

Development of English translations for use in a visualization system of classical Japanese poetic vocabulary

Hodošček Bor
Osaka University, Graduate School of Humanities
hodoscek.bor.hmt@osaka-u.ac.jp

Yamamoto Hilofumi
Tokyo Institute of Technology
yamagen@ila.titech.ac.jp

Summary

Utilized structured generation techniques with multiple open-source and closed-source LLMs to programmatically create English translations and word-level alignments of classical Japanese poetry. Evaluations demonstrate that chunk- and line-based translations achieve best quality while word-level translations remain useful for generating English tokens for the visualization of Japanese poetic vocabulary. Prompt-based word alignment was also employed to recover token mappings from line- and chunk-based translations. [URL https://github.com/borh/hachidaishu-translation](https://github.com/borh/hachidaishu-translation)

① Introduction and Motivation

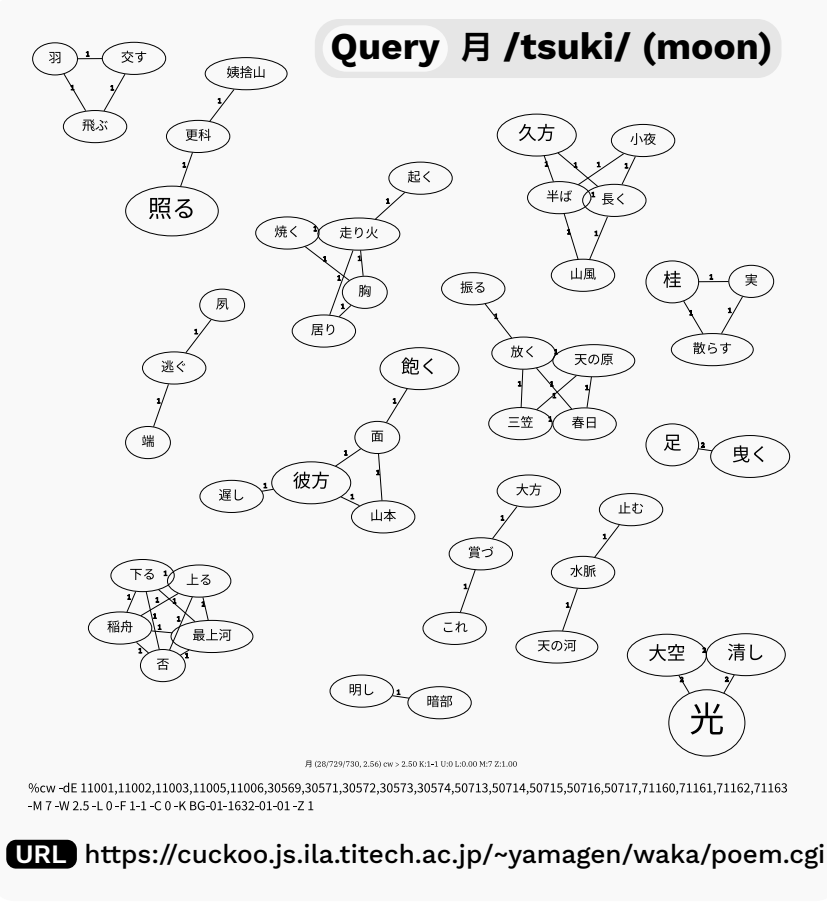
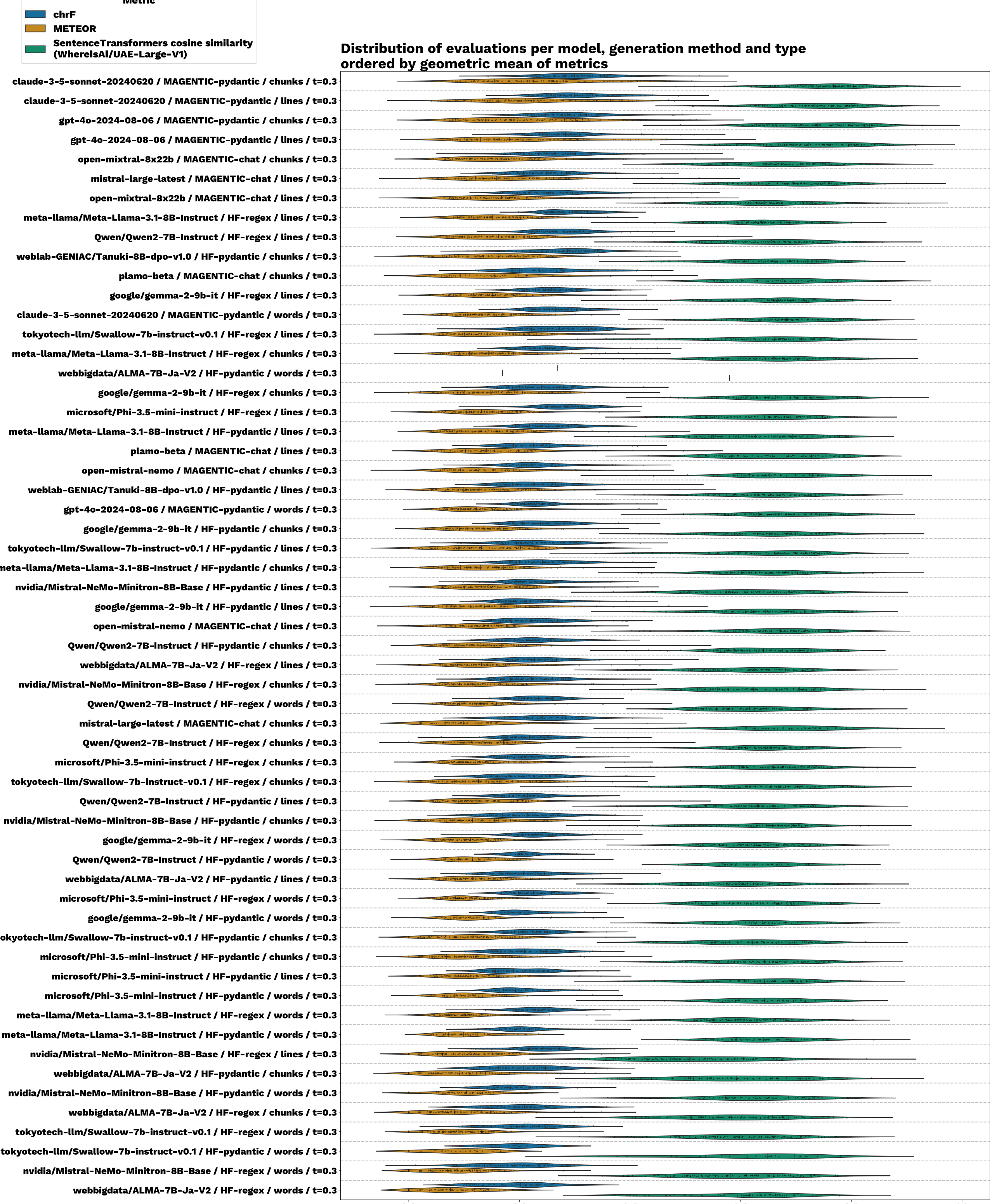
Working on visualizing lexical relationships in graphs showing Japanese terms and their corresponding English translations. During the manual development of an English token dictionary, we noticed that GitHub Copilot provided useful completion tokens but sometimes introduced inaccuracies. However, a human-in-the-loop approach was found to be highly effective in our pilot experiment.

