CMPT 825
Natural Language Processing

Angel Xuan Chang

angelxuanchang.github.io/nlp-class

Adapted from material by Anoop Sarkar

Some examples of NLP tasks

Identifying confusable drug names

G. Kondrak and B. Dorr

Table 4 Top 8 names that are most similar to *Toradol* according to the BI-SIM similarity measure, and the corresponding recall values

	Name	Score	+/−	Recall
1.	*Tramadol*	0.6875	+	0.25
2.	*Tobradex*	0.6250	−	0.25
3.	*Torecan*	0.5714	+	0.50
4.	*Stadol*	0.5714	−	0.50
5.	*Torsemide*	0.5000	−	0.50
6.	*Theraflu*	0.5000	−	0.50
7.	*Tegretol*	0.5000	+	0.75
8.	*Taxol*	0.5000	−	0.75

Word Segmentation (in Chinese)

北京大学生体育馆

- 北京 (Beijing) 大学生 (university students) 体育馆 (gym)
The gym for university students in Beijing.

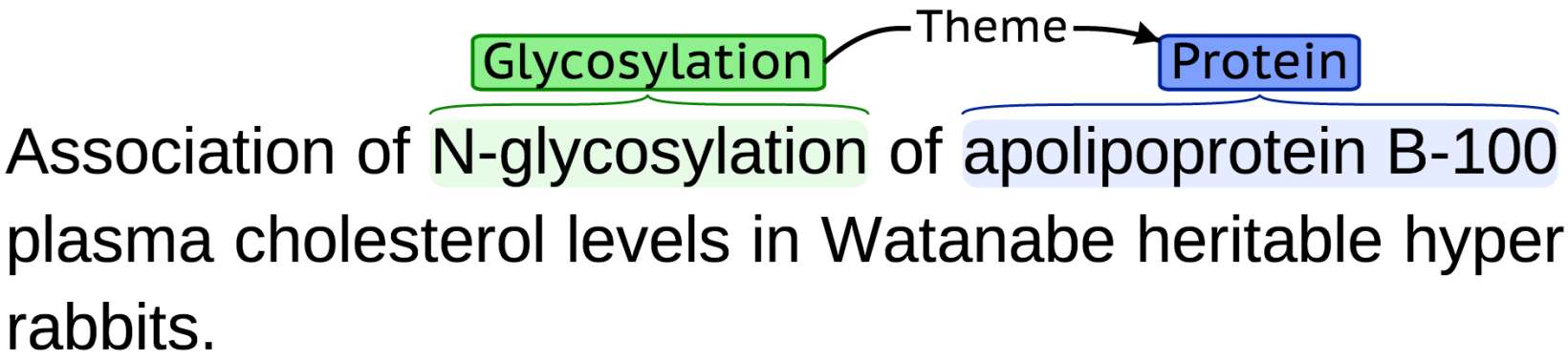
-北京大学 (Peking University) 生 (give birth to) 体育馆 (gym)
Peking University gave birth to the gym?

