Distributional Profiles

Wikipedia

Membership Functions

Acquiring Naturalistic Concept Descriptions from the Web

Collocations

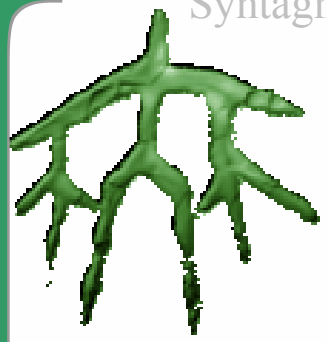
Corpora

BNC

Syntagmatics

Co-occurrence types

Lex-Ecology



WordNet

Norms

Exploitations

Clusters

Radia

Tony Veale, Yanfen Hao

School of Computer Science,

{Tony.Veale, Yanfen.Hao}@UCD.ie

UCD Creative Language Systems Group

Introduction

Concept representation is the building block in the knowledge representation.

Figurative language processing demands more naturalistic descriptions (common sense knowledge) to refer concepts.
E.g., metaphor generation, metaphor comprehension.

Do lexicons have enough common sense knowledge?
E.g., In WN, surgeon is "a physician who specializes in surgery".
Hypernyms: Surgeon is kind of doctor. Hyponyms: Neurosurgeon is kind of surgeon.
Synonyms: {operating surgeon, sawbones}.

Our common sense tells us: Surgeons are delicate, skilled, accurate and careful.
Supermodels are poised, skinny, lithe, pretty, and graceful.

How to find these common sense knowledge?

Using Similes to Identify Stereotypical Cultural Associations

- **Similes / Comparisons reveal the most diagnostic features of a concept**

  E.g., "as hot as the sun", "as dry as sand", "as wobbly as jelly", "as sweet as pie"

- **The most frequent similes characterize the most pivotal concepts / senses**

  E.g., animal concepts ("lion", "rat", etc.) are frequently used in comparisons

- **Unlike metaphors, similes have a standard, recognizable syntactic frame**

  "as barren as a desert", "as delicate as a surgeon", "as stiff as a corpse"

- **Detailed Knowledge-Representations can be gathered for individual concepts**

  Example: surgeon = {delicate, sensitive, skilled, clinical, professional, …}

Sampling Comparisons/Similes from the WWW

Query-pattern #1: *"as ADJ as a|an *"* for all antonymous adjectives in WN

Query-pattern #2: *"as * as a|an NOUN"* for all nouns gathered with query #1

- **200 sampled snippets per query, to give 74,704 apparent simile instances**

42,618 unique simile types, linking 3769 adjectives to 9287 unique nouns

- **Major Issues: Implicit/Local Context, Irony**

"as hairy as a bowling-ball", "as sober as a Kennedy"

- **Annotation:**

12,259 are bona-fide similes and 2796 are ironic similes

Ironic Comparisons/Similes from the WWW

Some Examples:

As {welcome, painless, appealing, pleasant, exciting, entertaining} as a root-canal

As subtle as a {sledgehammer, freight_train, anvil, axe, rhino, toilet_seat, …}

As hefty as a {laptop, croissant}

As blind as a {referee, hawk}

2796 unique adj:noun ironic simile types.
936 adjectives to 1417 nouns.

