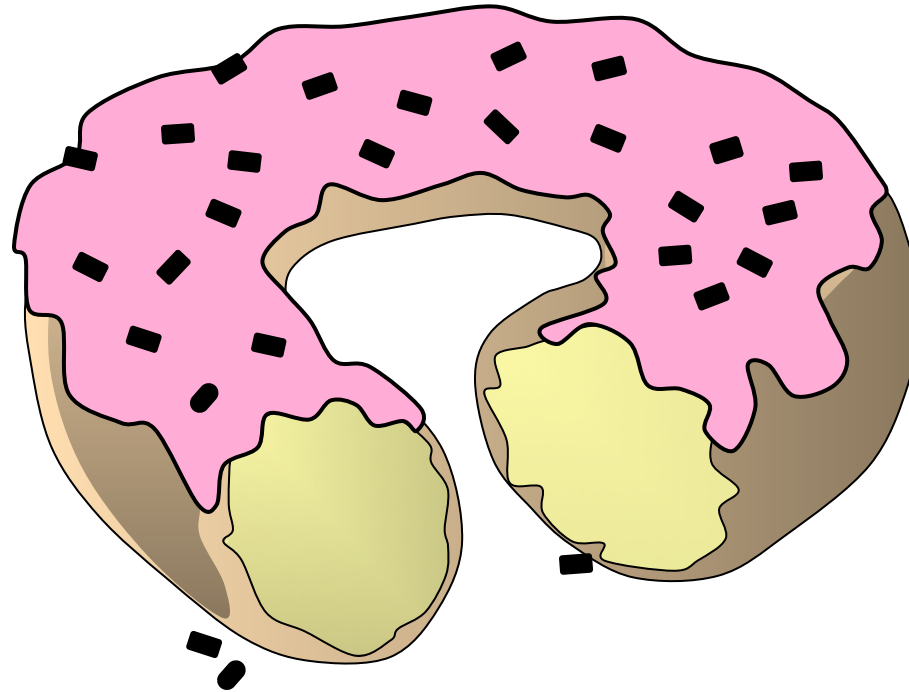


SIMILARITY JUDGMENTS & *DIVERGENT THINKING*



Tony Veale,
UCD & KAIST

A Simple Task (Simple enough for *Homer Simpson*)



Can you provide a good logical definition of a **DONUT**?

(One that matches all donuts but no non-donuts)

The Linnaean Taxonomy (after Carolus Linnaeus)

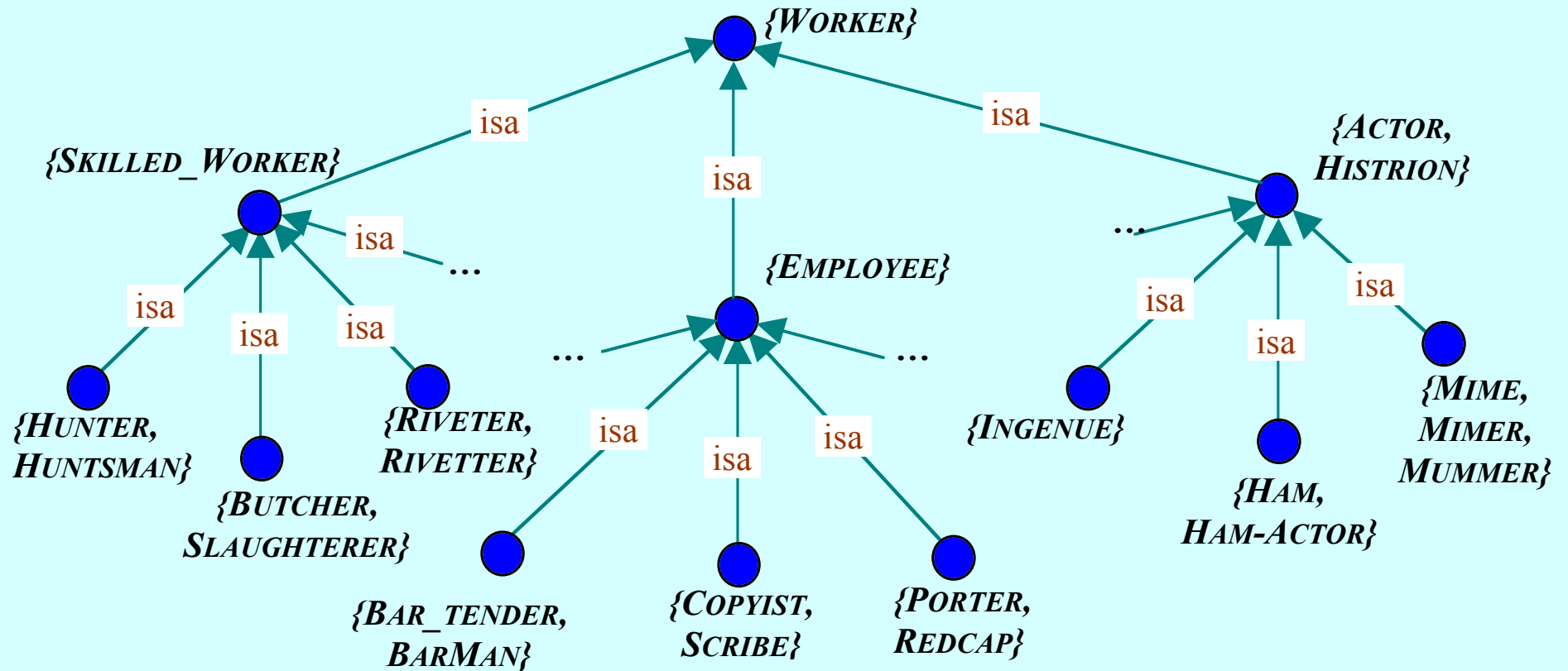
Common Name:	Human	Nine-Banded Armadillo
Kingdom:	Animalia	Animalia
Phylum:	Chordata	Chordata
Class:	Mammalia	Mammalia
Order:	Primata	Cingulata
Family:	Hominidae	Dasypodidae
Genus:	<i>Homo</i>	<i>Dasyus</i>
Species:	<i>Homo sapiens</i>	<i>Dasyus novemcinctus</i>

The Linnaean Taxonomy organizes all life into a series of **Taxa** (sing. **Taxon**)

From the highest (broadest) Taxon **Kingdom** to the lowest (most specific) **Species**

Above is a comparison of the Linnaean classifications of Humans and Armadillos

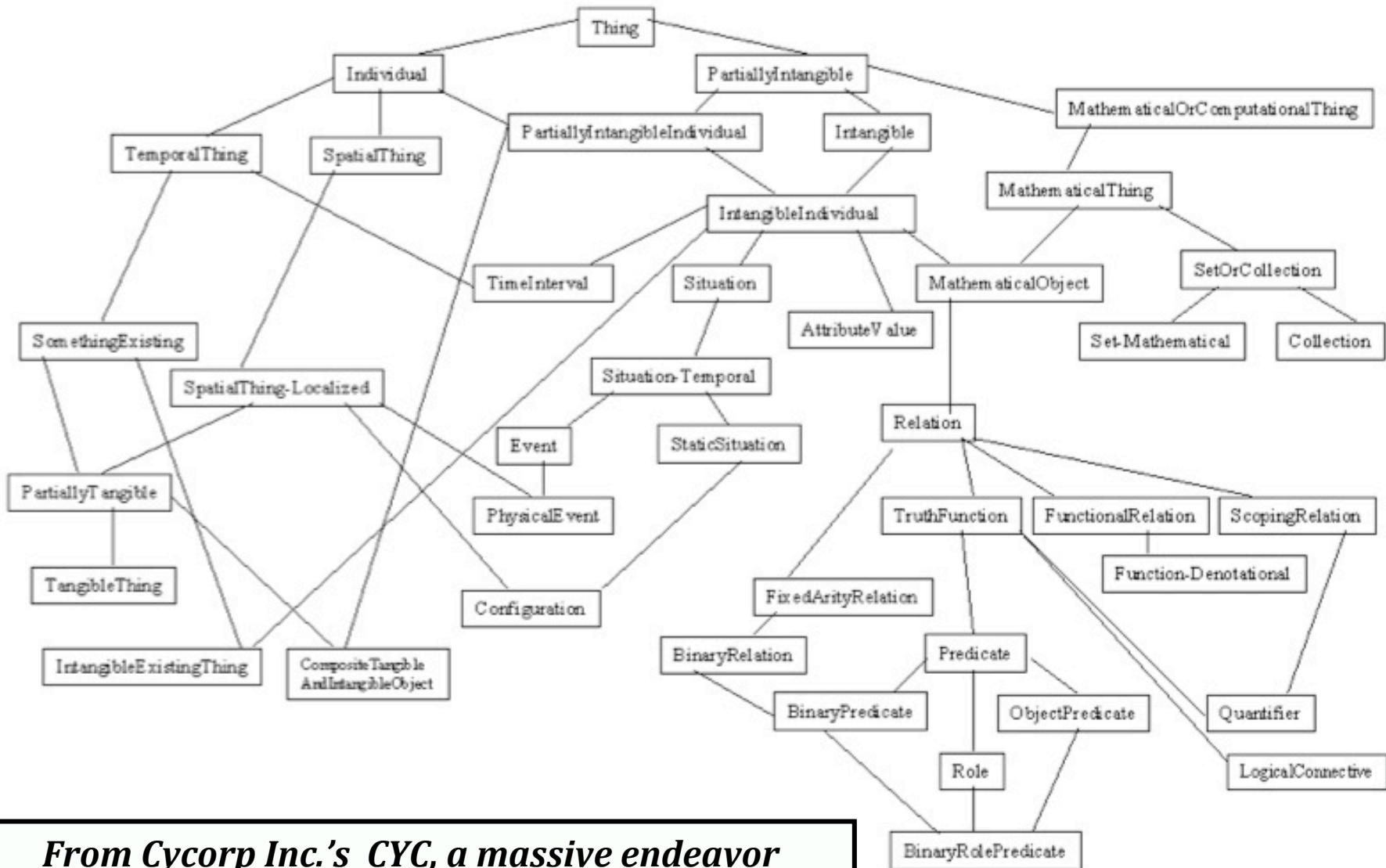
Vertical / Convergent Reasoning: WordNet



