Incongruent Headlines: Yet Another Way to Mislead Your Readers

Sophie Chesney[†]**, Maria Liakata**[‡]**, Massimo Poesio**[†] and **Matthew Purver**[†]

[†]Cognitive Science Research Group
School of Electronic Engineering and Computer Science
Queen Mary University of London, UK
`{initial.lastname}@qmul.ac.uk`

[‡]Department of Computer Science
University of Warwick, Coventry, UK
`m.liakata@warwick.ac.uk`

Abstract

This paper discusses the problem of incongruent headlines: those which do not accurately represent the information contained in the article with which they occur. We emphasise that this phenomenon should be considered separately from recognised problematic headline types such as clickbait and sensationalism, arguing that existing natural language processing (NLP) methods applied to these related concepts are not appropriate for the automatic detection of headline incongruence, as an analysis beyond stylistic traits is necessary. We therefore suggest a number of alternative methodologies that may be appropriate to the task at hand as a foundation for future work in this area. In addition, we provide an analysis of existing data sets which are related to this work, and motivate the need for a novel data set in this domain.

1 Introduction

The problem of mis- and disinformation in the media is the subject of much recent attention. This is often given the general label 'fake news' – but this term can refer to a number of distinct concepts, from fabricated or manipulated content to satire (Wardle, 2017), each of which might have very different requirements for a computational treatment. In this paper we highlight a specific problem within this realm, that of headline incongruence, show that it is distinct from problems considered within NLP so far, and discuss how it might be approached. Consider (1), taken from the *Express*

UK online newspaper[1] (Ecker et al., 2014):

(1) **Headline:** Air pollution now leading cause of lung cancer
Evidence within article: "We now know that outdoor air pollution is not only a major risk to health in general, but also ***a*** leading ***environmental*** cause of cancer deaths." Dr. Kurt Straif, of IARC [emphasis added]

As Ecker et al. (2014) highlight, this headline misleads the reader by overstating the claim made later in the article. First, omitting 'environmental' from the headline radically generalises the claim: a leading environmental cause may not be the leading cause, above all other causes. Second, omitting the indefinite determiner 'a' (as is common in English 'headlinese', Mårdh, 1980) allows a salient reading with an implicit definite article 'the', further exaggerating the claim.

The headline therefore significantly misrepresents the findings reported in the article itself. While the article reports these accurately, even quoting another source contradicting the exaggerated claim (*"...although air pollution increases the risk of developing lung cancer by a small amount, other things have a much bigger effect on our risk, particularly smoking"*), these nuances are lost in the headline. This seems particularly dangerous in the light of experimental work into reader behaviour: Ecker et al. (2014) show that even after reading the article in full, a reader is likely to be left with their initial impression gained from the headline; and Gabielkov et al. (2016) found that c.60% of shared URLs on Twitter are not clicked on before sharing, suggesting that in

---

[1]Tom Rawle (2013): http://www.express.co.uk/lifestyle/health/437473/Air-pollution-now-leading-cause-of-lung-cancer

many cases only headlines are read. Automatic detection of these misleading cases could therefore directly impact the spread of misinformation.

Indeed, the phenomenon is particularly noticeable on social media, partly due to the practice of using different headlines online. Official posts on social media from some sources include a different headline in the social media post preview than on the article itself, as demonstrated by (2), taken from the *Independent*'s Facebook page.

(2) **Social media post copy:** Enjoy it while you can
Social media headline[2]: Scientists have predicted the end of sex
Article headline[3]: Sex will be made unnecessary by 'designer babies', professor says
Evidence within article: Professor Henry Greely believes that in as little as 20 years, most children will be conceived in a laboratory, rather than through sexual intercourse.

This example shows a gradual increase in accuracy and detail, from the misleading social media post to the evidence within the article itself. The social media headline is incongruent with the details of the story, and this is exaggerated further when combined with the rest of the post. This clearly demonstrates that social media can be used to carefully market stories by exaggerating and twisting key elements of a story in the headline in conjunction with copy in the post itself.

