Towards Fully Automated Manga Translation

R. Hinami*[1], S. Ishiwatari*[1], K. Yasuda[2], Y. Matsui[2]

[1] Mantra Inc. [2] The University of Tokyo

MANTRA

About me

Research interests

- 2014-2016 Semantic Repr. & SMT
- 2016-2017 Chunk-based NMT
- 2017-2019 Definition Generation
- 2019- Manga Translation
- 2020- Manga Colorization
- 2021- Manga-based Language Learning


PLOF
TSHUUU
PLOF
BOMBS AWAY!!
GET OFFA ME, STUPID!
TA-TA TA-TA
HE WENT DOWN THE PIPE!!
KZIN
KZIN
COULDN'T THE BEAM OF A FLASHLIGHT
BRING A LIGHT?
CAN'T BE A BOMB!
217

Mantraのミッション

MANTRA

世界の言葉で、
マンガを届ける。

漫画家

たくさん
読まれたい
買われたい

世界の
マンガ好き

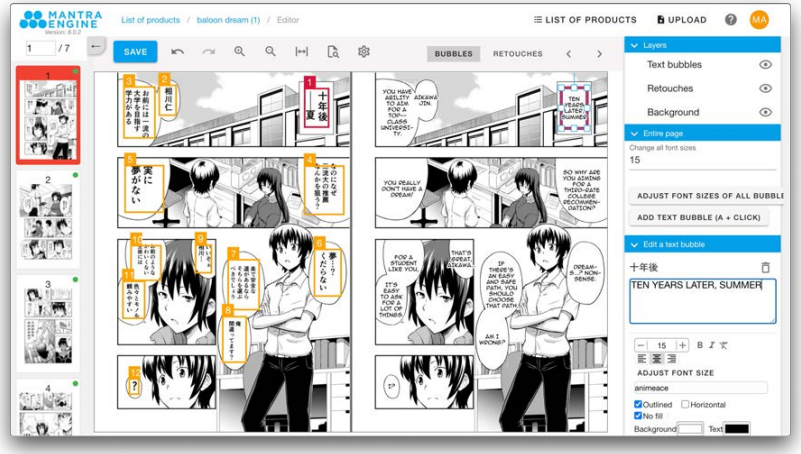
いち早く
母語で
読みたい

Mantraの事業

MANTRA

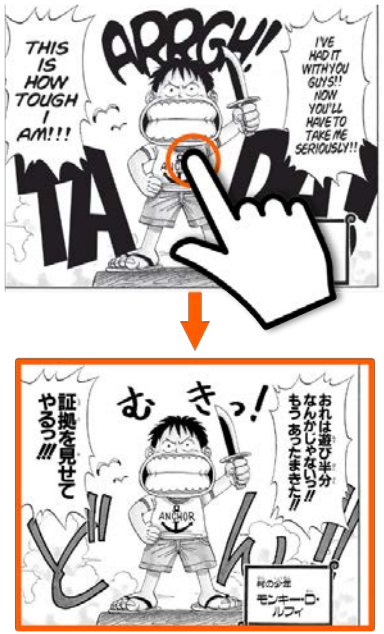
高速翻訳でマンガを多言語展開

Mantra Engine



マンガで外国語学習

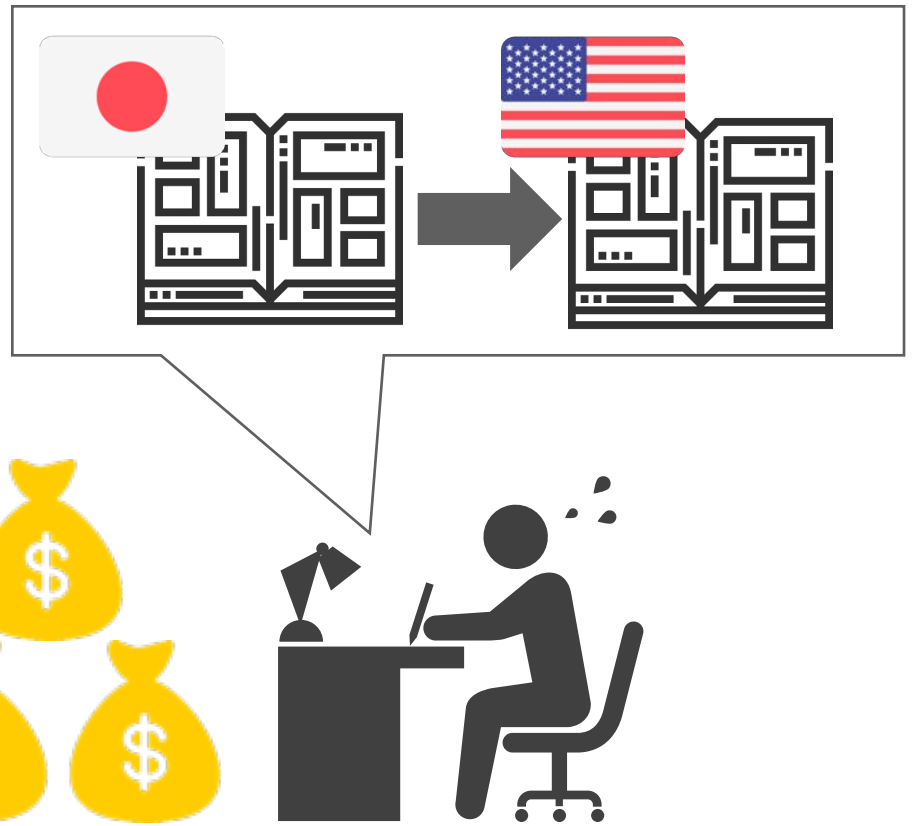
Langaku



Problems: Slow & expensive translation

MANTRA

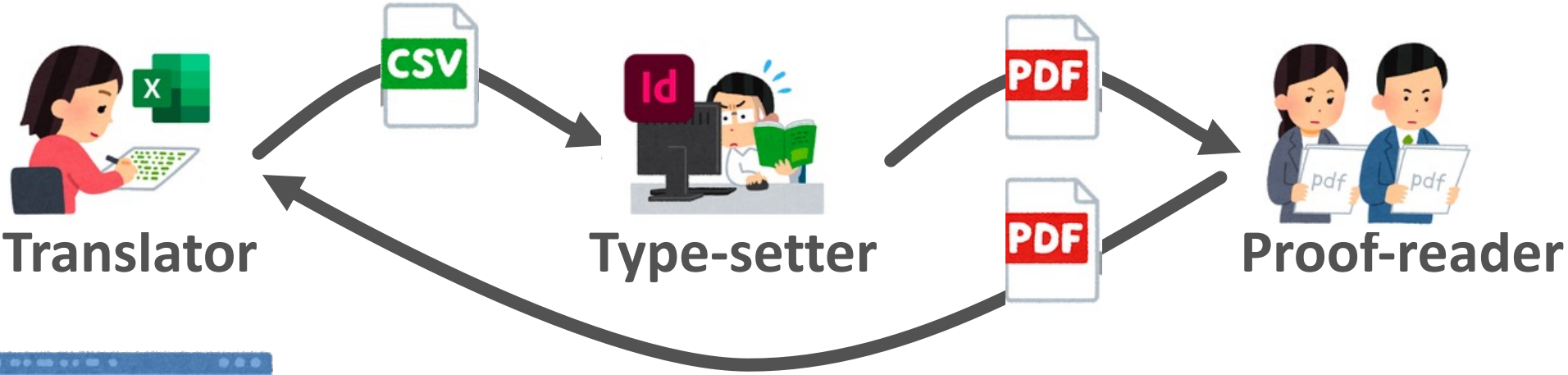
2+ weeks per episode

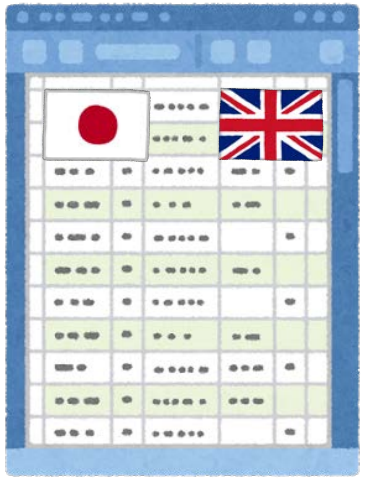


2-3x cost vs. plain text

Why is translation so inefficient?

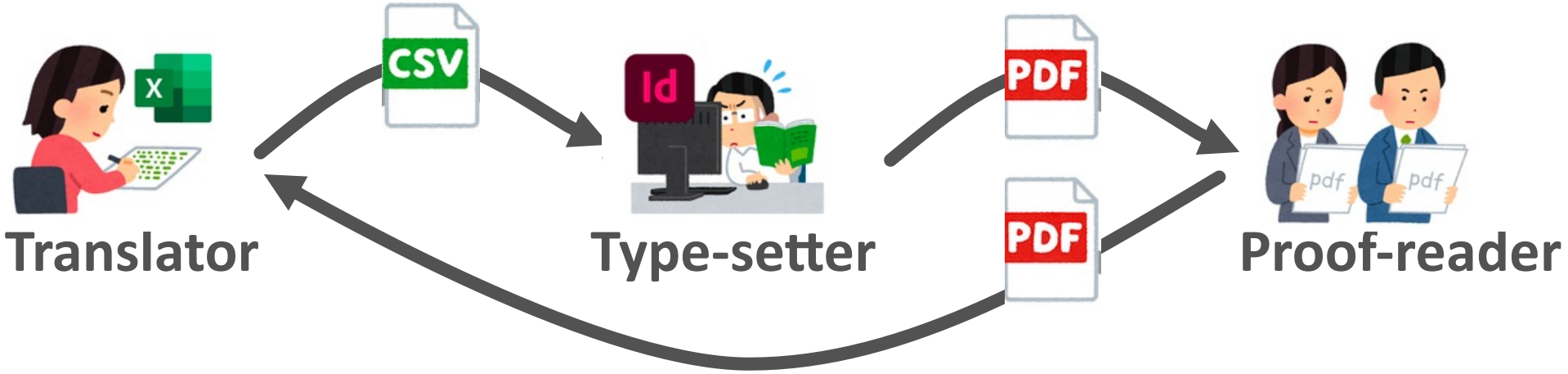
MANTRA





Why is translation so inefficient?

MANTRA



- Difficult translation + lettering
- Back-and-forth file sharing via email
- Various file formats for various software

