

THE LINGUISTIC AND TYPOLOGICAL FEATURES OF CLICKBAIT IN YOUTUBE VIDEO TITLES

Roy Kemm¹

Abstract

This exploratory study aims to identify which linguistic and typological features commonly associated with clickbait in online news headlines are indicative of clickbait in YouTube video titles. A comparative corpus analysis is conducted to compare YouTube video titles commonly associated with clickbait to titles not associated with clickbait. Results indicate that a majority of the typological and linguistic features associated with clickbait in online news headlines are found to be indicative of clickbait in YouTube video titles. However, the role which each of the features plays seems to differ to that of online news. The findings contribute to the understanding of clickbait in non-news contexts from a linguistics perspective, an area which has been relatively unexplored in the current literature.

Keywords: clickbait, YouTube titles, linguistics, digital media, clickbait detection, corpus analysis

Introduction

Deception has been a persistent phenomenon in online communication for decades (Hancock & Gonzalez, 2013). Currently, the most widespread deceptive practice may be clickbait (Zannettou et al., 2018): the act of convincing a user to click a link they otherwise would not so as to increase click-based advertising revenue (Biyani et al., 2016). Clickbait can frustrate users and crowd out higher quality content (Rony et al., 2017). The scale of the clickbait problem is evident in both the recent attention being paid to the phenomenon by the academic community (Tafesse, 2020) and in the efforts of online media outlets and social media platforms to detect and remove clickbait content.

Much of the precursory research surrounding clickbait has focused on its use in online news, with studies from a various academic fields having identified a number of linguistic and typological features associated with clickbait headlines (Alves et al., 2016; Chakraborty et al., 2016; Potthast et al., 2016; Qu et al., 2018; Tafesse, 2020). Yet, as noted by Scott (2021), "very little work has considered clickbait headlines from a linguistic or pragmatic perspective" (p.54). YouTube is the most visited social media site globally with an estimated 122 million daily users (Tafesse, 2020). Given the vast number of visi-

¹ Nago City Board of Education; University of Birmingham, email: r.kemm@mail.meio-u.ac.jp, ORCID: 0000-0002-5166-0426

tors YouTube attracts, it is no surprise that clickbait has become a pressing issue on the platform. The exploratory study presented in this paper attempts to determine which of the features of clickbait identified in the literature surrounding online news can be considered as indicators of clickbait in YouTube video titles.

Literature Review

Clickbait is a broad term which refers primarily to the way in which online content is advertised (Potthast et al., 2016; Qu et al., 2018). Clickbait is not a new phenomenon (Mormol, 2019), the technique of enticing readers through the use of exaggerated and sensationalistic headlines can be found in tabloid newspapers as far back as the 1980s (Chen et al., 2015). From a journalistic perspective, news headlines have been transformed by digital reporting into elements intended to attract reader attention rather than provide information (Jiang et al., 2019). The use of clickbait techniques seems to be spreading to include mainstream news media and entertainment content (Rony et al., 2017) and has been thrust into the spotlight by the recent proliferation of fake news (Zannettou et al., 2018). Yet, as noted by Potthast et al. (2018), the working mechanisms of clickbait are still not adequately understood.

Few studies to date have focused on clickbait in the context of YouTube despite its far-reaching popularity and immense user base. Much of the existing literature has focused on the classification and detection of clickbait in online news (Blom and Hansen, 2014; Potthast et al., 2016; Orosa et al., 2017; Rony et al., 2017; Mormol, 2019; Genç and Surer, 2021; Scott, 2021). In the context of YouTube, the advertising of a video is achieved primarily through the video title and thumbnail image; the most prominent pieces of information which users are presented with (Lopezosa et al., 2020; Tafesse, 2020). Additional YouTube video metadata such as the video description, tags, number of likes and dislikes and user comments, only available to the user after clicking the link, may still have an impact on the way in which search engines index a video (Tafesse, 2020) and thereby indirectly contribute to the promotion of clickbait videos.

Defining Clickbait

Despite the increasing attention being given to clickbait, there is still no agreed upon universal definition of what clickbait is (Kuiken et al., 2017). Mormol (2019) states that “the available descriptions of clickbait do not succeed in providing a coherent picture of the phenomenon in question” (p.3). Table 1 presents definitions of clickbait found in the literature.

Table 1. Definitions of Clickbait.

Definition	Context
“content of which the main purpose is to attract attention and encourage visitors to click on a link to a particular web-page” (Chen et al., 2015, p.15)	Online ‘soft’ news websites (headlines)
“a certain kind of web content advertisement that is designed to entice its readers into clicking an accompanying link” (Potthast et al., 2016, p.810)	Online news Twitter posts
News (headlines) which generate clicks via catchy, provocative, or sensationalist headlines aimed at exploiting the curiosity of users (Orosa et al., 2017)	Online news websites (headlines)
“a form of web content that employs writing formulas and linguistic techniques in headlines to trick readers into clicking links, but does not deliver on promises.” (Rony et al., 2017, p.232)	News headlines on Facebook

“a marketing instrument employed by many publishers on social media that entices and manipulates users to click on a certain link by using eye-catching teaser content, exaggerated descriptions, by omitting key information, or even outright deception – irrespective of whether users are actually interested in the content’s topic or not.” (Qu et al., 2018, p.1)	YouTube metadata (all metadata)
“online content that is specifically designed to entice a reader to click on a link but which offers very little reward for doing so.” (Scott, 2021, p.54)	Online ‘soft’ news websites (headlines)
Content which is “purposely designed to attract the user’s attention and make them curious to follow the link and read, view, or listen to the attached content” (Varshney & Vishwakarma, 2021, p.4212)	YouTube metadata (user comments)
“an advertisement aimed at attracting visitors’ attention and encouraging them to click on an attached link” (Elyashar et al., 2017, p.1)	Online news Twitter posts
an umbrella term, used to describe many different techniques, all with the common goal of attracting attention and invoking curiosity to get the reader to click on a headline (Kuiken et al., 2017)	Online news websites (headlines)

Source: own research.

