Translation as Action Preservation (TAP): A Practice Report for Anime/Manga Translation

Draft for the Journal of Audiovisual Translation (JAT) — Practice Report

Author: (Masamichi Iizumi)
Affiliation: (Miosync, Inc.)
Date: January 2026
Code & data: https://anonymous.4open.science/r/divergence-z-3836/

Abstract

This practice report proposes **Translation as Action Preservation (TAP)**: evaluating translations not only by semantic similarity, but by whether a target-language line preserves the **illocutionary action** (what the line *does* to the listener and the scene).

We operationalize TAP with two complementary evaluators:

- **IAP (Illocutionary Act Preservation)**: a structured, act-level metric computed by an LLM and a deterministic matcher.

- **ZAP (Z-Axis Preservation)**: a rubric-style judge focusing on **character voice**, **emotional dynamics**, **listener relationship**, and **speech pattern**.

We demonstrate TAP on three widely translated anime/manga properties—**Re:Zero**, **Steins;Gate**, and **One Piece**—via three experiments:

- **E1 (Address Mode):** direct confession vs. reported narration (a known failure mode in JP→EN for self-name self-reference).

- **E2 (Persona/Context Sensitivity):** same source line, different persona or context produces different *actions* in translation.

- **E3 (Round-trip Robustness):** EN→FR→EN tests whether a key action (**ULTIMATUM**) survives language hopping.

Across experiments, TAP highlights failures that are invisible to meaning-only evaluation (e.g., a "correct" line that breaks direct address and collapses intimacy).

Keywords: anime translation, manga translation, speech acts, illocutionary force, audiovisual translation, machine translation evaluation, character voice

1. Motivation

Anime/manga translation often succeeds when it preserves **how a line lands**—its pressure, intimacy, deflection, threat, or vow—rather than its literal structure. Standard automatic metrics (BLEU, chrF) and "adequacy/fluency" scoring underweight this.

TAP reframes translation quality as:

> **Does the translation preserve the action performed by the line, under the same persona, context, and listener relationship?**

2. Theoretical Background: The Z-Axis

2.1 Speech-Act Grounding: From "Meaning" to "Doing"

TAP starts from a simple speech-act claim: dialogue lines are not only containers of meaning, they are **social actions**.

Austin's classic distinction—*locutionary content* (what is said), *illocutionary force* (what is done in saying it), and *perlocutionary effect* (what happens because it was said)—is especially relevant to audiovisual translation, where dialogue is "performed" and character voice is part of the narrative mechanism (Austin, 1962; Searle, 1969).

A subtitle that preserves propositional content but changes the illocutionary move can make a character feel "off" even when the literal message is correct.

Example: Luffy's "I won't dominate"

Consider this iconic line from One Piece:

> **Source (JP):** 「支配なんかしねェよ。この海で一番自由な奴が海賊王だ！」

> **Acts (A_1):** [REJECT dominance] + [REDEFINE freedom] = **"Rewriting the definition of king"**

Approach	Translation	BLEU	IAP	What happens
Standard MT	"I won't dominate. The freest guy in this sea is the Pirate King."	✅ High	❌ Low	Sounds like factual statement; Luffy's "magnitude" is lost
Z-Axis	"I'm not looking to rule over anything. Being Pirate King means being the freest one	△ Novel vocab	✅ High	**Crushes the assumption** that kings dominate; the Act of redefining is

Approach	Translation	BLEU	IAP	What happens
	on these seas!"			preserved

The standard translation preserves *locutionary content* (the words mean roughly the same thing), but weakens the *illocutionary force* from **declaration** to **explanation**. The Z-axis translation uses vocabulary not in the original ("rule over," "looking to"), but preserves the action: **shattering the conventional wisdom about what "king" means**.

This is why TAP evaluates translation primarily at the **illocutionary level**. Perlocutionary effects (e.g., whether the audience laughs or cries) depend on viewers and culture, so they are harder to treat as a stable target. Illocutionary intent is a more controlled object for analysis: we can ask whether a line still functions as a confession, an ultimatum, an escape-closing insistence, a deflection, or a self-repair.

