

# The luavlna package

Michal Hoftich (michal.h21@gmail.com)

Miro Hrončok (miro@hroncok.cz)

Version v0.1d

2019-10-22

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Usage</b>	<b>2</b>
<b>3</b>	<b>Commands</b>	<b>3</b>
3.1	Split hyphens . . . . .	4
3.2	Turning off language switching . . . . .	4
3.3	Turning off processing . . . . .	5
3.4	Debugging commands . . . . .	5
<b>4</b>	<b>Lua module luavlna-langno.lua</b>	<b>5</b>
4.1	Recognized languages . . . . .	6
4.1.1	Lua <sub>TeX</sub> and Lua <sup>Λ</sup> <sub>TeX</sub> . . . . .	6
4.1.2	CSplain . . . . .	6
<b>5</b>	<b>License</b>	<b>7</b>

## 1 Introduction

This is a small package for plain Lua<sub>TeX</sub> and Lua<sup>Λ</sup><sub>TeX</sub>. 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`, `impnattypo` for `LuaHTeX`).

Another feature is inclusion of a non-breakable space after initials, like in personal names, after or before academic degrees and between numbers and units (SI and others).

The code is modified version of Patrick Gundlach's answer on `TeX.sx`<sup>1</sup>. 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:
% \chyph
\input luavlna
\preventsingledbugon
\input luaotfload.sty
\font\hello={name:Linux Libertine 0:+rlig;+clig;+liga;+tlig} at 12pt
\hsize=3in
\hello
Příliš 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 a titulů: M. J. Hegel, Ing. Běháková, Ph.D., Ž. Zíbrt.

Podpora jednotek: 100,5 MN\cdot{s}, 100.5 kJ, 200 μA, \$-1\$ dag,

---

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

12MiB, 1m\$^3\$/s.

Uvnitř matematiky by mělo být zpracování vypnuté:  $k \in \mathbb{N}$ .  
Pokračujeme v příkladu.

```
\preventsinglededebugoff
\bye
```

Příliš *žluťoučkový 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 zna-  
cích

Podpora iniciál a titulů: M. J. Hegel,  
Ing. Běháková, Ph.D., Ž. Zíbrt.

Podpora jednotek: 100,5MN·s, 100.5kJ,  
200μA, −1dag, 12MiB, 1m<sup>3</sup>/s.

Uvnitř matematiky by mělo být zpracování  
vypnuté:  $k \in \mathbb{N}$ . Pokračujeme v příkladu.

It is also possible to use the package with Lua<sup>La</sup>T<sub>E</sub>X, just use

```
\usepackage{luavlina}
```

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 Split hyphens

The hyphenated words like “je-li” should be be hyphenated as “je-/-li” according to the Czech typesetting rules. This behaviour can be enabled or disabled for a particular language using the following commands:

`\enablesplithyphens{<language name>}`

`\disablesplithyphens{<language name>}`

By default, it is enabled for the Czech language.

Example in action:

Sedlec-Prčice, modro-zelený, překladatel-  
-tlumočník, kuchař-číšník, propan-butan,  
Otýlie Sklenářová-Malá, František Jílek-  
-Oberpfalcer.

The similar behaviour can be achieved using the `\splithyphens` command provided by Czech language definitions for the Babel package, but it’s use is discouraged. It prevents use of the hyphen character in many situations. The solution provided by Luavlna should be safe.

### 3.2 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>}`

### 3.3 Turning off processing

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

```
\preventsingloff
```

To resume processing, use

```
\preventsingleon
```

You can also disable units and degrees processing:

```
\nounits
```

```
\noprededegrees
```

```
\nosufdegrees
```

### 3.4 Debugging commands

```
\preventsingleddebugon
```

```
\preventsingleddebugoff
```

Insert debugging marks on/off. Default off.

## 4 Lua module `luavlna-langno.lua`

When we process glyph nodes with Lua<sub>T</sub><sub>E</sub><sub>X</sub> 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.<sup>2</sup> These numbers are format dependent, majority of formats like Lua<sub>E</sub><sub>T</sub><sub>E</sub><sub>X</sub> 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<sub>E</sub><sub>T</sub><sub>E</sub><sub>X</sub>, Lua<sub>T</sub><sub>E</sub><sub>X</sub> and CSplain formats are supported at the moment.

---

<sup>2</sup>Language names are stored in <sub>T</sub><sub>E</sub><sub>X</sub>macros like `\l@langname`, but different formats use different naming of these macros

## 4.1 Recognized languages

### 4.1.1 Lua $\TeX$ and Lua $\LaTeX$

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

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

### 4.1.2 CSplain

Different method is used with the CSplain format. The `lua-hyphen.lan` file is parsed for the language codes.

afrikaans, ancientgreek, armenian, assamese, basque, bokmal, catalan, classiclatin, coptic, croatian, czech, danish, dutch, estoniak, finnish, french, galician, greek, hindi, hungarian, icelandic, indonesian, interlingua, irish, italian, kannada, kurmanji, latin, latvian, lithuanian, liturgicallatin, malayalam, marathi, mongolian, monogreek, ngerman, nynorsk, oriya, panjabi, pinyin, polish, portuguese, romanian, russian, sanskrit, slovak, slovenian, spanish, tamil, telugu, turkish, turkmen, ukenglish, ukrainian, uppersorbian, usenglishmax, welsh

## **5 License**

Permission is granted to copy, distribute and/or modify this software under the terms of the LaTeX Project Public License, version 1.3.