

Typesetting Vietnamese with VnTeX (and with the T_EX Gyre fonts too)

Hàn Thế Thành
River Valley Technologies
<http://river-valley.com>

Abstract

This paper attempts to give an overview of the topics related to typesetting Vietnamese with T_EX. There are no detailed instructions, but rather a comprehensive list of relevant issues and resources, so the reader can get an overall feeling of what is possible and where to look for further information for a quick start. A section is dedicated to a frequently asked question, i. e. how to write a few Vietnamese words in a non-Vietnamese document. I also take a close look at the Latin Modern and T_EX Gyre fonts from the perspective of a Vietnamese end user, hoping to provide some useful feedback to the authors of those fonts.

1 A brief introduction to the Vietnamese language

For convenience I repeat here a short introduction to the Vietnamese language from [1] (with small modifications):

Vietnamese is written with Latin letters and additional accents in a system called *Quốc Ngữ* (which can be translated to English as “national language”) developed by the French missionary Alexander de Rhodes. What separates Vietnamese from other languages typeset with Latin characters is that in Vietnamese some letters can have two accents. The total number of accented letters in Vietnamese (including uppercase and lowercase letters) is 134. Table 1 lists all accented lowercase Vietnamese letters.

Vietnamese accents can be divided into three kinds of diacritic marks: tone, vowel and consonant. Table 2 shows them all with examples.

The accents in Vietnamese play a very important role, since they are used to indicate different meanings and pronunciations of words. The same words with different accents can have a totally different meaning (though the pronunciations can sound very similar to a non-Vietnamese).

The vowel marks are more important than tone marks. Letters with different vowel marks are considered as distinguished letters, but a letter with different tone marks is considered as a single letter with different tones. Vowels have more impact on pronunciation than tones.

The official Vietnamese alphabet as taught in school has 29 letters in this collating order:

A Ă Â B C D Đ E Ê G H I K L M N
O Ô Ó P Q R S T U Ú V X Y
a ă â b c d đ e ê g h i k l m n
o ô ó p q r s t u ú v x y

In Vietnamese all words consist of single syllables, so they are often very short; hyphenation is not allowed at all.

The large number of Vietnamese accented letters has caused big problems in the past. Since there are 134 accented letters, we have 6 letters that don't fit within the upper 128 positions of an 8-bit encoding. So people invented various encodings to deal with this problem, using techniques like dropping uppercase letters, or placing some letters in slots 0–31, or removing some ASCII letters. None was perfect, until UTF-8 became widely used. Thanks to UTF-8 support in the recent L^AT_EX kernel, this is no longer a problem.

The way that punctuation marks are typeset is not consistent; sometimes people put a space before them, sometimes not. However, the case without a space before punctuation is dominant.

More information on the Vietnamese alphabet can be found in [9].

2 Introduction to VnTeX

VnTeX is a package to typeset Vietnamese with L^AT_EX and plain T_EX. VnTeX consists of the following:

- Vietnamese fonts;
- L^AT_EX support for Vietnamese: input encoding, font encoding, support for babel, a style file to activate Vietnamese, etc.;
- plain T_EX support;
- comprehensive font samples and test files;
- some brief documentation.

a	á	ạ	à	ả	ã
ă	ắ	ặ	ằ	ẳ	ẵ
â	ấ	ậ	ầ	ẩ	ẫ
e	é	ẹ	è	ẻ	ẽ
ê	ế	ệ	ề	ể	ễ
i	í	ị	ì	ỉ	ĩ
đ					

Table 1: List of all Vietnamese lowercase letters taking accents

<i>Vowel</i>	
breve	lăn
circumflex	tôi
horn	lươn
<i>Tone</i>	
acute	lá
grave	là
hook above	lả
tilde	lã
dot below	lạ
<i>Consonant</i>	
stroke	đi

Table 2: List of all Vietnamese diacritic marks

VnTeX tries to make typesetting Vietnamese as accessible and easy as typesetting English, and also tries to follow common L^AT_EX conventions, in order to minimize the chance of conflicts with other packages. The official L^AT_EX font encoding for Vietnamese is T5, made by Werner Lemberg and Vladimir Volovich. VnTeX supports various input encodings for Vietnamese. The recommended encoding to use with VnTeX is UTF-8. If for some reason UTF-8 cannot be used (for example your T_EX editor doesn't support it), VISCII is the recommended 8-bit encoding.

