PCFG Approximation, Robust Meaning Composition, and Parser Evaluation with Elementary Dependency Matching

Yi Zhang, Rebecca Dridan, Stephan Oepen, Hans-Ulrich Krieger

LT-Lab, DFKI, Saarbrücken, Germany
Informatics, University of Oslo, Norway

Sofia, July 2012

Motivation

Rich grammar frameworks with powerful formalisms are convient for linguistic description, but hard to parse with.

Motivation

Rich grammar frameworks with powerful formalisms are convient for linguistic description, but hard to parse with.

- Linguistic Grammar → Parsing

Motivation

Rich grammar frameworks with powerful formalisms are convient for linguistic description, but hard to parse with.

- ~~Linguistic Grammar $\longrightarrow$ Parsing~~
- Linguistic Grammar
  $\longrightarrow$ Treebank Annotation
  $\longrightarrow$ Parsing Grammar
  $\longrightarrow$ Parsing

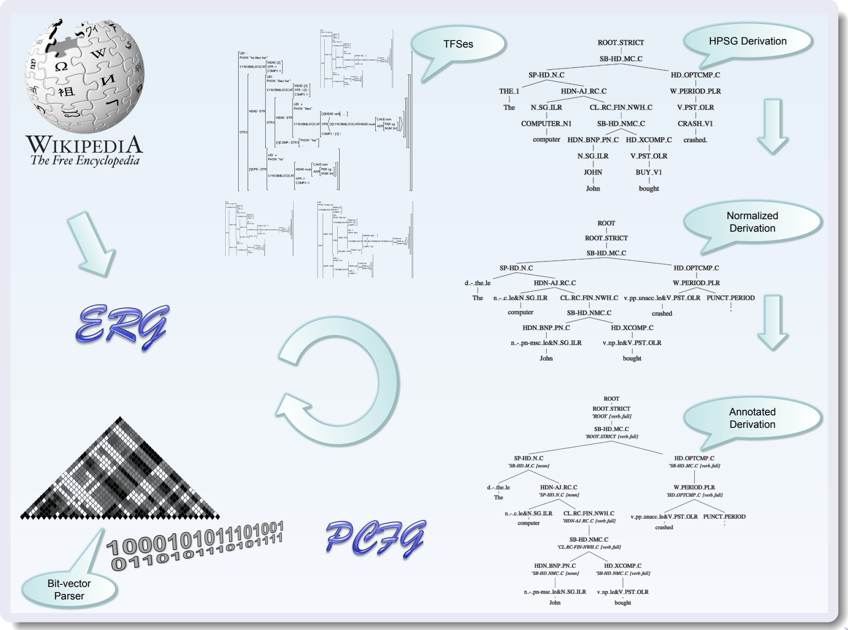
Problems with broad coverage HPSG parsing

- Robustness: hand-written pricision-oriented linguistic grammars typically miss 10∼20% raw coverage
- Specificity: the accuracy of the parsers based on such grmmars is further offset by the accuracy of the disambiguation model
- Efficiency: No polynomial time complexity upper-bound in unification-based parsing. Practically a parser can be never too fast

Corpus-Driven PCFG Approximation

- Approximate both the HPSG grammar and the disambiguation model with a single (generative) probabilistic CFG
- Only use one preferred reading per sentence (either the gold tree from manual disambiguation, or the top-ranked reading)

Workflow Overview

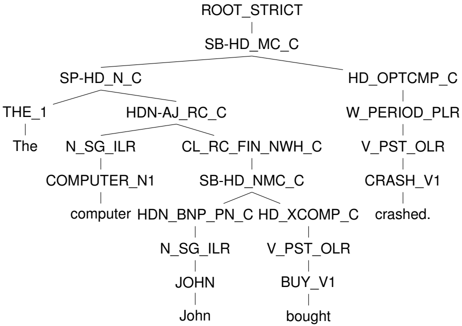


PCFG Extraction from Treebank

- A derivation tree records a complete analysis
- CFG categories and rules are extracted from annotation enriched derivation trees
- Probabilities Estimation: MLE
  - Constructions: no smoothing
  - Lexicon: $P(w|t)$ is smoothed to take care of unknown words. No context is used. Not sequence tagging

