

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

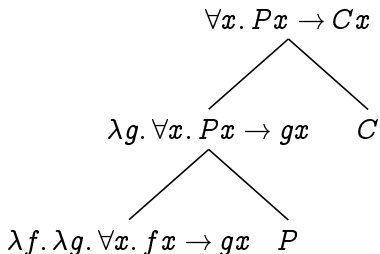
“In order to say what a meaning *is*,
we may first ask what a meaning *does*,
and then find something that does that.” —David Lewis

Truth, entailment

Every person cried. \models Every professor cried.

A person cried. $\not\models$ A professor cried.

Formal semantics



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

ambulance \sim battleship

ambulance \approx bookstore

Distributional semantics

	abandon	abdominal	ability	academic	accept	...
ambulance	27	10	50	17	130	...
battleship	35	0	32	1	25	...
bookstore	5	0	6	33	13	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Distributional semantics for entailment among words

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

“pointwise mutual information”

Distributional semantics for entailment among words

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

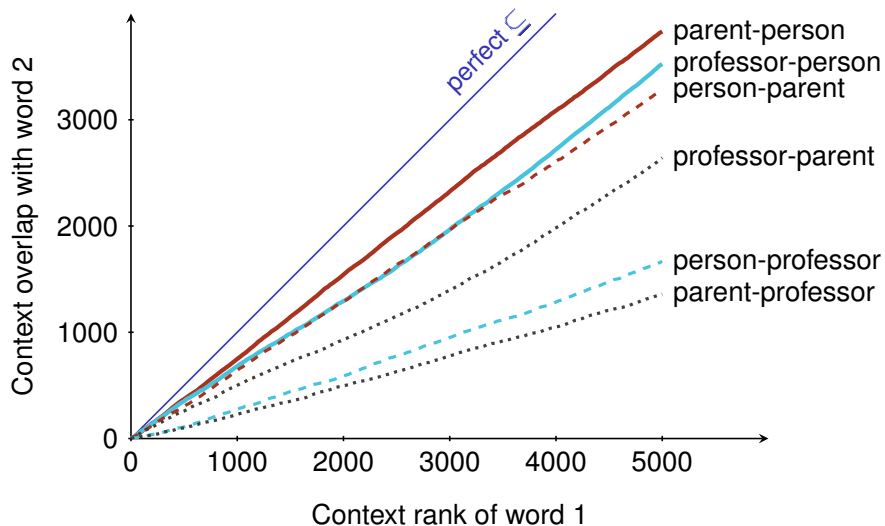
“pointwise mutual information”

parent argcount_n arglist_n arglist_j phane_n specity_n qdisc_n carthy_n
parents-to-be_n non-resident_j step-parent_n tc_n ballons_n
eliza_n symptons_n adoptive_j stepparent_n nonresident_j
home-school_n scabrid_n petiolule_n ...

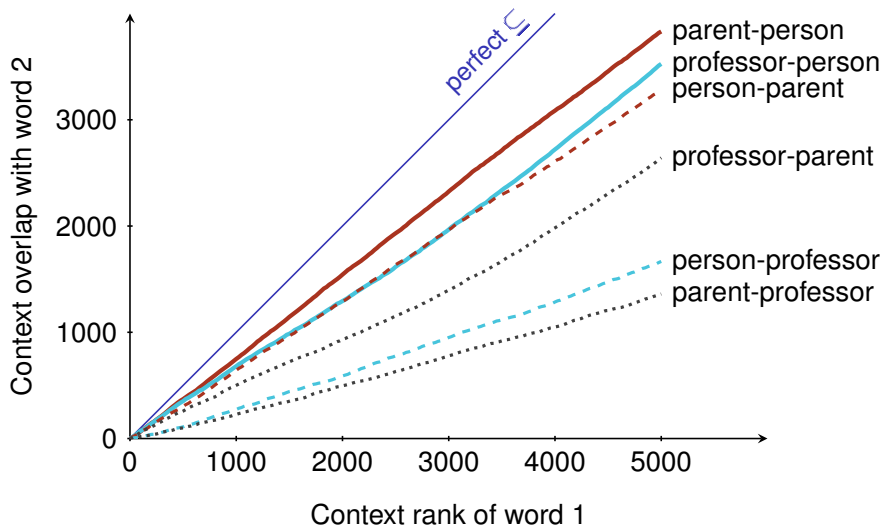
person anglia_n first-mentioned_j unascertained_j enure_v
deposit-taking_j bonis_n iconclass_j cotswolds_n aforesaid_n
haver_v foresaid_j gha_n sub-paragraphs_n enacted_j geest_j
non-medicinal_j sub-paragraph_n intimation_n arrestment_n
incumbrance_n ...

