Linguistic properties of multi-word passphrases

Joseph Bonneau, Ekaterina Shutova

jcb82,es407@cl.cam.ac.uk

UNIVERSITY OF CAMBRIDGE

Computer Laboratory

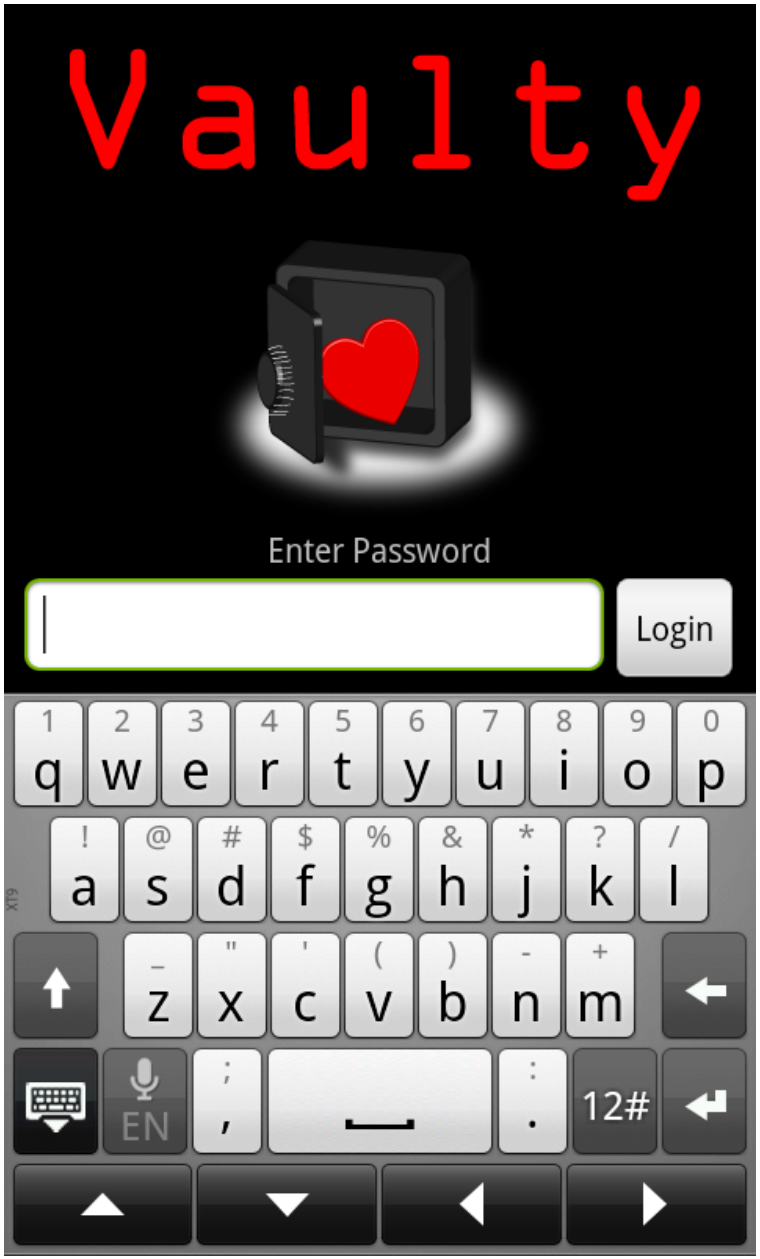
USEC Workshop on Usable Security 2012

Kralendijk, Bonaire, Netherlands

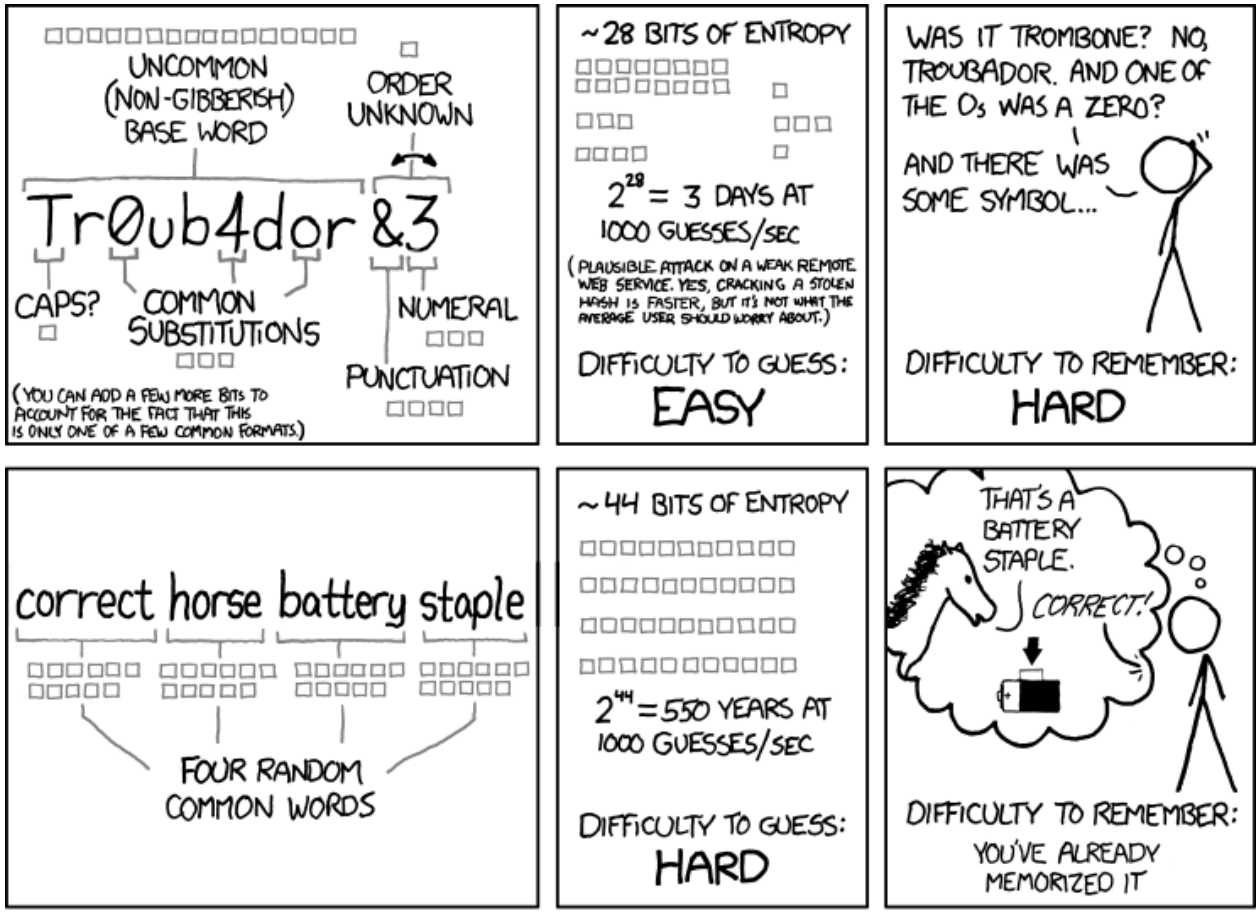
March 2, 2012