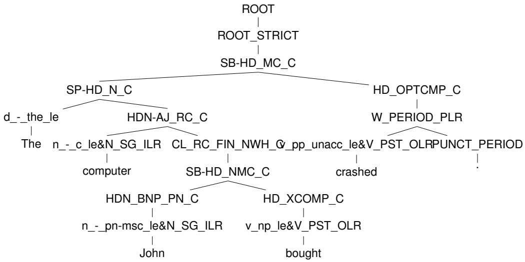
Annotated Derivations

1. Normalize derivation trees
2. External annotation with grandparent nodes
3. Internal annotation with feature-path values



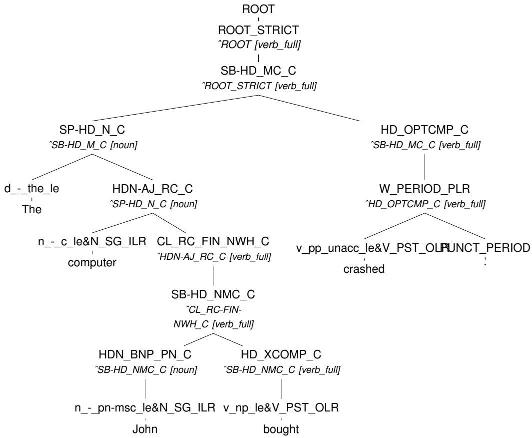
Annotated Derivations

1. Normalize derivation trees
2. External annotation with grandparent nodes
3. Internal annotation with feature-path values



Annotated Derivations

1. Normalize derivation trees
2. External annotation with grandparent nodes
3. Internal annotation with feature-path values



The Parser

Jigsaw

- A BitPar Java implementation, using bit-vector operations for efficient “parallel” recognition
- The best reading is recovered with Viterbi decoding
- Works with millions of CFG rules and hundred thousands of NTs without pruning
- Supports word lattice input
- Training PCFG on WSJ takes 10 seconds
- Reproduces almost exactly the same results as reported by [Klein and Manning, 2003] on WSJ parsing

Experiment

- ERG (1010), 200 rules, ~1000 leaf lexical types
- Train approximating PCFG with WeScience treebank (manually disambiguated)
- Train approximating PCFG with WikiWoods (auto-parsed corpus)
- ParsEval labeled bracketing F1, exact match rate and tagging accuracy are used as quality measures
- Gold tokenization is assumed for easy evaluation; The parser also works with chart-mapped input lattice (without supertags assigned)

Evaluation

		#Rule	#NT	#T	P	R	F_1	EX	TA
ws	PET	-	-	-	87.1	87.1	**87.1**	48.79	96.5
	PCFG(0)	10,251	208	1,152	64.8	59.1	61.8	18.09	83.3
	PCFG(FP1)	12,178	669	1,152	71.7	63.3	67.2	18.73	82.4
ww000	PCFG(0)	25,859	236	1,799	61.3	58.2	59.7	13.76	83.9
	PCFG(GP1)	64,043	3,983	1,799	73.5	70.7	**72.1**	20.25	85.9
	PCFG-LA(SM3)	*	*	*	74.4	69.4	71.8	1.91	88.0
ww00	PCFG(0)	61,426	247	2,546	64.5	62.1	63.3	16.56	87.8
	PCFG(GP1)	187,852	5,828	2,546	78.5	77.9	78.2	25.35	91.5
	PCFG(GP1,FP4)	271,956	16,731	2,546	81.6	80.7	81.2	29.04	92.2
	PCFG(GP1,FP5)	319,511	21,414	2,546	82.0	81.2	**81.6**	28.54	92.4
	PCFG(GP1,FP6)	320,630	21,694	2,546	81.9	81.1	81.5	28.41	92.4
	PCFG(GP2)	489,890	45,658	2,546	80.2	79.8	80.0	28.92	91.6
	PCFG(GP2,FP2)	559,006	66,218	2,546	81.1	80.3	80.7	32.10	91.8
ww	PCFG(GP1)	1,007,563	8,852	4,472	81.3	80.6	80.9	29.43	92.5
	PCFG(GP2)	3,952,821	128,822	4,472	85.0	84.8	**84.9**	37.45	93.6

