Hóng-Zì: A Chinese METAFONT

Javier Rodríguez Laguna
SISSA, Trieste (Italy)
jrlaguna@sissa.it
http://www.sissa.it/~jrlaguna

Abstract

Hóng-Zì (红字) is a new Chinese METAFONT, still at a very early stage of development, freely available at http://hongzi.sourceforge.net under the GPL. The structure is split into four levels of abstraction: strokes, radicals, distributions and full characters, such that the highest level description of the characters becomes truly simple.

1 Introduction

Chinese characters (*hanzi*) and their close relatives[1] appear to be specially suitable for treatment under METAFONT [KNU 86]. Their shapes, notwithstanding their complexity, are made up of a finite (though large) number of parts. But these parts are not just pasted inside a box, like letters in a word. The parts change their shape depending on the size and position where they are going to be inserted. METAFONT is an approach to font design which is especially suited for these shape modifications.

Nonetheless, the TEX community still lacks a full fledged Chinese METAFONT. Of course, it is easy to include CJKV text in a TEX document using packages which employ other types of fonts, but then we lack the tunability and high quality to which we are accustomed.

The first development of a Chinese METAFONT was performed by Gu Guoan and John Hobby in the METAFONT-79 dialect [GH 84], which is not compatible with the current version. In the November 1982 issue of Scientific American, Douglas Hofstadter described his own project, in collaboration with David Leake, called *Hàn-Zì* [HOF 82], [HOF 86]. Unfortunately, it was abandoned shortly after.

In 1993, Martin Dürst provided an interesting approach for a Chinese METAFONT [DÜR 93] which, as far as the present author knows, did not have a follow-up. The recent proposal of Candy L.K. Yiu and Wai Wong [YW 03] may be more fruitful. In it, the authors describe a language for the description of *hanzi*, known as *HanGlyph*, and a METAPOST program called CCSS (Chinese Character Synthesis System) which renders the characters. Although the approach introduced in this paper was developed independently, it shares many features with the ideas of Yiu and Wong.

Hóng-Zì (红字) was born as an attempt to overcome this gap in the TEX world. It is supported by SourceForge and its first release was published in May 2003 at http://hongzi.sourceforge.net. It is organized into four levels of abstraction, so that reading (and writing) the highest level files is easy even for a METAFONT-newbie. This way, non-experts should be able to help in its development.

The current version, described in this paper, is Hóng-Zì 0.5. Although this distribution contains only 126 characters, it has the potential to build many more. The characters are composed from 69 *radicals*, which are drawn using 32 different types of strokes. The basic mechanism is probably fixed now, although some details may change in subsequent versions. The main tasks are now (a) the development of more radicals and more characters and (b) the design of a nice TEX interface.

Section 2, which is the core of this paper, is a description of the abstraction levels of Hóng-Zì. This paper ends with a discussion of the open problems and future work.

2 Abstraction levels of Hóng-Zì

Hanzi are made up of parts which, going beyond the strict meaning of the term, we shall name *radicals*. (Strictly speaking, radicals are the 214 *hanzi* parts used by Chinese dictionaries for indexing.) Radicals are distributed in certain ways inside the character, always fitting a square box. (To learn more about the structure of *hanzi*, I recommend the books

[1] Chinese *hanzi* (both simplified and traditional), Japanese *kanji*, Korean *hanja* (and even *hangul*) and Vietnamese *chữ hán* and *chữ nôm*. Writing systems from these countries are usually termed CJKV text. See [LUN 98] for details.

[HEI 01] and [FAZ 86], and the program *Hanzi Master* (`hanzim`) [ROB 02].)

In summary, *characters* consist of *radicals*, *distributed* in certain ways. Radicals, in turn, consist of *strokes*. The distribution of Hóng-Zì contains (at this moment) four files, one for each abstraction level. Let us describe each one separately.

(a) Strokes. This is the lowest level, the one which only uses METAFONT primitives. There are 32 stroke functions defined in the current version. Each one requires either two, three or four parameters. Some examples: `point(cx,cy)` only needs two parameters, which correspond to the coordinates of the peak of the point. The stroke `hook_v(cx,cy,l)` (which draws a vertical hooked bar) requires three: the coordinate of the starting point and the length. And `down_right(cx,cy,lx,ly)` (which draws a stylized stroke going down and right) requires four: the coordinates of the starting point, and the width and height of a box which contains it.

