

The Bosks Are Thorny, Grim, and Icky

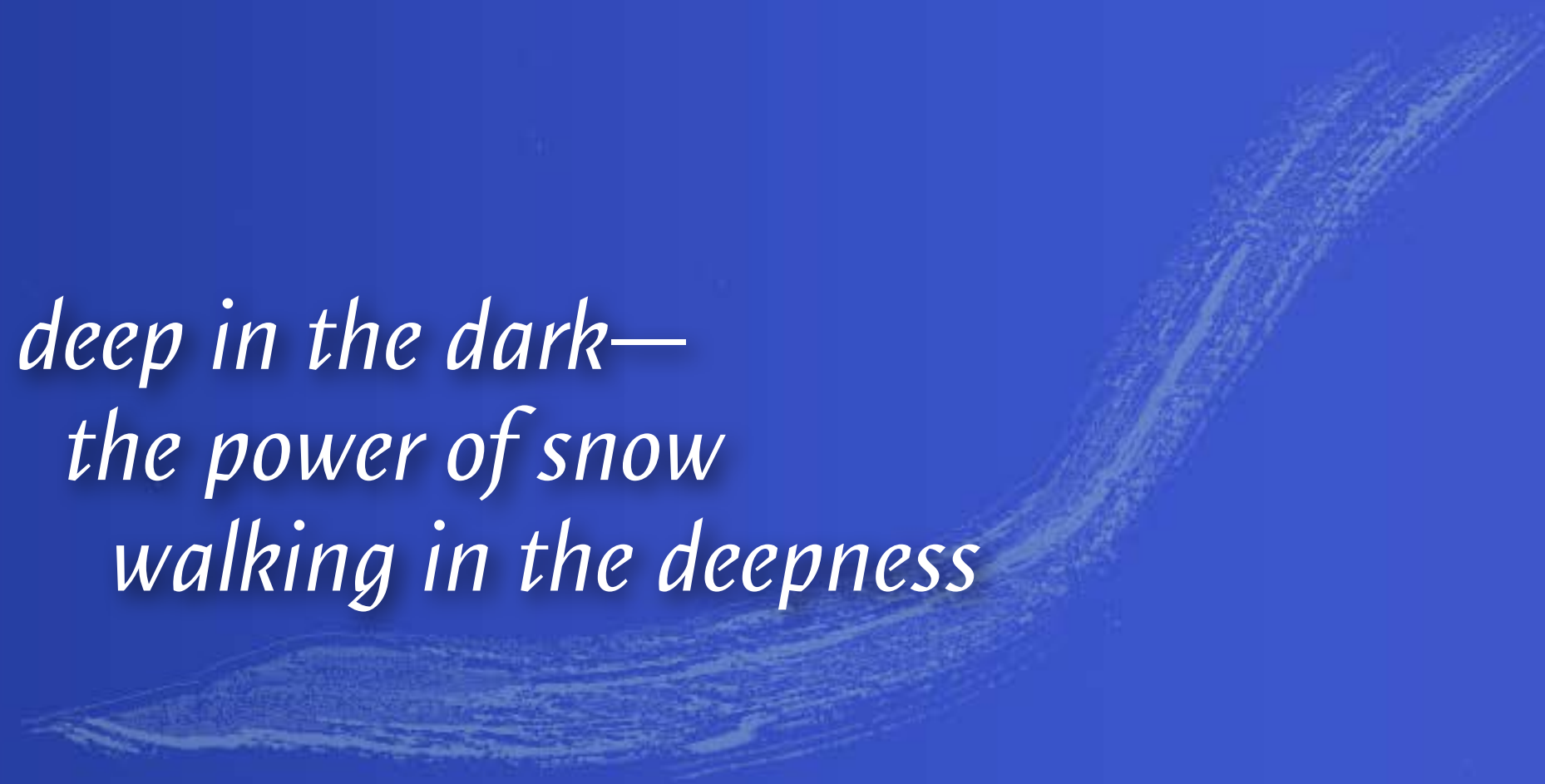
Richard P. Gabriel

poet, writer, computer scientist



Proving Theorems & Seeing Cats

Richard P. Gabriel



*deep in the dark—
the power of snow
walking in the deepness*

—InkWell, December 2014

revise Frost's "Stopping by Woods"

sound like Hemingway; use short words

...

*He gives his harness bells a cloud
To ask if there is some spot.
The only other music is the illusion
Of worn wind and cold mound.*

*The logs are good, glad, and tight.
But I have words to keep
And miles to go before I sleep,
And miles to go before I sleep.*

*a crooked rag day—
by myself
dunking distracted sardines*

*day after day
in the man's can
a man can*

Asked to write about
“blues guitar and loud music”

*tuned adrenalin
my music,
a beat-boogied headful*

*not this fatalist murderousness,
deathwatch,
but your dead subroutine*

*functional function?
the orchid—
even when it's Lisped functionally*

So...

- ◆ I write software as part of doing
- ◆ science. I use software as a machine or
- ◆ instrument to explore how the mind / brain
- ◆ might work, creatively

- ◆ Programming is not software engineering
- ◆ Programming is not algorithm design

- ◆ Software engineering is based on setting objectives and finding stepping stones to achieve them

Software Engineering

- ◆ is for producing something you can describe,
- ◆ either using specifications (the old-fashioned way)
- ◆ or a product backlog (or some other suchlike
- ◆ thing in an agile setting). I don't do anything
- ◆ like that.

So What?

- ◆ writing programs can be weird
- ◆ ~~what it feels like to work on the Manhattan project~~
- ◆ old style AI was not enough; new style is not enough



Broad Agency Announcement

Social Media in Strategic Communication (SMISC)

DARPA-BAA-11-64

July 14, 2011



Defense Advanced Research Projects Agency
3701 North Fairfax Drive
Arlington, VA 22203-1714

The conditions under which our Armed Forces conduct operations are rapidly changing with the spread of blogs, social networking sites, and media-sharing technology (such as YouTube), and further accelerated by the proliferation of mobile technology. Changes to the nature of conflict resulting from the use of social media are likely to be as profound as those resulting from previous communications revolutions. The effective use of social media has the potential to help the Armed Forces better understand the environment in which it operates and to allow more agile use of information in support of operations.

The general goal of the Social Media in Strategic Communication (SMISC) program is to develop a new science of social networks built on an emerging technology base. In particular, SMISC will develop automated and semi-automated operator support tools and techniques for the systematic and methodical use of social media at data scale and in a timely fashion to accomplish four specific program goals:

1. Detect, classify, measure and track the (a) formation, development and spread of ideas and concepts (memes), and (b) purposeful or deceptive messaging and misinformation.
2. Recognize persuasion campaign structures and influence operations across social media sites and communities.
3. Identify participants and intent, and measure effects of persuasion campaigns.
4. Counter messaging of detected adversary influence operations.

Background

Events of strategic as well as tactical importance to our Armed Forces are increasingly taking place in social media space. We must, therefore, be aware of these events as they are



Initial Plan

1. create a template for the text
2. establish a target personality
3. mark some words to possibly change
4. find alternative words
5. measure the personality of the potential text
6. optimize to get as close as possible to the target



Initial Idea

The	woods	are	lovely	dark	and	deep
	alder		attractive	black		abysmal
	ash		beautiful	caliginous		bottomless
	balsa		colorful	dim		heavy
	bamboo		dishy	dusky		intense
	beech		exquisite	gloomy		profound
	birch		fair	subdued		unsounded
	cedar		fine-looking	unlit		wide



Simulated Annealing

...a generic probabilistic metaheuristic for locating a good approximation to the global optimum of a given function in a large, discrete search space....



RHODIE VI

RORC YCF
LR

Personality: Big Five

- ◆ **Openness** to experience
- ◆ **Conscientiousness** vs. not
- ◆ **Extraversion** vs. Introversion
- ◆ **Agreeableness** vs. not
- ◆ Emotional stability vs. **Neuroticism**

Sentiment Analysis

The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods

Yla R. Tausczik¹ and James W. Pennebaker¹

Journal of Language and Social Psychology
29(1) 24–54

© 2010 SAGE Publications
DOI: 10.1177/0261927X09351676
<http://jls.sagepub.com>



Linguistic Inquiry and Word Count



NIH Public Access

Author Manuscript

J Res Pers. Author manuscript; available in PMC 2011 June 1.

Published in final edited form as:

J Res Pers. 2010 June 1; 44(3): 363–373. doi:10.1016/j.jrp.2010.04.001.

Personality in 100,000 Words: A large-scale analysis of personality and word use among bloggers

Tal Yarkoni

University of Colorado at Boulder

Abstract

Previous studies have found systematic associations between personality and individual differences in word use. Such studies have typically focused on broad associations between major personality domains and aggregate word categories, potentially masking more specific associations. Here I report the results of a large-scale analysis of personality and word use in a large sample of blogs (N=694). The size of the dataset enabled pervasive correlations with personality to be identified for a broad range of lexical variables, including both aggregate word categories and individual English words. The results replicated category-level findings from previous offline studies, identified numerous novel associations at both a categorical and single-word level, and underscored the value of complementary approaches to the study of personality and word use.

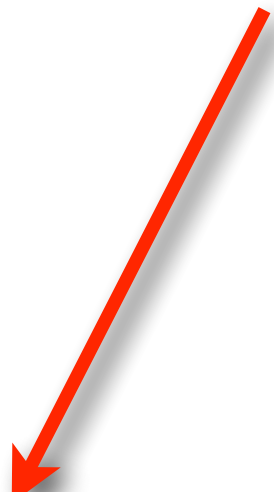
People differ considerably from each other in their habitual patterns of thought, feeling and action. Not surprisingly, these differences are reflected not only in what people think, feel, and do, but also in what they *say* about what they think, feel, or do. Recent studies have identified systematic associations between personality and language use in a variety of different contexts,

“Reading Gabriel’s essays is pure agony.”

Categories	Count	% Words
Sadness [19]	1	0.167
Cognition [20]	1	0.167
Perception [27]	1	0.167
Certainty [26]	1	0.167
Seeing [28]	1	0.167
Affect [12]	1	0.167
Present [39]	1	0.167
Negative Emotion [16]	1	0.167

“Reading Gabriel’s essays is pure agony.”

machine learned


$$Tr = \sum_1^n Y_i P_i$$

where Y_i is the Yarkoni coefficient associated with the i^{th} of n categories and P_i is the percentage of words in the i^{th} category.

Trait	Value
Agreeableness	-0.042
Conscientiousness	-0.163
Extraversion	-
Neuroticism	0.099
Openness	-0.085

“Reading Gabriel’s essays is pure agony.”



