataze: «ka min t'ii pagižęte'i, om néčit' mužjk, bardańke da furaškoiš». </s>

<s no="0680"> śid vähäine da tu'í iče hän príha. </s>

<s no="0681"> he kúze'ltasęi: «kedak śina bazarou líužid?». </s>

<s no="0682"> – «a kedak om?». </s>

<s no="0683"> – «ka mii naku kacuime'i, sanotas, da em t'ea, k'en om». </s>

<s no="0684"> a hän śiižut'í heit' kaikid ühtę da ozut'í: «vot, sanob, t'ii iče o'le'tęgi». </s>

</p>

</div4>

</div3>

## ПОЭЗИЙЫН ЙЄРАТЫМЕ ЙОЛТАШЫЖЕ

Миклай Казаковын шочмыжлан — 80 ий

Писатель-влакын шочмыштым (творчестве корныш шогалмыштым) ожнырак элыште, обществыште лийше вашталтыш дене, саманын ойыртеме дене кылдат ыле. Тиде шот гыч ончалаш гын, Миклай Казаковын мурпашаже икымше вичияш жапыште (1928—1932 ийлаште) тўзланаш, нерешташ тўоалын. Но ялысе марий рвезын кидышкыже ручкам социализм пашан шарлен толмыжо веле огыл кучыктен. Тудым шке йўк дене мурым ооарен ончаш тунамсе литераторна-влакын — С. Чавайнын, О. Шабдарын, Олык Ипайын, Йыван Кырлян — возымышт таратен. 1930 ий гычак Йыван Миклай "У вий" журнал дене кылым кучаш тўоалеш, тушко "Властьлан ваштареш ит кай" пьесыжым, почеламутлажым колтылеш. Нуныжо лудшо дек миен огыт шу. Тыгай пиал эн ондак "Ленин" почеламутшылан логалеш. Вуймутшо да содержанияйже тыгай огыт лий гын, ты серымыжат, векат, ош тўням ок уж ыле.

"Утларак терлаташ, илышым утларак шынен (шымлен) возаш тўоалат гын, пашашт вийнен кертеш", — вашештен "У вий" самырык авторлан. Тыге каласаш лиеш: келгын шынен возымаш поэтын творчествоштыже эреак тўоо верым налын. 1938 ийыште коло ияш рвезын икымше почеламут ойпогыжо лектеш. Вес ийжылан тудым Писатель ушем радамыш налыт. Тиде жап гыч Миклай Казаков марий поэзийын эн ончыл радамышкыже шогалеш да лўйкален пытарыме мурапшат-влакын кидышт гыч камвочшо сылнымут эстафетым нелталеш, пел курым чоло тудым ўшанлын кучен ошкылеш.

Кумло талук ончыч, пел курымаш юбилейжым палемдыме кечын, Николай Иванович каласен (рушла гыч кусарыдеак кодена): "Поэзия должна быть чистой и согревающей, как горячий источник-гейзер, точной и целеустремленной, как пуля, выпущенная в мишень, красивой, как букеты живых цветов, и звонкой, подобно биению сердца... По мере своих способностей я всегда придерживался этой линии". Возен кодымо кумало утла книга кокла гыч кажныжым лудмо годым ме тушто чон яндарлыкым да шокшо кумылым, серыпле йылмым да раш шонымашым, чонеш перныше моторлыкым, авторын йўлен пырткыше шўмжым шижына.

