Common Sense is Common,

Making Sense is Uncommon

Peter Norvig
Google

Leibniz (1646-1716)

The only way to rectify our reasonings is to make them as tangible as those of the Mathematicians, so that we can find our error at a glance, and when there are disputes among persons, we can simply say: *Let us calculate,* without further ado, to see who is right.

Les sens commun est fort rare.

Voltaire (1694-1778)

Simon Pierre Laplace (1749-1827)

Probability theory is nothing but common sense reduced to calculation.

Life's most important questions are, for the most part, nothing but probability problems.

George Boole (1815-1864)

There is not only a close analogy between the operations of the mind in general reasoning and its operations in the particular science of *Algebra*, but there is to a considerable extent an exact agreement in the laws by which the two classes of operations are conducted.

George Boole (1815-1864)

Probability is expectation founded upon partial knowledge. A perfect acquaintance with all the circumstances affecting the occurrence of an event would change expectation into certainty, and leave nether room nor demand for a theory of probabilities.

Bertrand Russell (1872-1970)

Since the world is what it is, it is clear that valid reasoning from sound principles cannot lead to error; but a principle may be so nearly true ... and yet may lead to [absurd] consequences ... *There is therefore a justification for common sense in philosophy* ... The theorist may retort that common sense is no more infallible than logic. But this retort, though made by Berkeley and Hume, would have been wholly foreign to Locke's intellectual temper.

Logic vs. Common Sense

Logic vs. Common Sense

What's an anagram of "Banach-Tarski"?

Logic vs. Common Sense

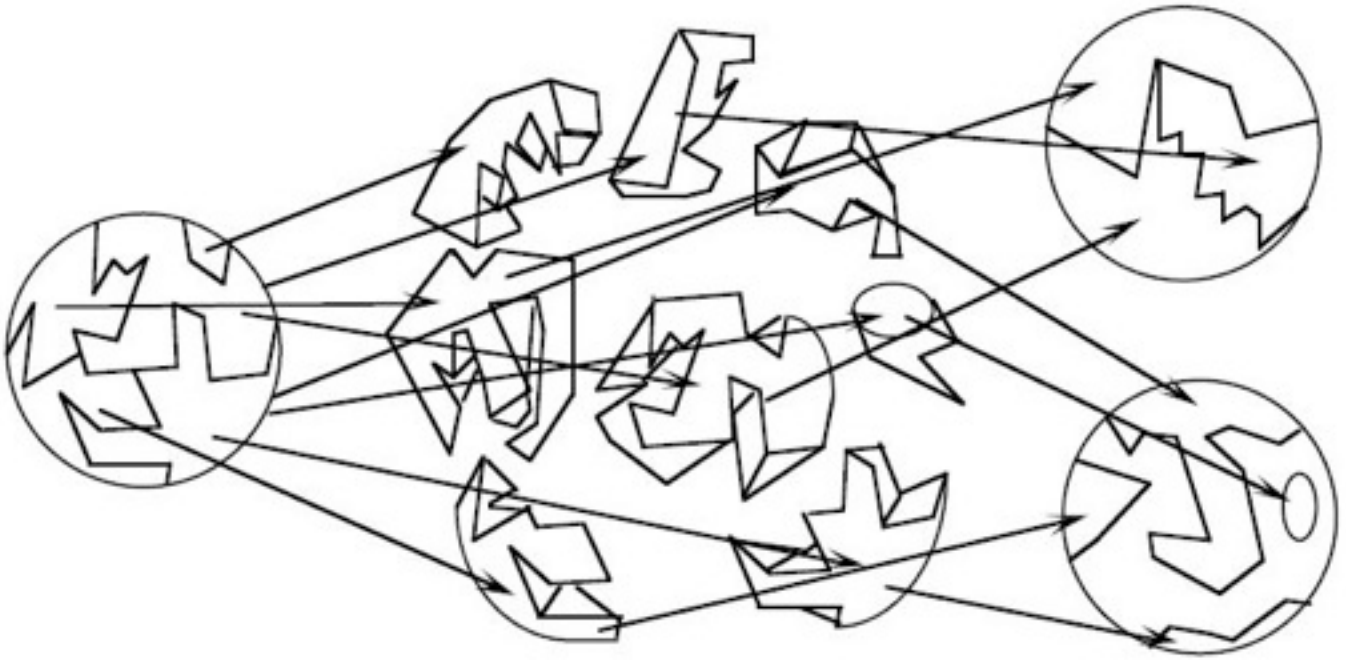
What's an anagram of "Banach-Tarski"?

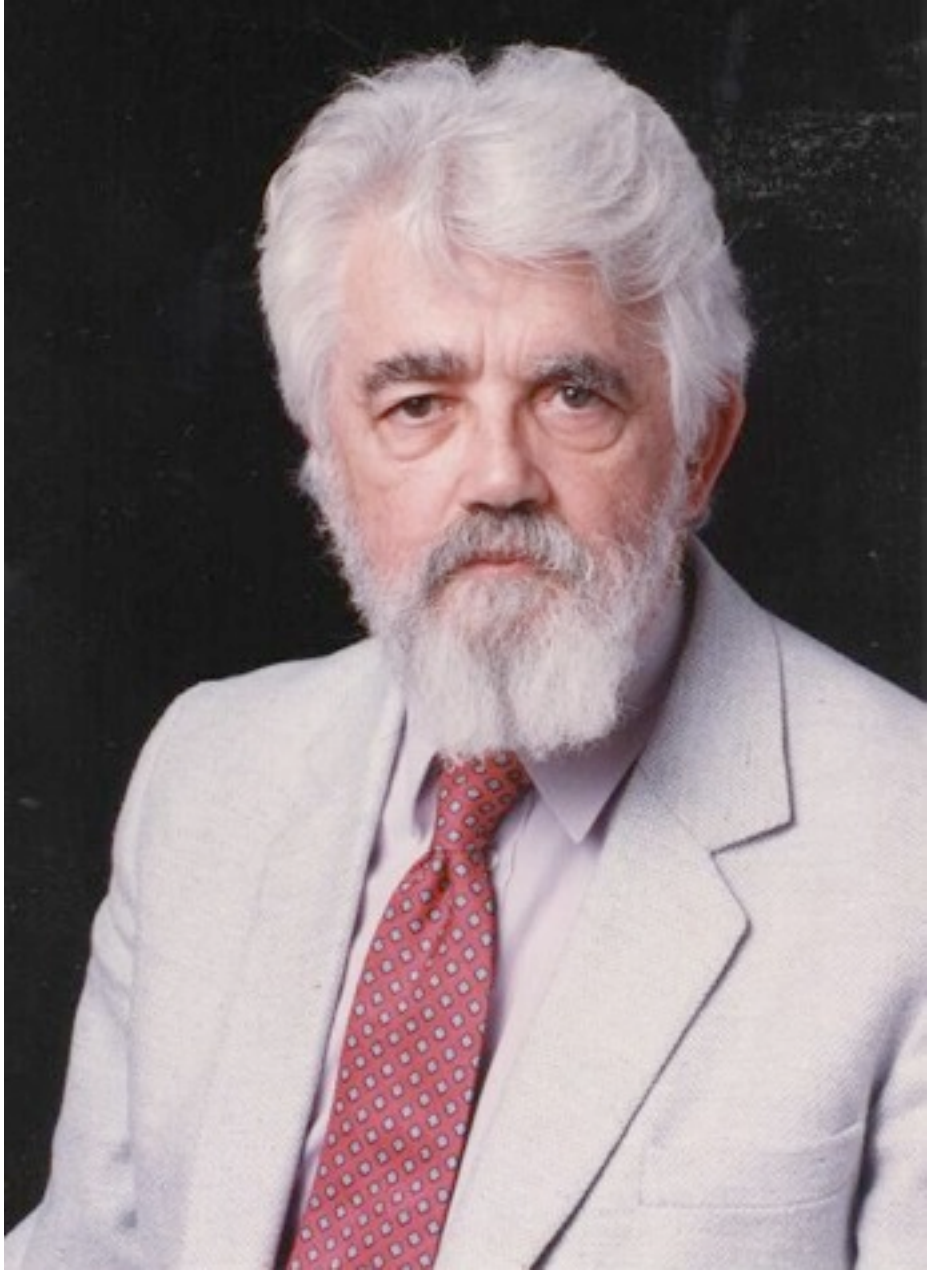
A: "Banach-Tarski Banach-Tarski"

Logic vs. Common Sense

What's an anagram of "Banach-Tarski"?

A: "Banach-Tarski Banach-Tarski"





John McCarthy (1927-)

A program has *common sense* if it automatically deduces for itself a sufficiently wide class of immediate consequences ...

Programs with Common Sense (1959)

Common Sense Defined

- Common Sense Knowledge is:
  - That which is commonly observed
  - Or easily derived from common sense

Empiricism
+
Rationalism

- Common sense is:
  - Applying common sense knowledge to reach good conclusions
    or take good actions

“Common sense” problems?

- Its largest airport is names for a WWII hero, its second largest, for a WWII battle.
  A: What is Chicago?
  (Trebek, 2010)
- What women have worn Chanel clothing to award ceremonies?
  What is the capital of Liechtenstein?
  What pathways are mediated by CD44?
  A: Best systems score in range of 77%
  (TREC, 2008)

Watson vs. Google

1	1	1	X	2	X	48	Questions Watson had correct	79%
X	4	1	1	X	X	13	Questions Watson had wrong	21%
1	1	X	7	X	2			
1	1	1	1	X	X	34	Questions Google had in top result	56%
1	1	X	1	1	1	14	Questions Google had in results 2-10	23%
						13	Questions Google didn't have in top 10 results	21%
5	1	X	1	1	1			
1	1	1	3	1	X			
4	1	3	2	1	2			
1	3	2	10	1	6			
1	1	1	X	1	1			
	Watson correct					#	Rank on Google	
	Watson correct, didn't answer					X	Not in first 10 results	
	Watson wrong, didn't answer							
	Watson wrong							

Watson, Jennings, Google-in-top-10: all 79%

"Google-proof" problems

- Joan made sure to thank Susan for all the help she had X? Who had X the help?

  X = given: Susan; X = received: Joan

- The large ball crashed through the table because it was made of X? What was made of X?

  X = steel: ball; X = styrofoam: table

(Levesque, 2011)

“Google-proof” problems

- The man lost his balance on the ladder. What happened as a result?
  A: He fell off the ladder.
  B: He climbed up the ladder.

- Human agreement: 96.5%
- “Google-proof?”
  Baseline word frequency methods: 58%

(Roemmele, Bejan, & Gordon, 2011)

Common Sense Defined

- Common Sense Knowledge is:
  - That which is commonly observed
  - Or easily derived from common sense

Empiricism
+
Rationalism

- Commons sense is:
  - Applying common sense knowledge to reach good conclusions
    or take good actions

That Which is Commonly Observed

- 15 million books
- 10 billion images
- 1 trillion web pages
- Trillions of user interactions per year
- 10 petabytes/year of video (40 hours/minute)

Making Sense with

Making Sense with

- Machine Reading

Making Sense with

- Machine Reading
- Machine Seeing/Watching

Making Sense with

- Machine Reading
- Machine Seeing/Watching
- Machine Hearing

Making Sense with

- Machine Reading
- Machine Seeing/Watching
- Machine Hearing
- Machine Observing

Making Sense with

- Machine Reading
- Machine Seeing/Watching
- Machine Hearing
- Machine Observing
- Machine Bribing

Look under the light first

Look under the light first

Look under the light first

Spread out from there
(Take baby steps)

Look under the light first

Spread out from there
(Take baby steps)