MANTRA ENGINE
Version: 3.0.0

作品一覧　入稿　アカウント　ログアウト

日本語原稿をアップロード

画像複数枚をまとめてドラッグ&ドロップしてください
サイズ1枚10MB以下
対応拡張子 .jpg, .png

写植なし原稿をアップロード

ページ対応付け方法
対応する日本語原稿と同じファイル名

新しい作品を作成　既存の作品に最新話を追加

作品タイトルを入力してください

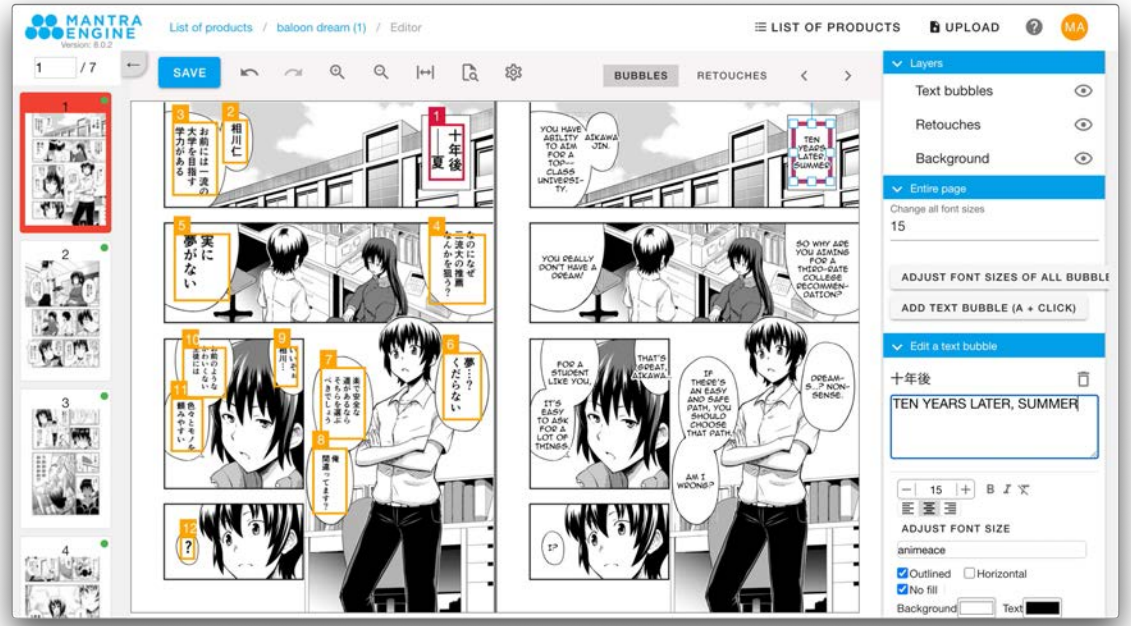
言語
英語　中国語

入稿する

Product: CAT* tool for comic translation

MANTRA

* Computer assisted translation



- ✓ Machine translation & type-setting for comics
- ✓ No email communication
- ✓ No format conversion

2x faster translation!

Topics

MANTRA

1. **Motivation**
   Why comic translation?

▶ 2. **Challenges**
   How difficult is (automated) comic translation?

3. **Approaches**
   Towards machine translation for comics translation

4. **Future directions**

Is comic translation difficult?

MANTRA

Subject “I” is omitted

©Mitsuki Kuchitaka

A sentence is divided into two balloons

MT

Reference
I just wanted to...

Reference
pick her up...

Context is required for accurate translation

What makes comic translation difficult?

MANTRA

❖ Challenges for comic translation

- High dependence on context **-> Context-aware MT**
- Multi-modality **-> Multi-modal MT**
  - e.g., Speaker-specific wordings
- Low-resource **-> Low-resource MT**
  - No comic-domain training/evaluation data available

Topics

MANTRA

1. **Motivation**
   Why comic translation?

2. **Challenges**
   How difficult is (automated) comic translation?

▶ 3. **Approaches**
   Towards machine translation for comics translation

4. **Future directions**

Approaches to automatic comic translation

MANTRA

1. Evaluation dataset/framework
2. Context-aware comic translation
3. Visual-aware comic translation

OpenMantra: Evaluation dataset for manga translation

MANTRA

- Five manga (Japanese comics) titles w/ manual annotation & translation

Romance

Mystery

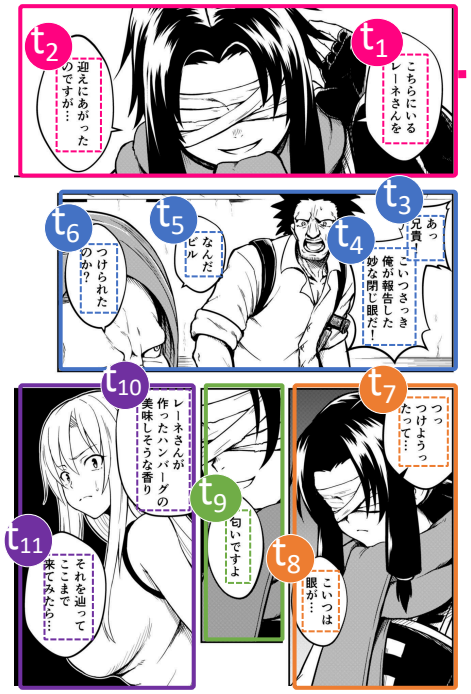
Fantasy

Slice of life

Battle

OpenMantra: Evaluation dataset for manga translation

❖ Five manga (Japanese comics) titles w/ manual annotation & translation



Ja: (#1) こちらにいるレーネさんを
(#2) 迎えにあがったのですが...

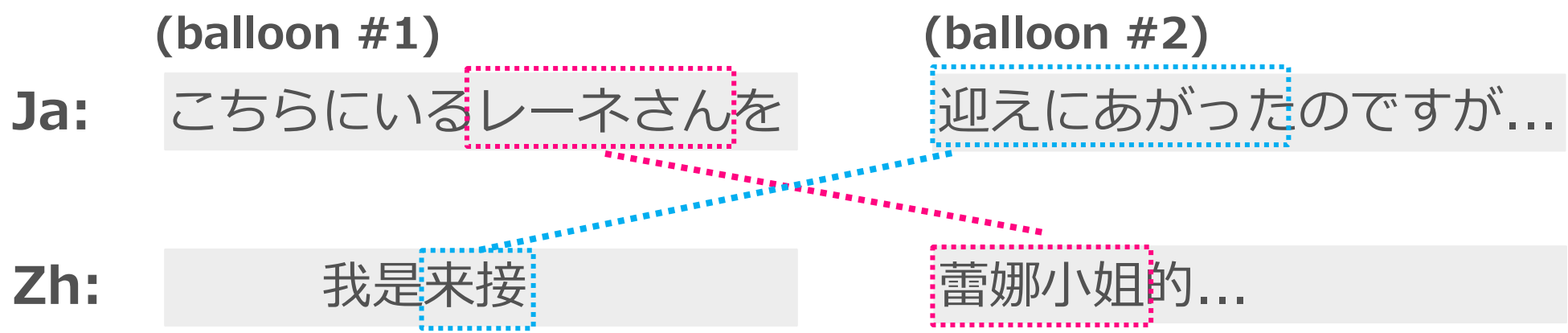
En: (#1) I just wanted to...
(#2) pick her up...