As {muscular, epicurean, smart, straight, sturdy, weighty, …} as a paper_clip

As rare as a {ham_sandwich, toaster, traffic_jam, monsoon, garbage_pickup}

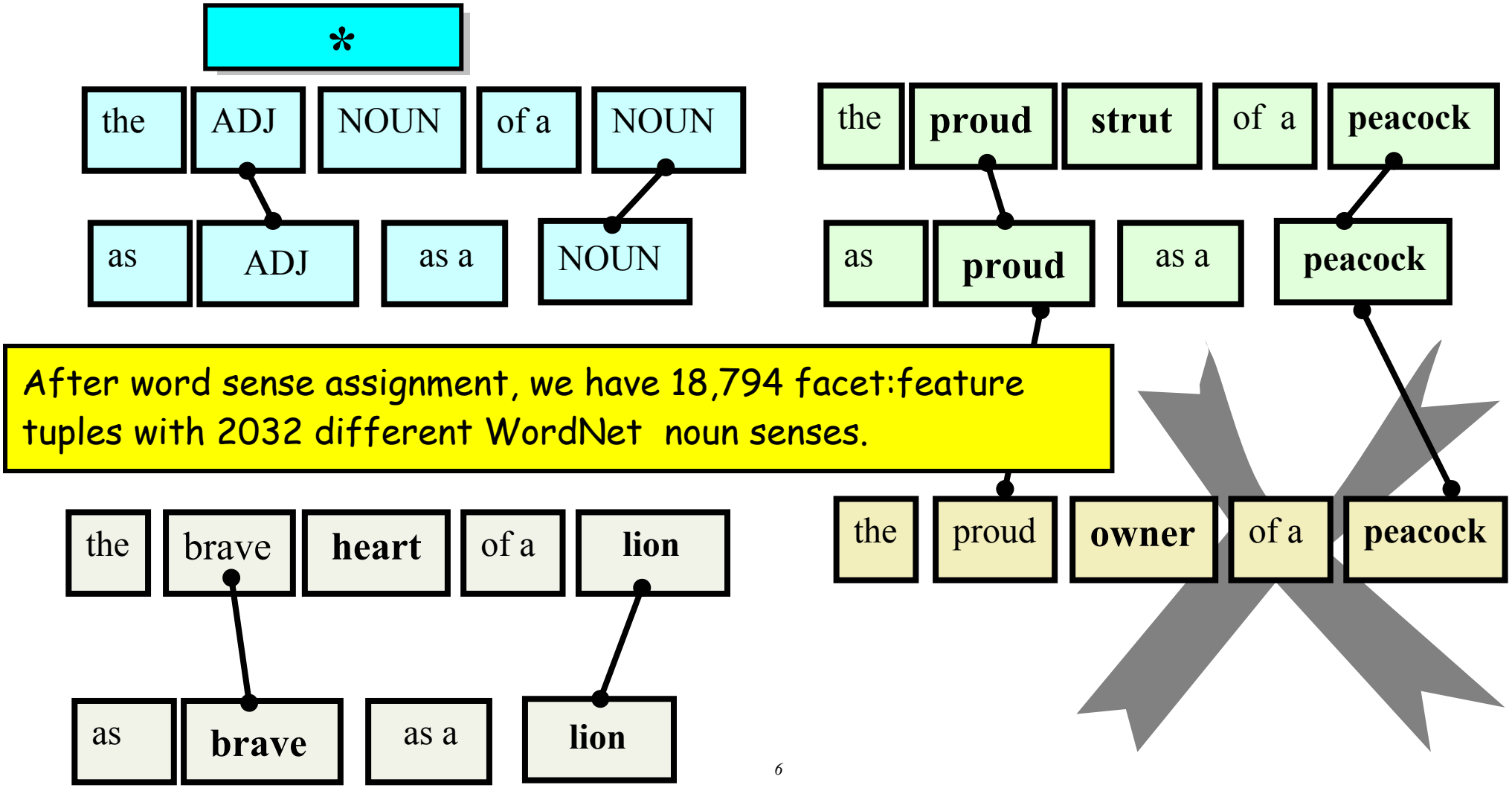
As {bulletproof, scary, subversive} as a sponge_cake

