Joint Syntacto-Discourse Parsing and the Syntacto-Discourse Treebank

Kai Zhao† Liang Huang {kzhao.hf, liang.huang.sh}@gmail.com

School of EECS, Oregon State University †Kai is now with Google Inc., New York

Oregon State University

Motivations

Most discourse parsers are pipelined (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))

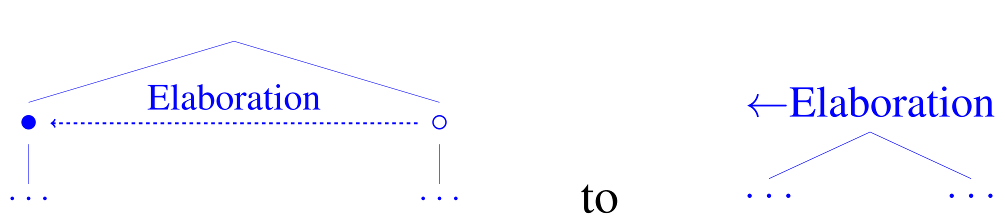
- Elementary Discourse Units (EDUs) as leaf nodes
- mostly binary branching
  - *nucleus* (•): core semantic meaning of the branching
  - *satellite* (○): semantically decorating nucleus
  - *relations*: e.g., "Purpose", "Background"
- multi-branching for conjunctions
  - e.g., "List", "Comparison"

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 trees, and discourse parses w/ 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