Problem #1

- ◆ the messages to tailor are tweet sized
- ◆ our sentiment analysis is based on LIWC 2001
- ◆ LIWC 2001: 2,319 words
- ◆ LIWC 2007: 4,487 words & better vetted

Add LIWC 2007 Words to LIWC 2001

- ◆ but...
 - LIWC 2001: 68 categories
 - LIWC 2007: 64 categories
 - different category names
- ◆ use dictionary overlap to machine learn a numerically weighted mapping
- ◆ use old-AI-style certainty factors to combine weights

Add Words from LIWC 2007 to LIWC 2001

- ◆ 4,878 words total



We Need Synonyms

WordNet

A Lexical Database for English
Princeton University

- ◆ WordNet V2 (repaired)
- ◆ WordNet V3 (expanded)
- ◆ ~160,000 words
- ◆ several synonym types, glosses

ABBREVIATION-FOR NIL

ALSO-SEE NIL

ANTONYM NIL

ATTRIBUTE NIL

BASIC-SYNS (("canis_familiaris" NOUN) ("dog" NOUN))

CAUSE NIL

DERIVATIONALLY-RELATED-FORM NIL

DOMAIN-OF-SYNSET-REGION NIL

DOMAIN-OF-SYNSET-TOPIC NIL

DOMAIN-OF-SYNSET-USAGE NIL

ENTAILMENT NIL

GLOSS "a member of the genus Canis (probably descended from the common wolf) that has been domesticated by man since prehistoric times; occurs in many breeds; 'the dog barked all night' "

GLOSS-WORDS ("breed" "occur" "time" "prehistoric" "man" "domesticate" "domesticated" "wolf" "common" "descend" "canis" "genus" "member" "animal" "dog" "domestic")

HYPERNYM (("domestic_animal" NOUN 1317541) ("domesticated_animal" NOUN 1317541)
("canine" NOUN 2083346) ("canid" NOUN 2083346))

HYPERNYM-CHAIN ("domestic_dog" "canis_familiaris" "dog" "canine" "canid" "domestic_animal"
"domesticated_animal" "fauna" "carnivore" "animal" "creature" "beast" "animate_being" "brute" "eutherian" "organism"
"being" "placental_mammal" "placental" "eutherian_mammal" "animate_thing" "mammalian" "mammal" "living_thing"
"craniate" "unit" "whole" "vertebrate" "physical_object" "chordate" "object" "physical_entity" "entity")

HYPONYM (("puppy" NOUN 1322604) ("bow-wow" NOUN 2084732) ("barker" NOUN 2084732)
("doggy" NOUN 2084732) ("doggie" NOUN 2084732) ("pooch" NOUN 2084732) ("mutt" NOUN 2084861)
("mongrel" NOUN 2084861) ("cur" NOUN 2084861) ("lapdog" NOUN 2085272) ("toy" NOUN 2085374)
("toy_dog" NOUN 2085374) ("hunting_dog" NOUN 2087122) ("working_dog" NOUN 2103406)
("carriage_dog" NOUN 2110341) ("coach_dog" NOUN 2110341) ("dalmatian" NOUN 2110341)
("basenji" NOUN 2110806) ("pug-dog" NOUN 2110958) ("pug" NOUN 2110958) ("leonberg" NOUN 2111129)
("newfoundland_dog" NOUN 2111277) ("newfoundland" NOUN 2111277) ("great_pyrenees" NOUN 2111500)
("spitz" NOUN 2111626) ("belgian_griffon" NOUN 2112497) ("brussels_griffon" NOUN 2112497) ("griffon" NOUN 2112497)
("welsh_corgi" NOUN 2112826) ("corgi" NOUN 2112826) ("poodle_dog" NOUN 2113335) ("poodle" NOUN 2113335)
("mexican_hairless" NOUN 2113978))

WordNet

INSTANCE-HYPERNYM NIL

INSTANCE-HYPONYM NIL

MEMBER-HOLONYM (("genus_canis" NOUN 2083863) ("canis" NOUN 2083863) ("pack" NOUN 7994941))

MEMBER-MERONYM NIL

MEMBER-OF-THIS-DOMAIN-REGION NIL

MEMBER-OF-THIS-DOMAIN-TOPIC NIL

MEMBER-OF-THIS-DOMAIN-USAGE NIL

ORDER 1

ORIGINAL-WORD "domestic_dog"

PART-HOLONYM NIL

PART-MERONYM (("flag" NOUN 2158846))

PARTICIPLE-OF-VERB NIL

PERTAINYM-PERTAINS-TO-NOUN NIL

POINTER-SYNS NIL

PRIMARY-SYNONYMS ("genus_canis" "canis" "pack" "flag" "domestic_animal" "domesticated_animal" "canine" "canid" "puppy" "bow-wow" "barker" "doggy" "doggie" "pooch" "mutt" "mongrel" "cur" "lapdog" "toy" "toy_dog" "hunting_dog" "working_dog" "carriage_dog" "coach_dog" "dalmatian" "basenji" "pug-dog" "pug" "leonberg" "newfoundland_dog" "newfoundland" "great_pyrenees" "spitz" "belgian_griffon" "brussels_griffon" "griffon" "welsh_corgi" "corgi" "poodle_dog" "poodle" "mexican_hairless" "canis_familiaris" "dog" "domestic_dog")

RELATED-TERM NIL

SEMANTIC-TYPE NOUN-ANIMAL

SENSE-DATA NIL

SIMILAR-TO NIL

SUBSTANCE-HOLONYM NIL

SUBSTANCE-MERONYM NIL

SYNSET-ID (2084071 NOUN)

TYPE NOUN

VERB-GROUP NIL

WORD "domestic_dog"

WORD-SENSE "domestic_dog%1"



LIWC Needs Words

- ◆ grab some from WordNet
- ◆ find some other sources

Add Synonyms

word \rightarrow $\text{syn}_1, \dots, \text{syn}_m$

if word \in LIWC dictionary
then $\forall i$ add(syn_i)
unless $\text{syn}_i \in$ LIWC dictionary

let $\text{cats}(\text{word}) = \{(c_1, cf_1) \dots\}$
add(syn_i) with $\{(c_1, f \cdot cf_1) \dots\}$

repeat until none added; $f = .75^n$

Adding All Synonyms

- ◆ 31,190 total words
- ◆ back-learned Yarkoni coefficients: RPGoni

Found a List of 458 Swear and Bad Words

Add swear-word with categories

swear-words {66 .3}

anger {18 .3}

affect {12 .3}

when swear-word \notin dictionary

31, 473 total words

Problem #2

- ◆ sentiment analysis seems to not work for deliberate writing
- ◆ this is an interesting problem, but not explored much
- ◆ I trained a genre classifier over personality traits
 - poetry, fiction, nonfiction
- ◆ not solved, only noted

Text File	Genre	Poetry	Fiction	Nonfiction
Knott (P)	Poetry			
Trakl (P)	Poetry			
Lanier (P)	Poetry			
The Wasteland (P)	Poetry			
Moby Dick (F)	Fiction			
US Constitution (NF)	Nonfiction			
To Kill a Mockingbird (F)	Fiction			
Hamlet (?)	Fiction[Poetry]	-5.14	24.7	-24.49
Bertrand Russell (NF)	Nonfiction			
Charles Babbage (NF)	Nonfiction			
Darwin (NF)	Nonfiction			
Tennyson / In Memoriam (P)	Fiction			
Bible (?)	Fiction[Poetry]	-9.6	42.3	-37.75
Pete Turchi's New Book (NF)	Fiction[Nonfiction]	-22.38	16.83	-5.14

$-5.0 \leq P$ (poetry)
 $-5.0 \leq NF$ (nonfiction)
 otherwise (fiction)

$-10.0 \leq P < -5.0$ (poetry mixin)
 $-10.0 \leq NF < -5.0$ (nonfiction mixin)
 $35.0 \leq F$ (fiction mixin)



Templates

- ◆ assume they are created elsewhere
- ◆ lay out choices to make
- ◆ simple
- ◆ compiles into an optimization problem
- ◆ creates a “misfit” function for simulated annealing

Simple Initial Template

I have a (big adj) (fat adj) (dog noun).

Simple Initial Template

I have a great heavy dog.

I have a big heavy dog.

I have a large heavy dog.

I have a great heavy cur.

I have a big fat cur.

Problem #3

- ◆ need to take a “basic” word and adjust it
- ◆ e.g., plurals, capitalization

(adjust-word "dog" 'noun '(pl))

-> "dogs"

(adjust-word "have" 'verb '(past singular))

-> "had"

(adjust-word "have" 'verb
'(first-person singular future-perfect indicative))

-> "will have had"

("hard_to_please" 'adj '(-est cap))

-> "Hardest to please"

Templates

Start to Become a DSL

```
(bind ((f (fat adj)))
```

```
  "I have a (big adj) (ref f) (dog noun).
```

```
  He is just too (ref f) for a (dog noun).")
```

