

The luavlna package

Michal Hoftich (michal.h21@gmail.com)

April 14, 2014

Contents

1	Introduction	1
2	Usage	2
3	Commands	3
3.1	Turning off language switching	4
3.2	Turning off processing	4
3.3	Debugging commands	4
4	Lua module langno.lua	4
4.1	Recognized languages	5
4.1.1	lua _{TeX} and lua _{LaTeX}	5
4.1.2	CSplain	5

1 Introduction

This is a small package for plain lua_{TeX} and lua_{LaTeX}. In some languages, like Czech or Polish, there should be no single letter words at the line end, according to the typographical norms. There exists some external commands (like `vlna`) or packages (`encxvlna` for `encTeX`, `xevlna` for `XeTeX`, `impnatty` for `luaLaTeX`).

Other feature of this package is including of non-breakable space after initials, like in personal names.

The code is modified version of Patrick Gundlach's answer on TeX.sx¹. The difference is that it is possible to specify which single letters should be taken into account for different languages.

2 Usage

The usage is simple:

```
\input ucode
\uselanguage{czech}
% in the case of luacsplain, use instead:
% \chyp
% but language code for Czech is different than in LaTeX or normal
% luaTeX, so you will need to set single letters with somethinh like:
% \singlechars{5}{AlIiVvOoUuSsZzKk}
\input luavlna
\preventsingleddebugon
\input luaotfload.sty
\font\hello={name:Linux Libertine 0:+rlig;+clig;+liga;+tlig} at 12pt
\hsize=3in
\hello
Přiliš žluťoučký kůň úpěl ďábelské ódy.
Text s krátkými souhláskami a samohláskami i dalšími jevy
z nabídky možností (v textu možnými).
```

I začátek odstavce je třeba řešit, i když výskyt zalomení není pravděpodobný.

Co třeba í znaky š diakritikou?

Různé možnosti [v závorkách <i jiných znacích

Podpora iniciál: G. F. Hegel, Č. Zíbrt

```
\preventsingleddebugoff
\bye
```

¹<http://tex.stackexchange.com/a/28128/2891>

Příliš *žlutoučký kůň* úpěl ďábelské ódy. Text s krátkými souhláskami a samohláskami i dalšími jevy z nabídky možností (v textu možnými).

I začátek odstavce je třeba řešit, i když výskyt zalomení není pravděpodobný.

Co třeba í znaky š diakritikou?

Různé možnosti [v závorkách <i>jiných značích

Podpora iniciál: G.F.Hegel, Č.Zíbrt

It is also possible to use the package with \LaTeX , just use

```
\usepackage{luavlna}
```

in the preamble.

3 Commands

```
\singlechars{<language name>}{<letters>}
```

Enable this feature for certain letters in given language.

Default values:

```
% only Czech and Slovak are supported out of the box
```

```
\singlechars{czech}{AIiVvOoUuSsZzKk}
```

```
\singlechars{slovak}{AIiVvOoUuSsZzKk}
```

```
\compoundinitials{<language name>}{<compounds>}
```

Declare compound letters for given language. Second argument should be comma separated list of compound letters, in exact form in which they can appear.

Default values:

```
\compoundinitials{czech}{Ch,CH}
```

3.1 Turning off language switching

By default, language of the nodes is taken into account. If you want to use settings for one language for a whole document, you can use following command:

```
\preventsinglelang{<language name>}
```

3.2 Turning off processing

If you want to stop processing of the spaces in the document you can use command

```
\preventsingleoff
```

To resume processing, use

```
\preventsingleon
```

3.3 Debugging commands

```
\preventsingledebugon
```

```
\preventsingledebugoff
```

Insert debugging marks on/off. Default off.

4 Lua module langno.lua

When we process glyph nodes with lua_{TeX} callbacks, there are `lang` fields available. These are numerical codes of languages, but no information about language names easily accesible from lua side is available.² These numbers are format dependent, majority of formats like lua_{LaTeX} use `language.dat` file provided by `babel` for assign numbers to languages, but for example `csplain` use its own system.

To allow easy setting of language dependent parameters, `langno` module was created. It's purpose is to translate language code to language name and the other way. lua_{LaTeX}, lua_{TeX} and CSplain formats are supported at the moment.

²Language names are stored in _{TeX}macros like `\l@langname`, but different formats use different naming of these macros

4.1 Recognized languages

4.1.1 `luaTeX` and `lua®TeX`

File `language.dat` is processed to load language names, aliases and assigned numbers. These language names are the same as supported by `babel` package.

UKenglish, USenglish, afrikaans, american, amharic, ancientgreek, arabic, armenian, assamese, basque, bengali, bokmal, british, bulgarian, catalan, coptic, croatian, czech, danish, dumylang, dutch, english, espanol, esperanto, estonian, ethiopic, farsi, finnish, francais, french, friulan, galician, geez, german, greek, gujarati, hindi, hungarian, ibycus, icelandic, indonesian, interlingua, irish, italian, kannada, kurmanji, latin, latvian, lithuanian, malayalam, marathi, mongolian, mongolianlmc, monogreek, ngerman, nohyphenation, norsk, norwegian, nynorsk, oriya, panjabi, patois, persian, piedmontese, pinyin, polish, polygreek, portuges, portuguese, romanian, romansh, russian, sanskrit, serbian, serbianc, slovak, slovene, slovenian, spanish, swedish, swissgerman, tamil, telugu, turkish, turkmen, ukenglish, ukrainian, uppersorbian, usenglish, usenglish-max, welsh

4.1.2 `CSplain`

Different method is used. File `hyphen.lan` is included in `CSplain`, where language numbers are assigned to ISO-639-1 or ISO-639-2 language codes. These language codes were then normalized to names used by `babel`, or standard English language names.

afrikaans, armenian, assamese, basque, bokmål, catalan, chinese, coptic, croatian, czech, danish, dutch, english, estonian, finnish, french, galician, german, greek, hindi, icelandic, indonesian, interlingua, irish, italian, kannada, kurdish, lao, latin, latvian, lithuanian, magyar, malayalam, marathi, mongolian, norsk, oriya, panjabi, polish, portuges, romanian, russian, sanskrit, slovak, slovenian, spanish, tamil, telugu, turkish, turkmen, ukrainian, upper-sorbian, welsh