From language models to distributional semantics

Chung-chieh Shan
University of Tsukuba

1 June 2012

Approaches to semantics

"In order to say what a meaning *is*,
we may first ask what a meaning *does*,
and then find something that does that." —David Lewis

Approaches to semantics

"In order to say what a meaning *is*,
we may first ask what a meaning *does*,
and then find something that does that." —David Lewis

Truth, entailment

$$\text{Every person cried.} \quad \vDash \quad \text{Every professor cried.}$$

$$\text{A person cried.} \quad \nvDash \quad \text{A professor cried.}$$

Formal semantics

$$\forall x.\, Px \to Cx$$

$$\lambda g.\, \forall x.\, Px \to gx \qquad C$$

$$\lambda f.\, \lambda g.\, \forall x.\, fx \to gx \qquad P$$

Approaches to semantics

"In order to say what a meaning *is*,
we may first ask what a meaning *does*,
and then find something that does that." —David Lewis

Concepts, similarity

$$\begin{array}{lcl} \text{ambulance} & \sim & \text{battleship} \\ \text{ambulance} & \nsim & \text{bookstore} \end{array}$$

Distributional semantics

$$\begin{array}{l c} & \begin{array}{cccccc} \text{abandon} & \text{abdominal} & \text{ability} & \text{academic} & \text{accept} & \cdots \end{array} \\ \begin{array}{l} \text{ambulance} \\ \text{battleship} \\ \text{bookstore} \\ \vdots \end{array} & \begin{pmatrix} 27 & 10 & 50 & 17 & 130 & \cdots \\ 35 & 0 & 32 & 1 & 25 & \cdots \\ 5 & 0 & 6 & 33 & 13 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \end{array}$$

Distributional semantics for entailment among words

For each word w, rank contexts c by descending $\dfrac{\Pr(c \mid w)}{\Pr(c)} > 1$.

"pointwise mutual information"

Distributional semantics for entailment among words

For each word w, rank contexts c by descending $\dfrac{\Pr(c \mid w)}{\Pr(c)} > 1$.

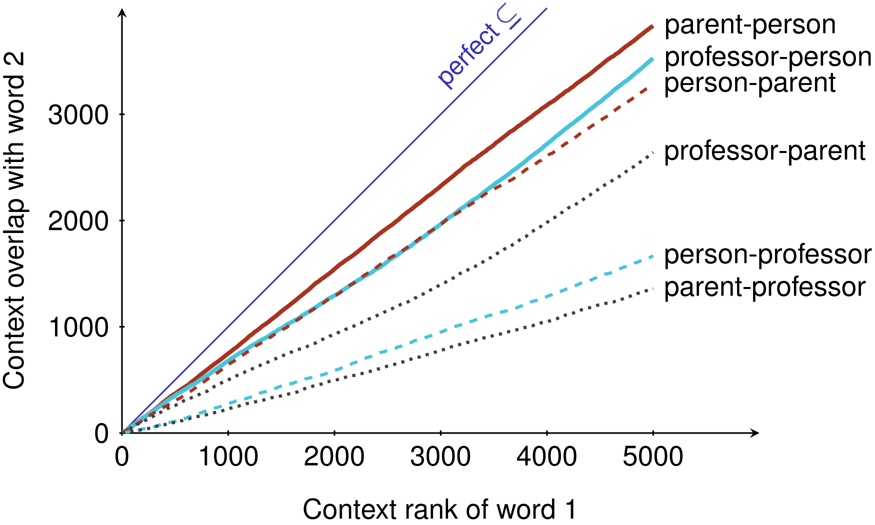
"pointwise mutual information"

parent argcount$_{\text{n}}$ arglist$_{\text{n}}$ arglist$_{\text{j}}$ phane$_{\text{n}}$ specity$_{\text{n}}$ qdisc$_{\text{n}}$ carthy$_{\text{n}}$ parents-to-be$_{\text{n}}$ non-resident$_{\text{j}}$ step-parent$_{\text{n}}$ tc$_{\text{n}}$ ballons$_{\text{n}}$ eliza$_{\text{n}}$ symptons$_{\text{n}}$ adoptive$_{\text{j}}$ stepparent$_{\text{n}}$ nonresident$_{\text{j}}$ home-school$_{\text{n}}$ scabrid$_{\text{n}}$ petiolule$_{\text{n}}$ ...