It is important to highlight, however, that this phenomenon is not limited to social media, nor to particular sectors of the press (e.g. tabloid press, press from certain political leanings). We found examples from across the political spectrum, as well as across multiple reputable mainstream sources cross-lingually. Consider examples (3)-(8),[4] which discuss a recent announcement by Volvo Cars on the production of electric cars. As with (1), the headlines consistently exaggerate the claims made in the original press release (8), varying from outright incongruence (3) to subtle quantifier scope ambiguity that leaves interpretation open (6)-(8).

(3) **Dagens Industri*[5]:** Volvo stops developing cars with internal combustion engines

(4) **Independent (Social media headline)[6]:** Petrol cars are dead, say Volvo

(5) **Sveriges Radio*[7]:** Volvo becomes electric car brand

(6) **Göteborgs Posten*[8]:** Volvo to only make electric cars

(7) **Reuters[9]:** Geely's Volvo to go all electric with new models from 2019

(8) **Volvo Cars Press Release[10]:** Volvo Cars to go all electric
Evidence from official press release: Volvo Cars, the premium car maker, has announced that every Volvo it launches from 2019 will have an electric motor, marking the historic end of cars that ***only*** have an internal combustion engine.

The story, which, from the headlines suggests that Volvo Cars will completely stop production of cars with internal combustion engines and only produce electric vehicles, circulated in the mainstream and automotive press. In fact, in-article evidence makes clear that, although all new vehicles produced after 2019 will contain some electric element, many will still contain some petrol or diesel component. Importantly, Volvo Cars CEO Håkan Samuelsson is quoted to say, "this announcement marks the end of the ***solely*** combustion engine-powered car", a nuance which is lost in the headlines above. Interestingly, these examples illustrate that headline incongruence can occur even in sources widely considered as reliable and reputable, such as Reuters (7), as well as in the very source of the story, as in the case of Volvo Cars' own press release (8).

Here, we consider whether existing definitions and NLP techniques can be applied to this phenomenon, and if not, how we may define it and approach its detection computationally. This has ap-

---

[2]https://www.facebook.com/TheIndependentOnline/posts/10154972799736636
[3]Will Worley (2016): http://www.independent.co.uk/news/science/sex-unnecessary-designer-babies-stanford-professor-says-a6957636.html
[4]Examples labelled with * are translated from Swedish, and independently verified by a native speaker.

[5]Håkan Matson (2017): http://www.di.se/bil/volvo-slutar-tillverka-bilar-med-forbranningsmotorer/
[6]https://www.facebook.com/TheIndependentOnline/posts/10154981271001636
[7]http://sverigesradio.se/sida/artikel.aspx?programid=83&artikel=6732573
[8]Ida Johansson (2017): http://www.gp.se/nyheter/ekonomi/volvo-ska-bara-tillverka-elbilar-1.4413878
[9]Niklas Pollard (2017): https://uk.reuters.com/article/us-volvocars-geely-electric-idUKKBN19Q0BJ
[10]https://www.media.volvocars.com/global/en-gb/media/pressreleases/210058/volvo-cars-to-go-all-electric

plications within news aggregation, as a means of weighting articles and informing readers, as well as potential in the incentivisation of journalistic values.

2 Existing Definitions

These cases, then, do not involve misinformation or fabricated content within the article, but rather properties of the headline and its relation to the content. In this section, we examine existing work into the description and classification of problematic types of headline.

