



An i18n crash course for the Indico Developer

Tomas Roun



Agenda for this presentation

- What is software translation and what are the things that need translating
- Software translation from a developer's point of view
- UI design with multiple languages
- Translation statistics (thanks to Dirk Hoffmann)

Why should we care about software translation?

Why should we care about software translation?

A Cellphone's Missing Dot Kills Two People, Puts Three More in Jail

By **Jesus Diaz** | Published April 21, 2008 | Comments (143)



The life of 20-year-old Emine, and her 24-year-old husband Ramazan Çalçoban was pretty much the normal life of any couple in a separation process. After deciding to split up, the two kept having bitter arguments over the cellphone, sending text messages to each other until one day Ramazan wrote "you change the topic every time you run out of arguments." That day, the lack of a single dot over a letter—product of a faulty localization of the cellphone's typing system—caused a chain of events that ended in a violent blood bath (Warning: offensive language ahead.)

Why should we care about software translation?

- Global userbase both at CERN and the UN
- A tool as complex as Indico requires a good level of English to use effectively

Language support in Indico

Indico already supports several languages



Language support in Indico

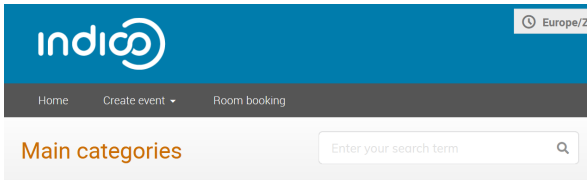
Indico already supports several languages



And lots more are being worked on!



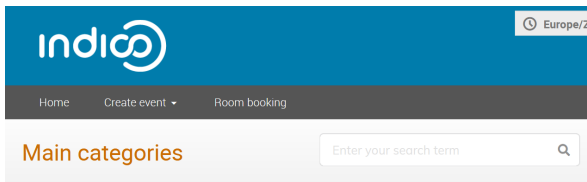
So how do we do it?



Welcome to Indico. The Indico tool allows you to manage complex conferences, workshops and meetings.

To start browsing, please select a category below.

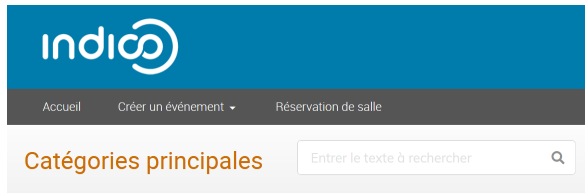
So how do we do it?



The screenshot shows the top navigation bar of the Indico website. The header is blue with the 'indico' logo on the left and a location dropdown menu on the right showing 'Europe/Z'. Below the header is a dark grey navigation bar with links for 'Home', 'Create event' (with a dropdown arrow), and 'Room booking'. Below this is a light grey section with the heading 'Main categories' on the left and a search input field on the right containing the placeholder text 'Enter your search term' and a magnifying glass icon.

Welcome to Indico. The Indico tool allows you to manage complex conferences, workshops and meetings.

To start browsing, please select a category below.



The screenshot shows the top navigation bar of the Indico website in French. The header is blue with the 'indico' logo on the left. Below the header is a dark grey navigation bar with links for 'Accueil', 'Créer un événement' (with a dropdown arrow), and 'Réservation de salle'. Below this is a light grey section with the heading 'Catégories principales' on the left and a search input field on the right containing the placeholder text 'Entrer le texte à rechercher' and a magnifying glass icon.

Bienvenue dans Indico. Le logiciel Indico permet de gérer des conférences complexes, des ateliers ou de simples réunions.

Pour commencer votre navigation, veuillez sélectionner une catégorie ci-dessous.

Internationalization & Localization

Internationalization (i18n)

Designing an application in a way that it can support multiple languages

- Detecting the user's language
- Support for regional and language preferences
- Extracting text for translation

Done by application developers

Internationalization & Localization

Internationalization (i18n)

Designing an application in a way that it can support multiple languages

- Detecting the user's language
- Support for regional and language preferences
- Extracting text for translation

Done by application developers

Localization (l10n)

Translating the text and other regional and language preferences such as:

- Date and time formats
- Number formats
- Currency
- And much more (we'll see later)

Done by translators

Examples



Shopping cart

Price: 49.99



Examples



Votre panier

Prix: 49.99



Examples



All text is translated, but what is the currency?



