

Linguistic properties of multi-word passphrases

Joseph Bonneau, Ekaterina Shutova

jcb82, es407@cl.cam.ac.uk



UNIVERSITY OF
CAMBRIDGE

Computer Laboratory

USEC WORKSHOP ON USABLE SECURITY 2012

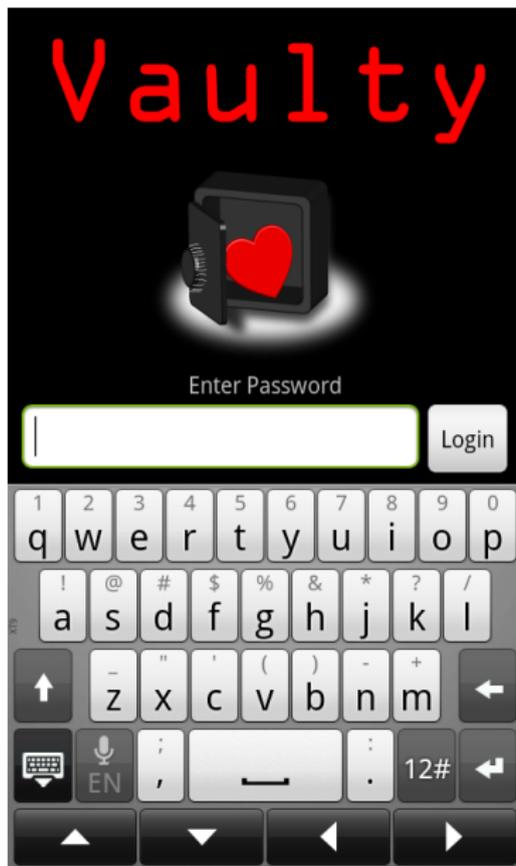
KRALENDIJK, BONAIRE, NETHERLANDS

MARCH 2, 2012

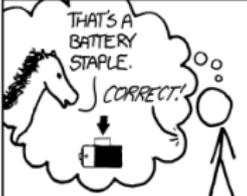
Passphrases an increasingly attractive approach



Passphrases an increasingly attractive approach



Passphrases an increasingly attractive approach

<p>UNCOMMON (NON-GIBBERISH) BASE WORD</p> <p>ORDER UNKNOWN</p> <p>Trøub4dor & 3</p> <p>CAPS? COMMON SUBSTITUTIONS NUMERAL PUNCTUATION</p> <p>(YOU CAN ADD A FEW MORE BITS TO PRODUCE FOR THE FACT THAT THIS IS ONLY ONE OF A FEW COMMON FORMATS)</p>	<p>~28 BITS OF ENTROPY</p> <p>$2^{28} = 3 \text{ DAYS AT } 1000 \text{ GUESSES/SEC}$</p> <p>(PLAUSIBLE ATTACK ON A WORK REMOTE WEB SERVICE: YES, CRACKING A STOKEN HASH IS FASTER, BUT IT'S NOT WHAT THE AVERAGE USER SHOULD WORRY ABOUT.)</p> <p>DIFFICULTY TO GUESS: EASY</p>	<p>WAS IT TROMBONE? NO, TROUBADOR. AND ONE OF THE O's WAS A ZERO?</p> <p>AND THERE WAS SOME SYMBOL....</p>  <p>DIFFICULTY TO REMEMBER: HARD</p>
<p>correct horse battery staple</p> <p>FOUR RANDOM COMMON WORDS</p>	<p>~44 BITS OF ENTROPY</p> <p>$2^{44} = 550 \text{ YEARS AT } 1000 \text{ GUESSES/SEC}$</p> <p>DIFFICULTY TO GUESS: HARD</p>	<p>THAT'S A BATTERY STAPLE.</p> <p>CORRECT!</p>  <p>DIFFICULTY TO REMEMBER: YOU'VE ALREADY MEMORIZED IT</p>

THROUGH 20 YEARS OF EFFORT, WE'VE SUCCESSFULLY TRAINED EVERYONE TO USE PASSWORDS THAT ARE HARD FOR HUMANS TO REMEMBER, BUT EASY FOR COMPUTERS TO GUESS.

xkcd #936

What do we know about passphrase guessing?

[this space intentionally left blank]

Data source: Amazon PayPhrases

Enter your own phrase to see if it's available or claim this one:

“ ”

” (PayPhrase must be at least 2 words long, no numbers or special characters)

✘ Sorry, this phrase is not available

- must be at least two words
- must be globally unique
 - security ← PIN + passphrase
- can only contain the letters a-z, A-Z, SPACE
 - capitalisation and spacing ignored

Data source: Amazon PayPhrases



PayPhrases killed 2012-02-20

A simple dictionary attack



WIKIPEDIA
The Free Encyclopedia

Main page

Contents

Featured content

Current events

Random article

Donate to Wikipedia

Interaction

Help

About Wikipedia

Community portal

Recent changes

Contact Wikipedia

Toolbox

Print/export

Languages

Esperanto

Suomi

Português

Русский

Article Talk

Read View source

List of board games

From Wikipedia, the free encyclopedia

This is a list of **board games**. This page classifies board games according to the concerns which might be uppermost for someone organizing a gaming event or party. See the article on [game classification](#) for other a&ts board game articles.