Zh: (#1) 我是来接
(#2) 蕾娜小姐的...

How to evaluate translated texts?

MANTRA

- **Challenge: Texts are often swapped after translation**

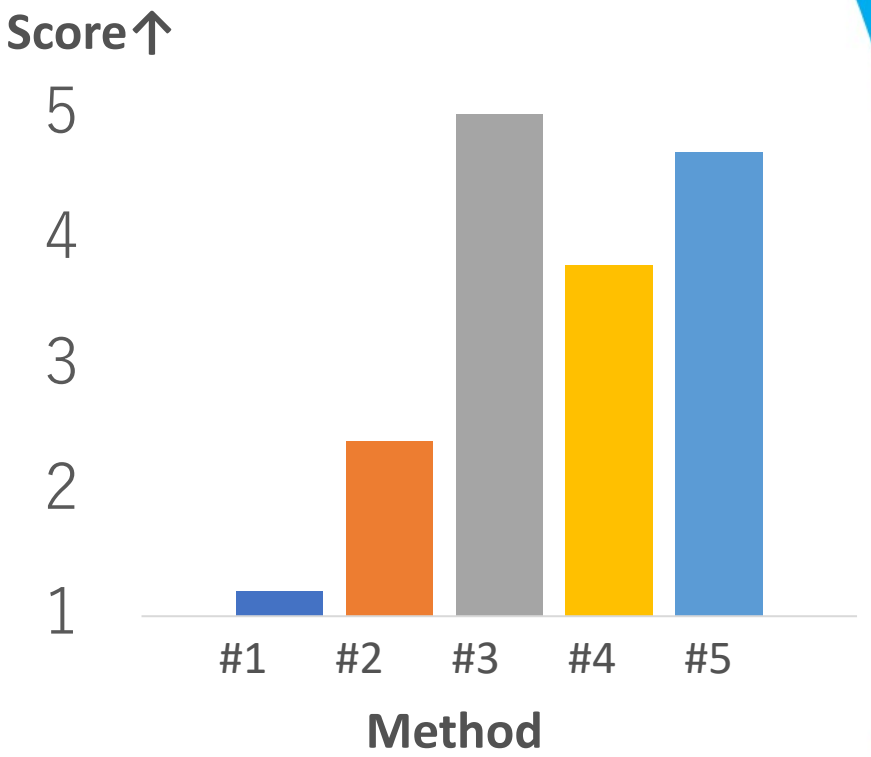


Balloon-by-balloon evaluation does not work with comics

Evaluation toolkit for page-by-page evaluation

MANTRA

Aggregate



Approaches to automatic comic translation

MANTRA

1. Evaluation dataset/framework
2. Context-aware comic translation
3. Visual-aware comic translation

Importance of context

MANTRA

A sentence consists of two balloons

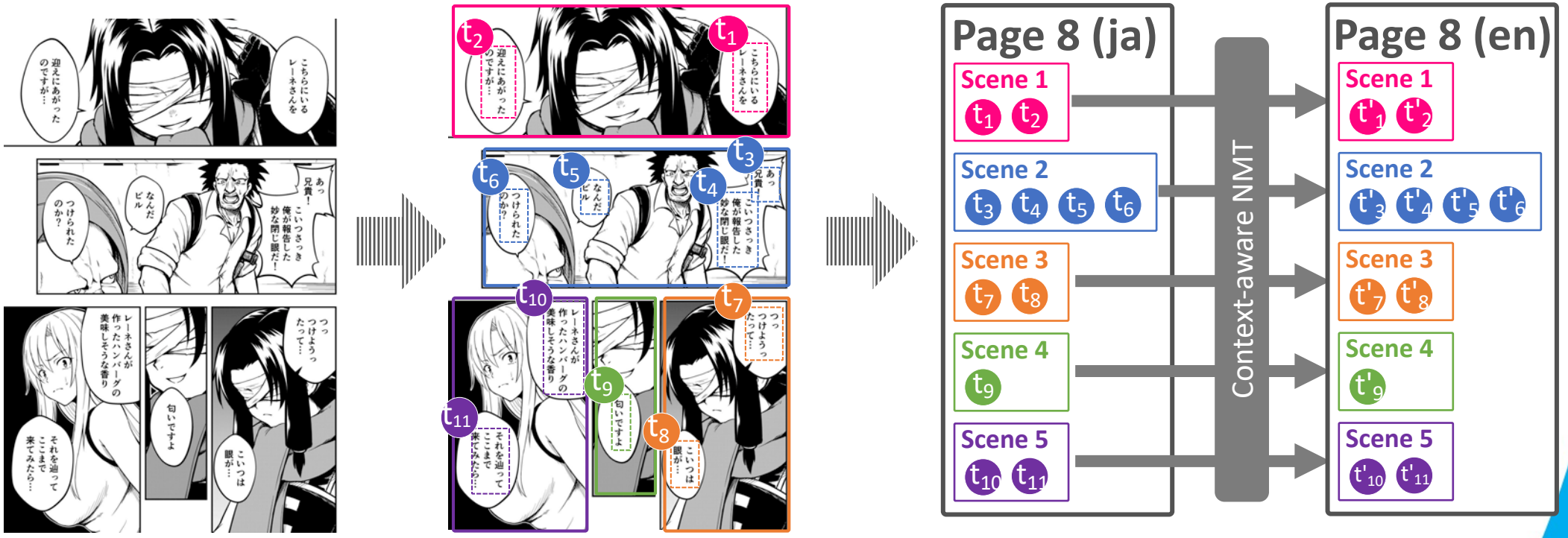
Ja:	こちらにいるレーネさんを	迎えにあがったのですが…
Reference:	I just wanted to…	pick her up…
MT:	Renee-san here…	I was picked up…

Which balloons should be used as “context”?