Examples



**Your appointment is confirmed
for 05/04/2023 at 12:00**



Exemples



**Votre rendez-vous est confirmé
pour le 05/04/2023 à 12h00**



Examples



**Votre rendez-vous est confirmé
pour le 05/04/2023 à 12h00**

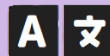
- Is the date **mm/dd** or **dd/mm**?
- What is the time zone?



Examples

Localization

Translation



Examples

Localization

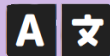
Date and time

dd/mm vs mm/dd

12-hour vs 24-hour

CET vs GMT

Translation



Examples

Localization

Date and time

dd/mm vs mm/dd

12-hour vs 24-hour

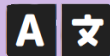
CET vs GMT

Numbers

1.5 vs 1,5

1,000,000 vs 1 000 000

Translation



Examples

Localization

Date and time

dd/mm vs mm/dd

12-hour vs 24-hour

CET vs GMT

Numbers

1.5 vs 1,5

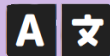
1,000,000 vs 1 000 000

Currency - €, \$, ¥

Phone numbers

Address formatting, Name formatting

Translation



Examples

Localization

Date and time

dd/mm vs mm/dd

12-hour vs 24-hour

CET vs GMT

Numbers

1.5 vs 1,5

1,000,000 vs 1 000 000

Currency - €, \$, ¥

Phone numbers

Address formatting, Name formatting

Units - km vs miles, pounds vs kgs

Pluralization

Writing direction - RTL, Vertical

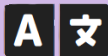
Punctuation - "" vs «»

Character sets - Unicode

Paper size & color

And much more...

Translation



What language preferences?

These definitions can even change over time!

What language preferences?

These definitions can even change over time!



Croatia just changed their currency from Kuna to Euro this year



Locale

All of these definitions together are called a **Locale**.

Some examples:

fr_FR - French (France)



fr_CH - French (Switzerland)



en_GB - English (United Kingdom)



en_US - English (United States)



...

Where do we get information about a specific locale?

Where do we get information about a specific locale?

The Unicode Common Locale Data Repository

“... largest and most extensive standard repository of locale data ...”

Freely available with many libraries built on top such as ICU for C/C++ & Java or babel for Python

unicode-org/**cldr**

The home of the Unicode Common Locale Data Repository



80 Contributors 0 Issues 676 Stars 331 Forks



Quick tour inside the Unicode CLDR project

```
1 {  
2   "sun": "dimanche",  
3   "mon": "lundi",  
4   "tue": "mardi",  
5   "wed": "mercredi",  
6   "thu": "jeudi",  
7   "fri": "vendredi",  
8   "sat": "samedi"  
9 }
```

Quick tour inside the Unicode CLDR project

```
1 {  
2     "sun": "dimanche",  
3     "mon": "lundi",  
4     "tue": "mardi",  
5     "wed": "mercredi",  
6     "thu": "jeudi",  
7     "fri": "vendredi",  
8     "sat": "samedi"  
9 }
```

```
1 {  
2     "midnight": "minuit",  
3     "noon": "midi",  
4     "morning1": "matin",  
5     "afternoon1": "après-midi",  
6     "evening1": "soir",  
7     "night1": "nuit"  
8 }
```

Quick tour inside the Unicode CLDR project

```
1 {  
2   "sun": "dimanche",  
3   "mon": "lundi",  
4   "tue": "mardi",  
5   "wed": "mercredi",  
6   "thu": "jeudi",  
7   "fri": "vendredi",  
8   "sat": "samedi"  
9 }
```

```
1 {  
2   "full": "EEEE d MMMM y G",  
3   "long": "d MMMM y G",  
4   "medium": "d MMM y G",  
5   "short": "dd/MM/y GGGGG"  
6 }
```

```
1 {  
2   "midnight": "minuit",  
3   "noon": "midi",  
4   "morning1": "matin",  
5   "afternoon1": "après-midi",  
6   "evening1": "soir",  
7   "night1": "nuit"  
8 }
```