See if we can reach horizon

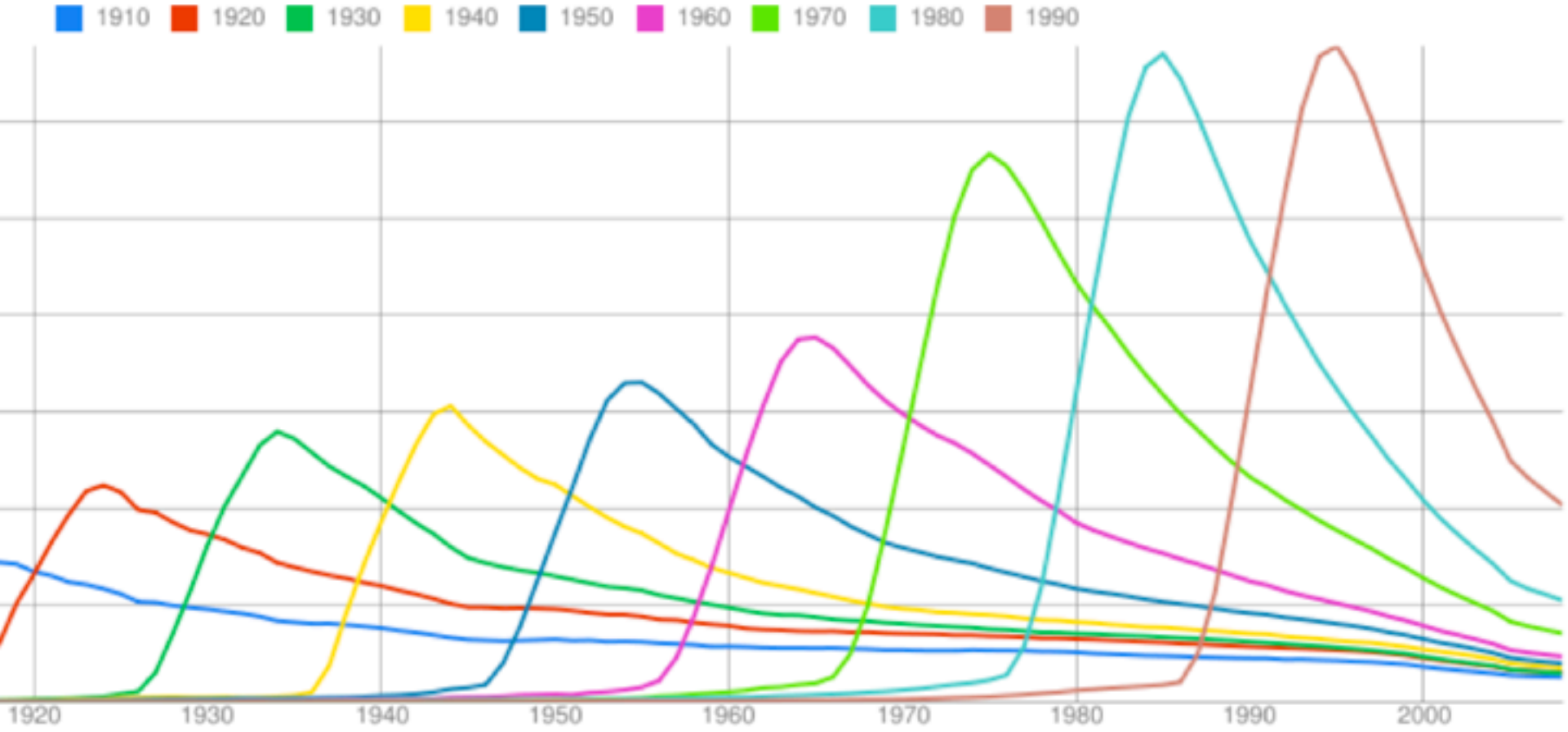
Spreading Out

- Ngram data
- Clusters
- Sentence alignments
- Part of speech tags
- Restricted parses
- Full parses
- Extracted entities
- Extracted attribute/value pairs
- Extracted relations
- Extracted inference rules

Spreading Out

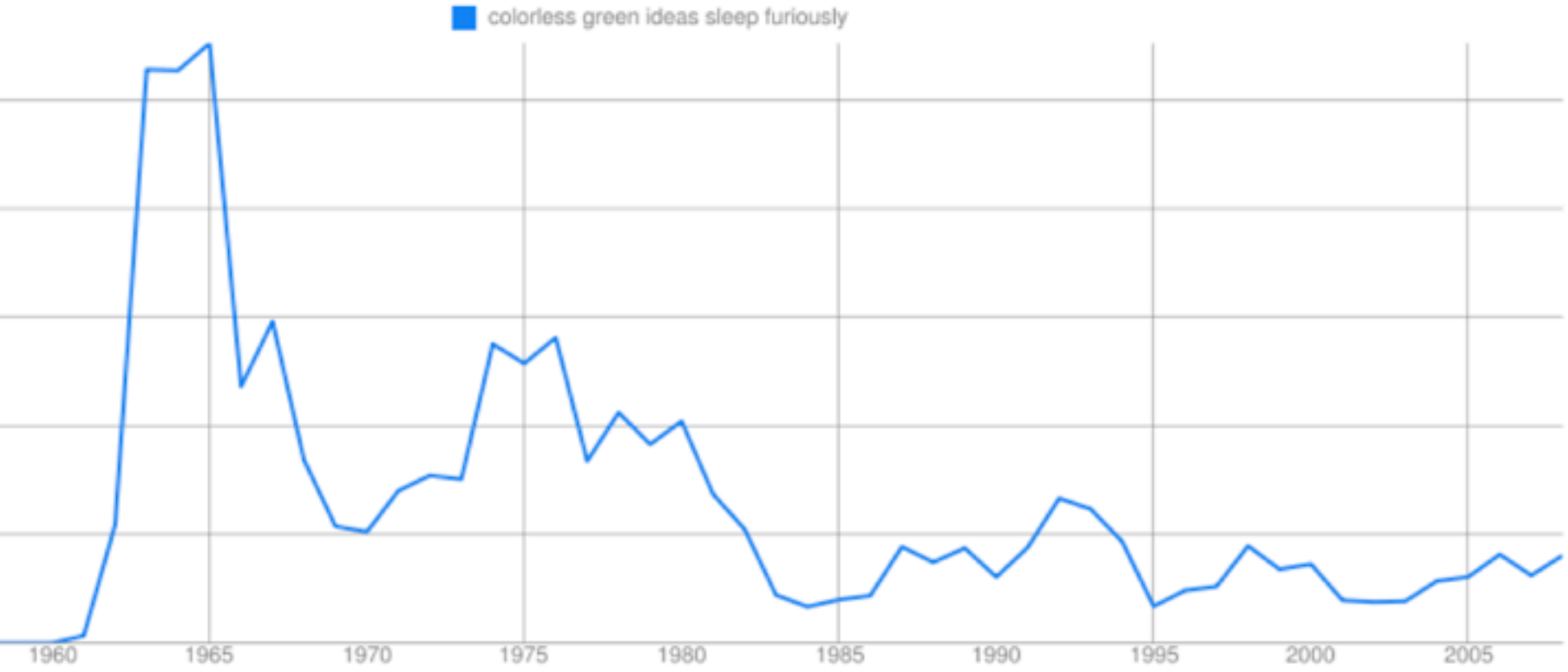
- Grouping similar images
- Partitioning into regions
- Separating into objects
- Grouping similar objects
- Identifying objects
- Detecting motion
- Interpreting action

1910, 1920, 1930, … , 1990

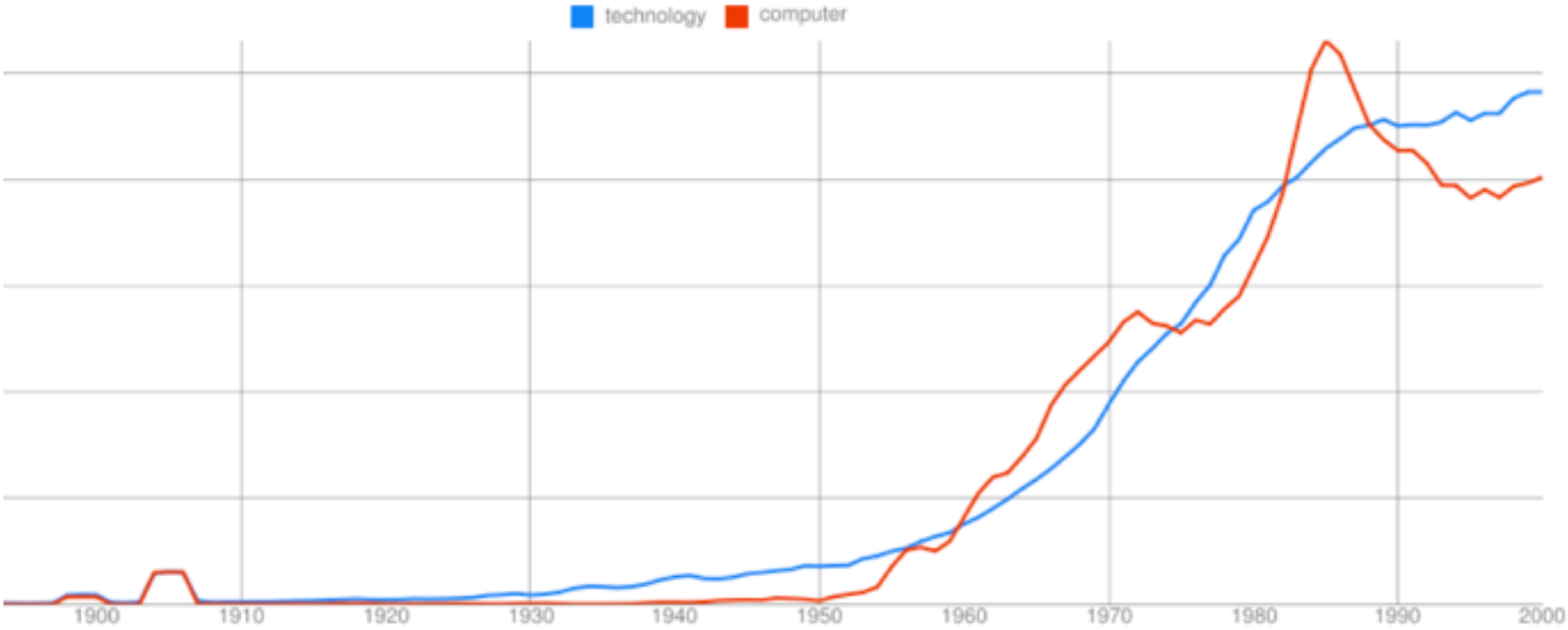


				Nobel Prize Win	
Full Name	Born	Died	milliDarwins	Bio	Chem
Bertrand Russell	1872	1970	1500		
Charles Darwin	1809	1882	1000		
Albert Einstein	1879	1955	878		
Lewis Carroll	1832	1898	479		
Claude Bernard	1813	1878	429		
Oliver Lodge	1851	1940	394		
Julian Huxley	1887	1975	350		
Karl Pearson	1857	1936	346		
Niels Bohr	1885	1962	289		
Alexander Graham Bell	1847	1922	274		
Max Planck	1858	1947	256		
Francis Galton	1822	1911	255		
Robert Oppenheimer	1904	1967	252		
Louis Pasteur	1822	1895	237		
Chaim Weizmann	1874	1952	236		
Alfred North Whitehead	1861	1947	229		
Marie Curie	1867	1934	189		✓