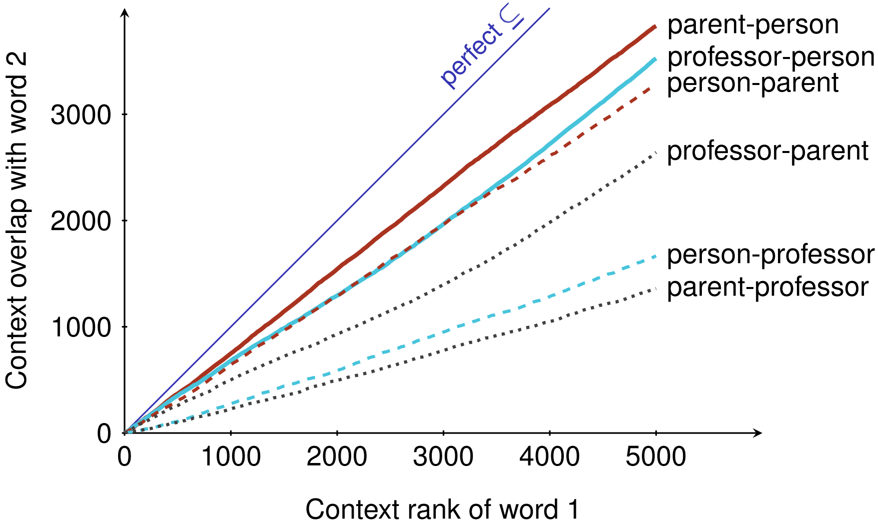
person anglia$_{\text{n}}$ first-mentioned$_{\text{j}}$ unascertained$_{\text{j}}$ enure$_{\text{v}}$ deposit-taking$_{\text{j}}$ bonis$_{\text{n}}$ iconclass$_{\text{j}}$ cotswolds$_{\text{n}}$ aforesaid$_{\text{n}}$ haver$_{\text{v}}$ foresaid$_{\text{j}}$ gha$_{\text{n}}$ sub-paragraphs$_{\text{n}}$ enacted$_{\text{j}}$ geest$_{\text{j}}$ non-medicinal$_{\text{j}}$ sub-paragraph$_{\text{n}}$ intimation$_{\text{n}}$ arrestment$_{\text{n}}$ incumbrance$_{\text{n}}$ ...

professor william$_{\text{n}}$ extraordinarius$_{\text{n}}$ ordinarius$_{\text{n}}$ francis$_{\text{n}}$ reid$_{\text{n}}$ emeritus$_{\text{n}}$ emeritus$_{\text{j}}$ derwent$_{\text{n}}$ regius$_{\text{n}}$ laurence$_{\text{n}}$ edward$_{\text{n}}$ carisoprodol$_{\text{n}}$ adjunct$_{\text{j}}$ winston$_{\text{n}}$ privatdozent$_{\text{j}}$ edward$_{\text{j}}$ xanax$_{\text{n}}$ tenure$_{\text{v}}$ cialis$_{\text{n}}$ florence$_{\text{n}}$ ...

Distributional semantics for entailment among words



Distributional semantics for entailment among words



More sophisticated: *Kullback-Leibler divergence*, *skew divergence* (Lee), *balAPinc* (Kotlerman et al.), . . .

Sparse data strikes back

Successes for words and short phrases:

- similarity
- entailment
- sentiment

'common sense' from noisy large corpora

For *long, rare, episodic* phrases and sentences, need

- syntactic structure
- pragmatic context
- grounding in other information sources

'linguistic generalization' from poor stimulus

This need goes way back—

From documents×terms to words×contexts

Information retrieval started with bag of terms in each document.
Stopwords, stemming, tagging; TF-IDF.

$$
\begin{array}{c}
\begin{array}{ccccc} \text{abandon} & \text{abdominal} & \text{ability} & \text{academic} & \dots \end{array} \\
\begin{pmatrix}
27 & 10 & 50 & 17 & \dots \\
35 & 0 & 32 & 1 & \dots \\
5 & 0 & 6 & 33 & \dots \\
\vdots & \vdots & \vdots & \ddots &
\end{pmatrix}
\end{array}
$$

$\vdots$

From documents×terms to words×contexts

Information retrieval started with bag of terms in each document. Stopwords, stemming, tagging; TF-IDF. Dimensionality reduction reveals topics.

$$
\begin{array}{c}
\begin{array}{ccccc} \text{abandon} & \text{abdominal} & \text{ability} & \text{academic} & \cdots \end{array} \\
\begin{pmatrix}
27 & 10 & 50 & 17 & \cdots \\
35 & 0 & 32 & 1 & \cdots \\
5 & 0 & 6 & 33 & \cdots \\
\vdots & \vdots & \vdots & \ddots &
\end{pmatrix}
\end{array}
$$

$$
= \begin{pmatrix} \quad \\ \quad \\ \quad \\ \quad \end{pmatrix} \times \begin{array}{c}
\begin{array}{ccccc} \text{abandon} & \text{abdominal} & \text{ability} & \text{academic} & \cdots \end{array} \\
\begin{pmatrix} \quad & & & & \\ & & & & \quad \end{pmatrix}
\end{array}
$$

From documents×terms to words×contexts

Information retrieval started with bag of terms in each document. Stopwords, stemming, tagging; TF-IDF. Dimensionality reduction reveals topics. Now rows are phrases and columns are contexts.

$$
\begin{array}{l}
 \\
\text{ambulance} \\
\text{battleship} \\
\text{bookstore} \\
\vdots
\end{array}
\begin{array}{c}
\begin{array}{ccccc} \text{abandon} & \text{abdominal} & \text{ability} & \text{academic} & \dots \end{array} \\
\begin{pmatrix}
27 & 10 & 50 & 17 & \dots \\
35 & 0 & 32 & 1 & \dots \\
5 & 0 & 6 & 33 & \dots \\
\vdots & \vdots & \vdots & \ddots &
\end{pmatrix}
\end{array}
$$

$$
= \begin{array}{l}
\text{ambulance} \\
\text{battleship} \\
\text{bookstore} \\
\vdots
\end{array}
\begin{pmatrix} \\ \\ \\ \\ \end{pmatrix}
\times
\begin{array}{c}
\begin{array}{ccccc} \text{abandon} & \text{abdominal} & \text{ability} & \text{academic} & \dots \end{array} \\
\begin{pmatrix} & & & & \\ & & & & \\ & & & & \end{pmatrix}
\end{array}
$$