professor william_n extraordinarius_n ordinarius_n francis_n reid_n
emeritus_n emeritus_j derwent_n regius_n laurence_n edward_n
carisoprodol_n adjunct_j winston_n privatdozent_j edward_j
xanax_n tenure_v cialis_n florence_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 \times terms to words \times contexts

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

	abandon	abdominal	ability	academic	...
	27	10	50	17	...
	35	0	32	1	...
	5	0	6	33	...
⋮	⋮	⋮	⋮	⋮	⋮

From documents \times terms to words \times contexts

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

$$\begin{array}{c} \text{document 1} \\ \text{document 2} \\ \text{document 3} \\ \vdots \end{array} \begin{pmatrix} \text{abandon} & \text{abdominal} & \text{ability} & \text{academic} & \dots \\ 27 & 10 & 50 & 17 & \dots \\ 35 & 0 & 32 & 1 & \dots \\ 5 & 0 & 6 & 33 & \dots \\ \vdots & \vdots & \vdots & \ddots & \end{pmatrix}$$
$$= \begin{array}{c} \text{document 1} \\ \text{document 2} \\ \text{document 3} \\ \vdots \end{array} \begin{pmatrix} \phantom{\text{abandon}} \\ \phantom{\text{abdominal}} \\ \phantom{\text{ability}} \\ \phantom{\text{academic}} \\ \end{pmatrix} \times \begin{pmatrix} \text{abandon} & \text{abdominal} & \text{ability} & \text{academic} & \dots \\ \phantom{\text{abandon}} & \phantom{\text{abdominal}} & \phantom{\text{ability}} & \phantom{\text{academic}} & \\ \phantom{\text{abandon}} & \phantom{\text{abdominal}} & \phantom{\text{ability}} & \phantom{\text{academic}} & \\ \phantom{\text{abandon}} & \phantom{\text{abdominal}} & \phantom{\text{ability}} & \phantom{\text{academic}} & \\ \phantom{\text{abandon}} & \phantom{\text{abdominal}} & \phantom{\text{ability}} & \phantom{\text{academic}} & \end{pmatrix}$$

From documents \times terms to words \times 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.**

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

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

Composite phrases?

Need syntactic structure: substitution? locality? compositionality?

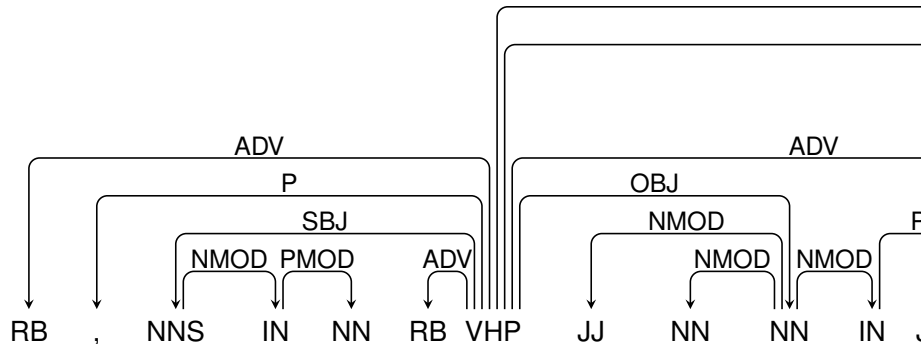
RB , NNS IN NN RB VHP JJ NN NN IN ,

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?