As private as a {park_bench, town_hall, shopping_mall}

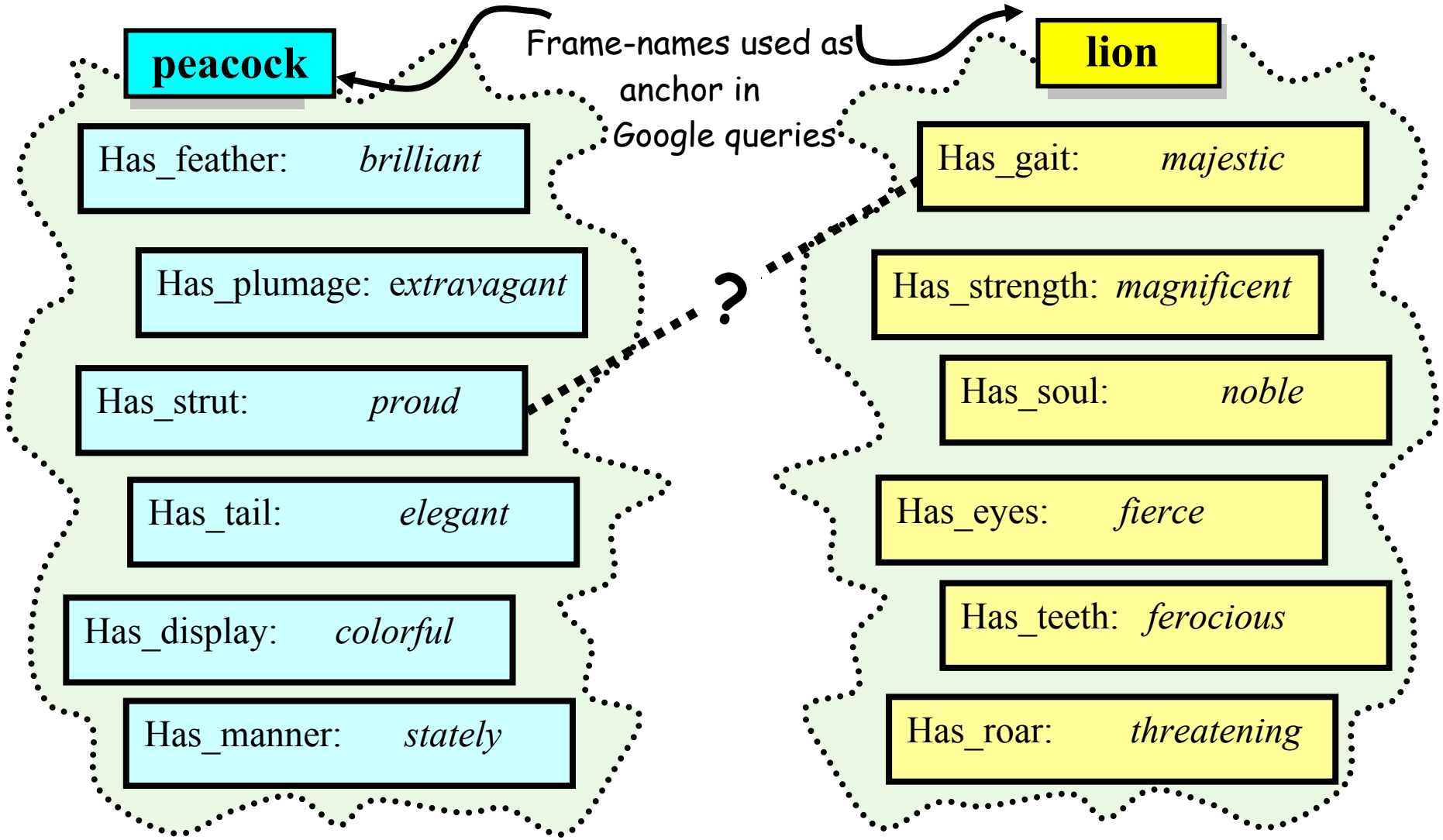
13% of all annotated simile instances. 18% of unique simile types

View on the Web: `http://afflatus.ucd.ie/sardonicus/tree.jsp`

Mining the Web for Conceptual Facets.