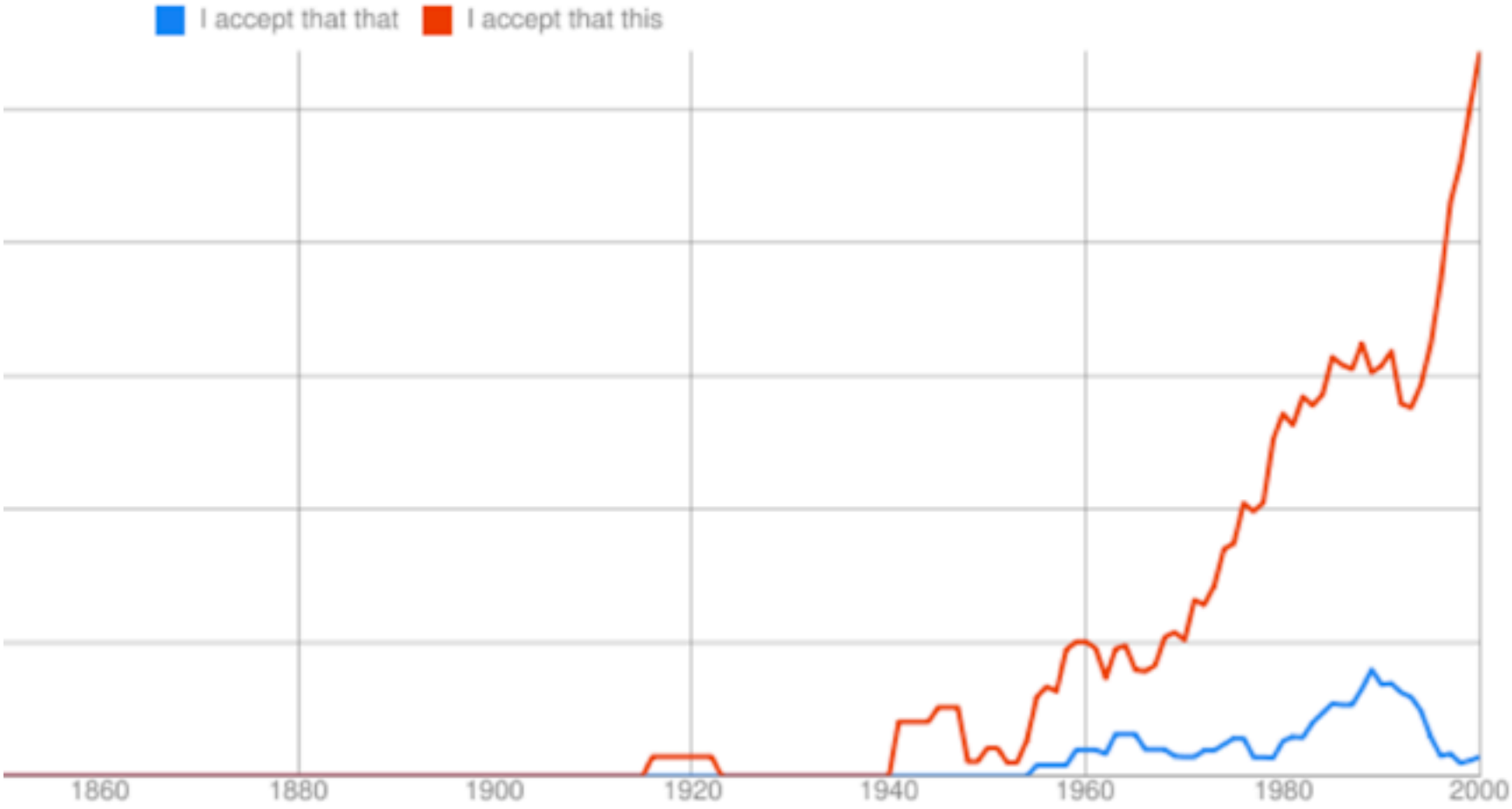
colorless green ideas sleep furiously



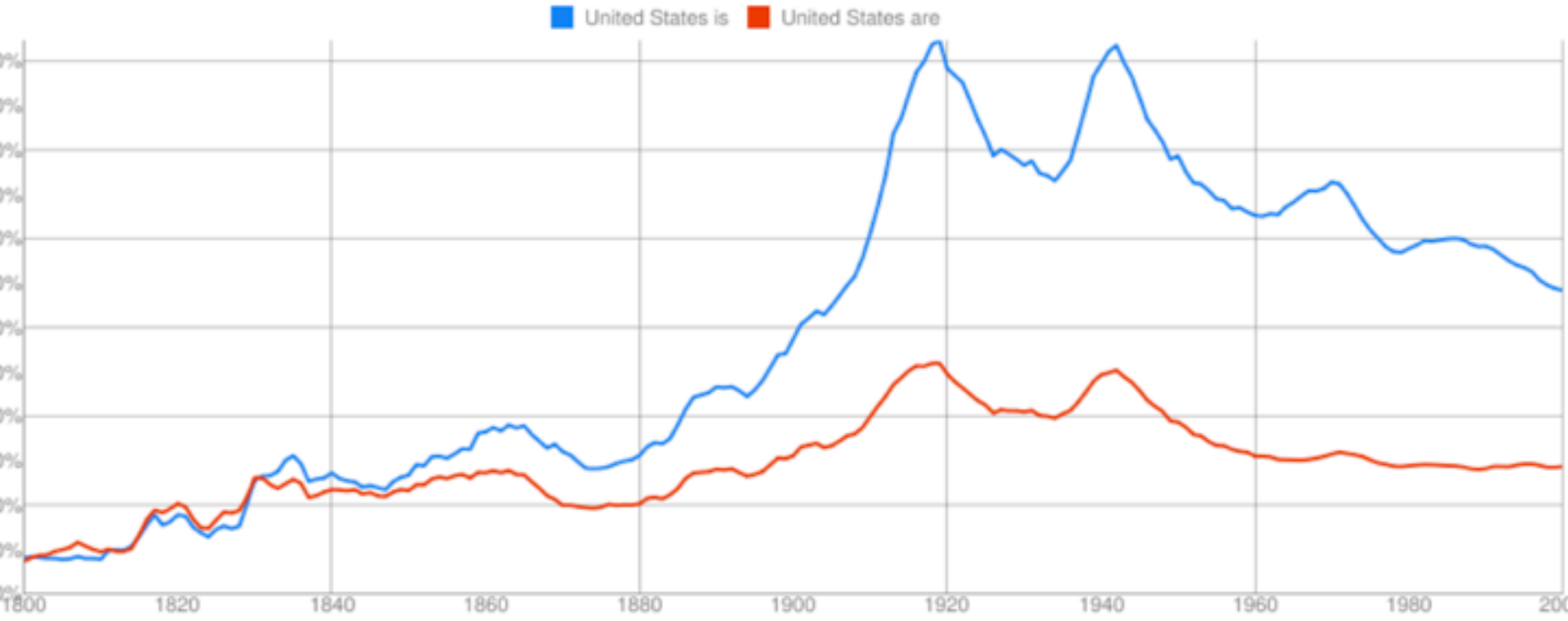
technology, computer

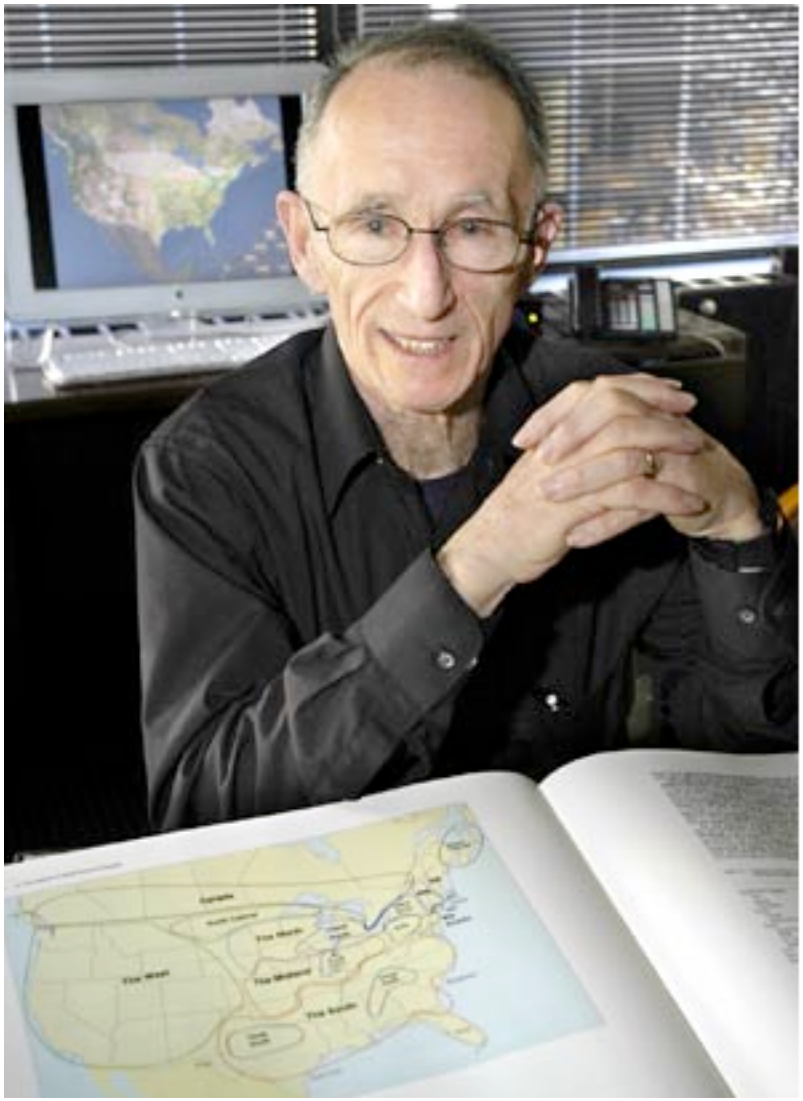


I accept that this/that



United States is/are





Bill Labov (1927-)

- If linguistics can be said to be any one thing it is the study of categories...

Feedback Discuss Terms of Use

Google™ Sets

Automatically create sets of items from a few examples.

Enter a few items from a set of things. (example)
Next, press *Large Set* or *Small Set* and we'll try to predict other items in the set.

- pablo picasso
- henri matisse
-
-
-

(clear all)

Large Set | Small Set (15 items or fewer)

Google Sets

henri matisse, pable picasso, vincent van gogh, claude monet, pablo picasso, salvador dali, edgar degas, paul cezanne, andy warhol, pierre auguste renoir, marc chagall, paul gauguin, mary cassatt, michelangelo, jackson pollock, camille pissarro, paul klee, georgia o'keeffe, norman rockwell, leonardo da vinci, edward hopper, frida kahlo, rembrandt, wassily kandinsky, rene magritte, francisco goya, raphael, georges seurat, giotto, henri rousseau, diego rivera, van gogh, botticelli, johannes vermeer, ansel adams, monet,

Predicted Items
commonsense reasoning
computer vision
computational intelligence
cognitive tutor
captcha
crossover genetic algorithm
computer assisted proof
connectionist expert system

Google Sets

Given “lion, tiger, bear” find:

bear, tiger, lion, elephant, monkey, giraffe, dog, cat, snake, horse, zebra, rabbit, wolf, dolphin, dragon, pig, frog, duck, cheetah, bird, cow, *cotton*, hippo, turtle, penguin, rat, gorilla, leopard, sheep, mouse, puppy, ox, rooster, fish, lamb, panda, *wood*, *musical*, *toddler*, fox, goat, deer, squirrel, koala, crocodile, hamster

Google Sets

Google Sets

Given “cat, man” find:

Google Sets

Given “cat, man” find:

cat, man, ls, cp, rm, mkdir, mv, cd, pwd, rmdir, chmod, ln, grep, touch, find, ps, chown, df, less, tar, chgrp, du, sort, date, echo, kill, tail, wc, mount, sed, passwd, dd, head, vi, which, gzip, lpr, who, file, su, umount, diff, cut, uname, exit, basename, clear, sleep, whoami, mknod

Google Sets

Google Sets

Tin Woodsman: Some, but mostly lions and tigers and bears.
Dorothy: Lions?
Scarecrow: And tigers?
Tin Woodsman: And bears.

Google Sets

Tin Woodsman: Some, but mostly lions and tigers and bears.
Dorothy: Lions?
Scarecrow: And tigers?
Tin Woodsman: And bears.

African cats:

- Lions
- Leopards
- Cheetahs

Google Sets

Tin Woodsman: Some, but mostly lions and tigers and bears.
Dorothy: Lions?
Scarecrow: And tigers?
Tin Woodsman: And bears.

African cats:
- Lions
- Leopards
- Cheetahs

Animal Toys (Ages 2-6):
- Bears
- Puppies
- Dolphins
- Wooden

Google Sets

Tin Woodsman: Some, but mostly lions and tigers and bears.
Dorothy: Lions?
Scarecrow: And tigers?
Tin Woodsman: And bears.

African cats:
- Lions
- Leopards
- Cheetahs

```
10:03:37 [cheetah pics]
10:05:51 [leopard pics]
```

Animal Toys (Ages 2-6):
- Bears
- Puppies
- Dolphins
- Wooden

Google Sets

Tin Woodsman: Some, but mostly lions and tigers and bears.
Dorothy: Lions?
Scarecrow: And tigers?
Tin Woodsman: And bears.

African cats:
- Lions
- Leopards

```
10:03:37 [cheetah pics]
10:05:51 [leopard pics]
```

es 2-6):

U.S. Senate Committee on Environment and Public Works
Exotic animals, **such as lions**, tigers, servals, monkeys, bears, snakes, iguanas, wolves, prairie dogs, and binturongs are being privately possessed as ...
epw.senate.gov/hearing_statements.cfm?id=213174 - 13k -

Subcategories, Attributes, Values

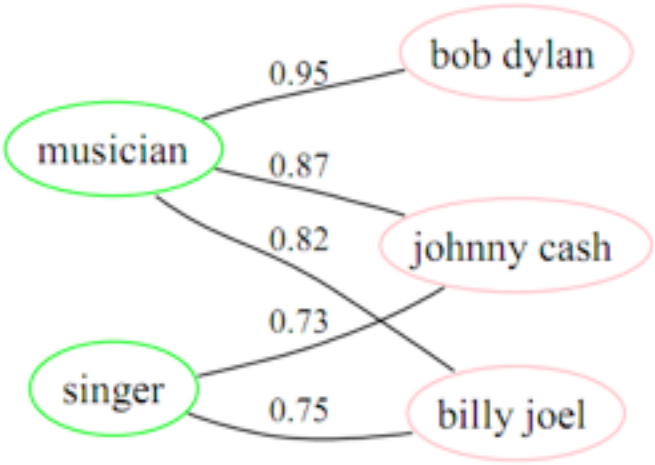
- From 10 (Person, BirthYear) examples, mine 1M pairs with 90% precision
- Mine 2M (instance, attribute) and (class, attribute) pairs, 60-80% precision

Instance	Run	Attributes		
		Precision		Top Ten Attributes
		@5	@10	
angioplasty	S_nC_n	0.00	0.00	*none extracted*
	S_yC_n	1.00	0.75	[history, types, complications, risks, definition, cost, pictures, symptoms, cen
	S_nC_y	1.00	0.95	[complications, cost, history, types, risks, purpose, definition, pictures, side tages]
	S_yC_y	1.00	0.75	[history, complications, types, cost, risks, definition, pictures, side effects, sy
	Per-class	1.00	0.90	[types, history, pictures, cost, principles, pros and cons, map, definition, methods]

(Pasca *et al.* 2008, 2011)

Class instances

- Mine 100Ms pairs



Class	Precision at 100 (non-A8 extractions)
Book Publishers	87.36
Federal Agencies	29.89
NFL Players	94.95
Scientific Journals	90.82
Mammal Species	84.27

Class	Size	Examples of Instances
Book Publishers	70	crown publishing, kluwer academic, prentice hall, puffin
Federal Agencies	161	catsa, dhs, dod, ex-im bank, fsis, iema, mema, nipc, nmfs, tdh, usdot
Mammals	956	armadillo, elephant shrews, long-tailed weasel, river otter, weddell seals, wild goat
NFL Players	180	aikman, deion sanders, fred taylor, jamal lewis, raghib ismail, troy vincent
Scientific Journals	265	biometrika, european economic review, nature genetics, neuroscience
Social Issues	210	gender inequality, lack of education, substandard housing, welfare dependency
Writers	5089	bronte sisters, hemingway, kipling, proust, torquato tasso, ungaretti, yeats

(Talukdar *et al.*, *EMNLP* 2008)