Чын талант тыматле койышан, шкенжым нелташ шоныдымо, чап да кугу лўмнер верч тыршыдыме лиеш. Казаковым лишыч палыше е%-влак ойлат да возат: 50—60-шо ийлаштат (тунам Сталинский премий лауреат да Марий АССР-ын калык поэтше лўмым налын, сылнымут пашажлан кок орден дене наградитлалтын), варажат кугешнылме, не-

- 1: 12—24
- 2:
- 3: POE1ZIJYN JO1RATYME JOLTAS1YZ1E
- 4:
- 5: Miklaj Kazakovyn s1oc1myz1lan — 80 ij
- 6:
- 7: Pisatelj1 -vlakyn s1oc1mys1tym (tvorc1estve kornys1 s1ogalmys1tym) oz1nyrak e1lys1te, obs2es-  
8: tvys1te lijs1e vas1taltys1 dene, samanyñ ojyrtemz1e dene kyldat yle. Tide s1ot gyc1 on-  
9: c1alas1 gyn, Miklaj Kazakovyn murpas1az1e ikyms1e vic1ia1s1 z1apys1te (1928—1932 ijlas1te)  
10: tu2zlanas1, neres1tas1 tu2n1alyn. No a1lyse marij rvezyn kidys1kyz1e ruc1kam socializm pa-  
11: s1an s1arlen tolmyz1o vele ogyl kuc1ykten. Tudym s1ke ju2k dene murym on1aren onc1as1 tu-  
12: namse literatorna-vlakyn — S.C1avajynyn, O.S1abdaryñ, Olyk Ipajyn, Jyvan Kyrla1n —  
13: vozymys1t taraten. 1930 ij gyc1ak Jyvan Miklaj "U vij" z1urnal dene kylym kuc1as1  
14: tu2n1ales1, tus1ko "Vlastj1lan vas1tares1 it kaj" pj1esyz1ym, poc1elamutlaz1ym koltyles1.  
15: Nunyz1o luds1o dek mien ogyt s1u. Tygaj pial e1n ondak "Lenin" poc1elamuts1ylan loga-  
16: les1. Vujmuts1o da soderz1anijz1e tygaj ogyt lij gyn, ty serymyz1at, vekat, os1 tu2na1m  
17: ok uz1 yle.
- 18: "Utlarak to1rlatas1, ils1ym utlarak s1ynen (s1ymlen) vozas1 tu2n1alat gyn, pas1as1t vij-  
19: nen kertes1", — vas1es1ten "U vij" samyryk avtorlan. Tyge kalasas1 lies1: kelgyn s1y-  
20: nen vozymas1 poe1tyñ tvorc1estvys1tyz1e e1reak tu2n1 verym nalyñ. 1938 ijys1te kolo ia1s1  
21: rvezyn ikyms1e poc1elamut oj pogyz1o lektes1. Ves ijz1ylan tudym Pisatelj1 us1em radamys1  
22: nalyt. Tide z1ap gyc1 Miklaj Kazakov marij poe1zijyn e1n onc1yl radamys1kyz1e s1ogales1  
23: da lu2jkalen pytaryme muraps1at-vlakyn kidys1t gyc1 kamvoc1s1o s1ynymut e1stafetym

<text id="Onchyko97\_163" src="122497">

<ref>12 — 24</ref>

<h>ПОЭЗИЙЫН ЙӖРАТЫМЕ ЙОЛТАШЫЖЕ</h>

<p> Миклай Казаковын шочмыжлан — 80 ий </p>

<bl/>

<p>

<s no="14956"> Писатель-влакын шочмыштым (творчестве корныш шогалмыштым) ожнырак элыште, обществыште лийше вашталтыш дене, саманын ойыртемыже дене кылдат ыле. </s>

<s no="14957"> Тиде шот гыч ончалаш гын, Миклай Казаковын мурпашаже икымше вичияш жапыште (1928 — 1932 ийлаште) тӱзланаш, нерешташ тӱнгалын. </s>

<s no="14958"> Но ялысе марий рвезын кидышкыже ручкам социализм пашан шарлен толмыжо веле огыл кучыктен. </s>

<s no="14959"> Тудым шке йӱк дене мурым онгарен ончаш тунамсе литераторна-влакын — С. Чавайнын, О. Шабдарын, Олык Ипайын, Йыван Кырлян — возымышт таратен. </s>