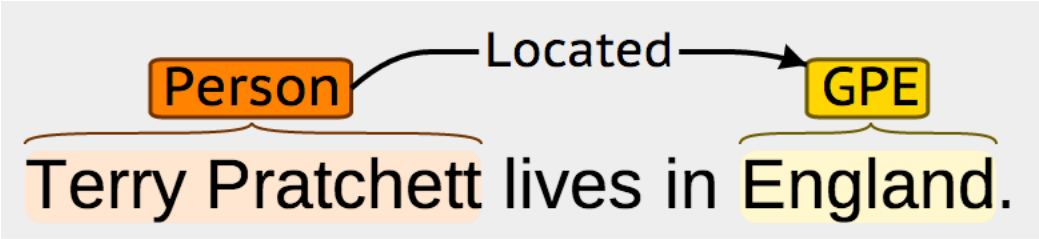
Information Extraction from Text

Finding named entities

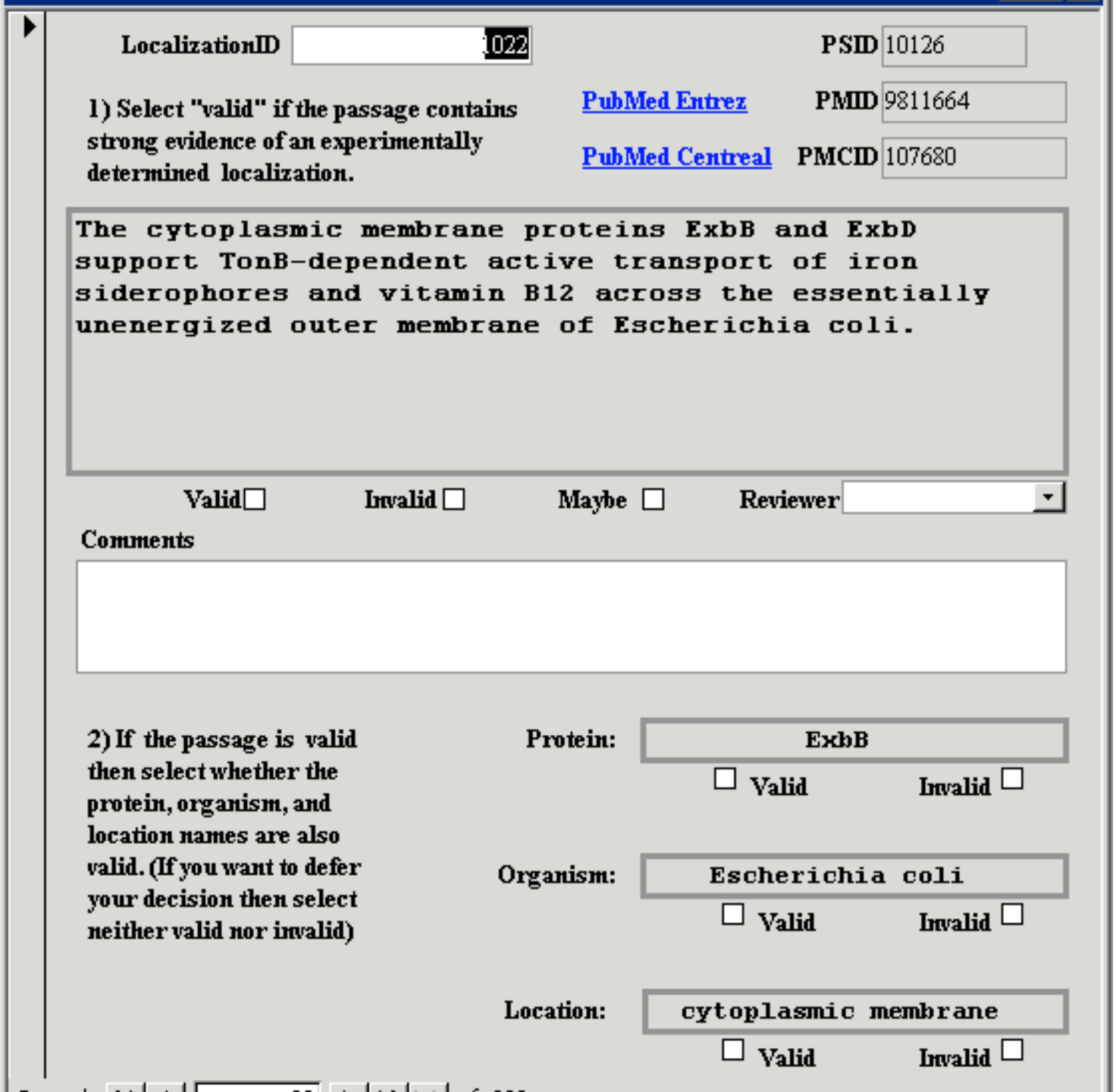
Así lo explicó hoy el presidente del Gobierno [ORG] español, José María Aznar [PER], en la conferencia de prensa con la que concluyó la XIII Cumbre Hispano-francesa [MISC], celebrada en Santander [LOC], con asistencia del presidente francés, Jacques Chirac [PER] ; del primer ministro, Lionel Jospin [PER], y trece miembros de ambos gabinetes.

Relation Extraction





Relation Extraction



PICO frames (Cochrane)

Intervention Participants
Randomized controlled study of chemoimmunotherapy with bestatin of acute nonlymphocytic leukemia in adults.

Intervention
A new immunomodulating agent, bestatin (INN: Ubenimex has low to has sig
immunological response.

ervention
A co estatin immunotherap othera
adul s performed.

After induction of complete remission, patients were randomized to the the co

Participants
Intervention
The 101 eligible cases (bestatin: 48, control: 53) were analyzed; the Intervention bestatin group achieved Outcomes longer remission than the contro

Outcomes
longer s

Though ficant in the up in the 15-49 yr a
was sig

Interventi tcomes
Bestatin is shown to be a clinically useful drug for immunotherapy of adult ANLL, since it has prolonged survival and remission

Outcomes
side-effects.