Parsing: Baby Steps

- Parse short sentences first; beats state of art in unsupervised parser learning

Corpus	*Sentences*	*POS Tokens*
WSJ1	159	159
WSJ2	499	839
WSJ3	876	1,970
WSJ4	1,394	4,042
WSJ5	2,008	7,112
WSJ6	2,745	11,534
WSJ7	3,623	17,680
WSJ8	4,730	26,536
WSJ9	5,938	37,408
WSJ10	7,422	52,248
WSJ11	8,856	68,022
WSJ12	10,500	87,750

Corpus	*Sentences*	*POS Tokens*
WSJ13	12,270	110,760
WSJ14	14,095	136,310
WSJ15	15,922	163,715
WSJ20	25,523	336,555
WSJ25	34,431	540,895
WSJ30	41,227	730,099
WSJ35	45,191	860,053
WSJ40	47,385	942,801
WSJ45	48,418	986,830
WSJ100	49,206	1,028,054
Section 23	2,353	48,201
Brown100	24,208	391,796

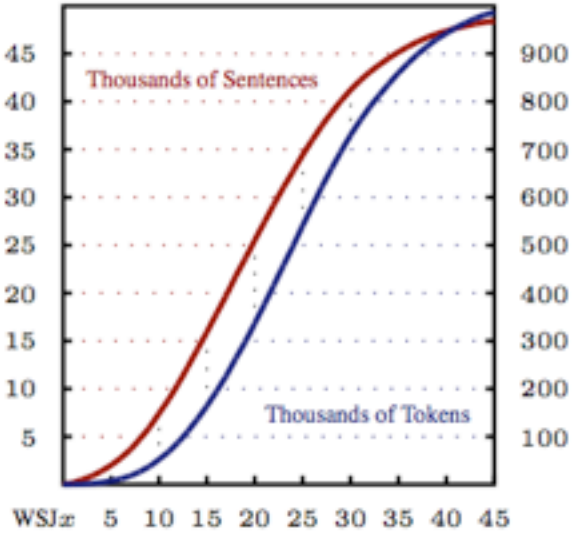


Figure 1: Sizes of WSJ$\{1, \ldots, 45, 100\}$, Section 23 of WSJ$^{\infty}$ and Brown100.

(Spitkovsky, Alshawi, Jurafsky, 2010)

Parsing: <a>Annotations

- Use HTML markup to beat state-of-art unsupervised parsers by 5%
  "in <a>an analysisof..."

	Count	Constituent Production	Frac	Sum
1	746	NP → NNP NNP	9.6%	
2	357	NP → NNP	4.6	14.3
3	266	NP → NP PP	3.4	17.7
4	183	NP → NNP NNP NNP	2.4	20.1
5	165	NP → DT NNP NNP	2.1	22.2
6	140	NP → NN	1.8	24.0
7	131	NP → DT NNP NNP NNP	1.7	25.7
8	130	NP → DT NN	1.7	27.4
9	127	NP → DT NNP NNP	1.6	29.0
10	109	S → NP VP	1.4	30.4
11	91	NP → DT NNP NNP NNP	1.2	31.6
12	82	NP → DT JJ NN	1.1	32.7
13	79	NP → NNS	1.0	33.7
14	65	NP → JJ NN	0.8	34.5
15	60	NP → NP NP	0.8	35.3
$BLOG_p$	+5,000	*more with Count* ≤ 60	64.7%	

	Count	Head-Outward Spawn	Frac	Sum
1	1,889	NNP	24.4%	
2	623	NN	8.1	32.5
3	470	DT NNP	6.1	38.6
4	458	DT NN	5.9	44.5
5	345	NNS	4.5	49.0
6	109	NNPS	1.4	50.4
7	98	VBG	1.3	51.6
8	96	NNP NNP NN	1.2	52.9
9	80	VBD	1.0	53.9
10	77	IN	1.0	54.9
11	74	VBN	1.0	55.9
12	73	DT JJ NN	0.9	56.8
13	71	VBZ	0.9	57.7
14	69	POS NNP	0.9	58.6
15	63	JJ	0.8	59.4
$BLOG_p$	+3,136	*more with Count* ≤ 62	40.6%	

(Spitkovsky, Jurafsky, Alshawi, 2010)

Translation: (Parentheticals)

- Generate 26M entry bilingual dictionary from parenthetical examples

(1) 美国智库布鲁金斯学会（Brookings Institution）专研跨大西洋恐怖主义的美欧中心研究部主任杰若米·夏皮罗（Jeremy Shapiro）却认为，...

(2) 消化性溃疡的症状往往与消化不良（indigestion），胃炎（gastritis）等其他胃部疾病症状相似.

(3) 殊不知美国是不会接受（not going to fly）这一想法的

(4) …当是一次式时，叫线性规划(linear programming).

(Lin *et al.*, 2008)

English	Wikipedia Translation	Parenthetical Translation
Pumping lemma	泵引理	引理[1]
Topic-prominent language	话题优先语言	突出性语言[1]
Yoido Full Gospel Church	汝矣岛纯福音教会	全备福音教会[1]
First Bulgarian Empire	第一保加利亚帝国	强大的保加利亚帝国[2]
Vespid	黄蜂	针对境内胡蜂[2]
Ibrahim Rugova	易卜拉欣·鲁戈瓦	鲁戈瓦[3]
Jerry West	杰里·韦斯特	威斯特[3]
Nicky Butt	尼基·巴特	巴特[3]
Benito Mussolini	贝尼托·墨索里尼	墨索里尼[3]
Ecology of Hong Kong	香港生态	本文介绍的*
Paracetamol	对乙酰氨基酚	扑热息痛*
Thermidor	热月	必杀*
Udo	独活	乌多
Public opinion	舆论	公众舆论

Parsing: Deterministic Parsing

- Corpus of (text, LF) pairs; aligned
- Train a dependency parser (85% accuracy)
- Use parser to create LF for input text
- "In 2002, Chirpy Systems stealthily acquired two profitable companies producing pet accessories"
  [acquired / stealthily : [in, ^, 2002],
  Chirpy+Systems,
  companies.two :profitable
  :[producing, ^, pet+accessories]]

(Alshawi, Chang, Ringgacrd, 2011)

Statistical Machine Translation

- Ngram model: P(E)
- Alignment model: P(E,F)

(Och *et al.*, 2007, *etc.*)

Translation from Ngrams (etc.)

- **EN**: We shall therefore say that a program has common sense if it automatically deduces for itself a sufficiently wide class of immediate consequences of anything it is told and what it already knows.
- **EN→ES**: Por lo tanto, diremos que un programa tiene sentido común, si se deduce automáticamente por sí mismo una clase suficientemente amplia de las consecuencias inmediatas de todo lo que se dice y lo que ya sabe.
- **EN→ES→EN**: Therefore, we say that a program has common sense if it automatically deduces for itself a sufficiently wide class of immediate consequences of all that is said and what you know.

Example for Italian - English

Piano della NASA per inviare una sonda sulla superficie lunare

L'ente spaziale americano, NASA, ha annunciato che la prossima missione sulla Luna non sarà limitata a orbitare intorno al satellite, ma che includerà anche il lancio di due veicoli spaziali che raggiungeranno la superficie lunare per mezzo di un atterraggio con schianto.

Il Lunar Reconnaissance Orbiter (LRO), la cui missione principale è quella di esplorare la Luna, invierà una navetta di appoggio e una sonda a impatto verso un cratere situato al polo sud lunare.

Sembra che il cratere sia ricco di idrogeno e forse ghiaccio.

Questa missione va vista nel contesto di una serie di iniziative che hanno lo scopo di riportare gli astronauti sulla Luna e, forse, di farceli rimanere per un periodo superiore rispetto a quello della missione Apollo.

NASA's Plan to Send a Probe on the Lunar Surface

The American space agency, NASA, announced that the next mission to the Moon will not be limited to orbit around the satellite, but also include the launch of two spacecraft that will reach the lunar surface by means of a crash landing.

The Lunar Reconnaissance Orbiter (LRO), whose mission is to explore the Moon, will send a shuttle to and support a probe into an impact crater located on the south lunar pole.

It seems that the crater is full of hydrogen and possibly ice.

This mission should be seen as part of a series of initiatives that aim to bring the astronauts on the moon and, perhaps, let us stay for a period longer than the Apollo mission.

Statistical Machine Translation

- Collect parallel texts

DE	EN
SEHR GEEHRTER GAST! KUNST, KULTUR UND KOMFORT IM HERZEN BERLIN.	DEAR GUESTS, ART, CULTURE AND LUXURY IN THE HEART OF BERLIN.
DIE ÖRTLICHE NETZSPANNUNG BETRÄGT 220/240 VOLT BEI 50 HERTZ.	THE LOCAL VOLTAGE IS 220/240 VOLTS 50 HZ.

Statistical Machine Translation

- Align

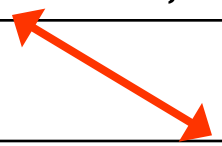
KUNST, KULTUR UND KOMFORT IM HERZEN BERLINS.

ART, CULTURE AND LUXURY IN THE HEART OF BERLIN.

Statistical Machine Translation

- Align

KUNST, KULTUR UND KOMFORT IM HERZEN BERLINS.



ART, CULTURE AND LUXURY IN THE HEART OF BERLIN.

Statistical Machine Translation

- Align

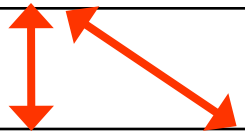
KUNST, KULTUR UND KOMFORT IM HERZEN BERLINS.

ART, CULTURE AND LUXURY IN THE HEART OF BERLIN.

Statistical Machine Translation

- Align

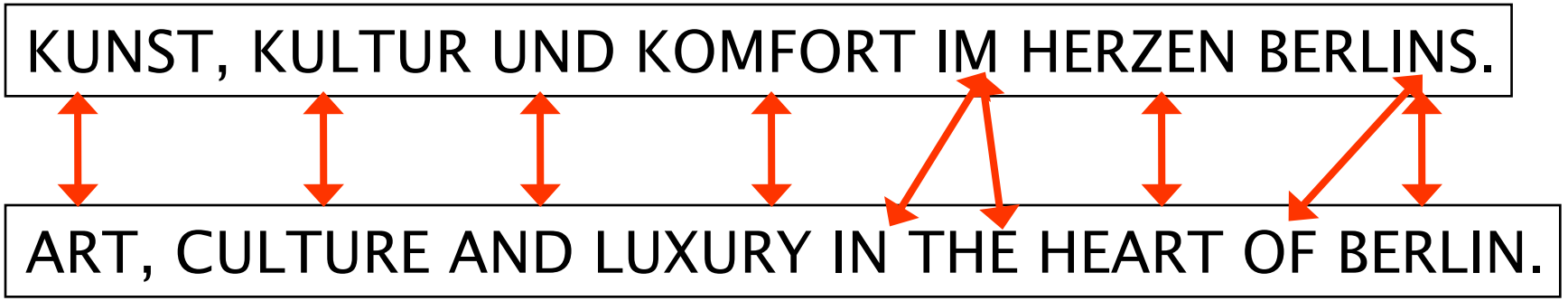
KUNST, KULTUR UND KOMFORT IM HERZEN BERLINS.



ART, CULTURE AND LUXURY IN THE HEART OF BERLIN.

Statistical Machine Translation

- Align



Categories

Bai Bi Bu Cai Cao Chang Chen Cheng Chou Chuang Cui Dai Deng Ding Du Duan Fan Fu Gao Ge Geng Gong Gu Guan Han Hou Hsiao Hsieh Hsu Hu Huang Huo Jiang Jiao Juan Kang Kuang Kuo Li Liang Liao Lin Liu Lu Luo Mao Meets Meng Mi Miao Mu Niu Pang Pi Pu Qian Qiao Qiu Qu Ren Run Shan Shang Shen Si Song Su Sui Sun Tan Tang Tian Tu Wang Wu Xie Xiong Xu Yang Yao Ye Yin Zeng Zhang Zhao Zheng Zhou Zhu Zhuang Zou
% PERCENT cents percent
approvals bonus cash concessions cooperatives credit disbursements dividends donations earnings emoluments entitlements expenditure expenditures fund funding funds grants income incomes inflation lending liquidity loan loans mortgage mortgages overhead payroll pension pensions portfolio profits protectionism quotas receipts receivables remittances remuneration rent rents returns revenue revenues salaries salary savings spending subscription subsidies subsidy surplus surpluses tax taxation taxes tonnage tuition turnover wage wages
Abby Abigail Agnes Alexandra Alice Amanda Amy Andrea Angela Ann Anna Anne Annette Becky Beth Betsy Bonnie Brenda Carla Carol Carole Caroline Carolyn Carrie Catherine Cathy Cheryl Christina Christine Cindy Claire Clare Claudia Colleen Cristina Cynthia Danielle Daphne Dawn Debbie Deb-

tallie Nina Nora Norma Olivia Pam Pamela Patricia Patti Paula Pauline Peggy Phyllis Rachel Rebecca Regina Renee Rita Roberta Rosemary Sabrina Sally Samantha Sarah Selena Sheila Shelley Sherry Shirley Sonia Stacy Stephanie Sue Susanne Suzanne Suzy Sylvia Tammy Teresa Teri Terri Theresa Tina Toni Tracey Ursula Valerie Vanessa Veronica Vicki Vivian Wendy Yolanda Yvonne
almonds apple apples asparagus avocado bacon bananas barley basil bean beans beets berries berry boneless broccoli cabbage carrot carrots celery cherries cherry chile chiles chili chilies chives cilantro citrus cranberries cranberry cucumber cucumbers dill doughnuts egg eggplant eggs elk evergreen fennel figs flowers fruit fruits garlic ginger grapefruit grasses herb herbs jalapeno Jell-O lemon lemons lettuce lime lions macaroni mango maple melon mint mozzarella mushrooms oak oaks olives onion onions orange oranges orchids oregano oyster parsley pasta pastries pea peach peaches peanuts pear pears peas pecan pecans perennials pickles pine pineapple pines plum pumpkin pumpkins raspberries raspberry rice rosemary roses sage salsa scallions scallops seasonings seaweed shallots shrimp shrubs spaghetti spices spinach strawberries strawberry thyme tomato tomatoes truffles tulips turtles walnut walnuts watermelon wildflowers zucchini
mid-April mid-August mid-December mid-February mid-January mid-July mid-June mid-March mid-May mid-November mid-October mid-September mid-afternoon midafternoon midmorning midsummer

Table 4: Examples of clusters

他 信 也 说 自己 仍然 是 总理，拒绝 辞职.

