

Making Type 1 fonts for Vietnamese

Vietnamese letters

- Vietnamese is written with Latin letters and a few more accents in a system by the Portuguese missionary Alexander Rhodes.
- What separates Vietnamese from other languages with Latin characters is that in Vietnamese some letters can have two accents.
- Vietnamese accents can be divided to three kinds of diacritic marks: tone, vowel and consonant.

Vietnamese lowercase letters

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

o	ó	ọ	ò	ỏ	õ
ô	ố	ộ	ồ	ỗ	ỗ
ơ	ớ	ợ	ờ	ở	ỡ
u	ú	ụ	ù	ủ	ũ
ư	ứ	ự	ừ	ử	ữ
y	ý	ỵ	ỳ	ỷ	ỹ
đ					

Vietnamese uppercase letters

A	Á	Ạ	À	Ả	Ã
Ă	Ắ	Ặ	Ằ	Ẵ	Ẳ
Â	Ấ	Ậ	Ầ	Ẻ	Ẻ
E	É	Ệ	È	Ẻ	Ễ
Ê	Ế	Ệ	Ề	Ễ	Ễ
I	Í	Ị	Ì	Ỉ	Ĩ

O	Ó	Ọ	Ò	Ỏ	Õ
Ô	Ố	Ộ	Ồ	Ổ	Ỗ
Ơ	Ớ	Ợ	Ờ	Ở	Ỡ
U	Ú	Ụ	Ù	Ủ	Û
Ư	Ứ	Ự	Ừ	Ử	Ữ
Y	Ý	Ỡ	Ỡ	Ỡ	Ỡ
Đ					

Vietnamese diacritic marks

Diacritic mark	Example
<i>Vowel</i>	
breve	băn khoản
circumflex	hôm nay
horn	Qui Nhơn
<i>Tone</i>	
acute	Lái Thiêu
grave	Bình Dương
hook above	Thủ Đức
tilde	dĩ vãng
dot below	học tập
<i>Consonant</i>	
stroke	đã đời

CMR-based fonts for Vietnamese: VNR

- Vietnamese letters are just like Latin letters, so it is natural to have VNR based on CMR sources.
- Several works have been done on this topic prior to VNR.
- The best among them was the package `vncmr` by Werner Lemberg.
- But I am not entirely happy with `vncmr` so I decided to write VNR by myself.
- A lot of ideas were borrowed from the `vncmr` package and other CMR-based fonts.

- In the beginning I wrote VNR fonts based on CMR sources.
- Later, Werner Lemberg and Vladimir Volovich convinced me that it would be better to switch to EC sources instead of CMR sources. The two main reasons are:
 1. It will be easier to include Vietnamese letters in CM-Super fonts;
 2. it will fix some problems in VNR fonts (mainly with encoding and some missing glyphs)
- However, both EC-like and CMR-like naming are supported, for a reason we will see later on.

It's good that I can use Vietnamese with $\text{T}_{\text{E}}\text{X}$, but I want PDF

- And PK fonts just look ugly in Acrobat Reader, so I need Type 1 fonts.
- My first attempt was to create a set of virtual fonts for Vietnamese, which refer to CMR fonts. The shortcoming is that some accents and letters look poor.
- My next attempt was to generate Type 1 version of VNR using $\text{T}_{\text{E}}\text{X}$ trace. The result is better, however there is still plenty of room for improvement compared with e.g. BlueSky fonts.

Efficient generating Type 1 format for VNR fonts

1. Take the accents from METAFONT sources and convert them to Type 1 format;
2. compose those accents and the English letters from BlueSky fonts to get accented letters.

Converting the accents to Type 1 format

- This can be done quickly using `TEXtrace`, but the result is not perfect.
- A more elaborate method is to use `METAPOST` and `MetaFog`, which gives better result.
- for each font, a number of 47 glyphs have to be converted, from which:
 - 22 glyphs are Vietnamese accents and letters;
 - 25 glyphs are those that are in T5 encoding but missing in BlueSky fonts.

