

**Brian Hsieh
Kristen Ingalz
Regex in Trados**

Language Pair: English ↔ Japanese

Summary: We identified 5 use cases for regex in Trados for the English-Japanese (EN-JA) language pair.

- 4 use cases identify potential or likely translation errors using the QA Checker
 - 1 use case identifies and fixes a translation error using the Find and Replace feature
-

QA Checker Use Case 1: Postal Codes

1. Background

- The Japanese postal code format is 〒XXX-XXXX
 - 〒 indicates the following numbers represent a postal code
 - X represents a number between 0 and 9

2. Problem: Japanese postal codes are often formatted incorrectly in source or target segments.

3. Regex Solution: 〒+\d{3}(\-|\-)\d{4}

4. Regex Description

- This regex will:
 - Search for Japanese postal codes in the described format in source or target segments.
 - Identify incorrect formatting in the corresponding segment.

5. Screenshot (Trados):

- This screenshot shows proper Japanese postal code formatting identified in the line 30 target segment.
- The QA check identified that this formatting was missing in the source segment.

The screenshot displays the Trados QA Checker interface. At the top, a 'Messages (15)' window lists various errors and warnings. The 'Messages' table has columns for Severity, Message, Origin, Document, and Segment. The messages include:

Severity	Message	Origin	Document	Segment
Warning	Finding common punctuation: target matches count 1.	QA Checker 3.0	Mod7 RegEx Email .txt.sdxliff	8
Warning	Finding common punctuation: target matches count 1.	QA Checker 3.0	Mod7 RegEx Email .txt.sdxliff	9
Warning	Finding common punctuation: target matches count 1.	QA Checker 3.0	Mod7 RegEx Email .txt.sdxliff	12
Warning	Finding common punctuation: target matches count 1.	QA Checker 3.0	Mod7 RegEx Email .txt.sdxliff	22
Warning	Finding common punctuation: target matches count 1.	QA Checker 3.0	Mod7 RegEx Email .txt.sdxliff	24
Warning	Find AMPM: source matches count 2, target matches count 1.	QA Checker 3.0	Mod7 RegEx Email .txt.sdxliff	18
Warning	Find AMPM: source matches count 2, target matches count 1.	QA Checker 3.0	Mod7 RegEx Email .txt.sdxliff	19
Warning	EN>JP Postal Code: source matches count 0, target matches count 1.	QA Checker 3.0	Mod7 RegEx Email .txt.sdxliff	30
Warning	Convert string of numbers to phone#: target matches count 1.	QA Checker 3.0	Mod7 RegEx Email .txt.sdxliff	25

Below the messages, the 'Translation Results - Mod7 RegEx Email' window shows a comparison between source and target segments. The source segment (line 30) is '38-39 Higashidaicho Kawashima' and the target segment (line 30) is '〒615-8140京都市西京区東代町川島38-39 日本'. The target segment is highlighted in blue, indicating it is the correct format. The source segment is highlighted in yellow, indicating it is incorrect. The message 'EN>JP Postal Code: source matches count 0, target matches count 1.' is shown in the messages window, indicating that the source segment does not match the target segment's postal code format.

QA Checker Use Case 2: AM/PM

1. Background

- The English language commonly identifies morning times as AM (eg, 2:45 AM) and afternoon/evening times as PM (eg, 6:45 PM).
- The Japanese language commonly uses military time
 - Morning times are represented in the format XX:XX in a sequence that starts at 00:00 (midnight) and extends to 11:59 (one minute before noon).
 - Evening times are represented in the format YY:YY in a sequence that starts at 12:00 (noon) and ends at 23:59 (one minute before midnight).

2. Problem:

- Japanese translations of English language formatted times appear in AM/PM format and not military format.
- English translations of Japanese formatted times appear in military format and not AM/PM format.

3. Regex Solution: `(\b(?:!w*(\w)\w*\1)[PM|pm]+\b)|(\b(?:!w*(\w)\w*\1)[AM|am]+\b)`

4. Regex Description: This regex will find times noted as AM, am, PM, or pm in source or target segments, allowing the translator to correct the corresponding source or target segment formatting easily.

5. Screenshot (Trados):

- This screenshot shows English language formatted times (PM) in the source and target segments.
 - The Japanese target segment should be formatted for military time.
- The QA correctly identifies that AM/PM was found in both source and target segments in line 18.
- The translator can easily edit the Japanese target segment to reflect military time format.