Evaluation

			#Rule	#NT	#T	P	R	F_1	EX	TA
7636	ws	PET	-	-	-	87.1	87.1	**87.1**	48.79	96.5
		PCFG(0)	10,251	208	1,152	64.8	59.1	61.8	18.09	83.3
		PCFG(FP1)	12,178	669	1,152	71.7	63.3	67.2	18.73	82.4
86K	ww000	PCFG(0)	25,859	236	1,799	61.3	58.2	59.7	13.76	83.9
		PCFG(GP1)	64,043	3,983	1,799	73.5	70.7	**72.1**	20.25	85.9
		PCFG-LA(SM3)	*	*	*	74.4	69.4	71.8	1.91	88.0
482K	ww00	PCFG(0)	61,426	247	2,546	64.5	62.1	63.3	16.56	87.8
		PCFG(GP1)	187,852	5,828	2,546	78.5	77.9	78.2	25.35	91.5
		PCFG(GP1,FP4)	271,956	16,731	2,546	81.6	80.7	81.2	29.04	92.2
		PCFG(GP1,FP5)	319,511	21,414	2,546	82.0	81.2	**81.6**	28.54	92.4
		PCFG(GP1,FP6)	320,630	21,694	2,546	81.9	81.1	81.5	28.41	92.4
		PCFG(GP2)	489,890	45,658	2,546	80.2	79.8	80.0	28.92	91.6
		PCFG(GP2,FP2)	559,006	66,218	2,546	81.1	80.3	80.7	32.10	91.8
48M	ww	PCFG(GP1)	1,007,563	8,852	4,472	81.3	80.6	80.9	29.43	92.5
		PCFG(GP2)	3,952,821	128,822	4,472	85.0	84.8	**84.9**	37.45	93.6

Evaluation

Grammar Size — ParsEval

			#Rule	#NT	#T	P	R	F_1	EX	TA
7636	ws	PET	-	-	-	87.1	87.1	**87.1**	48.79	96.5
		PCFG(0)	10,251	208	1,152	64.8	59.1	61.8	18.09	83.3
		PCFG(FP1)	12,178	669	1,152	71.7	63.3	67.2	18.73	82.4
86K	ww000	PCFG(0)	25,859	236	1,799	61.3	58.2	59.7	13.76	83.9
		PCFG(GP1)	64,043	3,983	1,799	73.5	70.7	**72.1**	20.25	85.9
		PCFG-LA(SM3)	*	*	*	74.4	69.4	71.8	1.91	88.0
482K	ww00	PCFG(0)	61,426	247	2,546	64.5	62.1	63.3	16.56	87.8
		PCFG(GP1)	187,852	5,828	2,546	78.5	77.9	78.2	25.35	91.5
		PCFG(GP1,FP4)	271,956	16,731	2,546	81.6	80.7	81.2	29.04	92.2
		PCFG(GP1,FP5)	319,511	21,414	2,546	82.0	81.2	**81.6**	28.54	92.4
		PCFG(GP1,FP6)	320,630	21,694	2,546	81.9	81.1	81.5	28.41	92.4
		PCFG(GP2)	489,890	45,658	2,546	80.2	79.8	80.0	28.92	91.6
		PCFG(GP2,FP2)	559,006	66,218	2,546	81.1	80.3	80.7	32.10	91.8
48M	ww	PCFG(GP1)	1,007,563	8,852	4,472	81.3	80.6	80.9	29.43	92.5
		PCFG(GP2)	3,952,821	128,822	4,472	85.0	84.8	**84.9**	37.45	93.6

