Unsupervised Frame Learning from Text

Brendan O'Connor

March 8, 2012

Data Analysis Project, Machine Learning Department

DAP committee: Noah Smith, Geoff Gordon, Jaime Carbonell

- Introduction: Frame and Scripts
- Models: Unsupervised Learning of Frames
- Datasets
- Experiments

Scripts

[Schank and Abelson, 1977]

John went to Lundy's. He ordered lobster.
He paid the check and left.

KS	CD
John **$RESTAURANT**	S1 John **PTRANS** John **to** restuarant
Customer = John	⇐E
Food = lobster	S1 John **MTRANS** | waitress **ATRANS** lobster **to** John | **to** waitress
Name = Lundy's	⇐rE
Location = Brooklyn	S1 John **INGEST** lobster **to** stomach
	⇐rE
	S1 John **ATRANS** money **to** management
	⇐rE
	S1 John **PTRANS** John from restaurant

Thus an entire story spanning many script and non-script-like events would be represented as a linked causal chain of Conceptual Dependency conceptualizations, some subset of which would be linked via the Script link to the scriptname that governs it at the Knowledge Structure level.

MUC (Message Understanding Conference)

The terrorists ***used*** *explosives against the town hall. El Comercio reported that alleged Shining Path members also* ***attacked*** *public facilities in huarpacha, Ambo, tomayquichua, and kichki. Municipal official Sergio Horna was seriously* ***wounded*** *in an explosion in Ambo.*

The entities from this document fill the following slots in a MUC-4 bombing template.

Perp: Shining Path members **Victim**: Sergio Horna
Target: public facilities **Instrument**: explosives

Learning the templates

Chambers and Jurafsky (2011)

- Unsupervised learning of event/role templates
  - Chambers and Jurafsky 2011
  - Uses ad-hoc clustering cascade

Kidnap Template (MUC-4)

Perpetrator *Person/Org* who releases, abducts, kidnaps, ambushes, holds, forces, captures, is imprisoned, frees

Target *Person/Org* who is kidnapped, is released, is freed, escapes, disappears, travels, is harmed, is threatened

Police *Person/Org* who rules out, negotiates, condemns, is pressured, finds, arrests, combs

Weapons Smuggling Template (NEW)

Perpetrator *Person/Org* who smuggles, is seized from, is captured, is detained

Police *Person/Org* who raids, seizes, captures, confiscates, detains, investigates

Instrument A *physical object* that is smuggled, is seized, is confiscated, is transported

Frame Semantics

Charles J. Fillmore (1982)

University of California, Berkeley

By the term 'frame' I have in mind any system of concepts related in such a way that to understand any one of them you have to understand the whole structure in which it fits; when one of the things in such a structure is introduced into a text, or into a conversation, all of the others are automatically made available. I intend the word 'frame' as used here to be a general cover term for the set of concepts variously known, in the literature on natural language understanding, as 'schema', 'script', 'scenario', 'ideational scaffolding', 'cognitive model', or 'folk theory'.[1]

Frame Semantics

- BLAME, ACCUSE, CRITICIZE
  - Judger
  - Defendant

The details of my description have been 'criticized' (see esp. McCawley 1975), but the point remains that we have here not just a group of individual words, but a 'domain' of vocabulary whose elements somehow presuppose a schematization of human judgment and behavior involving notions of worth, responsibility, judgment, etc., such that one would want to say that nobody can really understand the meanings of the words in that domain who does not understand the social institutions or the structures of experience which they presuppose.

FrameNet

https://framenet2.icsi.berkeley.edu/fnReports/data/frameIndex.xml?frame=Crime_scenario

Frame Index

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Abandonment
Abounding_with
Absorb_heat
Abundance
Abusing
Access_scenario
Accompaniment
Accomplishment
Accoutrements
Accuracy
Achieving_first
Active_substance
Activity
Activity_abandoned_state
Activity_done_state
Activity_finish
Activity_ongoing
Activity_pause
Activity_paused_state
Activity_prepare
Activity_ready_state

Crime_scenario

Lexical Unit Index

Definition:

A (putative) Crime is committed and comes to the attention of the Authorities. In response, there is a Criminal_investigation and (often) Arrest and criminal court proceedings. The Investigation, Arrest, and other parts of the Criminal_Process are pursued in order to find a Suspect (who then may enter the Criminal_process to become the Defendant) and determine if this Suspect matches the Perpetrator of the Crime, and also to determine if the Charges match the Crime. If the Suspect is deemed to have committed the Crime, then they are generally given some punishment commensurate with the Charges.

Semantic Type: Non-Lexical Frame

FEs:

Core:

Authorities []	The group which is responsible for the maintenance of law and order, and as such have been given the power to investigate Crimes, find Suspects and determine if a Suspect should be submitted to the Criminal_process.
Charge []	A description of a type of act that is not permissable according to the law of society.
Crime []	An act, generally intentional, that matches the description that belongs to an official Charge.
Perpetrator [] **Semantic Type:** Sentient	The individual that commits a Crime.
Suspect []	The individual which is under suspicion of having committed the Crime.

Frame theories

linguistics

Fillmore 1964
The Case for Case

Fillmore 1982
Frame Semantics

artificial intelligence

Minsky 1974
A Framework for Representing Knowledge

Rumelhart 1978
Schemata: The Building Blocks of Cognition

Schank and Abelson 1977
Scripts, Plans, Goals, Understanding

Datasets

FrameNet
VerbNet
PropBank

OntoNotes
GENIA

MUC
ACE

(~Supervised) Tasks

"Semantic Role Labeling"

"Template-Filling Information Extraction"

Slide made with Dipanjan Das

this is a horribly reductionist diagram, but there is a genuine bit of separation in these literatures. linguistics and AI are different areas. what we've been talking about with the semantic roles and such basically derives from Fillmore's classic theory of Case Grammar, with lots of other work by others through the years (Jackendoff, Levin, others i'm forgetting). the theories are nice, but to make it concrete you need to make datasets that computers can read. in this vein, ones you may have heard of include framenet, verbnet, propbank, and current work is on ontonotes. Then for any of these, you can analyze text and label it with its lexicon and labels. this is a structured prediction task, and it's called semantic role labeling.

but there's another theoretical tradition too -- frames, or sometimes called scripts. again lots of people working on this but one of the big names is roger schank; schank and abelson 1977 is the main book on it. i'll argue that it eventually evolved into what we now call "template-filling information extraction.", typified by the MUC competition and datasets. also ACE, and also the biomed IE corpus GENIA, though i think that one became more broad over the years.

anyways, the SRL and template-filling IE tasks are, as structured prediction problems, extremely similar. when you read the literature there are funny holes and stuff because people in different research communities tend to publish about different ones. however recent work has merged these strands more and more; both ontonotes and genia have multilevel annotations from syntactic to more semantic labels.

Frame theories

<u>linguistics</u>

Fillmore 1964
The Case for Case

Fillmore 1982
Frame Semantics

<u>artificial intelligence</u>

Minsky 1974
A Framework for Representing Knowledge

Rumelhart 1978
Schemata: The Building Blocks of Cognition

Schank and Abelson 1977
Scripts, Plans, Goals, Understanding

> Goal:
> Learn the frames:
> coherent sets of
> *actions*, *actors*, and *objects*

Slide made with Dipanjan Das

this is a horribly reductionist diagram, but there is a genuine bit of separation in these literatures. linguistics and AI are different areas. what we've been talking about with the semantic roles and such basically derives from Fillmore's classic theory of Case Grammar, with lots of other work by others through the years (Jackendoff, Levin, others i'm forgetting). the theories are nice, but to make it concrete you need to make datasets that computers can read. in this vein, ones you may have heard of include framenet, verbnet, propbank, and current work is on ontonotes. Then for any of these, you can analyze text and label it with its lexicon and labels. this is a structured prediction task, and it's called semantic role labeling.

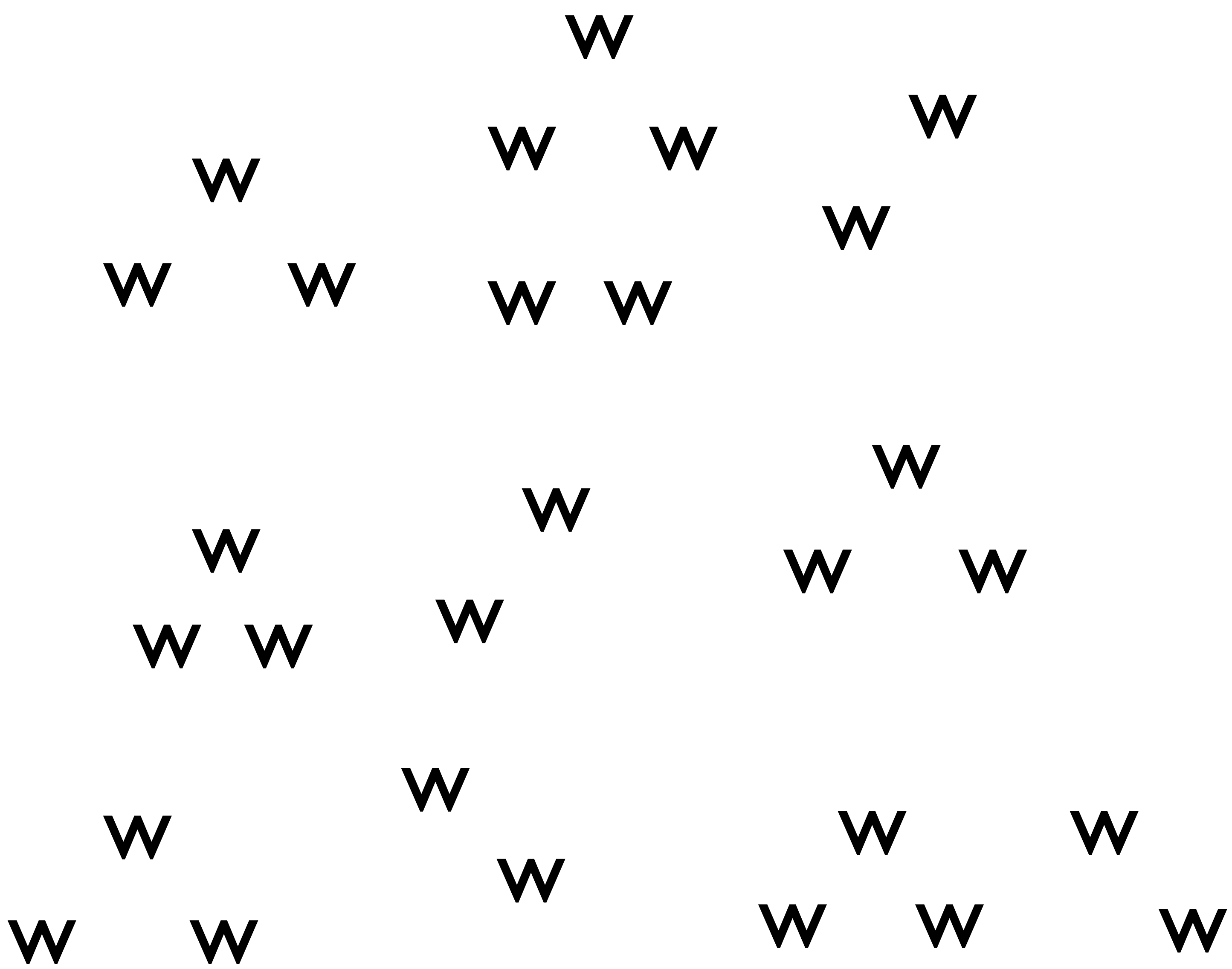
but there's another theoretical tradition too -- frames, or sometimes called scripts. again lots of people working on this but one of the big names is roger schank; schank and abelson 1977 is the main book on it. i'll argue that it eventually evolved into what we now call "template-filling information extraction.", typified by the MUC competition and datasets. also ACE, and also the biomed IE corpus GENIA, though i think that one became more broad over the years.

anyways, the SRL and template-filling IE tasks are, as structured prediction problems, extremely similar. when you read the literature there are funny holes and stuff because people in different research communities tend to publish about different ones. however recent work has merged these strands more and more; both ontonotes and genia have multilevel annotations from syntactic to more semantic labels.

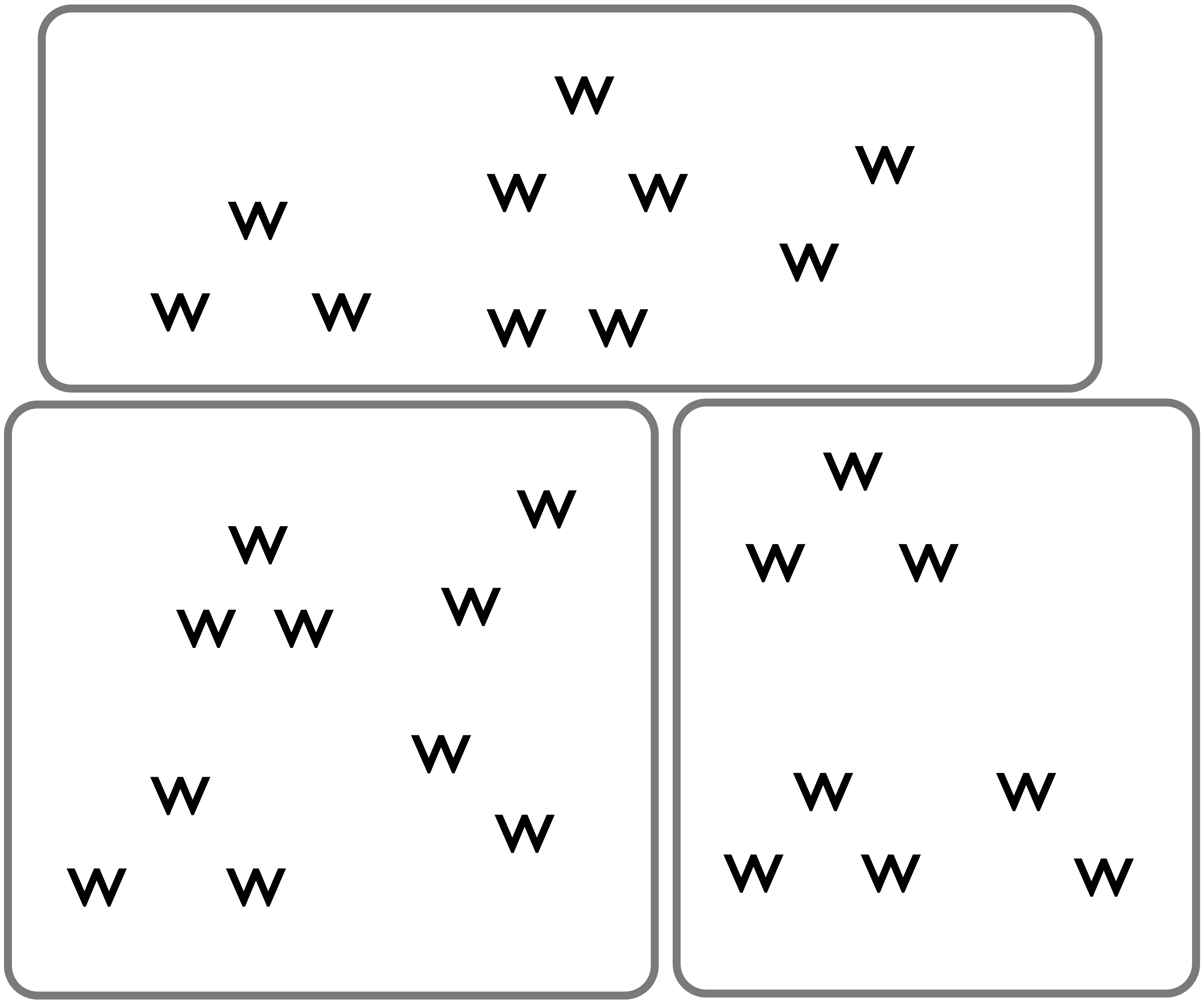
Models

- *LDA:* Topic-Word
- *Model 1:* Frame-Argument
- *Model 2:* Frame-Role
- (*Model 3:* LabeledLDA, metadata constraints)