Quick tour inside the Unicode CLDR project

```
1 {  
2   "sun": "dimanche",  
3   "mon": "lundi",  
4   "tue": "mardi",  
5   "wed": "mercredi",  
6   "thu": "jeudi",  
7   "fri": "vendredi",  
8   "sat": "samedi"  
9 }
```

```
1 {  
2   "full": "EEEE d MMMM y G",  
3   "long": "d MMMM y G",  
4   "medium": "d MMM y G",  
5   "short": "dd/MM/y GGGG"  
6 }
```

```
1 {  
2   "midnight": "minuit",  
3   "noon": "midi",  
4   "morning1": "matin",  
5   "afternoon1": "après-midi",  
6   "evening1": "soir",  
7   "night1": "nuit"  
8 }
```

```
1 { # Croatia  
2   "HRK": {  
3     "_from": "1994-05-30",  
4     "_to": "2023-01-15"  
5   },  
6   "EUR": {  
7     "_from": "2023-01-01"  
8   }  
9 }
```


Recap - Internationalization & Localization

Internationalization (i18n)

Designing an application in a way that it can support multiple languages

- Detecting the user's language
- Support for regional and language preferences
- Extracting text for translation

Done by application developers

Localization (l10n)

Translating the text and other regional and language preferences such as:

- Date and time formats
- Number formats
- Currency
- And much more (we'll see later)

Done by translators

Internationalization

How do we make this translatable?

```
1 form.title = "Create new meeting"
```

Internationalization

How do we make this translatable?

```
1 form.title = "Create new meeting"

1 if locale == "en_UK":
2     form.title = "Create new meeting"
3 else if locale == "fr_FR":
4     form.title = "Créer une nouvelle réunion"
5 else if locale == "es_ES":
6     form.title = "Crear nueva reunión"
7 # ... More locales here
```

We want to modify the source code as little as possible!

Use gettext!

- Standardized i18n system available in many programming languages
- Separates programming and translating - **gettext** works by looking up translations from an external list of translations

Use gettext!

- Standardized i18n system available in many programming languages
- Separates programming and translating - **gettext** works by looking up translations from an external list of translations

```
1 form.title = "Create new meeting"
```

Use gettext!

- Standardized i18n system available in many programming languages
- Separates programming and translating - **gettext** works by looking up translations from an external list of translations

```
1 form.title = "Create new meeting"
```

```
1 from gettext import gettext
2 form.title = gettext("Create new meeting")
3 # Créer une nouvelle réunion
```

gettext

So we just need to wrap all text in our application in a `gettext()` call?

Plurals

How do we translate this?

```
1 if n == 1:  
2     message = "The attachment has been uploaded"  
3 else:  
4     message = "{n} attachments have been uploaded".format(n)  
5  
6 # Example: n = 7 -> "7 attachments have been uploaded"
```


Plurals

How do we translate this?

```
1 if n == 1:
2     message = "The attachment has been uploaded"
3 else:
4     message = "{n} attachments have been uploaded".format(n)
5
6 # Example: n = 7 -> "7 attachments have been uploaded"
```

```
1 if n == 1:
2     message = gettext("The attachment has been uploaded")
3 else:
4     message = gettext("{n} attachments have been uploaded").
5     format(n)
```

Plurals



Plurals

Not every language has only one plural form.

[https://www.gnu.org/software/gettext/manual/html_node/Plural – forms.html](https://www.gnu.org/software/gettext/manual/html_node/Plural%20%2D%20forms.html)

Plurals

Not every language has only one plural form.

Polish has **2** plurals:

1 załącznik (= attachment)

2,3,4 załączniki

5-21 załączników

22-24 załączniki

25-31 załączników

...

Some languages like Arabic, Slovene and Irish have a **dual** - a special case for exactly two entities (objects, persons)

[https://www.gnu.org/software/gettext/manual/html_node/Plural – forms.html](https://www.gnu.org/software/gettext/manual/html_node/Plural%20forms.html)



Plurals

```
1 if n == 1:  
2     message = "The attachment has been uploaded"  
3 else:  
4     message = "{n} attachments have been uploaded".format(n)
```

Plurals

```
1 if n == 1:
2     message = "The attachment has been uploaded"
3 else:
4     message = "{n} attachments have been uploaded".format(n)
```

```
1 gettext("The attachment has been uploaded",
2         "{n} attachments have been uploaded", n).format(n)
3 # n = 1 -> "Załadowano ten załącznik"
4 # n = 2 -> "Załadowano 2 załączników"
5 # n = 5 -> "Załadowano 5 załączników"
```

gettext selects the appropriate plural based on the number n .