Although not apparently different, the definitions of clickbait offered in the literature often specify the nature of clickbait content in a contextually convenient manner. The aspects of clickbait universal to all the definitions in Table 1 are (a) clickbait has a deceptive intent, and (b) clickbait exploits the curiosity of readers to generate clicks.

The very nature of clickbait is deceptive, as it aims to convince a reader to click a link they might otherwise not click. The underlying deceptive intent of clickbait has been widely criticised (Alves et al., 2016; Biyani et al., 2016; Chakraborty et al., 2016) as the linked content often fails to satisfy expectations set by the headline. These criticisms are often based on a journalistic understanding of headlines; that of autonomous texts intended to inform readers of the basic content of an article (Dor, 2003). YouTube video titles, being unbound by journalistic standards, may well function to attract reader attention. However, as noted by Tafesse (2020), titles are still obliged to offer some measure of information as to the content of a video.

Deception in headlines can be understood through relevance theory (Sperber & Wilson, 1995, 2008), which proposes that readers will interpret a headline in a way which maximises its relevance to the individual through assumptions a reader naturally makes about the intended meaning of a headline. Ifantidou (2009), drawing on relevance theory, proposes that under informative, vague, or ambiguous headlines are perceived as more relevant by readers because they allow for more assumptions to be made. In the context of YouTube, vague or ambiguous titles which deemphasize the provision of information can be considered as clickbait.

From a cognitive perspective, the efficacy of clickbait can be explained by Loewenstein’s (1994) information-gap theory of curiosity. Loewenstein suggests that a reader’s curiosity is piqued by the highlighting of a gap in their knowledge, creating a sense of deprivation, and thereby motivating the reader to click a link (Chakraborty et al., 2016). Mormol (2019) explains this ‘curiosity gap’ phenomenon in the context of online news headlines by stating that “the determination to fill in the existing knowledge gap is enough to entice the reader into clicking on any headline which alludes to events with undisclosed outcomes” (p.3). Following Loewenstein’s theory, YouTube videos with titles which employ techniques to intentionally pique reader curiosity through the creation of a curiosity gap can be considered to be clickbait.

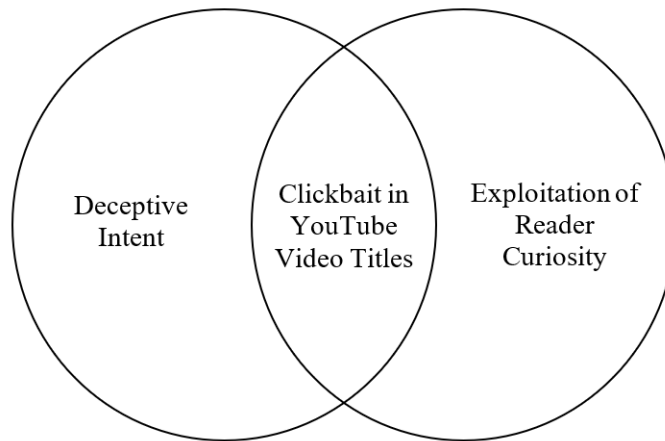


Figure 1. Relationship Between the Defining Principles of Clickbait. Source: own research.

As YouTube is a largely informal, creative platform unbound by the standards of news journalism, titles to be considered as clickbait must contain some measure of both defining principles of clickbait, as illustrated in Figure 1.

Features of Clickbait

Empirical research has revealed several linguistic and typological features of clickbait in news headlines. These features represent the mechanisms by which the definitive principles of clickbait are realised in a news headline. Table 2 lists only the most predominant and recurrent linguistic and typological features associated with clickbait present in the literature and is by no means exhaustive.

Table 2. Linguistic and Typological Features of Clickbait.

Clickbait Feature	Description
Overuse of personal and demonstrative pronouns (Qu et al., 2018; Scott, 2021)	Personal pronouns can create a false sense of relevance for the reader, while demonstrative pronouns exploit reader curiosity.
Overuse of superlative adjectives and intensifiers (Mormol, 2019; Scott, 2021)	Superlative adjectives often indicate exaggeration and sensationalism in headlines.
The use of forward-reference (Blom and Hansen, 2014)	Referring to undisclosed outcomes or information (and implied to be present on the landing page) creating an information-gap.
Emotional appeal (Lockwood, 2016; Tafesse, 2020)	Creating emotional appeal through the wording of a headline, often at the expense of providing information about the content.
Hyperbolic words and strong lexical units (Alves et al., 2016; Mormol, 2019)	Strong lexical units and hyperbole often indicate exaggeration contributing to the deceptive nature of a headline.
Using numerals (Kuiken et al., 2017; Mormol, 2019)	Numerals attract reader attention and contribute to creating an information gap, piquing reader curiosity.
Use of upper case (Biyani et al. 2016)	A typological feature which attracts reader attention and contributes to exaggeration in headlines.
Distinctive punctuation patterns (Chakraborty et al., 2016; Qu et al., 2018)	Punctuation which serves no purpose other than to attract attention to a headline.
Modals (Molina et al., 2021)	Modal verbs create a false sense of relevance as they often and contribute to the creation of an information-gap.
Shortened forms (Chakraborty et al., 2016)	The use of shortened word forms such as contractions, semantic symbols, numeronyms (l8r), and leet speak.