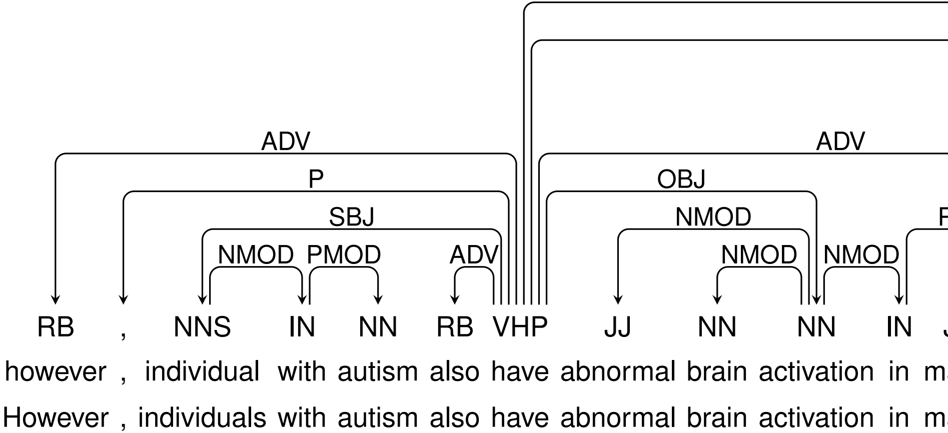
Composite phrases?

Need syntactic structure: substitution? locality? compositionality?

RB	,	NNS	IN	NN	RB	VHP	JJ	NN	NN	IN	J
however	,	individual	with	autism	also	have	abnormal	brain	activation	in	m
However	,	individuals	with	autism	also	have	abnormal	brain	activation	in	m

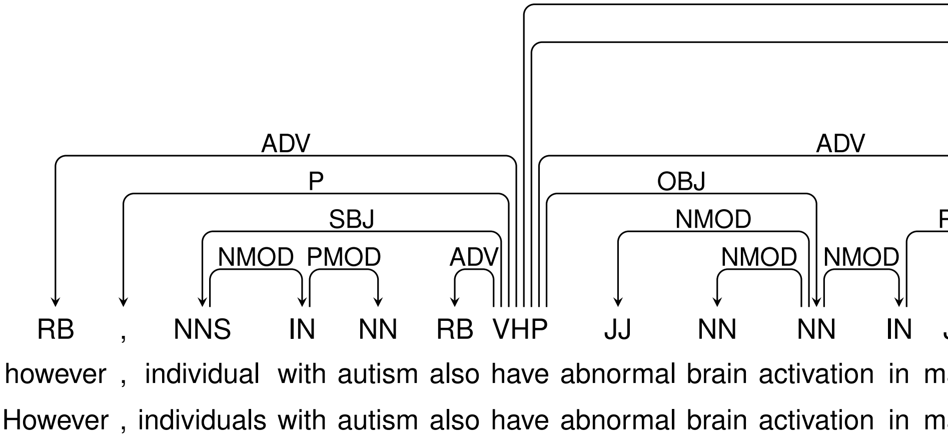
Composite phrases?

Need syntactic structure: substitution? locality? compositionality?



Composite phrases?

Need syntactic structure: substitution? locality? compositionality?



To cope with sparse data, NLP (parsing, translation, compression) applies linguistic insight (factoring, smoothing).

From language models to distributional semantics

A language model is a virtual infinite corpus:
not frequencies observed but probabilities estimated.

Let the distributional meaning of a phrase w be the probability distribution over its contexts c.

$$[\![w]\!] = \lambda c.\, \frac{\Pr(c[w])}{\sum_{c'} \Pr(c'[w])}$$

$$[\![\text{red army}]\!] = \lambda(l,r).\, \frac{\Pr(l \text{ red army } r)}{\sum_{(l',r')} \Pr(l' \text{ red army } r)}$$

$$[\![\text{red } w]\!] = \lambda(l,r).\, \frac{[\![w]\!](l \text{ red}, r)}{\sum_{(l',r')} [\![w]\!](l' \text{ red}, r')}$$

Probabilities from any model: bag of words, Markov, PCFG...
Pass the buck.

From language models to distributional semantics

A language model is a virtual infinite corpus:
not frequencies observed but probabilities estimated.

Let the distributional meaning of a phrase w be the probability distribution over its contexts c.

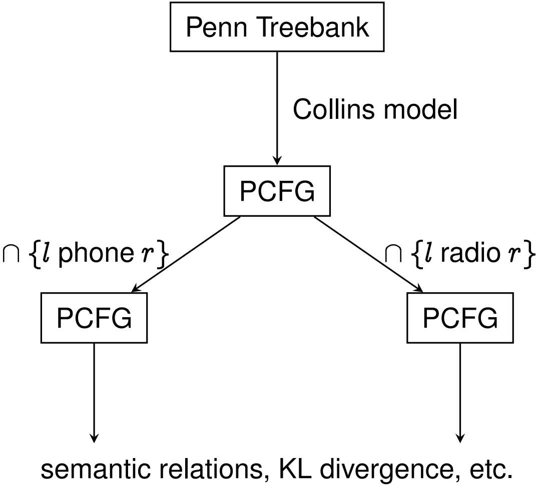
$$[\![w]\!] = \lambda c.\, \frac{\Pr(c[w])}{\sum_{c'} \Pr(c'[w])}$$

$$[\![\text{red army}]\!] = \lambda(l, r).\, \frac{\Pr(l \text{ red army } r)}{\sum_{(l',r')} \Pr(l' \text{ red army } r)}$$

$$[\![\text{red } w]\!] = \lambda(l, r).\, \frac{[\![w]\!](l \text{ red}, r)}{\sum_{(l',r')} [\![w]\!](l' \text{ red}, r')}$$

Probabilities from any model: bag of words, Markov, PCFG...
Pass the buck.