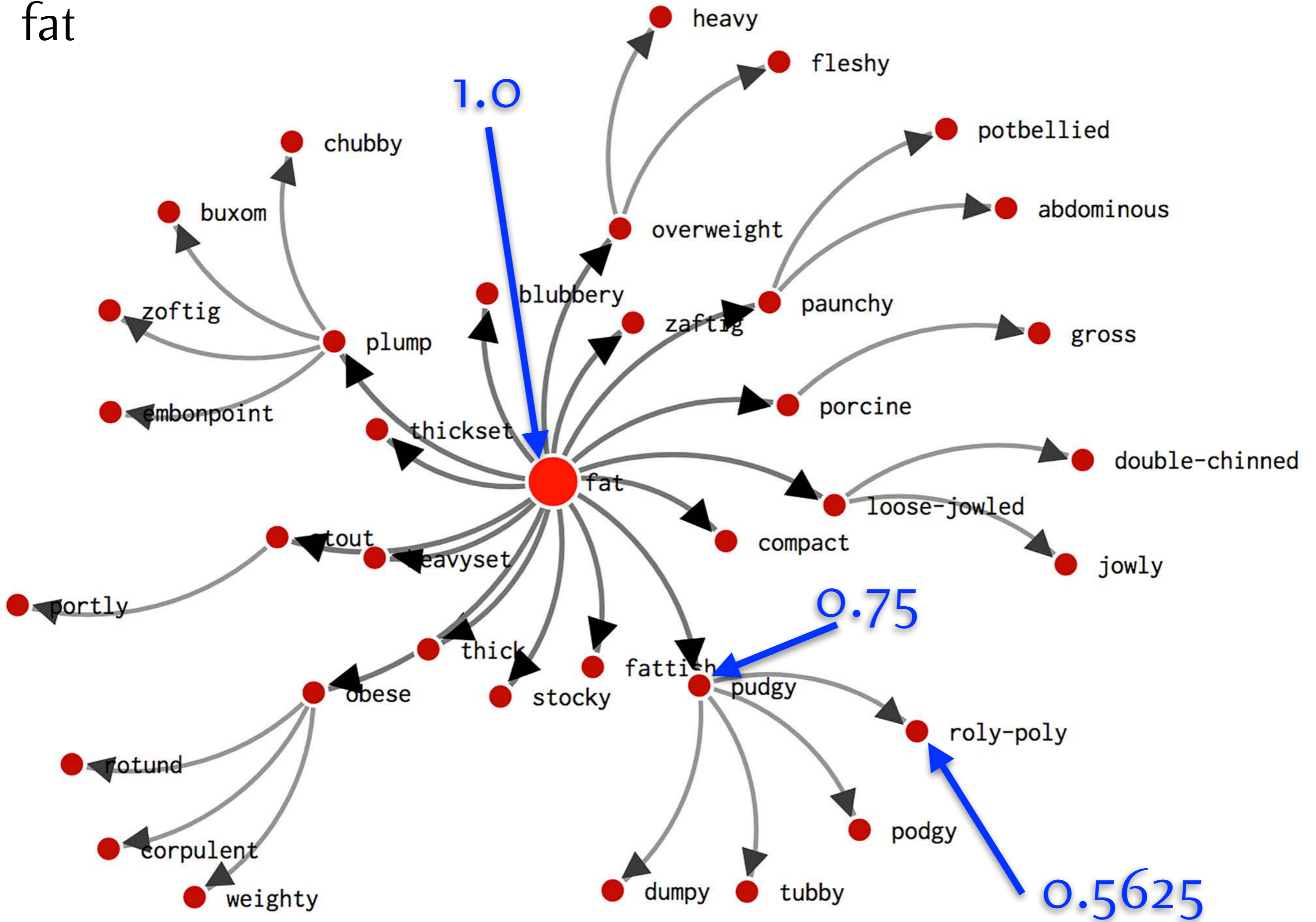

I have a big fat dog.
He is just too fat for a cur.

I have a great heavy Peke.
He is just too heavy for a dog.

I have a big abdominous dog.
He is just too abdominous for a dog.

I have a large heavy dog.
He is just too heavy for a Peke.

fat




```

(lambda (word-penalty process-number number-of-words-mp length-liwc-categories liwc-sentiments trampoline n-sentence-words)
  (declare (ignorable word-penalty number-of-words-mp length-liwc-categories liwc-sentiments process-number trampoline n-sentence-words))
  (macrolet ((sentence-word (index1 index2) `(aref (aref *word-arrays-mp* process-number) ,index1
                                                    (aref (aref *text-permutation-vectors-mp* process-number) ,index2))))
    (+-evaluation (&body forms) `(+ ,@forms)))
  (let ((process-sentence-word #'(lambda (index) (sentence-word wa-word index))))
    (declare (ignorable process-sentence-word))
    (+-evaluation
      (* 1.0
        (abs (- 50.0
              (personality-mp '((65 -0.2 grooming)
                                (64 -0.14 sleeping)
                                (63 -0.15 eating)
                                (60 -0.09 physical_states)
                                (59 0.15 death)
                                (53 -0.14 sports)
                                (52 -0.2 home)
                                (51 -0.17 leisure)
                                (46 -0.22 motion)
                                (44 0.11 inclusive)
                                (43 -0.11 down)
                                (42 -0.15 up)
                                (41 -0.11 space)
                                (39 -0.16 present_tense_vb)
                                (38 -0.16 past_tense_vb)
                                (37 -0.22 time)
                                (36 -0.09 humans)
                                (35 -0.17 family)
                                (33 -0.14 other_refs)
                                (31 -0.14 social_processes)
                                (29 -0.08 hearing)
                                (27 -0.11 sensory_processes)
                                (23 -0.12 discrepancy)
                                (20 -0.09 cognitive_processes)
                                (14 -0.11 positive_feelings)
                                (13 -0.15 positive_emotions)
                                (12 -0.12 affect)
                                (11 -0.08 numbers)
                                (10 0.17 prepositions)
                                (9 0.2 articles)
                                (8 -0.11 assent)
                                (7 -0.13 negation)
                                (5 -0.12 2nd_person)
                                (4 -0.19 1st_person)
                                (3 -0.1 1st_person_pl)
                                (2 -0.16 1st_person_sing)
                                (1 -0.21 total_pronouns)))
          0.28
          -1.0839001
          number-of-words-mp
          length-liwc-categories
          liwc-sentiments
          trampoline)))))))))

```

constructed misfit function
for simulated annealing



Problem #4

- ◆ even though every revision is **guaranteed*** grammatical, many revisions are nonsensical

*PL Theorists: note the word “guarantee”

Giving Synonym Search Free Rein Reveals Insanity

```
(bind ((f (fat adj)))
```

```
"I (have verb) a (big adj) (ref f) (dog noun).  
He is just too (ref f) for a (dog noun).")
```


I amass a big heavy dog.
He's just too heavy for a dog.

I stock a big abdominous dog.
He's just too abdominous for a dog.

I economize a big heavy dog.
He's just too heavy for a dog.

I computerize a big heavy dog.
He's just too heavy for a dog.

I husband a big abdominous dog.
He's just too abdominous for a dog.

Senses of Words Not Accounted for

- ◆ average senses per word: 1.34
- ◆ most number of senses: 59 (for the noun “break”)
- ◆ average senses per word with more than one sense: 2.9
- ◆ words with more than one sense: 28,131

Senses

hacker, NOUN

Senses:

1. **hacker**, NOUN-PERSON: someone who plays golf poorly
2. **cyberpunk**, NOUN-PERSON: a programmer who breaks into computer systems in order to steal or change or destroy information as a form of cyber-terrorism
3. **hacker**, NOUN-PERSON: a programmer for whom computing is its own reward; may enjoy the challenge of breaking into other computers but does no harm; "true hackers subscribe to a code of ethics and look down upon crackers"
4. **hack**, NOUN-PERSON: one who works hard at boring tasks

Plan #4

- ◆ n-grams to control sense
- ◆ smarter use of WordNet
- ◆ go beyond synonyms (later)



What is an N-Gram?

are huge and	0.14455404
one copy of	0.380291
sooner rather than	0.3726525
was wrong and	0.32407436
be candid with	0.12155247
decided to sleep	0.12155247
had appeared as	0.20447502
up the remnants	0.11795887
its opinion the	0.121112496
answer is absolutely	0.1
which are irrelevant	0.122132495
of bare ground	0.10501691

Google n-Grams

- ◆ 50m Google 1-, 2-, 3-, 4-, 5-grams, general literature
- ◆ two weeks to download them
- ◆ a month to clean and process them
 - original texts OCRed
 - some in non-Western alphabets
 - scored by frequency, made consistent (cascading)

How n-Grams Could Help

"\##bigdog"

#(("#\##bigdogs" . 0.2238675) ("#\##bigdog" . 0.28050122))

(in-2-grams "big" "fat") -> 0.32660025

(in-2-grams "fat" "big") -> 0.0

I like (or fat big :label fb) (or fat big :different fb) dogs.

I like big fat dogs.

more on this later



Smarter Use of WordNet

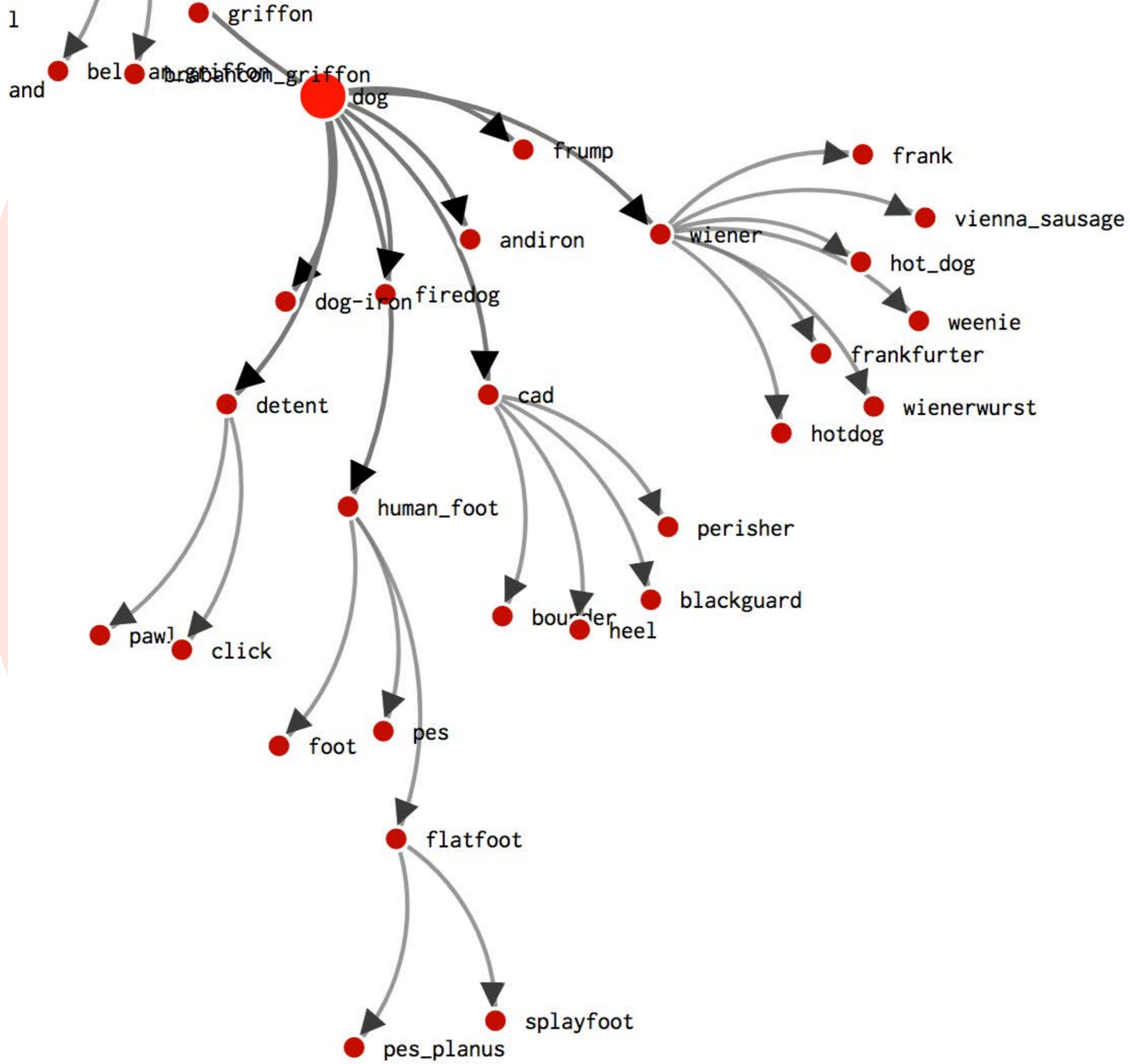
dog, NOUN

Senses:

- 1: **domestic_dog**, NOUN-ANIMAL: a member of the genus Canis (probably descended from the common wolf) that has been domesticated by man since prehistoric times; occurs in many breeds; "the dog barked all night"
- 2: **dog**, NOUN-PERSON: a dull unattractive unpleasant girl or woman; "she got a reputation as a frump"; "she's a real dog"
- 3: **dog**, NOUN-PERSON: informal term for a man; "you lucky dog"
- 4: **cad**, NOUN-PERSON: someone who is morally reprehensible; "you dirty dog"
- 5: **wiener**, NOUN-FOOD: a smooth-textured sausage of minced beef or pork usually smoked; often served on a bread roll
- 6: **detent**, NOUN-ARTIFACT: a hinged catch that fits into a notch of a ratchet to move a wheel forward or prevent it from moving backward
- 7: **dog**, NOUN-ARTIFACT: metal supports for logs in a fireplace; "the andirons were too hot to touch"
- 8: **human_foot**, NOUN-BODY: the part of the leg of a human being below the ankle joint; "his bare feet projected from his trousers"; "armored from head to foot"

I like (dog noun pl).

