Discovering Components of Meaning in English and Chinese using Predictive Language Models (Machine Learning)
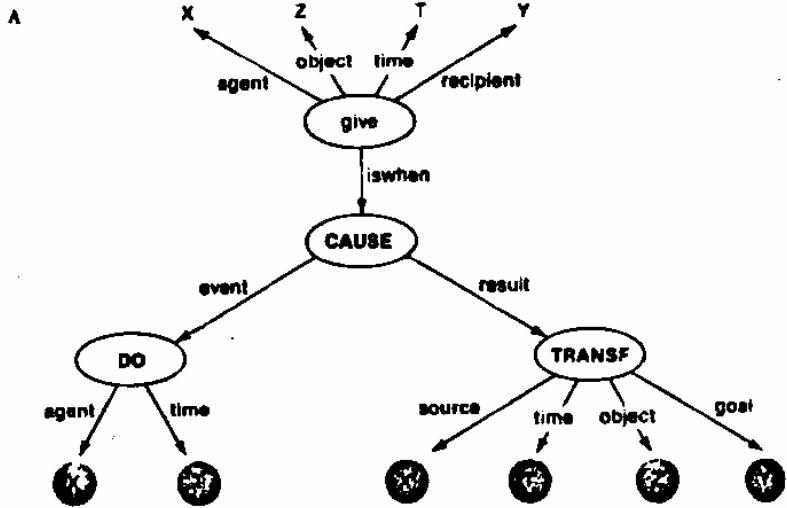
PHILLIP WOLFF

EMORY UNIVERSITY
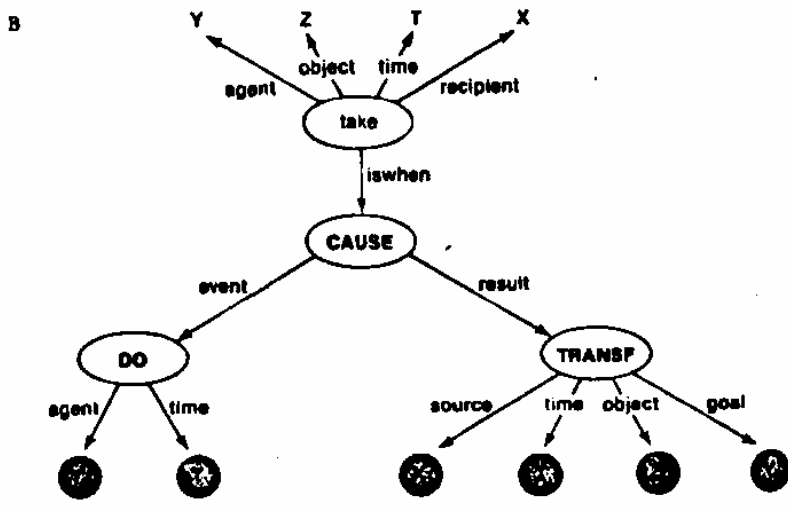PSYCHOLOGY AND LINGUISTICS

Components of meaning

Give   *Take*   Gentner (1975, 1981)
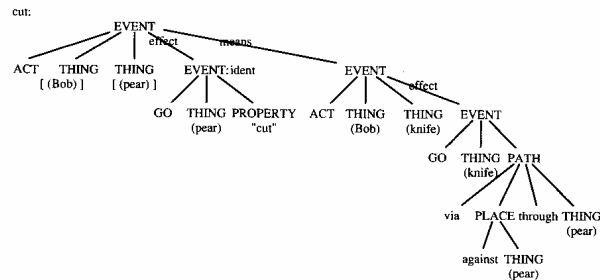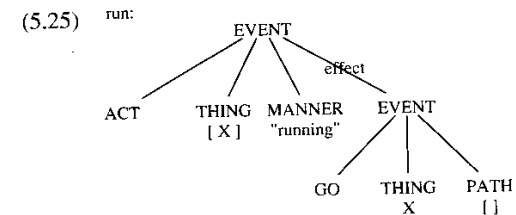
DO

CAUSE

TRANSF

Components of meaning

Pinker (1989)

cut – CAUSE, MEANS, GO

run – MANNER

Rappaport Hovav & Levin (1998)

a. Complex event schema:
$[[x\ \text{ACT}_{<MANNER>}]\ \text{CAUSE}\ [\text{BECOME}\ [y\ <RES\text{-}STATE>]]]$

b. Simple event schema:
$[x\ \text{ACT}_{<MANNER>}]$
$[x\ <STATE>]$
$[\text{BECOME}\ [x\ <STATE>]]$

Probing for components

"…it can be quite difficult to pin down the meanings of words using introspection alone" (Levin, 1993)

Syntactic behavior can be quite helpful

Hypothesis: *sink* encodes change-of-state

◦ The submarine sunk the boat.
◦ The boat sank.

Others do not (Two-argument activities)
◦ The torpedo hit the boat.
◦ *The boat hit.

Potential complications

Verbs of Cutting
- The gardener cut the branch.
- *The branch cut.

Verbs of Entity-Specific Change of State
- *The gardener bloomed the flowers.
- The flowers bloomed.

Distinguishing CoS verbs (L&R-H 1995)

Hypotheses about alternating change-of-state verbs (*Alexiadou, et al., 2005*)

- Basically **dyadic**; external argument is removed in intransitive entry

        vs.

- Basically **monadic**: external argument is added in transitive entry

Adicity of the non-alternating verbs (e.g., *bloom*)
- Basically **monadic?**

Decomposition Using Machine Learning

Extract component of meaning using machine learning

Traditional word embedding methods (**Word2vec** (Mikolov et al., 2013)

- Cosine tells gives similarity relations

  (*sink float fall* vs. *blossom flower decay*)

```
VB_say+nsubj+ccomp→       0.104814395→    0.03342392→    0.06842138→     0.022990467→   -0.022603553→
VB_say+nsubj→   0.08005889→      -0.06542914→   -0.115272105→   0.03740651→     0.024933169→    0.018
VB_have+nsubj+dobj→      -0.07441021→    -0.026547713→  0.000104090395→ -0.030716497→  0.03339178→
VB_be+expl+nsubj→        0.06403918→     -0.07272451→   -0.09833856→    -0.048553716→  0.054407034→
VB_say+ccomp+ccomp+nsubj→        0.002966348→    -0.007338829→  -0.019051638→  -0.06639934→    0.057
VB_say+nsubj+ccomp+ccomp→        0.13019288→     0.095864475→   0.036619864→   -0.08411098→    -0.02
VB_be+nsubj+ccomp→       0.0021583992→   -0.020407585→  -0.029457891→  -0.059884794→   0.0025561992→
VB_think+nsubj+ccomp→    0.04512→-0.041387927→   -0.03654735→    0.008637076→   -0.046887893→   0.037
VB_have+nsubj+xcomp→      -0.11415646→    -0.04197037→   -0.06601734→    -0.04287955→   0.051375393→
VB_be+nsubj→    -0.17823936→    -0.08829497→    0.03897092→    -0.051765792→  0.07382385→     -0.03
VB_be+nsubj+xcomp→       -0.04715332→    -0.050542176→  0.03144036→    -0.051531874→  -0.022707028→
```

Testing theories of meaning requires access to components of meaning (CAUSE CHANGE CONTACT)

Predictive Language Models

The New York Times (Oct. 18, 2018)

If a machines can read, it can write

- Robin Sloan (*Mr. Penumbra's 24-Hour Bookstore*):

- Sloan: *The bison are gathered around the canyon. …*
- Computer*: "by the bare sky."*

- Sloan*: The bison have been traveling for two years back and forth. …*
- Computer: *between the main range of the city.*
- Computer*: The slow-sweeping tug moved across the emerald harbor*.

