

TeX helps you learn Chinese character meanings

Alan Hoenig

My story begins about 6 or 7 years ago now when, on a whim, I decided to study Mandarin Chinese. I've had these whims often over the years, and I know that the more intense it is, the sooner it burns itself out sooner or later they burn themselves out — which is why this time I focused on the interesting stuff and neglected the dull material, which in this case was Chinese characters themselves and their meanings. But, days, months, and now years passed, and I stayed intrigued, so it was a mistake to have ignored them.

But studying Chinese characters turned out to be tough — too tough. I couldn't seem to remember more than a handful accurately, not enough to make any real headway.

I found this frustrating, but fortunately before ditching everything, I realized it couldn't hurt to apply one of the great lessons of Metafont. That is, rather than to actually *study* this material, I took a step back and thought about *how* to study this material. I looked around to see what other people had to say and what methods they used, and then I came up with the following scheme, one that seems to work pretty well.

I took as an initial pool the 2000 most frequently used characters. (I used the so-called simplified character set, the characters in official use by the People's Republic of China.) Then I imposed an order on them — not an alphabetical or numerical order, but one based on how complex each character is — how easy (or not) it is to write. I arranged them from simplest to increasingly complicated.

Then I used an induction-based learning scheme relying on three essential platforms:

1. The form and meaning of any character depends only on characters and components that appear earlier in the sequence.
2. You can remember this form and its meaning by means of relatively simple/straightforward mnemonic narratives which use prior characters (characters that are already known) as elements in this story. These characters are already known because they precede the current character in the sequence.
3. Finally, the initial item in this sequence must be easy to remember all by itself.

As far as the mnemonic stories go, anything — any kind of story connecting the components, any pun or play on words, and any kind of outlandish sce-



Figure 1: The three easiest Chinese characters to draw have meanings 'one', 'two', and 'three'.



Figure 2: A scepter-like or stick-like component which is not an independent character.

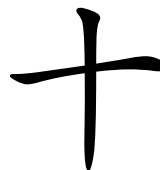


Figure 3: The character which means 'ten'.

nario — is legitimate as we create this defiantly ahistorical and unconventional method for remembering character meanings.

Let me show you how this works.

I think everyone will agree that the simplest character is the leftmost in figure 1, a single horizontal line. It means 'one', and it's pretty easy to remember its meaning. By great good fortune, the next two characters are equally easy. Chinese scribes build them up out of additional horizontals, and no one will be surprised when I tell you they mean 'two' and 'three', and I don't even have to tell you which is which!

But it's not possible to create any more characters if you limit yourself to horizontal strokes, so we need to add the first of about 100 components which combine with additional characters to form new ones (fig. 2). This first component is simply a vertical stroke, which looks like a stick. In a mnemonic story, we'll let this stick represent a simple hoe-like tool, or perhaps a scepter to represent the authority of some leader.

The next character in the sequence (fig. 3) is a simple cross, bar plus scepter. Its Chinese meaning is 'ten', and since it looks like a 't' it's easy to remember its meaning this way — 't' stands for 'ten'. But also, if you try to stand it up on its bottom point, it will naturally tip to the right or the left by

Figure 4: ‘Labor’ or ‘work’.

Figure 5: ‘Earth’ or ‘soil’; ‘dry’.

Figure 6: ‘King’.

45 degrees, forming an ‘X’, a Roman numeral whose value is 10.

In figure 4, someone has inserted the scepter between the two strokes of ‘two’, not so easy to do because it takes *work* to pry the bars apart. One meaning for this character is ‘labor’ or ‘work’.

Using the scepter as a tool, imagine it a primitive hoe we’ll use at planting time in the spring. The top bar of this next character shows the surface of the soil, and the bottom bar reminds the farmer of the depth to which he must plow. One meaning for this character is ‘earth’ or ‘soil’ (left part of fig. 5).

But what happens in time of drought? The dry soil becomes so soft and crumbly that it’s easy to plunge the hoe all way down until it’s flush with the top surface, like this. One meaning for this character is ‘dry’ (right part of fig. 5).

Finally, consider the character in figure 6. The ‘one’ is on top of ‘two’ and the pair are transfixed by a royal scepter, the symbol of leadership and authority. This character refers to a person who is the ‘one’ ‘to’ (sounds like ‘two’, get it?) ‘lead’—a king, which is the primary meaning for *this* character.

