

# Unicode Tools for Field Linguists

## A Brief Introduction

The accompanying CD-ROM contains (i) a Unicode TrueType font (**LangSerif**), (ii) an input tool for Unicode characters (**UniInput**), and (iii) a dictionary tool (**Dictool**) with a 1000-word comparative word list of five Chukchi-Kamchatkan languages encoded in UTF-8. The development of software was made possible by Grant-in-Aid for Scientific Research on Priority Areas (A) from Japan's Ministry of Education, Science, Sports and Culture for the project "Computer Corpora and Corpus Linguistic Tools for the Study of Endangered Languages" (#12039213).

The tools, UniInput and Dictool, are Tcl scripts. The instructions for software installation below are for those who work with Windows 2000/XP and have Tcl installed on their PC. Those who haven't are advised to ask their knowledgeable colleagues for advice or contact the authors at [ono\\_suz\\_mac@kmatsum.info](mailto:ono_suz_mac@kmatsum.info). The easiest way to introduce Tcl into your PC is download a distribution of Tcl called **ActiveTcl** from <http://www.activestate.com/Products/ActiveTcl/>.

Unless otherwise indicated, the files mentioned below are all located in the directory <source>.

### Installation of LangSerif (font)

Open [Control Panel] and click on [Fonts]. Drag **LangSerif.ttf** from the CD-ROM to the [Fonts] folder and drop it there. Once installed, this Unicode font enables you to use the IPA phonetic symbols used in the phonetic transcription of the Chukchi-Kamchatkan languages as well as the special Cyrillic characters used in the orthographies of these languages. One should note that two Cyrillic characters (ӡ, ӣ) used in Itelmen are yet to be defined in the Unicode Standard, so that they are placed in the Private Use Area (U+E000 and U+E001 respectively) in LangSerif.

### Installation of UniInput (input tool)

Create a directory (folder) for UniInput. Copy **uniinput\_1.0.tcl** and **uniinput\_config.dat** into that directory. Double-click on **uniinput\_1.0.tcl** to start UniInput.

English is the default language for displaying the menus, i.e., the content of the configuration file, **uniinput\_config.dat**, is a clone of **uniinput\_config\_en.dat**. If you prefer the menus to be displayed in Russian or Japanese, you may delete the default configuration file, copy **uniinput\_config\_ru.dat** for Russian or **uniinput\_config\_jp.dat** for Japanese from the CD-ROM into the UniInput directory, and rename it as **uniinput\_config.dat**.

### Installation of Dictool (dictionary tool)

Create a directory (folder) for Dictool. Copy the following four files into that directory: **dictool\_1.0.tcl**, **balloon.tcl**, **dictool\_config.dat**, and **goishuu.txt**. Double-click on **dictool\_1.0.tcl** to start Dictool.

The menus are in English by default, i.e., the content of the configuration file, **dictool\_config.dat**, is the same as the file **dictool\_config\_en.dat**. If you wish the menus to be displayed in Russian or Japanese, you may delete the default configuration file, copy **dictool\_config\_ru.dat** for Russian or **dictool\_config\_jp.dat** for Japanese into the Dictool directory from the CD-ROM, and rename it as **dictool\_config.dat**.

The Chukchi-Kamchatkan vocabulary data, **goishuu.txt** is an electronic version of the linguistic material published in the following work:

Megumi KUREBITO, ed., *Comparative Basic Vocabulary of the Chukchee-Kamchatkan Language Family* (Сравнительный базовый словарь языков чукотко-камчатской языковой семьи): 1 (ELPR Publications Series A2-011), 2001, Osaka, Japan.

### Software License

The use of the font, **LangSerif**, as well as the Tcl scripts, **uniinput\_1.0.tcl**, **dictool\_1.0.tcl** and **balloon.tcl**, is subject to the conditions known as GNU General Public License (GPL) (cf. <http://www.gnu.org/copyleft/gpl.html>).

For the use of the vocabulary data, **goishuu.txt**, consult its 'readme' file, **readme\_goishuu.txt**, in the CD-ROM.

### Inquiries

Inquiries about the software in this CD-ROM should be addressed to:  
**ono\_suz\_mac@kmatsum.info**

instead of the individual authors' private e-mail addresses. The scripts, the data and the font may be updated from time to time. The latest version of the software and vocabulary data will be available from:

<http://www2.create.human.nagoya-u.ac.jp/~reiji/ufiles/>

### About the authors

Kazuto MATSUMURA  
Project leader

Professor  
Graduate School of Humanities and Sociology  
University of Tokyo  
[kmatsum@tooyoo.l.u-tokyo.ac.jp](mailto:kmatsum@tooyoo.l.u-tokyo.ac.jp)

Reiji SUZUKI

Graduate student  
Graduate School of Human Informatics  
Nagoya University  
[reiji@info.human.nagoya-u.ac.jp](mailto:reiji@info.human.nagoya-u.ac.jp)

Chikako ONO

JSPS Research Fellow  
Graduate School of Humanities and Sociology  
University of Tokyo  
[chono@tooyoo.l.u-tokyo.ac.jp](mailto:chono@tooyoo.l.u-tokyo.ac.jp)