Internationalizing dates

Avoid hard-coded date formats

```
1 display_date = date.strftime("%m/%d/%y")  
2 # 03/07/2023
```

Internationalizing dates

Avoid hard-coded date formats

```
1 display_date = date.strftime("%m/%d/%y")  
2 # 03/07/2023
```

Use a library which supports locales instead

```
1 from babel.dates import format_date  
2 display_date = format_date(date, locale="fr_FR")  
3 # 7 mars 2023
```


Internationalizing dates

Avoid hard-coded date formats

```
1 display_date = date.strftime("%m/%d/%y")
2 # 03/07/2023
```

Use a library which supports locales instead

```
1 from babel.dates import format_date
2 display_date = format_date(date, locale="fr_FR")
3 # 7 mars 2023
```

You can also specify a different format

```
1 from babel.dates import format_date
2 display_date = format_date(date, locale="fr_FR", format="short")
3 # 07/03/2023
```

Internationalizing numbers

Avoid default number-string conversion

```
1 display_price = str(price) # 1099.98
```

Internationalizing numbers

Avoid default number-string conversion

```
1 display_price = str(price) # 1099.98
```

Use a library which supports locales instead

```
1 from babel.numbers import format_decimal
2 display_price = format_decimal(price, locale="cs_CZ")
3 # 1 099,98
```

Internationalizing numbers

Avoid default number-string conversion

```
1 display_price = str(price) # 1099.98
```

Use a library which supports locales instead

```
1 from babel.numbers import format_decimal
2 display_price = format_decimal(price, locale="cs_CZ")
3 # 1 099,98
```

There is a dedicated function for currencies

```
1 from babel.numbers import format_currency
2 display_price = format_currency(price, "USD", locale="en_US")
3 # $1,099.98
4 display_price = format_currency(price, "CZK", locale="cs_CZ")
5 # 1 099,98 Kč
```

i18n gotchas in Indico

Missing context

- "judge", "review", "display" - Noun or verb?
- "book" - A book or a booking?
- "Linked to", "New", "None" - Gendered adjectives
- "Manager will manually validate it" - Gendered pronouns

i18n gotchas in Indico

Use **pgettext**!

```
1 pgettext("noun", "judge") # juge
2 pgettext("verb", "judge") # juger
3
4 pgettext("room booking", "book") # réserver
5
6 pgettext("room", "Linked to") # Liée à
7 pgettext("password", "New") # Nouveau
8
9 pgettext("registration", "Manager will manually validate it")
10 # Un administrateur de l'événement doit la valider manuellement
```

i18n gotchas in Indico

Missing pluralization

- " contribution(s) "
- " minutes "
- " hours "
- "Field must contain between {min} and {max} words."
- "Please select between {min} and {max} options"

i18n gotchas in Indico

Fun with dates

- "ddd DD/MM" - Let your locale library handle date formats
- "Last {day}" (e.g. "Last Monday") - Gendered adjectives

i18n gotchas in Indico

String concatenation

- " will have their "

Intended use: "[All entries] will have their
[starting time] [changed]"

i18n gotchas in Indico

String concatenation

- " will have their "

Intended use: "[All entries] will have their [starting time] [changed]"



A Localization Horror Story: It Could Happen To You

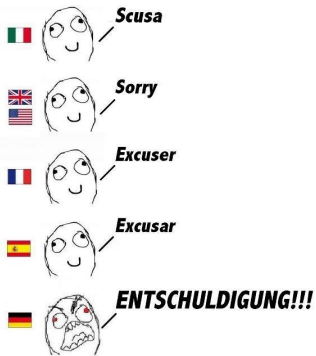
<https://metacpan.org/dist/Locale-Maketext/view/lib/Locale/Maketext/TPJ13.pod>

UI design for multilingual software

Translations might be significantly longer (or shorter) than the original text

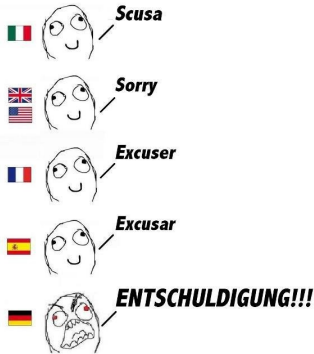
UI design for multilingual software

Translations might be significantly longer (or shorter) than the original text



UI design for multilingual software

Translations might be significantly longer (or shorter) than the original text



UI design for multilingual software

English → French

- "Last use" → "Utilisé pour la dernière fois"
- "Privacy" → "Protection de la vie privée"