Glyphs that need to be converted to Type 1 using MetaFog

Vietnamese accents and letters	Additional glyphs
dotbelow	quotesinglbase
hookabove	guilsingleft
vlgrave	guilsingright
vlacute	quotedblleft
vlcircumflex	quotedblright
vtilde	quotedblbase
vldotbelow	guillemotleft
vlbreve	guillemotright
vlhookabove	endash
vugrave	emdash
vuacute	cwm
vucircumflex	zeroinferior
vutilde	uni2423
vudotbelow	quotedbl
vubreve	dollar
vuhookabove	less
Ohorn	greater
Uhorn	backslash
ohorn	asciicircum
uhorn	underscore
Dcroat	braceleft
deroat	bar
	braceright
	asciitilde
	sfthyphen

Hinting the accents

- When MetaFog finished, hints for accents were auto-generated using FontLab.
- As the accent shapes are quite simple, hints auto-generated by FontLab are relatively reasonable.
- It's hard to do better without a great deal of experience in manual hinting.

Composing the accented letters

- The tool used in this step is FMP (Font Manipulation Package) from Y&Y.
- In this package, there is a tool that can create composite glyphs from existing glyphs in a font.
- To place the accents precisely, hooks were added into VNR sources so information about accent positioning is written to a log file.
- Some Perl scripts were made to extract the data from the log file and use them with FMP to compose accented letters.

What about the result size?

- A rough comparison on average font size gave the following result:

BlueSky fonts: 25 KB

VNR fonts generated by T_EXtrace: 70 KB

VNR fonts generated as described here: 40 KB

- Compactness is gained thanks to FMP, which puts all accents and English letters into subroutines so they can be reused without duplication.

And about the outlines and hints?

- The base part of accented letters has the same outlines and hints as in BlueSky fonts;
- the accent part of accented letters has the outlines produced by METAPOST and MetaFog, hints auto-generated by FontLab.