... And so on. I proceeded in this way to develop mnemonic stories for a total of 2178 characters.

I used T_EX, or rather the X_YT_EX variant of Jonathan Kew, to typeset these stories in a book, and figure 7 shows a (reduced) page from it. (Please visit www.EZChinese.com to download an extended excerpt from this book.) This book is organized as a sequence of numbered panels, where each panel dis-

UNIT 6, §§72-77		UP AND DOWN	
72	埋 埋 bury	■ earth, soil 9 ■ inside 71	mái
When are you part of the <i>earth’s insides</i> ? When they come to bury you!			
[10 STROKES RANK 1640]			
73	理 理 rational	■ king 11 ■ inside 71	lǐ
How does a king get to be in charge, at least in theory? The <i>king’s insides</i> supposedly endow him with superior logic and rationality , so much so that he deserves our fealty.			
[11 STROKES RANK 89]			
74	哩 哩 enumeration	■ mouth 41 ■ inside 71	lǐ
I’ve chosen a slightly non-standard use for this glyph. The enumeration I refer to has to do with listing items:			
Yīngwén lì, Déwén lì, Fǎwén lì, tā dōu huì shuō.			
<i>He can speak English, German, and French.</i>			
When speaking, you enumerate items by speaking each item (with your <i>mouth</i> , naturally) as it bubbles up from within (<i>inside</i>) you.			
[10 STROKES RANK 1949]			
75	唱 唱 sing	■ mouth 41 ■ flourishing 69	chàng
It’s possible to think of song as enhanced speaking, that is, the expressive powers of the singing mouth have <i>flourished</i> beyond normal.			
[11 STROKES RANK 1252]			
76	晶 晶 crystal	■ day, sun 63 ■ day, sun 63 ■ day, sun 63	jīng
BF The light on the facets of a crystal , so bright and brilliant, looks just like several <i>suns</i> .			
[12 STROKES RANK 1725]			
77	量 量 measure	■ dawn 65 ■ inside 71	liáng
At the break of <i>dawn</i> , visibility improves <i>inside</i> your room, making it easy to measure things by natural light.			
[12 STROKES RANK 241]			

Figure 7: A page from the book *Chinese Characters*.

plays the same organization. I begin with the panel number, followed by the current character displayed using two fonts. I don’t know how the Chinese characterize them, but I think of them as the Chinese equivalent of roman and of sans-serif. Also on this line in bold-face is the primary definition for the character, a geometric decomposition of the character into its predecessors, and the pinyin recipe for its primary pronunciation. When you write Chinese, you imagine that each character lies in a single square, and in these decompositions, blackened parts of a character square give the reader an idea of what parts of the finished character belong to which preceding characters. It’s straightforward for T_EX keep track of which panels belong to which definition, and to print this information with the component information. That way, it’s easy to refer to an earlier panel to remind yourself what they look like. The term *pinyin* refers to the official transcription method developed by and used in mainland China.

The mnemonic narrative itself forms the central part of the panel, and I depend on certain typographic cues in the story. Any words in **boldface** refer to the definition of the current character, and references to its components always appear in *italic*.

This is Linux Libertine ff ffi
 ffl fi fl Th ABCDE FGHIJ 0 1
 2 3 4
Bold *Italic Bold italic*

Figure 8: Linux Libertine: a small sample.

EZChinesey.com
 info@EZChinesey.com
 EZChinesey@gmail.com

Figure 9: Publishing venture and contact information.

Finally, there’s some other information just for fun—the stroke count and the frequency rank of each character.

Incidentally, the roman typeface I used for the book is the family Linux Libertine, an OpenType family designed by Philipp Poll and available for free use (fig. 8). It’s a snap to install these fonts for use by X_YTeX (at least it is on the Mac platform), and when you do so properly you get the entire suite of proper TeX behavior, including small caps, all the standard ligatures, all special German ligatures, and an intriguing ‘T-h’ ligature which you can see in the figure.

Anyway, the result is an actual book containing the stories for about a hundred components and twenty-one hundred some-odd characters. For more fun, and for the thrill of fulfilling a long-held dream, I decided to print and publish this book on my own.

