Unseen Patterns: Using Latent-Variable Models for Natural Language

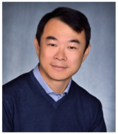
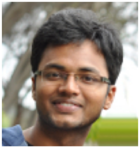
Shay Cohen
Institute for Language, Cognition and Computation
School of Informatics
University of Edinburgh

July 13, 2017

Thanks to...

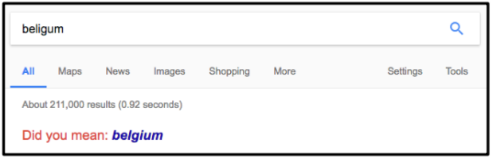





JOINT WORK WITH ALL MY COLLABORATORS – SEE REFERENCES

Natural Language Processing



Translate
Understand
Search
Organise
Extract information
Interact

Main Challenge: Ambiguity

Ambiguity: Natural language utterances have many possible analyses

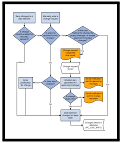
Need to prune out thousands of interpretations even for simple sentences (for example: parse trees)

Variability

Many surface forms for a single meaning:

There is a bird singing
A bird standing on a branch singing
A bird opening its mouth to sing
A black and yellow bird singing in nature
A Rufous Whistler singing
A bird with a white patch on its neck

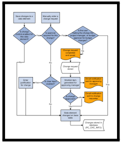
Approach to NLP



1980s - rule based systems

1990s and onwards - data-driven (machine learning)

Approach to NLP



1980s - rule based systems

1990s and onwards - data-driven (machine learning)

Challenge: The labeled data bottleneck

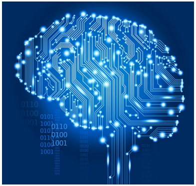
Labeled Data Bottleneck

Approach to NLP since 1990s: use labeled data. Leads to the labeled data bottleneck – never enough data

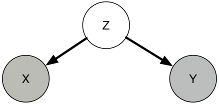
How to solve the labeled data bottleneck?

Ignore it

Unsupervised learning

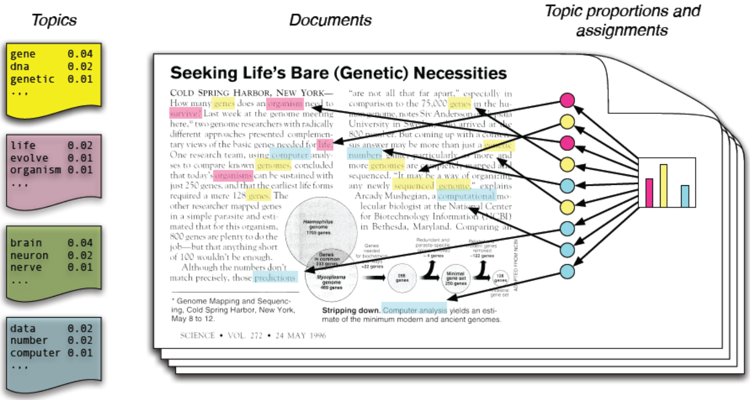


Latent-variable modelling



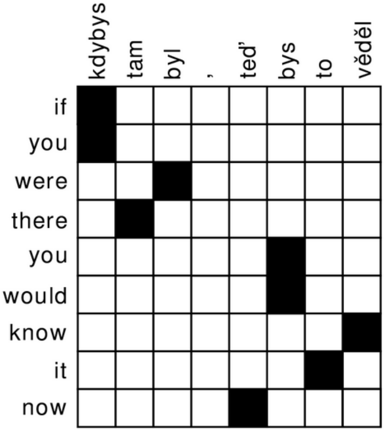
incomplete data

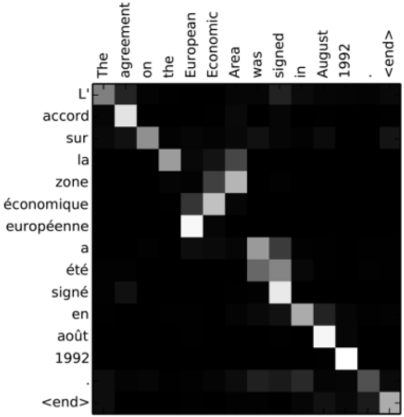
Topic Modeling



(Image from Blei, 2011)

Machine Translation





- Alignment is a hidden variable in translation models
- With deep learning, this is embodied in “attention” models

Bayesian Learning

With Bayesian inference, the parameters are a "latent" variable:

$$p(\theta, h \mid x) = \frac{p(\theta, h, x)}{\int_\theta \sum_h p(\theta, h, x)}$$

- Popularized latent-variable models (where structure is missing as well)
- Has been used for problems in morphology, word segmentation, syntax, semantics and others

This Talk in a Nutshell

How do we learn from incomplete data?

- The case of syntactic parsing
- Other uses of grammars for learning from incomplete data
- The canonical correlation principle and its uses

Why Parsing?

Do we need to work on parsing when we can build direct “transducers?” (such as with deep learning)

Why Parsing?

Do we need to work on parsing when we can build direct "transducers?" (such as with deep learning) **Yes!**

- We develop algorithms that generalize to structured prediction
- We see recent results that even with deep learning, incorporating parse structures can help applications such as machine translation (Bastings et al., 2017; Kim et al., 2017)
- We develop theories for syntax in language and test them empirically
- One of the classic problems that demonstrates so well ambiguity in natural language

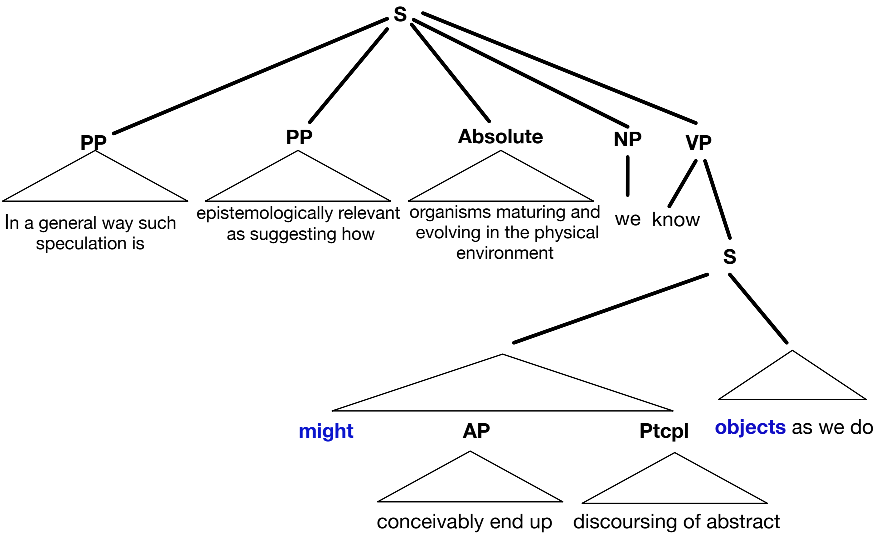
Ambiguity: Example from Abney (1996)

