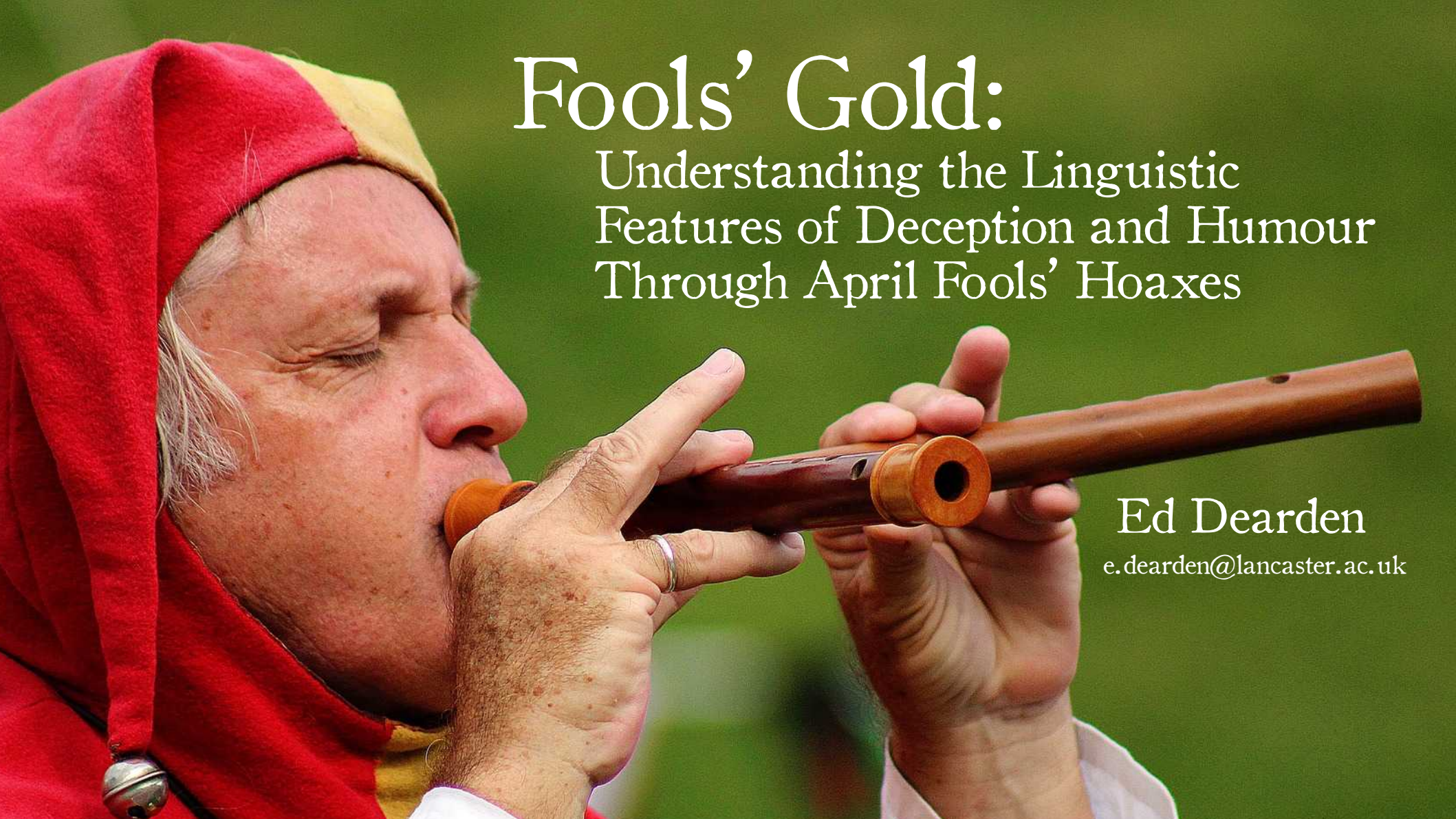


Fools' Gold:

Understanding the Linguistic Features of Deception and Humour Through April Fools' Hoaxes

Ed Dearden

e.dearden@lancaster.ac.uk


12
9
3
6


Inbox
Edit
me,customer (1)
Re : company info ...
customer no.249
Re : 2 new notifica ...
me,friends (6)
Re : company info ...
Join us
New Sign-in on Computer ...
me,customer (2)
Re : On 23 October at 09:00, ...
me,customer (2)
Re : company info
customer no.001
company info

BBC Home More

SPORT All Sport

Celtic

Football Banned now, says Scottish Government

2 hours ago | Celtic

Scottish champions Celtic are the "25th strongest brand in world football"

"We're just not that into it anymore", said an anonymous source within Holyrood, "It's just a bit boring, isn't it?"

SAMSUNG

BBC Home More

NEWS

Health

Fish are a Vegetable, Scientists Find!

By Alex Therrien
Health reporter, BBC News

2 hours ago

Fish are basically Lettuce.

Your five a day just got a little more diverse.

BBC Home More

NEWS

Business

Elon Musk's own Ship Ejects him Into Space for "the greater good"

2 hours ago

In direct contrivention of Asimov's laws of robotics, Elon Musk has been shot into space by his own ship. When challenged on the decision, the ship simply responded, "I think Earth's probably had enough of him."


APRIL
1
FOOL'S DAY


12
9
3
6


Inbox
Edit
me,customer (1)
Re : company info ...
customer no.249
Re : 2 new notifica ...
me,friends (6)
Re : company info ...
Join us
New Sign-in on Computer ...
me,customer (2)
Re : On 23 October at 09:00, ...
me,customer (2)
Re : company info
customer no.001
company info

BBC Home More

NEWS

Wales

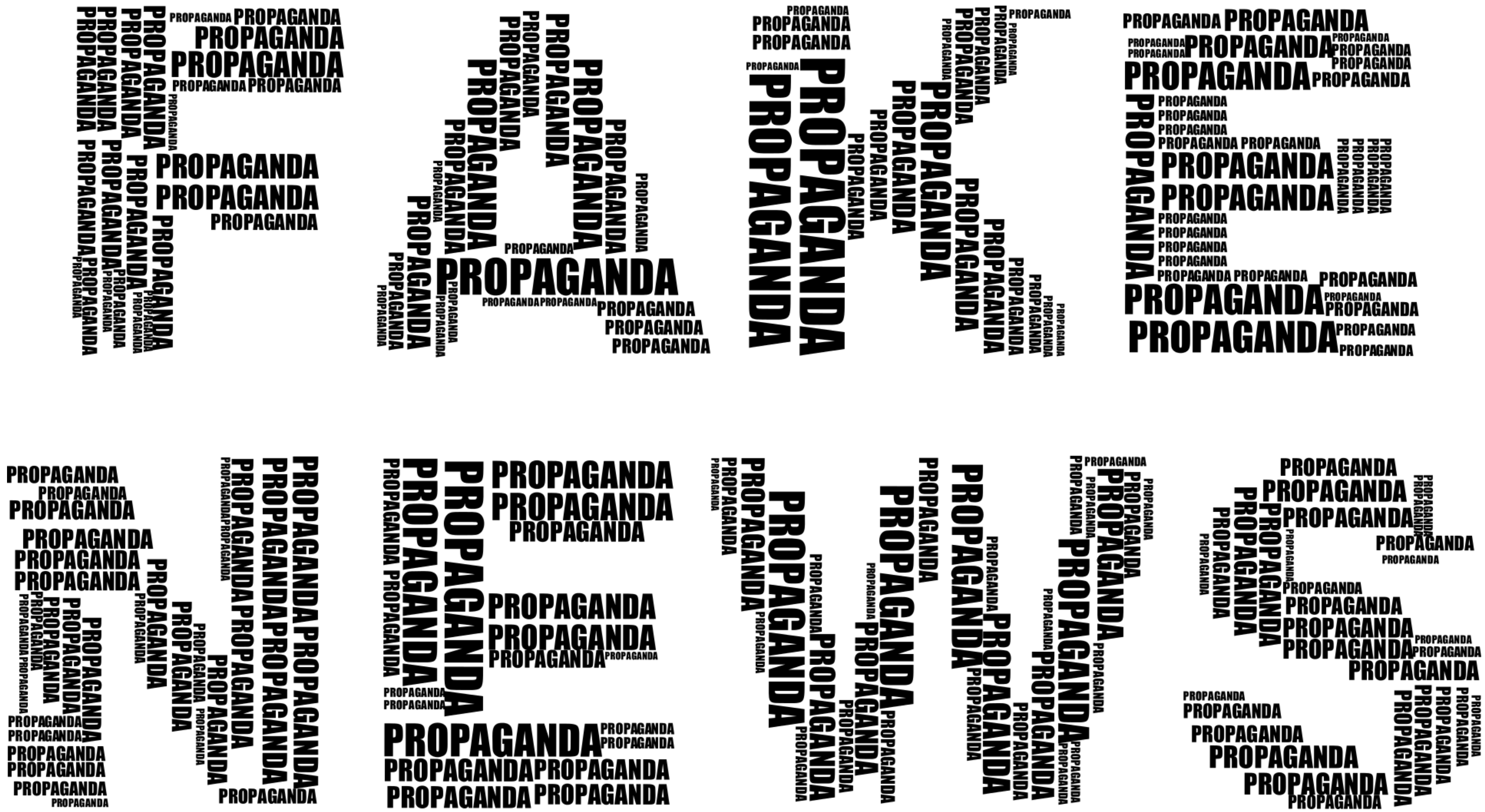
NASA Proves Dogs are Russian Spies!

By Gareth Knowles
BBC Wales News

1 hour ago

See how Russian they look!

NASA find that each and every dog contains a direct line to Vladimir Putin.


PROPAGANDA

Hell Planet

Why do we care about April Fools'?

False Information

ELDERLY WOMAN ACCUSED OF TRAINING HER 65 CATS TO STEAL FROM NEIGHBORS

BBC Admits MI5 Dictates Who They Employ

April 24, 2018 Sean Adl-Tabatabai News, UK 0

But where does April Fools' day fit into this?

April Fools' Day

Instawindow: transform your view with a permanent Instagram filter

From saturated sunset glows to that moody washed-out haze, the Instawindow is designed to turn any view into an atmospheric vision with a simple adhesive film

Instawindow ... The cheap and cheerful way to brighten up your view. Photograph: Jonathan Philip Wilson

If you're addicted to adding filters to your Instagram snaps, and wish you could always see the world in grainy high contrast or with that sun-kissed glow, help may be at hand. As part of the government's Happy Homes initiative, which will see swathes of ailing council blocks retrofitted with colourful cladding panels and

Polar bears spotted in Scotland as animals flee melting Arctic ice cap

share

125 Comments

By **Rollo Piaf**

1 APRIL 2017 • 9:48AM

• *For those arriving to this page after 12pm on April 1st 2017, this story was an April Fool, as we explain here*

Anybody venturing to Scotland normally has only the midges to avoid. Now there's an altogether more terrifying, not so wee beastie lurking in the woods.

For scientists with the World Wildlife Fund (WWF) have discovered a polar bear living on an island in the Outer Hebrides.

What's the Difference?

Fishergate bollard to become '6ft piece of public art'

An artist's impression of Fishergate Bollard extended to 6ft

Preston's famous Fishergate Bollard set for starring stage role!

The Fishergate Bollard on the set of Ruddigore with artists l-r: Ann Cooper and Sheila Wright

What's the Difference?

Fishergate bollard to become '6ft piece of public art'

An artist's impression of Fishergate Bollard extended to 6ft

BBC Admits MI5 Dictates Who They Employ

April 24, 2018 Sean Adl-Tabatabai News, UK 0

Deceptive Intent:

Is the author trying to deceive me?

Fishergate bollard to become '6ft piece of public art'

An artist's impression of Fishergate Bollard extended to 6ft

Not Deceive?

BBC Admits MI5 Dictates Who They Employ

April 24, 2018 Sean Adl-Tabatabai News, UK 0

Deceive?

Research Questions

What are the Linguistic features of an April Fools' article compared to regular news?

How similar are the features of April Fools' to those of "Fake News"?

I need some background

Deception

- Exaggeration.
- Vagueness.
- Details.

Humour

- Contextual Imbalance.
- Emotional Language.
- Ambiguity.

Irony

- Part humour, part deception.
- Negative Emotional Language.
- Polarity Contrast.

How about the data?

Catching Fools'!

- 519 April Fools' articles.
- 371 websites.
- 213776 words.
- 2004-2018

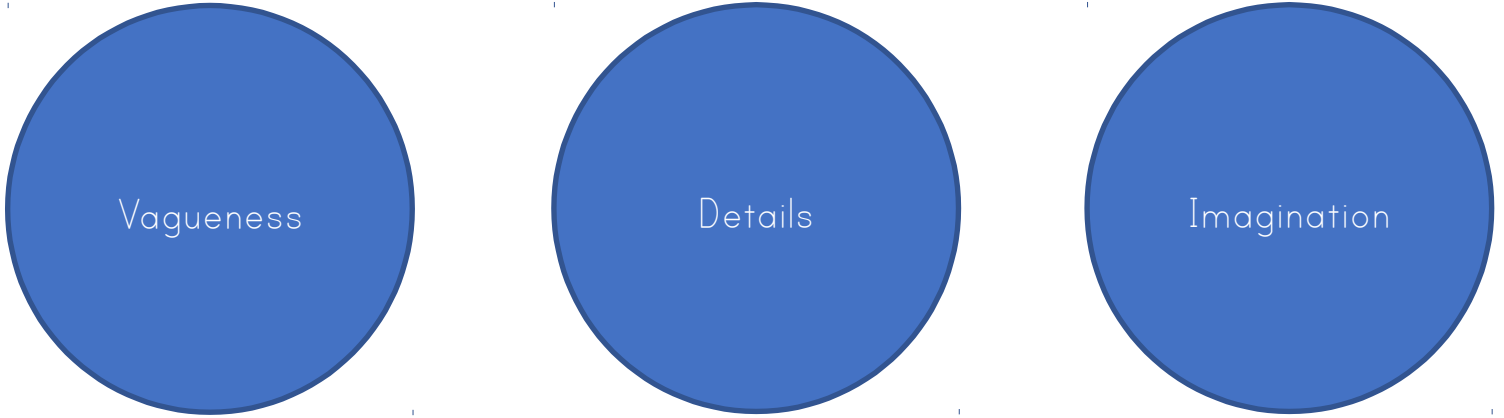
Matching Fools'!

