The `luavlna` package

Michal Hoftich (`michal.h21@gmail.com`)
Miro Hrončok (`miro@hroncok.cz`)

Version v0.1e
2019-11-06

Contents

1 Introduction **1**

2 Usage **2**

3 Package options **3**

4 Commands **4**
4.1 Split hyphens 4
4.2 Turning off language switching 5
4.3 Turning off processing 5
4.4 Debugging commands 5

5 Lua module `luavlna-langno.lua` **6**
5.1 Recognized languages 6
5.1.1 LuaTeX and LuaLaTeX 6
5.1.2 CSplain 7

6 License **7**

1 Introduction

This is a small package for plain LuaTeXand LuaLaTeX. 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 LuaLaTEX).

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[1]. 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
\preventsingledebugon
\input luaotfload.sty
\font\hello={name:Linux Libertine O:+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.
```

[1] `http://tex.stackexchange.com/a/28128/2891`

```
Podpora jednotek: 100,5 MN\cdot{}s, 100.5 kJ, 200 µA, $-1$ dag,
12 MiB, 1 m$^3$/s.

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

\preventsingledebugoff
\bye
```

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 značích
Podpora iniciál a titulů: M. J. Hegel, Ing. Běháková, Ph.D., Ž. Zíbrt.
Podpora jednotek: 100,5 MN·s, 100.5 kJ, 200 µA, −1 dag, 12 MiB, 1 m^3/s.
Uvnitř matematiky by mělo být zpracování vypnuté: $k \in \mathbb{N}$. Pokračujeme v příkladu.

It is also possible to use the package with LuaLaTeX, just use

```
\usepackage{luavlna}
```

in the preamble.

3 Package options

The `luavlna` package recognizes the following options

`noprocess` – disable the automatic document processing.

`noinitials` – disable processing of the initials.

`nounits` – disable processing of the SI units.

nopredegrees – disable processing of the the academic degrees before names.

nosufdegrees – disable processing of the the academic degrees after names.

debug – enable the debugging mode.

defaults – load default settings. Enabled by default.

4 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}
```

4.1 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 langauge definitions for the Babel package, but it's use is discouraged. It prevents use ot the hyphen character in many situations. The solution provided by Luavlna should be safe.

4.2 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*⟩`}`

4.3 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
```

You can also disable initials, units and degrees processing:

```
\noinitials
\nounits
\nopredegrees
\nosufdegrees
```

4.4 Debugging commands

```
\preventsingledebugon
\preventsingledebugoff
```

Insert debugging marks on/off. Default off.

5 Lua module `luavlna-langno.lua`

When we process glyph nodes with LuaTEX 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 LuaLATEXuse `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. LuaLATEX, LuaTEX and CSplain formats are supported at the moment.

5.1 Recognized languages

5.1.1 LuaTEX and LuaLATEX

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,

[2] Language names are stored in TEXmacros like `\l@langname`, but different formats use different naming of these macros

thai, turkish, turkmen, ukenglish, ukrainian, uppersorbian, usenglish, usenglishmax, welsh

5.1.2 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

6 License