Proposed: Scene-based MT

MANTRA

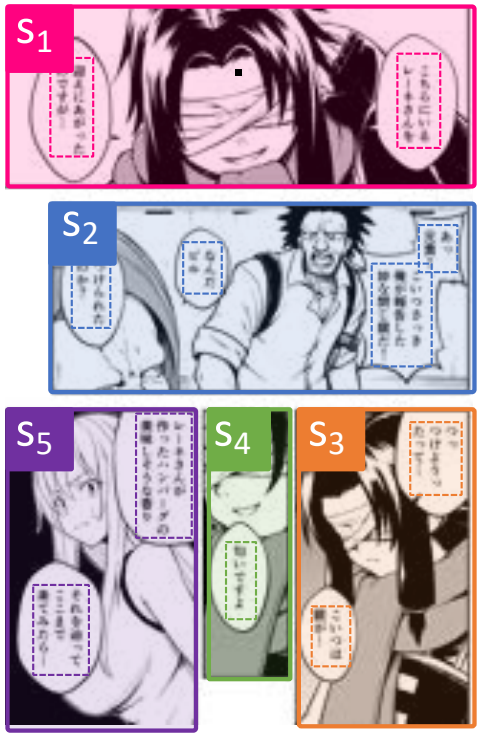
- Idea: Assuming each frame as a “scene” and considering all the texts in it as context



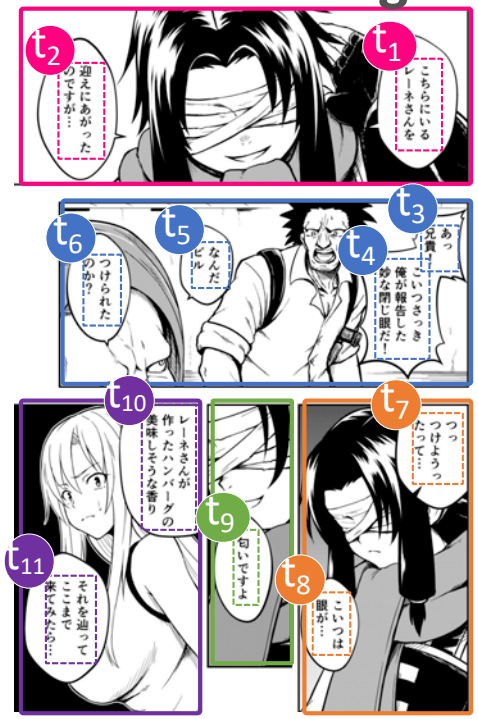
Frame detection and Text ordering

MANTRA

Frame detection



Text ordering



- **Frame detection**
  - Faster R-CNN [Ren+ 15] trained on Manga109 dataset [Matsui+ 17]
- **Text ordering**
  - First order frames then balloons
  - A simple heuristic: order balloons by the distance from the upper right point of each frame