To that end, my wife and I set up a small publishing company EZChinesey.com; can you guess its associated web site? I’m not exactly rolling in royalties, but running your own company turns out to be a great adventure with lots of unexpected twists and turns. I recommend it highly (fig. 9). By the way, I know I’m no typographer, and so I am desperately seeking feedback to improve the format of the character panels, and I hope anyone and everyone with better ideas will feel free to bring them to my attention. I encourage anybody with suggestions for improvements to please get in touch with me.

This book is the first of what I hope will be a series of ‘EZChinesey Guides’. A second one will appear later this summer, and it’ll be a guide for travelers to China who want to eat in local restaurants. Think of it as a menu translation guide, with translations for over 3000 Chinese menu items. Its format is quite different from that of my character volume; figure displays a typical page.

Eating Out in China	POULTRY AND EGGS [Quail]
9 Sautéed pigeon breast in lemon sauce	jīn níng rǔ gē pú .. 金柠乳鸽脯 802
10 Sautéed pigeon breast in XO sauce	XO jiàng bào rǔ gē pú .. XO 酱爆乳鸽脯 803
11 Sautéed sliced pigeon breast with vegetables	shí cài chǎo gē pú .. 时菜炒鸽脯 804
12 Steamed pigeon in lotus leaves	hé xiāng zhēng rǔ gē .. 荷香蒸乳鸽 805
13 Sautéed pigeon with mushrooms and ham	běi gū yún tuī zhēng rǔ gē .. 北菇云腿蒸乳鸽 806
14 Stewed pigeon with gastrodia tuber	tiān má dùn rǔ gē .. 天麻地乳鸽 807
16 Quail	
1 Braised quail breast with black bean sauce	dòu chí ān chún pú .. 豆豉鹌鹑脯 808
2 Crispy quail with sesame	sū xiāng ān chún .. 酥香鹌鹑 809
3 Sautéed quail breast with coriander	yuán bào ān chún pú .. 芫爆鹌鹑脯 810
4 Stewed quail with aweto	dōng cǎo huā dùn zhè gū .. 冬草花炖鹌鹑 811
17 Soups	
1 Braised chicken breast in pumpkin soup	jīn tāng huì jī pú .. 金汤烩鸡脯 812
2 Minced chicken soup	jī róng dòu huā .. 鸡茸豆花 813
18 Tomatoes	
1 Eggs and tomato wedges stir-fried	mù xū shì zǐ .. 木须柿子 814
F Beef Niú ròu lèi 牛肉类	
1 BBQ (Barbecue)	
1 Beef BBQ with spicy sauce	ē xiāng pò shí tou kǎo ròu .. 阿香婆石头烤肉 815
2 Beef ribs	
1 Pan-fried beef ribs with black peppers	hēi jiāo niú lèi gǔ .. 黑椒牛肋骨 816
2 Sautéed beef ribs	bào chǎo niú lèi gǔ .. 爆炒牛肋骨 817
3 Sautéed calf ribs with spicy salt	jiāo yán niú zǐ gǔ .. 椒盐牛仔骨 818
4 Sizzling calf ribs with papaya	tiě bǎn mù guā niú zǐ gǔ .. 铁板木瓜牛仔骨 819
5 Spare ribs with soy sauce and fermented soy beans	chǐ zhī pái gǔ .. 豉汁排骨 820
6 Steamed beef ribs in black bean sauce	chǐ zhī niú zǐ gǔ .. 豉汁牛仔骨 821
3 Beef tendon	
1 Braised beef tendon in brown sauce	hóng shāo niú tí jīn .. 红烧牛蹄筋 822
4 Beef	
1 Beef filet with tomato sauce, Chinese style	zhōng shì niú liú .. 中式牛柳 823
2 Beef steak with tomato sauce, Chinese style	zhōng shì niú pái .. 中式牛排 824
3 Beef tenderloin with mixed vegetables	shí jīn pà niú ròu .. 什锦扒牛肉 825
4 Beef braised in brown sauce	hóng shāo niú ròu .. 红烧牛肉 826
5 Beef curry	kā lǐ niú ròu .. 咖喱牛肉 827

Figure 10: A typical page from the book *Eating Out in China*.