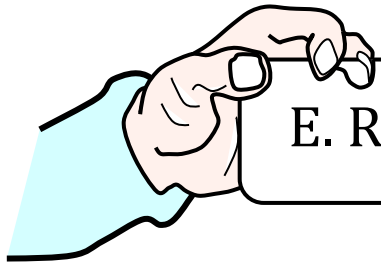
Words are grouped into **Synsets** (sets of synonyms) that *uniquely denote* senses

Synsets (**denoting senses**) are then structured via IS-A, Part-Whole relations

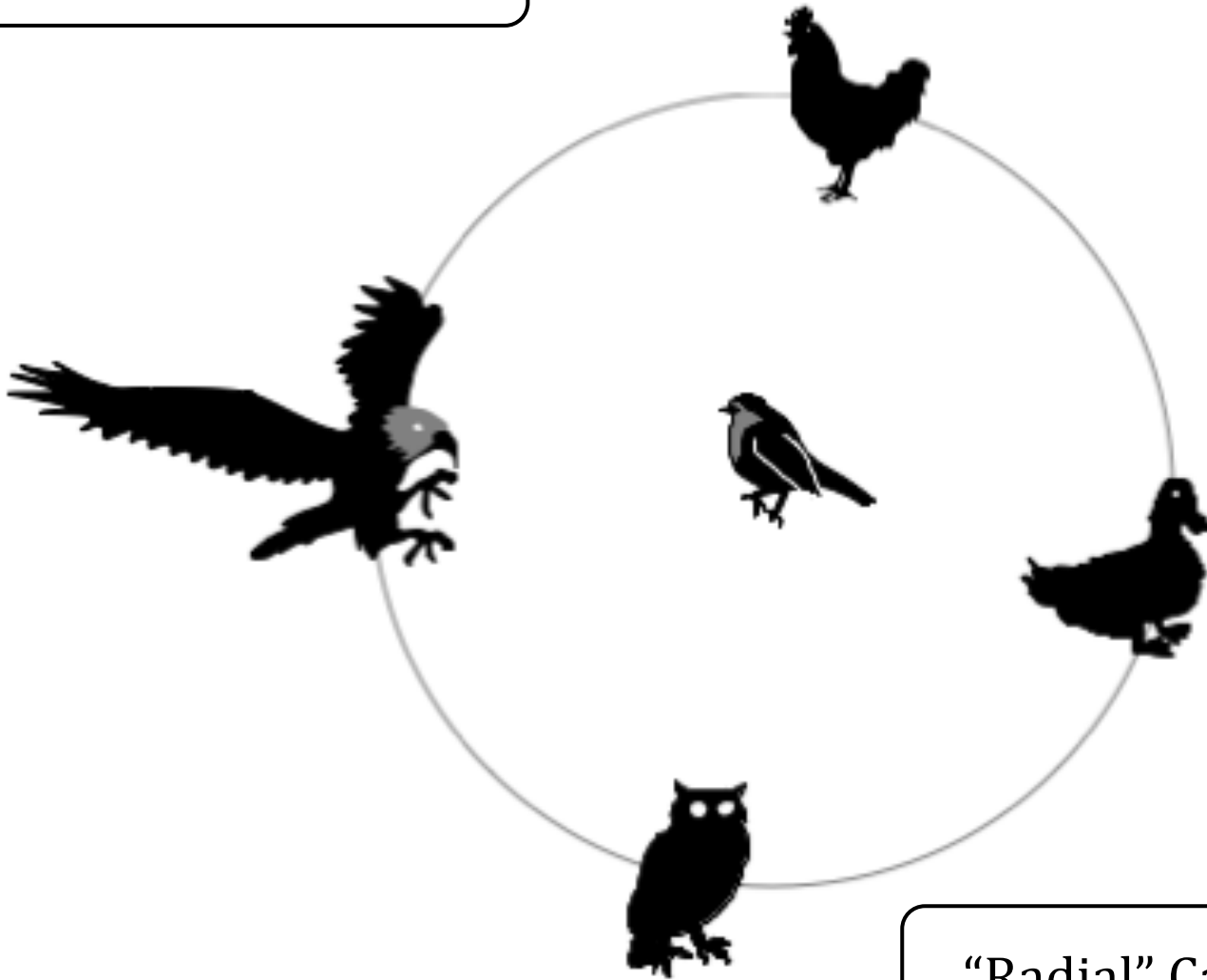
Upper Ontologies: The CYC Upper-Model



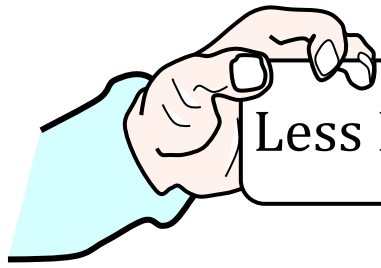
From Cycorp Inc.'s CYC, a massive endeavor



