Overview

- Semantic Analysis of Image-based Learner Sentences (SAILS) Corpus
  - 13,533 picture description task (PDT) responses
  - Both native (NS) & non-native speakers (NNS)
  - Annotated for five binary features
- Goal: Evaluate content of NNS sentences
  - Compare to gold standard (GS) of NS sentences
- Need: Adequate data, appropriately constrained
  - Large set of PDT responses
  - Varied task prompts & participant demographics
  - Annotation for content analysis

Participants

- 490 total participants
- 141 NNSs: students in intermediate & advanced ESL writing courses at IU
- 358 NNSs: 29 Native English Speakers (NSs)
- 29 Relates or friends of researchers (assumedly higher quality)
- 329 Crowdsourced Native Speakers (CNSs)
- Responses purchased via SurveyMonkey (assumedly lower quality)

PDT instructions

- Focus on the main action
- Respond in a complete sentence
- Multiple versions
  - Most participants completed 30 items
  - Roughly equal number of targeted & untargeted responses
  - NNSs provide one response per item
  - NSs provide two non-identical responses per item (more robust GS)

Picture Description Task

Table 1: Example PDT images with their targeted questions.

<table>
<thead>
<tr>
<th>Intransitive</th>
<th>Transitive</th>
<th>Ditransitive</th>
</tr>
</thead>
<tbody>
<tr>
<td>What is the woman doing?</td>
<td>What is the man doing?</td>
<td>What is happening?</td>
</tr>
</tbody>
</table>

Administrated as online survey (SurveyMonkey.com)

Responses

Table 2: First & second response counts for SAILS Corpus participant groups

<table>
<thead>
<tr>
<th>Group</th>
<th>NS</th>
<th>NNS</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>First</td>
<td>4290</td>
<td>0</td>
<td>4290</td>
</tr>
<tr>
<td>Second</td>
<td>4609</td>
<td>921</td>
<td>5530</td>
</tr>
<tr>
<td>Total</td>
<td>8909</td>
<td>921</td>
<td>9830</td>
</tr>
</tbody>
</table>

Table 3: TTRs for complete responses (not words), for full corpus

<table>
<thead>
<tr>
<th>Type-Token Ratios (TTRs)</th>
<th>Targeted</th>
<th>Unattended</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intransitives</td>
<td>0.902</td>
<td>0.947</td>
</tr>
<tr>
<td>Transitives</td>
<td>0.915</td>
<td>0.917</td>
</tr>
<tr>
<td>Ditransitives</td>
<td>0.835</td>
<td>0.842</td>
</tr>
</tbody>
</table>

Table 4: TTRs for complete responses, separated by first (R1) & second responses (R2)

<table>
<thead>
<tr>
<th>Type-Token Ratios (TTRs)</th>
<th>Targeted</th>
<th>Unattended</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intransitives</td>
<td>R1: 0.902, R2: 0.866</td>
<td></td>
</tr>
<tr>
<td>Transitives</td>
<td>R1: 0.915, R2: 0.917</td>
<td></td>
</tr>
<tr>
<td>Ditransitives</td>
<td>R1: 0.835, R2: 0.842</td>
<td></td>
</tr>
</tbody>
</table>

Table 5: Sample responses from development transitive item, with adjudicated annotations

Annotation Scheme

- Core Event (C): Does response capture the core event depicted in image?
- Verifiability (V): Does response contain true & verifiable info, based on image?
- Answerhood (A): Does response make a clear attempt to answer the question?
- Interpretability (I): Does response evoke clear mental image (even if different from PDT)?
- Grammaticality (G): Is response free from errors of spelling & grammar?

Annotation Examples

- What is the boy doing? (Targeted)
- The boy is eating pizza.
- The kid is eating pizza.
- The boy is eating his favorite food.
- A youngster anticipates the taste of pizza.
- What is happening? (Untargeted)
- The boy is eating pizza.
- The kid is eating pizza.
- The boy is eating his favorite food.
- A young boy anticipates the taste of pizza.
- The boy is happy.

Inter-Annotator Agreement

- Cohen’s kappas well above conventional 0.67 threshold for meaningful agreement
- Annotation scheme can be implemented reliably by following guidelines

Observations from Table 6

- Average yes rates (Agv Yes) show all features skew toward yes annotations
- Cohen’s kappas needed as measure of inter-annotator agreement
- Agreement slightly higher for untargeted than targeted items
- Guidelines less complicated for untargeted items
- Matches annotator reporting of easiest & hardest features to annotate

Accessing the SAILS Corpus

Download the entire annotated SAILS Corpus, PDTs, & annotation guidelines at:

https://github.com/sailscorpus/sails

SAILS corpus can be used for:
- Language testing & ICALL
- Question answering, dialog systems, pragmatic modeling, visual references
- Possibilities for expansion from other researchers:
  - New participants, items, approaches for processing