- 519 regular news articles.
- 240 Websites.
- 344927 Words.
- 2004-2018

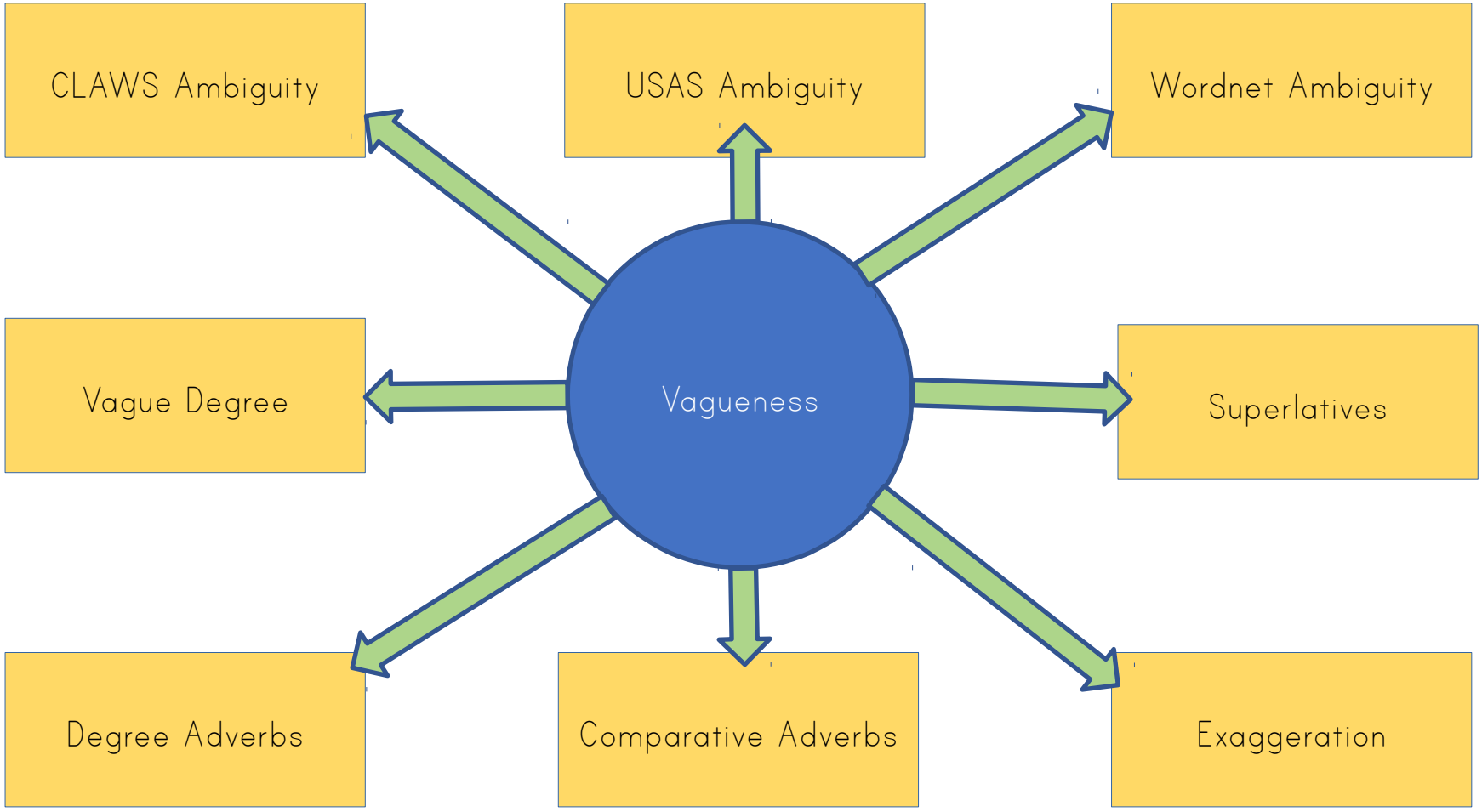
Fake News!

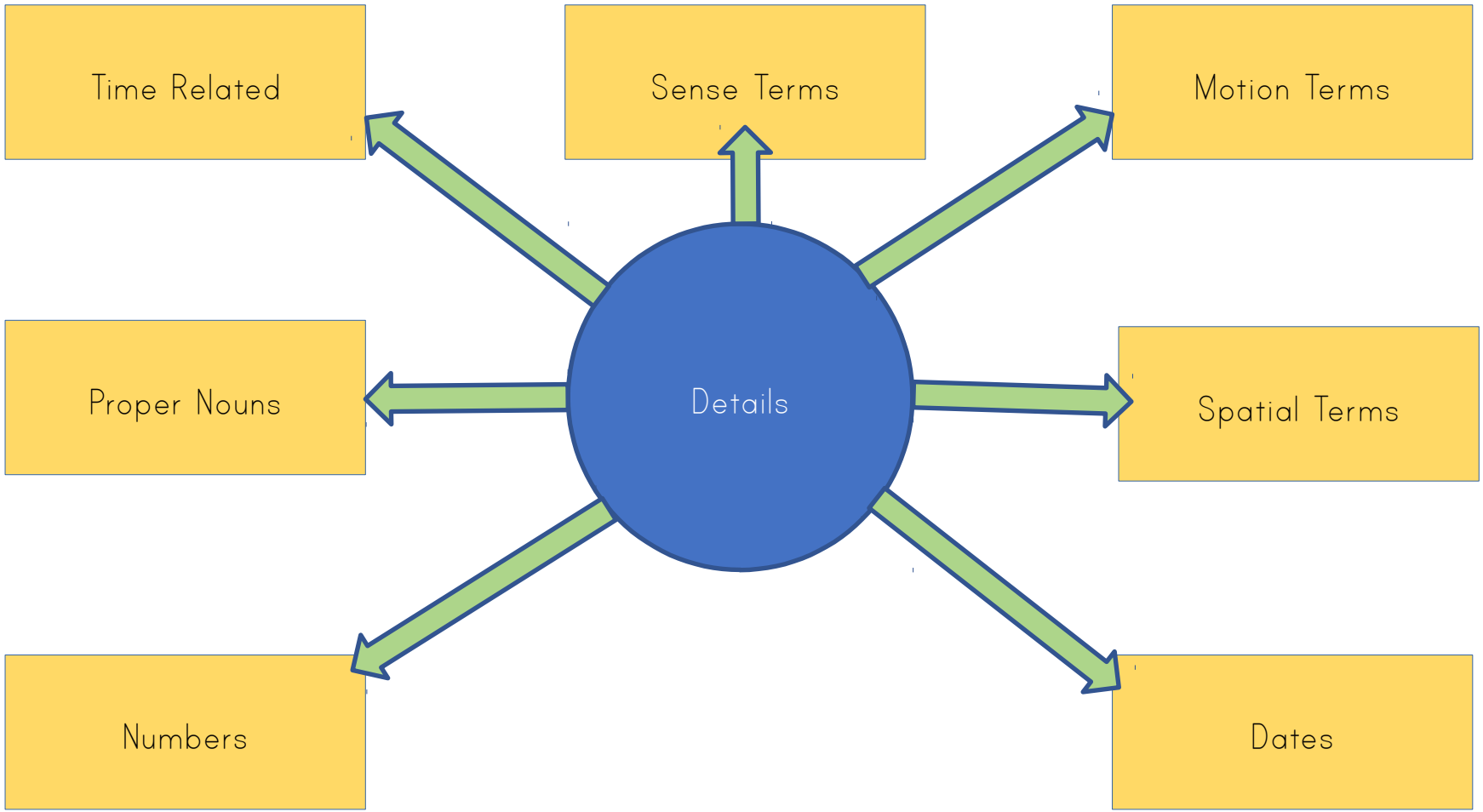
- Flagged as fake by Buzzfeed.
- 2016 Election.
- Horne and Adali, 2017.

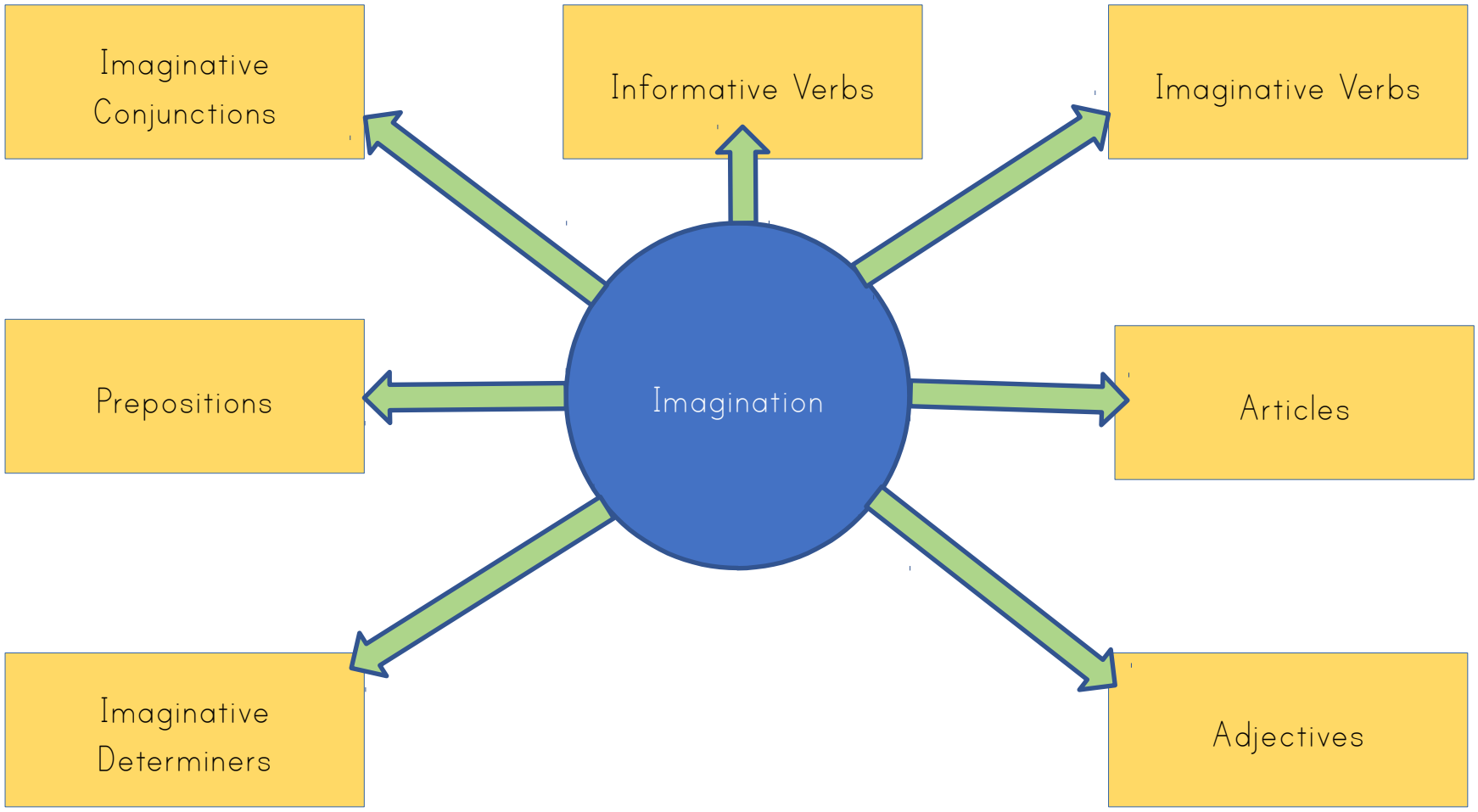
But what are you going to do with it?