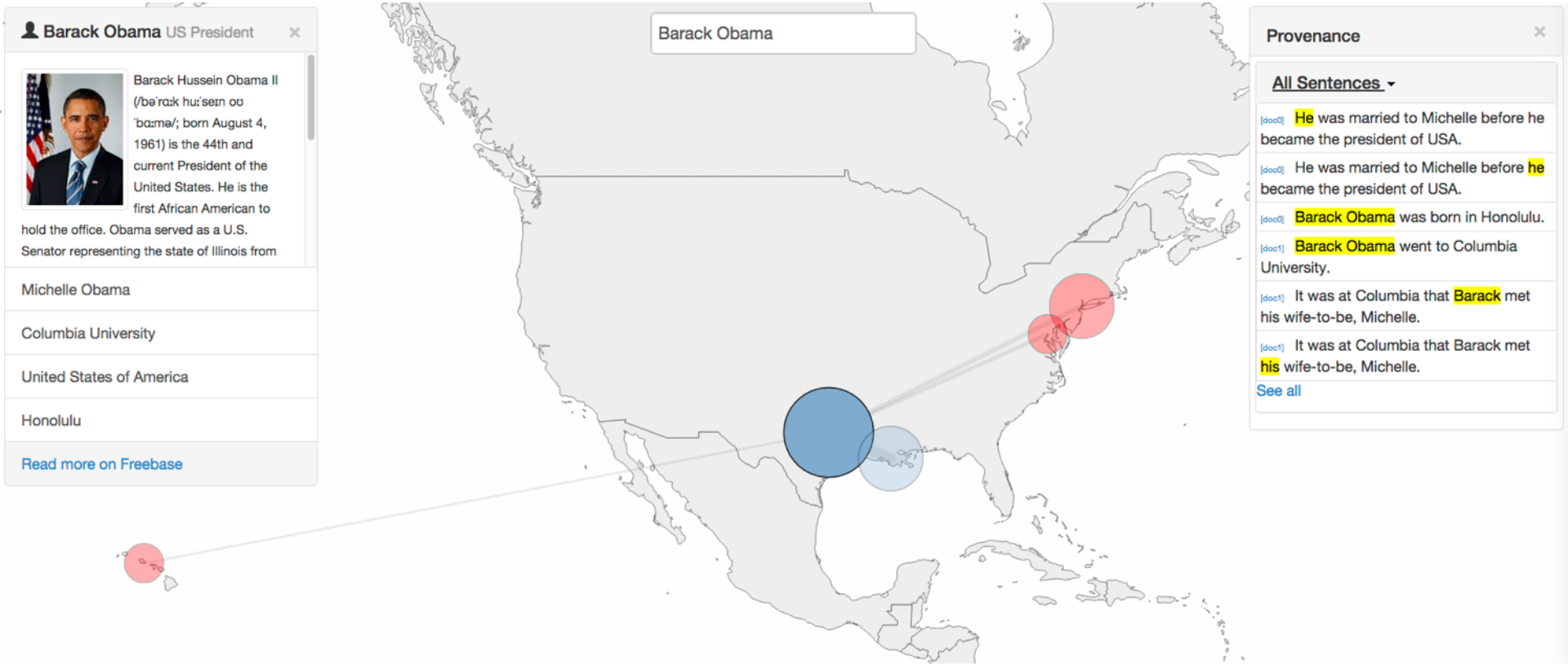
Intervention:
What is the intervention under consideration for participants?

Participants: Patient, Population or Problem
What are the characteristics of the patient or population (demographics, risk factors, pre-existing conditions, etc)?
What is the condition or disease of interest?

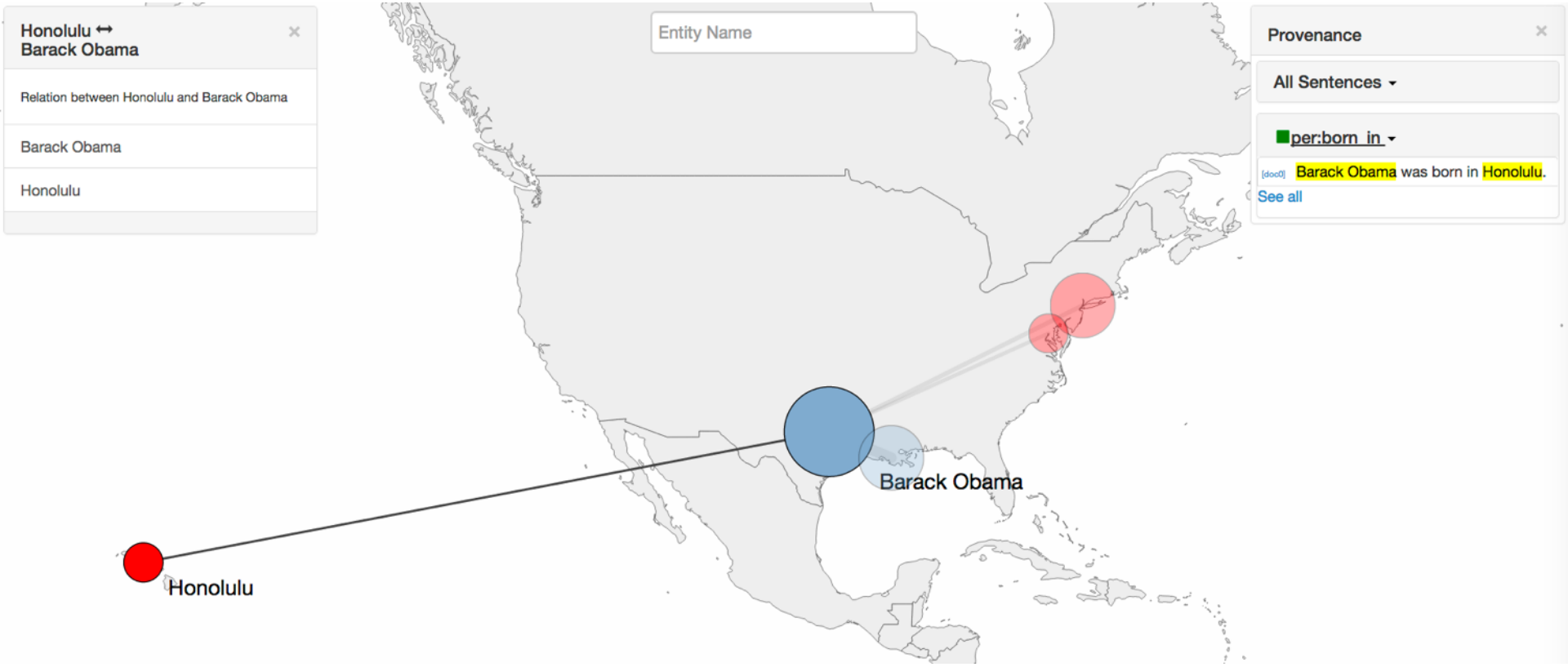
Comparison:
What is the alternative to the intervention (e.g. placebo, different drug, surgery)?