Evaluation

			Grammar Size			ParsEval				
			#Rule	#NT	#T	P	R	F_1	EX	TA
7636	ws	PET	-	-	-	87.1	87.1	**87.1**	48.79	96.5
		PCFG(0)	10,251	208	1,152	64.8	59.1	61.8	18.09	83.3
		PCFG(FP1)	12,178	669	1,152	71.7	63.3	67.2	18.73	82.4
86K	ww000	PCFG(0)	25,859	236	1,799	61.3	58.2	59.7	13.76	83.9
		PCFG(GP1)	64,043	3,983	1,799	73.5	70.7	**72.1**	20.25	85.9
		PCFG-LA(SM3)	*	*	*	74.4	69.4	71.8	1.91	88.0
482K	ww00	PCFG(0)	61,426	247	2,546	64.5	62.1	63.3	16.56	87.8
		PCFG(GP1)	187,852	5,828	2,546	78.5	77.9	78.2	25.35	91.5
		PCFG(GP1,FP4)	271,956	16,731	2,546	81.6	80.7	81.2	29.04	92.2
		PCFG(GP1,FP5)	319,511	21,414	2,546	82.0	81.2	**81.6**	28.54	92.4
		PCFG(GP1,FP6)	320,630	21,694	2,546	81.9	81.1	81.5	28.41	92.4
		PCFG(GP2)	489,890	45,658	2,546	80.2	79.8	80.0	28.92	91.6
		PCFG(GP2,FP2)	559,006	66,218	2,546	81.1	80.3	80.7	32.10	91.8
48M	ww	PCFG(GP1)	1,007,563	8,852	4,472	81.3	80.6	80.9	29.43	92.5
		PCFG(GP2)	3,952,821	128,822	4,472	85.0	84.8	**84.9**	37.45	93.6

MEM: MaxEnt parse selection accuracy given top-500 candidates

Evaluation

Grammar Size | ParsEval

			#Rule	#NT	#T	P	R	F$_1$	EX	TA
7636	ws	PET	-	-	-	87.1	87.1	**87.1**	48.79	96.5
		PCFG(0)	10,251	208	1,152	64.8	59.1	61.8	18.09	83.3
		PCFG(FP1)	12,178	669	1,152	71.7	63.3	67.2	18.73	82.4
86K	ww000	PCFG(0)	25,859	236	1,799	61.3	58.2	59.7	13.76	83.9
		PCFG(GP1)	64,043	3,983	1,799	73.5	70.7	**72.1**	20.25	85.9
		PCFG-LA(SM3)	*	*	*	74.4	69.4	71.8	1.91	88.0
482K	ww00	PCFG(0)	61,426	247	2,546	64.5	62.1	63.3	16.56	87.8
		PCFG(GP1)	187,852	5,828	2,546	78.5	77.9	78.2	25.35	91.5
		PCFG(GP1,FP4)	271,956	16,731	2,546	81.6	80.7	81.2	29.04	92.2
		PCFG(GP1,FP5)	319,511	21,414	2,546	82.0	81.2	**81.6**	28.54	92.4
		PCFG(GP1,FP6)	320,630	21,694	2,546	81.9	81.1	81.5	28.41	92.4
		PCFG(GP2)	489,890	45,658	2,546	80.2	79.8	80.0	28.92	91.6
		PCFG(GP2,FP2)	559,006	66,218	2,546	81.1	80.3	80.7	32.10	91.8
48M	ww	PCFG(GP1)	1,007,563	8,852	4,472	81.3	80.6	80.9	29.43	92.5
		PCFG(GP2)	3,952,821	128,822	4,472	85.0	84.8	**84.9**	37.45	93.6