Interrogative forms (Lockwood, 2016; Kuiken et al., 2017)	Questions create an information-gap and may also contribute to a false sense of relevance.
Prioritization of familiar (and trending) vocabulary (Ifantidou, 2009)	The use of familiar and trending words attracts reader attention and can create a false sense of relevance.

The features of clickbait listed in Table 2 above will be examined in order to answer the following research question: To what extent do the linguistic and typological features of clickbait, identified in the literature surrounding online news, indicate clickbait in YouTube video titles?

Method

A comparative corpus analysis using Sketch Engine is employed to identify which of the linguistic and typological features associated with clickbait headlines are indicators of clickbait in YouTube video titles. Two corpora, consisting of video titles taken from channels associated with the use of clickbait (CB) (N=1000) and channels not associated with the use of clickbait (NCB) (N=1000) were compiled. The 100 most recent video titles were selected from each of ten CB channels and ten NCB channels across five genres of content, illustrated in Table 3. Channels were selected on the basis of the extent to which the most recent titles aligned with the defining principles of clickbait discussed above, and on user comments associated with each channel.

Table 3. Corpora Information.

Clickbait Channels				
Channel Name		Genre	Total Words (N=7,870)	Total Tokens (N=10,431)
Troom Troom	CB1	DIY & Crafts	845	999
5-Minute Crafts	CB2	DIY & Crafts	988	1,191
AliA	CB3	Gaming	758	1,329
PewDiePie	CB4	Gaming	574	723
Morgz	CB5	People & Lifestyle	775	1,143
RiceGum	CB6	People & Lifestyle	721	950
Mrwhostheboss	CB7	Science & Technology	623	840
Linus Tech Tips	CB8	Science & Technology	751	968
Dahr Mann	CB9	Other	1,107	1,334
Top5Central	CB10	Other	728	954

Non-clickbait Channels				
Channel Name		Genre	Total Words (N=6,774)	Total Tokens (N=8,228)
DIY Creators	NCB1	DIY & Crafts	723	845
I Like To Make Stuff	NCB2	DIY & Crafts	775	965
IGN	NCB3	Gaming	785	1,020
xisumavoid	NCB4	Gaming	603	827
Jason Vlogs	NCB5	People & Lifestyle	766	784
Vegabrothers	NCB6	People & Lifestyle	803	973
Vsauce	NCB7	Science & Technology	412	551
Mark Rober	NCB8	Science & Technology	691	874
Ozzy Man Reviews	NCB9	Other	390	468

Tom Scott	NCB10	Other	826	921
-----------	-------	-------	-----	-----

Note: 'Tokens' includes all words, punctuation marks, and symbols.

Qu et al. (2018) state that YouTube content creators “are more or less forced to employ clickbait to avoid their content from being lost among all the catchy titles” (p.1). Thus, to allow for a fairer representation of the nature and extent of clickbait use in YouTube video titles, the choice was taken to use video titles from channels associated with clickbait rather than selecting clickbait video titles individually.

Results and Analysis

Of the twelve linguistic and typological features of clickbait identified in online news, ten were found to be at least partially indicative of clickbait in YouTube video titles. Interrogative forms were found to be more a product of the type of content, being far more prominent in the science and technology sub-corpora than any other. Shortened word forms, including both contractions and non-standard shortenings, were found to be statistically insignificant in the clickbait corpus relative to the non-clickbait corpus.

Personal and Demonstrative Pronouns

Personal and demonstrative pronouns are identified by Chakraborty et al. (2016) and Scott (2021) as indicators of clickbait in online news headlines. Table 4 lists the most statistically significant personal and demonstrative pronouns found in the corpora.

Table 4. Distribution of Personal and Demonstrative Pronouns.

Personal pronoun	Log-likelihood	Use in the clickbait corpus relative to the non-clickbait corpus
1st person singular subject (I)	+26.24	Significant overuse
1st person singular object (me)	+21.79	Significant overuse
1st person singular poss. (my)	+64.4	Significant overuse
1st person plural subject (we)	-0.32	No significance
1st person plural object (us)	+5.2	Significant overuse
2nd person (you)	+34.47	Significant overuse
3rd person singular neutral (it)	+54.41	Significant overuse
3rd person plural subject (they)	+15.57	Significant overuse
Plural near (these)	+4.21	Significant overuse
Singular near (this)	+28.48	Significant overuse

Personal pronouns can contribute to the clickbait nature of a title in a number of ways. The use of the unresolved pronoun “she” in (1) creates an information gap, leaving the reader wondering who “she” is.

(1) She Surprises Me For Valentines Day!!! (CB6)

(2) How THIS Instagram story kills your phone. (CB7)

The personal pronoun “your” in (2) directly addresses the reader. Scott (2021) notes that “presenting the information as directly related to the reader, gives the writer another opportunity to create an information gap” (p. 62). The information-gap is compounded by the use of the demonstrative pronoun “this”, further piquing reader curiosity. The use of personal and demonstrative pronouns can be considered an indicator of clickbait in YouTube video titles.

Superlative Adjectives and Intensifiers

Superlative adjectives and intensifiers are associated with extremes, and thereby contribute to exaggeration in video titles, a feature consistent with clickbait (Scott, 2021). Table 5 shows a significant overuse of all types of adjectives and adverbs in the clickbait corpus.

Table 5. Distribution of Adjectives and Adverbs.