2.1 Clickbait

Headlines have traditionally been characterised as short, 'telegram'-like, and maximally informative summaries of the article with which they appear (Van Dijk, 1988; Cotter, 2010). They appear to follow a particular condensed grammar commonly referred to as 'headlinese' (Mårdh, 1980; De Lange, 2008), and are often carefully constructed to attract the attention of a reader (Bell, 1984; Ecker et al., 2014). In extreme cases this results in 'clickbait'-style headlines, characteristic of tabloids and online-native digital media sites such as *Buzzfeed*[11], expressly designed to withhold information to entice the reader to read on, or in most cases, to click. A recent press release by Facebook[12] describes clickbait as *"headlines that intentionally leave out crucial information, or mislead people, forcing people to click to find out the answer"* – see (9)-(11):

(9) You'll Never Believe Who Tripped and Fell on the Red Carpet. . . (Facebook)

(10) Here's What Happens When You Put A Few Little Kids In A Room With 2 Dolls In 2 Different Colors. (Chen et al., 2015)

(11) Which Real Housewife Are You Based On Your Birth Month (Chakraborty et al., 2016)

Clickbait shows characteristic stylistic and lexical features: 'forward-referencing', heavy use of demonstrative pronouns, adverbs, interrogatives, and imperatives (Blom and Hansen, 2015), as well as extensive use of personal pronouns (e.g. 'you'), numbers, and celebrity references (Chen et al., 2015). These features can therefore be used within standard NLP methodologies: Chakraborty et al.

---

[11]www.buzzfeed.com
[12]https://newsroom.fb.com/news/2017/05/news-feed-fyi-new-updates-to-reduce-clickbait-headlines/

(2016) achieved 93% classification accuracy on a corpus including 7,500 English clickbait headlines using a set of 14 such features in a Support Vector Machine (SVM) classifier.

Returning to our example (1), however, although the headline does withhold information and thereby misleads, it does not fulfil our expectation of a clickbait headline. Most importantly, it does not 'force' the reader to click to find out the conclusions of the story, but rather delivers a *misleading* conclusion up front in the headline which (likely purposefully) misinforms the reader on the details in order to frame the facts in a certain light. Consequently, it lacks the typical observable features of clickbait (e.g. forward-referencing, demonstrative pronouns, numbers, etc.), and is therefore unlikely to be detected through these stylometric means. It is therefore rather more subtle than archetypal clickbait as targeted by the methods suggested by Chen et al. (2015); Chakraborty et al. (2016).

2.2 Sensationalism

Some examples labelled as clickbait, however, have a different approach to engaging readers. Chen et al. (2015) also identify the use of affective language and action words, associated with emotional engagement, as in (12):

(12) The first lady of swearing! How a ten-year-old Michelle Obama lost out on a 'best camper' award because she wouldn't stop cursing (Daily Mail, Chen et al., 2015)

While Chen et al. (2015) refer to this example as 'clickbaiting', this arguably introduces a complexity and inconsistency into their definition. This example does not *force* the reader to click by withholding information or using leading language, but instead uses techniques more traditionally considered as *sensationalism*, to dramatise an otherwise non-dramatic story.

Though many definitions exist, sensationalism can be considered as "the presentation of stories in a way that is intended to provoke public interest or excitement, *at the expense of accuracy*" (Oxford Dictionary Online). Sensationalist news is generally considered negatively in the journalism literature (see e.g. Wang and Cohen, 2009), as content which "triggers emotion for the reader (Vettehen et al., 2008) and treats an issue in a predominantly tabloid-like way" (Kilgo et al., 2016). Although traditionally associated with certain topics

e.g. sex, scandal, crime and disaster (Grabe et al., 2001; Vettehen et al., 2008), recent work suggests that it is now just as likely with political stories (Kilgo et al., 2016). Examples (13)-(15) (Molek-Kozakowska, 2013, originally *Daily Mail*) show the characteristic use of exaggeration, emotive language, and punctuation, and cover a range of topics including health, crime, and education:

(13) A sausage a day could lead to cancer: Pancreatic cancer warning over processed meat

(14) Rise of the hugger mugger: Sociable thieves who cuddle while they rob

(15) £100 to play truant! Schools accused of bribing worst pupils to stay away when Ofsted inspectors call

Molek-Kozakowska (2013) views sensationalism as a *discourse strategy* used to repackage information in a more exciting, extraordinary or interesting way, via the presence of several discourse *illocutions* (e.g. exposing, speculating, generalising, warning, and extolling).[13] Based on this view, Hoffman and Justicz (2016) propose a method for automatic sensationalism detection in scientific reporting, training a supervised Maximum Entropy classifier on a corpus of 500 annotated news records, with bag-of-words TF.IDF document vectorisation[14]. They achieve an average accuracy of 73% over 200 validation instances. Crucially, headline and article were not treated separately, so any nuances between the two components will not be captured in this model.

