Trollet: A Tool for Multilingual Grammar Development

Pavel Mihaylov <bin@bash.info>

14th June 2006

NTNU
Innovation and Creativity

Introduction

- Most of the grammar development software originated in the English-speaking world

  – often lacks essential support for scripts other than basic Latin
  – antiquated font handling or support for legacy encodings only

- Such issues prevent many people from using native scripts in their grammars.

- Addressing these problems in the grammar development platform LKB and finding a solution

NTNU
Innovation and Creativity

The problems in LKB

- Supports well only West European languages

- Limited support for displaying a few other scripts

- The developers focus on other issues and will not invest more time in improving the multilingual support

- Uses Unicode internally but talks to the outside world through a Motif-based interface

  – Very limited Unicode support (e.g. no input)
  – Bad font support (ugly or no rendering of many scripts)

How LKB works
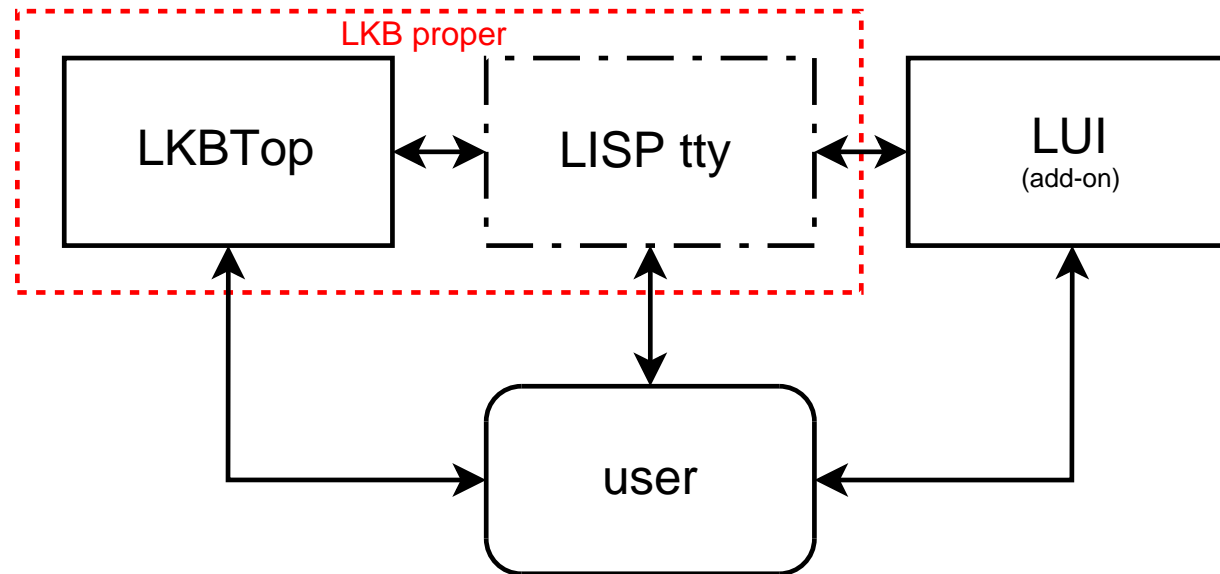
- LKB is written in LISP

- LUI was written later in C with the intention to provide better font support

A closer look at LUI

- Displays trees and AVMs

- Written in C using standard X11 font rendering

  – X11 font rendering is obsolete
  – can be ported to another font renderer (e.g. Pango)

- Has some issues with UTF-8

Porting LUI to Pango

- Pango is the font shaper and renderer behind GNOME

  - can display most scripts in use today, including many historical ones
  - provides clear antialiased rendering

- Porting problems

  - different API
  - different font handling (e.g. physical vs. logical approach)