Part-of-speech	Log-likelihood	Use in the clickbait corpus relative to the non-clickbait corpus
General adjectives	+30.66	Significant overuse
Comparative adjectives	+13.47	Significant overuse
Superlative adjectives	+11.57	Significant overuse
General adverbs	+35.41	Significant overuse
Comparative adverbs	+23.24	Significant overuse
Superlative adverbs	+12.16	Significant overuse
Intensifiers (degree adverbs)	+14.93	Significant overuse

This finding stands in contrast to Scott (2021) who found a significant underuse of only general and comparative adjectives in clickbait headlines. This discrepancy may be explained by the informal nature of YouTube, where video titles can employ any number of linguistic and typological features such as unique punctuation, numerals, and an abundance of symbols to attract reader attention. News headlines, however, may be limited to more traditional choices such as the use of a superlative over a general adjective.

(3) Zenfone 8 – The most boring INCREDIBLE phone ever. (CB7)

(4) Fortnite’s ACTUALLY doing *THIS*! (CB3)

In (3) the intensifier “most” together with the adverb “ever” serves to exaggerate the degree to which the phone is both “boring” and “incredible”. This hyperbolic exaggeration is a typical indicator of clickbait. In (4) the adverb “actually” serves to amplify the information-gap formed by the use of the unresolved demonstrative pronoun “this” by adding an element of disbelief. The use of adjectives and adverbs to exaggerate a title is indicative of clickbait.

Forward-reference

Forward-reference is a linguistic device whereby reference is made to upcoming parts of a text. Blom & Hansen (2014) distinguish two forms of forward-reference: (a) discourse deixis, which refers to forthcoming parts of the discourse relative to the current location as in (5) where the interrogative structure implicitly refers to an answer provided in the associated video, and (b) cataphora which refers to an upcoming part of the sentence as in (6) where “this YouTuber” refers cataphorically to the YouTube channel H3H3 in parenthesis.

(5) Can They Watercool Better Than Me? (CB8)

(6) This YouTuber is Lying To You (H3H3) (CB6)

Blom & Hansen (2014) state that “since discourse deixis and cataphora do not carry any full content in headlines, they must, tentatively, be considered a device primarily used for creating anticipation and making readers click” (p.89). As shown in Figure 2, forward-reference is a prominent feature in video titles of channels associated with clickbait.

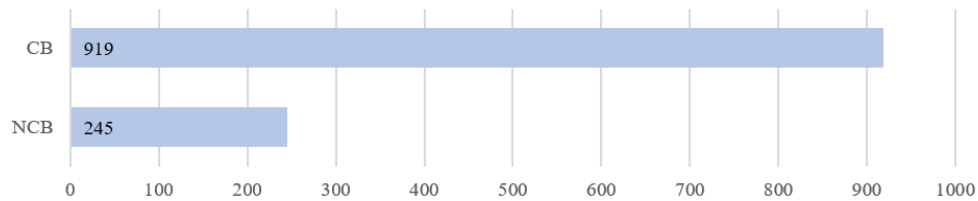


Figure 2. Video Titles Employing Forward-referencing Techniques (N=1000).

The results differ from those obtained by Blom & Hansen (2014) who identified forward-referencing in only 17.2% of online news headlines. This discrepancy could be explained in part by the fact that sites which generate revenue through advertising rather than subscription charges, such as YouTube, are predisposed to clickbait (Blom & Hansen, 2014; Chakraborty et al., 2016). Forward-referencing is a strong indicator of clickbait in YouTube titles.

Emotional Appeal

In a study of online news headlines Orosa et al. (2017) concluded that clickbait headlines are characterised by the way in which they “privilege sensations and emotions over information” (p. 1267). The results of the analysis shown in Figure 3 indicate that emotional appeal is present in a much larger proportion of clickbait video titles than non-clickbait video titles.

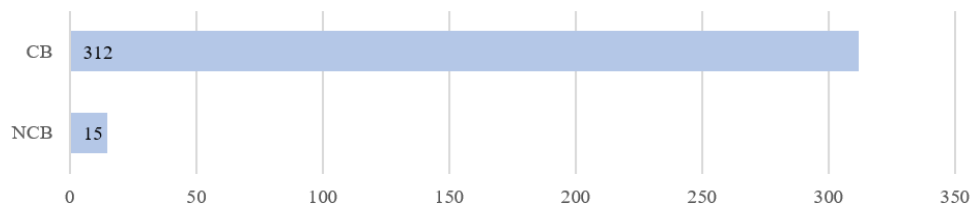


Figure 3. YouTube Titles Prioritising Emotional Appeal (N=1000).

Although several different techniques may be employed to elicit emotions on the part of the reader, it is the prioritisation of emotional appeal over the provision of information which determines whether a title can be considered clickbait.

(7) Homeless Man’s LIFE IS CHANGED, What Happens Is Shocking (CB9)

(8) Why Everyone In China Hates Me... (CB6)

In (7) the writer evokes contradictory emotions without providing any information of the content of the video, a common technique used to create emotional appeal in a title (Mormol, 2019). Tafesse (2020) shows that framing an emotive title negatively, as in (8), generates more clicks than framing a title positively.

Hyperbole and Strong Lexical Units

The use of strong lexical units and hyperbole function to exaggerate or sensationalise a title (Chakraborty et al., 2016). Strong lexical units and hyperbole are words or phrases which act to exaggerate a title beyond necessity. Examples of strong lexical units may be words associated with danger (“warning”), death (“kill”, “massacre”), violence (“rape”, “assault”), sexuality (“sexy”, “nude”), profanity, and attention-grabbing words such as “free”, or “new”. Figure 4. illustrates that hyperbole and strong lexical units are a more common feature in clickbait titles.