Outcomes:
quality of life, change in clinical status, morbidity, adverse effects, complications

Knowledge Graphs from Text



http://sameersingh.org/er-visualizer/



http://sameersingh.org/er-visualizer/

Text Summarization

SQuASH: SFU QA Summarization System

Input: 25 news articles, Complex question **Output:** 250-word summary

Q. Describe developments in the movement for the independence of Quebec from Canada.

A. Canadian Prime Minister Jean Chretien has ruled out the possibility that the French-speaking province of Quebec could declare independence even if a majority of Quebeckers vote for secession in the next referendum. Premier Lucien Bouchard and the separatist Parti Quebecois have retained control of Canada's largest province, possibly setting the stage for a bitter confrontation with the rest of Canada, and possibly another referendum on independence. Canada's Supreme Court managed to rule on the question of Quebec secession without infuriating leaders on either side of a bitter ideological divide. ...

by *ent423* , *ent261* correspondent updated 9:49 pm et , thu march 19 , 2015 (*ent261*) a *ent114* was killed in a parachute accident in *ent45* , *ent85* , near *ent312* , a *ent119* official told *ent261* on wednesday . he was identified thursday as special warfare operator 3rd class *ent23* , 29 , of *ent187* , *ent265* . `` *ent23* distinguished himself consistently throughout his career . he was the epitome of the quiet professional in all facets of his life , and he leaves an inspiring legacy of natural tenacity and focused

. . .

ent119 identifies deceased sailor as **X** , who leaves behind a wife

by *ent270* , *ent223* updated 9:35 am et , mon march 2 , 2015 (*ent223*) *ent63* went familial for fall at its fashion show in *ent231* on sunday , dedicating its collection to `` mamma '' with nary a pair of `` mom jeans '' in sight . *ent164* and *ent21* , who are behind the *ent196* brand , sent models down the runway in decidedly feminine dresses and skirts adorned with roses , lace and even embroidered doodles by the designers ' own nieces and nephews . many of the looks featured saccharine needlework phrases like `` i love you ,

. . .

X dedicated their fall fashion show to moms

CNN/Daily News dataset:
https://arxiv.org/pdf/1506.03340.pdf

Headline Generation

Headline A: US launches air raids in Somalia

Headline B: Somalia says dozens killed in US attack

Headline C: Many dead after US strike in Somalia

Headline D: US Launches New Attacks in Somalia

Headline E: US strikes terrorist targets in Somalia

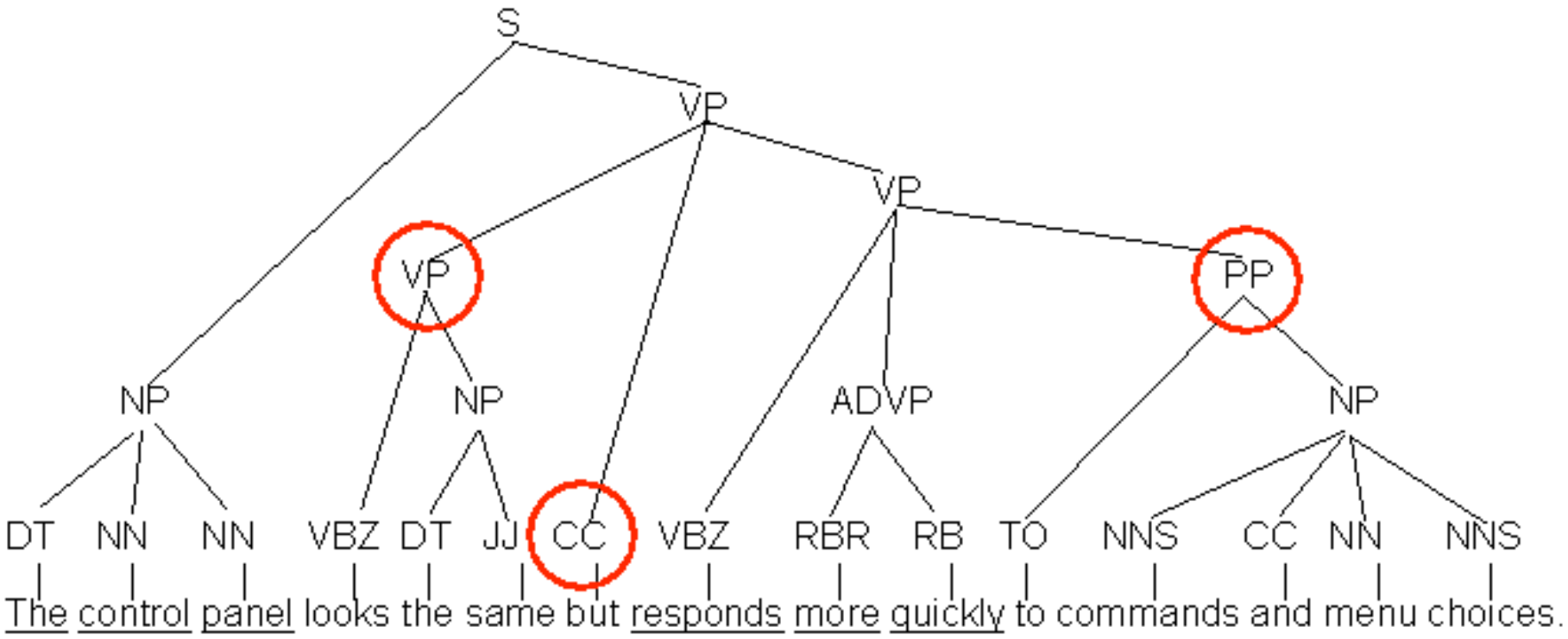
Cluster of headlines for an event on Google News