Font display before and after
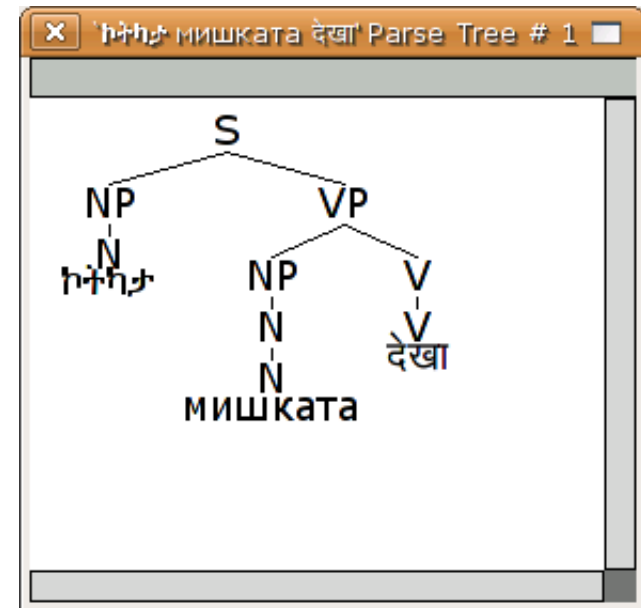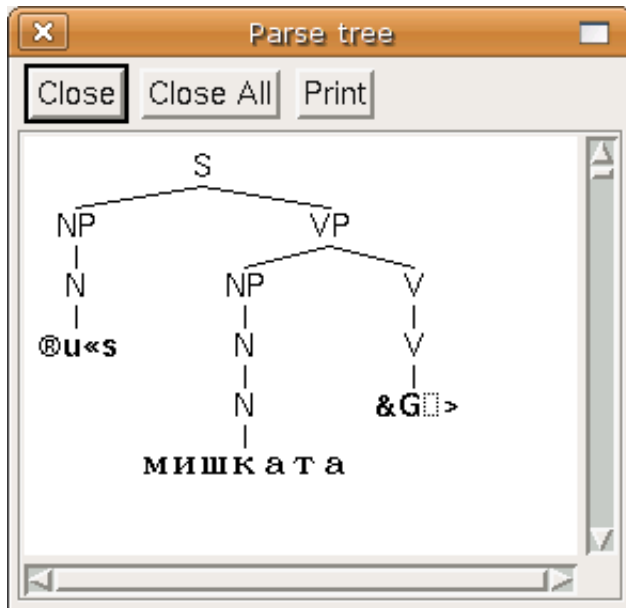
Unicode input

- LKB Top is written in LISP using Motif

  - cannot be made to support Unicode
 - parsing Unicode strings is possible only through calling a LISP function via the tty

Solution: Make another interface and ignore LKB Top

- Provide the same functionality as LKB Top

- Make this interface communicate with LKB by calling LISP functions via the tty

TROndheim LingLab Engineering Tool (Trollet)

- Written in Perl using GTK+ 2

  – excellent Unicode support, including input
  – easy to extend
  – talks to LKB through the LISP tty

- Some LKB functions are rewritten

  – use a Unix domain socket to talk to Trollet for all GUI matters

Trollet + LKB
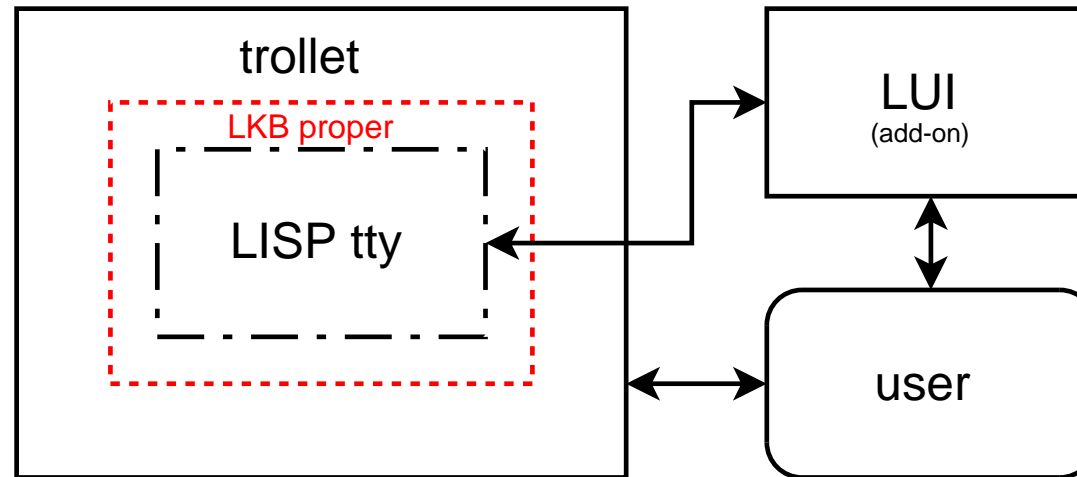
- LKB is embedded in Trollet and mostly invisible for the user

  – no need for running LKB in a terminal or Emacs

- Trollet has replaced the LKB Top window

NTNU
Innovation and Creativity

9

References

1. LKB: `http://www.delph-in.net/lkb/`

2. The Unicode standard: `http://www.unicode.org/`

3. Motif: `http://www.opengroup.org/motif/`

4. Pango: `http://www.pango.org/`

5. Perl and GTK+ 2.x: `http://gtk2-perl.sourceforge.net/`