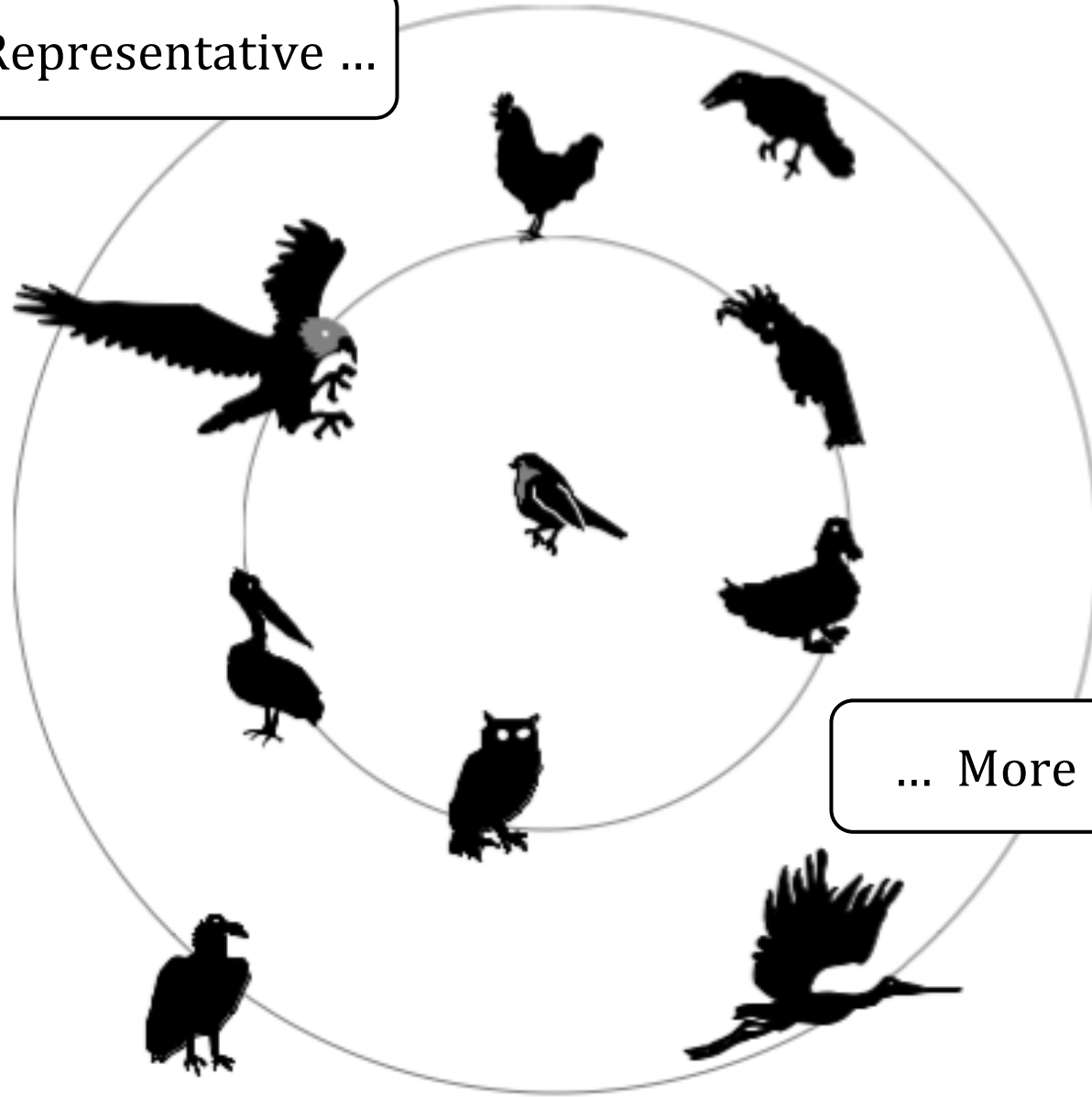
E. Rosch, G. Lakoff



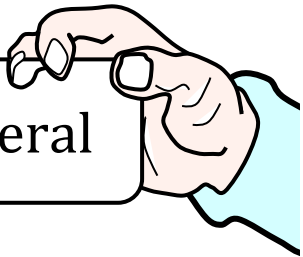
“Radial” Categories



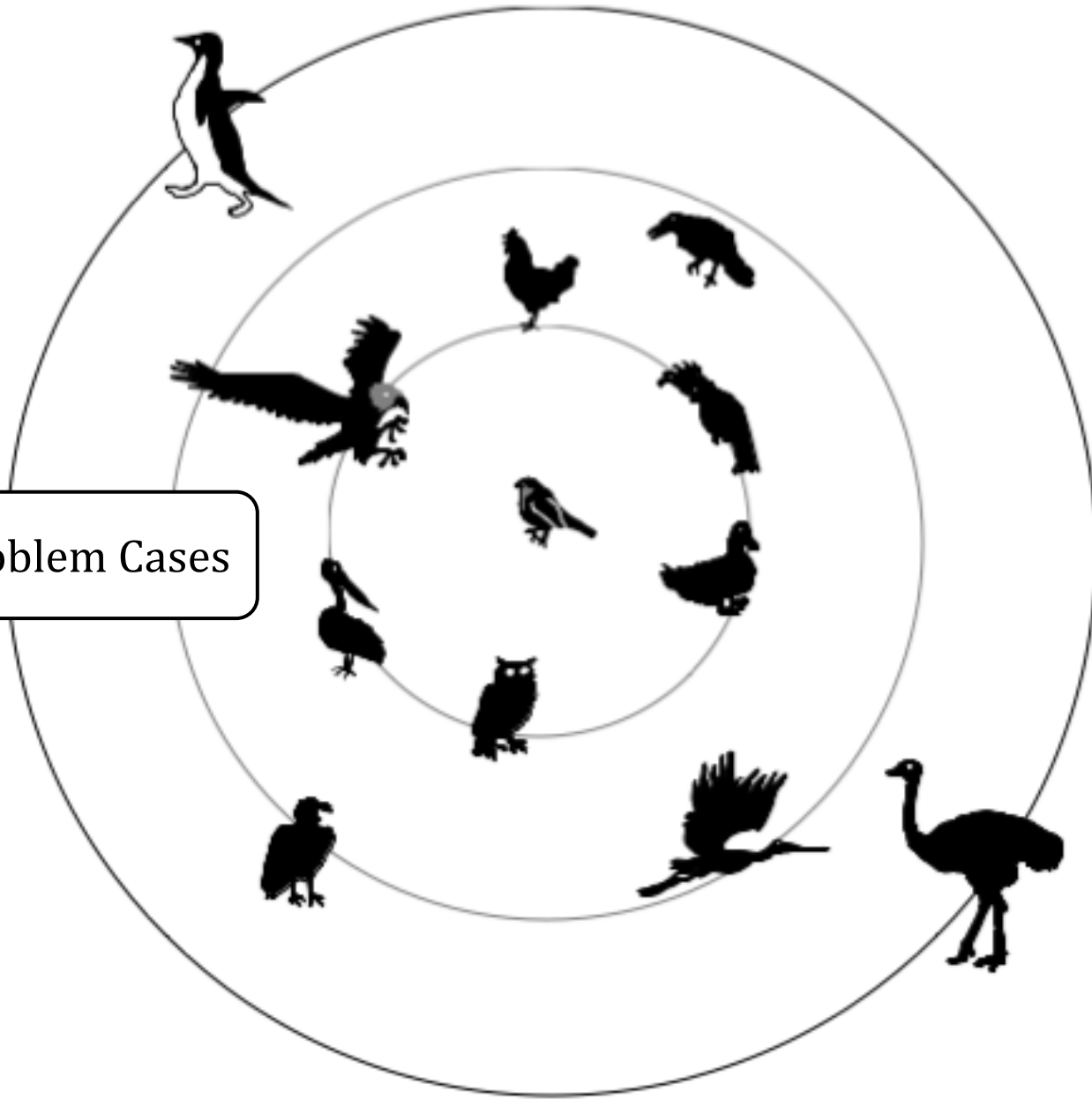
Less Representative ...