Headline Generation

Headline Candidate	Score
Bush to sign of	-22.614
Bush to sign bill on	-26.652
Bush to sign of the	-26.835
the House of The Internet gambling	-29.946
The bill of the Internet gambling	-29.982
Bush to end of the Internet gambling	-32.576
Bush to sign bill on the Internet gambling	-35.746
Bush to sign bill on the Internet gambling law	-39.710
Bush to end of the Internet gambling on The Senate bill	-46.988
Bush to sign bill on the Internet gambling site of The law	-50.912

Table 5.9: Top Headlines for "Law on Internet Gambling" news story

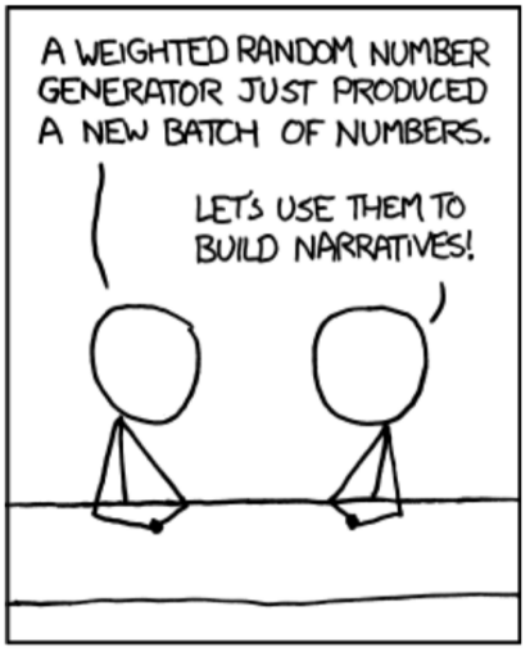
Sentence Compression



Natural Language Generation

TEAM	WIN	LOSS	PTS	FG_PCT	RB	AS ...
Heat	11	12	103	49	47	27
Hawks	7	15	95	43	33	20

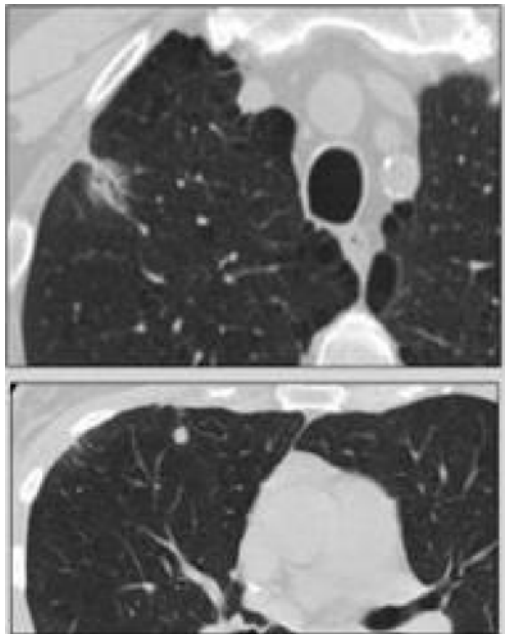
PLAYER	AS	RB	PT	FG	FGA	CITY ...
Tyler Johnson	5	2	27	8	16	Miami
Dwight Howard	11	17	23	9	11	Atlanta
Paul Millsap	2	9	21	8		
Goran Dragic	4	2	21	8		
				7		
				8		
				3		



Natural Language Generation

The Atlanta Hawks defeated the Miami Heat, 103 - 95, at Philips Arena on Wednesday. Atlanta was in desperate need of a win and they were able to take care of a shorthanded Miami team here. Defense was key for the Hawks, as they held the Heat to 42 percent shooting and forced them to commit 16 turnovers. Atlanta also dominated in the paint, winning the rebounding battle, 47 - 34, and outscoring them in the paint 58 - 26. The Hawks shot 49 percent from the field and assisted on 27 of their 43 made baskets. This was a near wire-to-wire win for the Hawks, as Miami held just one lead in the first five minutes. Miami (7 - 15) are as beat-up as anyone right now and it's taking a toll on the heavily used starters. Hassan Whiteside really struggled in this game, as he amassed eight points, 12 rebounds and one blocks on 4 - of - 12 shooting ...