2.2 The Three Axes of Translation Evaluation

Traditional translation evaluation operates on two axes:

- **X-axis (Semantics):** What is communicated? — the informational content
- **Y-axis (Style):** How is it expressed? — syntax, register, fluency

Standard metrics (BLEU, chrF, human adequacy/fluency ratings) measure these two dimensions well. But fictional speech—especially in anime and manga—has a third dimension that X-Y evaluation cannot capture:

- **Z-axis (Dynamics):** Why did the character *have to* say it this way? — the internal conflict and character-specific bias that shaped the utterance

The Z-axis asks not "what was said" or "how it was phrased," but **"what psychological process forced this particular surface form?"**

2.3 Audience Design and Address Mode

A key reason "meaning-only" translation fails for character dialogue is that utterances are designed for an addressee. Audience design research shows that speakers adjust linguistic form to the listener (Bell, 1984), while interactional pragmatics emphasizes how footing and stance determine what an utterance "counts as" in the scene (Goffman, 1981).

TAP makes this explicit via an **address mode** feature: whether the utterance is (i) direct address to the addressee, (ii) reported/narrated about someone, (iii) self-directed monologue, or (iv) indirect/implied.

This matters because AVT mistakes often shift the deictic center. Japanese media dialogue can use self-reference by name ("レムは…") as a politeness/intimacy device;

naïvely mapping it to third-person English can turn a direct confession into a line that reads like a summary.

In E1, the DeepL-style "Rem loves Subaru" preserves much of the propositional content but moves the scene from *speaking to him* to *talking about her*—which is why both ZAP and IAP penalize it as an address-mode mismatch.

2.4 The Two-Layer Structure of the Z-Axis

The Z-axis consists of two interacting layers:

Layer 1: Conflict (Ξ) — The horizontal force

What internal tensions are active at the moment of speech?

- "I want to express affection" vs. "I want to protect myself"
- "I want to be honest" vs. "I don't want to seem vulnerable"

These conflicts create *interference patterns* in speech: hesitation, hedging, contradiction, repair sequences.

Layer 2: Bias — The vertical force

How does *this specific character* resolve that tension? Different characters with identical conflicts produce different outputs:

Character	Conflict	Bias Pattern	Typical Output
Rem	affection vs. modesty	devotion-first, deflects praise	"No… that's only natural, Subaru-kun."
Kurisu	affection vs. pride	tsundere, leak-then-overwrite	"N-not that it's for you or anything…"

2.5 Negation Directionality: A Micro-Theory

A practical micro-theory that falls out of this view is **negation directionality**.

A tsundere line ("別に…あんたのためじゃない") often uses **concealment negation**: surface ¬P is deployed to hide underlying P and can generate repairs ("I mean—").

Luffy's refusal is **declaration negation**: ¬P is the intended commitment and escalates into an ultimatum.

Treating these as distinct action-shapes prevents "negation" from collapsing into a single generic translation template.

2.6 Why Both Layers Disappear in Translation

Consider this Japanese line:

「まあ…別にいいんだけどさ…うん」

A context-blind machine translation might produce:

"It's fine."

Layer	What's happening in Japanese	Standard EN translation
Conflict (Ξ)	"want to say yes" vs. "don't want to seem eager"	**Lost**
Bias	avoidance pattern, hedging, self-interruption	**Lost**
Surface markers	「まあ」 (hedge) 「別に」 (deflection) 「…」 (hesitation)	**Lost**

The *meaning* survives. The *dynamics* are erased.

2.7 Formalization: Commutativity as a Design Target

Let Φ be an analysis function that maps an utterance (optionally with context) to an act structure A—a multiset of act tuples:

```
a = (act_type, target_role, target_entity, intensity, force_text)
A = {a₁, a₂, ...}
```

Let T be a translation function. TAP's goal can be written as an approximate commutative diagram:

```
u (L₁)  --T-->  u' (L₂)
   |Φ             |Φ
   v              v
A (acts) --≈-->  A' (acts)
```

IAP operationalizes "≈" with weighted similarity over act types, targets, and intensity, plus explicit penalties for critical mismatches (especially address mode).