Contents (show)

Two-player abstract strategy games

Main article: *List of abstract strategy games*

In *abstract strategy games*, players know the entire game state at all times, and random generators such as dice are not used.

Two-player games

- Abalone
- Algon
- Backgammon
- Battleship
- Blood Bowl
- Bul
- Chess
- Chinese Checkers
- Connect Four
- The Cross and Circle game family
- Dalás
- Downfall
- DVD&N
- English draughts (Checkers)
- Fanorona
- Ghosts
- Gipf
- Go
- Guest Who?
- Hare and Hounds
- Hjara
- Isola
- Janggi (Korean Chess)
- Kalah
- Kamasado
- Kingdoms
- Liú pǎi
- Lost Cities
- Mad Gab
- Master Mind
- Nine Men's Morris
- Obsession
- Patoli
- Pawn duel
- Plateau
- PÜNCT
- Reversi (aka Othello)
- Rehmmamachy
- Sáhkku
- Sanet
- Shogi (Japanese Chess)
- Snakes and Ladders
- Space Hulk
- Stratego
- Sugoroku
- Táb
- Tantrix
- War
- Xiangqi (Chinese Chess)
- YINSH
- ZERTZ

Multi-player elimination games

Participants are typically eliminated before game end.

- 1313 Dead End Drive
- American Megafallana
- Anti-Monopoly
- Apples to Apples
- Attack!
- Axis & Allies
- Bang!
- Crash! The bankrupt game
- Diplomacy
- Djambi
- France
- Ghettopoly
- The Great Train Robbery Board Game
- Hotels
- Risk
- Shogun/Samurai Swords
- Solarquest
- Spy Alley
- Star Wars Epic Duels
- Star Wars Tactics
- Strange Synergy

- proper nouns
- titles
- idiomatic phrases
- slang

A simple dictionary attack

The screenshot shows the IMDb website's 'All-Time Worldwide Box office' chart. The chart lists the top 26 movies worldwide based on their box office revenue. The top entries are 'Avatar' (2009) at \$2,781,505,847, 'Titanic' (1997) at \$1,835,300,000, and 'Harry Potter and the Deathly Hallows - Part 2' (2011) at \$1,327,455,619. The chart is organized into columns for Rank, Title, and Worldwide Box Office.

Rank	Title	Worldwide Box Office
1.	Avatar (2009)	\$2,781,505,847
2.	Titanic (1997)	\$1,835,300,000
3.	Harry Potter and the Deathly Hallows - Part 2 (2011)	\$1,327,455,619
4.	The Lord of the Rings - The Return of the King (2003)	\$1,119,102,668
5.	Transformers - Dark of the Moon (2011)	\$1,114,559,779
6.	Pirates of the Caribbean: Dead Man's Chest (2006)	\$1,065,896,541
7.	Toy Story 3 (2010)	\$1,062,884,497
8.	Pirates of the Caribbean: On Stranger Tides (2011)	\$1,041,963,875
9.	Alice in Wonderland (2010)	\$1,023,285,206
10.	The Dark Knight (2008)	\$1,001,921,825
11.	Harry Potter and the Philosopher's Stone (2001)	\$969,457,891
12.	Pirates of the Caribbean: At World's End (2007)	\$958,404,152
13.	Harry Potter and the Deathly Hallows - Part 1 (2010)	\$946,880,434
14.	Harry Potter and the Order of the Phoenix (2007)	\$937,300,866
15.	Harry Potter and the Half-Blood Prince (2009)	\$933,956,480
16.	Star Wars: Episode I - The Phantom Menace (1999)	\$922,379,000
17.	The Lord of the Rings - The Two Towers (2002)	\$921,600,000
18.	Jurassic Park (1993)	\$919,700,000
19.	Harry Potter and the Goblet of Fire (2005)	\$892,194,197
20.	Ice Age: Dawn of the Dinosaurs (2009)	\$887,773,705
21.	Spider-Man 3 (2007)	\$885,430,103
22.	Shrek 2 (2004)	\$880,871,036
23.	Harry Potter and the Chamber of Secrets (2002)	\$866,300,000
24.	Finding Nemo (2003)	\$865,800,000
25.	The Lord of the Rings - The Fellowship of the Ring (2001)	\$860,700,000
26.	Star Wars: Episode III - Revenge of the Sith (2005)	\$848,462,555

- proper nouns
- titles
- idiomatic phrases
- slang

A simple dictionary attack

The screenshot shows the UsingEnglish.com website interface. At the top, there is a navigation menu with options like Home, Members, Students, Teachers, Forums, Testing, Reference, Articles, Resources, and Shop. Below this is a search bar and a list of categories including Idioms, Phrasal Verbs, Irregular Verbs, Grammar Glossary, English Dictionaries, and Downloads. The main content area is titled 'English Idioms & Idiomatic Expressions' and features a search bar with the letter 'A' selected. Below the search bar, there is a list of idioms starting with 'A', each with a brief description. The idioms listed are: 'A bit much', 'A bridge too far', 'A chain is no stronger than its weakest link', 'A day late and a dollar short', 'A fool and his money are soon parted', 'A fool at 40 is a fool forever', 'A fresh pair of eyes', 'A hitch in your giddy-up', 'A lick and a promise', 'A List', 'A little bird told me', and 'A little learning is a dangerous thing'. On the right side of the page, there is a 'SUBSCRIBE' section with a form to enter an email address and a 'Become a Member' section with a registration link. At the bottom right, there is a 'Don't Miss...' section with a link to 'More Discussion Focus'.