A third volume will be similar to the one I’ve spoken about today, but it will deal with the traditional characters, instead of the simplified ones in the current book.

I’m anxious to explore a different format for this upcoming volume, the traditional character volume, and I’d like to show a preliminary version to you, again with the goal of soliciting suggestions for improvement (fig. 11). New information includes the stroke order diagrams you see, which help when you review characters you’ve learned. Perhaps you recall this format is slightly different from the panels in figure 7.

I’ve typeset a sample panel twice, showcasing some interesting typefaces I’ve discovered. These fonts are from the collection that Google seems to

唐 Tang dynasty ■ age 689 (altered) □ mouth 55
 之 产 产 产 产 唐 唐 唐 唐_{10str}
 §690
 táng The great Tang dynasty was a high point of Chinese civilization. The
 Rank 1687 hallmark of this great age was the *verbal* arts, particularly poetry.

唐 Tang dynasty ■ age 689 (altered) □ mouth 55
 之 产 产 产 产 唐 唐 唐 唐_{10str}
 §690
 táng The great Tang dynasty was a high point of Chinese civilization. The
 Rank 1687 hallmark of this great age was the *verbal* arts, particularly poetry.

Figure 11: Proposed format changes, showing the use of Old Standard (top) and Droid Serif fonts (bottom).

be assembling to make the Web a more beautiful place, and, I suppose, to improve legibility on small screens under the control of their Android operating system. The ones I’ve tried have worked out-of-the-box with X_YTEX.

In the upper sample in figure 11, I’ve used the ‘Old Standard’ font by Alexei Kryukov. Despite its name, it seems to be the first T_EX-aware font I know of to be truly in the Monotype Modern Roman family, a set of families long out of favor but very beautiful and readable. Monotype Modern fonts do live on in the Computer Modern fonts most of us use all the time (Monotype Modern 8a served as model and muse for Computer Modern).

The second sample uses Droid Serif and Droid Sans Serif fonts, developed by Steve Matteson. Chinese reference works tend to be very dense with respect to information presented, and I think I can get away with smaller font sizes as a result of the enhanced legibility of these fonts.

It was only at the conclusion of producing this first book that I realized I learned an important lesson from it. At the risk of making old T_EX hands, let me pass this insight on in hopes that they will be useful to you, especially to any newbies reading this.

At the outset, I viewed this project as if it were a simple letter or article writ large, and that’s where I went wrong. It is far better to view this project as one centered around data management, where typesetting and whatever else occupies the periphery. Here’s why.

This project — and similar ones such as catalogs, dictionaries and encyclopedias, collections of letters — comprise a list of records, each of which

```
record
definition Tang dynasty enddefinition
character 唐 endcharacter
pinyin tang2 endpinyin
rank 1687 endrank
story ... endstory
endrecord
```

Figure 12: Very simple and abbreviated data record format for “Chinese Characters”.

has a more-or-less identical structure. In my case, each record corresponds to a Chinese character, and we can imagine the structure for these records is something like the one in figure 12.

The typeset format that you envision on the first day of a project is never the one you finally end up using. Yet if you forgo the data format in favor of creating a T_EX typesetting command right from the get-go, you’ll find it challenging to force a command created for an original format to easily satisfy the needs of a possibly very different format. It’s much better to create the sequence of records with an accompanying script in a language like Perl or Python to translate the data file into a T_EX source file. One big win here is that it’s straightforward to alter this script to accommodate changes in T_EX commands when I change formats. That way, your T_EX source more closely parallels your book’s format and makes the T_EX coding much simpler and more straightforward. You can’t use such a script to make typesetting decisions, but there’s usually plenty of other stuff for which Perl works just fine.

Moreover, you may need to generate auxiliary materials, material not typically the kind that, say, L^AT_EX knows how to generate. In my case, such materials include flashcard files, review material, and graded reading practice, and in other contexts might include price lists, special-purpose indices, and glossaries. In my experience with EZChinese, it’s much easier to generate this lot from well-organized data files rather than from messy-looking T_EX source.

Well, there are no further comments I need to make. Thanks for your attention during this article, most of which was just shameless self-promotion!

◇ Alan Hoenig
 EZChinese.com