French → English

- "Gérer l'événement" → "Switch to the management area of this event"
- "S'inscrire" → "Apply for participation"

Examples from Indico



Examples from Indico



Examples from Indico

Export

Events at hand

or

Categories

or

Everything

Export

Blížící se události

nebo Kategorie

nebo Všechno

Exportieren

Aktuelle Veranstaltungen

oder Kategorien

oder Alles



Examples from Indico

Subnets

Type to add subnet

Examples from Indico

Subnets

Type to add subnet

Підмережі

Напишіть, аби дода



Examples from Indico

 Switch to display view

 Settings

 Protection

 Privacy

Examples from Indico

 Switch to display view

 Settings

 Protection

 Privacy

 Basculer vers la page ...

 Paramètres

 Protection

 Protection de la vie p...



Some tips

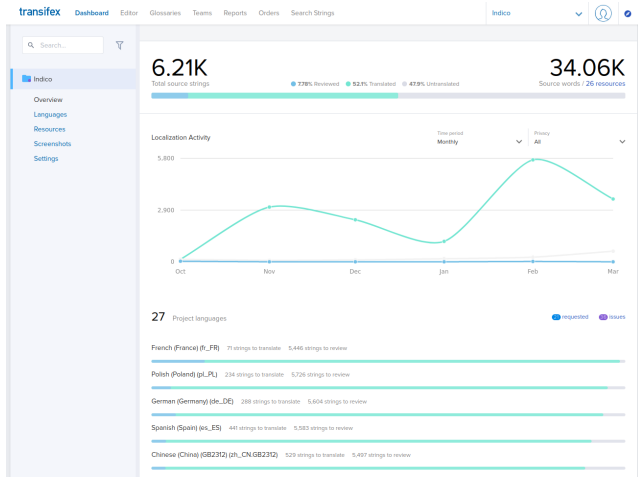
- For every translatable string, try doubling the length
- Use Indico in your language! Even if it's not officially supported, you can get it from [Transifex](#)

Recap

- Internationalization & Localization
- Localization is not just text translation
- Internationalization using gettext and babel
- UI design for multiple languages

