

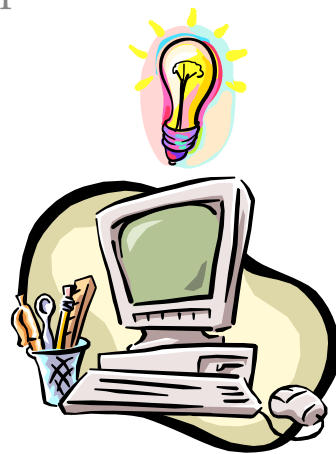
# Lecture 25: Language Games

CS442: Great Insights in Computer Science  
Michael L. Littman, Spring 2006

## Long-standing Debate

What can computers do, in principle?

- compute?
- think?
- learn?
- make judgments?
- exhibit creativity?
- feel?



Are these things even separable?

# Chess is One Thing...

The *drosophila* of AI. Intellectual game, taxes the human mind.

Simple in a sense:

- Choices, decision of winning/losing, all known in advance.
- Can live in the world of pure logic.
- No external influences.

**Closed World Assumption**

# Real Life is Another

Navigating cross campus:

- No single right answer.
- Trade off energy, time, safety, etc.
- Affects other aspects of life.
- Changes over time!

Learning, judgment, necessary.

**Open World Reality**

# Language Games

Like other games:

- Evaluation process clean.
- Challenging (and fun!) for people.

Unlike logical games:

- No closed world assumption.
- “AI-complete”!

Machine performance far from humans’.

## Example



Down

44 Telephone call?

44
R
E
N
G
S

Is this move legal? Is it wise?

# Word Games

Super-human performance common:

- Scrabble™: Maven, near-perfect (Sheppard 02)
- Boggle™: millisecond solutions (Boyan 98)
- Hangman (Littman 00)
  - 99.97% 9-letter words under 5 guesses
  - 1.35 misses on average

# Trivial Pursuit™

Race around board, answer questions.

Categories: Geography, Entertainment, History, Literature, Science, Sports



# Wigwam

QA via AQUA (Abney et al. 00)

- back off: word match in order helps score.
- “When was Amelia Earhart’s last flight?”
  - 1937, 1897 (birth), 1997 (reenactment)
- Named entities only, 100G of web pages

Move selection via MDP (Littman 00)

- Estimate category accuracy.
- Minimize expected turns to finish.

## Wigwam’s Knowledge

	wigwam	me	buff	web
arts & literature	.3	.6	.6	.9
entertainment	.3	.3	.5	.9
science & nature	.2	.7	.7	.7
geography	.1	.2	.4	.9
history	.1	.2	.5	.9
sports & leisure	.025	.6	.7	.4
<b>~turns/game</b>	<b>414</b>	<b>48</b>	<b>22</b>	<b>8</b>

# Who Wants to Be a Millionaire

“You know, we ought to enter her in one of those TV quiz shows. We could make a fortune.” (Danny Dunn in Williams & Abrashkin 58)

Multiple choice questions, increasing difficulty

1. 100, 200, 300, 500, 1000
2. 2000, 4000, 8000, 16000, 32000
3. 64000, 125000, 250000, 500000, 1000000

# Question Answering Approach

Choose highest ranked choice.

- 75%, 68%, 56% (Clarke, Cormack & Lynam 01)

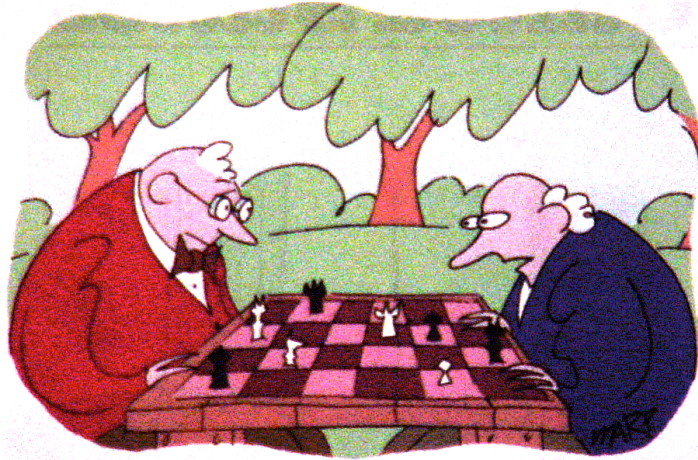
Expected value (always go on):

- \$3,689
- Most value due to (rare) \$1M.