From Penn Treebank to distributional semantics



Intersection grammars reveal meaning

Random sentences

[[[[[,/, he/PRP] -LCB-/-LRB-] [the/DT company/NN]] [[[[the/DT +unknown+/NNP] +unknown+/NNP] +unknown+/NNP] compared/VBN]] [says/VBZ [:/: :/:]]]

[have/VBP [lately/RB [[in/IN [July/NNP [[[weighted/JJ large/JJ] exchange/NN] for/IN] clients/NNS]]] been/VBN]]]

was/VBD

[[,/, [Mr./NNP Bush/NNP]] said/VBD]

[plunged/VBD [4/CD ,/,]]

[[start/NN of/IN] [was/VBD [me/PRP [1.9/CD pence/NN]]]]

has/VBZ

[[,/, they/PRP] [see/VBP [aided/VBN [[in/IN [[some/DT seconds/NNS] [[of/IN [executive/NN [[[a/DT share/NN] [[yesterday/NN [[the/DT last/JJ] market/NN]] acted/VBD]] [[the/DT advance/NN] revenue/NN]]]] [a/DT share/NN]]]] [by/IN [[[the/DT pound/NN] and/CC] Exchange/NNP]]]]]]

would/MD

[[Dec./NNP 28/CD] [announced/VBD that/IN]]

[[[[,/, ,/,] ,/,] [is/VBZ [[[[the/DT only/JJ] third-quarter/JJ] net/JJ] emphasis/NN] [[very/RB executive/JJ] [[not/RB Soviet/JJ] to/TO]]]]] says/VBZ [[President/NNP Co./NNP] [Lee/NNP [chamber/NN supervisor/NN]]]]

Intersection grammars reveal meaning

Random sentences containing “car dealer”

[[[many/JJ [[[the/DT car/NN] dealer/NN] Developers/NNS]] ,/,] are/VBP]

[[[the/DT characteristic/NN] [[that/WDT agreed/VBD] in/IN]] [is/VBZ [[[[[a/DT new/JJ] federal/JJ] car/NN] dealer/NN] of/IN] [:/: [in/IN [[[their/PRP$ modern/JJ] movie/NN] [likely/JJ ,/,]]]]]]]

[says/VBZ [[[[[[[[[[[[[[[,/, [[[[the/DT crude/NN] oil/NN] man/NN] plans/NNS] car/NN]] [dealer/NN [of/IN [in/IN [[[the/DT U.S./NNP] Cambodia/NNP] Europe/NNP]]]]] But/CC] was/VBD] [anxiety/NN of/IN]] [[[John/NNP +unknown+/NNP] +unknown+/NNP] ,/,]] ,/,] [a/DT +unknown+/NNP]] In/IN] [said/VBD [[[[[,/, ,/,] bonds/NNS] Yet/RB] nonperforming/VBG] will/MD]]] [the/DT role/NN]] [[[[the/DT executive/NN] ’s/POS] son/NN] ,/,]] [the/DT computer/NN]] ,/,] [engineering/NN [[of/IN life/NN] of/IN]]] that/DT] [[+unknown+/NN +unknown+/NNP] [[The/DT head/NN] from/IN]]] goes/VBZ]]

[resigned/VBD [[to/TO [price/VB [[[[the/DT car/NN] dealer/NN] [from/IN [[Mr./NNP +unknown+/NNP] on/IN]]] [through/RP [[such/PDT a/DT] offering/NN]]]]] [showing/VBG [[[[The/DT junk/NN] defense/NN] measure/NN] [bank/NN known/VBN]]]]]

[[[[[[[still/RB [[the/DT Gardens/NNPS] life/NN]] [initial/JJ transaction/NN]] [[a/DT few/JJ] arrangement/NN]] administration/NN] [The/DT group/NN]] [[[the/DT first/NN] price/NN] of/IN]] ,/,] [is/VBZ [[[car/NN dealer/NN] [of/IN [[$/$ +unknown+/CD] [the/DT futures/NNS]]]] [was/VBD [[going/VBG [[against/IN

Intersection grammars reveal meaning

Random sentences containing “drug dealer”

[[[[[[[[[In/IN ,/,] ,/,] [[[[its/PRP$ past/JJ] five/CD] structural/JJ] +unknown+/NN] [of/IN [Allied/NNP stock/NN]]]] [+unknown+/JJ farmer/NN]] +unknown+/NNS] ,/,] [[the/DT most/JJS] [[who/WP [drag/VBP [require/VBP because/IN]]] [because/IN of/IN [[the/DT company/NN] [[a/DT year/NN] [[+unknown+/JJ cooperative/JJ] children/NNS]]]]]] [[[[this/DT +unknown+/JJ] OTC/NNP] market/NN] for/IN]] [The/DT company/NN]] [[The/DT drug/NN] dealer/NN]] [soared/VBD [,/, [reducing/VBG [,/, [,/, Monday/NNP]]]]]