In a general way such speculation is epistemologically relevant, as suggesting how organisms maturing and evolving in the physical environment we know might conceivably end up discoursing of abstract objects as we do

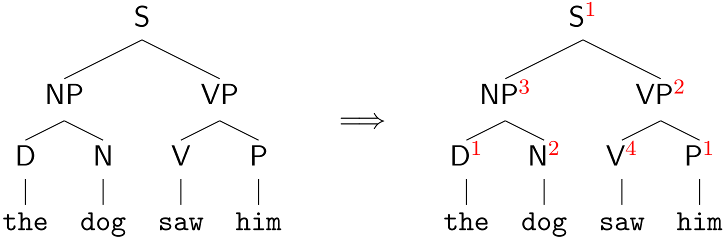
(Quine, 1960, p. 123)

- Should be interpreted: organisms might end up ...

Ambiguity Revisited

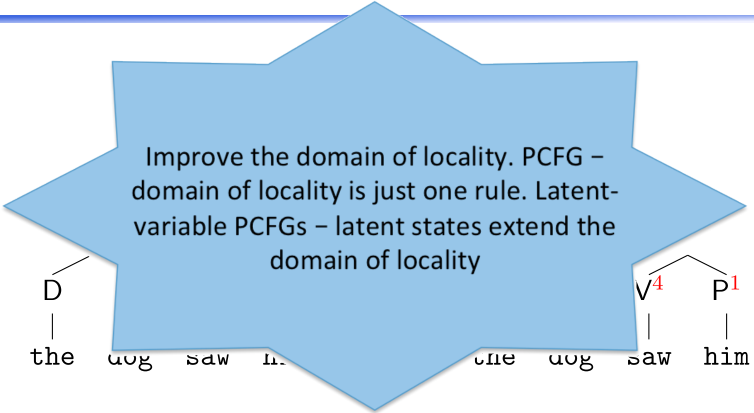


Latent-State Syntax (Matsuzaki et al., 2005; Prescher, 2005; Petrov et al., 2006)



Improves the accuracy of a PCFG model from $\sim 70\%$ to $\sim 90\%$.

Latent-State Syntax (Matsuzaki et al., 2005; Prescher, 2005; Petrov et al., 2006)

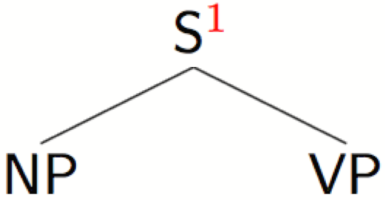


Improves the accuracy of a PCFG model from $\sim 70\%$ to $\sim 90\%$.

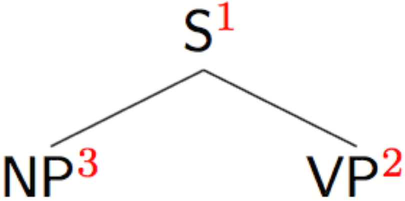
Generative Process

S^1

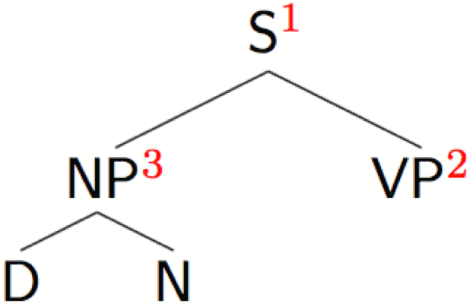
Generative Process



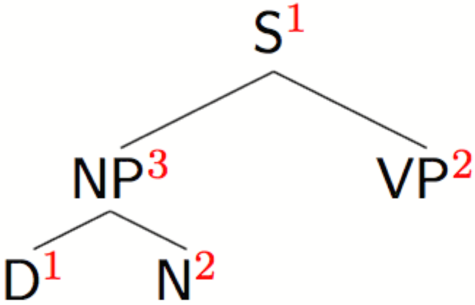
Generative Process



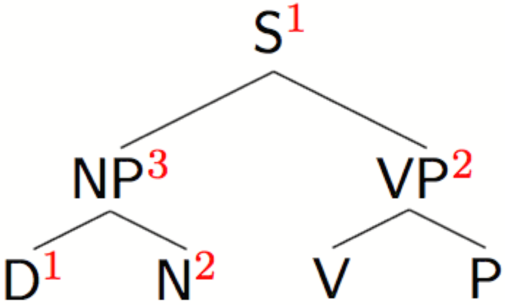
Generative Process



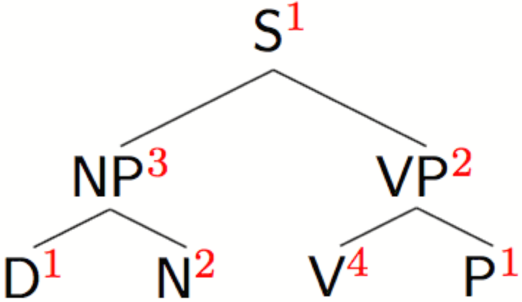
Generative Process



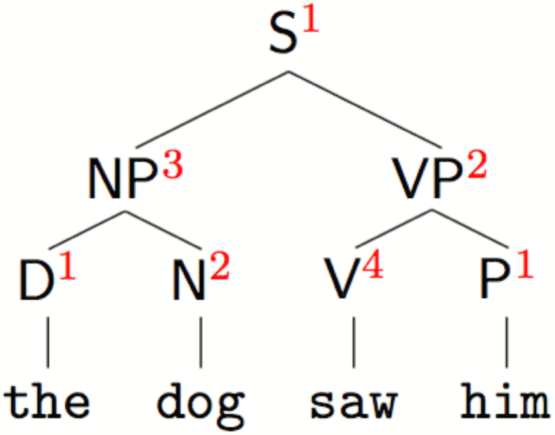
Generative Process



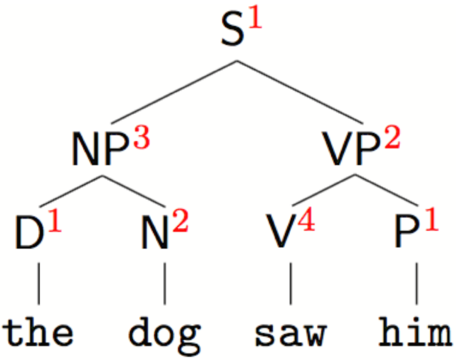
Generative Process



Generative Process



Generative Process

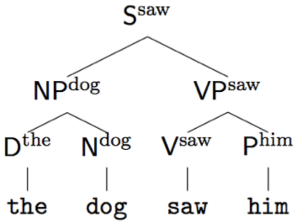


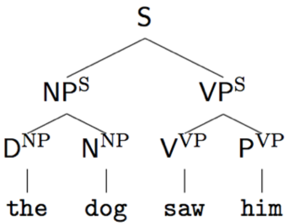
- Derivational process is similar to that of PCFG together with contextual information