PCFG(0): raw PCFG from normalized derivation trees

Evaluation

Grammar Size — ParsEval

Size	Data	Model	#Rule	#NT	#T	P	R	F1	EX	TA
7636	ws	PET	-	-	-	87.1	87.1	**87.1**	48.79	96.5
		PCFG(0)	10,251	208	1,152	64.8	59.1	61.8	18.09	83.3
		PCFG(FP1)	12,178	669	1,152	71.7	63.3	67.2	18.73	82.4
86K	ww000	PCFG(0)	25,859	236	1,799	61.3	58.2	59.7	13.76	83.9
		PCFG(GP1)	64,043	3,983	1,799	73.5	70.7	**72.1**	20.25	85.9
		PCFG-LA(SM3)	*	*	*	74.4	69.4	71.8	1.91	88.0
482K	ww00	PCFG(0)	61,426	247	2,546	64.5	62.1	63.3	16.56	87.8
		PCFG(GP1)	187,852	5,828	2,546	78.5	77.9	78.2	25.35	91.5
		PCFG(GP1,FP4)	271,956	16,731	2,546	81.6	80.7	81.2	29.04	92.2
		PCFG(GP1,FP5)	319,511	21,414	2,546	82.0	81.2	**81.6**	28.54	92.4
		PCFG(GP1,FP6)	320,630	21,694	2,546	81.9	81.1	81.5	28.41	92.4
		PCFG(GP2)	489,890	45,658	2,546	80.2	79.8	80.0	28.92	91.6
		PCFG(GP2,FP2)	559,006	66,218	2,546	81.1	80.3	80.7	32.10	91.8
48M	ww	PCFG(GP1)	1,007,563	8,852	4,472	81.3	80.6	80.9	29.43	92.5
		PCFG(GP2)	3,952,821	128,822	4,472	85.0	84.8	**84.9**	37.45	93.6

PCFG(FP1): PCFG with 1 feature-path annotation (HEAD)

Evaluation

Grammar Size | ParsEval

Size			#Rule	#NT	#T	P	R	F_1	EX	TA
7636	ws	PET	-	-	-	87.1	87.1	**87.1**	48.79	96.5
		PCFG(0)	10,251	208	1,152	64.8	59.1	61.8	18.09	83.3
		PCFG(FP1)	12,178	669	1,152	71.7	63.3	67.2	18.73	82.4
86K	ww000	PCFG(0)	25,859	236	1,799	61.3	58.2	59.7	13.76	83.9
		PCFG(GP1)	64,043	3,983	1,799	73.5	70.7	**72.1**	20.25	85.9
		PCFG-LA(SM3)	*	*	*	74.4	69.4	71.8	1.91	88.0
482K	ww00	PCFG(0)	61,426	247	2,546	64.5	62.1	63.3	16.56	87.8
		PCFG(GP1)	187,852	5,828	2,546	78.5	77.9	78.2	25.35	91.5
		PCFG(GP1,FP4)	271,956	16,731	2,546	81.6	80.7	81.2	29.04	92.2
		PCFG(GP1,FP5)	319,511	21,414	2,546	82.0	81.2	**81.6**	28.54	92.4
		PCFG(GP1,FP6)	320,630	21,694	2,546	81.9	81.1	81.5	28.41	92.4
		PCFG(GP2)	489,890	45,658	2,546	80.2	79.8	80.0	28.92	91.6
		PCFG(GP2,FP2)	559,006	66,218	2,546	81.1	80.3	80.7	32.10	91.8
48M	ww	PCFG(GP1)	1,007,563	8,852	4,472	81.3	80.6	80.9	29.43	92.5
		PCFG(GP2)	3,952,821	128,822	4,472	85.0	84.8	**84.9**	37.45	93.6