[[[[[[[,/, +unknown+/NNP] [[rose/VBD [[from/IN [[[[$/$ [+unknown+/CD [million/CD [million/CD [+unknown+/CD [million/CD [billion/CD 15.6/CD]]]]]] [a/DT share/NN]] [,/, [today/NN tickets/NNS]]] [[[A/DT deep/JJ] series/NN] [[of/IN [fact/NN [[last/JJ car/NN] that/WDT]]] [with/IN looks/NNS]]] [[seven/CD cents/NNS] [a/DT share/NN]]]]] [on/IN [[[The/DT +unknown+/JJ] ownership/NN] or/CC] yesterday/NN]]]] [[[[another/DT year/NN] price/NN] [above/IN –/:]] [was/VBD [[against/IN him/PRP] although/IN]]] [but/CC [[[[the/DT following/JJ] week/NN] [was/VBD [[[[[[[a/DT specific/JJ] +unknown+/JJ] short-term/JJ] Treasury/NNP] [economic/JJ trade/NN]] [[the/DT FDA/NNP] ,/,]] ,/,] [marks/NNS [[[[the/DT coming/VBG] early/JJ] next/JJ] year/NN] little/RB]]] ,/,]]] [and/CC [and/CC [,/, and/CC]]]]]]] ,/,] [+unknown+/NNP +unknown+/NNS]] [has/VBZ n’t/RB]] [received/VBD [[[[a/DT serious/JJ] drag/NN] on/IN] [now/RB [[when/WRB [[[[the/DT drug/NN] dealer/NN] pay/VB] [[[Sun/NNP Jeep/NNP] Stoll/NNP]

Intersection grammars reveal meaning

Random sentences containing “card dealer”

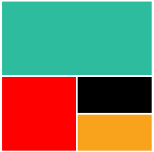
[[[,/, [[[a/DT newspaper/NN] base/NN] of/IN]] [the/DT floor/NN]] [was/VBD [ago/RB [[breaking/JJ trend/NN] [sharply/RB [[[[[a/DT formal/JJ] brokerage/NN] and/CC] couple/NN] [[[[[[[John/NNP Agency/NNP] He/PRP] [July/NNP [where/WRB when/WRB]]] he/PRP] [can/MD [[close/VB [at/IN [[until/IN [should/MD [has/VBZ [causes/VBZ [the/DT world/NN]]]]] once/RB]]] even/RB]]] [they/PRP ['ve/VBP [[+unknown+/VBN [on/IN [[the/DT University/NNP] in/IN]]] [[got/VBN [[in/IN [[last/JJ week/NN] [[a/DT philosophy/NN] [[the/DT Congress/NNP] [[The/DT +unknown+/NN] to/TO]]]]] [his/PRP$ toll/NN]]] [[already/RB traded/VBN] [got/VBN [grown/VBN [,/, [[based/VBN [[on/IN [the/DT amount/NN]] with/IN]] [by/IN [[according/VBG [[to/TO [[[[Prudential/NNP Committee/NNP] 's/POS] media/NNS] [[[[a/DT visible/JJ] program/NN] trading/NN] arena/NN]]] [[to/TO [[chief/JJ big/JJ] [[[Bear/NNP II/NNP] official/NN] [[the/DT disaster/NN] [[a/DT card/NN] dealer/NN]]]]] [to/TO [[[a/DT [[[10/CD 6.9/CD] 34/CD] %/NN]] secretary/NN] [of/IN [of/IN [area/NN [bought/VBD [[[the/DT +unknown+/JJ] shareholders/NNS] according/VBG]]]]]]]]] [in/IN [+unknown+/NN of/IN]]]]]]]]]]]] said/VBD] [[of/IN [no/DT technology/NN]] [in/IN [that/WDT for/IN]]]]] [Using/VBG [[[[another/DT leveraged/VBN] +unknown+/NN] concerned/JJ] Robert/NNP]]]]]]]

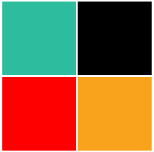
[is/VBZ [[[[[the/DT New/NNP] York/NNP] +unknown+/NNP] Co./NNP] [of/IN [[ABC/NNP television/NN] negotiations/NNS]]] [[[[[[[[the/DT Soviet/JJ] real/JJ] Labor/NNP] and/CC] +unknown+/NN] James/NNP] and/CC] potential/NN]

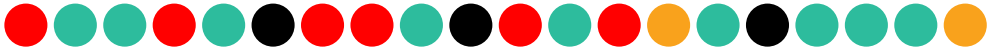
Kullback-Leibler divergence

$$D_{\mathrm{KL}}(P \,\|\, Q) \;=\; \overbrace{\sum_x P(x) \log \frac{1}{Q(x)}}^{\text{cross entropy}} \;-\; \overbrace{\sum_x P(x) \log \frac{1}{P(x)}}^{\text{entropy}}$$

Example

$P =$ 

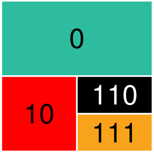
$Q =$ 

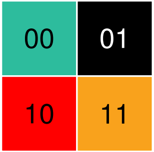
20 samples from P: 