- We read the grammar off the treebank, but not the latent states

Evolution of L-PCFGs

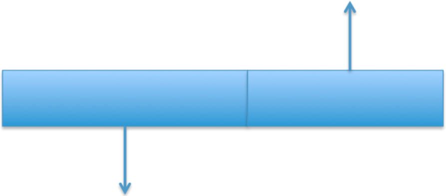
1997 - Charniak, Collins
Lexicalized grammars
1998 – Johnson
Treebank transformations
1999 – Eisner
Bilexical grammars





Evolution of L-PCFGs

2004 - Klein and Manning
Linguistic annotations



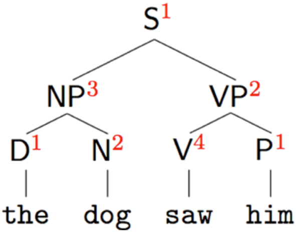
1997 - Charniak, Collins
Lexicalized grammars
1998 – Johnson
Treebank transformations
1999 – Eisner
Bilexical grammars

Evolution of L-PCFGs

2004 - Klein and Manning
Linguistic annotations

1997 - Charniak, Collins
Lexicalized grammars
1998 – Johnson
Treebank transformations
1999 – Eisner
Bilexical grammars

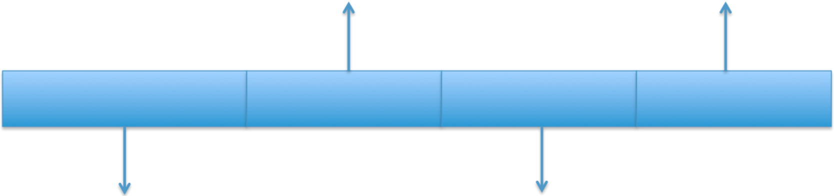
2005 - Matsuzaki et al.
2005 - Prescher
Latent annotations



Evolution of L-PCFGs

2004 - Klein and Manning
Linguistic annotations

2006 - Petrov et al.
Coarse-to-fine methods



1997 - Charniak, Collins
Lexicalized grammars
1998 – Johnson
Treebank transformations
1999 – Eisner
Bilexical grammars

2005 - Matsuzaki et al.
2005 - Prescher
Latent annotations

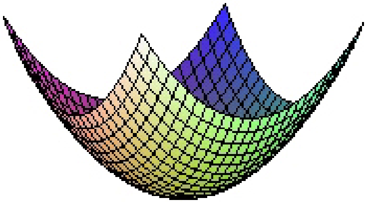
The Estimation Problem

Goal: Given a treebank, estimate rule probabilities, including for latent states.

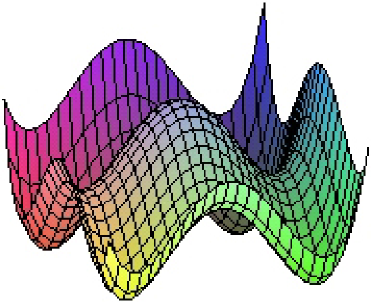
Traditional way: use the expectation-maximization (EM) algorithm:

- E-step - infer values for latent states using dynamic programming
- M-step - re-estimate the model parameters based on the values inferred

Local Maxima with EM



Convex optimization



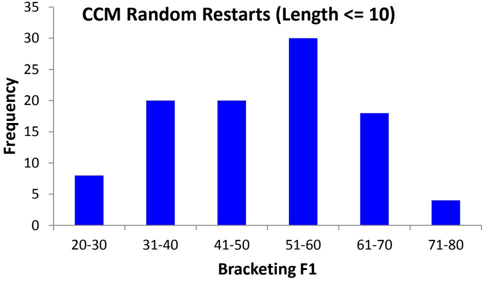
Non-convex optimization

EM finds a local maximum of a non-convex objective

Especially problematic with unsupervised learning

How Problematic are Local Maxima?

For unsupervised learning, local maxima are a very serious problem:

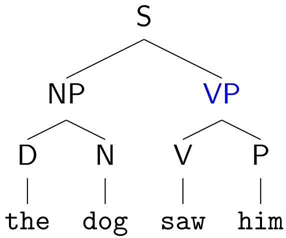


For deep learning, can also be a problem. For L-PCFGs, variability is smaller

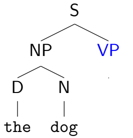
Depends on the problem and the model

Basic Intuition

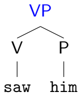
At node VP:



Outside tree $o =$



Inside tree $t =$



Conditionally independent given the label and the hidden state

$$p(o, t|\text{VP}, h) = p(o|\text{VP}, h) \times p(t|\text{VP}, h)$$

Cross-Covariance Matrix

Create a cross-covariance matrix and apply singular value decomposition to get the latent space:

$$
\begin{array}{c c}
 & \begin{array}{cccc} \text{outside tree 1} & & & \text{outside tree 10} \end{array} \\
\begin{array}{c} \text{inside tree 1} \\ \\ \\ \text{inside tree 10} \end{array} &
\begin{pmatrix}
1 & 0 & \dots & 1 \\
0 & 1 & \dots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
1 & 0 & \dots & 1
\end{pmatrix}
\end{array}
$$

Based on the method of moments – set up a set of equations that mix moments and parameters and have a unique solution

Previous Work

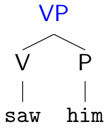
The idea of using a co-ocurrence matrix to extract latent information is an old idea. It has been used for:

- Learning hidden Markov models and finite state automata (Hsu et al., 2012; Balle et al., 2013)
- Learning word embeddings (Dhillon et al., 2011)
- Learning dependency and other types of grammars (Bailly et al., 2010; Luque et al., 2012; Dhillon et al., 2012)
- Learning document-topic structure (Anandkumar et al., 2012)

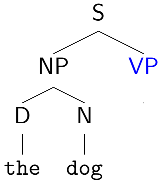
Much of this work falls under the use of **canonical correlation analysis** (Hotelling, 1935)

Feature Functions

Need to define feature functions for inside and outside trees

$\phi($ 

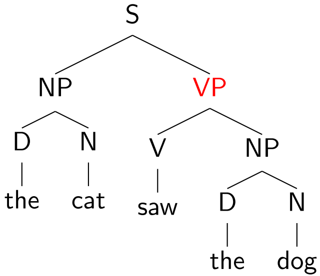
$) = (0,\ldots,1,\ldots,1,0,1,0)$

$\psi($ 

$) = (0,\ldots,1,\ldots,0,0,0,1)$

Inside Features Used

Consider the VP node in the following tree:

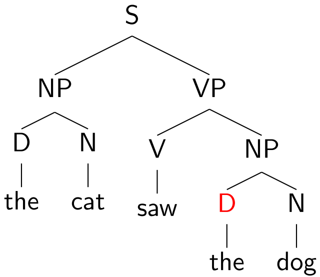


The inside features consist of:

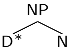
- The pairs `(VP, V)` and `(VP, NP)`
- The rule `VP` $\rightarrow$ `V NP`
- The tree fragment `(VP (V saw) NP)`
- The tree fragment `(VP V (NP D N))`
- The pair of head part-of-speech tag with VP: `(VP, V)`
- The width of the subtree spanned by VP: `(VP, 2)`

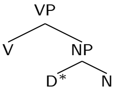
Outside Features Used

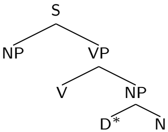
Consider the D node in the following tree:



The outside features consist of:

- The fragments 

, 

and 

- The pair `(D, NP)` and triplet `(D, NP, VP)`
- The pair of head part-of-speech tag with D: `(D, N)`
- The widths of the spans left and right to D: `(D, 3)` and `(D, 1)`

Outside Features Used

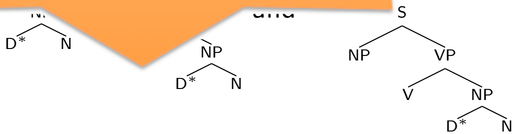
Consider the D node in the following tree:

S

As such, the algorithm is another step in the evolution of L-PCFGs – throw in all the information and local context from previous work and automatically distill it into latent states!

The outside fe

- The fragr



- The pair (D, NP) and triplet (D, NP, VP)
- The pair of head part-of-speech tag with D: (D, N)
- The widths of the spans left and right to D: (D, 3) and (D, 1)

Final Results on Multilingual Parsing

Narayan and Cohen (2016):

language	Berkeley	Spectral	
		Cluster	SVD
Basque	74.7	**81.4**	80.5
French	**80.4**	75.6	79.1
German	**78.3**	76.0	78.2
Hebrew	87.0	87.2	**89.0**
Hungarian	85.2	88.4	**89.2**
Korean	78.6	78.4	**80.0**
Polish	86.8	91.2	**91.8**
Swedish	80.6	79.4	**80.9**

Parsing is far from being solved in the multilingual setting

What Do We Learn?

Closed-word tags essentially do lexicalization:

State	Frequent words
IN (preposition)	
0	of $\times 323$
1	about $\times 248$
2	than $\times 661$, as $\times 648$, because $\times 209$
3	from $\times 313$, at $\times 324$
4	into $\times 178$
5	over $\times 122$
6	Under $\times 127$

What Do We Learn?

State	Frequent words
`DT` (determiners)	
0	These $\times 105$
1	Some $\times 204$
2	that $\times 190$
3	both $\times 102$
4	any $\times 613$
5	the $\times 574$
6	those $\times 247$, all $\times 242$
7	all $\times 105$
8	another $\times 276$, no $\times 211$

What Do We Learn?

State	Frequent words
`CD` (numbers)	
0	8 $\times 132$
1	million $\times 451$, billion $\times 248$
`RB` (adverb)	
0	up $\times 175$
1	as $\times 271$
2	not $\times 490$, n't $\times 2695$
3	not $\times 236$
4	only $\times 159$
5	well $\times 129$

What Do We Learn?

State	Frequent words
`CC` (conjunction)	
0	But $\times 255$
1	and $\times 101$
2	and $\times 218$
3	But $\times 196$
4	or $\times 162$
5	and $\times 478$

What Do We Learn?

Example latent for NP:

- "James McCall , vice president , materials , at Battelle , a technology and management-research giant based in Columbus , Ohio"
- "Frank Kline Jr. , partner in Lambda Funds , a Beverly Hills , Calif. , venture capital concern"
- "Allen Hadhazy , senior analyst at the Institute for Econometric Research , Fort Lauderdale , Fla. , which publishes the New Issues newsletter on IPOs"
- "a group of investment banks headed by First Boston Corp. and co-managed by Goldman , Sachs & Co. , Merrill Lynch Capital Markets , Morgan Stanley & Co. , and Salomon Brothers Inc"
- "Charles J. O'Connell , deputy district director in Los Angeles of the California Department of Transportation , nicknamed Caltrans"
- "Francis J. McNeil , who , as deputy assistant secretary of state for inter-American affairs , first ran across reports about Mr. Noriega in 1977"

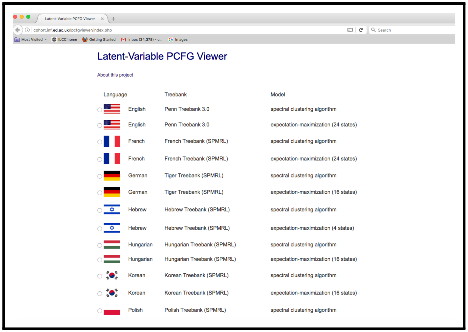
What Do We Learn?

Example latent state for NP:
"Aug. 30 , 1988", "Aug. 31 , 1987", "Dec. 31 , 1988", "Oct. 16 , 1996", "Oct. 1 , 1999", "Oct. 1 , 2019", "Nov. 8 , 1996", "Oct. 15 , 1999", "April 30 , 1988", "Nov. 8 , 1994"
Another state:

- "AMERICAN BUILDING MAINTENANCE INDUSTRIES Inc. , San Francisco , provider of maintenance services , annual revenue of $ 582 million , NYSE ,"
- "DIASONICS INC. , South San Francisco , maker of magnetic resonance imaging equipment , annual sales of $ 281 million , Amex ,"
- "EVEREX SYSTEMS INC. , Fremont , maker of personal computers and peripherals , annual sales of $ 377 million , OTC ,"
- "ANTHEM ELECTRONICS INC. , San Jose , distributor of electronic parts , annual sales of about $ 300 million , NYSE ,"

LPCFGViewer

If you are interested in further looking at such patterns for other languages (English, French, German, Hebrew, Hungarian, Korean, Polish, Swedish, Basque), consider visiting `http://cohort.inf.ed.ac.uk/lpcfgviewer/index.php`.