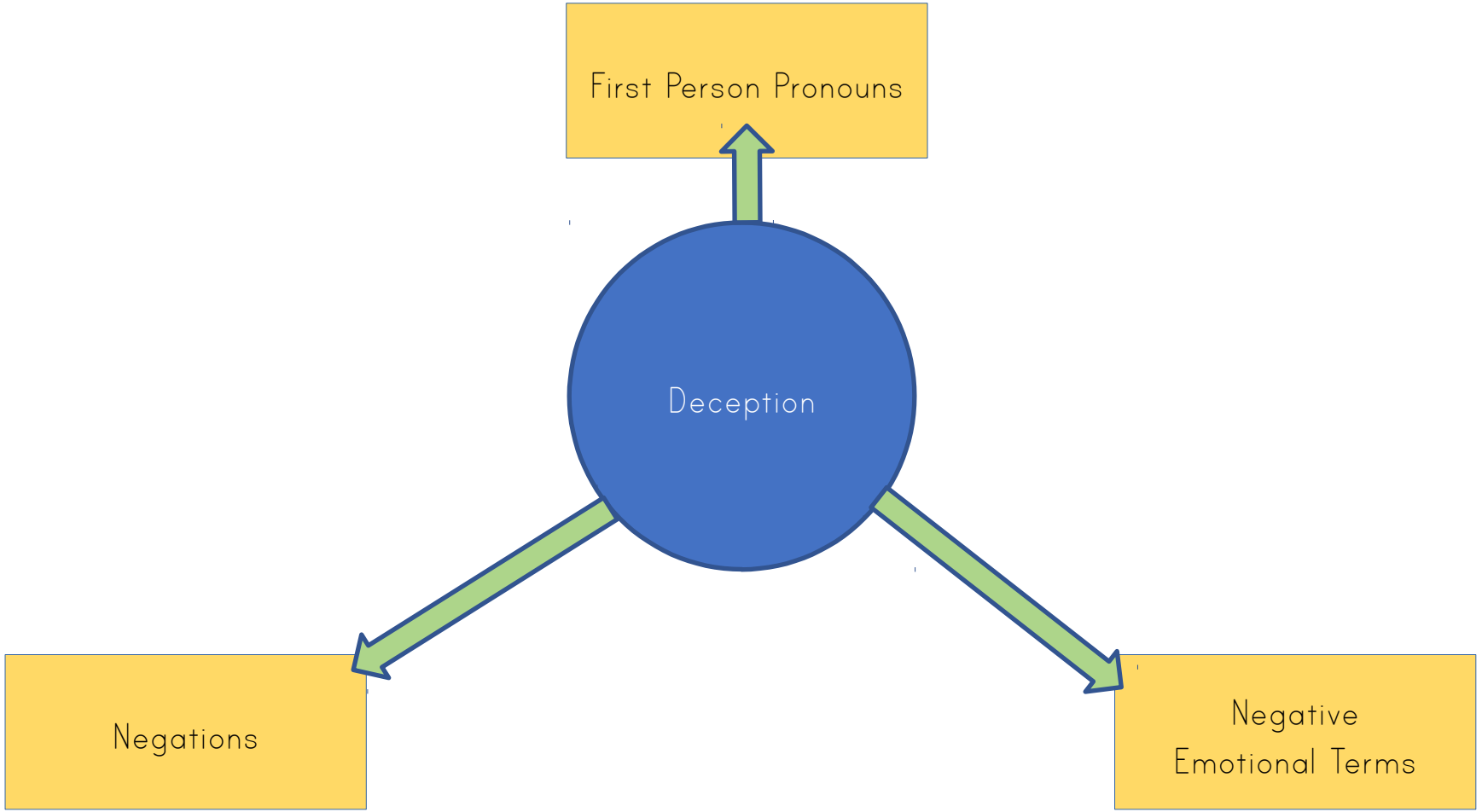


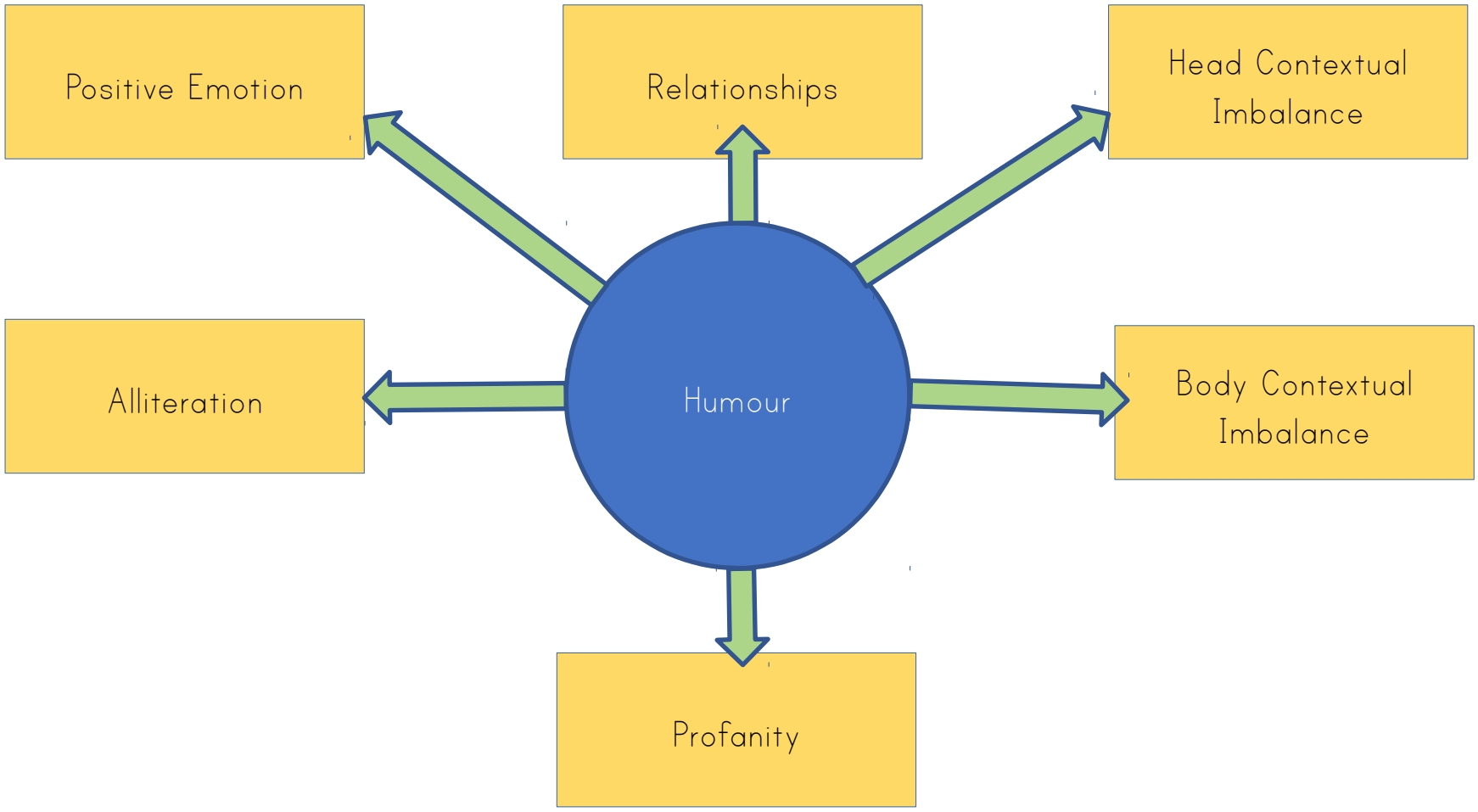
Building a feature set

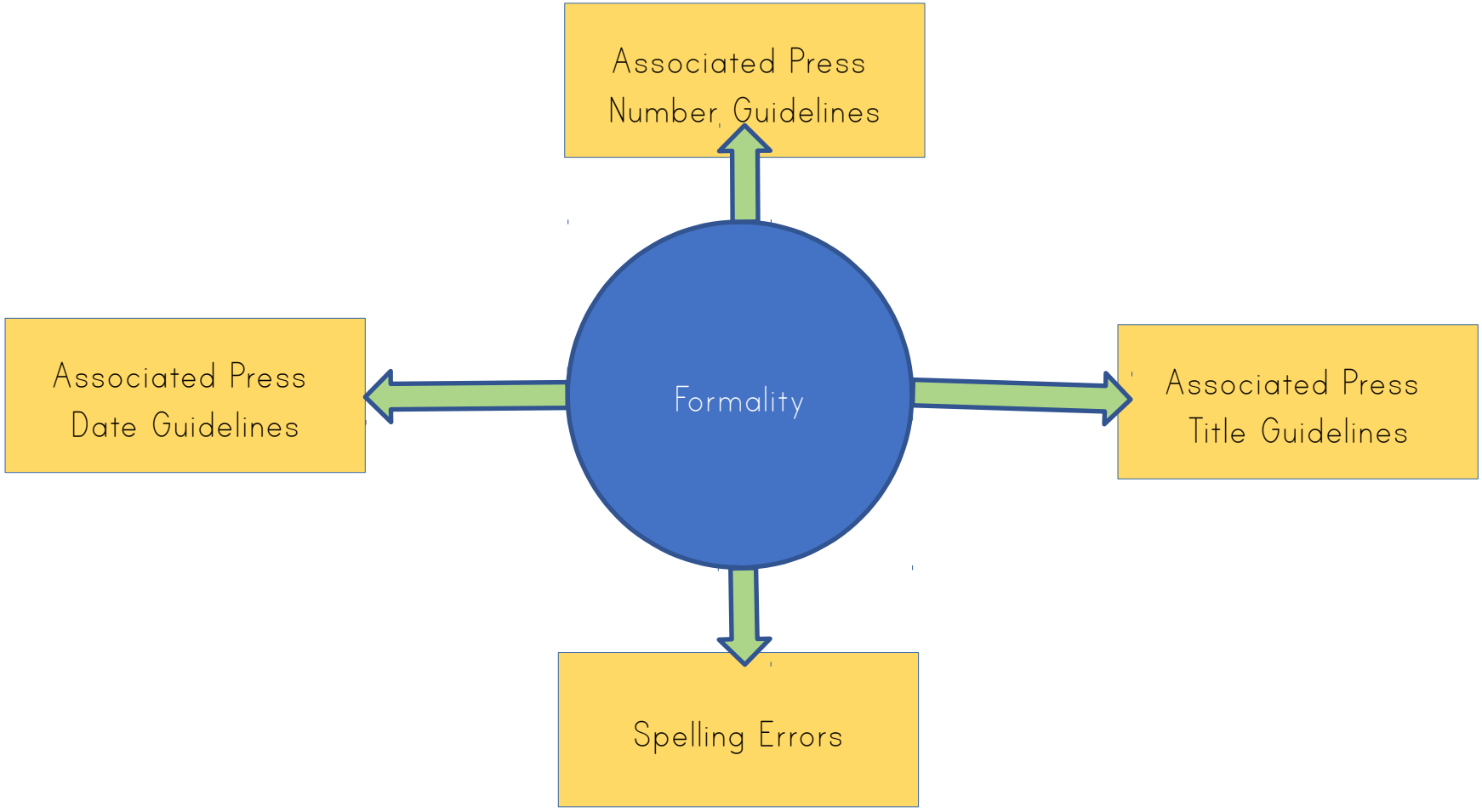

CLAWS Ambiguity
USAS Ambiguity
Wordnet Ambiguity
Vague Degree
Vagueness
Superlatives
Degree Adverbs
Comparative Adverbs
Exaggeration

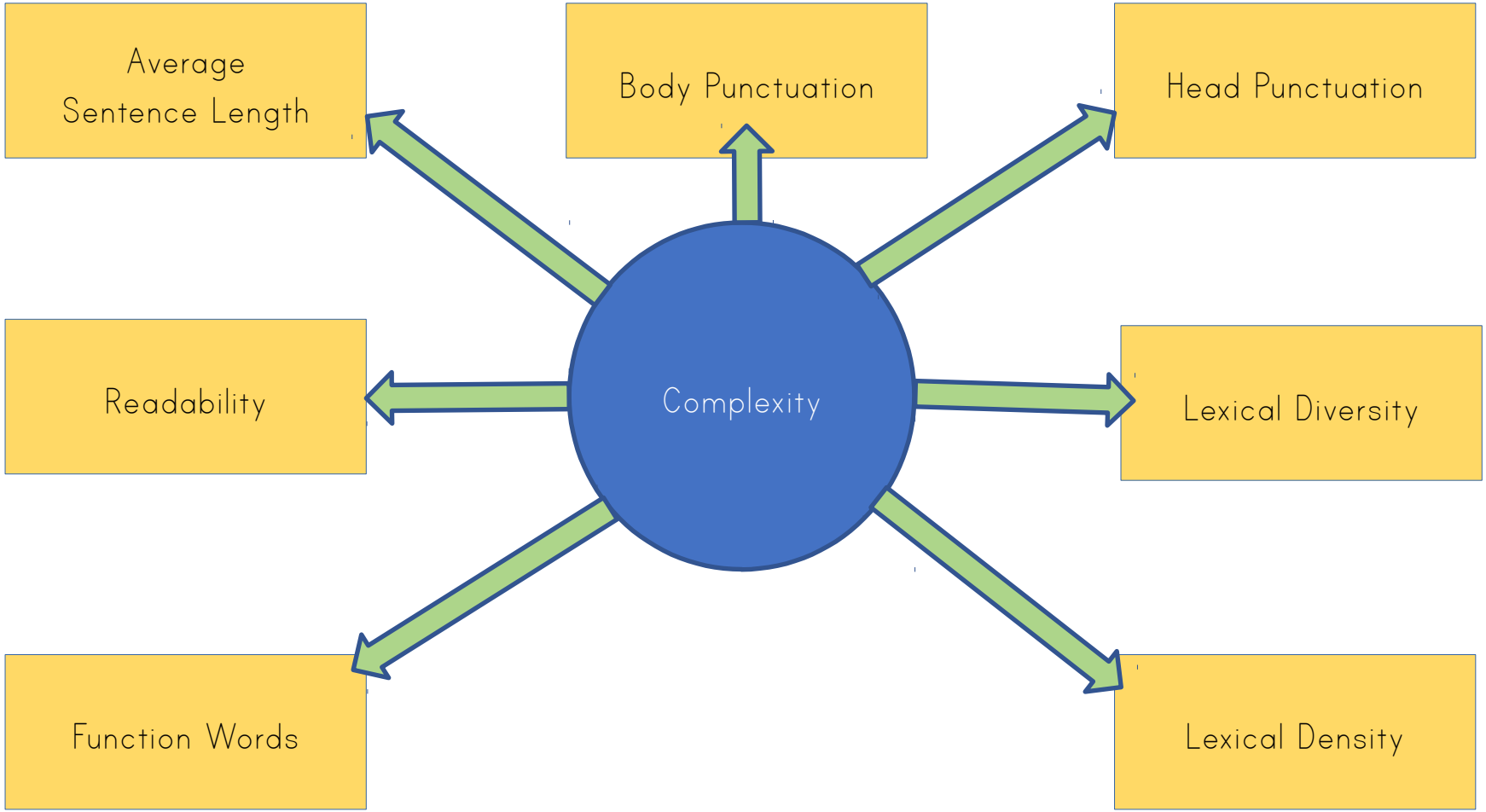

Time Related
Sense Terms
Motion Terms
Proper Nouns
Details
Spatial Terms
Numbers
Dates