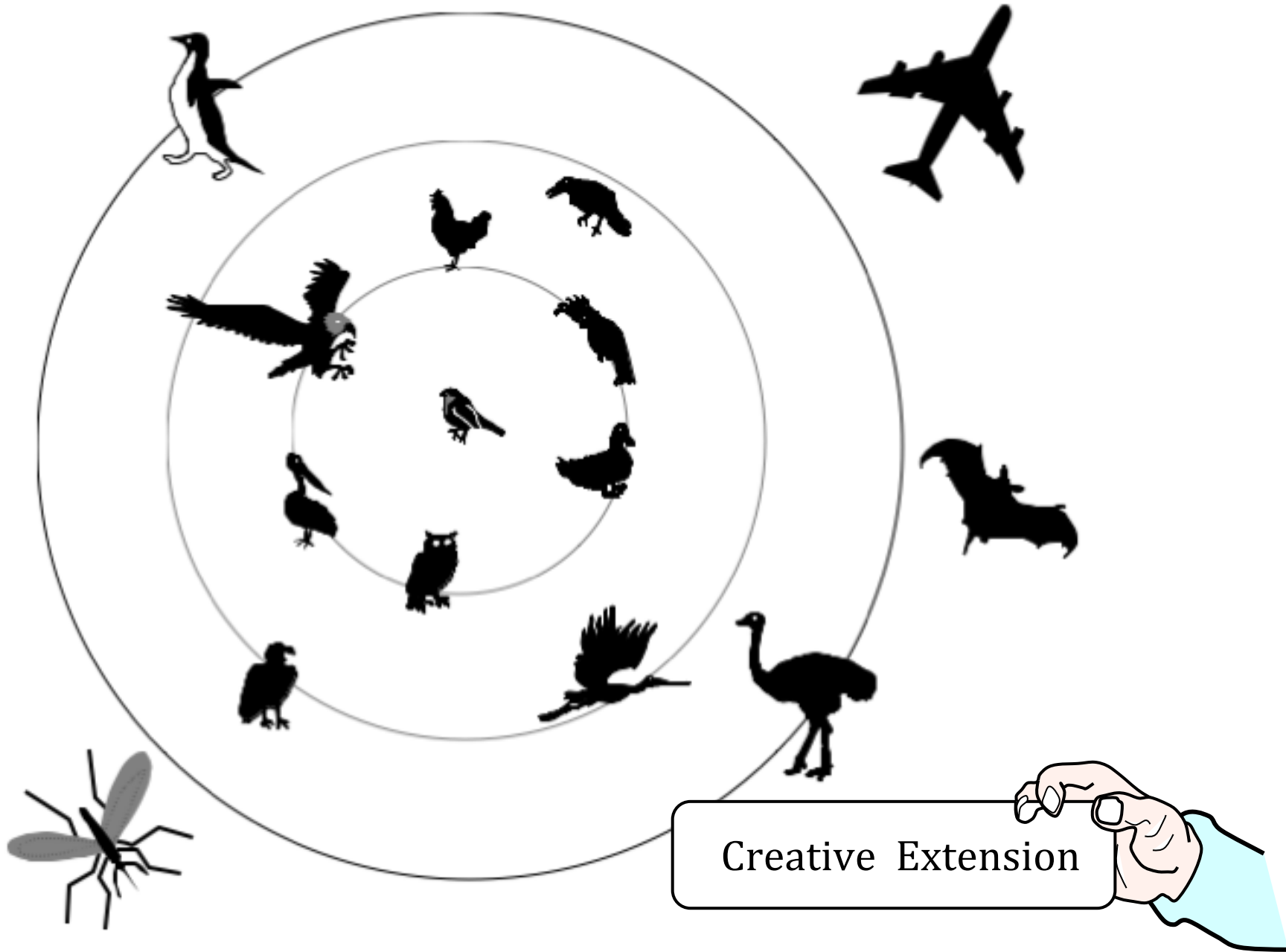


... More Peripheral



Atypical Problem Cases





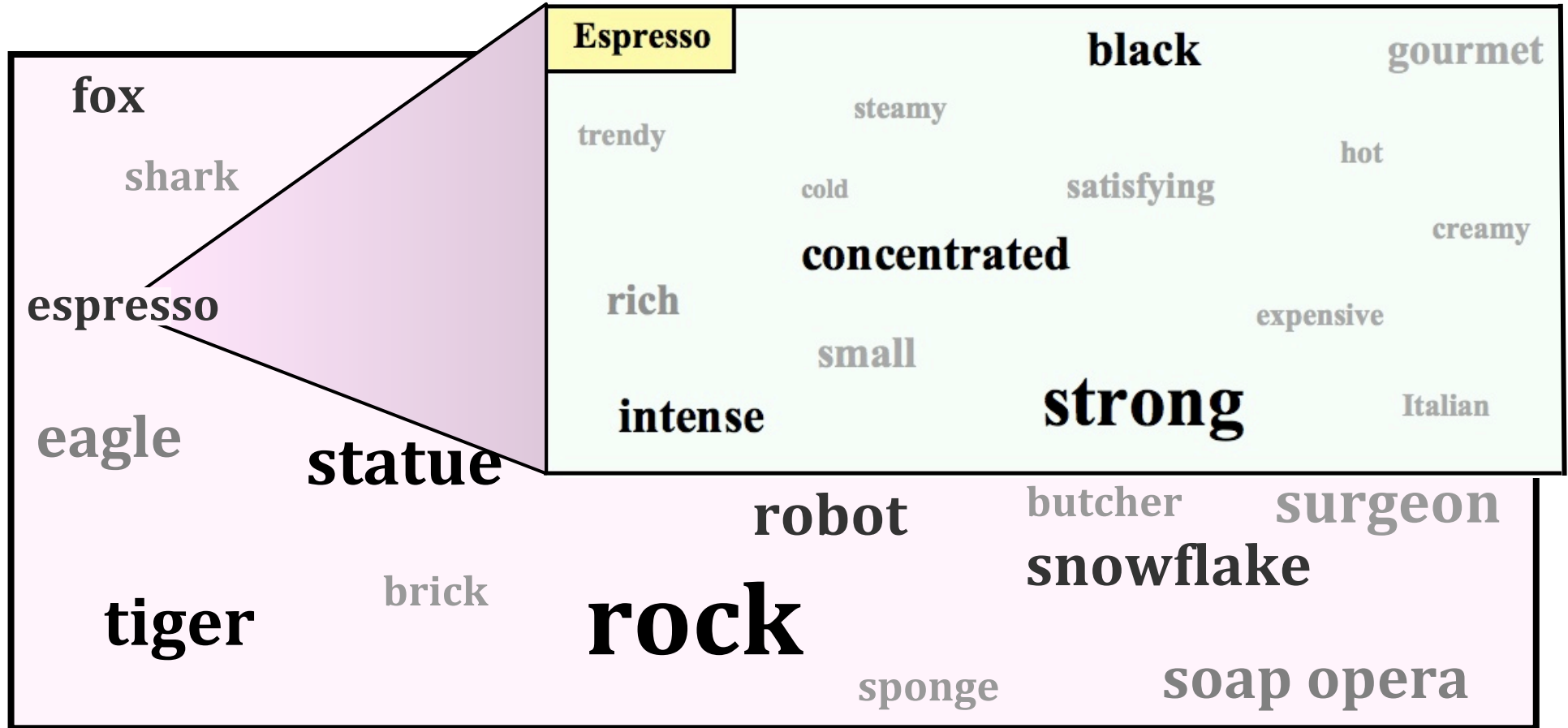
Creative Extension



creativity

subjective:skill, personal:skill, professionalism, subjective:measure, personal:attribute,
basic:skill, essential:skill, soft:skill, professional:attribute, mental:ability, humanistic:attribute, spiritual:attribute,
natural:attribute, entrepreneurial:skill, academic:ability, subjective:thing, musical:ability, important:attribute, key:skill,
psychological:attribute, individual:skill, natural:ability, personal:motive, cognitive:skill, nonverbal:skill,
important:skill, social:attribute, psychological:attitude, desirable:attribute, entrepreneurial:attribute, abstract:skill,
intellectual:ability, social:skill,
diverse:attribute, cognitive:power, commercial:skill, technical:skill, mental:faculty, intellectual:skill, noble:attribute, artistic:skill, athletic:skill, feminine:attribute, feminine:skill,
wonderful:thing, behavioural:skill, individual:motivation, mental:attribute, positive:attribute, vital:skill, interpersonal:skill,
professional:skill, valuable:skill, positive:attitude, individual:attribute, spiritual:power, essential:attribute, cognitive:ability,
mental:skill, educational:skill,

Acquiring **Stereotypical Ideas**: Harvest *Simile* patterns from Web



Use Web query pattern

“ as * a | an as * ”

to harvest 10,000s of similes



from Google 3-grams ...



decaying **like** a zombie
(70 Google Web hits)



brainwashed **like** zombie
(451 Google Web hits)

a	devastating	zombie
a	shambling	zombie
	rotting	zombie
	ravaged	zombie
a	brainwashed	zombie
a	drooling	zombie
a	freaking	
an	attacking	
a	crazed	zombie
an	obsessed	zombie
an	infected	zombie
	marauding	zombie
	disturbed	zombie
	reanimated	zombie
a	flying	zombie
a	flaming	zombie
a	decaying	zombie
an	unexpected	zombie

Baby (163 properties)

{*delicate, squalling, weeping, baptized, adopted, startled, attentive, blessed, teeny, rocked, adorable, whining, bundled, toothless, placid, expected, rescued, treasured, new, sleepy, indulged, slumbering, weaned, supple, helpless, small, sleeping, mute, animated, vulnerable, wailing, cradled, kicking, soft, rested, bellowing, blameless, grinning, screaming, tiny, cherished, reliant, thriving, loveable, guileless, harmless, inexperienced, dribbling, unthreatening, nursed, angelic, bawling, beaming, naked, spoiled, scared, weak, squirming, blubbering, contented, smiling, wiggling, mewling, blubbing, sniffing, overtired, dimpled, loving, dear, tired, powerless, bewildered, peaceful, distressed, naive, wee, soiled, sucking, fussy, gurgling, vaccinated, pure, heartwarming, pouting, constipated, drooling, quiet, wiggly, lovable, bare, weaning, suckling, cute, bald, whimpering, tender, pampered, incontinent, fleshy, charming, dependent, artless, fussing, flabby, babbling, warm, giddy, crawling, snoozing, hairless, cuddled, sweet, sobbing, squealing, wrapped, cooing, swaddled, laughing, toddling, fragile, innocent, moaning, gentle, terrified, precious, cranky, giggling, confused, cuddly, fat, ignorant, snoring, young, howling, screeching, shrieking, trusting, shivering, napping, resting, frightened, fresh, loved, demanding, chubby, adored, appealing, happy, tame, relaxed, wriggly, rocking, wriggling, conceived, clean, content, smooth, crying, submissive, bumbling, pink, sniveling, orphaned*}