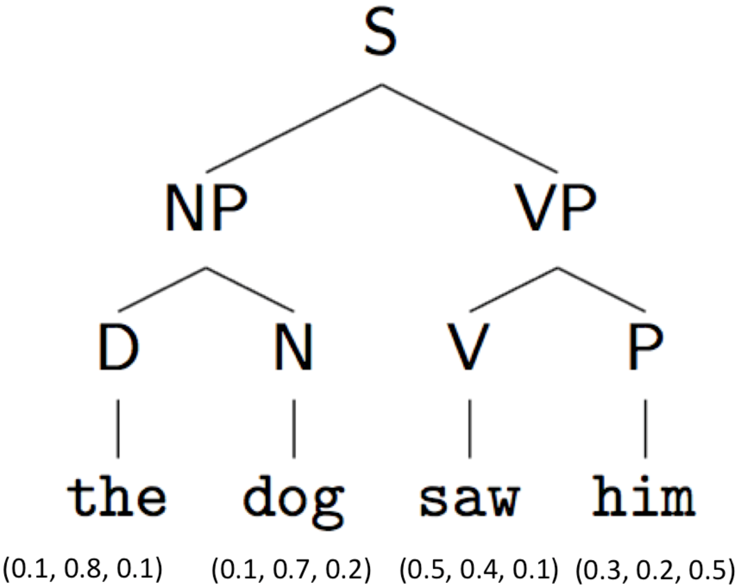


This Talk in a Nutshell

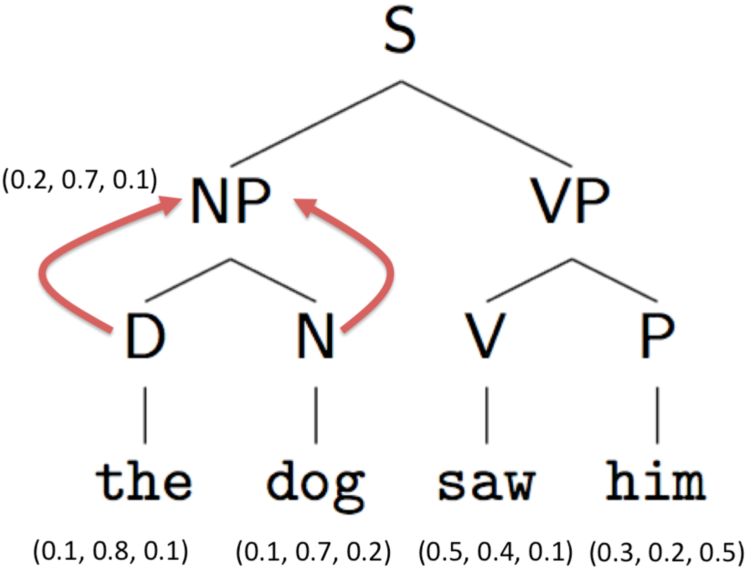
How do we learn from incomplete data?

- The case of syntactic parsing
- Other uses of grammars for learning from incomplete data
- The canonical correlation principle and its uses

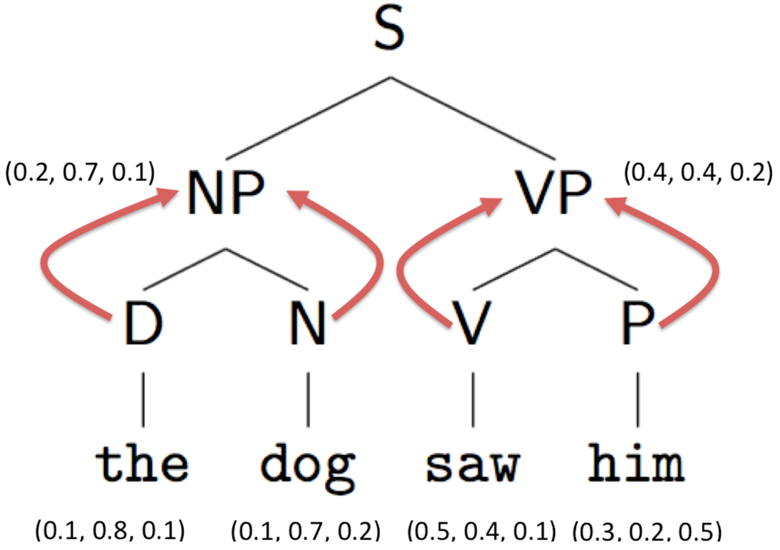
A Different Perspective



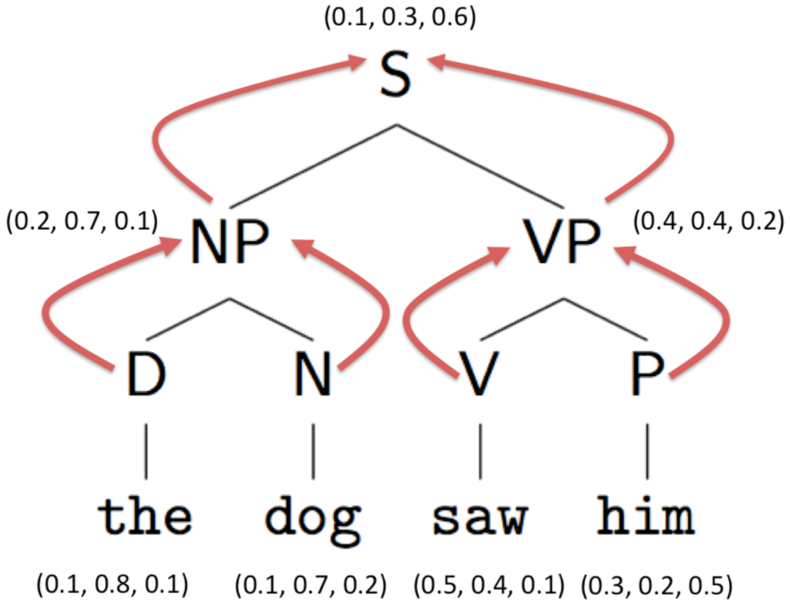
A Different Perspective



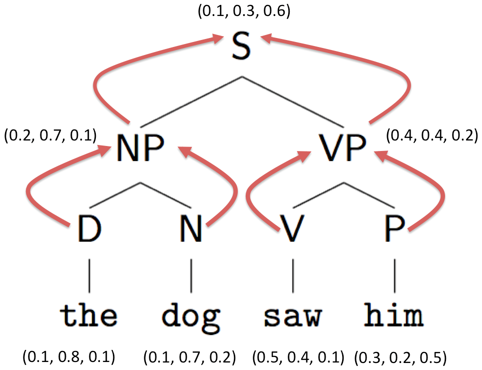
A Different Perspective



A Different Perspective

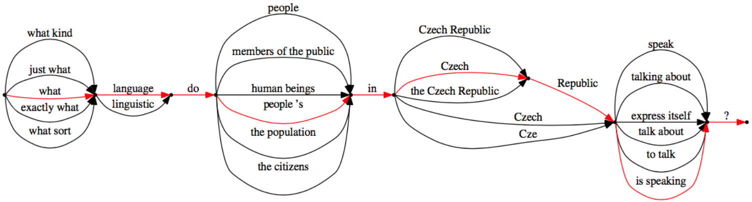


A Different Perspective

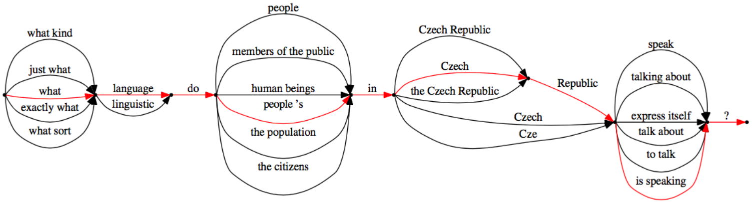


- Related to neural network models with grammars (Socher et al., 2010; Socher et al., 2013)
- Also related to compositional distributional semantics (Baroni and Bernardi, 2014; Grefenstette and Sadrzadeh, 2010; Coecke et al., 2010)