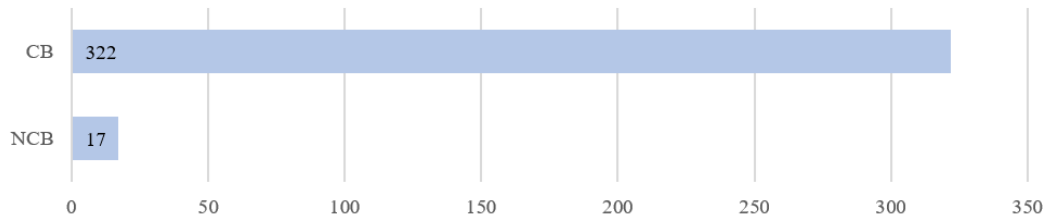


Figure 4. Titles Containing Hyperbole and Strong Lexical Units (N=1000).

Given that attracting reader attention may be an acceptable function of video titles, the use of hyperbole and strong lexical units may not be as indicative of clickbait as in news headlines. However, it is often the case that the sensationalism created by hyperbole and strong lexical units can contribute to the deceptive nature of YouTube titles.

(9) Tik Toks That Will Melt your Brain (CB4)

(10) *WARNING* – Fortnite is in DANGER! (HELP) (CB3)

In (9) the use of the hyperbolic phrase melt your brain acts to pique reader curiosity while offering no information about the content of the video, as does the strong lexical units “warning” and “danger” in (10). This use of strong lexical units and hyperbole may indicate clickbait.

Numerals

The use of numbers attracts reader attention and contributes to creating an information-gap (Potthast et al., 2016; Kuiken et al., 2017). The most common use of numbers in clickbait titles may be ‘listicles’ (Vijgen, 2014), which are “articles that are simply lists or rankings” (Sadri, 2019). Numeral occurrences were divided into three categories: (a) listicles, (b) monetary amounts, and (c) other uses including proper nouns (iPhone 12), acronyms (3D, 4K) and technical terms (240p).

Table 6. Distribution of Numerals.

Numeral Usage	Log-likelihood	Use in the clickbait corpus, compared to the non-clickbait corpus
List	+73.3	Significant overuse
Amount (monetary)	+28.6	Significant overuse
Other	-63.68	Significant underuse

As indicated in Table 6, lists and monetary amounts showed significant overuse in the clickbait corpus. Numerals are often combined with other techniques as a way of amplifying the curiosity of the reader. In (11) the use of numerals is combined with hyperbolic word “insane”, superlative adjective “most”, adverb “ever”, and the use of capitalisation.

(11) Top 10 MOST INSANE Sandcastles EVER BUILT (CB10)

(12) \$100 ROAD TRIP vs \$10,000 ROAD TRIP – Challenge (CB5)

The use of monetary amounts in (12) immediately attracts reader attention while the preposition vs deictically refers to upcoming parts of the discourse. The significant underuse of numbers in the ‘other’ uses category suggests that numbers may indicate clickbait under certain conditions only.

Capitalisation

The use of upper-case in YouTube titles is far more prominent in clickbait titles than in non-clickbait titles, as illustrated in Table 7. Accepted uses of capitalisation including

acronyms (DIY, ATM) and proper nouns (FIFA, NASA) were not included in the analysis.

Table 7. The Use of Capitalisation in Each Corpus

Corpus	Number of titles containing capitalised words (N=1000)	Total number of capitalised words in each corpus	% Capitalised words in each corpus
Clickbait	548	1,664	21.14%
Non-Clickbait	104	254	3.74%

Mormol (2019) states that capitalisation as a feature of clickbait “lures the reader by means of diversified typography that includes a mixture of upper- and lower-case”. Of the 548 titles containing upper-case words in the clickbait corpus, only 36 were capitalised in their entirety.

(13) Top 5 LUXURIOUS Airplane Seats YOU WON’T BELIEVE EXIST. (CB10)

In (13) the choice of capitalised words acts to highlight the superlative adjective “luxurious” as well as the hyperbolic phrase “you won’t believe exist”. In this example, the capitalised words are those which contribute to the clickbait nature of a title. Capitalisation is not clickbait in and of itself as its use does not contribute to the creation of a knowledge-gap or a false sense of relevance. Rather, the primary purpose of capitalisation seems to be to attract reader attention. The overrepresentation of capitalised words in the clickbait corpus suggests that capitalisation may indicate the presence of other clickbait features in YouTube titles.

Punctuation and Symbols

A wide variety of punctuation and symbols were identified in the corpora. The most significant of which are presented in Table 8. Many of these typological features are seemingly used to attract reader attention.

Table 8. Distribution of Punctuation and Symbols.

Punctuation	Log-likelihood	Use in the clickbait corpus, compared to the non-clickbait corpus
*	+ 234.66	Significant overuse
...	+ 56.98	Significant overuse
!	+ 159.58	Significant overuse
?	- 5.80	Significant underuse
/	- 0.38	No significance
()	+ 11.31	Significant overuse
+	+ 4.37	Significant overuse
=	+ 8.54	Significant overuse
#	- 1.79	No significance
\$	+ 27.24	Significant overuse
&	- 31.90	Significant underuse
“its”	- 0.97	No Significance
	- 24.42	Significant underuse
@	+ 10.47	Significant overuse
Emoji	+ 32.57	Significant overuse
Other visual elements	- 8.19	Significant underuse

The most significant of the punctuation marks and symbols are the asterisk, exclamation mark, and ellipsis, which appear to be used to draw attention to the title and

accentuate other clickbait features.

(14) It's **ACTUALLY** happening...! (Fortnite ICON Skin) (CB3)

In (14) the capitalised adverb “actually”, framed in asterisks, adds an exaggerated element of surprise which is further emphasised by the use of an exclamation mark. The use of ellipsis indicates an unfinished thought and can be considered a type of forward-referencing device used to contribute to the curiosity gap initially created by the unresolved pronoun “it”. Certain punctuation and symbols in YouTube video titles may act as amplifying devices, exaggerating, and highlighting other clickbait features.