Data structure in generative text models



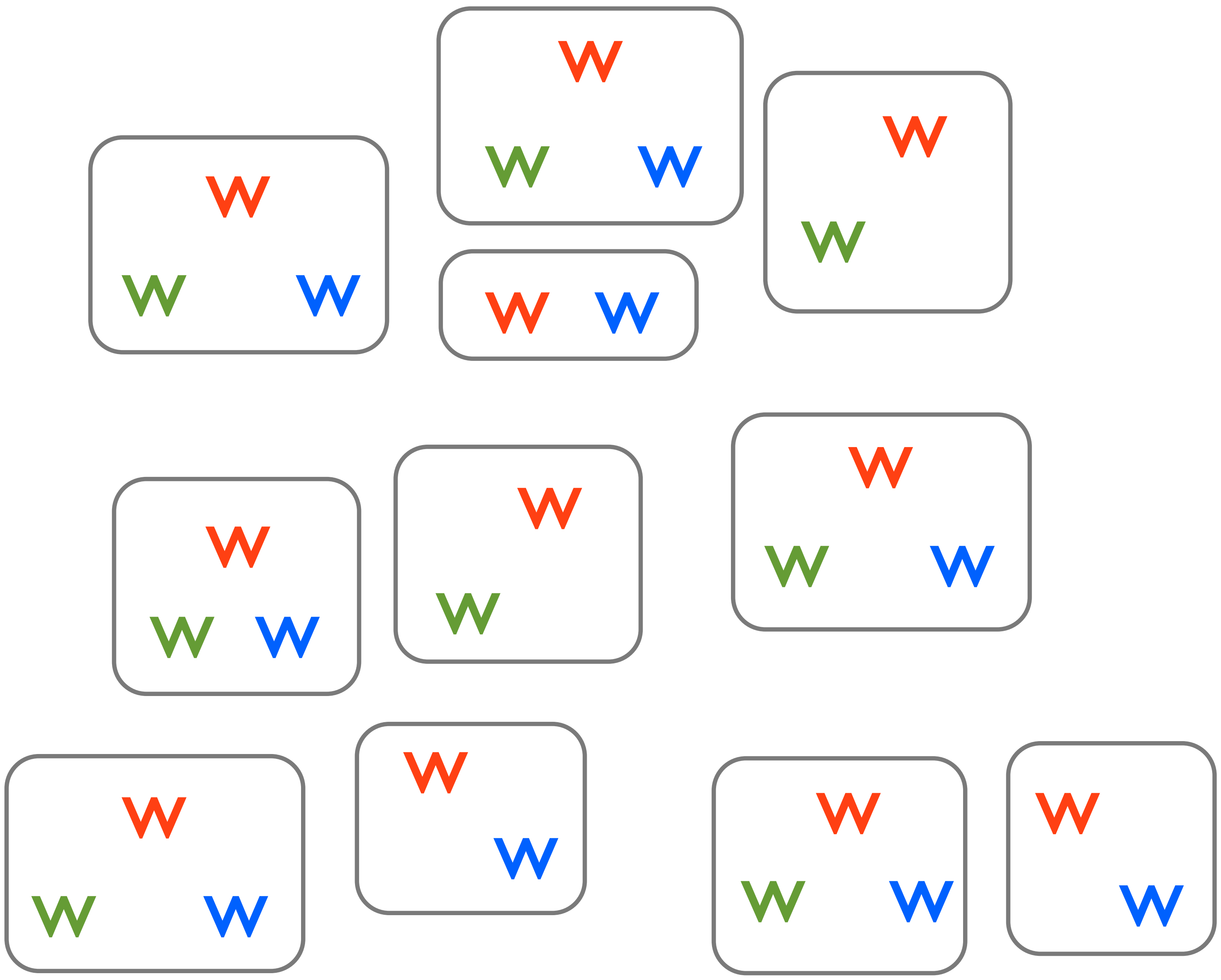
Data structure in generative text models



Documents

Hoffman 99, Blei 03

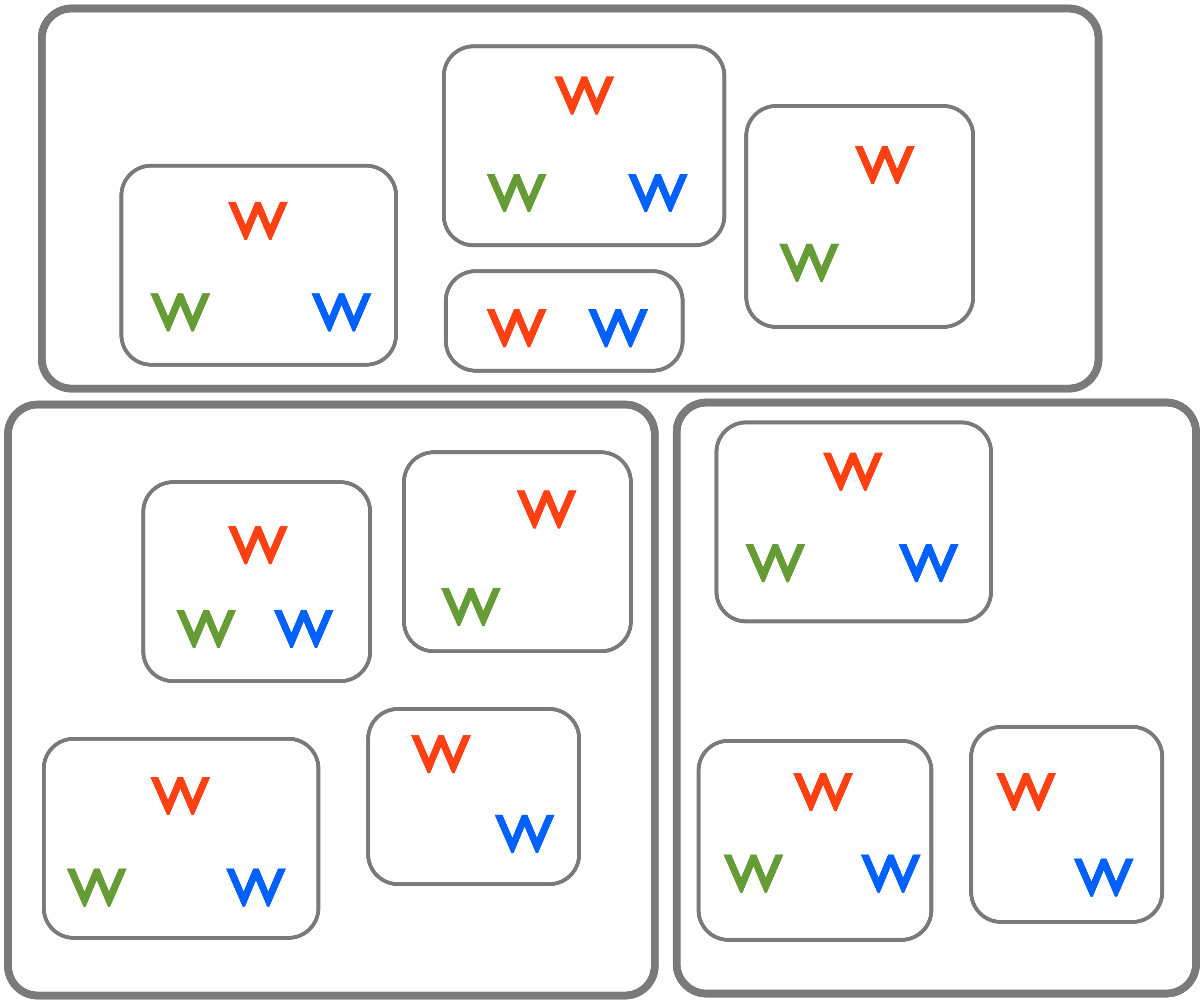
Data structure in generative text models



Documents
Hoffman 99, Blei 03

Syntactic Tuples
verb subject object
Pereira 93, Rooth 98
O Seaghdan 10/11
Ritter 10