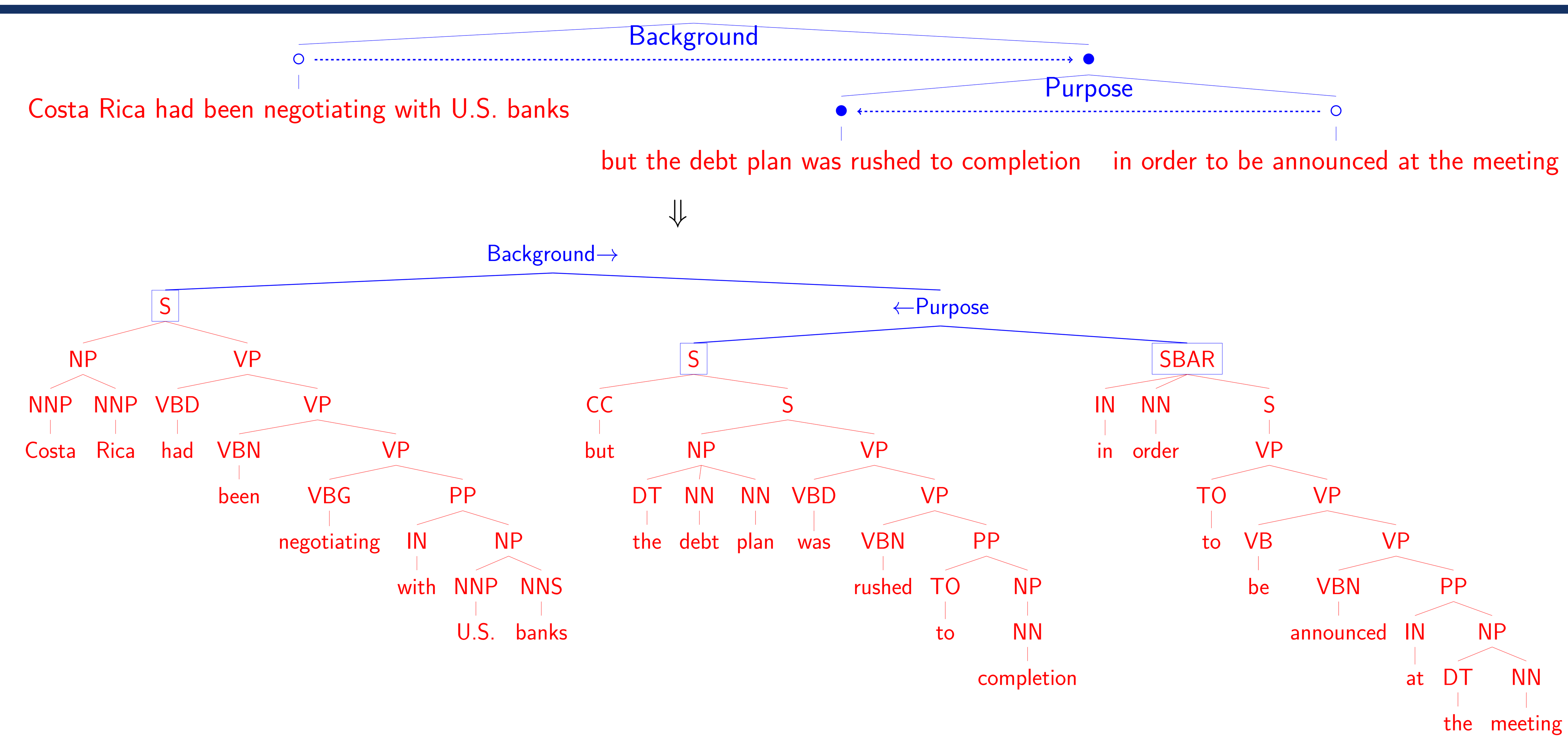


Figure 1: Examples of the RST discourse treebank and our syntacto-discourse treebank (PTB-RST).

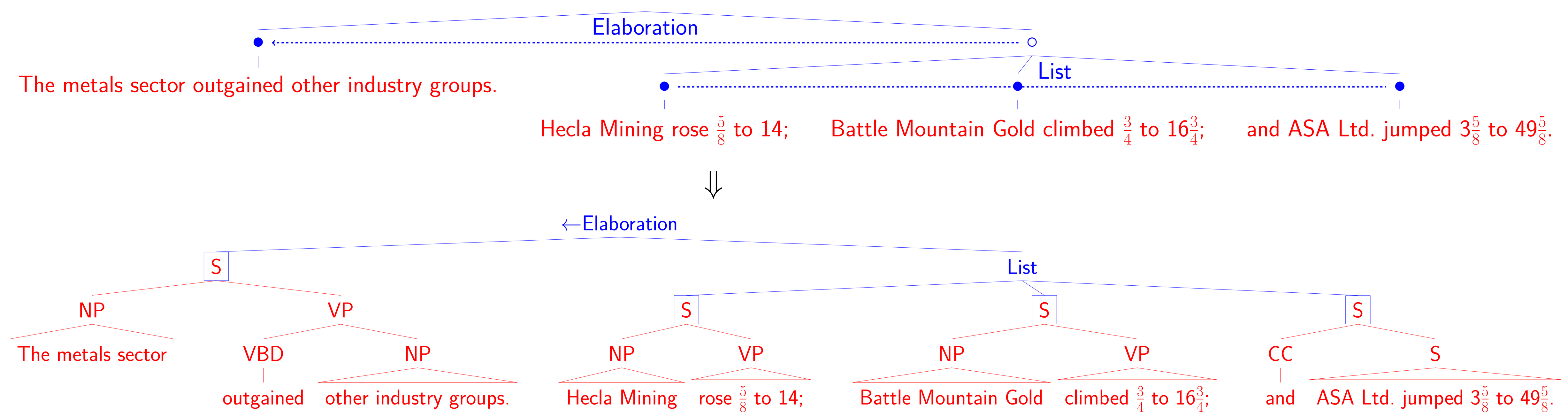
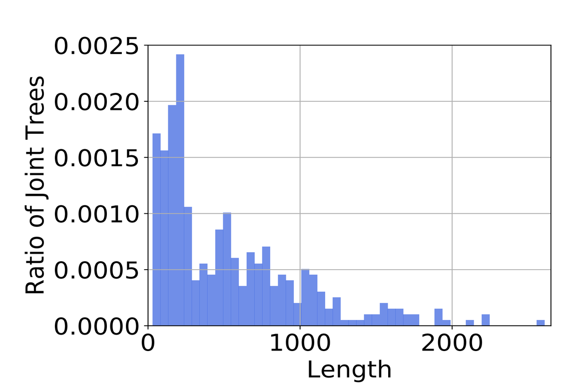


Figure 2: Another example of RST ⇒ PTB-RST, demonstrating a discourse tree over two sentences and a non-binary relation (List).