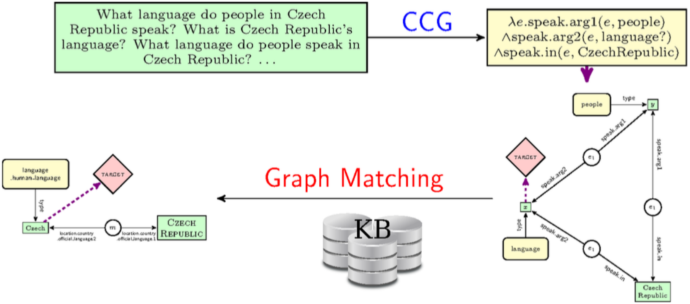
Question Answering (Narayan et al., 2016)



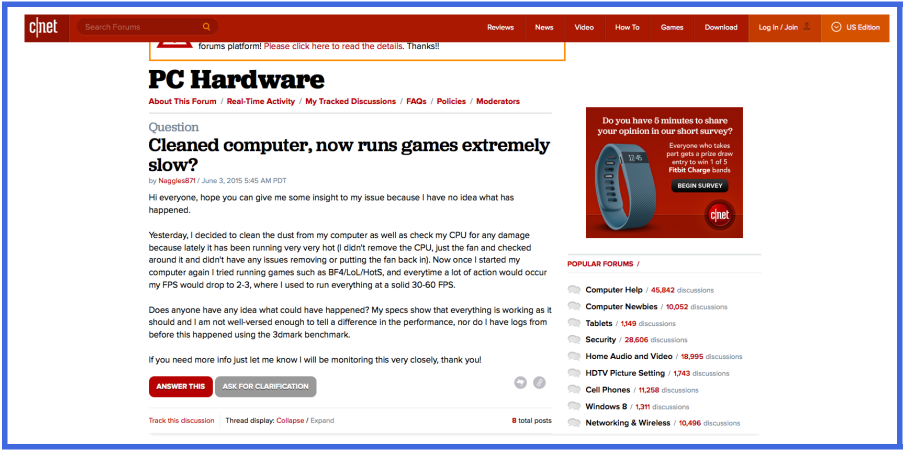
Question Answering (Narayan et al., 2016)



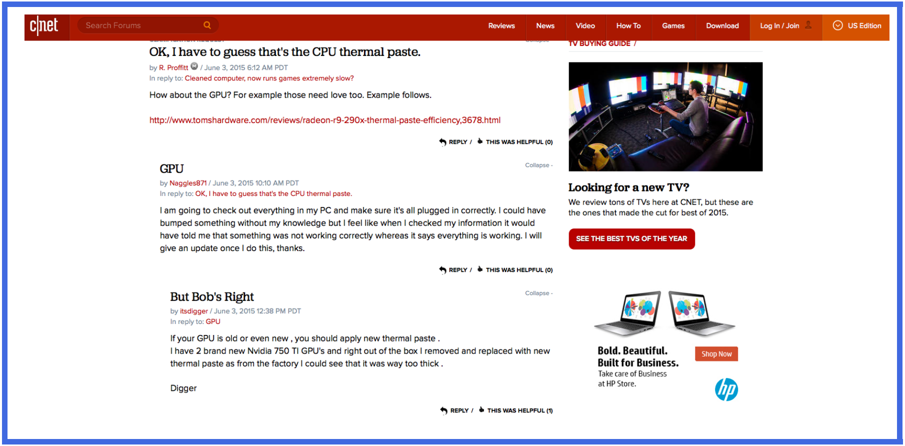
What language do people in Czech Republic speak?



Discussion Forums



Discussion Forums


cnet Search Forums Reviews News Video How To Games Download Log In / Join US Edition

OK, I have to guess that's the CPU thermal paste.

by R. Proffitt / June 3, 2015 6:12 AM PDT
In reply to: Cleaned computer, now runs games extremely slow?

How about the GPU? For example those need love too. Example follows.

http://www.tomshardware.com/reviews/radeon-r9-290x-thermal-paste-efficiency,3678.html

REPLY / THIS WAS HELPFUL (0)

GPU Collapse -

by Naggles871 / June 3, 2015 10:10 AM PDT
In reply to: OK, I have to guess that's the CPU thermal paste.

I am going to check out everything in my PC and make sure it's all plugged in correctly. I could have bumped something without my knowledge but I feel like when I checked my information it would have told me that something was not working correctly whereas it says everything is working. I will give an update once I do this, thanks.

REPLY / THIS WAS HELPFUL (0)

But Bob's Right Collapse -

by itsdigger / June 3, 2015 12:38 PM PDT
In reply to: GPU

If your GPU is old or even new , you should apply new thermal paste .
I have 2 brand new Nvidia 750 TI GPU's and right out of the box I removed and replaced with new thermal paste as from the factory I could see that it was way too thick .

Digger

REPLY / THIS WAS HELPFUL (1)

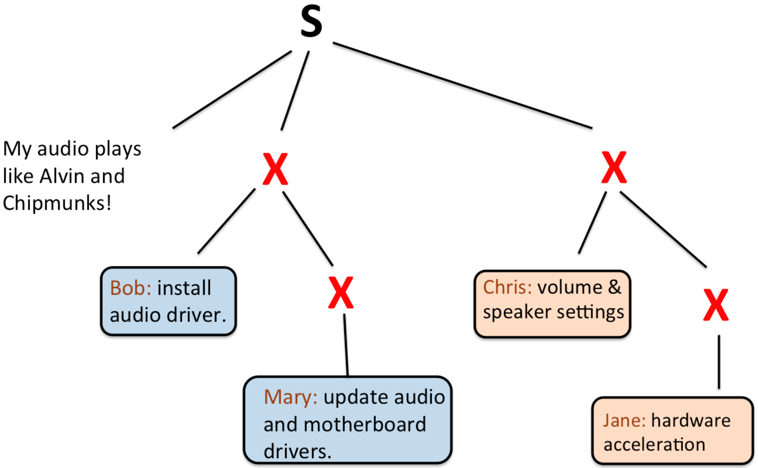
p_0	Bob:	When I play a recorded video on my camera, it looks and sounds fine. On my computer, it plays at a really fast rate and sounds like Alvin and the Chipmunks!
p_1	Kate:	I'd find and install the machine's latest audio driver.
p_2	Mary:	The motherboard supplies the clocks for audio feedback. So update the audio and motherboard drivers.
p_3	Chris:	Another fine mess in audio is volume and speaker settings. You checked these?
p_4	Jane:	Yes, under speaker settings, look for hardware acceleration. Turning it off worked for me.
p_5	Matt:	Audio drivers are at this link. Rather than just audio drivers, I would also just do all drivers.