The first version was released in 2000. The current version is 3.1.5 and was released in January 2007. VnTeX is being maintained by Hàn Thế Thành, Werner Lemberg and Reinhard Kotucha. The official website of VnTeX is [2].

VnTeX is already part of T_EX Live and MiK_TE_X, which frees most users from the need to install it manually.

ConT_EXt also has support for Vietnamese (partly imported from VnTeX) and a number of Vietnamese users.

3 How to typeset just a few Vietnamese words

This section tries to answer the question that has often been asked: *How can I typeset just a few Vietnamese words in my L^AT_EX document, which is written in English (German/Polish/French/...)?*

The answer depends very much on the particular scenario; however, I assume that you are in a hurry, you don't want to bother with issues like how to display and write Vietnamese in your T_EX editor. You have only a few Vietnamese words in your L^AT_EX file and you would like to see them properly displayed

in your final PDF or PostScript file.

- As the very first requirement, you must have some minimal L^AT_EX support for Vietnamese:
 - Check whether you have VnTeX installed. VnTeX is included in t_EX, MiK_TE_X and T_EX Live.
 - If you don't have VnTeX installed and you are using Latin Modern or T_EX Gyre fonts, then you can just download a single file <http://vntex.sf.net/download/vntex-support/t5enc.def> and put it in the directory where your L^AT_EX source is.
 - Or, if you feel that you need VnTeX, or want to give it a try, or you are not using Latin Modern or T_EX Gyre fonts but a font available in VnTeX only, you can try to download and install VnTeX by following the instructions (in English) at [3].
- If all the above fails, try to get help from someone else to solve at least one of those issues.
- Make sure you have the package `fontenc` loaded with T5 encoding. For example, if your document contains European language(s) only, then you should have a line saying


```
\usepackage[T5,T1]{fontenc}
```

 in your preamble.
- Here's an example of how to input Vietnamese words:


```
{\fontencoding{T5}\selectfont
Ti\'\ecircumflex{}ng Vi\d\ecircumflex{}t}
```

 which gives the output `Tiếng Việt`.

5. If the font family you are using doesn't support T5 encoding, you might want to use another family by saying something like

```
\fontencoding{T5}\fontfamily{cmr}
\selectfont
```

Instead of `cmr` (Computer Modern fonts) you can use `lmr` (Latin Modern fonts), or any font family that supports T5 encoding. For a complete list, see [4].

6. Table 3 contains all Vietnamese letters for your reference.
7. If you have quite a lot of Vietnamese words it can be somewhat tedious to translate them to the above form (often called the L^AT_EX Internal Character Representation — LICR). On Windows you can use the package <http://vntex.sourceforge.net/download/vntex-support/tovntex.zip> to translate text in the clipboard from VIQR or UTF-8 to LICR by one key press.

The same (or close) convenience could be made for Unix/Linux users, but at somewhat higher cost due to deficiencies of Unix-like systems. So if you don't use Windows then you are out of luck, sorry. However, if you use Vim, you can still download the package mentioned above, and use the `vim` script inside the zip archive to do the conversion. If you want to make this easier for Unix users then let me know.

In the future, a web page might be set up to provide online conversion from UTF-8 text to LICR.

4 A quick start for VnTeX users

This section describes the very first steps for those who want to use VnTeX. In contrast to the previous section, the requirement here is that you have VnTeX installed, your editor can properly display Vietnamese text and you can input Vietnamese text. If your setup doesn't meet these requirements, here are some quick tips:

1. for Windows users:
 - install MiKTeX;
 - use a text editor that supports UTF-8: TeX-Maker with a Unicode font like Courier New;
 - use the Unikey keyboard driver to input Vietnamese letters.
2. for Unix users:
 - install TeX Live 2007;
 - use a text editor that supports UTF-8: TeX-Maker, Emacs or Kile, with a Unicode font like Courier New;

- use the XUniKey or Xvnc keyboard driver to input Vietnamese letters (this is not needed for Emacs which has its own input methods).

Supposing that you have VnTeX installed and can read/write Vietnamese with your editor, we can continue now.