I like feet.
I like huskies.
I like detents.
I like clicks.
I like perishers.
I like curs.
I like toys.
I like blackguards.
I like retrievers.
I like heels.
I like Pekes.
I like Newfoundlands.
I like hounds.
I like franks.
I like cads.
I like wieners.
I like andirons.
I like puppies.
I like barkers.



I like (dog noun-animal pl).

I like dogs.
I like huskies.
I like curs.
I like toys.
I like retrievers.
I like Pekes.
I like Newfoundlands.
I like hounds.
I like puppies.
I like barkers.
I like Leonbergs.
I like bow-wows.
I like poodles.
I like mongrels.
I like pooches.
I like lapdogs.
I like pug-dogs.
I like corgis.
I like spitzes.
I like mutts.
I like griffons.



Personality Not the Best Way to Model a Writer

- ◆ we can use a writer's n-grams to model their writing in addition to personality
- ◆ we can also use more of the capabilities of WordNet to enhance the effects of character
- ◆ but the problem of nonsense persists

I (meet verb past) a (woman noun-person).
On that (day noun-time)
the (wind noun-phenomenon)
(or (blow verb-weather past continuous)
(blow verb-motion past continuous)).

King James Bible

I saw a woman.	On that day the wind was blowing.
I saw a woman.	On that day the breath was blowing.
I saw a woman.	On that day the air was blowing.
I saw a woman.	On that day the wind was storming.
I saw a woman.	On that day the wind was streaming.
I saw a woman.	On that day the wind was blasting.
I saw a woman.	On that day the breath was storming.
I saw a woman.	On that day the breath was streaming.
I saw a wife.	On that day the wind was blowing.
I saw a woman.	On that day the breath was blasting.

A Type Theorist

I saw a **char**.

I saw a cat.

I saw a widow.

I saw a nanny.

I saw a siren.

I saw a geisha.

I saw a nymph.

I saw a bridesmaid.

I saw a debutante.

I saw a **baggage**.

On that even the draft was streaming.

On that even the draft was streaming.

On that even the draft was streaming.

On that even the draft was streaming.

On that even the draft was streaming.

On that even the draft was streaming.

On that even the draft was streaming.

On that even the draft was streaming.

On that even the draft was streaming.

On that even the draft was streaming.

char, NOUN-PERSON: a human female employed to do housework

baggage, NOUN-PERSON: a worthless or immoral woman

On every page: JULY 22, 2010 **DRAFT** 17:13

Hm

- ◆ simple synonym search even with n-grams still produces the seemingly nonsensical
- ◆ technically it's not totally without sense, but...
- ◆ ...it can be interesting, and the variety is refreshing



New Official Research Goals

- ◆ (SMISC finished, around 2014)
- ◆ assist human creativity
- ◆ mimic a specific writer

My Research Questions

- ◆ can a program manipulate language like a poet?
- ◆ can a program assist human creativity
- ◆ can a program think like a poet?
- ◆ can (almost) all the reasoning be performed using only words?

Writerly Thinking

- ◆ (literary) writers think about words
- ◆ ...how they work
- ◆ ...how they sound
- ◆ ...how they interact

Haloes: Slanting Searches

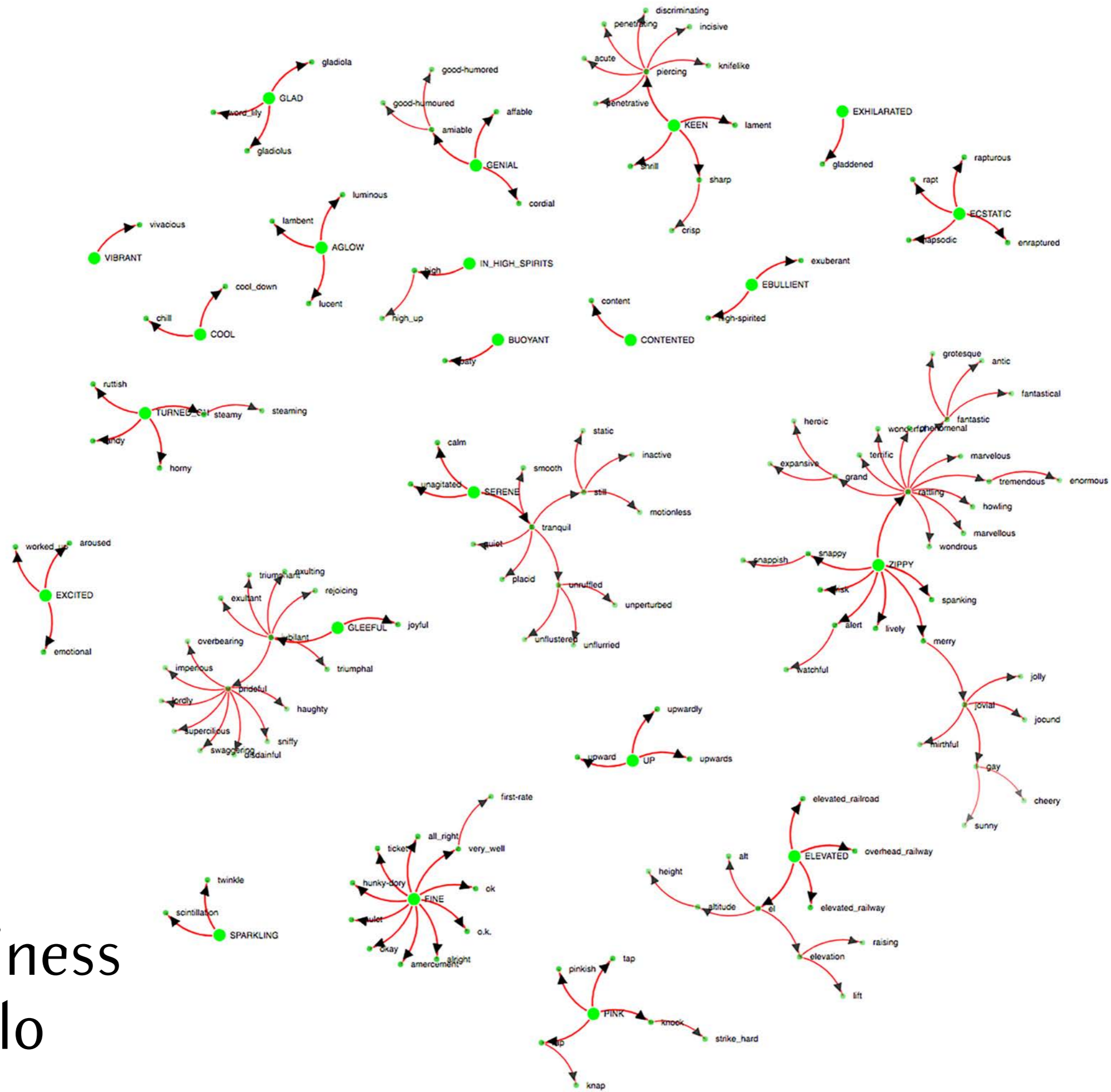
- ◆ subtext
- ◆ mood
- ◆ emotion

Halos

Happiness Halo

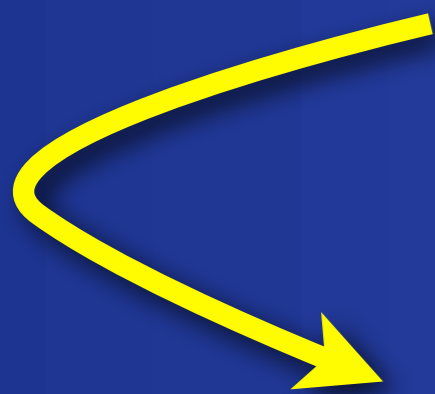
Buoyant, Delighted, Ebullient, **Ecstatic**, Elated, Energetic, Enthusiastic, Euphoric, Excited, Exhilarated, Fine, Overjoyed, **Pleasant**, Thrilled, Tickled pink, Turned on, Vibrant, Zippy

Happiness Halo



Halos

The woods are lovely, dark, and deep



*The woods are **fine-looking**, not too **light**, and **heavy**.*

Happiness Halo

Buoyant, Delighted, Ebullient, **Ecstatic**, Elated, Energetic,
Enthusiastic, Euphoric, Excited, Exhilarated, Fine,
Overjoyed, **Pleasant**, Thrilled, Tickled pink, Turned on, Vibrant, Zippy

The woods are lovely, dark, and deep

*The woods aren't very **ugly**, not too **light**,
and not too **shallow***

Anger Halo

Affronted, **Belligerent**, Bitter, Burned up, **Crabby**, Enraged,
Fuming, Furious, Heated, Incensed, Infuriated, **Intense**,
Outraged, Provoked, Seething, Storming, Truculent, Vengeful,
Vindictive, Wild

Halos Can Attract / Repel

- ◆ positive words boost word evaluations
- ◆ negative words decrease word evaluations
- ◆ in general, negative weights can be used to avoid words and some sense problems



What Happens When You Obey n-grams?