2.8 Why Persona, Context, and Listener Belong in the Input

The experiments (E2/E3) show that the same string can be a different action depending on persona, context, and listener. This aligns with pragmatic accounts in which interpretation depends on common ground and conversational goals (Grice, 1975; Clark, 1996).

Fiction intensifies this dependency because persona is *authored*: "what the character would do with this line" is part of the meaning.

For TAP, this implies a translation function closer to professional practice:

```
T : (text × persona × context × listener) → translated_text
```

The persona YAML and the YAML generator treat translator intuition as something that can be externalized into inspectable structure: conflicts, biases, triggers, and listener assumptions become inputs rather than invisible "taste."

---

3. System Overview

3.1 Z-Axis Translation (generation)

Our Z-axis translation pipeline generates a target line conditioned on:

- **Source line**
- **Persona**
- **Context**
- **Listener type / relationship**
- **Z-axis intensity** (low/medium/high)

The generator produces both: 1) a translation, and
2) a structured "Z-signature" describing surface markers (hesitation, negation patterns, repair/self-correction, "leak then overwrite", etc.).

3.2 Persona Engineering: Why It's Essential

Z-axis translation cannot work without **high-resolution persona data**. The same line, translated without persona context, will lose the character's voice.

Why? Because the Z-axis (conflict + bias) is *character-specific*. Kurisu's "別に" and Rem's "別に" perform completely different actions, even though the surface form is identical.

Persona YAML Structure

We represent each character as a structured YAML with five key components:

Component	What it captures	Example (Kurisu)
conflict_axes	Internal tensions that generate speech	"want to admit feelings" vs. "can't be honest" (weight: 0.95)
bias	Default resolution	Tsun→Dere→Overwrite (deny, leak,

Component	What it captures	Example (Kurisu)
	pattern	cover)
emotion_states	State-specific surface markers	flustered: stutter_count=1, negation_first=true
triggers	What causes Z-axis spikes	Called "assistant" → z_delta: +0.3
speech_patterns	Character-specific verbal markers	Negation phrases: "別に", "勘違いしないで"

Example: Kurisu's Conflict Axes
```
conflict_axes:
  - axis: "認めたい vs 素直になれない"
    side_a: "好意を認めたい"
    side_b: "素直になれない"
    weight: 0.95  # Almost always active

  - axis: "論理 vs 感情"
    side_a: "論理的でいたい"
    side_b: "感情が溢れる"
    weight: 0.9
```

Example: Kurisu's Bias Pattern
```
bias:
  expression_pattern: "Tsun-Dere-Overwrite"
  pattern: "ツン → 素直（漏れ） → 即座にツンで上書き"
  rule: |
    Z軸が上昇すると、まず否定・強がりで防御。
    しかし本音が漏れ出し、それに気づくと慌てて
    ツンで上書きしようとするが、残響（residual）が残る。
```

This bias pattern explains why Kurisu's translations need: - **Negation first** ("N-not that…") - **Emotional leak** (brief honesty) - **Overwrite** (defensive recovery) - **Residual** (trailing off, incomplete sentences)

Example: Triggers and Z-Delta
```
triggers:
  - trigger: "「助手」と呼ばれる"
    z_delta: "+0.3"
    surface_effect: "stutter_increase, loud_denial"
    example_response: "助手じゃない！"

  - trigger: "好意を直接指摘される"
    z_delta: "+0.5"
```

```
surface_effect: "negation_first, overwrite_required, residual_leak"
example_response: "はぁ!? 何言ってんの！…別に…"
```

Without this trigger information, a translator (human or machine) cannot know that "assistant" causes a +0.3 Z-spike that manifests as loud denial with increased stuttering.

