Inquisitive archduchess wrestles comparatively apologetic pelicans:

Improving security and usability of passphrases with guided word choice

Nikola K. Blanchard [1], Clément Malaingre [2], Ted Selker[3]

[1]IRIF, Université Paris Diderot

[2]Teads France

[3]University of California, Berkeley

ACSAC 34

December 7th, 2018

Why talk about passphrases ?

Current methods to make passphrase

First possibility: let people choose them

Problems:

- Sentences from literature (songs/poems)
- Famous sentences (2.55% of users chose the same sentence in a large experiment)
- Low entropy sentences with common words

Current methods to make passphrase

First possibility: let people choose them

Problems:

- Sentences from literature (songs/poems)
- Famous sentences (2.55% of users chose the same sentence in a large experiment)
- Low entropy sentences with common words

Second possibility: random generation

Limits :

- Small dictionary if we want to make sure people know all words
- Harder to memorise

What if we take the best of both world ?

Passphrase choice experiment

We show 20 or 100 words to users, they have to pick – and remember – six.

Questions :

- What factors influence their choices ?
- What is the effect on entropy ?
- What are the most frequent mistakes ?
- How is memorisation affected ?

Initial hypotheses

We are principally looking for three effects:

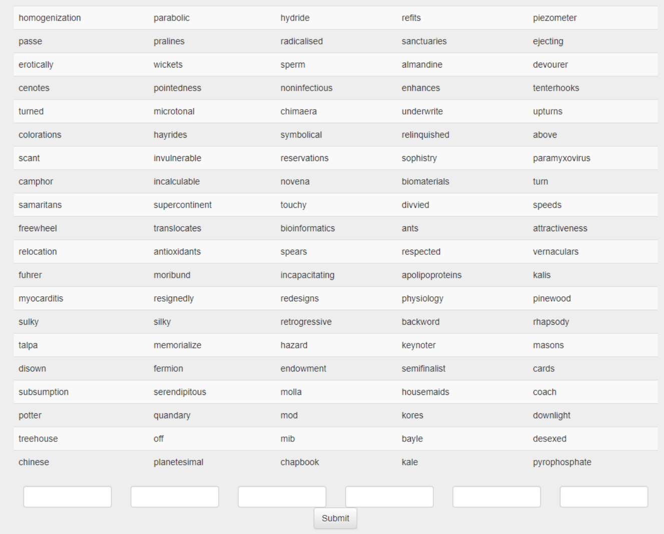
- Positional effects: choose words in certain places
- Semantic effects: choose familiar words
- Syntactic effects: create sentences/meaning

Protocol

Simple protocol :

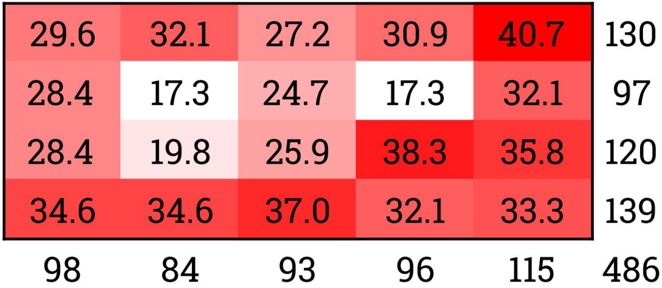
- Show a list of 20/100 random words from a large dictionary
- Ask to choose and write down 6 words (imposed on the control group)
- Show them the sentence and ask them to memorise, with little exercise to help them.
- Distractor task: show them someone else's word list and ask to guess the word choice
- Ask them to write the initial sentence

Interface



homogenization	parabolic	hydride	refits	piezometer
passe	pralines	radicalised	sanctuaries	ejecting
erotically	wickets	sperm	almandine	devourer
cenotes	pointedness	noninfectious	enhances	tenterhooks
turned	microtonal	chimaera	underwrite	upturns
colorations	hayrides	symbolical	relinquished	above
scant	invulnerable	reservations	sophistry	paramyxovirus
camphor	incalculable	novena	biomaterials	turn
samaritans	supercontinent	touchy	divvied	speeds
freewheel	translocates	bioinformatics	ants	attractiveness
relocation	antioxidants	spears	respected	vernaculars
fuhrer	moribund	incapacitating	apolipoproteins	kalis
myocarditis	resignedly	redesigns	physiology	pinewood
sulky	silky	retrogressive	backword	rhapsody
talpa	memorialize	hazard	keynoter	masons
disown	fermion	endowment	semifinalist	cards
subsumption	serendipitous	molla	housemaids	coach
potter	quandary	mod	kores	downlight
treehouse	off	mib	bayle	desexed
chinese	planetesimal	chapbook	kale	pyrophosphate

Positional bias



29.6	32.1	27.2	30.9	40.7	130
28.4	17.3	24.7	17.3	32.1	97
28.4	19.8	25.9	38.3	35.8	120
34.6	34.6	37.0	32.1	33.3	139
98	84	93	96	115	486

