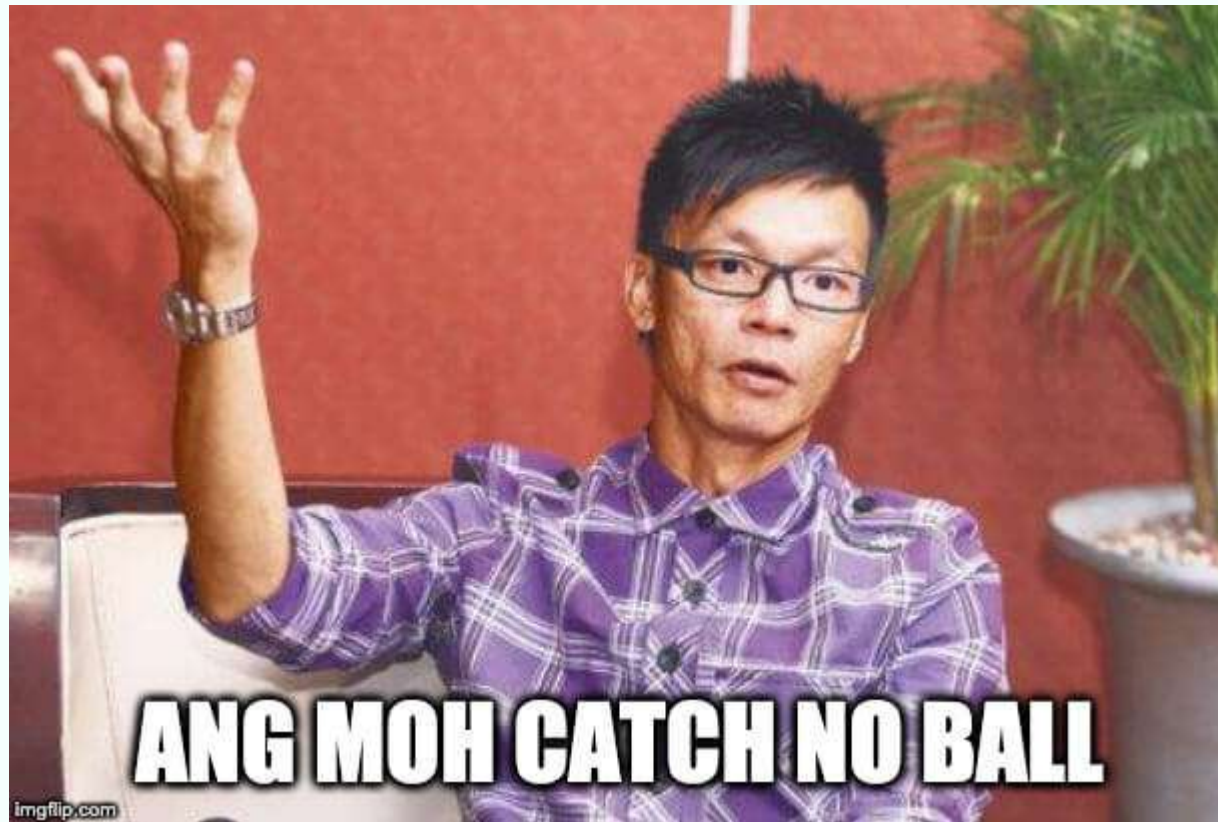


Do your regional websites talk funny?

By Chen Hui Jing / @hj_chen

Singlish-o-meter







陈	慧	晶
Chen	Hui	Jing



@hj_chen



“Internationalisation is the design and development of a product, application or document that enables easy localisation for target audiences that vary in culture, region, or language.”

ANYONE. CAUGHT. SHOPLIFTER WILL
BE SHOT 100 TIMES THE VALUE OF THE
STOLEN ITEM (S)
ขโมยสินค้าปรับ 100 เท่า



一旦失窃要报警，切莫姑息又养奸
If you are stolen, call the police at once.
上海市公安局城市轨道交通分局
Urban Mass Transportation Branch Shanghai Public Security Bureau
www.english.com

남자화장실은
매장 외부 우측 계단으로~
A man toilet buries you
to external right side stairs~

โปรดอย่าทิ้งก้นบุหรืลงในโถส้วม
PLEASE DO NOT DROP BUTT IN THE TOILET BOWL

The “Billion” problem

The Chinese numeral system (also used by Japanese and Korean) has specific words for large numbers as per the traditional Chinese grouping of 10,000.