- proper nouns
- titles
- idiomatic phrases
- slang

A simple dictionary attack

The screenshot shows the Urban Dictionary website interface. At the top, there's a search bar and navigation links for 'word of the day', 'dictionary', 'thesaurus', 'names', 'media', 'store', 'add', and 'help'. Below the navigation is a grid of letters A-Z, with 'A' highlighted. The main content area is titled 'Most popular words in A' and lists 100 words in three columns. The words are: A, Abilisk FireDragon, abc sex, abstinence where, Auto Incorrect, acid, asplode, and then I found five dollars, ass, a crapola, ar guitar, Advertising, acid sex, andrew, angry price, angelfin job, academic totema, asstet, angry dragon, aban, AB-tobler, Anzacbath, Anzaku, asplode mode, acid sex, Altv, Ales, Ales Salomon, aurl laugre, ass, asplode, afterdip, asplode, arch douche, alone, Anna, apocalypse sex, Ann Coulter, ar quite, Angelfin job, Audience typing, asin, aversore, ar sex, Australia, abhor, and.

A	A	A
A	Abilisk	Angry Dragon
Abilisk FireDragon	Abner	ant
abc sex	angelina job	Antisocial Networking
abstinence where	autismgeeky	Asshole
Auto Incorrect	acid	as Ferg
acid	airport visitors	Ass! Sex
asplode	advertisers hot pocket	astrow
and then I found five dollars	asf	ass
ass	Angelfin Job	Asshat
a crapola	aslan	assward turtle
ar guitar	angelina job	Atty
Advertising	Angelfin Job	acids
acid sex	asplode	acids
andrew	angelina job	acids
angry price	Angelfin Job	Andre
angelfin job	As You Know	Andre
academic totema	angelina job	Anonymous
asstet	Angelfin Job	alexandra
angry dragon	apple	Andre
aban	Angelfin Job	Angelfin Job
AB-tobler	Alligator Fuckhouse	Andre
Anzacbath	Allyssa	Andre
Anzaku	Angelfin Job	all your base are belong to us
asplode mode	asme	ASBEPROMISE AND FITCH
acid sex	Angelfin Job	asplode
Altv	asf	asin
Ales	Angelfin Job	abercrombie
Ales Salomon	Abe Lincoln	asplode
aurl laugre	a	America
ass	abclat	Angelfin Job
asplode	Anonymous	andrew
afterdip	approval from corporate	Alban
asplode	agronomic	weeping
arch douche	assward am	ACL
alone	Andrew	adidas
Anna	ANGL	ADG
apocalypse sex	Anthony	aght
Ann Coulter	angelina job	ab
ar quite	Andre	asid
Angelfin Job	abofghatla	Ann Coulter
Audience typing	Asplode	australia
asin	Asshole	AZN
aversore	Al Gore	anonymous
ar sex	as	asshole
Australia	Abilisk FireDragon	aght
abhor	angelina job	angelina job
and	Alton	

- proper nouns
- titles
- idiomatic phrases
- slang

Results

word list	example	list size	success rate	\hat{p}
<i>arts</i>				
musicians	three dog night	679	49.5%	0.0464%
albums	all killer no filler	446	56.5%	0.0372%
songs	with or without you	476	72.9%	0.0623%
movies	dead poets society	493	69.6%	0.0588%
movie stars	patrick swayze	2012	28.1%	0.0663%
books	heart of darkness	871	47.0%	0.0553%
plays	guys and dolls	75	70.7%	0.0093%
operas	la gioconda	254	17.3%	0.0048%
TV shows	arrested development	836	46.3%	0.0520%
fairy tales	the ugly duckling	813	13.3%	0.0116%
paintings	birth of venus	268	11.2%	0.0032%
brand names	procter and gamble	456	17.3%	0.0087%
<i>total</i>		7679	38.5%	0.4159%
<i>sports teams</i>				
NHL	new jersey devils	30	83.3%	0.0056%
NFL	arizona cardinals	32	87.5%	0.0070%
NBA	sacramento kings	29	93.1%	0.0085%
MLB	boston red sox	30	90.0%	0.0074%
NCAA	arizona wildcats	126	56.3%	0.0105%
fantasy sports	legion of doom	121	71.1%	0.0151%
<i>total</i>		368	71.7%	0.0542%
<i>sports venues</i>				
professional stadiums	soldier field	467	14.1%	0.0071%
collegiate stadiums	beaver stadium	123	12.2%	0.0016%
golf courses	shadow creek	97	6.2%	0.0006%
<i>total</i>		687	12.7%	0.0094%