Passphrases an increasingly attractive approach

Passphrases an increasingly attractive approach



Passphrases an increasingly attractive approach

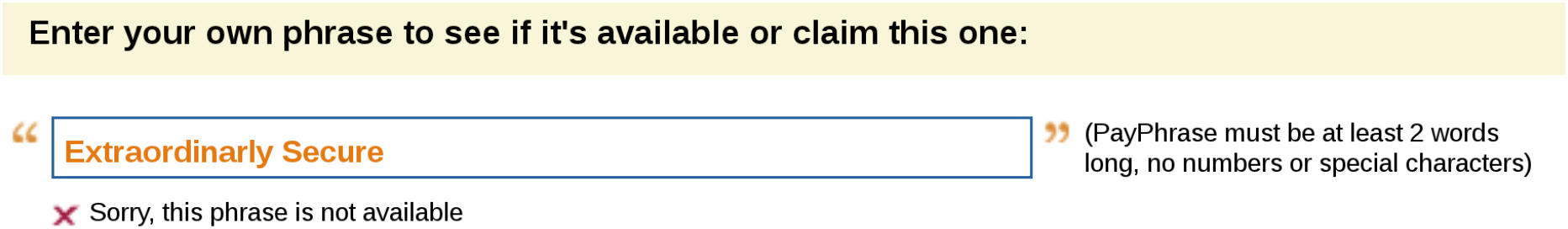


xkcd #936

What do we know about passphrase guessing?

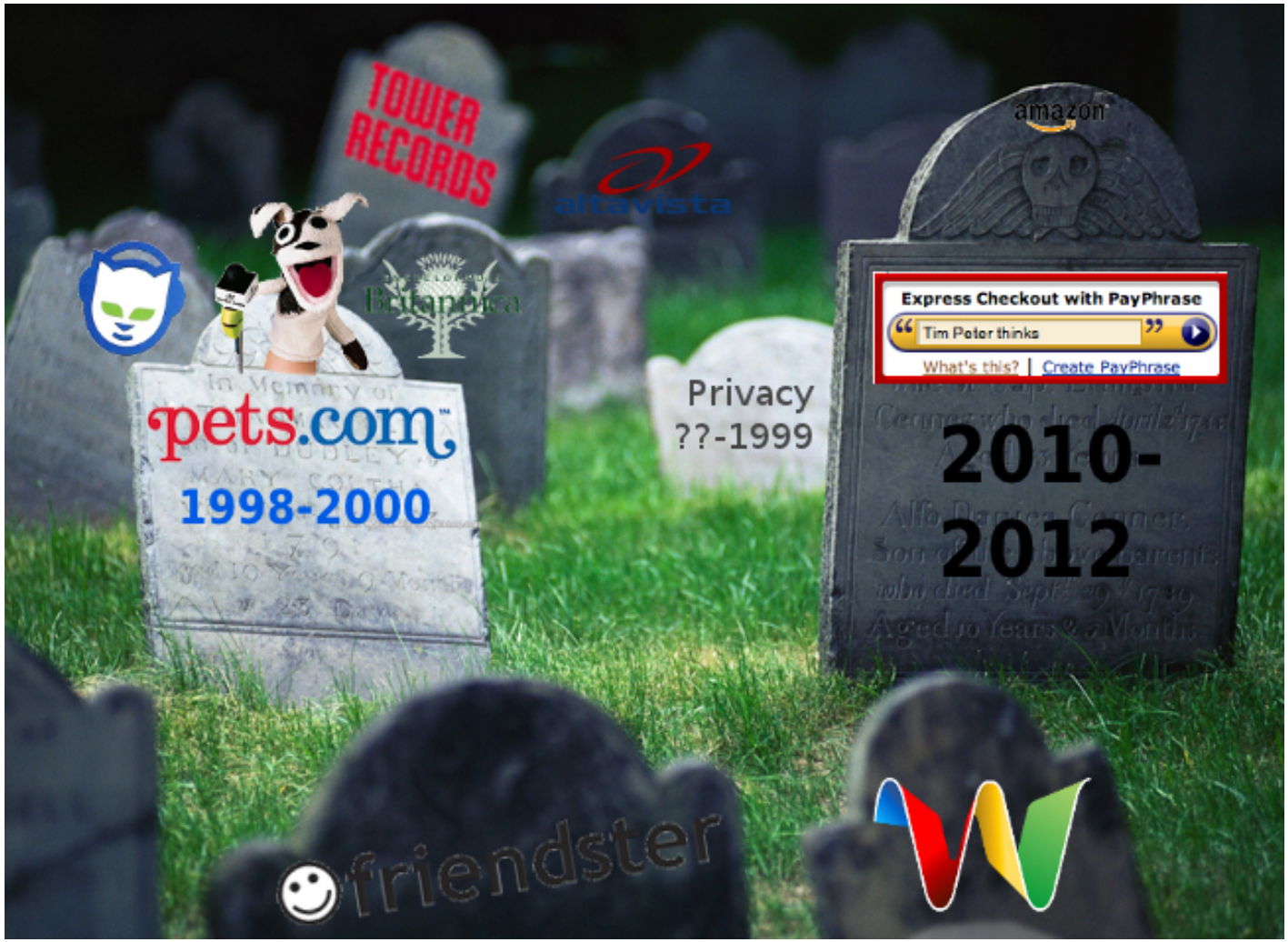
[this space intentionally left blank]

