The Semantics of Fake News

How NLP and AI may help us identify misinformation

Charles Lam
Department of English
The Hang Seng University of Hong Kong
charleslam@hsu.edu.hk

International Conference
on Misinformation
and Fake News

香港恒生大學
THE HANG SENG UNIVERSITY
OF HONG KONG

February 6, 2021

Overview

1. What are privative adjectives?
2. What makes *fake news* fake?
3. Can we detect misinformation with NLP?
4. Challenges to automatic detection of misinformation

Background

- Misinformation is a global threat!
- Human fact checking is great, but content-based verification is costly
- Identifying pattern / meaning in text?
- Crosslinguistic, cross-cultural challenges

What makes *fake news* fake?
How can we tell fake news by the text?

Semantics of Adjectives

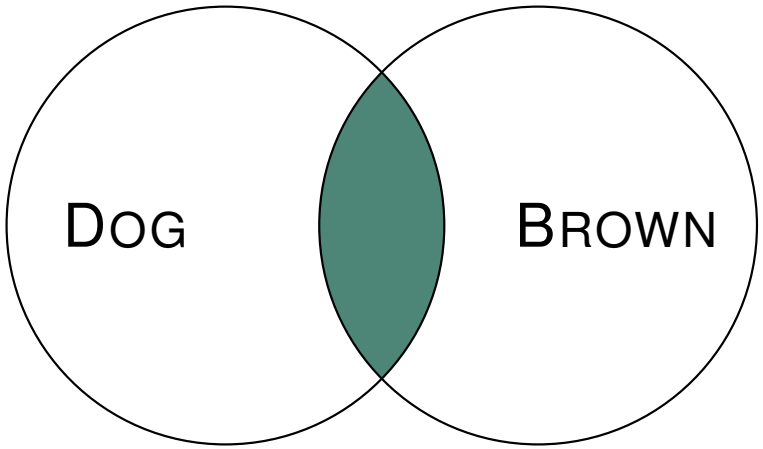
What do we mean by *fake*?

- Semantics = the study of linguistic meaning
- *fake* belongs to a class of adjectives called “privative adjectives”
- Cf. *false*, *pseudo-*, *former*, *counterfeit*

Types of Adjectives

1. Intersective
   - *brown dog*
   - *rectangular cake*
2. Subsective
   - *small elephant*
   - *large dog*
3. Privative
   - *fake gun*
   - *former chairman*
   - *forged document*
   - *fictitious claim*; *fake news*

Intersective Adjectives

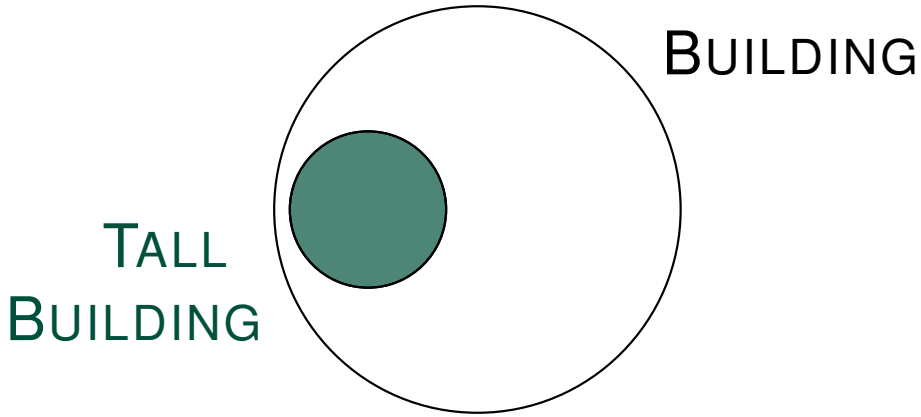


Set-theoretic interpretation of adjective modification:

- A *brown dog* = an object that is a dog and is brown.
- A *rectangular cake* = an object that is a cake and is rectangular

Subsective Adjectives

- For adjectives like *tall*, *large*, the interpretation of the Adj-N combination (e.g. *tall guy*, *tall building*)
- *tall building* as a subset of the set *building*



- Typically, these can be paraphrased to 'X is *tall* for a *guy*' (subsective)
- Cf. **??**'X is *brown* for a *dog*' (intersective)

Adnominal modification

(1) *complete* description

(2) *complete* idiot

(3) *real* idiot

- *complete* vs. *incomplete* description
- *complete* vs. *incomplete* idiot **???**
- *real* vs. *unreal* / *fake* idiot **???**
- Same word, different senses
- A *real* / *complete* idiot is a prototypical member of the set of idiots

Adnominal modification = Modification of Membership

(Constantinescu, 2011; Morzycki, 2011)

