Twitter Thread by zev handel



zev handel
@ZevHandel

Last month I presented seven sentences in seven different languages, all written in a form of the Chinese-character script. The challenge was to identify the languages and, if possible, provide a

> \U0001f17b\U0001f130\U0001f17d\U0001f176\U0001f184\U0001f130\U0001f176\U0001f174
> \U0001f180\U0001f184\U0001f178\U0001f189
>
> The following sentences are in seven different languages, all written in Chinese-character script (or a modification of it). Can you identify the languages?
>
> Sentences are in thread.
>
> (1/3)
>
> — zev handel (@ZevHandel) December 17, 2020

Here again are those seven sentences:

1) ■■■■■■■■■■■■
2) ■■■■■■■■■■■■■■■
3) ■■■■■■■■■■■■■■■
4) ■■■■■■■■
5) ■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■
6) ■■■■■■■■■
7) ■■■■■■■■■■■■■■

Six of those seven sentences are historically attested. One is not: I invented #7. I'm going to dive into an exploration of that seventh sentence in today's thread.

Sentence #7 is an English-language sentence written sinographically — that is, using graphs that originate in the Chinese script. I didn't do this for fun (even though it is fun), or as a proposal for a new way to write English.
https://t.co/W9WNof0edI

> 7) \u4eca\u5929\u611b\u665a\u7279\u8a9e\u5154\u5403\u4e8c\u9b5a\u4f5b\u5348\u98ef \u2013 Modern English
> Today I want you to each two fish for lunch.

> That this sentence is a written form of English is undeniable, as the sentence is made up entirely of English words following the rules of English grammar. 23/
>
> — zev handel (@ZevHandel) December 21, 2020

I did it as a thought experiment. Why? Because thinking about how the modern Chinese script might be adapted to write modern English can give us valuable insights into historical instances of script borrowing, like those that took place centuries ago in Japan, Korea, and Vietnam.

So here's the amazing thing about Sentence #7 (■■■■■■■■■■■■■■■):

It superficially looks like nonsense. I didn't explain anything about it, I just put it out there. And yet quite a few people deciphered it without any hints, and pretty quickly too. How did they do it?

Understanding this written sentence requires (a) ability to read Standard Written Chinese and pronounce it aloud in Modern Mandarin, and (b) knowledge of modern English.

■■■■■■■■■■■■■■
"Today I want you to each two fish for lunch."

What's amazing is that anyone with those two skills who spends a bit of time with the sentence is able to crack the code without too much difficulty. This tells us something about the way Chinese characters function and how they can be repurposed to write different languages.

Let's first see how this works in the English sentence, and then we can talk about the implications for how Korean, Japanese, and Vietnamese speakers first figured out how to commit their spoken languages to writing using Chinese characters.

First, I'll add some spaces to the written sentence to provide some structure for our discussion:
■■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■■

The sentence is made up of eleven units, each composed of one or two characters.

If you can read Chinese and speak Mandarin, then each unit has two conventialized properties: a pronunciation and a meaning. We can schematize this relationship this way:

Character(s): {pronunciation, meaning}

I'll give pronunciations in pinyin, and meanings in English.

■■ : {j■nti■n, 'today'}
■ : {ài, 'love'}
■ : {w■n, 'late'}

■ : {tè, ‘special’}
■ : {y■, ‘spoken language’}
■ : {tù, ‘rabbit’}
■ : {ch■, ‘eat’}
■ : {èr, ‘two’}
■ : {yú, ‘fish’}
■ : {fó, ‘Buddha’}
■■ : {w■fàn, ‘lunch’}

If for each unit you focus on just the pronunciation or just the meaning, ignoring the other property, then you can get a coherent English sentence. Like this:

TODAY ài w■ntè y■ tù EAT TWO FISH fó LUNCH >
“Today I want you to eat two fish for lunch.”

I have to point out again how amazing it is that people figured this out. I did not say that this sentence was in English. I did not explain that each graph is used for either its sound or its meaning (but not both). I gave no guidance at all on the writing technique I was using.

And still people figured it out. HOW?

Well, it turns out that humans have a natural, intuitive ability to split apart those pronunciation-meaning pairs. Even little kids do it without prompting.

For literate Mandarin speakers, the meanings and pronunciations come easily. For English speakers, knowledge of the lexicon and grammar of English constrains the possible interpretations of the Chinese characters.

For someone who knows both languages, the result is an ability to read the sentence — EVEN WITHOUT BEING TOLD HOW TO DO IT.

You’ve probably guessed by now that this is how Chinese characters were adapted to write spoken languages that had never before been written. And this is indeed how ancient forms of spoken Japanese, Korean, and Vietnamese came to be written down for the first time in history.

But that’s only the first part of the story.

