

Keigo Identification and Proficiency Assessment

Proposed Approach:

1. User Input:

- Accept Japanese text input from the user.

Example Input:

"田中さんは先生です。私はご飯を食べていただきます。"

2. Text Conversion:

- Convert the text to **Romaji** using tools like **MeCab** and **Romkan** for easier processing.

Converted Text (Romaji):

"Tanaka-san wa sensei desu watashi wa gohan o tabete itadakimasu"

3. Keigo Identification:

- Use **regular expressions** to identify **Keigo forms** (Sonkeigo, Kenjougo, Teineigo) and keep a **count** of each type.
- **Sonkeigo** (尊敬語): Polite forms used to elevate others
- **Kenjougo** (謙譲語): Humble forms used to lower oneself .
- **Teineigo** (丁寧語): Polite speech used in formal situations.

Example Analysis and Frequencies:

Identified Keigo:

- Sonkeigo: 2 occurrences (sensei-sama, o-tabete itadakimasu)
- Kenjougo: 1 occurrence (watashi)
- Teineigo: 2 occurrences (desu, masu)

4. LLM Evaluation:

LLM Prompts:

1. **Prompt 1:** "Is the Keigo usage contextually correct?"
Response: Simple **Yes** or **No**.
2. **Prompt 2:** "Explain why the Keigo usage is correct or incorrect."
Response: A detailed explanation of the correctness of the Keigo usage.

Example LLM Response:

1. **LLM Answer (Prompt 1): "No"**
2. **LLM Reasoning (Prompt 2):** *"In polite Japanese, 'sensei' itself is a respectful term, and adding '-sama' is redundant unless addressing someone in an extremely formal or ceremonial context. The use of 'itadakimasu' is correct for humility when referring to receiving food, and 'watashi' is appropriate for Kenjougo in this context. However, 'sensei-sama' is an overuse of honorifics."*

5. Proficiency Categorization:

Based on the identification of different Keigo forms, determine the user's **proficiency level**:

- **Basic:** The sentence contains **Teineigo** (e.g., **masu, desu**).
- **Intermediate:** The sentence contains **Teineigo** and **Sonkeigo** (e.g., **sama, masu**).
- **Advanced:** The sentence contains **Teineigo**, **Sonkeigo**, and **Kenjougo** (e.g., **sama, masu, itadakimasu**).

Proficiency for Example:

Proficiency Level: **Advanced**

Explanation: All three types of Keigo (Sonkeigo, Kenjougo, Teineigo) were present.

6. Scoring System:

The score will be based on three major dimensions:

1. **Proficiency:** Basic, Intermediate, or Advanced.
2. **Accuracy:** How correctly the user has applied the Keigo, as evaluated by the LLM.
3. **Frequency:** How often each type of Keigo is used, indicating mastery.

SCORES

1. Proficiency Evaluation (0-10 points):

Proficiency is based on the presence and correct usage of **Teineigo**, **Sonkeigo**, and **Kenjougo** in the user's input. To evaluate this:

Proficiency Scoring:

- **Basic** (0-3 points): The user includes **Teineigo** (e.g., "desu", "masu").
 - **Score Range:**
 - 1 point for occasional Teineigo use. (count \leq 1)
 - 2-3 points for consistent use. (count $>$ 1)
- **Intermediate** (4-7 points): The user includes both **Teineigo** and **Sonkeigo** (e.g., honorifics like "-sama", respectful verbs like "o-tabete itadakimasu").
 - **Score Range:**
 - 4 points: Teineigo is consistent, and there's minimal Sonkeigo use.
 - 5-6 points: Teineigo is consistent, Sonkeigo is occasionally used correctly.
 - 7 points: Good balance of Teineigo and Sonkeigo.
- **Advanced** (8-10 points): The user uses all three forms: **Teineigo**, **Sonkeigo**, and **Kenjougo** (e.g., humility in self-referencing like "watashi" with Sonkeigo for others).
 - **Score Range:**
 - 8 points: Teineigo, Sonkeigo are present; Kenjougo may have slight errors.
 - 9 points: All forms used correctly but some minor redundant or awkward constructions.
 - 10 points: Perfect use of Teineigo, Sonkeigo, and Kenjougo without any redundancy or misuse.

2. Accuracy/ Relevance Evaluation

Since Keigo usage is cultural, subjective and not a grammar use, there cannot be a score to measure the accuracy. However, we can make use of the LLM reasoning as a metric to explain the relevance of Keigo Usage in the given answer of the candidate.

Accuracy Scoring:

- **Yes/No Response:** First, the LLM gives a binary answer: **Yes** or **No** for whether the usage is correct.
- **Follow-up reasoning:** The LLM then explains why the Keigo usage is correct or incorrect.

3. Frequency Evaluation (0-5 points):

Frequency relates to the **balanced** use of different Keigo forms. The idea is to assess whether the user is overusing or underusing specific types of Keigo, or maintaining a natural and balanced flow.

Ideal Frequency:

- **Balanced frequency** means that the user uses **Teineigo** frequently enough for polite speech, but also incorporates **Sonkeigo** and **Kenjougo** appropriately without overloading the sentence.

Frequency Scoring:

- **Full Marks (5 points):** Balanced use of Teineigo, Sonkeigo, and Kenjougo.
 - Example: 3 occurrences of Teineigo, 2 occurrences of Sonkeigo, 1 occurrence of Kenjougo (3:2:1 ratio is ideal).
- **Moderate Score (3-4 points):** Slight overuse of one form or underuse of another, but the sentence still feels natural.
 - Example: Teineigo is used frequently, but Sonkeigo is missing or used too much.
- **Low Score (0-2 points):** Overuse or underuse of a form, or very few forms are present.
 - Example: Teineigo is the only form present, no Sonkeigo or Kenjougo.