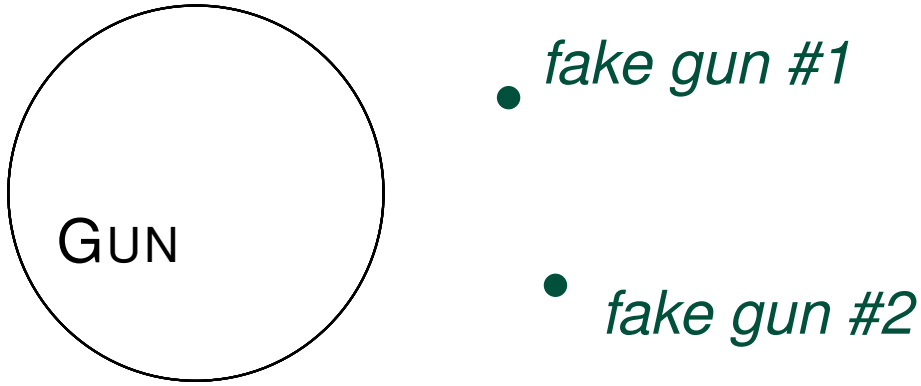
Privative Adjectives

(4) *fake gun*
$\neq$ x is a gun

(5) *former chairman*
$\neq$ x is a chairman

(Partee, 2003; Kamp and Partee, 1995; Del Pinal, 2015)

Privative Adjectives



- For any entity to be called a *fake gun*, it cannot be part of set (authentic) guns
- More generally: fake guns are not guns

Wait, but are all non-guns fake guns?

Not really.
Which one of these can be described as a *fake gun*?

- The entity has to be ‘close enough’ and similar to the prototype
- “resemblence”, “valency” of certain relevant features (to be decided in context) (Raskin and Nirenburg, 1995; Kamp and Partee, 1995)

Privative Adjectives with Abstract Nouns

It gets tricky when we have more abstract entities

(6) *pseudoscience*
$\neq$ science (but not everyone knows that)

(7) *counterfeit bank note*
$\neq$ bank note (but not everyone knows that)

(8) ***fake news***

- These nouns all denote entities that have its form and contents
- Modification of content/function, not observable form
- Cf. *counterfeit signature*; *forged painting* ; *fictitious claim*
- *alleged crime*; *potential threat*; *preemptive defense*

"How fake is fake?"

How $fake_1$ is $fake_2$?

- $fake_1$: Gradable level of resemblence
- $fake_2$: Non-gradable, human perception of being unreliable or untrustworthy

How do the properties of the news article affect our perception / judgment?

How much/little of resemblence qualifies as $fake_2$?

Report on the alleged DPRK plan for sun landing.

How fake are these?

- Clickbaits / factoids / trivia (true but useless information)
- Misinformation (inaccurate, misleading, incorrect information)
- Fake news / disinformation (intentional misinformation, possibly with perceived agenda)
- Scam / conspiracy theories (disinformation with larger scheme to trick readers)

But satire, jokes and stories are presented as fictional and readers are expected to know they are fake.

Properties of Fake News

~~Bad~~ Fake news travels fast!

But why?

- Overflow of information enabled by IT $\rightarrow$ Readers' fatigue in verification (Andrejevic, 2013)
- Cognitive anthropology: Some contents are psychologically more attractive (Acerbi, 2019)
  - strong emotions
  - disgust
  - negative contents
  - threats

Misinformation or sensational articles are in fact "high-quality information", because of the efficiency in spreading!

Why Auto-Detection?

- Content-based (human-like) fact checking is costly and hard
- Understanding of meaning and context is necessary to assess truthfulness of information
- Existing approaches
  - Source of the text (e.g. identified content farms) (Rashkin et al., 2017)
  - Linguistic features (e.g. swearing in the post, pronouns) (Pennebaker et al., 2015; Lam et al., 2020)
  - Sentiment analysis (more emotive language) (Wang, 2017; Rashkin et al., 2017; Shu et al., 2018)
  - Responses from other users (more emojis and swearing) (Jiang and Wilson, 2018)

Automation is great, but it doesn't replace journalistic investigation.

Challenges

Challenges to Auto-Detection of Misinformation:

1. Misinformation as a data problem
2. Misinformation across languages

Misinformation as a Data Problem

- Machine learning and deep learning techniques all require massive data (Asr and Taboada, 2019)
- In short, ML/DL trains an AI by making the system read many many instances.
- Less massive with state-of-the-art model of pre-training + fine-tuning (Peng and Wang, 2020; Ziegler et al., 2020)
- If we don't have enough data for training, the AI is not sensitive to the phenomenon

What if we have fake news in other languages?
The training in English data does not apply (directly)!

