

# Pseudo-localization in **A**ction

Xiang Xu, Erwin Hom, Mike McKenna  
PayPal, Inc.









































## i18n Team at PayPal

---





















- works with developers
- teaches i18n
  - a set of best practices
  - used to build global-ready products

## Where PayPal does business

### Africa

 Algeria	 Angola	 Benin	 Botswana	 Burkina Faso
 Burundi	 Cameroon	 Cape Verde	 Chad	 Comoros
 Cote d'Ivoire	 Democratic Republic of the Congo	 Djibouti	 Egypt	 Eritrea
 Ethiopia	 Gabon Republic	 Gambia	 Guinea	 Guinea-Bissau
 Kenya	 Lesotho	 Madagascar	 Malawi	 Mali
 Mauritania	 Mauritius	 Mayotte	 Morocco	 Mozambique
 Namibia	 Niger	 Nigeria	 Republic of the Congo	 Reunion
 Rwanda	 Saint Helena	 Sao Tome and Principe	 Senegal	 Seychelles
 Sierra Leone	 Somalia	 South Africa	 Swaziland	 Tanzania
 Togo	 Tunisia	 Uganda	 Zambia	 Zimbabwe

### Americas

 Anguilla	 Antigua and Barbuda	 Argentina	 Aruba	 Bahamas
 Barbados	 Belize	 Bermuda	 Bolivia	 Brazil
 British Virgin Islands	 Canada	 Cayman Islands	 Chile	 Colombia
 Costa Rica	 Dominica	 Dominican Republic	 Ecuador	 El Salvador

## Important Terms

- **Internationalization (i18n)**
  - Developer Activity
- **Localization (L10n)**
  - Customization of Content
  - Translation
- **Localizability Testing**
  - Unit Test performed by Developers
  - App localization-friendly?
- **Pseudo-localization**
  - simulates localization

## Topics

---

- Pseudo-localization
  - A Brief Intro
- Pseudo-localization at PayPal
- Next Steps

---

November 5, 2014

Internationalization & Unicode Conference #38

5



## Problems encountered by Localization

---

- UI can't accommodate localized text
- UI messages contains grammatical errors or looks like a ransom note. 😞
- English shows up in the middle of a French sentence
- App stops working !!!
- Font Issues 🍂 → □ □

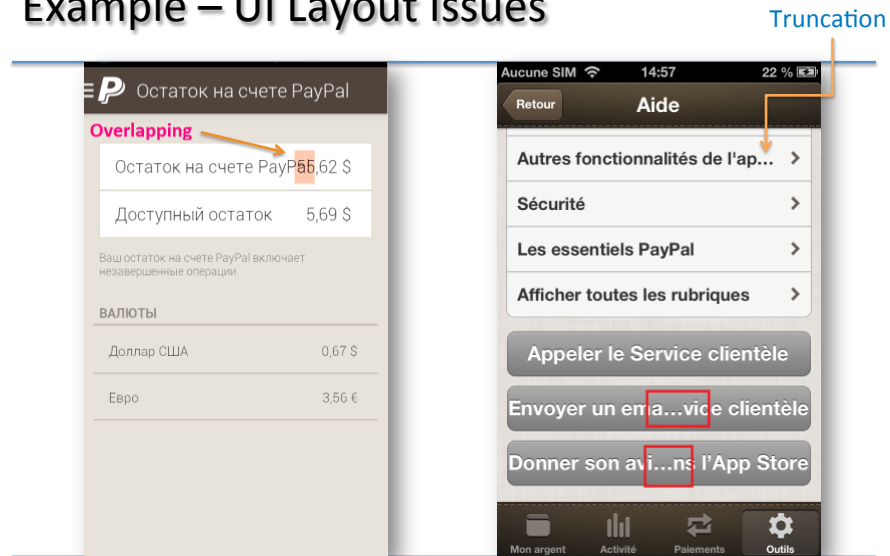
---

November 5, 2014

Internationalization & Unicode Conference #38

6

## Example – UI Layout Issues



November 5, 2014

Internationalization &amp; Unicode Conference #38

7



November 5, 2014

Internationalization &amp; Unicode Conference #38

8



## Pseudo-localization

---

A way to simulate translation while preserving readability for testing.

Character Substitution

Sentence → ©ēntēncē

Text Expansion

©ēntēncē \*\*\* \*\*\*\*\* \*

Bracketing

«©ēntēncē \*\*\* \*\*\*\*\* \*»

## Example: Pseudo-localized String Resource

### Good Source:

Msg1=Your account will be closed in {number} days.

Character substitution

### Target:

«Yōŭr āccōũnt wīll bē clōsēd īn [3] dāys. \*\*\* \*\*\*\*\* \*»

Text Expansion

### Bad Source:

Part1=Your account will be closed in

Part2= days.