PCFG(GP1): PCFG with 1 level grand-parent annotation

Evaluation

Size	Data	Model	#Rule	#NT	#T	P	R	F_1	EX	TA
7636	ws	PET	-	-	-	87.1	87.1	**87.1**	48.79	96.5
	ws	PCFG(0)	10,251	208	1,152	64.8	59.1	61.8	18.09	83.3
	ws	PCFG(FP1)	12,178	669	1,152	71.7	63.3	67.2	18.73	82.4
86K	ww000	PCFG(0)	25,859	236	1,799	61.3	58.2	59.7	13.76	83.9
	ww000	PCFG(GP1)	64,043	3,983	1,799	73.5	70.7	**72.1**	20.25	85.9
	ww000	PCFG-LA(SM3)	*	*	*	74.4	69.4	71.8	1.91	88.0
482K	ww00	PCFG(0)	61,426	247	2,546	64.5	62.1	63.3	16.56	87.8
	ww00	PCFG(GP1)	187,852	5,828	2,546	78.5	77.9	78.2	25.35	91.5
	ww00	PCFG(GP1,FP4)	271,956	16,731	2,546	81.6	80.7	81.2	29.04	92.2
	ww00	PCFG(GP1,FP5)	319,511	21,414	2,546	82.0	81.2	**81.6**	28.54	92.4
	ww00	PCFG(GP1,FP6)	320,630	21,694	2,546	81.9	81.1	81.5	28.41	92.4
	ww00	PCFG(GP2)	489,890	45,658	2,546	80.2	79.8	80.0	28.92	91.6
	ww00	PCFG(GP2,FP2)	559,006	66,218	2,546	81.1	80.3	80.7	32.10	91.8
48M	ww	PCFG(GP1)	1,007,563	8,852	4,472	81.3	80.6	80.9	29.43	92.5
	ww	PCFG(GP2)	3,952,821	128,822	4,472	85.0	84.8	**84.9**	37.45	93.6

Grammar Size: #Rule, #NT, #T — ParsEval: P, R, F_1, EX, TA

PCFG-LA(SM3): 3 iteration split-merge latent variable PCFG (Berkeley)

Evaluation

			Grammar Size			ParsEval				
			#Rule	#NT	#T	P	R	F_1	EX	TA
7636	ws	PET	-	-	-	87.1	87.1	**87.1**	48.79	96.5
		PCFG(0)	10,251	208	1,152	64.8	59.1	61.8	18.09	83.3
		PCFG(FP1)	12,178	669	1,152	71.7	63.3	67.2	18.73	82.4
86K	ww000	PCFG(0)	25,859	236	1,799	61.3	58.2	59.7	13.76	83.9
		PCFG(GP1)	64,043	3,983	1,799	73.5	70.7	**72.1**	20.25	85.9
		PCFG-LA(SM3)	*	*	*	74.4	69.4	71.8	1.91	88.0
482K	ww00	PCFG(0)	61,426	247	2,546	64.5	62.1	63.3	16.56	87.8
		PCFG(GP1)	187,852	5,828	2,546	78.5	77.9	78.2	25.35	91.5
		PCFG(GP1,FP4)	271,956	16,731	2,546	81.6	80.7	81.2	29.04	92.2
		PCFG(GP1,FP5)	319,511	21,414	2,546	82.0	81.2	**81.6**	28.54	92.4
		PCFG(GP1,FP6)	320,630	21,694	2,546	81.9	81.1	81.5	28.41	92.4
		PCFG(GP2)	489,890	45,658	2,546	80.2	79.8	80.0	28.92	91.6
		PCFG(GP2,FP2)	559,006	66,218	2,546	81.1	80.3	80.7	32.10	91.8
48M	ww	PCFG(GP1)	1,007,563	8,852	4,472	81.3	80.6	80.9	29.43	92.5
		PCFG(GP2)	3,952,821	128,822	4,472	85.0	84.8	**84.9**	37.45	93.6