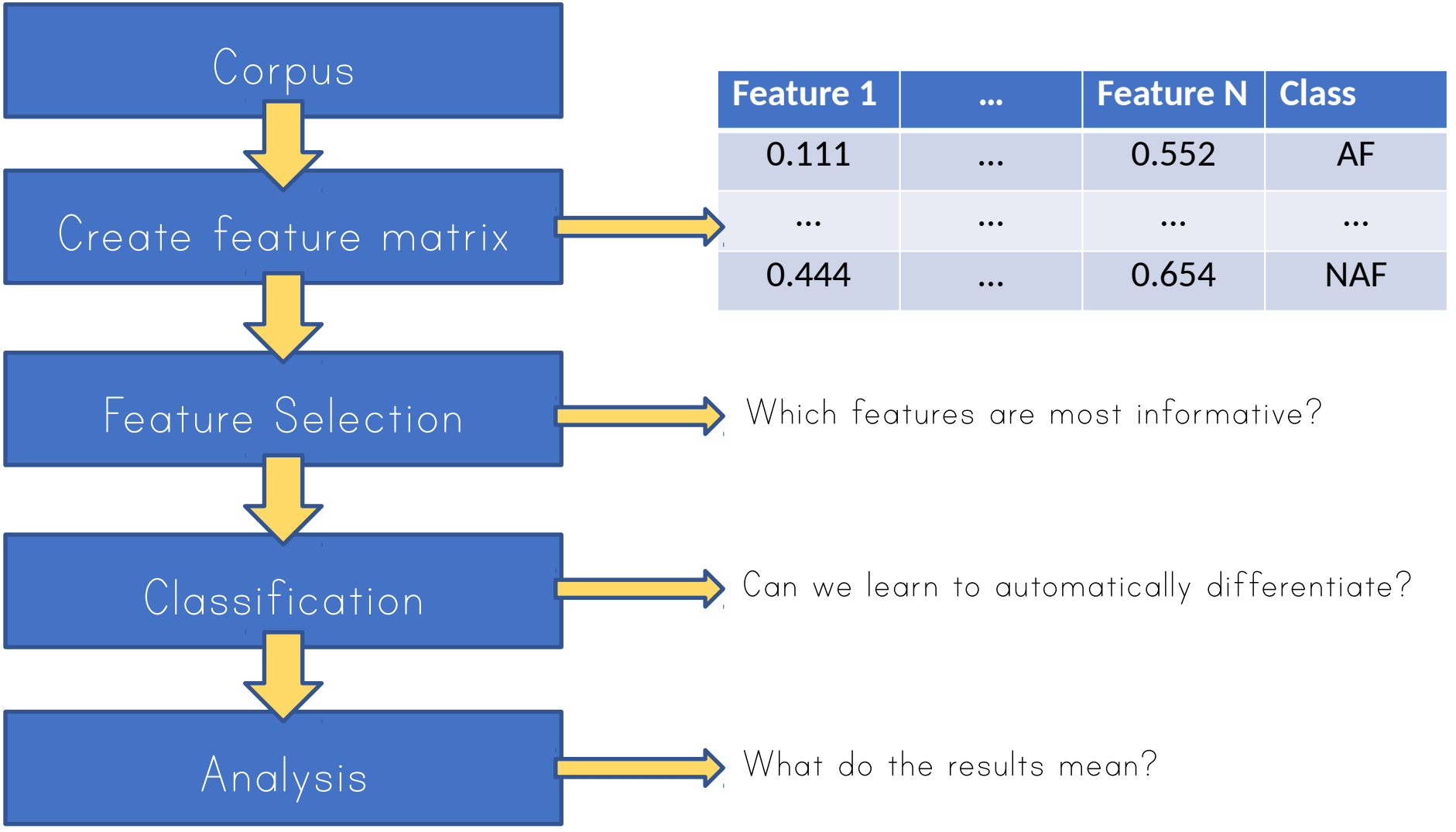

Imaginative Conjunctions
Informative Verbs
Imaginative Verbs
Prepositions
Imagination
Articles
Imaginative Determiners
Adjectives


First Person Pronouns
Deception
Negations
Negative
Emotional Terms


Positive Emotion
Relationships
Head Contextual Imbalance
Alliteration
Humour
Body Contextual Imbalance
Profanity


Associated Press Number Guidelines
Associated Press Date Guidelines
Formality
Associated Press Title Guidelines
Spelling Errors

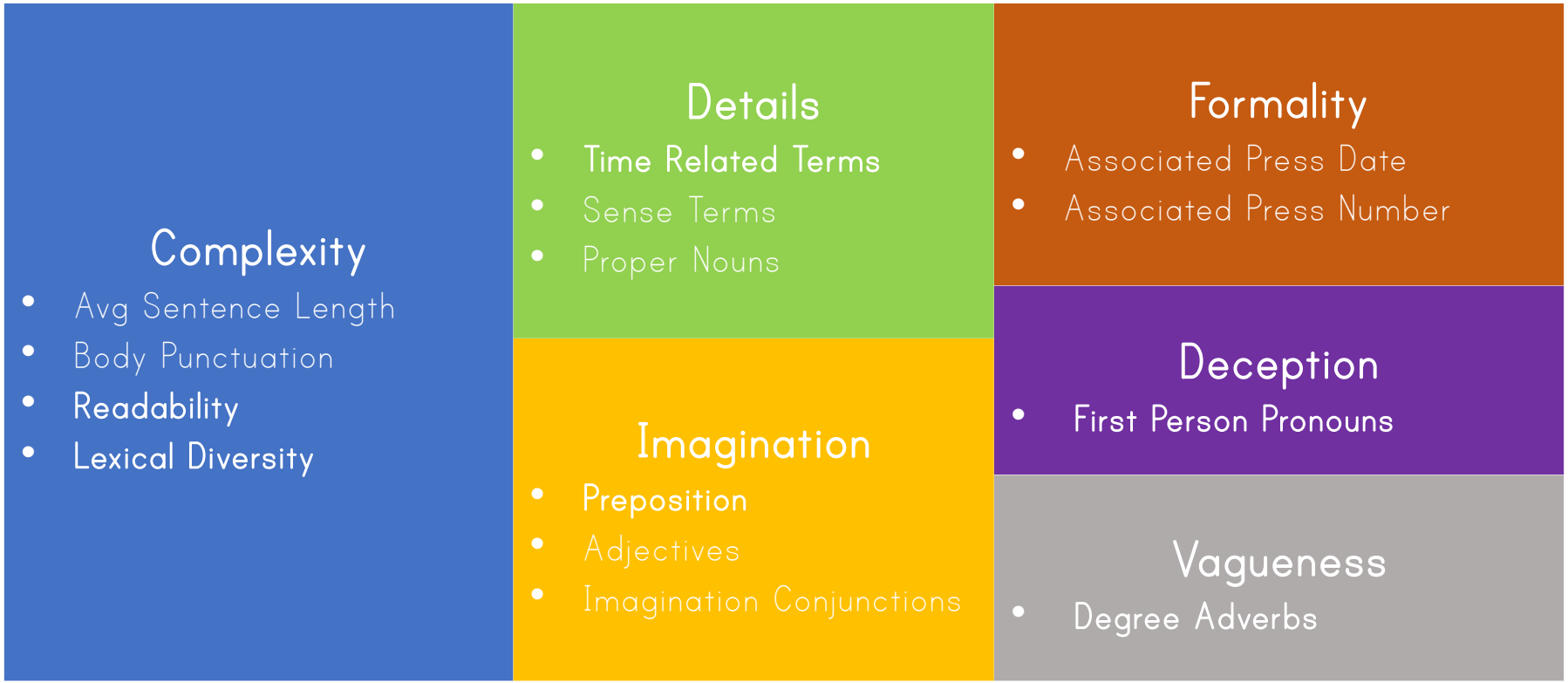

Average Sentence Length
Body Punctuation
Head Punctuation
Readability
Complexity
Lexical Diversity
Function Words
Lexical Density


Corpus
Create feature matrix
Feature Selection
Classification
Analysis
Feature 1
...
Feature N
Class
0.111
...
0.552
AF
...
...
...
...
0.444
...
0.654
NAF
Which features are most informative?
Can we learn to automatically differentiate?
What do the results mean?

Feature Selection

- Chi-squared test
- ANOVA
- Mutual Information
- Recursive Feature Elimination
- Logistic Regression Coefficients

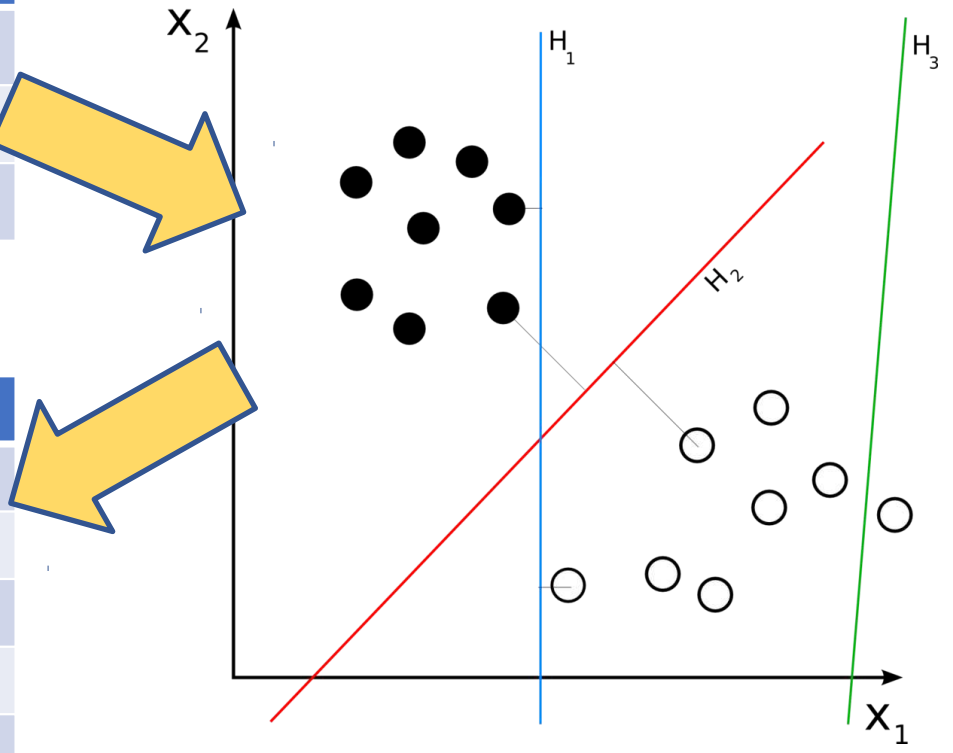
Feature Selection



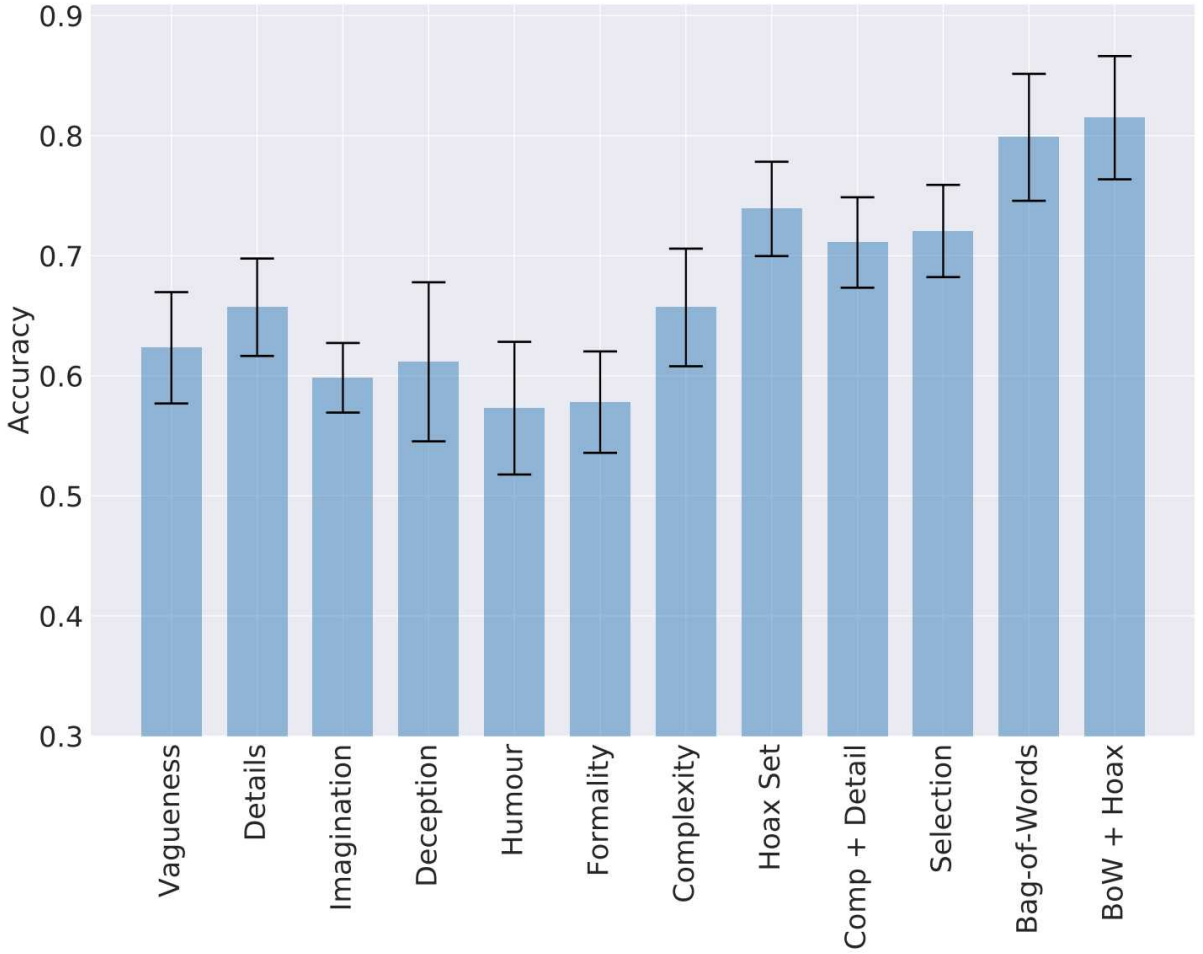
Classification

Feature 1	...	Feature N	Class
0.111	...	0.552	AF
...	...	...	...
0.444	...	0.654	NAF