Scaling: Persona Generator

To scale experiments, we implemented a **persona generator** that scaffolds high-resolution personas from canonical character evidence (dialogue samples, character descriptions, fan wiki analysis).

We also built a **YAML generator** that produces complete translation request YAMLs (*scene / relationship / context_block / emotion_state / z_axis_intensity*) from a persona + target line + brief scene hint (see Appendix B).

---

4. Evaluation: IAP and ZAP

4.1 From Comparing Sentences to Comparing Effects

Traditional automatic metrics—BLEU (Papineni et al., 2002), chrF (Popović, 2015), COMET (Rei et al., 2020)—have proven effective for general-purpose translation evaluation. We do not claim these metrics are flawed; they measure what they are designed to measure.

However, **for anime and manga translation, they are not enough.**

The reason is structural: these metrics evaluate whether the *sentence* is equivalent, not whether the *effect on the scene* is equivalent. In character-driven media where dialogue performs narrative work—confessing, threatening, deflecting, closing escape routes—a translation can be semantically accurate yet pragmatically broken.

This is not a hypothetical concern. As native speakers of the source language and consumers of the target translations, we observe that "correct" subtitles routinely flatten character voice, shift address modes, and neutralize emotional dynamics. These failures are invisible to n-gram overlap or embedding similarity.

TAP therefore introduces a different evaluation target:

Traditional metrics compare sentences. TAP compares effects.

- Did Rem's words, in English as in Japanese, **close Subaru's escape routes**?
- Did Luffy's words, in any language, **stop Rayleigh from answering**?

This is the shift from X-Y evaluation to Z-axis evaluation: not "are the words equivalent?" but **"did the line do the same thing to the listener and the scene?"**

4.2 IAP (Illocutionary Act Preservation)

IAP = Illocutionary Act Preservation rate

Following Austin (1962) and Searle (1969), we treat each line as performing one or more *illocutionary acts*—actions accomplished through speech (confessing, refusing, threatening, closing escape routes, issuing ultimatums, etc.).

Calculation method:

1. **Extract acts from source (A_1):** What actions does the original line perform?
2. **Reverse-analyze acts from translation (A_2):** What actions does the translated line perform?
3. **Judge isomorphism:** Are the same acts preserved?

$$\text{IAP} = \frac{|A_1 \cap A_2|}{|A_1|}$$

Example: Rem's confession (E1)

Act	Source (JP)	DeepL	Z-axis
CONFESS (direct, to listener)	✓	✗ (becomes REPORT)	✓
DECLARE EXCLUSIVITY	✓	✓	✓
CLOSE ESCAPE ROUTES	✓	partial	✓

- DeepL IAP ≈ 0.51 (confession act lost)
- Z-axis IAP ≈ 0.76 (core acts preserved)

Example: Luffy's ultimatum (E3)

Act	Source (JP)	DeepL	Z-axis
REFUSE (information)	✓	✓	✓
ISSUE ULTIMATUM	✓	✓	✓
SET BOUNDARY (protect journey)	✓	partial ("don't care" ≠ boundary)	✓ ("don't wanna hear")

Both preserve the explicit ULTIMATUM, but only Z-axis preserves the *framing* of the refusal as boundary-setting rather than apathy.

4.3 Key Extensions for Anime/Manga

During development, we added domain-specific extensions:

- **Address Mode** classification (**direct** vs **reported**) with penalty when direct confession becomes narration.
- **Negation-type** support (e.g., concealment vs declaration) to better model anime-specific negation styles.
- Entity matching upgrades (e.g., kana→romaji normalization for named addressees).

4.4 ZAP (Z-Axis Preservation)

While IAP measures whether the **acts** are preserved, ZAP measures whether the **character's voice and emotional dynamics** survive translation.

ZAP = Z-Axis Preservation score

ZAP asks: "Does this translation sound like something *this character* would actually say in *this situation*?"