Data structure in generative text models



Documents

Hoffman 99, Blei 03

Syntactic Tuples

verb subject object

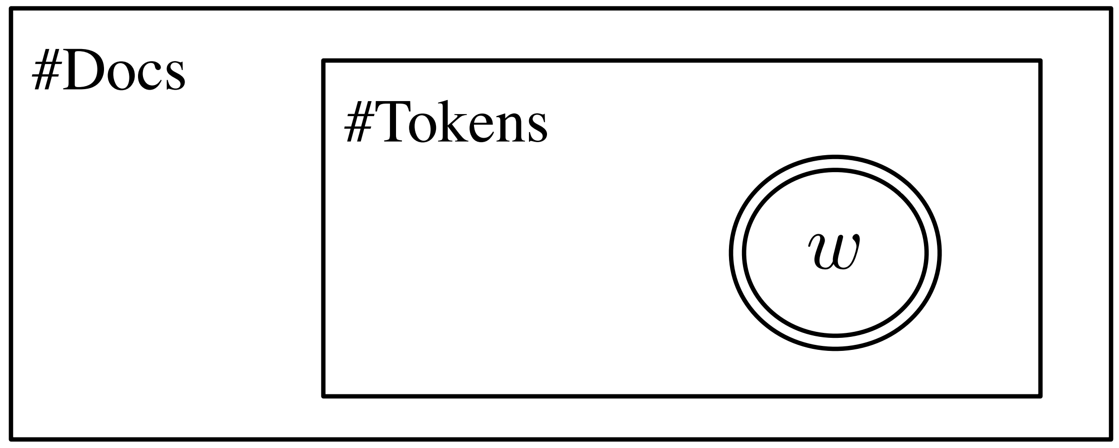
Pereira 93, Rooth 98

O Seaghdan 10/11

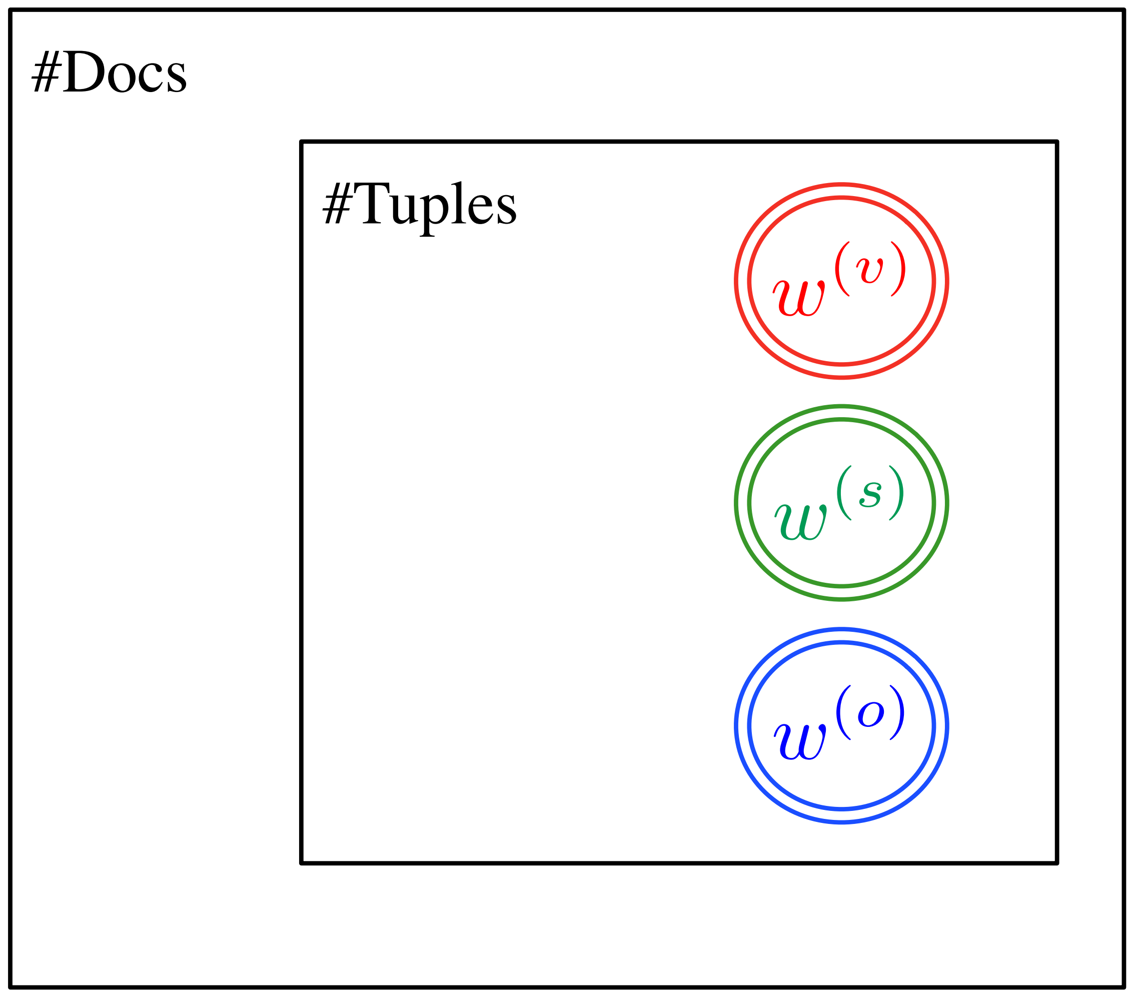
Ritter 10

This work

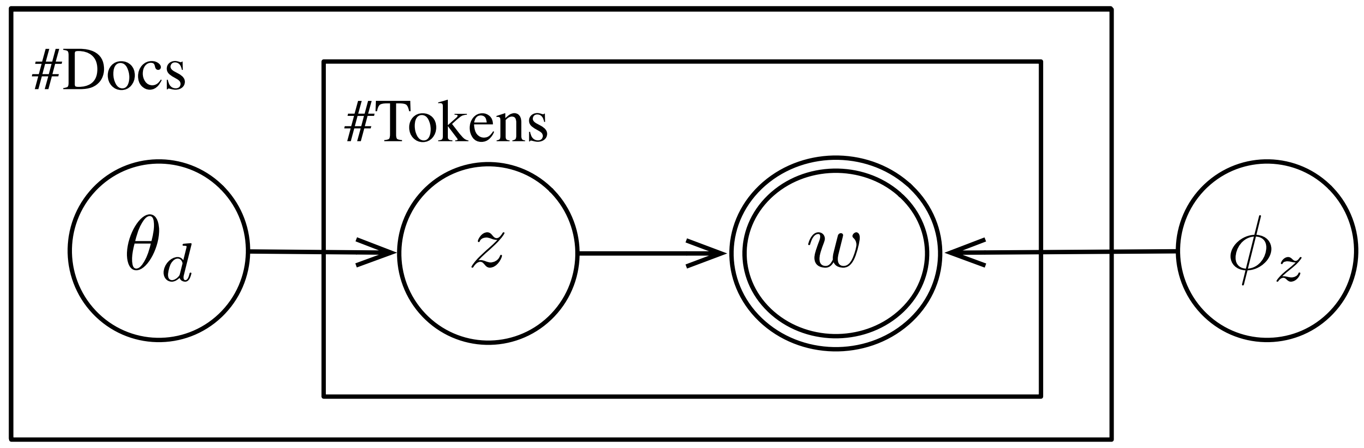
LDA



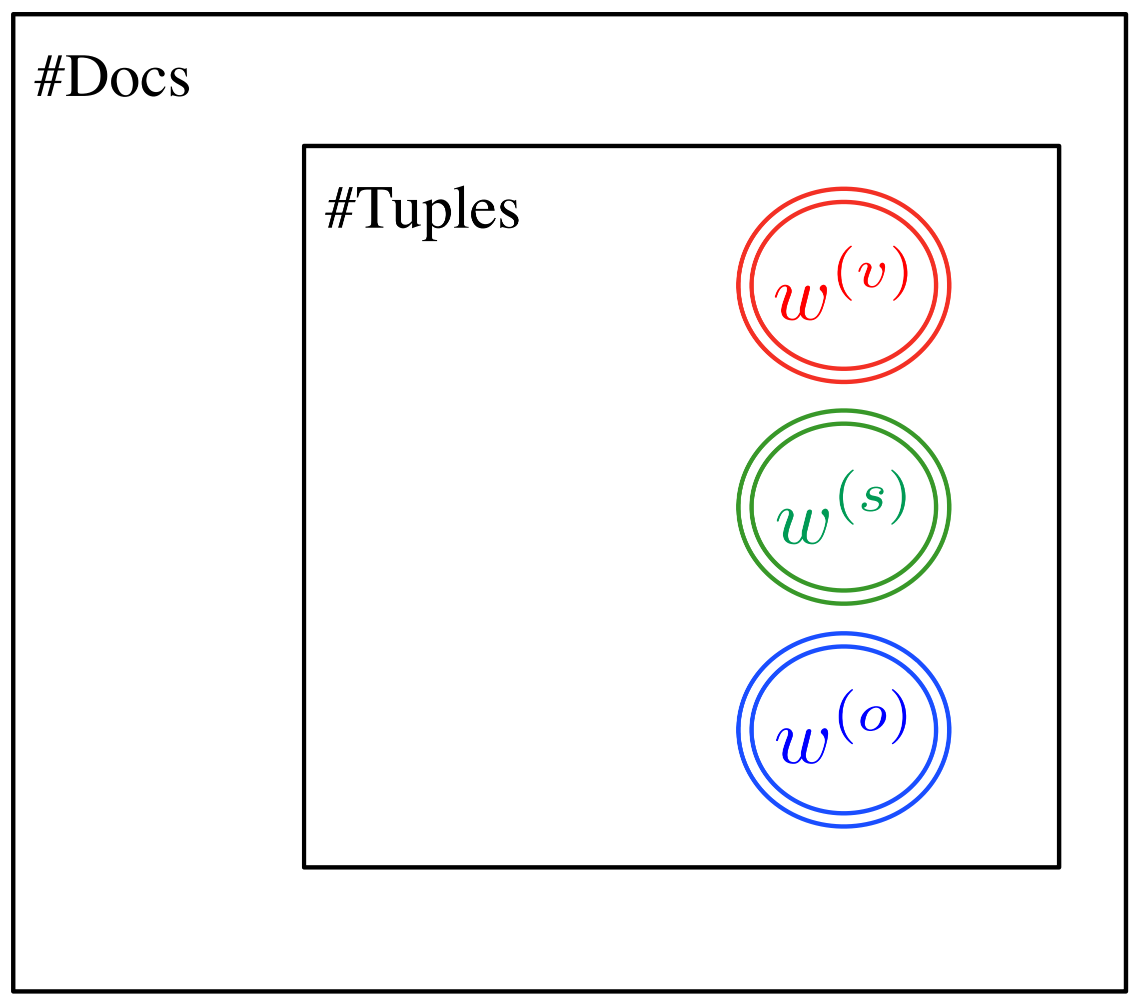
Model 1: Frame-Argument



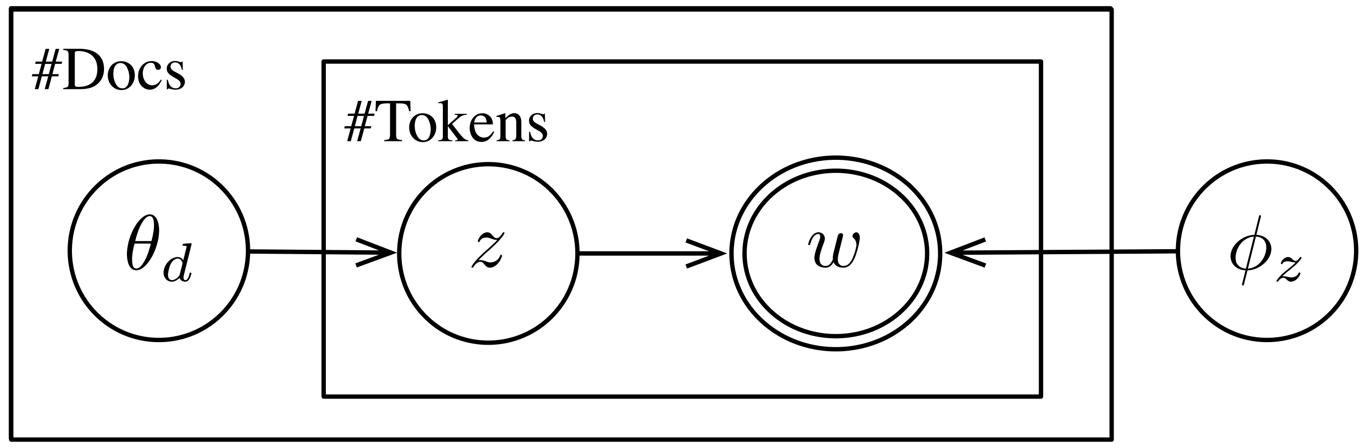
LDA



Model 1: Frame-Argument



LDA



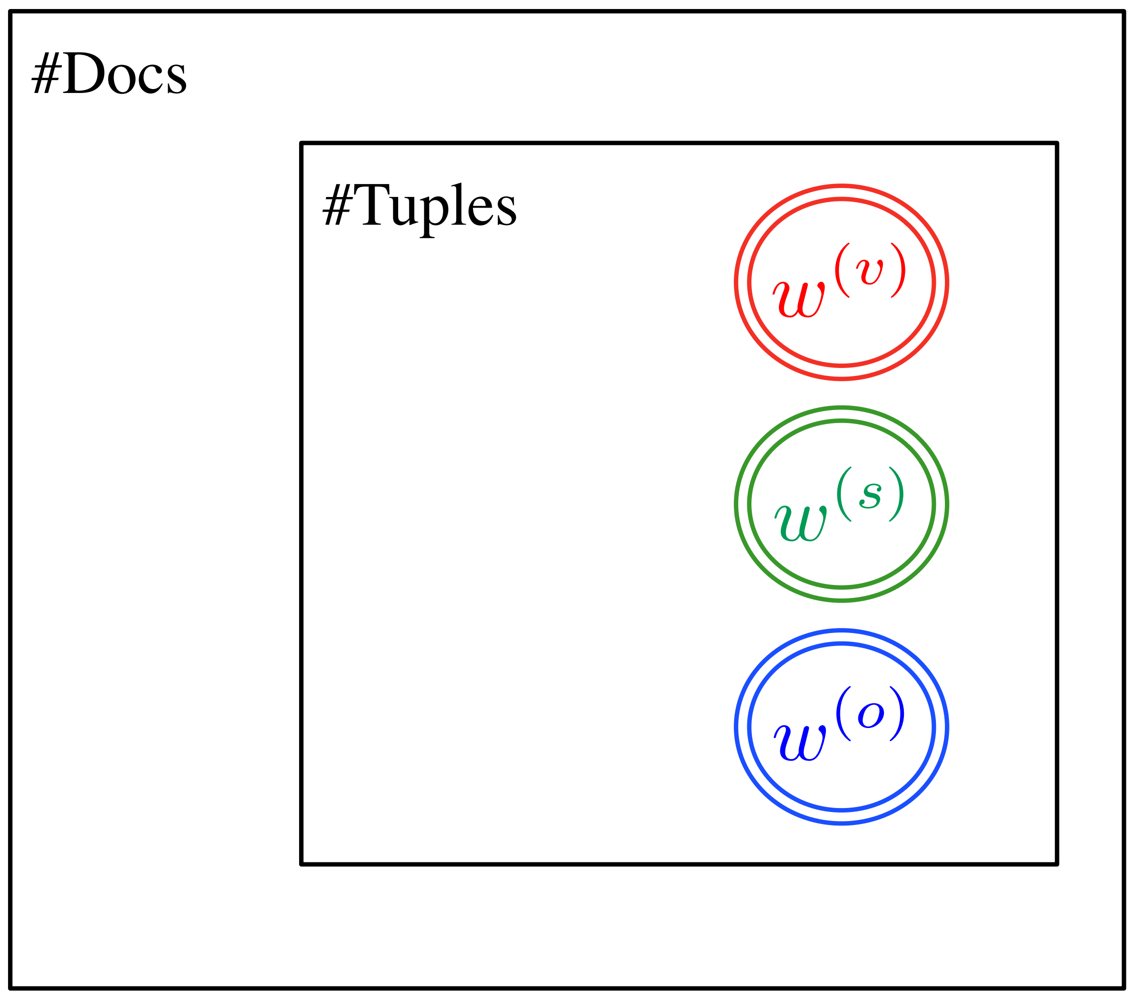
Lexicon

$\phi_k \sim Dir(\beta)$

K word multinomials

K "topics"

Model 1: Frame-Argument



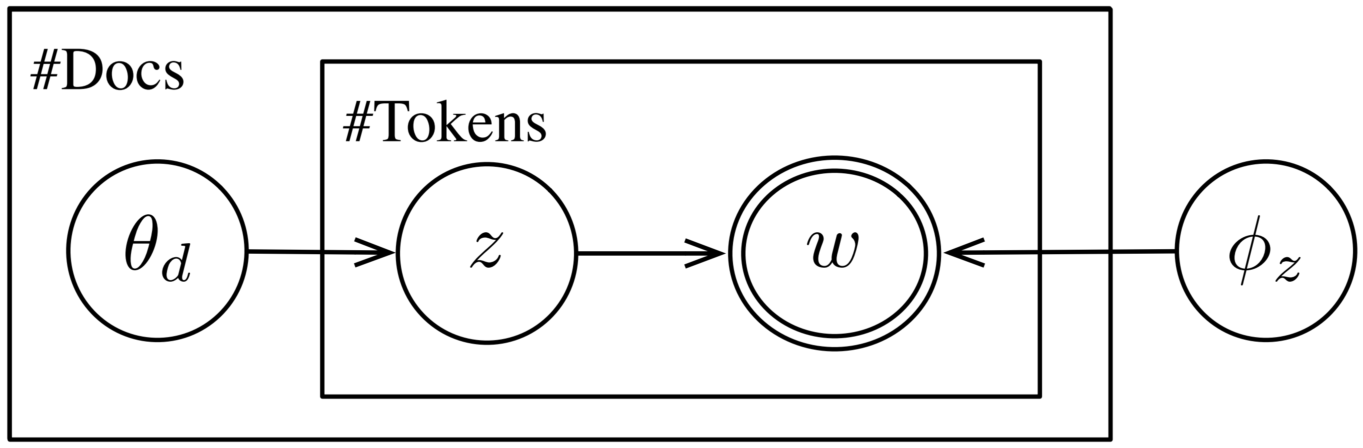
Docs & Text

$\theta_d \sim Dir(\alpha)$

$z \sim \theta_d$

$w \sim \phi_z$

LDA



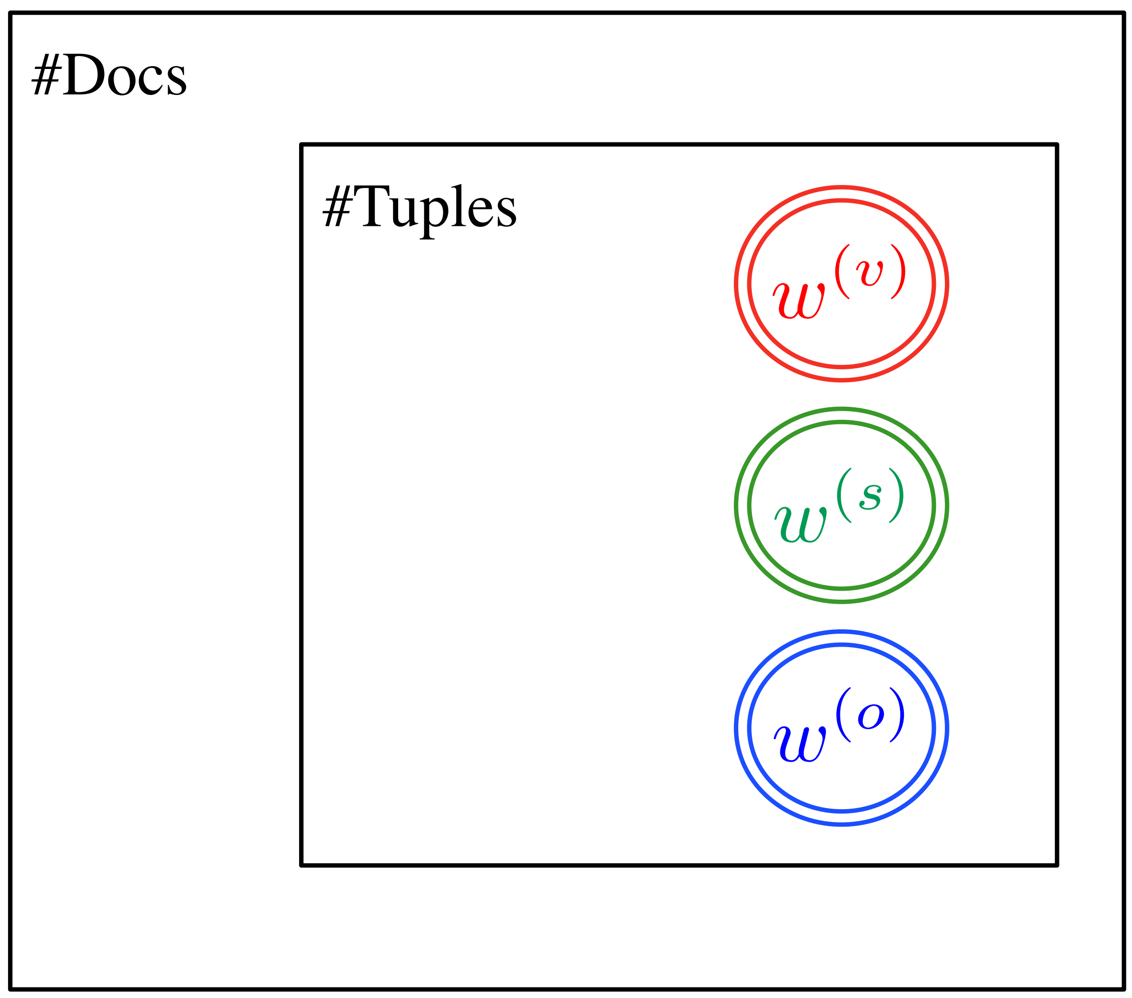
Lexicon

$\phi_k \sim Dir(\beta)$

K word multinomials

K "topics"