People:

- \$97,357

# Why Crosswords?



**“We should try crossword puzzles.  
Crossword puzzles are babe magnets.”**

NYT, Saturday, October 10<sup>th</sup>, 1998

1	2	3	4	5	6		7	8	9	10	11	12	13	14				
H	O	T	O	N	E		P	A	L	O	M	I	N	O				
15	A	S	I	M	O	V		16	I	S	O	L	A	T	E	D		
17	S	L	E	E	V	E		18	T	H	W	A	R	T	E	D		
19	T	I	G	G	E	R		20	C	O	R	N	Y					
21	O	N	S	A	L	E		22	A	R	I	D		23	J	A	M	
						26	E	S	P	I	E	S		28	L	O	G	O
29	S	E	A	O	T	T	E	R		33	K	E	E	N	O	N		
35	A	B	B	O	T		36	A	N	A		38	U	S	A	G	E	
39	B	O	O	Z	E	S		41	S	N	A	P	S	H	O	T		
43	E	N	V	Y		44	P	L	I	N	T	H						
46	R	Y	E		47	H	I	E	S		48	T	E	A	S	E	T	
					53	K	A	R	E	L		54	I	M	P	A	L	E
55	M	A	R	I	N	A	R	A			58	M	I	A	S	M	A	
59	A	B	E	R	D	E	E	N			60	E	S	C	H	E	R	
61	J	U	N	K	Y	A	R	D			62	S	M	E	A	R	Y	

# Variety of Clue Types

Thesaurus	Cut off	<u>I</u> <u>S</u> <u>O</u> <u>L</u> <u>A</u> <u>T</u> <u>E</u> <u>D</u>
Puns & Wordplay	Monk's head?	<u>A</u> <u>B</u> <u>B</u> <u>O</u> <u>T</u>
Arts & Literature	"Foundation Trilogy" author	<u>A</u> <u>S</u> <u>I</u> <u>M</u> <u>O</u> <u>V</u>
Popular Culture	Pal of Pooh	<u>T</u> <u>I</u> <u>G</u> <u>G</u> <u>E</u> <u>R</u>
Encyclopedic	Mountain known locally as Chomolungma	<u>E</u> <u>V</u> <u>E</u> <u>R</u> <u>E</u> <u>S</u> <u>T</u>
Crosswordese	Kind of coal or coat	<u>P</u> <u>E</u> <u>A</u>

# PROVERB: System Design

Candidate generation (Keim et al. 99)

- Like information retrieval: clue implies target
- Variety of approaches used simultaneously

Merging

- Like meta search engine: create master list

Grid filling (Shazeer et al. 99)

- Like constraint satisfaction: fit answer to grid

*Probabilities are the common language*

# Modules: ClueDB

Nymph pursuer: SATYR Bugs pursuer: ELMER

Nymph chaser: SATYR Place for an ace: SLEEVE

Highball ingredient: RYE Highball ingredient: ICE

**$X$  chaser  $\rightarrow$   $X$  pursuer**

**exact:** Highball ingredient: RYE

**partial:** Ace place?: SLEEVE

**TransModule:** Bugs chaser: ELMER

Also [Dijkstra\[1-2\]](#), [d\[1-2\]c](#), [lsicwdb](#)

# Modules: Other DBs

Database modules: Transform clue to DB query.

**imdb:** Warner of Hollywood: OLAND

**wordnet:** Fruitless: ARID

Syntactic: Variations of fill-in-the-blanks.

also [blanks\\_{books, geo, movies, music, quotes}](#), [kindof](#)

**blanks\_movies:** “Heavens \_\_\_!”: ABOVE

Web search: Not used in experimental system.

**google:** “The Way To Natural Beauty” author, 1980: TIEGS

Also [rogetsyns](#), [geo](#), [writers](#), [compass](#), [myth](#),

[altavista](#), [yahoo](#), [infoseek](#), [EbModule](#), [lsiency](#), etc.