③ Results: translations evaluation

Translations are assessed comparing to our golden translation dataset (Rodd, L. R., & Henkenius, M. C. (Eds.). (1996). *Kokinshū: A Collection of Poems Ancient and Modern*. Cheng & Tsui.). We randomly sample 100 translations and use three different translation granularities (word, chunk and line) to perform a total of a maximum 300 translations per model below. (Model with / in their name are using their HuggingFace Models identifier; the rest are API-based models). Models were chosen from a variety of best-performing as well as specialist models (webbigdata/ALMA-7B-Ja-V2 for Japanese-English translation).

Evaluation metrics employed:

chrF: surface-level metric; F-score of character-level n-grams in translations
METEOR: considers semantic equivalence through synonym matching and stemming
SentenceTransformers cosine similarity (Wherels/UAE-Large-V1): sentence embeddings capture semantic similarity



② Pilot Experiment

Translations were produced using ChatGPT-4, Copilot, Copilot with human assistance, and human translations. Findings: ChatGPT-4 offered natural and easy-to-understand translations but sometimes diverged too far from the source content. Copilot was more accurate but produced less natural translations. With human refinement, it balanced accuracy and fluency. Human translations provided the best literary nuance but were costly and time-consuming.

Problem: Manual translation of Japanese tokens into English is highly context-sensitive and time-consuming.
Proposed solution: Leverage structured generation with LLMs to programmatically assign English tokens.

We use a previously open-sourced dataset of the Hachidaishu, a collection of eight Japanese imperial anthologies (ca. 905-1205). The Hachidaishu database consists of *waka* poems of 5-7-5-7-7 syllabus length.