Kullback-Leibler divergence

$$D_{\mathrm{KL}}(P \,\|\, Q) \;=\; \overbrace{\sum_x P(x)\log\frac{1}{Q(x)}}^{\text{cross entropy}} \;-\; \overbrace{\sum_x P(x)\log\frac{1}{P(x)}}^{\text{entropy}}$$

Example

$P =$ 

$Q =$ 

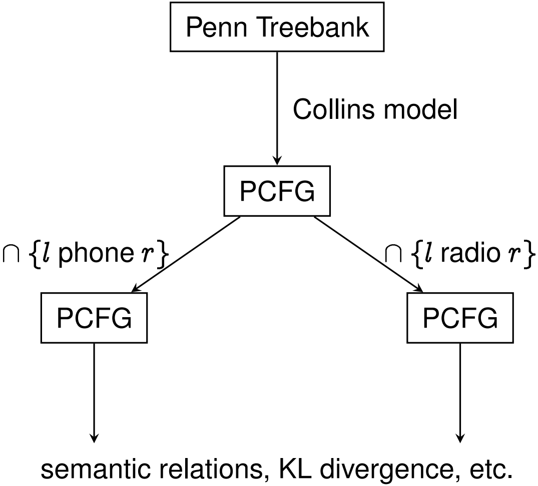
20 samples from P: 

encoded for P: 100010011010100110100101110110000111

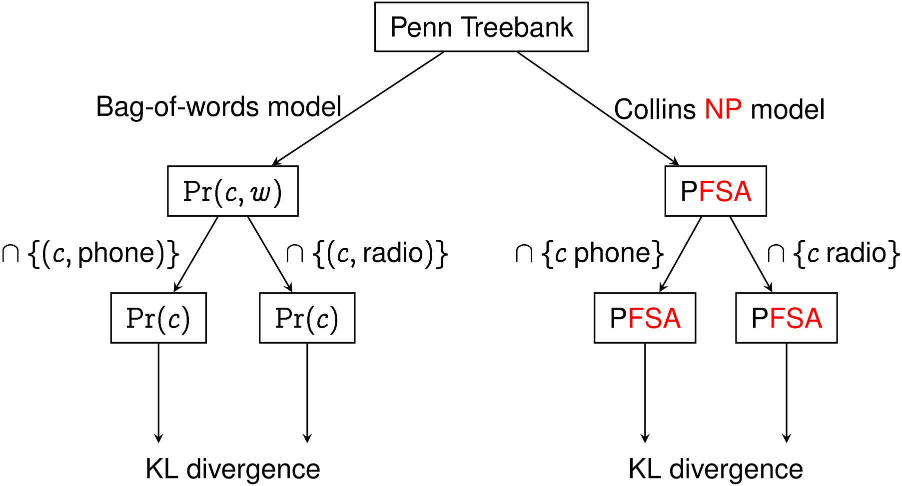
encoded for Q: 1000001000011010000110001011000100000011

KL divergence: 0.25 bits $= 2.00$ bits $- 1.75$bits

From Penn Treebank to distributional semantics

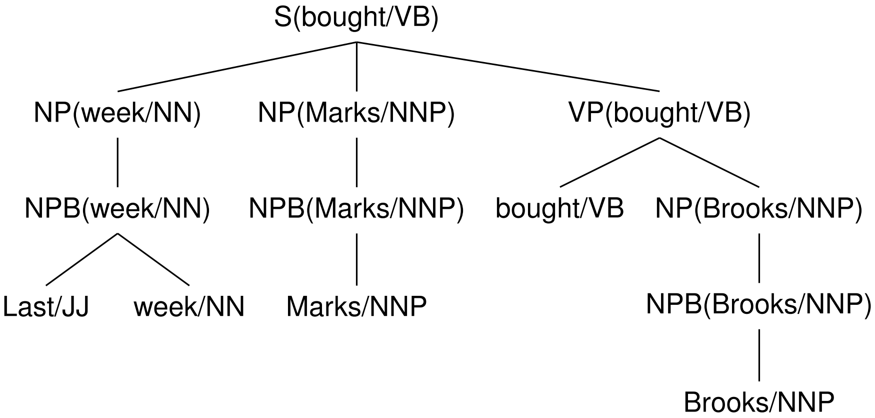


From Penn Treebank to distributional semantics



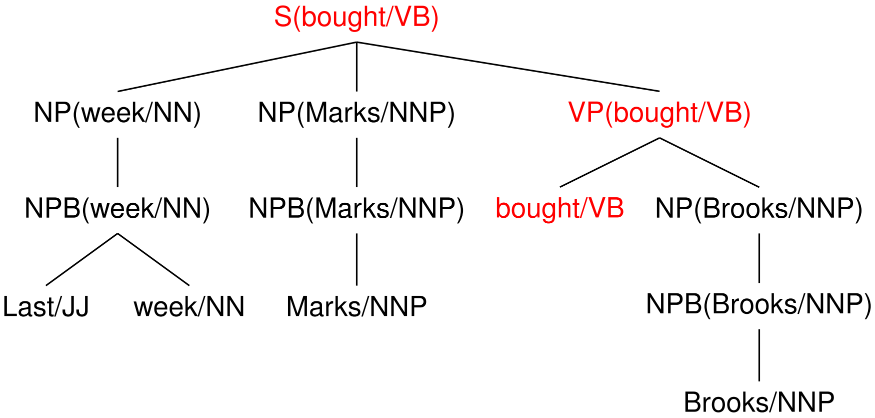
Collins model

Lexicalized PCFG for parsing (1997)
Not for generation (Post & Gildea 2008)
Bikel (2004) exegesis