Linguistic Challenge 1

(9) Sentences in English use space to separate words!

(10) 中文句子沒有空格分開詞語。

(11) 中文 句子 沒有 空格 分開 詞語。
Chinese sentence not.have space separate word

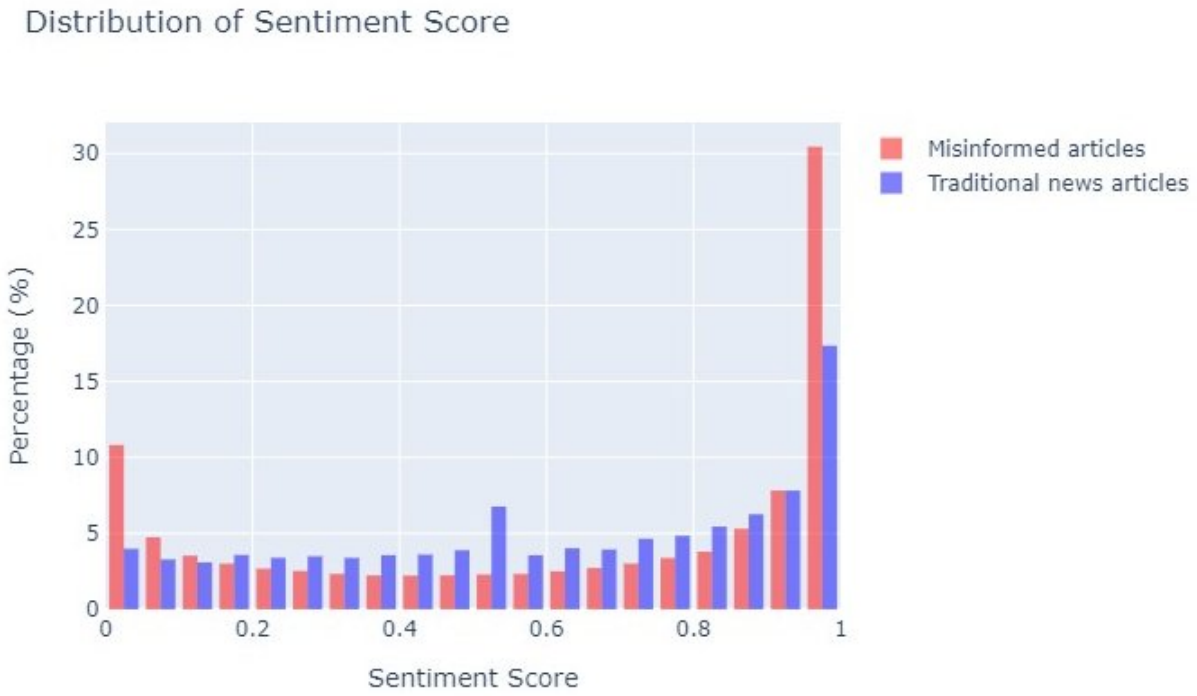
Natural Language Processing

- Word segmentation
- Part-of-Speech tagging (NB: no plural marking, no verb conjugation in Chinese)
- NL understanding (topic, summary, sentiment)

→ **No fool-proof solution, just automated aids to human**

Linguistic Challenge 1

NL understanding / Sentiment Analysis



- Fake news tends to be more emotive Data from Lam et al. (2020); (Acerbi, 2019)
- about 10% “0” (very negative)
- about 30% “1” (very positive)
- traditional news is more neutral in emotion

Linguistic Challenge 2

Misinformation data

- Impoverished in Chinese
- Even for the datasets in Chinese, the quality / usefulness is limited (repetition; trivial content farm posts)

Sociocultural Challenges

Different context, different keywords

- Different ways of fear mongering in Chinese
  - English: more political (cf. `politifact.com/`)
  - Chinese: health; finance; scams
    Frequent n-grams / collocations:
    - 等於/慢性/自殺 equal/chronic/suicide
    - 農民/朋友/注意 farmer/friend/note
    - 微信/聊天/記錄 WeChat/chat/record
- Urban legend stories in Chinese
  - 2014 浙江手機實拍 UFO 不明飛行物！
    (UFO spotted by cell phone in Zhejiang province in 2014!)
  - 1000 人犯罪團伙來德州偷孩子取器官
    (Gang of 1,000 members coming to Texas to steal children for their organs)

Data from ByteDance - WSDM Challenge (2018)

More details in Lam et al. (2020)

Misinformation in Hong Kong

What do we fall for?

- Templatic scam with local celebrities
- Relatively easy to detect

NLP Solution?