A minimal example looks as follows:

```
\documentclass{report}
\usepackage[utf8]{vietnam}
\begin{document}
Tiếng Việt
\end{document}
```

If the document contains multiple languages, you can use `babel`:

```
\documentclass{report}
\usepackage[english,vietnam]{babel}
\usepackage[utf8]{inputenc}
\begin{document}
\selectlanguage{english}
English text
\selectlanguage{vietnam}
Tiếng Việt
\end{document}
```

Once you get started with those examples, you might be interested in some practical issues like:

- how to use VnTeX with Scientific WorkPlace,
- how to create PDF that can be searched, cut or pasted,
- how to create PDF with Unicode bookmarks,
- how to use MakeIndex with VnTeX,
- how to convert L^AT_EX to HTML using TeX4ht.

All those issues are solved and described at the VnTeX website [2].

5 A survey of fonts that can be used with VnTeX

VnTeX comes with quite a large set of fonts. Apart from that, many existing fonts also support Vietnamese. Most of them are made by the Polish font experts who are working on the Latin Modern and TeX Gyre font projects.

5.1 Fonts coming with VnTeX

Many freely available TeX fonts have their Vietnamese counterpart coming with VnTeX:

- VNR: based on Computer Modern, the default TeX fonts;
- URWVN: based on URW fonts, the free version of 35 standard PostScript fonts;
- Vn CM Bright: based on CM Bright, a sans serif font derived from Computer Modern fonts;

À \A	Í \I	Ý \Y	ị \d{i}
Ả \h{A}	Ị \d{I}	Ỳ \d{Y}	ò \o
Ã \~A	Ò \o	à \a	Ỗ \h{fo}
Á \'A	Ỏ \h{0}	ả \h{a}	õ \o
Ạ \d{A}	Õ \~0	ã \~a	ó \o
Ằ \Abreve	Ố \o	á \a	ơ \d{fo}
Ằ' \Abreve	Ọ \d{0}	ạ \d{a}	ô \ocircumflex
Ằ̃ \hAbreve	Ỗ \0circumflex	ă \abreve	ồ \ocircumflex
Ằ̃' \Abreve	Ỗ' \0circumflex	ã \a	ỗ \h\ocircumflex
Ằ̃̃ \Abreve	Ỗ̃ \h\0circumflex	ã̃ \h\abreve	ỗ̃ \ocircumflex
Ằ̃̃' \Abreve	Ỗ̃' \h\0circumflex	ã̃̃ \~\abreve	ỗ̃̃ \ocircumflex
Ằ̃̃̃ \d\Abreve	Ỗ̃̃̃ \~\0circumflex	ã̃̃̃ \'\abreve	ộ \d\ocircumflex
Â \Acircumflex	Ỗ̃̃̃' \'\0circumflex	ẵ \d\abreve	ơ \ohorn
Â' \Acircumflex	Ỗ̃̃̃̃ \d\0circumflex	â \acircumflex	ờ \'\ohorn
Ẫ \h\Acircumflex	Ỗ̃̃̃̃' \h\0circumflex	ẫ \a	ỡ \h\ohorn
Ẫ' \Acircumflex	Ỗ̃̃̃̃̃ \h\0circumflex	ẫ̃ \h\acircumflex	ỡ̃ \ohorn
Ẫ̃ \Acircumflex	Ỗ̃̃̃̃̃' \h\0circumflex	ẫ̃̃ \~\acircumflex	ỡ̃' \ohorn
Ẫ̃' \Acircumflex	Ỗ̃̃̃̃̃̃ \~\0circumflex	ẫ̃̃̃ \'\acircumflex	ợ \d\ohorn
Ẫ̃̃ \d\Acircumflex	Ỗ̃̃̃̃̃̃' \'\0circumflex	ậ \d\acircumflex	ù \u
Đ \DJ	Ỗ̃̃̃̃̃̃̃ \d\0circumflex	đ \dj	ủ \h{u}
È \E	Ỗ̃̃̃̃̃̃̃' \h\0circumflex	è \e	ũ̃ \u
Ẻ \h{E}	Ỗ̃̃̃̃̃̃̃̃ \h\0circumflex	ẻ \h{e}	ú \u
Ẻ̃ \~E	Ỗ̃̃̃̃̃̃̃̃' \h\0circumflex	ẽ \e	ụ \d{u}
Ẻ' \E	Ỗ̃̃̃̃̃̃̃̃̃ \h\0circumflex	é \e	ư \uhorn
Ẹ̉ \d{E}	Ỗ̃̃̃̃̃̃̃̃̃' \h\0circumflex	ẹ \d{e}	ừ \'\uhorn
Ê \Ecircumflex	Ỗ̃̃̃̃̃̃̃̃̃̃ \h\0circumflex	ê \ecircumflex	ử \h\uhorn
Ê' \Ecircumflex	Ỗ̃̃̃̃̃̃̃̃̃̃' \h\0circumflex	ề \'\ecircumflex	ử̃ \ohorn
Ễ \h\Ecircumflex	Ỗ̃̃̃̃̃̃̃̃̃̃̃ \h\0circumflex	ề̃ \h\ecircumflex	ử̃' \uhorn
Ễ' \Ecircumflex	Ỗ̃̃̃̃̃̃̃̃̃̃̃' \h\0circumflex	ề̃̃ \~\ecircumflex	ử̃̃ \d\uhorn
Ễ̃ \Ecircumflex	Ỗ̃̃̃̃̃̃̃̃̃̃̃̃ \h\0circumflex	ề̃̃̃ \'\ecircumflex	ỳ \y
Ễ̃' \Ecircumflex	Ỗ̃̃̃̃̃̃̃̃̃̃̃̃' \h\0circumflex	ễ̃̃̃ \d\ecircumflex	ỷ \h{y}
Ễ̃̃ \d\Ecircumflex	Ỗ̃̃̃̃̃̃̃̃̃̃̃̃̃ \h\0circumflex	ì \i	ỷ̃ \y
Ì \I	Ỗ̃̃̃̃̃̃̃̃̃̃̃̃̃' \h\0circumflex	ỉ \h{i}	ỷ' \y
Ỉ \h{I}	Ỗ̃̃̃̃̃̃̃̃̃̃̃̃̃̃ \h\0circumflex	ĩ \i	ỵ \d{y}
Ĩ \~I	Ỗ̃̃̃̃̃̃̃̃̃̃̃̃̃̃' \h\0circumflex	Ý \Y	
	Ỗ̃̃̃̃̃̃̃̃̃̃̃̃̃̃̃ \h\0circumflex	Ỳ \h{Y}	
	Ỗ̃̃̃̃̃̃̃̃̃̃̃̃̃̃̃' \h\0circumflex	Ỷ \h{Y}	
	Ỗ̃̃̃̃̃̃̃̃̃̃̃̃̃̃̃̃ \h\0circumflex	Ỵ \Y	