Data source: Amazon PayPhrases



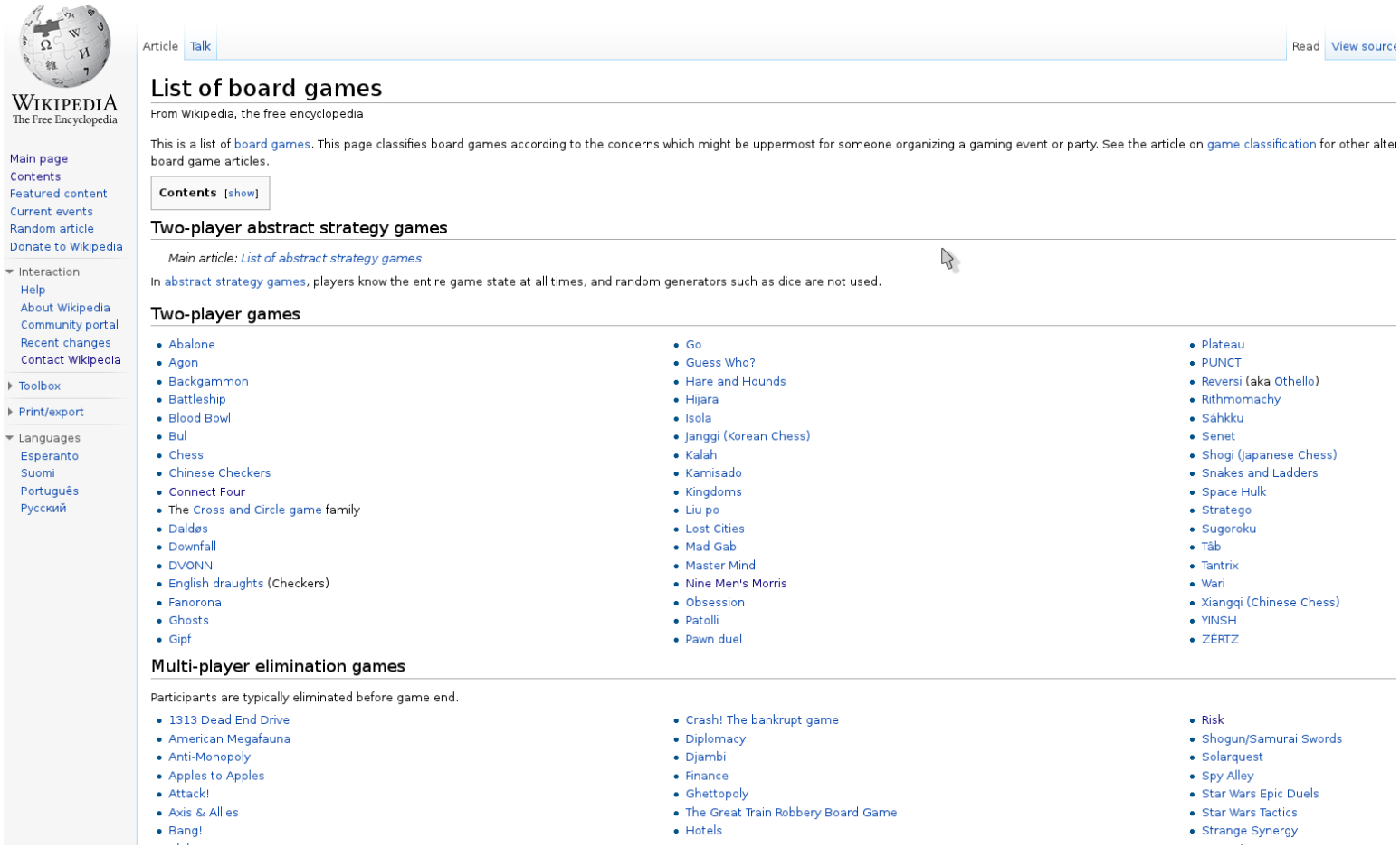
- must be at least two words
- must be globally unique
  - security $\leftarrow$ PIN + passphrase
- can only contain the letters a-z, A-Z, SPACE
  - capitalisation and spacing ignored

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 alter 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
- Agon
- Backgammon
- Battleship
- Blood Bowl
- Bul
- Chess
- Chinese Checkers
- Connect Four
- The Cross and Circle game family
- Dalđøs
- Downfall
- DVONN
- English draughts (Checkers)
- Fanorona
- Ghosts
- Gipf
- Go
- Guess Who?
- Hare and Hounds
- Hijara
- Isola
- Janggi (Korean Chess)
- Kalah
- Kamisado
- Kingdoms
- Liu po
- Lost Cities
- Mad Gab
- Master Mind
- Nine Men's Morris
- Obsession
- Patolli
- Pawn duel
- Plateau
- PÜNCT
- Reversi (aka Othello)
- Rithmomachy
- Sáhkku
- Senet
- Shogi (Japanese Chess)
- Snakes and Ladders
- Space Hulk
- Stratego
- Sugoroku
- Tâb
- Tantrix
- Wari
- Xiangqi (Chinese Chess)
- YINSH
- ZÈRTZ