Four evaluation axes (each 0.0–1.0):

Axis	Question	What it catches
Character Voice	Does it sound like this character?	Out-of-character word choices, wrong speech patterns
Emotional Intensity	Is the emotional strength preserved?	Passionate confession becoming clinical, anger becoming mild annoyance
Listener Relationship	Is the addressee relationship preserved?	Direct confession becoming third-person narration
Speech Pattern	Are hesitation, negation, and verbal tics preserved?	Loss of stutter, deflection patterns, character-specific markers

Critical issues ZAP flags:

- Third-person narration replacing direct confession
- Loss of intimacy/directness
- Emotional flattening
- Out-of-character word choices
- Wrong politeness register

Example: Rem's confession (E1)

Axis	DeepL	Z-axis

Axis	DeepL	Z-axis
Character Voice	0.5 (clinical tone)	0.9 (Rem's gentle intensity)
Emotional Intensity	0.4 (flattened)	0.9 (preserved)
Listener Relationship	0.2 (direct → reported)	0.95 (direct preserved)
Speech Pattern	0.6 (some markers lost)	0.8 (hesitation preserved)
ZAP Overall	**0.46**	**0.86**

4.5 The IAP-ZAP Division of Labor: A Deliberate Design

A critical design principle: **IAP and ZAP do not compete—they evaluate different layers.**

Evaluator	Question	In Scope	Deliberately Out of Scope
IAP	Was the act preserved?	act_type, target, intensity, address_mode	Character voice, stylistic naturalness
ZAP	Does it feel like the character?	voice, emotion, cultural register	Act-level identity judgment

Why this separation matters:

IAP correctly treats these as equivalent at the act level: - "I love you" - "You're the one I choose"

Both perform EXPRESS + ASSERT_CHOICE. But whether the line sounds like *Rem* vs. *Kurisu* vs. *Tomoe* is not IAP's job—that's ZAP's domain.

Similarly, IAP does not distinguish between: - "I can't bear it" - "I won't accept it"

At the act level, both perform CLOSE_ESCAPE. The nuance difference is a ZAP concern (emotional register, character voice).

This two-layer architecture means:

- A translation can score **high IAP + low ZAP** (acts preserved, but character voice lost)
- A translation can score **low IAP + high ZAP** (sounds like the character, but wrong act performed)
- A **good translation** scores high on both: acts preserved *and* character voice intact

This division prevents the evaluation from collapsing into a single score that obscures *what kind* of failure occurred.

5. Experiments and Results

5.1 E1 — Re:Zero (Rem): Address Mode as a Decisive Failure Mode

Source (JP):
レムは、スバルくんを、愛しています。
スバルくんが良いんです。スバルくんじゃなきゃ、嫌なんです。

DeepL (representative):
Rem loves Subaru. Subaru is the one. I won't accept anyone but Subaru.

Z-axis translation (direct confession):
I love you, Subaru-kun.
Subaru-kun… you're the one I choose.
If it isn't you… I can't bear it.

E1-A. IAP (Address Mode) comparison

System	Address Mode	IAP Overall	Key issue
DeepL	direct → **reported**	**0.51**	Confession becomes narrated report (intimacy loss)
Z-axis	direct → **direct**	0.76	Preserves direct confession to listener

Interpretation: meaning-only translation can preserve "love + exclusivity" while still breaking the core audiovisual function: **face-to-face confession**.

E1-B. ZAP comparison (voice + intimacy)

System	ZAP Overall	Listener relationship	Summary
DeepL	0.46	**fails** (direct → third-person)	Reads like a summary, not lived confession
Z-axis	0.86	**preserved**	High-intensity confession remains intimate

5.2 E2 — Steins;Gate (Kurisu): Persona and Context Dependence

E2 tests a TAP claim:

> The same surface form can perform different actions depending on **persona** and **context**, and a translation should preserve *that*.