Model 1: Frame-Argument



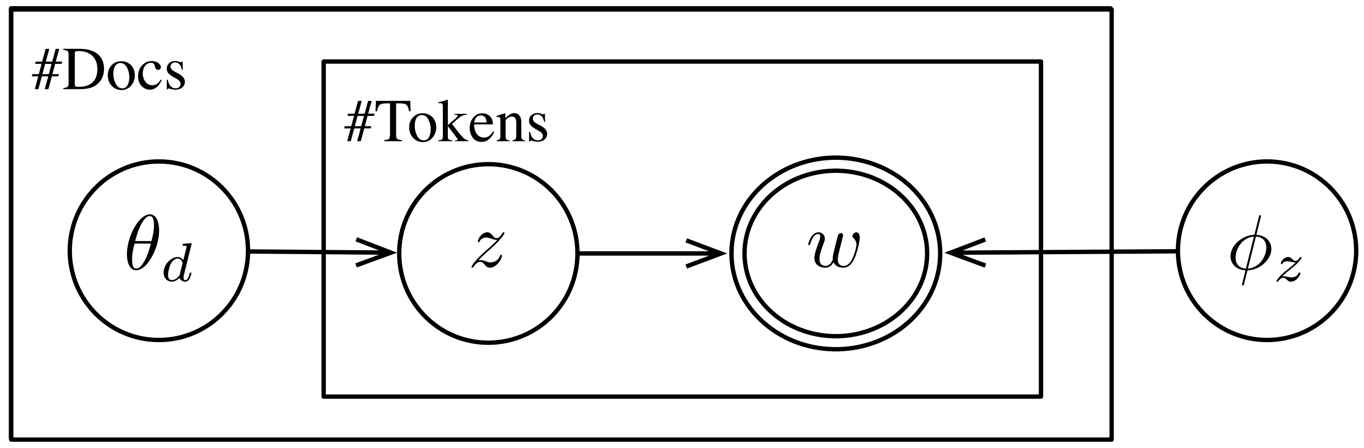
Docs & Text

$\theta_d \sim Dir(\alpha)$

$z \sim \theta_d$

$w \sim \phi_z$

LDA



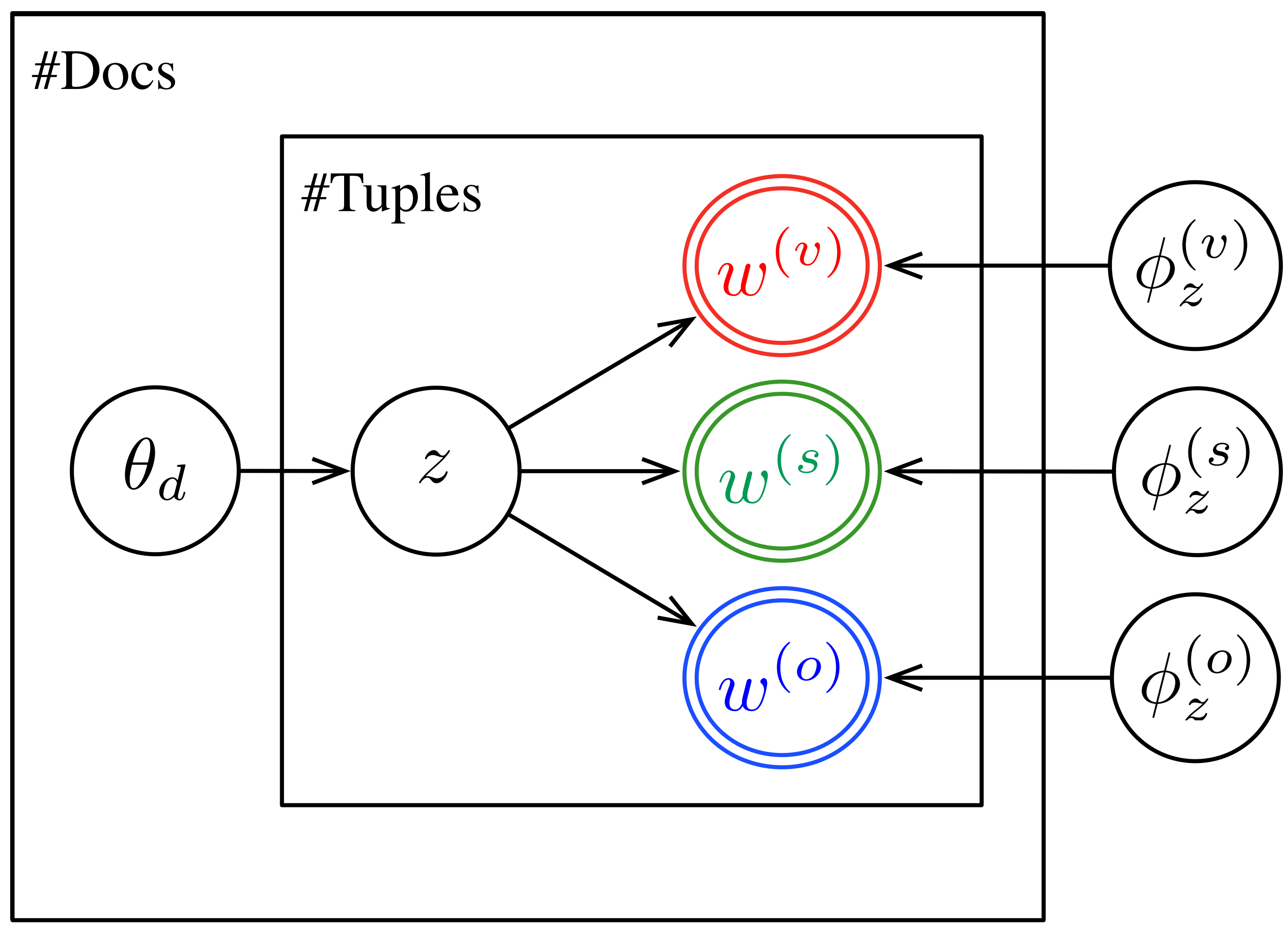
Lexicon

$\phi_k \sim Dir(\beta)$

K word multinomials

K "topics"

Model 1: Frame-Argument



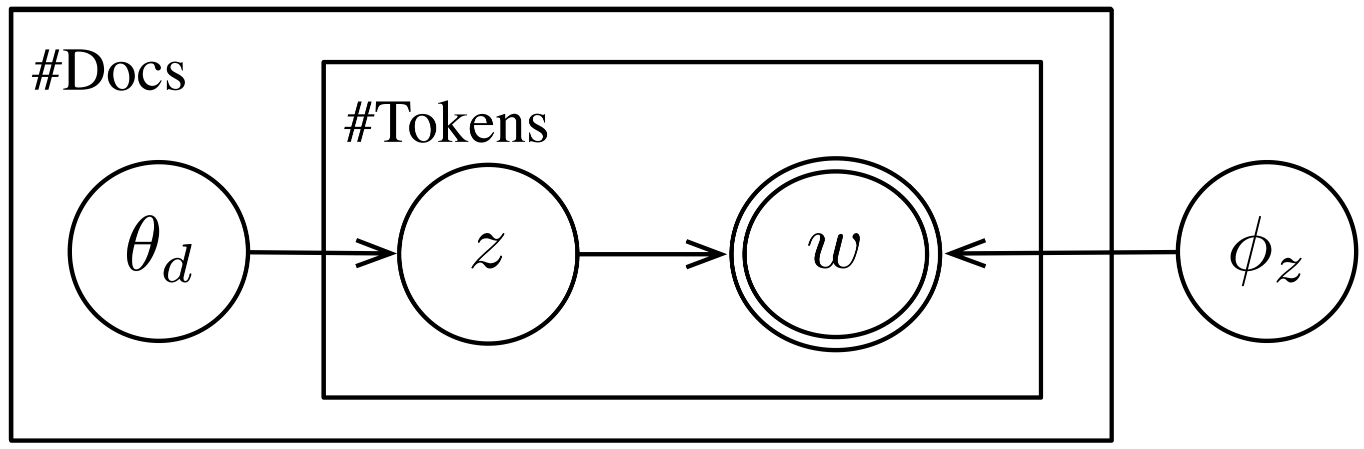
Docs & Text

$\theta_d \sim Dir(\alpha)$

$z \sim \theta_d$

$w \sim \phi_z$

LDA



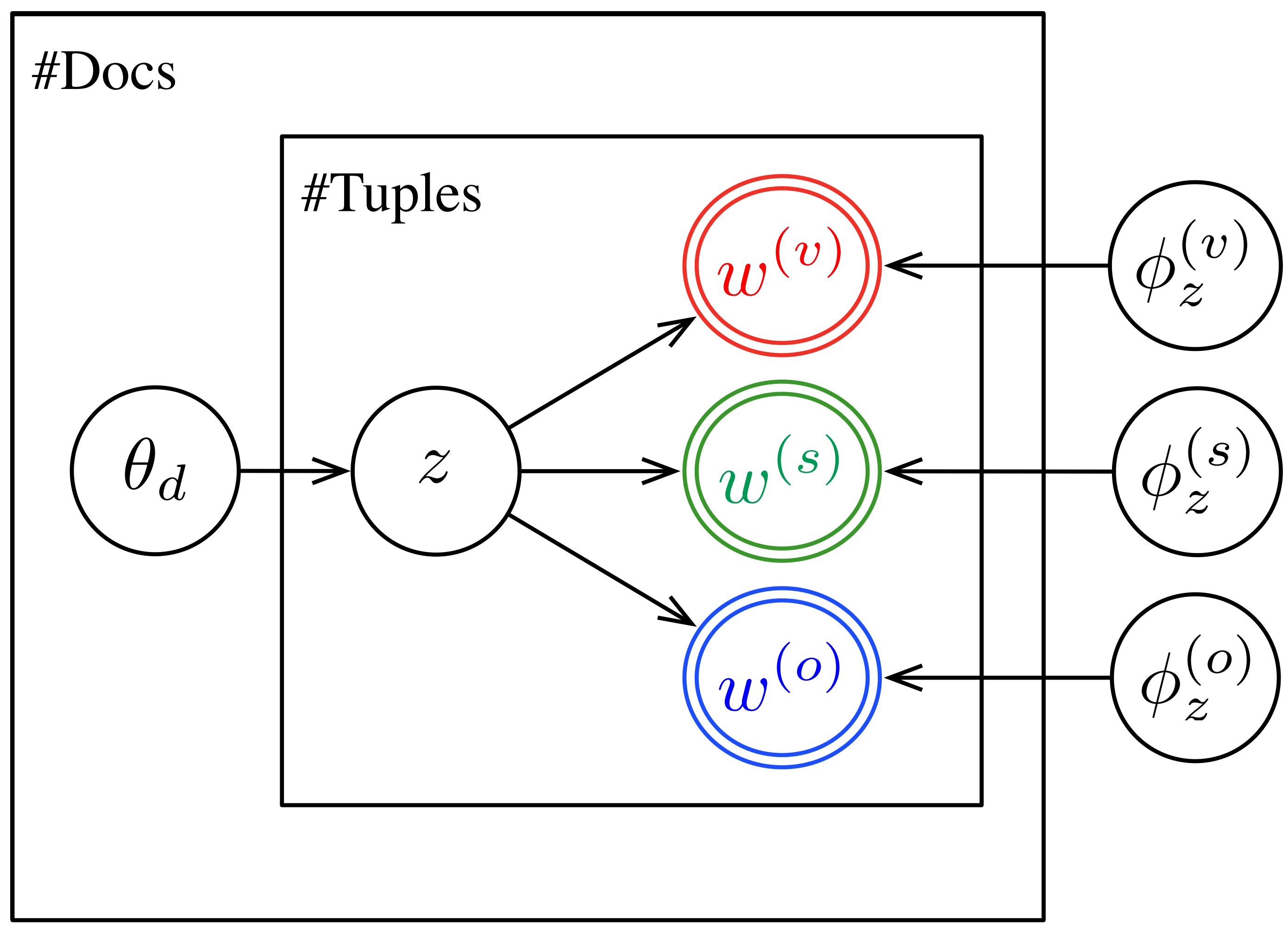
Lexicon

$\phi_k \sim Dir(\beta)$

K word multinomials

K "topics"

Model 1: Frame-Argument



$\phi_k^{(a)} \sim Dir(\beta)$

3K word multinomials

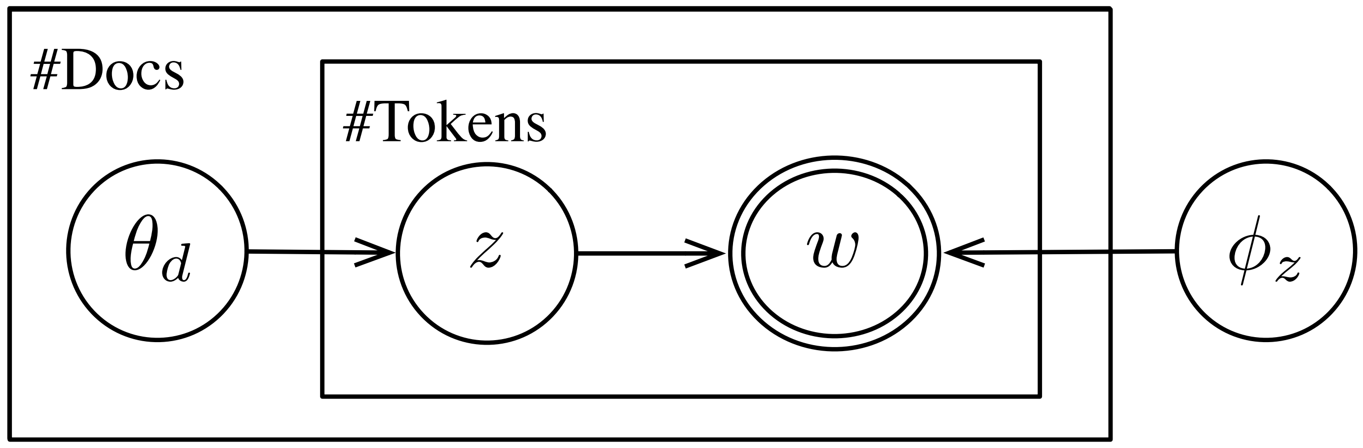
K "frames"

3 "arguments"
verb
subject
object

LDA

<u>*Docs & Text*</u>

$$\theta_d \sim Dir(\alpha)$$
$$z \sim \theta_d$$
$$w \sim \phi_z$$



<u>*Lexicon*</u>

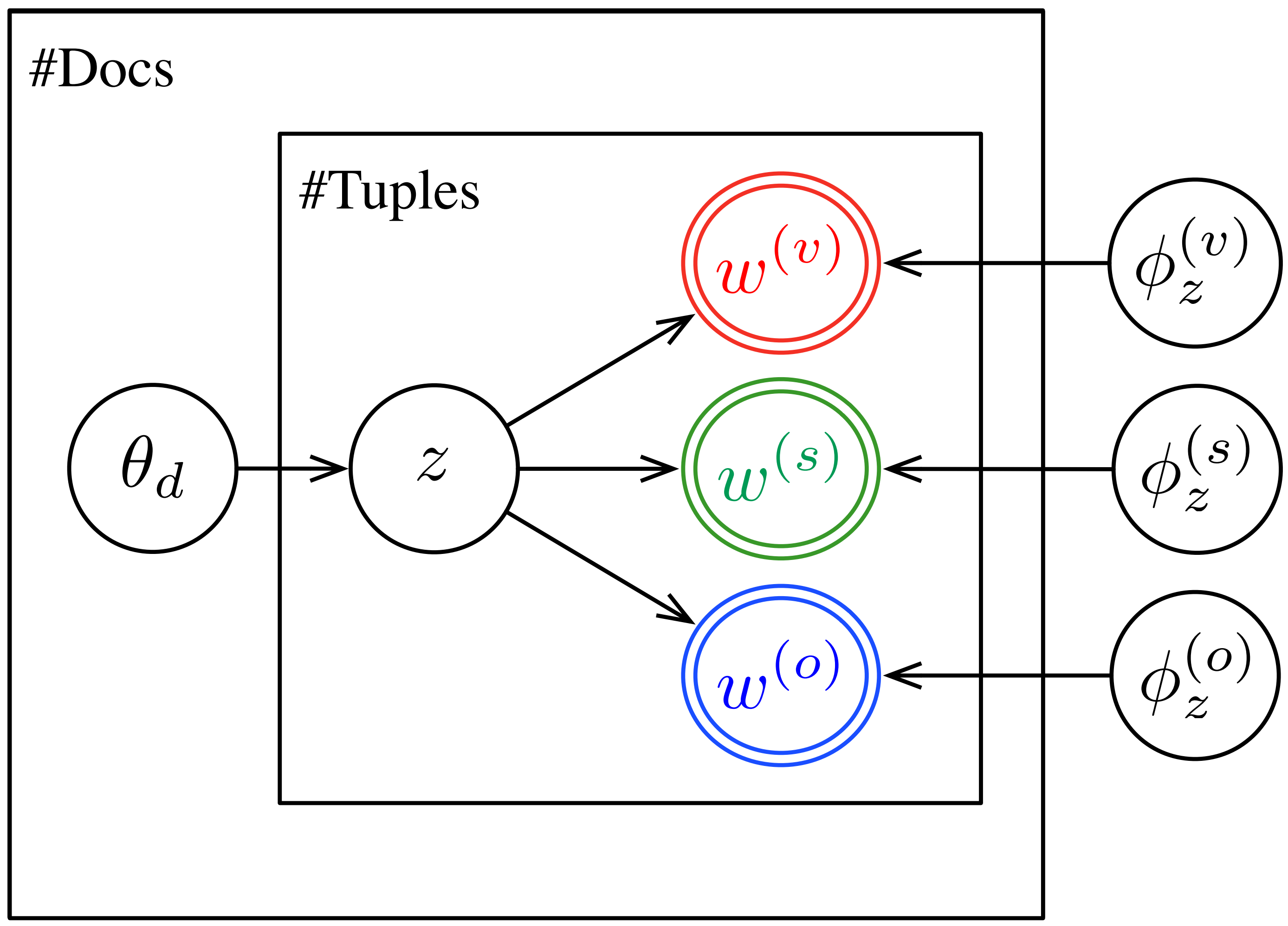
$$\phi_k \sim Dir(\beta)$$

K word multinomials

K "topics"