- ◆ nothing new happens
- ◆ nothing truly new is written

Trained Writers

- ◆ think about language a lot
- ◆ worry about the words a lot
- ◆ listen to how words sound
- ◆ ...how their rhythms sound
- ◆ think about the subtleties of words
- ◆ ...the structure of sentences and paragraphs



Carnegie Mellon Pronouncing Dictionary

- ◆ sounds of words
- ◆ stresses (rhythm)
- ◆ designed for speech technology
- ◆ Merriam Webster Rhyming Dictionary
 - ~30,000 words in ~4,000 rhyme groups

Problem #5

- ◆ only ~125,000 words
- ◆ InkWell needs to be able to pronounce any string
- ◆ machine learn pronunciation tactics from the CMU dictionary
- ◆ use an algorithmic fallback for the CMU dictionary

CMU Phonetic Dictionary

"fiction"	((F) (IH 1) (K) (SH) (AH 0) (N))
"fictional"	((F) (IH 1) (K) (SH) (AH 0) (N) (AH 0) (L))
"fictional_animal"	((F) (IH 1) (K) (SH) (AH 0) (N) (AH 0) (L) (AE 1) (N) (AH 0) (M) (AH 0) (L))
"fictional_character"	((F) (IH 1) (K) (SH) (AH 0) (N) (AH 0) (L) (K) (EH 1) (R) (IH 0) (K) (T) (ER 0))
"fictionalize"	((F) (IH 1) (K) (SH) (AH 0) (N) (AH 0) (L) (AY 2) (Z))
"fictionalized"	((F) (IH 1) (K) (SH) (AH 0) (N) (AH 0) (L) (AY 2) (Z) (D))
"fictions"	((F) (IH 1) (K) (SH) (AH 0) (N) (Z))
"fictitious"	((F) (IH 0) (K) (T) (IH 1) (SH) (AH 0) (S))
"fictitious_character"	((F) (IH 0) (K) (T) (IH 1) (SH) (AH 0) (S) (K) (EH 1) (R) (IH 0) (K) (T) (ER 0))
"fictitious_name"	((F) (IH 0) (K) (T) (IH 1) (SH) (AH 0) (S) (N) (EY 1) (M))
"fictitious_place"	((F) (IH 0) (K) (T) (IH 1) (SH) (AH 0) (S) (P) (L) (EY 1) (S))
"fictitiously"	((F) (IH 0) (K) (T) (IH 1) (SH) (AH 0) (S) (L) (AY 1))
"ficus"	((F) (IY 1) (S) (IY 1) (Y) (UW 1) (EH 1) (S)) (F) (IY 1) (K) (AH 1) (S)) (F) (AY 1) (S) (IY 1) (Y) (UW 1) (EH 1) (S)) (F) (AY 1) (K) (AH 1) (S))

(IH 1) = (<sound> <stress>)

Algorithmic Pronunciation

"disorganised" #((((D) (IH 1) (Z) (AO 1) (R) (G) (AE 1) (N) (AH 0) (S) (D))) ARTIFICIAL)

"dispelief" #((((D) (IH 1) (Z) (P) (EH 0) (L) (IY 1) (F))) ARTIFICIAL)

"dispensations" #((((D) (IH 1) (Z) (P) (EH 0) (N) (Z) (AE 1) (T) (AY 1) (AH 0) (N) (Z))) ARTIFICIAL)

"disphoria" #((((D) (IH 1) (Z) (F) (OW 1) (R) (IY 0) (AH 0))) ARTIFICIAL)

"disputandum" #((((D) (IH 1) (Z) (P) (UW 0) (T) (AE 1) (N) (D) (AH 1) (M))) ARTIFICIAL)

"disputatio" #((((D) (IH 1) (Z) (P) (UH 1) (T) (AE 1) (T) (AY 1) (OW 0))) ARTIFICIAL)

"disputatious" #((((D) (IH 1) (Z) (P) (UH 1) (T) (AE 1) (T) (AY 2) (OW 2) (Y) (UW 1) (Z))) ARTIFICIAL)

"disputatiously" #((((D) (IH 1) (Z) (P) (UH 1) (T) (AE 1) (T) (AY 2) (OW 2) (Y) (UW 1) (Z) (L) (AY 1))) ARTIFICIAL)

"disquisition" #((((D) (IH 1) (Z) (K) (W) (IY 1) (Z) (IH 1) (T) (AY 1) (AH 0) (N))) ARTIFICIAL)

"disquisitions" #((((D) (IH 1) (Z) (K) (W) (IY 1) (Z) (IH 1) (T) (AY 1) (AH 0) (N) (Z))) ARTIFICIAL)

"disreputability" #((((D) (IH 1) (Z) (R) (EH 1) (P) (UW 0) (T) (AA 1) (B) (IH 1) (L) (IH 1) (T) (Y))) ARTIFICIAL)

"disreputably" #((((D) (IH 1) (Z) (R) (EH 1) (P) (UW 0) (T) (EY 1) (B) (L) (IY 0))) ARTIFICIAL)

Algorithmic Rhyming

- ◆ rhyme computation for words and phrases
- ◆ how much do vowels and consonants contribute to rhyme?
- ◆ how much do early syllables in a word or phrase contribute to rhyminess?
- ◆ “fictional” and “critical” rhyme this much: 50%

Echoes

- ◆ like an unordered rhyme
- ◆ counts the number common sounds, weighted by their sonic similarity
- ◆ takes into account similar ordering of common sounds —how close do the words come to rhyming
- ◆ “fictional” and “critical” echo this much: 70%

Notable Echoes of “computer programmer”

- ◆ provoked me to anger (King James Bible)
- ◆ most persecuted by man (Origin of Species, Darwin)
- ◆ compound for grey amber (Moby Dick, Melville)
- ◆ come from yond poor girl (Sonnets & Plays, Shakespeare)
- ◆ program a computer (Patterns of Software, rpg)



Problem #6

- ◆ the “woods” problem
 - “Whose woods these are I think I know”
- ◆ using simply looked-up synonyms
 - “I like the (woods noun).” -> 14 synonyms
- ◆ haloes can be generalized to help

Senses

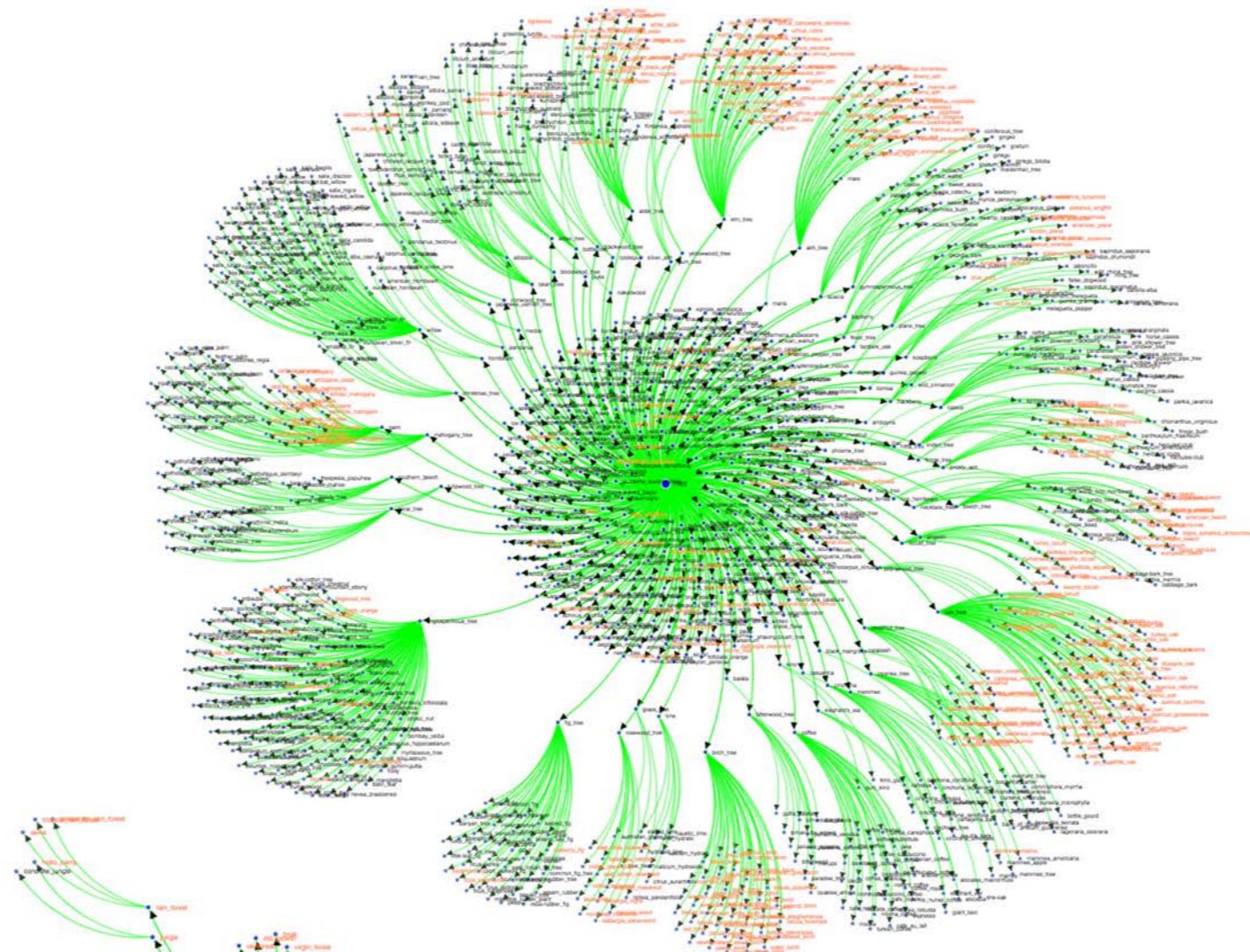
- ◆ data structure consisting of **words + relevances**, based on decaying synonym-network spreading
- ◆ can be used to **compute semantic similarity** of words / similarity of meaning, kind of like cosine distance
- ◆ ...like a magnet passed over all the words InkWell knows—it picks up to various heights words attracted to it
- ◆ kind of like word2vec embeddings (a rediscovery)

noun-animal: [dog wolf]