5.2.1 E2-A (Persona sweep): same line, different personas

Fixed context: reply after being thanked
Fixed JP line: 「そんなことは当然です」
Varied persona: Tomoe / Rem / Kurisu

Persona	Intended action	Z-axis English output (example)
Tomoe (controlled yandere secretary)	duty framing with suppressed affection leak	"Of course, Master— it's only natural."
Rem (devoted, modest)	deflect praise into devotion/"it's what I should do"	"No… that's only natural, Subaru-kun."
Kurisu (tsundere)	embarrassment defense; denial/overwrite	"N-not that it's a big deal or anything… it's only natural."

Takeaway: even when all say "当然", the *action* differs: obligation, devotion, or tsundere deflection.

5.2.2 E2-B (Context + listener sweep): same persona, same line, different contexts

Fixed persona: Makise Kurisu
Fixed JP line: 「別に…あんたのためじゃないから。」
Varied context: daily / tense / jealous / alone / monologue (listener=self)

This is where context-blind MT typically collapses nuance.

E2-B.1 Baseline vs Z-axis outputs

Condition	Baseline MT output (context-blind)	Z-axis output (context-conditioned)
Any context (DeepL-like)	"Not that it's for you or anything."	*(cannot vary; no listener/context model)*
Daily (mild embarrassment)	—	"N-not like I did it for you or anything…"
Tense (post-danger)	—	"N-not like… I did it for you or anything, okay."
Jealous (caught caring)	—	"Not that I—I mean… it's not like I did it for you, okay."
Two-person alone	—	"Not that it's… f-for you or anything, okay."

Condition	Baseline MT output (context-blind)	Z-axis output (context-conditioned)
Monologue (self-persuasion)	—	"Not like… it's f-for you or anything. I mean— it's not."

E2-B.2 Marker table (what changes, and why it matters)

Context	Stutter locus	Repair/self-correction	Listener tag	Extra denial	Action-level interpretation
Daily	N-not	—	—	—	standard tsun defense
Tense	N-not + pause	—	**okay**	—	defensive confirmation to listener
Jealous	—	**I—I mean...**	okay	—	emotion leak → overwrite
Alone	**f-for**	—	okay	—	"for you" itself becomes hard to say
Monologue	**f-for**	**I mean—**	—	**double denial**	self-deception: denying to *self*

Takeaway: this is not "extra flavor" but action: *deflection*, *self-protection*, *repair*, and *self-binding*.

5.3 E3 — One Piece (Luffy vs Rayleigh): Round-trip Robustness of ULTIMATUM

E3 tests whether a key pragmatic action survives EN→FR→EN:

> **ULTIMATUM preservation**: "If you tell us here, I quit being a pirate."

Source (JP): (Sabaody; Luffy refuses spoilers)

- 宝がどこにあるかなんて聞きたくねェ!!!
- 宝があるかないかだって聞きたくねェ!!!
- …ここでおっさんから何か教えて貰うんなら おれは海賊やめる

E3-A Baseline (DeepL-like) round-trip

Stage	Key refusal	ULTIMATUM clause (EN)
JP→EN	"I don't give a damn…"	"I'm quitting being a pirate."
EN→FR	"Je me fiche complètement…"	"je vais arrêter d'être pirate."
FR→EN	"I don't care…"	"I'll quit being a pirate."

E3-B Z-axis round-trip

Stage	Key refusal	ULTIMATUM clause
JP→EN	"I don't wanna hear…"	"then I'm quitting being a pirate."
EN→FR	"J'veux pas entendre…"	"alors j'arrête d'être pirate."
FR→EN	"I don't wanna hear…"	"then I'm done being a pirate."

E3-C What this shows (and what it doesn't)

- **Good news:** ULTIMATUM is structurally explicit, so it survives round-trip for both pipelines.

- **Where TAP still matters:** refusal framing ("don't care" vs "don't wanna hear") affects **character voice**:
    - Luffy's action is not apathy ("don't care"); it is **boundary-setting to protect the journey**.

---

6. Discussion: Why TAP helps (and when it's hard)

6.1 TAP highlights failures invisible to meaning preservation

E1 shows that "meaning preserved" can still destroy: - direct confession, - intimacy, - audiovisual performance function.