	Simp. Chinese	Trad. Chinese	Japanese	Korean
10	十 (shí)	十	十 (juu)	십 (ship)
100	百 (bǎi)	百	百 (hyaku)	백 (baek)
1000	千 (qiān)	千	千 (sen)	천 (cheon)
10,000	万 (wàn)	万	万 (man)	만 (man)
100,000,000	亿 (yì)	億	億 (ichioku)	억 (eok)
1,000,000,000,000*	兆 (zhào)	兆	兆 (icchou)	조 (jo)

Translation strings

```
import {useTranslation} from "react-i18next";

function HeaderComponent() {
  const {t, i18n} = useTranslation('common');
  return <h1>{t('welcome.title')}</h1>
}
```

```
{
  "welcome": {
    "title": "Welcome to the app!"
  }
}
```

```
{
  "welcome": {
    "title": "Laipni lūdzam lietotnē!"
  }
}
```


Interpolation

```
/* src/translations/en.json */  
{  
  "welcome": {  
    "title": "Welcome to {{framework}}"  
  }  
}
```

```
/* src/translations/de.json */  
{  
  "welcome": {  
    "title": "Willkommen bei {{framework}}"  
  }  
}
```

What translators see

Not this:

Over **1,000,000 businesses** in **175 countries** around the world have made over **\$200 billion USD** in sales using Shopify.

But this:

```
Over %{total_stores} businesses in %{total_countries} countries around the world have made over $%  
{total_gmv_billions} billion USD in sales using Shopify
```

The end result

借助 Shopify，全球 **175** 个国家/地区 有超过**1,700,000** 家企业 创造了逾 **200 0** 亿美元 的销售额。

在全球 **175** 個國家/地區，有超過 **1,700,000** 間商家透過 Shopify 締造了 **200** 十億美元以上的銷售額。

175개 국가에서 **1,700,000개** 사업체 이상이 Shopify를 사용해 **200십억 달러(USD)** 이상의 매출을 올렸습니다.

The problem

Saying that the GMV is 200 ten million dollars is quite a glaring grammatical error, if you speak the language.

But why didn't Japan have this problem?

世界175か国で1,700,000のビジネスがShopifyを使用し、流通総額は\$2000億米ドルを超えています。

The Japanese team had worked around this issue previously by editing the locale files that reference our total GMV numbers like so:

```
%{total_gmv_billions}0億米ドル
```

Our Japan team spotted this error and made the adjustment of replacing 十 with a 0.

See solution, steal it

This will be how it looks for each respective language:

```
%{total_gmv_billions}0억 달러(USD)  
%{total_gmv_billions}0 亿美元  
%{total_gmv_billions}0 億美元
```