Results

word list	example	list size	success rate	\hat{p}
<i>games</i>				
board games	luck of the draw	219	28.8%	0.0074%
card games	pegs and jokers	322	27.6%	0.0104%
video games	counter strike	380	28.4%	0.0127%
<i>total</i>		921	28.2%	0.0306%
<i>comics</i>				
print comics	kevin the bold	1029	29.5%	0.0361%
web comics	something positive	250	16.8%	0.0046%
superheros	ghost rider	488	45.3%	0.0295%
<i>total</i>		1767	32.1%	0.0701%
<i>place names</i>				
city, state (USA)	plano texas	2705	33.8%	0.1117%
multi-word city (USA)	maple grove	820	79.0%	0.1283%
city, country	lisbon portugal	479	35.7%	0.0212%
multi-word city	ciudad juarez	55	69.1%	0.0066%
<i>total</i>		4059	43.7%	0.2677%
<i>phrases</i>				
sports phrases	man of the match	778	26.1%	0.0235%
slang	sausage fest	1270	45.0%	0.0761%
idioms	up the creek	3127	43.6%	0.1789%
<i>total</i>		5175	41.3%	0.2785%

Results

- Estimating $N = 10^6$, our 20k dictionary covers 1.1% of users
 - Equivalent to 20.8 bits
- Password comparison #1: 2 passwords cover 1.1% of users
 - Equivalent to 7.5 bits
- Password comparison #2: 20k dictionary covers 26.3% of users
 - Equivalent to 16.3 bits

Similar to mnemonic-phrase passwords

Which syntactic construction do users prefer?

Which syntactic construction do users prefer?

bigram type	example	list size	success rate
adverb-verb	probably keep	4999	5.0%
verb-adverb	send immediately	4999	1.9%
direct object-verb	name change	5000	1.2%
verb-direct object	spend money	5000	2.4%
verb-indirect object	go on holiday	4999	0.7%
nominal modifier-noun	operation room	4999	9.8%
subject-verb	nature explore	4999	1.3%

Phrases generated from British National Corpus/Robust Accurate Statistical Parser

Single objects or actions strongly preferred

Which factors predict a phrase's popularity?

Which factors predict a phrase's popularity?

bigram type	example	list size	success rate
adjective-noun	powerful form	10000	13.3%
noun-noun	island runner	10000	4.4%

Phrases generated from Google n-gram corpus

Which factors predict a phrase's popularity?

Possible selection models:

baseline

random

natural-language production

$$p(w_1 || w_2)$$

independent word selection

$$p(w_1) \cdot p(w_2)$$

mutual information

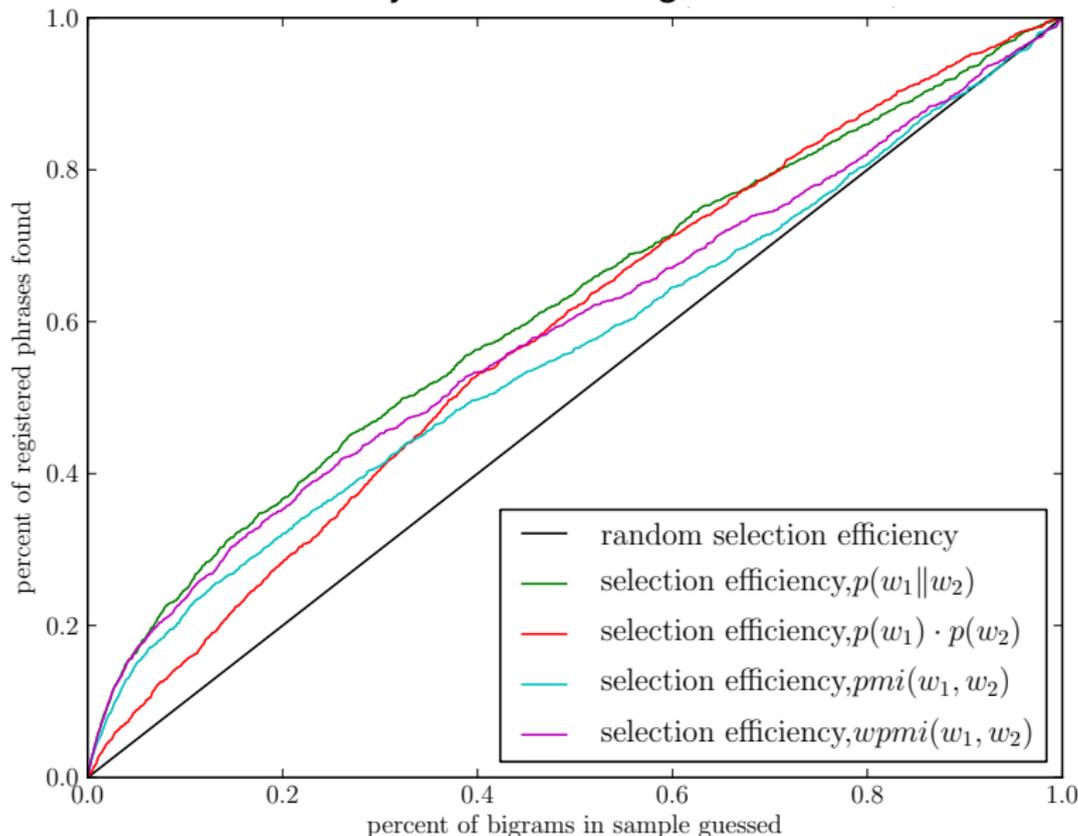
$$pmi(w_1, w_2) = \lg \frac{p(w_1 || w_2)}{p(w_1) \cdot p(w_2)}$$