Again, though, while our example (1) does satisfy several aspects of the definitions of sensationalism discussed here (e.g. warning, use of emotive content), it does not do so through the typical stylistic traits seen in (13)-(15). The vocabulary is not particularly inflammatory or emotive, nor is the structure typical of sensationalism. This defines the precise difficulty with the detection of incongruence in headlines this paper aims to highlight: incongruent headlines do not necessarily adhere to an identifiable style in their surface form, but rather must be identified in relation to the text they represent. This presents significant problems for the NLP approaches so far discussed.

---

[13] As Molek-Kozakowska (2013) used only one news source (the *Daily Mail*), this list may be specific to this particular newspaper's voice and/or the knowledge, subjectivity and demographic range of the annotators.

[14] See Hoffman and Justicz (2016, Appendices 1-4).

3 Incongruent Headlines: Suggested Methodology

The relationship between a headline and the article with which it appears can be conceptualised in a number of ways. We propose novel methods of incongruence detection which would explore varying aspects of the phenomenon, based on existing work in other areas. It is clear from the cross-source examples (3)-(8) that relying on source information alone is unlikely to be sufficient in determining headline incongruence, given that this phenomenon does not seem to be strictly limited to one section of the press. However, in conjunction with other methodology, the source of the headline-article pair may well prove to be a useful feature in the broader classification process, which we will explore experimentally in future work.

Arguably, the task of headline incongruence detection is best approached in parts: to analyse complex relationships between a headline and an entire news article is likely to be extremely difficult, not least because of their very different lengths and levels of linguistic complexity. This could therefore be facilitated with the extraction of key quotes (Pouliquen et al., 2007) or claims (Vlachos and Riedel, 2015; Thorne and Vlachos, 2017). Alternatively, one could automatically generate the statistically 'best' headline for an article using existing title and headline generation and summarisation methods (e.g. Banko et al. (2000); Zajic et al. (2002); Dorr et al. (2003)), and evaluate how far away the existing headline is from this in terms of a number of criteria, such as lexical choices, syntactic structure, length, tonality (sentiment or emotion), and so on.

It may also be interesting to explore existing work on argument analysis: for example, Stab and Gurevych (2017) explore methods for the identification of arguments supported by insufficient evidence. This could be viewed as very close to the task of the detection of incongruent headlines, where the headline represents an argument which is not supported by claims in the text. Further, we could approach incongruence as a semantic issue and look to existing work on contradiction (De Marneffe et al., 2008), contrast (Harabagiu et al., 2006) and entailment recognition (Levy et al., 2013). In doing so, we may well discover several sub-types of incongruence which may fall into different semantic categories.

Finally, stance detection (Augenstein et al., 2016; Mohammad et al., 2016) has been applied in the Fake News Challenge (FNC-1)[15] as a means of exploring whether different articles agree or disagree with a given headline or claim, to aid in the task of fact checking. Stance is certainly relevant to task of incongruence detection, but we argue that it is not sufficient for our task, as the headline-article relation may be incongruent in ways separate from (dis)agreement. Beyond the headline-article pair itself, however, stance detection could be used to analyse engagement and interaction with an article on social media, given that early indications suggest that users are compelled to alert others when they notice that a headline is misleading.

4 Existing Data

A number of data sets are available which address related tasks, but none seem directly suited to the incongruence problem. The Clickbait Challenge[16] released a data set of 2495 social media posts (titles, articles and social media copy), labelled on a four-point scale (not/slightly/considerably/heavily clickbaiting) through crowdsourcing. Although precise guidelines for the annotation process are not provided, it seems that the organisers follow a definition of clickbait similar to those discussed in Section 2.1, in which posts are "designed to entice readers into clicking an accompanying link". As already emphasised, this differs from the concept of headline incongruence described here, and we do not expect this annotation to be useful for our task; however, as a source of paired titles and articles it may provide useful raw data.