BERT

Bidirectional **E**ncoder **R**epresentations from **T**ransformsers (Devlin et al., 2018)

Architecture:
- 12 than 24 layers
- Deep learning

Features:
- Context sensitive word representations
- Uses context before and after a word (bidirectional)
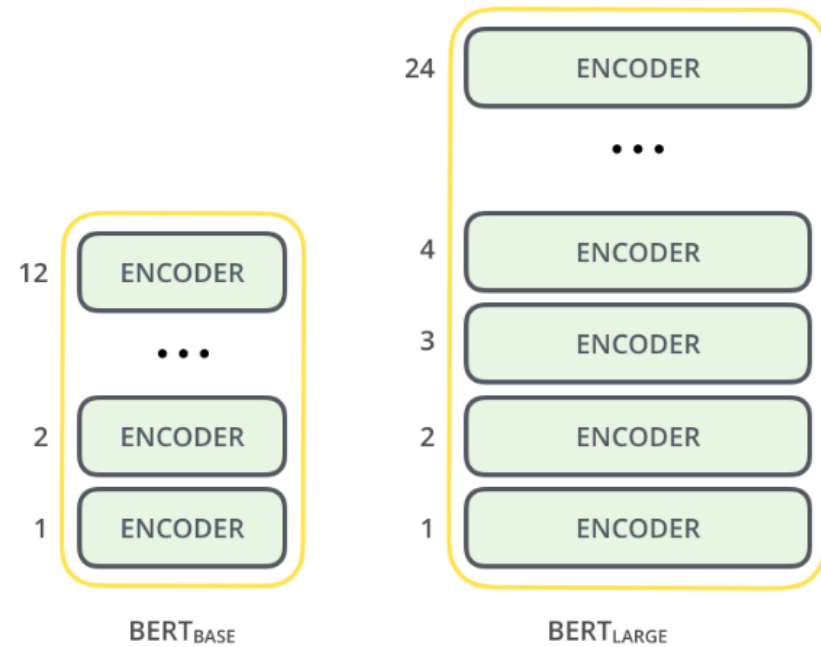
BERT training: predict masked word

Training: 3.3 billion word corpus
- ◦ BooksCorpus (800 million words)
- ◦ English Wikipedia (2.5 billion words)

Use the output of the masked word's position to predict the masked word

	Possible classes: All English words
0.1%	Aardvark
…	…
10%	Improvisation
…	…
0%	Zyzzyva

FFNN + Softmax

1 2 3 4 5 6 7 8 ••• 512

BERT

Randomly mask 15% of tokens

1   2   3   4   5   6   7   8   •••   512
[CLS] Let's stick to   [MASK]   in   this   skit

Input

[CLS] Let's stick to improvisation in this skit
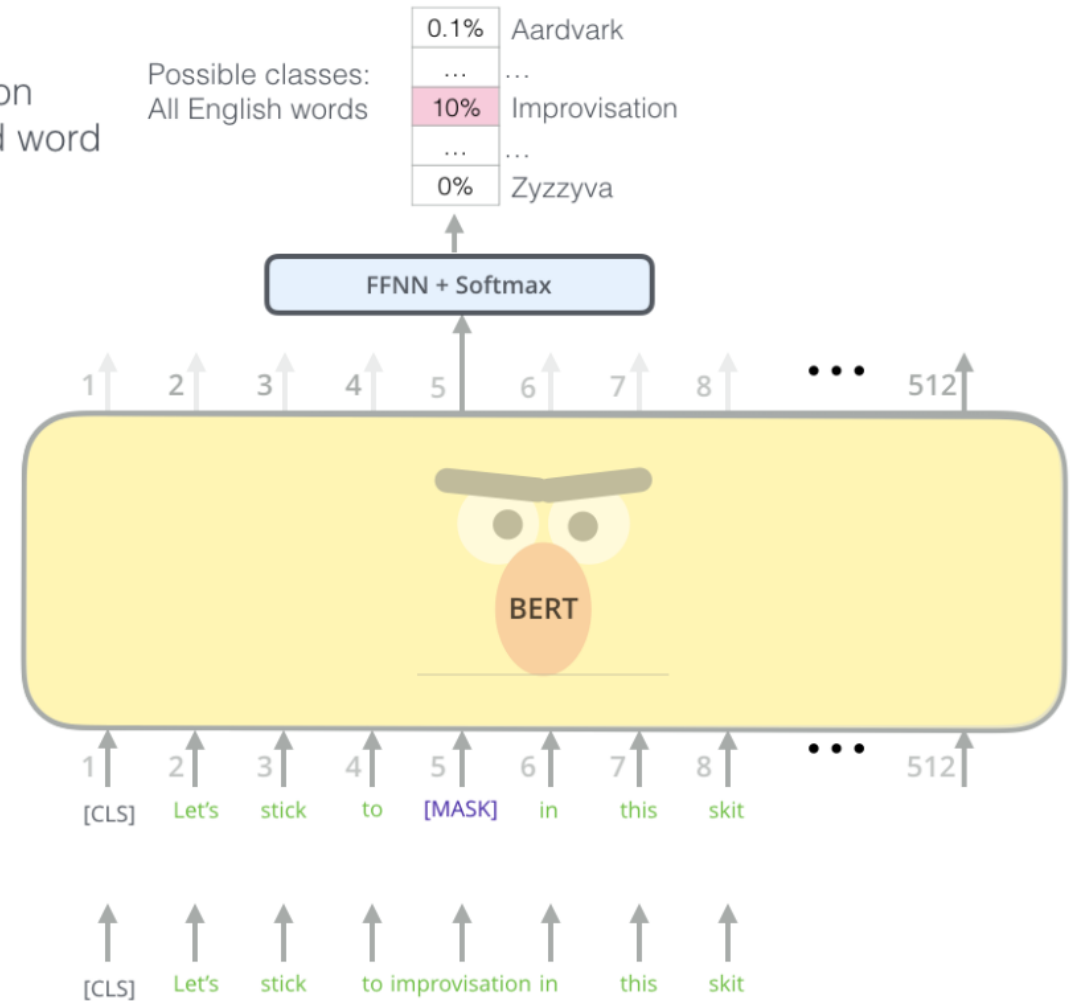
BERT training: predicting next sentence

[CLS] The man went to the store . [SEP] The man bought a gallon of milk .[SEP]

[CLS] He bought a gallon of milk . [SEP] The man went to the store . [SEP]

Predict likelihood
that sentence B
belongs after
sentence A

| 1% | IsNext |
| 99% | NotNext |

FFNN + Softmax

1 2 3 4 5 6 7 8 ... 512

BERT

Tokenized
Input

1 2 ... 512
[CLS] the man [MASK] to the store [SEP]

