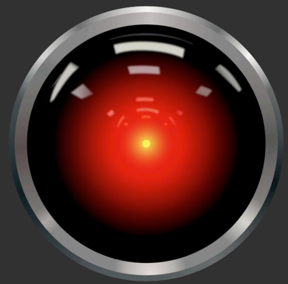
Language, Mind, and Vision

- Learning to Read Deception
- Learning to Describe the Visual World

Yejin Choi
Computer Science & Engineering
UNIVERSITY *of* WASHINGTON

Natural Language Processing (NLP)

- a quick overview



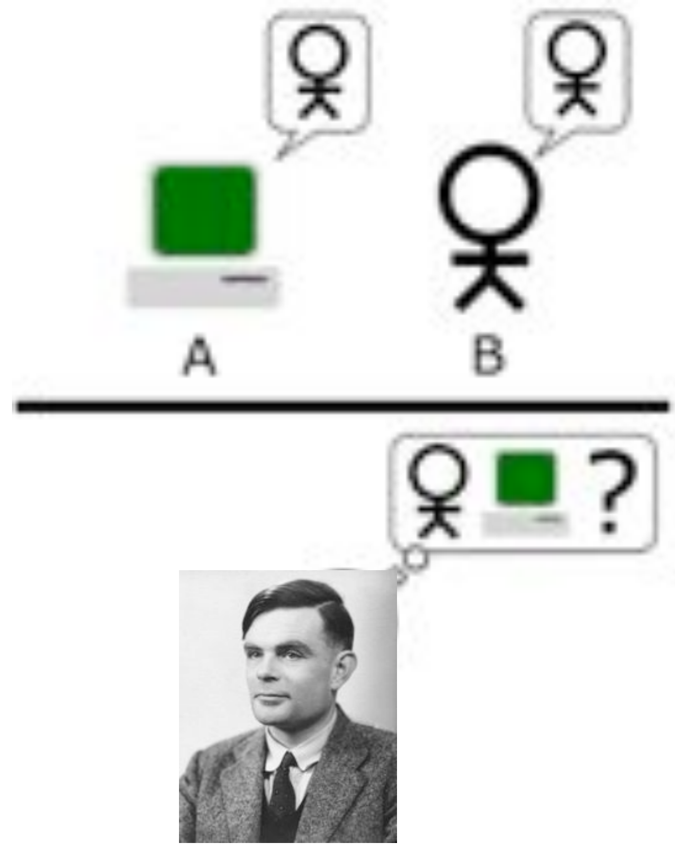
What is NLP?

Fundamental goal: *deep* understand of human language

– Not just string processing or keyword matching!

What is NLP?

- Simple: spelling correction, text categorization…
- Complex: speech recognition, machine translation, dialog interfaces, question answering…
- Unknown: human-level comprehension (is this just NLP?)



Semantic Ambiguity

At last, a computer that understands you like your mother.

- Direct Meanings:
  - It understands you like your mother (does) [presumably well]
  - It understands (that) you like your mother
  - It understands you like (it understands) your mother
- But there are other possibilities, e.g. mother could mean:
  - a woman who has given birth to a child
  - a stringy slimy substance consisting of yeast cells and bacteria; is added to cider or wine to produce vinegar
- Context matters, e.g. what if previous sentence was:
  - Wow, Amazon predicted that you would need to order a big batch of new vinegar brewing ingredients. ☺

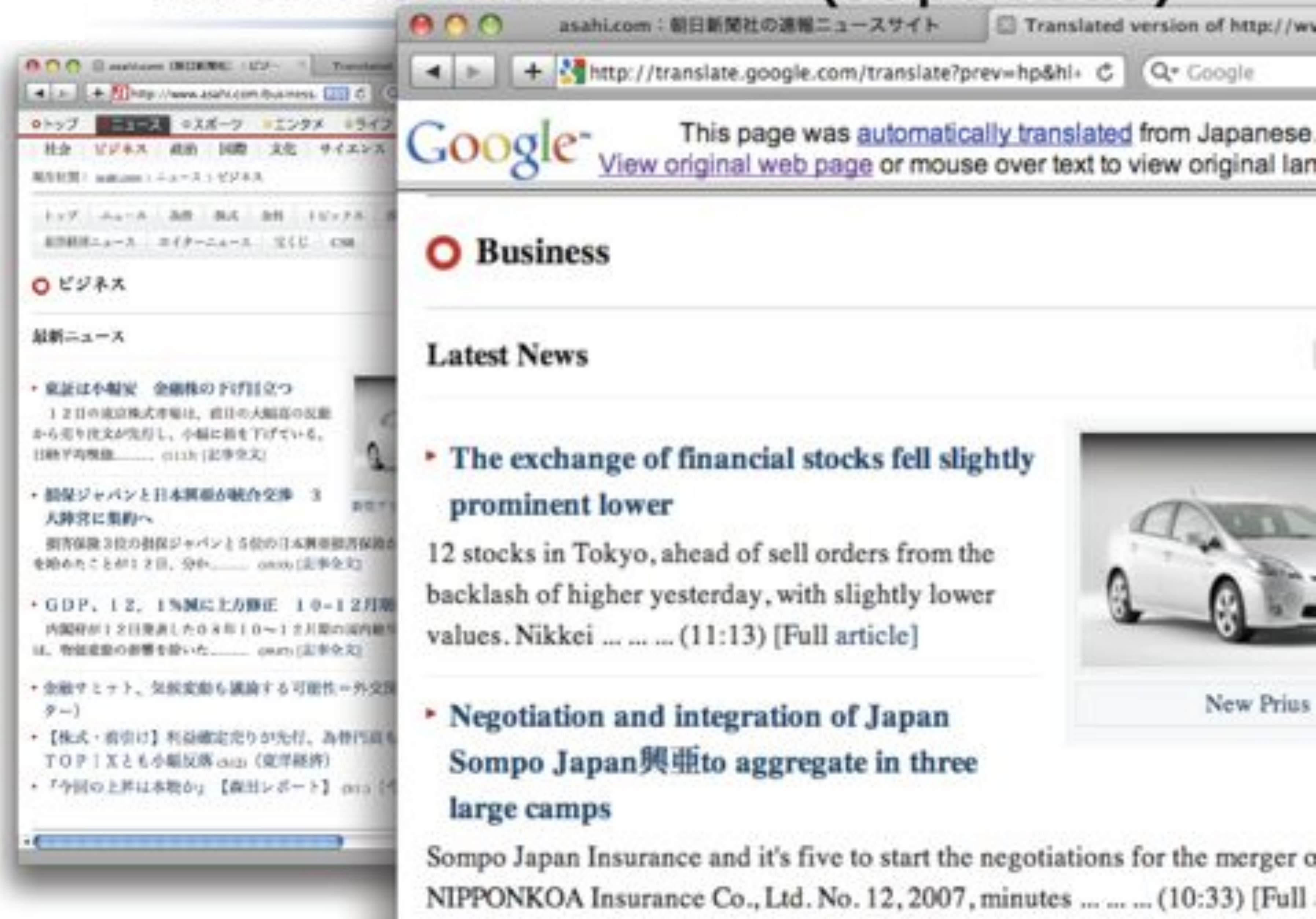
[Example from L. Lee]

A phone that understands our questions

Jeopardy! World Champion

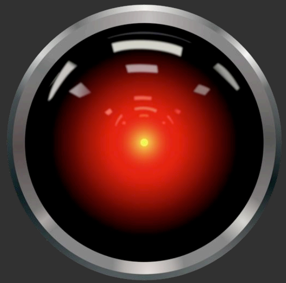
US Cities: Its largest airport is named for a World War II hero; its second largest, for a World War II battle.

Machine Translation (Japanese)



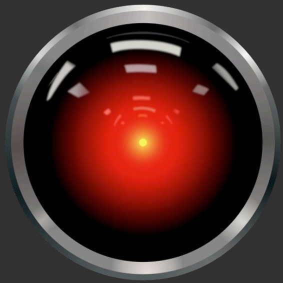
Natural Language Processing (NLP)

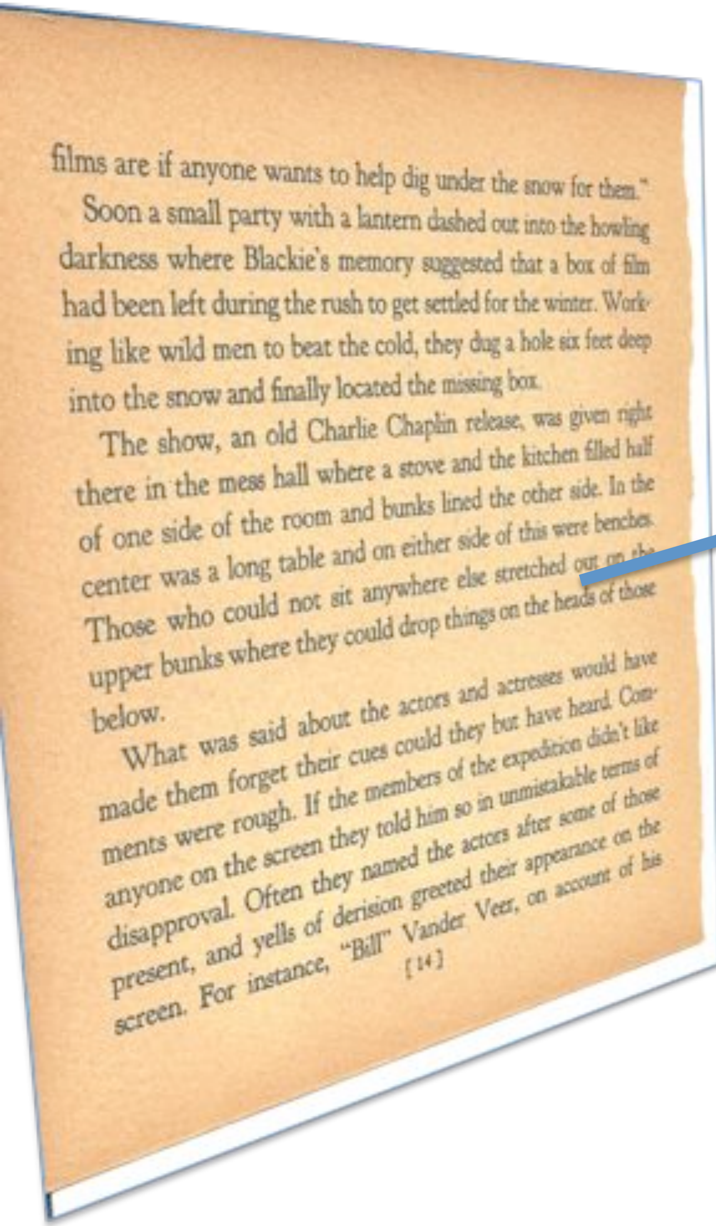
- a quick overview



Natural Language Processing (NLP)

- recent research (of our own)





films are if anyone wants to help dig under the snow for them."

Soon a small party with a lantern dashed out into the howling darkness where Blackie's memory suggested that a box of film had been left during the rush to get settled for the winter. Working like wild men to beat the cold, they dug a hole six feet deep into the snow and finally located the missing box.

The show, an old Charlie Chaplin release, was given right there in the mess hall where a stove and the kitchen filled half of one side of the room and bunks lined the other side. In the center was a long table and on either side of this were benches. Those who could not sit anywhere else stretched out on the upper bunks where they could drop things on the heads of those below.

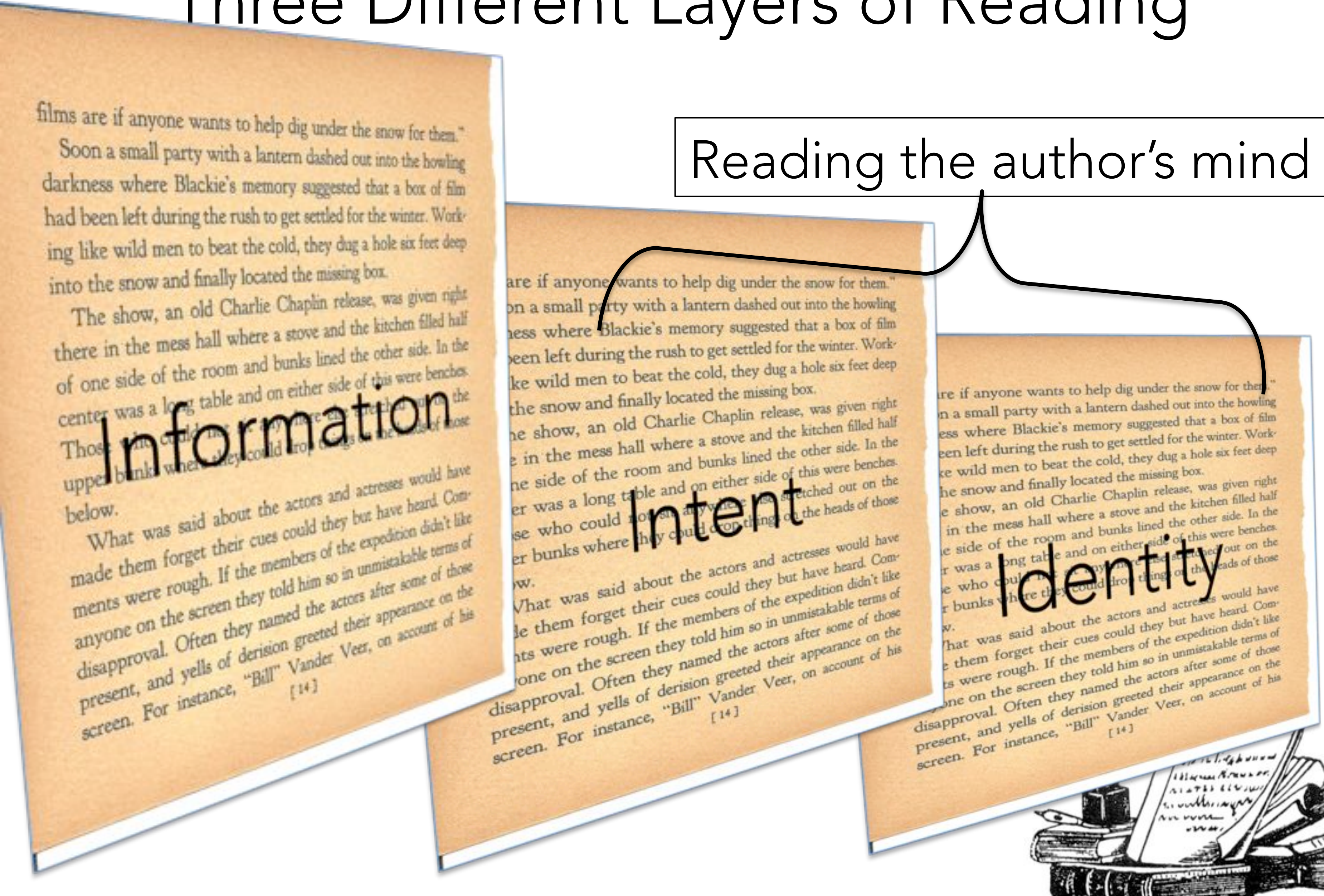
What was said about the actors and actresses would have made them forget their cues could they but have heard. Comments were rough. If the members of the expedition didn't like anyone on the screen they told him so in unmistakable terms of disapproval. Often they named the actors after some of those present, and yells of derision greeted their appearance on the screen. For instance, "Bill" Vander Veer, on account of his

[14]

What text understanding is really about?

Three Different Layers of Reading

Reading the author's mind



films are if anyone wants to help dig under the snow for them."

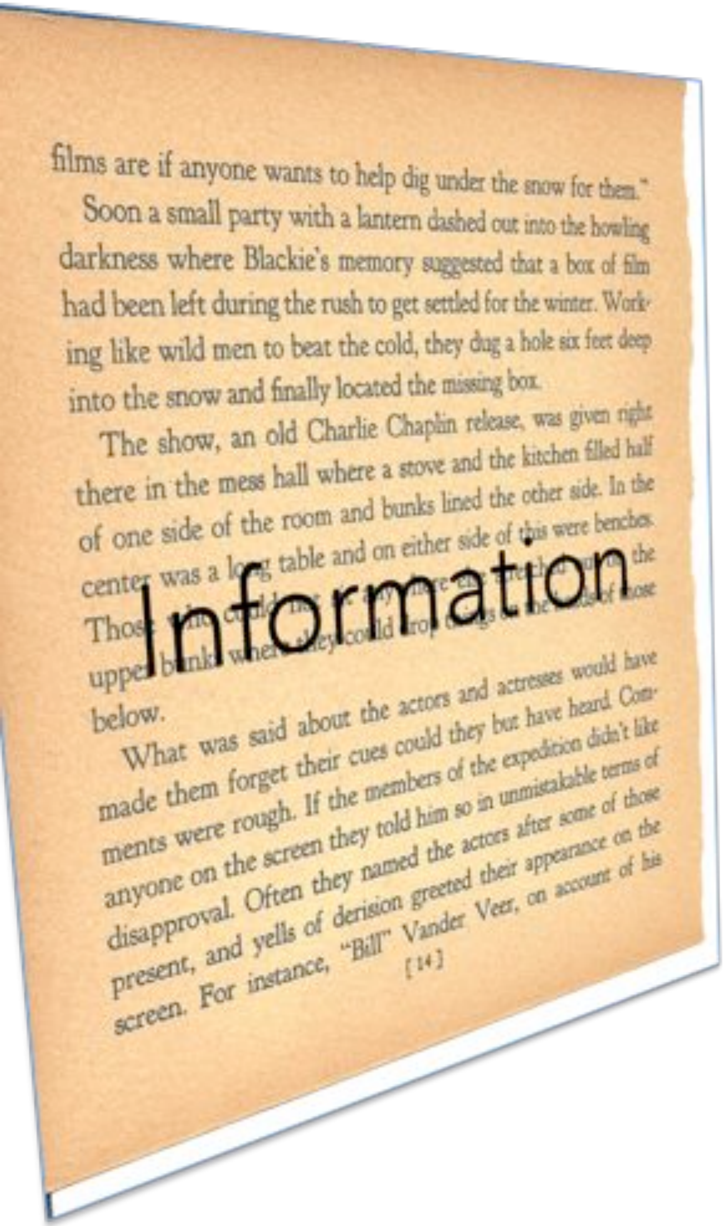
Soon a small party with a lantern dashed out into the howling darkness where Blackie's memory suggested that a box of film had been left during the rush to get settled for the winter. Working like wild men to beat the cold, they dug a hole six feet deep into the snow and finally located the missing box.

The show, an old Charlie Chaplin release, was given right there in the mess hall where a stove and the kitchen filled half of one side of the room and bunks lined the other side. In the center was a long table and on either side of this were benches. Those who could [illegible] stretched out on the upper bunks where they could drop things on the heads of those below.

What was said about the actors and actresses would have made them forget their cues could they but have heard. Comments were rough. If the members of the expedition didn't like anyone on the screen they told him so in unmistakable terms of disapproval. Often they named the actors after some of those present, and yells of derision greeted their appearance on the screen. For instance, "Bill" Vander Veer, on account of his

[14]

Information



films are if anyone wants to help dig under the snow for them."

Soon a small party with a lantern dashed out into the howling darkness where Blackie's memory suggested that a box of film had been left during the rush to get settled for the winter. Working like wild men to beat the cold, they dug a hole six feet deep into the snow and finally located the missing box.

The show, an old Charlie Chaplin release, was given right there in the mess hall where a stove and the kitchen filled half of one side of the room and bunks lined the other side. In the center was a long table and on either side of this were benches. Those [illegible] upper bunks where they could drop things on the heads of those below.

What was said about the actors and actresses would have made them forget their cues could they but have heard. Comments were rough. If the members of the expedition didn't like anyone on the screen they told him so in unmistakable terms of disapproval. Often they named the actors after some of those present, and yells of derision greeted their appearance on the screen. For instance, "Bill" Vander Veer, on account of his

[14]


"All the News That's Fit to Print"

The New York Times

Late Edition

NEW YORK, TUESDAY, OCTOBER 20, 1987

STOCKS PLUNGE 508 POINTS, A DROP OF 22.6%; 604 MILLION VOLUME NEARLY DOUBLES RECORD

U.S. Ships Shell Iran Installation In Gulf Reprisal

Offshore Target Termed a Base for Gunboats

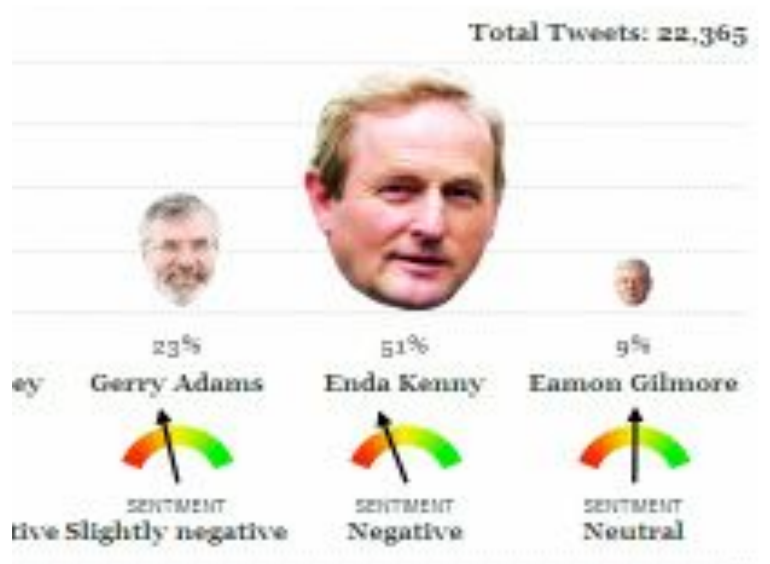
A Huge Blow to the Five-Year Bull Market

Dow's Record Fall

WORLDWIDE IMPACT

Frenzied Trading Raises Fears of Recession — Tape 2 Hours Late

Does 1987 Equal 1929?



Intent

films are if anyone wants to help dig under the snow for them."

Soon a small party with a lantern dashed out into the howling darkness where Blackie's memory suggested that a box of film had been left during the rush to get settled for the winter. Working like wild men to beat the cold, they dug a hole six feet deep into the snow and finally located the missing box.

The show, an old Charlie Chaplin release, was given right there in the mess hall where a stove and the kitchen filled half of one side of the room and bunks lined the other side. In the center was a long table and on either side of this were benches. Those who could not sit anywhere else stretched out on the upper bunks where they could drop things on the heads of those below.

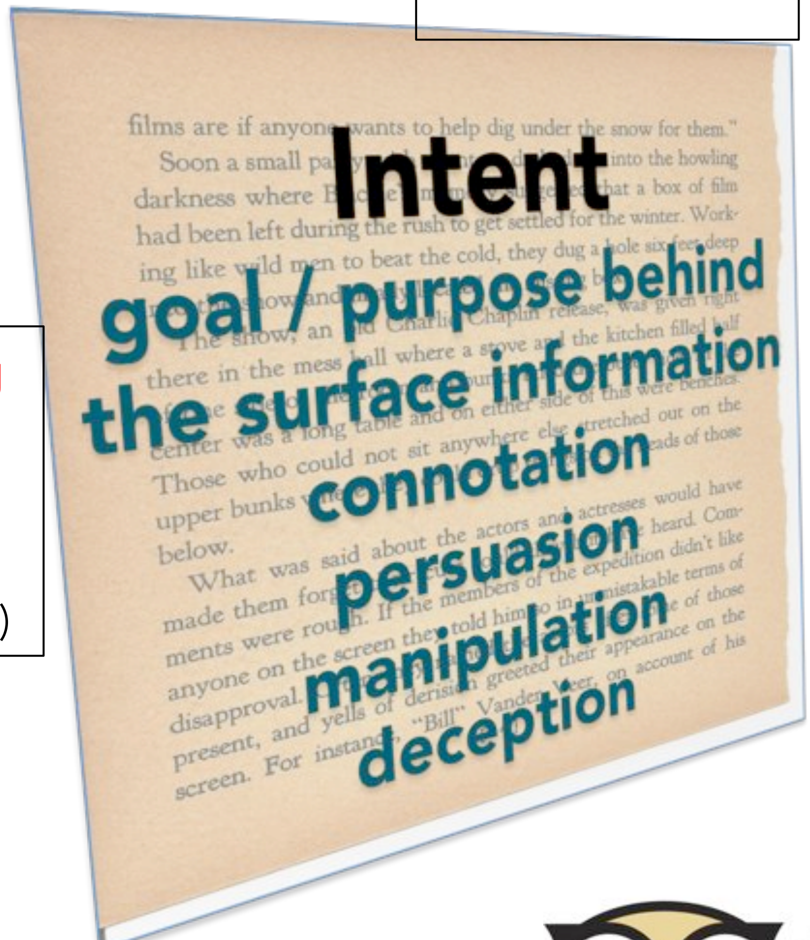
What was said about the actors and actresses would have made them forget their cues could they but have heard. Comments were rough. If the members of the expedition didn't like anyone on the screen they told him so in unmistakable terms of disapproval. Often they named the actors after some of those present, and yells of derision greeted their appearance on the screen. For instance, "Bill" Vander Veer, on account of his

dodging
(Nguyen et al 2013)

hedging
(Choi et al. 2012)
(Ganter and Strube, 2009)
(Kilicoglu and Bergler 2008)

"Eunsol" Choi

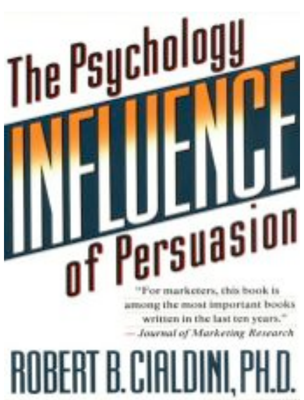
framing in media & political discourse
(Yano et al., 2010)
(Recasens et al., 2013)



syntactic packaging
"My toy broke"
instead of
"I broke my toy"
(Greene and Resnik 2009)

deception

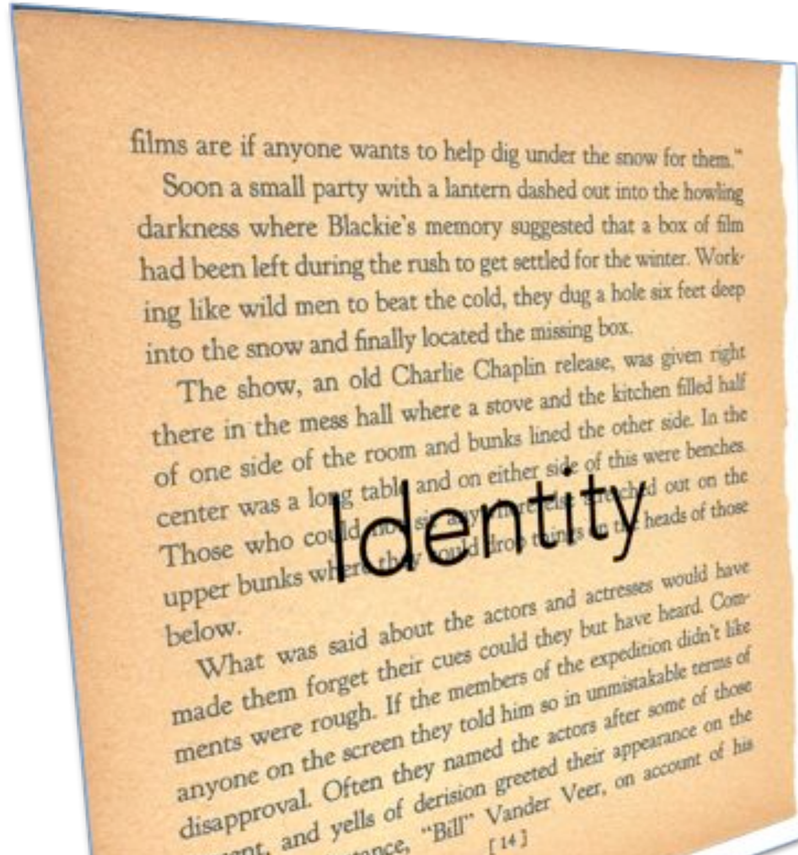
fake online reviews



yelp

tripadvisor

Identity



films are if anyone wants to help dig under the snow for them."