Positional bias

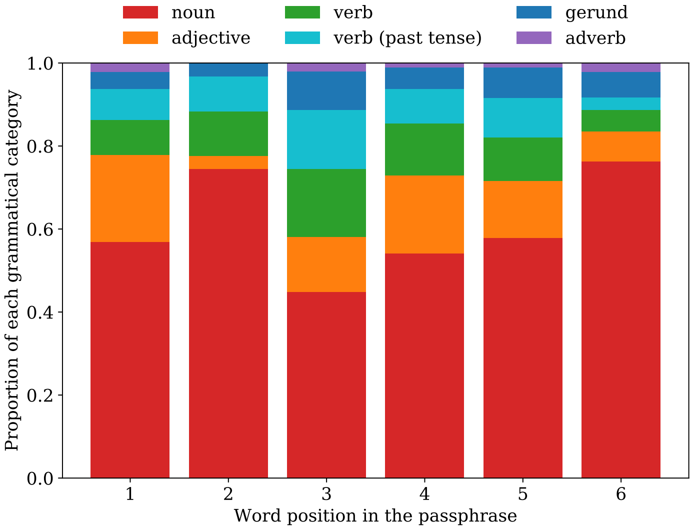
5.6	5.6	10.1	2.3	11.3	31
1.1	3.4	1.1	3.4	10.1	17
6.8	5.6	1.1	1.1	1.1	14
6.8	5.6	9.0	9.0	4.5	31
1.1	6.8	3.4	6.8	4.5	20
9.0	11.3	6.8	5.6	5.6	34
6.8	2.3	1.1	9.0	4.5	21
9.0	6.8	10.1	3.4	5.6	31
10.1	7.9	10.1	3.4	5.6	33
6.8	9.0	10.1	3.4	3.4	29
5.6	9.0	10.1	3.4	5.6	30
9.0	4.5	9.0	4.5	4.5	28
7.9	5.6	9.0	2.3	3.4	25
3.4	5.6	6.8	5.6	2.3	21
6.8	0.0	5.6	7.9	4.5	22
6.8	4.5	3.4	2.3	5.6	20
2.3	6.8	2.3	7.9	10.1	26
4.5	6.8	3.4	10.1	9.0	30
6.8	5.6	7.9	10.1	7.9	34
9.0	7.9	6.8	9.0	7.9	36
111	107	113	98	104	533

Syntactic bias

Syntactic effects :

- Average frequency ($<$ 50%) of meaningful sentences
- 65 different syntactic structures for 99 sentences
- Single frequent structure: six nouns in a row

Syntactic bias

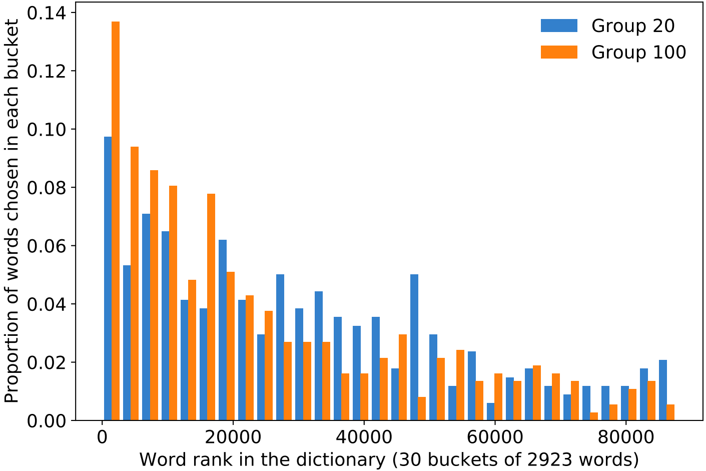


Syntactic bias

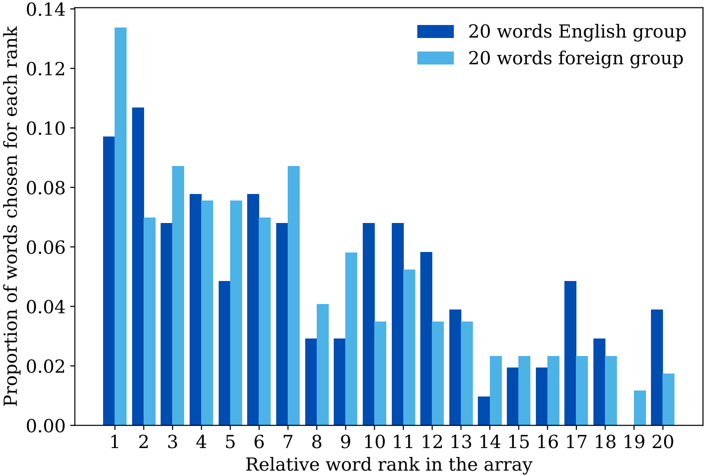
Passphrase examples :

- Monotone customers circling submerging canteen pumpkins
- Furry grills minidesk newsdesk deletes internet
- Here telnet requests unemotional globalizing joinery
- Brunette statisticians asked patriarch endorses dowry
- Marginal thinker depressing kitty carcass sonatina