Prototype	Shape
`point(cx,cy);`	丶
`point_sized(cx,cy,lx,ly);`	丶
`point_ne_sized(cx,cy,lx,ly);`	丶
`bar_v(cx,cy,l);`	丨 一
`down_left(cx,cy,lx,ly);`	丿 丿
`down_right(cx, cy, lx, ly);`	乀 乀
`hook_v(cx, cy, l);`	亅 一
`hook_ob(cx, cy, lx, ly);`	乚
`corner_tr(cx, cy, lx, ly);`	冂
`corner_hook(cx, cy, lx, ly);`	冂
`vert_raise(cx, cy, l);`	亅
`nu_stroke(cx, cy, lx, ly);`	𠂉
`angle(cx, cy, lx, ly);`	𠂉
`ell(cx, cy, lx, ly);`	乚
`ye_stroke(cx, cy, lx, ly);`	𠂉
`three(cx, cy, lx, ly);`	𠃌
`three_hook(cx, cy, lx, ly);`	𠃌
`spoon_stroke(cx, cy, lx, ly);`	𠃍

Table 1. Some of the stroke functions defined in Hóng-Zì. Strokes with more than one shape are polymorphic.

Some strokes are *polymorphic*; for instance, the shape of a `down_right(0f,10f,8f,2f)` is not just a rescaling of a `down_right(0f,10f,8f,8f)`. The most important parameter which decides the shape is (usually) the aspect ratio, $A :=$ height/width.

Table 1 shows some of the stroke functions defined so far, along with some of the possible output. The complete list is in the file `strokes.mf` in the current distribution.

(b) Radicals. The term *radical* is extended in this work to cover any combination of strokes we regard as a unit inside a *hanzi*. Each radical function takes four parameters: coordinates of the upper-left corner, width and height. Radicals are given "meaningful" names in English, such as `child` (子) or `eye` (目). It is not always easy to find an appropriate name for a given radical.

Many radicals are also polymorphic. For example, let us consider the radical `water`. if the aspect ratio is $A > 2$, its appearance is: 氵. But if $A \leq 2$, it looks like the full character: 水. Table 2 shows some of the radicals defined in Hóng-Zì, as described in file `radicals.mf`.

Radical name	Rendered form
`water`	水 氵
`roof`	宀
`child`	子
`woman`	女
`tongue`	舌
`omen`	水
`person`	人 亻
`heart`	心 忄 ⺗
`sun`	日
`grass`	艹
`moon`	月
`fire`	火 灬
`ghost`	鬼
`way`	辶
`knife`	刀 刂
`rice`	米

Table 2. A few of Hóng-Zì's radicals. Polymorphic radicals show some of their possible renditions.

(c) Distributions. A number of binary operators have been defined which are useful for distributing radicals on the character. E.g.: the `H(·,·)` operator composes a single radical out of two, distributing them horizontally. Each radical get 50% of the width. It is similar to the TEX expression `\hbox{X Y}`, but taking into account the fact that radicals have no natural height or width.

If some space must be left between the radicals, we may use `H_(·,·)`, which leaves 10% of blank space. The `Hl` series, which splits horizontally and gives more space to the left part) is depicted in full. We have three equivalent series: `Hr` (horizontal splitting, more space to the right part), `Vu` (vertical splitting, more space to the upper part) and `Vd` (vertical splitting, more space to the lower part).

There are a few other distribution operators which do not fall into any of the above four series: the L-box `L(X,Y)` operator, used in 道 *dāo* (road); the inner-box `I(X,Y)` operator, used in 国 *guó* (country); and the `Is(X,Y)` operator, which leaves a smaller inner box, as in 问 *wèn* (question).

Table 3 shows some of the distribution operators created so far. They are defined in the source file `distributions.mf`.

Of course, these operators may be composed, thus providing a way to describe complex characters, as we shall discuss in the following section.

(d) Full characters. All the lower layers are designed to make the highest level description of characters easy. As an example, the Hóng-Zì code for 明 *míng* (luminosity), as expressed in our file `hongzi.mf`, is just

```
H_rr(sun)(moon)
```

This line is to be read as follows: "Make up a character whose left part is the radical for sun and whose right part is the radical for moon. Give much more space to the right part and leave some space between them". In the real code, this line is enclosed by a `zi`...`iz` pair.