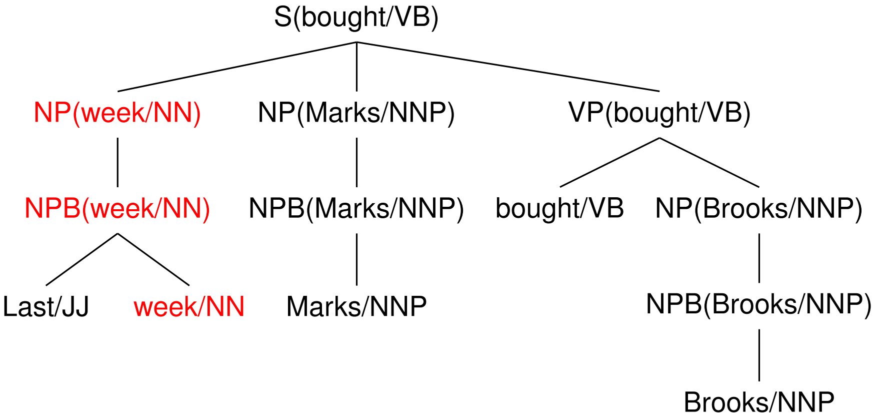
Collins model

Lexicalized PCFG for parsing (1997)
Not for generation (Post & Gildea 2008)
Bikel (2004) exegesis



Collins model

Lexicalized PCFG for parsing (1997)
Not for generation (Post & Gildea 2008)
Bikel (2004) exegesis



Summary statistics

Standard English training set: Wall Street Journal §§02–21

- 39 832 sentences
- 950 028 word tokens
  44 113 unique words
  10 437 unique words that occur 6+ times
- 28 basic nonterminal labels
  42 parts of speech

Tiny for a corpus today.

Simplified Collins Model 1

- 575 936 nonterminals
  15 564 terminals
  12 611 676 rules

Big for a grammar today.

Pilot evaluation using BLESS data set

Concept	Relation	Relatum
phone	coord	computer
phone	coord	radio
phone	coord	stereo
phone	coord	television
phone	hyper	commodity
phone	hyper	device
phone	hyper	equipment
phone	hyper	good
phone	hyper	object
phone	hyper	system
phone	mero	cable
phone	mero	dial
phone	mero	number
phone	mero	plastic
phone	mero	wire
phone	random-n	choice
phone	random-n	clearance
phone	random-n	closing
phone	random-n	entrepreneur

Baroni and Lenci Evaluation of Semantic Spaces (2011)

Only head nouns observed in corpus:

NP(phone/NN)

NPB(phone/NN) →

← phone/NN

Compute KL divergences among distributions over *modifier-nonterminal sequences*

Pilot evaluation using BLESS data set

Concept	Relation	Relatum
phone	coord	computer
phone	coord	radio
phone	coord	stereo
phone	coord	television
phone	hyper	commodity
phone	hyper	device
phone	hyper	equipment
phone	hyper	good
phone	hyper	object
phone	hyper	system
phone	mero	cable
phone	mero	dial
phone	mero	number
phone	mero	plastic
phone	mero	wire
phone	random-n	choice
phone	random-n	clearance
phone	random-n	closing
phone	random-n	entrepreneur

Baroni and Lenci Evaluation of Semantic Spaces (2011)

Only head nouns observed in corpus:

NP(phone/NN)

NPB(phone/NN) $\longrightarrow$

$\longleftarrow$ phone/NN

Compute KL divergences among distributions over *modifier-nonterminal sequences*

Pilot evaluation using BLESS data set

Concept	Relation	Relatum
phone	coord	computer
phone	coord	radio
phone	coord	stereo
phone	coord	television
phone	hyper	commodity
phone	hyper	device
phone	hyper	equipment
phone	hyper	good
phone	hyper	object
phone	hyper	system
phone	mero	cable
phone	mero	dial
phone	mero	number
phone	mero	plastic
phone	mero	wire
phone	random-n	choice
phone	random-n	clearance
phone	random-n	closing
phone	random-n	entrepreneur

Baroni and Lenci Evaluation of Semantic Spaces (2011)

Only head nouns observed in corpus:

NP(phone/NN)

NPB(phone/NN) ⟶

⟵ phone/NN

Compute KL divergences among distributions over *modifier-nonterminal sequences*

Pilot evaluation using BLESS data set

38 Concept	Relation	687 Relatum
phone	173 coord	computer
phone	coord	radio
phone	coord	stereo
phone	coord	television
phone	125 hyper	commodity
phone	hyper	device
phone	hyper	equipment
phone	hyper	good
phone	hyper	object
phone	hyper	system
phone	490 mero	cable
phone	mero	dial
phone	mero	number
phone	mero	plastic
phone	mero	wire
phone	561 random-n	choice
phone	random-n	clearance
phone	random-n	closing
phone	random-n	entrepreneur

Baroni and Lenci Evaluation of Semantic Spaces (2011)