<s no="14960"> 1930 ий гычак Йыван Миклай "У вий" журнал дене кылым кучаш тӱнгалеш, тушко "Властьлан ваштарешит кай" пьесыжым, почеламутлажым колтылеш. </s>

<s no="14961"> Нуныжо лудшо дек миен огыт шу. </s>

<s no="14962"> Тыгай пиал эн ондак "Ленин" почеламутшылан логалеш. </s>

<s no="14963"> Вуймутшо да содержанияйже тыгай огыт лий гын, ты серымыжат, векат, ош тӱням ок уж ыле. </s>

</p>

<p>

<s no="14964"> "Утларак тӱрлаташ, илышым утларак шынен (шымлен) возаш тӱнгалат гын, пашашт вийнен кертеш", — вашештен "У вий" самырык авторлан. </s>

<s no="14965"> Тыге каласаш лиеш: келгын шынен возымаш поэтын творчествыштыже эреак тӱн верым налын. </s>

<s no="14966"> 1938 ийыште коло ияш рвезын икымше почеламут ойпогыжо лектеш. </s>

<s no="14967"> Вес ийжылан тудым Писатель ушем радамыш налыт. </s>

Seisinnatsel tärkamise ja õitsemisise aial hakkab nüüd ka Postimees omma semet külvi tarvis kandma. Ehk kül temma surem hool peab ollema, sedda kõnneleba, mis igga mehhele mele järrel ja mokka pärast on ning kasso ja keelmatta rõmo sadab: ommeti teab ta ette ärra, et üks igga kord keikide melepärrast ei sa ja pallub sepärrast: võtke kannatuses wasto, mis armastuses pakkutakse! Uus wanker jookseb esmalt ikka nattuke raskest ja Postimees, kuida näete, tulleb jo ka alles esimest korda. Ta tõttab liina poolt, kõrtst mõda kirriko pole, kus igga mees omma lehti otab. Uega mõda harjume mollemad: temma kulutama, teie kuulma. Andko siis Summal teie ja temmale omma rahho ja kallist terwist.

Keist päiwist ja aegust, mis hilja mõda läinud, polle meil paljo tarwis räkida, nad on meil muidogi alles meles. Monne südda tufsub ja kõrwad ugawad weel praego alles sest sõda karrast ja mürrast, mis nüüd tänno Summalale, jubba selja tagga on. — Jah, monnesuggused ja sured asjad on Jssand meile saatnud. Olleme hirmo ja kartust, kurbdust ja ahhastust näinud; olleme ka õnne ja rõmo maitšnud, keik sedda mõda, kuida se meile hea olnud, kui moistime armo andede eest tännada ja karristamisise witsale suud anda. Raskest pãrwad on meil kül olnud: Merri finni, Ma finni, toit kallis, tenistus kašsin, kurwad pãrwad, paljo ahhastust! Ugga kas teate mišk-pãrrast? Keik Summala karristamisise hãrwad: „Põõrge pattust ja parrandage meelt!“ ja keik Temma armo annid maenitšewad: „Tundke ja maitške, kui hea Jssand on ja andke Temmale tänno!“ Ugga nisuggust Summala hirmo- ja armojutlust olleme ka saggedaste kõrwast sisse, teisest wãlja lastnud ja seãlt se tulnud, et armas taewa tabike piddanud ka wahhel kimpõ sidduma ja meile monne nisugguse rapso andma, mis luusse mõus. Ai, ai, se olli kibbe, wãgga kibbe! Jah, sedda ta olli ja piddi ka ollema; agga kibbe wits peab jo head lapsed kaswatama ja seddasamma tahhab ka taewane Jssa, kui omma õllekãtte lãinud lapsi monni kord ras-