Article	Prediction	Truth
1	AF	AF
2	NAF	AF
...	...	...
n-1	NAF	NAF
n	AF	NAF



Classification Accuracies for all Feature Sets



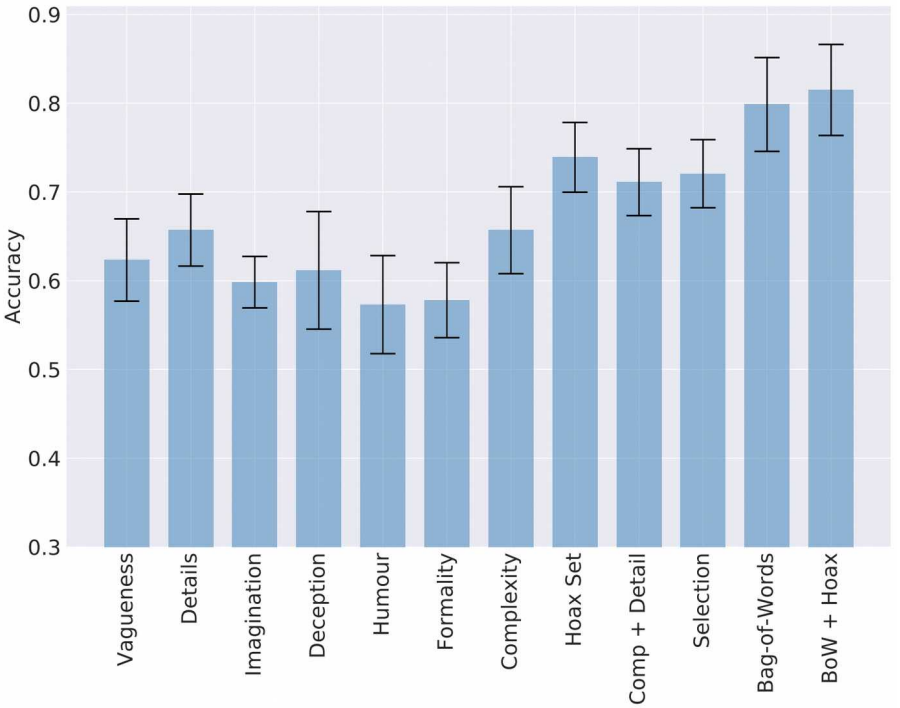
Hoax Set: 74%

Bag-of-Words: 80%

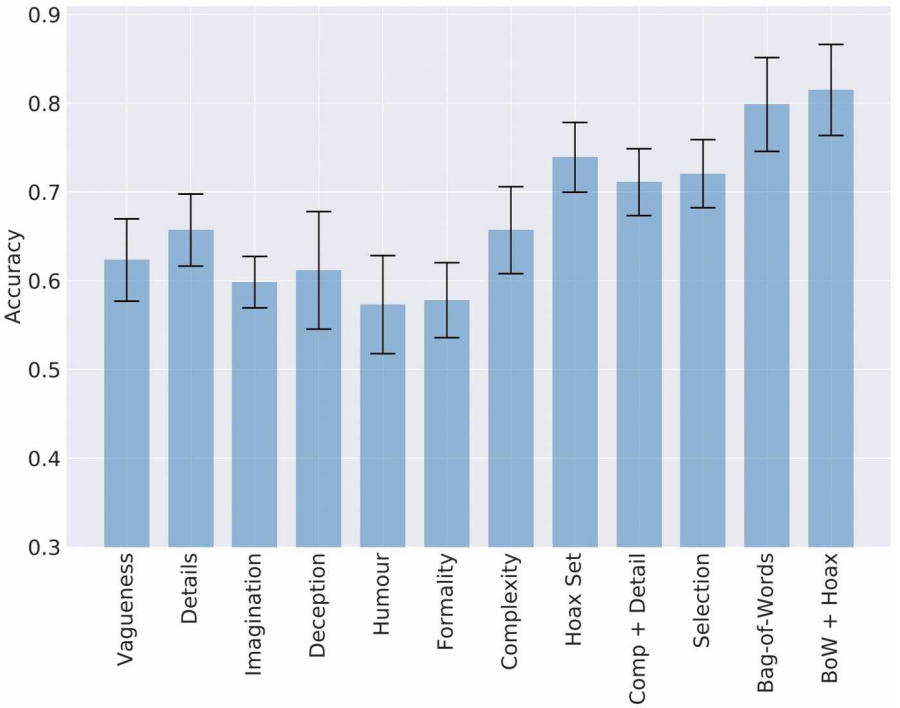
Complexity + Detail: 71%

What are we seeing so far?

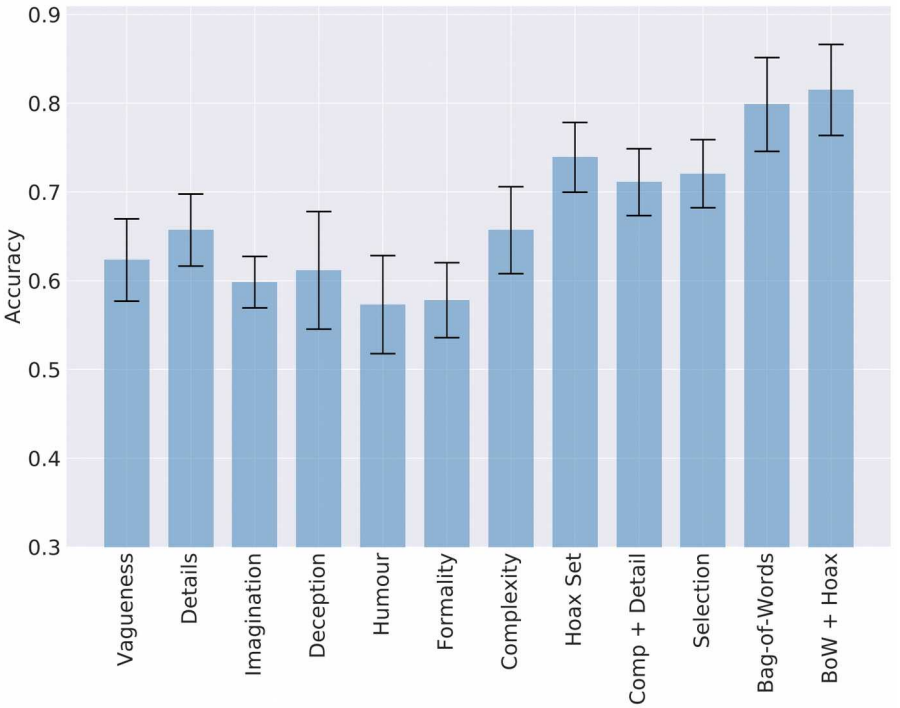
Our feature set can differentiate between hoax and genuine.



Most individual feature groups don't do so well.



Complexity and
Detail are Important.



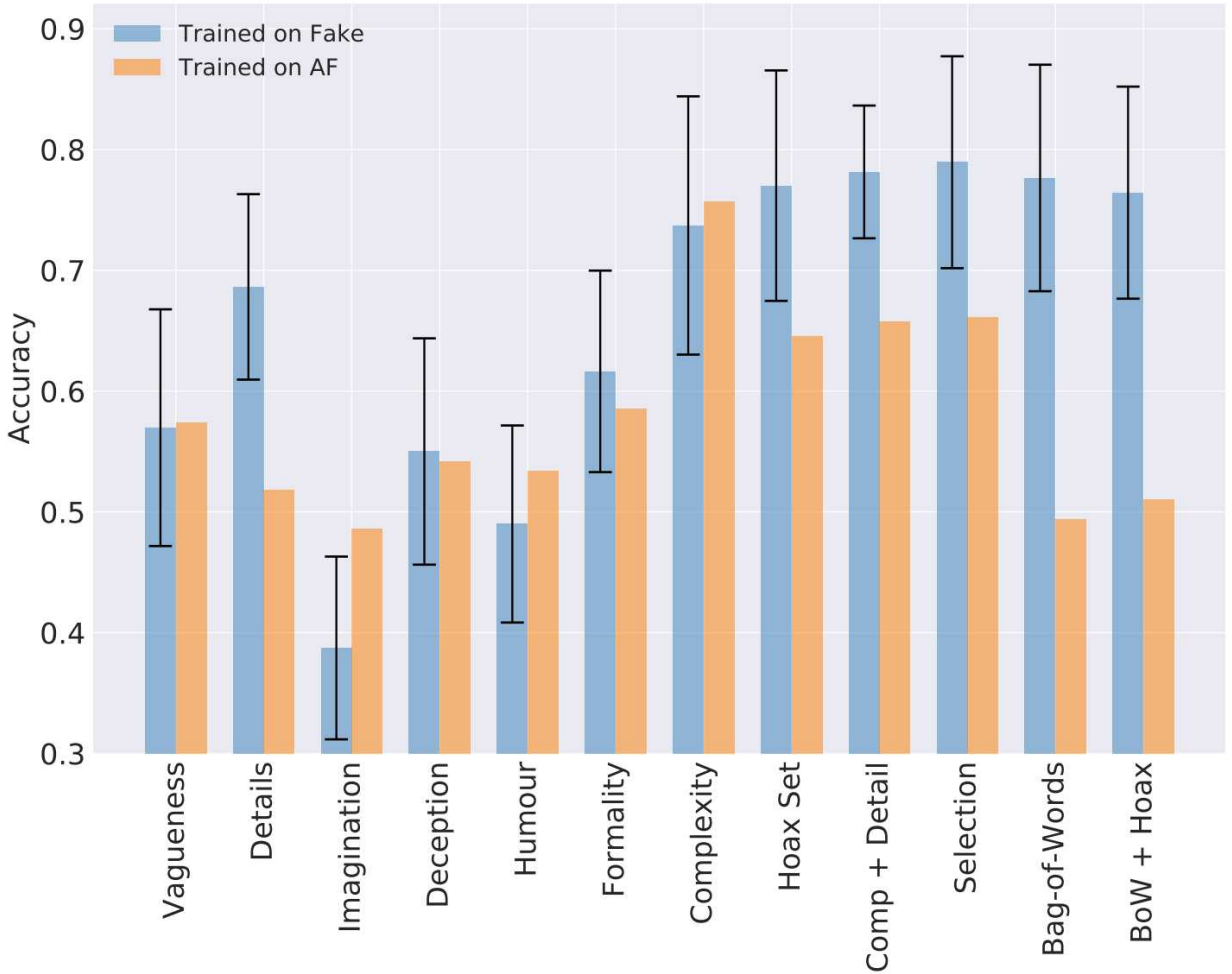
How does this compare to Fake News?

