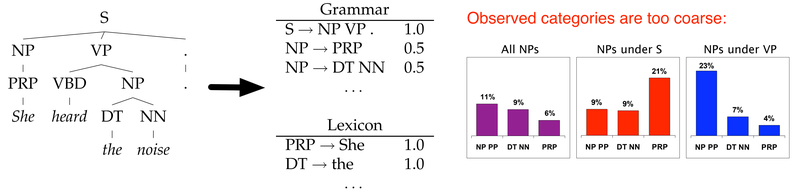
Learning and Inference for Hierarchically Split Probabilistic Context-Free Grammars

Slav Petrov and Dan Klein
University of California, Berkeley

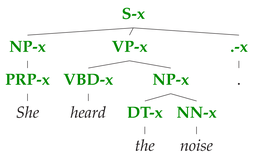
Learning

Grammar Extraction:



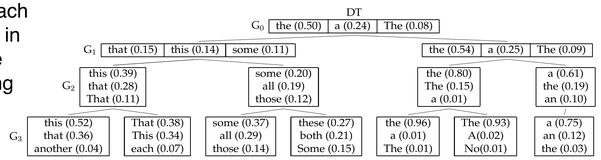
Adaptive Grammar Refinement:

Split each category in k subcategories and fit grammar with the EM algorithm.



Hierarchical Splitting:

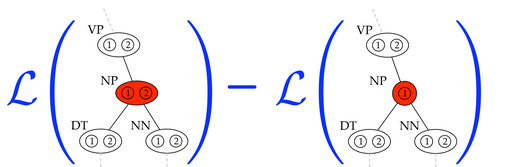
Repeatedly split each annotation symbol in two and retrain the grammar, initializing with the previous grammar.



Evolution of the DT tag during hierarchical splitting and merging.
Shown are the top three words for each subcategory and their respective probability.

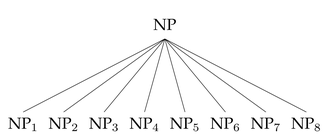
Merging:

Roll back the least useful splits in order to allocate complexity only where needed.



Smoothing:

Reduce overfitting by shrinking the productions of each subcategory towards their common base category.

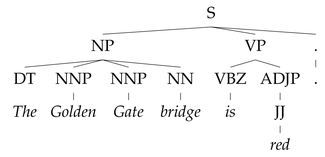


Reference:
Slav Petrov, Leon Barrett, Romain Thibaux and Dan Klein,
"Learning accurate, compact and interpretable tree annotation", in **ACL-COLING '06**

Inference

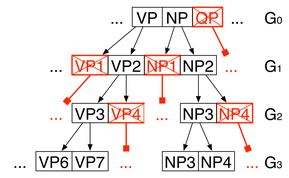
Parsing:

The Golden Gate bridge is red.

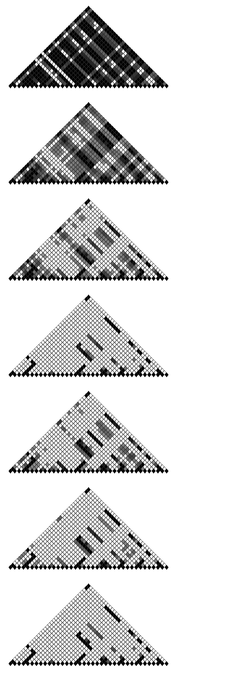


Parse Efficiency:

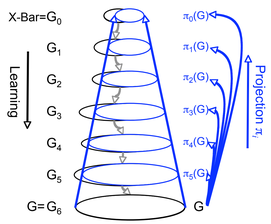
Rapidly pre-parse the sentence in a hierarchical coarse-to-fine fashion pruning away unlikely chart items.



Bracket posterior probabilities during coarse-to-fine decoding

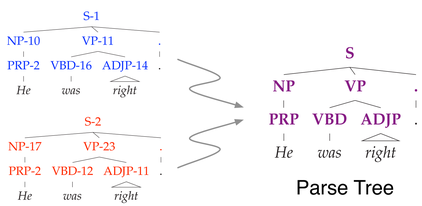


Compute grammars of intermediate complexity by projecting the most refined grammar.



Parse Selection:

Use a variational approximation to select the tree with the maximum number of expected correct rules (since computing the best parse tree is intractable and selecting the best derivation is a poor approximation).



Parse Derivations

Parse Tree

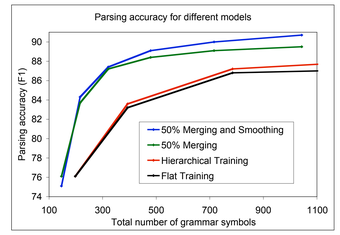
Reference:
Slav Petrov and Dan Klein,
"Improved Inference for Unlexicalized Parsing", in **NAACL-HLT '07**

Results

General technique for learning refined, structured models when only the trace of a complex underlying process is observed.
Learns compact and accurate grammars from a treebank without additional human input.
Gives best known parsing accuracy on a variety of languages, while being extremely efficient.