A text sample of VNR10

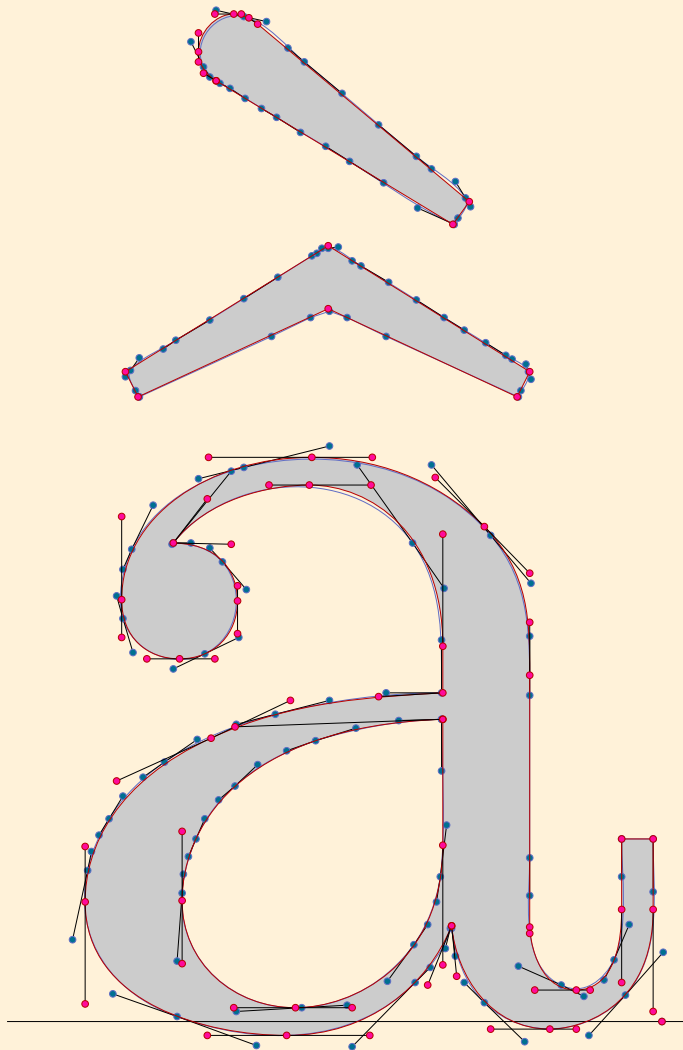
Chí Phèo

Hắn vừa đi vừa chửi. Bao giờ cũng thế, cứ rượu xong là hắn chửi. Bắt đầu chửi trời. Có hề gì? Trời có của riêng nhà nào? Rồi hắn chửi đời. Thế cũng chẳng sao: đời là tất cả nhưng chẳng là ai. Tức mình hắn chửi ngay tất cả làng Vũ Đại. Nhưng cả làng Vũ Đại ai cũng nhủ, “*Chắc nó trừ mình ra!*” Không ai lên tiếng cả. Tức thật! Ô! Thế này thì tức thật! Tức chết đi được mất! Đã thế, hắn phải chửi cha đứa nào không chửi nhau với hắn. Nhưng cũng không ai ra điều. Mẹ kiếp! Thế thì có phí rượu không? Thế thì có khổ hắn không? Không biết đứa chết mẹ nào đẻ ra thân hắn cho hắn khổ đến nông nổi này? A ha! Phải đấy, hắn cứ thế mà chửi, hắn chửi đứa chết mẹ nào đẻ ra thân hắn, đẻ ra cái thằng Chí Phèo! Hắn nghiến răng vào mà chửi cái đứa đã đẻ ra Chí Phèo. Nhưng mà biết đứa nào đã đẻ ra Chí Phèo? Có trời mà biết! Hắn không biết, cả làng Vũ Đại cũng không ai biết...

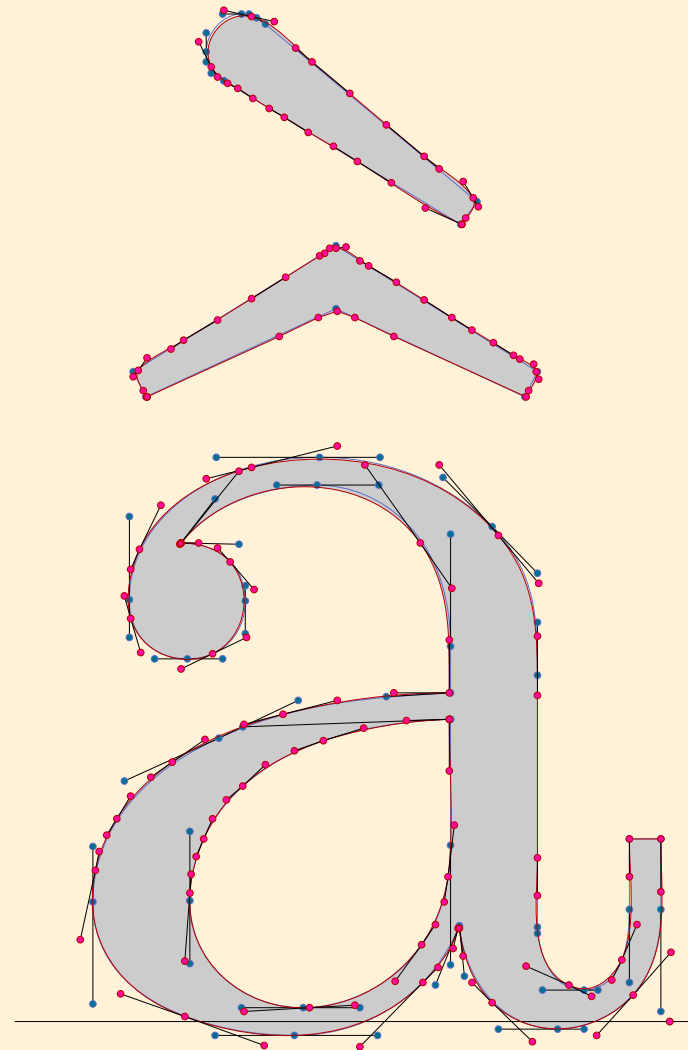
Anything missing?