p_0	Bob:	When I play a recorded video on my camera, it looks and sounds fine. On my computer, it plays at a really fast rate and sounds like Alvin and the Chipmunks!
p_1	Kate:	I'd find and install the machine's latest audio driver.
p_2	Mary:	The motherboard supplies the clocks for audio feedback. So update the audio and motherboard drivers.
p_3	Chris:	Another fine mess in audio is volume and speaker settings. You checked these?
p_4	Jane:	Yes, under speaker settings, look for hardware acceleration. Turning it off worked for me.
p_5	Matt:	Audio drivers are at this link. Rather than just audio drivers, I would also just do all drivers.

p_0	Bob:	When I play a recorded video on my camera, it looks and sounds fine. On my computer, it plays at a really fast rate and sounds like Alvin and the Chipmunks!
p_1	Kate:	I'd find and install the machine's latest audio driver.
p_2	Mary:	The motherboard supplies the clocks for audio feedback. So update the audio and motherboard drivers.
p_3	Chris:	Another fine mess in audio is volume and speaker settings. You checked these?
p_4	Jane:	Yes, under speaker settings, look for hardware acceleration. Turning it off worked for me.
p_5	Matt:	Audio drivers are at this link. Rather than just audio drivers, I would also just do all drivers.

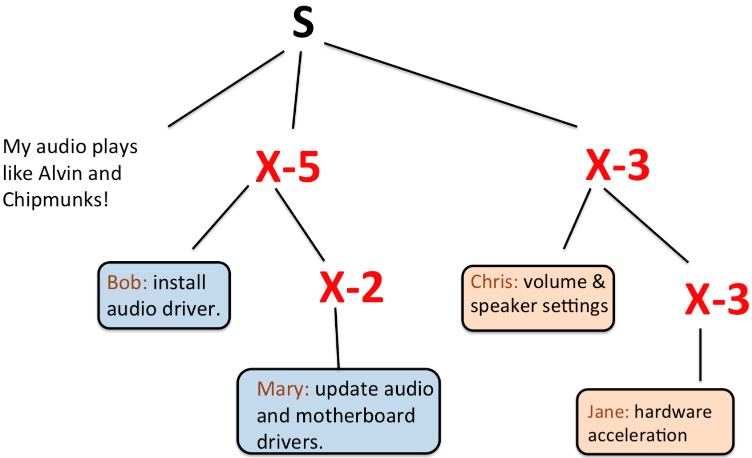
p_0	Bob:	When I play a recorded video on my camera, it looks and sounds fine. On my computer, it plays at a really fast rate and sounds like Alvin and the Chipmunks!
p_1	Kate:	I'd find and install the machine's latest audio driver.
p_2	Mary:	The motherboard supplies the clocks for audio feedback. So update the audio and motherboard drivers.
p_3	Chris:	Another fine mess in audio is volume and speaker settings. You checked these?
p_4	Jane:	Yes, under speaker settings, look for hardware acceleration. Turning it off worked for me.
p_5	Matt:	Audio drivers are at this link. Rather than just audio drivers, I would also just do all drivers.

Conversation Trees



(Louis and Cohen, 2015)

Conversation Trees



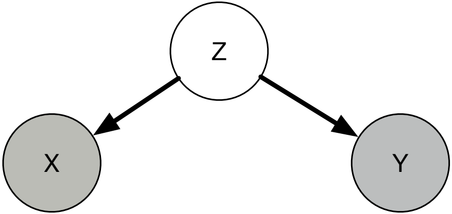
(Louis and Cohen, 2015)

This Talk in a Nutshell

How do we learn from incomplete data?

- The case of syntactic parsing
- Other uses of grammars for learning from incomplete data
- The canonical correlation principle and its uses

Canonical Correlation Analysis



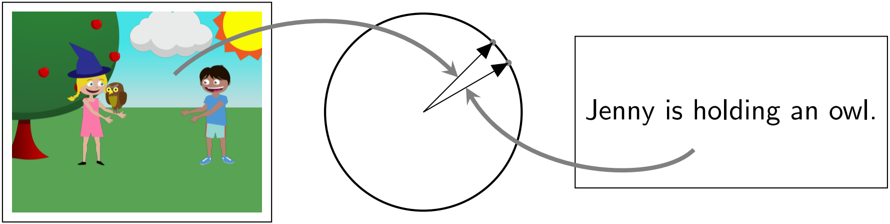
- Assume a "confounding" variable that explains two separate views
- Correlation between x and y gives z – the two are independent given z
- In the case of L-PCFGs: x and y are inside and outside trees
- Where else can this principle be used?

Word Embeddings

$$
\begin{array}{cc}
 & \begin{array}{cccc} \text{the} \star \text{chased} & & & \text{a} \star \text{ran} \end{array} \\
\begin{array}{c} \text{mouse} \\ \\ \\ \text{cat} \end{array} &
\begin{pmatrix}
1 & 0 & \dots & 1 \\
0 & 1 & \dots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
1 & 0 & \dots & 1
\end{pmatrix}
\end{array}
$$

- Co-occurrence matrix of words and contexts ("the *cat* chased", "the *mouse* chased")
- Apply CCA on this matrix to get word embeddings (Dhillon et al., 2011)
- Inject prior knowledge into matrix (Osborne et al., 2016)

Canonical Correlation Inference

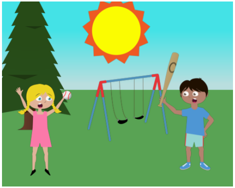


- Requires also generation (using sampling techniques)
- The probability of text we sample is proportional to the "similarity" of the text to the image

Example Predictions

Good predictions:

mike and jenny are camping



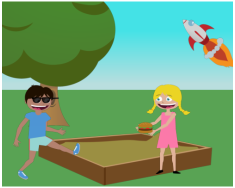
mike is holding a bat

jenny is throwing the frisbee

Bad predictions:

mike is kicking a blass

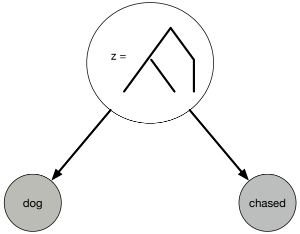
jenny wants the bear



the rocket is behind mike

Unsupervised Parsing

The dog, true to form, chased the cat.