他 信 也 说 自己 仍然 是 总理，拒绝 辞职.

他
he
He
other
his
him
other
that he
he was
him to
he is
he has
of his

他 信 也 说 自己 仍然 是 总理，拒绝 辞职.

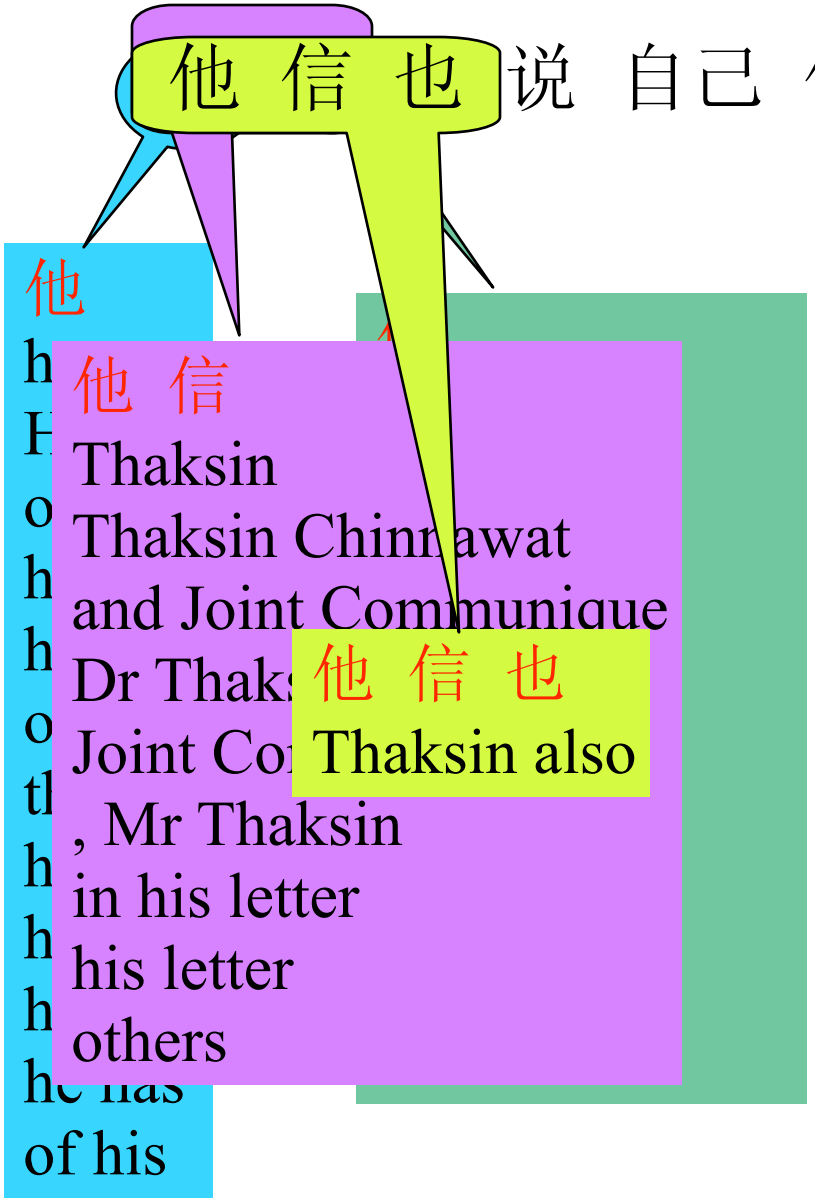
他
he
He
other
his
him
other
that he
he was
him to
he is
he has
of his

信
letter
trust
letters
believe

signal
a letter
believe that
letter of
confidence

他 信也 说 自己 仍然 是 总理，拒绝 辞职.

他 信 也 说 自己 仍然 是 总理，拒绝 辞职.



他 信 也 说 自己 仍 然 是 总 理 ， 拒 绝 辞 职 .

他
h
H
o
h
h
o
t
h
h
h
h
he has
of his

他 信
Thaksin
Thaksin Chinnawat
and Joint Communique
Dr Thaks
Joint Co
, Mr Thaksin
in his letter
his letter
others

他 信 也
Thaksin also

说 自己
themselves
said that
say they
said he
say that
said they
themselves
saying that he
would say that
said that she had
saying that he has

他 信 也 说 自己 仍 然 是 总理，拒 绝 辞 职.

他
h
H
o
h
h
o
t
h
h
h
he has
of his

他 信
Thaksin
Thaksin Chinnawat
and Joint Communique
Dr Thaks
Joint Co
, Mr Thaksin
in his letter
his letter
others

他 信 也
Thaksin also

说 自
themse
said tha
say the
said he
say that
said the
themse
saying
would s
said that she had
saying that he has

仍 然 是
continues to be
are still the main
would still be
continued to be
remains one of
remains one
continues to be the
still is
remains an area
still viewed by
are always one of

他 信 也 说 自己 仍 然 是 总理，拒 绝 辞 职.

他
h
H
o
h
h
o
t
h
h
h
he has
of his

他 信
Thaksin
Thaksin Chinnawat
and Joint Communique
Dr Thaks
Joint Co
, Mr Thaksin
in his letter
his letter
others

他 信 也
Thaksin also

说 自
themse
said tha
say the
said he
say that
said the
themse
saying
would s
said that she had
saying that he has

仍 然 是
contin
are sti
would
contin
remains one of
remains one
continues to be the
still is
remains an area
still viewed by
are always one of

是 总理
Prime Minister
the Prime Minister
is the Prime Minister

他 信 也 说 自己 仍 然 是 总理 ， 拒 绝 辞 职 .

他
h
H
o
h
h
o
t
h
h
h
he has
of his

他 信
Thaksin
Thaksin Chinnawat
and Joint Communique
Dr Thaks
Joint Co
, Mr Thaksin
in his letter
his letter
others

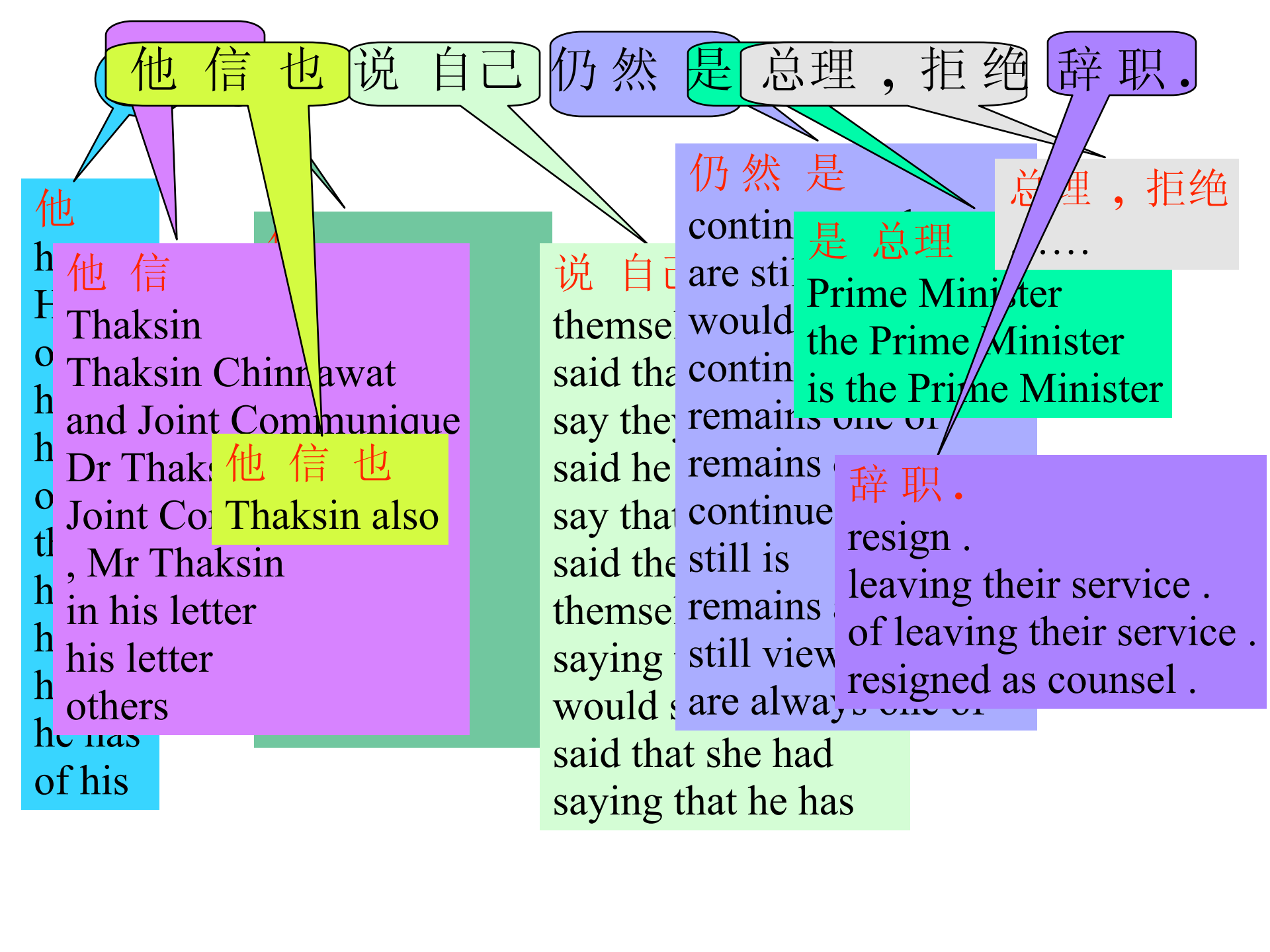
他 信 也
Thaksin also

说 自己
themse
said tha
say the
said he
say that
said the
themse
saying
would s
said that she had
saying that he has

仍 然 是
contin
are sti
would
contin
remains one of
remains one
continues to be the
still is
remains an area
still viewed by
are always one of

是 总理
Prime Minister
the Prime Minister
is the Prime Minister

总理 ， 拒绝
……


他 信 也
说 自己
仍 然 是
总理，拒绝
辞 职.
他
of his
他 信
Thaksin
Thaksin Chinnawat
and Joint Communique
Dr Thaks
Joint Co
, Mr Thaksin
in his letter
his letter
others
他 信 也
Thaksin also
说 自己
themse
said tha
say the
said he
say that
said the
themse
saying
would s
said that she had
saying that he has
仍 然 是
contin
are sti
would
contin
remains
remains
continue
still is
remains
still view
are alway
总理，拒绝
....
是 总理
Prime Minister
the Prime Minister
is the Prime Minister
辞 职.
resign .
leaving their service .
of leaving their service .
resigned as counsel .

Laudato sie, mi Signore cum tucte le Tue creature,
spetialmente messor lo frate **Sole**,
lo qual è iorno, et allumini noi per lui.