Model 1: Frame-Argument

$$\theta_d \sim Dir(\alpha)$$
$$z \sim \theta_d$$
$$w^{(a)} \sim \phi_z^{(a)}$$



$$\phi_k^{(a)} \sim Dir(\beta)$$

3K word multinomials

K "frames"

3 "arguments"
verb
subject
object

LDA parameters

__Documents__: doc -> topics θ_d

doc 1: 1 2 3 □ 5 6 7 □

__Lexicon__: topic -> words $\phi_k^{(a)} \sim Dir(10/26)$

1 [a b c d e f g h i j k l m n o p q r s t u v w x y z]

2 [a b c d e f g h i j k l m n o p q r s t u v w x y z]

Model 1 (Frame-Arg)

<u>Documents</u>: doc -> topics θ_d

doc 1: $\boxed{1}$ $\boxed{2}$ $\boxed{3}$ $\boxed{}$ $\boxed{5}$ $\boxed{6}$ $\boxed{7}$ $\boxed{}$

<u>Lexicon</u>: (frame,arg) -> words $\phi_k^{(a)} \sim Dir(10/26)$

$\boxed{1}$

verb position [a b c d f g h i l m n o r s u v w x y z]

subject position [b c d g h i k l m n p q r t u v w x z]

object position [a b d e f g h i j k l m n r s t v y]

Model 1 (Frame-Arg) result

(Data: news stories about crime)

Frame f=66

13,392 (1%) sites

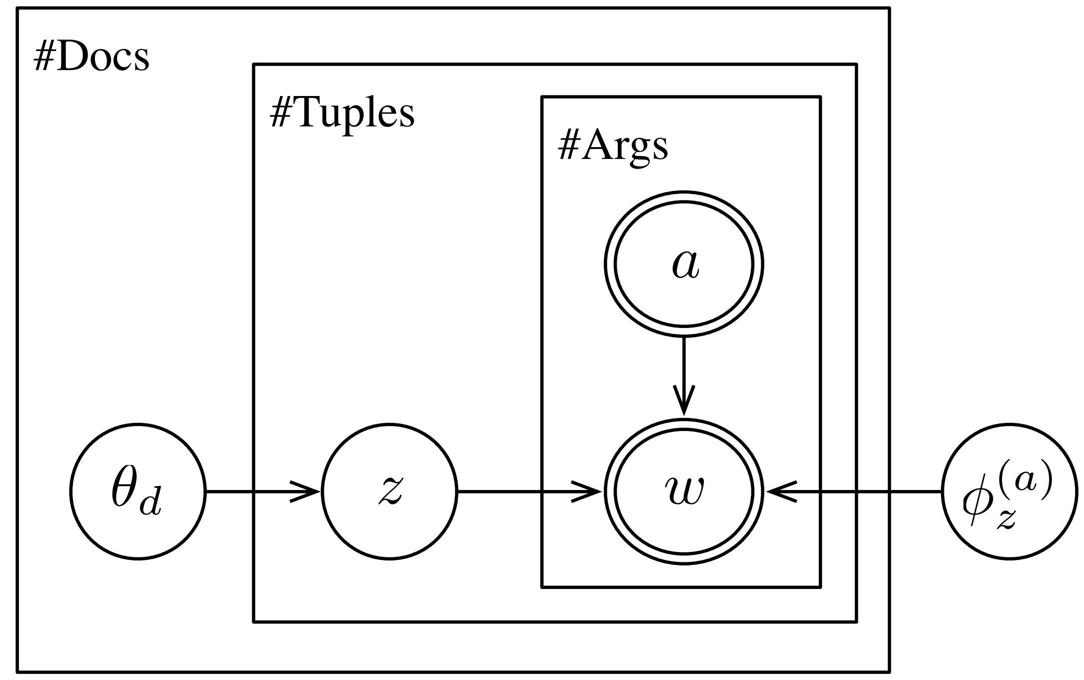
$\phi_k^{(v)}$	$\phi_k^{(s)}$	$\phi_k^{(o)}$
present hear have cite give offer support make use include prove call introduce find admit challenge provide contradict produce corroborate incriminate review describe play question show consider OOV believe discuss dispute allow attack read reject discredit deny say accept fabricate obtain elicit confirm suppress turn rebut take establish examine recant (0.620 mass)	NONE(0.67) prosecutor lawyer OOV prosecution jury evidence defense witness report police judge testimony investigator government juror defendant attorney trial woman court case statement officer state team expert official investigation account Milosevic side tape other inquiry agent supporter record Gotti detective authority accuser Government member office Judges Puccio article tribunal Smith (0.885 mass)	evidence testimony statement case account witness argument confession story claim credibility conversation tape assertion report charge guilt contention OOV defense fact role detail allegation version innocence accusation finding theory information videotape word picture transcript anything motive inconsistency summation suggestion conclusion description part effort document truth expert admission involvement plea remark (0.680 mass)

Frame f=89

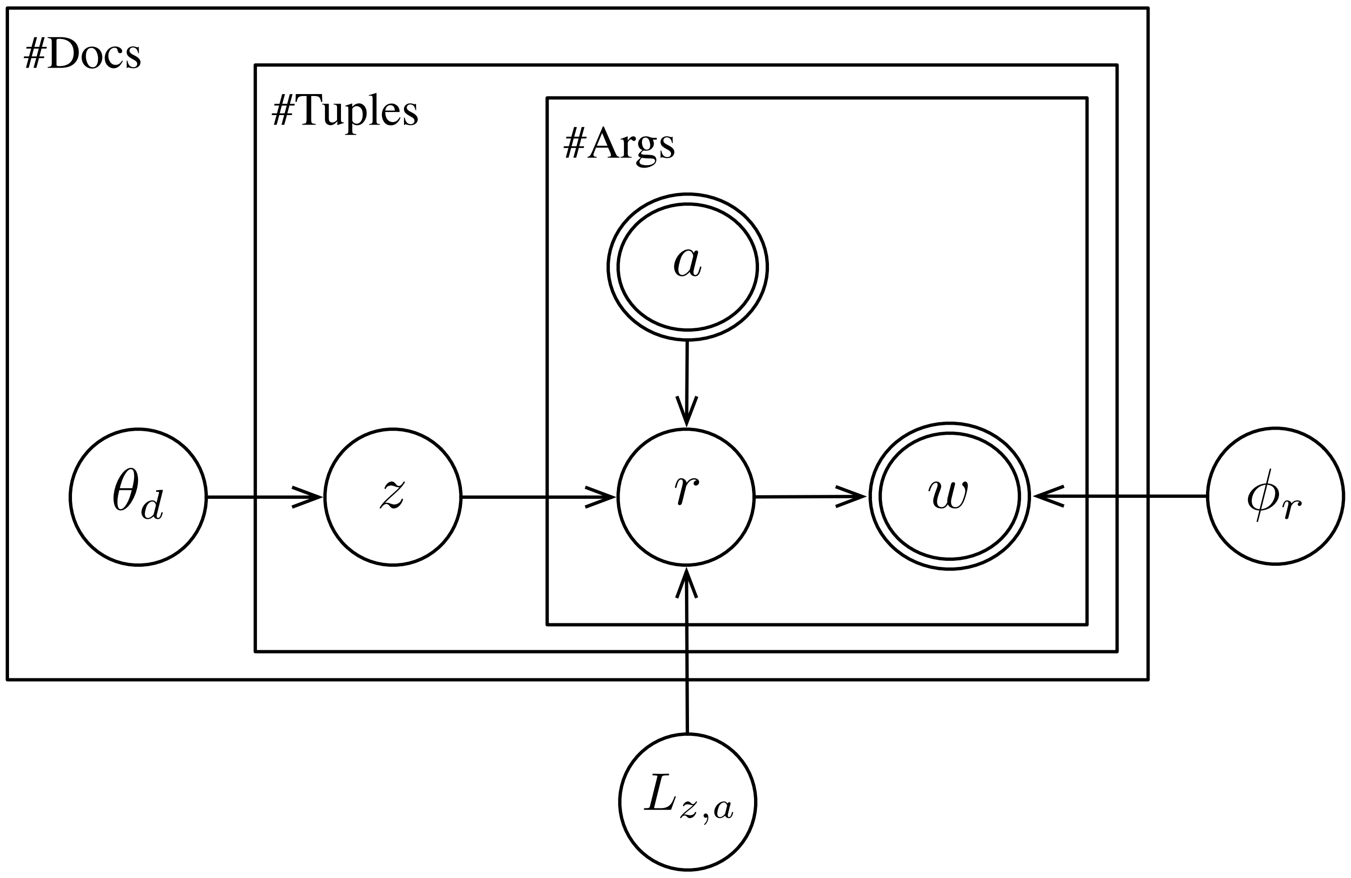
12,551 (1%) sites

have give lose take gain use exercise maintain retain lack win assume regain seize hold deny claim establish restore exert limit wield grant abuse enjoy seek bear share get increase relinquish assert OOV show accept strengthen keep recognize resign demonstrate expand earn extend undermine overstep build provide lend sever cede (0.823 mass)	NONE(0.62) OOV court judge government people Congress States city prosecutor state police tribunal officer official Milosevic member agency man family defendant force proposal group authority president commission board parent Department Giuliani Bush Government party Washington leader Gotti organization trial citizen Americans decision Judges Council country campaign office crime Democrats woman (0.795 mass)	power control authority right responsibility jurisdiction support position influence discretion tie access role OOV chance reputation effect rights interest opportunity ability confidence custody impact office status job credibility connection experience obligation link option advantage post leadership duty legitimacy majority knowledge title trust case respect independence benefit time sense seat license (0.637 mass)

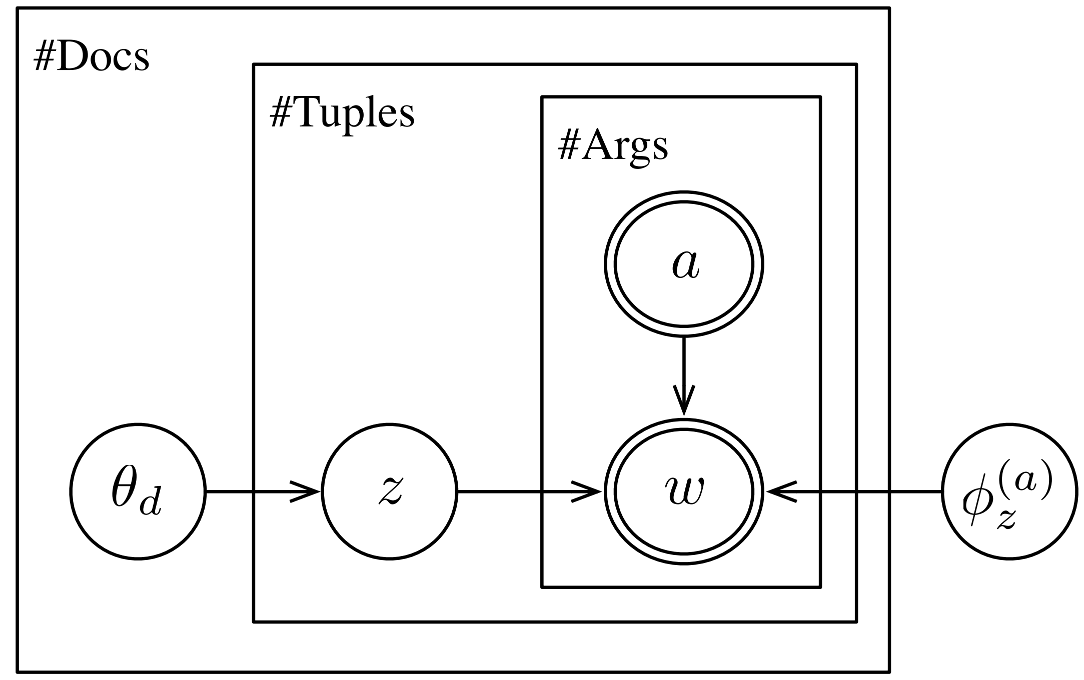
Model 1: Frame-Argument



Model 2: Frame-Role



Model 1: Frame-Argument



$$L_{k,a} \sim Dir(\gamma_a)$$
$$r \sim L_{z,a}$$
$$w \sim \phi_r$$

Introduce <u>roles</u>: shared across frames

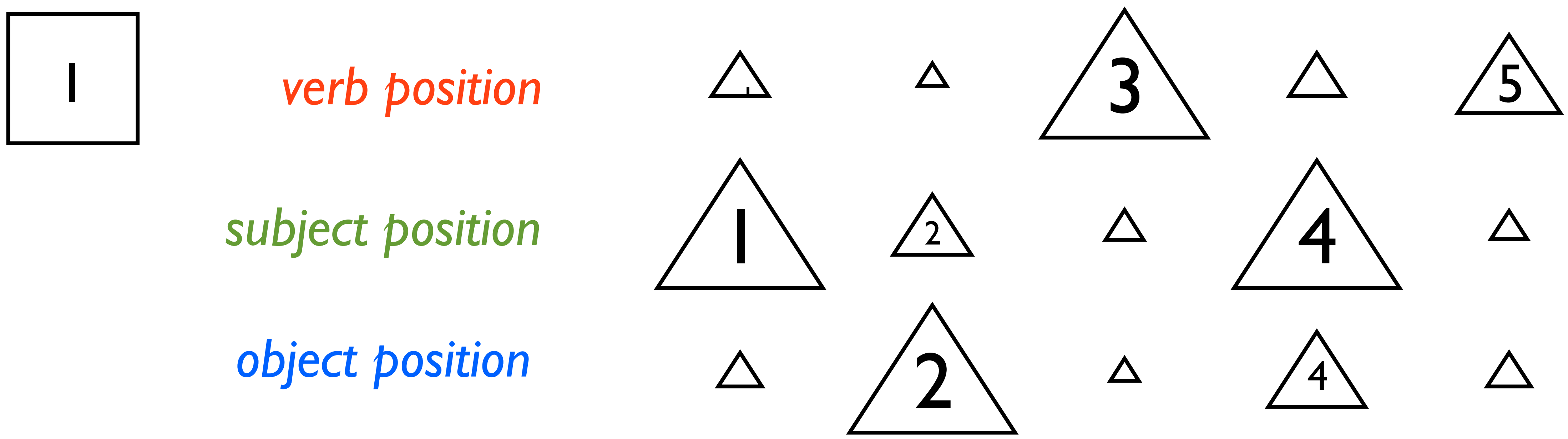
- roles can have different argument positions, in different frames
- roles are word classes