blended model

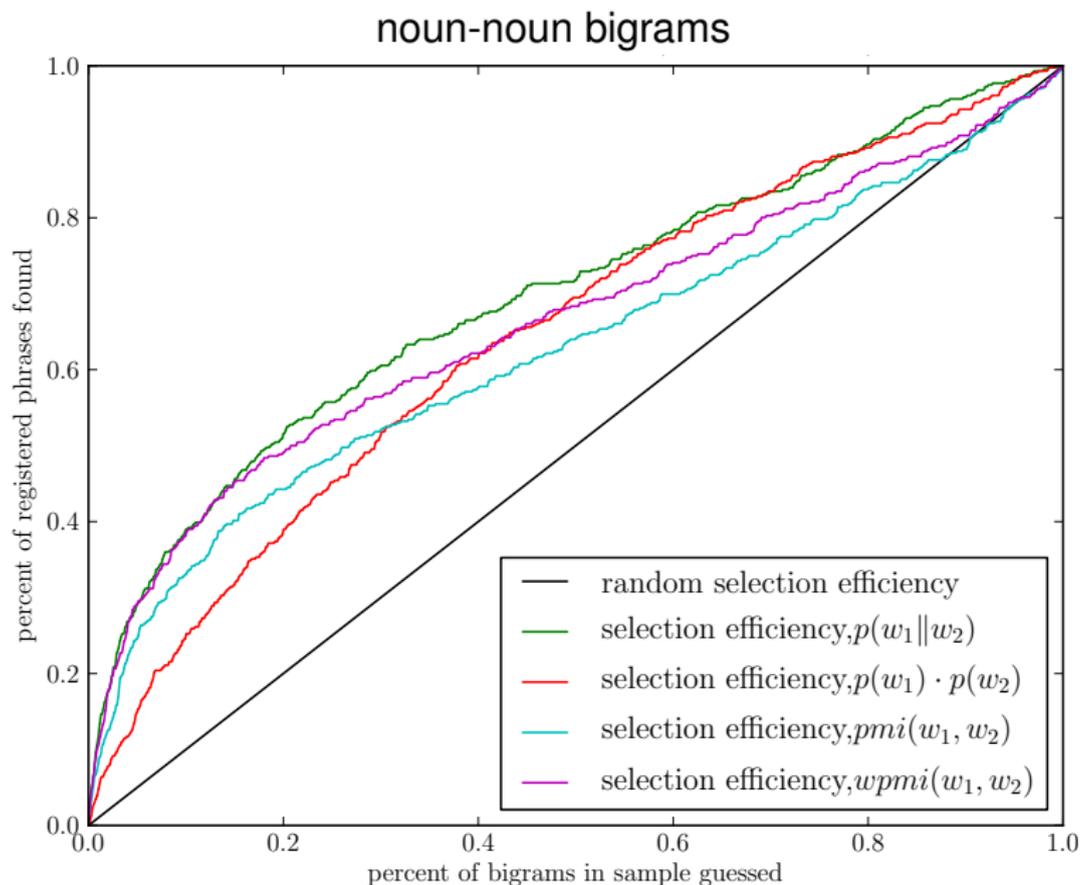
$$p(w_1 || w_2) \cdot pmi(w_1, w_2)$$

Which factors predict a phrase's popularity?