The screenshot displays the Trados QA Checker interface. At the top, it shows 6 Errors and 9 Warnings. A table lists the errors, with line 18 highlighted. Below the table, the source and target text for lines 18 and 19 are shown. In line 18, the source text is '12/26/2023 NH5828 TPE (1:35 PM) - KIX (5:15 PM)' and the target text is '2023/12/26 NH5828 TPE (1:35 PM) - KIX (5:15 PM)'. In line 19, the source text is '12/30/2023 NH32 ITM (4:00 PM) - HND (5:17 PM)' and the target text is '2023/12/30 NH32 ITM (4:00 PM) - HND (5:17 PM)'. Red boxes highlight the 'PM' in the source and target of line 18.

Severity	Message	Origin	Document	Segment
▲	Finding common punctuation: target matches count 1.	QA Checker 3.0	Mod7 Regex Email btt.sdxliff	8
▲	Finding common punctuation: target matches count 1.	QA Checker 3.0	Mod7 Regex Email btt.sdxliff	9
▲	Finding common punctuation: target matches count 1.	QA Checker 3.0	Mod7 Regex Email btt.sdxliff	12
▲	Finding common punctuation: target matches count 1.	QA Checker 3.0	Mod7 Regex Email btt.sdxliff	22
▲	Finding common punctuation: target matches count 1.	QA Checker 3.0	Mod7 Regex Email btt.sdxliff	24
▲	Find AM/PM: source matches count 2, target matches count.	QA Checker 3.0	Mod7 Regex Email btt.sdxliff	18
▲	Find AM/PM: source matches count 2, target matches count.	QA Checker 3.0	Mod7 Regex Email btt.sdxliff	19
▲	EN>JP Postal Code: source matches count 0, target matches.	QA Checker 3.0	Mod7 Regex Email btt.sdxliff	30
▲	Convert string of numbers to phone#: target matches count 1.	QA Checker 3.0	Mod7 Regex Email btt.sdxliff	25

Line	Source	Target
18	12/26/2023 NH5828 TPE (1:35 PM) - KIX (5:15 PM)	2023/12/26 NH5828 TPE (1:35 PM) - KIX (5:15 PM)
19	12/30/2023 NH32 ITM (4:00 PM) - HND (5:17 PM)	2023/12/30 NH32 ITM (4:00 PM) - HND (5:17 PM)

QA Checker Use Case 3: Phone Numbers

1. Background

- Japanese telephone numbers use a XXX-XXXX-XXXX format (3 digits - 4 digits - 4 digits) where X represents a number between 0 and 9.
- In translations, Japanese telephone numbers frequently appear as a string of 11 digits not separated by dashes into XXX-XXXX-XXXX format.

2. Problem: Japanese telephone numbers should appear in the correct format for accuracy and ease of understanding.

3. Regex Solution: `(\d{9,11})`

4. Regex Description This regex will find strings of 11 numbers not separated by dashes. This allows the translator to easily identify number strings that may need to be adjusted into the Japanese telephone number format.

5. Screenshot (Trados):

- This screenshot shows a string of 11 digits not separated by dashes in the line 18 target segment.
- The string represents a Japanese phone number, which should be separated by dashes into XXX-XXXX-XXXX format.
- The QA checker identifies a string of 11 digits has been found in the line 18 target segment.
- The translator can easily edit the string so it reflects the correct format.

The screenshot displays the Trados QA Checker interface. At the top, a status bar shows '12/26/2023 NH5828 TPE (1:35 PM) - KIX (5:15 PM)' with a 99% match rate. Below this, a table lists 15 messages (errors) with columns for Severity, Message, Origin, Document, and Segment. The messages are categorized as 'Finding common punctuation: target matches count 1' and 'Finding common punctuation: target matches count 2'. The bottom part of the screenshot shows a translation segment with the source text 'Is your phone number still 090-6566-7223?' and the target text '電話番号は09065667223のままですか？'. The phone number '090-6566-7223' in the source text is highlighted with a red box, and the corresponding target text '09065667223' is also highlighted with a red box.