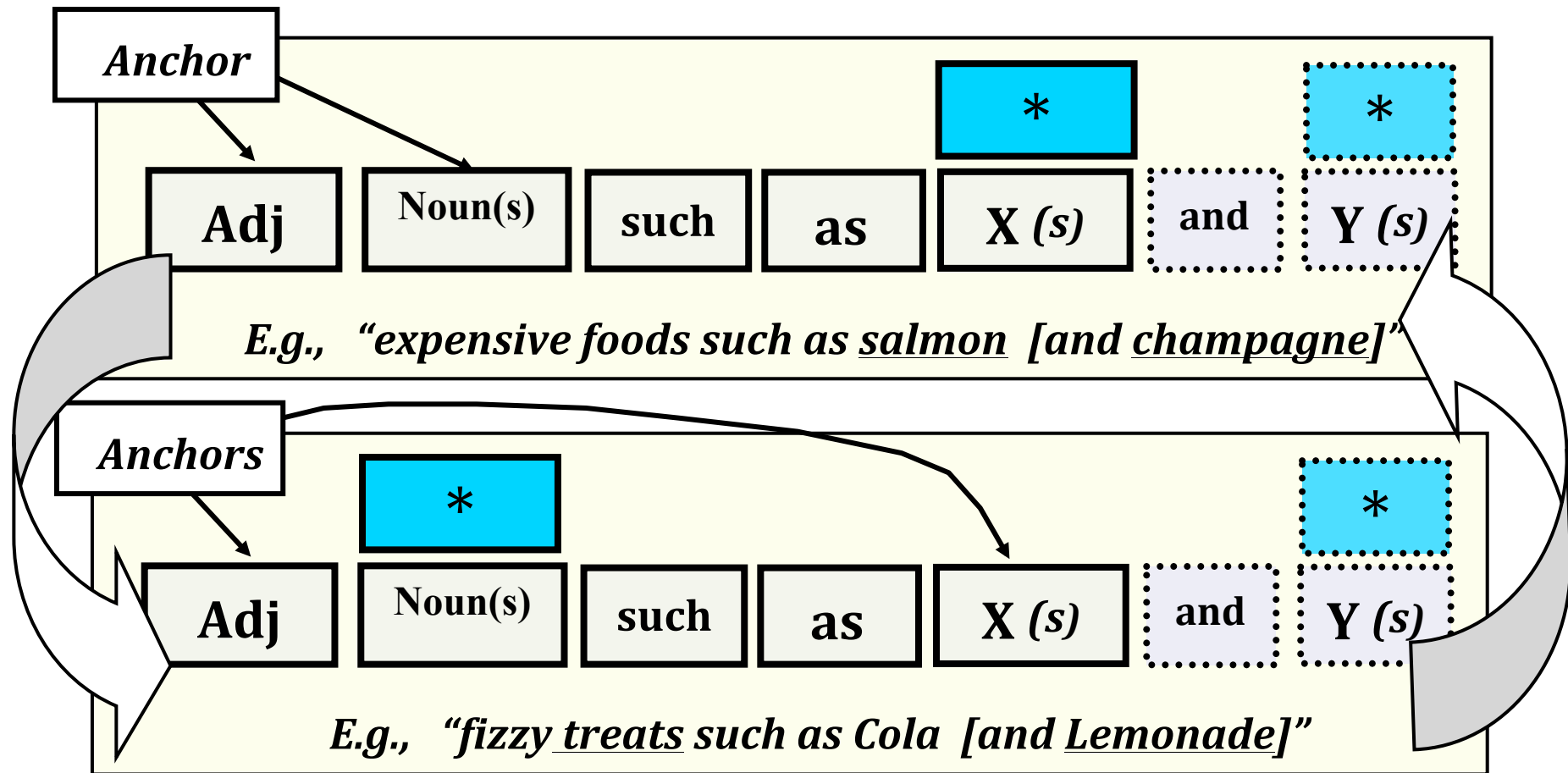
Overall: 9479 Stereotypes, 7898 properties, 75,000 mappings

A Taxonomy as A Pool of Triples: How to obtain the largest Pool?



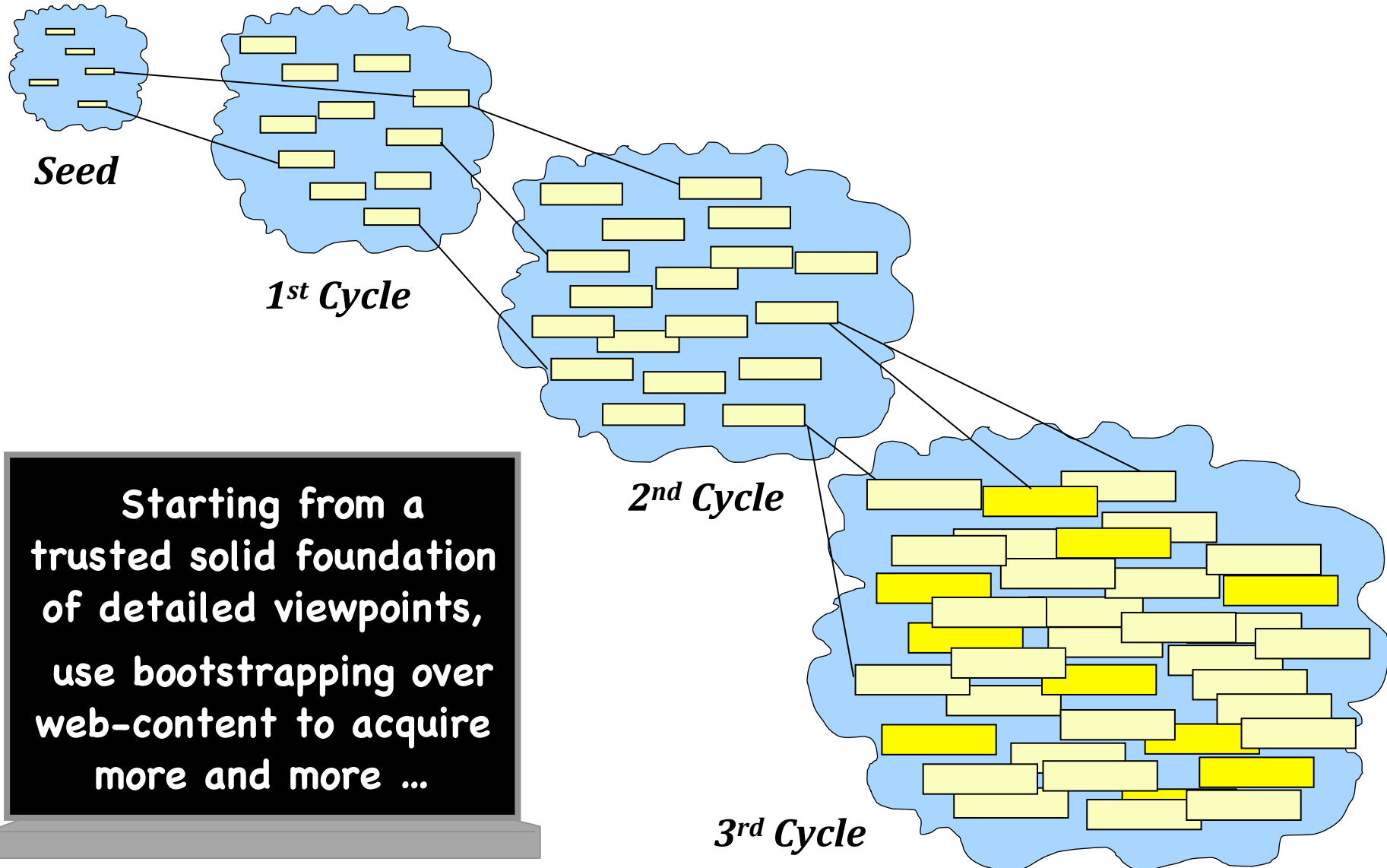
Collecting Diverse Fine-Grained Perspectives via Lateral Thinking

Acquiring informative fine-grained categorizations such as:
cola is a carbonated drink or *tofu is a boring food*



Bootstrapping on web is a controlled form of Divergent / Lateral Thinking

Lateral Thinking **on the Web**: Divergent Growth of Knowledge



Seed # 1 (of 3) : WordNet Glosses

Shallow parse the textual glosses associated with individual WordNet senses



<espresso, black, coffee>

<espresso, strong, coffee>



E.g., Espresso “strong black coffee brewed by forcing stream through ...”