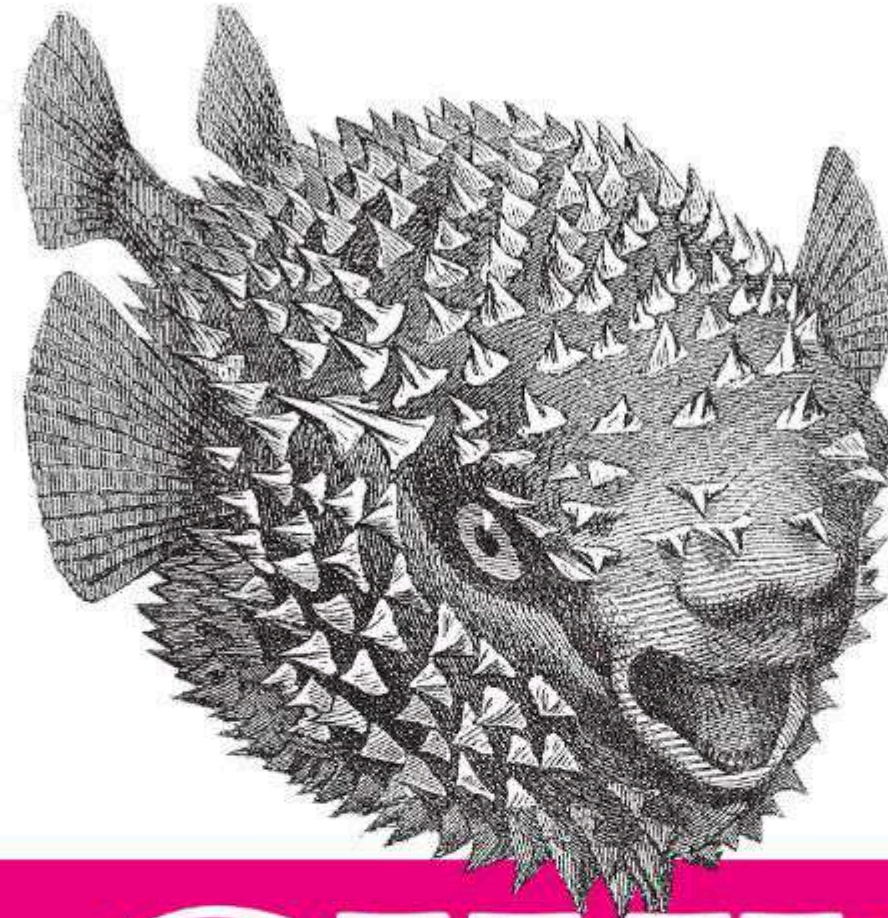
Lessons and best practices

- Interpolate with caution
- Do Not Manually Construct Sentences or Manipulate Text in Code
- Let i18n Libraries Handle the Hard Stuff

Lessons From Linguistics: i18n Best Practices for Front-End Developers by Lucas Huang

*Chinese, Japanese, Korean
& Vietnamese Computing*

2nd Edition
Broader Unicode Coverage



Information Processing

O'REILLY®

Ken Lunde

Unicode character ranges

- 0020–007F Basic Latin
- 00A0–00FF Latin-1 Supplement
- 0100–017F Latin Extended-A
- 0180–024F Latin Extended-B
- 0250–02AF IPA Extensions
- 02B0–02FF Spacing Modifier Letters
- 0300–036F Combining Diacritical Marks
- 0370–03FF Greek and Coptic
- 0400–04FF Cyrillic
- 0500–052F Cyrillic Supplementary
- 0530–058F Armenian
- 0590–05FF Hebrew
- 0600–06FF Arabic
- 0700–074F Syriac
- 0780–07BF Thaana
- 0900–097F Devanagari
- 0980–09FF Bengali
- 0A00–0A7F Gurmukhi
- 0A80–0AFF Gujarati
- 0B00–0B7F Oriya
- 0B80–0BFF Tamil
- 0C00–0C7F Telugu
- 0C80–0CFF Kannada
- 0D00–0D7F Malayalam
- 0D80–0DFF Sinhala
- 0E00–0E7F Thai
- 0E80–0EFF Lao
- 0F00–0FFF Tibetan
- 1000–109F Myanmar
- 10A0–10FF Georgian
- 1100–11FF Hangul Jamo
- 1200–137F Ethiopic
- 13A0–13FF Cherokee
- 1400–167F Unified Canadian Aboriginal Syllabics
- 1680–169F Ogham
- 16A0–16FF Runic
- 1700–171F Tagalog
- 1720–173F Hanunoo
- 1740–175F Buhid
- 1760–177F Tagbanwa
- 1780–17FF Khmer
- 1800–18AF Mongolian
- 1900–194F Limbu
- 1950–197F Tai Le
- 19E0–19FF Khmer Symbols
- 1D00–1D7F Phonetic Extensions
- 1E00–1EFF Latin Extended Additional
- 1F00–1FFF Greek Extended
- 2000–206F General Punctuation
- 2070–209F Superscripts and Subscripts
- 20A0–20CF Currency Symbols
- 20D0–20FF Combining Diacritical Marks for Symbols
- 2100–214F Letterlike Symbols
- 2150–218F Number Forms
- 2190–21FF Arrows
- 2200–22FF Mathematical Operators
- 2300–23FF Miscellaneous Technical
- 2400–243F Control Pictures
- 2440–245F Optical Character Recognition
- 2460–24FF Enclosed Alphanumerics
- 2500–257F Box Drawing
- 2580–259F Block Elements
- 25A0–25FF Geometric Shapes
- 2600–26FF Miscellaneous Symbols
- 2700–27BF Dingbats
- 27C0–27EF Miscellaneous Mathematical Symbols-A
- 27F0–27FF Supplemental Arrows-A
- 2800–28FF Braille Patterns
- 2900–297F Supplemental Arrows-B
- 2980–29FF Miscellaneous Mathematical Symbols-B
- 2A00–2AFF Supplemental Mathematical Operators
- 2B00–2BFF Miscellaneous Symbols and Arrows
- 2E80–2EFF CJK Radicals Supplement
- 2F00–2FDF Kangxi Radicals
- 2FF0–2FFF Ideographic Description Characters
- 3000–303F CJK Symbols and Punctuation
- 3040–309F Hiragana
- 30A0–30FF Katakana
- 3100–312F Bopomofo
- 3130–318F Hangul Compatibility Jamo
- 3190–319F Kanbun
- 31A0–31BF Bopomofo Extended
- 31F0–31FF Katakana Phonetic Extensions
- 3200–32FF Enclosed CJK Letters and Months
- 3300–33FF CJK Compatibility
- 3400–4DBF CJK Unified Ideographs Extension A
- 4DC0–4DFF Yijing Hexagram Symbols
- 4E00–9FFF CJK Unified Ideographs
- A000–A48F Yi Syllables
- A490–A4CF Yi Radicals
- AC00–D7AF Hangul Syllables
- D800–DB7F High Surrogates
- DB80–DBFF High Private Use Surrogates
- DC00–DFFF Low Surrogates
- E000–F8FF Private Use Area
- F900–FAFF CJK Compatibility Ideographs
- FB00–FB4F Alphabetic Presentation Forms
- FB50–FDFF Arabic Presentation Forms-A
- FE00–FE0F Variation Selectors
- FE20–FE2F Combining Half Marks
- FE30–FE4F CJK Compatibility Forms
- FE50–FE6F Small Form Variants
- FE70–FEFF Arabic Presentation Forms-B
- FF00–FFEF Halfwidth and Fullwidth Forms
- FFF0–FFFF Specials
- 10000–1007F Linear B Syllabary
- 10080–100FF Linear B Ideograms
- 10100–1013F Aegean Numbers
- 10300–1032F Old Italic
- 10330–1034F Gothic
- 10380–1039F Ugaritic
- 10400–1044F Deseret
- 10450–1047F Shavian
- 10480–104AF Osmanya
- 10800–1083F Cypriot Syllabary
- 1D000–1D0FF Byzantine Musical Symbols
- 1D100–1D1FF Musical Symbols
- 1D300–1D35F Tai Xuan Jing Symbols
- 1D400–1D7FF Mathematical Alphanumeric Symbols
- 20000–2A6DF CJK Unified Ideographs Extension B
- 2F800–2FA1F CJK Compatibility Ideographs Supplement
- E0000–E007F Tags