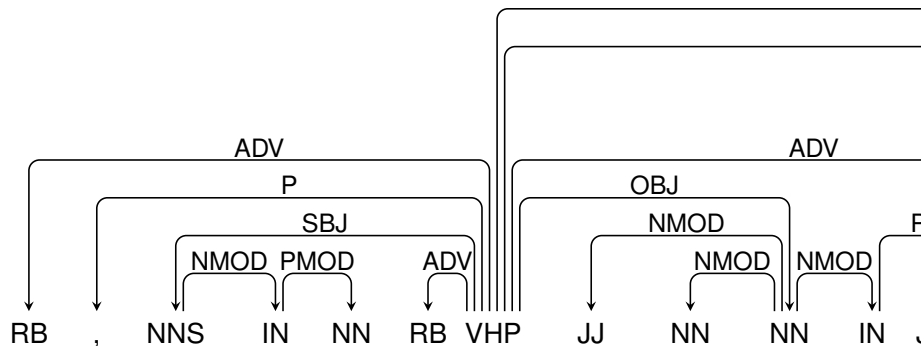


however, individual with autism also have abnormal brain activation in m

However, individuals with autism also have abnormal brain activation in m

Composite phrases?

Need syntactic structure: substitution? locality? compositionality?



however , individual with autism also have abnormal brain activation in m...

However , individuals with autism also have abnormal brain activation in m...

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 .

$$\llbracket w \rrbracket = \lambda c. \frac{\text{Pr}(c[w])}{\sum_{c'} \text{Pr}(c'[w])}$$

$$\llbracket \text{red army} \rrbracket = \lambda(l, r). \frac{\text{Pr}(l \text{ red army } r)}{\sum_{(l', r')} \text{Pr}(l' \text{ red army } r')}$$

$$\llbracket \text{red } w \rrbracket = \lambda(l, r). \frac{\llbracket w \rrbracket(l \text{ red}, r)}{\sum_{(l', r')} \llbracket w \rrbracket(l' \text{ red}, r')}$$

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 .

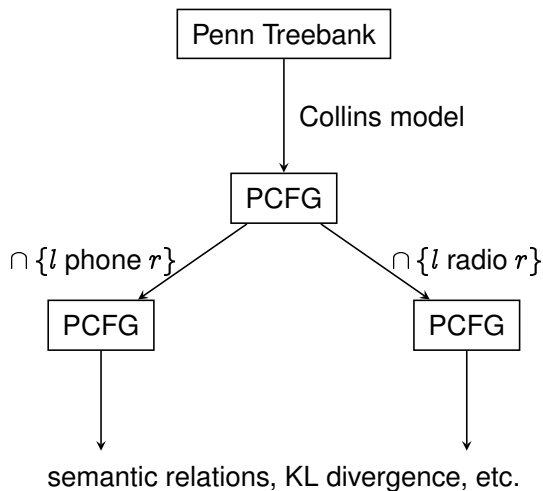
$$\llbracket w \rrbracket = \lambda c. \frac{\text{Pr}(c[w])}{\sum_{c'} \text{Pr}(c'[w])}$$

$$\llbracket \text{red army} \rrbracket = \lambda(l, r). \frac{\text{Pr}(l \text{ red army } r)}{\sum_{(l', r')} \text{Pr}(l' \text{ red army } r')}$$

$$\llbracket \text{red } w \rrbracket = \lambda(l, r). \frac{\llbracket w \rrbracket(l \text{ red}, r)}{\sum_{(l', r')} \llbracket w \rrbracket(l' \text{ red}, r')}$$

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]
[entered/VBD [[MCI/NNP] +unknown+/NNP] U.S./NNP]]] hurt/VB]]]]] [[the/DT