Seed # 2 (of 3) : ConceptNet Propositions (IS-A)

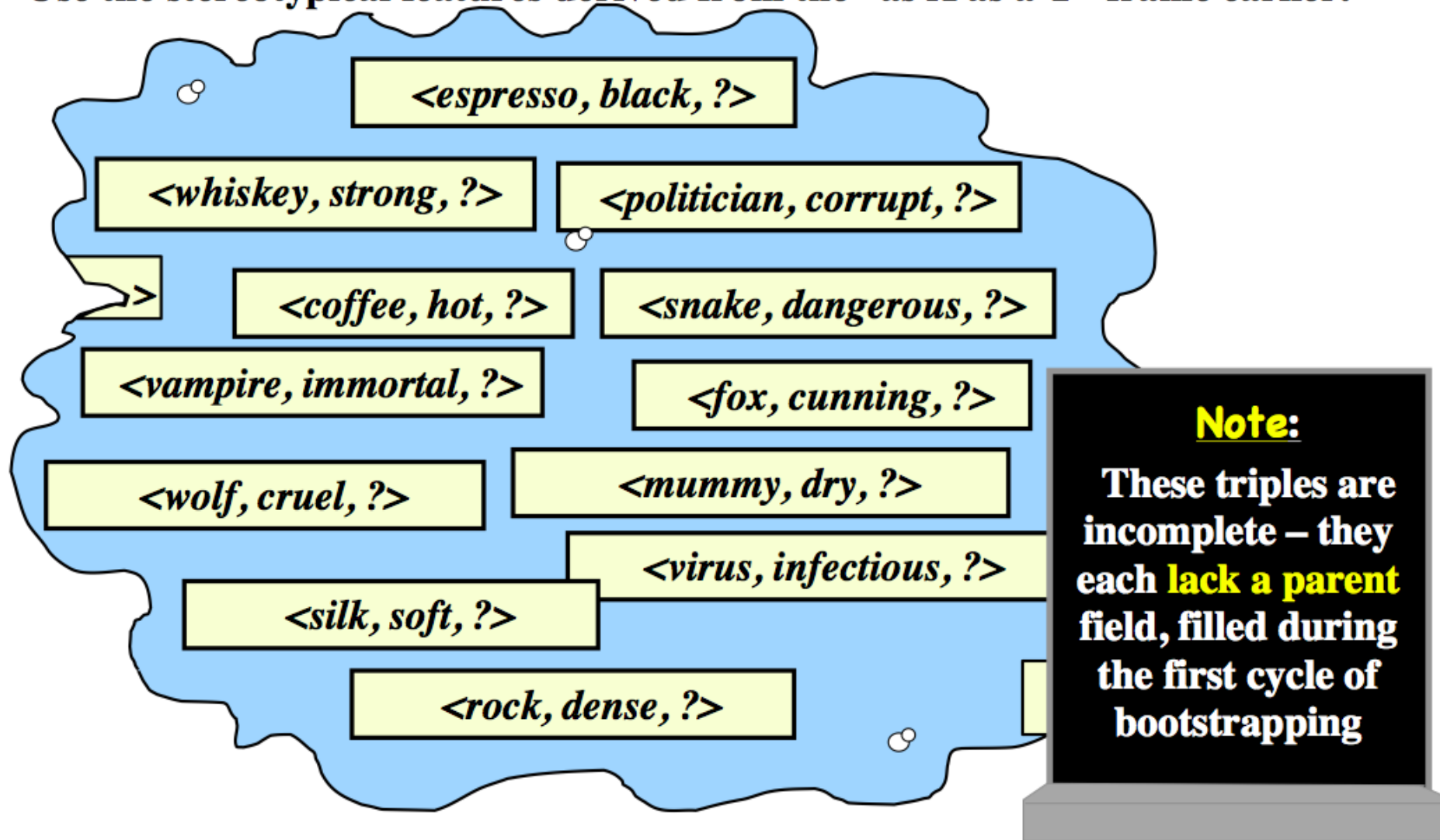
Filter ConceptNet IS-A propositions to obtain only the most plausible ones

- (IsA "bagdad" "capital")
- (IsA "bagel" "bread")
- (IsA "bagel" "breakfast food")**
- (IsA "bagel" "chewy kind")
- (IsA "bagel" "doughnut")
- (IsA "bagel" "food")
- (IsA "bagel" "good food")**
- (IsA "bagel" "pastry")
- (IsA "bagel" "roll")
- (IsA "bagel" "round bread")**
- (IsA "bagel" "torus")
- (IsA "bagpipes" "musical instrument")**
- (IsA "bagpipes" "scottish instrument")**
- (IsA "bagpipes" "traditional irish")
- (IsA "bahrain" "island")
- (IsA "bald eagle" "national bird")**
- (IsA "bald eagle" "national symbol")
- (IsA "bald eagle" "national emblem")
- (IsA "bald eagle" "rare bird")**
- (IsA "ballpoint pen" "english channel")
- (IsA "bambi" "cute character")**
- (IsA "bambi" "ditzy name")
- (IsA "bambi" "pejorative name")
- (IsA "balloon" "rubber sack")
- (IsA "balloon" "expensive sport")
- (IsA "banana" "yellow fruit")**
- (IsA "banjo" "stringed instrument")**
- (IsA "baseball bat" "long round")
- (IsA "barn" "large structure")**
- (IsA "baseball" "american tradition")

Find triples with Adj-Noun genus where Wordnet agrees with Noun part

Seed # 3 (of 3) : Simile-derived Associations

Use the stereotypical features derived from the “as X as a Y” frame earlier:



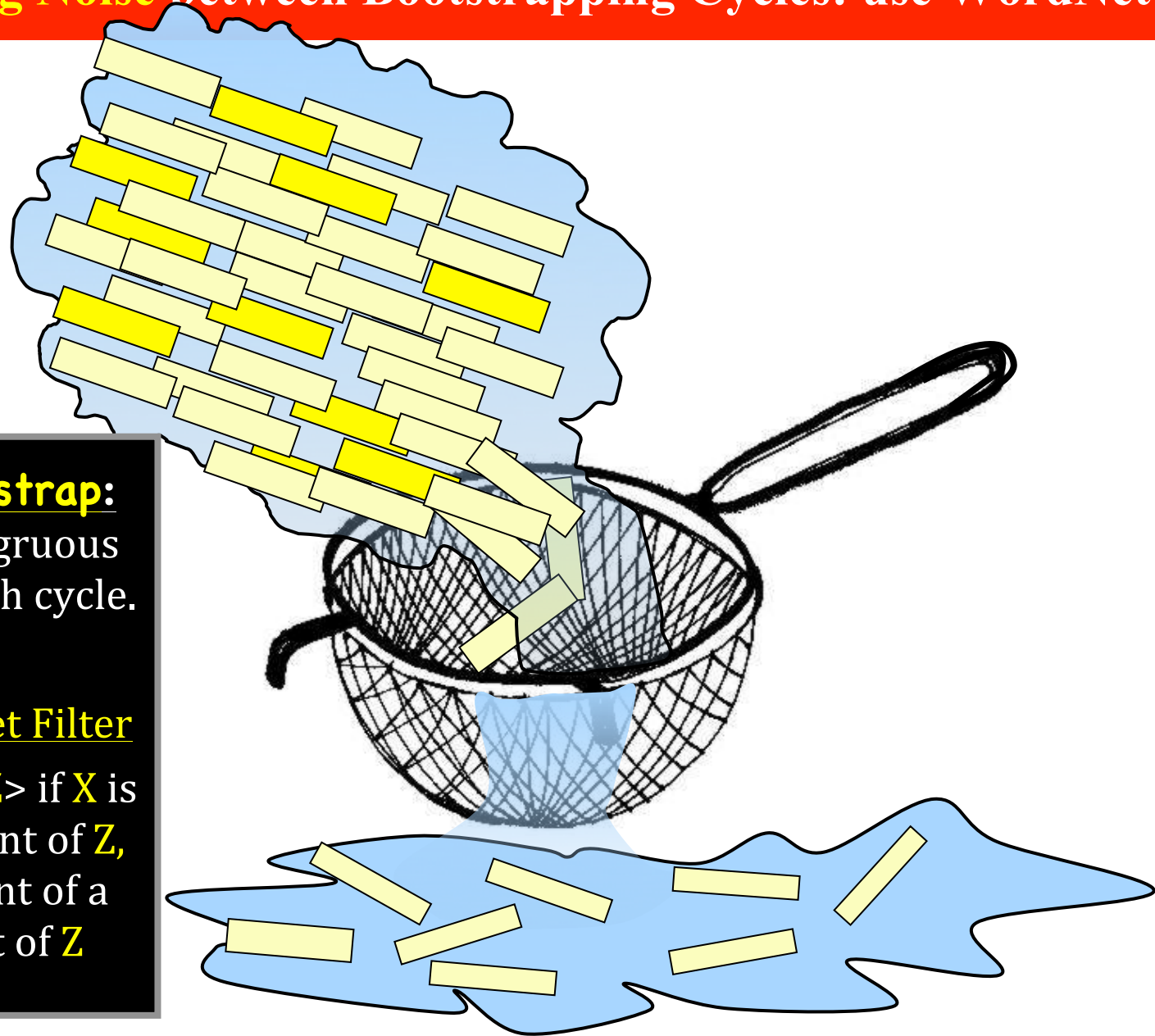
Removing Noise between Bootstrapping Cycles: use WordNet