Because … reading this way may be simple, but it’s also hard. It’s slow. It’s tiring. You can almost feel the cognitive load weighing on your brain as you work through the sentence. Why?

Because you are testing multiple possibilities for each character, constructing partial English sentences in your mental workspace, then redoing and refiguring each time you hit a dead end.

It sure would be easier if there was some indication of which characters were being used for their sound value and which characters were being used for their meaning value, wouldn’t it? It would take a lot of the guessing and second-guessing out

of the process.

Does “■ : {èr, ‘two’}” write the English word ‘are’ or the English word ‘two’?

Does “■ : {yú, ‘fish’}” write the English word ‘you’ or the English word ‘fish’?

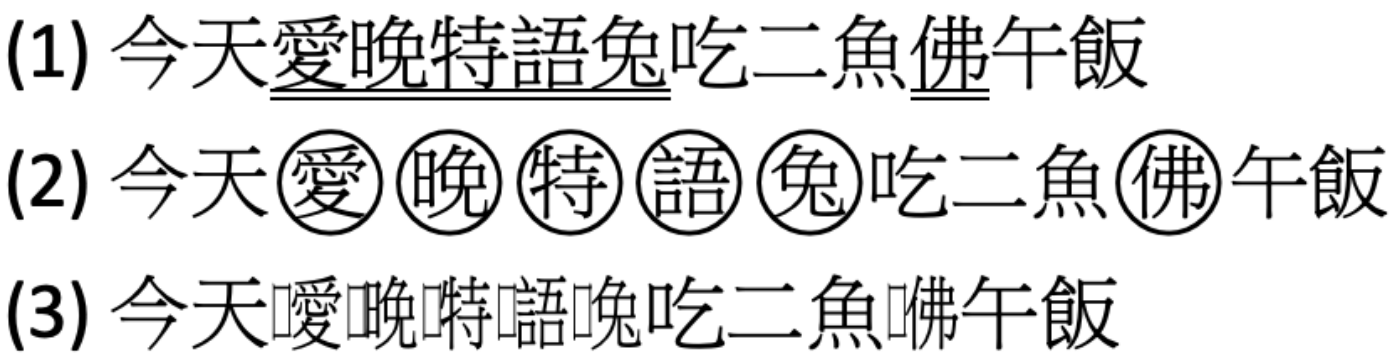
Does “■ : {tù, ‘rabbit’}” write the English word ‘to’, ‘too’, ‘two’, or ‘rabbit’?

Each time you encounter one of these characters, you have to pause to figure out which possibility best fits the context of the partial English sentence you’ve deciphered so far.

This is the problem of ambiguity — and it was a major factor in motivating further adaptations of Chinese script to write down spoken languages other than Chinese.

You can probably think of many ways to visually indicate which characters should be read for their meaning and which should be read for their sound. Unfortunately, the formatting limitations of Twitter make it difficult to illustrate them in text. I’ll provide pictures.

You might imagine using some kind of visual marking, like (1) underlining or (2) enclosing, that a writer could use to indicate to readers which characters are used for their pronunciation. Or maybe (3) putting a special semantic indicator like ■ next to such characters.



There are still other conceivable techniques for graphic differentiation that perhaps don’t come to your mind as quickly.

For example, you might simplify or abbreviate those graphs used for their sound value so that they become visually distinctive from those used for their meaning value, perhaps changing their whole look and feel to be more angular, or (as in (4)) more curvy.



And then there are techniques that aren't graphic at all. You might come to the conclusion that it's really the grammatical bits of English (plural -s, verbal -ing, prepositions, pronouns) that are best suited to sound-based representation ...

and the more lexically solid bits (verbs, nouns) that are easier to represent using characters for their meaning. In this way, the very grammatical structure of English would help you to parse the correct interpretation of each character.

If you were to do that for Sentence #7, it would lead you to adapt the characters this way:
meaning-based adaptation: want, eat, two, fish, dinner
sound-based adaptation: I, you, to, for

Or … and now things get really interesting … you might even decide that the best way to avoid ambiguity entirely is to specify both the sound and meaning of the target English word *at the same time*! How could you do this?

Well, consider the English word 'two'. Its meaning could be represented by the character ■ : {èr, 'two'}. Its sound could be represented by the character ■ : {tù, 'rabbit'}. What if we stacked or combined the two characters into a single representation of the English word?

That is, ■■■ or ■■■. That's a cool way to represent {/tu/, 'two'}! Not only is it unambiguous (it can't possibly write 'rabbit', 'to', 'too' or 'are'), it also isn't a Chinese character, so it's clear that what is being written here is not a Chinese word but an English one.



Once again you can probably guess where this is going! Gotta stop here for now, so we'll pick all this up next time and see exactly how these few simple ideas explain how written forms of spoken Japanese, Korean, and Vietnamese first developed.