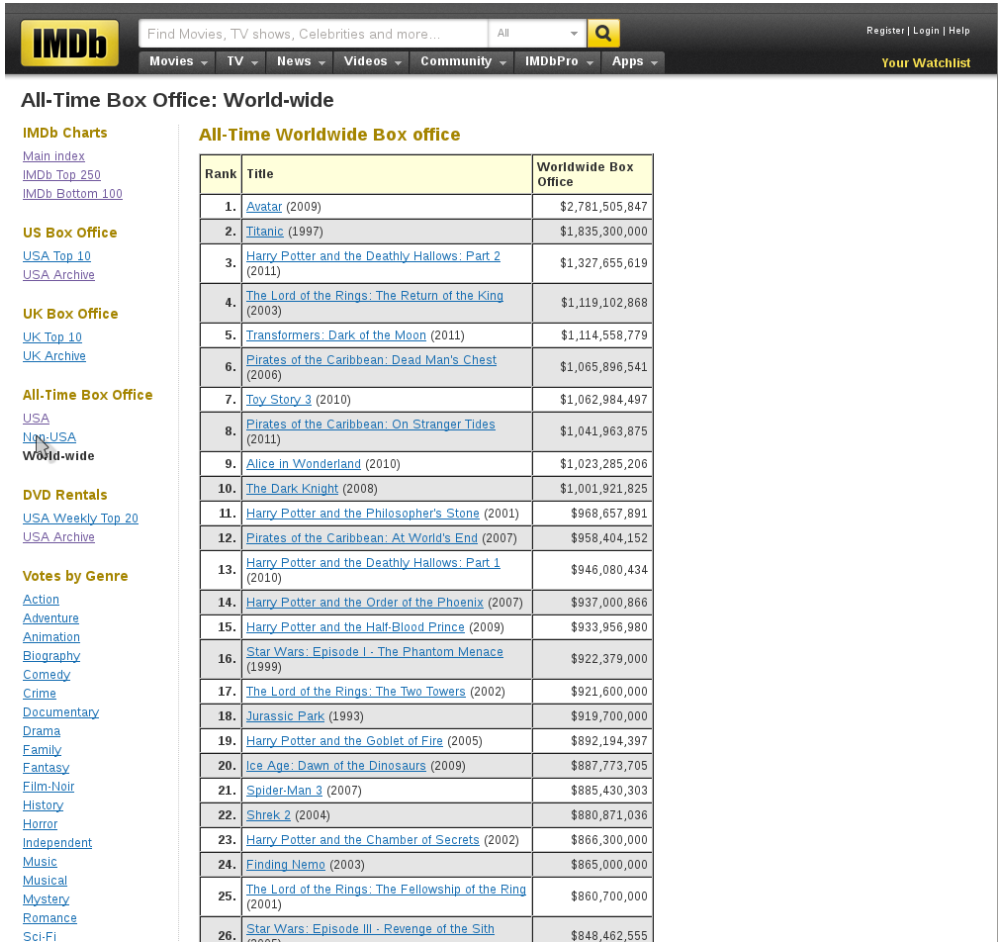
Multi-player elimination games

Participants are typically eliminated before game end.

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

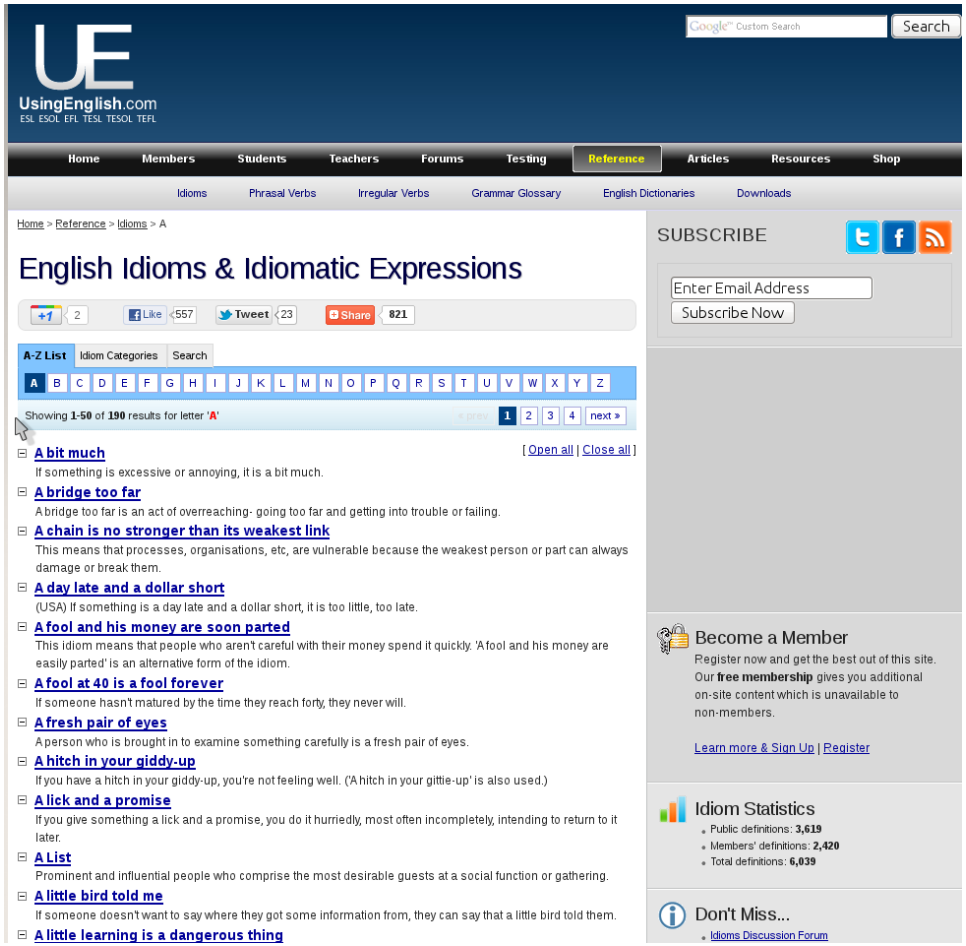
- proper nouns
- titles
- idiomatic phrases
- slang

A simple dictionary attack



- proper nouns
- titles
- idiomatic phrases
- slang

A simple dictionary attack



- proper nouns
- titles
- idiomatic phrases
- slang

A simple dictionary attack

- proper nouns
- titles
- idiomatic phrases
- slang

Results

word list	example	list size	success rate	$\hat{p}$
	arts			
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%
total		7679	38.5%	0.4159%
	sports teams			
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%
total		368	71.7%	0.0542%
	sports venues			
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%
total		687	12.7%	0.0094%

Results

word list	example	list size	success rate	$\hat{p}$
	games			
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%
total		921	28.2%	0.0306%
	comics			
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%
total		1767	32.1%	0.0701%
	place names			
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%
total		4059	43.7%	0.2677%
	phrases			
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%
total		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%