Input

[CLS] the man [MASK] to the store [SEP] penguin [MASK] are flightless birds [SEP]
Sentence A   Sentence B
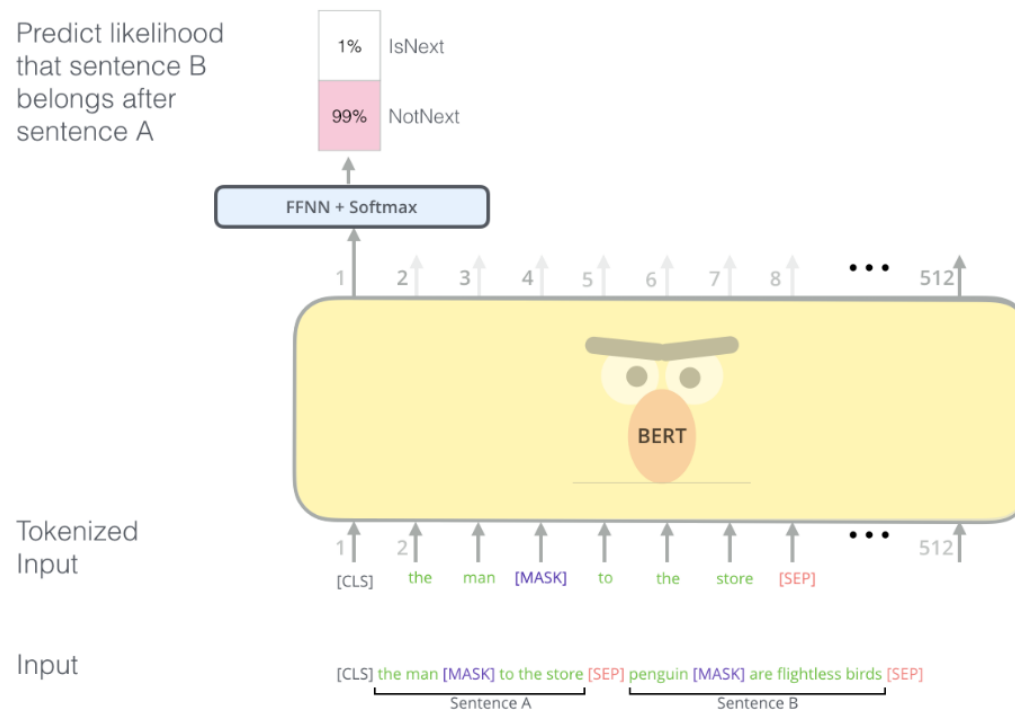
BERT: interrogating

- The man [MASK] a gallon of milk.   [MASk] = drank, grabbed, bought

- The man went to the store . He [MASK] a gallon of milk.   [MASK] = bought, lifted, grabbed, drank

- [CLS] The man went to the store . [SEP] The man bought a gallon of milk .[SEP]
  [CLS] He bought a gallon of milk . [SEP] The man went to the store . [SEP]

BERT: Syntactic Abilities

Goldberg (2019)

◦ the game that the guard hates **is** bad

◦ the game that the guard hates **are** bad

◦ [CLS] the game that the guard hates [MASK] bad .

BERT: Syntactic Abilities

Goldberg (2019)

	BERT Base	BERT Large	LSTM (M&L)	Humans (M&L)	# Pairs (# M&L Pairs)
SUBJECT-VERB AGREEMENT:					
Simple	1.00	1.00	0.94	0.96	120 (140)
In a sentential complement	0.83	0.86	0.99	0.93	1440 (1680)
Short VP coordination	0.89	0.86	0.90	0.82	720 (840)
Long VP coordination	0.98	0.97	0.61	0.82	400 (400)
Across a prepositional phrase	0.85	0.85	0.57	0.85	19440 (22400)
Across a subject relative clause	0.84	0.85	0.56	0.88	9600 (11200)
Across an object relative clause	0.89	0.85	0.50	0.85	19680 (22400)
Across an object relative (no *that*)	0.86	0.81	0.52	0.82	19680 (22400)
In an object relative clause	0.95	0.99	0.84	0.78	15960 (22400)
In an object relative (no *that*)	0.79	0.82	0.71	0.79	15960 (22400)
REFLEXIVE ANAPHORA:					
Simple	0.94	0.92	0.83	0.96	280 (280)
In a sentential complement	0.89	0.86	0.86	0.91	3360 (3360)
Across a relative clause	0.80	0.76	0.55	0.87	22400 (22400)

BERT: Performance on NLP tasks

Rank	Name	Model	URL	Score
1	GLUE Human Baselines	GLUE Human Baselines	↗	87.1
2	Microsoft D365 AI & MSR AI	MT-DNN-ensemble	↗	84.2
3	王玮	ALICE large (Alibaba DAMO NLP)		83.9
4	Stanford Hazy Research	Snorkel MeTaL	↗	83.2
5	张倬胜	SemBERT	↗	82.9
6	Anonymous Anonymous	BERT + BAM	↗	82.3
7	Jason Phang	BERT on STILTs	↗	82.0
8	Jacob Devlin	BERT: 24-layers, 16-heads, 1024-h	↗	80.5
9	Neil Houlsby	BERT + Single-task Adapters	↗	80.2
10	Alec Radford	Singletask Pretrain Transformer	↗	72.8
11	GLUE Baselines	BiLSTM+ELMo+Attn	↗	70.0
		BiLSTM+ELMo	↗	67.7

Single Task BiLSTM+ELMo+Attn	↗	66.5
Single Task BiLSTM+ELMo	↗	66.4
GenSen	↗	66.1
BiLSTM+Attn	↗	65.6
BiLSTM	↗	64.2
InferSent	↗	63.9
Single Task BiLSTM	↗	63.7
Single Task BiLSTM+CoVe	↗	63.6
BiLSTM+CoVe+Attn	↗	63.1
Single Task BiLSTM+CoVe+Attn	↗	63.1
BiLSTM+CoVe	↗	62.9
Single Task BiLSTM+Attn	↗	62.8
DisSent	↗	61.9
Skip-Thought	↗	61.3
CBOW		58.6

Analysis 1: BERT's knowledge of argument structure

Four verb classes of transitive verbs (Levin, 1993):

Change-of-state: *broke opened melted sunk froze*

Cut verbs: *cut carved crushed sliced*

Hit verbs: *hit kicked slapped pounded*

Touch verbs: *touch patted stroked nudged*

Diathesis alternations

Middle alternation
- Jerry broke the vase.
- The vase broke easily.

Body-part possessor ascension alternation
- Bill cut his thumb.
- Bill cut himself on the thumb.

Conative alternation
- Rachel cut the wood.
- Rachel cut at wood.

Causative
- Marj opened the window.
- The window opened.

Levin (1993)

	ChangeState	Cut	Hit	Touch
Middle	yes	yes	no	no
Body-part	no	yes	yes	yes
Conative	no	yes	yes	no
Causative	yes	no	no	no

Method

Use BERT to predict syntactic behavior of verbs

	CoS	Cut	Hit	Touch
Middle	yes	yes	no	no
Body-part	no	yes	yes	yes
Conative	no	yes	yes	no
Causative	yes	no	no	no

- Middle: An object [MASK] easily .
- Body-part Possessor: An object [MASK] another object on its side .
- Conative: An object [MASK] at another object .
- Inchoative: An object [MASK] .