Soon a small party with a lantern dashed out into the howling darkness where Blackie's memory suggested that a box of film had been left during the rush to get settled for the winter. Working like wild men to beat the cold, they dug a hole six feet deep into the snow and finally located the missing box.

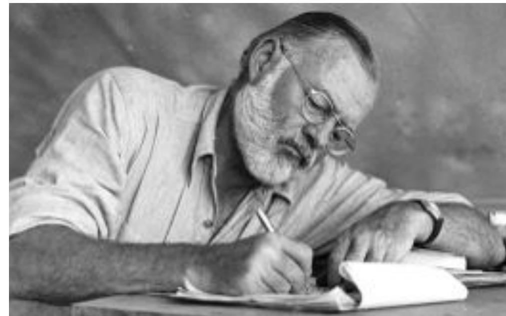
The show, an old Charlie Chaplin release, was given right there in the mess hall where a stove and the kitchen filled half of one side of the room and bunks lined the other side. In the center was a lo[illegible] tabl[illegible] and on either si[illegible]e of this were benches. Those who co[illegible] [illegible]d out on the upper bunks wh[illegible] [illegible] heads of those below.

What was said about the actors and actresses would have made them forget their cues could they but have heard. Comments were rough. If the members of the expedition didn't like anyone on the screen they told him so in unmistakable terms of disapproval. Often they named the actors after some of those present, and yells of derision greeted their appearance on the screen. For instance, "Bill" Vander Veer, on account of his

[14]


Capote

Hempel


Hemingway


Woolf

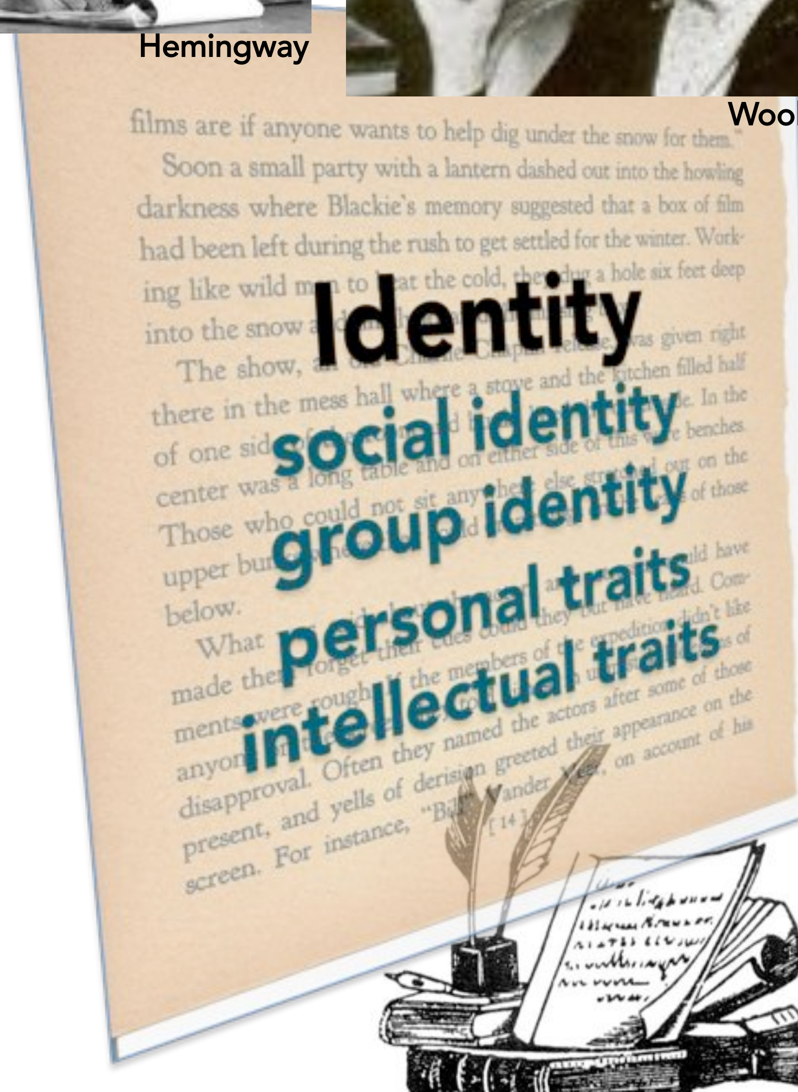
authorship verification

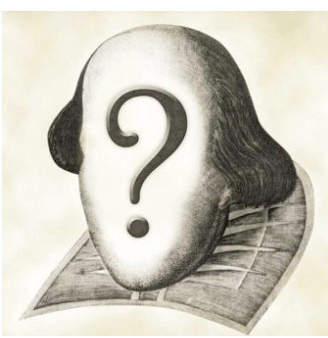
authorship obfuscation

demographics: gender, nationality, age, vocation

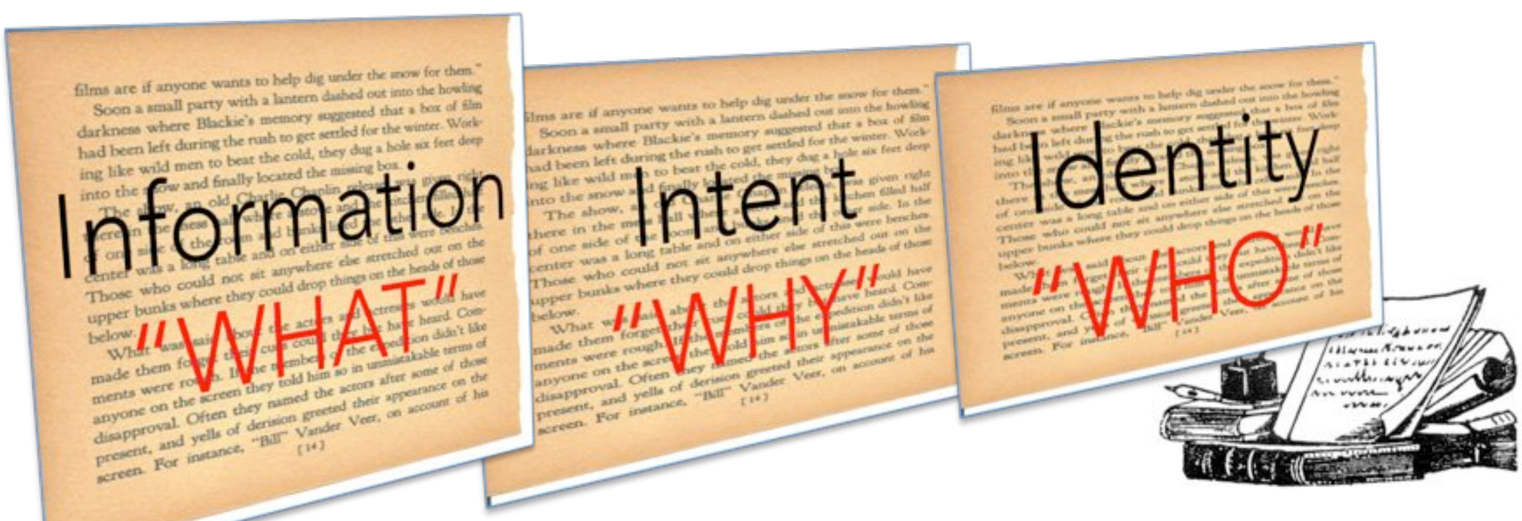
personality, psychological state: happy, authoritative, depressed…

intellectual traits & development: literary success





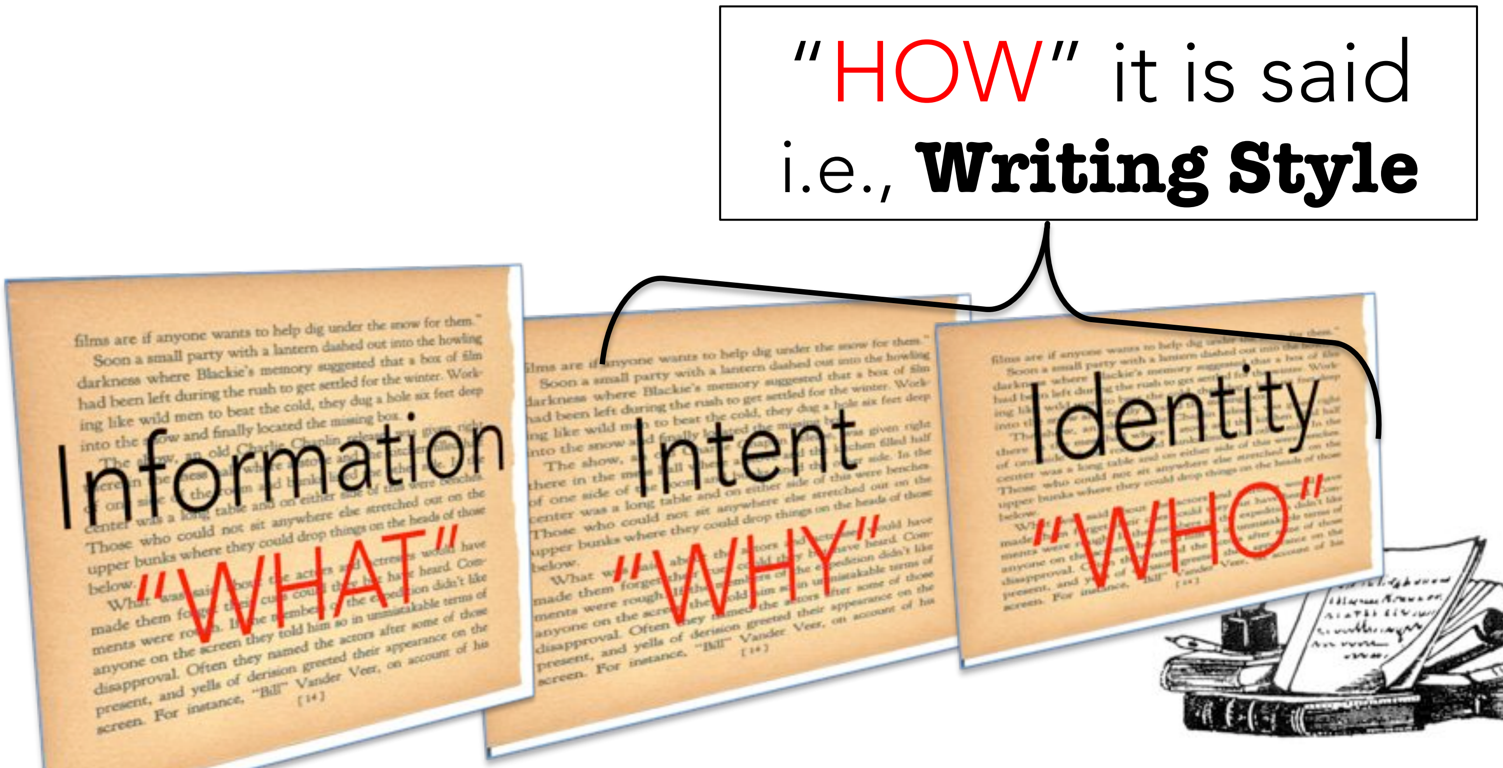
From Language to the Mind



From Language to the Mind

Is it even possible? (without full semantic understanding)

- It is more about "HOW" it is said than "WHAT" is said.



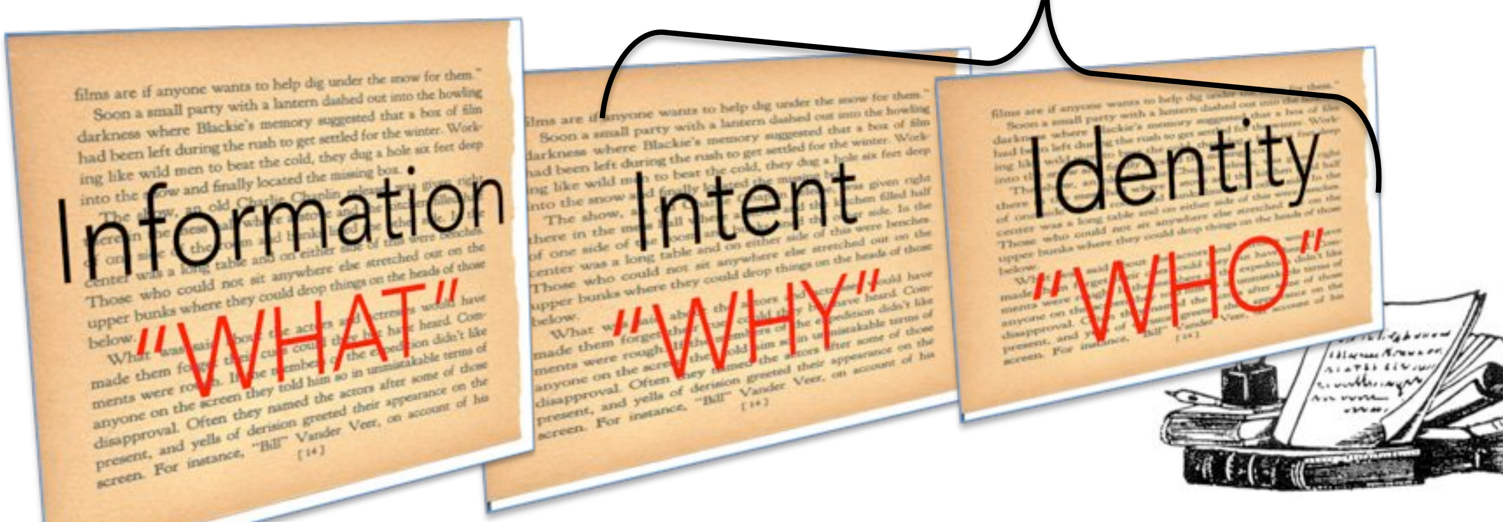
From Language to the Mind

Is it even possible? (without full semantic understanding)

- It is more about "HOW" it is said than "WHAT" is said.
- We --humans– also often rely on "overall impression".

Computers at times can do better than humans!

"HOW" it is said
i.e., **Writing Style**



What is “Writing Style” ?

Research Papers? *New York Time*

Blog Post

“So how can you spot a f
difficult, but with some te
signs:”

Research Paper (**ACL**, 2011)

“To obtain a deeper understanding of the nature of
deceptive reviews, we examin
potentially complementary fra

The New York Times

“As online retailers increasingly depend on reviews as a sales tool, an industry of fibbers and promoters has sprung up to buy and sell raves for a pittance.”

What is "Writing Style" ?

Genre Categorization:

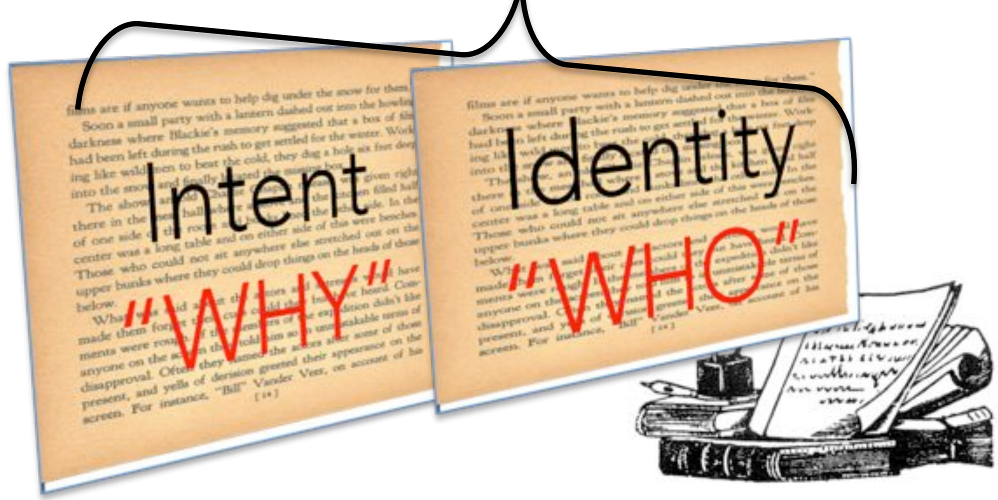
Petrenz and Webber, 2011; Finn et al., 2006; Argamon et al., 2003; Kessler et al., 1997

Authorship Attribution:

Holmes 1985, Raghavan et al., 2010; Koppel and Shler, 2004; Gamon, 2004;

"HOW" it is said

i.e., **Writing Style**



Many more possibilities…

Swanson and Charniak, 2012; Xu et al., 2012; Iyyer et al., 2014; Hardisty et al., 2010

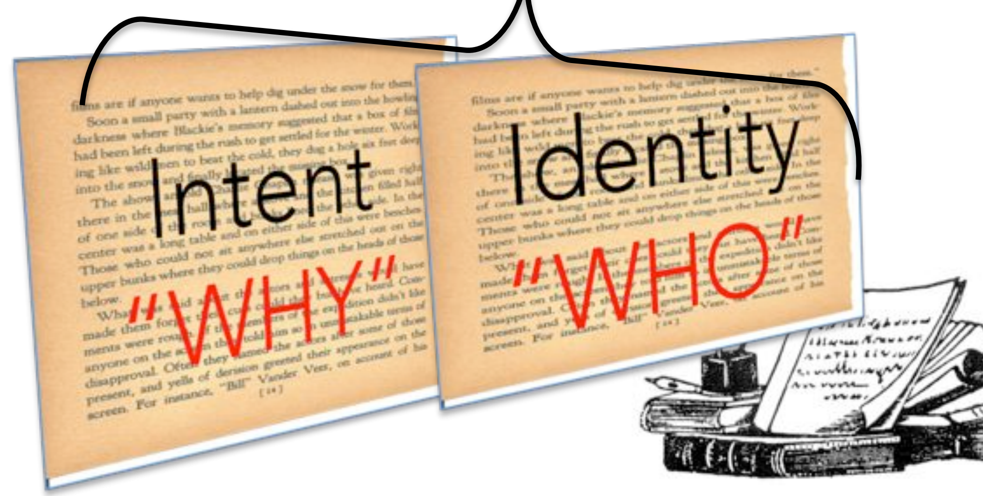
From Language to the Mind

Unconventional Case Studies:

I. Deceptive Reviews (ACL 2011)

II. Success of Novels (EMNLP 2013)

"HOW" it is said
i.e., **Writing Style**



Motivation

Online reviews

= shopping tool

Commercial impact

→ potential target for deceptive reviews

"My husband and I stayed at the James Chicago Hotel for our anniversary. This place is fantastic! We knew as soon as we arrived we made the right choice! The rooms are BEAUTIFUL and the staff very attentive and wonderful! The area of the hotel is great, since I love to shop I couldn't ask for more! We will definitely be back to Chicago and we will for sure be back to the James Chicago."

Deceptive or Truthful?

"My husband and I stayed at the James Chicago Hotel for our anniversary. This place is fantastic! We knew as soon as we arrived we made the right choice! The rooms are BEAUTIFUL and the staff very attentive and wonderful! The area of the hotel is great, since I love to shop I couldn't ask for more! We will definitely be back to Chicago and we will for sure be back to the James Chicago."

"I have stayed at many hotels traveling for both business and pleasure and I can honestly say that The James is tops. The service at the hotel is first class. The rooms are modern and very comfortable. The location is perfect within walking distance to all of the great sights and restaurants. Highly recommend to both business travellers and couples."

Gathering Data

- ~~Label existing reviews?~~
  - Can’t manually do this

Gathering Data

- ~~Label existing reviews?~~
  - Can't manually do this

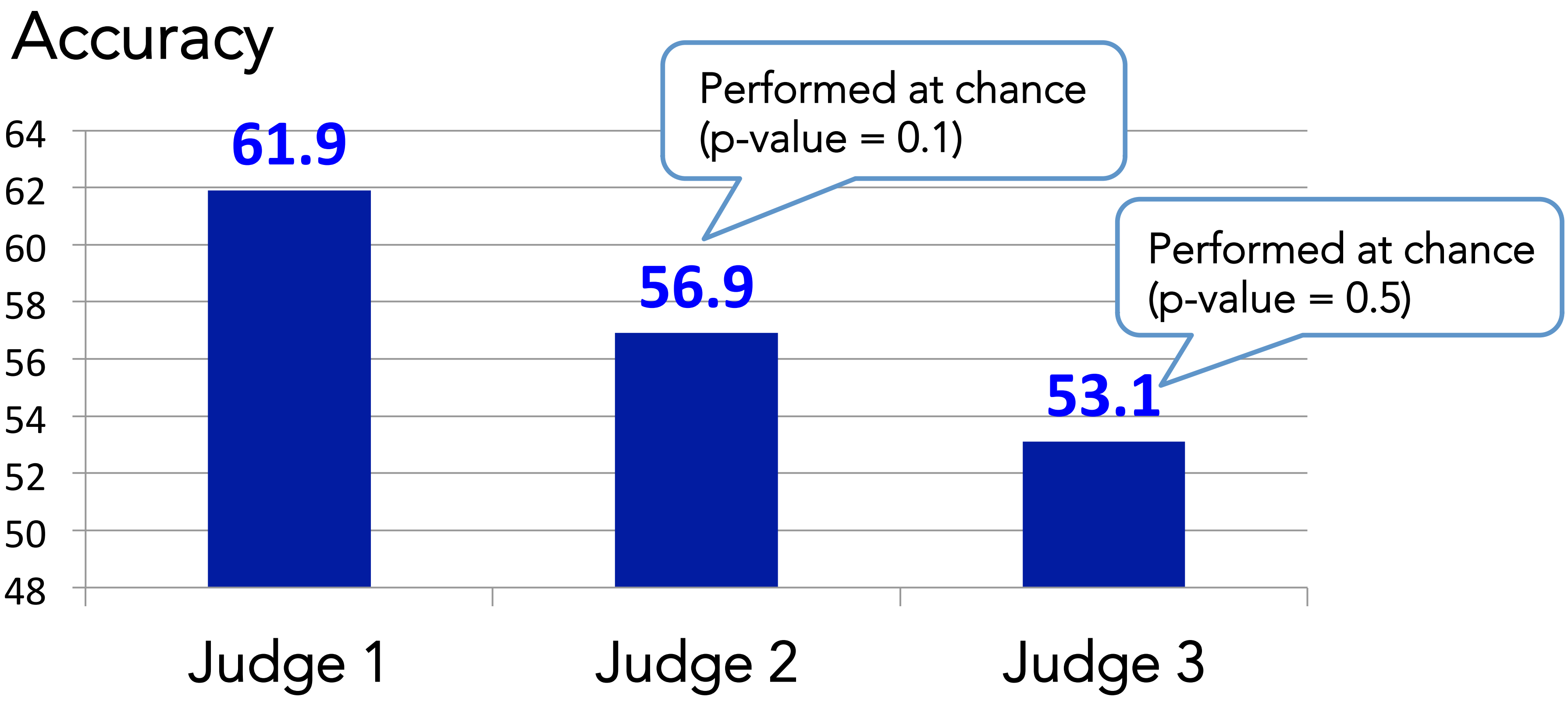
❑ Instead, create new reviews
  - By hiring people to write fake positive reviews
  - Amazon Mechanical Turk
    - 20 hotels
    - 20 reviews / hotel
    - Offer $1 / review
    - 400 reviews

How good are humans in detecting deceptive reviews?