Classifying Fakes

1. One classifier trained on Fake News.

2. Second Classifier trained on April Fools' and tested on Fake News.

Classification Accuracies for Fake News

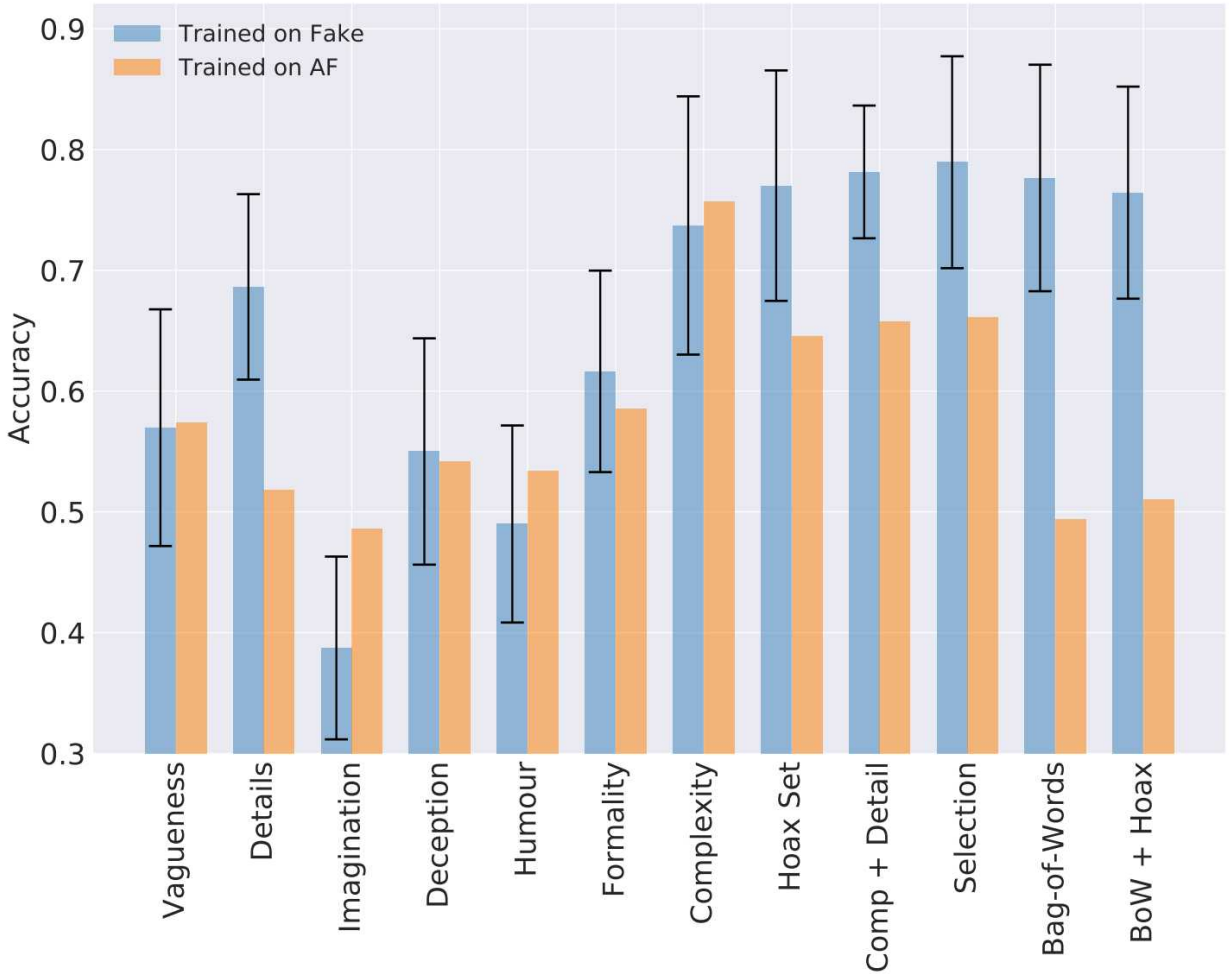


Hoax Set: 76.9%

Bag-of-Words: 77.7%

Complexity + Detail: 78.1%

Classification Accuracies for Fake News



Hoax Set: 64.5%

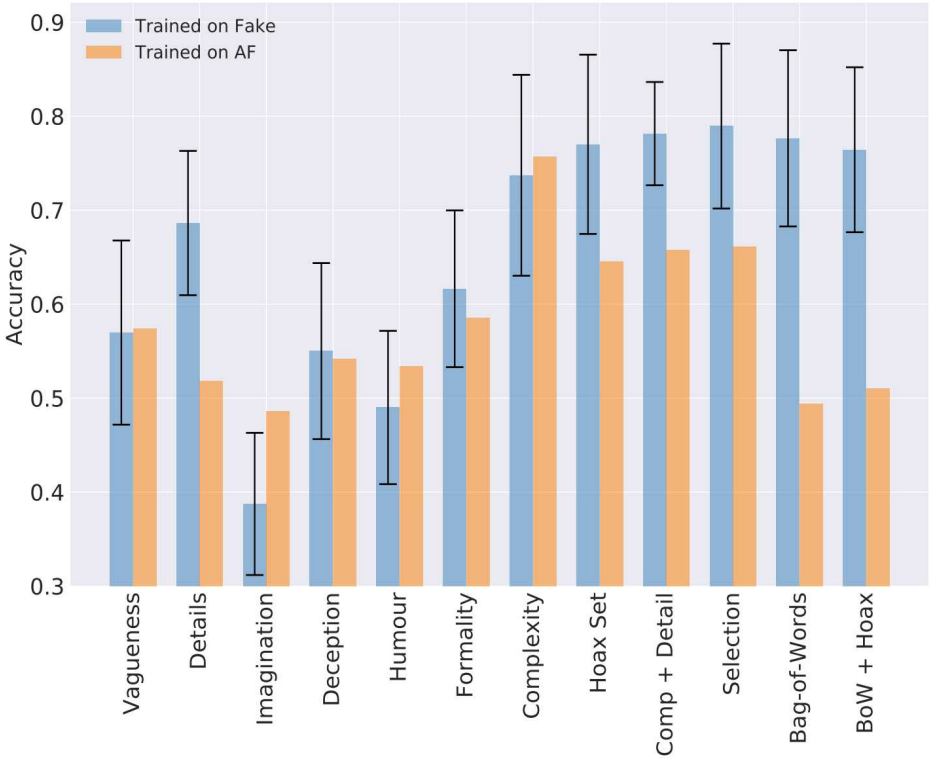
Bag-of-Words: 49.4%

Complexity + Detail: 65.7%

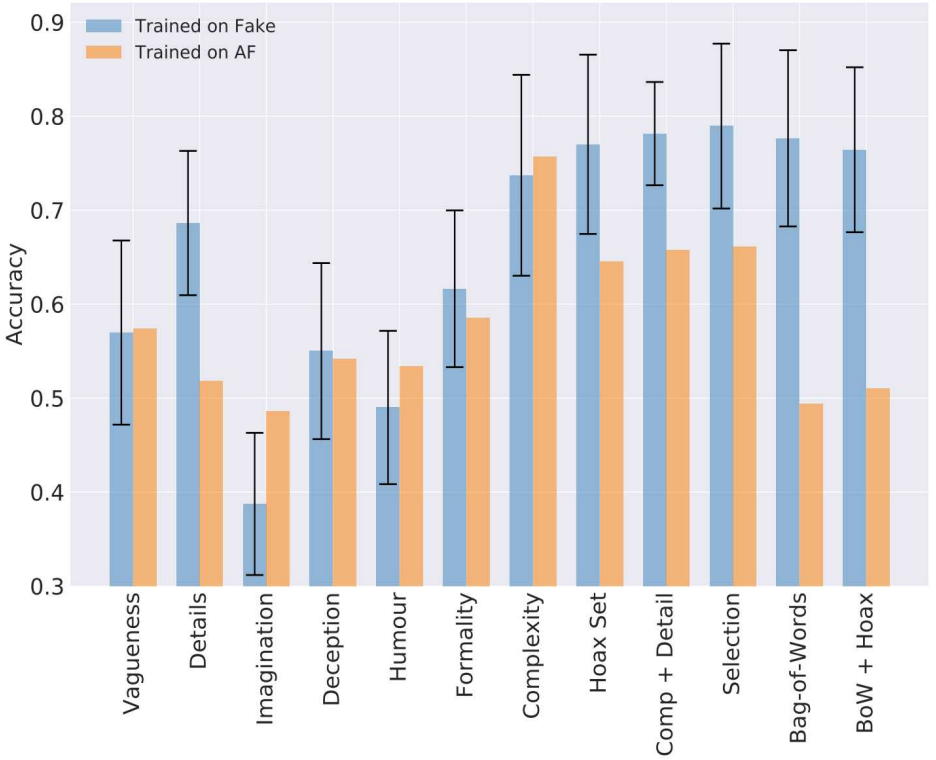
Complexity: 75.7%

What does this suggest?

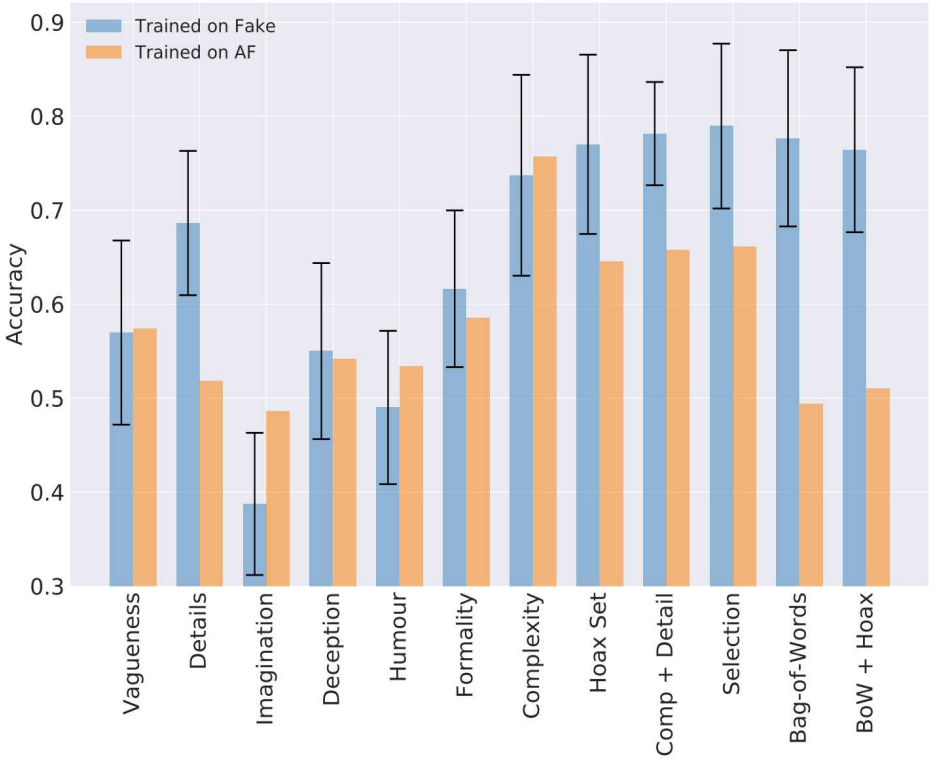
Our feature set differentiates fake news similarly well to April Fools'.



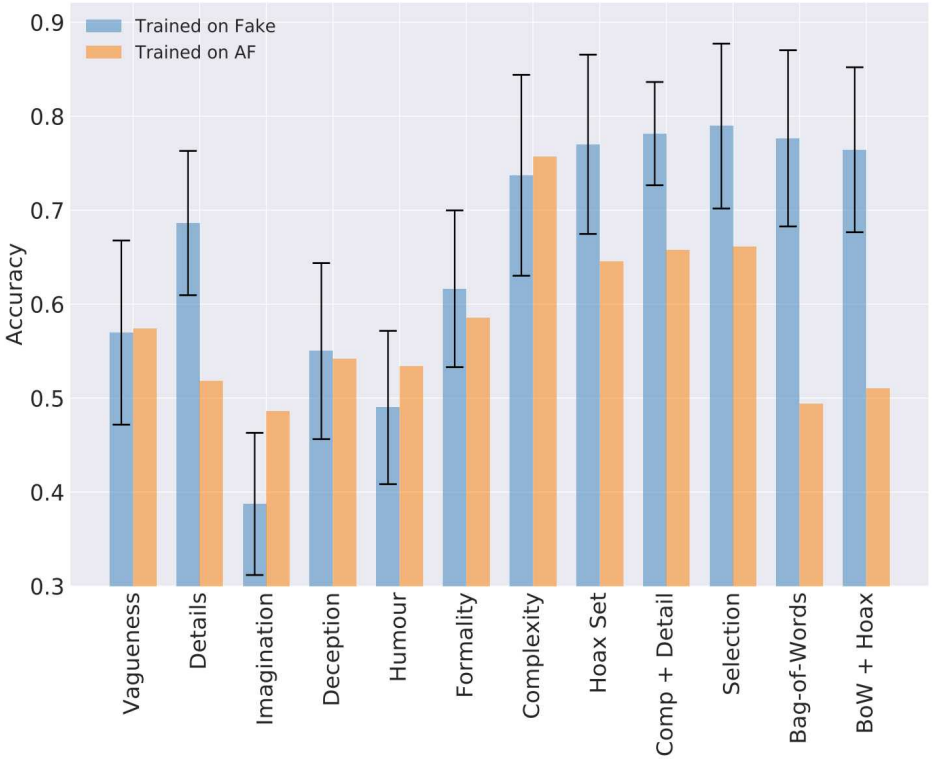
Some feature groups perform much worse.



Complexity and Detail remain the most important feature groups.

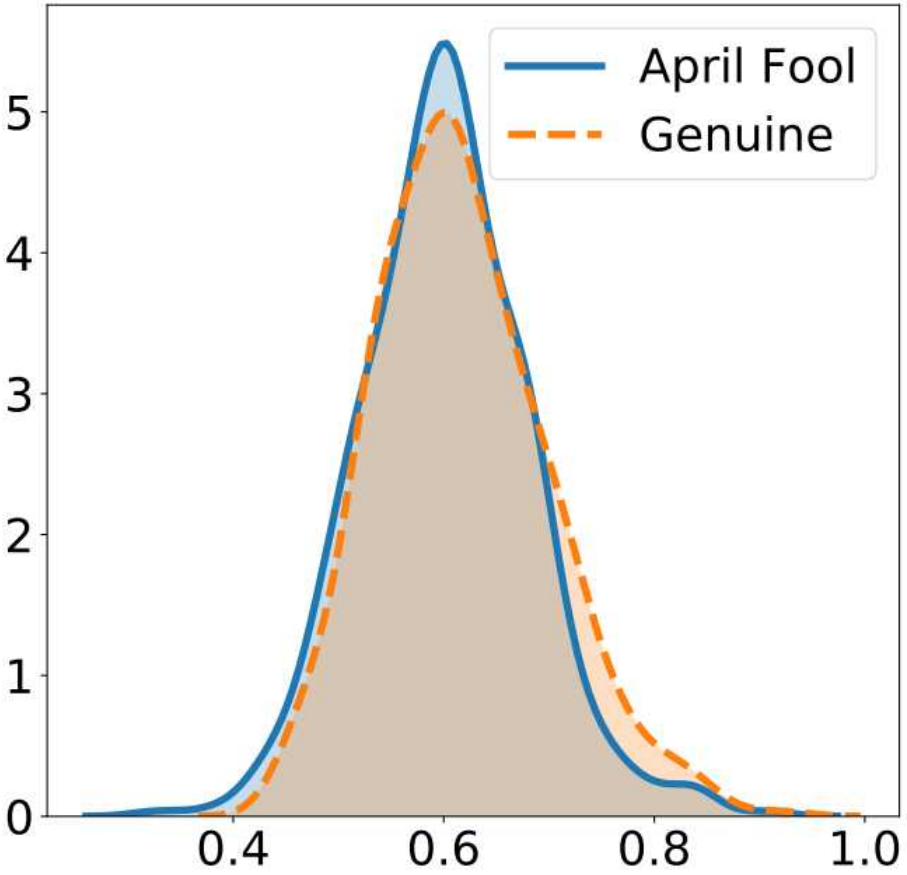


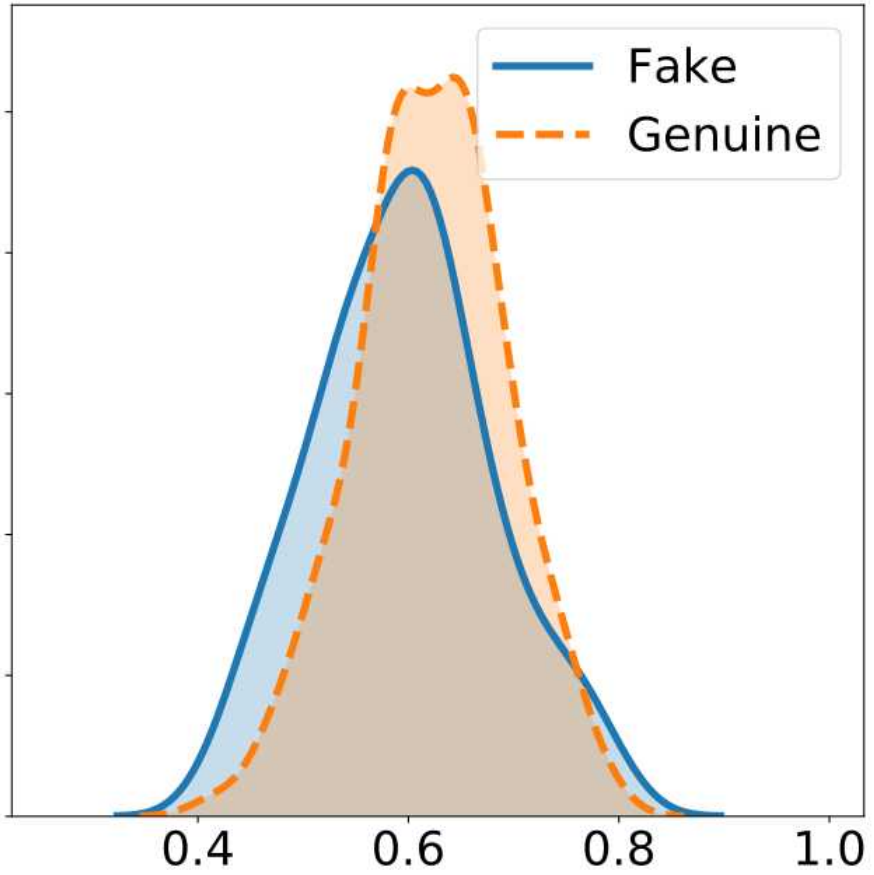
Our classifier trained on AF seems to work (to some extent) on Fake News.