Logged rankings for each verb obtained from BERT (minus 10)

Results

	CoS	Cut	Hit	Touch
Middle	yes	yes	no	no
Body-part	no	yes	yes	yes
Conative	no	yes	yes	no
Causative	yes	no	no	no
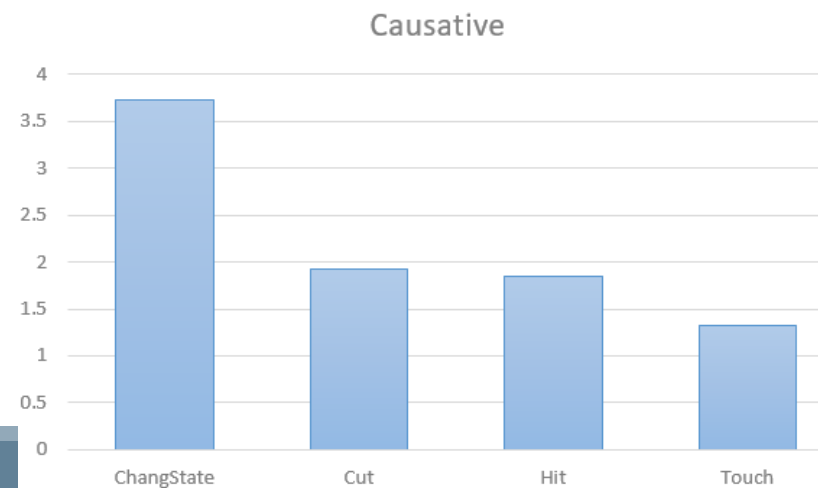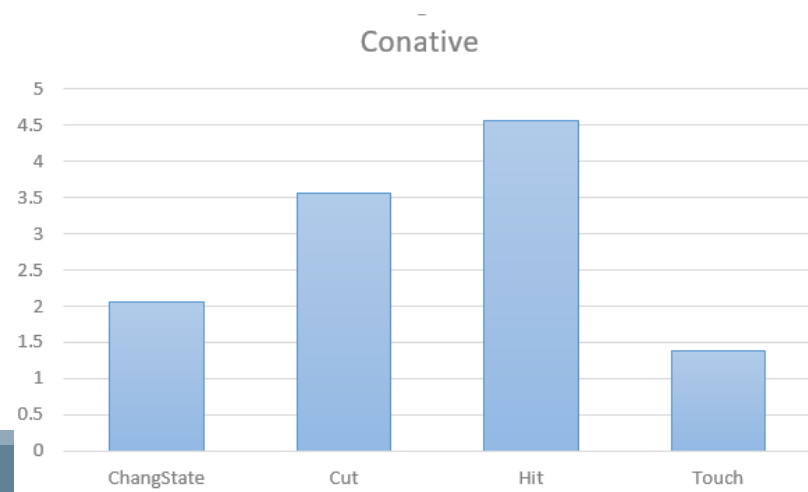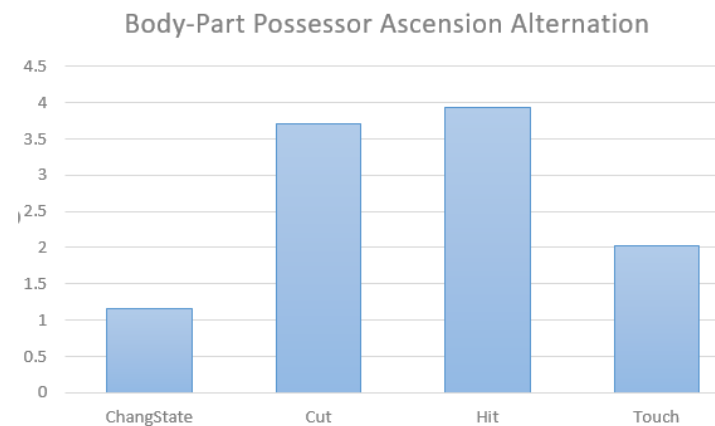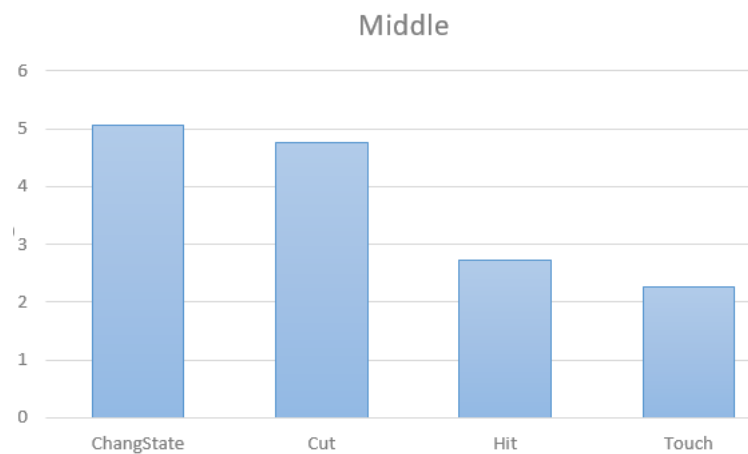
Verb classes?

Reduce dimensions using t-SNE (Maaten & Hinton, 2008)

Use K-means++ to identify clusters

Choose k (number of clusters) using Silhouette Coefficient

cut
sliced
carved crushed

touched
stroked
nudged
patted
pounded

froze
sunk
opened
melted

hit
slapped
kicked

broke_

Analysis 2: Extracting components of meaning

Example: An object *touched* another object . The object ____ ____ ____ something .

→ made contact with

Find the primitives for each verb

Look to see if there are shared primitives among class members that differentiate that class from another class

Predictions

	ChangeState	Cut	Hit	Touch	
Middle	yes	yes	no	no	CHANGE
Body-part	no	yes	yes	yes	CONTACT
Conative	no	yes	yes	no	CONTACT + MOTION
Causative	yes	no	no	no	CAUSE

Results: frequently occurring components

caused change to

made contact with

put pressure on

made impact with

made marks with

caused injury to

put stress to

put pressure to

moved nearer to

changed shape into

Components for change-of-state verbs

Break

caused change to

made contact with

lost contact with

gave form with

caused damage to

gave effect to

caused injury to

put stress to

Open

caused change to

made contact with

lost contact with

gave form with

changed shape with

created room with

gave room to

took care of

took notice of

Sink

caused change to

made contact on

lost contact with

gave form to

poured light into

held liquid from

changed shape into

shifted shape into

contained liquid from

Melt

caused change to

made contact on

lost contact with

gave form to

absorbed heat from

became liquid or

turned liquid to

added energy to

lost energy to

Components *cut* verbs of

Cut	*Carve*	*Slice*	*Crush*
caused change to	**changed shape into**	**caused changes to**	**caused changes to**
made contact with	makes form into	**made contact with**	**made contact with**
divided parts of	creates forms of	cut cuts to	**lost contact with**
makes marks on	made shape into	caused wounds on	put weight on
creates damage to	turned stone into	cut tissue on	made impact on
split bones with	transformed shape into	caused injury on	caused impact to
being divides from	transformed stone into	created wounds to	forced contact with
split pieces of	shaped stone into	touched tissue on	dropped weight on
split pieces from			