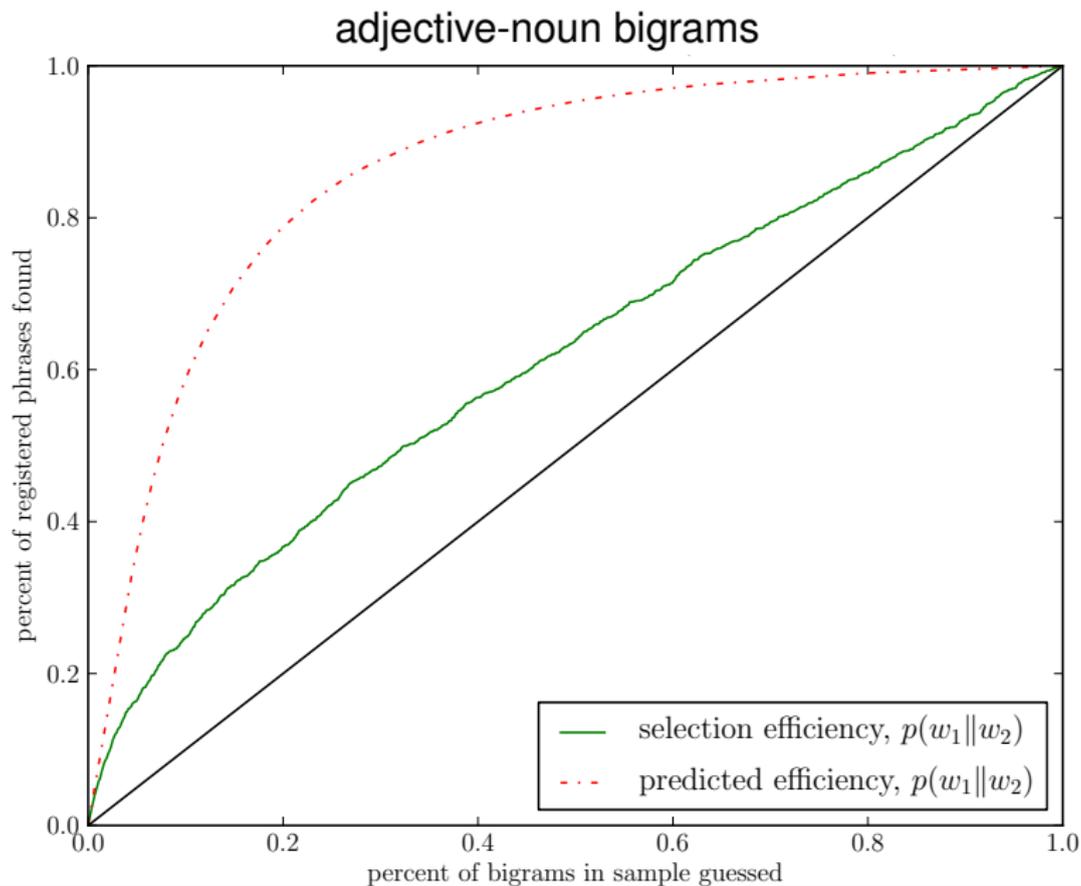
adjective-noun bigrams



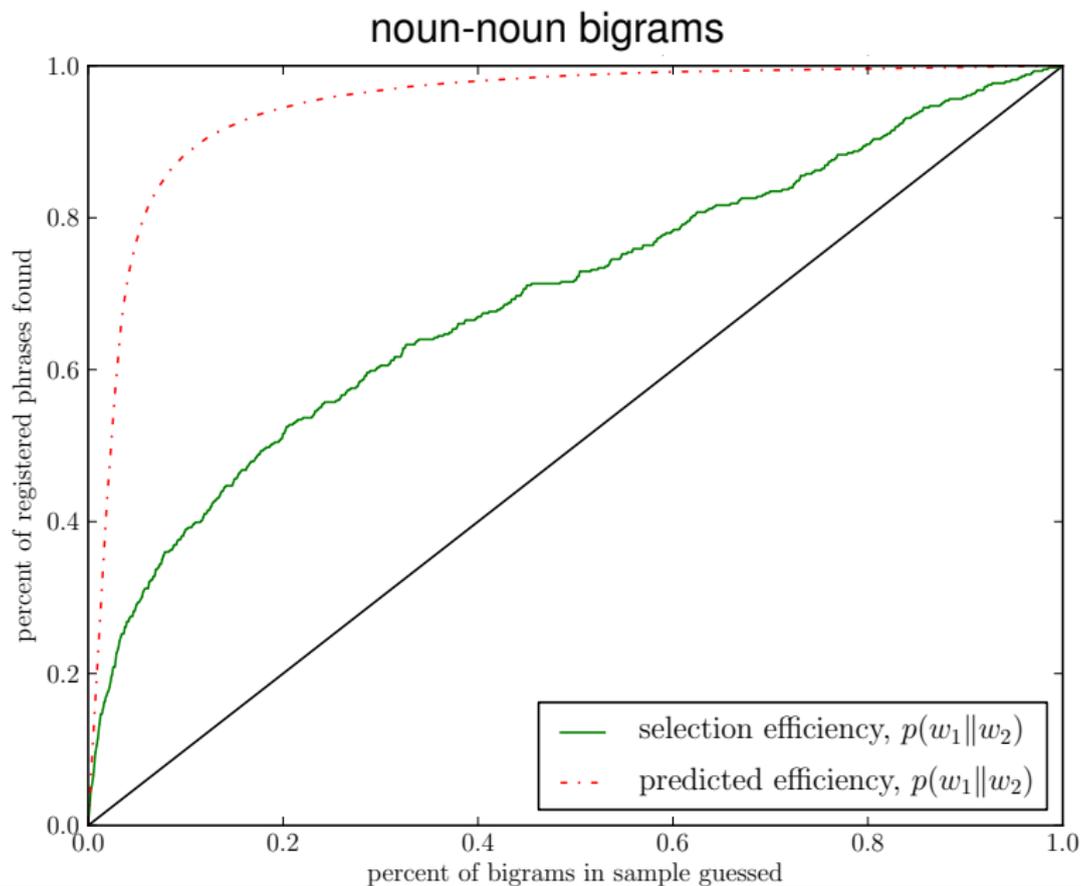
Which factors predict a phrase's popularity?



Which factors predict a phrase's popularity?