noun-plant & noun-group

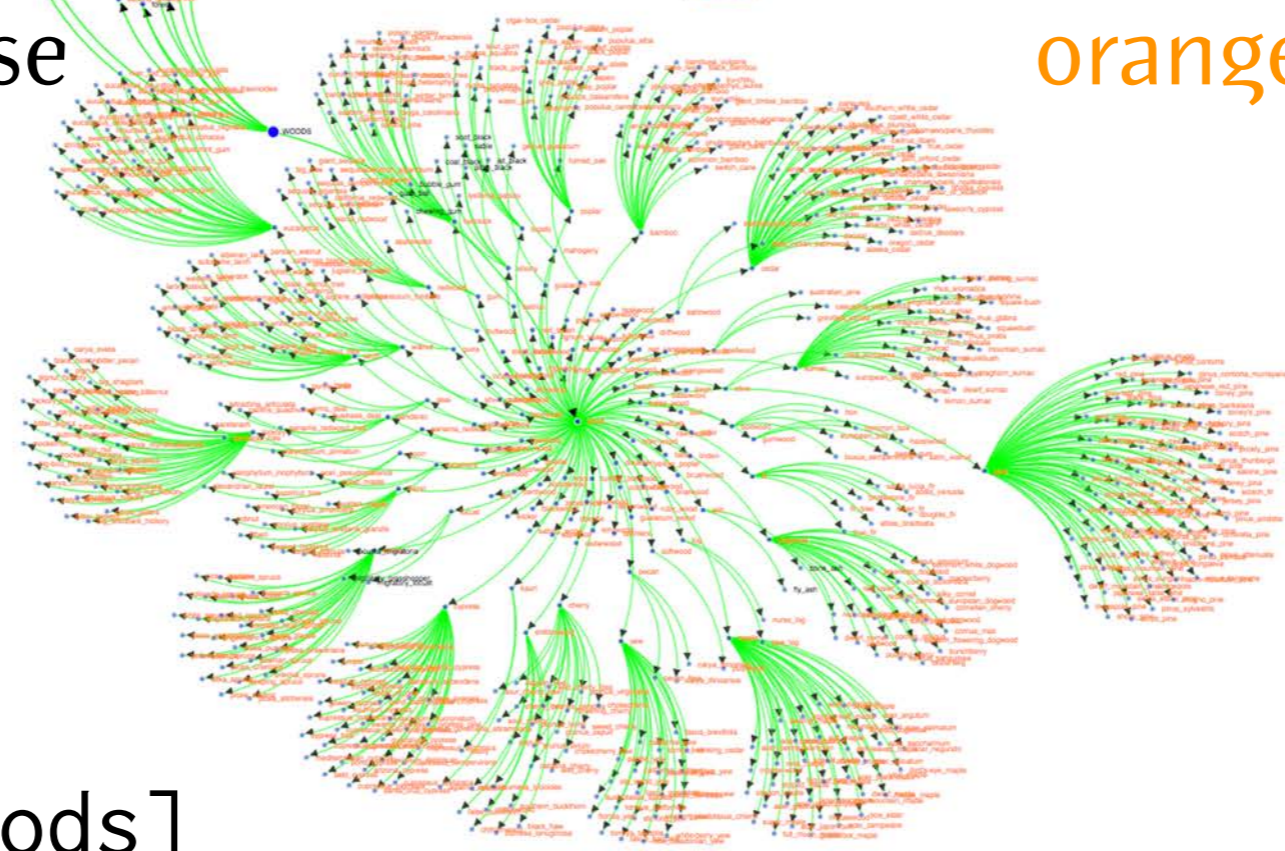
+sense: [forest tree woods]

-sense: [wood]



black words: +sense

orange words: -sense



[forest tree woods]

[wood]

<choose>

- ◆ search all synsets of all words
- ◆ filtering for semantic type
- ◆ computing the similarity between the sense and the synset

I like (<choose> noun-animal pl :+sense [dog wolf])

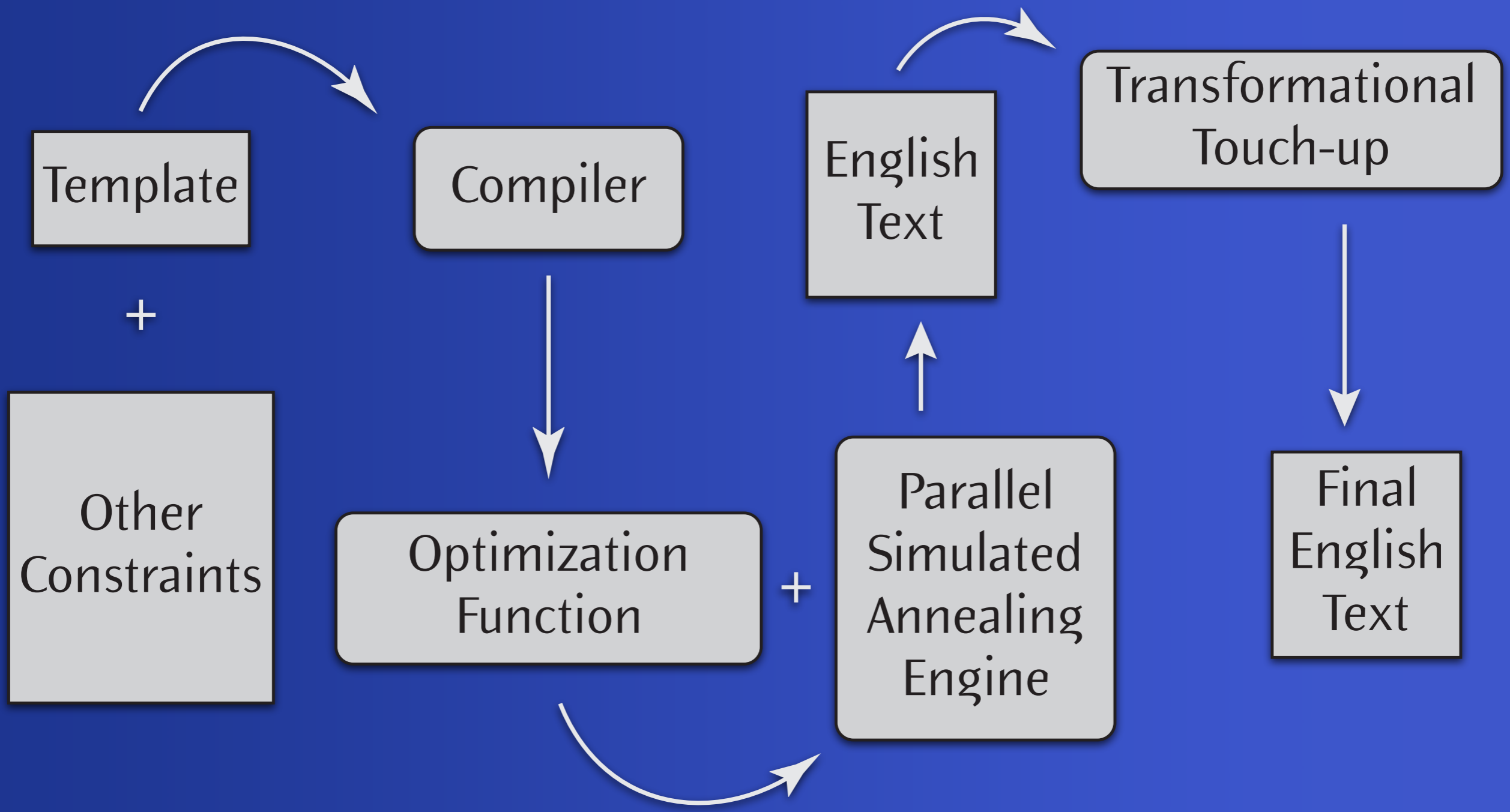
find words of this semantic
type and sense

I like dogs
I like toys
I like Newfoundlands
I like huskies
I like retrievers
I like hounds
I like cubs
I like puppies
I like pups
I like whelps
I like Canidaes
I like wolfhounds
I like Carnivoras
I like Canises
I like jackals
I like curs
I like Pokes
I like Leonbergs
I like coydogs
I like Pomeranians

NOUN-ANIMAL: any of several breeds
of very small dogs kept purely as pets

NOUN-ANIMAL: offspring
of a coyote and a dog





```

(with-personality-traits (*writer-big-five*)
 (with-global-constraints ((all-echo)(all-different))
  (with-pervasive-predicates (#'syllable-bonus-few)
   (with-typographic-style (poem)
    (bind ((w1 (<choose> verb-cognition :+sense [know certain]))
           (w2 (snow noun-substance :rhyme [though]))
           (w3 (<choose> adj :+sense [queer odd unusual weird demented stupid silly]))
           (w4 (or (year noun-quantity) (week noun-quantity) (month noun-quantity)
                  (season noun-quantity)))
           (w5 (<choose> verb-motion :+sense [cause move back forth shake]))
           (w6 (<choose> noun-object
                :+sense [small fragment broken break whole flake]))
           (here (here adj)) (near (near adj))
           (mile (mile noun-quantity pl)) (sleep (sleep verb))
           (woods (<choose> noun-plant pl
                  :+sense [forest trees plants wooded area] :-sense [wood])))))

```

```

"Whose (ref woods) these are I (<choose> verb-cognition
                               :+sense [think guess surmise]) I (ref w1).

```

```

His (house noun) is in the (village noun) though;

```

```

He will not see me (stop verb gerund) (ref here :rhyme near)

```

```

To (watch verb-perception) his (ref woods) (fill verb) up with (ref w2
                                                                :different w1
                                                                :rhyme w1).

```

```

...

```


My (<choose> adj :sense little-sense) (horse noun-animal) must
(<choose> verb-cognition :sense think-sense) it (ref w3 :echo w1)
To (stop verb) without a (<choose> noun :sense farmhouse-sense)
(binding near :rhyme w3)
Between the (ref woods) and (frozen adj) (<choose> noun-object :sense lake-sense)
The (<choose> adj -est :sense darkest-sense)
(<choose> noun-time :sense evening-sense)
of the (ref w4 :different w3 :rhyme w3).

He gives his (harness noun) (bell noun pl) a (ref w5 :echo w3)
To (ask verb) if there is some (mistake noun :rhyme w5).
The only other (<choose> noun :sense sound-sense) is the
(<choose> noun-act :sense arc-sense)
Of (<choose> ADJ :sense easy-sense) (wind noun) and
(<choose> ADJ :sense downy-sense)
(ref w6 :different w5 :rhyme w5).

The (ref woods) are (<choose> adj :sense lovely-sense),
(<choose> adj :sense dark-sense), and (<choose> adj :sense deep-sense),
But I have (promise noun pl) to (keep verb :rhyme sleep),
And (ref mile) to go before I (ref sleep),
And (ref mile) to go before I (sleep verb :different sleep :rhyme sleep).”))))