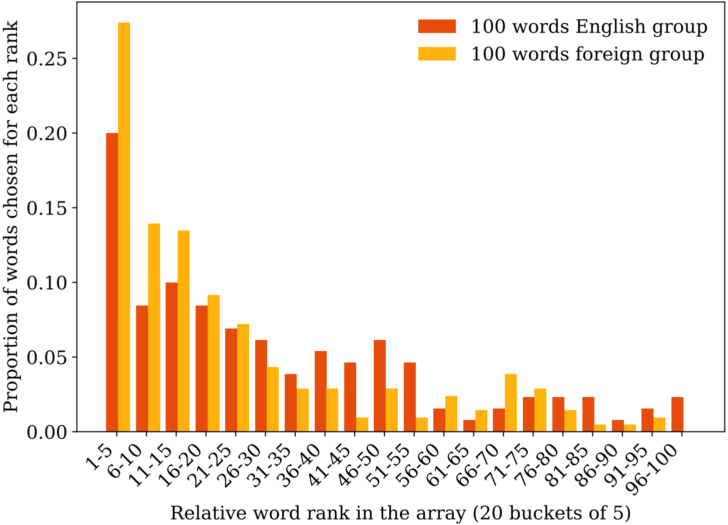
Semantic bias



Semantic bias



Semantic bias



Choosing models

Three main models to analyse user's choice

Uniform : every word with equal probability

Smallest : Take the six most frequent words from the list shown

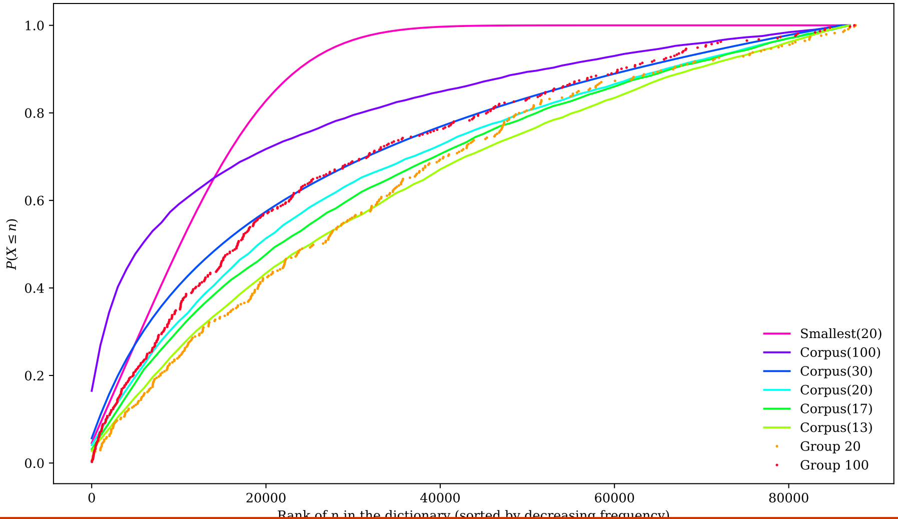
Corpus : every word taken with probability proportional to its use in natural language. The word of rank r_k is taken with probability :

$$\frac{\frac{1}{r_k}}{\sum_{i=1}^{n} \frac{1}{r_i}}$$

Entropy comparison

Strategy	Entropy (bits)	Strategy	Entropy
Uniform(87,691)	16.42	*Smallest*(20)	12.55
Corpus(13)	16.25	*Uniform*(5,000)	12.29
Corpus(17)	16.15	*Uniform*(2,000)	10.97
Corpus(20)	16.10	*Smallest*(100)	10.69
Corpus(30)	15.92	*Corpus*(300,000)	8.94
Corpus(100)	15.32	*Corpus*(87,691)	8.20
Uniform(10,000)	13.29		

Entropy curves



Error comparison

Section	Correct	Typo	Variant	Order	Miss	Wrong
1:20	19/47	6	8	6	26	5
1:100	26/51	10	5	3	16	4
Control	6/26	11	11	10	31	12
2:20	14/29	1	2	8	0	3
2:100	15/26	4	2	3	1	4

Conclusion

Passphrase choice method

Advantage with 100-word list:

- Secure: 97% of maximal entropy, 30% increase over uniform with limited dictionary
- Memorable: error rate divided by 4
- Lightweight: <1MB tool, can and should be used inside a browser

Passphrase choice method

Advantage with 100-word list:

- Secure: 97% of maximal entropy, 30% increase over uniform with limited dictionary
- Memorable: error rate divided by 4
- Lightweight: <1MB tool, can and should be used inside a browser

Limitations:

- Requires more testing for long-term memory
- Depends on the user's will

Questions

Questions:

- What is the optimal number of words to show ?
- Is it interesting to take even bigger dictionaries ?
- Can this method be applied to languages with small vocabularies (Esperanto)
- What is the best way to model user choice ?