Natural Language Generation



Primary finding: 23x18 mm nodule in right upper node

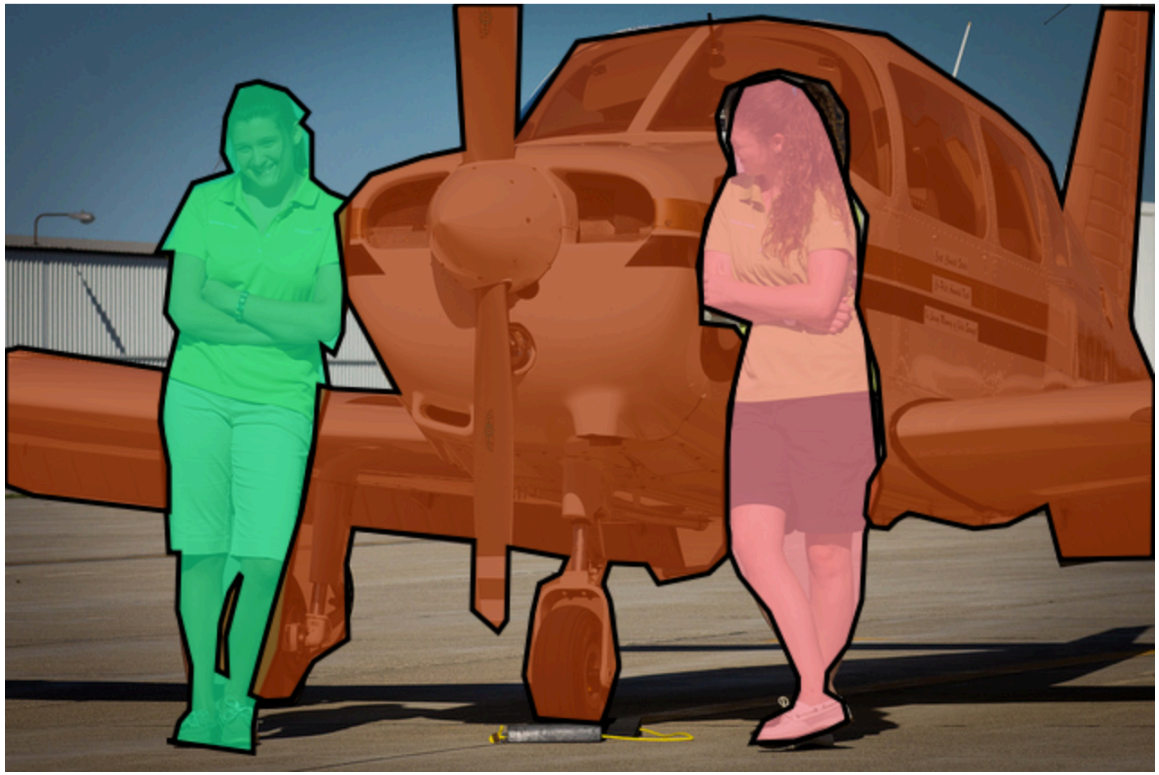
Features: Mixed solid and ground glass attenuation. >2cm from carina. No atelectasis. Contains visceral pleura. Satellite micronodule in same lung.

Additional Findings: 15mm right paratracheal node; 10mm subcarinal node. Small right pleural effusion. No visible metastases. Subpleural interstitial fibrosis.

Conclusion: Probable primary lung cancer. Clinical stage: Stage IV. Other differential: Fungal infection.

Natural Language Generation

MS Coco Dataset



- two ladies in polo shirts are leaning against an airplane.
- two girls standing next to a propeller of a small plane.
- two girls in lime green polo shirts leaning against a small propellor aircraft.
- two girls are standing near the propellers of an airplane.
- two women standing near the front of a plane

Translation

Machine Translation

MT uses parallel corpora to automatically learn a translation

SOURCE: 目前 , 某些 西方 国家 已经 宣布 终止 对 津巴布韦 的 经济援助 .

H1: at present , some western nations have already announced their termination of economic aid to zimbabwe .
H2: at present , certain western countries have already suspended their economic aids to zimbabwe .
H3: so far , some western countries have declared ending economic aid to zimbabwe .
H4: some western countries have already halted economic aid to zinbarbwe at present .

SYSTEM: at present , some western countries have announced the* end* of the* financial* assistance* to zimbabwe .
Hn: different human translators

Open Source Machine Translation! openmt.net

Chatbots and Dialog Agents

ubuntuaddicted	what's my ip?	[02:59]
DF3D2	k11: so I reinstalled fglrx manually, and startx just keeps saying "no protocol specified"	[02:59]
nahtnam	ubuntuaddicted: Are you in europe?	[03:00]
xtpeeps	Anyone can introduce me some interest channel of irc:p THX	[03:00]
timwis	hey guys, just did a fresh install on a Lenovo yoga to Pro, and I'm getting Wi-Fi is disabled	[03:01]
DF3D2	k11: and time out in locking the .Xauthority file	[03:01]
Bashing-om	DF3D2: Before you rebooted, did you do -> sudo amdconfig --initial <- ??	[03:01]
timwis	this article suggests I modify ideapad-laptop.c but it doesn't seem to exist on the filesystem wireless-lan/	[03:01]
xangua	!alis | xtpeeps	[03:01]
ubottu	xtpeeps: alis is a services bot that can help you find channels. Read "/msg alis help list" usage: /msg alis list #ubuntu* or /msg alis list *http*	[03:01]
DF3D2	Bashing-om: yes	[03:01]
ubuntuaddicted	nahtnam, no, why?	[03:01]
DF3D2	Bashing-om: I also did rm -r ~/.Xauthority as I saw suggested on the web, didn't help	[03:02]
cfhowlett	timwis, yep. only took me 3 years to learn. hit the windows wifi switch but experiment	[03:02]
cfhowlett	timwis, ctrl, alt, shift and super keys are all candidates	[03:03]
timwis	that article actually suggests that with the Lenovo laptops there's a problem beyond that	[03:04]
timwis	what is the super key?	[03:04]
cryptodan	the windows key	[03:04]
cfhowlett	timwis, aka "windows" key	[03:04]
timwis	ah! super indeed	[03:04]
somsip	timwis: windows key, or mod key, between left ctrl and left alt usually	[03:04]