# Modules: Backstops

Word lists: Ignore clue, return all words.

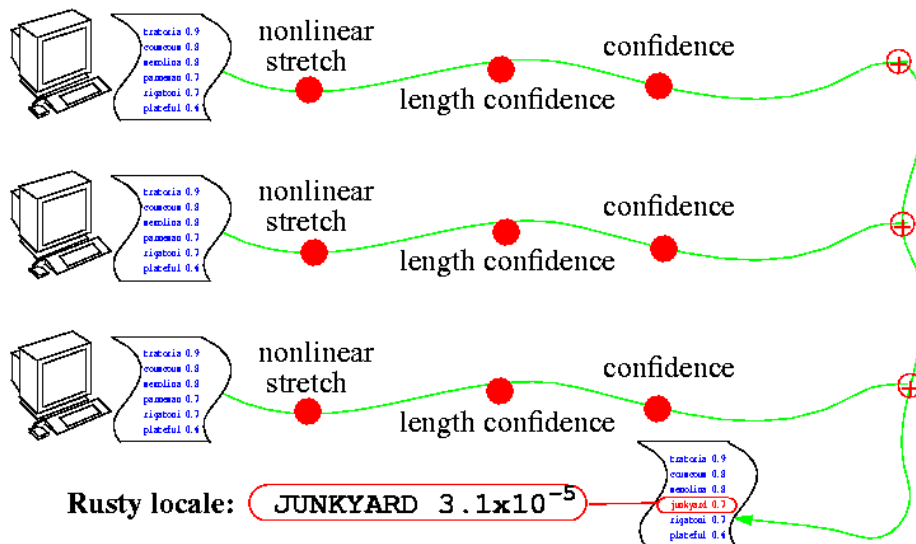
**wordList**: 10,000 words, perhaps: NOVELETTE

Implicit modules: Probability distributions over all strings of words (e.g., **bigram**).

**segmenter**: 1934 Hall and Nordhoff adventure novel: PITCAIRNISLAND

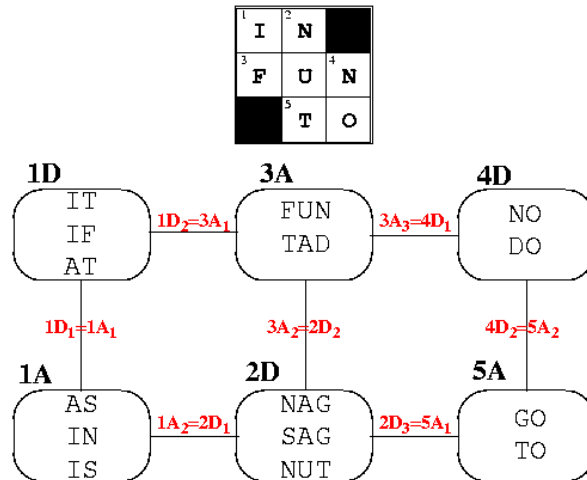
Also **bigWordList**, **wordList**, **DbList**

# Merging Candidate Lists



Hillclimbing used to tune the knobs to maximize the probability assigned to correct answers.

# Grid Filling and CSPs



Standard NP-hard search problem. Crossword puzzles are a classic (and misleading!) example.

# CSPs and IR

Domain from ranked candidate list?

Tortellini topping:

TRATORIA, COUSCOUS, SEMOLINA, **PARMESAN**,

RIGATONI, PLATEFUL, FORDLTDS, SCOTTIES,

ASPIRINS, MACARONI, FROSTING, RYEBREAD,

STREUSEL, LASAGNAS, GRIFTERS, BAKERIES,...

**MARINARA**, REDMEATS, VESUVIUS, ...

Standard recall / precision tradeoff.

# Basic Probability Idea

- Instead of having to choose a cutoff line between *correct* and *incorrect*, fill in the grid so that higher ranked choices are preferred.
- Maximize the *expected overlap between* our solution and a randomly generated solution that fits the probabilities.
- Hard problem, but can be approximated well in practice.

# Solution Grid

- Estimate the probability of each letter in each position in the grid.
- Iteratively improve the estimate by taking more context from the puzzle into account.
- Darkness of writing conveys the probability. Red squares have the most likely letter incorrect.