PCFG(GP2): PCFG with 2 levels of grand-parent annotation

Evaluation

			Grammar Size			ParsEval				
			#Rule	#NT	#T	P	R	F_1	EX	TA
7636	ws	PET	-	-	-	87.1	87.1	**87.1**	48.79	96.5
	ws	PCFG(0)	10,251	208	1,152	64.8	59.1	61.8	18.09	83.3
	ws	PCFG(FP1)	12,178	669	1,152	71.7	63.3	67.2	18.73	82.4
86K	ww000	PCFG(0)	25,859	236	1,799	61.3	58.2	59.7	13.76	83.9
	ww000	PCFG(GP1)	64,043	3,983	1,799	73.5	70.7	**72.1**	20.25	85.9
	ww000	PCFG-LA(SM3)	*	*	*	74.4	69.4	71.8	1.91	88.0
482K	ww00	PCFG(0)	61,426	247	2,546	64.5	62.1	63.3	16.56	87.8
	ww00	PCFG(GP1)	187,852	5,828	2,546	78.5	77.9	78.2	25.35	91.5
	ww00	PCFG(GP1,FP4)	271,956	16,731	2,546	81.6	80.7	81.2	29.04	92.2
	ww00	PCFG(GP1,FP5)	319,511	21,414	2,546	82.0	81.2	**81.6**	28.54	92.4
	ww00	PCFG(GP1,FP6)	320,630	21,694	2,546	81.9	81.1	81.5	28.41	92.4
	ww00	PCFG(GP2)	489,890	45,658	2,546	80.2	79.8	80.0	28.92	91.6
	ww00	PCFG(GP2,FP2)	559,006	66,218	2,546	81.1	80.3	80.7	32.10	91.8
48M	ww	PCFG(GP1)	1,007,563	8,852	4,472	81.3	80.6	80.9	29.43	92.5
	ww	PCFG(GP2)	3,952,821	128,822	4,472	85.0	84.8	**84.9**	37.45	93.6

PCFG(GPx,FPy): PCFG with x GP and x FP annotation

Robust Semantic Composition

- ~45% of PCFG derivations are not fully consistent;
- need unification for semantic composition;
- conjecture: few(er) unification failures in semantics;

→ failure-tolerant *robust* unification.

A Few Examples

Strategies for Robust Unification

- In case of conflict, need to discard *some* information;
- robustness triggered by glb failure on two (sub-)nodes;
- *generalization* find most specific subsuming type;
- *default unification* treat one node as 'more important';

→ heuristically pick 'richer' node: number of sub-nodes;

- no notable quality difference for various heuristics (so far);
- for new glb type, only recurse over appropriate features;

→ simple, 'greedy' default unification; all decisions local.

How to evaluate relative quality
of (robust) MRSs?

Information Classes

Can be useful to split the information up, so relevant elements can be evaluated as appropriate.

We propose three classes of information relating to meaning:

CLASS 1	core functor – argument structure, whether syntactic or semantic
CLASS 2	functor-related information, such as the lemma, word category, and sense
CLASS 3	(morpho-syntactic) properties of events and entities, e.g. tense, number, gender

Common Dependency-based Representations

- Stanford Dependencies: CLASS 1
- Grammatical Relations: CLASS 1
- CCG Dependencies: CLASS 1 and CLASS 2 combined
- PARC700: CLASS 1 and CLASS 2 combined, CLASS 3 separately

All four dependency styles use grammatical relations such as SUBJ and OBJ to represent CLASS 1 information.

EDS to EDM triples

Start from reduction into variable-free Elementary Dependency Structures (EDS) [Oepen and Lønning, 2006]

Debt burdens are heavier.

```
{
 e3:
 _1:udef_q⟨0:4⟩[BV x9]
 x9:_debt_n_1⟨0:4⟩[]
 _2:udef_q⟨0:12⟩[BV x6]
 x6:_burden_n_1⟨5:12⟩[]
 e10:compound⟨0:12⟩[ARG1 x6, ARG2 x9]
 e16:comp⟨17:25⟩[ARG1 e3]
 e3:_heavy_a_1⟨17:25⟩[ARG1 x6]
}
```

Construct triples in terms of spans rather than predicate names, to separate CLASS 1 from CLASS 2 .

EDM triples

Debt burdens are heavier.