- 80 truthful and 80 deceptive reviews
- 3 undergraduate judges

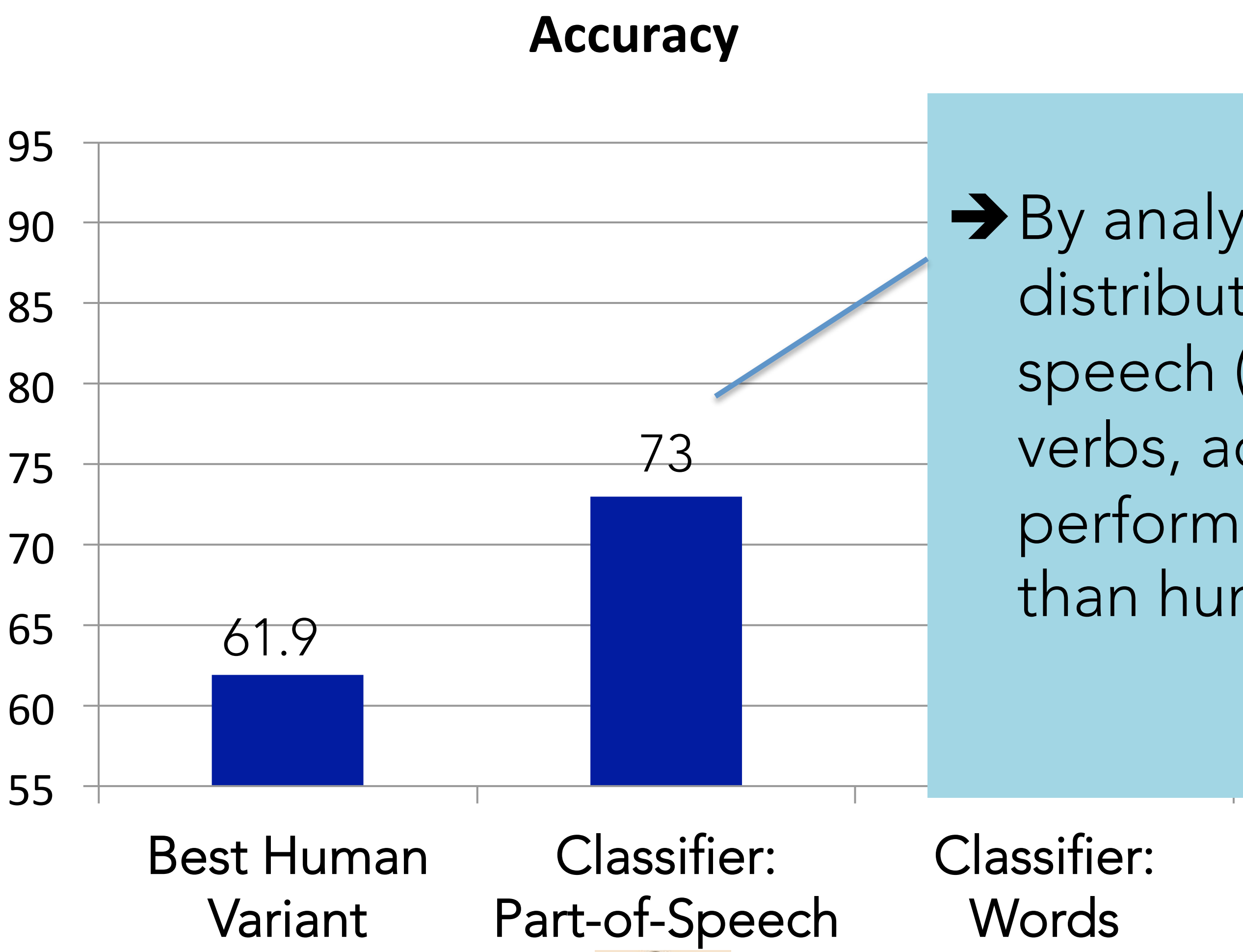
Human Performance

➔ Aligns with previous studies in deception literature:
humans typically perform barely better than chance.
trained experts may perform at ~70%

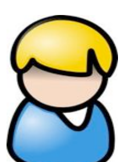


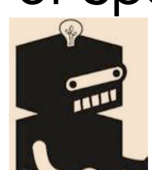
How Well Can Computers Do?

Classifier Performance (SVM with 5-fold CV)

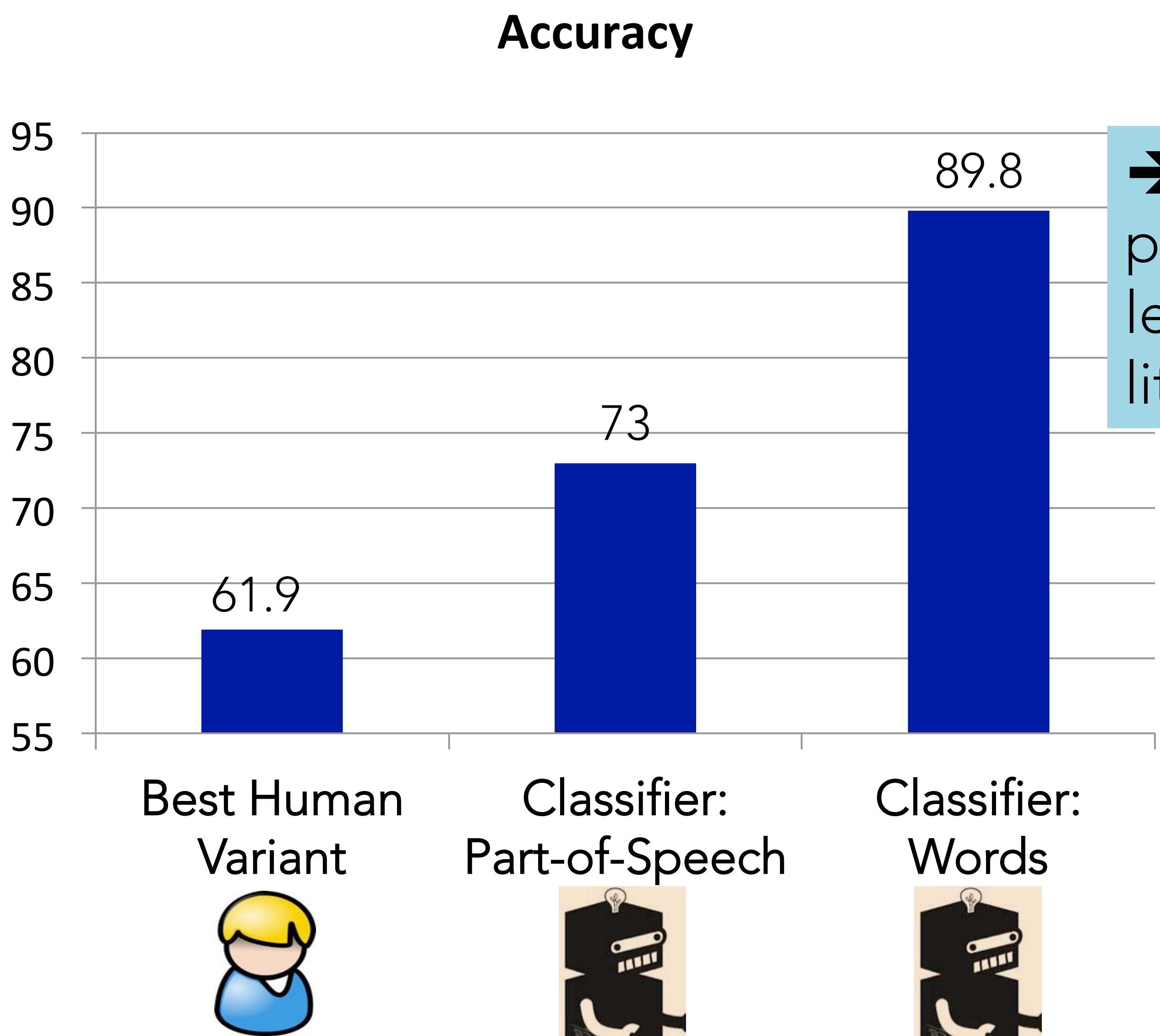


➔ By analyzing *only* the distribution of part-of-speech (e.g., nouns, verbs, adjectives), already performs much better than human judges!





Classifier Performance (SVM with 5-fold CV)



→ No human performs at this level in deception literature!

Data-driven Discovery of Insights
into
Deceptive Writings

<table>
<tr><th colspan="3">TRUTHFUL/INFORMATIVE</th><th colspan="3">DECEPTIVE/IMAGINATIVE</th></tr>
<tr><th>Category</th><th>Variant</th><th>Weight</th><th>Category</th><th>Variant</th><th>Weight</th></tr>
<tr><td rowspan="4">NOUNS</td><td>Singular</td><td>0.008</td><td rowspan="7">VERBS</td><td>Base</td><td>-0.057</td></tr>
<tr><td>Plural</td><td>0.002</td><td>Past tense</td><td>0.041</td></tr>
<tr><td>Proper, singular</td><td>-0.041</td><td>Present participle</td><td>-0.089</td></tr>
<tr><td>Proper, plural</td><td>0.091</td><td>Singular, present</td><td>-0.031</td></tr>
<tr><td rowspan="3">ADJECTIVES</td><td>General</td><td>0.002</td><td rowspan="2">Third person singular, present</td><td rowspan="2">0.026</td></tr>
<tr><td>Comparative</td><td>0.058</td></tr>
<tr><td>Superlative</td><td>-0.164</td><td>Modal</td><td>-0.063</td></tr>
<tr><td>PREPOSITIONS</td><td>General</td><td>0.064</td><td rowspan="2">ADVERBS</td><td>General</td><td>0.001</td></tr>
<tr><td>DETERMINERS</td><td>General</td><td>0.009</td><td>Comparative</td><td>-0.035</td></tr>
<tr><td>COORD. CONJ.</td><td>General</td><td>0.094</td><td rowspan="2">PRONOUNS</td><td>Personal</td><td>-0.098</td></tr>
<tr><td>VERBS</td><td>Past participle</td><td>0.053</td><td>Possessive</td><td>-0.303</td></tr>
<tr><td>ADVERBS</td><td>Superlative</td><td>-0.094</td><td>PRE-DETERMINERS</td><td>General</td><td>0.017</td></tr>
</table>

Informative writing (left) --- nouns, adjectives, prepositions
Imaginative writing (right) --- verbs, adverbs, pronouns
Rayson et. al. (2001)

TRUTHFUL/INFORMATIVE		
Category	**Variant**	**Weight**
NOUNS	Singular	0.008
	Plural	0.002
	Proper, singular	**-0.041**
	Proper, plural	0.091
ADJECTIVES	General	0.002
	Comparative	0.058
	Superlative	**-0.164**
PREPOSITIONS	General	0.064
DETERMINERS	General	0.009
COORD. CONJ.	General	0.094
VERBS	Past participle	0.053
ADVERBS	Superlative	**-0.094**

DECEPTIVE/IMAGINATIVE		
Category	**Variant**	**Weight**
VERBS	Base	-0.057
	Past tense	**0.041**
	Present participle	-0.089
	Singular, present	-0.031
	Third person singular, present	**0.026**
	Modal	-0.063
ADVERBS	General	**0.001**
	Comparative	-0.035
PRONOUNS	Personal	-0.098
	Possessive	-0.303
PRE-DETERMINERS	General	**0.017**

Truthful Reviews

≈

Informative Writing (Journalism)

Deceptive Reviews

≈

Imaginative Writing (Novels)

The New York Times

STRONG DECEPTIVE INDICATORS

A focus on who they were with
In this example, "My husband;" also words like "family."

Greater use of first-person singular
Fake reviews tend to use "I" and "me" more often.

Direct mention of where they stayed
Hotel and city names were less common in truthful reviews, which focus more on details about the hotel itself, like "small" or "bathroom."

"My husband and I stayed in the [*hotel name*] Chicago and had a very nice stay! The rooms were large and comfortable. The view of Lake Michigan from our room was gorgeous. Room service was really good and quick, eating in the room looking at that view, awesome! The pool was really nice but we didn't get a chance to use it. Great location for all of the downtown Chicago attractions such as theaters and museums. Very friendly staff and knowledgable, you can't go wrong staying here."

SLIGHT DECEPTIVE INDICATORS

High adverb use
"Very" and "really" are both used twice; "here" is used once.

High verb use
"Get", "go", "use", "can't", "didn't", "eating", "had", "looking", "stayed", "was" (three times), "were."

Use of "!" and positive emotion
Deceptive reviews tend to use exclamation points, while truthful reviews used more punctuation of other kinds, including "$."

The New York Times

STRONG DECEPTIVE INDICATORS

A focus on who they were with
In this example, "My husband;" also words like "family."

Greater use of first-person singular
Fake reviews tend to use "I" and "me" more often.

Direct mention of where they stayed
Hotel and city names were less common in truthful reviews, which focus more on details about the hotel itself, like "small" or "bathroom."

"My husband and I stayed in the [*hotel name*] Chicago and had a very nice stay! The rooms were large and comfortable. The view of Lake Michigan from our room was gorgeous. Room service was really good and quick, eating in the room looking at that view, awesome! The pool was really nice but we didn't get a chance to use it. Great location for all of the downtown Chicago attractions such as theaters and museums. Very friendly staff and knowledgeable, you can't go wrong staying here."

- lack of spatial, sensorial details (Vrij et al., 2009)
- lack of descriptive adjectives: low, small, shiny
- less use of prepositions

The New York Times

STRONG DECEPTIVE INDICATORS

A focus on who they were with
In this example, "My husband;" also words like "family."

Greater use of first-person singular
Fake reviews tend to use "I" and "me" more often.

Direct mention of where they stayed
Hotel and city names were less common in truthful reviews, which focus more on details about the hotel itself, like "small" or "bathroom."

"My husband and I stayed in the [*hotel name*] Chicago a... had a very nice stay! The rooms were large and co... The view of Lake Michigan from our room wa... ...ervice was really good and quick,

instead, story telling:

-- why they were there: "vacation", "business", "anniversary"
-- whom they were with: "husband", "family"

"really" are both used twice; "here" is used once.

"eating", "had", "looking", "stayed", "was" (three times), "were."

to use exclamation points, while truthful reviews used more punctuation of other kinds, including "$."

The New York Times

STRONG DECEPTIVE INDICATORS

- exaggeration, words over the top: "fantastic", "luxurious", "gorgeous", "awesome"
- superlatives: "the most", "best", "ever"
- certainty: "absolutely", "definitely", "for sure"

...re less ...ws, which ...out the hotel ...room."

and had a very nice stay! The ro... and comfortable. The view of Lake ...higan from our room was gorgeous. Room service was really good and quick, eating in the room looking at that view, awesome! The pool was really nice but we didn't get a chance to use it. Great location for all of the downtown Chicago attractions such as theaters and museums. Very friendly staff and knowledgable, you can't go wrong staying here."

SLIGHT DECEPTIVE INDICATORS

High adverb use
"Very" and "really" are both used twice; "here" is used once.

High verb use
"Get", "go", "use", "can't", "didn't", "eating", "had", "looking", "stayed", "was" (three times), "were."

Use of "!" and positive emotion
Deceptive reviews tend to use exclamation points, while truthful reviews used more punctuation of other kinds, including "$."

The New York Times

STRONG DECEPTIVE INDICATORS

A focus on who they were with
In this example, "My husband;" also words like "family."

Greater use of first-person singular
Fake reviews tend to use "I" and "me" more often.

Direct mention of where they stayed
Hotel and city names were less common in truthful reviews, which focus more on details about the hotel itself, like "small" or "bathroom."

"My husband and I stayed in the *[hotel name]* Chicago and had a very nice stay! The rooms were large and comfo ... view of Lake Michigan from our room was ... vice was really good and quick,

Increased level of "first person singular"
"I", "me", "my", "mine"

In contrast to psychological distancing (Newman et al., 2003)
➔ deception cues are domain dependent

"really" are both used twice; "here" is used once.

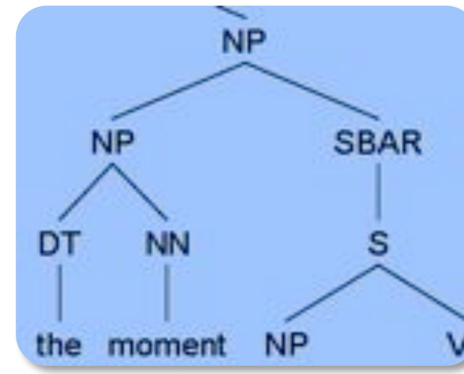
"eating", "had", "looking", "stayed", "was" (three times), "were."

to use exclamation points, while truthful reviews used more punctuation of other kinds, including "$."

What happened after then (= 2011) ?

1. We built better detection models

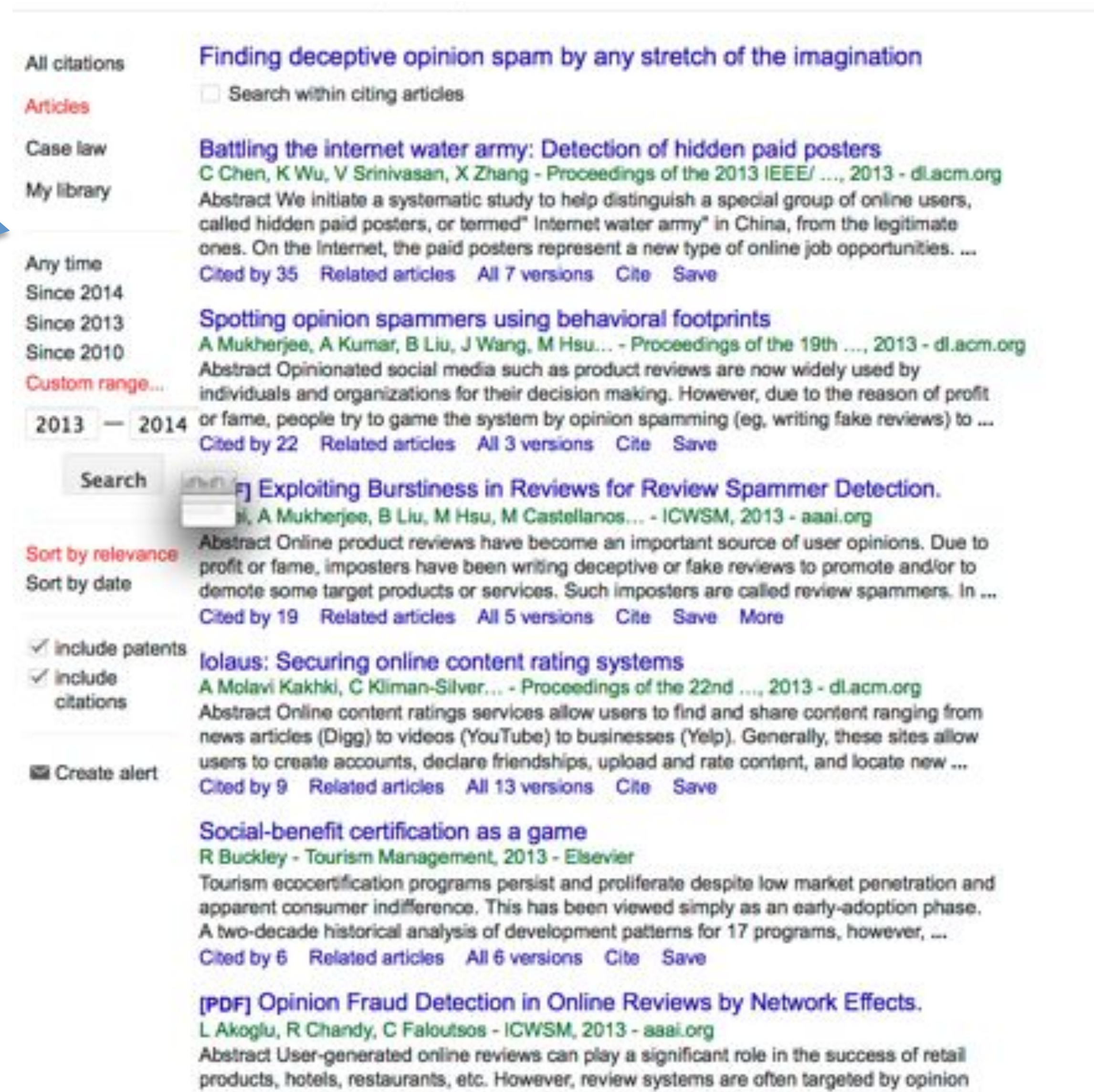
① Syntax Improves Deception Detection
(Feng et al., **ACL 2012**)
--- 3 product review dataset
---1 essay dataset (Mihalcea and Strapparava (2009))



② Natural V.S. Distorted Distributions of Opinions
(Feng et al., **ICWSM 2012,** best paper runner up)

2. We excited other researchers

185 citations


All citations
Articles
Case law
My library

Any time
Since 2014
Since 2013
Since 2010
Custom range...
2013 — 2014
Search

Sort by relevance
Sort by date

include patents
include citations

Create alert

Finding deceptive opinion spam by any stretch of the imagination
Search within citing articles

Battling the internet water army: Detection of hidden paid posters
C Chen, K Wu, V Srinivasan, X Zhang - Proceedings of the 2013 IEEE/ ..., 2013 - dl.acm.org
Abstract We initiate a systematic study to help distinguish a special group of online users, called hidden paid posters, or termed" Internet water army" in China, from the legitimate ones. On the Internet, the paid posters represent a new type of online job opportunities. ...
Cited by 35 Related articles All 7 versions Cite Save

Spotting opinion spammers using behavioral footprints
A Mukherjee, A Kumar, B Liu, J Wang, M Hsu... - Proceedings of the 19th ..., 2013 - dl.acm.org
Abstract Opinionated social media such as product reviews are now widely used by individuals and organizations for their decision making. However, due to the reason of profit or fame, people try to game the system by opinion spamming (eg, writing fake reviews) to ...
Cited by 22 Related articles All 3 versions Cite Save

Exploiting Burstiness in Reviews for Review Spammer Detection.
A Mukherjee, B Liu, M Hsu, M Castellanos... - ICWSM, 2013 - aaai.org
Abstract Online product reviews have become an important source of user opinions. Due to profit or fame, imposters have been writing deceptive or fake reviews to promote and/or to demote some target products or services. Such imposters are called review spammers. In ...
Cited by 19 Related articles All 5 versions Cite Save More

Iolaus: Securing online content rating systems
A Molavi Kakhki, C Kliman-Silver... - Proceedings of the 22nd ..., 2013 - dl.acm.org
Abstract Online content ratings services allow users to find and share content ranging from news articles (Digg) to videos (YouTube) to businesses (Yelp). Generally, these sites allow users to create accounts, declare friendships, upload and rate content, and locate new ...
Cited by 9 Related articles All 13 versions Cite Save

Social-benefit certification as a game
R Buckley - Tourism Management, 2013 - Elsevier
Tourism ecocertification programs persist and proliferate despite low market penetration and apparent consumer indifference. This has been viewed simply as an early-adoption phase. A two-decade historical analysis of development patterns for 17 programs, however, ...
Cited by 6 Related articles All 6 versions Cite Save

[PDF] Opinion Fraud Detection in Online Reviews by Network Effects.
L Akoglu, R Chandy, C Faloutsos - ICWSM, 2013 - aaai.org
Abstract User-generated online reviews can play a significant role in the success of retail products, hotels, restaurants, etc. However, review systems are often targeted by opinion

3. Been featured by media outlets

(Highlights 2011-2014)

- [ACL 2011] Finding Deceptive Opinion Spam by Any Stretch of the Imagination.
- [ICWSM 2012] Distributional Footprints of Deceptive Product Reviews.
- [EMNLP 2013] Where Not to Eat? Improving Public Policy by Predicting Hygiene...

4. We hope NLP for Social Good

- When our work was first published in 2011, no clear legal regulations against fake reviews.
- Not any more! New York law enforcement charged 19 firms $350,000 for facilitating fake reviews (Sep 2013).
  - (not based on automatic detection)

theguardian

News | US | World | Sports | Comment | Culture | Business | Money

News › World news › New York

Fake online reviews crackdown in New York sees 19 companies fined

Attorney general set up a fake yoghurt shop in Brooklyn to ensnare fake online review companies, fined a total of $350,000

Dominic Rushe in New York
Follow @dominicru Follow @guardian
theguardian.com, Monday 23 September 2013 14.42 EDT

Conclusion (Part I – Deception)

- Learning to read the "intent" of the author, even a hidden one.
- Humans not good at this task.
- Computers can at times perform better than humans, even without full blown semantic understanding.
- Data-driven discovery of insights to complement hypothesis-driven research

Ganganath, Jurafsky, McFarland (EMNLP 2009)
➔ computers predict flirtation intention better than humans can, despite humans having access to vastly richer information (visual features, gesture, etc.).