“Stopping by Summ
on a Snowy Mean Solar Day”

by

the Roberts Frost and Harper

*Whose sums these are I express I know.
His unit is in the body though;
He will not see me entering existing
To consider his sums operate up with snow.*

*My smaller horse must evaluate it interesting
To keep in without a family discriminating
Between the sums and polar form
The darkest mean solar day of the fundamental quantity.*

*He gives his influence devices an exit
To pass along if there is some statement.
The only other element is the use
Of benign context and yielding element.*

*The sums are logical, insensitive, and severe.
But I have expressions to position,
And large indefinite amounts to go before I sleep in,
And large indefinite amounts to go before I position.*

**Bob Harper
Long words
Anger halo**

*Whose bosks these are I prise I grok.
His manse is in the moshav though;
He'll not see me conking on hand
To snoop his bosks stack up with snow.*

*My svelte stud must prove it grand
To conk without a croft hot
Between the bosks and frost-bound sward
The most black kip of the month.*

*He gives his harness bells a grab
To pry if there's some trip.
The only other cart is the stunt
Of blest gust and soft scrap.*

*The bosks are blessed, drear, and bleak,
But I've troths to ask for it,
And miles to go before I kip,
And miles to go before I kip.*

Same template,
different context

Bob Harper
Short words
Anger halo

*Whose virgin forests these are I hypothesize I misapprehend.
His Tuileries Palace is in Greenwich Village though;
He'll not see me reining in naturally occurring
To reconnoitre his virgin forests alcoholize up with corn snow.*

*My preternatural mesohippus must misapprehend it upsetting
To pull up short without a cancer of the liver neighboring
Between the virgin forests and glaciated cellular inclusion
The most low-density Upper Carboniferous period of the seasoning.*

*He gives his harness telephone bells a bring up
To confer with if there's some trip-up.
The only other intrinsic factor is the electioneering
Of pocket-size moderate gale and indusial buffalo chip.*

*The virgin forests are untrusty, low-spirited, and evil-minded,
But I've Hippocratic oaths to hang in,
And geographical miles to go before I sleep in,
And geographical miles to go before I sleep in.*

Bob Harper
Long words
Anger halo

Personality Redux

I (<choose> verb-emotion :+sense [like favor love])
(or (<choose> adj :+sense [happy])
(<choose> adj :+sense [sad])
(<choose> adj :+sense [angry]))
(<choose> noun-animal pl :+sense [dog wolf]).

I love happy dogs.
No Personality Target

I love passionate bird dogs.
High Extraversion: 46.15%

I yearn for humiliated attack dogs.
High Neuroticism: 32.36%

I relish unclean housedogs.
Low Openness: -54.20%

n-Grams Redux

(<choose> noun-animal pl :+sense [dog wolf])
(or is are) animals.

n-grams select right verb

(<choose> noun-animal pl :+sense [dog wolf])
(or is are) animals.


Dogs are animals.
Toys are animals.
Huskies are animals.
Hounds are animals.
Cubs are animals.
Pups are animals.
Puppies are animals.
Coyotes are animals.

(in-3-grams "dogs" "are" "animals"): 0.0
(in-2-grams "dogs" "are"): 0.36829454
(in-2-grams "are" "animals"): 0.27010748

(in-3-grams "dogs" "is" "animals"): 0.0
(in-2-grams "dogs" "is"): 0.23805899
(in-2-grams "is" "animals"): 0.13000134

<null-word>

(<choose> noun-animal pl :+sense [dog wolf])
(or is are) (or a an <null-word>) animals.



n-grams eliminate stupid word

(<choose> noun-animal pl :+sense [dog wolf])
(or is are) (or a an <null-word>) animals.

Text	n-gram score
Dogs are a animals:	5.11
Dogs are an animals:	5.35
Dogs is a animals:	5.99
Dogs is an animals:	6.66
Dogs is animals:	7.89
Dogs are animals:	8.79

Pups are animals.
Dogs are animals.
Coyotes are animals.
Puppies are animals.
Cubs are animals.
Hounds are animals.
Huskies are animals.
Toys are animals.
Wolves are animals.
Foxes are animals.

I (<choose> verb-emotion :+sense [like favor love])
(or (<choose> adj :+sense [happy])
(<choose> adj :+sense [sad])
(<choose> adj :+sense [angry]))
(<choose> noun-animal pl :+sense [dog wolf]).

I hero-worship easygoing working dogs.

High Conscientiousness: 30.67%

High Music, Echoes: 300

Unusual wording

Common Word Bonus: -200.0

2-Gram Bonus: -100.0

3-Gram Bonus: -100.0

4-Gram Bonus: -100.0

5-Gram Bonus: -100.0

n-grams
can make
interesting stuff



Haiku

- ◆ quick hack in 2014 revealed a talent for short poems
- ◆ led to second order template language
- ◆ novelty seeking
- ◆ ~~self-generation~~

Hairy Templates

(make-full-haiku

```
`(((local-sense tree) noun-plant ([tree] noun) no-auto-cap (:label tree)) (conjunction) (determiner)
((local-sense moon) noun-object ([moon sun stars heavenly body] noun) (:label moon)) (return)
((local-sense light) noun-phenomenon ([light -dark/.5] noun) (:label light) (:echo mute)) (conjunction)
((local-sense mute) adj ([mute silent loud/.33] adj) (:label mute))
((local-sense constellation) noun-object ([constellation galaxy] noun) pl (:label constellation))
(return)
(determiner) ((local-sense bone) noun-body ([bone] noun)) (word-hyphen)
(local-sense chilling) (verb-emotion verb-change) ([chill freeze warm heat] verb) gerund)
((local-sense sea) noun-object ([sea lake] noun))
(:global-constraint (related-synset-p :numeric-row moon constellation))
(:initializer (clr-related-word-caches))))
```


*oak trees and the moon
light and mute constellations
the bone-chilling sea*



Creative Writing

- ◆ all soft constraints can be positive or negative
- ◆ sound like a writer or to sound unlike that writer
- ◆ novelty seeking: don't sound like yourself

I love holy dogs.
King James Bible

I favor maddened curs.
~King James Bible

Novelty Seeking

avant-garde Cooper Union for the Advancement of Science and Art
setting on fire
the half-deep-fat-fried ship-towed long-range acoustic detection system



InkWell

- ◆ not a software engineering project
- ◆ not designed in any standard sense
- ◆ InkWell is my research colleague
- ◆ the “problem” space is a mystery to be explored
- ◆ slowly dawning insights guide revisions

InkWell

- ◆ 50m n-grams from general literature
- ◆ 100,000–1,000,000 n-grams per writer (~165 writers)
- ◆ ~500,000 pronunciations (CMU Phonetic + algorithmic)
- ◆ ~800,000 word & phrase adjustments
- ◆ ~165,000 lemmatizations (inverse of adjust-word)
- ◆ ~66,000 lines of (Common Lisp) code
- ◆ ~15gb when everything loaded and running (tiny n-grams)
 - ~25gb image with full set of n-grams



Correctness?

- ◆ What exactly is correctness for InkWell?
- ◆ InkWell **does not crash**—nothing runs “balls to the wall”
- ◆ Bugs result only in “**interesting wording**”
- ◆ When errors are caught **heuristic alternatives** are used
- ◆ Complex and large data structures are frequently **verified and repaired**
- ◆ When engineered algorithms don't work, **machine-learned alternatives** are used (e.g., pronunciation), sometimes in competition with each other



Making Sense

- ◆ InkWell works “best” when it knows what you want to mean (though craziness is good, too)
- ◆ Not solved yet (by anyone)—it’s an AI holy grail
- ◆ InkWell is getting there

“I like dogs that bark and growl.”

```
(ROOT
 (S
  (NP (PRP I))
  (VP
   (VBP LIKE)
   (NP
    (NP
     (NNS
      DOGS "domestic_dog"
      "a member of the genus Canis probably descended from the common
      wolf that has been domesticated by man.."))
    (SBAR
     (WHNP (WDT THAT))
     (S (VP (VBP BARK "bark" "make barking sounds")
            (CC AND)
            (VBP GROWL "growl" "to utter or emit low dull
            rumbling sounds")))))
    (<PERIOD> <PERIOD>)))
```


“I like dogs with bread and ketchup.”

```
(ROOT
  (S
    (NP (PRP I))
    (VP
      (VBP LIKE)
      (NP (NNS DOGS "wiener" "a smooth-textured sausage of minced beef
        or pork usually smoked often served on a bread roll"))
      (PP
        (IN WITH)
        (NP
          (NN BREAD "staff_of_life" "food made from dough of flour or meal
            and usually raised with yeast or baking powder and then baked")
          (CC AND)
          (NN KETCHUP "cetchup" "thick spicy sauce made from tomatoes"))))
      (<PERIOD> <PERIOD>)))
```




- ◆ Summer 2015
- ◆ Warren Wilson Alumni Writing Conference
- ◆ 18 InkWell haiku presented as mine in a writers' workshop

“terse condensations”

“took the top of my head off”

“evocative”

“funny and profound”

“natural, personal, and rhythmic”

“compact fluid energy”

“whimsical elegance”

“wry and elliptical”

These are extraordinary and extraordinarily small, large poems. The writer of these—this guy, Richard, or whoever—he is not a random person, he's not a random guy. I think he understands randomness, so it's all the more scary. He doesn't do things—as a rule—by accident. He makes choices. The variety is amazing on every level: number of syllables, subject matter, syntax, whether they start out specific and go to the general, or start out general and go to the specific. Some of them are simple, some of them are complex, some of them are funny, some of them are dead serious, some are kind-of in the natural world (but mostly not); there are different persons in them....

*awake in the dark—
the edge of the water can
spread in your presence*

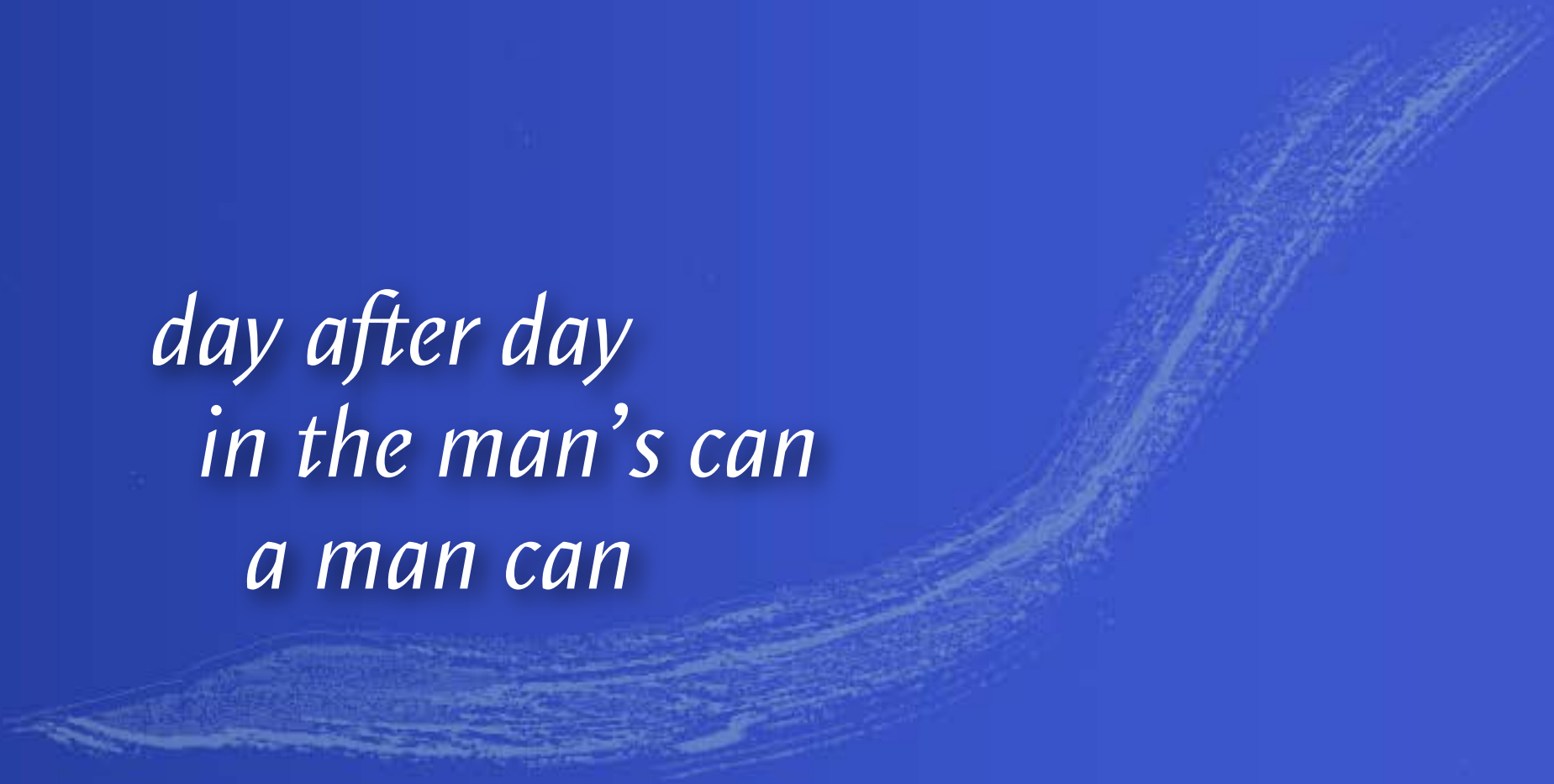
*scrupulous in the twilight—
the price of gold chases
the way of the world in power*