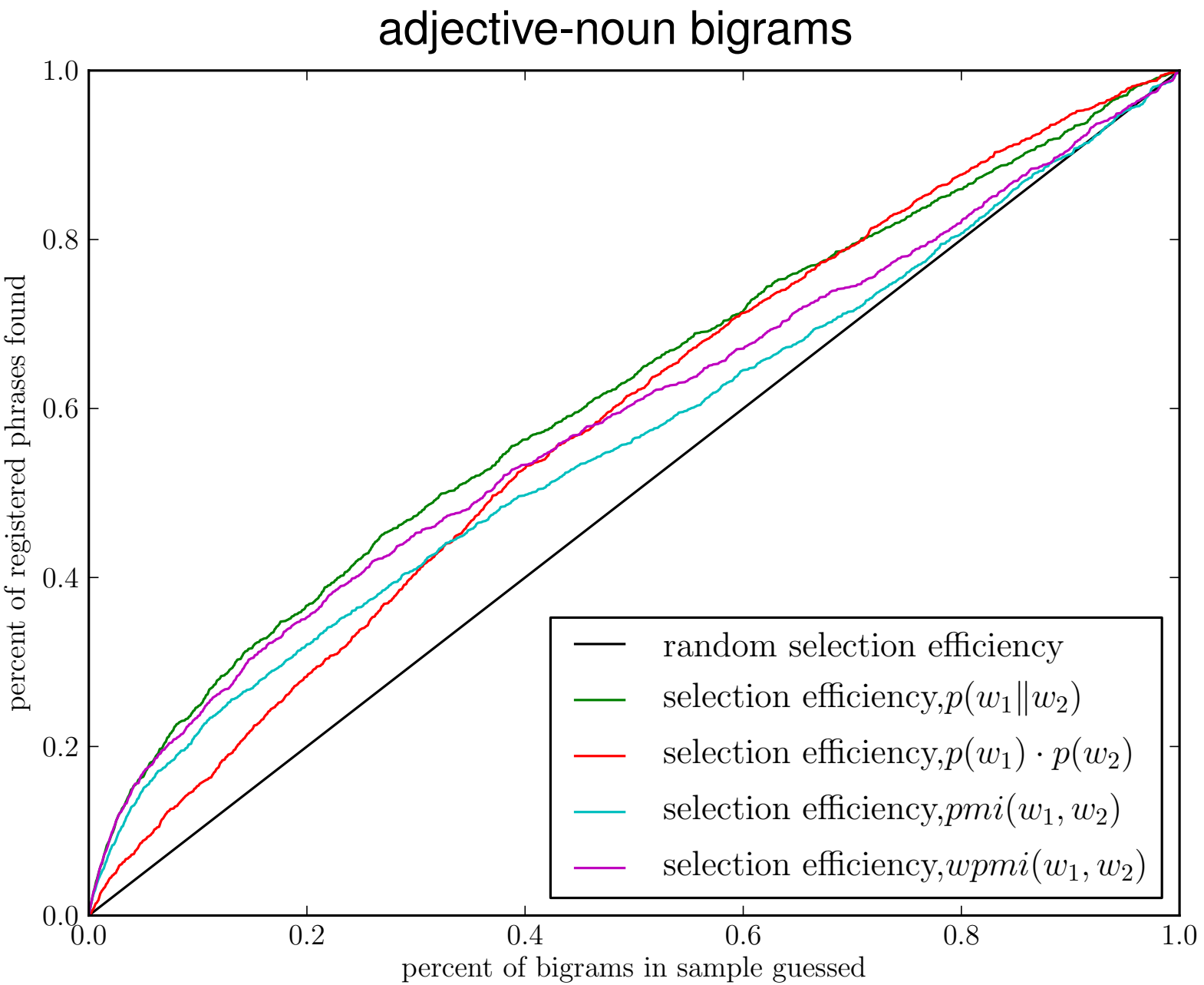
Phrases generated from Google n-gram corpus

Which factors predict a phrase's popularity?

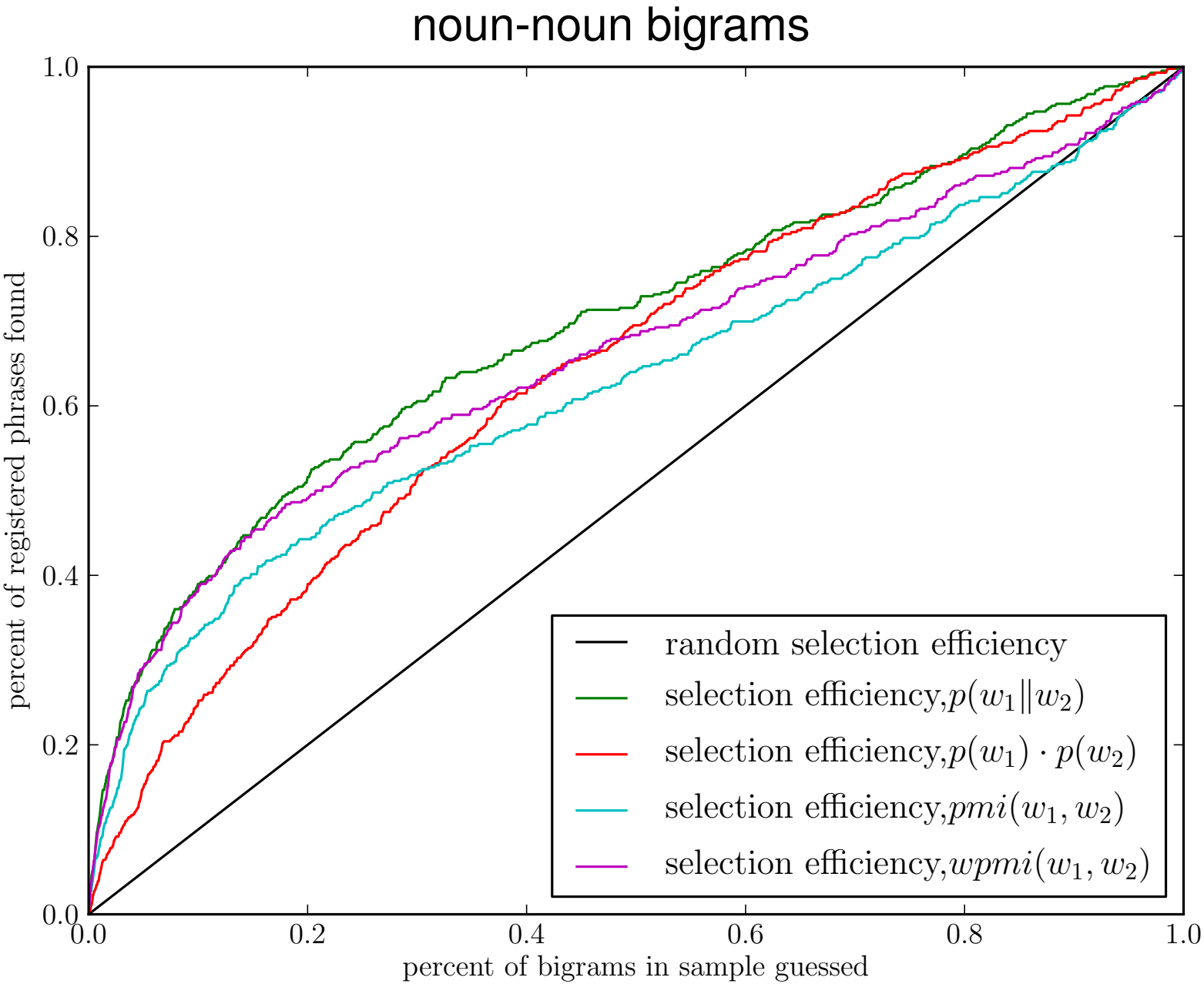
Possible selection models:

baseline	random
natural-language production	$p(w_1 \mid\mid w_2)$
independent word selection	$p(w_1) \cdot p(w_2)$
mutual information	$pmi(w_1, w_2) = \lg \frac{p(w_1 \mid\mid w_2)}{p(w_1) \cdot p(w_2)}$
blended model	$p(w_1 \mid\mid w_2) \cdot pmi(w_1, w_2)$