Filtered Bootstrap:

Remove incongruous triples after each cycle.

Coarse WordNet Filter

Remove $\langle X, Y, Z \rangle$ if X is not a descendant of Z , or a descendant of a direct parent of Z



Comparing our Three Seeds: Size and Coverage



WordNet

triples: 51,314
terms: 12,227
features: 2,305



ConceptNet

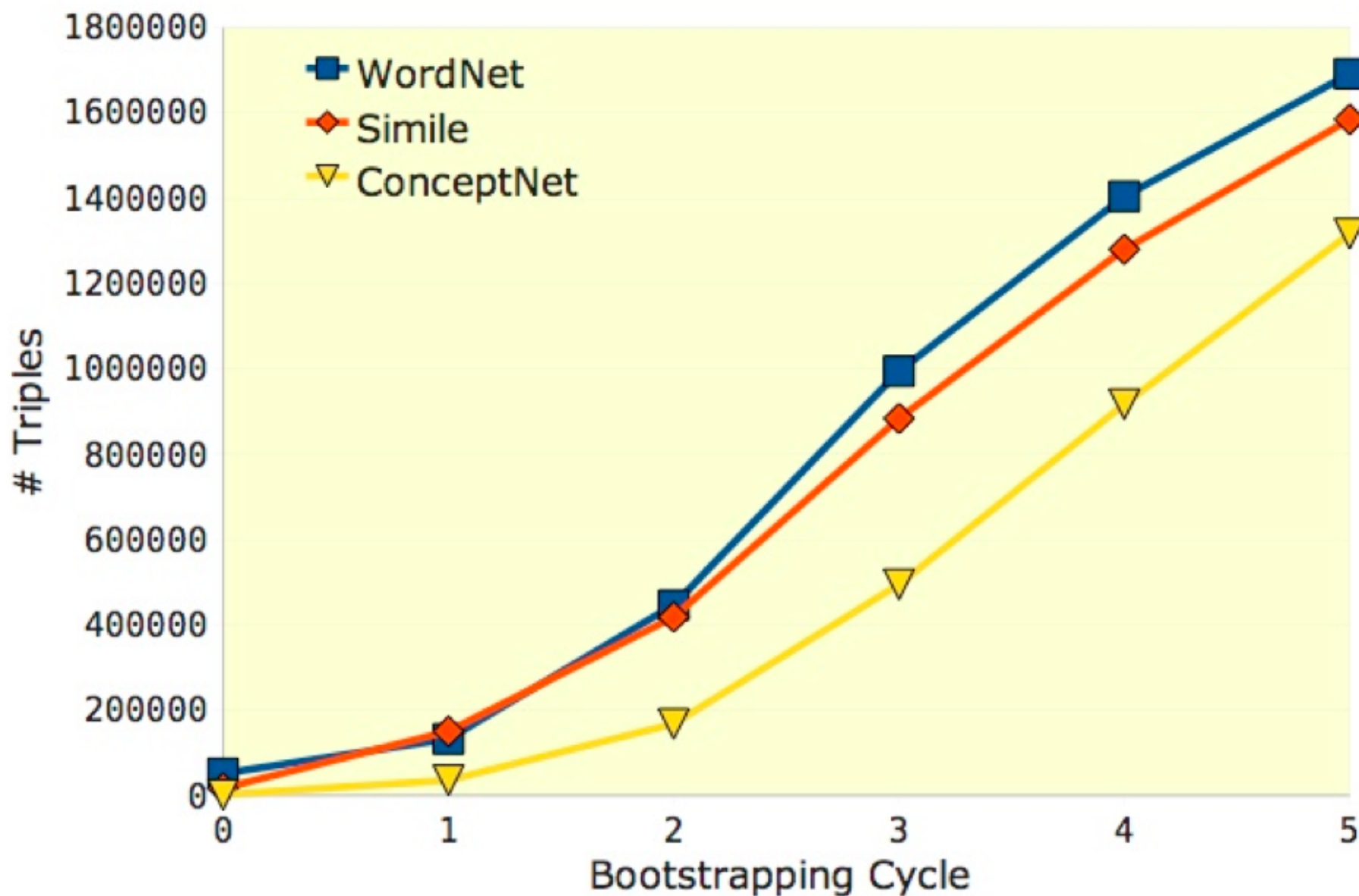
triples: 1,808
terms: 1,133
features: 550