Components for *hit* verbs

Hit

made contacts by

made contact with

made impact on

made touch with

made contacts on

found contact with

put damages on

entered touch with

Kick

made contacts by

made contact with

put pressure on

made impact on

placed pressure on

made touch with

found impact with

found contact with

Slap

made contacts by

made contact with

put pressure on

made impact on

left marks on

placed pressure on

made touch with

moved causes hitting

Pound

made contacts by

made contact with

put pressure on

made impact on

made contacts on

caused damage against

Components for touch verbs

Touch	Nudge	Pat	Stroke
put pressure on	**put pressure on**	**put pressure on**	**put pressure on**
gave form to	**gave form to**	**gave form to**	**gave form to**
changed form with	made impact with	hit causes hitting	changed shape with
placed pressure on	made touch with	gave form to	gave form to
added meaning to	came nearer with	kept slapping with	hit causes hitting
entered contact with	moved nearer to	hitting causes hitting	made contacts with
brushed anything with	placed weight on	shifted shape with	hitting causes hitting

Components by verb class

	ChangeState	Cut	Hit	Touch	Levin (1993)	BERT
Middle	yes	yes	no	no	CHANGE	CAUSE+CHANGE
Body-part	no	yes	yes	yes	CONTACT	- LOST CONTACT WITH
Conative	no	yes	yes	no	CONTACT + MOTION	?
Causative	yes	no	no	no	CAUSE	LOST CONTACT WITH
With/Against	no	no	yes	no		MADE CONTACT BY
??	no	no	no	yes		PUT PRESSURE ON
Conative?	yes	no	no	yes		GAVE FORM

Paula hit the stick against/on the fence.

Paula hit the fence with the stick.

Analysis 3: components of meaning for intransitives

Question: What components of meaning are stored in the verb's root as opposed to the syntax?

Causative: CAUSE+CHANGE

Verbs of entity-specific change of state: *decay bloom flower blossom blush*
- Transitive: not available
- Intransitive: CHANGE

Verbs of change of state: *broke opened melted sunk froze*
- Transitive: CAUSE+CHANGE
- Intransitive: CAUSE+CHANGE? CHANGE?

Predictions

Verbs of Change of State
- Inherently dyadic → CAUSE+CHANGE in the intransitive
- Inherently monadic → CHANGE in the intransitive

Verbs of entity-specific change of state
- Inherently monadic → CHANGE in the intransitive

Methods

Determine components in the intransitive

Verbs of change of state:
◦ *broke opened melted sunk froze*

Verbs of entity-specific change of state:
◦ *decay bloom flower blossom blush*

An object verbed .

The object ____ ____ ____ .

Entity-specific change of state verbs

Bloom

becomes,material,itself

becomes,fruit,again

produces,fruit,again

contains,energy,itself

produces,energy,itself

produces,seed,again

produces,energy,itself

see,change,again

cause,lives,decay

cause,cycle,decay

Decay

becomes,material,completely

loses,energy,itself

loses,state,itself

become,future,completely

becomes,death,again

becomes,energy,completely

lose,colour,again

loses,material,itself

cause,death,again

cause,material,itself

Flower

grows,fruit,again

produces,fruit,again

becoming,fruit,again

produces,seed,again

growing,fruit,again

becomes,fragment,decay

produce,change,again

cause,fragment,decay

blush

turn,colors,again

becomes,consciousness,completely

became,color,again

contains,energy,itself

becomes,appearance,disappear

turn,appearance,disappear

glow,change,again

cause,colour,again

Change of state verbs

Broke

changed,value,automatically

changed,position,automatically

changed,status,automatically

lost,value,automatically

caused,motion,inside

caused,movement,instead

caused,force,instead

received,damage,instead

received,force,instead

received,movement,instead

Open

changed,shape,automatically

changed,position,automatically

took,shape,automatically

changed,behavior,slightly

caused,momentum,within

caused,power,within

caused,weight,within

received,energy,slightly

received,properties,slightly

Sink

changed,course,slightly

changed,behavior,slightly

changed,properties,slightly

suffered,damages,there

caused,alarm,there

caused,damages,there

caused,confusion,there

received,energy,slightly

Melt

changed,behavior,slightly

changed,appearance,slightly

changed,properties,slightly

received,heat,only

cause,state,itself

absorbed,energy,slightly

received,damage,only

received,energy,slightly

Results

Both CoS and Entity-specific CoS verbs
- CAUSE, CHANGE
- "ITSELF" (disconnected process)

CoS verbs only
- RECEIVE ENERGY

Analysis 5: What are the possible constructions and what do they mean?

Method

- ◦ Find the most common sentence types
- ◦ Find verbs that most often fill those types

Corpora

25 years, 42,833,581 sentence; 957,290,668 raw words

4 weeks, 123,732,422 sentences; 2,200,871,688 raw words

3 days, 43,665,391, 1,025,479,505 raw words

Sample posts from Reddit

1512086400   Always inviting you to go out to eat. My spoiled rich Malaysian friend causally asks me to get $25 ramen.

1512086400   Seriously? What planet do you live on? If the woman who accused Matt Lauer of harassing her at the Olympics was not on the trip to the Olympics, or if she claimed he had done something to her during a meeting, but either he or she was not at the meeting. What good would an alibi be if you couldn't prove that you weren't in a place a crime was committed? Please.

1512086403   You paid too much for it. Amazon is showing $10.97 for me.

Parsing

Stanford parser Klein & Manning (2003)
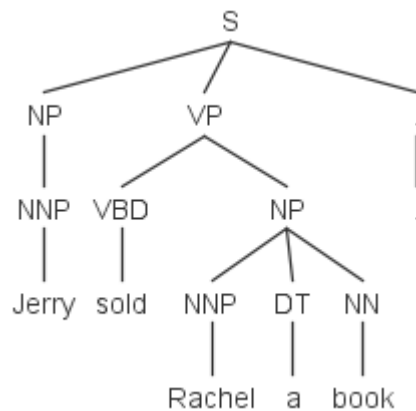
Dependency Parsing (Tesnière(1959)
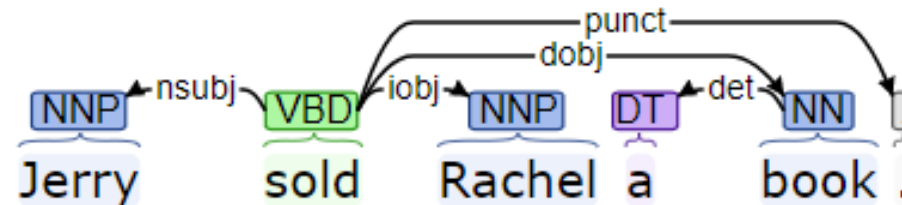
VB_sold+nsubj+iobj+dobj

http://corenlp.run/

Dependency parses

Intuition linguistic units, e.g. words, are connected by directed links

Common dependencies

- **nsubj** – subject
- **dobj** – direct object
- **iobj** – indirect object
- **ccomp** - clausal complement of a verb or adjective
- **xcomp** - open clausal complement of a verb or an adjective is a predicative or clausal complement without its own subject
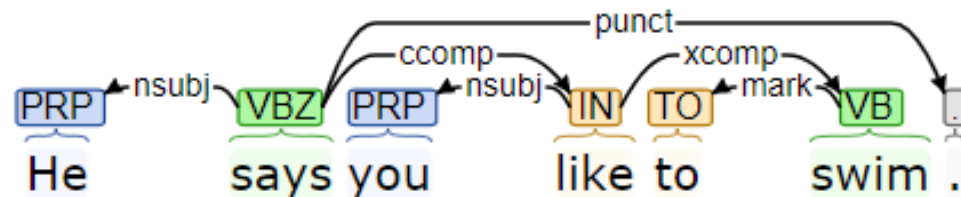
Dependency parses