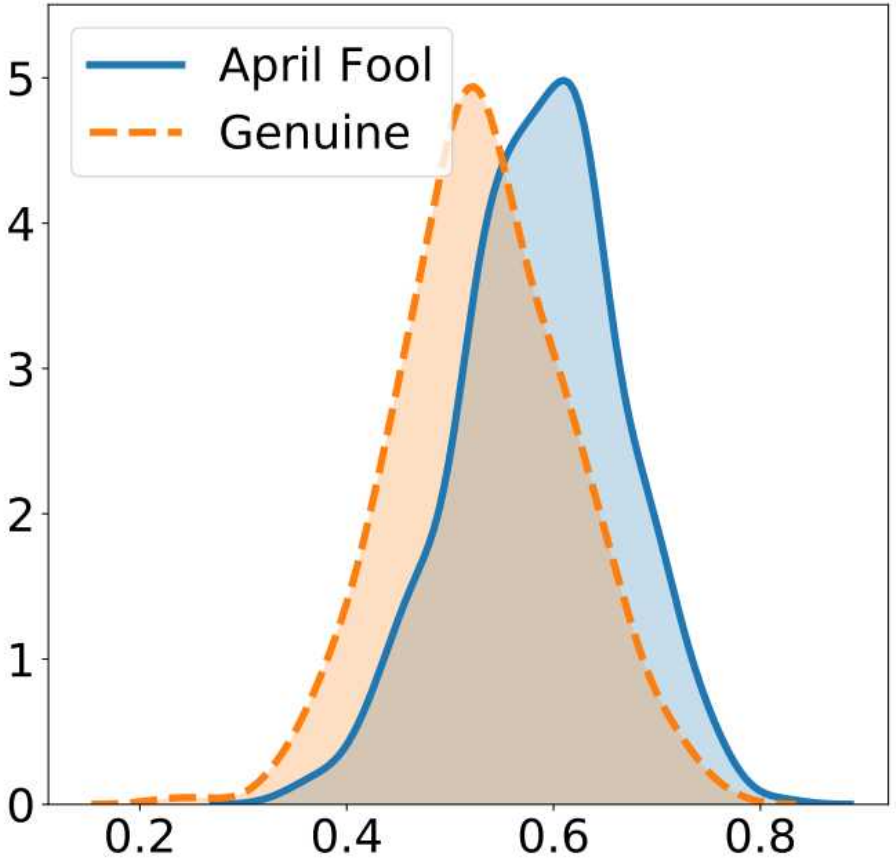
But what does the data say?