Modals

Although the effectiveness of modal verbs to pique reader curiosity is limited relative to other clickbait features, they are suggested to act as supporting devices to attract reader attention (Molina et al., 2021). Only those modals which were found in both corpora, listed in Table 9, were considered.

Table 9. Distribution of Modal Verbs

Modal	Log likelihood	Use in the clickbait corpus, compared to the non-clickbait corpus
Will	+ 20.06	Significant overuse
Can	+ 1.19	No significance
Need	+ 10.40	Significant overuse
Must	+ 3.74	No significance
Should	- 0.06	No significance

The modal “will” is classified as a strong modal of certainty, implying certainty of outcome, and the modal “need” is classified as a strong modal of necessity (Depraetere and Reed, 2020). These modals occurred most often as parts of hyperbolic phrases.

(15) Inventions That Will CHANGE YOUR LIFE! (CB4)

(16) 16 PARENTING HACKS YOU NEED TO KNOW (CB2)

In (15), “will” implies certainty relative to the hyperbolic phrase “change your life”, which is directed at the reader through the use of the pronoun “your”. The use of “need” in (16) implies necessity in the hyperbolic phrase “need to know”. The information-gap created by the deictic forward referencing noun phrase “16 parenting hacks” is directed at the reader through the use of the pronoun “you”. These examples show how strong modal verbs may act as a supporting device to other clickbait features such as hyperbole and forward-reference.

Familiar Vocabulary

Ifantidou (2009) identifies a reader preference for semantically familiar vocabulary in news headlines. Similarly, in an online context where certain words or phrases may rise and fall in popularity, ‘trending’ words may act to attract reader attention (McCulloch, 2019). A word frequency analysis revealed lexical items overrepresented in the clickbait corpus. Stop words (common use words such as ‘a’, ‘an’, ‘the’ and ‘is’) as well as words overrepresented in any single sub-corpus (such as ‘game’ or ‘update’ in the gaming sub-corpus) were not considered.

Table 10. Trending Vocabulary in the Clickbait Corpus.

Word	Number of occurrences	% of corpus word count
Fortnite	110	1.39
hack*	78	0.99
challenge	72	0.91
vs	58	0.73
TikTok	25	0.31
unboxing	18	0.22

*The noun *hack* refers to a novel solution to an everyday problem, no instances of *hack* as a verb were found.

As illustrated in Table 10, the most frequent word was Fortnite a popular online game. Both Fortnite and TikTok are illustrations of how clickbait titles may include socially relevant vocabulary as a lure to attract reader attention. The nouns “hack”, “challenge”, and “unboxing” also refer to popular online trends.

(17) Last To Stop Playing AMONG US Wins \$10,000 – Challenge (CB5)

(18) BOYS vs GIRLS || Cool TikTok hacks #shorts (CB2)

Example (17) illustrates a common occurrence in the clickbait corpus whereby the word “challenge” seems to be tagged onto a title to attract reader attention. The preposition “vs”, as in (18), is used in noun phrases as a type of deictic forward-referencing device (also seen in (12) above). Deliberately chosen familiar vocabulary can draw reader attention to other clickbait features which may be present.

Discussion

The various features of clickbait play different roles in the overall effect of a clickbait title. It may be useful to distinguish the extent to which certain features may be indicative of clickbait in YouTube titles. The results suggest three categories; (a) indicative, the strongest indicators of clickbait, (b) partially indicative, features indicative of clickbait depending on their usage or type, and (c) facilitative, features which serve to draw reader attention to a title or act to highlight clickbait features within a title, but which are not clickbait in and of themselves. Table 11 offers a summary of the roles the various clickbait features may play in YouTube titles.

Table 11. Clickbait Features in YouTube Titles.

Clickbait Feature	Relevance to YouTube Titles	Category
Overuse of personal and demonstrative pronouns	Personal and demonstrative pronouns serve to create a knowledge gap and a false sense of relevance for the reader.	Indicative
Overuse of superlative adjectives and intensifiers	Superlative adjectives and intensifiers exaggerate titles and thereby contribute to the deceptive nature of clickbait titles.	Indicative
The use of forward-reference	Forward reference piques reader curiosity by creating an information gap.	Indicative
Emotional appeal	Indicative of clickbait when the elicitation of reader emotion is prioritised over the provision of information.	Partially Indicative
Hyperbolic words and strong lexical units	Hyperbole and strong lexical units act to exaggerate titles and pique reader curiosity.	Indicative
Using numerals	Cardinal numbers used in lists are a strong indicator of clickbait. Monetary amounts are a facilitative feature. Other uses of numbers are not indicative of clickbait.	Partially Indicative/ Facilitative

Use of upper case	The use of capitalised words attracts reader attention, particularly when upper- and lower-case are mixed in a single title. Often used to highlight other features of clickbait.	Facilitative
Distinctive punctuation patterns	Distinctive punctuation acts to draw reader attention to the title, and often to highlight particular words and phrases.	Facilitative
Modals	Partial indicator of clickbait. Strong modals of necessity and certainty are indicators of clickbait.	Partially Indicative
Shortened forms	Not an indicator of clickbait in YouTube titles.	n/a
Interrogative forms	Not an indicator of clickbait in YouTube titles.	n/a
Prioritization of familiar (<i>trending</i>) vocabulary (Ifantidou, 2009)	Bait phrases are an indicator of clickbait. Trending vocabulary can be an indicator of clickbait when used primarily to draw reader attention to a title.	Partially Indicative

The linguistic and typological features of clickbait rarely appear in isolation (Scott, 2021). Majority of the video titles in the clickbait corpus contained a variety of clickbait techniques. Mormol (2019) proposes that “a possible explanation for this might be related to the overuse of various clickbait constructions, which become fossilized over time and consequently, their effectiveness decreases” (p. 9). Hence, by combining several techniques in one title, the overall clickbait effect can be amplified.