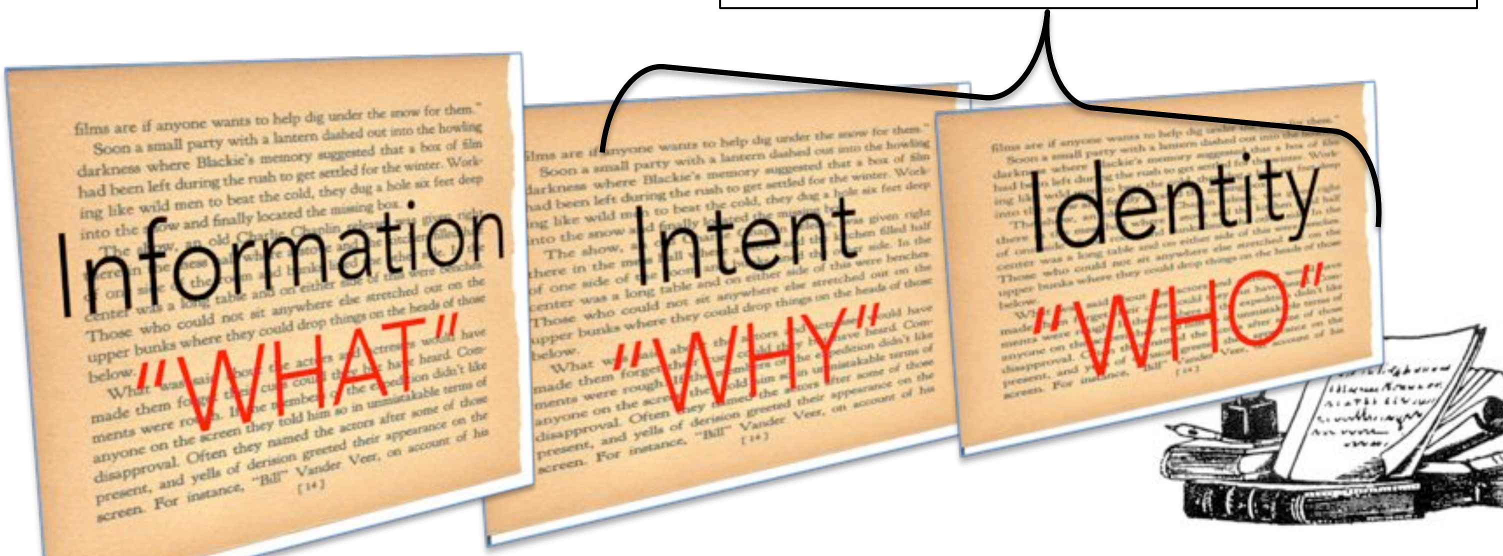
From Language to the Mind

Unconventional Case Studies:

I. Deceptive Reviews (ACL 2011)

II. Success of Novels (EMNLP 2013)

"HOW" it is said
i.e., **Writing Style**

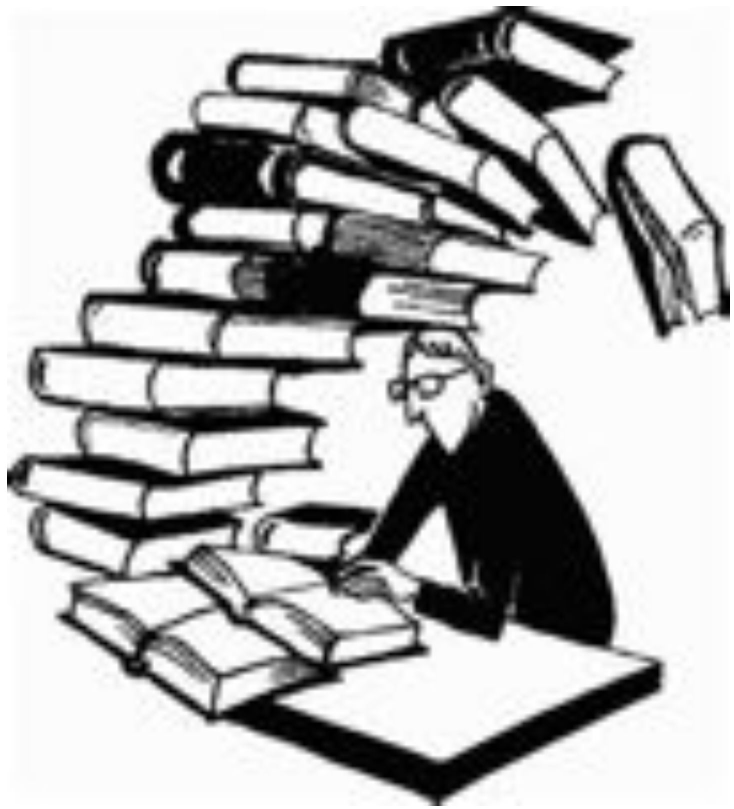


Predicting the success of novels

Novelty

Style of writing

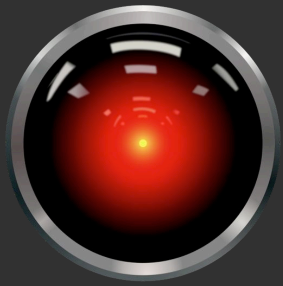
Story line



Social context

Luck !

Describing the Visual World in *Natural Language*

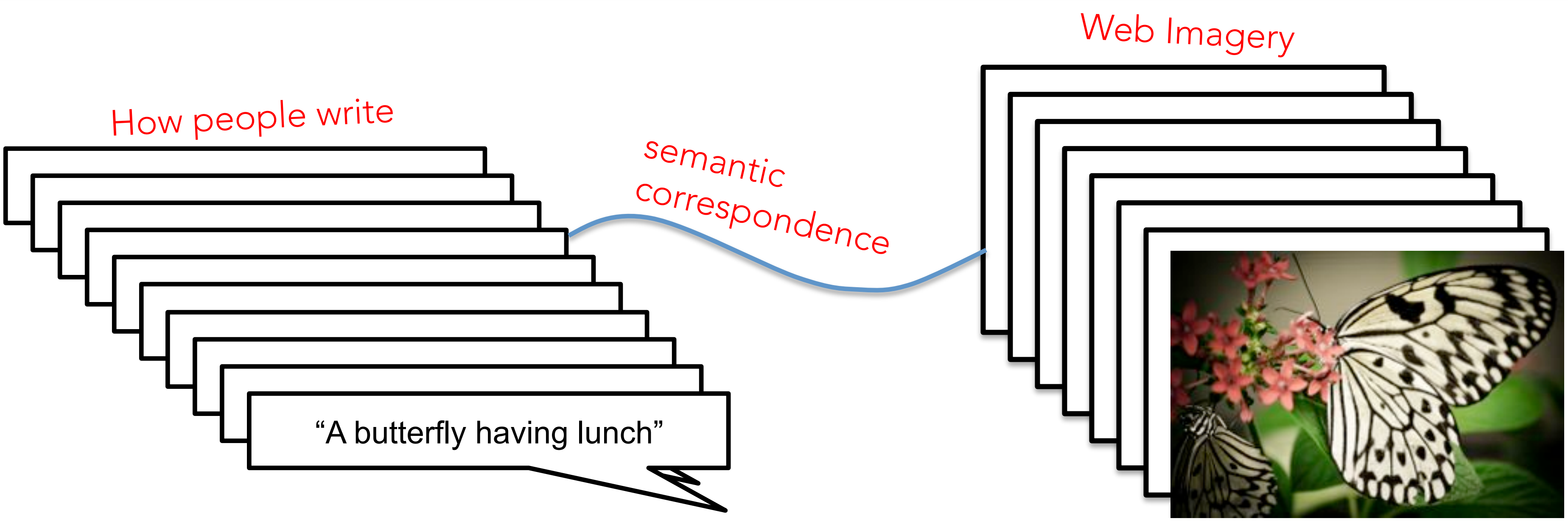


Task:
Learning to Describe Images in Natural Language

Two approaches:

I. `BabyTalk` Formulaic image description
   - CVPR 2011

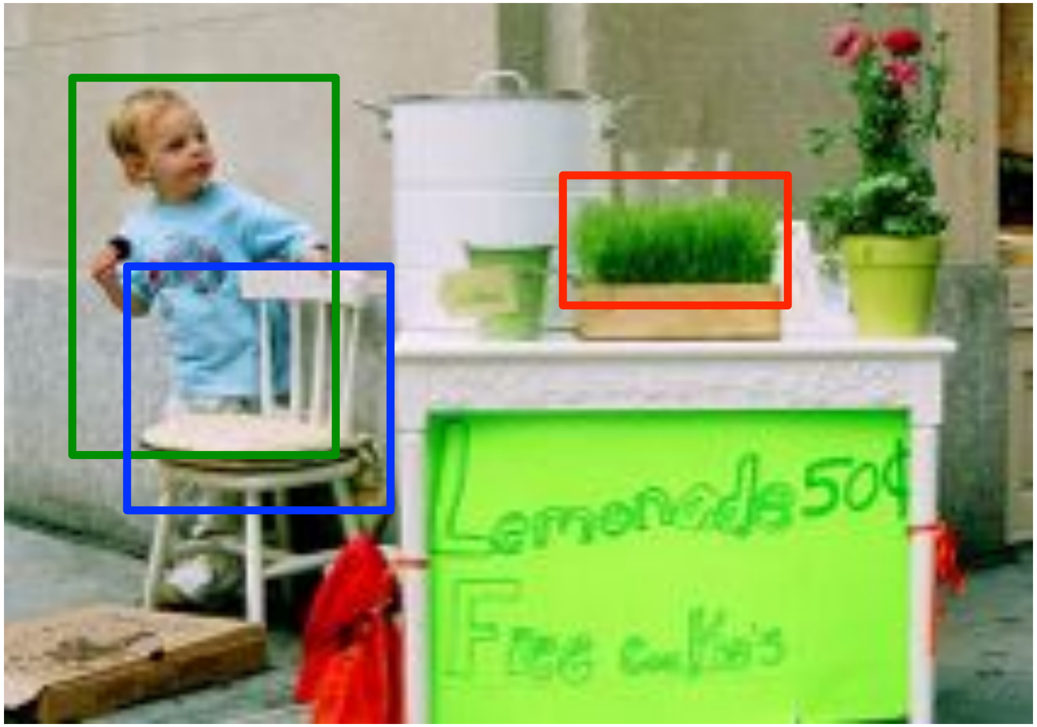
II. `TreeTalk` Expressive image description
   - TACL 2014 (in submission), ACL 2013, ACL 2012



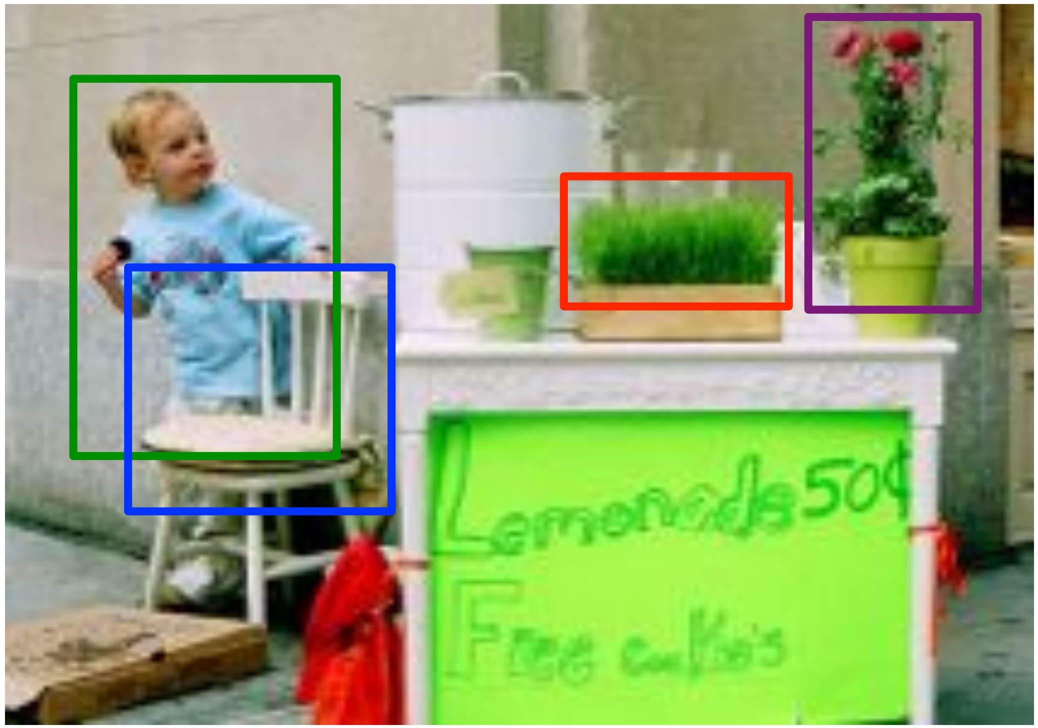

Lemonade 50¢

“This picture shows one person,

“This picture shows one person, one grass,



“This picture shows one person, one grass, one chair,



“This picture shows one person, one grass, one chair, and one potted plant.

“This picture shows one person, one grass, one chair, and one potted plant. The person is near the green grass,

"This picture shows one person, one grass, one chair, and one potted plant. The person is near the green grass, and in the chair.

“This picture shows one person, one grass, one chair, and one potted plant. The person is near the green grass, and in the chair. The green grass is by the chair, and near the potted plant.”

Methodology Overview

Input Imag

a) dog

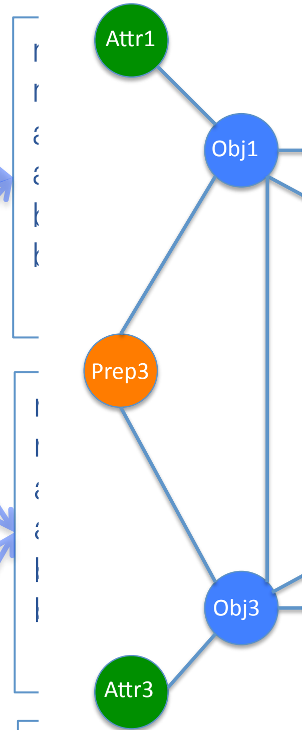
b) person

c) sofa

Extract Objects/stuff

Predict prepositions

Predict attributes



This is a photograph of one person and one brown sofa and one dog. The person is against the brown sofa. And the dog is near the person, and beside the brown sofa.

<<null,person_b>,against,<brown,sofa_c>>
<<null,dog_a>,near,<null,person_b>>
<<null,dog_a>,beside,<brown,sofa_c>>

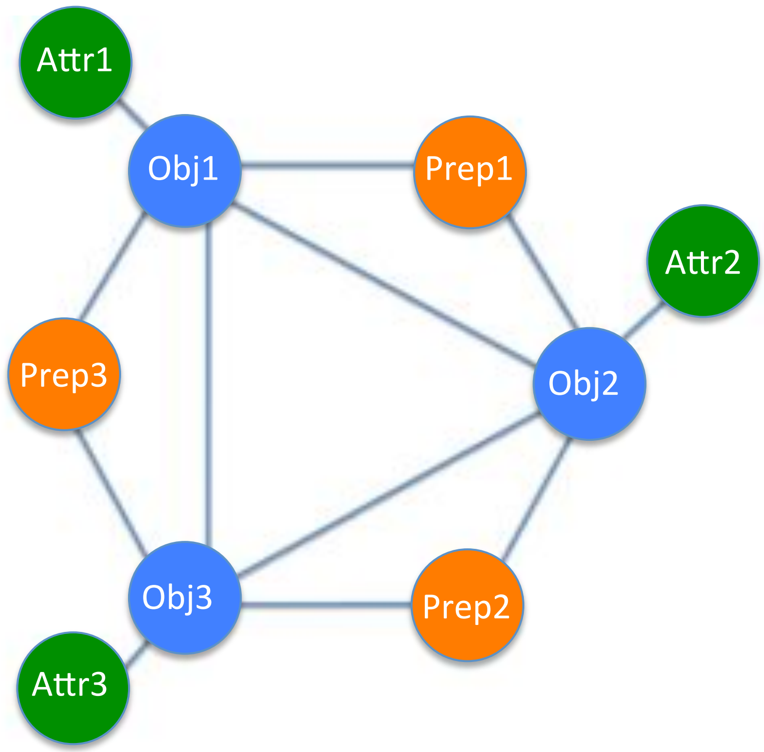
beside(c,b) .19

...

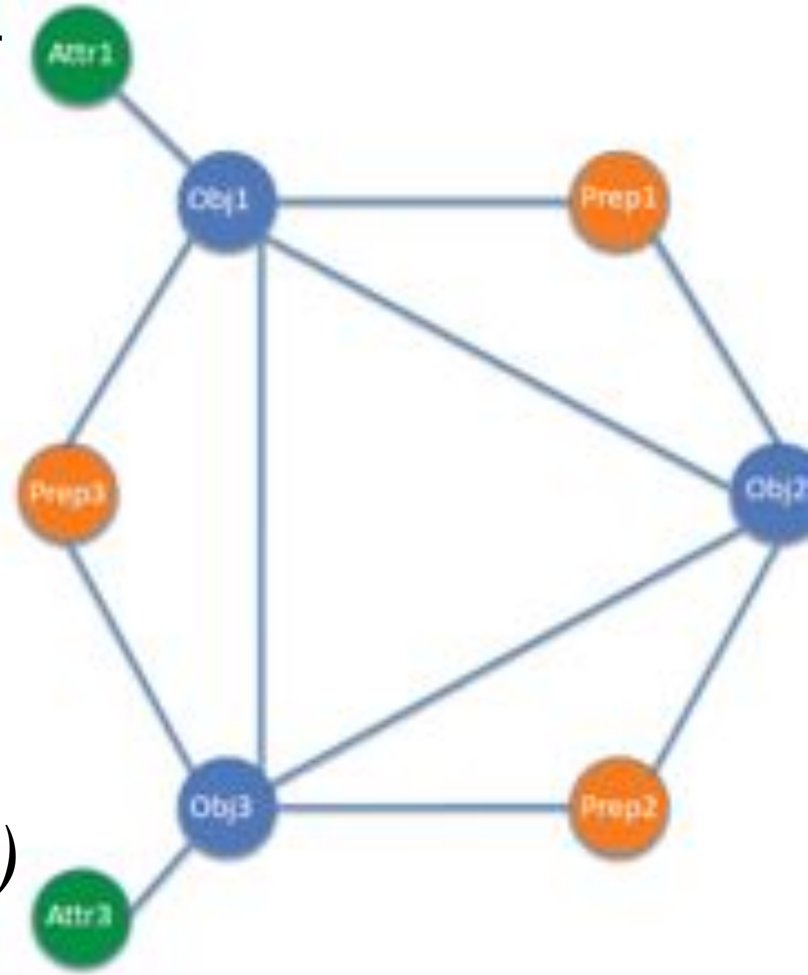
Generate natural language description

Predict labeling – vision potentials smoothed with text potentials

Conditional Random Fields (CRF)



Potential Functions for CRF



unary potentials

$$\psi(\text{object}\_i)$$

$$\psi(\text{attribute}\_i)$$

$$\psi(\text{preposition}\_ij)$$

relational (binary & ternary) potentials

$$\psi(\text{attribute}\_i\,,\text{object}\_i)$$

$$\psi(\text{object}\_i,\text{preposition}\_ij,\text{object}\_j)$$

Potential Functions for CRF

Practical challenge of relational potentials:

➔

observing all possible combinations of variables unlikely (limited corpus with detailed visual annotations)

~~unary potentials~~

$$\psi(\text{object}\_i)$$

$$\psi(\text{attribute}\_i)$$

$$\psi(\text{preposition}\_ij)$$

visual potentials

relational (binary & ternary) potentials

$$\psi(\text{attribute}\_i, \text{object}\_i)$$

$$\psi(\text{object}\_i, \text{preposition}\_ij, \text{object}\_j)$$

textual potentials

Computer vs Human Generated Caption

Computer: "This picture shows one person, one grass, one chair, and one potted plant. The person is near the green grass, and in the chair. The green grass is by the chair, and near the potted plant."

Human (UIUC Pascal dataset):

A. A Lemonaide stand **is manned by** a blonde child with a cookie.
B. A small child at a lemonade and cookie stand **on a city corner.**
C. Young child behind lemonade stand **eating a cookie.**

(1) formulaic, robotic and unnatural
(2) limited semantic expressiveness, especially, no verb except "be" verb

Computer: "This picture shows one person, one grass, one chair, and one potted plant. The person is near the green grass, and in the chair. The green grass is by the chair, and near the potted plant."

How can we reduce the gap between these two?

Human (UIUC Pascal dataset):

A. A Lemonaide stand **is manned by** a blonde child with a cookie.
B. A small child at a lemonade and cookie stand **on a city corner.**
C. Young child behind lemonade stand **eating a cookie.**

Web in 1995


MONEY & INVESTING UPDATE
from THE WALL STREET JOURNAL.
Front Page
STOCKS
U.S.
Small U.S.
Americas
Asia
Europe
Heard on the Street
Credit Markets
Foreign Exchange
Commodities
Mutual Funds
Wednesday, September 6, 1995
What's News —
Business and Finance
MARKETS DIARY
5 p.m. EDT
Computer Shares Lift Stocks Again; Bonds Are Weak
By DAVE PETTIT
Money & Investing Update


amazon.com
Welcome to Amazon.com Books!
One million titles, consistently low prices.
(If you explore just one thing, make it our personal notification service. We think it's very cool!)
SPOTLIGHT! -- AUGUST 16TH
These are the books we love, offered at Amazon.com low prices. The spotlight moves EVERY day so please come often.
ONE MILLION TITLES
Search Amazon.com's million title catalog by author, subject, title, keyword, and more... Or take a look at the books we recommend in over 20 categories... Check out our customer reviews and the award winners from the Hugo and Nebula to the Pulitzer and Nobel... and bestsellers are 30% off the publishers list...
EYES & EDITORS, A PERSONAL NOTIFICATION SERVICE

Web Today: Increasingly Visual

-- social media, news media, online shopping

flickr

Pinterest

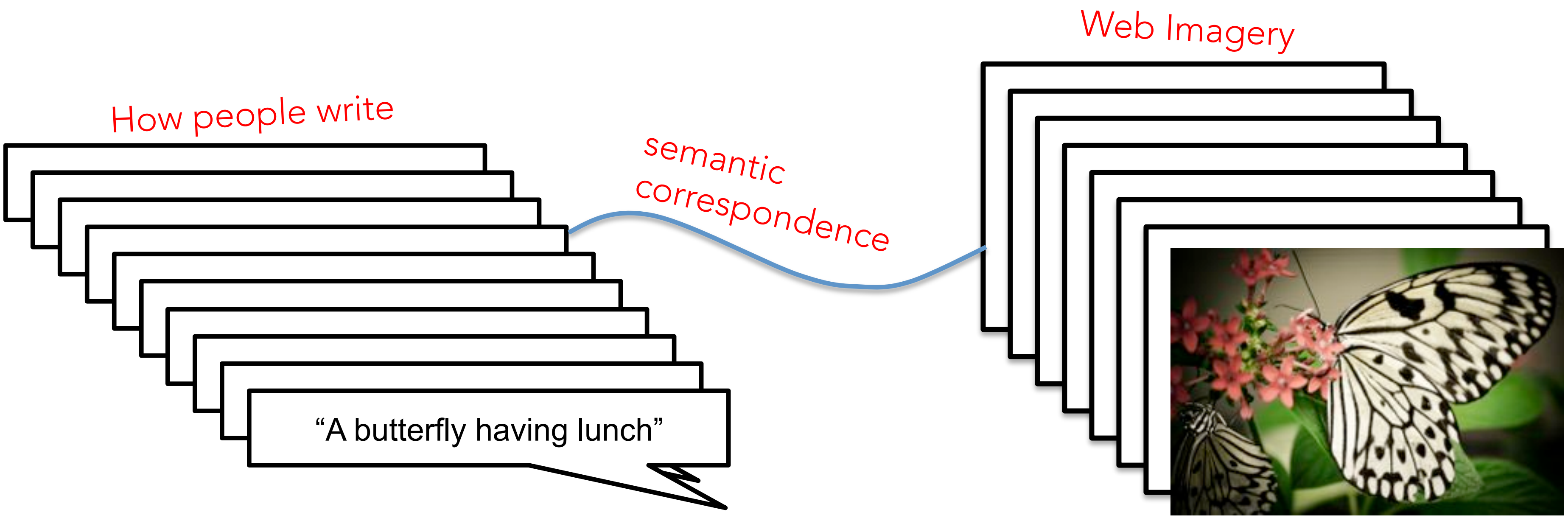
- Facebook.com has over 250 billion images uploaded as of Jun 2013
- 1.15 billion users uploading 350 million images a day on average

Task:
Learning to Describe Images in Natural Language

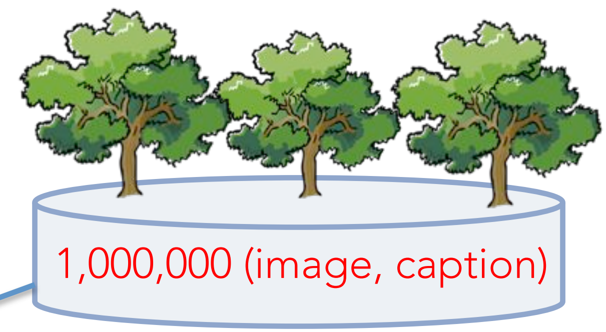
Two approaches:

I. `BabyTalk` Formulaic image description
- CVPR 2011

II. `TreeTalk` Expressive image description
- TACL 2014 (in submission), ACL 2013, ACL 2012



Operational Overview



SBU Captioned Photo Dataset
(Ordonez et al. 2011)

Given a query image (& an object)