Table 3: Vietnamese alphabet

- Vn Concrete: based on the CM Concrete font;
- TXTTVN: based on TXTT, a very nice typewriter font from the `txfonts` package;
- Vntopia: based on Adobe Utopia;
- Vn Charter: based on Bitstream Charter;
- Vn URW Grotesk: based on URW Grotesk font, a heavy font suitable for displayed text;
- Vn Garamond: based on URW Garamond;

- Vn Classico: based on URW Classico (also known as Optima).

In addition to the above, $\text{VnT}_{\text{E}}\text{X}$ also contains $\text{T}_{\text{E}}\text{X}$ support for some TrueType fonts (MS Core fonts, ArevSan, ComicSans).

5.2 Fonts from other sources

The fonts made by Bogusław Jackowski and Janusz

Nowacki have excellent support for many languages including Vietnamese. The fonts come ready-to-use and cover all Vietnamese letters. Since the fonts are so well known, I will only list them here without description:

- Latin Modern fonts,
- TeX Gyre fonts (Bonum, Pagella, Schola, Termes, Heros),
- Antykwa Toruńska,
- Iwona,
- Kurier.

Antykwa Półtawskiego is also a very nice font made by our Polish font experts but doesn't support Vietnamese yet. Therefore I take this opportunity to repeat my plea to add Vietnamese support to this font :).

Starting with version 7.0, Adobe Reader comes with two nice fonts: Minion and Myriad. These fonts have full support for Vietnamese. However, the use of these fonts is restricted to Adobe Reader only. It's not even allowed to embed a subset of the fonts in a PDF or PostScript file. Using a recent version of pdfTeX, it is possible to use the fonts without embedding them (but doing so implies that the PDF output must be viewed or printed with Adobe Reader 7.0 or newer).