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

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

Example



20 samples from P : ●

Kullback-Leibler divergence

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

Example



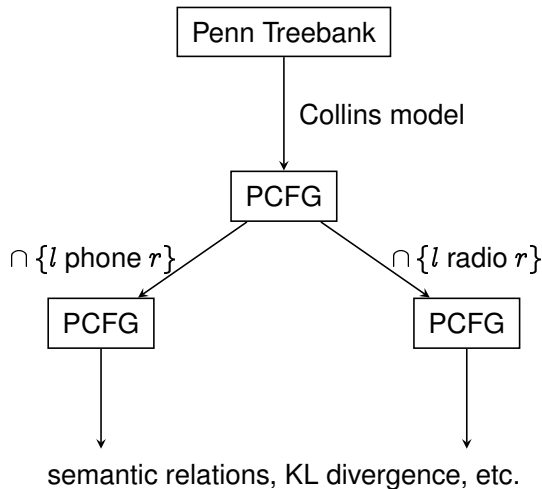
20 samples from P : ●●●●●●●●●●●●●●●●●●

encoded for P : 100010011010100110100101110110000111

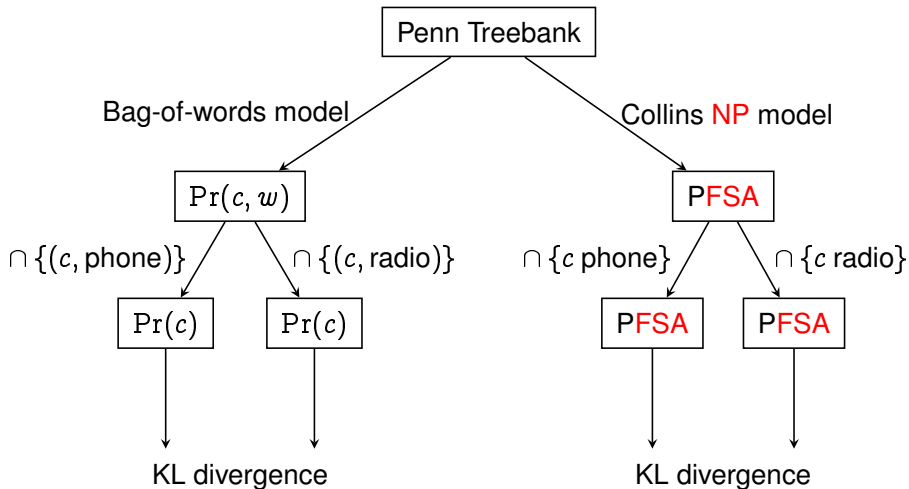
encoded for Q : 1000001000011010000110001011000100000011

KL divergence: 0.25 bits = 2.00 bits – 1.75bits

From Penn Treebank to distributional semantics



From Penn Treebank to distributional semantics

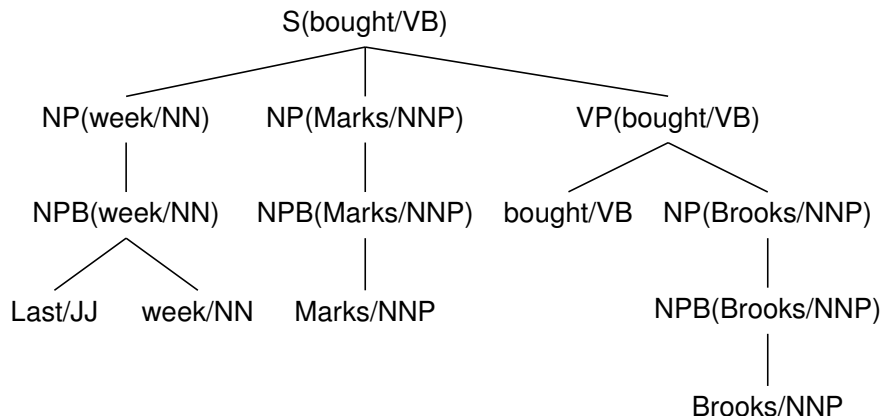


Collins model

Lexicalized PCFG for parsing (1997)

Not for generation (Post & Gildea 2008)

Bikel (2004) exegesis

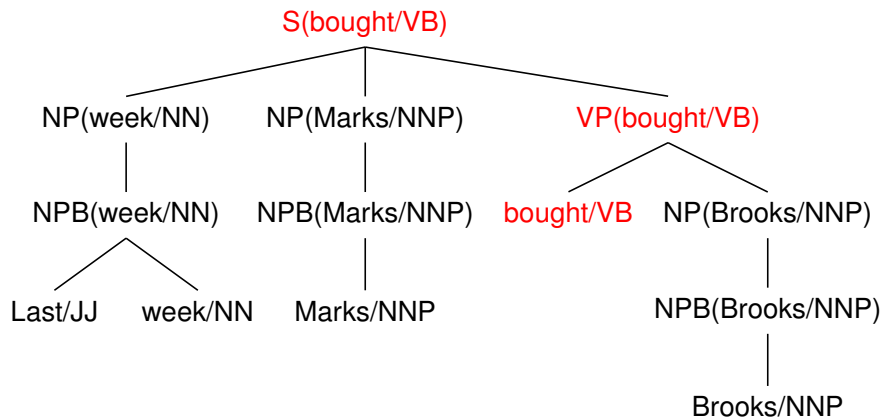


Collins model

Lexicalized PCFG for parsing (1997)