① **Harvest** tree branches

② **Compose** a new tree by combining tree branches

Description Generation

Object appearance → NP: the dirty sheep

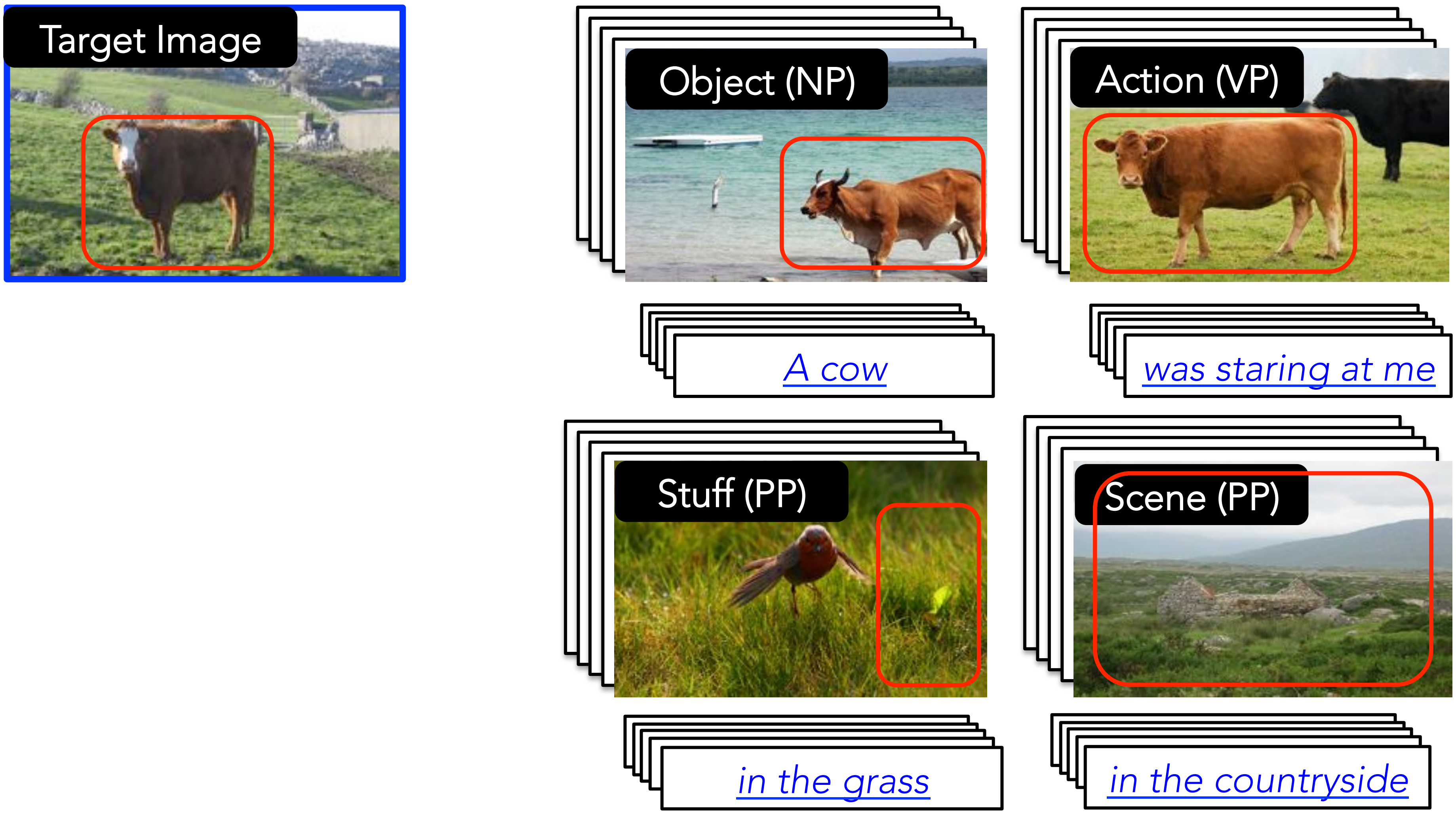
Object pose → VP: meandered along a desolate road

Scene appearance → PP: in the highlands of Scotland

Region appearance & relationship → PP: through frozen grass

Example Composition: the dirty sheep meandered along a desolate road in the highlands of Scotland through frozen grass

Input to Sentence Composition :=

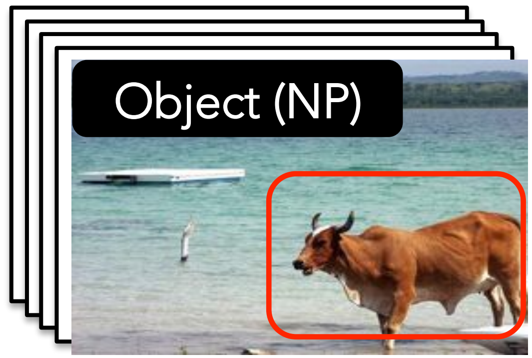


Sentence Composition :=

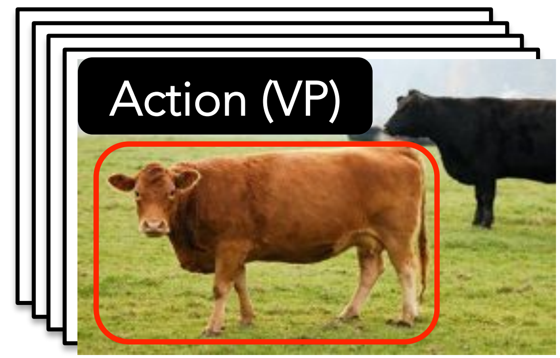
1. Select a subset of harvested phrases
2. Decide the ordering of the selected phrases

A cow
in the grass
was staring at me
in the countryside

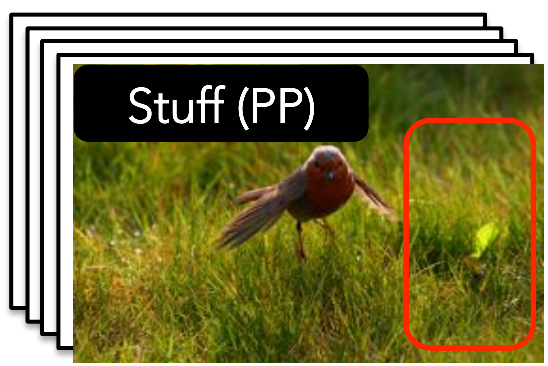
A cow
was staring at me
~~in the grass~~
in the countryside



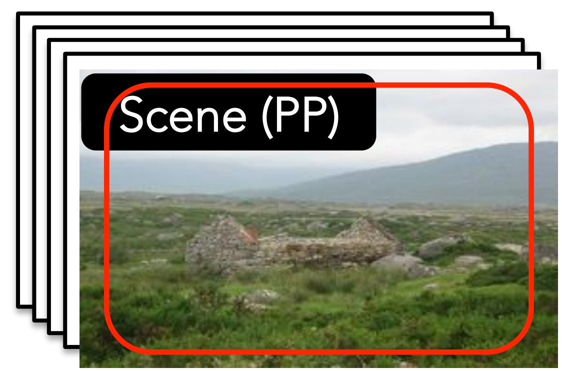
A cow



was staring at me



in the grass



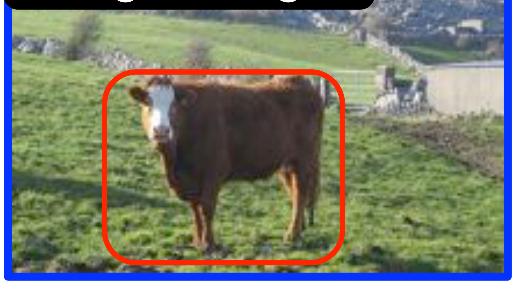
in the countryside

Sentence Composition :=

1. Select a subset of harvested phrases
2. Decide the ordering of the selected phrases

Tree Structure --- Probabilistic Context Free Grammars (PCFG)

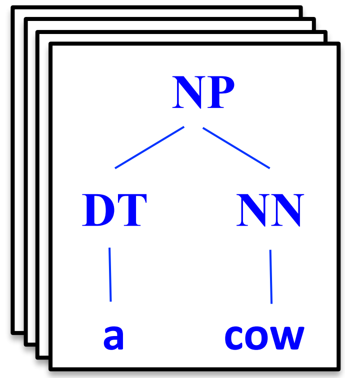
Target Image



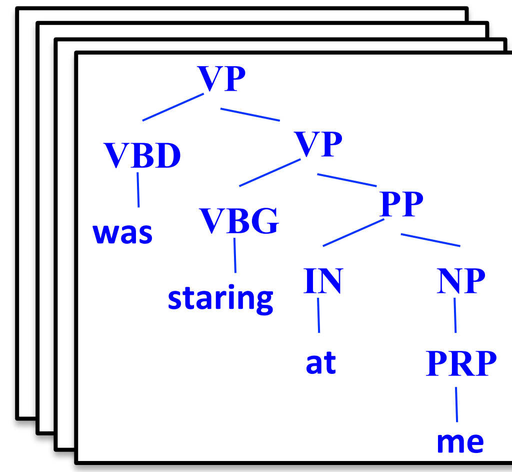
A cow
in the grass
was staring at me
in the countryside

A cow
was staring at me
~~in the grass~~
in the countryside

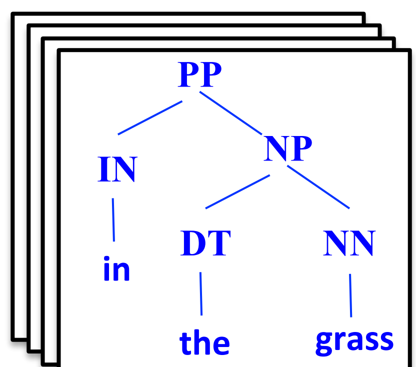
Object (NP)



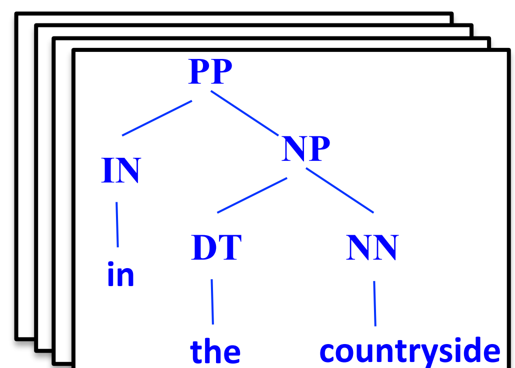
Action (VP)



Stuff (PP)

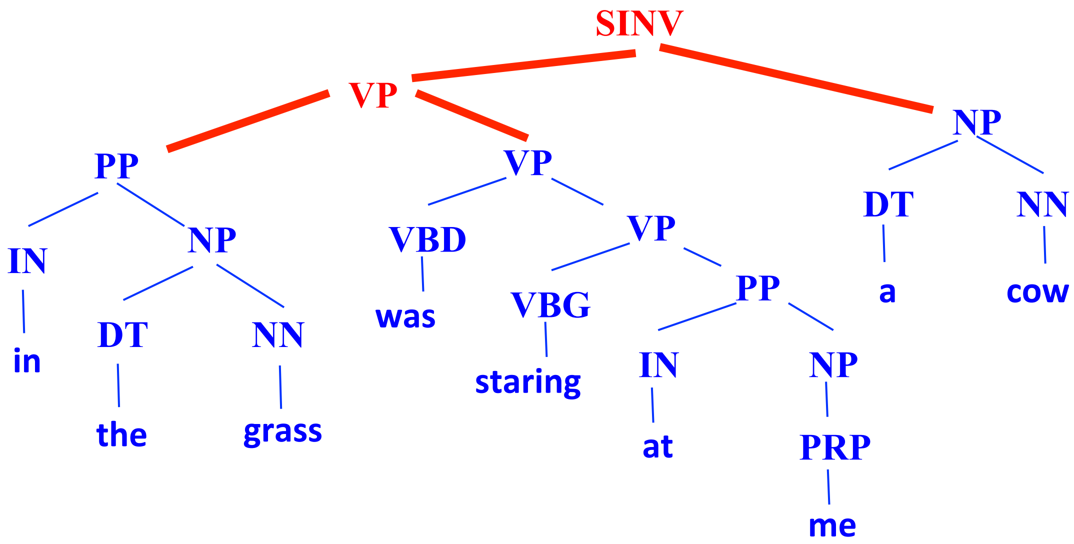


Scene (PP)



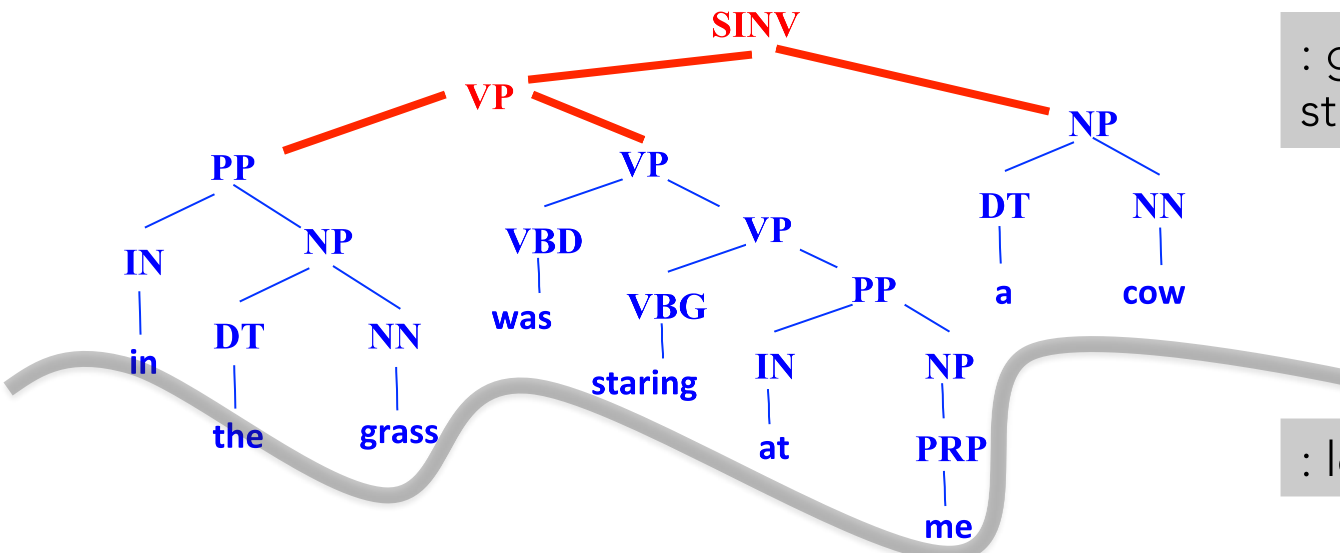
Sentence Composition :=

In the grass --- was staring at me --- a cow



Sentence Composition :=

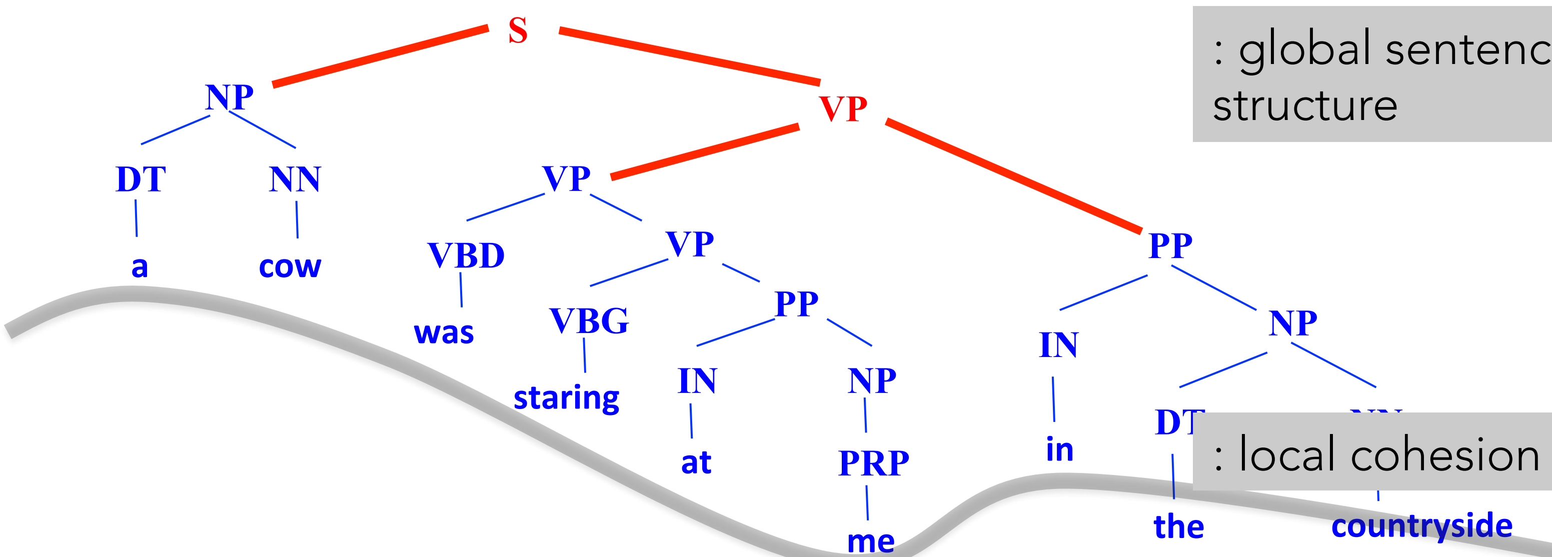
In the grass --- was staring at me --- a cow



: global sentence structure

: local cohesion

A cow --- was staring at me --- in the countryside

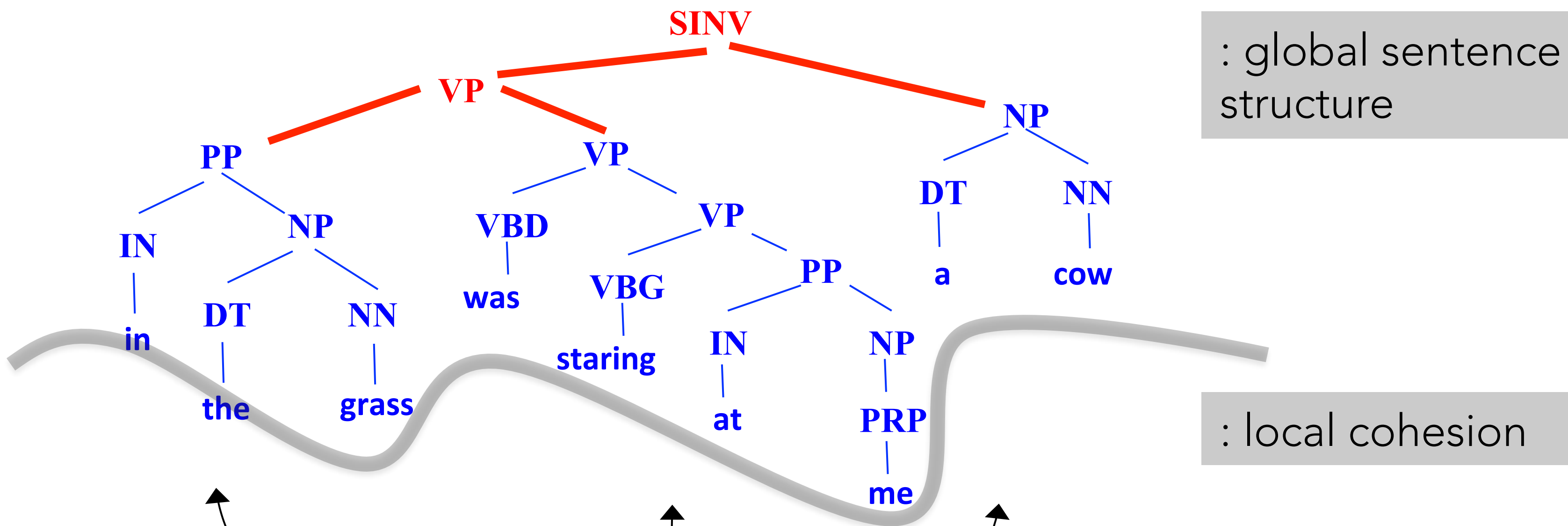


: global sentence structure

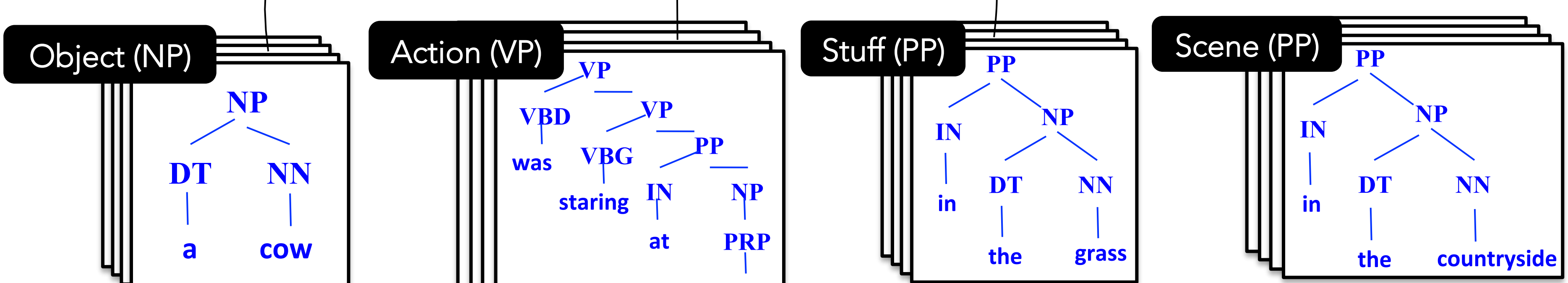
: local cohesion

Sentence Composition :=

In the grass --- was staring at me --- a cow



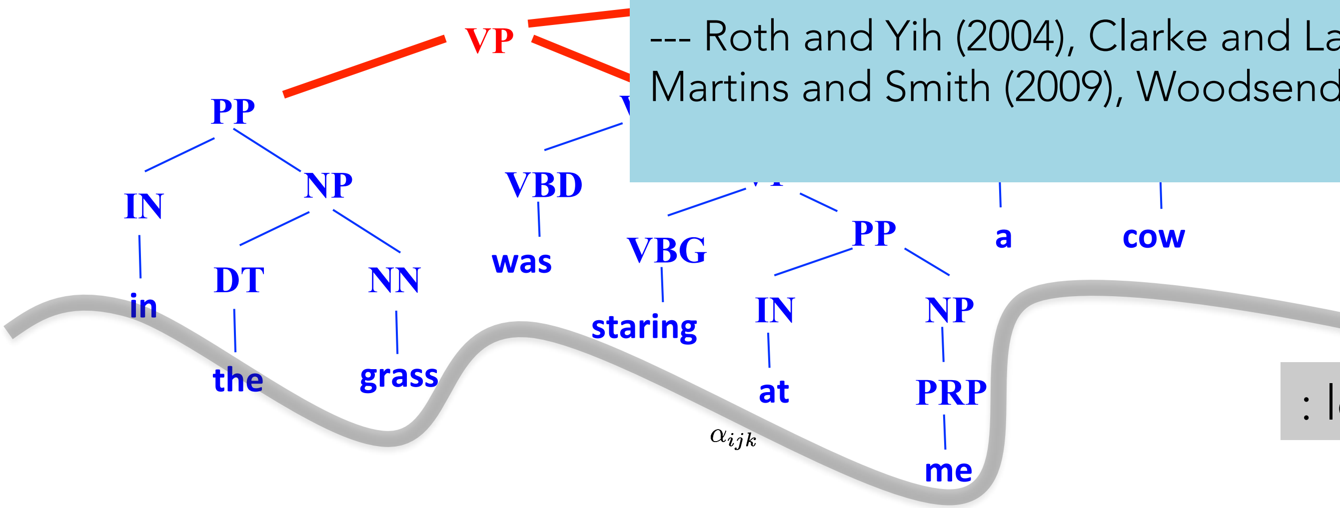
➔ different from parsing because we must consider different choices of subtree selection and re-ordering simultaneously



Sentence Composition as **Constraint Optimization** using **Integer Linear Programming**

In the grass --- was star

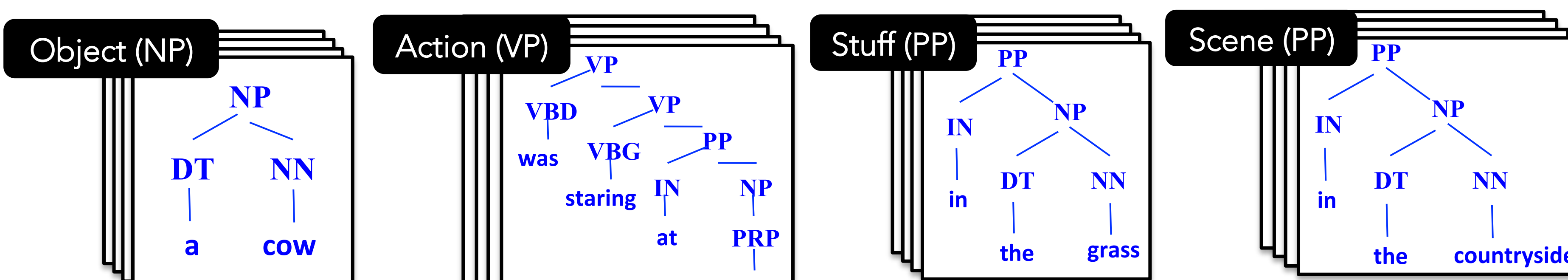
--- Roth and Yih (2004), Clarke and Lapata (2006), Martins and Smith (2009), Woodsend and Lapata(2010)