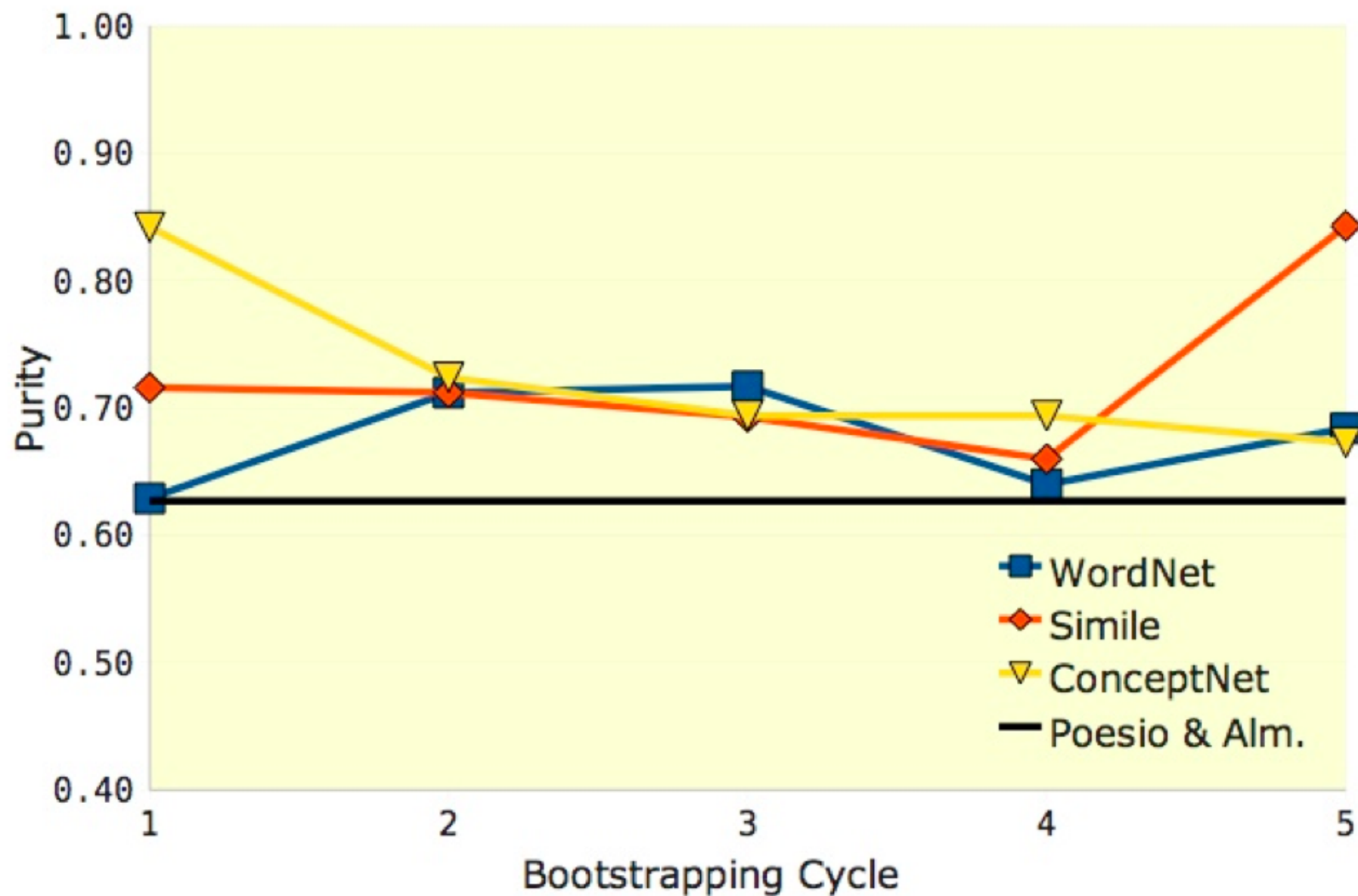
Similes

triples: 16,688
terms: 6,512
features: 1,172

Bootstrapping Results: Growth of Structure over 5 Cycles



Bootstrapping Results: Change in Precision over 5 Cycles





Welcome to Thesaurus Rex!

[Go Back](#) [See XML](#) => [See Disambiguation Page](#)

Top-Ranked Fine-Grained Categories of *cola*: =>see poetic categories

colored:beverage, common:food, nice:thing, **soft:drink,** carbonated:liquid, cool:beverage, **acidic:beverage,** **dark:beverage,**
cold:soda, **acidic:drink,** **dark:liquid,** **dark:drink,** colored:drink, cold:drink,
coloured:beverage, dark:substance, **stimulating:drink,**
dark:food, **sugary:food,** **carbonated:drink,** **simple:mixer,** **fizzy:drink,**
sugary:item, clear:liquid, colored:fluid, clear:beverage, **sugary:drink,** clear:drink, **acidic:liquid,**
carbonated:beverage,
colored:soda, coloured:drink, cold:beverage, traditional:beverage, nutritional:value, **colored:substance,** **carbonated:soda,** **dark:soda,**
popular:beverage, acidic:fluid, sensitive:beverage, prepared:drink,
acidic:material, sweet:drink, nonalcoholic:drink, **salty:fluid,** **stimulating:beverage,**
refreshing:beverage, forbidden:food, **colored:liquid,** common:item, traditional:drink, **dark:mixer,** mixed:drink,
acidic:food,
effervescent:beverage, sugary:liquid, common:drink, common:beverage, carbonated:mixer, industrial:drink,

Shared Category for *coffee* and *juice*

acidic_beverage, acidic_fluid, **acidic_food,** acidic_level,
acidic_substance,
additional_beverage, adult_food, **alcoholic_beverage,** american_food, aqueous_beverage, aqueous_medium, **basic_foodstuff,**
basic_item, basic_provisions, carbonated_beverage, **clear_fluid,** **cold_beverage,**
cold_item, **colored_beverage,** colored_fluid, commercial_beverage, common_food,
common_item, complimentary_beverage,
consumable_beverage,
convenient_item, conventional_beverage, cool_food, delicate_food, delicious_food, **desirable_food,** diverse_food,
dried_food,
essential_provisions, **everyday_beverage,** **everyday_item,** **favorite_beverage,**
favorite_food, favorite_item, favourite_beverage, **fluid_food,** forgotten_item, **free_beverage,** hazardous_food,
healthy_beverage, **healthy_food,** **hot_beverage,** hot_item, important_food, inexpensive_item,
irritating_food, key_food,
larger_quantity, leafy_substance, **light_food,** liquid_beverage, **liquid_food,** liquid_form, liquid_material,
liquid_system,
minute_item, natural_beverage, necessary_item, neutral_beverage, new_beverage, non-alcoholic_beverage,
nonalcoholic_beverage,

Shared Category for *coffee* and *tobacco*

acidic_food, addictive_substance, agricultural_item, agricultural_material, agricultural_plant, american_item, aversive_substance, bad_thing, basic_material, benign_substance, commercial_plant, common_substance, complex_mixture, consumable_item, consumable_substance, dark_substance, dietary_substance, diverse_material, dry_item, dry_material, economic_plant, essential_item, everyday_item, everyday_substance, exempt_item, exotic_food, exotic_item, exotic_plant, external_agent, external_substance, familiar_substance, favorite_item, foreign_material, foreign_substance, harmful_substance, homogenous_material, illegal_substance, important_item, important_plant, industrial_plant, intoxicating_substance, irritating_substance, leafy_substance, legal_intoxicant, legal_substance, legal_thing, medicinal_plant, medicinal_substance, narcotic_substance, natural_material, natural_substance, necessary_item, new_substance, organic_material, organic_matter, outside_agent, particulate_material, perishable_item, personal_item, pleasurable_substance, poisonous_plant, poisonous_substance, psychoactive_plant, psychoactive_substance, raw_material, recreational_substance,

<i>Word Pair</i>	<i>Miller & Charles</i>	<i>T-Rex (all cats)</i>	<i>T-Rex (popular cats)</i>
<i>car - automobile</i>	3.92	1	1
<i>gem - jewel</i>	3.84	1	1
<i>journey - voyage</i>	3.84	0.847	0.847
<i>boy - lad</i>	3.76	0.801	0.801
<i>coast - shore</i>	3.7	0.829	0.831
<i>asylum - madhouse</i>	3.61	0.888	0.893
<i>magician - wizard</i>	3.5	1	1
<i>midday - noon</i>	3.42	1	1
<i>furnace - stove</i>	3.11	0.660	0.665
<i>food - fruit</i>	3.08	0.590	0.609
<i>bird - cock</i>	3.05	0.728	0.728
<i>bird - crane</i>	2.97	0.626	0.626
<i>tool - implement</i>	2.95	0.962	0.963
<i>brother - monk</i>	2.82	0.838	0.841
<i>crane - implement</i>	1.68	0.344	0.344
<i>lad - brother</i>	1.66	0.260	0.260
<i>journey - car</i>	1.16	0.014	0.014
<i>monk - oracle</i>	1.1	0.176	0.176
<i>food - rooster</i>	0.89	0.046	0.046
<i>coast - hill</i>	0.87	0.596	0.386
<i>forest - graveyard</i>	0.84	0.219	0.229
<i>monk - slave</i>	0.55	0.284	0.284
<i>coast - forest</i>	0.42	0.521	0.279
<i>lad - wizard</i>	0.42	0.293	0.293
<i>chord - smile</i>	0.13	0.078	0.078
<i>glass - magician</i>	0.11	0.117	0.117
<i>rooster - voyage</i>	0.08	0.035	0.008
<i>noon - string</i>	0.08	0.008	0.035
Correlation	1	0.895	0.932

adverse_event, bad_event, bad_thing, catastrophic_event, changing_event, charged_event,
critical_event, destructive_thing,
devastating_event, disruptive_event, distressing_event, domestic_conflict, domestic_event,
dramatic_event,
economic_event, emotional_event, environmental_event, experienced_event, external_event,
extraordinary_event, financial_event, identifiable_event, immoral_act, important_event, intense_event,
legal_event,
major_conflict, major_event, negative_event, ordinary_event, outside_event, painful_event, past_event,
rare_event, recent_event, severe_conflict,
severe_event,
significant_event, single_event, social_event, social_occurrence, specified_event, stressful_event,
sudden_event, surrounding_event, traumatic_event, unanticipated_event, unavoidable_event,
uncontrollable_event, undesirable_event,
unexpected_event, unexpected_occurrence, unforeseeable_event,
unforeseen_event, unfortunate_event,
unpleasant_event,
unpleasant_thing, untoward_event, unusual_event,



Questions
?