Interactive demo and download at http://nlp.cs.berkeley.edu

Parser	≤ 40 words F_1	all F_1
ENGLISH		
Charniak & Johnson '05	90.1	89.6
This Work	**90.6**	**90.1**
GERMAN		
Dubey '05	76.3	-
This Work	**80.8**	**80.1**
CHINESE		
Chiang & Bikel '02	80.0	76.6
This Work	**86.3**	**83.4**



Learned grammars are compact and interpretable:

VBZ			
VBZ-0	gives	sells	takes
VBZ-1	comes	goes	works
VBZ-2	includes	owns	is
VBZ-3	puts	provides	takes
VBZ-4	says	adds	says
VBZ-5	believes	means	thinks
VBZ-6	expects	makes	calls
VBZ-7	plans	expects	wants
VBZ-8	is	's	gets
VBZ-9	's	is	remains
VBZ-10	has	's	is
VBZ-11	does	Is	Does

NNP			
NNP-0	Jr.	Goldman	INC.
NNP-1	Bush	Noriega	Peters
NNP-2	J.	E.	L.
NNP-3	York	Francisco	Street
NNP-4	Inc	Exchange	Co
NNP-5	Inc.	Corp.	Co.
NNP-6	Stock	Exchange	York
NNP-7	Corp.	Inc.	Group
NNP-8	Congress	Japan	IBM
NNP-9	Friday	September	August
NNP-10	Shearson	D.	Ford
NNP-11	U.S.	Treasury	Senate
NNP-12	John	Robert	James
NNP-13	Mr.	Ms.	President
NNP-14	Oct.	Nov.	Sept.
NNP-15	New	San	Wall

JJS			
JJS-0	largest	latest	biggest
JJS-1	least	best	worst
JJS-2	most	Most	least

DT			
DT-0	the	The	a
DT-1	A	An	Another
DT-2	The	No	This
DT-3	The	Some	These
DT-4	all	those	some
DT-5	some	these	both
DT-6	That	This	each
DT-7	this	that	each
DT-8	the	The	a
DT-9	no	any	some
DT-10	an	a	the
DT-11	a	this	the

CD			
CD-0	1	50	100
CD-1	8.50	15	1.2
CD-2	8	10	20
CD-3	1	30	31
CD-4	1989	1990	1988
CD-5	1988	1987	1990
CD-6	two	three	five
CD-7	one	One	Three
CD-8	12	34	14
CD-9	78	58	34
CD-10	one	two	three
CD-11	million	billion	trillion

PRP			
PRP-0	It	He	I
PRP-1	it	he	they
PRP-2	it	them	him

RBR			
RBR-0	further	lower	higher
RBR-1	more	less	More
RBR-2	earlier	Earlier	later

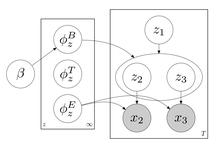
IN			
IN-0	In	With	After
IN-1	In	For	At
IN-2	in	for	on
IN-3	of	for	on
IN-4	from	on	with
IN-5	at	for	by
IN-6	by	in	with
IN-7	for	with	on
IN-8	If	While	As
IN-9	because	if	while
IN-10	whether	if	That
IN-11	that	like	whether
IN-12	about	over	between
IN-13	as	de	Up
IN-14	than	ago	until
IN-15	out	up	down

RB			
RB-0	recently	previously	still
RB-1	here	back	now
RB-2	very	highly	relatively
RB-3	so	too	as
RB-4	also	now	still
RB-5	however	Now	However
RB-6	much	far	enough
RB-7	even	well	then
RB-8	as	about	nearly
RB-9	only	just	almost
RB-10	ago	earlier	later
RB-11	rather	instead	because
RB-12	back	close	ahead
RB-13	up	down	off
RB-14	not	Not	maybe
RB-15	n't	not	also

The most frequent three words in the subcategories of several part-of-speech tags.

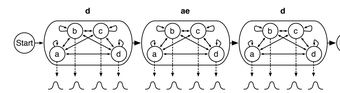
Extensions

Hierarchical Dirichlet Processes as a nonparametric Bayesian alternative to split and merge:



Reference:
Percy Liang, Slav Petrov, Michael Jordan and Dan Klein,
"The Infinite PCFG using Hierarchical Dirichlet Processes",
in **EMNLP-CoNLL '07**

Automatic refinement of acoustic models learns phone-internal structure as well as phone-external context:



Reference:
Slav Petrov, Adam Pauls and Dan Klein,
"Learning Structured Models for Phone Recognition",
in **EMNLP-CoNLL '07**