tem kipputab. Meid on Temma hilja aego ka nenda karristanud ja Temma kãssi on raske meie peãl olnud; agga Ta on sedda mürris-tamisise pilwe jãlle mõda saatnud ning kallis rahho pãike paistab ja rõmustab meie sũbdant. Meie armolinne Keisri Herra, Summal pittendago temma pãiwi, on jãlle rahho teinud ning werrewallaminne on mõdas ja sõda hirm ja ahhastus lõpnud. Ugga nüüd kũssigem isseennast, kas olleme ka sedda wããrt olnud? kas olleme nõudnud rahhus Summalat tenida, ehk rahhus patto teha? Ma mõtlen, meil on kũllalt wasto rindo lãa ja tõlneriga palluda: „Summal olle meile waeste pattuštele ka eddespiddi armolinne!“ — Andkem siis Summalale tãnno ja kitust, kes ennam on annud ja ikka annab, kui wããrt olleme, pallume ja moistame. On siis rahho meie maal, olgo siis ka rahho maias ja sũddames. Kartkem Summalat, auustagem kunningast ja ellagem nüüd eddespiddi omma kalli Ewangeliumi Lutterusise usso sees nenda, kui wagga ristinnimeste kohhus on! — Ja nüüd rãgime mudest sũndinud asjust.

Keige Wennerigi „Seitungides“ seisab allamal seisaw Keisri Herra kãe kirri ja awwalik kulutus:

„Summala armust

**Meie Aleksander Teine,**

Wennema Keiser, Polama  
kunningas, Somema Suurwũrst,  
n. t. f. n. t. f. n. t. f.

anname keige omma tru allamattele teãda:

Sel 29mal Aprilli k. p. sai Meie armastud abbiķasa, Keisri Proua Maria Aleksandrowna, õhhe poiaga mahha, kel nimme pandi Sergei. Seddasinnast omma keiserlikko suggu kaswamist Summala õnnistusse ueks mãrgiks piddades, mis Meie ja Meie keisriviigi õlle on wãljawallatud, anname sedda rõmust asja keige omma tru allamattele teãda. Meie teame, et nemmad keik Meiega õhhes sũddamelikko palwesid taewa ja Ma Jssanda pole saatwad, et se uestesũndinud woiks hoi-

<h3> Sannumed ommalt maalt. </h3>

<p>

<s> Tänno Jummalale, uus kewwade on jälle kä, päsokestes lendamas, lehhed puus, rohhi on tärkanud, lojuste kellad köllawad waino peäl, karjatsed löwad lulli ja käggo hüab oksa peält ka sekka; igga põllomees on omma semet lotusse peäle põldo külwanud ja kes Jummalat tundis ja kartis, on pärrast wimast äästamist weel mütsi kätte wahhele wötnud ja öhkanud: Issand! nüüd on keik Sinno kääs ja wääs. </s>

<s> Minna kündsin ja külwasin, Sinna kasta ja kaswata; Sulle auuks ja mulle jouuks. </s>

</p>

<p>

<s> Selsinnatsel tärkamisse ja öitsemisses aial hakkab nüüd ka Postimees omma semet külwi tarwis kandma. </s>

<s> Ehk kül temma surem hool peab ollema, sedda kõnneleda, mis igga mehhele mele järrel ja mokka pärrast on ning kasso ja keelmatta römo sadab: ommeti teab ta ette ärra, et üks igga kord keikide melepärrast ei sa ja pallub sepärrast: wötke kannatusses wasto, mis armastusses pakkutakse! </s>

<s> Uus wanker jookseb esmalt ikka nattuke raskestes ja Postimees, kuidas näete, tulleb jo ka alles esimest korda. </s>

<s> Ta töttab linna poolt, kõrtsist möda kirriko pole, kus igga mees omma lehti otab. </s>

<s> Aega möda harjume mollemad: temma kulutama, teie kuulma. </s>

<s> Andko siis Jummal teile ja temmale omma rahho ja kallist terwist. </s>

</p>

<p>

<s> Neist päiwist ja aegust, mis hilja möda läinud, polle meil paljo tarwis räkida, nad on meil muidogi alles meles. </s>

<s> Monne südda tuksub ja körwad ugawad weel praego alles sest söa kärrast ja mürrast, mis nüüd tänno Jummalale, jubba selja tagga on. — </s>

<s> Jah, monnesuggused ja sured asjad on Issand meile saatnud.