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 wutg ∼ 5k tokens; lengths of the discourses vary from 45 to 2,607



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 anywhere
- alternate between structural (sh, comb) and label (label$_X$, nolabel) actions
- after structural actions, keep branching point k,
  - k will be used later in determing the relations b/w EDUs
  - k disappears after label action
- nolabel makes binarization of the discourse/constituency tree unnecessary

Deductive System

input $w_0 \ldots w_{n-1}$

axiom $\langle {}_{-1}\triangle_0 \rangle : (0, \emptyset)$ goal $\langle {}_{-1}\triangle_0\triangle_n \rangle : (\_, t)$

sh $\dfrac{\langle \ldots {}_i\triangle_j \rangle : (c, t)}{\langle \ldots {}_i\triangle_j\triangle_{j+1} \rangle : (c + sc_{\mathsf{sh}}(i, j), t)}\ j < n$

comb $\dfrac{\langle \ldots {}_i\triangle_k\triangle_j \rangle : (c, t)}{\langle \ldots {}_i\triangle^k_j \rangle : (c + sc_{\mathsf{comb}}(i, k, j), t)}$

label$_X$ $\dfrac{\langle \ldots {}_i\triangle^k_j \rangle : (c, t)}{\langle \ldots {}_i\triangle_j \rangle : (c + sc_{\mathsf{label}_X}(i, k, j), t \cup \{{}_iX_j\})}$

nolabel $\dfrac{\langle \ldots {}_i\triangle^k_j \rangle : (c, t)}{\langle \ldots {}_i\triangle_j \rangle : (c + sc_{\mathsf{nolabel}}(i, k, j), t)}$

Recurrent Neural Models

- bi-directional LSTM in Cross & Huang (2016)
- no explicit discourse/syntactic tree structures represented in features
- span boundaries LSTM representations are passed to FF network to calc. likelihoods of actions/labels

Training & Emipirical Evaluation

Settings

- use "training with exploration" & **dynamic oracle**
- set most hyperparams based on Cross & Huang 2016
- use higher β ($= 0.8$) to discourage exploration
  - lower β leads to more diversions to wrong trajectories for larger discourse trees

End-to-End Comparison (F1 scores)

	description	synt. feats	seg.	struct.	+nuc.	+rel.
Bach+ '12	segment. only	Stanford	95.1	-	-	-
Hernault+ '10	end-to-end pipe.	PTB	94.0	72.3	59.1	47.3
joint syntacto-discourse parsing		-	**95.4**	**78.8**	**65.0**	**52.2**

Comparison w/ Gold Segmentation (F1 scores)

		syntactic feats	struct.	+nuc.	+rel.
	human annotation	-	88.7	77.7	65.8
sparse	Hernault et al. 2010	Penn Treebank	83.0	68.4	54.8
	Joty et al. 2013	Charniak (retrained)	82.7	68.4	55.7
	Joty + Moschitti 2014	Charniak (retrained)	-	-	57.3
	Feng & Hirst 2014	Stanford	85.7	71.0	58.2
	Heilman + Sagae 2015	ZPar (retraied)	83.5	68.1	55.1
	Wang et al. 2017	Stanford	**86.0**	**72.4**	59.7
neural	Li et al. 2014	Stanford	82.4	69.2	56.8
	+ sparse features		84.0	70.8	58.6
	Ji & Eisenstein 2014	MALT	80.5	68.6	58.3
	+ sparse features		81.6	71.1	**61.8**
	span-based disc. parsing	-	84.2	67.7	56.0