Not for generation (Post & Gildea 2008)

Bikel (2004) exegesis

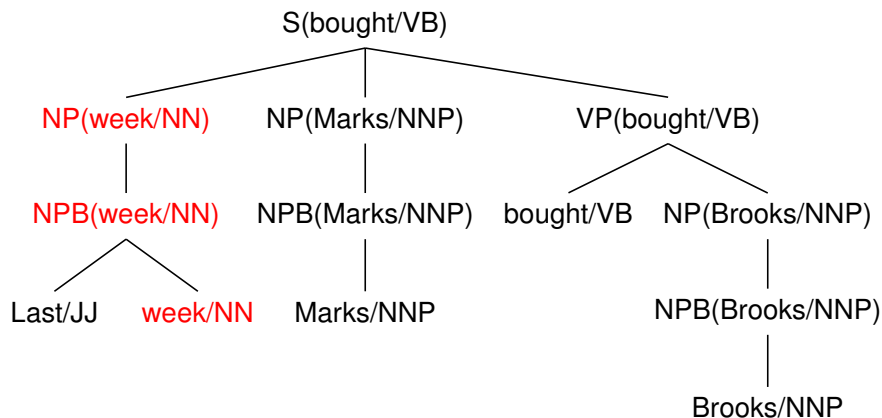


Collins model

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

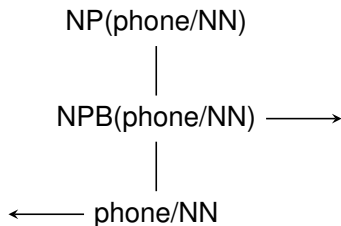
Big for a grammar today.

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:



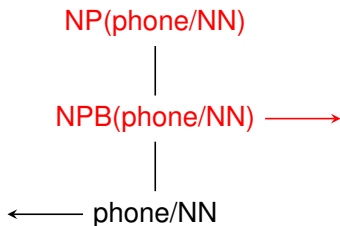
Compute KL divergences among distributions over *modifier-nonterminal sequences*

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:



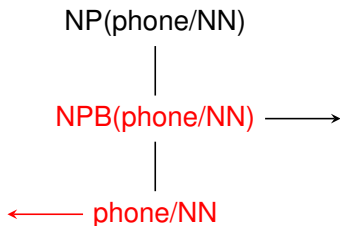
Compute KL divergences among distributions over *modifier-nonterminal sequences*

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:



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

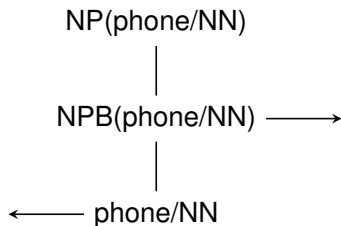
Pilot evaluation using BLESS data set

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)

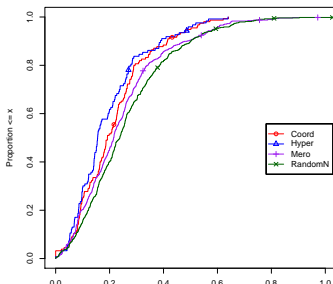
Only head nouns observed in corpus:



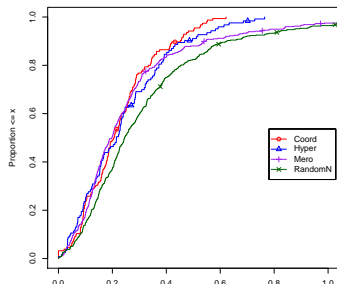
Compute KL divergences among distributions over *modifier-nonterminal sequences*

$D_{KL}(\text{Concept} \parallel \text{Relatum})$ $D_{KL}(\text{Relatum} \parallel \text{Concept})$

NP
 |
 NPB →

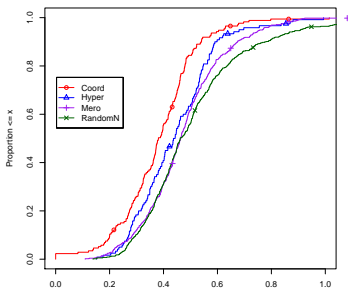


r_np_npb_KL_by Relation (Kruskal-Wallis rank sum test $p=1.73002e-05$)
 n:1288 m:61