Bracketing

### Target:

«Yōŭr āccōũnt wīll bē clōsēd īn \*\*\* \*\*\*\*\* \*»3«dāys. \*\*\* \*\*\*\*\*»

Text Fragmentation

November 5, 2014

Internationalization & Unicode Conference #38

11

## Pseudo-localization can reveal...

### ...these i18n issues:

1. UI can't accommodate localized text
2. UI messages contains grammatical errors or look like a ransom note. 😞
3. English shows up in the middle of a French sentence.
4. App stopped working !!!
5. Font Issues 🍊 → □ □

### Resolution

1. Make sure text wrapping is enabled.
2. Make sure each message is contained in a single string resource.
3. Expose Text for Translation
4. Private was exposed for translation. (DO NOT TRANSLATE)
5. Provide Font with adequate support.

November 5, 2014

Internationalization & Unicode Conference #38

12

## Challenges

---

- Content comes in many different formats



- Global products support many languages
- Developers not i18n-aware

---


November 5, 2014

Internationalization & Unicode Conference #38

13

## Developing a Pseudo-localization Tool

---

- Run it on the source content 
- Simulate 3 types of languages:
  - European, CJK, Right-to-Left
- Tool designed to be extensible
  - New Content File Types
  - Support for Complex Languages (i.e., Thai)
- Easily integrates into our CI (Continuous Integration) process

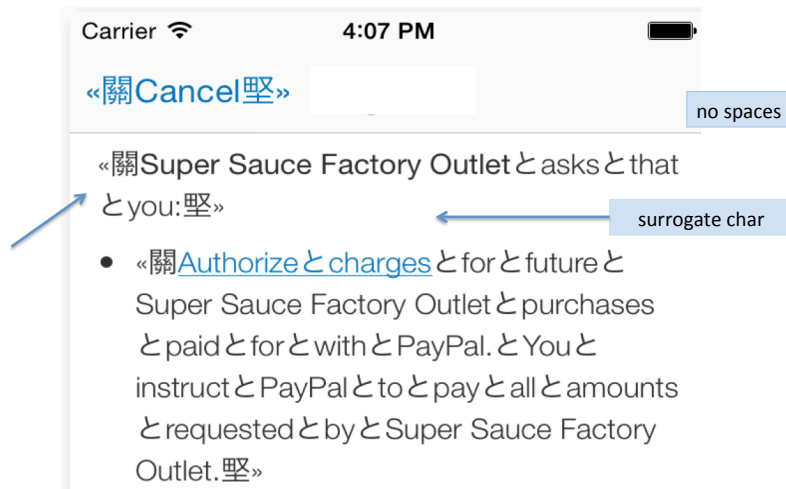
---

November 5, 2014

Internationalization & Unicode Conference #38

14

## Pseudo CJK localization

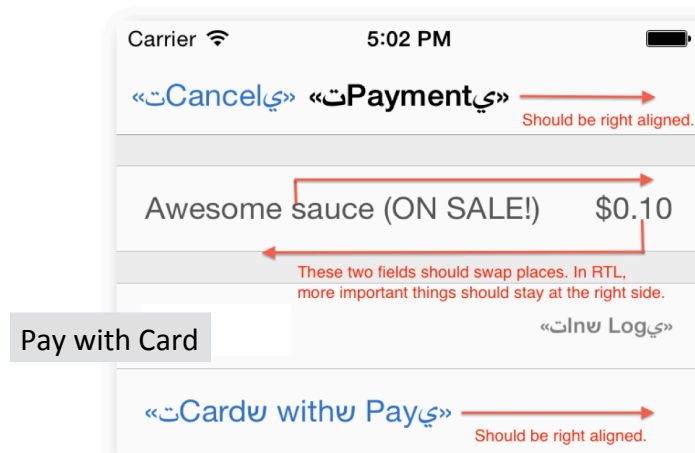


November 5, 2014

Internationalization &amp; Unicode Conference #38

15

## Pseudo RTL localization



November 5, 2014

Internationalization &amp; Unicode Conference #38

16



## Education & Processes

---

- Provide training for:
  - Developers, IQA, LQA, and Visual Designers
- Track Pseudo-localization adoptions and feedback

Before content is accepted for localization:

- Request evidence of Pseudo-localization

November 5, 2014

Internationalization & Unicode Conference #38

17

## Benefiting from Pseudo-localization

---

- Catches Bugs Early
- Saves the company money.
- Improved customer experience globally.



November 5, 2014

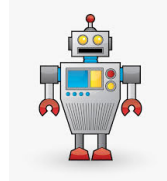
Internationalization & Unicode Conference #38

18

## Next Steps for Pseudo-localization

---

- Combine with Machine Translation



- Integrate with Framework



---

November 5, 2014

Internationalization & Unicode Conference #38

19

## Combining with Machine Translation

---

### Pseudo-localization

- ~~1. Substitute Characters~~
- ~~2. Text Expansion~~
3. Bracketing

Your account will be closed in {number} days.