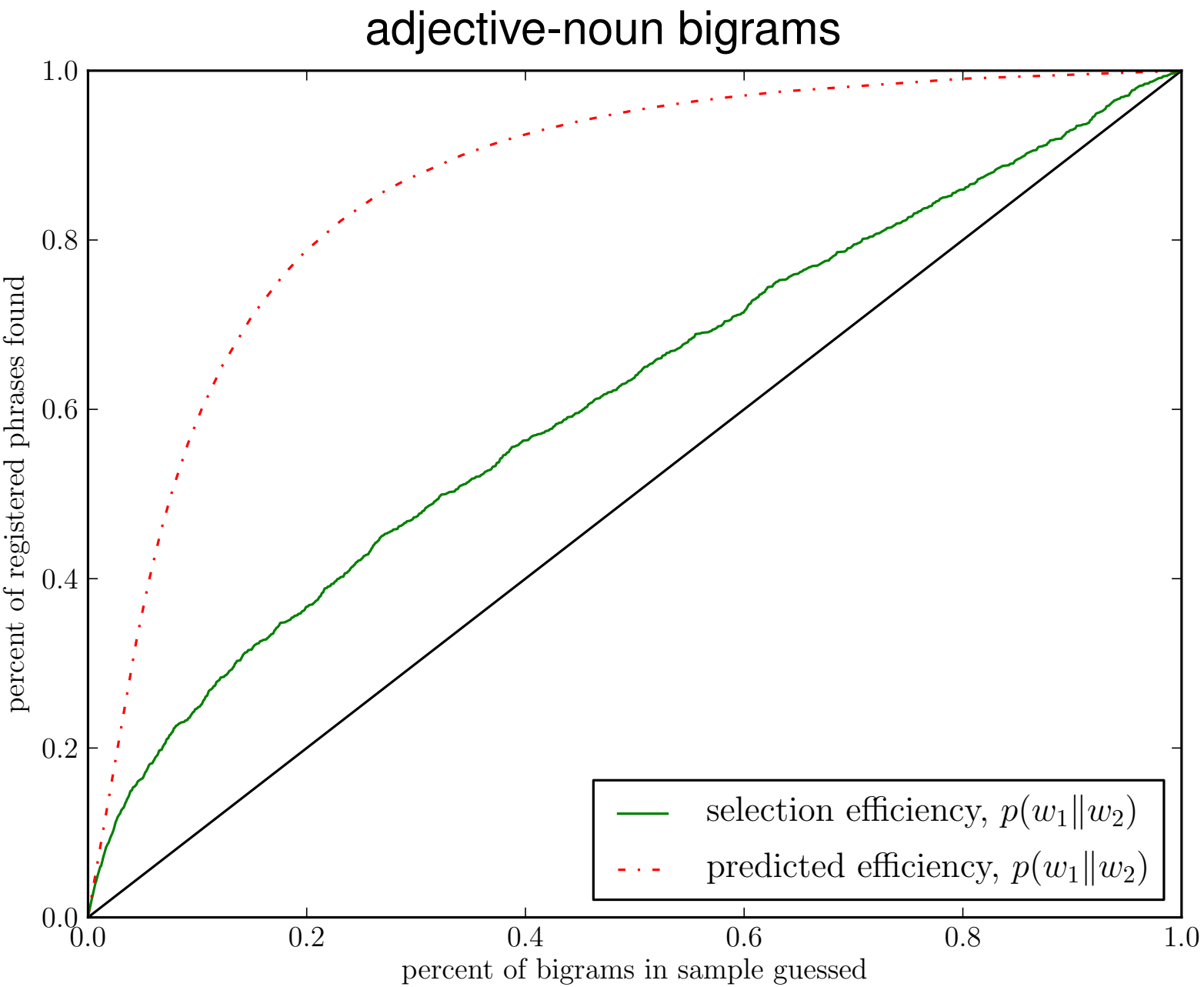
Which factors predict a phrase's popularity?



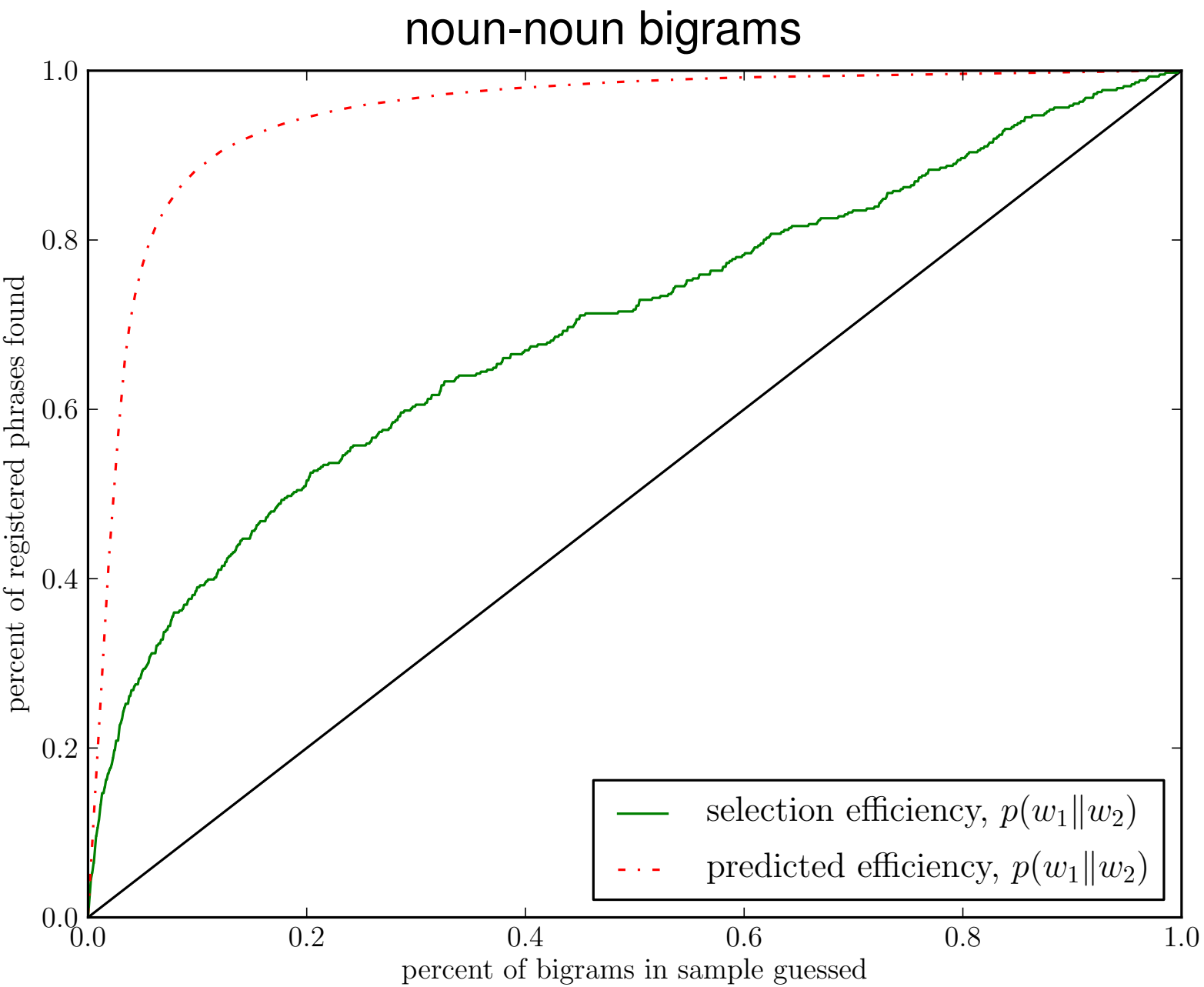
Which factors predict a phrase's popularity?