In (19), the indicative clickbait features of deictic forward-reference, achieved by the phrase “what happens next” and the hyperbolic phrase “will shock you”, are supported by the use of the strong modal of certainty “will”. The use of mixed upper- and lower-case words and an exclamation mark draws the readers’ initial attention to the title.

(19) RICH PEOPLE Taught Lessons By POOR PEOPLE, What Happens Next WILL SHOCK YOU! (CB9)

(20) 10 LUCKIEST People ON THE PLANET!

In (20) the indicative clickbait features are the use of the superlative adjective “luckiest” which is emphasised by capitalisation and supported using a listicle structure. The hyperbolic phrase “on the planet” is capitalised and emphasised with an exclamation mark. These examples illustrate the complex relationship which various clickbait features may have in relation to each other.

Conclusion

This paper presents an exploratory study of the linguistic and typological features of clickbait in YouTube video titles. It is revealed that not all the features of clickbait identified in the literature are indicative of clickbait in the context of YouTube. The study is limited by the size of the data set and the narrow focus of the study. A larger data set may allow for more generalizable results and a more accurate analysis. The focus on clickbait features previously identified in the literature does not allow for the uncovering of clickbait features which may be unique to YouTube. Many questions are left unanswered regarding clickbait on YouTube. The interplay between the individual clickbait features warrants further exploration, and the relationship between thumbnail images and titles is yet unexplored.

In a recent video (Muller, 2021), popular YouTube channel Veritasium suggests that clickbait is becoming more scientific in its approach. The advent of YouTube Analytics, a function whereby the view count of a video can be tracked in real time, allows content creators to experiment with different titles until a spike in views is observed. Veritasium contends that practices such as this are driving clickbait forward to ever-more sophisticated techniques, which unless identified and understood, will become increasingly difficult to stop. It is hoped that an approach such as the one presented in this paper may

aid in the further understanding of clickbait use in non-news online contexts.

References

- ALVES, L., ANTUNES, N., AGRICI, O., SOUSA, C., & RAMOS, C. (2016). Click Bait: You won't believe what happens next!. *Fronteiras: Journal of Social, Technological and Environmental Science*, 5(2), 196–213. <https://doi.org/10.21664/2238-8869.2016V5I2.P196-213>
- BIYANI, P., TSIOUTSIOLIKLIS, K., & BLACKMER, J. (2016, February 21). "8 Amazing secrets for getting more clicks": Detecting clickbaits in news streams using article informality. [Conference proceedings]. *Proceedings of the 13th AAAI Conference on Artificial Intelligence* (pp. 94-100). AAAI-16, Phoenix, Arizona. AAAI Press. <https://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/11807/11569>
- BLOM, J. N. & HANSEN, K. R. (2014). Click bait: Forward-reference as lure in online news headlines. *Journal of Pragmatics*, 76, 87–100. <https://doi.org/10.1016/j.pragma.2014.11.010>
- CHAKRABORTY, A., PARANJAPE, A., SOURYA, K., & NILOY, G. (2016). Stop Clickbait: Detecting and preventing clickbaits in online news media". [Conference proceedings]. *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 9–16). 2016 IEEE, San Francisco, California. IEEE. <https://doi.org/10.48550/arXiv.1610.09786>
- CHEN, Y., CONROY, N. J., & RUBIN, V. L. (2015). Misleading online content: Recognising clickbait as "False News". [Conference proceedings]. *Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection* (pp. 15-19). 2015 ACM, Seattle, Washington. ACM. <https://doi.org/10.1145/2823465.2823467>
- DEPRAETERE, I., & REED, S. (2020). Mood and modality in English. In B. Aarts, A. McMahon, & Hinrichs, L. (Eds.), *The Handbook of English Linguistics* (pp. 269–290). Blackwell.
- DOR, D. (2003). On newspaper headlines as relevance optimizers. *Journal of Pragmatics*, 35(5), 695–721. [https://doi.org/10.1016/S0378-2166\(02\)00134-0](https://doi.org/10.1016/S0378-2166(02)00134-0)
- ELYASHAR, A., BENDAHAN, J., & PUZIS, R. (2017). *Detecting clickbait in online social media: You won't believe how we did it*. <http://arxiv.org/abs/1710.06699>
- GENÇ, Ş., & SURER, E. (2021). ClickbaitTR: Dataset for clickbait detection from Turkish news sites and social media with a comparative analysis via machine learning algorithms. *Journal of Information Science*, 1–20. <http://doi.org/10.1177/01655515211007746>
- HANCOCK, J., & GONZALEZ, A. (2013). Deception in computer-mediated communication. In S. C. Herring, D. Stein, & T. Virtanen (Eds.), *Pragmatics of Computer-Mediated Communication* (pp. 363–383). De Gruyter Mouton.
- IFANTIDOU, E. (2009). Newspaper headlines and relevance: ad hoc concepts in ad hoc contexts. *Journal of Pragmatics*, 41(4), 699–720. <https://doi.org/10.1016/j.pragma.2008.10.016>
- JIANG, T., GOU, Q., XU, Y., ZHAO, Y., & FU, S. (2019). What prompts users to click on news headlines? A click-stream data analysis of the effects of news recency and popularity. In N. G. Taylor, C. Christian-Lamb, M. H. Martin, & B. Nardi (Eds.), *Information in Contemporary Society* (pp. 539–546). Springer International Publishing.
- KUIKEN, J., SCHUTH, A., SPITTERS, M., & MARX, M. (2017). Effective headlines of newspaper articles in a digital environment. *Digital Journalism*, 5, 1300–1314. <https://doi.org/10.1080/21670811.2017.1279978>
- LOCKWOOD, G. (2016). Academic clickbait: Articles with positively-framed titles, interesting phrasing, and no wordplay get more attention online. *The Winnower*, 3, 1–13. <https://doi.org/10.15200/winn.146723.36330>
- LOEWENSTEIN, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin*, 116(1), 75–98. <https://psycnet.apa.org/doi/10.1037/0033-2909.116.1.75>
- LOPEZOSA, C., ORDUNA-MALEA, E., & PÉREZ-MONTORO, M. (2020). Making video news visible: Identifying the optimization strategies of the cybermedia on YouTube using web metrics. *Journalism Practice*, 14(4), 465–482. <https://doi.org/10.1080/17512786.2019.1628657>
- MCCULLOCH, G. (2019). *Because Internet: Understanding the New Rules of Language*. Riverhead Books.
- MOLINA, M. D., SUNDAR, S.S., RONY, M.M.U., HASSAN, N., LE, T., & LEE, D. (2021). Does clickbait actually attract more clicks? Three clickbait studies you must read. [Conference proceedings]. In *CHI Conference on Human Factors in Computing Science* (pp. 1-19). CHI 2021, Yokohama, Japan. ACM. <https://doi.org/10.1145/3411764.3445753>
- MORMOL, P. (2019). "I urge you to see this...". Clickbait as one of the dominant features of contemporary online headlines. *Social Communication*, 5(2), 1–10. <http://dx.doi.org/10.2478/sc-2019-0004>