https://github.com/rkadlec/ubuntu-ranking-dataset-creator

Paraphrasing

- open borders imply increasing racial fragmentation in ***european countries*** .
- open borders imply increasing racial fragmentation in *the countries of europe* .
- open borders imply increasing racial fragmentation in *european states* .
- open borders imply increasing racial fragmentation in *europe* .
- open borders imply increasing racial fragmentation in *european nations* .
- open borders imply increasing racial fragmentation in *the european countries* .

Why is paraphrasing useful?

Natural Language Inference (NLI)

#	Premise	Hypothesis	Label
1	ALT , AST , and lactate were elevated as noted above	patient has abnormal lfts	entailment
2	Chest x-ray showed mild congestive heart failure	The patient complains of cough	neutral
3	During hospitalization , patient became progressively more dyspnic requiring BiPAP and then a NRB	The patient is on room air	contradiction
4	She was not able to speak , but appeared to comprehend well	Patient had aphasia	entailment
5	T1DM : x 7yrs , h/o DKA x 6 attributed to poor medication compliance , last A1c [** 3-23 **] : 13.3 % 2	The patient maintains strict glucose control	contradiction
6	Had an ultimately negative esophagogastroduodenoscopy and colonoscopy	Patient has no pain	neutral
7	Aorta is mildly tortuous and calcified .	the aorta is normal	contradiction

Samples from the MedNLI Corpus

Sentiment

http://www.machinedlearnings.com/2011/01/happiness-is-warm-tweet.html

Sentiment detection

Annotate tweets using labels from http://en.wikipedia.org/wiki/List_of_emoticons

10 Happiest Tweets

- @WRiTExMiND no doubt! <--guess who I got tht from? Bwahaha anyway doe I like surprising people it's kinda my thing so ur welcome! And hi :)
- @skvillain yeh wiz is dope, got his own lil wave poppin! I'm fuccin wid big sean too he signed to kanye label g.o.o.d music
- And @pumahbeatz opened for @MarshaAmbrosius & blazed! So proud of him! Go bro! & Marsha was absolutely amazing! Awesome night all around. =)
- Awesome! RT @robscoms: Great 24 hours with nephews. Watched Tron, homemade mac & cheese for dinner, Wii, pancakes & Despicable Me this am!
- Good Morning 2 U Too RT @mzmonique718: Morninggggg twitt birds!...up and getting ready for church...have a good day and LETS GO GIANTS!
- Goodmorning #cleveland, have a blessed day stay focused and be productive and thank god for life
- AMEN!!!>>>RT @DrSanlare: Daddy looks soooo good!!! God is amazing! To GOD be the glory and victory #TeamJesus Glad I serve an awesome God
- AGREED!! RT @ILoveElizCruz: Amen to dat... We're some awesome people! RT @itsVonnell_Mars: @ILoveElizCruz gotta love my sign lol
- #word thanks! :) RT @Steph0e: @IBtunes HAppy Birthday love!!! =) still a fan of ya movement... yay you get another year to be dope!!! YES!!
- Happy bday isaannRT @isan_coy: Selamatt ulang tahun yaaa RT @Phitz_bow: Selamat siangg RT @isan_coy: Slamat pagiiii

http://www.machinedlearnings.com/2011/01/happiness-is-warm-tweet.html

Sentiment detection

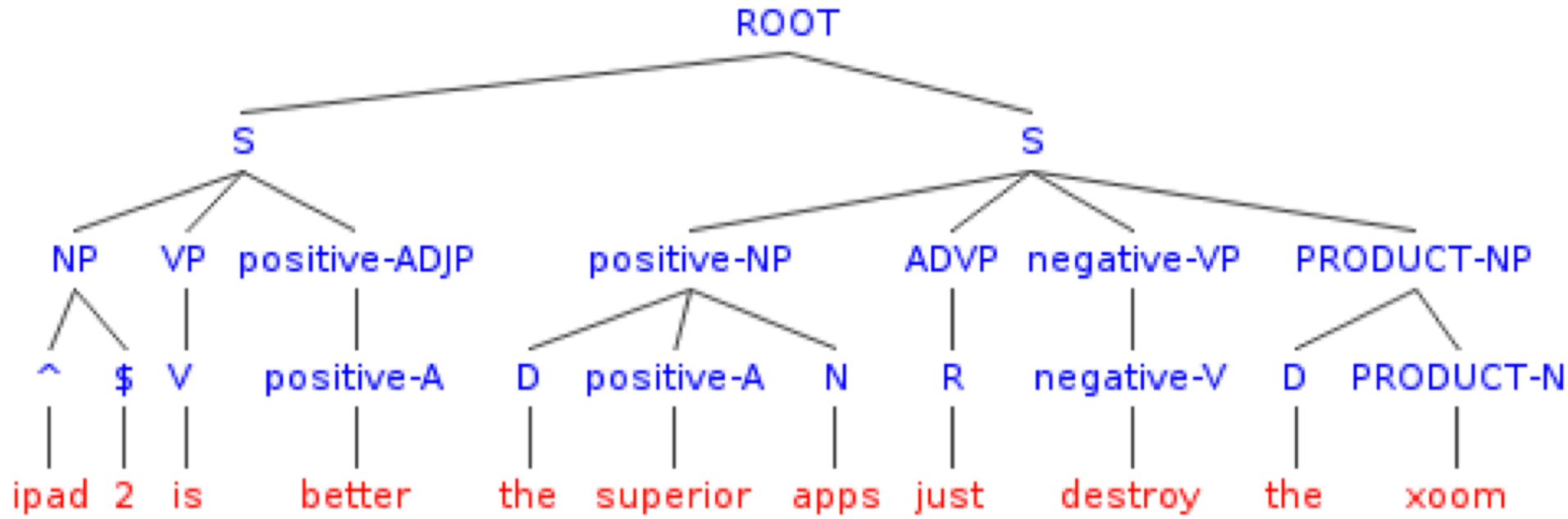
Annotate tweets using labels from http://en.wikipedia.org/wiki/List_of_emoticons