- **nmod:to** – a noun functioning as a non-core (oblique) argument or adjunct
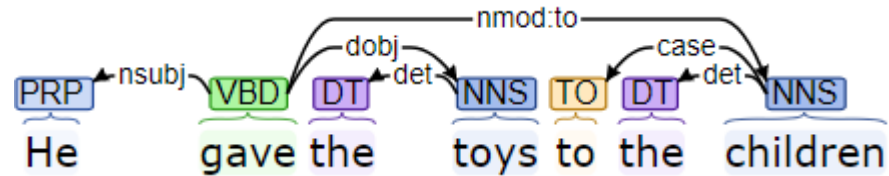
- **nmod:for**
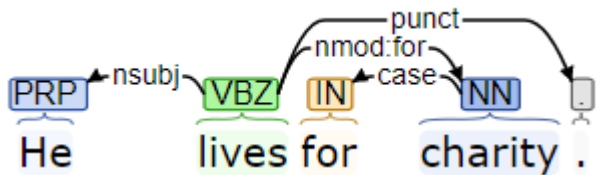
- **compound:prt -** phrasal verb particle
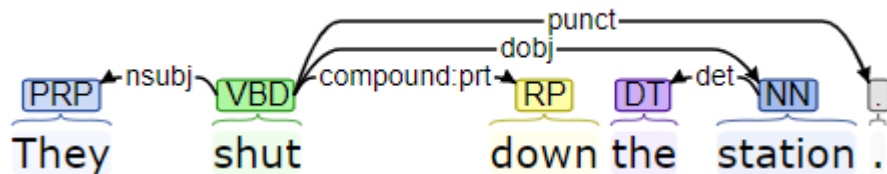
Most frequent dependencies in English

'auxpass','advmod','nsubj','aux','mark','advcl','dobj','ccomp','xcomp','iobj','nmod','nmod:above','nmod:against','nmod:agent','nmod:among','nmod:at','nmod:before','nmod:beneath','nmod:between','nmod:by','nmod:down','nmod:except','nmod:for','nmod:from','nmod:in','nmod:including','nmod:into','nmod:npmod','nmod:of','nmod:on','nmod:out_of','nmod:over','nmod:per','nmod:poss','nmod:than','nmod:through','nmod:tmod','nmod:to','nmod:under', 'nmod:until','nmod:upon','nmod:with','compound:prt','neg','expl','cc','det','cc:preconj','mwe','discourse','csubj','case','nmod:poss','csubjpass','nmod:poss','csubjpass','acl','appos','cop','nummod','acl:relcl','det:predet'

Re-expressing sentences

1. Lemmatize the words

2. Add part-of-speech markers

3. Attach dependencies governed by the verb ono the verb

- We live on borrowed time.
- PRP_we VB_live+nsubj+nmod:on IN_on VB_borrow NN_time .

```
NNP_Brother IN_of NNP_John NNP_Boeckels ._.↓
JJ_funeral IN_on NNP_Monday CD_9 NNP_AM IN_from DT_the NNP_Volk N
JJ_funeral NN_liturgy CD_10 VB_be+nsubj+nmod:at IN_at NNP_St. NNF
NN_burial NNP_Calvary NNP_Cemetery ,_, NNP_Patterson ,_, NNP_New
VB_visit+dobj+dobj+dobj NNP_Saturday CD_7-9 CC_and NNP_Sunday CD_
DT_the NN_victim ,_, NNP_Dr. NNP_Victoria NNP_P. NNP_Shoaf ,_, CD
DT_the NN_friend ,_, WP$_whose NN_name VB_be VB_withhold+auxpass
```

Basic sentences

◦ Find most frequent dependency patterns for each verb (N = 3574 with n > 500)

◦ Combine across verbs

Results: Basic sentences

Reddit	TRANSITIVE	1	nsubj+dobj	VB_have	VB_get	VB_do	VB_make	VB_love	VB_need	VB_take	VB_use	VB_see
NYT	TRANSITIVE	1	nsubj+dobj	VB_have	VB_include	VB_make	VB_take	VB_do	VB_see	VB_get	VB_use	VB_need
Reddit	THINKING	2	nsubj+ccomp	VB_think	VB_be	VB_know	VB_say	VB_mean	VB_make	VB_hope	VB_guess	VB_see
NYT	THINKING	2	nsubj+ccomp	VB_say	VB_be	VB_think	VB_know	VB_believe	VB_add	VB_help	VB_suggest	VB_show
Reddit	INTRANSITIVE	3	nsubj	VB_be	VB_do	VB_feel	VB_look	VB_sound	VB_happen	VB_go	VB_see	VB_agree
NYT	INTRANSITIVE	3	nsubj	VB_say	VB_be	VB_do	VB_add	VB_come	VB_look	VB_go 32751	VB_feel	VB_write
Reddit	WANTING	4	nsubj+xcomp	VB_want	VB_need	VB_seem	VB_make	VB_look	VB_try	VB_feel	VB_use	VB_keep
NYT	WANTING	4	nsubj+xcomp	VB_be	VB_want	VB_seem	VB_make	VB_become	VB_need	VB_try	VB_appear	VB_begin
Reddit	MOVING	5	nsubj+advmod	VB_go	VB_come	VB_work	VB_run	VB_move	VB_fall	VB_walk	VB_stay	VB_stand
NYT	MOVING	12	nsubj+advmod	VB_go	VB_move	VB_fall	VB_stand	VB_exist	VB_walk	VB_stay	VB_matter	VB_step
Reddit	DIRECTIONS	6	nsubj+nmod:to	VB_go	VB_lead	VB_listen	VB_refer	VB_apply	VB_move	VB_belong	VB_respond	VB_reply
NYT	DIRECTIONS	6	nsubj+nmod:to	VB_lead	VB_go	VB_belong	VB_refer	VB_point	VB_return	VB_contribute	VB_move	VB_amount
Reddit	GIVING	7	nsubj+iobj+dobj	VB_give	VB_send	VB_owe	VB_hand					
NYT	GIVING	11	nsubj+iobj+dobj	VB_give	VB_hand	VB_award	VB_fine					
Reddit	INTERACTING	8	nsubj+nmod:with	VB_agree	VB_come	VB_disagree	VB_deal	VB_stick	VB_interact	VB_mess	VB_combine	VB_compete
NYT	INTERACTING	15	nsubj+nmod:with	VB_deal	VB_end	VB_finish	VB_disagree	VB_coincide	VB_compare	VB_compete	VB_interfere	VB_contrast
Reddit	CAUSING/EXPECT	9	nsubj+dobj+xcomp	VB_allow	VB_expect	VB_cause	VB_force	VB_encourage	VB_enable	VB_convince	VB_urge	VB_prompt
NYT	CAUSING/EXPECT	7	nsubj+dobj+xcomp	VB_allow	VB_expect	VB_ask	VB_require	VB_urge	VB_force	VB_enable	VB_encourage	VB_order
Reddit	COMMUNICATING	10	nsubj+dobj+ccomp	VB_tell	VB_assure	VB_convince	VB_amaze	VB_inform	VB_sadden			
NYT	COMMUNICATING	9	nsubj+dobj+ccomp	VB_tell	VB_remind	VB_assure	VB_inform	VB_convince				
Reddit	STAYING	11	nsubj+nmod:in	VB_live	VB_result	VB_stay	VB_sit	VB_stick	VB_invest	VB_specialize	VB_participate	VB_engage
NYT	STAYING	5	nsubj+nmod:in	VB_live	VB_die	VB_specialize	VB_result	VB_end	VB_occur	VB_participate	VB_sit	VB_stand
Reddit	CHANGING	14	nsubj+compound:prt+dobj	VB_set	VB_pick	VB_strike	VB_pull	VB_lay	VB_shut	VB_hand	VB_knock	VB_wipe
NYT	CHANGING	16	nsubj+compound:prt+dobj	VB_bring	VB_pick	VB_turn	VB_shut	VB_mess	VB_wipe	VB_slow	VB_tear	VB_knock