</s>



Tartus, 1867.

Trükkitud ja müüa S. Laakmanni juures

saani peavad sinu püüsid trahvits mustad olema, ja ei pea jõest ka millasgi enam wet jooma, vaid wihma-tilkadega rahul olema, teada ja nagu kass hiiresti kinni saab tabama, ja pead üksi siis laulma, kui tõised linnud wihma-tuleku eest peitu lähevad!" — Sest aeast saani olla, Gesti ennemuistese juttu järele, peoleol mustad püüsid jalas, ja ta laulab üksi siis, kui wihm tulemas on. Ka kisenab ta siis sagedaste nagu kass, mikspärast teada rahwa suus ka wihm-kassiks nimetatakse. — Aga ma räägin teile ühe seesuguse peoleo elu-lou otsast otsani ära.

Kord elas metsas üks lind peoleo, kes ilusam oli kui kõik muud linnud. Tema kuub oli nagu selgest kullast ja tema tiivad mustast sametist.

Senni, kui weel lapse-põlmes elas, sööstiwad teada tema wanemad rõõwikute ja färbestega; wahel seta sai ta ka mõne mardika ja pühapääwa praeks ühe suure ämbliku. Döpit, leppalind ja muud metsa-mooselandid on terve eluaeg otsa seesuguse toiduga rahul. N'ad koristawad palju paha loomi ära, on väga tululikud ja igamees armastab neid selle eest.

Meie kulla-watti peoleole ei tahtnud aga seesugune toit kauema aea peale enam mottamööda olla. Kui ta suuremaks sai ja oma eest ise pidi muretsema, mõtles ta: „Kärkse lindsud ja lahjad rõõwikud! need wõiwad küll inetu böpiku ja nälgind linawäästrikule mõnufaks toiduks olla; minusuguse mehele, kes paremaid suu-täisi kätte wõib saada, on n'ad liig waesed!"

Nõnda lendab ta siis aida wislapuu otsa ja hakkab seal hootaanikat studeerima. Kõige paremad marjad otsib ta omale wälja, halvemad jättab ta warblastele. Siis läheb ta reisi peale Itaalia maale, läheb Rooma ja Naapoli linna ja muud kohte waatama, nagu suurte herrade wiis on. Seal riisub ta talupoegade wiinamäed ja atad puhkaks, korjab kõige magusamad marjad wälja ja warastab ka nende küpsaks saanud wiigimarjad nina alt ära. Temast saab nõnda meister warguse ameti peale, ning on seal juures ikka väga etewaatlif, nagu suur frants peab olema. Sgale poole wälguwad tema silmad ja üksigi krabin ei jää t'ast kuulmata. Näeb üht poisikest kwiiga ehk pealegi mõnd meest püssiga tulewat, siis punub ta aegsaste minema ja aia-peremees ei leia muud, kui paljaid warsti ja teri maast.

Aga ka marja-wargale ei jää nuhtlus tulemata. Seal istub ta ühel pääwal puu otsas ja on ütlemata õnnelik nende