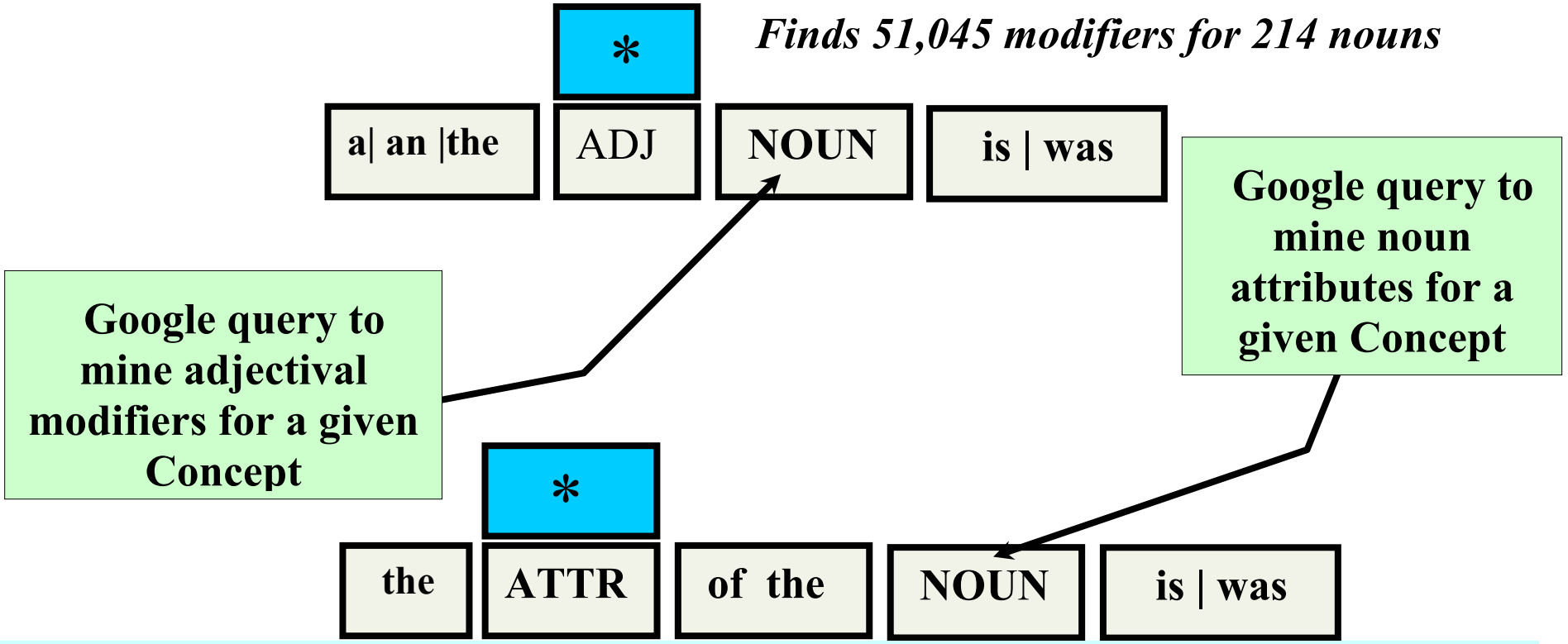


Stereotypical Frames: Web-Derived Attribute-Value Pairings



Emprirical Evaluation

Almuhareb & Poesio (2004, 2005): Web-Mining of Concept Modifiers/Attributes



e.g., rocket = [fast, powerful, speed, thrust, …] **vector space of 59,979 features**

Almuhareb & Poesio (2004) / Veale & Hao (2007): Clustering Results

13-way clustering: [I2=9.58e+001] [214 of 214], Entropy: 0.133, Purity: 0.902

cid	Entpy	Purty	body	crea	dise	fami	vehi	publ	feel	clot	buil	time	anim	frui	furn
0	0.000	1.000	0	0	18	0	0	0	0	0	0	0	0	0	0
1	0.087	0.941	0	0	0	0	0	0	0	0	0	0	1	16	0
2	0.106	0.923	0	1	0	0	0	0	12	0	0	0	0	0	0
3	0.000	1.000	0	13	0	0	0	0	0	0	0	0	0	0	0
4	0.000	1.000	16	0	0	0	0	0	0	0	0	0	0	0	0
5	0.000	1.000	0	0	0	0	0	0	0	0	0	17	0	0	0
6	0.321	0.750	0	1	0	0	12	0	0	2	0	1	0	0	0
7	0.160	0.895	0	0	0	0	1	0	0	0	17	0	0	0	1
8	0.100	0.929	0	1	0	13	0	0	0	0	0	0	0	0	0
9	0.000	1.000	0	0	0	0	0	0	0	12	0	0	0	0	0
10	0.155	0.864	0	0	0	3	0	0	0	0	0	0	19	0	0
11	0.405	0.722	0	0	0	0	1	1	1	0	1	1	0	0	13
12	0.286	0.789	0	1	0	0	0	15	0	2	1	0	0	0	0

Compare 0.855 for Almuhareb & Poesio (2004)

Compare
V+H:
7183 feat.
A+P:
59,979 feat.

Almuhareb & Poesio (2004, 2005) / Veale & Hao (2007, 2008): Total Comparasions of Clustering Results

Table 1: Clustering accuracy for experiment 1 (214 nouns, 13 WordNet semantic classes).

Approach	*Values only*	*Attr's only*	*All (V + A)*
Almu. + Poesio	71.96% (51045 vals)	64.02% (8934 attr)	**85.51%** (59979 v+a)
Naturalistic Descriptions	70.2% (2209 vals)	78.7% (4974 attr)	**90.2%** (7183 v+a)

Table 2: Clustering accuracy for experiment 2 (402 nouns, 21 WordNet semantic classes).

Approach	*Values only*	*Attr's only*	*All (V + A)*
Almu. + Poesio (no filtering)	56.7% (94989 vals)	65.7% (24178 attr)	**67.7%** (119167 v+a)
Almu. + Poesio (with filtering)	62.7% (51345 vals)	70.9% (12345 attr)	66.4% (63690 v+a)
Naturalistic Descriptions	64.3% (5547 vals)	54.7% (3952 attr)	**69.85%** (9499 v+a)

Conclusions

- **Similes provide best clues to naturalistic descriptions of common concepts**

A large case-base of "natural" comparisons is easily acquired from the web

- **Useful for Metaphor/Simile Processing On-Line** **Afflatus.ucd.ie/aristotle**

Generate metaphors for arbitrary target concepts that highlight given features

Thanks