Observations

Source	Category	#	Pattern									
Reddit	TRANSITIVE	1	nsubj+dobj	VB_have	VB_get	VB_do	VB_make	VB_love	VB_need	VB_take	VB_use	VB_see
NYT	TRANSITIVE	1	nsubj+dobj	VB_have	VB_include	VB_make	VB_take	VB_do	VB_see	VB_get	VB_use	VB_need
Reddit	THINKING	2	nsubj+ccomp	VB_think	VB_be	VB_know	VB_say	VB_mean	VB_hope	VB_guess	VB_see	
NYT	THINKING	2	nsubj+ccomp	VB_say	VB_be	VB_think	VB_know	VB_believe	VB_add	VB_help	VB_suggest	VB_show
Reddit	INTRANSITIVE	3	nsubj	VB_be	VB_do	VB_feel	VB_look	VB_sound	VB_happen	VB_go	VB_see	VB_agree
NYT	INTRANSITIVE	3	nsubj	VB_say	VB_be	VB_do	VB_add	VB_come	VB_look	VB_go 32751	VB_feel	VB_write
Reddit	WANTING	4	nsubj+xcomp	VB_want	VB_need	VB_seem	VB_make	VB_look	VB_try	VB_feel	VB_use	VB_keep
NYT	WANTING	4	nsubj+xcomp	VB_be	VB_want	VB_seem	VB_make	VB_become	VB_need	VB_try	VB_appear	VB_begin
Reddit	MOVING	5	nsubj+advmod	VB_go	VB_come	VB_work	VB_run	VB_move	VB_fall	VB_walk	VB_stay	VB_stand
NYT	MOVING	12	nsubj+advmod	VB_go	VB_move	VB_fall	VB_stand	VB_exist	VB_walk	VB_stay	VB_matter	VB_step
Reddit	DIRECTIONS	6	nsubj+nmod:to	VB_go	VB_lead	VB_listen	VB_refer	VB_apply	VB_move	VB_belong	VB_respond	VB_reply
NYT	DIRECTIONS	6	nsubj+nmod:to	VB_lead	VB_go	VB_belong	VB_refer	VB_point	VB_return	VB_contribute	VB_move	VB_amount
Reddit	GIVING	7	nsubj+iobj+dobj	VB_give	VB_send	VB_owe	VB_hand					
NYT	GIVING	11	nsubj+iobj+dobj	VB_give	VB_hand	VB_award	VB_fine					
Reddit	INTERACTING	8	nsubj+nmod:with	VB_agree	VB_come	VB_disagree	VB_deal	VB_stick	VB_interact	VB_mess	VB_combine	VB_compete
NYT	INTERACTING	15	nsubj+nmod:with	VB_deal	VB_end	VB_finish	VB_disagree	VB_coincide	VB_compare	VB_compete	VB_interfere	VB_contrast
Reddit	CAUSING/EXPECT	9	nsubj+dobj+xcomp	VB_allow	VB_expect	VB_cause	VB_force	VB_encourage	VB_enable	VB_convince	VB_urge	VB_prompt
NYT	CAUSING/EXPECT	7	nsubj+dobj+xcomp	VB_allow	VB_expect	VB_ask	VB_require	VB_urge	VB_force	VB_enable	VB_encourage	VB_order
Reddit	COMMUNICATING	10	nsubj+dobj+ccomp	VB_tell	VB_assure	VB_convince	VB_amaze	VB_inform	VB_sadden			
NYT	COMMUNICATING	9	nsubj+dobj+ccomp	VB_tell	VB_remind	VB_assure	VB_inform	VB_convince				
Reddit	STAYING	11	nsubj+nmod:in	VB_live	VB_result	VB_stay	VB_sit	VB_stick	VB_invest	VB_specialize	VB_participate	VB_engage
NYT	STAYING	5	nsubj+nmod:in	VB_live	VB_die	VB_specialize	VB_result	VB_end	VB_occur	VB_participate	VB_sit	VB_stand
Reddit	CHANGING	14	nsubj+compound:prt+dobj	VB_set	VB_pick	VB_strike	VB_pull	VB_lay	VB_shut	VB_hand	VB_knock	VB_wipe
NYT	CHANGING	16	nsubj+compound:prt+dobj	VB_bring	VB_pick	VB_turn	VB_shut	VB_mess	VB_wipe	VB_slow	VB_tear	VB_knock

1. NYT and Reddit results align closely
   ◦ Share most similar basic sentences
   ◦ Share most frequent verbs for these sentences

2. Eleven categories
   ◦ TRANSITIVE, THINKING, INTRANSITIVE, WANTING, MOVING, DIRECTION, GIVING, INTERACTING, CAUSING, COMMUNICATING, CHANGING

3. CAUSING (nsubj+dobj+xcomp) and WANTING (nsubj+xcomp) highly related

4. THINKING (nsubj+ccomp) and WANTING (nsubj+xcomp) highly related

5. STAYING (nsubj+nmod:(in)), MOVING (nsubj+nmod:(to)), and INTERACTING (nsubj+nmod:with) related

6. TRANSITIVE and INTRANSITIVE related: cover full range of topics

7. Could reflect major ontological distinctions in the English verb system

8. Could point to conceptual primitives

9. These results will likely not change

Are the categories conceptual?

Mandarin

- pro-drop language

Dependencies
◦ "dep","advmod:loc","nmod:topic","nsubjpass", "aux:modal", "appos", "discourse", "parataxis:prnmod", "aux:asp","nmod:poss","name","nmod:range","amod","auxpass","amod:ordmod","mark:clf","advmod","ccomp","KILL", "dobj","punct","aux:prtmod","advmod:dvp","acl","aux:ba","conj","nmod:tmod","nmod:prep","etc","cop" "nsubj","advcl:loc","case","det","xcomp","nmod","cc","advmod:rcomp","nummod","mark","neg","nsubj:xsubj", "compound:vc","nmod:assmod","compound:nn"

Weibo

- much like Twitter

- lots of ads

Method

Space

"我有一个小秘密偷偷的告诉你哦"↓
""你的小苹果被我吃掉了！"//@夏影:啊啊啊为什么小熊猫能这么甜//@醋味儿小给:啾咪//@r__i__
"【声控福利】糖醋里脊 这道菜教你什么叫"宽油劝退！" 现在是#美食101#最终C位争夺战，你们的每
"超喜欢小姐姐//@日本流行每日速报:转发投票，送我最爱的美食博主上C位！#美食101#之#美食C位
"姐妹们记住了！！！ #电视剧知否知否应是绿肥红瘦#"↓