Йоңылыш лият, {үдыр}ем. ос6b\_8392  
 Миссә он пикку тўттө ни – кушто {үдыр}ем? (финнла). ос6a\_12663  
 – Нимом удам ыштен отыл, {үдыр}ем! – нюслен пелешткала Тайса. ос6a\_13233  
 ым төрлаш мыйын сомыл огыл, шке {үдыр}ем-влакат илымышт годым койышым онч ос6b\_2151  
 гай илыш пўралтын огыл гын, кеч {үдыр}ем-моторем пиалан лиеш ... ос6a\_13235  
     Мыйым {үдыр}ем-шамыч шинчат, нелеш огыт нал, – ос6a\_3814  
     Кокымшо {үдыр}емже ялыштак, марлан лектын. ос6с\_2115  
 Пўрышын пўрен пуымо шочшемлан – {үдыр}емлан, венгымлан – мужыранашышт пўры ос6a\_12788  
     – А мый шонышым: {үдыр}емым иктаж-могай бусурманлан али ан ос6a\_12837  
 Икмыняр ий гыч изирак {үдыр}емым товар дене руэн пуштыч. ос6b\_2282  
 а: "Тыйын омсалондемет кўкшө, а {үдыр}емын капше изи. ос6b\_2059  
     Кугурак {үдыр}емын, Веруш да Ивук кок пачаш коча- ос6с\_3164  
 н – Мылам чорт посана! – Төршта {үдыр}ет ос6a\_1683  
     Кум гана кыралтат – {үдыр}ет дене "кутыраш" уремыш лектат. ос6b\_3416  
 неш, – тендан ойда почеш Марина {үдыр}ет дене йўдым пырля эртарышна. ос6a\_7050  
     Туде самырык {үдыр}ет огыл: мом шонет, тудым ыштылаш о ос6b\_2806  
     Саде {үдыр}ет телефон деке лап лие, пучым пылы ос6a\_1020  
     – Ок кўл, ок кўл {үдыр}ет! ос6b\_2720  
 Рвезе-влакет дене кузе, ала {үдыр}етат уло? ос6с\_12917  
     Кушто {үдыр}етше? – өрмалген шогальым. ос6a\_4560  
     Йолташ {үдыр}етым ала-кө солалтенат каен ... ос6b\_7994  
 о гын, рвезырак годымак келшыше {үдыр}етым ит мучыштаре. ос6a\_5344

# OpenType font JLOT-Fluralic

**JLOT-Fluralic** is an OpenType font, which has both plain and italic styles and which include among normal character ranges – Latin, IPA, Greek, and Cyrillic – all necessary Uralic Phonetic Alphabet (UPA) characters found in the Unicode version 4.0. Please read more about UPA and Unicode at the article: [www.dkuug.dk/jtc1/sc2/wg2/docs/n2442.pdf](http://www.dkuug.dk/jtc1/sc2/wg2/docs/n2442.pdf). JLOT-Fluralic contains CFF (Compact Font Format) PostScript based glyphs and special glyph substitution tables (GSUB) for high quality typography. JLOT-Fluralic is suited best for publishing purposes and for importing phonetic data from e.g. XLM database into page layout program. A single Unicode character of JLOT-Fluralic can contain even nearly twenty typographic variants.

**Supported operating systems:** Mac OS 8.6 or newer and Windows 95 or newer using Adobe Type Manager (ATM); directly without ATM: Mac OS X and Windows 2000 or newer. Please read more about the OpenType technology and system requirements on <http://www.adobe.com/type/opentype>. There you can also find detailed instructions for OpenType font installation (*OTGuide.pdf*). There is yet no operating system, that supports main OpenType typographic features on a system level. Microsoft uses now OpenType for e.g. Indic script support, but it has announced, that the next main Windows release (“Longhorn”, available maybe 2006 or 2007) will support also OpenType typographic features on system level (<http://longhorn.msdn.microsoft.com>). Apple has not announced, whether it will support OpenType font features on system level. Now it has own type technology, AAT (Apple Advanced Typography), but it is commonly predicted, that also Apple will support natively OpenType features sooner or later. Mac OS 10.4 (“Tiger”), published in spring 2005, has ability to convert some basic OpenType features into AAT features. There is also a font engine project called FreeType for achieving better font support for e.g. Linux environment (<http://freetype.sourceforge.net/freetype2>).