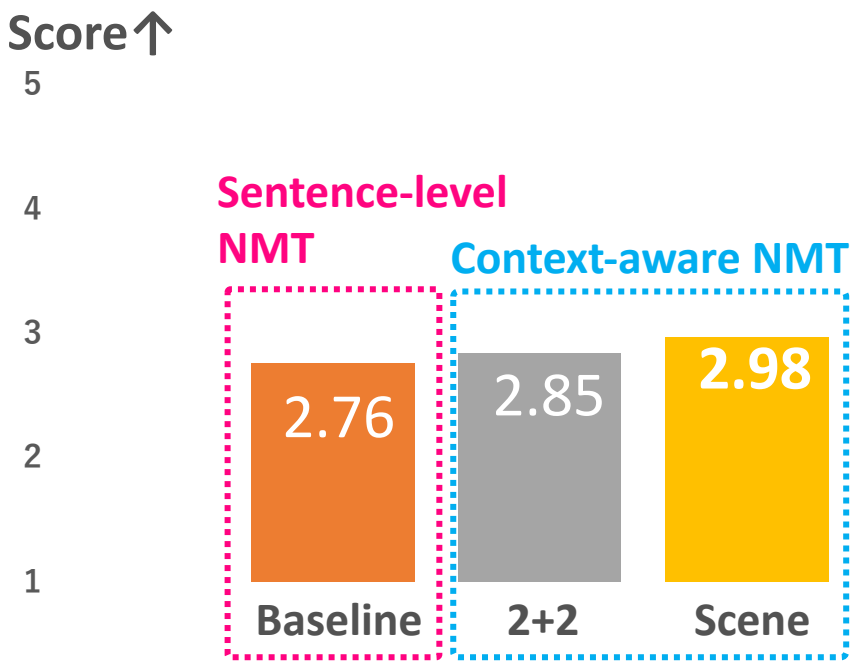
Context-aware NMT v. sentence-level NMT

MANTRA

❖ Experimental settings

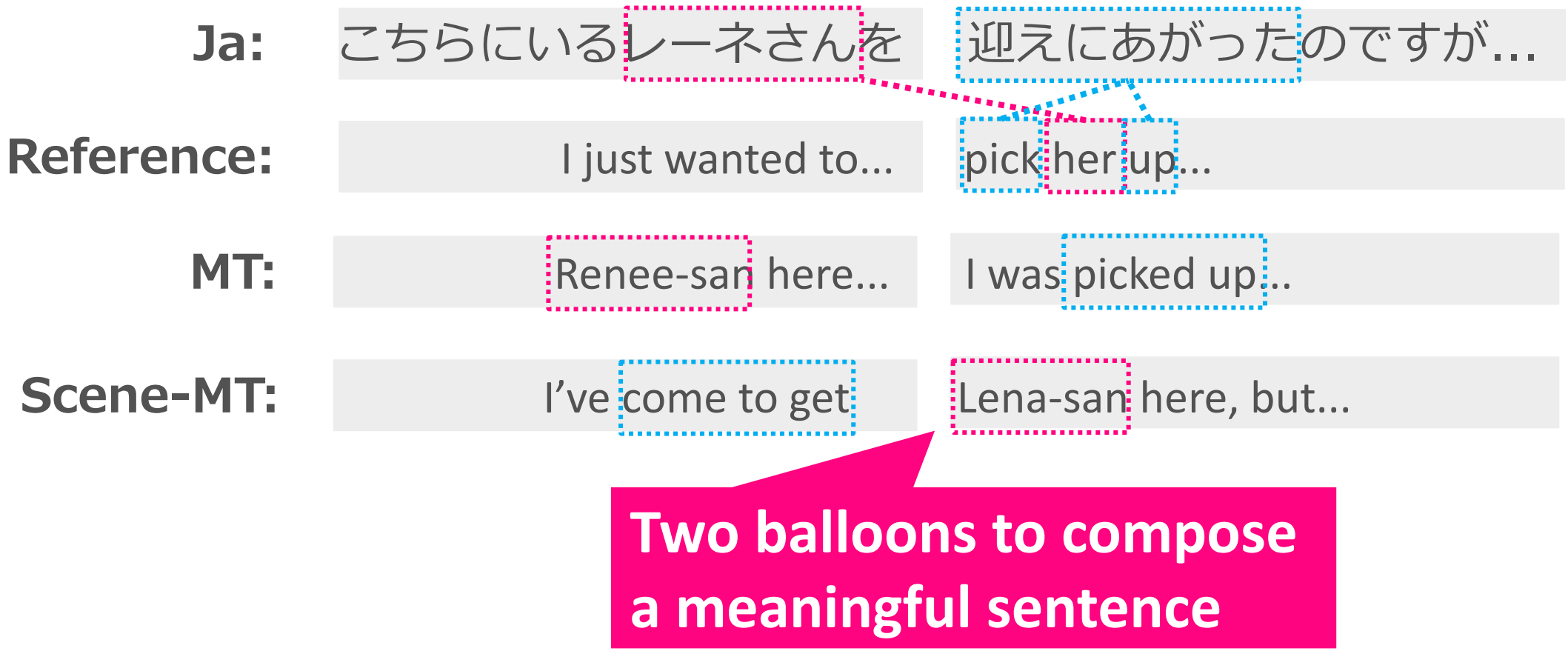
- **Baseline** (sentence-level NMT)
  - Transformer (big) [Vaswani+ 17]
  - Trained data: manga corpus (4M)
- **2+2** [Tiedemann & Scherrer 17]
  - Simply concatenate 2 sentences
- **Scene-based NMT (Proposed)**
  - Concatenate all the text in each frame

❖ Manual evaluation results



How context works in the scene-NMT

MANTRA



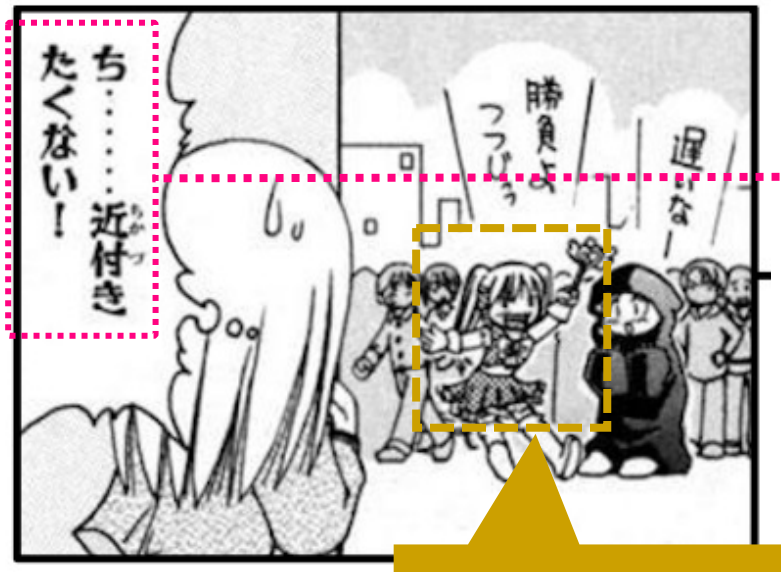
Approaches to automatic comic translation

MANTRA

1. Evaluation dataset/framework
2. Context-aware comic translation
3. Visual-aware comic translation

Importance of visual information

MANTRA



©Satoshi Arai

"him" ?

No subject nor object!

Ja: 近づき たく ない

get close to / want to / not

MT: I... I don't want to go near him! ☹

Can we combine visual information into scene-NMT?

Proposed: Visual-aware Scene-based NMT

- Idea: Extracting visual information for each scene and input the extracted tags to NMT