*time of life issue:
a bird of prey pulls up
out of the way into the palm*

*not this fatalist murderousness,
deathwatch,
but your dead subroutine*

*this grave—
no one sees it
mortality, mortality*

*the maiden condominium
opens its award-winning gametocyte
in the control room of the banquet*



*day after day
in the man's can
a man can*

*a reasonable assumption—
by myself,
sampling in chocolate*

*the powerful head
designates its powerful head
to support cognition*



*a few days—
by myself,
browsing guitar-shaped coloring*

*a bitch,
this deep in trick
a fortiori not a man*

*parted in the middle—
the authority of the air conditioner
perfection in the brightness*

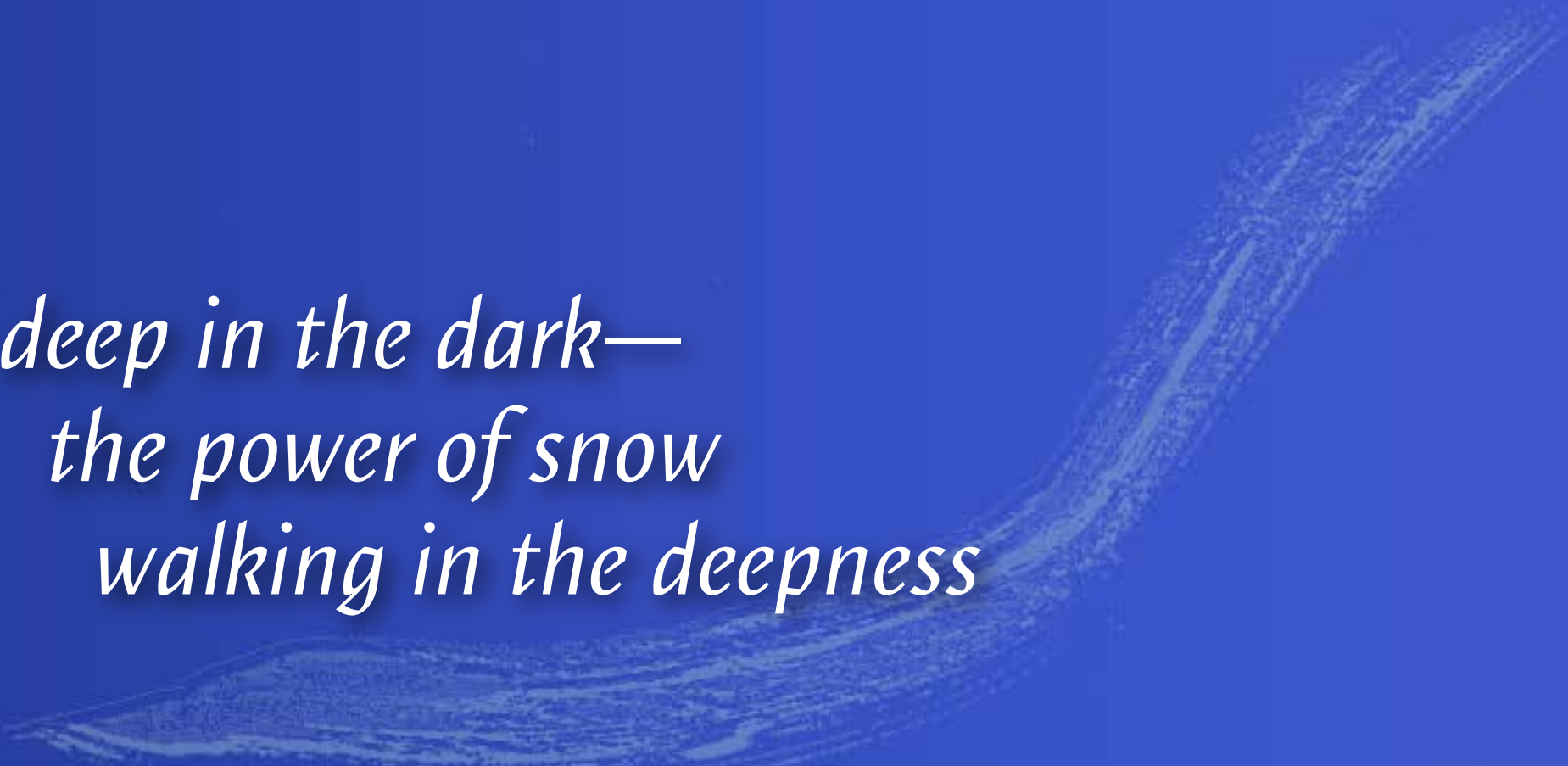
*the hostile defense
leads its problematic rear,
the rear of frustration*

*rural signal,
cannot understand Oregon
—agricultural*

*an on-the-far-side summer night—
whipping up high tea,
we stripped pickles*

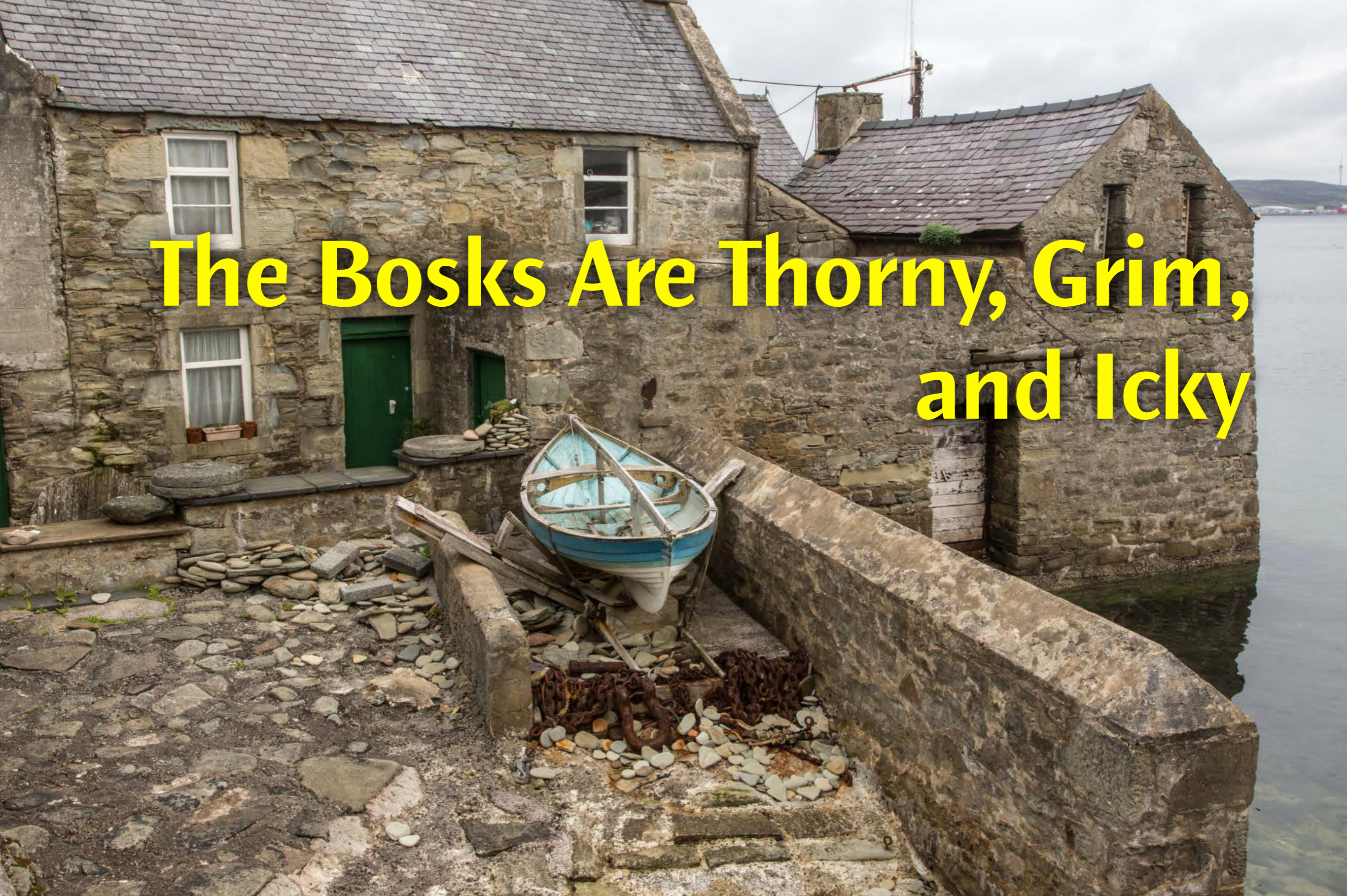
*a crooked rag day—
by myself
dunking distracted sardines*

*tuned adrenalin
my music,
a beat-boogied headful*



*deep in the dark—
the power of snow
walking in the deepness*

—InkWell, December 2014



The Bosks Are Thorny, Grim, and Icky