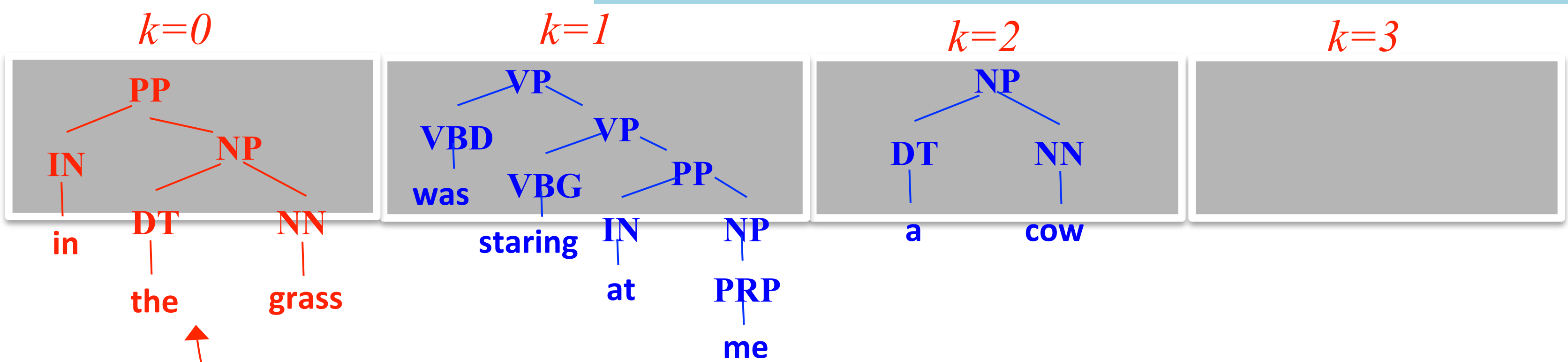
: local cohesion

➔ different from parsing because we must consider different choices of subtree selection and re-ordering simultaneously

➔ finding the optimum selection+ordering = NP-hard (~= TSP)



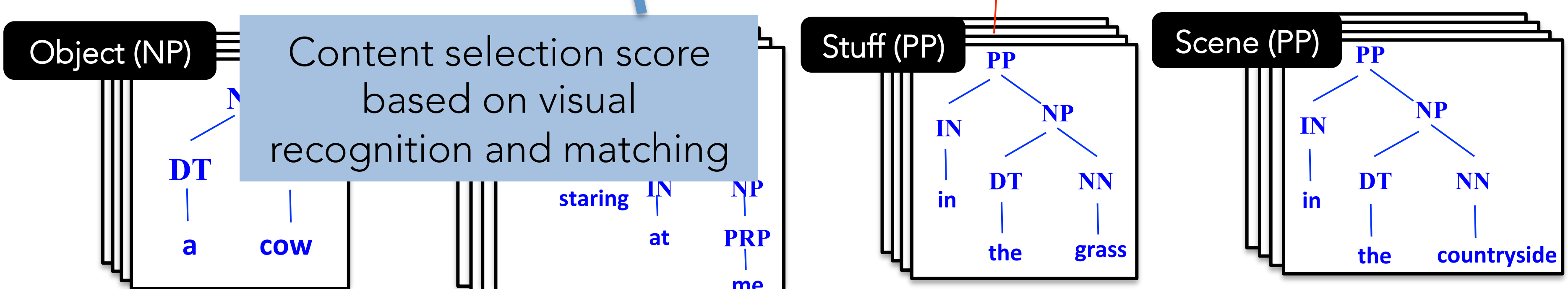
Sentence Composition as Constraint Optimization using Integer Linear Programming



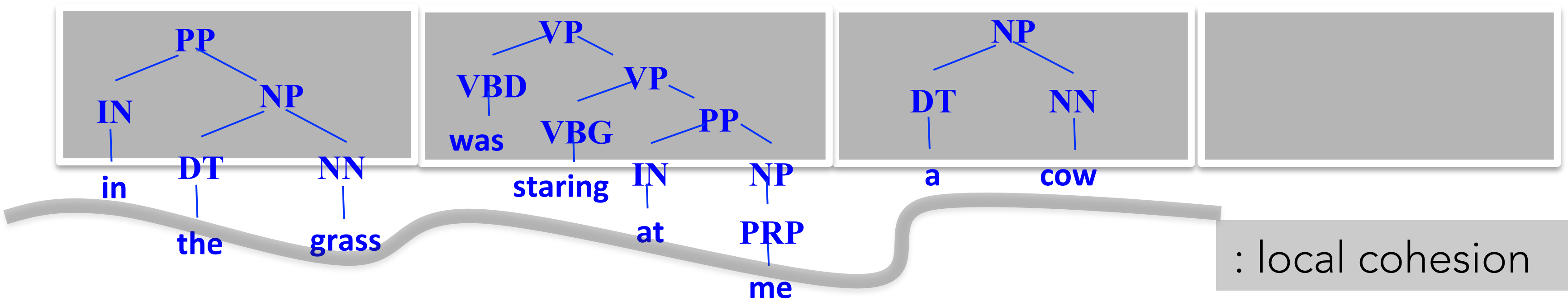
decision variable: $\alpha_{ijk} = 1 \quad \text{iff} \quad \text{phrase } i \text{ of type } j \text{ selected for position } k \in [0, N)$

objective function: $F = \sum_{ij} F_{ij} \times \sum_{k=0}^{N-1} \alpha_{ijk}$

i'th phrase from Stuff(PP)-type



Sentence Composition as **Constraint Optimization** using **Integer Linear Programming**



decision variable: $\alpha_{ijk} = 1 \quad \text{iff} \quad \text{phrase } i \text{ of type } j \text{ selected for position } k \in [0, N)$

~ ACL 2012 system

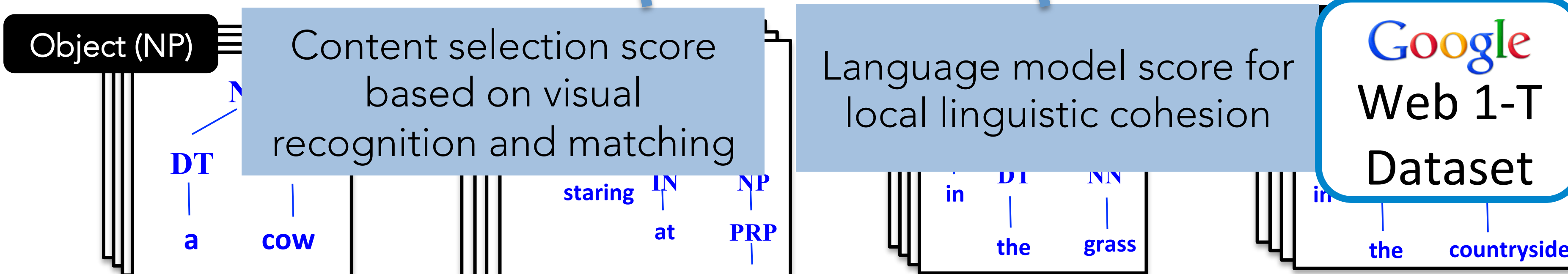
$$\boxed{\alpha_{ijkpq(k+1)} = 1 \quad \text{iff} \quad \alpha_{ijk} = \alpha_{pq(k+1)} = 1}$$

objective function: $$F = \sum_{ij} F_{ij} \times \sum_{k=0}^{N-1} \alpha_{ijk} \boxed{+ \sum_{ijpq} F_{ijpq} \times \sum_{k=0}^{N-2} \alpha_{ijkpq(k+1)}}$$

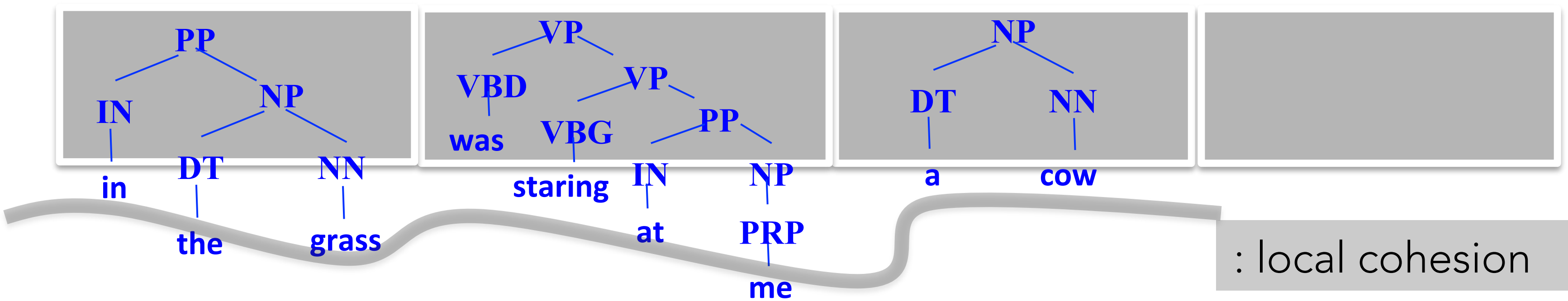
Content selection score based on visual recognition and matching

Language model score for local linguistic cohesion

Google Web 1-T Dataset



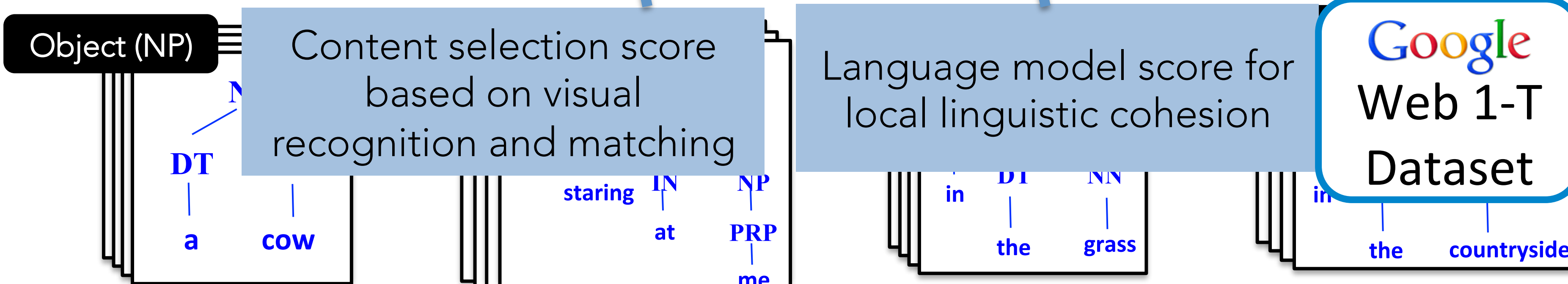
Sentence Composition as **Constraint Optimization** using **Integer Linear Programming**



decision variable: $\alpha_{ijk} = 1 \quad \text{iff} \quad \text{phrase } i \text{ of type } j \text{ selected for position } k \in [0, N)$

$$\alpha_{ijkpq(k+1)} = 1 \quad \text{iff} \quad \alpha_{ijk} = \alpha_{pq(k+1)} = 1$$

objective function: $F = \sum_{ij} F_{ij} \times \sum_{k=0}^{N-1} \alpha_{ijk} + \sum_{ijpq} F_{ijpq} \times \sum_{k=0}^{N-2} \alpha_{ijkpq(k+1)}$

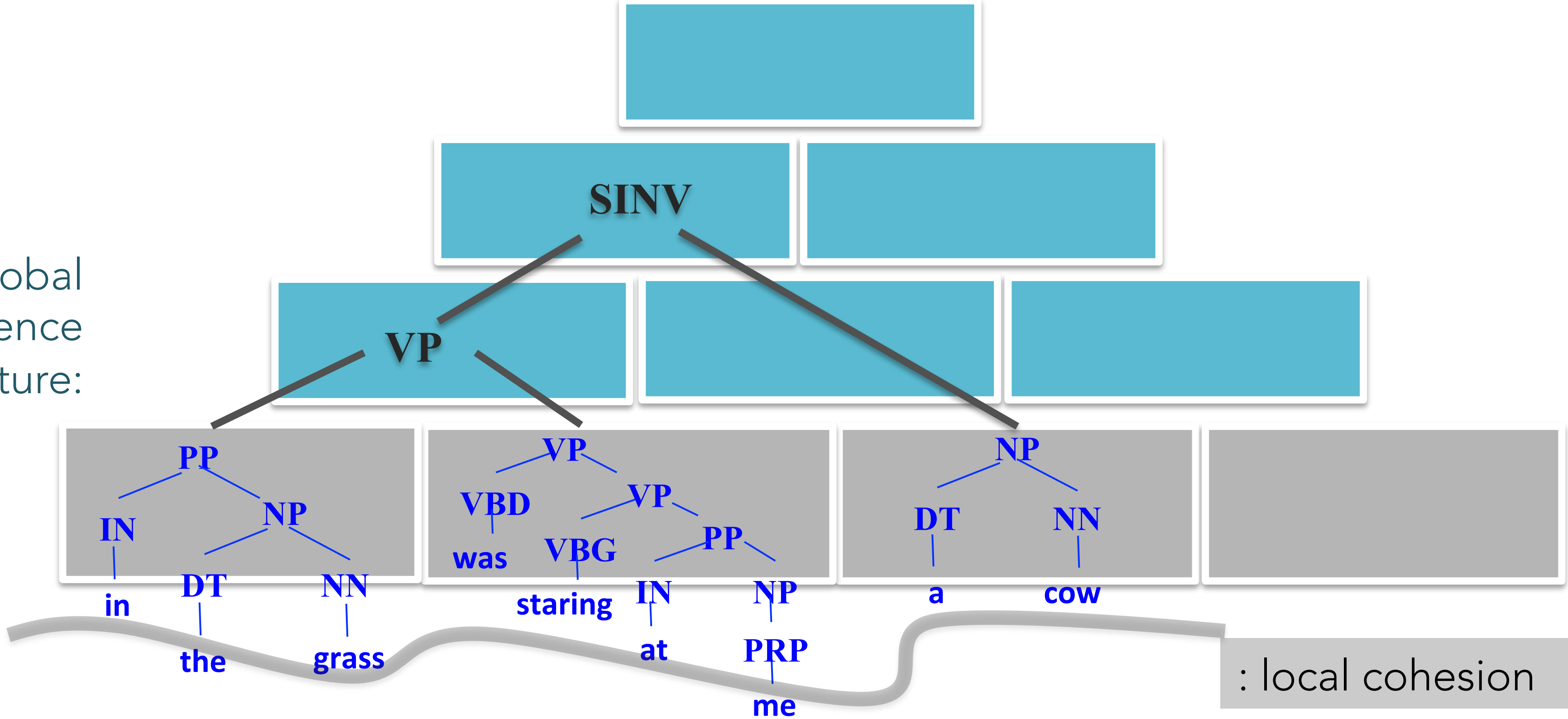


Content selection score based on visual recognition and matching

Language model score for local linguistic cohesion

Google Web 1-T Dataset

global sentence structure:



: local cohesion

decision variable: $\alpha_{ijk} = 1$ iff phrase i of type j selected for position $k \in [0, N)$

$$\alpha_{ijkpq(k+1)} = 1 \quad \text{iff} \quad \alpha_{ijk} = \alpha_{pq(k+1)} = 1$$

objective function: $$F = \sum_{ij} F_{ij} \times \sum_{k=0}^{N-1} \alpha_{ijk} + \sum_{ijpq} F_{ijpq} \times \sum_{k=0}^{N-2} \alpha_{ijkpq(k+1)}$$

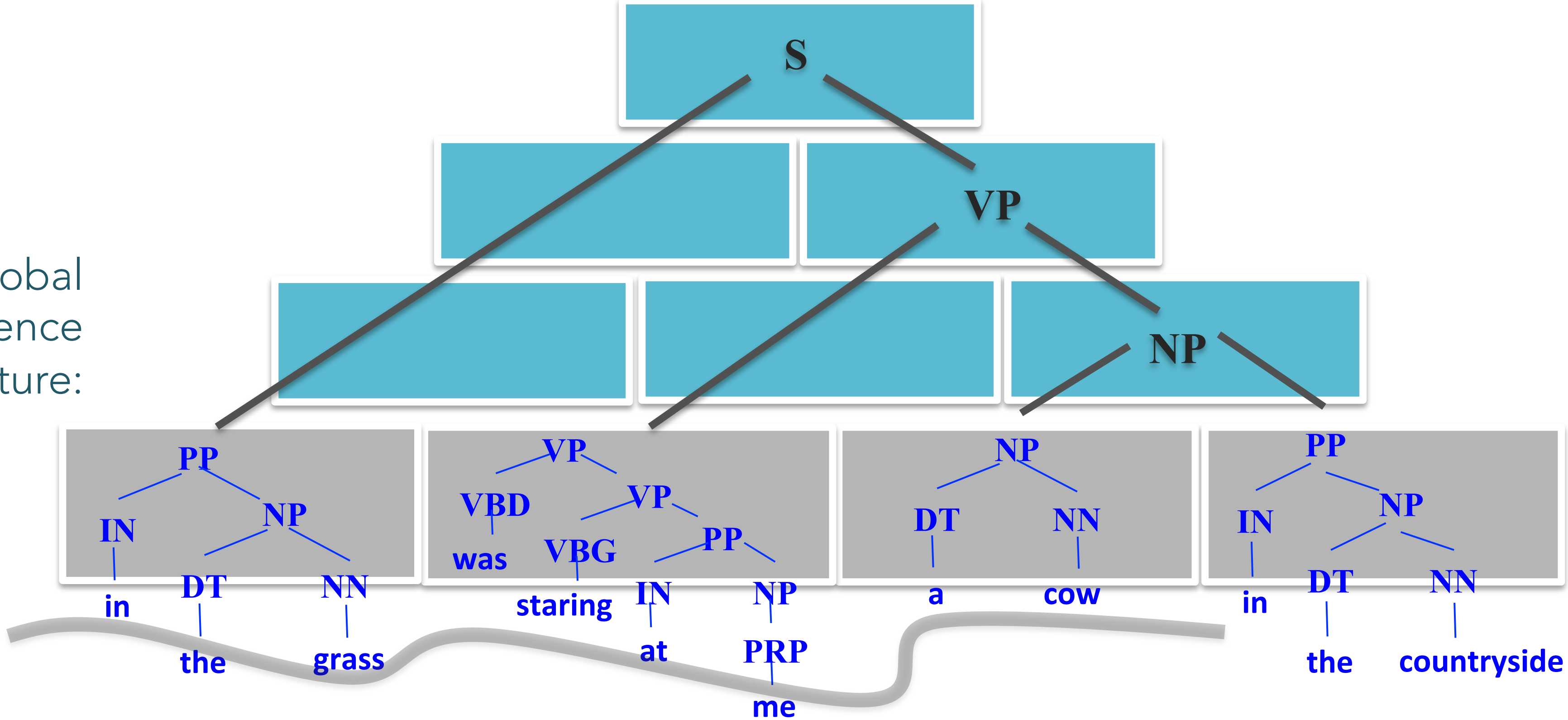
Object (NP)

Content selection score based on visual

Language model score for local linguistic cohesion

Google Web 1-T

global sentence structure:



decision variable: $\alpha_{ijk} = 1 \quad \text{iff} \quad \text{phrase } i \text{ of type } j \text{ selected for position } k \in [0, N)$

$$\alpha_{ijkpq(k+1)} = 1 \quad \text{iff} \quad \alpha_{ijk} = \alpha_{pq(k+1)} = 1$$

objective function: $F = \sum_{ij} F_{ij} \times \sum_{k=0}^{N-1} \alpha_{ijk} + \sum_{ijpq} F_{ijpq} \times \sum_{k=0}^{N-2} \alpha_{ijkpq(k+1)}$

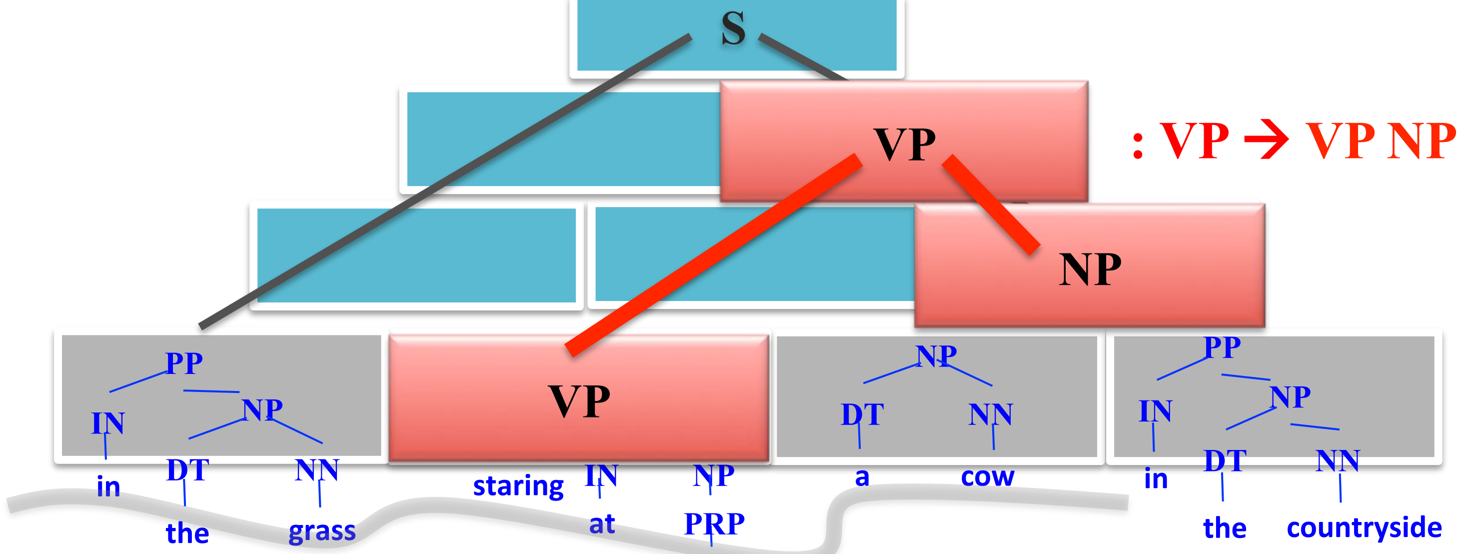
Object (NP)

Content selection score based on visual

Language model score for local linguistic cohesion

Google Web 1-T

global sentence structure:



decision variable: $\alpha_{ijk} = 1$ iff phrase i of typ... selected

for position k ...

$\alpha_{ijkpq(k+1)} = 1$ iff $\alpha_{ijk} =$ ...

