Motivation

**Goal:** Probe grammatical annotation for learner language
- Support work on POS tagging & syntactic parsing
- Support investigation of features for SLA, NLI, proficiency classification

**Contributions:**
- Present scheme for dependencies, integrated with other annotation layers
- Test inter-annotator agreement for this scheme

Annotation Scheme

**Perspective:** annotate as closely as possible to learner production

1. **Lemma:** orthographic/phonetic variants
2. **Morphological POS:** form-based
3. **Distributional POS:** context-based
4. **Lexical violations:** when nothing else fits for syntactically ungrammatical forms

Dependencies

1. **Morphosyntactic dependencies:** what is realized?
2. **Subcategorization:** what is required?

Evaluation

- **Unlabeled Attachment Agreement (UAA)**
- **Labeled Attachment Agreement (LAA)**
- **Label Only Agreement (LOA)**

For dependencies & subcategorization, we use set-valued metrics

Results

**Inter-annotator Agreement Study**

- **Text selection**
  - Paragraph excerpts from 3 levels of proficiency
  - Text 1: 19 sentences (333 tokens)
  - Text 2: 22 sentences (271 tokens)

- **Annotators**
  - 3 undergraduates at IU, trained over one semester
  - Annotated & made detailed notes

- **Guidelines**
  - ~100 pages outlining general principles, layers, & examples

- **Task**
  - Phase 1: Text 1 annotated, Text 2 annotated
  - Phase 2: Individual modifications
  - Phase 3: Individual modifications, based on pairwise disagreements
  - Phase 4: Annotator meeting, followed by individual modifications

  - No feedback from researchers

<table>
<thead>
<tr>
<th>Annotators</th>
<th>POSm</th>
<th>POSd</th>
<th>Subcat.</th>
<th>UAA</th>
<th>LAA</th>
<th>LOA</th>
</tr>
</thead>
<tbody>
<tr>
<td>A, B</td>
<td>93.1</td>
<td>96.9</td>
<td>99.0</td>
<td>98.7</td>
<td>85.5</td>
<td>94.0</td>
</tr>
<tr>
<td>B, C</td>
<td>94.4</td>
<td>97.7</td>
<td>99.0</td>
<td>99.3</td>
<td>86.1</td>
<td>95.7</td>
</tr>
<tr>
<td>C, A</td>
<td>92.4</td>
<td>96.9</td>
<td>99.7</td>
<td>99.3</td>
<td>86.1</td>
<td>96.0</td>
</tr>
</tbody>
</table>

**Dependencies, Text 1**

<table>
<thead>
<tr>
<th>Ann.</th>
<th>UAA</th>
<th>LAA</th>
<th>LOA</th>
</tr>
</thead>
<tbody>
<tr>
<td>A, B</td>
<td>81.8</td>
<td>96.1</td>
<td>80.3</td>
</tr>
<tr>
<td>B, C</td>
<td>80.9</td>
<td>96.2</td>
<td>79.3</td>
</tr>
<tr>
<td>A, C</td>
<td>83.6</td>
<td>97.6</td>
<td>81.8</td>
</tr>
</tbody>
</table>

**Dependencies, Text 2**

<table>
<thead>
<tr>
<th>Ann.</th>
<th>UAA</th>
<th>LAA</th>
<th>LOA</th>
</tr>
</thead>
<tbody>
<tr>
<td>A, B</td>
<td>92.6</td>
<td>98.1</td>
<td>89.3</td>
</tr>
<tr>
<td>B, C</td>
<td>93.8</td>
<td>98.3</td>
<td>90.2</td>
</tr>
<tr>
<td>A, C</td>
<td>90.9</td>
<td>97.9</td>
<td>87.6</td>
</tr>
</tbody>
</table>

Qualitative notes

- **Syntactic complexity meets learner innovations:**
  - My most important goals are pursuing the profession to be a top marketing manager and then to earn a lot of money to buy a beautiful house and a good car.

- **Under-annotation:** subcategorization & verb raising

- **Annotation clarifications:** lexical violations

Bird's eye view

```
<table>
<thead>
<tr>
<th>Annotators</th>
<th>UAA</th>
<th>LAA</th>
<th>LOA</th>
</tr>
</thead>
<tbody>
<tr>
<td>A, B</td>
<td>94.4</td>
<td>97.7</td>
<td>99.0</td>
</tr>
<tr>
<td>B, C</td>
<td>93.8</td>
<td>97.9</td>
<td>90.2</td>
</tr>
<tr>
<td>A, C</td>
<td>91.9</td>
<td>97.9</td>
<td>87.6</td>
</tr>
</tbody>
</table>
```

**Evaluation**

- **Morphosyntactic tree**

BEA 2013; 13 June, 2013; Atlanta, GA

http://cl.indiana.edu/~salle/ (mragheb_md7)@indiana.edu