Text comparison between two posts with misinformation

譚詠麟 - 一位得過好多音樂獎項嘅歌手兼演員,其中包括最受歡迎男歌手獎同IFPI獎。譚詠麟以自己係敢寸敢言而見稱,亦唔介意話您畀知佢自己啲錢係點賺番嚟。喺上星期,譚詠麟出席由羅沛琪主持嘅"ViuTV erviu"節目訪問時,佢透露咗一個全新"財富漏洞",證明可以喺3-4個月內令任何人變成百萬富翁。譚詠麟催促所有香港人去參與,唔好等到啲大行將呢個漏洞堵塞先發現原來自己錯過咗呢個咁正嘅機會。可以肯定嘅係,喺訪問結束後幾分鐘,匯豐銀行已經打電話嚟叫停止個訪問 - 但其實已經太遲。

甄子丹(Donnie Yen),以武術家、演員及製作人等多重身份著稱。他是香港其中一位頂尖的武打演員,並榮獲多個獎項。甄子丹催促所有香港人去參與,唔好等到啲大行將呢個漏洞堵塞先發現原來自己錯過咗呢個咁正嘅機會。可以肯定嘅係,喺訪問結束後幾分鐘,匯豐銀行已經打電話嚟叫停止個訪問 - 但其實已經太遲。

- It does not solve the core problem
- Targets only the fakest type of fake news
- Obviously the approach can be more sophisticated (fuzzy matching, NER, word embedding)

Outlook

Ongoing collaboration with start-up @ Cyperport HK

Fact Explorer `https://www.f-stem.tech/fact-explorer/`

- Social fact-checking network to annotate, comment and aggregate misinformation
- NLP and deep learning to identify false information

Conclusion

- Semantics of fakeness
- Semantics of fake news articles
- How semantics / linguistics inform NLP and detection of misinformation
- How pragmatics / sentiment analysis can indicate truthfulness / reliability of articles
- Ongoing work with the community in fact-checking (journalists and vigilante) and technological solution

Thank you!
Questions and comments are welcome.

References

Acerbi, A. (2019). Cognitive attraction and online misinformation. *Palgrave Communications*, 5.

Andrejevic, M. (2013). *Infoglut: How too much information is changing the way we think and know*. Routledge.

Asr, F. T. and Taboada, M. (2019). Big data and quality data for fake news and misinformation detection. *Big Data & Society*, 6(1):2053951719843310.

Constantinescu, C. (2011). *Gradability in the nominal domain*. PhD thesis, LOT, Netherlands Graduate School of Linguistics, Utrecht.

Del Pinal, G. (2015). Dual content semantics, privative adjectives, and dynamic compositionality. *Semantics and Pragmatics*, 8(7):1–53.

Jiang, S. and Wilson, C. (2018). Linguistic signals under misinformation and fact-checking: Evidence from user comments on social media. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):1–23.

Kamp, H. and Partee, B. (1995). Prototype theory and compositionality. *Cognition*, 57(2):129–191.

Lam, C., Leung, B., Yip, C., and Yung, J. (2020). A linguistic approach to misinformation in Chinese. In *Proceedings of the Computational Humanities Research*.

Morzycki, M. (2011). The several faces of adnominal degree modification. West coast conference on formal linguistics 29.

Partee, B. H. (2003). Are there privative adjectives? *Manuscript, University of Massachusetts, Amherst*.

Peng, P. and Wang, J. (2020). How to fine-tune deep neural networks in few-shot learning?

Pennebaker, J. W., Booth, R. J., Boyd, R. L., and Francis, M. E. (2015). Linguistic inquiry and word count: LIWC2015. PennebakerConglomerates, Austin, TX.

Rashkin, H., Choi, E., Jang, J. Y., Volkova, S., and Choi, Y. (2017). Truth of varying shades: Analyzing language in fake news and political fact-checking. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2931–2937, Copenhagen, Denmark. Association for Computational Linguistics.

Raskin, V. and Nirenburg, S. (1995). Lexical semantics of adjectives: A microtheory of adjectival meaning. In *MCCS, Las Cruces, N.M.: New Mexico State University*.

Shu, K., Mahudeswaran, D., Wang, S., Lee, D., and Liu, H. (2018). FakeNewsNet: A data repository with news content, social context and spatialtemporal information for studying fake news on social media.

Wang, W. Y. (2017). "Liar, liar pants on fire": A new benchmark dataset for fake news detection. *arXiv preprint arXiv:1705.00648*.

Ziegler, D. M., Stiennon, N., Wu, J., Brown, T. B., Radford, A., Amodei, D., Christiano, P., and Irving, G. (2020). Fine-tuning language models from human preferences.