- MULLER, D. (2021, August 17). *Clickbait is unreasonably effective*. <https://www.youtube.com/watch?v=S2x-HZPH5Sng>
- OROSA, B. G., SANTORUN, S. G., & GARCÍA, X. L. (2017). Use of clickbait in the online news media of the 28 EU member countries. *Revista Latina de Comunicación Social*, 72, 1261–1277. <https://doi.org/10.4185/RLCS-2017-1218en>
- POTTHAST, M., GOLLUB, T., KOMOLOSSY, K., SCHUSTER, S., WEIGMANN, M., FERNANDEZ, E.P.G., HAGEN, M., & STEIN, B. (2018). Crowdsourcing a large corpus of clickbait on Twitter. [Conference proceedings]. In *Proceedings of the 27th International Conference on Computational Linguistics* (pp. 1498–1507). ICCL 2018, Santa Fe, New Mexico. ICCL.
- POTTHAST, M., KÖPSEL, S., STEIN, B. AND HAGAN, M. (2016). Clickbait detection. In Information Retrieval (pp. 810–817). [Conference proceedings]. *European Conference on Information Retrieval*, Padua, Italy. Springer International Publishing.
- QU, J., HIBBACH, A.M., GOLLUB, T., & POTTHAST, M. (2018). Towards crowdsourcing clickbait labels for YouTube videos. [Conference proceedings]. In *Proceedings of the The 6th AAAI Conference on Human Computation and Crowdsourcing* (pp. 1–4). HCOMP 2018, Zürich: CEUR-WS.
- RONY, M. M. U., HASSAN, N. AND YOUSUF, M. (2017). Diving deep into clickbaits: Who use them to what extents in which topics with what effects?. [Conference proceedings]. In *Proceedings of the 2017 IEEE/ACM International Conference* (pp. 232–239). 2017 IEEE/ACM International Conference, Sydney, Australia, ACM.
- SADRI, S. R. (2019). Listicles and the modern news article: comparing the perceived credibility of listicles and traditional articles among millennial media consumers. *Atlantic Journal of Communication*, 27(2), 83–98. <https://doi.org/10.1080/15456870.2019.1574794>
- SCOTT, K. (2021). You won't believe what's in this paper! Clickbait, relevance and the curiosity gap. *Journal of Pragmatics*, 175, 53–66. <https://doi.org/10.1016/j.pragma.2020.12.023>
- SPERBER, D., & WILSON, D. (1995). *Relevance: Communication and cognition*. Blackwell.
- SPERBER, D., & WILSON, D. (2008). A deflationary account of metaphors. In R.W. Gibbs (Ed.), *The Cambridge Handbook of Metaphor and Thought* (pp. 84–105). Cambridge University Press.
- TAFESSE, W. (2020). YouTube marketing: How marketers' video optimisation practices influence video views. *Internet Research*, 30(6), 1689–1707. <https://doi.org/10.1108/INTR-10-2019-0406>
- VARSHNEY, D., & VISHWAKARMA, D. K. (2021). A unified approach for detection of clickbait videos on YouTube using cognitive evidence. *Applied Intelligence*, 51, 4214–4235. <https://doi.org/10.1007/s10489-020-02057-9>
- VIJGEN, B. (2014). The listicle: an exploring research on an interesting sharable new media phenomenon. *Studia Universitatis Babeş-Bolyai – Ephemeres*, 59(1), 103–122. <http://studia.ubbcluj.ro/download/pdf/894.pdf>
- ZANNETTOU, S., CHATZIS, S., PAPANAMOU, K., & SIRIVIANOS, M. (2018). The good, the bad and the bait: Detecting and characterising clickbait on YouTube. In *Security and Privacy Workshops* (pp. 63–69). [Symposium]. IEEE Symposium on Security and Privacy Workshops, San Francisco, California.