Motivations

Most discourse parsers are pipeline (rather than end-to-end), sophisticated, not self-contained:
- they assume gold segmentations (EDUs);
- they use external parsers for syntactic features.

Here we propose:

- **Syntacto-Discourse Treebank**: a combined representation of the constituency and discourse trees
  - facilitates parsing at both levels w/o explicit conversion
  - a joint treebank based on Penn Treebank and RST Treebank
- the first end-to-end discourse parser
  - jointly parses at constituency and discourse levels.
  - do not use any explicit syntactic features.
  - no need to do binarization.

Combined Represenation & Treebank

**RST Discourse Tree** (Fig. 1 (a))
- Elementary Discourse Units (EDUs) as leaf nodes
- mostly binary branching
  - nucleus (∗): core semantic meaning of the branching
  - satellite (∗): semantically decorating nucleus
  - relations: e.g., “Purpose,” “Background”
- multi-branching for conjunctions
  - e.g., “List,” “Comparison”

Combined Representation

- low-level lexical and syntactic info greatly help determining EDUs, structures, and relations.
  - previously from pre-trained tools
  - we directly determine the segmentations, syntactic, and discourse parses w/o a single joint parser.
  - trained on combined trees of constituency and discourse.

Step 1: Convert RST tree to constituency tree format

- binary branching: use relation + nucleus/satellite direction as label of the parent
- multi-branching: use the relation as the label

Step 2: Replace the leaf EDUs with syntactic (sub)trees

- in most cases, one EDU aligns to one single (sub)tree
- when one EDU corresponds to multiple (sub)trees, we take the lowest common ancestor as parent node

**PTB-RST Treebank**

The first Syntacto-Discourse treebank
- joining RST Treebank with PTB Treebank
- training set: 347 joint trees with ~17k tokens;
- lengths of the discourses vary from 30 to 2,199
- testing set: 38 trees w/ ~5k tokens; lengths of the discourses vary from 45 to 2,607

### End-to-End Training (F1 scores)

<table>
<thead>
<tr>
<th>Bacher et al. (2013)</th>
<th>PTB-RST-EDU</th>
<th>58.9</th>
<th>57.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hehman et al. (2014)</td>
<td>joint syntacto-discourse parsing</td>
<td>77.9</td>
<td>73.7</td>
</tr>
</tbody>
</table>

**Joint Syntacto-Discourse Parsing**

- linear-time parsing due to substantially longer input
- greedy parsing
- span-based parsing (Cross & Huang 2016)
  - stack maintains spans instead of any subtrees
  - no tree structure representations anymore
- alternate between structural (sh, comb) and label (label₁, nolabel) actions
- after structural actions, keep branching point k,
  - k will be used later in determining the relations b/w EDUs
  - k disappears after label action
- nolabel makes binarization of the discourse/constituency tree unnecessary