wordcereal

an english grammar-aware passphrase word list

nathan bloomfield

grand meetup 2017 – whistler

disclosure

prior art

- http://copleytech.com/pass-phrases/ (powered by WP.org!)
- http://philzimmermann.com/docs/PGP_word_list.pdf
- https://www.eff.org/deeplinks/2016/07/new-wordlists-random-passphrases

let’s talk about passphrases

let’s talk about passphrases

correct horse battery staple

let’s talk about passphrases

correct horse battery staple

memorable > short

let’s talk about passphrases

correct horse battery staple

memorable > short

grammar?

let's talk about passphrases

scary pop-press articles say grammar is bad, citing this paper:

https://www.cs.cmu.edu/~agrao/paper/Effect_of_Grammar_on_Security_of_Long_Passwords.pdf

let's talk about passphrases

scary pop-press articles say grammar is bad, citing this paper:

https://www.cs.cmu.edu/~agrao/paper/Effect_of_Grammar_on_Security_of_Long_Passwords.pdf

let's talk about passphrases

any reversible map from bits to sentences
will losslessly encode entropy

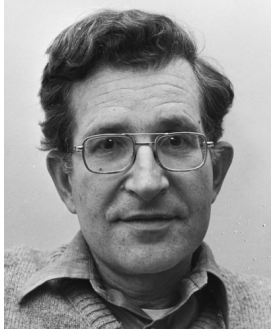
let's talk about passphrases

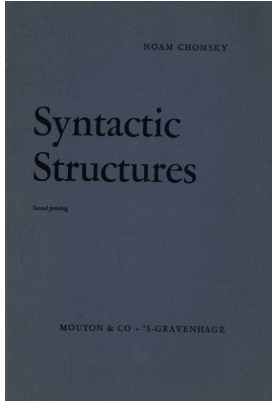
any reversible map from bits to sentences
will losslessly encode entropy

compromise: may need a *longer* phrase

how to construct an arbitrary sentence

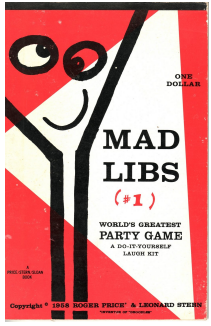
generative grammars (Chomsky, 1957)



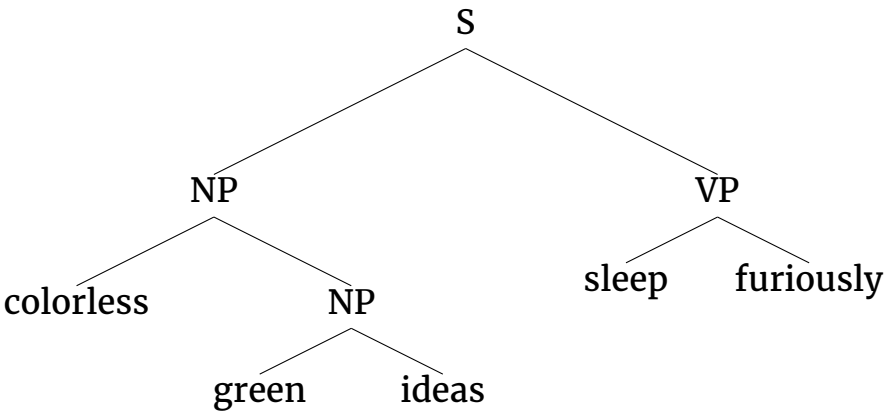


how to construct an arbitrary sentence

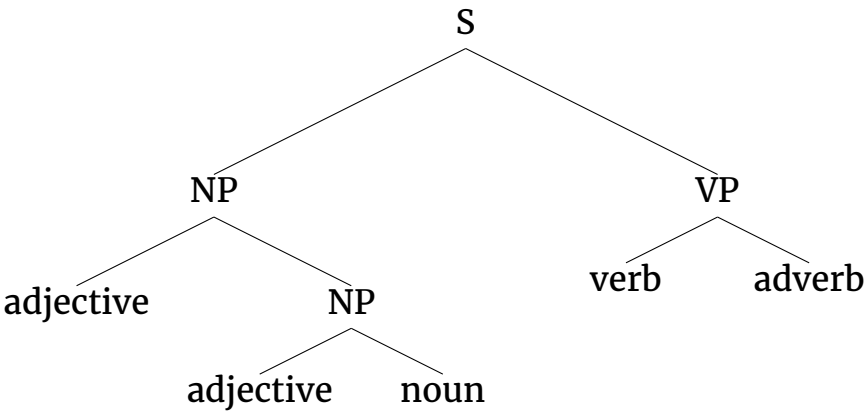
mad libs (Stern and Price, 1953)



how to construct an arbitrary sentence



how to construct an arbitrary sentence



wordcereal

proof of concept

`https://github.com/nbloomf/wordcereal`

wordcereal

proof of concept

`https://github.com/nbloomf/wordcereal`

- 512 each of plural nouns, transitive verbs, adjectives, and adverbs (8 bits/word)
- 8 each of prepositions, determiners, and conjunctions (2 bits/word)

combinatorial properties

- prefix free (and suffix free)
- unique initial 5-grams
- initial 3-grams decode uniquely
- edit distance ≥ 3

semantic properties

- avoid vulgar words
- avoid marginalizing, “-ist” words
- avoid proper nouns and trademarks

phonetic properties

- phonetically prefix free (and suffix free)
- phonetic edit distance ≥ 1 (this is fuzzy)
- avoid words that are homophones
- two lists, divided on syllable count parity (even/odd)

examples

waterfalls deduplicate gossipy channels (32 bits)

those pathways vilify several schematics
throughout few crayons (40 bits)

schematics familiarize two aberrations and
level references lampoon lozenges (60 bits)