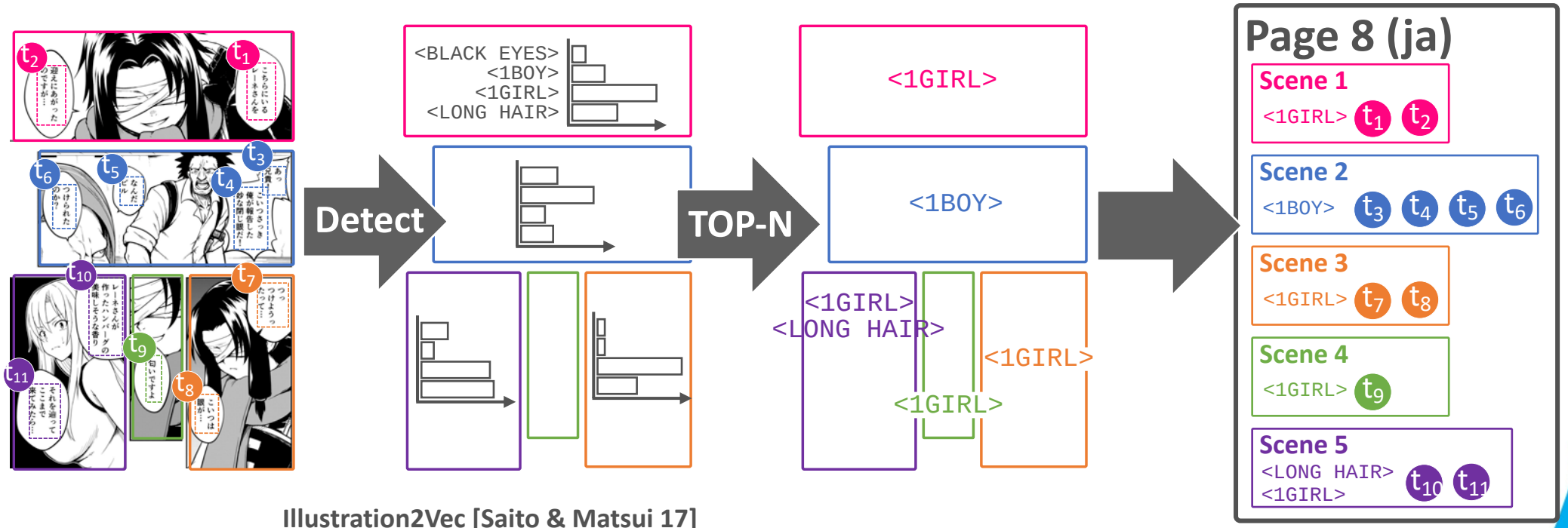
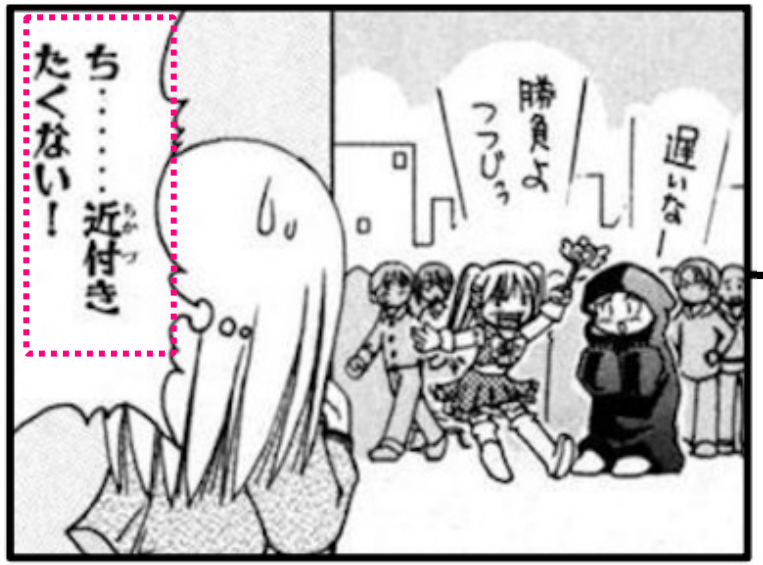


Illustration2Vec [Saito & Matsui 17]

Example of visual-aware MT

MANTRA



©Satoshi Arai

Extracted visual tags

<MULTIPLE_GIRLS>
<SCHOOL_UNIFORM>
<LONGHAIR>
<SERAFUKU>
<TWINTAILS>
<SHORTHAIR>

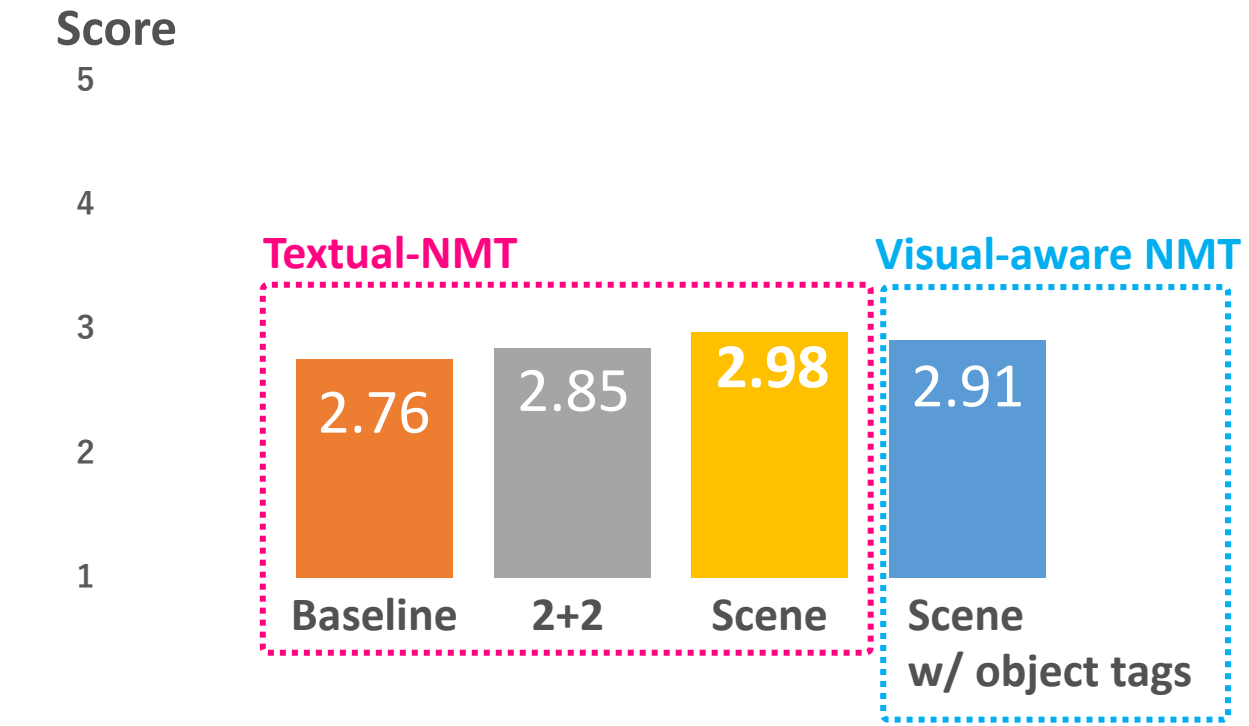
w/o visual tags: I… I don’t want to go near **him**! ☹