A more complex character is 凉 *liáng* (cold),

```
Hrr(ice)(Vdd(above)(Vd(box)(small)))
```

METAFONT reads this complex line *downwards*, like this: it first prepares a big box and splits it horizontally (giving more space to the right part, `Hrr`). It fills the left part with the `ice` radical. The right part, on the other hand, is processed further. First, it is split vertically (giving much more space to the lower part, `Vdd`). In the upper *half*, it renders the `above` radical. The lower part is again split (giving slightly more space to the lower part, `Vd`), and there the `box` and `small` radicals are rendered.

Distribution operator	Effect
`H(X,Y)`	
`H_(X,Y)`	
`Hl(X,Y)`	
`H_l(X,Y)`	
`Hll(X,Y)`	
`H_ll(X,Y)`	
`V(X,Y)`	
`V_(X,Y)`	
`Vu(X,Y)`	
`L(X,Y)`	
`I(X,Y)`	
`Is(X,Y)`	

Table 3. Some of Hóng-Zì's distribution operators.

Table 4 shows the code for some full characters, ordered according to their complexity.

3 Open problems and future work

The basic structure of the project seems to be reasonably fixed, although some issues of design must still faced. For example, regarding the distribution operators, perhaps we should replace the `H`, `Hl`, `Hll` series with a single operator containing an extra parameter. Also, the character descriptions should make up a sort of dictionary, containing the shape

Hanzi	Pinyin	English
女	*nǚ*	woman
`woman`		
水	*shǔi*	water
`water`		
好	*hǎo*	good
`H(woman)(child)`		
安	*ān*	peace
`Vdd(roof)(woman)`		
迷	*mí*	lost
`L(way)(rice)`		
谜	*mí*	riddle
`Hrr(word)(L(way)(rice))`		
孬	*nāo*	not good
`Vd(no)(H(woman)(child))`		
花	*huā*	flower
`Vdd(grass)(H(person)(spoon))`		
影	*yǐng*	shadow
`H_ll(Vdd(sun)(Vdd(above)(Vd(box)(small))))(beard)`		

Table 4. Character descriptions as they appear in the `hongzi.mf` file. They become quite easy at the highest level. Even complex characters such as 影 are explained in a single line.

description, pinyin, English translation and some compounds.

But the main problems of Hóng-Zì at this stage are that we must have (a) many more radicals and characters; (b) calligraphic improvements and different sets of strokes; and (c) a nice TEX interface. Therefore, we need and welcome new volunteers for collaboration! Deep knowledge of METAFONT is not needed as much as deep knowledge of the Chinese language, as can be seen from the high-level structure. And of course, extensions to other scripts in the CJKV family would be very nice.

The project web page, `http://hongzi.sf.net`, aspires to contain a full-fledged introduction both to the Chinese language for non-native speakers, and to CJKV typing.

Suggestions, comments, criticisms and, most of all, offers to help, are most welcome.

Acknowledgements

The author acknowledges Douglas R. Hofstadter and the organizers of TUG'05 for their encouragement and very useful comments. This project is supported by SourceForge (`http://sourceforge.net`).

Bibliography

[DÜR 93] Martin J. Dürst, *Coordinate-independent font description using Kanji as an example*, Electronic Publishing **6**(3), 133–143 (1993).

[FAZ 86] Edoardo Fazzioli, *Caratteri cinesi*, Mondadori (1986).

[GH 84] John D. Hobby and Gu Guoan, *A Chinese Meta-Font*, *TUGboat* **5**(2), 119–136 (1984).
`http://cm.bell-labs.com/who/hobby/pubs.html`

[HEI 01] James W. Heisig, *Remembering the Kanji*, Tokyo: Japan Publishing Co. (2001).

[HOF 82] Douglas R. Hofstadter, *Changes in default words and images, engendered by rising consciousness*, Sci. Am., Nov. 1982.

[HOF 86] Douglas R. Hofstadter, *Metafont, metamathematics and metaphysics*, in *Metamagical themas* (chapter 13), Penguin Books (1986).

[KNU 86] Donald E. Knuth, *The METAFONTbook*, Addison-Wesley (1986).

[LUN 98] Ken Lunde, CJKV *information processing*, O'Reilly (1998).

[ROB 02] Adrian Robert, *Hanzi Master, a Chinese character learning-aid program* (2002).
`http://zakros.ucsd.edu/~arobert/hanzim.html`

[YW 03] Candy L.K. Yiu and Wai Wong, *Chinese character synthesis using METAPOST*, *TUGboat* **24**(1), 85–93 (2003).