Severity	Message	Origin	Document	Segment
Warning	Finding common punctuation: target matches count 2	QA Checker 3.0	Mod7 Regex Email: btt.sdxliff	22
Warning	Finding common punctuation: target matches count 1	QA Checker 3.0	Mod7 Regex Email: btt.sdxliff	8
Warning	Finding common punctuation: target matches count 1	QA Checker 3.0	Mod7 Regex Email: btt.sdxliff	9
Warning	Finding common punctuation: target matches count 1	QA Checker 3.0	Mod7 Regex Email: btt.sdxliff	12
Warning	Finding common punctuation: target matches count 1	QA Checker 3.0	Mod7 Regex Email: btt.sdxliff	24
Warning	Find AM/PM: source matches count 2, target matches count 1	QA Checker 3.0	Mod7 Regex Email: btt.sdxliff	18
Warning	Find AM/PM: source matches count 2, target matches count 1	QA Checker 3.0	Mod7 Regex Email: btt.sdxliff	19
Warning	EN-JP Postal Code: source matches count 0, target matches count 1	QA Checker 3.0	Mod7 Regex Email: btt.sdxliff	30
Warning	Convert string of numbers to phone#: target matches count 1	QA Checker 3.0	Mod7 Regex Email: btt.sdxliff	25

Translation Results - Mod7 Regex Email | Comments(0) | Messages(15)

25 Is your phone number still 090-6566-7223? 電話番号は09065667223のままですか？

26 While I'm at it, I just want to confirm I have the correct address for your parents. ついでに、あなたのご両親の住所が正しいかどうか確認させていただきます。

QA Checker Use Case 4: Random Punctuation

1. Background

- The English and Japanese languages use different punctuation styles.
 - English comma , vs Japanese comma 、
 - English period . vs Japanese period 。

2. Problem:

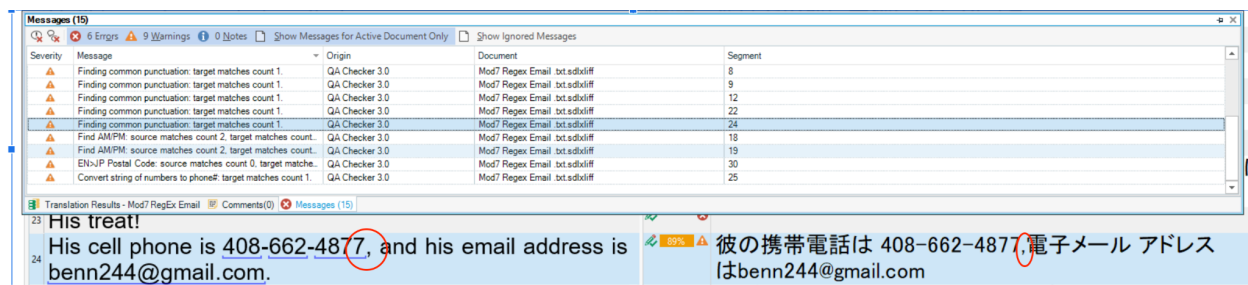
- Japanese translations of English text often include stray English-style punctuation.
 - Incorrect comma or period style
 - In case of typos, English-style periods appearing at the beginning of sentences

3. Regex Solution: \.,\$|^\.|\,

4. Regex Description This regex will find English-style periods and commas that appear in both the source and target segments.

5. Screenshot (Trados):

- This screenshot shows English-style commas appearing in both the source (English) and target (Japanese) segments.
- The QA checker indicates common punctuation has been identified in both source and target segments in line 24.
- The translator can easily edit the target Japanese segment to reflect the correct Japanese-style comma.



Find and Replace Use Case 1: Random Punctuation

1. Background
 - The English and Japanese languages use different period styles.
 - English period . vs Japanese period 。
6. Problem: Japanese translations of English text often include stray English-style periods.
7. “Find and Replace” Regex Solution: \.\$ **[Note: in computers using Japanese language settings, the \ symbol appears as the yen (¥) symbol, but still functions as the \ symbol in Regex].**
8. Regex Description This regex will find English-style periods in Japanese segment and will replace with the Japanese-style period if “Replace” is clicked from the “Find and Replace” window in Trados.
9. Screenshot (Trados):
 - This screenshot shows an English-style period at the end of a sentence in the Japanese target segment.
 - The “Find and Replace” feature, when set up as indicated in the screenshot, highlights the above error.
 - When “Replace” is clicked at the bottom of the “Find and Replace” window, the English-style period in the Japanese target segment will be replaced with a Japanese-style period.