**Supported applications:** All modern applications can recognize installed OpenType fonts, but only applications which can interpret the glyph substitution feature (GSUB) can handle complicated phonetic text properly. Only Redlex’ word processor Mellel and Adobe’s InDesign and CS versions of other Adobe’s application can handle properly typographical features needed for sophisticated UPA output. This is possible, because Redlex and Adobe have made own font engines and these applications don’t rely on - by now inadequate - text services provided by operating systems. We may expect, that core applications on forthcoming Microsoft’s Longhorn will support typographic OpenType features. If you are using JLOT-Fluralic or any other OpenType font on applications, which don’t support typographic OpenType features or even Unicode, you can access only font’s standard characters without automatic typographic rendering and variants.

**Using JLOT-Fluralic in UPA notations.** The Unicode standard requires, that the base character proceed the diacritics. If there are more than one diacritics above or below the base character, the nearest are written first (see The Unicode Standard 4.0; chapter 2.10). JLOT-Fluralic has intelligent glyph substitution instructions, which require 1) that the diacritics below a base characters are

entered first, then the diacritics above a base characters; and 2) that the nearest diacritics to the base character are entered first. For example, if you want to write  $\xi$ , you should enter first the base character U+0073 (small s), then U+031C (first diacritic below), then U+032C (lower diacritic below), then U+030C (caron above) and then U+0301 (acute above). This order is crucial, because the substitution tables inside JLOT-Fluralic require this typing order.

**Notes: 1)** Please note, that the combining acute accent U+0301 is typographically positioned after some base characters (*d f l t β δ*), but it is still a variant of U+0301. It is written in normal order also in these cases; e.g.  $\bar{d}$  is written U+0064 (small d), then U+031C (coronal accent below), then U+0301 (acute above) and last U+0304 (macron above). **2)** If you are writing U+032D and U+0355, you can write them on ambiguous order, because the substitution tables inside JLOT-Fluralic convert these diacritics to U+0356 (e.g.  $\bar{q}$ ) in all cases. Also U+0355 and U+032C on ambiguous order are treated similarly (e.g.  $\bar{q}$ ), although this combination has not own Unicode value. **3)** You need not to use dotless *i* and *j* characters before overlaying diacritics; these will be generated automatically without affecting to proper Unicode values (e.g. *j* + U+0306 will produce  $\check{j}$  without dot).

**The number of supported glyph variants.** One of the most important UPA features is the ability to use more than one diacritics by a base letter. JLOT-Fluralic contain glyph substitution instructions and diacritic variants for three levels above the ordinary base characters (e.g.  $\overset{\sim}{s}$ ) and for two levels below the ordinary base characters (e.g.  $\underset{\sim}{s}$ ; but three levels with double accents  $\underset{\sim}{\underset{\sim}{q}}$ ), two levels above the small superscript characters (e.g.  $n^{\hat{s}}$ ) and two levels below the small superscript characters (e.g.  $p^{\hat{s}}$ ). However, on the third level above the base letter only diacritics denoting to the phonetic length (U+0300, U+0302, U+0304, U+0306, U+0350) and to the palatalization (U+0301). In Uralistic practice this should be enough even for the most complex phonetic notation. Under the small subscript characters (e.g.  $n_i$ ) there is no need for diacritics and these are not supported. You can use basic diacritics also with ordinary *ä, å, ö, and ü* (eg.  $\bar{ä} \grave{ä} \bar{ö} \grave{ö}$ ), because the combinations are needed for some purposes, although not always according to the UPA principles.

**Other typographic features.** Standard Russian alphabet does not use accented vowels for denoting the stress. Also the Unicode standard doesn't contain combined glyphs for these. Using combinations of base vowel + combining acute (U+0301) produce normally typographically poor results (like:  $\bar{А}\bar{а}\bar{Е}\bar{е}\bar{И}\bar{и}\bar{О}\bar{о}\bar{У}\bar{у}\bar{Ы}\bar{ы}\bar{Э}\bar{э}\bar{Ю}\bar{ю}\bar{Я}\bar{я}$ ), but JLOT-Fluralic has an OpenType feature for proper acute positioning for all Russian accented vowels:  $\bar{А}\bar{а}\bar{Е}\bar{е}\bar{И}\bar{и}\bar{О}\bar{о}\bar{У}\bar{у}\bar{Ы}\bar{ы}\bar{Э}\bar{э}\bar{Ю}\bar{ю}\bar{Я}\bar{я}$ .