- The VNR sources are EC-based, while the English letters from BlueSky fonts are CMR-based.
- There **are** differences between glyphs in EC and CMR fonts.
- Most users, however, will not notice it.

A closer look at the outlines



Red (front): MetaFog + BlueSky
Blue (back): TeXtrace



Red (front): TeXtrace
Blue (back): MetaFog + BlueSky

Related works

- The same method was applied to add Vietnamese letters into the URW fonts.
- The result is a package called URWVN.
- URWVN fonts are very similar to Type 1 VNR fonts regarding technical aspects:
 - They are very compact; the average size of URWVN fonts is 35 KB;
 - the base part of accented letters has the same outlines and hints as in URW fonts;
 - the accent part of accented letters was drawn manually, hints auto-generated by FontLab.

A text sample of Times Roman

Chí Phèo

Hắn vừa đi vừa chửi. Bao giờ cũng thế, cứ rượu xong là hắn chửi. Bắt đầu chửi trời. Có hề gì? Trời có của riêng nhà nào? Rồi hắn chửi đời. Thế cũng chẳng sao: đời là tất cả nhưng chẳng là ai. Tức mình hắn chửi ngay tất cả làng Vũ Đại. Nhưng cả làng Vũ Đại ai cũng nhủ, “*Chắc nó trừ mình ra!*” Không ai lên tiếng cả. Tức thật! Ô! Thế này thì tức thật! Tức chết đi được mất! Đã thế, hắn phải chửi cha đứa nào không chửi nhau với hắn. Nhưng cũng không ai ra điều. Mẹ kiếp! Thế thì có phí rượu không? Thế thì có khổ hắn không? Không biết đứa chết mẹ nào đẻ ra thân hắn cho hắn khổ đến nông nỗi này? A ha! Phải đây, hắn cứ thế mà chửi, hắn chửi đứa chết mẹ nào đẻ ra thân hắn, đẻ ra cái thằng Chí Phèo! Hắn nghiền răng vào mà chửi cái đứa đã đẻ ra Chí Phèo. Nhưng mà biết đứa nào đã đẻ ra Chí Phèo? Có trời mà biết! Hắn không biết, cả làng Vũ Đại cũng không ai biết. . .

How VNR and URWVN fonts look?

- I am not a type designer,
- therefore the aesthetic aspect of Vietnamese letters (regarding accent shapes and positioning) is open to discussion.
- What I have done is heavily based on what I learnt from other Vietnamese fonts and comments from various people.
- It's not to say that I take no responsibility for VNR fonts.
- If something looks bad, it's my fault of course.

Future works

- Convert these fonts to other formats so they can be used with other applications.
- The accents shapes of VNR fonts can still be improved (in METAFONT sources):
 - It is better to have two separate versions for each accent (for uppercase and lowercase).
 - This is important for Vietnamese, as many letters have double accents, which causes such letters to be very high.
 - Uppercase letters should have wider and lower accents than their lowercase counterparts.

- The same method can be applied to add Vietnamese letters into other Type 1 fonts.
- I consider adding Vietnamese letters into the following fonts:
 1. The Utopia family from Adobe;
 2. the Charter family from Bitstream;
 3. the Lucidux family from Bigelow & Holmes and Y&Y.

Acknowledgments

- Werner Lemberg for creating the vncmr package from which I learnt a lot;
- Tom Kacvinsky for donating a copy of FMP;
- Richard Kinch for creating MetaFog and donating an evaluation copy;
- Paul Watry for donating a copy of FontLab;
- William Adams for correcting my paper;
- and folks from the conference organizers and from all around the world for their great help so I could come to TUG 2003.