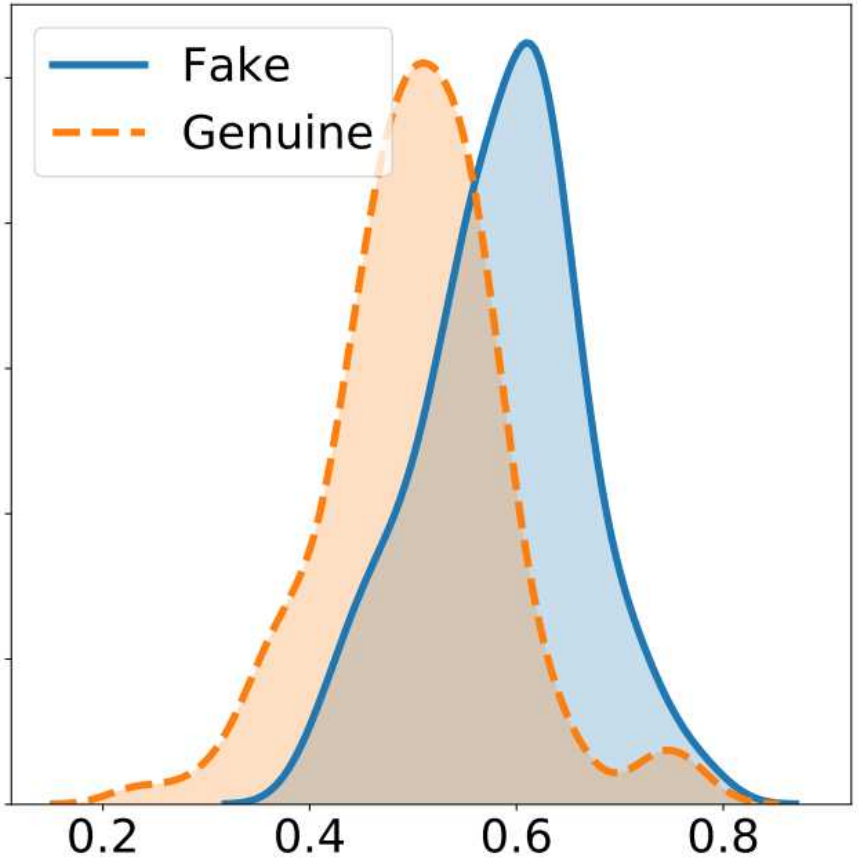
Readability *(Complexity)*



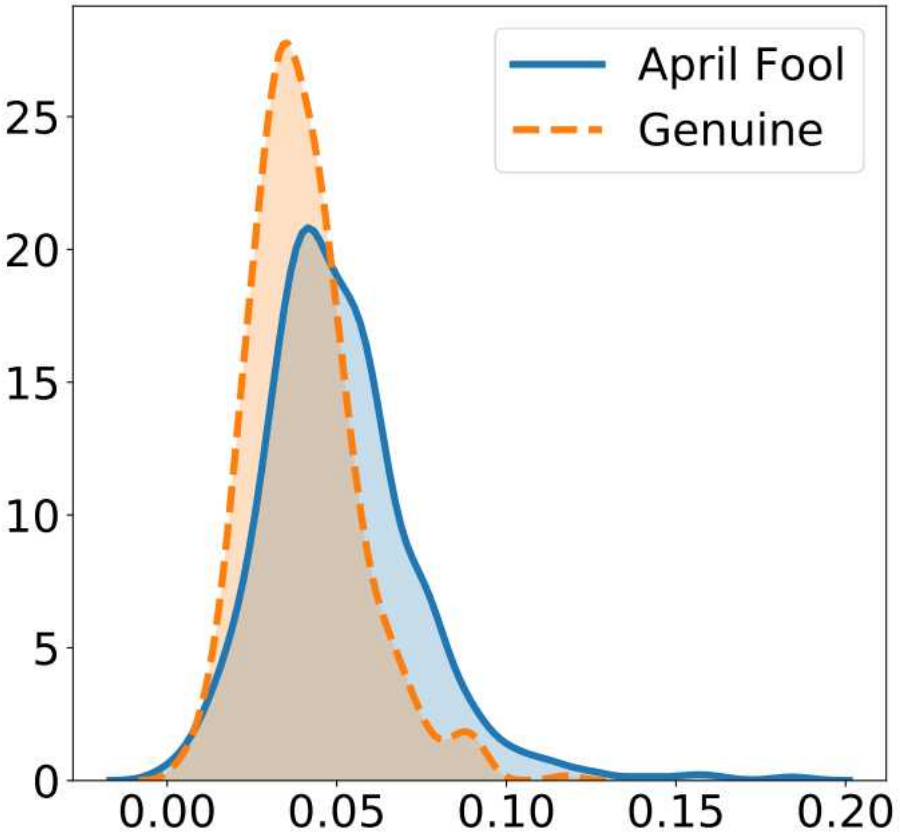


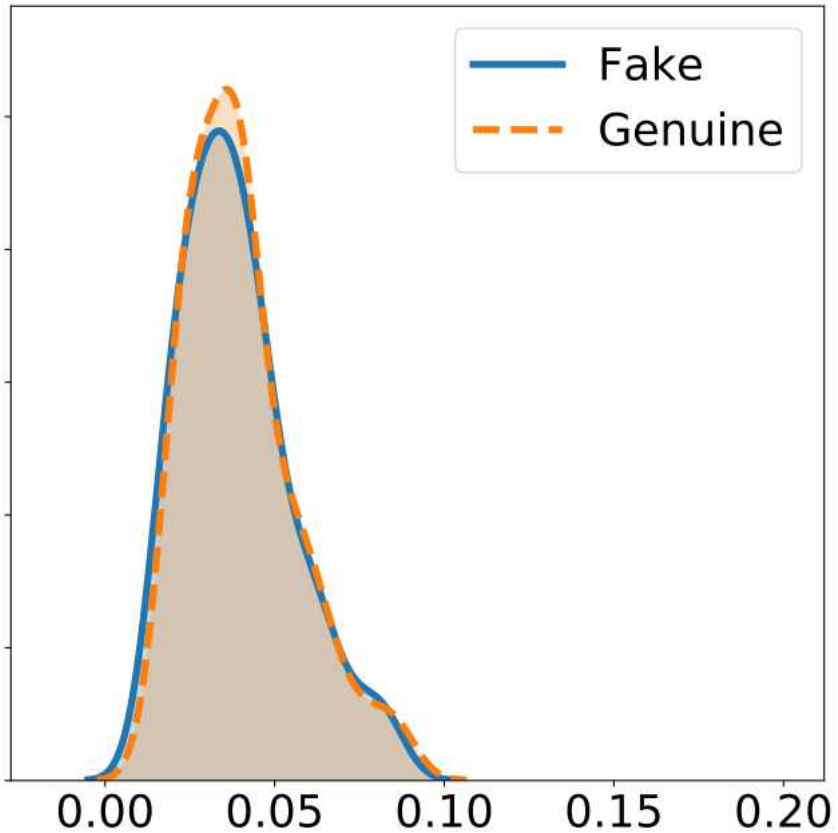
Lexical Diversity *(Complexity)*



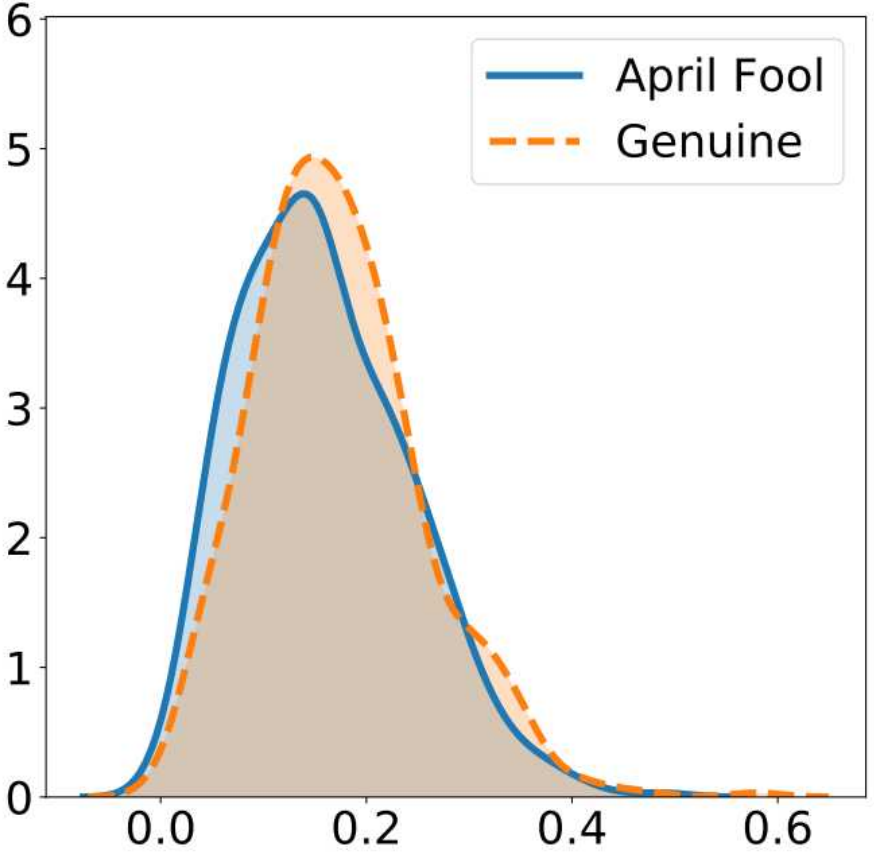


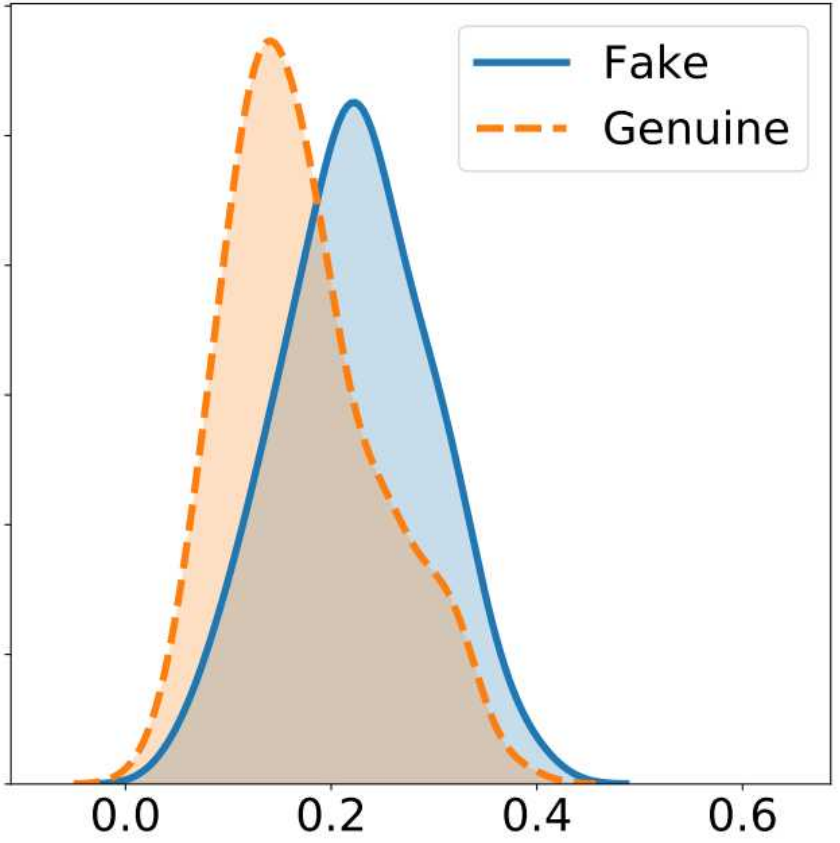
Time Related Vocabulary *(Detail)*



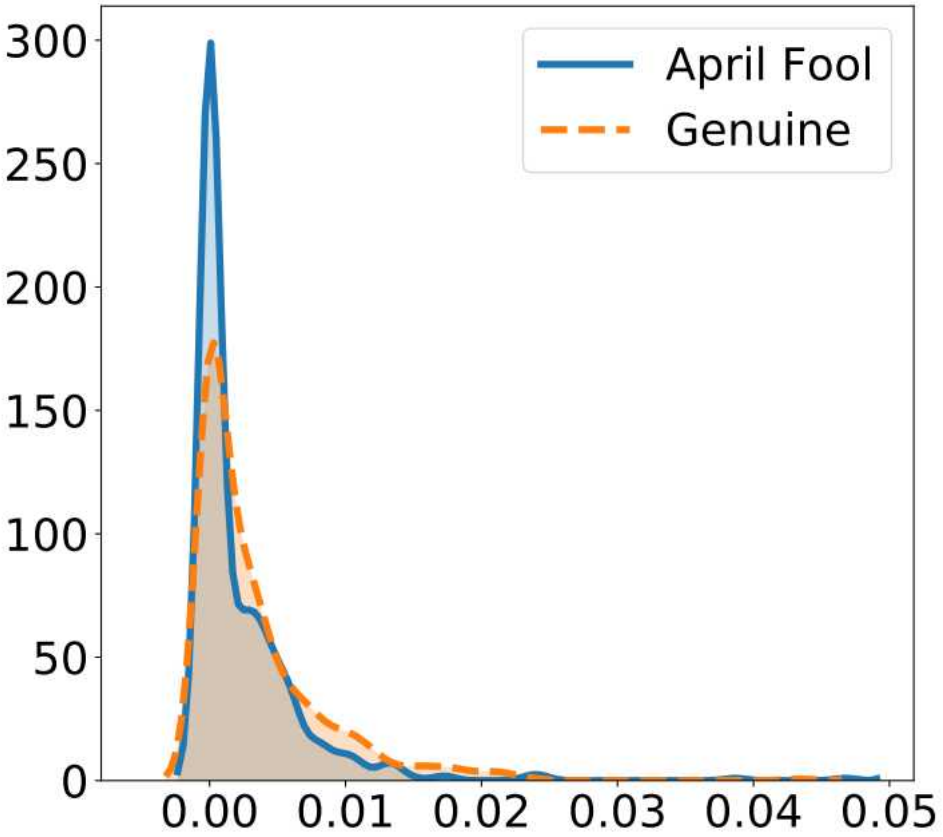


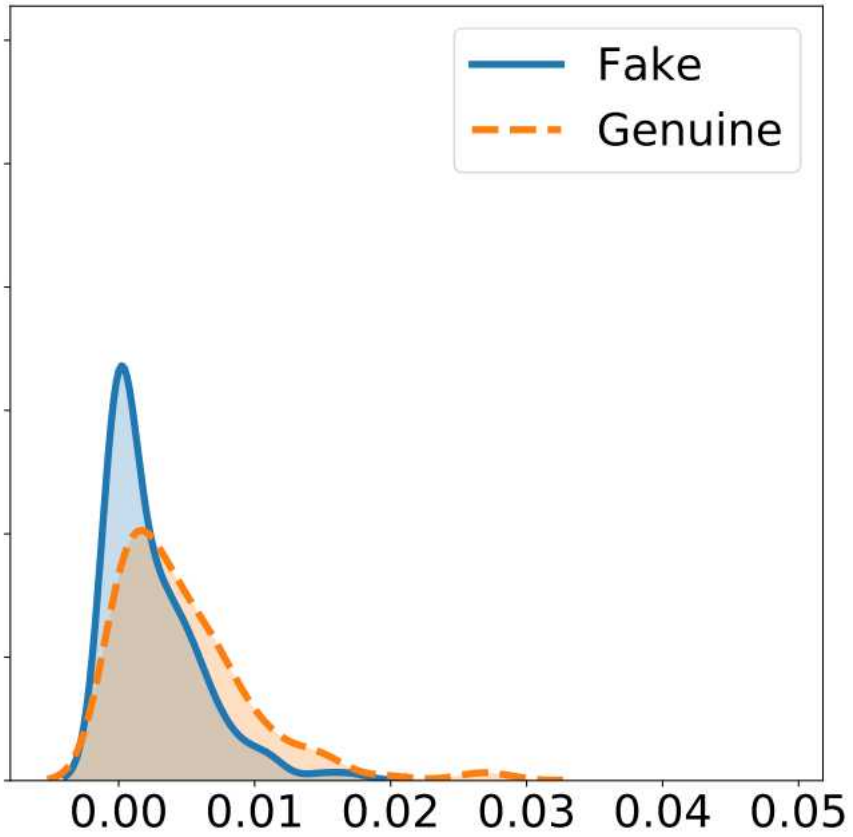
Proper Nouns *(Detail)*



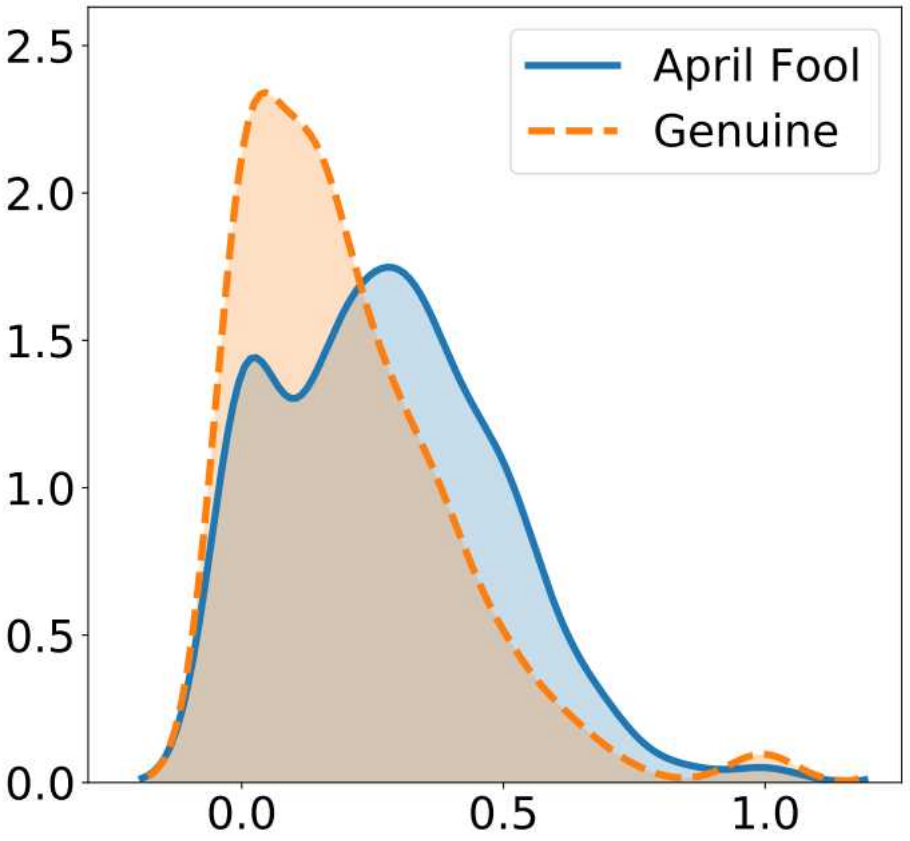


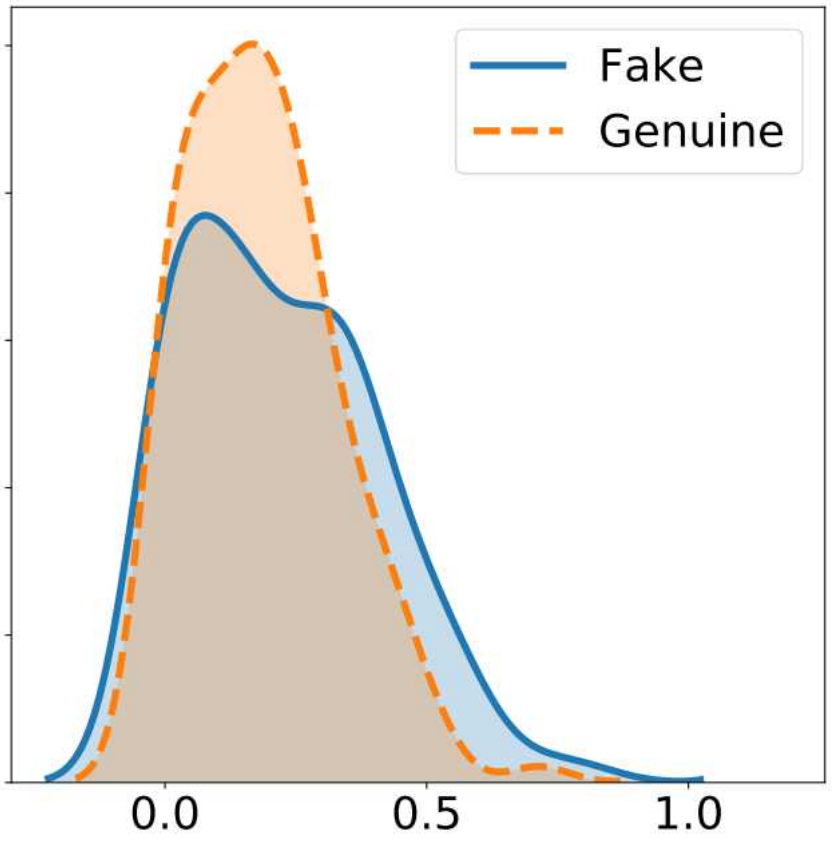
Dates *(Detail)*





First Person Pronouns *(Deception)*





Can you sum it all up?

Conclusions – Part 1

- **Created a new corpus of April Fools' hoaxes.**
- Used features from deception, humour, and irony detection to classify hoaxes with moderate success.
- Showed that features relating to complexity and detail seem to be the most important.

Conclusions – Part 1

- Created a new corpus of April Fools' hoaxes.
- **Used features from deception, humour, and irony detection to classify hoaxes with moderate success.**
- Showed that features relating to complexity and detail seem to be the most important.

Conclusions – Part 1

- Created a new corpus of April Fools' hoaxes.
- Used features from deception, humour, and irony detection to classify hoaxes with moderate success.
- **Showed that features relating to complexity and detail seem to be the most important.**

Conclusions – Part 2

- **Found that similar features are useful in identifying April Fools' and Fake News.**
- Some of these features manifest themselves similarly for both AF Hoaxes and Fake News.

Conclusions – Part 2

- Found that similar features are useful in identifying April Fools' and Fake News.
- **Some of these features manifest themselves similarly for both AF Hoaxes and Fake News.**

Future Work


Sus Bonos de la Libertad ayudarán á dar fin con esto
Your
LIBERTY BOND
Will Help Stop This


He's one of the busiest men in town. While his door may say Office Hours 2 to 4, he's actually on call 24 hours a day.
The doctor is a scientist, a diplomat, and a friendly sympathetic human being all in one, no matter how long and hard his schedule.
According to a recent Nationwide survey:
MORE DOCTORS SMOKE CAMELS
THAN ANY OTHER CIGARETTE
DOCTORS in every branch of medicine—113,597 in all—were queried in this nationwide study of cigarette preference. Three leading research organizations made the survey. The gist of the query was—What cigarette do you smoke, Doctor?
The brand named most was Camel!
The rich, full flavor and cool mildness of Camel's superb blend of costlier tobaccos seem to have the same appeal to the smoking tastes of doctors as to millions of other smokers. If you are a Camel smoker, this preference among doctors will hardly surprise you. If you're not—well, try Camels now.
CAMEL
Your "T-Zone" Will Tell You...
T for Taste...
T for Throat...
that's your proving ground for any cigarette. See if Camels don't suit your "T-Zone" to a "T."
CAMELS Costlier Tobaccos


If the sun is 93 million miles away
Nick Havok
How are the clouds behind it?

e.dearden@lancaster.ac.uk