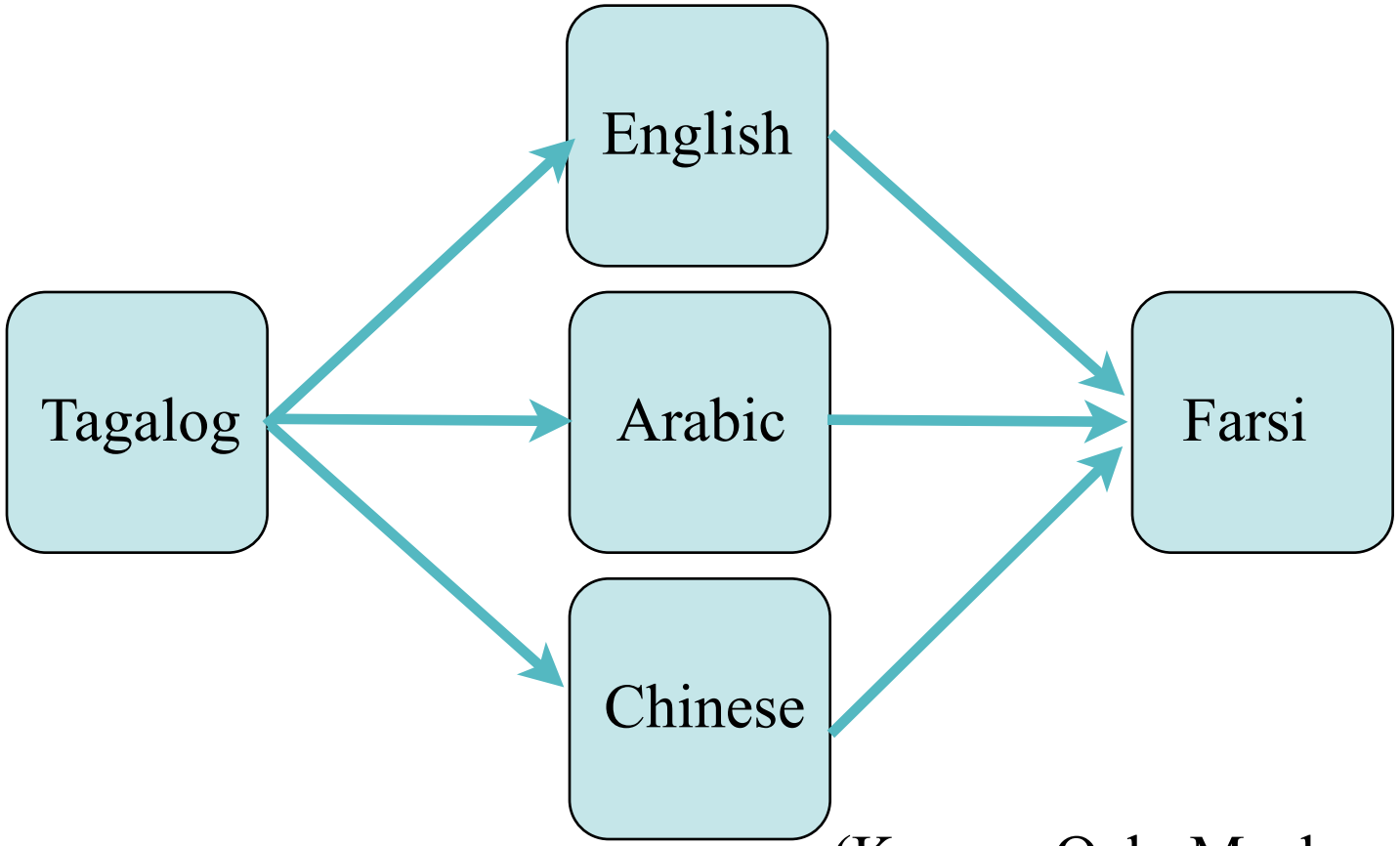
Et ellu è bellu e radiante cum grande splendore:
de Te, Altissimo, porta significatione.
Laudato si, mi Signore, per sora **Luna** e le stelle

Translating Poetry?
Eng(Wilde) ⇒ Fr(Human) ⇒ Eng(MT)

Wilde's original	**Our translation**
He did not wring his hands, as do Those witless men who dare To try to rear the changeling Hope In the cave of black Despair: He only looked upon the sun, And drank the morning air.	Without hands twisted like these men, Poor men without hope, dare To nourish hope in our vault Of desperation there And looked toward the sun, drink cool Until the evening air.
With slouch and swing around the ring We trod the Fool's Parade! We did not care: we knew we were The Devil's Own Brigade: And shaven head and feet of lead Make a merry masquerade.	We are in our circle we Dragged like the Fools' Parade! It mattered little, since we were The Devil's sad Brigade: A shaved head and the feet of lead Regardless gay charade!

(Genzel, Uszkoreit, Och, 2010)

Union of Languages: Practical



(Kumar, Och, Macherey, 2007)

Union of Languages: Interlingua?

ES: sillon

FR: fauteuil

EN: chair

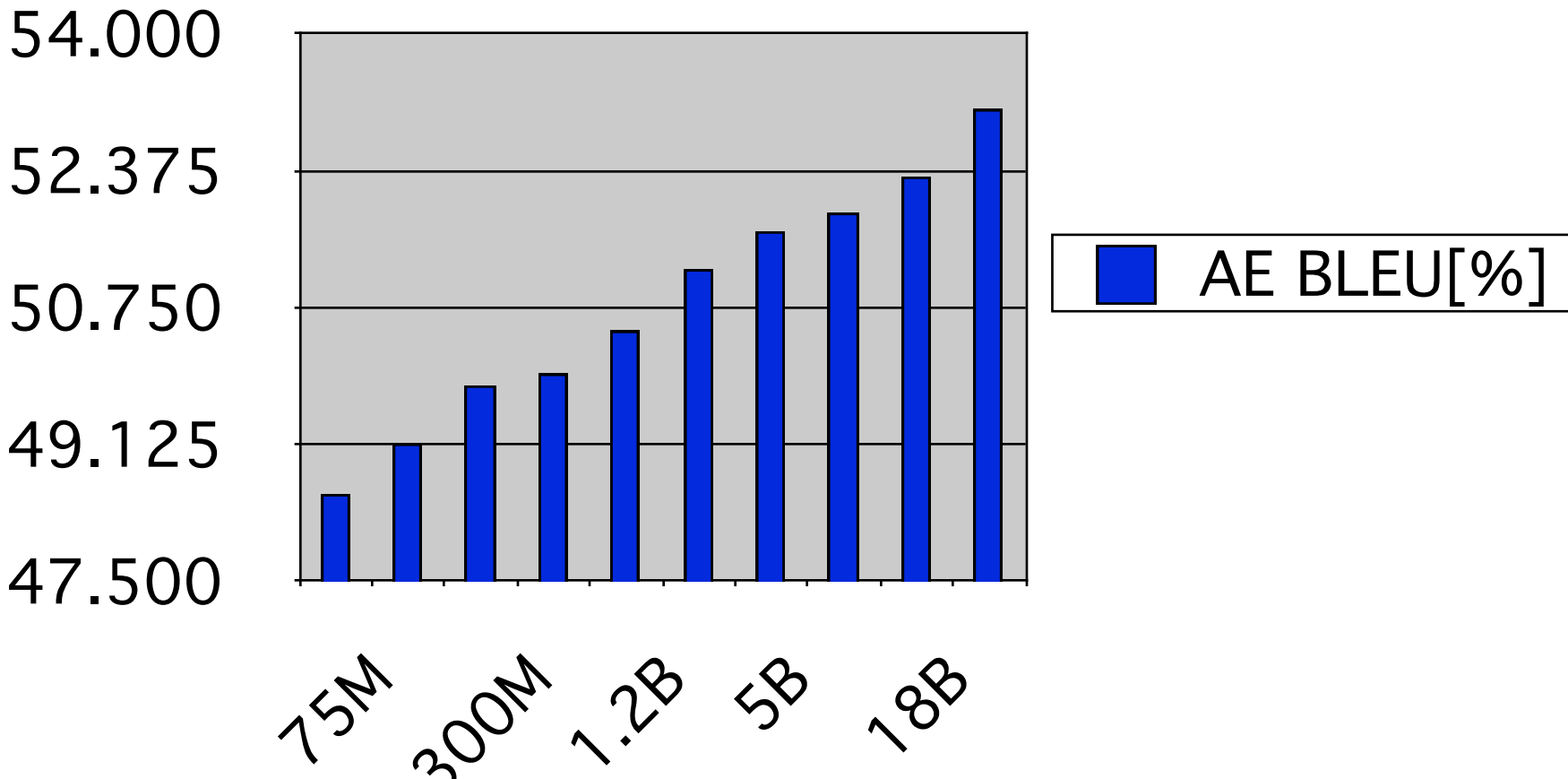
Extraction vs Annotation

- Annotation as a practical strategy
  - Minimize the re-computation of petabytes
- Annotation as a theoretical strategy
  - Addresses well-known problems:
  - Context
  - Qualification
  - Elaboration
  - Nonmonotonicity

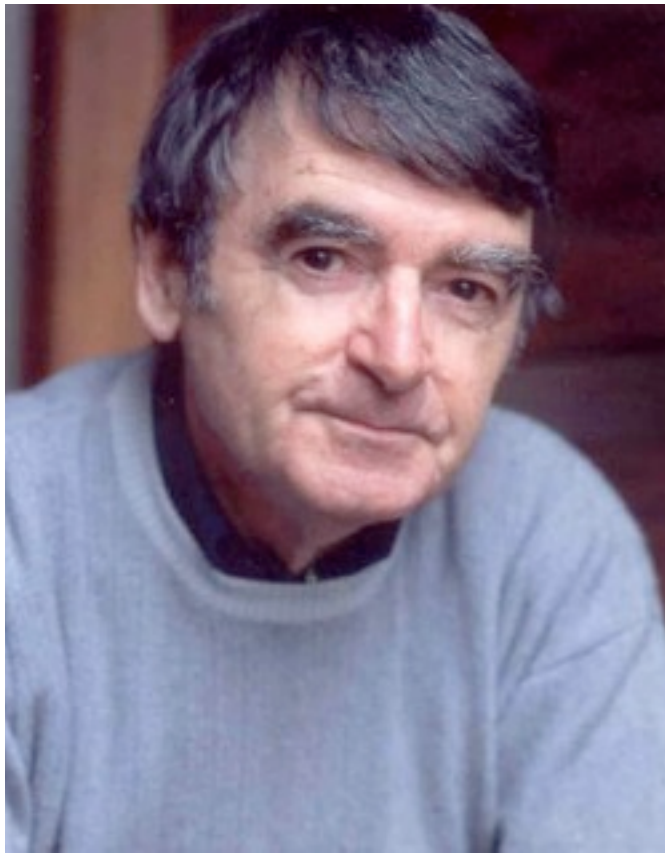
"Who are you and what do you want?" he asked gruffly, standing in front of them and towering tall above Gandalf.

Q: *Is Beorn taller than Gandalf?*

More Data, More Better



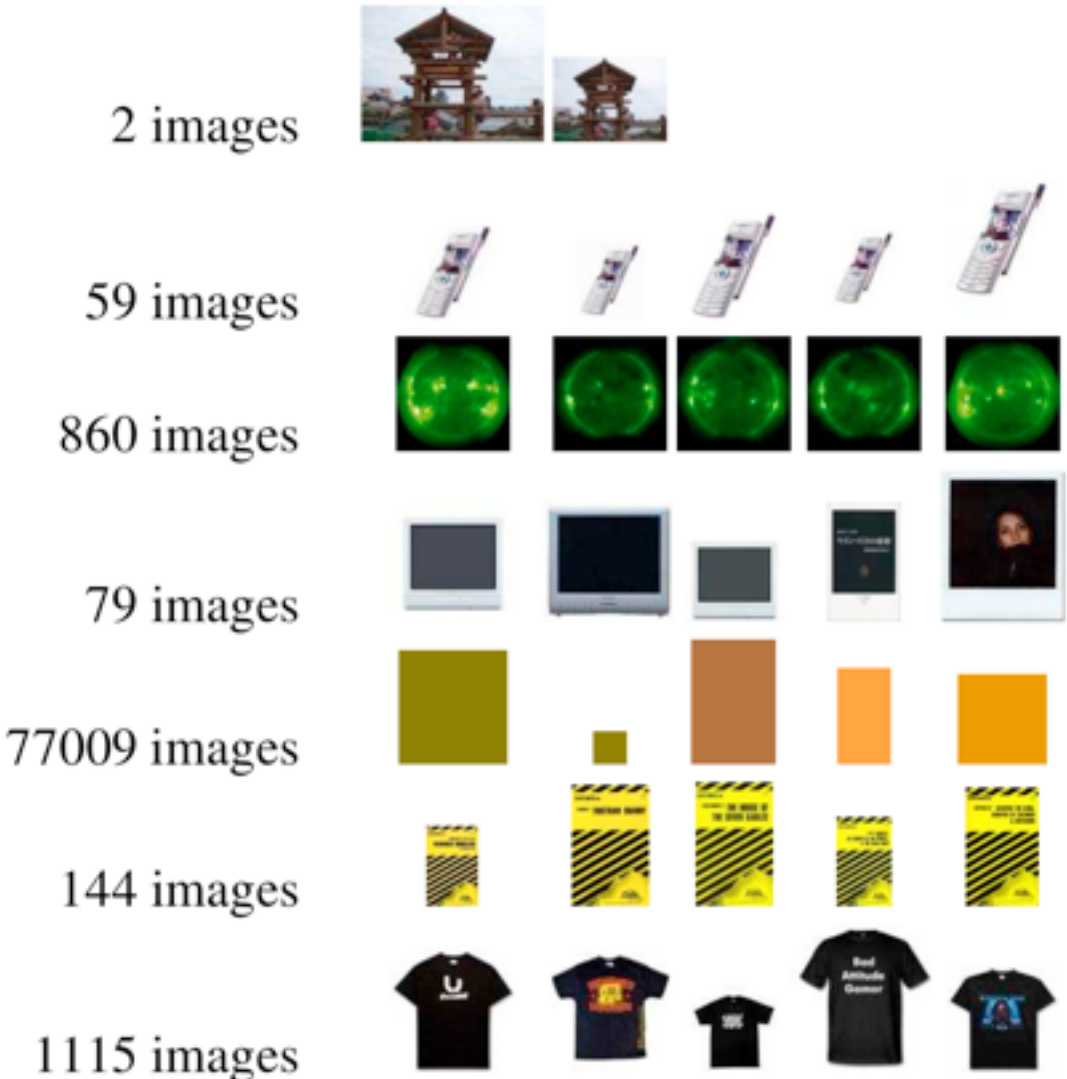
Arabic to English, Five-gram language model,
no count-cutoff, integrated into search