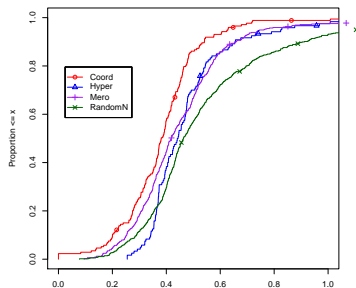


r_np_npb_LK_by Relation (Kruskal-Wallis rank sum test $p=5.88196e-06$)
 n:1288 m:61

NPB
 |
 ← NNS



l_npb_KL_by Relation (Kruskal-Wallis rank sum test $p=2.54141e-13$)
 n:1340 m:9

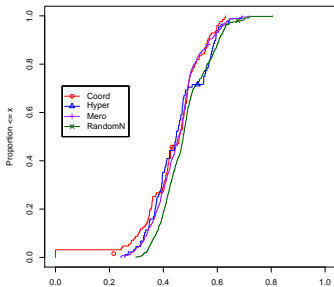


l_npb_LK_by Relation (Kruskal-Wallis rank sum test $p=1.0453e-15$)
 n:1340 m:9

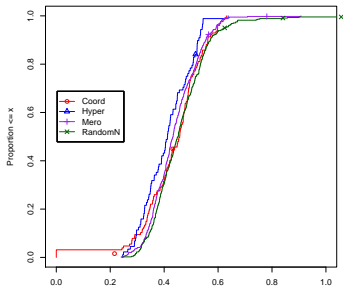
$D_{KL}(\text{Concept} \parallel \text{Relatum})$

$D_{KL}(\text{Relatum} \parallel \text{Concept})$

Bag of words



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

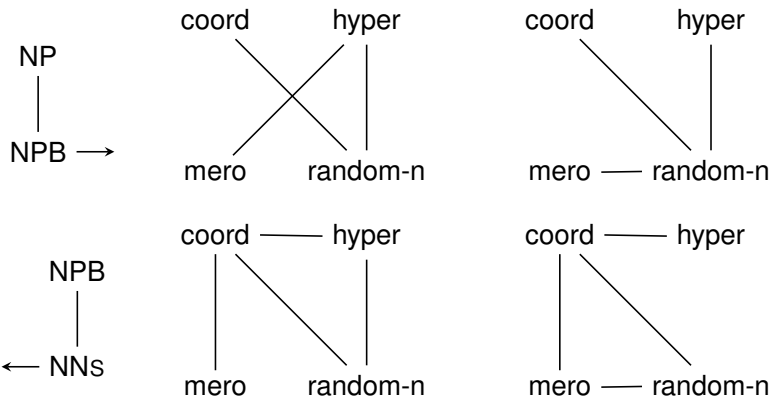


BagOfWords_LK by Relation (Kruskal-Wallis rank sum test $p=0.000968842$)
n:1048 m:0

Mann-Whitney-Wilcoxon rank sum test

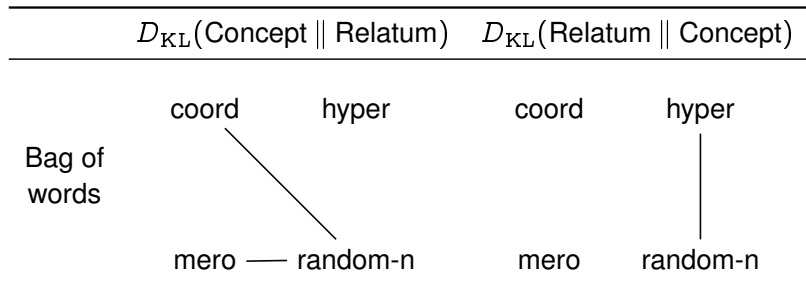
Edges indicate $p < .01$

$D_{KL}(\text{Concept} \parallel \text{Relatum})$ $D_{KL}(\text{Relatum} \parallel \text{Concept})$



Mann-Whitney-Wilcoxon rank sum test

Edges indicate $p < .01$



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