«您的帐户将在【3】天内被关闭。」

---

November 5, 2014

Internationalization & Unicode Conference #38

20

## Example: Machine Translated String Resource

**Good Source:**

Msg1=Your account will be closed in {number} days.

**Target:**

«您的帐户将在【3】天内被关闭。」

(Your account will be closed in 3 days.)

**Bad Source:**

Part1=Your account will be closed in

Part2= days.

Bracketing

**Target:**

«您的帐户将被关闭»3«天。」

( Your account will be closed for 3 days.)

November 5, 2014

Internationalization & Unicode Conference #38

21

## Summary

- Pseudo-localization is an effective way to do Localizability Testing.
- A well-designed Pseudo-localization tool is essential.
- Developers, QA, and Visual Designers must be aware of Pseudo-localization.
- It will save you time and headaches.
- Most importantly ....

November 5, 2014

Internationalization & Unicode Conference #38

22



**Ⓟ̣sēũdσ-  
lσcālīzātīσn**

November 5, 2014 Internationalization & Unicode Conference #38 23

## Q&A

---

[xiangxu@paypal.com](mailto:xiangxu@paypal.com)  
[ehom@paypal.com](mailto:ehom@paypal.com)  
[mimckenna@paypal.com](mailto:mimckenna@paypal.com)



## References

---

- [Pseudo-localization](#) (Wikipedia)
- [Pseudo-localization](#) tool (Google)
- Globalization Step-by-Step: [Localizability Testing](#) (Microsoft)
- [Providing for UI Expansion](#) (IBM Globalization)
- [Localizability Testing](#) (Lingobit)



## APPENDIX

## Text Expansion

Number of Characters in Text	Additional Physical Space Required
1 to 4	200%
5 to 10	150%
11 to 20	90%
21 to 30	70%
31 to 50	50%
51 to 70	35%
Over 70	30%

November 5, 2014

Internationalization &amp; Unicode Conference #38

27

## Character Substitution Table

Source Character	Replacement Character	Unicode Code-point	Name
A	Ⓐ	U+24B6	Circled Latin Capital Letter A
B	Ⓑ	U+24B7	Circled Latin Capital Letter B
C	Ⓒ	U+24B8	Circled Latin Capital Letter C
D	Ⓓ	U+24B9	Circled Latin Capital Letter D
E	Ⓔ	U+24BA	Circled Latin Capital Letter E
F	Ⓕ	U+24BB	Circled Latin Capital Letter F
G	Ⓖ	U+24BC	Circled Latin Capital Letter G
H	Ⓖ	U+24BD	Circled Latin Capital Letter H
I	Ⓖ	U+24BE	Circled Latin Capital Letter I
J	Ⓖ	U+24BF	Circled Latin Capital Letter J
K	Ⓖ	U+24C0	Circled Latin Capital Letter K

November 5, 2014

Internationalization &amp; Unicode Conference #38

28

## Character Substitution Table (cont'd)

Source Character	Replacement Character	Unicode Code-point	Name
L	Ⓛ	U+24C1	Circled Latin Capital Letter L
M	Ⓜ	U+24C2	Circled Latin Capital Letter M
N	Ⓝ	U+24C3	Circled Latin Capital Letter N
O	Ⓞ	U+24C4	Circled Latin Capital Letter O
P	Ⓟ	U+24C5	Circled Latin Capital Letter P
Q	Ⓠ	U+24C6	Circled Latin Capital Letter Q
R	Ⓡ	U+24C7	Circled Latin Capital Letter R
S	Ⓢ	U+24C8	Circled Latin Capital Letter S
T	Ⓣ	U+24C9	Circled Latin Capital Letter T
U	Ⓤ	U+24CA	Circled Latin Capital Letter U
V	Ⓥ	U+24CB	Circled Latin Capital Letter V

November 5, 2014

Internationalization &amp; Unicode Conference #38

29

## Character Substitution Table (cont'd)

Source Character	Replacement Character	Unicode Code-point	Name
W	Ⓦ	U+24CC	Circled Latin Capital Letter W
X	Ⓧ	U+24CD	Circled Latin Capital Letter X
Y	Ⓨ	U+24CE	Circled Latin Capital Letter Y
Z	Ⓩ	U+24CF	Circled Latin Capital Letter Z
a	ă	U+0103	Latin Small Letter A with Breve
e	ē	U+0113	Latin Small Letter E with Macron
i	ĩ	U+0129	Latin Small Letter I with Tilde
o	ơ	U+01a1	Latin Small Letter O with Horn
u	ũ	U+0169	Latin Small Letter U with Tilde

November 5, 2014

Internationalization &amp; Unicode Conference #38

30