David Lodge (1935-)

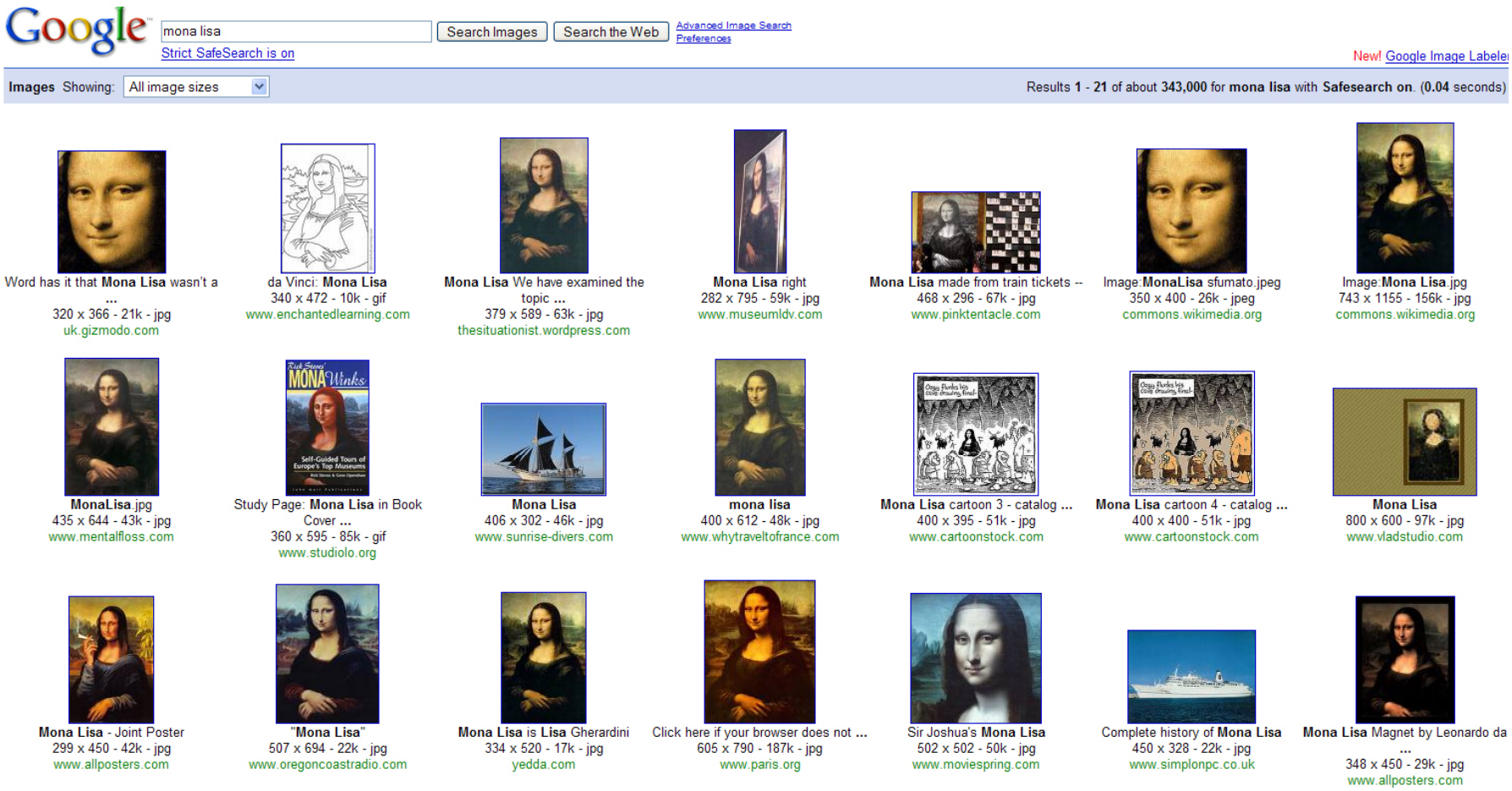
Literature is mostly about having sex, and not much about having babies. Life is the other way round.

Finding Near Duplicates



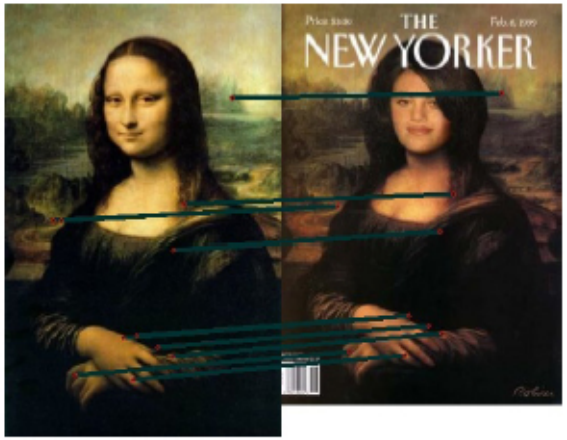
(Liu, Rosenberg, Rowley, 2007)

Finding Canonical Images

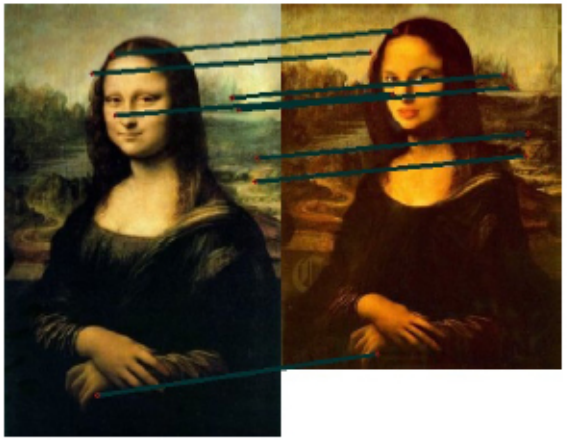


(Jing, Baluja, Rowley, 2007)