" 我 有 一个 小 秘密 偷偷 的 告诉 你 哦 "↓
" " 你 的 小 苹果 被 我 吃 掉 了 ！ "//@ 夏影 : 啊啊 啊 为什么 小熊猫 能 这么 甜//@ 醋味儿
" 【 声控 福利 】 糖醋 里脊 这 道 菜 教 你 什么 叫 " 宽 油 劝退 ！ " 现在 是# 美食 101# 最终 (
"超 喜欢 小 姐姐 //@ 日本 流行 每日 速报 : 转发 投票 ， 送 我 最爱 的 美食 博主 上 C 位 !#
" 姐妹 们 记住 了 ！ !!# 电视剧 知 否 知 否 应 是 绿肥红瘦 #"↓
" 奶奶 说 得 对 //@ Az___s : 记住 了 !//@ YouTube 上 的 搬运工 : //@ 安生 的 大白 : 记 着
" 唯美 的 日出 / 日落 ？ ？ 📷? 摄影师 : BenMulder# 你 拍 过 最 好看 的 日出 日落 #"↓
" 好好 看 !! ! //@ 时尚 达人 强子 : 好看 //@ 这 只 兔子 叫 DOKI : 天啊 好 美 ! //@ 毕 恐
" " 他 不 会 有 愧疚 ， 之后 遇到 的 人 足够 好 的 话 ， 他 甚至 不 会 回忆 起 你 。 " "↓

Method

Parse

PN_我 AD_正在 VV_使用+nsubj+advmod+dobj+conj+discourse PU_" NR_赵 NN_丽颖+punct+compound:nn PU_" PU_# AD_微博
NR_ PU_" PU_" NR_骁 NN_龙+nmod:assmod CD_636 CC_和 CD_6 NN_GB+dep JJ_超大 NN_内存+compound:nn+conj+cc+compoun
NT_4月 NT_1日 NT_...+compound:nn+compound:nn NN_全文+compound:nn PU_" PU_" PU_# NN_魅 NN_蓝 NN_E3+compound:nn
NT_现在+case LC_起 PU_, AD_即 VV_可 VV_下单+advmod:loc+advmod+aux:modal+ccomp+conj VV_购买+dobj PU_# NR_vivoA
JJ_全新 NN_科技+amod NN_时尚 NN_单品+compound:nn+compound:nn PU_, VV_想+nsubj+ccomp+conj VV_要 VV_猛戳+xcomp |
NT_4月 NT_9日+compound:nn VV_开奖+nmod:tmod PU_！↓
NR_ PU_" PU_" AD_还 VA_行+dep+advmod+conj PU_, AD_就 VC_是 NN_电 VE_没 DER_得 VA_快+advmod+cop+dep+dep+dep PL
AD_不过 VV_充+mark DEC_的 NR_也快@ NN_神秘+acl+nmod:assmod VV_念念+ccomp VV_寻+dobj AD_@hly VA_甜蜜+advmod+mar
PU # NR 吴 VV 亦凡+punct+nsubj PU # PU .↓

Results: Basic sentences

Category	Relations									
TRANSITIVE	nsubj+dobj dobj	VB_have VV_查看 190749 check	VB_get VV_看 183277 look	VB_do VV_吃 181851 eat	VB_make VV_做 177976 do	VB_love VV_转发 172109 retweet	VB_need VV_拍 139608 take a photo/sla	VB_take VV_去 110659 go	VB_use VV_爱 89223 love	VB_see VV_打 66881 hit
WANTING	nsubj+xcomp ccomp	VB_be ccomp want	VB_want VV_想 227424 hope	VB_seem VV_希望 121909 like	VB_make VV_喜欢 77371 see	VB_become VV_看到 52834 know	VB_need VV_知道 51532 feel	VB_try VV_感觉 50541 follow	VB_appear VV_关注 49209 think	VB_begin VV_觉得 48008 see
THINKING	nsubj+ccomp nsubj+ccomp	VB_say VV_说 353654 say	VB_be VV_称 93126 claim	VB_think VV_表示 73823 claim	VB_know VV_觉得 54297 think	VB_believe VV_认为 51947 think	VB_add VV_知道 46960 know	VB_help VV_显示 32775 show	VB_suggest VV_发现 30001 found out	VB_show VV_叫 26069 call
CAUSE	nsubj+dobj+xcomp dobj+ccomp	VB_allow VV_让 468634 let	VB_expect VV_请 76482 invite/please	VB_ask VV_给 44723 give	VB_require VV_使 42283 let/cause	VB_urge VV_帮 34911 help	VB_force VV_祝 32972 hope	VB_enable VV_要 31200 ask for/demand	VB_encourage VV_令 30844 cause/command	VB_order VV_教 23882 teach
INTRANSITIVE	nsubj nsubj	VB_say VV_受伤 20841 be injured	VB_be VV_死亡 14149 die	VB_do VV_加油 12363 cheer	VB_add VV_发生 11459 happen	VB_come VV_感 10661 feel	VB_look VV_拍# 9088 take a photo/sla	VB_go 32751 VV_十足 8430 complete	VB_feel VV_疼 7272 hurt	VB_write VV_采访 6891 interview
TRANSITIVE	nsubj+dobj nsubj+dobj	VB_have VV_拍 188703 take a photo/slap	VB_get VV_叫 35314 call	VB_do VV_在 14025 be in/at somepla	VB_make VV_发生 12552 happen	VB_love VV_拍# 11506 take a photo/sl	VB_need VV_成 10925 become	VB_take VV_供 9348 give	VB_use VV_问 8728 ask	VB_see VV_切 8490 cut
COMMUNICATING	nsubj+dobj+ccomp nsubj+dobj+ccomp	VB_tell VV_告诉 20478 tell	VB_remind VV_叫 17972 call	VB_assure VV_问 16946 ask	VB_inform VV_邀 1847 invite	VB_convince VV_视 1762 see	VV_喊 1506 shout	VV_劝 748 persuade	VV_指责 642 accuse	VV_敦促 293 urge
DIRECTIONS	nsubj+nmod:to?? xcomp	VB_lead VV_旅行 26648 travel	VB_go VV_收听 7712 listen to	VB_belong VV_巡山 3161	VB_refer VV_上班 1609 go to work	VB_point VV_旅游 1562 travel	VB_return VV_恨 1340 hate	VB_contribute VV_上学 1090 go to school	VB_move VV_举手 1056 raise hand	VB_amount VV_上课 890 go to class
MOVING	nsubj+advmod nsubj+case	VB_go VV_结束 8474 end/stop	VB_move VV_来去 2270 come and go	VB_fall VV_赶到 1531 arrive	VB_stand VV_烧开 1524 boil	VB_exist VV_煮开 1203 come	VB_walk VV_来临 1024 broadcast	VB_stay VV_播出 627 build up	VB_matter VV_建成 450 arrive	VB_step VV_到场 360

Conclusions

Machine learning methods can be used to
- explore argument structure
- Identify components of meaning

Transitivity
- May be licensed by a component of meaning implying RECEPTION of energy

Syntactic patterns/constructions

Have frequently occurring meanings

Are stable across corpora

Are consistent across languages → reflect basic conceptual distinctions