Language model score for global parse tree structure

$\beta_{ijs} = 1$ iff cell ij of the matrix is assigned with PCFG tag s

α_{ijk}

$\beta_{ijkr} = 1$ iff $\beta_{ijh} = \beta_{ikp} = \beta_{(k+1)jq} = 1$

$$+ \sum_{ij} \sum_{k=i}^{j-1} \sum_{r \in R} F_r \times \beta_{ijkr}$$

objective function: $$F = \sum_{ij} F_{ij} \times \sum_{k=0}^{N-1} \alpha_{ijk} + \sum_{ijpq} F_{ijpq} \times \sum_{k=0}^{N-2} \alpha_{ijkpq(k+1)}$$

Object (NP)

Content selection score based on visual

Language model score for local linguistic cohesion

Google Web 1-T

Sentence Composition as **Constraint Optimization** using **Integer Linear Programming**

Constraints:

Consistency between
sequence variables ------ α_{ijk}
& tree leaf variables ----- β_{ijs}

$$\forall_{ijk}, \alpha_{ijk} \le \sum_{s \in S^j} \beta_{kks}$$

$$\forall_k, \sum_{ij} \alpha_{ijk} = \sum_{s \in S} \beta_{kks}$$

Valid PCFG parse tree

$$\forall_{ij}, \sum_{s \in S} \beta_{ijs} \le 1$$

$$R_h = \{r \in R : r = h \to pq\}$$

$$\forall_{i,j>i,h}, \quad \beta_{ijh} = \sum_{k=i}^{j-1} \sum_{r \in R_h} \beta_{ijkr}$$

$$\forall_{k \in [1,N)}, \sum_{s \in S} \beta_{kks} \le \sum_{t=k}^{N-1} \sum_{s \in S} \beta_{0ts}$$

$$\forall ij \sum_k \gamma_{ijk} \le 1$$

Objective function:

$$F = \sum_{ij} F_{ij} \times \sum_{k=0}^{N-1} \alpha_{ijk}$$

(Content selection ~ Visual Rec)

$$+ \sum_{ijpq} F_{ijpq} \times \sum_{k=0}^{N-2} \alpha_{ijkpq(k+1)}$$

(Sequential cohesion ~ Lang Model)

$$+ \sum_{ij} \sum_{k=i}^{j-1} \sum_{r \in R} F_r \times \beta_{ijkr}$$

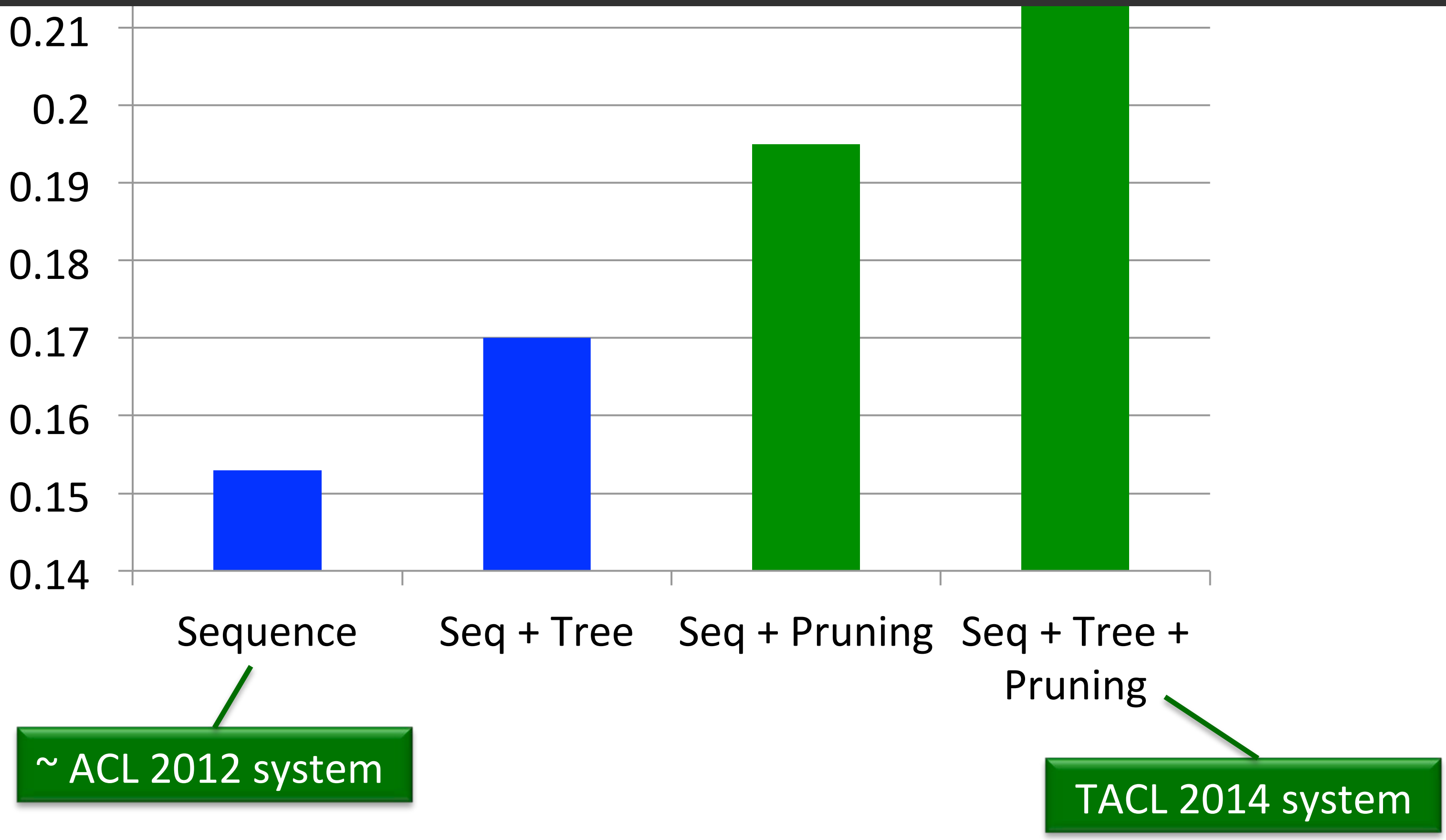
(Tree structure ~ PCFG Model)

Decision variable:

α_{ijk}	$\alpha_{ijkpq(k+1)}$	(Sequential)
β_{ijs}	β_{ijkr}	(Tree structure)

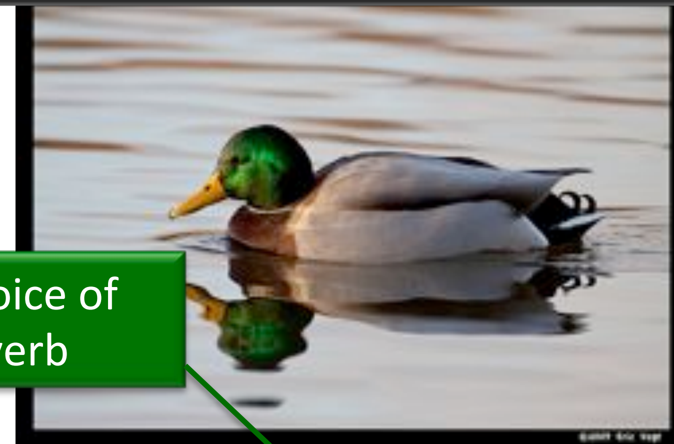
Machine Caption VS Human Caption (forced choice w/ Amazon Mechanical Turk)

- Final system (seq + tree + pruning): 24% win



Good Examples

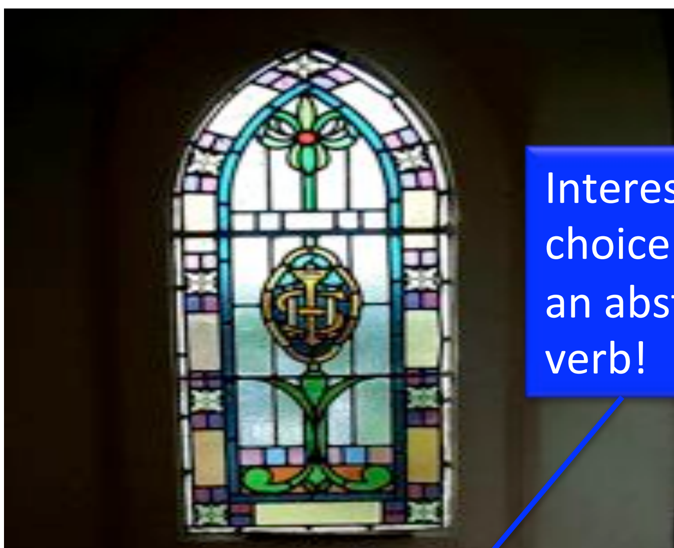
The flower was so vivid and attractive.



correct choice of an action verb

The duck sitting in the water.

Highly expressive!

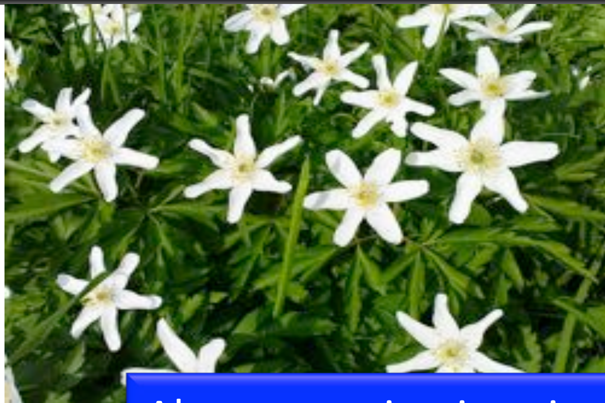


Interesting choice of an abstract verb!

This window depicts the church.

Blue flowers are running rampant in my garden.

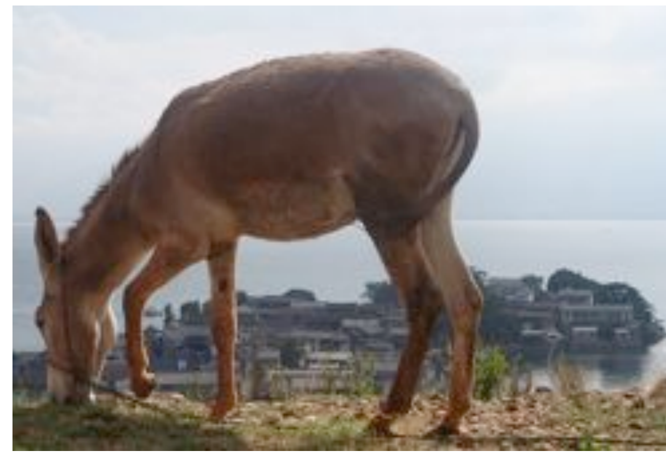
Mini Turing Test: our system wins in ~ 24 % cases!



Spring in a white dress.

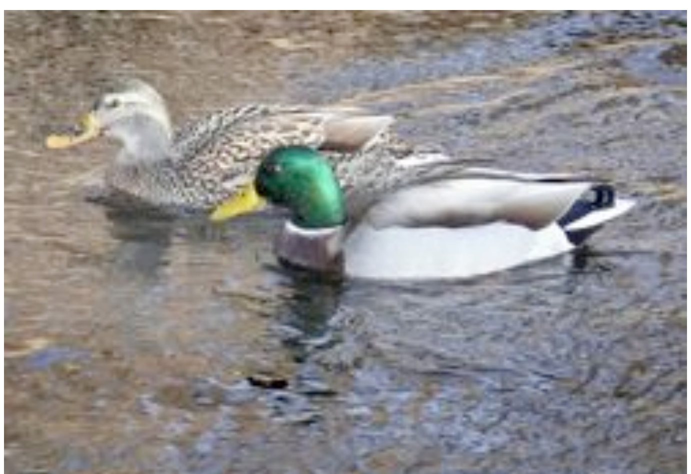
Blue flowers have no scent. Small white flowers have no idea what they are.

Almost poetic, situationally relevant



Scenes around the lake on my bike ride.

This horse walking along the road as we drove by.



Maybe the most common bird in the neighborhood, not just the most common water fowl in the neighborhood!

The duck was having a feast.

Examples with Mistakes

The couch is definitely bigger than it looks in this photo.

Yellow ball suspended in water.

Incorrect Object Recognition

Incorrect Scene Matching

Incorrect Composition

My cat laying in my duffel bag.

A high chair in the trees.

Examples with Mistakes

A cat looking for a home.
The other cats are making the computer room. ???

The castle *known for being the home of Hamlet in the Shakespeare play.*

Acknowledgements

My PhD — Ritwik Banerjee, Song Feng, Jun Seok Kang, Polina Kuznetsova

Other PhD — Vikas Ashok, Ritwik Bose, Jianfu Chen, Vicente Ordonez, Karl Stratos, Siming Li,

MS — Manoj Harpalani, Kailash Gajulapalli, Anupam Gogar, Rohith Menon, Ruchita Sarawgi, Sandesh Singh, Longfei Xing, Girish Kulkarni, Sagnik Dhar, Visruth Premraj

Undergrad — Alyssa Mensch, **Jesse Dodge**

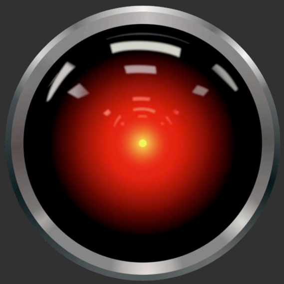
Professor — Hal Daumé III, Jia Deng, Alex Berg, Tamara Berg, Claire Cardie, Jeffrey Hancock, Rob Johnson, Michael Luca

Industry — Margaret Mitchell, Sujith Ravi, Ravi Kumar, Michael Hart, Myle Ott, Amit Goyal

Natural Language Processing
Artificial Intelligence
Machine Learning
Computer Vision

(algorithms
+ statistics
+ probabilities
+ programming
+ ...)

Question?

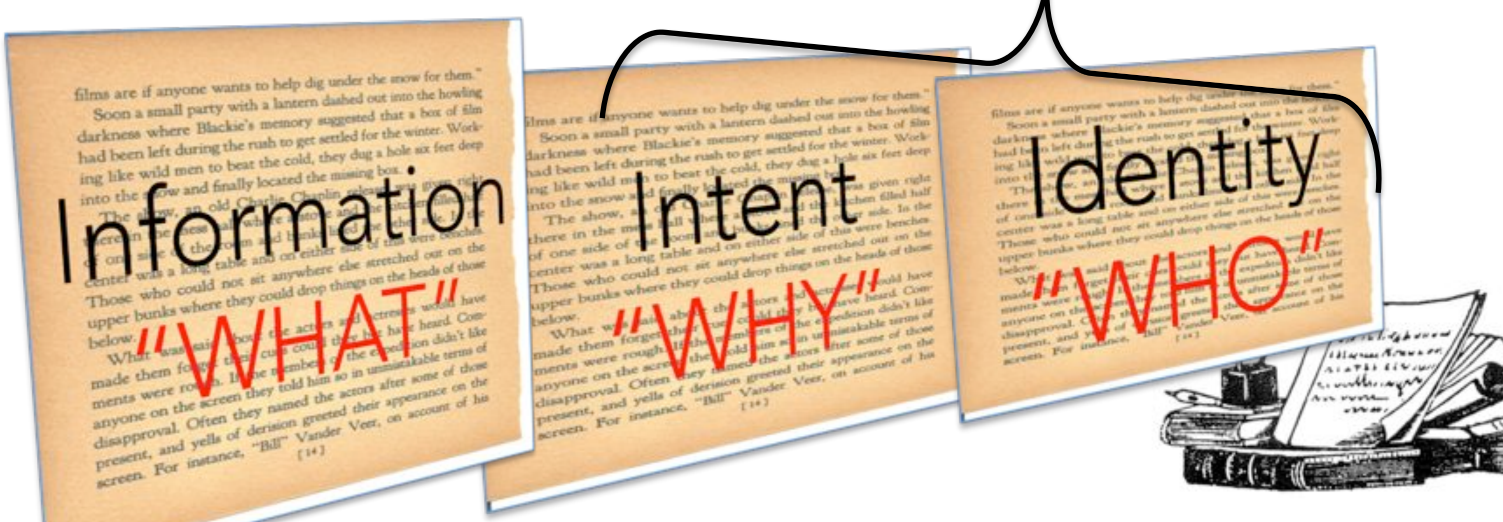


From Language to the Mind

Unconventional Case Studies:

I. Deceptive Reviews (ACL 2011)
II. Success of Novels (EMNLP 2013)

"HOW" it is said
i.e., **Writing Style**

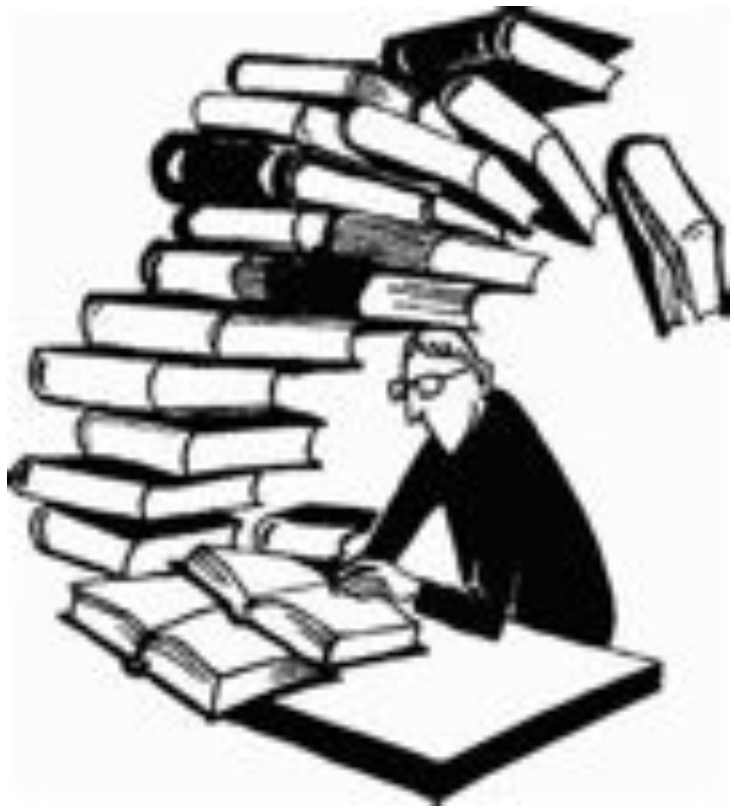


Predicting the success of novels

Novelty

Style of writing

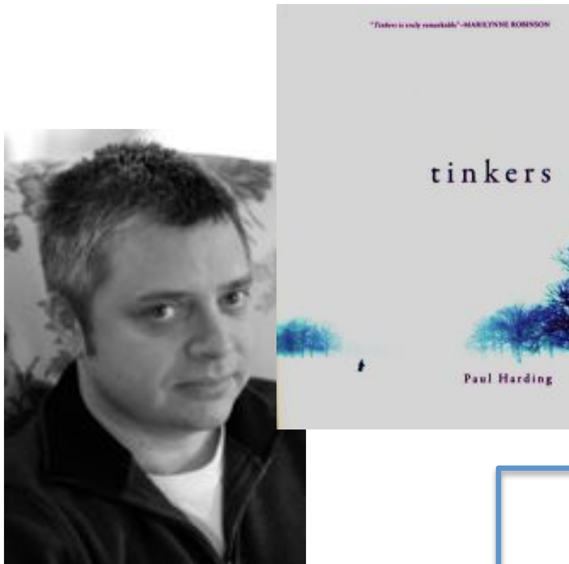
Story line



Social context

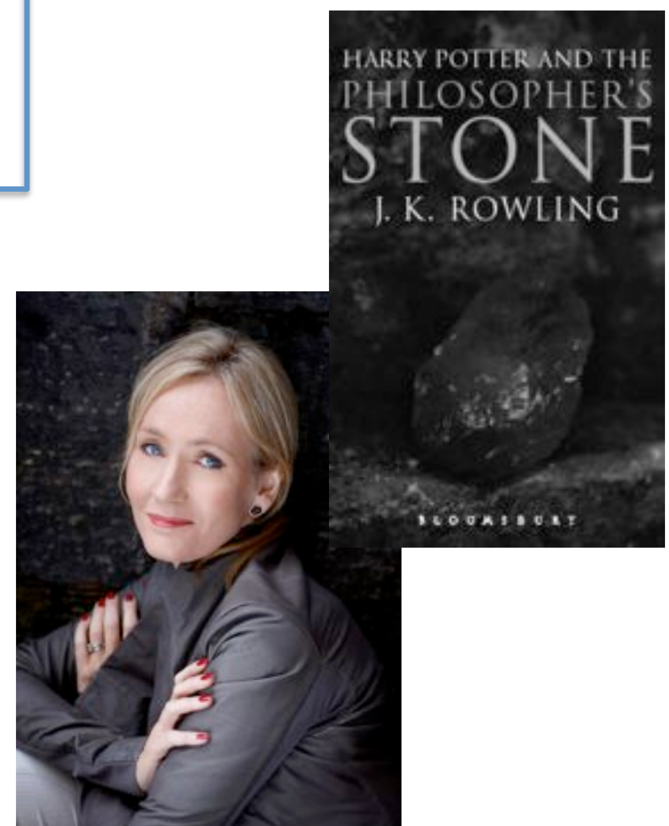
Luck !

Publishers do make mistakes



Paul Harding's "Tinkers" that won 2010 Pulitzer Prize for Fiction was rejected couple times before publication.

Rejected ~12 times before publication.



Can Computers Predict the Success of Novels without Really Reading the Book?

- based <u>only</u> on writing style
- stylistic correlates of successful novels?

How to
define success

How to
quantify success

Popularity v.s. Literary Quality

THE NEW YORK TIMES BOOK REVIEW

Best Sellers

FICTION

1 **MINDSTRETCH**, by Pamela McLaughlin. (Warner, $24.95.) Trang Martinez suspects her Pilates instructor may also be a vicious serial killer. — 1

2 **SAGEKNIGHTS OF DARKHORN**, by Gerry Banion. (Morrow, $26.95.) Astrid Soulblighter attempts to reclaim the throne from the wicked Scarking clan. The fifteenth volume of the "Bloodrealms" series. — 1 3

3 **THE BALTHAZAR TABLET**, by Tim Drew. (Doubleday, $24.95.) The murder of a cardinal leads a Yale professor and an underwear model to the Middle East, where they uncover clues to a conspiracy kept hidden by the Shriners. — 3 58

4 **GREAT FISH**, by Liz Martin. (Simon & Schuster, $23.95.) The Biblical story of Jonah, retold from the point of view of the whale. — 5 18

5 **NICK BOYLE'S SHOCK BLADE: LYNCHPIN**, by Simon Moskowitz. (Broadman & Holman, $24.99.) After a coup by Admiral Chao threatens to destroy the Internet, the ShockBlade team is forced to ally with their Chinese rivals. — 1

NONFICTION

1 **CRACKED LIKE TEETH**, by Dexter Eagan. (Morrow, $25.95.) A memoir of petty crime, drunken brawls, and recovery, by a writer who was addicted to paint thinner by age nine. — 1

2 **EMPANADAS IN WORCESTER**, by James Wirzbicki. (Farar, Straus & Giroux, $27.50.) Traveling from Khartoum to Madras to Rhode Island, a commentator for CNN suggests globalization means a stranger but friendlier world in the 21st century. — 26

3 **WRONG: THE LIBERAL PLAN TO HIJACK YOUR LIFE AND PERVERT YOUR KIDS**, by Katie Crispin. (ReganBooks/HarperCollins, $25.95.) The host of TV's "Smashmouth" takes aim at "Hollywood mind-molesters," "media jihadis," public school teachers, and others. — 1

4 **NEEDS IMPROVEMENT IN ALL AREAS**, by Margot Kilby with Sean Boyland. (ReaganBooks/HarperCollins, $29.95.) An attack on President George W. Bush, written by his former kindergarden teacher. — 3 4

5 **JOCKSTRAPS AIN'T FOR EATING**, by J. D. Preggerson. (St. Martin's, $29.95.) The former — 7 2

Project Gutenberg Downloads

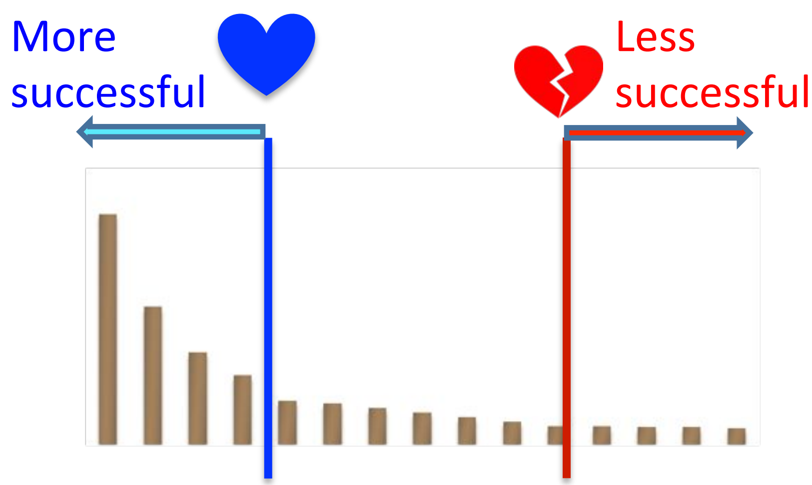
2013-10-10

last 7 days

last 30 days

Dataset

- Project Gutenberg
  - free ebooks.
  - Title, author, genre, download count.
- 50 books per class, 8 genres.



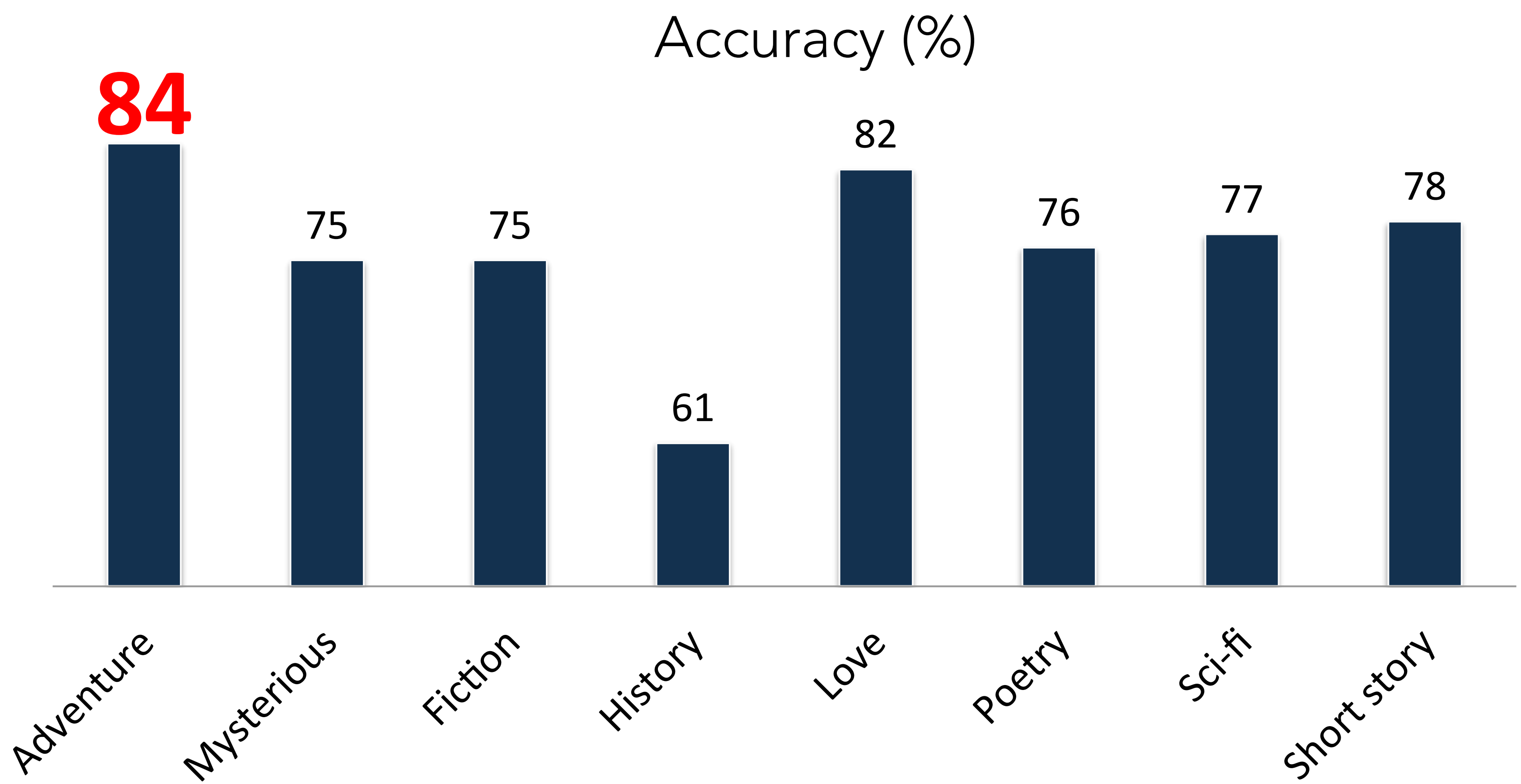
Dataset

- Project Gutenberg
  - offers over *40,000* free ebooks.
  - Title, author, genre, download count.
- 50 books per class, 8 genres.
- <=2 books per author.