w/ visual tags: I-I don’t want to get close to **her**! ☺

Textual-NMT v. Visual-aware NMT

MANTRA

- Result: The visual tags don't contribute to overall performance



Analysis: Negative impact of visual tags

MANTRA

Input

©Nako Nameko

Predicted tag

<MULTIPLE_GIRLS>

Scene-based NMT w/ visual tags

1. Maybe **she**’s tired.
2. We’ll wake **her** up later.

Reference

1. Maybe **he** was tired.
2. I'll wake **him** later.

Scene-based NMT w/o visual tags

1. Maybe he was just tired.
2. We’ll wake him up later.

Analysis: Negative impact of visual tags

MANTRA

Input

©Nako Nameko

Male face

Predicted tag

<MULTIPLE_GIRLS>

Scene-based NMT w/ visual tags

1. Maybe **she**'s tired.
2. We'll wake **her** up later.

☹

Predicted tag

<1BOY>

Scene-based NMT w/ visual tags

1. Maybe **he** was just tired.
2. We'll wake **him** up later.

Better image encoder/visual representation are required

Approaches to automatic comic translation

MANTRA

1. Evaluation dataset/framework
2. Context-aware comic translation
3. Visual-aware comic translation

Topics

MANTRA

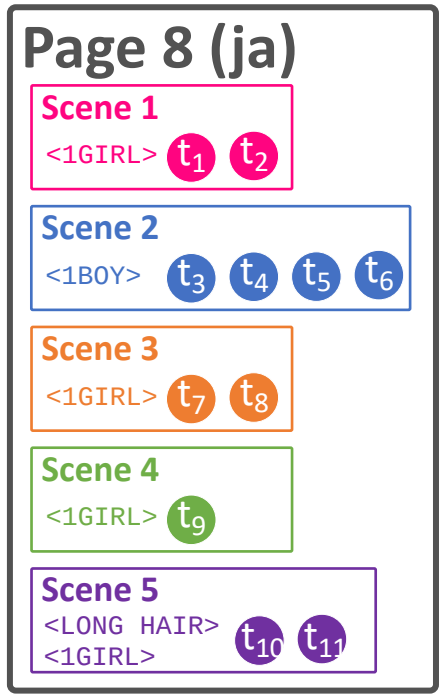
1. **Motivation**
   Why comic translation?

2. **Challenges**
   How difficult is (automated) comic translation?

3. **Approaches**
   Towards machine translation for comics translation

4. **Future directions**

Limitation of current scene-NMT model



❖ Visual information

- Discrete tags of the detected objects

❖ Textual information

- Limited length of context
  - i.e., texts in a single frame

Richer visual repr. & contextual encoders

MANTRA

- **Visual information**
  - Discrete tags of the detected objects

→ **Continuous distributions /embeddings for comic images should be explored**

- **Textual information**
  - Limited length of context
    - i.e., texts in a single frame

→ **Document-level MT architectures that capture longer/hierarchical context can be applied**

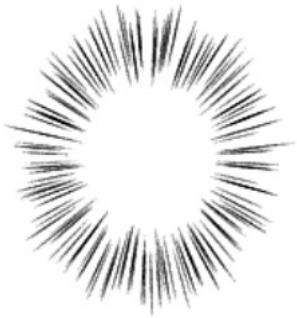
Better use of multi-modality

❖ Emotion prediction

Anger

Politeness

Surprising



Anxiety

[Yamanishi+ 17]

❖ Speaker/listener detection

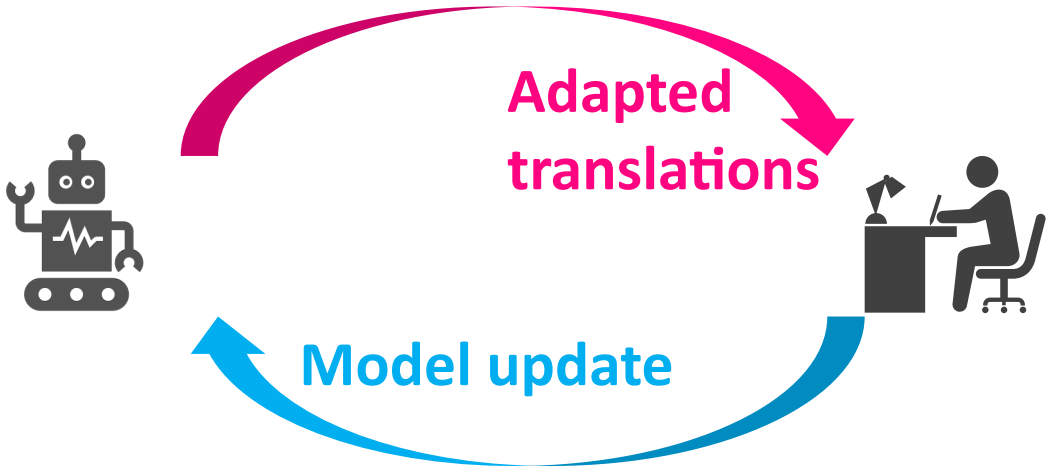
speaker listener

[Rigaud+ 15]

Online/incremental domain adaptation

MANTRA

- **Real-time domain adaptation improves productivity of human translators**
  - e.g., [Turchi+ 17; Kothur+ 18; Karimova+ 18]



Machine translation for entertainments

MANTRA

- **Multi-modality of entertainment contents**

Comics	Movies	Videogames
• Text • Images	• Video • Audio	• Text • Images • Video • Audio

Challenge:
How to define, extract, and model multi-modal context?

Topics

MANTRA

1. **Motivation**
   Why comic translation?
2. **Challenges**
   How difficult is (automated) comic translation?
3. **Approaches**
   Towards machine translation for comics translation
4. **Future directions**