Piotrkowicz et al. (2016) present a corpus of 11,980 *Guardian News* headlines automatically annotated with news values (prominence, sentiment, superlativeness, proximity, surprise, and uniqueness). Although this corpus does not contain a target class in line with headline incongruence, it may provide useful insight in the feature extraction process.

The Fake News Challenge (FNC-1) has released a corpus of headline-article pairs which are annotated with one of the following four stances:

Agrees: *The body text agrees with the headline.*
Disagrees: *The body text disagrees with the headline.*

Discusses: *The body text discuss the same topic as the headline, but does not take a position.*
Unrelated: *The body text discusses a different topic than the headline.*

Built on the data set described in Ferreira and Vlachos (2016), which is collected from rumour tracking website, Emergent[17], the corpus contains approximately 50,000 annotated headline-body pairs. A manual analysis of the first 50 body IDs led to a number of observations on the applicability of this data set to the problem of headline incongruence. Firstly, the 'headline' in a pair is the claim from the original post on the website, and is as such not necessarily a gold-standard headline. In addition, a single 'headline' can occur with multiple article bodies, and vice versa, which means that the original relation between the two is not captured. In our task, we are particularly interested in how a headline is utilised to (mis)represent the information in an article; it is therefore important that the data we use reflects these subtle connections and disconnections, a feature that may be lost when pairing a headline (or claim) with an article body at random. The *unrelated* class in this data set is therefore unlikely to be relevant, as it appears to reflect a random shuffling of headline-body pairs. The *disagree* class represents contradictions between headline and body, which is too strong a notion of incongruence for our purposes; disagreement represents a direct contrast, whereas incongruence can be a subtle exaggeration or misrepresentation of facts but need not represent an opposing view. If this data set contains incongruent headline-body pairs by our definition, it appears that they are not in line with the existing labels, therefore it cannot be used in its current form.

5 Conclusions

The paper discusses incongruent headlines and how we may approach their automatic detection using existing NLP methods, motivated by experimental evidence on reader behaviour (Ecker et al., 2014; Gabielkov et al., 2016). We emphasise that headline incongruence, as seen in example (1), cannot be approached through methodology applied to related concepts like clickbait and sensationalism, as these use headline-specific stylometric features, and do not consider any deeper semantic relation between headline and text that

---

[15]http://www.fakenewschallenge.org/
[16]http://www.clickbait-challenge.org/

[17]http://www.emergent.info/

would be critical to the task at hand. We consequently suggest a number of potential approaches for this task, based on existing work in summarisation and headline generation, stance detection, claim and quote extraction, as well as argument analysis. Finally, we discuss a number of existing data sets, but demonstrate that, in their current forms, none are appropriate for the task discussed here. This therefore motivates the need for a novel data set in this domain, which lays the foundation for the next stages of our future work.

References

I. Augenstein, A. Vlachos, and K. Bontcheva. 2016. Any-target stance detection on twitter with autoencoders. In *Proc. SemEval*.

M. Banko, V. O. Mittal, and M. J. Witbrock. 2000. Headline generation based on statistical translation. In *Proc. ACL*.

A. Bell. 1984. 'Good copy-bad news'. the syntax and semantics of news editing. *Applied Sociolinguistics* pages 73–116.

J. N. Blom and K. R. Hansen. 2015. Click bait: Forward-reference as lure in online news headlines. *Journal of Pragmatics* 76:87–100.

A. Chakraborty, B. Paranjape, S. Kakarla, and N. Ganguly. 2016. Stop clickbait: Detecting and preventing clickbaits in online news media. In *Proc. IEEE/ACM ASONAM*.

Y. Chen, N. Conroy, and V. Rubin. 2015. Misleading online content: Recognizing clickbait as false news. In *Proc. ACM Multimodal Deception Detection*.