Which factors predict a phrase's popularity?

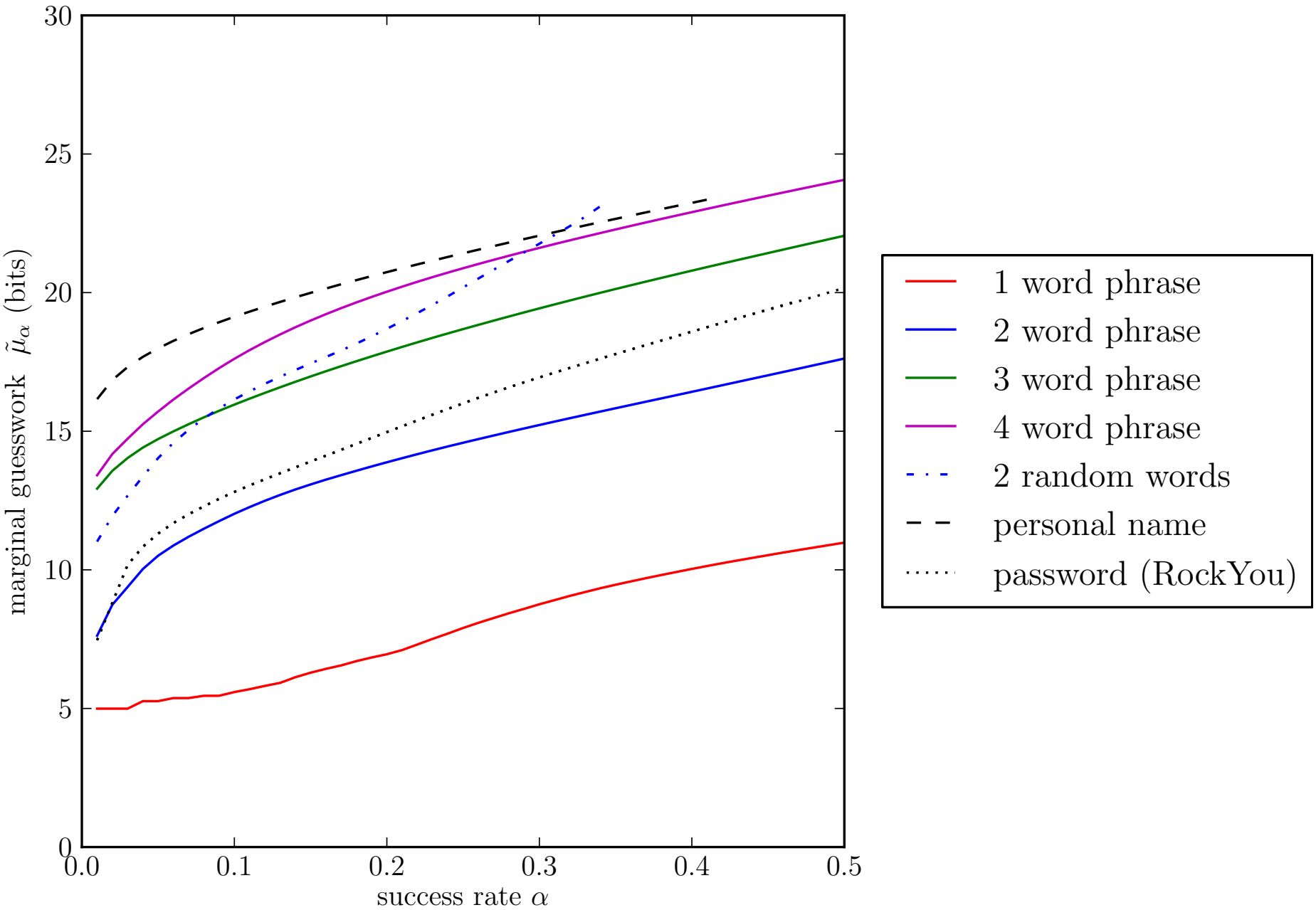


Which factors predict a phrase's popularity?



Are natural language phrases difficult to guess?