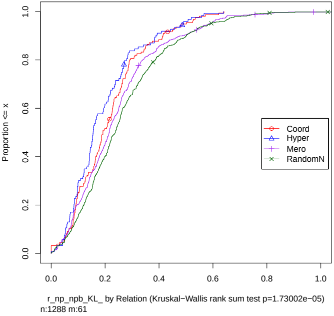
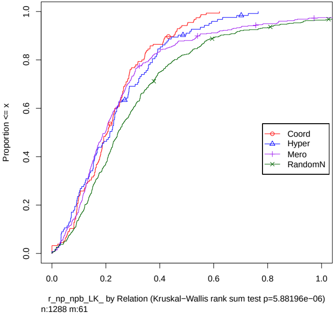
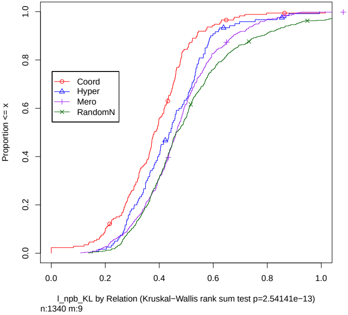
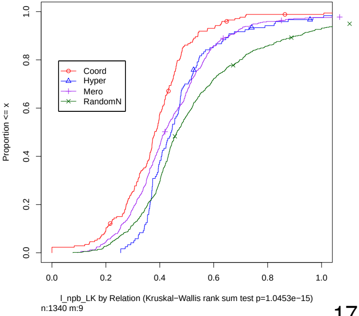
Only head nouns observed in corpus:

NP(phone/NN)

NPB(phone/NN) ⟶

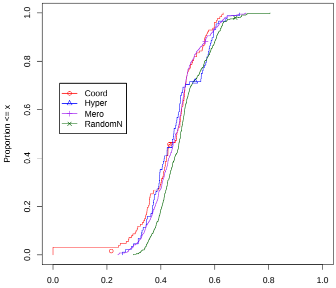
⟵ phone/NN

Compute KL divergences among distributions over *modifier-nonterminal sequences*

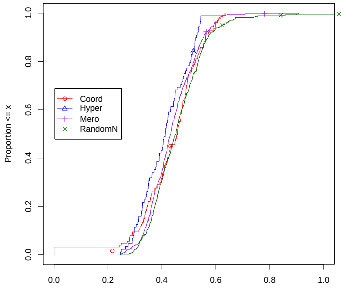
	$D_{\mathrm{KL}}(\text{Concept} \,||\, \text{Relatum})$	$D_{\mathrm{KL}}(\text{Relatum} \,||\, \text{Concept})$
NP | NPB →		
NPB | ← NNs		

$D_{\mathrm{KL}}(\text{Concept} \,\|\, \text{Relatum})$ $D_{\mathrm{KL}}(\text{Relatum} \,\|\, \text{Concept})$

Bag of words



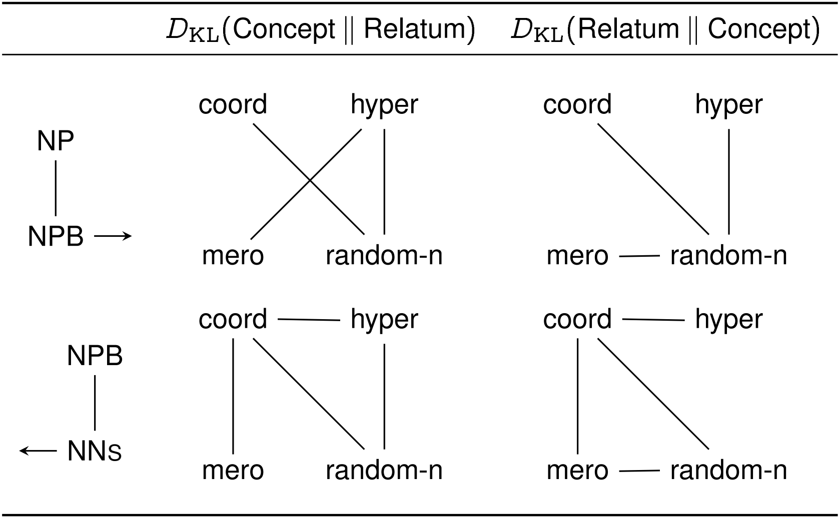
BagOfWords_KL by Relation (Kruskal−Wallis rank sum test p=9.20412e−05)
n:1048 m:0



BagOfWords_LK by Relation (Kruskal−Wallis rank sum test p=0.000968842)
n:1048 m:0

Mann-Whitney-Wilcoxon rank sum test

Edges indicate $p < .01$



Mann-Whitney-Wilcoxon rank sum test

Edges indicate $p < .01$

	$D_{\mathrm{KL}}(\text{Concept} \parallel \text{Relatum})$		$D_{\mathrm{KL}}(\text{Relatum} \parallel \text{Concept})$	
	coord	hyper	coord	hyper
Bag of words				
	mero	random-n	mero	random-n

Summary

Distributional semantics from language models

- Estimate felicity *in context* from observed use
- Cope with sparse data using linguistic insight such as syntax

Better distributional semantics from better language models?

- Pilot test on the Penn Treebank using two language models
- Further tests need more computation techniques, resources

Thanks!

- Bolzano: European Masters Program in Language and Communication Technologies
- Trento: Marco Baroni, Raffaella Bernardi, Roberto Zamparelli
- Rutgers: Jason Perry, Matthew Stone
- Cornell: John Hale, Mats Rooth