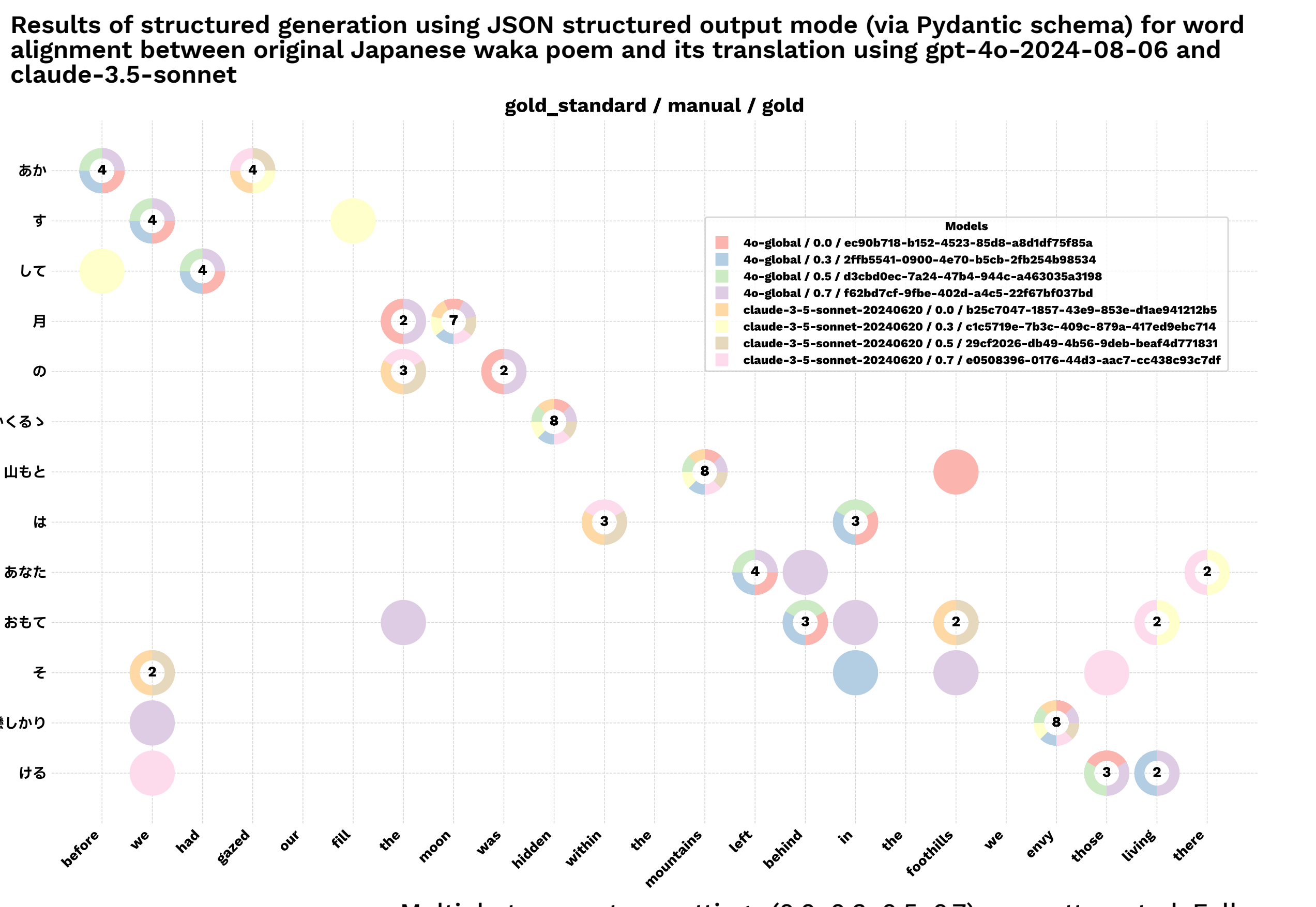
Structured generation overview

```
SYSTEM PROMPT = """You are an expert Japanese-English poetry translator. Your task is to faithfully translate given Japanese waka poems into English. Ensure that each translation is between 16 and 48 words long. Provide only the translated poem without any additional notes or formatting unless ordered to do so."""
```

```
class WordSchema(BaseModel):  
    original: str = Field(description="Original Japanese word")  
    translated: str = Field(description="Translated English word")  
    words: List[WordSchema] = Field(  
        description="List of original-translated word pairs."  
    )  
magentic.prompt("""[Instructions]  
Translate the following waka poem into English, word-by-word.  
Output the translation for each word using at most 3 English words.  
## Japanese  
(waka)  
## Japanese-English word pairs  
""")  
def magentic_words_prompt_fn(instructions: str, waka: str) -> WordSchema: ...
```

```
magentic.chat_prompt(  
    magentic.SystemMessage(SYSTEM_PROMPT),  
    magentic.UserMessage(ONE_SHOT_EXAMPLE),  
    magentic.AssistantResponse(ONE_SHOT_RESPONSE),  
    magentic.UserMessage("Translate the following [...]").  
)
```