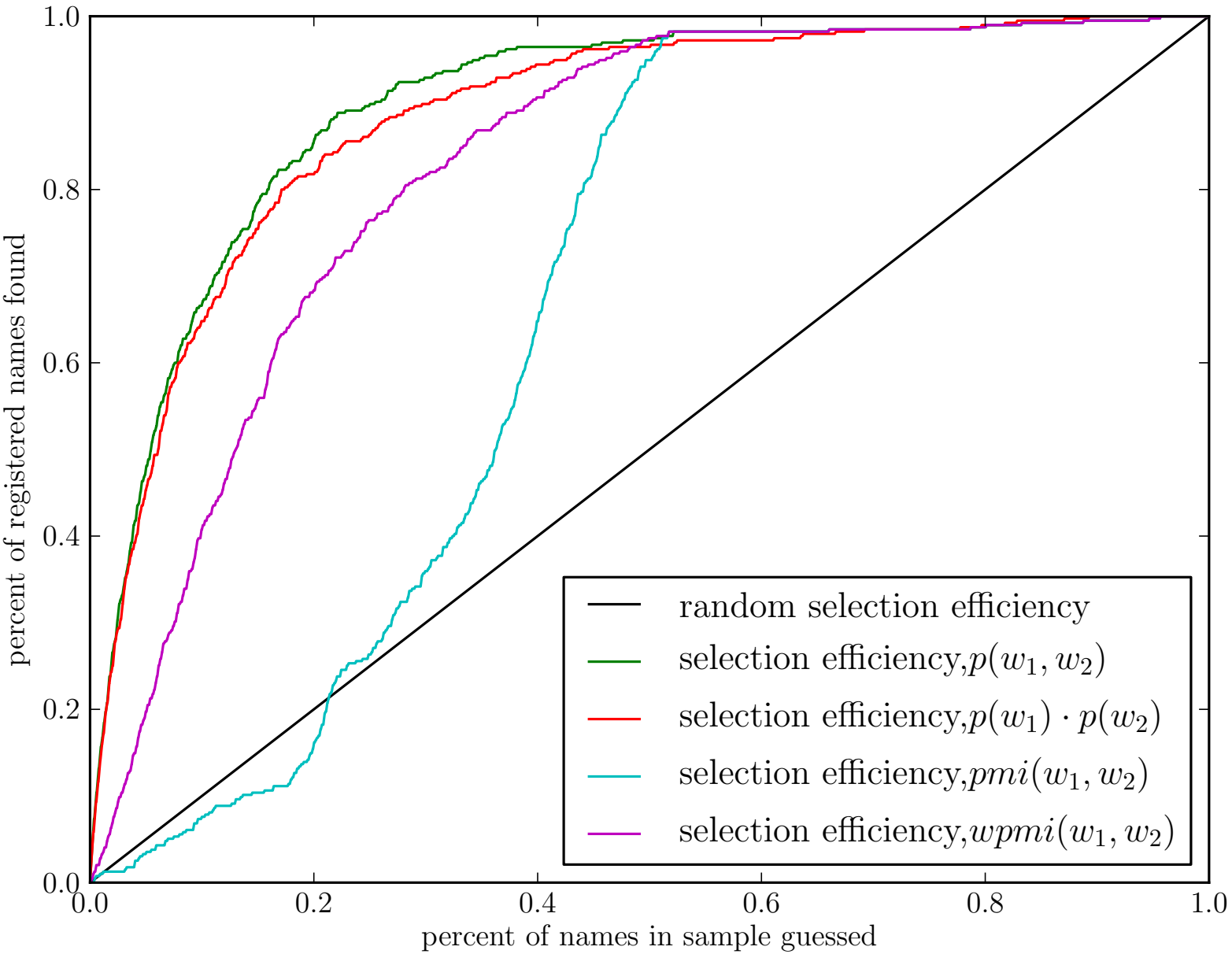
Are natural language phrases difficult to guess?



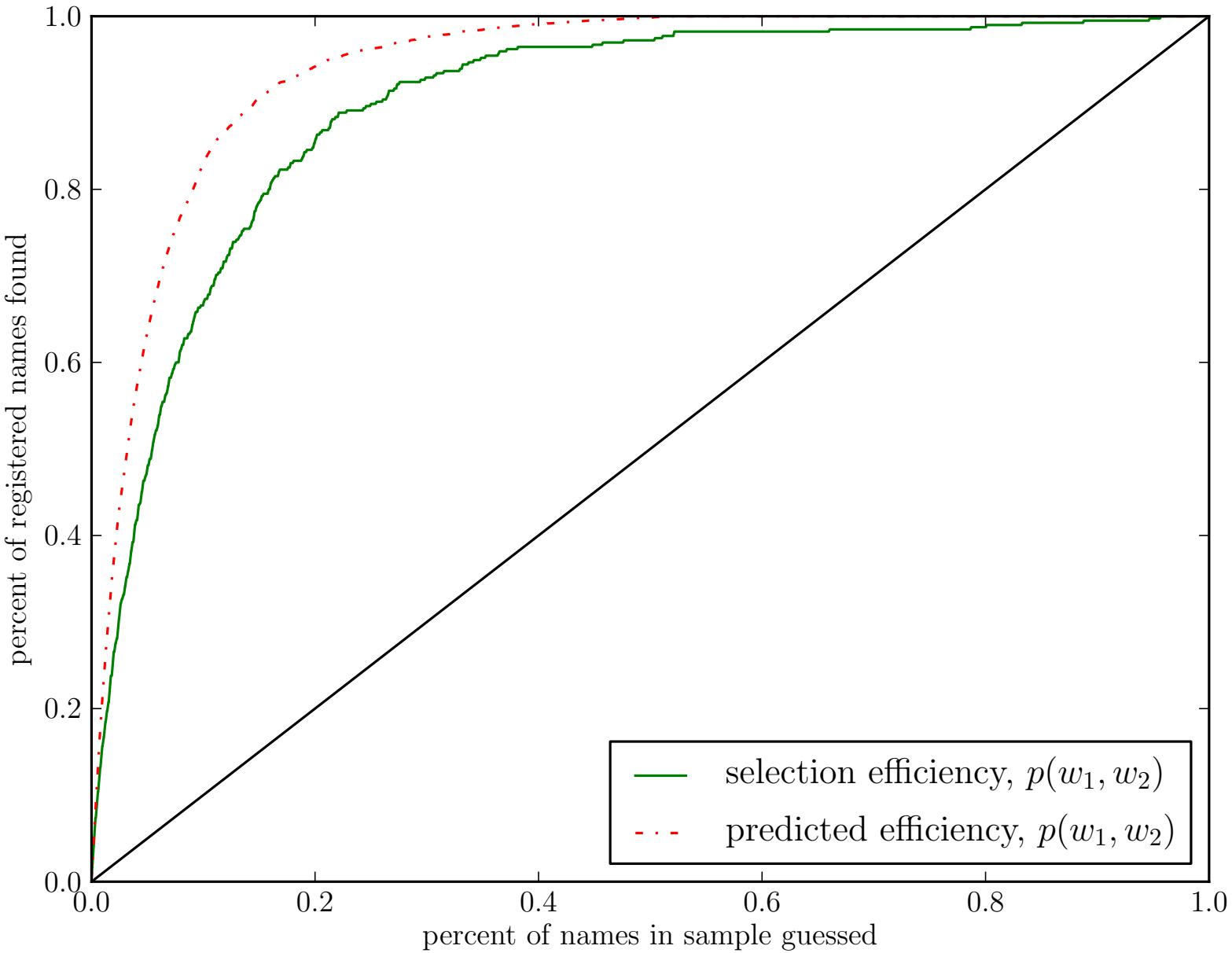
Thank you

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:

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