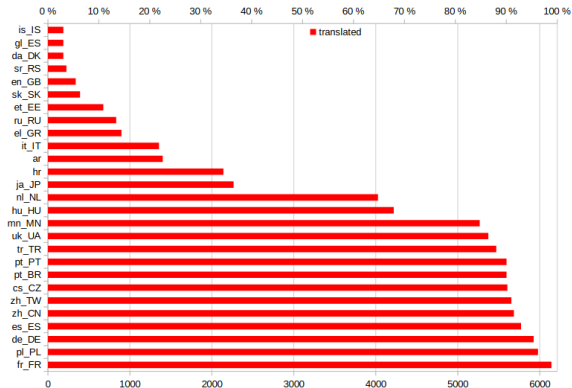


Some translation statistics



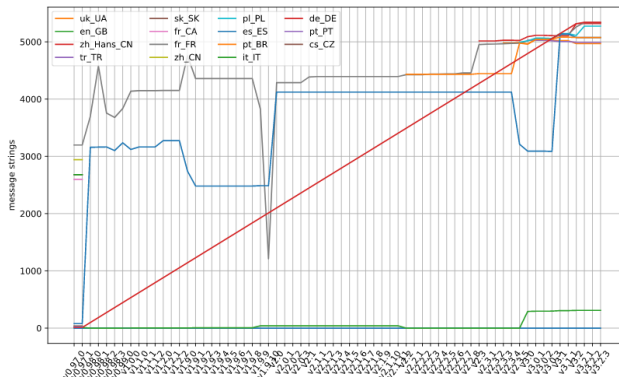
Some translation statistics

Fraction of the 6214 character chains translated per language



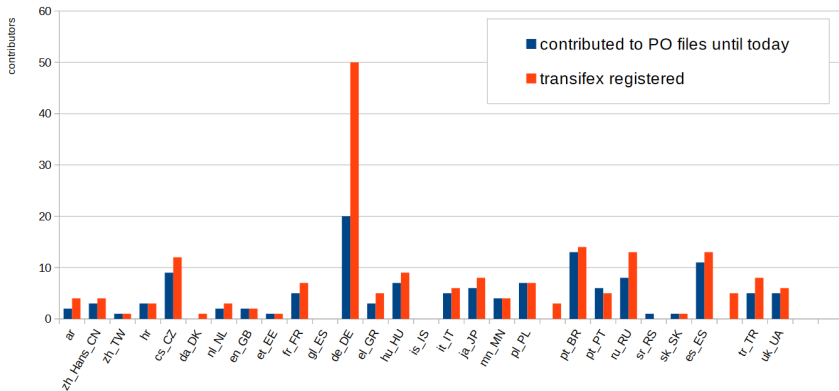
Some translation statistics

Translated character chains per Indico tag in github



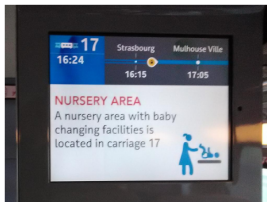
Some translation statistics

Translation team members: registered in Transifex vs. contributing according to PO files



Some translation statistics

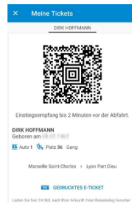
SNCF-DB



SNCF

Transifex (March 20th)

Indico team 191 membres qui traduit le projet



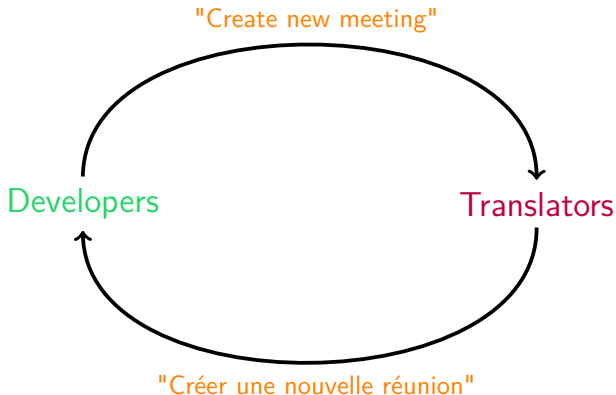


Backup slides

Internationalization & Localization

Not a one and done thing

Software translation is a continuous cycle which follows the evolution of the app



gettext

gettext can generate the list of translatable strings automatically

```
1 #: indico/web/.../protection.py:39
2 msgid "Event"
3 msgstr ""
4
5 #: indico/web/...base.html:56
6 msgid "Search"
7 msgstr ""
8
9 #: indico/modules/.../admin_settings.html:4
10 msgid "Payment"
11 msgstr ""
```

gettext

A translator just needs to fill in the msgstr

```
1 #: indico/web/.../protection.py:39
2 msgid "Event"
3 msgstr "Événement"
4
5 #: indico/web/...base.html:56
6 msgid "Search"
7 msgstr "Recherche"
8
9 #: indico/modules/.../admin_settings.html:4
10 msgid "Payment"
11 msgstr "Paieiment"
```

Plurals

```
1 #: indico/modules/attachments/.../base.py:120
2 msgid "The attachment has been uploaded"
3 msgid_plural "{n} attachments have been uploaded"
4 msgstr[0] ""
5 msgstr[1] ""
```

```
1 #: indico/modules/attachments/.../base.py:120
2 msgid "The attachment has been uploaded"
3 msgid_plural "{n} attachments have been uploaded"
4 msgstr[0] "Załadowano ten załącznik"
5 msgstr[1] "załadowano {n} załączników"
6 msgstr[2] "załadowano {n} załączników"
7 msgstr[3] "załadowano {n} załączników"
```

Plurals

Unicode CLDR definition for Polish

```
1 {  
2     "pluralRule-count-one": "i = 1 and v = 0 @integer 1",  
3     "pluralRule-count-few": "v = 0 and i % 10 = 2..4 and i % 100 != 12..14  
4     @integer 2~4, 22~24, 32~34, 42~44, 52~54, 62, 102, 1002,      ",  
5     "pluralRule-count-many": "v = 0 and i != 1 and i % 10 = 0..1 or v = 0 and  
6     i % 10 = 5..9 or v = 0 and i % 100 = 12..14 @integer 0, 5~19, 100, 1000,  
7     10000, 100000, 1000000,      ",  
8     "pluralRule-count-other": "      @decimal 0.0~1.5, 10.0, 100.0, 1000.0,  
9     10000.0, 100000.0, 1000000.0,      "  
10 }
```