(Parikh et al., 2014)

- Very challenging problem
- Sensitive to local maxima with existing techniques such as EM
- What if the tree for each pair of words in the sentence is a latent, confounding, hierarchical variable?

Conclusion

Latent-variable grammars are useful for problems outside of syntax

- Their symbolic component is interpretable
- Their probabilistic component helps reasoning under uncertainty
- Latent variables help detect unseen patterns

I have shown you how grammars can be used for several problems, and also how the principle behind learning latent-variable grammars can be used for other problems.

References I

Steven Abney.
Statistical methods and linguistics.
The balancing act: Combining symbolic and statistical approaches to language, pages 1–26, 1996.

Anima Anandkumar, Dean P Foster, Daniel J Hsu, Sham M Kakade, and Yi-Kai Liu.
A spectral algorithm for latent dirichlet allocation.
In *Advances in Neural Information Processing Systems*, pages 917–925, 2012.

R. Bailly, A. Habrard, and F. Denis.
A spectral approach for probabilistic grammatical inference on trees.
In *Proceedings of ALT*, 2010.

References II

Borja Balle, Xavier Carreras, Franco M Luque, and Ariadna Quattoni.
Spectral learning of weighted automata.
Machine Learning, 96(1-2):33–63, 2014.

Marco Baroni, Raffaela Bernardi, and Roberto Zamparelli.
Frege in space: A program of compositional distributional semantics.
LiLT (Linguistic Issues in Language Technology), 9, 2014.

Joost Bastings, Ivan Titov, Wilker Aziz, Diego Marcheggiani, and Khalil Sima'an.
Graph convolutional encoders for syntax-aware neural machine translation.
arXiv preprint arXiv:1704.04675, 2017.

References III

D. M. Blei, A. Ng, and M. Jordan.
Latent Dirichlet allocation.
Journal of Machine Learning Research, 3:993–1022, 2003.

Bob Coecke, Mehrnoosh Sadrzadeh, and Stephen Clark.
Mathematical foundations for a compositional distributional model of meaning.
arXiv preprint arXiv:1003.4394, 2010.

S. B. Cohen.
Latent-variable PCFGs: Background and applications.
In *Proceedings of ACL*, 2017.

S. B. Cohen, K. Stratos, M. Collins, D. F. Foster, and L. Ungar.
Spectral learning of latent-variable PCFGs.
In *Proceedings of ACL*, 2012.

References IV

S. B. Cohen, K. Stratos, M. Collins, D. P. Foster, and L. Ungar.
Experiments with spectral learning of latent-variable PCFGs.
In *Proceedings of NAACL*, 2013.

Shay B. Cohen.
Bayesian Analysis in Natural Language Processing.
Synthesis Lectures on Human Language Technologies. Morgan and Claypool, 2016.

M. Collins.
Head-driven statistical models for natural language processing.
Computational Linguistics, 29:589–637, 2003.

A. Dempster, N. Laird, and D. Rubin.
Maximum likelihood estimation from incomplete data via the EM algorithm.
Journal of the Royal Statistical Society B, 39:1–38, 1977.

References V

P. S. Dhillon, J. Rodu, M. Collins, D. P. Foster, and L. H. Ungar.
Spectral dependency parsing with latent variables.
In *Proceedings of CoNLL-EMNLP*, 2012.

Paramveer S Dhillon, Dean P Foster, and Lyle H Ungar.
Eigenwords: Spectral word embeddings.
Journal of Machine Learning Research, 16:3035–3078, 2015.

J. Eisner and G. Satta.
Efficient parsing for bilexical context-free grammars and head automaton grammars.
In *Proceedings of ACL*, 1999.

References VI

Edward Grefenstette and Mehrnoosh Sadrzadeh.
Experimental support for a categorical compositional distributional model of meaning.
In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1394–1404. Association for Computational Linguistics, 2011.

H Hotelling.
Canonical correlation analysis (cca).
Journal of Educational Psychology, 1935.

D. Hsu, S. M. Kakade, and T. Zhang.
A spectral algorithm for learning hidden Markov models.
Journal of Computer and System Sciences, 78(5):1460–1480, 2012.

References VII

Helen Jiang, Nikos Papasarantopoulos, and Shay B. Cohen.
Canonical correlation inference for mapping abstract scenes to text.
Technical report, 2016.

M. Johnson.
PCFG models of linguistic tree representations.
Computational Linguistics, 24(4):613–632, 1998.

Yoon Kim, Carl Denton, Luong Hoang, and Alexander M Rush.
Structured attention networks.
arXiv preprint arXiv:1702.00887, 2017.

D. Klein and C. D. Manning.
Accurate unlexicalized parsing.
In *Proceedings of ACL*, 2003.

References VIII

A. Louis and S. B. Cohen.
Conversation trees: A grammar model for topic structure in forums.
In *Proceedings of EMNLP*, 2015.

F. M. Luque, A. Quattoni, B. Balle, and X. Carreras.
Spectral learning for non-deterministic dependency parsing.
In *Proceedings of EACL*, 2012.

T. Matsuzaki, Y. Miyao, and J. Tsujii.
Probabilistic CFG with latent annotations.
In *Proceedings of ACL*, 2005.

S. Narayan and S. B. Cohen.
Optimizing spectral learning for parsing.
In *Proceedings of ACL*, 2016.

References IX

Shashi Narayan, Siva Reddy, and Shay B. Cohen.
Paraphrase generation from latent-variable pcfgs for semantic parsing.
In *Proceedings of INLG*, 2015.

A. P. Parikh, S. B. Cohen, and E. Xing.
Spectral unsupervised parsing with additive tree metrics.
In *Proceedings of ACL*, 2014.

S. Petrov, L. Barrett, R. Thibaux, and D. Klein.
Learning accurate, compact, and interpretable tree annotation.
In *Proceedings of COLING-ACL*, 2006.

D. Prescher.
Inducing head-driven PCFGs with latent heads: Refining a tree-bank grammar for parsing.
In *Proceedings of ECML*, 2005.

References X

- A. Saluja, C. Dyer, and S. B. Cohen.
  Latent-variable synchronous CFGs for hierarchical translation.
  In *Proceedings of EMNLP*, 2014.
- R. Socher, J. Bauer, C. D. Manning, and A. Y. Ng.
  Parsing with compositional vector grammars.
  In *Proceedings of ACL*, 2013.
- R. Socher, C. D. Manning, and A. Y. Ng.
  Learning continuous phrase representations and syntactic parsing with recursive neural networks.
  In *Proceedings of the NIPS Deep Learning and Unsupervised Feature Learning Workshop*, 2010.