Fools’ Gold: Understanding the Linguistic Features of Deception and Humour Through April Fools’ Hoaxes

Ed Dearden
e.dearden@lancaster.ac.uk
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?”
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.
Elon Musk’s own Ship Ejects him Into Space for "the greater good"

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
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.
Hell Planet
Why do we care about April Fools’?
False Information

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

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

BBC Admits MI5 Dictates Who They Employ
But where does April Fools’ day fit into this?
Instawindow: transform your view with a permanent Instagram filter

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

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 older council blocks retrofitted with colourful cladding panels and

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

Polar bears spotted in Scotland

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’

BBC Admits MI5 Dictates Who They Employ

April 24, 2018
Seán Adl-Tantawi
News, UK
0
Deceptive Intent:
Is the author trying to deceive me?

Not Deceive? 

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.

Making my own costume this year for company annual dinner. Watch this space for sewing tips and creative ideas 😂😅 #not #prayforsofea
How about the data?
Catching Fools’!

• 519 April Fools’ articles.
• 371 websites.
• 213776 words.
• 2004-2018
Matching Fools’!

- 519 regular news articles.
- 240 Websites.
- 344,927 Words.
- 2004-2018
Fake News!

- Flagged as fake by Buzzfeed.
- 2016 Election.
But what are you going to do with it?
Building a feature set

Vagueness
Details
Imagination
Deception
Humour
Formality
Complexity
Deception

First Person Pronouns

Negations

Negative Emotional Terms
Formality

Associated Press
Number Guidelines

Associated Press
Date Guidelines

Associated Press
Title Guidelines

Spelling Errors
Which features are most informative?

Can we learn to automatically differentiate?

What do the results mean?

<table>
<thead>
<tr>
<th>Feature 1</th>
<th>...</th>
<th>Feature N</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.111</td>
<td>...</td>
<td>0.552</td>
<td>AF</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>0.444</td>
<td>...</td>
<td>0.654</td>
<td>NAF</td>
</tr>
</tbody>
</table>
Feature Selection

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

Complexity
- Avg Sentence Length
- Body Punctuation
- Readability
- Lexical Diversity

Details
- Time Related Terms
- Sense Terms
- Proper Nouns

Imagination
- Preposition
- Adjectives
- Imagination Conjunctions

Formality
- Associated Press Date
- Associated Press Number

Deception
- First Person Pronouns

Vagueness
- Degree Adverbs
Classification

<table>
<thead>
<tr>
<th>Feature 1</th>
<th>...</th>
<th>Feature N</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.111</td>
<td>...</td>
<td>0.552</td>
<td>AF</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>0.444</td>
<td>...</td>
<td>0.654</td>
<td>NAF</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Article</th>
<th>Prediction</th>
<th>Truth</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>AF</td>
<td>AF</td>
</tr>
<tr>
<td>2</td>
<td>NAF</td>
<td>AF</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>n-1</td>
<td>NAF</td>
<td>NAF</td>
</tr>
<tr>
<td>n</td>
<td>AF</td>
<td>NAF</td>
</tr>
</tbody>
</table>
Classification Accuracies for all Feature Sets

- Hoax Set: 74%
- Bag-of-Words: 80%
- Complexity: 71%
  + Detail
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: 78.1% + Detail
Classification Accuracies for Fake News

Hoax Set: 64.5%
Bag-of-Words: 49.4%
Complexity: 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)

![Graph showing Readability (Complexity) for April Fool and Genuine, and Fake and Genuine.]
Lexical Diversity (Complexity)
Time Related Vocabulary (Detail)
Proper Nouns (Detail)

![Graph 1: April Fool vs Genuine](image1)

![Graph 2: Fake vs Genuine](image2)
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

If the sun is 93 million miles away

How are the clouds behind it?
Questions?

Thanks for listening!

e.dearden@lancaster.ac.uk