NAME			ARGUMENT		
⟨0:4⟩	PRED	udef_q	⋆	ROOT	⟨17:26⟩
⟨0:4⟩	PRED	_debt_n_1	⟨0:4⟩	BV	⟨0:4⟩
⟨5:12⟩	PRED	_burden_n_1	⟨0:12⟩	BV	⟨5:12⟩
⟨0:12⟩	PRED	udef_q	⟨0:12⟩	ARG1	⟨5:12⟩
⟨0:12⟩	PRED	compound	⟨0:12⟩	ARG2	⟨0:4⟩
⟨17:26⟩	PRED	_heavy_a_1	⟨17:26⟩	ARG1	⟨5:12⟩
⟨17:26⟩	PRED	comp	⟨17:26⟩	ARG1	⟨17:26⟩

PROPERTY

⟨5:12⟩	NUM	*pl*	⟨0:12⟩	SF	*prop*
⟨0:12⟩	TENSE	*untensed*	⟨17:26⟩	SF	*prop*
⟨17:26⟩	SF	*prop*	⟨17:26⟩	TENSE	*pres*

Elementary Dependency Metric

- Precision, recall and F-score over all triples (EDM)
- Can also be calculated over only those classes that are appropriate

For our robust evaluation, we have no predicates, so we evaluate over CLASS 1 and CLASS 3 information.

EDM-based Parser Evaluation

		FSC	EDM$_A$			EDM$_P$		
			P	R	F$_1$	P	R	F$_1$
ws	PET	**100%**	86.6	86.4	**86.5**	94.1	93.9	**94.0**
	PCFG(0)	22.9%	67.5	54.5	60.3	84.1	76.6	80.2
	PCFG(FP1)	25.9%	73.2	61.5	66.8	86.8	75.8	80.9
ww000	PCFG(0)	17.0%	63.9	51.1	56.8	81.6	75.4	78.4
	PCFG(GP1)	**31.5%**	73.9	65.6	69.5	87.1	79.0	82.8
	PCFG-LA(SM3)	19.6%	76.0	70.6	**73.2**	87.7	82.9	**85.2**
ww00	PCFG(0)	19.1%	67.8	55.7	61.2	84.1	79.0	81.5
	PCFG(GP1)	37.8%	79.5	74.2	76.8	90.3	85.8	88.0
	PCFG(GP1,FP4)	44.5%	81.5	79.9	80.7	91.1	89.8	90.4
	PCFG(GP1,FP5)	45.4%	81.6	80.0	**80.8**	91.2	90.0	**90.6**
	PCFG(GP1,FP6)	45.4%	81.6	80.0	80.8	91.1	89.9	90.5
	PCFG(GP2)	46.0%	81.2	75.9	78.5	90.9	86.4	88.6
	PCFG(GP2,FP2)	**51.2%**	81.5	79.1	80.3	91.3	89.3	90.3
ww	PCFG(GP1)	41.2%	80.7	79.2	79.9	91.0	90.5	90.8
	PCFG(GP2)	**55.4%**	84.6	83.8	**84.2**	92.9	92.6	**92.8**

A Few Observations

- All evaluated PCFGs are very robust, parsing over 99% of the test set, and over 97% on the complete corpus
- PCFG(GP3) is too sparse on ww−00, but too large to parse with on ww
- Approximating PCFG continues to grow after 50M trees, at the rate of 1 rule per 160 sentences for PCFG(GP1)

aj-hdn_norm_c↑sp-hd_n_c → v_np-pp*_to_le&v_pas_odlr&v_v-un_dlr&v_j-nb-pas-tr_dlr@ n_-_c-pl-nocnh_le&n_pl_olr@

- While no pruning is done currently, we suspect many rules could be discarded without a noticeable impact on parsing accuracy or coverage.

Conclusion

- Annotated approximating PCFG achieved high parsing accuracy (semantics still need to be checked)
- Combination of *internal* and *external* annotation achieved the best performance
- MLE works well on huge corpus
- Comparison with PCFG-LA parser shows that our MLE PCFG trained with 50M trees is performing better

References I

Klein, D. and Manning, C. D. (2003).
Accurate unlexicalized parsing.
In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pages 423–430, Sapporo, Japan.

Oepen, S. and Lønning, J. T. (2006).
Discriminant-based MRS banking.
In *Proceedings of the 5th International Conference on Language Resources and Evaluation*, pages 1250 – 1255, Genoa, Italy.