5.3 Typesetting Vietnamese and math

As mentioned above, VnTeX tries to make typesetting Vietnamese as easy as typesetting English. When it comes to math typesetting, this holds true as well. Most of the fonts listed in the Free Math Font Survey [7] either have a Vietnamese version available with VnTeX, or support Vietnamese themselves. Therefore, the samples listed in the survey also apply to Vietnamese. There is even a Vietnamese translation of the survey [8].

6 Comments on Latin Modern and TeX Gyre from a user perspective

The Latin Modern and TeX Gyre fonts try to make life easier for TeX maintainers as well as TeX users. Therefore I think it is important for the authors to get feedback not only from font experts and TeX maintainers, but also from end users. In this section I would like to comment on the fonts from the perspective of a Vietnamese TeX user.

Ease of use: excellent, since the fonts are already included in major TeX distributions, with everything needed for use. In other words, it cannot be easier for someone who wants to use them or give them a try. However, it seems that people

are not very interested in trying, since the benefits of switching to those new fonts are not that clear for an end user. Perhaps a short document explaining the benefits to switch to those new fonts would be helpful.

Quality: the Vietnamese letters look very good in general:

- accents are very consistent regarding shape and placement;
- in most cases the accents look similar to or better than those in VNR and URWVN fonts;
- kerning data are added for Vietnamese letters as well;
- the widths of letters like \mathring{U} , \mathring{O} , \mathring{u} , \mathring{o} are adjusted properly;
- the quality of hints and outlines are far better than in URWVN fonts.

Issues: still, there are some issues regarding accent shapes and placement that need to be discussed and perhaps re-considered if applicable:

- sometimes the accents don't seem to have the same (or optically compatible) darkness; especially the grave accent is often darker than the others in italic fonts;
- the placement of acute and grave over circumflex does not seem optimal to Vietnamese users, especially the combination of grave over circumflex. It is preferable to have the grave rather on the right side of the circumflex, instead of the center.
- sometimes the accents are placed too close to the base letter or to each other, making some letters look too "crowded". This is more visible at small sizes, or low resolution.

7 Acknowledgments

I would like to thank all the people who have contributed to VnTeX in various ways. I try to list here the most significant contributors that I can recall, in no particular order:

- Vladimir Volovich for work on the T5 encoding and supporting T5 encoding in his `cmapp` package,
- Werner Lemberg for major L^ATeX support in VnTeX and also for the `vncmr` package, from which VnTeX was derived,
- Reinhard Kotucha for maintaining the VnTeX package,
- Bogusław Jackowski and Janusz M. Nowacki for their excellent support of Vietnamese in their fonts,

- Nguyễn Đại Quý (also known as `vn Penguin`) for supporting the `vntex.org` domain and many various issues related to VnTeX, including dedicating a box on his forum [5] to VnTeX,
- Thái Phú Khánh Hòa for the Vietnamese translation of the Free Math Font Survey and the VnTeX logo at the VnTeX website [2].

References

- [1] Hàn Thế Thành, *Making Type 1 fonts for Vietnamese*, *TUGboat* 24:1, Proc. of the 24th Annual Meeting and Conference of the T_EX Users Group, pp. 69–84.
- [2] The official VnTeX website: `vntex.sf.net`.
- [3] The latest version of VnTeX and installation instructions (in English) are available at: <http://vntex.sf.net/download/vntex>
- [4] The complete font samples that can be used with VnTeX are available at <http://vntex.sf.net/fonts/samples>. The NFSS and TFM names of all fonts are included, making it easy to use a particular font.
- [5] The VnOSS forum has a box dedicated to T_EX and Vietnamese:
<http://forum.vnoss.org/viewforum.php?id=10>
- [6] Another web site with a forum useful for Vietnamese T_EX users:
<http://viettug.org>
- [7] The Free Math Font Survey is available at:
http://ctan.tug.org/tex-archive/info/Free_Math_Font_Survey/survey.html
- [8] The Vietnamese translation of the Free Math Font Survey is available at:
http://ctan.org/tex-archive/info/Free_Math_Font_Survey/vn/survey-vn.pdf
- [9] The definition of the Vietnamese alphabet at Wikipedia: http://en.wikipedia.org/wiki/Quoc_ngu