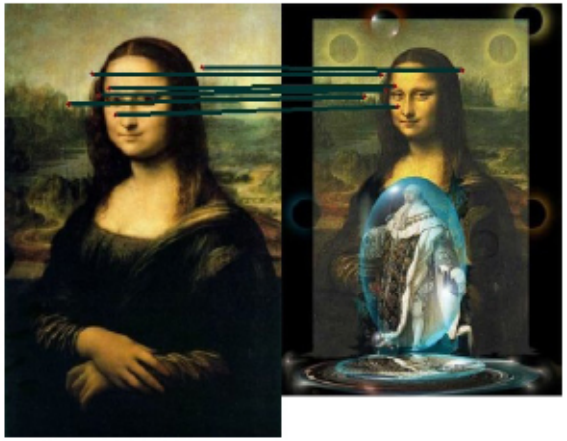
Compare low-level features



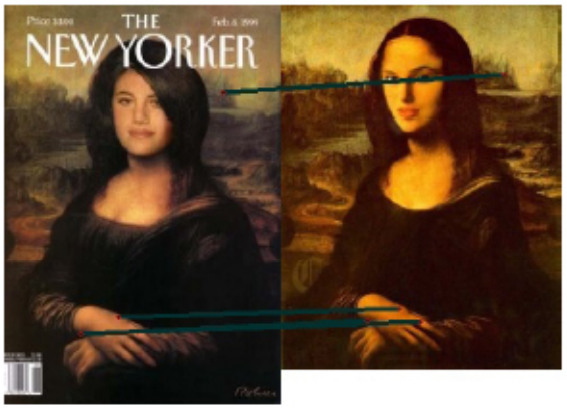
(a) A v.s. B



(b) A v.s. C

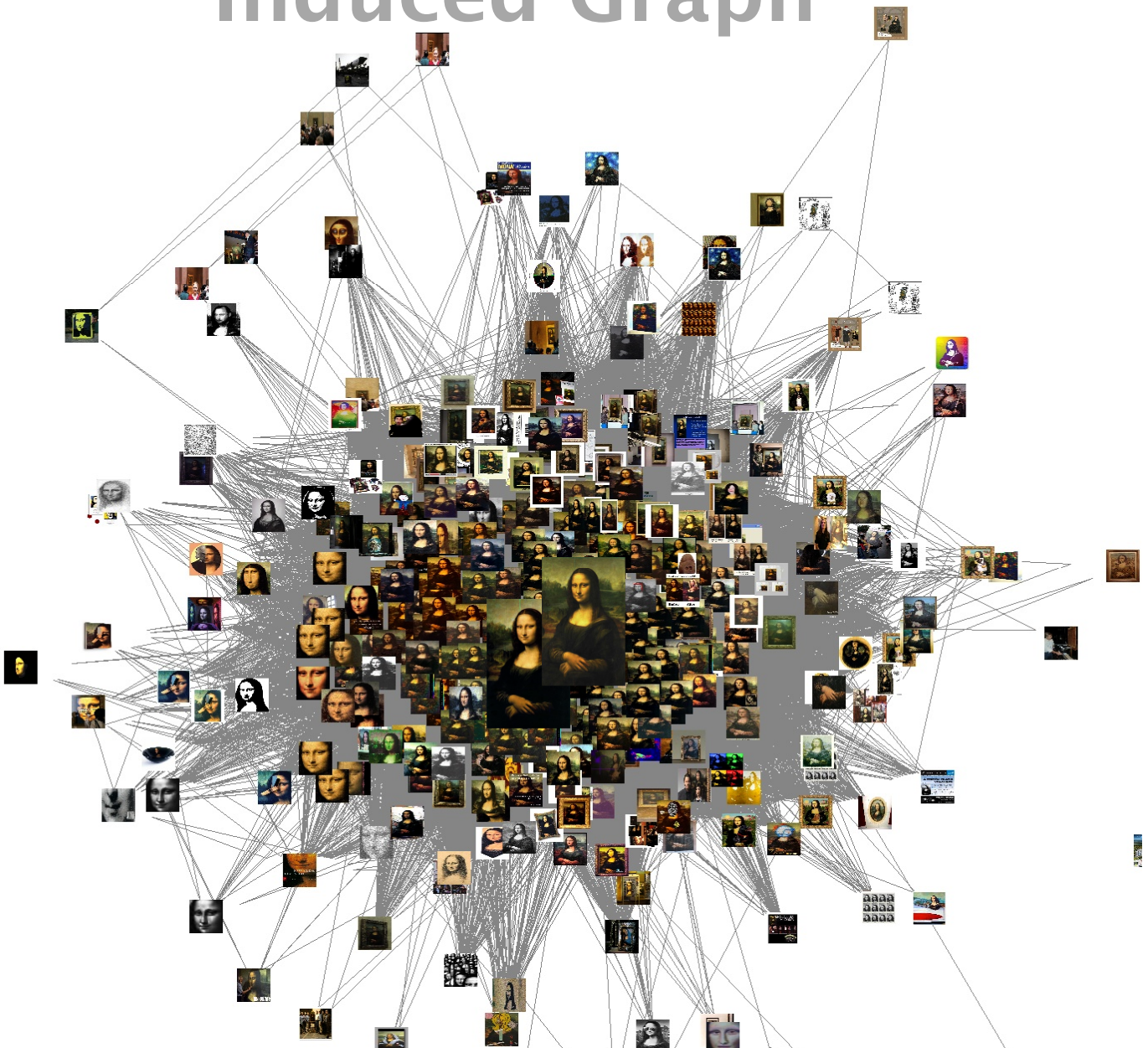


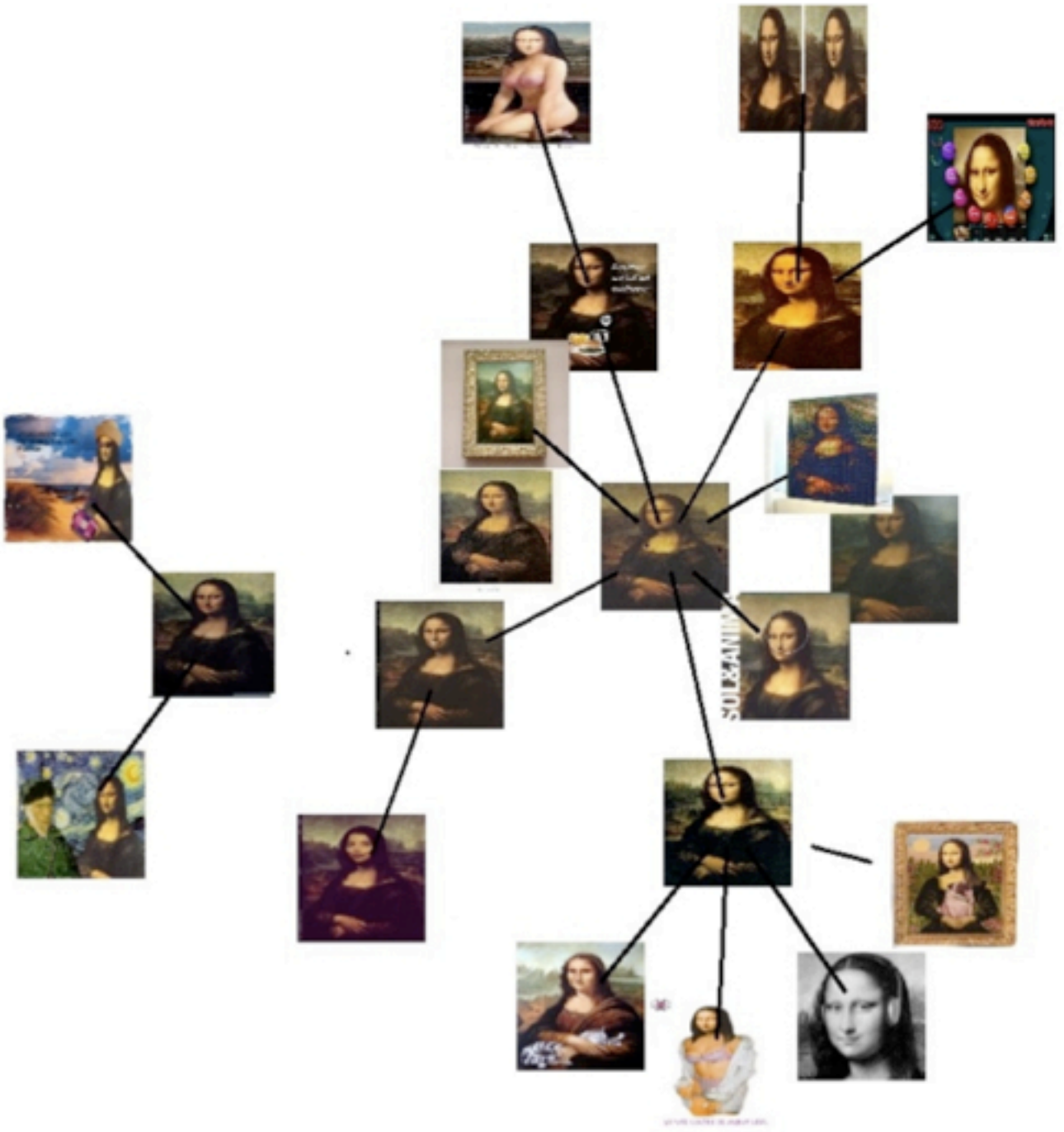
(c) A v.s. D



(d) B v.s. C

Induced Graph




SOLDANIS

SIFT Features

YouTube dog|dogs Search Browse

Search results for **dog|dogs**

About **1,190,000** results

Search options

Think This Is Funny?

See What Happens When Real **Dogs** Take Matters In To Their Own Hands.

by ThePurinaNetwork | **563,131 views**

Promoted Videos

Dog adopts panther and wolf cubs

First pictures of miracle **dog** nursing a nuzzling baby panther and wolf cub. . Follow us on twitter at twitter.com

by itnnews | 20 hours ago | **677 views**

NEW HD

From The Field: **Dog** Show In Louisville

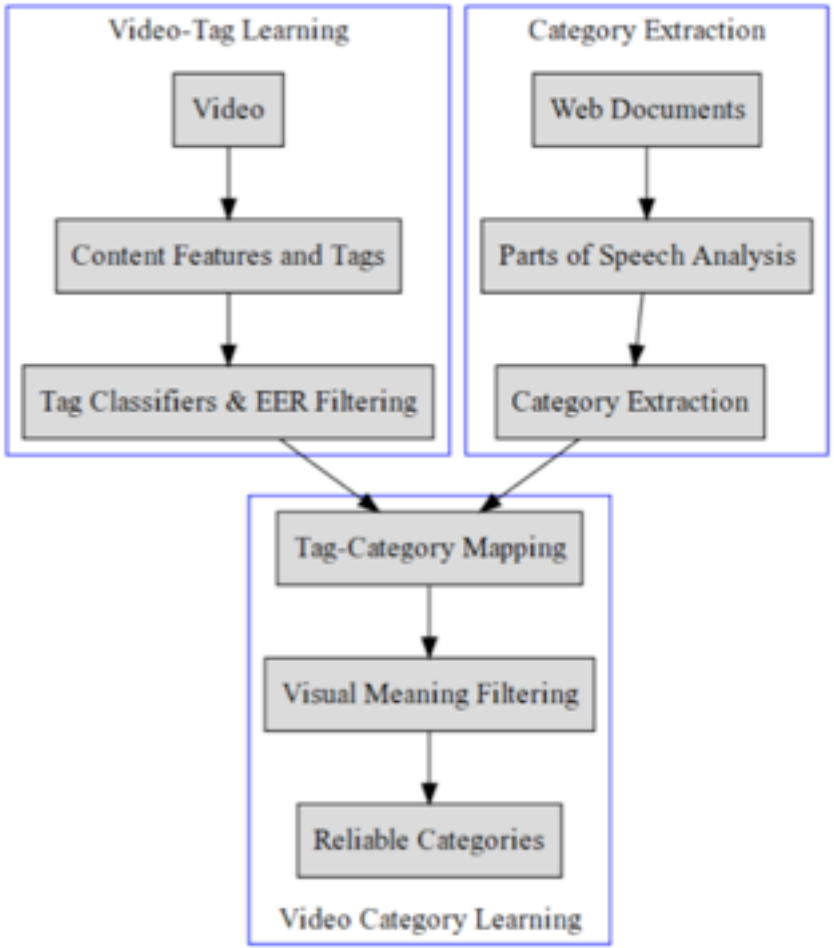
The AKC Kentuckiana Cluster All Breed **Dog** Show was in town this week. Watch video from the show!

by wlkytv | 18 hours ago | **1 views**

NEW

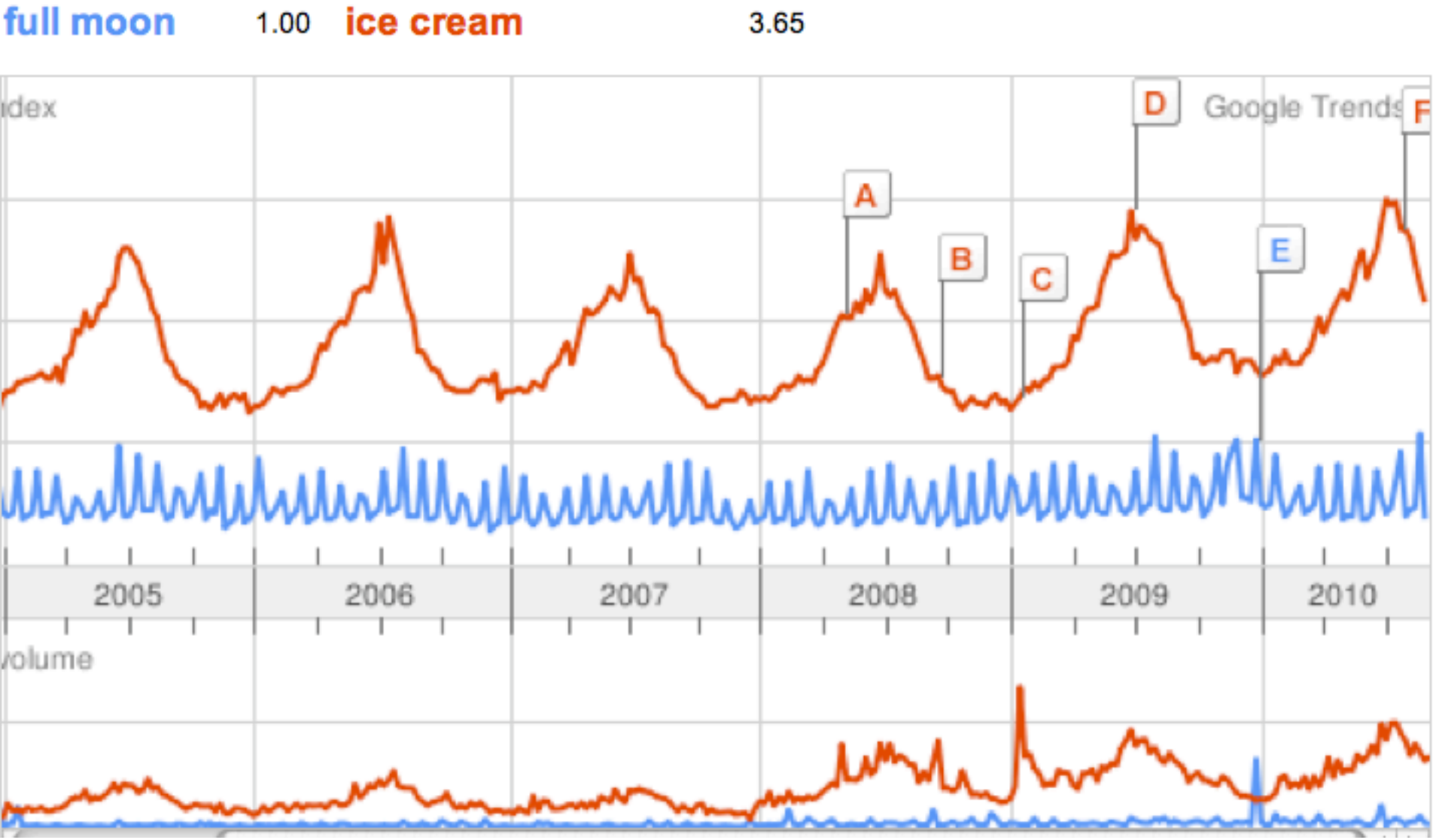
Automated tag learning with subtags

- Learn tag model 84% (prior result 57%)



Learning from users

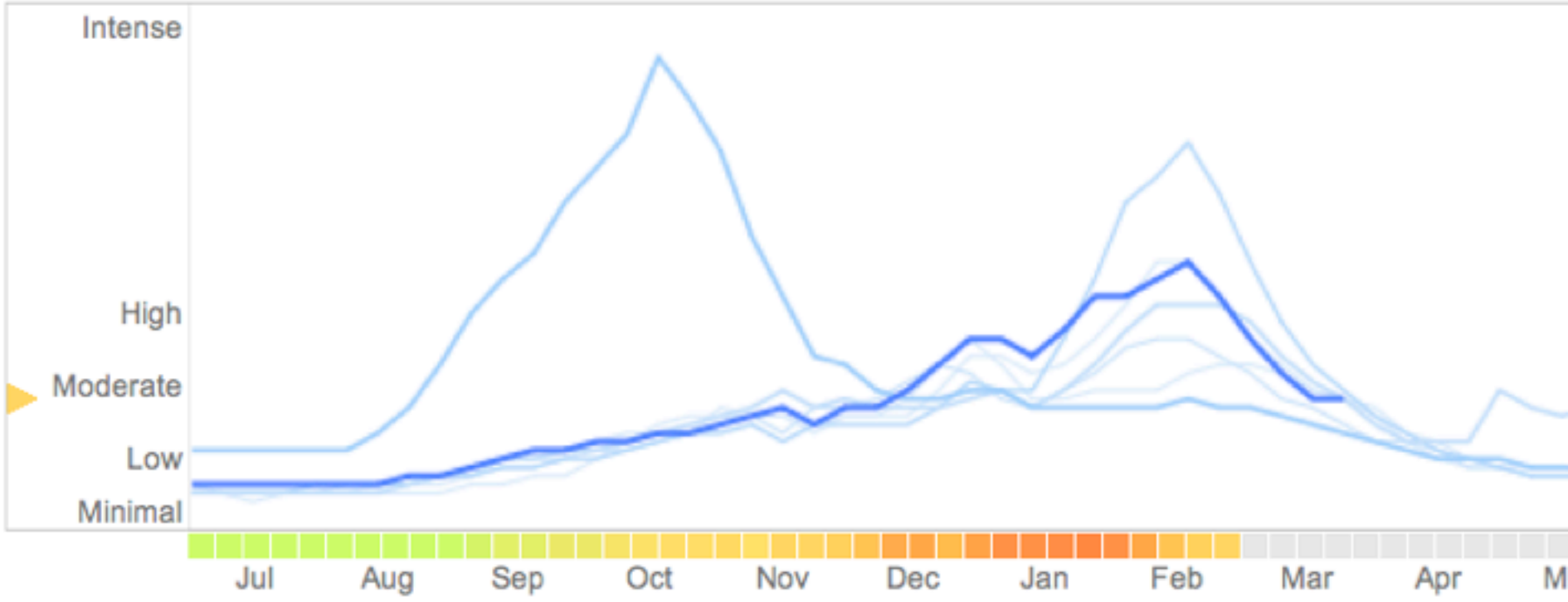
Orbital Mechanics with no Telescope or F = ma



Explore flu trends - United States

We've found that certain search terms are good indicators of flu activity. Google Flu Tr
aggregated Google search data to estimate flu activity. Learn more »

National



Google Image Labeler

label pass

r partner has suggested 17 labels.

zoom out

off-limits

lights
road
red
highway
light

my labels

taillights
curve
dark
sky
buildings
cars

(von Ahn and Dabbish, 2004)

Digitizing Books One Word at a Time

- HOME
- WHAT IS reCAPTCHA
  - DIGITIZATION ACCURACY
  - WHAT IS A CAPTCHA
  - SECURITY
- GET reCAPTCHA
- MY ACCOUNT
- EMAIL PROTECTION

Submit

The words above come from scanned books.
By typing them, you help to digitize old texts.