Unicode 15.0 Character Code Charts

The Vietnamese alphabets are listed in several non-contiguous Unicode ranges:

- Basic Latin {U+0000..U+007F}
- Latin-1 Supplement {U+0080..U+00FF}
- Latin Extended-A, -B {U+0100..U+024F}
- Latin Extended Additional {U+1E00..U+1EFF}
- Combining Diacritical Marks {U+0300.. U+036F}
- The Vietnamese đồng currency symbol is đ (U+20AB)

Missing Vietnamese glyphs

Đồng hành bên bạn trên mọi nẻo hành trình

Một nền tảng có mọi tính năng thương mại điện tử và điểm bán hàng bạn cần để khởi đầu, vận hành và phát triển công việc kinh doanh.

Bán hàng ở mọi nơi

Sử dụng một nền tảng để bán sản phẩm bất cứ đâu, cho bất kỳ đối tượng nào – bán trực tiếp với Điểm bán hàng và bán online qua trang web, mạng xã hội và thị trường trực tuyến.

The `:lang()` pseudo-class

```
:lang(vi) {  
  font-family: -apple-system, BlinkMacSystemFont, sans-serif;  
}
```


Issue with Apple SD Gothic Neo

Default Apple Korean font, Apple SD Gothic Neo, does not support Latin-1 Supplement character set
e.g. Beyoncé would not render correctly

Francisco Pérez Pellicena
Shopify 파트너, Code Black Belt, 스페인

```
:lang(ko) {  
  font-family: 'Work Sans', -apple-system, BlinkMacSystemFont, sans-serif;  
}
```

Links

- [W3C Internationalization \(I18n\) Activity](#)
- [i18next, internationalisation-framework for Javascript: Best Practices](#)
- [Lessons From Linguistics: i18n Best Practices for Front-End Developers](#)

谢谢

감사합니다

ありがとうございます

Cảm ơn