C. Cotter. 2010. News talk. *Investigating the Language of Journalism. CUP, Cambridge* .

J. De Lange. 2008. *Article Omission in Headlines and Child Language: A Processing Approach*. Netherlands Graduate School of Linguistics.

M.-C. De Marneffe, A. N. Rafferty, and C. D. Manning. 2008. Finding contradictions in text. In *ACL*.

B. Dorr, D. Zajic, and R. Schwartz. 2003. Hedge trimmer: A parse-and-trim approach to headline generation. In *Proc. HLT-NAACL*.

U. K. Ecker, S. Lewandowsky, E. P. Chang, and R. Pillai. 2014. The effects of subtle misinformation in news headlines. *Journal of Experimental Psychology: Applied* 20(4):323.

W. Ferreira and A. Vlachos. 2016. Emergent: a novel data-set for stance classification. In *NAACL-HLT*.

M. Gabielkov, A. Ramachandran, A. Chaintreau, and A. Legout. 2016. Social Clicks: What and Who Gets Read on Twitter? In *ACM SIGMETRICS / IFIP Performance 2016*.

M. E. Grabe, S. Zhou, and B. Barnett. 2001. Explicating sensationalism in television news: Content and the bells and whistles of form. *Journal of Broadcasting & Electronic Media* 45(4):635–655.

S. Harabagiu, A. Hickl, and F. Lacatusu. 2006. Negation, contrast and contradiction in text processing. In *AAAI*.

S. J. Hoffman and V. Justicz. 2016. Automatically quantifying the scientific quality and sensationalism of news records mentioning pandemics: validating a maximum entropy machine-learning model. *Journal of Clinical Epidemiology* 75:47–55.

D. K. Kilgo, S. Harlow, V. García-Perdomo, and R. Salaverría. 2016. A new sensation? an international exploration of sensationalism and social media recommendations in online news publications. *Journalism* .

O. Levy, T. Zesch, I. Dagan, and I. Gurevych. 2013. Recognizing partial textual entailment. In *ACL*.

I. Mårdh. 1980. *Headlinese: On the grammar of English front page headlines*, volume 58. Liberläromedel/Gleerup.

S. M. Mohammad, S. Kiritchenko, P. Sobhani, X. Zhu, and C. Cherry. 2016. Semeval-2016 task 6: Detecting stance in tweets. *Proceedings of SemEval* 16.

K. Molek-Kozakowska. 2013. Towards a pragmalinguistic framework for the study of sensationalism in news headlines. *Discourse & Communication* 7(2):173–197.

A. Piotrkowicz, V. Dimitrova, and K. Markert. 2016. Automatic extraction of news values from headline text. In *Proceedings of EACL*.

B. Pouliquen, R. Steinberger, and C. Best. 2007. Automatic detection of quotations in multilingual news. In *Proc. RANLP*.

C. Stab and I. Gurevych. 2017. Recognizing insufficiently supported arguments in argumentative essays. In *Proc. EACL*.

J. Thorne and A. Vlachos. 2017. An extensible framework for verification of numerical claims.

T. A. Van Dijk. 1988. *News as discourse*. Revised edition. Routledge (2013).

P. H. Vettehen, K. Nuijten, and A. Peeters. 2008. Explaining effects of sensationalism on liking of television news stories: The role of emotional arousal. *Communication Research* 35(3):319–338.

A. Vlachos and S. Riedel. 2015. Identification and verification of simple claims about statistical properties. In *Proc. EMNLP*.

T.-L. Wang and A. A. Cohen. 2009. Factors affecting viewers perceptions of sensationalism in television news: A survey study in Taiwan. *Issues and Studies* 45(2):125–157.

C. Wardle. 2017. Fake news. it's complicated. *First Draft News* Retrieved from https://firstdraft-news.com/fake-news-complicated/ on 04/05/2017.

D. Zajic, B. Dorr, and R. Schwartz. 2002. Automatic headline generation for newspaper stories. In *Workshop on Automatic Summarization*.