We built our CCLQA (EN-JA/JA-JA) system based on the technologies used in our past NTCIR systems.

- DEFINITION, BIOGRAPHY, and EVENT questions ⇒ we reused our definition module for QAC-4.
- RELATIONSHIP questions ⇒ we developed a new module based on our why-QA approach for QAC-4.
- Other questions ⇒ we used a simple sentence retriever based on BM25.
- English questions are translated into Japanese using translation dictionaries.
- Our EN-JA system performed rather poorly, but our JA-JA system showed reasonable performance.

### System Architecture

**Translation Module**

If the dictionary does not have the target, we use the left-to-right longest match.

Example: "Next Generation Network"

1. "Next Generation Network" is consulted
2. If the dictionary does not have this entry, "Next Generation" is consulted ⇒ We get "次世代".
3. The remaining "Network" is consulted ⇒ we get "ネットワーク".

As a result, we get "次世代ネットワーク".

**Definition Module**

Definition module extracts adnominal/adverbial modifiers for target X using TGREP2.

Example: “President Suharto (スハルト大統領)” ⇒ 32年間にわたって,人口約2億人のインドネシアを牛耳ってきたスハルト大統領 (President Suharto, who has been ruling Indonesia with the population of 200 million for 32 years)

**Relation Module**

We train a classifier that detects the mention of relationship from the EDR corpus using BACT (classifier of trees).

Positive Examples: sentences that have semantic categories corresponding to "関係 (relationship)"
Negative Examples: others

**Failure Analysis**

Failures are mainly caused by the fragility of our English question analyzer and dictionary-based translator.

1. Inappropriate dictionary look-up results “Martina Navratilova” ⇒ “マルティナナヴラティロワ” (no occurrence in the target corpus.)
2. No entry in E/J dictionaries “Suharto” ⇒ “スハルト” E-J transliteration module can save this case.
3. Transliteration fails in some cases "embryonic stem cells" ⇒ “ES 細胞”

### Performance of our JA-JA/EN-JA Systems

<table>
<thead>
<tr>
<th></th>
<th>DEF</th>
<th>BIO</th>
<th>REL</th>
<th>EVT</th>
<th>ALL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Our</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JA-JA</td>
<td>0.289</td>
<td>0.179</td>
<td>0.221</td>
<td>0.092</td>
<td>0.187</td>
</tr>
<tr>
<td><strong>Best</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JA-JA</td>
<td>0.420</td>
<td>0.190</td>
<td>0.233</td>
<td>0.094</td>
<td>0.220</td>
</tr>
<tr>
<td><strong>Our</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EN-JA</td>
<td>0.170</td>
<td>0.093</td>
<td>0.048</td>
<td>0.002</td>
<td>0.068</td>
</tr>
</tbody>
</table>

### Definition Module

Definition module extracts adnominal/adverbial modifiers for target X using TGREP2.

Example:

President Suharto (スハルト大統領)
32年間にわたって,人口約2億人のインドネシアを牛耳ってきたスハルト大統領 (President Suharto, who has been ruling Indonesia with the population of 200 million for 32 years)

### Relation Module

We train a classifier that detects the mention of relationship from the EDR corpus using BACT (classifier of trees).

Positive Examples: sentences that have semantic categories corresponding to "関係 (relationship)"
Negative Examples: others

**Scoring of an answer candidate (C):**

$$\text{score}(C) = \text{score}_{\text{rel}}(C) + \text{score}_{\text{sim}}(C)$$

BACT’s score indicating how likely relationship is expressed in C

Word-overlap-based similarity between the question and C