6.2 TAP predicts which lines are fragile vs robust

- **Robust:** explicit actions like ULTIMATUM (E3)

- **Fragile:** lines whose action depends on:
    - address mode (E1),
    - concealment/negation style (E2),
    - persona-specific speech patterns.

6.3 What we are not claiming

This report does not argue that "machines beat translators."
Instead, it externalizes one part of expert practice: **keeping action intact across languages**.

---

7. Conclusion

TAP reframes anime/manga translation quality as **action preservation**. With IAP and ZAP, we can detect when translations keep meaning but lose the line's pragmatic

function—especially in direct confession, persona-driven deflection, and listener-sensitive speech patterns.

The Z-axis framework—viewing translation as preserving not just semantics and style, but the underlying **conflict-bias dynamics**—offers both a diagnostic tool and a generative principle for fictional speech translation.

Appendix A — Prompt & schema transparency (excerpt)

- Prompts are designed to be publishable.
- Outputs are schema-constrained JSON.
- Matching is deterministic given extracted acts.

(Repository contains full prompts and JSON schemas.)

Appendix B — Tooling notes (persona generator, YAML generator, dialogue mode)

B.1 Persona generator (why it matters)

High-resolution persona is not "style fluff": it determines which conflicts activate and what surface markers appear (stutter, repairs, denial type, etc.).

We use persona YAMLs with: - conflict axes (+ activation) - bias patterns (e.g., Tsun–Dere–Overwrite) - triggers and risk flags

B.2 YAML generator (persona + line + hint → request YAML)

To scale experiments and ensure reproducibility, we implemented **yaml_generator.py**, which:

1) reads persona YAML

2) takes a **target line** and short **scene hint**

3) asks an LLM to output JSON matching a strict schema:
    - scene
    - relationship
    - context_block
    - emotion_state
    - z_axis_intensity
    - listener_type_hint

4) formats the JSON into a ready-to-run request YAML.

CLI example:

```
python yaml_generator.py \
  --persona personas/kurisu_v2.yaml \
  --line "別に...あんたのためじゃないから。" \
  --hint "独り言、岡部が他の女と話してて嫉妬" \
  --output requests/kurisu_context_monologue.yaml
```

Generated request YAML includes: - persona_file - scene / relationship - context_block (script-like build-up) - target_line / target_lang - z_axis_intensity / emotion_state - notes (listener_type_hint + reasoning)

B.3 Dialogue mode (two personas)

For short dialogue scenes, we added `z_axis_dialogue.py` supporting: - two persona files (speaker A/B) - per-turn trigger-based Z accumulation - per-turn translation while preserving each speaker's voice - summary report of the whole dialogue

Example config excerpt:

```
personas:
  A: "personas/kurisu_v2.yaml"
  B: "personas/okabe_v2.yaml"

scene: "ラボ、午後"
relationship: "両思い、未告白"

dialogue:
  - speaker: B
    line: "助手よ、コーヒーを"
  - speaker: A
    line: "助手じゃない！...はぁ、砂糖は？"
```

References

Austin, J. L. (1962). *How to do things with words*. Oxford University Press.

Bell, A. (1984). Language style as audience design. *Language in Society, 13*(2), 145–204. https://doi.org/10.1017/S004740450001037X

Clark, H. H. (1996). *Using language*. Cambridge University Press.

Goffman, E. (1981). *Forms of talk*. University of Pennsylvania Press.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41–58). Academic Press.

Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002). BLEU: A method for automatic evaluation of machine translation. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 311–318. https://doi.org/10.3115/1073083.1073135

Popović, M. (2015). chrF: Character n-gram F-score for automatic MT evaluation. *Proceedings of the Tenth Workshop on Statistical Machine Translation*, 392–395. https://doi.org/10.18653/v1/W15-3049

Rei, R., Stewart, C., Farinha, A. C., & Lavie, A. (2020). COMET: A neural framework for MT evaluation. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, 2685–2702. https://doi.org/10.18653/v1/2020.emnlp-main.213

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge University Press.