~~**Authorship attribution**~~

Prediction:

(based on best performing features, 5-fold CV with SVM)



Average accuracy: 77.2%

This is Surprising Because…

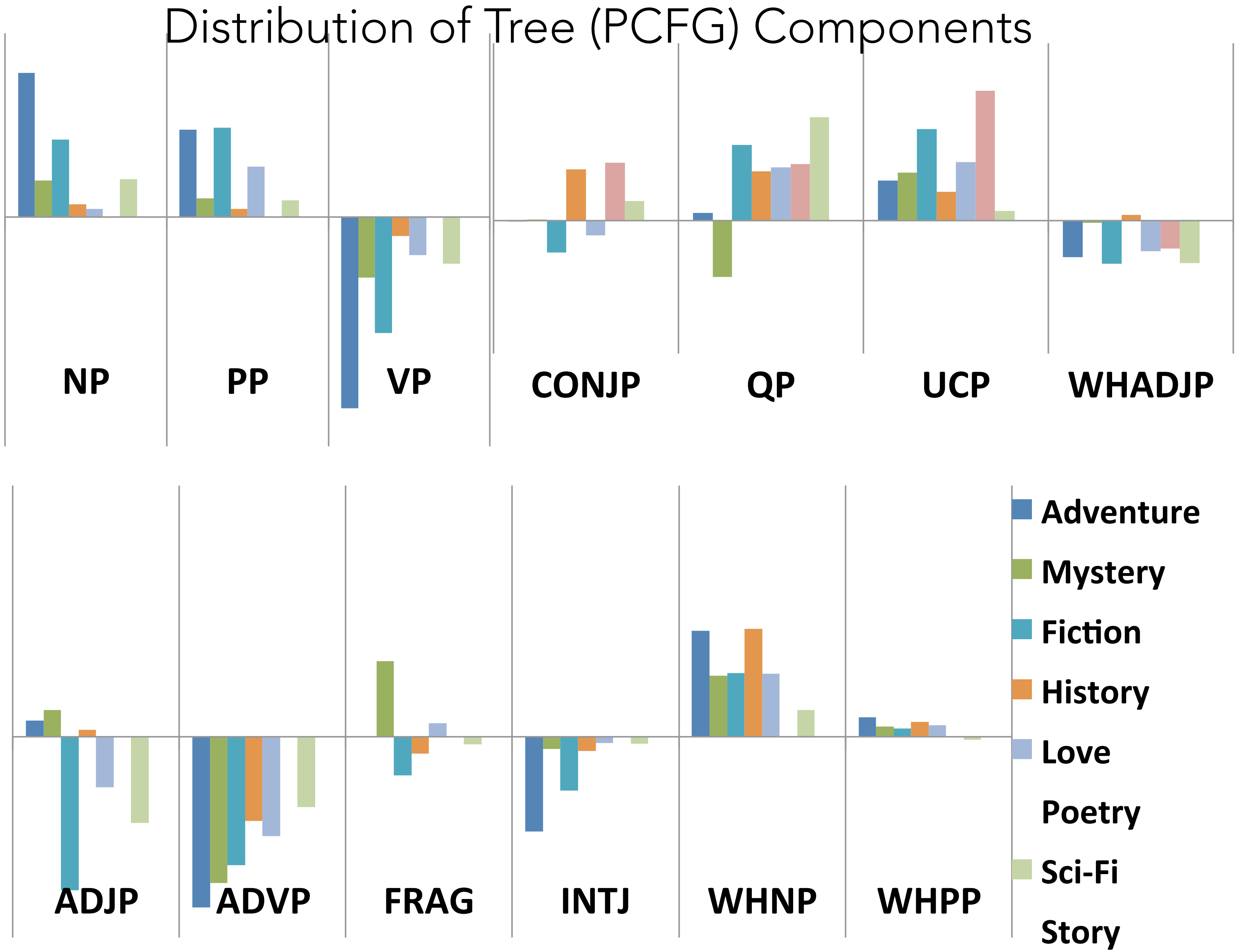
- Not considering any other influencing factor, not actually understanding the story, only looking at writing styles

- Different writers have wildly different writing styles. Should there even be stylistic commonalities shared by those different individuals?

- Testing : only the books by previously unseen authors (who presumably have his/her own unique writing style)

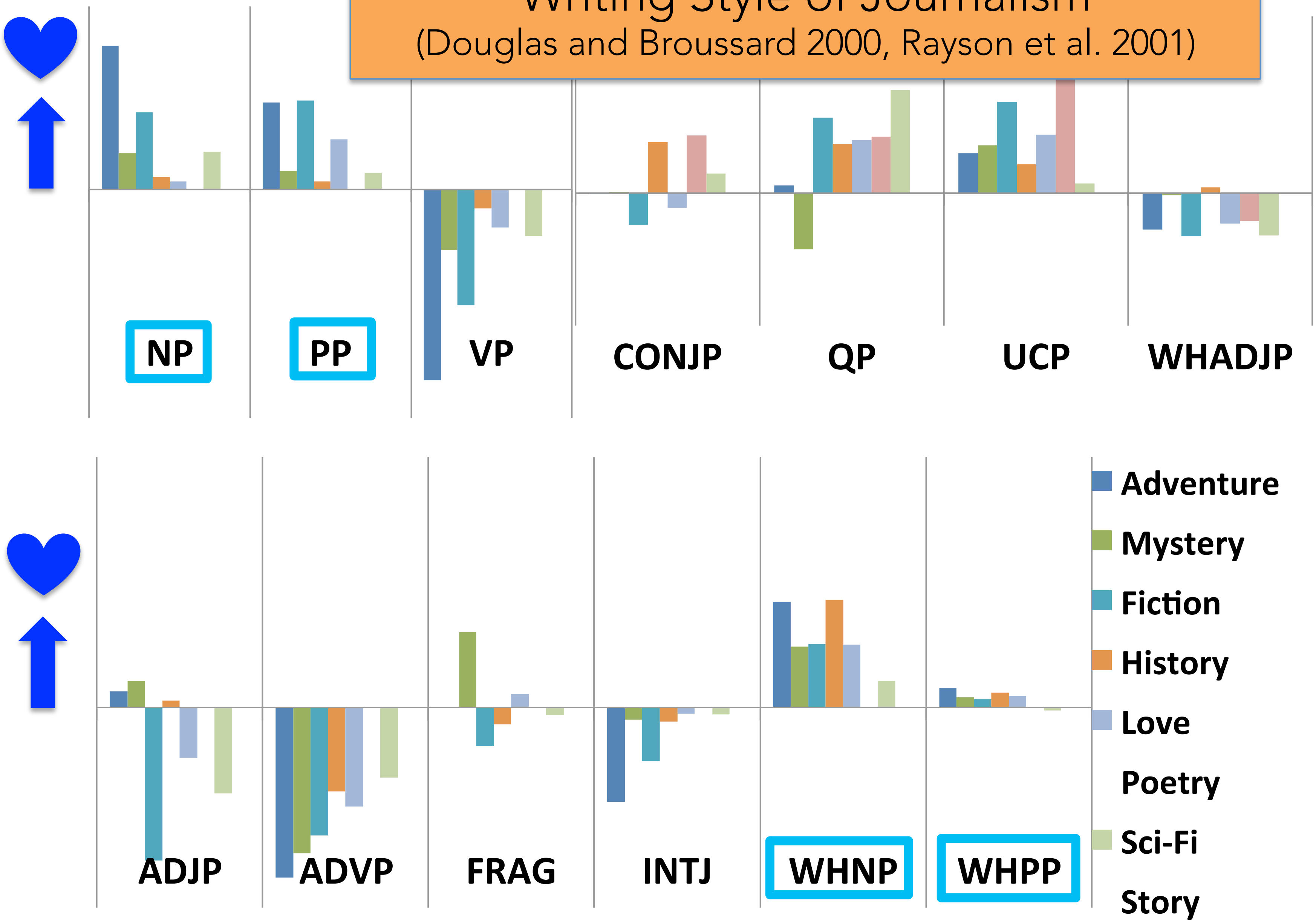
Secret Elements in Successful Novels

(only as correlates, not to be confused as causality)

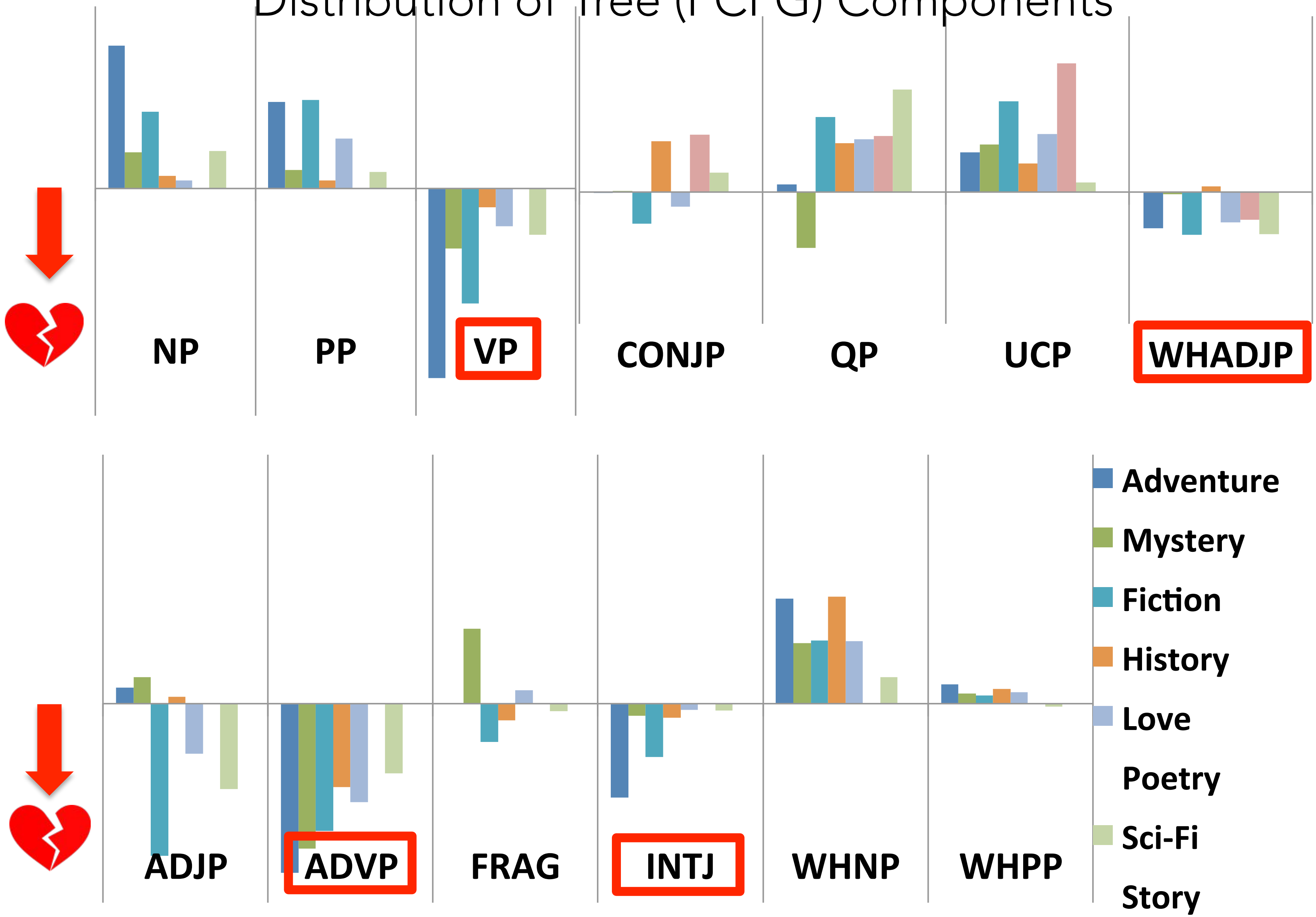


Distribution of Tree (PCFG) Components
NP
PP
VP
CONJP
QP
UCP
WHADJP
ADJP
ADVP
FRAG
INTJ
WHNP
WHPP
Adventure
Mystery
Fiction
History
Love
Poetry
Sci-Fi
Story

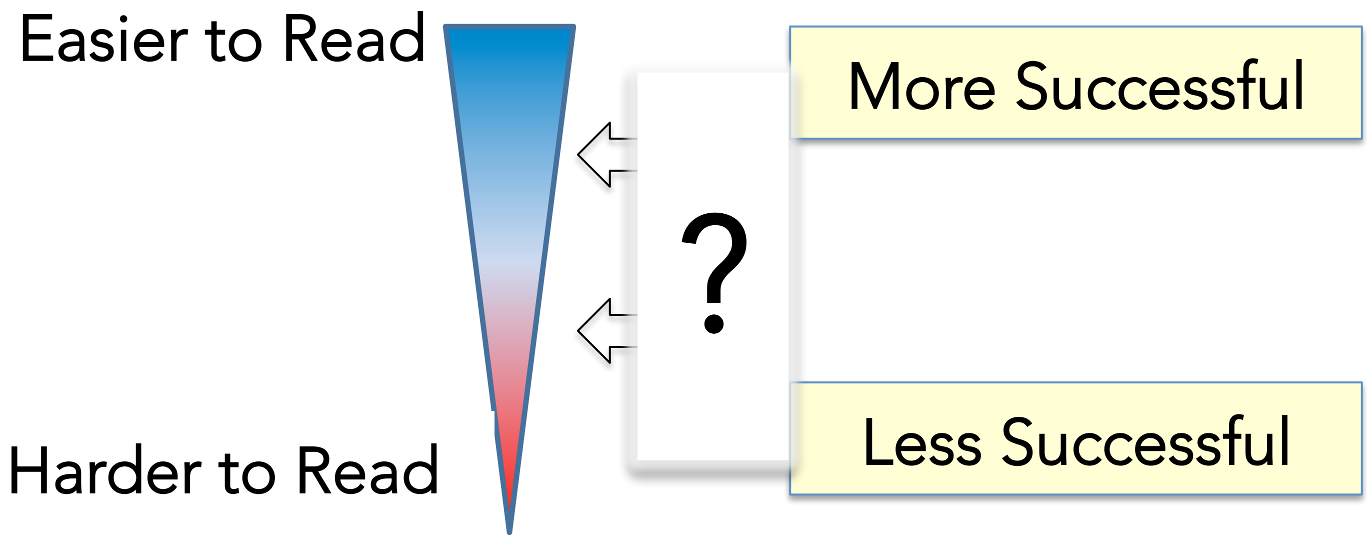
Writing Style of Journalism
(Douglas and Broussard 2000, Rayson et al. 2001)



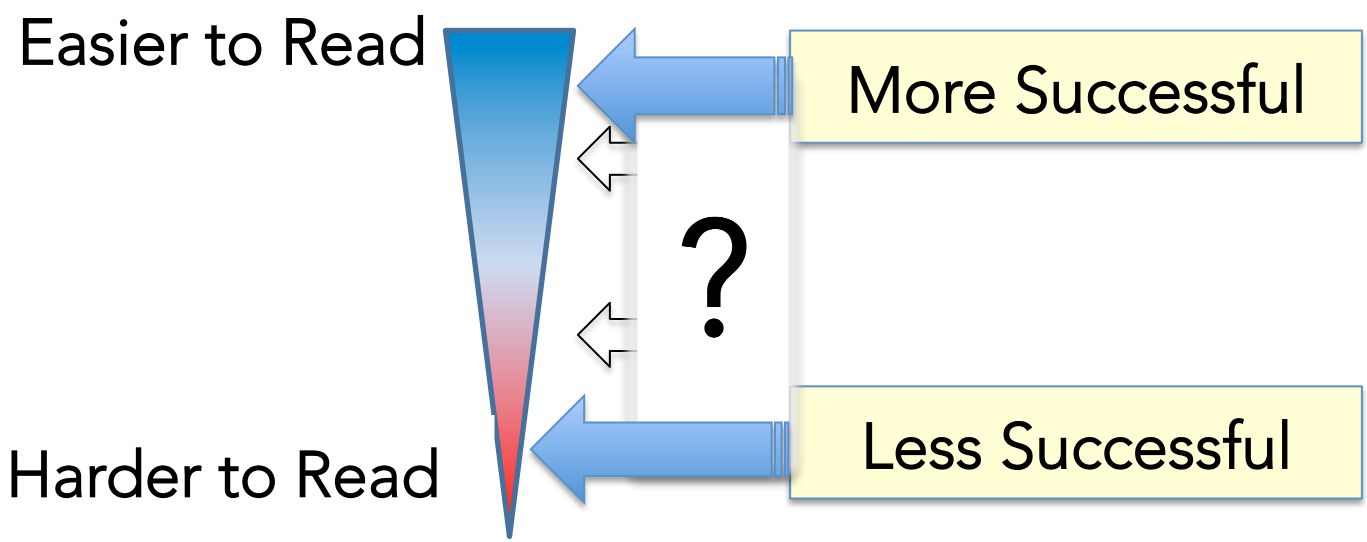
Distribution of Tree (PCFG) Components



Readability & Literary Success

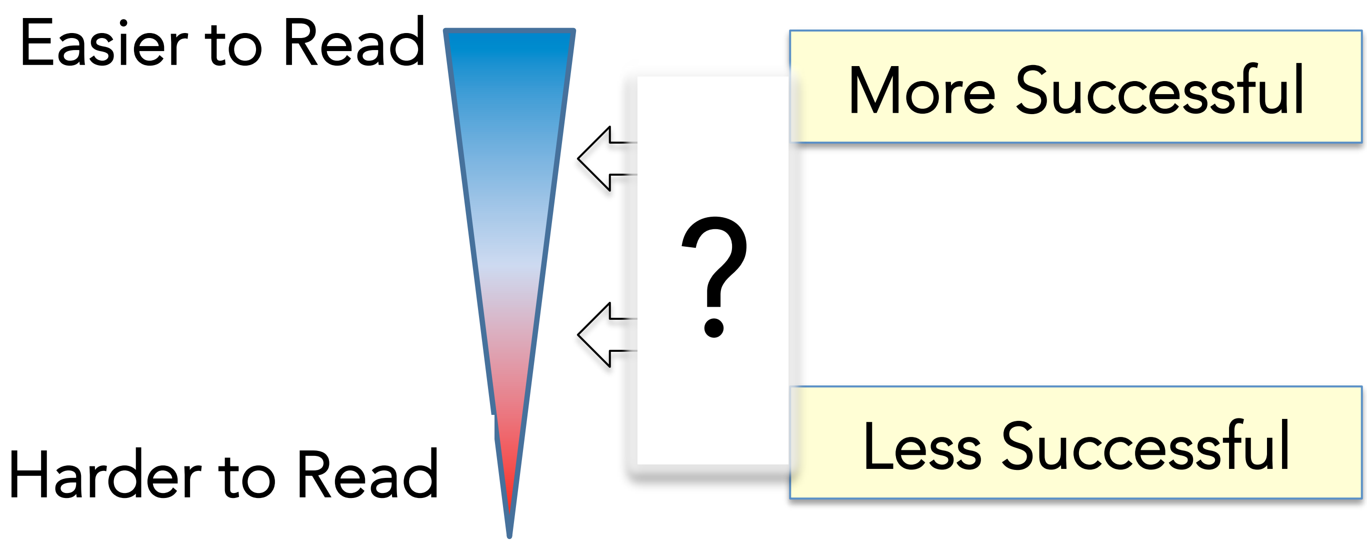


Readability & ~~Literary Success~~ Success in Academic Journals (best paper awards)

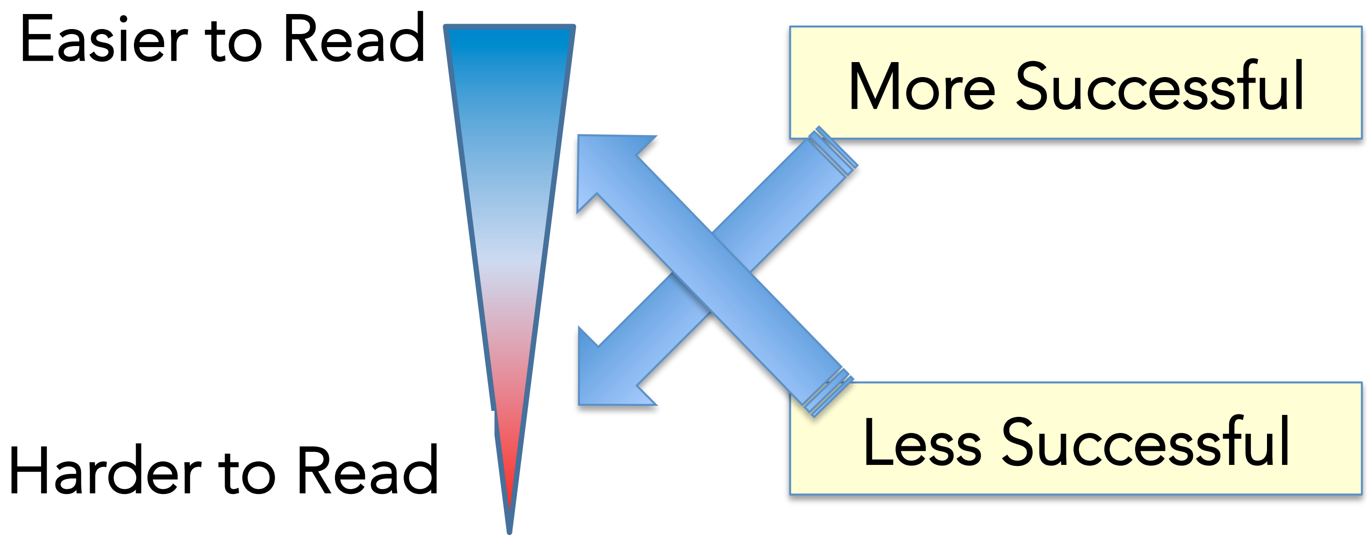


Sawyer et al (2008) @ Journal of Marketing

Readability & Literary Success



Readability & Literary Success



1. Increased use of VP= better readability (Pitler and Nenkova (2008)
2. Readability Indices:

METRIC	More Successful	Less Successful
FOG index	9.88	9.80
Flesch index	87.48	87.64

Insights into Lexical Choices (w.r.t. Adventure Genre)

Less successful: explicit telling

- verbs that are **explicitly descriptive** of actions and emotions: want, went, took, promise, cry, shout, jump, glare, urge
- **extreme** words: never, very, breathless, absolutely, perfectly
- **cliche**: love (desires, affair), body parts (face, arms, skin), obvious locations (beach, room, boat, avenue)

More successful: implicit showing

- verbs that describe **thought-processing**: recognized, remembered

  except for "*think*", which is a more direct and general word
- verbs for **reports** or **quotes**: said
- **prepositions**: up, into, out, after, in, within
- **discourse connectives:** and, which, though, that, as, after

From Language to the Mind

Unconventional Case Studies:

I. Deceptive Reviews
(ACL 2011)

II. Success of Novels
(EMNLP 2013)

intellectual traits
(~ cognitive identity)

Bibliography (2011 – 2013)

I. Deception & Public Opinion

- EMNLP 2013 Where Not to Eat? Improving Public Policy by Predicting Hygiene...
- ICWSM 2012 Distributional Footprints of Deceptive Product Reviews.
- ACL 2012 Syntactic Stylometry for Deception Detection
- ACL 2011 Finding Deceptive Opinion Spam by Any Stretch of the Imagination.

II. Authorship & Writing Style

- EMNLP 2012 Characterizing Stylistic Elements in Syntactic Structure.
- CoNLL 2011 Gender Attribution: Tracing Stylometric Evidence Beyond Topic...
- ACL 2011 Language of Vandalism: Improving Wikipedia Vandalism Detection..

III. Connotation

- ACL 2013 Connotation Lexicon: A Dash of Sentiment Beneath the Surface Meaning.
- EMNLP 2011 Learning General Connotation of Words using Graph-based Algorithms.

IV. Literary Success & Linguistic Creativity

- EMNLP 2013 Success with Style: Using Writing Style to Predict the Success of Novels.
- EMNLP 2013 Understanding and Quantifying Creativity in Lexical Composition.

Research Outlook

1. Many more surprising and impactful applications
   --- yet to be discovered, formulated, and explored!
2. Computers may at times perform better than humans.
3. NLP for Digital Humanities (... and for Humanities)
   --- Data-driven discovery of insights vs. hypothesis-driven