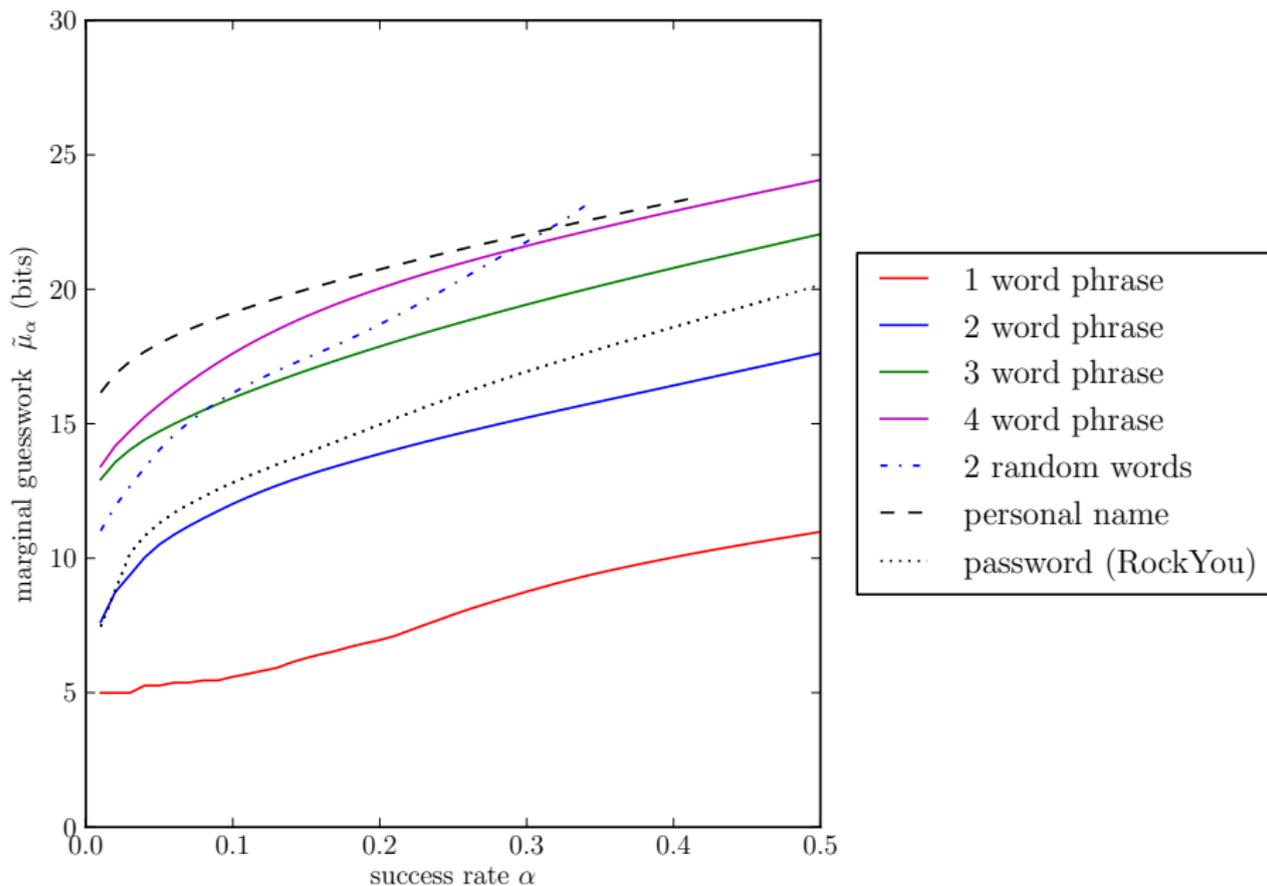


Which factors predict a phrase's popularity?



Are natural language phrases difficult to guess?