10 Saddest Tweets

- Migraine, sore throat, cough & stomach pains. Why me God?
- Ik moet werken omg !! Ik lig nog in bed en ben zo moe .. Moet alleen opstaan en tis koud buitn :(
- I Feel Horrible ' My Voice Is Gone Nd I'm Coughing Every 5 Minutes ' I Hate Feeling Like This :-/
- SMFH !!! Stomach Hurting ; Aggy ; Upset ; Tired ;; Madd Mixxy Shyt Yo !
- Worrying about my dad got me feeling sick I hate this!! I wish I could solve all these problems but I am only 1 person & can do so much..
- Malam2 menggigil+ga bs napas+sakit kepala....badan remuk redam *I miss my husband's hug....#nangismanja#
- Waking up with a sore throat = no bueno. Hoping someone didn't get me ill and it's just from sleeping. D:
- Aaaa ini tenggorokan gak enak, idung gatel bgt bawaannya pengen bersin terus. Calon2 mau sakit nih -___-
- I'm scared of being alone, I can't see to breathe when I am lost in this dream, I need you to hold me?
- Why the hell is suzie so afraid of evelyn! Smfh no bitch is gonna hav me scared I dnt see it being possible its not!

Opinion Mining

- Fine-grained sentiment
- For example: SenTube: sentiment and opinion mining from YouTube comments
  - *iPad 2 is better. the superior apps just destroy the xoom.*



Question Anwering

Question Answering

Passage Sentence

In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under gravity.

Question

What causes precipitation to fall?

Answer Candidate

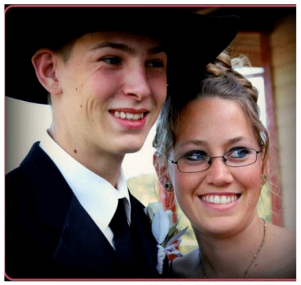
gravity

Visual Question Answering

Who is wearing glasses?

man

woman



Where is the child sitting?

fridge

arms

Is the umbrella upside down?

yes

no

How many children are in the bed?

2

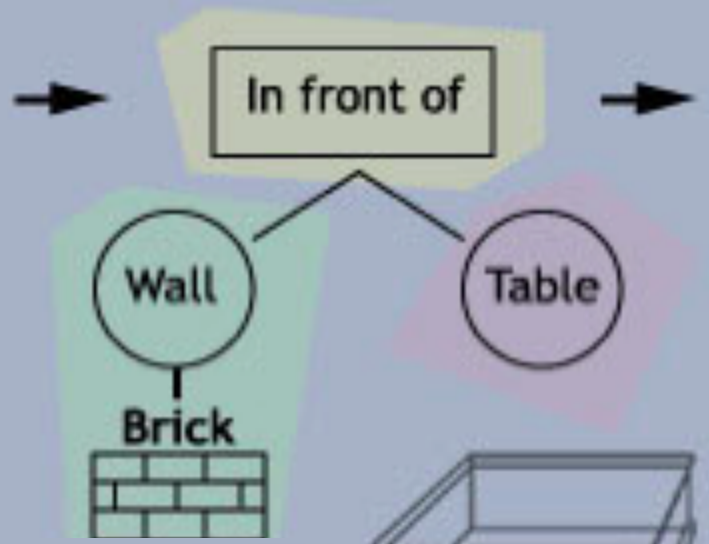
1

Holy Grail: Understanding Language

- Can we *generate* language from our knowledge of language?
- Can we convert a natural language utterance into a *model* (or some other fancy logic thing)
- Can we map it into a *database*?
- Can we map it into a *mental picture* (or a *real* one?)
- Demo: WordsEye (from Richard Sproat's group at AT&T)

Text to semantic model to image

The vase is on the Richard Sproat coffee table. The table is in front of the brick wall. The Van Gogh picture is on the wall. The Matisse sofa is next to the table. Mary is sitting on the sofa. She is playing the violin. She is wearing a straw hat.




details
details
The Devil is in the details
© 2000, AT&T Corp. Produced by Wordseye (TM)