K frames, *R* roles

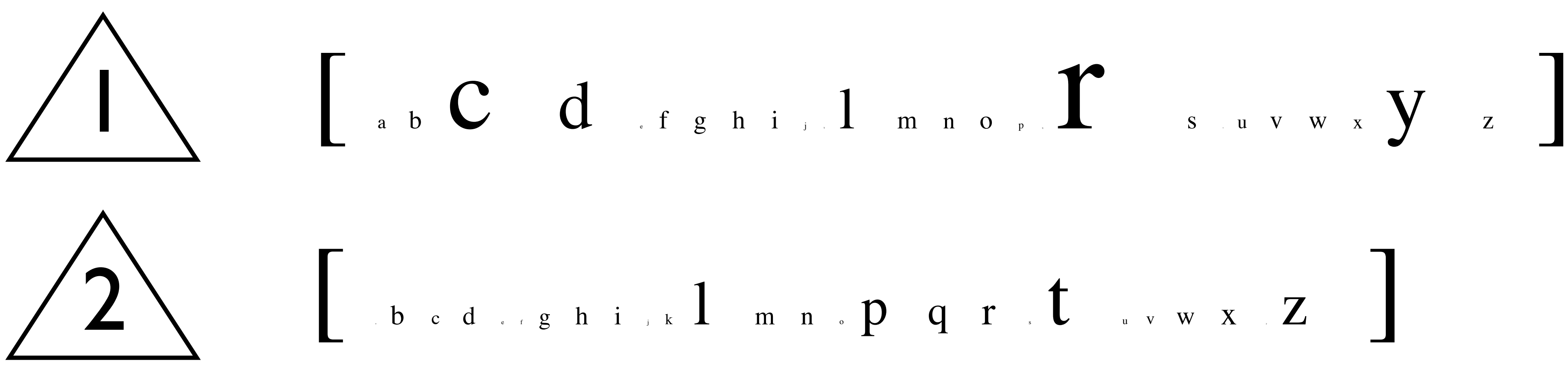
$L_{k,a}$: frame-role "linker"

Model 2 (Frame-Role)

Linker: (frame,arg) -> roles $L_{k,a} \sim Dir(\gamma_a)$



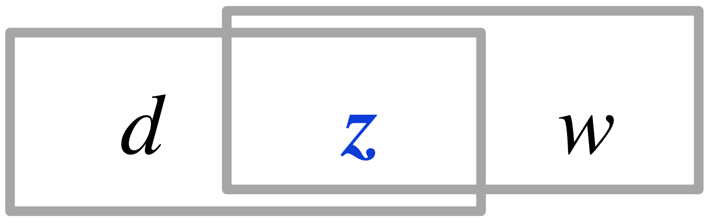
Roles: role -> words $\phi_k \sim Dir(\beta)$



Inference

Collapsed Gibbs sampling

(only showing discrete variables)



Document-Word LDA

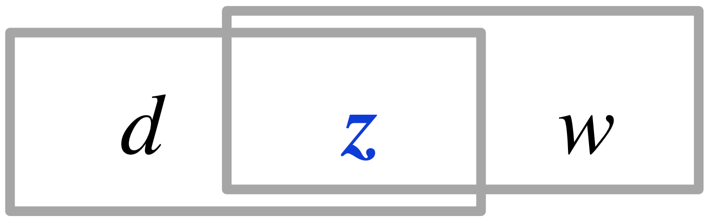
$$p(\mathbf{z} \mid d, w) \;\propto\; p(\mathbf{z} \mid d)\; p(w \mid \mathbf{z})$$


```
p(z_i=\mathbf{\textcolor{blue}{z}}\mid z_{-i}, w, d; \alpha,\beta)\ \propto\
  \frac{ C[\mathbf{\textcolor{blue}{z}}, d_i] + \alpha }{C[d_i] + \alpha_0}\
  \frac{ C[w_i, \mathbf{\textcolor{blue}{z}}] + \beta  }{C[z] + \beta_0}
```

Inference

Collapsed Gibbs sampling

(only showing discrete variables)



Document-Word LDA

$$p(\mathbf{z} \mid d, w) \ \propto \ p(\mathbf{z} \mid d) \ p(w \mid \mathbf{z})$$

$$p(z_i = \mathbf{z} \mid z_{-i}, w, d; \alpha, \beta) \ \propto \ \frac{C[\mathbf{z}, d_i] + \alpha}{C[d_i] + \alpha_0} \ \frac{C[w_i, \mathbf{z}] + \beta}{C[\mathbf{z}] + \beta_0}$$

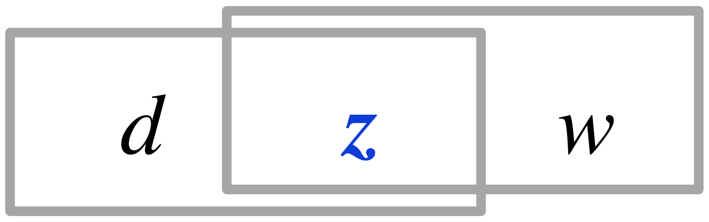
Thursday, March 8, 2012

```
p(z_i=\mathbf{\textcolor{blue}{z}}\mid z_{-i}, w, d; \alpha,\beta)\ \propto\
  \frac{ C[\mathbf{\textcolor{blue}{z}}, d_i] + \alpha }{C[d_i] + \alpha_0}\
  \frac{ C[w_i, \mathbf{\textcolor{blue}{z}}] + \beta  }{C[z] + \beta_0}
```

Inference

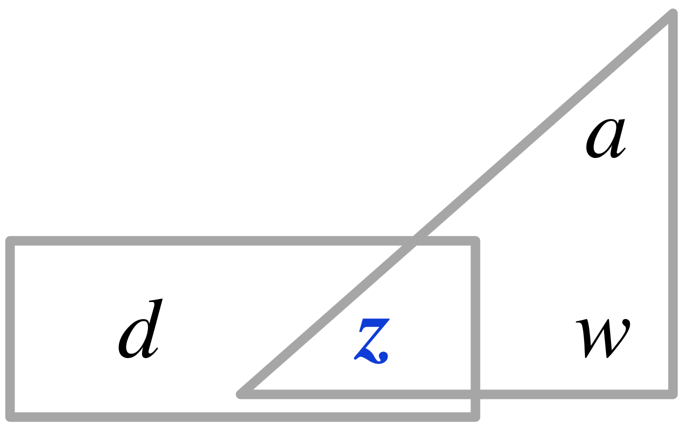
Collapsed Gibbs sampling

(only showing discrete variables)



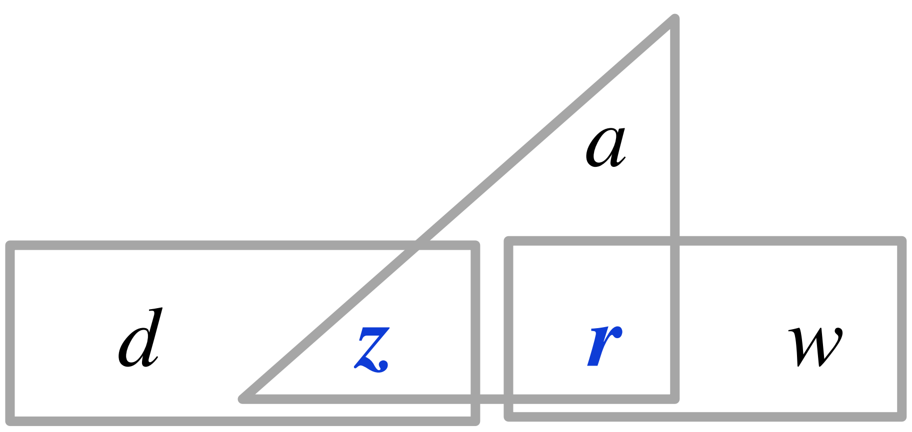
Document-Word LDA

$$p(\mathbf{z} \mid d, w) \;\propto\; p(\mathbf{z} \mid d)\; p(w \mid \mathbf{z})$$



Frame-Argument (Model 1)

$$p(\mathbf{z} \mid d, w, a) \;\propto\; p(\mathbf{z} \mid d) \prod_{a} p(w^{(a)} \mid \mathbf{z}, a)$$



Frame-Role (Model 2)

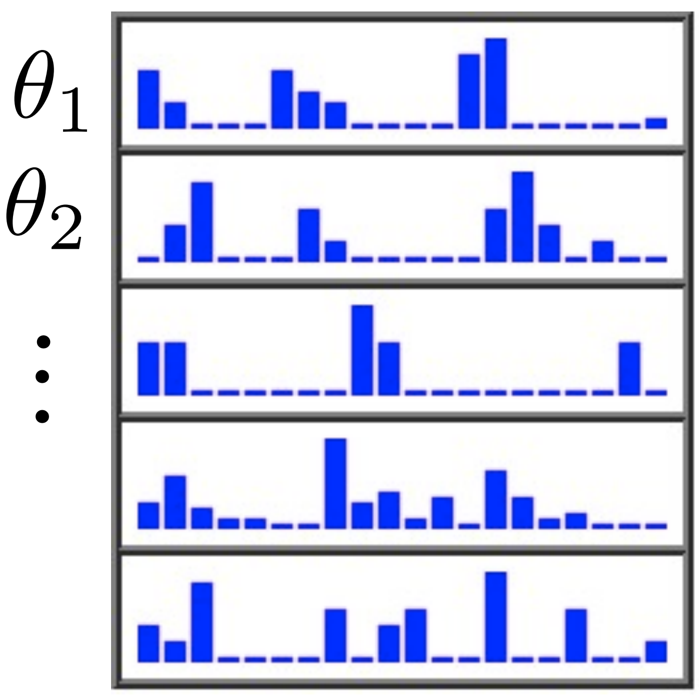
$$p(\mathbf{z} \mid d, r, a) \;\propto\; p(\mathbf{z} \mid d) \prod_{a} p(r^{(a)} \mid \mathbf{z}, a)$$

$$p(\mathbf{r^{(a)}} \mid z, w^{(a)}, a) \;\propto\; p(\mathbf{r^{(a)}} \mid z, a)\; p(w^{(a)} \mid \mathbf{r^{(a)}})$$

Concentration resampling

$$\theta \sim Dir(\alpha = high)$$

$\ldots$



Rest of model causes sparser theta's than implied by alpha

Thursday, March 8, 2012

p(\beta\mid\text{everything else})

Concentration resampling

Better likelihood with

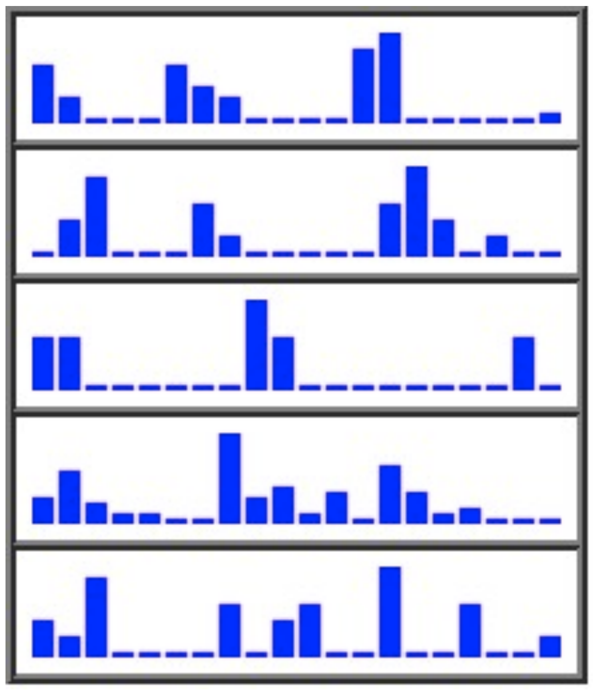
$$\theta \sim Dir(\alpha = high) \longrightarrow \alpha = low$$

$\ldots$

θ_1

θ_2

$\vdots$



Rest of model causes sparser theta's than implied by alpha

Previous work shows concentration optimization/ inference makes a large difference (Asuncion, Wallach, Johnson, ..)

p(\beta\mid\text{everything else})

Concentration resampling

Better likelihood with

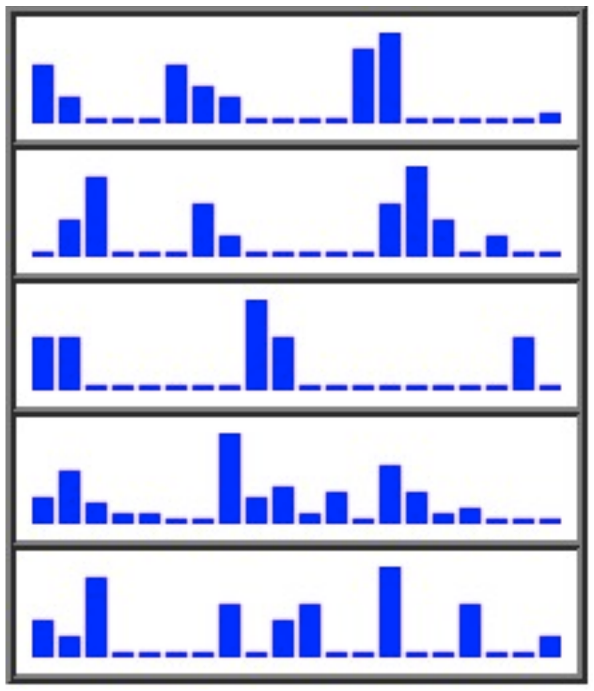
$$\theta \sim Dir(\alpha = high) \longrightarrow \alpha = low$$

$\ldots$

θ_1

θ_2

$\vdots$



Rest of model causes sparser theta's than implied by alpha

Previous work shows concentration optimization/ inference makes a large difference (Asuncion, Wallach, Johnson, ..)