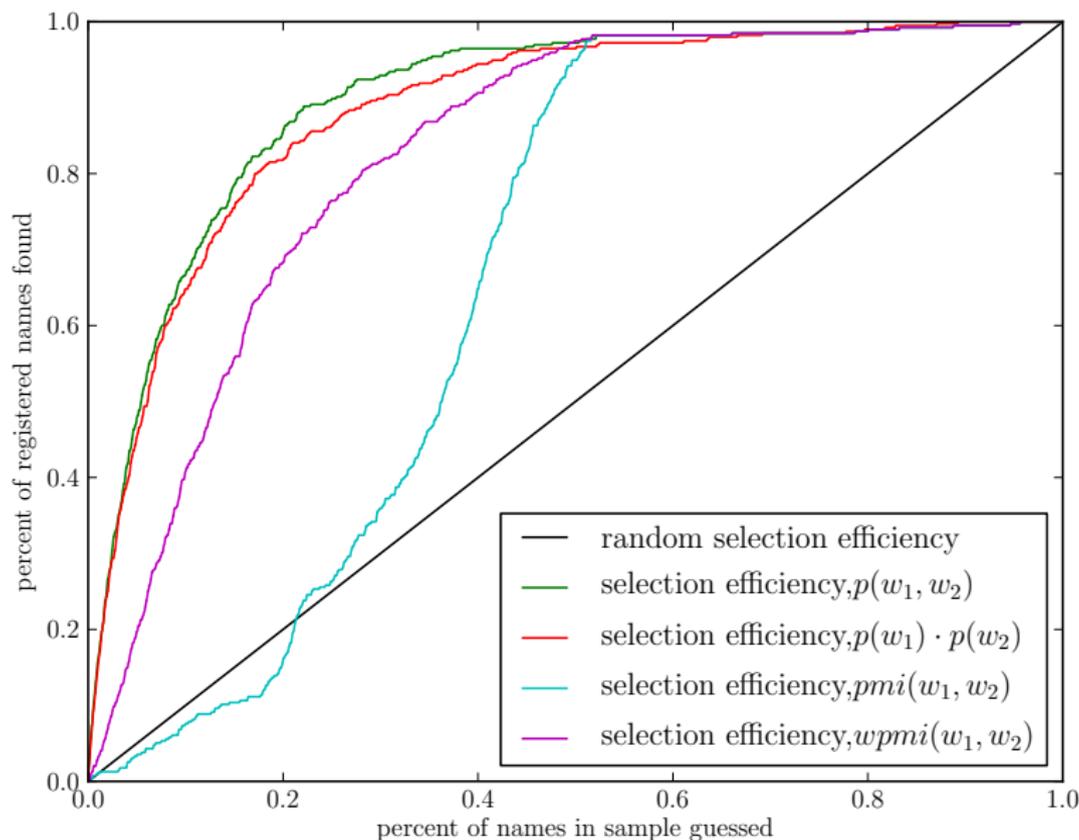
Are natural language phrases difficult to guess?



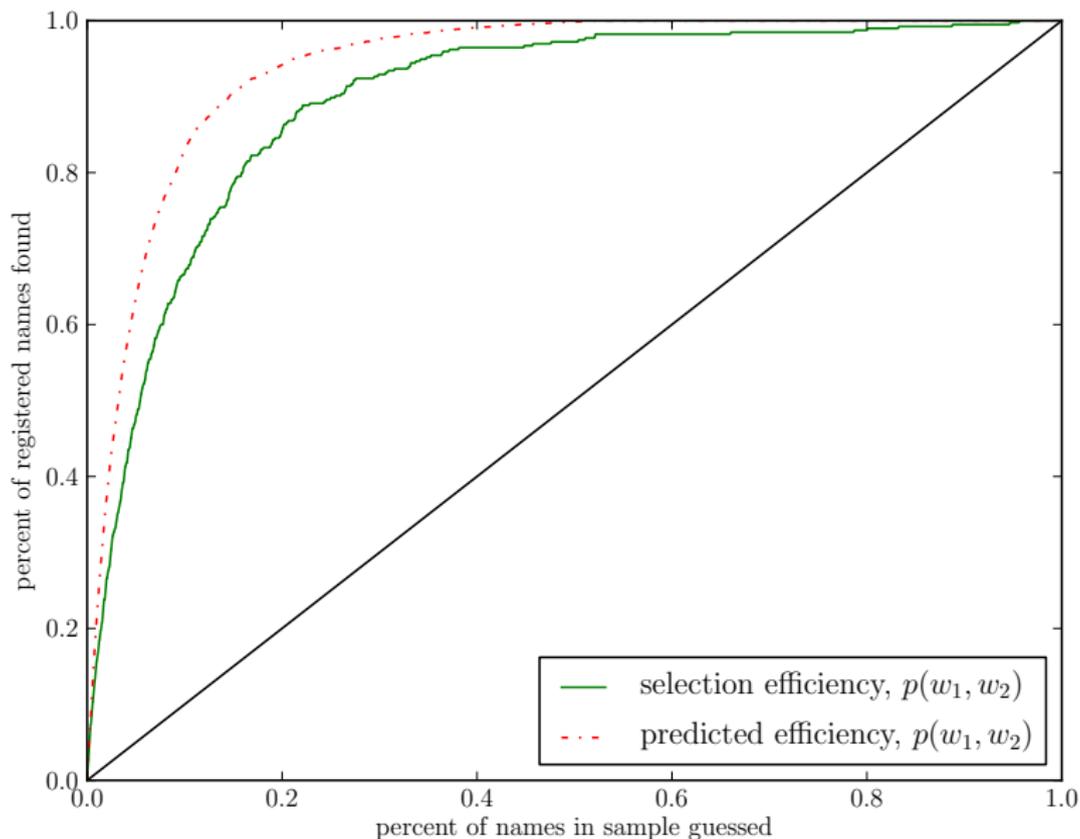
Thank you

jcb82@cl.cam.ac.uk

Similar results for names



Similar results for names



Estimating the probability of each class of phrase

Assumptions:

- N total phrases
- n phrases in this class of equal probability
- k were selected

Estimating the probability of each class of phrase

Assumptions:

- N total phrases
- n phrases in this class of equal probability
- k were selected

solve as a weighted coupon collector's problem:

$$\hat{p} = \frac{\prod_{j=1}^k \frac{n}{n-j}}{N \cdot n}$$