# Grid Filling

S	O	A	L	E	E		H	E	T	A	L	I	E	S
A	S	I	M	O	R		A	R	U	C	A	E	E	
S	I	R	E	R	E		C	A	R	C	E	T	E	
E	E	N	O	R	E		L	I	D	E	E			
S	R	E	E	E	D		A	A	I	N		A	A	M
						A	E	T	N	E		L	O	O
P	E	T	O		E	E	E	D		R	E	A	O	N
A	B	B	O	T		A	N	E		A	I	T	E	E
D	O	Z	E	S		O	N	R		T	E	E	T	
E	R	V	Y		P	A	R	N	T	E				
R	Y	E		H	I	E	S			R	E	A	A	E
				K	A	R	E	L		I	M	P	A	L
P	A	R	M	E	S	A	N			M	E	A	O	M
S	A	R	M	I	E	E				E	A	C	E	E
C	R	E	O	T	A	R	S			N	E	E	R	E

# Grid Filling

S	O	A	L	E	E		H	E	T	A	L	I	E	S
A	S	I	M	O	R		A	R	U	C	A	E	E	
S	I	R	E	R	E		C	A	R	C	E	T	E	
E	E	N	O	R	E		L	I	D	E	E			
S	R	E	E	E	D		A	A	I	N		A	A	M
						A	E	T	N	E		L	O	O
P	E	T	O		E	E	E	D		R	E	A	O	N
A	B	B	O	T		A	N	E		A	I	T	E	E
D	O	Z	E	S		O	N	R		T	E	E	T	
E	R	V	Y		P	A	R	N	T	E				
R	Y	E		H	I	E	S			R	E	A	A	E
				K	A	R	E	L		I	M	P	A	L
P	A	R	M	E	S	A	N			M	E	A	O	M
S	A	R	M	I	E	E				E	A	C	E	E
C	R	E	O	T	A	R	S			N	E	E	R	E

# Grid Filling

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

# Grid Filling

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

# Grid Filling

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

# Grid Filling

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

# Grid Filling

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

# Grid Filling

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

# Grid Filling

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

# Grid Filling

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

# Grid Filling

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

Final: 88% words, 97%

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

# PROVERB Results

Test collection (370 puzzles, @~15 min.)

- 95% words, 98% letters, 46% puzzles
- NYT: 89.5% (95.5% MTW, 85.0% TFSS)
- Ablation: ClueDB only 88%, no ClueDB 27%

American Crossword Puzzle Tournament

- 1998: 190/251, 80% words (vs. 100%)
  - tricks: letter pairs, words in single square
- 1999: 147/261, 75% words
  - tricks: Home is near: ALASKA

# TOEFL Synonyms

Used in college applications.

fish

- (a) scale
- (b) angle
- (c) swim
- (d) dredge

# Comparisons, Modules

- Landauer & Dumais (1997): LSA
  - 64.40% (52.90–74.80%)
- non-native speakers
  - 64.50% (53.01–74.88%)
- Turney (2001): PMI-IR
  - 73.75% (62.71–82.96%)
- Jarmasz & Szpakowicz (2002): Thesaurus
  - 78.75% (68.17–87.11%)
- Terra & Clarke (2003): web search
  - 81.25% (70.97–89.11%)
- Turney, Littman, Bigham, Shnayder (2003): modules
  - 97.50% (91.26–99.70%)

# Verbal Analogies

Used in college boards (SATs, GREs), and as an intelligence test. (Not any more.)

cat : meow ::

(a) mouse : scamper

(b) bird : peck

(c) dog : bark

(d) horse : groom

(e) lion : scratch

45% (thesaurus, web, connecting words).

# Wrap Up

Modular language-game systems.

PROVERB:

- Human-competitive performance.
- Components theoretically motivated.
- Probabilistically grounded.
- Driven by training data, not rules.

# In a Nutshell

- “Judgment” here comes from a comparison of choices to a large database of examples.
- Good choices “fit better” with existing examples. No formal rules.
- Performance improvements come from more examples, not more computation.

# Next Time

- Polynomiography