Solution: resample

$$p(\alpha \mid \text{everything else})$$

$$p(\beta \mid \text{everything else})$$

$$p(\gamma \mid \text{everything else})$$

every 50 iterations

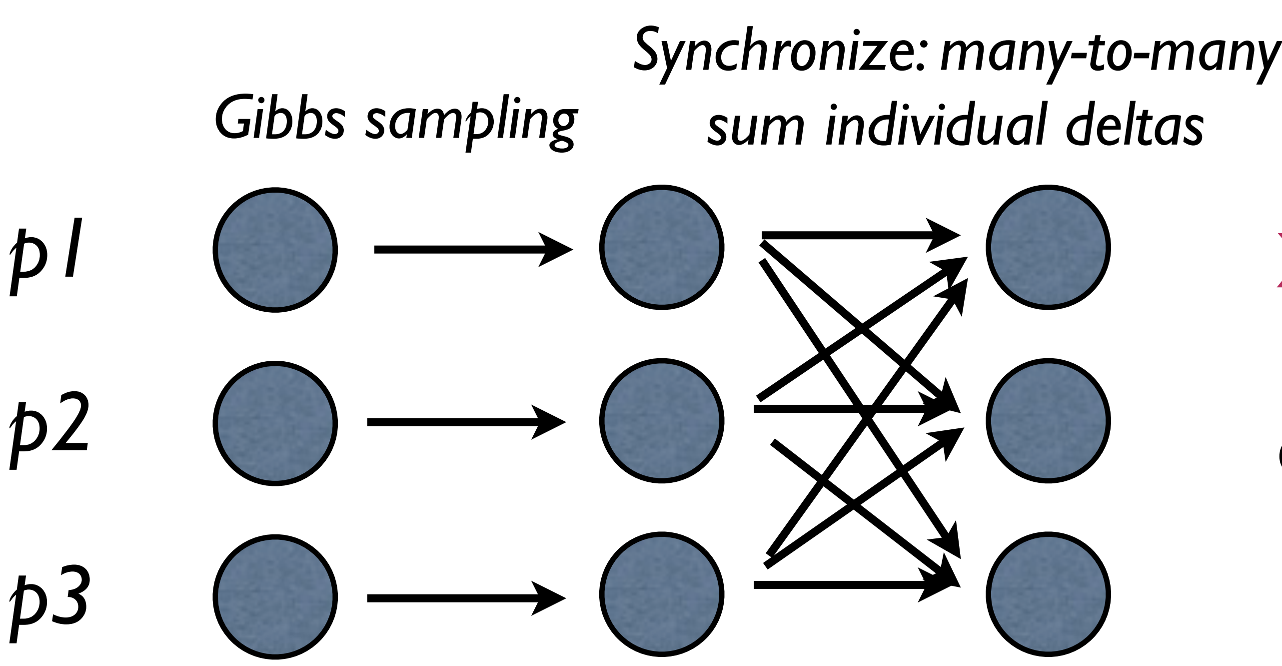
[Using slice sampling (Neal 2003): like MH but less fiddly]

Thursday, March 8, 2012

p(\beta\mid\text{everything else})

Parallelization

- Processors use stale counts, occasionally synchronize (~Newman 2007, etc.)
    - Provably non-ergodic! But likelihood seems to be going up.
- MPI -- Message-Passing Interface -- is great!



[John Langford wrote half a paper on AllReduce]

```
Xnew = np.empty(
    Xcur.shape, dtype=Xcur.dtype)
comm.Allreduce(
    Xcur, Xnew, MPI.SUM)
```

- Implementation: Python/C/NumPy/mpi4py
- Pittsburgh Supercomputing Center's *Blacklight* machine (16 to 256 or more?? cores)

Datasets

- Want (1) easy-to-parse, (2) coherent topics
- CrimeNYT
  - From the New York Times Annotated Corpus
    [Sandhaus 2007]
    1.8 M articles, 1987-2007, with manual labels
  - Select articles having one label containing "crime" or "criminal"
    27,117 articles (20M words)
- Penn Treebank: gold standard parses
  - Wall Street Journal (late 80's?) (1.2M words)
    [Marcus 1993]
  - Brown corpus: literature, essays (460k words)
    [Kucera and Francis 1964]

CrimeNYT sample

count	category label
48,645	crime and criminals
9,497	sex crimes
6,304	sentences (criminal)
3,892	war crimes, genocide and crimes against humanity
2,818	organized crime

Table 1: Most common category labels matching query

1987-05-05	JURY SELECTION MAJOR HURDLE IN TRIAL THAT MAY LAST YEARS
1988-02-25	Moslem Patrol Helps Cut Crime in Brooklyn
1991-10-22	GUILTY PLEAS SET IN U.S. COAL CASE
2001-05-30	4 GUILTY IN TERROR BOMBINGS OF 2 U.S. EMBASSIES IN AFRICA; JURY TO WEIGH 2 EXECUTIONS
2001-10-17	A Rush for Cipro, and the Global Ripples
2003-08-09	World Briefing — Europe: Northern Ireland: Fund For Bomb Lawsuits
2003-10-03	Bryant's Accuser Won't Have to Testify
2004-07-18	Despite Appeals, Chaos Still Stalks the Sudanese
2005-04-01	World Briefing — Europe: France: Longer Prison Term In Graft Case

Table 2: Sample of headlines from the dataset.

Preprocessing: SVO extraction

- Stanford CoreNLP
  - Sentence splitting, tokenization, part-of-speech tagging, lemmatization, named entity recognition, phrase structure parsing, dependency extraction

Lemmatization

- Part-of-speech-aware stemming: English inflectional morphology
- Smart with names
- *morpha* tool, Univ. Sussex (copied within Stanford NLP)
- Compare: lowercase + Porter stemmer

POS	*Word*	*Lemma*	*Porter Stem*
WRB	When	when	when
PRP	you	you	you
VBD	walked	walk	walk
IN	in	in	in
DT	that	that	that
NN	day	day	day
,	,	,	,
PRP	you	you	you
RB	almost	almost	almost
VBD	shot	**shoot**	**shot**
PRP	me	**I**	**me**

Lemmatization

- Part-of-speech-aware stemming: English inflectional morphology
- Smart with names
- *morpha* tool, Univ. Sussex (copied within Stanford NLP)
- Compare: lowercase + Porter stemmer

POS	*Word*	*Lemma*	*Porter Stem*
DT	That	**that**	**that**
VBD	was	**be**	**wa**
RB	quite	**quite**	**quit**
DT	an	**a**	**an**
NN	accomplishment	**accomplishment**	**accomplish**
,	,	,	,
VBN	given	**give**	**given**
IN	that	that	that
IN	for	for	for
NNS	years	year	year
,	,	,	,
NN	law	law	law
NN	enforcement	**enforcement**	**enforc**
NNS	officials	**official**	**offici**
VBD	were	**be**	**were**

Parsing and VSO extraction

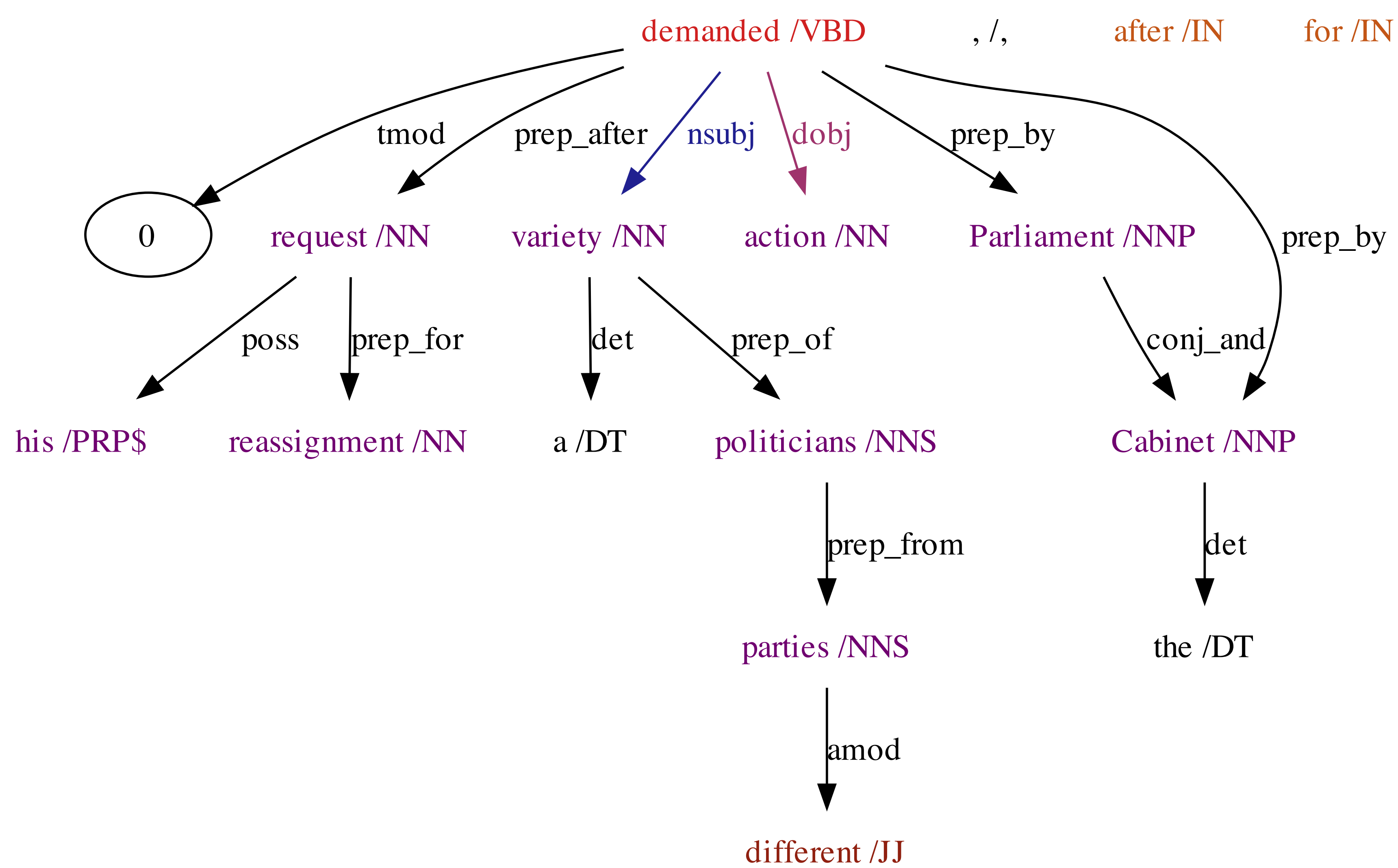
Text

Today , after his request for reassignment , a **variety** of politicians from different parties **demanded** **action** by Parliament and the Cabinet .

Constituent parse

(ROOT (S (NP-TMP (NN Today)) (, ,) (PP (IN after) (NP (NP (PRP$ his) (NN request)) (PP (IN for) (NP (NN reassignment))))) (, ,) (NP (NP (DT a) (NN variety)) (PP (IN of) (NP (NP (NNS politicians)) (PP (IN from) (NP (JJ different) (NNS parties)))))) (VP (VBD demanded) (NP (NN action)) (PP (IN by) (NP (NP (NNP Parliament)) (CC and) (NP (DT the) (NNP Cabinet))))) (. .)))

Syntactic dependencies



VSO tuple extraction

`1988.08.01.0166742 demand variety action`

$$\left(d,\ w^{(v)},\ w^{(s)},\ w^{(o)}\right)$$

"variety" vs "politicians"
"variety" vs "demanded"

Preprocessing Results

Dataset		#Docs	#Sentences	#Word tokens	#VSO tuples
CrimeNYT	parsed	27,150	788,906	20,411,164	1,252,720
Treebank: WSJ	preparsed	2,312	49,208	1,173,766	77,629
Treebank: Brown	preparsed	192	24,243	459,148	26,584

Tuple completeness (CrimeNYT)

Full: (V, S, O)	241,169	19.3%
Partial: (V, S, _)	536,353	38.0%
Partial: (V, _, O)	475,198	42.8%
Total	1,252,720	

In 1979 , policy makers did enact a modest amendment to the law , mainly to [reduce]_v the (penalties)_o for marijuana-related offenses .

Thursday, March 8, 2012

```
~/sem/semdoc/data/crime % pv  crime2.semtuple | awk '{print $4,$5,$6}' | awk
'$2=="NONE"{snone += 1} $3=="NONE"{onone += 1} $2!="NONE" && $3!="NONE"{ vso +=
1}  END{print "vso", vso, "  snone",snone,"  onone",onone}'
vso 241169   snone 475198   onone 536353
```

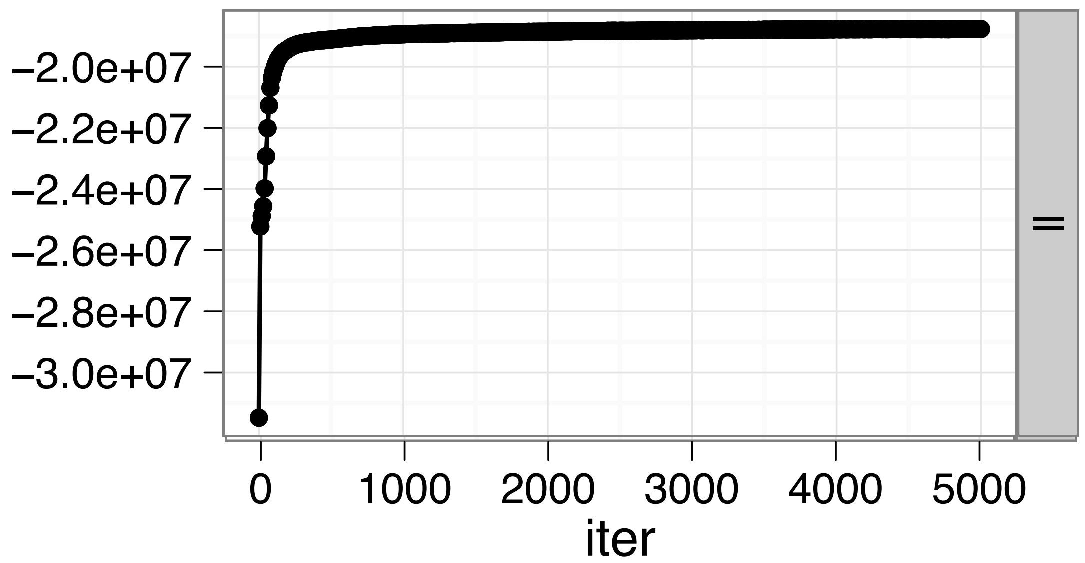
Experiments

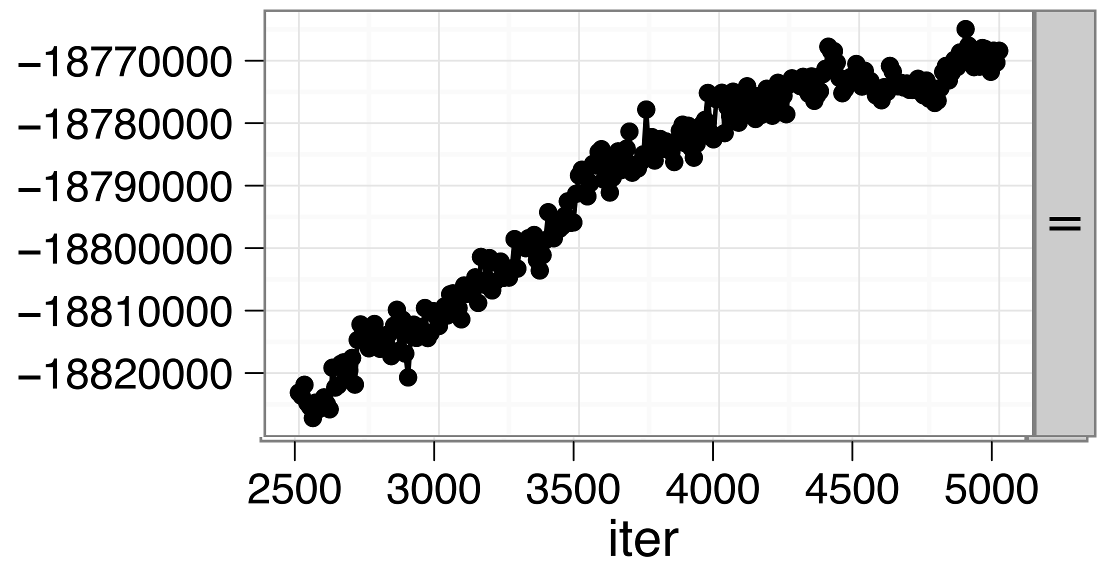
- Only use 10,000 most common words
  - Faster, though loses a lot
  - No "stopword" removal -- already filtered to content words
- 5000 Gibbs sampling iterations
  - CrimeNYT on 1 CPU: ~1 day
  - CrimeNYT on 16 CPU's: a few hours

Is my MCMC done yet?