Example Scoring

1. Proficiency Score (Out of 10):

- Identified Keigo:
 - Sonkeigo: 2 occurrences (sensei-sama, o-tabete itadakimasu)
 - Kenjougo: 1 occurrence (watashi)
 - Teineigo: 2 occurrences (desu, masu)

Since the sentence contains all three forms (Sonkeigo, Kenjougo, and Teineigo), the proficiency level is Advanced. However, the overuse of Sonkeigo (in "sensei-sama") leads to a minor deduction.

- **Proficiency Score: 8/10**

2. Accuracy Evaluation:

- LLM Response (Prompt 1): No
- LLM Reasoning (Prompt 2):
"In polite Japanese, 'sensei' itself is a respectful term, and adding '-sama' is redundant unless addressing someone in an extremely formal or ceremonial context. The use of 'itadakimasu' is correct for humility when referring to receiving food, and 'watashi' is appropriate for Kenjougo in this context. However, 'sensei-sama' is an overuse of honorifics."

3. Frequency Score (Out of 5):

- Ideal Ratio (3:2:1)
The actual identified frequencies are Teineigo: 2, Sonkeigo: 2, Kenjougo: 1, which fits the ideal 3:2:1 ratio reasonably well. However, the overuse of Sonkeigo ("sensei-sama") slightly disrupts this balance.
- **Frequency Score: 4/5**

COMMON REGEX PATTERNS IDENTIFIED

Regular Teineigo patterns

teineigo_patterns = [

`r"\bo-\w+"`, # Words starting with "o-" (e.g., o-hanashi)

`r"\bgo-\w+"`, # Words starting with "go-" (e.g., go-zonji)

`r"\bw+-masu$"`, # Words ending with "-masu" (e.g., o-kai ni narimasu)

`r"\bw+-desu$"`, # Words ending with "-desu" (e.g., Tanaka-san wa sensei desu)

`r"\bw+-gozaimasu$"`, # Words ending with "-gozaimasu" (e.g., gozaimasu)

`r"\bw+-orareru$"`, # Words ending with "-orareru" (e.g., o-orareru)

`r"\bw+-irassharu$"`, # Words ending with "-irassharu" (e.g., o-irassharu)

`r"\bw+-o-tomai suru$"`, # Words ending with "-o-tomai suru" (e.g., o-tomai shimasu)

]

Regular Sonkeigo patterns

```
sonkeigo_patterns = [  
    r"\b\w+-sama$", # Words ending with "-sama" (e.g., sensei-sama)  
    r"\b\w+-san$", # Words ending with "-san" (e.g., Tanaka-san)  
    r"\b\w+-sama$", # Words ending with "-sama" (e.g., o-kudasai)  
    r"\b\w+-rasareru$", # Words ending with "-rasareru" (e.g., o-kake-rasareru)  
    r"\b\w+-te itadaku$", # Words ending with "-te itadaku" (e.g., tabete itadaku)  
    r"\b\w+-ninaru$", # Words ending with "-ninaru" (e.g., o-kai ni naru)  
]
```

Regular Kenjougo patterns

```
kenjougo_patterns = [  
    r"\bo-\w+", # Words starting with "o-" (e.g., o-negai)  
    r"\bgo-\w+", # Words starting with "go-" (e.g., go-meiwaku)  
    r"\bmai-\w+", # Words starting with "mai-" (e.g., mai-rimasu)  
    r"\b\w+-suru$", # Words ending with "-suru" (e.g., tabemasu)  
    r"\b\w+-sasete itadaku$", # Words ending with "-sasete itadaku" (e.g., kakasete itadaku)  
    r"\b\w+-itasu$", # Words ending with "-itasu" (e.g., moi-itasu)  
    r"\b\w+-ageru$", # Words ending with "-ageru" (e.g., moi-ageru)  
    r"\b\w+-moshi-ageru$", # Words ending with "-moshi-ageru" (e.g., o-moshi-ageru)  
    r"\b\w+-zonjimasu$", # Words ending with "-zonjimasu" (e.g., zonjimasu)  
]
```

Irregular Teineigo patterns

```
irregular_teineigo_patterns = []
```

Irregular Sonkeigo patterns

```
irregular_sonkeigo_patterns = [  
    r"\b\w+-rareru$", # Irregular Sonkeigo form (e.g., o-kake-rareru)  
    r"\b\w+-iru$", # Irregular Sonkeigo form (e.g., o-ide-iru)  
    r"\b\w+-ninaru$", # Irregular Sonkeigo form (e.g., o-kai ni naru)  
    r"\b\w+-sha$", # Irregular Sonkeigo form for company names (e.g., gakusha)  
]
```

Irregular Kenjougo patterns

```
irregular_kenjougo_patterns = [  
    r"\b\w+-mairu$", # Irregular Kenjougo form (e.g., mairimasu)  
    r"\b\w+-ukagau$", # Irregular Kenjougo form (e.g., o-ukagai suru)  
    r"\b\w+-jimasu$", # Irregular Kenjougo form (e.g., zonjimasu)  
    r"\b\w+-ageru$", # Irregular Kenjougo form (e.g., moi-ageru)  
    r"\b\w+-sha$", # Irregular Kenjougo form for company names (e.g., heisha)  
    r"\b\w+-meshiagaru$", # Irregular Kenjougo form (e.g., o-meshiagari kudasai)  
    r"\b\w+-haidoku$", # Irregular Kenjougo form (e.g., o-haidoku suru)  
]
```