**Acknowledgments.** JL-types Ky created JLOT-Fluralic for the Department of Asian and Pacific Linguistics of the University of Tokyo as a part of project JSPS Grant-in-Aid for Scientific Research (A) #15202008 (project leader: professor Kazuto Matsumura). Klaas Ruppel's (Research Institute for the Languages of Finland) contribution on the UPA font project has been indispensable. JL-types Ky is responsible of all possible failures and mistakes in JLOT-Fluralic. One main obstacle for broader testing in practise is the lack of Unicode encoded phonetic text materials. It's completely possible, that the font's character repertoire or typographical details need to be improved in the

future. It's even obvious, that Unicode version 4.1 does not contain all possible UPA characters.

**Text examples.** On the next page there are some of the most complicated UPA examples, that can be found. Examples are found from e.g. Lagercrantz (MSFOu 112), Kannisto (MSFOu 101), Karjalainen (MSFOu 157), Uotila (MSFOu 186), and Lehtisalo (MSFOu 90) and from Unicode working document N2419. Because UPA is used always in italic, there are no example texts on Roman (upright) face of JLOT-Fluralic, although it would be technically and typographically possible.

## Some examples in JLOT-Fluralic

*nŷo·rrtl'ł'ε eddēn nēren addžē sud<sup>š</sup>χβō'ł'Γ γāsäχ<sub>φ</sub> nārūjz t'almèn t'ala,tšat uăxtâōjz  
jill'ibq,t'ı'ı' χăëβ' χōtš' χun,tšē·rkkart<sup>o</sup> öη,kè·l' öη,kje·l' à.η,kpl klož'nt'tâ  
knaššât puškùd\_ēdriksâbât taràs p<sup>#</sup>rid'ku hyl<sup>u</sup>kemε ful<sup>#</sup>k.ijT fŷö·L,kie kor'čie kořĥčie  
mâ·škam mο·škam p:zn<sup>d</sup>z:zn p:ñ<sup>d</sup>z:zn pü·ndzο p:ñ<sup>d</sup>žö pü·ndžâ ηατηοδà ατηοδὰς tānōrη<sup>θ</sup>  
jυρo,kbtāš d'ΞΛoknuη tōηηātās*

*nŷo·rrtl'ł'ε eddēn nēren addžē sud<sup>š</sup>χβō'ł'Γ γāsäχ<sub>φ</sub> nārūjz  
t'almèn t'ala,tšat uăxtâōjz jill'ibq,t'ı'ı' χăëβ' χōtš' χun,tšē·rkkart<sup>o</sup>  
öη,kè·l' öη,kje·l' à.η,kpl klož'nt'tâ  
knaššât puškùd\_ēdriksâbât taràs p<sup>#</sup>rid'ku hyl<sup>u</sup>kemε ful<sup>#</sup>k.ijT  
fŷö·L,kie kor'čie kořĥčie mâ·škam mο·škam p:zn<sup>d</sup>z:zn p:ñ<sup>d</sup>z:zn  
pü·ndzο p:ñ<sup>d</sup>žö pü·ndžâ ηατηοδà ατηοδὰς tānōrη<sup>θ</sup> jυρo,kbtāš  
d'ΞΛoknuη tōηηātās*



© JL-types Ky and the Department of Asian and Pacific Linguistics of the University of Tokyo 2003-2005.  
The Department of Asian and Pacific Linguistics of the University of Tokyo has rights to distribute JLOT-Fluralic.