CrimeNYT K=20 R=20

$$p(z)p(r|z)p(w|z)$$

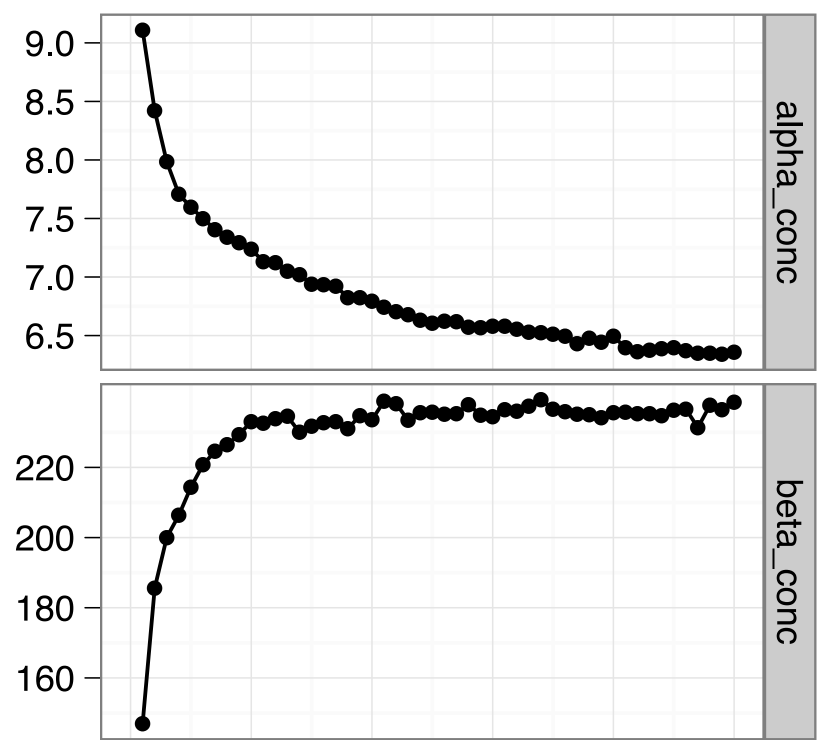




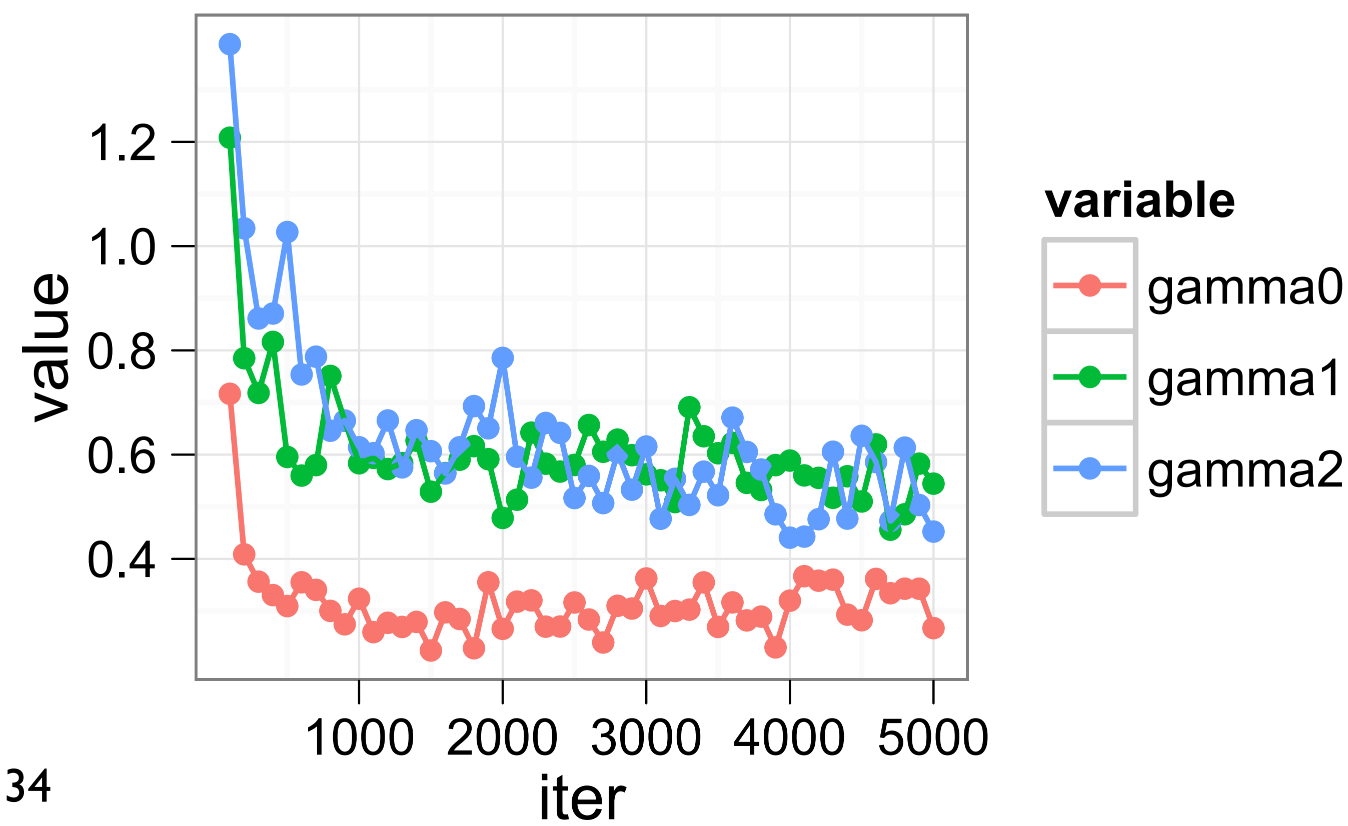
Minimal requirement: log-likelihood shouldn't be increasing.... This might be an early stop...
Concentrations might like OK:

α

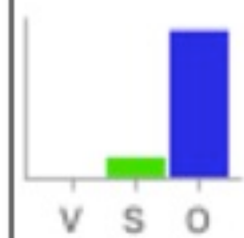
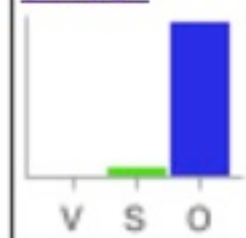
β



γ

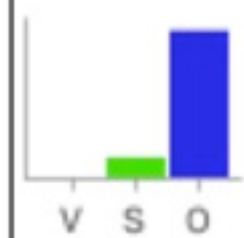
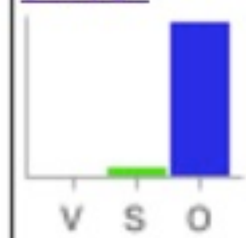


CrimeNYT *K=100 R=100*

$w \sim \phi_r$	$C[r,w]$	$C[f,a,r]$	
r=58 gun drug weapon OOV cocaine firearm handgun marijuana crack property money heroin amount card ounce pound alcohol car dealer document record item goods copy sale pistol rifle fare computer cigarette passport cash worth narcotic pornography gram force material evidence possession drinking number knife bag trade arm dollar thousand party quantity (0.762 mass)	10,901	 [0, 1300, 9601]	f=41  [0, 453, 8913]

CrimeNYT *K=100 R=100*

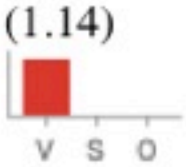
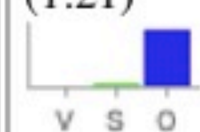
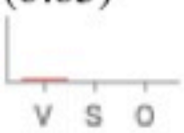
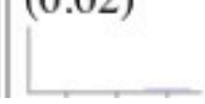
$w \sim \phi_r$ $\quad C[r, w]$ $\qquad C[f, a, r]$

r=58 gun drug weapon OOV cocaine firearm handgun marijuana crack property money heroin amount card ounce pound alcohol car dealer document record item goods copy sale pistol rifle fare computer cigarette passport cash worth narcotic pornography gram force material evidence possession drinking number knife bag trade arm dollar thousand party quantity (0.762 mass)	10,901	 [0, 1300, 9601]	f=41  [0, 453, 8913]

f=41

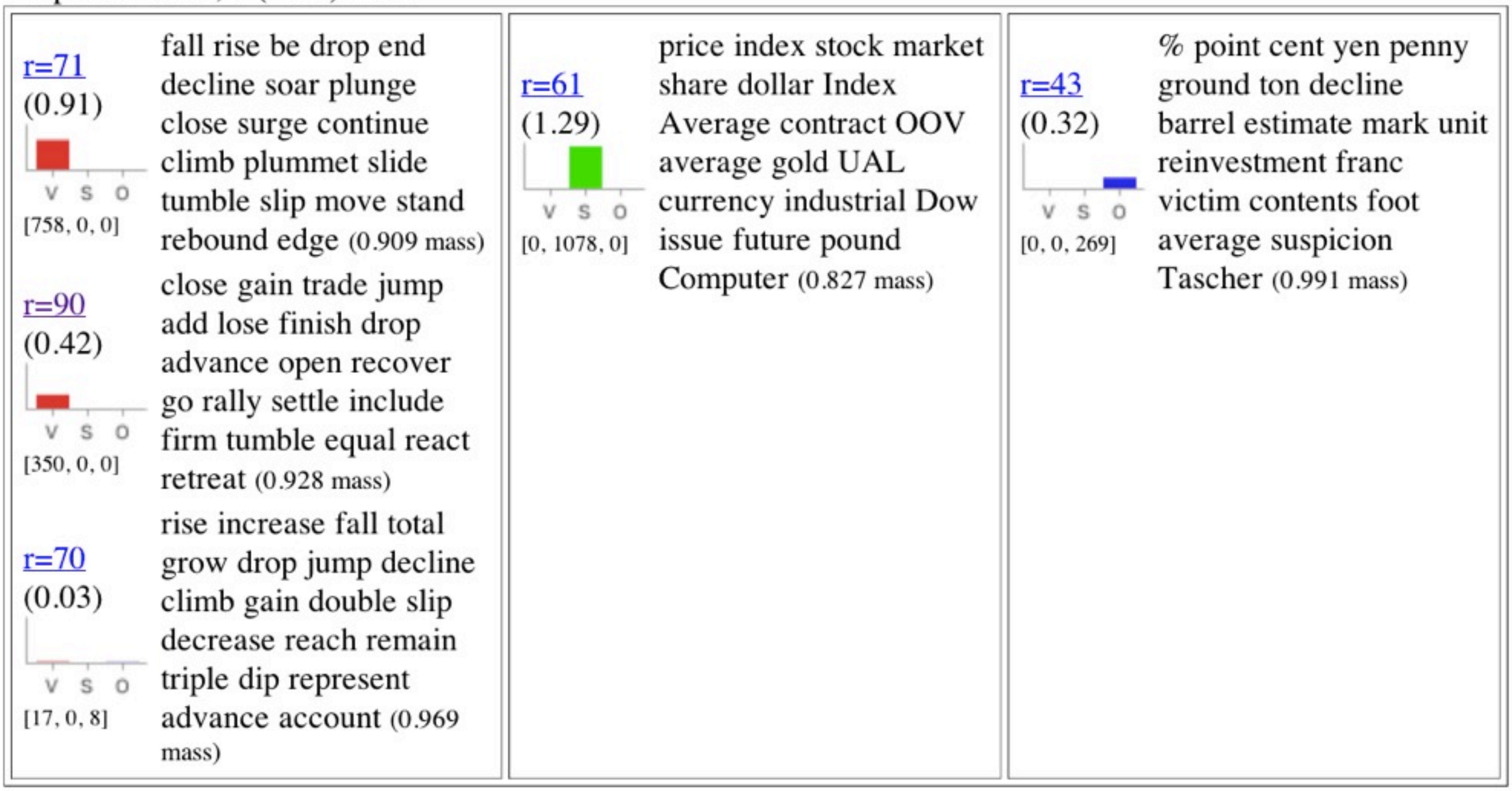
$r \sim L_{f,a}$ $\qquad C[f, a, r]$

31 top-99% roles, 20 (>1%) roles

r=96 (1.14)  [8830, 0, 0]	use sell carry buy find get match possess seize distribute steal keep involve contain identify obtain have collect give provide (0.541 mass)	r=64 (0.09)  [0, 660, 0]	officer member man prisoner guard agent OOV police criminal someone dealer soldier gang suspect worker detainee driver group killer resident (0.573 mass)	r=58 (1.21)  [0, 453, 8913]	gun drug weapon OOV cocaine firearm handgun marijuana crack property money heroin amount card ounce pound alcohol car dealer document (0.599 mass)
r=97 (0.05)  [399, 0, 0]	pay make run raise use buy sell get receive do collect control own accept extort involve lose operate obtain keep (0.448 mass)	r=28 (0.05)  [0, 363, 0]	police investigator OOV authority official detective neighbor witness resident officer Gigante agent chief newspaper report man F.B.I. worker sergeant Bureau (0.833 mass)	r=78 (0.03)  [0, 0, 215]	information evidence OOV dna test sample camera system computer record device material technology datum fingerprint database testing suspect image file (0.383 mass)
r=23 (0.04)  [327, 0, 0]	take give spend get receive have pay endanger draw need save offer earn serve hold provide lose attract deserve find (0.893 mass)	r=17 (0.05)	people OOV one man everyone other guy someone anyone nobody everybody person	r=54 (0.02) 	home gun car shot room street house apartment knife hand clothes train bag bus

WSJ *K=100 R=100*

5 top-99% roles, 6 (>1%) roles



r=71 (0.91) V S O [758, 0, 0]	fall rise be drop end decline soar plunge close surge continue climb plummet slide tumble slip move stand rebound edge (0.909 mass)
r=90 (0.42) V S O [350, 0, 0]	close gain trade jump add lose finish drop advance open recover go rally settle include firm tumble equal react retreat (0.928 mass)
r=70 (0.03) V S O [17, 0, 8]	rise increase fall total grow drop jump decline climb gain double slip decrease reach remain triple dip represent advance account (0.969 mass)
r=61 (1.29) V S O [0, 1078, 0]	price index stock market share dollar Index Average contract OOV average gold UAL currency industrial Dow issue future pound Computer (0.827 mass)
r=43 (0.32) V S O [0, 0, 269]	% point cent yen penny ground ton decline barrel estimate mark unit reinvestment franc victim contents foot average suspicion Tascher (0.991 mass)

Many ToDo's

- Evaluation!
  - Perplexities
  - Compare to FrameNet (and/or MUC?)
    - "Match-a-Linguist"
  - Human qualitative evaluation
    - semantic coherence, word similarity judgments [Chang et al 2009, Rubenstein and Goodenough 1965]
    - "Match-a-Human"
- External task?

Many ToDo's

- Better linguistics
  - More arguments, e.g. adjunct roles: "in", "on", spatial/temporal/instrument use ... real semantic role labeling
  - Noun types, coreference ...
- Incorporate document metadata
  - Plugs into hundreds of topic models using time, space, labels, etc.
- Model selection
  - K, R ?? Likelihood seems to vary
  - Non-parametric (DP / PYP) priors?
- Large-scale inference
  - 27k out 1.8 M New York Times
  - 5x more news articles out there (Gigaword)
  - 1000x more Twitter, blogs, Web data
    - Requires advances in part-of-speech tagging and parsing?

Acknowledgments

- DAP Committee: Noah Smith, Geoff Gordon, Jaime Carbonell
- Many conversations with labmates (slides with Dipanjan Das...)
- Pittsburgh Supercomputing Center