④ Results: source-target word alignment



Examples with high divergence between chrF, METEOR and ST (percentile)

- A** weblab-GENIAC/Tanuki-8B-dpo-v1.0 / HF-pydantic / lines
chrF = 0.15 (0.05) METEOR = 0.12 (0.46) ST = 0.77 (0.90)
かは風の / すしくもある / かうちよする / 浪とよもにや / 秋は立らん
MT The wind, like a gentle whisper, stirs the waves and autumn stands firm
GT how cool the breeze / from the river I feel the / refreshing chill / of near-approaching autumn / roll in with each rising wave
- B** meta-llama/Meta-Llama-3.1-8B-Instruct / HF-regex / chunks
chrF = 0.18 (0.15) METEOR = 0.04 (0.04) ST = 0.80 (0.95)
春風は / 花のあたりを / まきてふけ / 心づからや / うつろふとみん
MT In the spring wind / flowers / gently / from the heart / to be distracted from my thoughts and be
GT oh spring breeze please do / not draw near these petals as / you waft for I would / like to know whether it is / their own wish that they should fall
- C** mistral-large-latest / MAGENTIC-chat / chunks
chrF = 0.30 (0.82) METEOR = 0.04 (0.06) ST = 0.80 (0.96)
契りけん / 心そつらき / セタの / 年に一たひ / 逢はあふかは
MT Bound by a promise / Heart yearning, so cruel / On Tanabata night / Once a year we meet
GT our one promised night / of the year will it come to / pass Tanabata / oh too cruel is the young / heart that would deny me this
- D** claude-3.5-sonnet-20240620 / MAGENTIC-pydantic / chunks
chrF = 0.53
色よりも / かこそあはれと / おもほゆれ / たか袖ふれし / やとの梅そも
MT More than its color, / it's the fragrance that moves me: / the plum blossoms / at the house where once / your sleeve brushed against mine.
GT more than the color / of the flower the fragrance / delights my senses-- / whose scented sleeve brushed against / the plum blossoms near my house
- E** Qwen/Qwen2-7B-Instruct / HF-regex / lines
METEOR = 0.56
新しき / 年の始に / かしくこそ / 千年をかねて / たのしきをつめ
At the start of the new year, we anticipate a thousand years of joy and fulfillment, nestled within the freshness of the moment, yet deeply rooted in tradition and hope for the future, embodying the eternal cycle of life and, at the beginning / of this auspicious new year / we anticipate / the joys of a thousand years / by piling stacks of firewood
- F** claude-3.5-sonnet-20240620 / MAGENTIC-pydantic / chunks
ST = 0.93
かは風の / すしくもある / かうちよする / 浪とよもにや / 秋は立らん
MT The cool river breeze / How refreshing it feels / As waves lap the shore / Together they herald / The arrival of autumn
GT how cool the breeze / from the river I feel the / refreshing chill / of near-approaching autumn / roll in with each rising wave

⑤ Discussion

In general, frontier models (claude-3.5-sonnet and gpt-4o) outperformed smaller models, but preclude the use of regex-based generation methods that offer greater control over the exact tokens generated. Regex-based generations were generally superior to JSON schema structured output on open models usable through the Outlines structured text generation library, but their use is not supported in most API-based models.

Automated evaluations show moderate correlation but also disagreement and need further analysis to determine the most useful metric (or combinations of metrics) to use for guiding the creation of English word tokens.

⑦ Limitations and Future Work

Expanding models' context through the integration of detailed linguistic information such as part-of-speech (POS) data and pre-developed semantic labels from previous work (github:yamagen/hachidaishu-pos) is likely to help. Also, expanding the use of few-shot prompting (only used for chat-based API use currently) will likely improve format conformance for JSON schema based models. Assessing the optimal alignment with the Leipzig glossing rules when integrating tokens remains an open question. Future work will explore this issue as well as perform contamination checks on the original Japanese waka and their English translations, and expand the golden translation dataset to other established translations.